Ф. П. ВАСИЛЬЕВ

МЕТОДЫ ОПТИМИЗАЦИИ

Москва
Факториал Пресс
2002

УДК 519.6 (075.8)
ББК 22.19
В 19

Васильев Ф. П.
В 19 **Методы оптимизации.** — М.: Издательство «Факториал Пресс», 2002. — 824 с.
ISBN 5-88688-056-9.

Книга содержит численные методы решения задач оптимизации. Приводятся теоретическое обоснование и краткие характеристики этих методов. Рассматриваются задачи минимизации функций в конечномерных и бесконечномерных пространствах, а также задачи оптимального управления процессами, описываемыми системами обыкновенных дифференциальных уравнений и уравнений в частных производных.

Для студентов вузов по специальности «Прикладная математика», и специалистов, связанных с решением задач оптимизации.

Научное издание

Васильев Фёдор Павлович

МЕТОДЫ ОПТИМИЗАЦИИ

Формат 70×100/16. Усл. печ. л. 67. Бумага офсетная № 1. Гарнитура литературная. Подписано к печати 25.12.2001. Тираж 500 экз. Заказ № 5182.

Издательство «Факториал Пресс», 117449, Москва, а/я 331; ЛР ИД № 00316 от 22.10.1999.
e/mail: factorial@mail.compnet.ru; http://www.compnet.ru/factorial.

Отпечатано с готовых диапозитивов издательства «Факториал Пресс» в ППП типографии «Наука» Академиздатцентра «Наука» РАН. 121099, Москва Г-99, Шубинский пер., 6.

Оригинал-макет подготовлен с использованием издательской системы LaTeX.

ISBN 5-88688-056-9

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ

Первые задачи геометрического содержания, связанные с отысканием наименьших и наибольших величин, появились еще в древние времена. Развитие промышленности в XVII — XVIII веках привело к необходимости исследования более сложных задач на экстремум и к появлению вариационного исчисления. Однако лишь в XX веке при огромном размахе производства и осознании ограниченности ресурсов Земли во весь рост встала задача оптимального использования энергии, материалов, рабочего времени, большую актуальность приобрели вопросы наилучшего в том или ином смысле управления различными процессами физики, техники, экономики и др. Сюда относятся, например, задача организации производства с целью получения максимальной прибыли при заданных затратах ресурсов, задача управления системой гидростанций и водохранилищ с целью получения максимального количества электроэнергии, задача о космическом перелете из одной точки пространства в другую наибыстрейшим образом или с наименьшей затратой энергии, задача о быстрейшем нагреве или остывании металла до заданного температурного режима, задача о наилучшем гашении вибраций и многие другие задачи.

Потребности развития самой вычислительной математики также привели к необходимости исследования таких задач на максимум и минимум, как, например, задачи наилучшего приближения функций, оптимального выбора параметров итерационного процесса или узлов интерполирования, минимизации невязки уравнений и т. д.

На математическом языке такие задачи могут быть сформулированы как задачи отыскания экстремума (максимума или минимума) некоторой функции или функционала $J(u)$, выражающего собой качество (цену) управления u из заданного множества U некоторого пространства. Требование принадлежности управления u некоторому множеству U выражает собой ограничения, обычно вытекающие из законов сохранения, ограниченности наличных ресурсов, возможностей технической реализации управления, нежелательности каких-либо запрещенных (аварийных) состояний и т. п. Задачи отыскания экстремума функции $J(u)$ на множестве U принято называть экстремальными задачами. Заметим, что задача максимизации функционала $J(u)$ на множестве U эквивалентна задаче минимизации функционала $-J(u)$ на том же множестве U, поэтому можно ограничиться рассмотрением задач минимизации.

В настоящее время теория экстремальных задач обогатилась фундаментальными результатами, появились ее новые разделы, такие как линейное, выпуклое, стохастическое программирование, оптимальное управление и др. Потребности практики способствовали бурному развитию методов приближенного решения экстремальных задач. Появление быстродействующих электронных вычислительных машин (ЭВМ) сделало возможным эффективное решение многих важных прикладных экстремальных задач, которые ранее из-за своей сложности представлялись недоступными.

В настоящей книге излагаются элементы теории экстремальных задач, а также основы наиболее часто используемых на практике методов приближенного решения экстремальных задач, теоретическое обоснование и краткая характеристика этих методов. Книга написана как учебное пособие

для студентов факультетов и отделений прикладной математики университетов, технических вузов. В основу книги положен курс лекций по численным методам решения экстремальных задач, который автор в течение ряда лет читает на факультете вычислительной математики и кибернетики Московского университета.

Заманчиво было бы изложить теорию и методы минимизации сразу в общем виде на языке функционального анализа, охватив при этом как частный случай многие методы минимизации функций конечного числа переменных. Однако такой способ изложения, несмотря на свою привлекательность и удобства для читателя-математика, видимо, все же труден для первого знакомства с предметом, не говоря уже о том, что он не может отразить всю специфику конечномерных задач.

Поэтому автор, стремясь сделать книгу доступной широкому кругу читателей, впервые знакомящихся с теорией и методами решения экстремальных задач, разделил ее на две части.

В части I излагаются методы минимизации функций конечного числа переменных (главы 1–5), а также рассматриваются задачи оптимального управления процессами, описываемыми системами обыкновенных дифференциальных уравнений (главы 6, 7). Этот материал для своего понимания требует лишь знания основ математического анализа, линейной алгебры, теории обыкновенных дифференциальных уравнений в объеме стандартных курсов технических вузов.

В части II книги рассматриваются задачи оптимизации на множествах из бесконечномерных функциональных пространств, и посвящена задачам оптимального управления процессами, описываемыми как обыкновенными дифференциальными уравнениями, так и уравнениями с частными производными (глава 8), методам решения неустойчивых задач оптимизации (глава 9), проблемами аппроксимации бесконечномерных задач оптимизации (глава 10). Для понимания содержания части II желательно знание элементов функционального анализа в объеме программ, обычно изучаемых в университетах и технических вузах с повышенной математической подготовкой. Однако следует оговориться, что отсутствие знаний по функциональному анализу не будет мешать пониманию и усвоению излагаемых в книге основ методов и приложений к конкретным классам экстремальных задач, если читатель будет готов принять некоторые приводимые в книге утверждения не в самой общей их форме.

Настоящую книгу можно считать очередным изданием сразу двух предыдущих книг автора «Численные методы решения экстремальных задач» (М.: Наука, 1980 г. — 1-е издание, 1988 г. — 2-е издание) и «Методы решения экстремальных задач» (М.: Наука, 1981 г.), причем часть I ранее излагалась в первой из названных книг, часть II — во второй книге. В предлагаемом переиздании указанных книг сохранена прежняя их структура, но содержание существенно переработано и дополнено. Заново написаны главы 2, 9, добавлены новые параграфы: §§ 6, 11, 16 в главе 5, § 4 в главе 6, §§ 10, 11 в главе 8, §§ 2, 7, 13 в главе 9, § 7 в главе 10, существенно обновлено содержание § 5 главы 3, § 13 главы 5, §§ 2, 3 главы 8, улучшено изложение материала во многих других параграфах, исправлены замеченные ошибки, неточности, опечатки.

В § 2 главы 2 впервые в учебной литературе изложены новые необходимые условия экстремума второго порядка, принадлежащие А. В. Арутюнову. В главе 3 дано простое обоснование антициклина (§ 3), теория двойственности в линейном программировании изложена, опираясь лишь на симплекс-метод (§ 5). Приведено элементарное доказательство принципа максимума

Понтрягина для весьма общей задачи оптимального управления, включая задачи с фазовыми ограничениями (§ § 3, 4 главы 6). Часть текста, которая содержит материал, дополняющий и расширяющий основное содержание книги, напечатана петитом и при первом чтении может быть опущена.

По рассматриваемым в книге проблемам оптимизации имеется обширная библиография, насчитывающая много тысяч названий. Список литературы, приведенный в конце книги, содержит лишь некоторые работы, в основном, российских математиков, а также переведенные на русский язык книги и монографии зарубежных авторов, которые были или непосредственно использованы в книге или близко примыкают к ней, дополняя и расширяя ее содержание.

Нумерация формул, теорем, лемм, определений, упражнений в каждом параграфе самостоятельная; ссылки на материалы, расположенные в пределах данного параграфа, нумеруются одним числом, вне данного параграфа, но в пределах данной главы — двумя числами, вне данной главы — тремя числами. Так, например, теорема 3 из § 2 главы 4 в пределах этого параграфа именуется просто теоремой 3, в других параграфах 4-й главы — теоремой 2.3, в других главах — теоремой 4.2.3. Аналогично параграфы при ссылках на них в пределах данной главы нумеруются одним числом, а вне этой главы — двумя числами: первое число означает номер главы, второе — номер параграфа.

Работа по подготовке настоящей книги к изданию не могла бы быть выполнена без помощи многих моих коллег и читателей. Автор глубоко признателен А. С. Антипину, А. В. Арутюнову, Е. Г. Белоусову, Н. А. Бобылеву, Н. Л. Григоренко, М. И. Зеликину, Л. Ф. Зеликиной, А. Ф. Измаилову, А. С. Ильинскому, Х. Д. Икрамову, А. З. Ишмухаметову, М. Йовановичу, А. С. Леонову, А. Недич, М. С. Никольскому, О. Обрадовичу, Н. М. Попову, М. М. Потапову, А. В. Разгулину, В. А. Срочко, А. А. Станевичюсу, М. Ф. Сухинину, А. В. Тимохову, В. В. Федорову, А. А. Шананину, В. Янковичу, В. Ячимовичу, М. Ячимовичу, которые своими советами, предложениями, замечаниями способствовали улучшению содержания книги. Особенно благодарен А. В. Арутюнову, без помощи которого автор не смог бы написать §§ 2.4, 5.16, 6.4 в их настоящем виде, А. З. Ишмухаметову за помощь при написании § 10.7.

В столь бурно развивающейся области, как теория и методы решения экстремальных задач, очень трудно создать учебное пособие, которое обладало бы определенной завершенностью и было бы свободным от недостатков, и поэтому автор будет признателен читателям за критические замечания по содержанию книги.

Ф. П. Васильев

Часть I

КОНЕЧНОМЕРНЫЕ ЗАДАЧИ МИНИМИЗАЦИИ. ПРИНЦИП МАКСИМУМА. ДИНАМИЧЕСКОЕ ПРОГРАММИРОВАНИЕ

ГЛАВА 1

Методы минимизации функций одной переменной

С задачами минимизации функций одной переменной мы впервые сталкиваемся при изучении начальных глав математического анализа и решаем их методами дифференциального исчисления. Может показаться, что эти задачи относятся к достаточно простым и методы их решения хорошо разработаны и изучены. Однако это не совсем так. Методы дифференциального исчисления находят ограниченное применение и далеко не всегда удобны для реализации на современных ЭВМ. Хотя в последние десятилетия появились другие методы, более удобные для использования на ЭВМ, требующие меньшего объема вычислительного труда, но тем не менее эту область экстремальных задач нельзя считать завершенной. Работы, посвященные новым методам минимизации функций одной переменной, продолжают появляться на страницах математических книг и журналов. Мы здесь остановимся на некоторых наиболее известных методах, достаточно хорошо проявивших себя на практике. Другие методы минимизации функций одной переменной читатель найдет, например, в [76; 148; 193; 214; 257; 590; 662; 671; 681; 684; 709; 738; 755].

§ 1. Постановка задачи

Пусть $\mathbb{R} = \{x: -\infty < x < \infty\}$ — числовая ось, X — некоторое множество из $\mathbb{R}$, $f(x)$ — функция, определенная на множестве X и принимающая во всех точках $x \in X$ конечные значения. Примерами множеств из $\mathbb{R}$ являются: отрезок $[a, b] = \{x \in \mathbb{R}: a \leqslant x \leqslant b\}$, интервал $(a, b) = \{x \in \mathbb{R}: a < x < b\}$, полуинтервалы $[a, b) = \{x \in \mathbb{R}: a \leqslant x < b\}$, $(a, b] = \{x \in \mathbb{R}: a < x \leqslant b\}$, где a, b — заданные числа. Будем рассматривать задачу минимизации функции $f(x)$ на множестве X. Начнем с того, что уточним постановку этой задачи. Для этого сначала напомним некоторые определения из классического математического анализа.

Определение 1. Точку $x_* \in X$ называют *точкой минимума* функции $f(x)$ на множестве X, если $f(x_*) \leqslant f(x)$ для всех $x \in X$; величину $f(x_*)$ называют *наименьшим* или *минимальным значением* $f(x)$ на X и обозначают $\min\limits_{x \in X} f(x) = f(x_*)$. Множество всех точек минимума $f(x)$ на X будем обозначать через X_*.

В зависимости от свойств множества X и функции $f(x)$ множество X_* может содержать одну, несколько или даже бесконечно много точек, а также возможны случаи, когда X_* пусто.

Пример 1. Пусть $f(x) = \sin^2(\pi/x)$ при $x \neq 0$ и $f(0) = 0$. На множестве $X = \{x: 1 \leqslant x \leqslant 2\}$ минимальное значение $f(x)$ равно нулю, множество X_* состоит из единственной точки $x_* = 1$. Если $X = \{x: 1/3 \leqslant x \leqslant 1\}$, то X_* содержит три точки: 1/3, 1/2, 1; если $X = \{x: 0 < x \leqslant 1\}$, то $X_* = \{x: x = 1/n, n = 1, 2, \ldots\}$ — счетное множество. В случае $X = \{x: 2 \leqslant x < \infty\}$,

функция $f(x)$ не имеет наименьшего значения на X. В самом деле, какую бы точку $x \in X$ ни взять, найдется точка $v \in X$ (например, $v = k$ при достаточно большом k) такая, что $f(x) > f(v)$. Это значит, что X_* пусто.

Пример 2. Функция $f(x) = |x| + |x-1| - 1$ на $X = \{x\colon |x| \leqslant 1\}$ принимает свое наименьшее значение, равное нулю, во всех точках отрезка $X_* = \{x\colon 0 \leqslant x \leqslant 1\}$. Если $X = \{x\colon 1 \leqslant x \leqslant 2\}$, то X_* содержит одну точку $x_* = 1$; если $X = \{x\colon 1 < x \leqslant 2\}$, то $X_* = \varnothing$.

Пример 3. Пусть $f(x) = x$ при $x \neq 0$ и $f(0) = 1$. На множествах $X = \{x\colon 0 \leqslant x \leqslant 1\}$ или $X = \{x\colon 0 < x \leqslant 1\}$ эта функция не имеет наименьшего значения, т. е. $X_* = \varnothing$.

Пример 4. Пусть $f(x) = \ln x$, $X = \{x\colon 0 < x \leqslant 1\}$. Здесь $X_* = \varnothing$, так как во всех точках из X функция принимает конечные значения, а для последовательности $x_k = 1/k$ $(k = 1, 2\ldots)$ имеем $\lim\limits_{k\to\infty} f(x_k) = -\infty$.

Определение 2. Функция $f(x)$ называется *ограниченной снизу* на множестве X, если существует такое число M, что $f(x) \geqslant M$ для всех $x \in X$. Функция $f(x)$ *не ограничена снизу* на X, если существует последовательность $\{x_k\} \in X$, для которой $\lim\limits_{k\to\infty} f(x_k) = -\infty$.

В примерах 1–3 функции ограничены снизу на рассматриваемых множествах, а в примере 4 функция не ограничена.

В тех случаях, когда $X_* = \varnothing$, естественным обобщением понятия наименьшего значения функции является понятие нижней грани функции.

Определение 3. Пусть функция $f(x)$ ограничена снизу на множестве X. Тогда число f_* называют *нижней гранью* $f(x)$ на X, если: 1) $f_* \leqslant f(x)$ при всех $x \in X$; 2) для любого сколь угодно малого числа $\varepsilon > 0$ найдется точка $x_\varepsilon \in X$, для которой $f(x_\varepsilon) < f* + \varepsilon$. Если функция $f(x)$ не ограничена снизу на X, то в качестве нижней грани $f(x)$ на X принимается $f_* = -\infty$. Нижнюю грань $f(x)$ на X обозначают через $\inf\limits_{x\in X} f(x) = f_*$.

В примерах 1–3 $f_* = 0$, а в примере 4 $f_* = -\infty$.

Если $X_* \neq \varnothing$, то, очевидно, нижняя грань $f(x)$ на X совпадает с наименьшим значением этой функции на X, т. е. $\inf\limits_{x\in X} f(x) = \min\limits_{x\in X} f(x)$. В этом случае говорят, что функция $f(x)$ на X достигает своей нижней грани. Подчеркнем, что $\inf\limits_{x\in X} f(x) = f_*$ всегда существует, а $\min\limits_{x\in X} f(x)$, как мы видели из примеров 1–4, не всегда имеет смысл. Введем еще два определения.

Определение 4. Последовательность $\{f(x_k)\} \in X$ называется *минимизирующей* для функции $f(x)$ на множестве X, если

$$\lim_{k\to\infty} f(x_k) = \inf_{x\in X} f(x) = f_*.$$

Из определения и существования нижней грани следует, что минимизирующая последовательность всегда существует.

Определение 5. Скажем, что последовательность $\{x_k\}$ *сходится* к непустому множеству X, если $\lim\limits_{k\to\infty} \rho(x_k, X) = 0$, где $\rho(x_k, X) = \inf\limits_{x\in X} |x_k - x|$ — расстояние от точки x_k до множества X.

Заметим, что если $X_* \neq \varnothing$, то всегда существует минимизирующая последовательность, сходящаяся к X_*; например, можно взять стационарную последовательность $x_k = x_*$ $(k = 1, 2, \ldots)$, где x_* — какая-либо точка из X_*. Однако не следует думать, что при $X_* \neq \varnothing$ любая минимизирующая последовательность будет сходиться к X_*.

Пример 5. Пусть $f(x) = \dfrac{x^2}{1+x^4}$, $X = \mathbb{R}$. Очевидно, здесь $f_* = 0$ и множество X_* состоит из единственной точки $x_* = 0$. Последовательность $x_k = k$ $(k = 1, 2, \ldots)$ является минимизирующей, так как $\lim\limits_{k\to\infty} f(x_k) = 0$, но $\rho(x_k, X_*) = k$ не стремится к нулю.

Теперь перейдем к формулировке задачи минимизации функции $f(x)$ на множестве X. Здесь возможны различные подходы. Обычно принято различать задачи двух типов. К *первому типу* относят задачи, в которых требуется определить величину $f_* = \inf\limits_{x\in X} f(x)$. Сразу же подчеркнем, что в задачах первого типа неважно, будет ли множество X_* точек минимума $f(x)$ на X непустым или оно пусто. Ко *второму типу* задач относят те задачи, у которых множество X_* непусто и требуется наряду с f_* найти какую-либо точку $x_* \in X_*$. Здесь возможно дальнейшее уточнение постановки задачи. Можно искать точку $x_* \in X_*$, обладающую каким-либо дополнительным свойством (например, ближайшую к началу координат). Бывают ситуации, когда важно найти все множество X_* или какую-либо ее часть, представляющую, скажем пересечение X_* с некоторым заданным множеством.

Заметим, что получить точное решение задачи первого или второго типа удается лишь в редких случаях. Поэтому на практике при решении задач первого типа обычно строят какую-либо минимизирующую последовательность $\{x_k\}$ для функции $f(x)$ на X и затем в качестве приближения для f_* берут величину $f(x_k)$ при достаточно большом k. Аналогично для приближенного решения задач второго типа достаточно построить минимизирующую последовательность $\{x_k\}$, которая сходится ко множеству X_* в смысле определения 5, и в качестве приближения для f_* и точки $x_* \in X_*$ взять соответственно величину $f(x_k)$ и точку x_k при достаточно большом k.

Как показывает пример 5, в отличие от задач первого типа не всякая минимизирующая последовательность может быть использована для получения приближенного решения задач второго типа. Построение минимизирующих последовательностей, сходящихся ко множеству X_*, в общем случае требует привлечения специальных методов. В настоящей главе будем рассматривать лишь такие задачи второго типа, у которых любая минимизирующая последовательность сходится к X_*. Один такой класс задач дается следующей теоремой, называемой *теоремой Вейерштрасса* [327; 350; 352; 534].

Теорема 1. *Пусть X — замкнутое ограниченное множество из $\mathbb{R}$, функция $f(x)$ непрерывна на X. Тогда $f(x)$ ограничена снизу на X, множество X_* точек минимума $f(x)$ на X непусто, замкнуто и любая минимизирующая последовательность $\{x_k\}$ сходится к X_*.*

Несколько более общий факт будет установлен в § 2.1, из которого также будет следовать теорема 1. Предлагаем читателю вернуться к примерам 1–5 и выяснить, в каких случаях и какое из условий теоремы 1 нарушено и к чему это приводит.

Возможна и более широкая постановка задач минимизации второго типа — когда ищутся не только точки минимума в смысле определения 1, но и точки так называемого локального минимума.

Определение 6. Точка $v_* \in X$ называется *точкой локального минимума* функции $f(x)$ на множестве X со значением $c = f(v_*)$, если существует такое число $\alpha > 0$, что $f(v_*) \leqslant f(x)$ для всех $x \in X \cap \{x\colon |x - v_*| < \alpha\} = O_\alpha(v_*)$. Если при некотором $\alpha > 0$ равенство $f(v_*) = f(x)$ для $x \in O_\alpha(v_*)$ возможно только при $x = v_*$, то v_* называют точкой *строгого локального минимума*.

Для функции, график которой изображен на рис 1.1, точки x_0, x_2, x_4 являются точками строгого локального минимума, а в точках, удовлетворяющих неравенствам $x_5 < x \leqslant x_6$ и $x_8 \leqslant x \leqslant x_9$, реализуется нестрогий локальный минимум. Функция из примера 1 в точках $x_k = 1/k$ $(k = \pm 1, \pm 2)$ имеет строгий локальный минимум на $X = \mathbb{R}$, а в точке $x_* = 0$ — нестрогий локальный минимум.

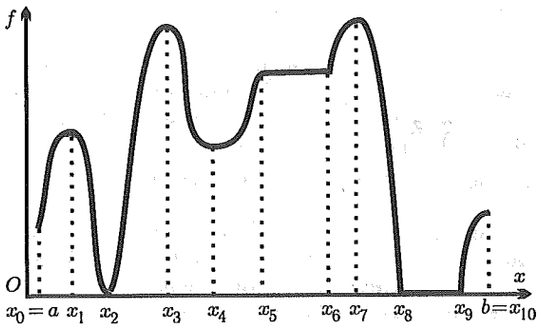

Рис. 1.1

Точки локального минимума, в которых минимум достигается в смысле определения 1, часто называют *точками глобального* или *абсолютного минимума* функции $f(x)$ на множестве X.

Выделим класс функций, у которых все точки локального минимума являются точками глобального минимума.

Определение 7. Функцию $f(x)$ назовем *унимодальной* на отрезке $X = [a, b]$, если она непрерывна на $[a, b]$ и существуют числа α, β $(a \leqslant \alpha \leqslant \beta \leqslant b)$ такие, что: 1) $f(x)$ строго монотонно убывает при $a \leqslant x \leqslant \alpha$ (если $a < \alpha$); 2) $f(x)$ строго монотонно возрастает при $\beta \leqslant x \leqslant b$ (если $\beta < b$); 3) $f(x) = f_* = \inf\limits_{x \in X} f(x)$ при $\alpha \leqslant x \leqslant \beta$, так что $X_* = [\alpha, \beta]$. Случаи, когда один или два из отрезков $[a, \alpha]$, $[\alpha, \beta]$, $[\beta, b]$ вырождаются в точку, здесь не исключаются. В частности, если $\alpha = \beta$, то $f(x)$ назовем *строго унимодальной* на отрезке $[a, b]$.

Функция из примера 2 унимодальна на любом отрезке $[a, b]$; функция из примера 1 строго унимодальна на $[2/3, 2]$, но не будет унимодальной на отрезке $[1/2, 2]$.

Нетрудно видеть, что если функция $f(x)$ унимодальна на $[a, b]$, то она остается унимодальной и на любом отрезке $[c, d] \subseteq [a, b]$.

В заключение кратко остановимся на задаче максимизации функции.

Определение 8. Функция $f(x)$ называется *ограниченной сверху* на множестве X, если существует такое число B, что $f(x) \leqslant B$ при всех $x \in X$. Функция $f(x)$ *не ограничена сверху* на X, если существует последовательность $\{x_k\} \in X$, для которой $\lim\limits_{k \to \infty} f(x_k) = +\infty$. Функцию $f(x)$ называют *ограниченной* на X, если она ограничена на X сверху и снизу.

Определение 9. Если функция $f(x)$ ограничена сверху на X, то число f^* называется *верхней гранью* $f(x)$ на X в том случае, когда: 1) $f(x) \leqslant f^*$ для всех $x \in X$; 2) для любого числа $\varepsilon > 0$ найдется такая точка $x_\varepsilon \in X$, что $f(x_\varepsilon) > f^* - \varepsilon$. Если $f(x)$ не ограничена сверху на X, то по определению принимается $f^* = \infty$. Последовательность $\{x_k\} \in X$ называется *максимизирующей* для $f(x)$ на X, если $\lim\limits_{k \to \infty} f(x_k) = f^*$. Если существует такая точка $x^* \in X$, что $f(x^*) = f^*$, то x^* называется *точкой максимума* $f(x)$ на X, а величина $f(x^*)$ — *наибольшим* или *максимальным значением* $f(x)$ на X. Множество точек максимума $f(x)$ на X будем обозначать через X^*, верхнюю грань — через $f^* = \sup\limits_{x \in X} f(x)$.

Заметим, что верхняя грань и максимизирующая последовательность всегда существуют, а максимальное значение может не существовать. Если выполнены условия теоремы 1, то $f^* < \infty$, $X^* \neq \varnothing$ и любая максимизирующая последовательность $\{x_k\}$ сходится к X^*.

В задачах максимизации также можно различать задачи двух типов: в задачах *первого типа* ищется величина f^*, а в задачах *второго типа* ищется f^* и какая-либо точка $x^* \in X^*$. Нетрудно видеть, что

$$\sup_{x \in X} f(x) = -\inf_{x \in X} (-f(x)),$$

причем любая точка максимума и любая максимизирующая последовательность для $f(x)$ на X является точкой минимума и соответственно минимизирующей последовательностью для функции $-f(x)$ на X. Это значит, что любая задача максимизации функции $f(x)$ на X равносильна задаче минимизации функции $-f(x)$ на том же множестве X. Поэтому мы можем ограничиться изучением лишь задач минимизации.

Наконец, немного о точках локального максимума.

Определение 10. Точка $v^* \in X$ называется *точкой локального максимума* функции $f(x)$ на множестве X, если существует такое число $\alpha > 0$, что $f(v^*) \geqslant f(x)$ для всех $x \in X \cap \{x: |x - v^*| < \alpha\} = O_\alpha(v^*)$. Если при некотором $\alpha > 0$ равенство $f(v^*) = f(x)$ для $x \in O_\alpha(v^*)$ возможно только при $x = v^*$, то v^* называют *точкой строгого локального максимума*.

Для функции, график которой изображен на рис. 1.1, точки x_1, x_3, x_7, x_{10} являются точками строгого локального максимума, а в точках, удовлетворяющих неравенствам $x_5 \leqslant x < x_6$ и $x_8 < x < x_9$, реализуется нестрогий локальный максимум; x_3 — точка глобального максимума.

Множество всех точек локального минимума и максимума функции на множестве X принято называть *точками локального экстремума* функции на этом множестве или, проще, *точками экстремума*.

Для обозначения задач минимизации (или максимизации) функции $f(x)$ на множестве X часто пользуются следующей краткой символической записью:

$$f(x) \to \inf, \quad x \in X \qquad (f(x) \to \sup, \quad x \in X),$$

при необходимости дополнительно уточняя постановку задачи; при этом минимизируемую (максимизируемую) функцию $f(x)$ называют *целевой функцией*, множество X — *допустимым множеством*.

Упражнения

1. Построить минимизирующую и максимизирующую последовательности для функции $f(x) = \operatorname{arctg} x$ на $X = \mathbb{R}$. Достигает ли функция своих нижней и верхней граней на $\mathbb{R}$?

2. Пусть $f(x) = |x^2 - 1|$ при $x \neq 1$ и $f(1) = 1$. Найти множество X_* точек минимума $f(x)$ на $X = \mathbb{R}$. Можно ли утверждать, что любая минимизирующая последовательность для этой функции будет сходиться к X_*?

3. Найти все точки локального экстремума функции $f(x) = |||x^2 - 1| - 1| - 1|$ на отрезке $[a, b]$ при различных a, b. При каких a, b эта функция будет унимодальной на $[a, b]$?

4. Выяснить, на каких отрезках будут унимодальными функции $f(x) = e^x$, $f(x) = x^2$, $f(x) = -x^2$, $f(x) = \sqrt{|x|}$, $f(x) = \cos x$.

5. Если функция $G(v)$ унимодальна на отрезке $[c, d]$, то функция $f(x) = G((d - c)(x - a)/(b - a) + c)$ унимодальна на отрезке $[a, b]$. Доказать.

6. Доказать, что линейная функция $f(x)=Ax+B$, где A, B — постоянные, $A\neq 0$, достигает своего минимума и максимума на отрезке $[a, b]$ только при $x=a$ или $x=b$.

7. Найти минимум функции $f(x)=\max\limits_{0\leqslant t\leqslant 1}|t^2-xt|$ на множествах $X=\mathbb{R}$ и $X=\{x\colon 1\leqslant x<\infty\}$.

§ 2. Классический метод

Под классическим методом будем подразумевать тот подход к поиску точек экстремума функции, который основан на дифференциальном исчислении и подробно описан в учебниках по математическому анализу [327; 350; 352; 534]. Мы здесь лишь вкратце остановимся на этом методе.

Пусть функция $f(x)$ кусочно непрерывна и кусочно гладка на отрезке $[a, b]$. Это значит, что на $[a, b]$ может существовать лишь конечное число точек, в которых $f(x)$ либо терпит разрыв первого рода, либо непрерывна, но не имеет производной. Тогда, как известно, точками экстремума функции $f(x)$ на $[a, b]$ могут быть лишь те точки, в которых выполняется одно из следующих условий: 1) либо $f(x)$ терпит разрыв; 2) либо $f(x)$ непрерывна, но производная $f'(x)$ не существует; 3) либо производная $f'(x)$ существует и равна нулю; 4) либо $x=a$ или $x=b$. Такие точки принято называть *точками, подозрительными на экстремум*.

Поиск точек экстремума функции начинают с нахождения всех точек, подозрительных на экстремум. После того как такие точки найдены, проводят дополнительное исследование и отбирают среди них те, которые являются точками локального минимума или максимума. Для этого обычно исследуют знак первой производной $f'(x)$ в окрестности (или соответствующей полуокрестности точек разрыва и граничных точек $x=a$, $x=b$) подозрительной точки. Для того чтобы подозрительная точка $v\in[a, b]$ была точкой локального минимума, достаточно, чтобы $\lim\limits_{x\to v-0}f(x)\geqslant f(v)$, $\lim\limits_{x\to v+0}f(x)\geqslant f(v)$ и при некотором $\alpha>0$ на множествах $[a, b]\cap(v, v+\alpha)=O_\alpha^+(v)$, $[a, b]\cap(v-\alpha, v)=O_\alpha^-(v)$ существовала производная $f'(x)$, причем $f'(x)>0$ при $x\in O_\alpha^+(v)$ и $f'(x)<0$ при $x\in O_\alpha^-(v)$. Если же $\lim\limits_{x\to v-0}f(x)\leqslant f(v)$, $\lim\limits_{x\to v+0}f(x)\leqslant f(v)$ и $f'(x)<0$ при $x\in O_\alpha^+(v)$, $f'(x)>0$ при $x\in O_\alpha^-(v)$, то v — точка локального максимума.

В тех случаях, когда удается вычислить в подозрительной точке производные второго и более высокого порядков, то их также можно использовать для исследования поведения функции в окрестности этой точки. А именно, пусть известны производные $f'(v), \ldots, f^{(n)}(v)$, причем $f^{(i)}(v)=0$ $(i=1, \ldots, n-1)$, а $f^{(n)}(v)\neq 0$ $(n\geqslant 1)$. Если n — четное число, то в случае $f^{(n)}(v)>0$ в точке v реализуется локальный минимум, а в случае $f^{(n)}(v)<0$ — локальный максимум. Если n нечетно, то при $a<v<b$ в точке v не может быть локального минимума или максимума; при $v=a$ $(v=b)$ в случае $f^{(n)}(v)>0$ в точке v имеем локальный минимум (максимум), а в случае $f^{(n)}(v)<0$ — локальный максимум (минимум).

Чтобы найти глобальный минимум (максимум) функции $f(x)$ на $[a, b]$, нужно перебрать все точки локального минимума (максимума) на $[a, b]$ и среди них выбрать точку с наименьшим (наибольшим) значением функции, если таковое существует. Если вместо отрезка $[a, b]$ имеем дело со множеством $X=\{x\colon a\leqslant x<\infty\}$, или $X=\{x\colon -\infty<x\leqslant b\}$, или $X=\mathbb{R}$, то наряду с вышеописанными исследованиями нужно также изучить поведение функции при $x\to\infty$ или $x\to-\infty$.

Классический метод исследования функции на экстремум следует использовать во всех тех случаях, когда достаточно просто удается выявить все подозрительные на экстремум точки и реализовать описанную выше схему отбора экстремальных точек. К великому сожалению, классический метод имеет весьма ограниченное применение. Дело в том, что вычисление производной $f'(x)$ в практических задачах зачастую является непростым делом. Например, может оказаться, что значения функции $f(x)$ определяются из наблюдений или каких-либо физических экспериментов, и получить информацию о ее производной крайне трудно. Но даже в тех случаях, когда производную все же удается вычислить, решение уравнения $f'(x)=0$ и выявление других точек, подозрительных на экстремум, может быть связано с серьезными трудностями. Поэтому важно иметь также и другие методы поиска экстремума, не требующие вычисления производных, более удобные для реализации на современных ЭВМ.

Упражнения

1. Найти точки экстремума функции $f(x)=\sin^3 x+\cos^3 x$ на отрезках $[0, 3\pi/4]$, $[0, 2\pi]$.

2. Пусть $f(x)=(1+e^{1/x})^{-1}$ при $x\neq 0$ и $f(0)=0$. Найти точки экстремума этой функции на отрезках $[0, 1]$, $[-1, 0]$, $[-1, 1]$, $[1, 2]$ и на $\mathbb{R}$.

3. Пусть непрерывная на отрезке $[a, b]$ функция $f(x)$ в точке v $(a<v<b)$ имеет строгий локальный минимум. Можно ли утверждать, что существует число $\alpha>0$ такое, что $f(x)$ монотонно убывает при $v-\alpha<x<v$ и монотонно возрастает при $v<x<v+\alpha$? Рассмотреть функцию $f(x)=2x^2+x^2\sin(1/x)$ $(x\neq 0)$, $f(0)=0$ на $[-1, 1]$. Исследовать случай, когда $f(x)$ имеет на $[a, b]$ конечное число точек локального экстремума.

4. Пусть функция $f(x)$ определена на $[a, b]$ и дважды дифференцируема в точке $v\in[a, b]$. Доказать, что если $a<v<b$ и в точке v реализуется локальный минимум $f(x)$, то необходимо, чтобы $f''(x)\geqslant 0$. Будет ли верным это утверждение, если $v=a$ или $v=b$? Будет ли оно верным, если $v=a$ или $v=b$ и, кроме того $f'(v)=0$? Рассмотреть функции $f(x)=-x^2$, $f(x)=\cos x$ на $[-\pi, \pi]$.

5. Пусть функция $f(x)$ определена на $[a, b]$ и в точке $v\in[a, b]$ имеет n производных $(n\geqslant 2)$, причем известно, что $f^{(i)}(v)=0$ при $i=1, \ldots, n-1$ и $f^{(n)}(v)\neq 0$. Доказать, что если v — точка локального минимума и $a<v<b$, то n — четное число и $f^{(n)}(v)>0$. Что изменится, если $v=a$ или $v=b$?

6. Пусть функция $f(x)$ аналитична на $[a, b]$, т. е. ряд Тейлора этой функции сходится к $f(x)$ во всех точках $[a, b]$. Может ли эта функция иметь на $[a, b]$ бесконечное число точек локального экстремума?

7. Пусть функция $f(x)$ определена на $[a, b]$ и в точке v имеет производные всех порядков. Можно ли утверждать, что если v — точка локального минимума, то $f^{(n)}(v)\neq 0$ при каком-либо $n\geqslant 1$? Рассмотреть функцию $f(x)=e^{-1/x^2}$ $(x\neq 0)$, $f(0)=0$ в точке $v=0$. Что изменится, если функция $f(x)$ аналитична на $[a, b]$?

8. Пусть функция $f(x)$ дифференцируема на $[a, b]$ и в точке $v\in[a, b]$ достигает своей нижней грани на $[a, b]$. Доказать, что тогда необходимо, чтобы $f'(v)(x-v)\geqslant 0$ при всех $x\in[a, b]$. Будет ли выполнение этого условия достаточно для того, чтобы в точке v достигалась нижняя грань $f(x)$ на $[a, b]$?

§ 3. Метод деления отрезка пополам

Простейшим методом минимизации функции одной переменной, не требующим вычисления производной, является метод деления отрезка пополам. Опишем его, предполагая, что минимизируемая функция $f(x)$ унимодальна на отрезке $[a, b]$. Поиск минимума $f(x)$ на $[a, b]$ начинается с выбора двух

точек $x_1=(a+b-\delta)/2$ и $x_2=(a+b+\delta)/2$, где δ — постоянная, являющаяся параметром метода, $0<\delta<b-a$. Величина δ выбирается вычислителем и может определяться целесообразным количеством верных десятичных знаков при задании аргумента x. В частности, ясно, что δ не может быть меньше машинного нуля ЭВМ, используемой при решении рассматриваемой задачи. Точки x_1, x_2 расположены симметрично на отрезке $[a, b]$ относительно его середины и при малых δ делят его почти пополам — этим и объясняется название метода.

После выбора точек x_1, x_2 вычисляются значения $f(x_1), f(x_2)$ и сравниваются между собой. Если $f(x_1)\leqslant f(x_2)$, то полагают $a_1=a$, $b_1=x_2$; если же $f(x_1)>f(x_2)$, то полагают $a_1=x_1$, $b_1=b$. Поскольку $f(x)$ унимодальна на $[a, b]$, то ясно, что отрезок $[a_1, b_1]$ имеет общую точку с множеством X_* точек минимума $f(x)$ на $[a, b]$ и его длина равна

$$b_1-a_1=(b-a-\delta)/2+\delta.$$

Пусть отрезок $[a_{k-1}, b_{k-1}]$, имеющий непустое пересечение с X_*, уже известен, и пусть $b_{k-1}-a_{k-1}=(b-a-\delta)/2^{k-1}+\delta>\delta$ $(k\geqslant 2)$. Тогда берем точки $x_{2k-1}=(a_{k-1}+b_{k-1}-\delta)/2$, $x_{2k}=(a_{k-1}+b_{k-1}+\delta)/2$, расположенные на отрезке $[a_{k-1}, b_{k-1}]$, симметрично относительно его середины, и вычисляем значения $f(x_{2k-1}), f(x_{2k})$. Если $f(x_{2k-1})\leqslant f(x_{2k})$, то полагаем $a_k=a_{k-1}$, $b_k=x_{2k}$; если же $f(x_{2k-1})>f(x_{2k})$, то полагаем $a_k=x_{2k-1}$, $b_k=b_{k-1}$. Длина получившегося отрезка $[a_k, b_k]$ равна $b_k-a_k=(b-a-\delta)/2^k+\delta>\delta$ и $[a_k, b_k]\cap X_*\neq\varnothing$.

Если количество вычислений значений минимизируемой функции ничем не ограничено, то описанный процесс деления отрезка пополам можно продолжать до тех пор, пока не получится отрезок $[a_k, b_k]$ длины $b_k-a_k<\varepsilon$, где ε — заданная точность, $\varepsilon>\delta$. Отсюда имеем, что $k>\log_2((b-a-\delta)/(\varepsilon-\delta))$. Поскольку каждое деление пополам требует двух вычислений значений функции, то для достижения точности $b_k-a_k<\varepsilon$ требуется всего $n=2k>2\log_2((b-a-\delta)/(\varepsilon-\delta))$ таких вычислений.

После определения отрезка $[a_k, b_k]$ в качестве приближения ко множеству X_* можно взять точку $\bar{x}_n=x_{2k-1}$ при $f(x_{2k-1})\leqslant f(x_{2k})$ и $\bar{x}_n=x_{2k}$ при $f(x_{2k-1})>f(x_{2k})$, а значит $f(\bar{x}_n)$ может служить приближением для $f_*=\inf\limits_{x\in[a,b]} f(x)$. При таком выборе приближения для X_* будет допущена погрешность $\rho(\bar{x}_n, X_*)\leqslant\max\{b_k-\bar{x}_n; \bar{x}_n-a_k\}=(b-a-\delta)/2^k$. Если не требовать того, чтобы значение функции, принимаемое за приближение к f_*, было вычислено непременно в той же точке, которая служит приближением к X_*, то вместо $\bar{x}_n$ можно взять точку $v_n=(a_k+b_k)/2$ с меньшей погрешностью $\rho(\bar{x}_n, X_*)\leqslant(b_k-a_k)/2=(b-a-\delta)/2^{k+1}+\delta/2$ (здесь $k=n/2$ и δ достаточно мало).

Конечно, и в этом случае можно бы провести еще одно дополнительное вычисление значения функции в точке v_n и принять $f(v_n)\approx f_*$. Однако заметим, что на практике нередко встречаются функции, нахождение значения которых в каждой точке связано с большим объемом вычислений или дорогостоящими экспериментами, наблюдениями, — понятно, что здесь приходится дорожить каждым вычислением значения минимизируемой функции. В таких ситуациях возможно даже, что число n, определяющее количество вычислений значений функции, заранее жестко задано и превышение его недопустимо.

Из предыдущего следует, что методом деления отрезка пополам с помощью $n=2k$ вычислений значений функции можно определить точку минимума унимодальной функции на отрезке $[a, b]$ в лучшем случае с точностью $\approx(b-a)2^{-1-n/2}$. Возникает вопрос, не существует ли методов, позволяющих с помощью того же числа вычислений значений функции решить задачу минимизации унимодальной функции поточнее? Оказывается, такие методы есть. Один из них будет описан в § 4.

В заключение отметим, что метод деления отрезка пополам без изменений можно применять для минимизации функций, не являющихся унимодальными. Однако в этом случае нельзя гарантировать, что найденное решение будет достаточно хорошим приближением к глобальному минимуму.

§ 4. Метод золотого сечения. Симметричные методы

Перейдем к описанию метода минимизации унимодальной функции на отрезке, столь же простого, как метод деления отрезка пополам, но позволяющего решить задачу с требуемой точностью при меньшем количестве вычислений значений функции. Речь пойдет о методе золотого сечения.

1. Как известно, *золотым сечением* отрезка называется деление отрезка на две неравные части так, чтобы отношение длины всего отрезка к длине большей части равнялось отношению длины большей части к длине меньшей части отрезка.

Нетрудно проверить, что золотое сечение отрезка $[a, b]$ производится двумя точками $x_1=a+(3-\sqrt{5})(b-a)/2=a+(b-a)\,0{,}381966011\ldots$ и $x_2=a+(\sqrt{5}-1)(b-a)/2=a+(b-a)\,0{,}618033989\ldots$, расположенными симметрично относительно середины отрезка, причем $a<x_1<x_2<b$, $(b-a)/(b-x_1)=(b-x_1)/(x_1-a)=(b-a)/(x_2-a)=(x_2-a)/(b-x_2)=(\sqrt{5}+1)/2=1{,}618033989\ldots$

Замечательно здесь то, что точка x_1 в свою очередь производит золотое сечение отрезка $[a, x_2]$, так как $x_2-x_1<x_1-a=b-x_2$ и $(x_2-a)/(x_1-a)=(x_1-a)/(x_2-x_1)$. Аналогично точка x_2 производит золотое сечение отрезка $[x_1, b]$. Опираясь на это свойство золотого сечения, можно предложить следующий метод минимизации унимодальной функции $f(x)$ на отрезке $[a, b]$.

Положим $a_1=a$, $b_1=b$. На отрезке $[a_1, b_1]$ возьмем точки x_1, x_2, производящие золотое сечение, и вычислим значения $f(x_1), f(x_2)$. Далее, если $f(x_1)\leqslant f(x_2)$, то примем $a_2=a_1$, $b_2=x_2$, $\bar{x}_2=x_1$; если же $f(x_1)>f(x_2)$, то примем $a_2=x_1$, $b_2=b_1$, $\bar{x}_2=x_2$. Поскольку функция $f(x)$ унимодальна на $[a, b]$, то отрезок $[a_2, b_2]$ имеет хотя бы одну общую точку с множеством X_* точек минимума $f(x)$ на $[a, b]$. Кроме того, $b_2-a_2=(\sqrt{5}-1)(b-a)/2$ и весьма важно то, что внутри $[a_2, b_2]$ содержится точка $\bar{x}_2$ с вычисленным значением $f(\bar{x}_2)=\min\{f(x_1); f(x_2)\}$, которая производит золотое сечение отрезка $[a_2, b_2]$.

Пусть уже определены точки $x_1, \ldots, x_{n-1}$, вычислены значения $f(x_1), \ldots, f(x_{n-1})$, найден отрезок $[a_{n-1}, b_{n-1}]$ такой, что $[a_{n-1}, b_{n-1}]\cap X_*\neq\varnothing$, $b_{n-1}-a_{n-1}=((\sqrt{5}-1)/2)^{n-2}(b-a)$, и известна точка $\bar{x}_{n-1}$, производящая золотое сечение отрезка $[a_{n-1}, b_{n-1}]$ и такая, что $f(\bar{x}_{n-1})=\min\limits_{1\leqslant i\leqslant n-1} f(x_i)$ $(n\geqslant 2)$. Тогда в качестве следующей точки возьмем точку $x_n=a_{n-1}+b_{n-1}-\bar{x}_{n-1}$, также производящую золотое сечение отрезка $[a_{n-1}, b_{n-1}]$, вычислим значение $f(x_n)$.

Пусть для определенности $a_{n-1} < x_n < \bar{x}_{n-1} < b_{n-1}$ (случай $\bar{x}_{n-1} < x_n$ рассматривается аналогично). Если $f(x_n) \leqslant f(\bar{x}_{n-1})$, то полагаем $a_n = a_{n-1}$, $b_n = \bar{x}_{n-1}$, $\bar{x}_n = x_n$; если же $f(x_n) > f(\bar{x}_{n-1})$, то полагаем $a_n = x_n$, $b_n = b_{n-1}$, $\bar{x}_n = \bar{x}_{n-1}$. Новый отрезок $[a_n, b_n]$ таков, что $[a_n, b_n] \cap X_* \neq \varnothing$, $b_n - a_n = ((\sqrt{5}-1)/2)^{n-1}(b-a)$, точка $\bar{x}_n$ производит золотое сечение $[a_n, b_n]$ и $f(\bar{x}_n) = \min\{f(x_n); f(\bar{x}_{n-1})\} = \min\limits_{1 \leqslant i \leqslant n} f(x_i)$.

Если число вычислений значений $f(x)$ заранее не ограничено, то описанный процесс можно продолжать, например, до тех пор, пока не выполнится неравенство $b_n - a_n < \varepsilon$, где ε — заданная точность. Если же число вычислений значений функции $f(x)$ заранее жестко задано и равно n, то процесс на этом заканчивается и в качестве решения задачи второго типа (см. § 1) можно принять пару $f(\bar{x}_n)$, $\bar{x}_n$, где $f(\bar{x}_n)$ является приближением для $f_* = \inf\limits_{x \in [a,\, b]} f(x)$, а точка $\bar{x}_n$ служит приближением для множества X_* с погрешностью

$$\rho(\bar{x}_n, X_*) \leqslant \max\{b_n - \bar{x}_n; \bar{x}_n - a_n\} = \tfrac{1}{2}(\sqrt{5}-1)(b_n - a_n) = \left(\frac{\sqrt{5}-1}{2}\right)^n (b-a) = A_n.$$

Вспомним, что с помощью метода деления отрезка пополам за $n = 2k$ вычислений значений функции $f(x)$ в аналогичном случае мы получили точку $\bar{x}_n$ с погрешностью

$$\rho(\bar{x}_n, X_*) \leqslant 2^{-n/2}(b - a - \delta) < 2^{-n/2}(b-a) = B_n.$$

Отсюда имеем $A_n/B_n = (2\sqrt{2}/(\sqrt{5}+1))^n \approx (0,87\ldots)^n$ — видно, что уже при небольших n преимущество метода золотого сечения перед методом деления отрезка пополам становится ощутимым.

2. Обсудим возможности численной реализации метода золотого сечения на ЭВМ. Заметим, что число $\sqrt{5}$ на ЭВМ неизбежно будет задаваться приближенно, поэтому первая точка $x_1 = a + (3 - \sqrt{5})(b-a)/2$ будет найдена с некоторой погрешностью. Посмотрим, как повлияет эта погрешность на результаты последующих шагов метода золотого сечения. Обозначим $\Delta_n = b_n - a_n = ((\sqrt{5}-1)/2)^{n-1}(b-a)$ $(n = 1, 2, \ldots)$. Нетрудно проверить, что Δ_n является решением конечно-разностного уравнения $\Delta_{n-2} = \Delta_{n-1} + \Delta_n$, или

$$\Delta_n = \Delta_{n-2} - \Delta_{n-1}, \qquad n = 3, 4, \ldots, \tag{1}$$

с начальными условиями $\Delta_1 = b - a$, $\Delta_2 = b - x_1$.

Как известно, линейно независимые частные решения этого уравнения имеют вид τ_1^n и τ_2^n $(n = 1, 2, \ldots)$, где $\tau_1 = (\sqrt{5}-1)/2$, $\tau_2 = -(\sqrt{5}+1)/2$ — корни характеристического уравнения $\tau^2 + \tau - 1 = 0$, а любое решение уравнения (1) представимо в виде

$$\Delta_n = A\tau_1^n + B\tau_2^n, \qquad n = 1, 2, \ldots, \tag{2}$$

где постоянные A и B однозначно определяются начальными условиями из линейной системы

$$A\tau_1 + B\tau_2 = \Delta_1, \quad A\tau_1^2 + B\tau_2^2 = \Delta_2. \tag{3}$$

При $\Delta_1 = b - a$, $\Delta_2 = b - x_1$ из (3) имеем $A = 2(b-a)/(\sqrt{5}-1)$, $B = 0$, и понятно что формула (2) в этом случае дает уже известное нам решение $\Delta_n = \tau_1^{n-1}(b-a)$. Однако точка x_1 задана с погрешностью, поэтому

в системе (3) вместо точного значения Δ_2 придется взять приближенное $\widetilde{\Delta}_2 = \Delta_2 + \delta$. Тогда постоянные A, B из (3) определятся с соответствующими погрешностями: $\widetilde{A} = A + \delta_1$, $\widetilde{B} = B + \delta_2$, и вместо (2) с точными A, B будем иметь $\widetilde{\Delta}_n = \widetilde{A}\tau_1^n + \widetilde{B}\tau_2^n$ $(n = 1, 2, \ldots)$. Поскольку $0 < \tau_1 = 0,6\ldots < 1$, $|\tau_2| = 1,6\ldots > 1$, то погрешность $|\Delta_n - \widetilde{\Delta}_n| = |\delta_1\tau_1^n + \delta_2\tau_2^n|$ с возрастанием n будет расти очень быстро. Это значит, что уже при не очень больших n отрезок $[a_n, b_n]$ и точки $\bar{x}_n$, $x_{n+1} = a_n + b_n - \bar{x}_n$ будут сильно отличаться от тех, которые получились бы при работе с точными данными. Численные эксперименты на ЭВМ также подтверждают, что метод золотого сечения в описанном выше виде практически неприменим уже при небольших n.

Как же быть? К счастью, имеется достаточно простая модификация метода золотого сечения, позволяющая избежать слишком быстрого возрастания погрешностей при определении точек x_n $(n \geqslant 2)$. А именно на каждом отрезке $[a_n, b_n]$, содержащем точку $\bar{x}_n$ с предыдущего шага, при выборе следующей точки x_{n+1} нужно остерегаться пользоваться формулой $x_{n+1} = a_n + b_n - \bar{x}_n$, и вместо этого лучше непосредственно произвести золотое сечение отрезка $[a_n, b_n]$ и в качестве x_{n+1} взять ту из точек $a_n + (3 - \sqrt{5})(b_n - a_n)/2$, $a_n + (\sqrt{5}-1)(b_n - a_n)/2$, которая наиболее удалена от $\bar{x}_n$ (здесь под $\sqrt{5}$ подразумевается какое-либо подходящее приближение этого числа). Конечно, после такой модификации метод золотого сечения, вообще говоря, теряет свойство симметричности и, быть может, уже не так красив, но зато вполне годится для приложений. Нетрудно видеть, что этот метод может применяться и без априорного знания о том, что минимизируемая функция унимодальна, но в этом случае полученное решение может оказаться далеким от глобального минимума.

3. Метод золотого сечения относится к классу так называемых *симметричных методов*. Дадим краткое описание произвольного симметричного метода минимизации функции $f(x)$ на отрезке $[a, b]$.

Первый шаг: на $[a, b]$ задается точка x_1 $(a < x_1 < b)$, полагается $a_1 = a$, $b_1 = b$, $\bar{x}_1 = x_1$ и вычисляется $f(x_1)$. Пусть уже сделано $n-1$ шагов $(n \geqslant 2)$ и найдены отрезок $[a_{n-1}, b_{n-1}]$ и точка $\bar{x}_{n-1}$ $(a_{n-1} < \bar{x}_{n-1} < b_{n-1})$ с вычисленным значением $f(\bar{x}_{n-1})$, причем $\bar{x}_{n-1} \neq (a_{n-1} + b_{n-1})/2$. Тогда на следующем n-м шаге берется точка $x_n = a_{n-1} + b_{n-1} - \bar{x}_{n-1}$, расположенная внутри $[a_{n-1}, b_{n-1}]$, симметрично точке $\bar{x}_{n-1}$ относительно середины этого отрезка — отсюда происходит название методов. Затем вычисляется значение $f(x_n)$ и сравнивается с $f(\bar{x}_{n-1})$. Пусть для определенности $\bar{x}_{n-1} < x_n$ (случай $x_n < \bar{x}_{n-1}$ рассматривается аналогично). Тогда при $f(\bar{x}_{n-1}) \leqslant f(x_n)$ полагается $a_n = a_{n-1}$, $b_n = x_n$, $\bar{x}_n = \bar{x}_{n-1}$; если же $f(\bar{x}_{n-1}) > f(x_n)$, то $a_n = \bar{x}_{n-1}$, $b_n = b_{n-1}$, $\bar{x}_n = x_n$. Если $\bar{x}_n \neq (a_n + b_n)/2$, то процесс может быть продолжен дальше. Может оказаться, что $\bar{x}_n = (a_n + b_n)/2$, — в этом случае процесс заканчивается; при необходимости на $[a_n, b_n]$ можно продолжать поиск минимума аналогичным методом, начиная с выбора новой начальной точки $\bar{x}_n \neq (a_n + b_n)/2$.

Из описания симметричного метода видно, что всякий симметричный метод полностью определяется заданием отрезка $[a, b]$ и первой точки x_1 $(a < x_1 < b)$. Отсюда следует, что в качестве другой характеристики симметричного метода можно взять длины Δ_n отрезков $[a_n, b_n]$ $(n = 1, 2, \ldots)$, где $\Delta_1 = b - a$, $\Delta_2 = \max\{b - x_1; x_1 - a\}$. Очевидно, $\Delta_{n+1} \geqslant \Delta_n/2$ при всех $n \geqslant 1$. Как видим, симметричные методы весьма просты и, пожалуй, даже изящны. Однако все эти методы страдают тем же недостатком, что и метод золотого сечения: погрешность, допущенная в задании первой точки x_1, приводит к быстрому накапливанию погрешностей на дальнейших шагах, и уже при не очень больших n результаты будут сильно отличаться от тех, которые могли бы получиться при точной реализации симметричного метода с точными исходными данными.

Если симметричный метод таков, что для $\Delta_n = b_n - a_n$ выполнено условие

$$\Delta_n/2 < \Delta_{n+1} \leqslant 2\Delta_n/3, \qquad n = 1, \ldots, N, \tag{4}$$

при некотором $N > 1$, то Δ_n будут удовлетворять конечно-разностному уравнению (1) при $n = 2, \ldots, N$, и исследование поведения погрешностей в этом случае может быть проведено так же, как это было сделано выше для метода золотого сечения. Чтобы избежать слишком быстрого роста погрешностей в симметричных методах со свойством (4), на каждом отрезке $[a_n, b_n]$ $(n = 2, \ldots, N)$, содержащем точку $\bar{x}_n$ с предыдущего шага, следующую точку x_{n+1} нужно определять не по формуле $x_{n+1} = a_n + b_n - \bar{x}_n$, а лучше принять за x_{n+1} ту из точек $a_n + \tau(b_n - a_n)$, $a_n + (1-\tau)(b_n - a_n)$ $(\tau = (\Delta_2 + \delta)/\Delta_1)$, которая наиболее удалена от $\bar{x}_n$.

Упражнения

1. Найти наименьшее n, начиная с которого точность метода золотого сечения больше точности метода деления отрезка пополам в 2 раза; в 10 раз.

2. Написать конечно-разностные уравнения для длин Δ_n отрезков $[a_n, b_n]$, получаемых симметричным методом, для случая, когда на каких-то шагах метода нарушается условие (4).

§ 5. Об оптимальных методах

1. В тех случаях, когда вычисление значений функции связано со значительными затратами, большую ценность приобретают экономичные или, как их еще называют, *оптимальные методы*, позволяющие решить задачу минимизации с требуемой точностью на основе вычислений значений минимизируемой функции как можно в меньшем числе точек, а также тесно связанные с ними методы, гарантирующие наилучшую точность при жестко заданном количестве вычислений значений минимизируемой функции. В связи с этим возникают вопросы, что такое оптимальные методы, существуют ли такие методы, как их строить? Абсолютно наилучший метод, пригодный для минимизации всех функций, вряд ли существует, и на поставленные вопросы можно попытаться ответить лишь при определенных ограничениях на рассматриваемые методы, функции и постановки задач минимизации.

Предположим, что нам задан некоторый класс функций Q, зафиксирована какая-либо постановка задачи минимизации функций из этого класса (например, задача первого или второго типа из § 1) и указано множество методов P, позволяющих решить поставленную задачу минимизации. Пусть $\Delta(f, p)$ — погрешность решения рассматриваемой задачи минимизации для функции $f = f(x) \in Q$ с помощью метода $p \in P$. Ясно, что, минимизируя одним и тем же методом p различные функции из Q, мы будем получать, вообще говоря, различные погрешности: для некоторых «хороших» функций из Q эта погрешность может оказаться равной нулю, а для других «плохих» функций из Q погрешность может быть значительной. Имеет смысл считать метод $p_1 \in P$ лучше метода $p_2 \in P$, если погрешность метода p_1 даже для самых «плохих» (для p_1) функций из Q будет меньше погрешности метода p_2 для «плохих» (для p_2) функций из Q. В связи с этим представляется разумным ввести величину $\delta(p) = \sup\limits_{f \in Q} \Delta(f, p)$, выражающую собой погрешность метода p при минимизации самой «плохой» (для p) функции из Q.

Определение 1. Величину $\delta(p) = \sup\limits_{f \in Q} \Delta(f, p)$ назовем *гарантированной точностью метода $p \in P$ на классе функций Q*. Скажем, что

метод $p_1 \in P$ лучше метода $p_2 \in P$ на классе Q, если $\delta(p_1) < \delta(p_2)$. Метод $p_* \in P$ назовем *оптимальным методом на классе Q*, если $\delta(p_*) = \inf\limits_{p \in P} \delta(p) = \delta(p_*)$, а величину $\delta(p_*)$ — *наилучшей гарантированной точностью методов P на классе Q*. Если для некоторого метода $p_\varepsilon \in P$ выполняется неравенство $\delta(p_\varepsilon) \leqslant \delta(p_*) + \varepsilon$, то метод p_ε назовем *ε-оптимальным на классе Q*.

Вопросы существования оптимальных и ε-оптимальных методов, возможности их построения для различных множеств методов P, классов функций Q и постановок задач минимизации, а также другие возможные подходы к проблеме выбора оптимальных методов изучались, например, в [74; 140; 148; 193; 214; 218; 374; 523; 671; 681; 684; 704; 709; 755].

2. Здесь мы кратко остановимся на оптимальных методах решения задачи минимизации функций из класса Q, состоящего из всех унимодальных функций на отрезке $[a, b]$. Ограничимся рассмотрением множества P методов минимизации, использующих лишь значения функции, считая при этом, что число n вычислений значений минимизируемой функции заранее задано. Будем предполагать, что в описание каждого метода p_n из P входит задание правила выбора точек $x_1, \ldots, x_n$ из отрезка $[a, b]$, вычисление значений $f(x_1), \ldots, f(x_n)$ минимизируемой функции $f(x) \in Q$, выделение из точек $x_1, \ldots, x_n$ такой точки $\bar{x}_n$, для которой $f(\bar{x}_n) = \min\limits_{1 \leqslant i \leqslant n} f(x_i)$, и определение отрезка $[a_n, b_n]$, где в качестве a_n, b_n берутся ближайшие слева или справа к $\bar{x}_n$ точки среди $x_1, \ldots, x_n, a, b$ (возможности $x_i = a$ или $x_i = b$ не исключаются).

Таким образом, применяя конкретный метод $p_n \in P$ к конкретной функции $f(x) \in Q$, в результате получаем отрезок $[a_n, b_n]$ и точку $\bar{x}_n \in [a_n, b_n]$ с вычисленным значением $f(\bar{x}_n) = \min\limits_{1 \leqslant i \leqslant n} f(x_i)$. Из определения унимодальной функции и построения отрезка $[a_n, b_n]$ следует неравенство $f(x) \geqslant f(\bar{x}_n)$ при всех $x \in [a, b] \setminus [a_n, b_n]$, так что

$$f_* = \inf_{a \leqslant x \leqslant b} f(x) = \inf_{a_n \leqslant x \leqslant b_n} f(x), \quad X_* \cap [a_n, b_n] \neq \varnothing. \tag{1}$$

В качестве приближения для f_* обычно берут величину $f(\bar{x}_n)$, а в качестве приближения к множеству X_* можно взять любую точку x_{n_*} из отрезка $[a_n, b_n]$ — на практике часто принимают $x_{n_*} = \bar{x}_n$ или $x_{n_*} = (a_n + b_n)/2$.

Отрезок $[a_n, b_n]$ принято называть *отрезком локализации минимума* функции $f(x)$ на отрезке $[a, b]$. Из (1) следует что расстояние от любой точки $x_{n_*} \in [a_n, b_n]$ до множества X_* не превышает длины отрезка локализации $b_n - a_n$:

$$\rho(x_{n_*}, X_*) = \inf_{x \in X_*} |x_{n_*} - x| \leqslant b_n - a_n. \tag{2}$$

Величину $\Delta(f, p_n) = b_n - a_n$ можно принять за погрешность решения задачи минимизации функции $f(x) \in Q$ методом $p_n \in P$. Согласно (2), чем меньше погрешность $\Delta(f, p_n)$, тем точнее будет определено приближение x_{n_*} к X_* и, следовательно, тем лучше метод p_n. Для точного определения наилучшего или близкого к нему метода нам остается еще уточнить правило выбора точек $x_1, \ldots, x_n$, в которых вычисляются значения минимизируемой функции. Здесь принято различать два типа методов: пассивные методы и последовательные методы.

Если все точки $x_1, \ldots, x_n$ метода p_n выбираются одновременно до начала вычислений и в дальнейшем уже не меняются, то такой метод называют *пассивным*. Если в методе p_n точки $x_1, \ldots, x_n$ выбираются последовательно отдельными порциями, причем при выборе каждой очередной порции учитываются результаты предыдущих вычислений и проводится уточнение отрезка локализации минимума, то такой метод называется *последовательным*.

Примером пассивного метода является *метод равномерного перебора*. В этом методе точки $x_1, \ldots, x_n$ выбираются по правилу: $x_i = x_1 + ih$ $(i = 1, \ldots, n)$, где $h > 0$ — шаг метода, x_1 — заданная точка из $[a, b]$, $x_1 - a \leqslant h$ (например, $x_1 = a$ или $x_1 = a + h/2$), и, кроме того, $nh \leqslant b - x_1 < (n+1)h$.

Примерами последовательного метода служат методы деления отрезка пополам, золотого сечения.

Пассивный метод является частным случаем последовательного метода, когда все n точек выбираются сразу в первой же порции. Поэтому нетрудно понять, что последовательные методы, вообще говоря, обладают большей гибкостью и гораздо точнее пассивных методов. Однако отсюда не следует, что пассивные методы вовсе не находят применения. Такие методы

весьма полезны, когда можно вести параллельные вычисления, используя, например, многопроцессорные ЭВМ. В тех случаях, когда значения минимизируемой функции определяются из физического эксперимента, условия проведения таких экспериментов также могут сделать необходимым применение пассивных методов.

Таким образом, n-точечные (т. е. использующие вычисление значений функции в n точках) последовательные и пассивные методы описаны. Если в определении 1 принять, что Q — класс унимодальных функций на отрезке $[a, b]$, $P = P_n$ — множество всех n-точечных последовательных (или пассивных) методов, $\Delta(f, p_n) = b_n - a_n$ — длина отрезка локализации минимума, полученного минимумом p_n для функции $f = f(x) \in Q$, то придем к определениям гарантированной точности, оптимального и ε-оптимального последовательного (пассивного) метода для унимодальных функций. Кратко рассмотрим вопросы существования и построения оптимальных методов для таких функций.

3. Сначала остановимся на пассивных методах. Пусть $p_n = \{x_1, \ldots, x_n\}$ — какой-либо пассивный метод, $a = x_0 \leqslant x_1 < \ldots < x_n \leqslant b = x_{n+1}$. Применяя его к какой-либо функции $f(x) \in Q$, получаем отрезок $[x_{i-1}, x_{i+1}]$ локализации минимума этой функции, так что погрешность метода p_n здесь будет равна $\Delta(f, p_n) = x_{i+1} - x_{i-1}$. Поэтому $\delta(p_n) = \sup\limits_{f \in Q} \Delta(f, p_n) \leqslant \max\limits_{1 \leqslant i \leqslant n} (x_{i+1} - x_{i-1})$. Пусть $\max\limits_{1 \leqslant i \leqslant n} (x_{i+1} - x_{i-1}) = x_{k+1} - x_{k-1}$. Возьмем функцию $f_k = f_k(x) = |x - x_k|$. Это строго унимодальная функция достигает своей нижней грани на отрезке $[a, b]$ в точке $x_* = x_k \in [x_{k-1}, x_{k+1}]$, причем $\Delta(f_k, p_n) = x_{k+1} - x_{k-1}$. Следовательно, гарантированная погрешность метода p_n на классе Q равна $\delta(p_n) = \max\limits_{1 \leqslant i \leqslant n} (x_{i+1} - x_{i-1})$. Для получения оптимального пассивного метода остается выяснить, достигается ли нижняя грань $\inf\limits_{p_n \in P_n} \delta(p_n) = \delta_*$, где P_n — множество всех пассивных методов, и если достигается, то на каком методе $p_n \in P_n$. Оказывается, здесь нужно различать случаи четного и нечетного n.

Теорема 1. *При всех нечетных $n = 2m + 1$ $(m \geqslant 0)$ существует бесконечно много оптимальных пассивных методов на классе Q; наилучшая гарантированная точность пассивных методов P_{2m+1} на этом классе равна $(b - a)/(m + 1)$.*

Доказательство. Возьмем пассивный метод $p_{n_*} = \{v_1, \ldots, v_n\}$, где $v_{2i} = a + i(b - a)/(m + 1)$ $(i = 1, \ldots, m)$, а точки v_{2i+1} $(i = 0, 1, \ldots, m)$ расположены на отрезке $[a, b]$ произвольно, лишь бы $v_{2i-1} < v_{2i} < v_{2i+1}$, $v_{2i+1} - v_{2i-1} \leqslant (b - a)/(m+1)$. Очевидно, $\delta(p_{n_*}) = (b - a)/(m + 1)$. С другой стороны, для любого пассивного метода $p_n = \{x_1, \ldots, x_n\}$ $(x_0 = a \leqslant x_1 < \ldots < x_n \leqslant b = x_{n+1}$, $n = 2m + 1)$ имеем $\delta(p_n) = \max\limits_{1 \leqslant i \leqslant n} (x_{i+1} - x_{i-1}) \geqslant \max\{b - x_{2m}, x_{2m} - x_{2m-2}, \ldots, x_4 - x_2, x_2 - a\} \geqslant (b - a)/(m+1)$. Следовательно, методы p_{n_*} оптимальны и $\delta_* = (b - a)/(m + 1)$. □

Теорема 2. *При всех четных $n = 2m$ $(m \geqslant 1)$ оптимального пассивного метода на классе Q не существует; наилучшая гарантированная точность пассивных методов P_{2n} на этом классе равна $(b - a)/(m + 1)$. В качестве ε-оптимального метода можно взять $p_{n\varepsilon} = \{v_1, \ldots, v_n\}$, где $v_{2i-1} = a + i(b - a)/(m+1) - \varepsilon$, $v_{2i} = a + i(b - a)/(m+1) + \varepsilon$ $(i = 1, \ldots, m$, $0 < \varepsilon < (b - a)/(2(m + 1)))$.*

Доказательство. Сначала убедимся в том, что $\delta(p_n) > (b - a)/(m + 1)$ для любого пассивного метода $p_n = \{x_1, \ldots, x_n\}$ $(x_0 = a \leqslant x_1 < \ldots < x_n \leqslant b = x_{n+1}$, $n = 2m)$. Обозначим $\bar{x} = a + (b - a)/(m + 1)$. Имеются две возможности: либо $x_2 > \bar{x}$, либо $x_2 \leqslant \bar{x}$. Если $x_2 > \bar{x}$, то $\delta(p_n) = \max\limits_{1 \leqslant i \leqslant n} (x_{i+1} - x_{i-1}) \geqslant x_2 - a > \bar{x} - a = (b - a)/(m + 1)$. Если же $x_2 \leqslant \bar{x}$, то $\delta(p_n) = \max\limits_{1 \leqslant i \leqslant n} (x_{i+1} - x_{i-1}) > x_2 - a$. В самом деле, если бы $\max\limits_{1 \leqslant i \leqslant n} (x_{i+1} - x_{i-1}) \leqslant x_2 - a$, то $x_{2i+1} - x_{2i-1} \leqslant x_2 - a$ $(i = 1, \ldots, m)$ и, кроме того, $x_1 - a < x_2 - a$. Сложив эти неравенства, придем к противоречивому неравенству $b - a < (m + 1)(x_2 - a) \leqslant (m + 1)(\bar{x} - a) = b - a$.

Таким образом, при $x_2 < \bar{x}$ имеем $\delta(p_n) = \max\limits_{2 \leqslant i \leqslant n} (x_{i+1} - x_{i-1}) = \delta(p'_{n-1})$, где p'_{n-1} — пассивный метод на отрезке $[x_1, b]$, составленный из точек $x_2, \ldots, x_n$ метода p_n. Но $n - 1 = 2m - 1$ — нечетное число, поэтому, применяя теорему 1 к отрезку $[x_1, b]$, имеем $\delta(p'_{n-1}) \geqslant (b - x_1)/m$. Тогда $\delta(p_n) = \delta(p'_{n-1}) \geqslant (b - x_1)/m > (b - \bar{x})/m = (b - a)/(m+1)$. Тем самым доказано, что $\delta(p_n) > (b - a)/(m + 1)$ при всех $p_n \in P_{2m}$. С другой стороны, $\delta(p_{n\varepsilon}) = (b - a)/(m + 1) + \varepsilon$ для всех ε $(0 < \varepsilon < (b - a)/(2(m + 1)))$. Следовательно, $\delta_* = (b - a)/(m + 1)$. □

Из теорем 1, 2 вытекает, что предпочтительнее пользоваться пассивными методами с четным числом $n = 2m$ точек, поскольку в случае $n = 2m + 1$ наилучшая гарантированная точность остается такой же, как и при $n = 2m$.

4. Перейдем к рассмотрению последовательных методов минимизации унимодальных функций на $[a, b]$. Здесь нам понадобятся знаменитые *числа Фибоначчи*, которые, как известно [193], определяются соотношениями

$$F_{n+2} = F_{n+1} + F_n, \quad n = 1, 2, \ldots, \quad F_1 = F_2 = 1.$$

С помощью индукции легко показать, что n-е число Фибоначчи представимо в виде

$$F_n = \left[\left(\frac{1 + \sqrt{5}}{2}\right)^n - \left(\frac{1 - \sqrt{5}}{2}\right)^n\right] \frac{1}{\sqrt{5}}, \qquad n = 1, 2, \ldots \tag{3}$$

Используя числа F_n, построим n-точечный последовательный метод, который принято называть *методом Фибоначчи*. Этот метод относится к классу симметричных методов, описанных в § 4, и определяется заданием на отрезке $[a, b]$ точки $x_1 = a + (b - a)F_{n-1}/F_{n+1}$ или симметричной ей точки $x_2 = a + b - x_1 = a + (b - a)F_n/F_{n+1}$. С помощью индукции нетрудно показать, что такой симметричный метод обладает свойством (4.4) и на k-м шаге $(k < n)$, когда проведены вычисления значений функции в точках $x_1, \ldots, x_k$, приводит к отрезку локализации минимума $[a_k, b_k]$ длиной

$$\Delta_k = b_k - a_k = (b - a)F_{n-k+2}/F_{n+1},$$

причем точка $\bar{x}_k$ $(a_k < \bar{x}_k < b_k)$ с вычисленным значением $f(\bar{x}_k) = \min\limits_{1 \leqslant i \leqslant k} f(x_i)$ совпадает с одной из точек

$$\begin{aligned} x'_k &= a_k + (b_k - a_k)\frac{F_{n-k}}{F_{n-k+2}} = a_k + (b - a)\frac{F_{n-k}}{F_{n+1}}, \\ x''_k &= a_k + (b_k - a_k)\frac{F_{n-k+1}}{F_{n-k+2}} = a_k + (b - a)\frac{F_{n-k+1}}{F_{n+1}} = a_k + b_k - x'_k, \end{aligned} \tag{4}$$

расположенных на отрезке $[a_k, b_k]$ симметрично относительно его середины.

Как видно из (4), при $k = n - 1$ точки x'_{n-1}, x''_{n-1} совпадают. Это означает, что при $k = n - 1$ первая часть процесса заканчивается вычислением значения функции в точке x_{n-1} и определением отрезка локализации минимума $[a_{n-1}, b_{n-1}]$ длины $b_{n-1} - a_{n-1} = (b - a)F_3/F_{n+1} = 2(b - a)/F_{n+1}$, причем точка $x'_{n-1} = x''_{n-1} = \bar{x}_{n-1}$ совпадает с серединой отрезка $[a_{n-1}, b_{n-1}]$. В заключение, несколько нарушая симметричность процесса, последнее n-е вычисление значения минимизируемой функции $f(x)$ проводится в точке $x_n = \bar{x}_{n-1} + \varepsilon$ (или $x_n = \bar{x}_{n-1} - \varepsilon$), где $0 < \varepsilon < (b - a)/F_{n+1}$, и отрезок $[a_n, b_n]$ локализации минимума определяется по формулам $a_n = a_{n-1}$, $b_n = \bar{x}_{n-1} + \varepsilon$ при $f(\bar{x}_{n-1}) \leqslant f(\bar{x}_{n-1} + \varepsilon)$ и $a_n = \bar{x}_{n-1}$, $b_n = b_{n-1}$ при $f(\bar{x}_{n-1}) > f(\bar{x}_{n-1} + \varepsilon)$, так что в худшем случае $b_n - a_n = (b - a)/F_{n+1} + \varepsilon$. Описанный метод обозначим через Φ_n.

Теорема 3. *При всех $n > 1$ оптимального последовательного метода на классе унимодальных функций не существует; наилучшая гарантированная точность последовательных методов на этом классе равна $(b - a)/F_{n+1}$. В качестве ε-оптимального метода можно взять метод Фибоначчи Φ_n.*

Доказательство этой теоремы можно найти в [148; 374; 684].

Заметим, что число F_{n-1}/F_{n+1}, вообще говоря, является бесконечной периодической десятичной дробью, поэтому первая точка x_1 метода Φ_n будет задаваться на ЭВМ приближенно. Во избежание быстрого роста погрешности из-за неточности задания первой точки на практике нужно пользоваться модификацией метода Φ_n, описанной в § 4 для симметричных методов в общем случае.

Следует подчеркнуть, что метод Φ_n для своей реализации требует, чтобы число n вычислений значений минимизируемой функции было задано заранее — выбор первой точки в этом методе невозможен без знания n. В тех случаях, когда число n по каким-либо причинам не может быть задано заранее, можно применять метод золотого сечения, не требующий для своей реализации априорного знания n.

Для сравнения вспомним, что методом золотого сечения за n вычислений значений функции мы получали отрезок $[a_n, b_n]$ локализации минимума длины $b_n - a_n = ((\sqrt{5} - 1)/2)^{n-1}(b - a) = (2/(\sqrt{5} + 1))^{n-1}(b - a)$. С учетом формулы $F_{n+1} \approx ((\sqrt{5} + 1)/2)^{n+1}/\sqrt{5}$, вытекающей из (3) при больших n, для метода Φ_n получаем отрезок локализации минимума, длина которого близка к $(b - a)/F_{n+1} \approx (2/(\sqrt{5} + 1))^{n+1}(b - a)/\sqrt{5}$. Отсюда следует, что метод золотого сечения хуже метода Φ_n при больших n всего в $((\sqrt{5} + 1)/2)^2/\sqrt{5} = 1,1708\ldots$ раз, т. е. на классе унимодальных функций метод золотого сечения близок к оптимальным методам. Интересно

также заметить, что

$$\lim_{n\to\infty}\frac{F_{n-1}}{F_{n+1}}=\frac{3-\sqrt{5}}{2},\quad \lim_{n\to\infty}\frac{F_n}{F_{n+1}}=\frac{\sqrt{5}-1}{2},$$

т. е. при достаточно больших n начальные точки x_1, x_2 методов Фибоначчи и золотого сечения практически совпадают.

Упражнения

1. Найти гарантированную на классе унимодальных функций точность последовательного (или пассивного) n-точечного метода p_n, если в качестве погрешности метода p_n при минимизации функции $f=f(x)$ принята величина $\Delta(f,p_n)=|f_*-f(\bar{x}_n)|$ [755].

2. Сравнить оптимальные и ε-оптимальные пассивные методы на классе унимодальных функций с методом деления отрезка пополам.

3. Указать все точки метода Φ_n на отрезке $[0,1]$ при $n=2,3,4,5$.

4. Применить метод Φ_5 к функциям $f(x)=x$, $f(x)=|x-1|$ на отрезке $[0,2]$.

5. Найти наименьшее n, для которого точность метода золотого сечения хуже точности метода Фибоначчи в 2 раза.

6. Доказать, что число Фибоначчи F_n является ближайшим целым числом к $((1+\sqrt{5})/2)^n/\sqrt{5}$.

7. Доказать, что решение уравнения (4.1) представимо в виде $\Delta_n=(-1)^nF_{n-1}\Delta_2+(-1)^{n-1}F_{n-2}\Delta_1$ $(n=3,4,\ldots)$. Отсюда вывести закон изменения погрешности величины Δ_n, если Δ_1, Δ_2 заданы неточно.

8. Доказать, что последовательность $\{F_{2m}/F_{2m+1}\}$ сходится к $\tau_1=(\sqrt{5}-1)/2$, монотонно возрастая, а $\{F_{2m-1}/F_{2m}\}$ сходится к τ_1 монотонно убывая.

9. Используя утверждения упражнений 7, 8, доказать, что метод золотого сечения является единственным симметричным методом, удовлетворяющим условию (4.4) при всех $n=1,2\ldots$

10. Пусть дан симметричный метод с начальными отрезками Δ_1, Δ_2, пусть $N\geqslant 2$ — заданное натуральное число. Используя утверждения упражнений 7, 8, указать промежуток изменения отношения Δ_2/Δ_1, чтобы метод удовлетворял условию (4.4) при всех $n=1,\ldots,N$.

11. Пусть дан некоторый симметричный метод, удовлетворяющий условию (4.4) при $n=1$. Используя утверждения упражнений 7, 8, указать максимальное число N, при котором условие (4.4) выполняется для всех $n=2,\ldots,N$.

§ 6. Метод ломаных

Описанные выше методы часто приходится применять без априорного знания о том, что минимизируемая функция является унимодальной. Однако в этом случае погрешности в определении минимального значения и точек минимума функции могут быть значительными. Например, применение этих методов к минимизации непрерывных на отрезке функций приведет, вообще говоря, лишь в окрестность точки локального минимума, в которой значение функции может сильно отличаться от искомого минимального значения на отрезке. Поэтому представляется важной разработка методов поиска глобального минимума, позволяющих строить минимизирующие последовательности и получить приближенное решение задач минимизации первого и второго типов (см. § 1) для функций, не обязательно унимодальных.

Здесь мы рассмотрим один из таких методов для класса функций, удовлетворяющих условию Липшица.

Определение 1. Говорят, что функция $f(x)$ удовлетворяет *условию Липшица* на отрезке $[a,b]$, если существует постоянная $L>0$ такая, что

$$|f(x)-f(y)|\leqslant L|x-y| \qquad \forall x,y\in[a,b]. \tag{1}$$

Постоянную L называют *постоянной Липшица* функции $f(x)$ на $[a,b]$.

Условие (1) имеет простой геометрический смысл: оно означает, что угловой коэффициент (тангенс угла наклона) $|f(x)-f(y)|\cdot|x-y|^{-1}$ хорды, соединяющей точки $(x,f(x))$ и $(y,f(y))$ графика функции, не превышает постоянной L для всех точек $x,y\in[a,b]$. Из (1) следует, что функция $f(x)$ непрерывна на отрезке $[a,b]$, так что по теореме 1.1 множество X_* точек минимума $f(x)$ на $[a,b]$ непусто.

Теорема 1. *Пусть функция $f(x)$ непрерывна на отрезке $[a,b]$ и на каждом отрезке $[a_i,a_{i+1}]$ $(i=1,\ldots,m)$, где $a_1=a$, $a_{m+1}=b$, удовлетворяет условию (1) с постоянной L_i. Тогда $f(x)$ удовлетворяет условию (1) на всем отрезке с постоянной $L=\max\limits_{1\leqslant i\leqslant m}L_i$.*

Доказательство. Возьмем две произвольные точки $x,y\in[a,b]$. Пусть $a_{p-1}\leqslant x\leqslant a_p$, $a_s\leqslant y\leqslant a_{s+1}$ при некоторых p, s. Тогда

$$|f(x)-f(y)|=|f(x)-f(a_p)+\sum_{i=p}^{s-1}(f(a_i)-f(a_{i+1}))+f(a_s)-$$
$$-f(y)|\leqslant L_{p-1}|x-a_p|+\Big|\sum_{i=p}^{s-1}L_i(a_{i+1}-a_i)\Big|+L_s|a_s-y|\leqslant L|x-y|. \quad \square$$

Теорема 2. *Пусть функция $f(x)$ дифференцируема на отрезке $[a,b]$ и ее производная $f'(x)$ ограничена на этом отрезке. Тогда $f(x)$ удовлетворяет условию (1) с постоянной $=\sup\limits_{x\in[a,b]}|f'(x)|$.*

Доказательство. По формуле конечных приращений для любых $x,y\in[a,b]$ имеем $f(x)-f(y)=f'(y+\theta(x-y))(x-y)$ $(0<\theta<1)$. Отсюда и из ограниченности $f'(x)$ следует утверждение теоремы. $\square$

Пусть функция $f(x)$ удовлетворяет условию (1) на отрезке $[a,b]$. Зафиксируем какую-либо точку $y\in[a,b]$ и определим функцию $g(x,y)=f(y)-L|x-y|$ переменной x $(a\leqslant x\leqslant b)$. Очевидно, функция $g(x,y)$ кусочно линейна на $[a,b]$, и график ее представляет собой ломаную линию, составленную из отрезков двух прямых, имеющих угловые коэффициенты L и $-L$ и пересекающихся в точке $(y,f(y))$. Кроме того, в силу условия (1)

$$f(x)-g(x,y)\geqslant(L-|f(x)-f(y)||x-y|^{-1})|x-y|\geqslant 0, \quad x\neq y,$$

т. е.

$$g(x,y)=f(y)-L|x-y|\leqslant f(x) \qquad \forall x\in[a,b], \tag{2}$$

причем $g(y,y)=f(y)$. Это значит, что график функции $f(x)$ лежит выше ломаной $g(x,y)$ при всех $x\in[a,b]$ и имеет с ней общую точку $(y,f(y))$.

Свойство (2) ломаной $g(x,y)$ можно использовать для построения следующего метода [257], который назовем *методом ломаных*. Этот метод начинается с выбора произвольной точки $x_0\in[a,b]$ и составления функции $g(x,x_0)=f(x_0)-L|x-x_0|=p_0(x)$. Следующая точка x_1 определяется из условий $p_0(x_1)=\min\limits_{x\in[a,b]}p_0(x)$ $(x_1\in[a,b])$; очевидно, $x_1=a$ или $x_1=b$. Далее

берется новая функция $p_1(x) = \max\{g(x, x_1); p_0(x)\}$, и очередная точка x_2 находится из условий $p_1(x_2) = \min\limits_{x \in [a, b]} p_1(x)$ ($x_2 \in [a, b]$) и т. д. (рис. 1.2).

Пусть точки $x_0, x_1, \ldots, x_n$ ($n \geqslant 1$) уже известны. Тогда составляется функция

$$p_n(x) = \max\{g(x, x_n), p_{n-1}(x)\} = \max_{0 \leqslant i \leqslant n} g(x, x_i),$$

и следующая точка x_{n+1} определяется условиями

$$p_n(x_{n+1}) = \min_{x \in [a, b]} p_n(x), \quad x_{n+1} \in [a, b]. \tag{3}$$

Если минимум $p_n(x)$ на $[a, b]$ достигается в нескольких точках, то в качестве x_{n+1} можно взять любую из них.

Метод ломаных описан. Очевидно, $p_n(x)$ является кусочно линейной функцией и график ее представляет собой непрерывную ломаную линию, состоящую из отрезков прямых с угловыми наклонами L или $-L$. Из теоремы 1 следует, что $p_n(x)$ удовлетворяет условию (1) с той же постоянной L, что и функция $f(x)$. Ясно также, что

$$p_{n-1}(x) = \max_{0 \leqslant i \leqslant n-1} g(x, x_i) \leqslant \max_{0 \leqslant i \leqslant n} g(x, x_i) = p_n(x), \quad x \in [a, b]. \tag{4}$$

Кроме того, согласно (2) функция $g(x, x_i) \leqslant f(x)$ ($x \in [a, b]$) для всех $i = 0, 1, \ldots, n$, поэтому

$$p_n(x) \leqslant f(x), \ x \in [a, b], \quad n = 0, 1, \ldots \tag{5}$$

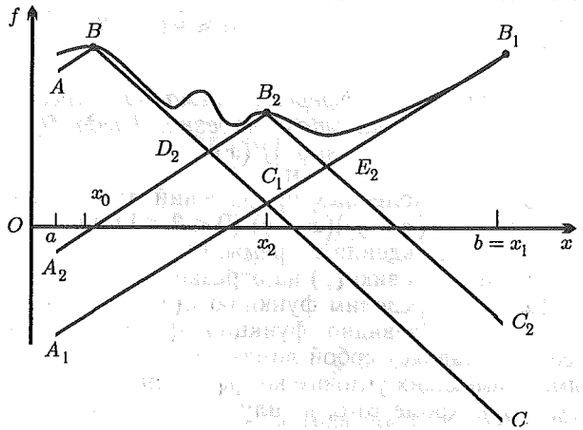

Рис. 1.2. ABC — график $p_0(x) = g(x, x_0)$, A_1B_1 — график $p_0(x) = g(x, x_1)$, ABC_1B_1 — график $p_1(x)$, $A_2B_2C_2$ — график $g(x, x_2)$, $ABD_2B_2E_2B_1$ — график $p_2(x)$

Таким образом, на каждом шаге метода ломаных задача минимизации функции $f(x)$ заменяется более простой задачей минимизации кусочно линейной функции $p_n(x)$, которая приближает $f(x)$ снизу, причем согласно (4) $\{p_n(x)\}$ монотонно возрастает. Докажем теперь, что при неограниченном увеличении n метод ломаных сходится.

Теорема 3. *Пусть $f(x)$ — произвольная функция, удовлетворяющая на отрезке $[a, b]$ условию (1). Тогда последовательность $\{x_n\}$, полученная с помощью описанного метода ломаных, такова, что:*

1) $\lim\limits_{n \to \infty} f(x_n) = \lim\limits_{n \to \infty} p_n(x_{n+1}) = f_* = \inf\limits_{x \in [a, b]} f(x)$, *причем справедлива оценка*

$$0 \leqslant f(x_{n+1}) - f_* \leqslant f(x_{n+1}) - p_n(x_{n+1}), \quad n = 0, 1, \ldots; \tag{6}$$

2) *$\{x_n\}$ сходится к множеству X_* точек минимума $f(x)$ на $[a, b]$, т. е.* $\lim\limits_{n \to \infty} \rho(x_n, X_*) = 0$.

Доказательство. Возьмем произвольную точку $x_* \in X_*$. С учетом условий (3) и неравенств (4), (5) имеем $p_{n-1}(x_n) = \min\limits_{x \in [a, b]} p_{n-1}(x) \leqslant$ $\leqslant p_{n-1}(x_{n+1}) \leqslant p_n(x_{n+1}) = \min\limits_{x \in [a, b]} p_n(x) \leqslant p_n(x_*) \leqslant f(x_*) = f_*$, т. е. последовательность $\{p_n(x_{n+1})\}$ монотонно возрастает и ограничена сверху. Отсюда сразу следует оценка (6) и существование предела $\lim\limits_{n \to \infty} p_n(x_{n+1}) = p_* \leqslant f_*$. Покажем, что $p_* = f_*$.

Последовательность $\{x_n\}$ ограничена и по теореме Больцано — Вейерштрасса обладает хотя бы одной предельной точкой. Пусть v_* — какая-либо предельная точка последовательности $\{x_n\}$. Тогда существует подпоследовательность $\{x_{n_k}\}$, сходящаяся к v_*, причем можем считать, что $n_1 < \ldots < n_{k-1} < n_k < \ldots$. Заметим, что $f(x_i) = g(x_i, x_i) \leqslant p_n(x_i) \leqslant f(x_i)$, т. е. $f(x_i) = p_n(x_i)$ при всех $i = 0, 1, \ldots, n$. Тогда $0 \leqslant p_n(x_i) - \min\limits_{x \in [a, b]} p_n(x) = f(x_i) -$ $- p_n(x_{n+1}) = p_n(x_i) - p_n(x_{n+1}) \leqslant L|x_i - x_{n+1}|$ при любом n и $i = 0, 1, \ldots, n$. Принимая здесь $n = n_k - 1$, $i = n_{k-1} \leqslant n_k - 1$, получаем $0 \leqslant f(x_{n_{k-1}}) -$ $- p_{n_k - 1}(x_{n_k}) \leqslant L|x_{n_{k-1}} - x_{n_k}|$ ($k \geqslant 2$). Отсюда при $k \to \infty$ имеем $f_* \leqslant f(v_*) =$ $= \lim\limits_{k \to \infty} f(x_{n_{k-1}}) = \lim\limits_{k \to \infty} p_{n_k - 1}(x_{n_k}) = p_* \leqslant f_*$, т. е. $\lim\limits_{k \to \infty} f(x_{n_k}) = \lim\limits_{k \to \infty} p_{n_k - 1}(x_{n_k}) =$ $= p_* = f_*$. Пользуясь тем, что рассуждения проведены для произвольной предельной точки v_* последовательности $\{x_n\}$, убеждаемся в справедливости первого утверждения теоремы. Второе утверждение следует из теоремы 1.1. □

Таким образом, с помощью метода ломаных можно получить решение задач минимизации первого и второго типов для функций, удовлетворяющих условию (1). Проста и удобна для практического использования формула (6), дающая оценку неизвестной погрешности $f(x_{n+1}) - f_*$ через известные величины, вычисляемые в процессе реализации метода ломаных. Этот метод не требует унимодальности минимизируемой функции, и, более того, функция может иметь сколь угодно точек локального экстремума на рассматриваемом отрезке. На каждом шаге метода ломаных нужно минимизировать кусочно линейную функцию $p_n(x)$, что может быть сделано простым перебором известных вершин ломаной $p_n(x)$, причем здесь перебор существенно упрощается благодаря тому, что ломаная $p_n(x)$ отличается от ломаной $p_{n-1}(x)$ не более чем двумя новыми вершинами. К достоинству метода относится и то, что он сходится при любом выборе начальной точки x_0.

К недостаткам метода ломаных следует отнести то, что с увеличением числа шагов n растет требуемый объем памяти ЭВМ для хранения координат вершин ломаной $p_n(x)$. В § 7 будет рассмотрен другой метод, по своей идее близкий к методу ломаных, но предъявляющий менее жесткие требования к объему памяти и более удобный для реализации на ЭВМ.

Следует также отметить, что метод ломаных невозможно реализовать без знания постоянной L из условия (1). На практике оценку для L получают, вычисляя угловые коэффициенты некоторого числа хорд, соединяющих точки графика минимизируемой функции. Здесь полезно иметь в виду, что если $u < v < w$, то

$$|f(w) - f(u)|/(w - u) \leqslant \max\{|f(w) - f(v)|/(w - v); |f(v) - f(u)|/(v - u)\}, \tag{7}$$

т. е. при добавлении новой точки на отрезке $[u, w]$ появляется новая хорда с неменьшим угловым коэффициентом.

Для доказательства (7) нужно рассмотреть два случая, когда неравенство

$$f(v) \geqslant (f(w) - f(u))(v-u)/(w-u) + f(u) \tag{8}$$

выполняется и когда оно не выполняется. Если $f(w) \geqslant f(u)$ и (8) выполняется, то $(f(v)-f(u))/(v-u) \geqslant (f(w)-f(u))/(w-u) \geqslant 0$; если $f(w) \geqslant f(u)$ и (8) не выполняется, то $(f(w)-f(v))/(w-v) \geqslant (f(w)-f(u))/(w-u) \geqslant 0$. Аналогично доказывается (7) в случае $f(w) < f(u)$.

Пусть $a = v_0 < v_1 < \ldots < v_m = b$; обозначим $L_m = \max\limits_{1 \leqslant i \leqslant m} |f(v_i) - f(v_{i-1})| \cdot |v_i - v_{i-1}|^{-1}$. Ясно, что $L_m \leqslant L$. Пусть при каждом $m \geqslant 1$ величина L_{m+1} вычисляется по точкам $a = w_0 < w_1 < \ldots < w_{m+1} = b$, полученным добавлением к точкам $v_0, v_1, \ldots, v_m$ одной новой точки. Тогда согласно (7) имеем $L_m \leqslant L_{m+1} \leqslant L$ $(m \geqslant 1)$. Это значит, что с возрастанием m величины L_m все лучше и лучше приближают L снизу. Если $\max\limits_{0 \leqslant i \leqslant m} |v_i - v_{i-1}| \to 0$ при $m \to \infty$, то $\lim\limits_{m \to \infty} L_m = L$. Приведенные соображения могут помочь в получении оценки для L. При определении L могут быть полезны теоремы 1,2.

Следует заметить, что использование завышенной оценки для L ухудшает скорость сходимости метода ломаных, приводит к излишне большому количеству вычислений значений минимизируемой функции. Если же пользоваться заниженной оценкой для L, то метод может привести к неправильному определению приближения минимального значения.

Упражнения

1. Привести пример функции, которая удовлетворяет условию (1), но не является унимодальной.

2. Можно ли утверждать, что всякая унимодальная на отрезке $[a, b]$ функция удовлетворяет условию (1) на $[a, b]$? Рассмотреть пример функции $f(x) = \sqrt{x}$ на $[0, 1]$.

3. Рассмотреть первые шесть шагов метода ломаных для функции $f(x) = ||x^2 - 1| - 1|$ на отрезке $[-2, 2]$ при различном выборе начальной точки x_0.

4. Выяснить, как ведет себя метод ломаных при минимизации функции $f(x) \equiv 1$ на отрезке $[0, 1]$.

5. Пусть $f(x) = a_n x^n + \ldots + a_1 x + a_0$ — многочлен n-й степени на отрезке $[a, b]$, где $0 < a < b$. Обозначим

$$A^+ = \{i: 0 \leqslant i \leqslant n, a_i > 0\}, \quad A^- = \{i: 0 \leqslant i \leqslant n, a_i < 0\},$$
$$f_+(x) = \sum_{i \in A^+} a_i x^i, \quad f_-(x) = \sum_{i \in A^-} |a_i| x^i.$$

Доказать, что $f(x) = f_+(x) - f_-(x)$, а в качестве постоянной L из условия (1) для функции $f(x)$ на $[a, b]$ можно взять величину $\max\{|f'_+(b) - f'_-(a)|; |f'_+(a) - f'_-(b)|\}$ [214].

§ 7. Методы покрытий

1. Обозначим через $Q(L)$ класс функций, удовлетворяющих условию Липшица (6.1) на отрезке $[a, b]$ с одной и той же для всех функций этого класса постоянной $L > 0$. Для функций $f = f(x) \in Q(L)$ будем рассматривать задачу минимизации первого типа, когда ищется величина $f_* = \inf\limits_{x \in [a, b]} f(x)$. Для решения этой задачи будем пользоваться методами p_n, которые заключаются в выборе точек $x_1, \ldots, x_n$ $(a \leqslant x_1 < \ldots < x_n \leqslant b)$, вычислении значений функции $f(x_1), \ldots, f(x_n)$ и определении величины $f(x_k) = \min\limits_{1 \leqslant i \leqslant n} f(x_i)$, принимаемой за приближение к f_*.

Возникает вопрос: как выбрать метод $p_n = \{x_1, \ldots, x_n\}$, чтобы

$$\min_{1 \leqslant i \leqslant n} f(x_i) \leqslant f_* + \varepsilon \qquad \forall f(x) \in Q(L), \tag{1}$$

где $\varepsilon > 0$ — заданная точность? Ниже будет изложено несколько методов решения поставленной задачи (1). В каждом из этих методов определенным образом строится некоторая система отрезков, покрывающих исходный отрезок $[a, b]$, и вычисляются значения функции в подходящим образом выбранных точках этих отрезков. Поэтому излагаемые ниже методы принято называть *методами покрытий*.

2. Простейшим методом p_n для решения задачи (1) может служить *метод равномерного перебора*, когда точки $x_1, \ldots, x_n$ выбираются по правилу

$$\begin{aligned} &x_1 = a + h/2, \quad x_2 = x_1 + h, \quad \ldots, \quad x_{i+1} = x_i + h = x_1 + ih, \ldots, \\ &x_{n-1} = x_1 + (n-2)h, \quad x_n = \min\{x_1 + (n-1)h; b\}, \end{aligned} \tag{2}$$

где $h = 2\varepsilon/L$ — шаг метода, а число n определяется условием $x_{n-1} < b - h/2 \leqslant x_1 + (n-1)h$.

Теорема 1. *Метод равномерного перебора* (2) *решает задачу* (1) *на классе* $Q(L)$. *Если* $h > 2\varepsilon/L$, *то существует функция* $f(x) \in Q(L)$, *для которой метод* (2) *не решает задачу* (1).

Доказательство. Пусть $f = f(x)$ — произвольная функция из $Q(L)$. С учетом неравенства (6.2) для любого $x \in [x_i - h/2, x_i + h/2]$ имеем $f(x) \geqslant f(x_i) - L|x - x_i| \geqslant f(x_i) - Lh/2 \geqslant \min\limits_{1 \leqslant i \leqslant n} f(x_i) - \varepsilon$ при всех $i = 1, \ldots, n$. Поскольку система отрезков $[x_i - h/2, x_i + h/2]$ $(i = 1, \ldots, n)$ покрывает весь отрезок $[a, b]$, т. е. всякая точка x из $[a, b]$ принадлежит одному из отрезков этой системы, то из предыдущего неравенства следует, что $f(x) \geqslant \min\limits_{1 \leqslant i \leqslant n} f(x_i) - \varepsilon$ для всех $x \in [a, b]$. Поэтому $f_* \geqslant \min\limits_{1 \leqslant i \leqslant n} f(x_i) - \varepsilon$ для любой функции $f = f(x) \in Q(L)$, что равносильно неравенству (1). Если $h > 2\varepsilon/L$, то, например, для функции $f(x) = Lx$ метод (2) дает $\min\limits_{1 \leqslant i \leqslant n} (Lx_i) - La = Lh/2 > \varepsilon$. □

3. Метод равномерного перебора (2) относится к пассивным методам, когда точки $x_1, \ldots, x_n$ задаются все одновременно до начала вычислений значений функции. На классе $Q(L)$ можно предложить такой же простой, но более эффективный последовательный метод перебора, когда выбор точки x_i при каждом $i > 2$ производится с учетом вычислений значения функции в предыдущих точках $x_1, \ldots, x_{i-1}$, и задачу (1) удается решить, вообще говоря, за меньшее количество вычислений значений функции, чем методом (2). А именно, следуя [286], положим

$$\begin{aligned} &x_1 = a + h/2, \quad x_{i+1} = x_i + h + (f(x_i) - F_i)/L, \quad i = 1, \ldots, n-2, \\ &x_n = \min\{x_{n-1} + h + (f(x_{n-1}) - F_{n-1})/L; b\}, \end{aligned} \tag{3}$$

где $h = 2\varepsilon/L$, $F_i = \min\limits_{1 \leqslant j \leqslant i} f(x_j)$, а число n определяется условием $x_{n-1} < b - h/2 \leqslant x_{n-1} + h + (f(x_{n-1}) - F_{n-1})/L$.

Теорема 2. *Метод последовательного перебора* (3) *решает задачу* (1) *на классе* $Q(L)$.

Доказательство. Пусть $f = f(x)$ — произвольная функция из $Q(L)$. С учетом неравенства (6.2) для всех $x \in [x_i,\ x_i + h/2 + (f(x_i) - F_i)/L]$ имеем $f(x) \geqslant f(x_i) - L\,(h/2 + (f(x_i) - F_i)/L) = F_i - L\,h/2 \geqslant F_n - \varepsilon$. Аналогично для всех $x \in [x_i - h/2,\ x_i]$ получим $f(x) \geqslant f(x_i) - L\,h/2 \geqslant F_n - \varepsilon$. Поскольку система отрезков $[x_i - h/2,\ x_i + h/2 + (f(x_i) - F_i)/L]$ $(i = 1, \ldots, n)$ покрывает весь отрезок $[a, b]$, то из предыдущих неравенств следует, что $f(x) \geqslant F_n - \varepsilon$ при всех $x \in [a, b]$. Тогда $f_* \geqslant F_n - \varepsilon$ при всех $f = f(x) \in Q(L)$, что равносильно неравенству (1). □

В худшем случае, когда, например, функция $f(x)$ постоянна или монотонно убывает на $[a, b]$ и, следовательно, $F_i = \min\limits_{1 \leqslant j \leqslant i} f(x_j) = f(x_i)$, метод (3) превращается в метод (2), и для решения задачи (1) тогда потребуется $N_1 \approx (b - a)L/(2\varepsilon)$ вычислений значений функции. В самом лучшем случае, когда $f(x) = A + L(x - B)$, где A, B — постоянные и, следовательно, $F_i = f(x_i)$, $x_i = x_1 + (2^{i-1} - 1)h$ $(i = 1, \ldots, n - 2)$, для решения задачи (1) понадобится всего $N_2 \approx 1 + \log_2 (b - a)L/(2\varepsilon)$ вычислений значений функции. И вообще, если $f(x_i) > F_i$ при каком-либо i, то $x_{i+1} - x_i > h$, и поэтому число n вычислений значений функции, необходимое для решения задачи (1), будет, вообще говоря, меньше N_1 и больше N_2.

Заметим также, что метод (3) идейно примыкает к методу ломаных из § 6, но метод (3) выгодно отличается простотой реализации и не требует большой машинной памяти. Недостатком метода (3), как и метода ломаных, является необходимость априорного знания постоянной L из условия (6.1).

4. Следуя [590], изложим еще один вариант метода покрытий для решения задачи (1). Пусть зафиксирована сетка точек (2) с шагом $h = 2\varepsilon/L$. Выберем две произвольные точки v_1, v_2 этой сетки, вычислим значения $f(v_1)$, $f(v_2)$ минимизируемой функции и положим $F_1 = f(v_1)$, $F_2 = \min\{F_1;\ f(v_2)\} = \min\{f(v_1), f(v_2)\}$. Имеются две возможности: либо $F_2 = f(v_2) < F_1$, либо $F_2 = F_1 \leqslant f(v_2)$. Если $F_2 < F_1$, то из дальнейших рассмотрений исключаем точку v_1 вместе с теми точками x_j сетки (2), для которых $|x_j - v_1| \leqslant \dfrac{F_1 - F_2}{L}$, не вычисляя значений $f(x_j)$. Если $F_2 = F_1$, то исключаем точку v_2 вместе с точками x_j сетки (2), для которых $|x_j - v_2| \leqslant \dfrac{f(v_2) - F_1}{L}$. Начальный шаг метода описан. Опишем общий шаг. Пусть в точках $v_1, v_2, \ldots, v_k$ сетки (2) уже вычислены значения $f(v_1), f(v_2), \ldots, f(v_k)$, найдена величина $F_k = \min\{F_{k-1};\ f(v_k)\} = \min\limits_{1 \leqslant i \leqslant k} f(v_i)$, и пусть v_{j_k} та из точек $v_1, v_2, \ldots, v_k$, в которой $F_k = f(v_{j_k}) = \min\limits_{1 \leqslant i \leqslant k} f(v_i)$. Далее возьмем любую точку v_{k+1} сетки (2), которая на предыдущих шагах не исключалась и в которой еще не вычислялось значение функции $f(x)$. Вычислим $f(v_{k+1})$ и величину $F_{k+1} = \min\{F_k;\ f(v_{k+1})\} = \min\limits_{1 \leqslant i \leqslant k+1} f(v_i)$. Имеются две возможности: либо $F_{k+1} = f(v_{k+1}) < F_k$, либо $F_{k+1} = F_k \leqslant f(v_{k+1})$. В первом случае, когда $F_{k+1} < F_k$, из дальнейших рассмотрений исключим точку v_{j_k} и вместе с нею те точки x_j сетки (2), для которых

$$|x_j - v_{j_k}| \leqslant \frac{F_k - F_{k+1}}{L}. \tag{4}$$

Заметим, что некоторые из этих точек могли оказаться исключенными уже на предыдущих шагах. Для нас здесь важно лишь то, что среди исключенных точек заведомо нет таких, в которых значение функции $f(x)$ было бы меньше, чем F_{k+1}. В самом деле, прежде всего $f(v_{j_k}) = F_k > F_{k+1}$. Для остальных исключенных точек x_j имеем: $f(x_j) - F_{k+1} = f(x_j) - f(v_{j_k}) + F_k - F_{k+1} \geqslant -L\,|x_j - v_{j_k}| + F_k - F_{k+1} \geqslant 0$ в силу (6.2) и (4). Таким образом, без дополнительных вычислений значений функции $f(x)$ мы сумели выяснить, что исключенные точки не являются перспективными с точки зрения получения в них значений функции, меньших F_{k+1}.

Рассмотрим вторую возможность, когда $F_{k+1} = F_k \leqslant f(v_{k+1})$. Тогда из дальнейшего перебора исключаем точку v_{k+1} вместе с точками x_j сетки (2), для которых

$$|x_j - v_{k+1}| \leqslant \frac{f(v_{k+1}) - F_k}{L}. \tag{5}$$

Нетрудно убедиться, что и в этом случае в исключенных точках значения функции не могут быть меньше F_{k+1}. В самом деле, здесь $f(x_j) - F_{k+1} = f(x_j) - F_k = f(x_j) - f(v_{k+1}) + f(v_{k+1}) - F_k \geqslant -L\,|x_j - v_{k+1}| + f(v_{k+1}) - F_k \geqslant 0$ в силу (6.1) и (5). Общий шаг метода описан.

Так как на каждом шаге метода берется новая точка сетки (2), которая еще не исключена из перебора и в которой значение функции $f(x)$ еще не вычислялось, то ясно, что на каком-то шаге такие точки будут исчерпаны и описанный процесс закончится за N шагов, $N \leqslant n$, перебором точек $v_1, v_2, \ldots, v_N$ сетки (2) и вычислением $F_N = \min\limits_{1 \leqslant i \leqslant N} f(v_i) = \min\limits_{1 \leqslant i \leqslant n} f(x_i)$.

Теорема 3. *Пусть сетка точек* $\{x_1, \ldots, x_n\}$ *определены согласно* (2), *пусть* $f(x)$ *— произвольная функция из класса* $Q(L)$. *Тогда найденная методом последовательного перебора* (4), (5) *величина* $F_N = \min\limits_{1 \leqslant i \leqslant n} F(x_i)$ *решает задачу* (1).

Доказательство. Поскольку система отрезков $[x_i - h/2,\ x_i + h/2]$, $i = 1, \ldots, n$, образует покрытие отрезка $[a, b]$, то для любой точки $x \in [a, b]$ найдется точка x_j сетки (2) такая, что $|x - x_j| \leqslant h/2$. Тогда $f(x) = f(x) - f(x_j) + f(x_j) \geqslant -L\,|x - x_j| + F_N \geqslant -L\,h/2 + F_N = F_N - \varepsilon$ для любого $x \in [a, b]$. Следовательно, $f_* \geqslant F_N - \varepsilon$, т. е. выполняется неравенство (1). Теорема 3 доказана. □

Метод (4), (5), как и метод (3), в худшем случае может превратиться в метод простого перебора точек сетки (2). В то же время ясно, что для многих функций $f(x) \in Q(L)$ этот метод гораздо эффективнее метода простого перебора, так как если величины $F_k - F_{k+1}$, $f(v_{k+1}) - F_k$ в (4), (5) достаточно большие, то многие точки сетки (2) могут оказаться исключенными из перебора без вычисления в них значений функции.

К методу покрытий мы еще вернемся в § 5.13.

Упражнения

1. Пусть одним из вышеописанных методов покрытий найден $\min\limits_{1 \leqslant i \leqslant n} f(x_i) = f(x_k)$. Можно ли принять x_k за приближение к множеству X_*? Оценить погрешность $\rho(x_k, X_*)$ для метода (2) на классе $Q(L)$; рассмотреть функцию $f(x) = L(x - a) - \varepsilon/2$ при $a \leqslant x \leqslant a + \varepsilon/L$, $f(x) = \varepsilon(b - x)/(2(b - a - \varepsilon/L))$ при $a + \varepsilon/L \leqslant x \leqslant b$, где $\varepsilon > 0$ — малое число. Оценить $\rho(x_k, X_*)$ для методов (3) и (4), (5) на классе $Q(L)$.

2. Найти оптимальный пассивный и оптимальный последовательный методы на классе функций $Q(L)$ [218; 671; 755].

§ 8. Выпуклые функции одной переменной

Рассмотрим класс функций, для которых существует более эффективный вариант метода ломаных, когда ломаные составляются из отрезков касательных и лучше аппроксимируют минимизируемую функцию. Речь идет о выпуклых функциях, играющих важную роль в теории экстремальных задач.

Определение 1. Функция $f(x)$, определенная на отрезке $[a, b]$, называется *выпуклой* на этом отрезке, если

$$f(\alpha u + (1-\alpha)v) \leqslant \alpha f(u) + (1-\alpha)f(v) \qquad (1)$$

при всех $u, v \in [a, b]$, $\alpha \in [0, 1]$.

Когда α пробегает отрезок $[0, 1]$, точки $(\alpha u+(1-\alpha)v, \alpha f(u)+(1-\alpha)f(v))$ на плоскости переменных (x, f) пробегают хорду AB, соединяющую точки $A = (u, f(u))$ и $B = (v, f(v))$ на графике функции $f = f(x)$. Поэтому неравенство (1) имеет простой геометрический смысл: график выпуклой функции на любом отрезке $[u, v] \subseteq [a, b]$ находится не выше хорды, соединяющей точки графика $(u, f(u))$ и $(v, f(v))$ (рис. 1.3). Примерами функций, выпуклых на любом отрезке, могут служить функции $f(x) = x^2$, $f(x) = |x|$, $f(x) = x$. Наряду с выпуклыми функциями в литературе рассматривают вогнутые функции.

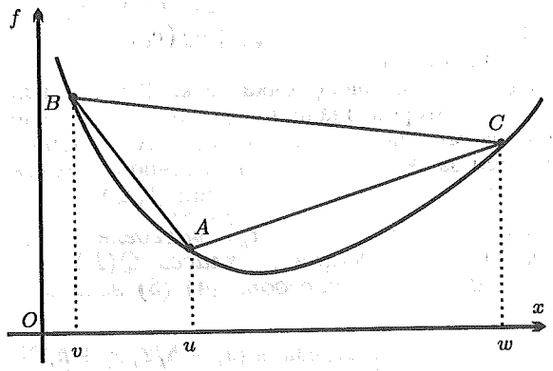

Рис. 1.3

Определение 2. Функция $f(x)$ называется *вогнутой* на отрезке $[a, b]$, если

$$f(\alpha u + (1-\alpha)v) \geqslant \alpha f(u) + (1-\alpha)f(v)$$

при всех $u, v \in [a, b]$, $\alpha \in [0, 1]$.

Между выпуклыми и вогнутыми функциями существует простая связь: если $f(x)$ вогнута на $[a, b]$, то $-f(x)$ выпукла на этом же отрезке. Учитывая эту связь, достаточно ограничиться изучением свойств выпуклых функций.

Теорема 1. *Для выпуклости функции $f(x)$ на отрезке $[a, b]$ необходимо и достаточно, чтобы*

$$(f(u)-f(v))/(u-v) \leqslant (f(w)-f(v))/(w-v) \leqslant (f(w)-f(u))/(w-u) \qquad (2)$$

при всех u, v, w, $a \leqslant v < u < w \leqslant b$.

Доказательство. Необходимость. Пусть функция $f(x)$ выпукла на $[a, b]$. Нетрудно проверить, что $u = \alpha v + (1-\alpha)w$, где $\alpha = (w-u)/(w-v)$ $(0 < \alpha < 1)$. Отсюда с учетом выпуклости функции $f(x)$ имеем $f(u) \leqslant (w-u)f(v)/(w-v) + (1-(w-u)/(w-v))f(w)$, или

$$(w-v)f(u) \leqslant (w-u)f(v) + (u-v)f(w).$$

Последнее неравенство можно переписать в двоякой форме:

$$(w-v)(f(u)-f(v)) \leqslant (u-v)(f(w)-f(v)),$$

или

$$(w-u)(f(w)-f(v)) \leqslant (w-v)(f(w)-f(u)),$$

откуда будет следовать (2).

Достаточность. Пусть $f(x)$ удовлетворяет одному из неравенств (2). Отправляясь от этого неравенства и проделав предыдущие преобразования в обратном порядке, убеждаемся в том, что $f(x)$ выпукла на отрезке $[a, b]$. □

Нетрудно понять геометрический смысл неравенств (2) (см. рис. 1.3), если вспомнить, что $(f(u)-f(v))/(u-v)$ представляет собой угловой коэффициент хорды AB, соединяющей точки $A = (u, f(u))$ и $B = (v, f(v))$ на графике функции $f = f(x)$.

Теорема 2. *Выпуклая на отрезке $[a, b]$ функция $f(x)$ в каждой внутренней точке u отрезка $[a, b]$ непрерывна и имеет конечную правую производную $\lim\limits_{h \to +0} (f(u+h)-f(u))/h = f'(u+0)$, конечную левую производную $\lim\limits_{\tau \to +0} (f(u)-f(u-\tau))/\tau = f'(u-0)$, причем $f'(u-0) \leqslant f'(u+0)$ при всех $u \in (a, b)$.*

Доказательство. Из теоремы 1 следует, что

$$(f(u)-f(u-\tau))/\tau \leqslant (f(u)-f(u-h))/h \leqslant$$
$$\leqslant (f(u+h)-f(u))/h \leqslant (f(u+\tau)-f(u))/\tau \qquad (3)$$

при всех τ, h, лишь бы $0 < h < \tau$ и точки $u, u \pm h, u \pm \tau \in (a, b)$ (рис. 1.4). Неравенства (3) означают, что величина $(f(u+h)-f(u))/h$ монотонно убывает при убывании h и ограничена снизу, например, величиной $(f(u)-f(u+\tau))/\tau$, не зависящей от h. Отсюда следует существование правой производной $f'(u+0)$. Аналогично доказывается существование левой производной $f'(u-0)$. Из (3) при $h \to +0$ получаем неравенство $f'(u-0) \leqslant f'(u+0)$. Из существования левой и правой производных следует непрерывность функции $f(x)$ при всех значениях $x \in (a, b)$. □

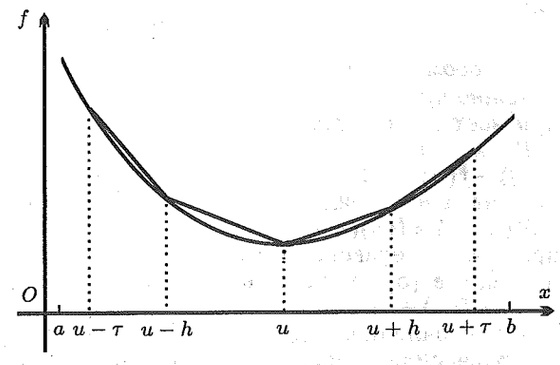

Рис. 1.4

Заметим, что на концах отрезка $[a, b]$ выпуклая функция может не иметь соответствующей односторонней производной и, более того, здесь она может терпеть разрыв.

Пример 1. Пусть $f(x) = x$ при $0 < x < 1$, $f(0) = f(1) = 2$. Очевидно, эта функция выпукла на $[0, 1]$, но на концах отрезка терпит разрывы.

Пример 2. Функция $f(x) = -\sqrt{1-x^2}$ выпукла и непрерывна на отрезке $[-1, 1]$, но на концах отрезка не имеет конечных производных $f'(1-0)$, $f(-1+0)$.

Теорема 3. *Пусть функция $f(x)$ выпукла на отрезке $[a, b]$ и имеет конечные производные $f'(a+0)$, $f'(b-0)$. Тогда*

$$f'(a+0)(u-v) \leqslant f(u)-f(v) \leqslant f'(b-0)(u-v) \tag{4}$$

при всех u, v $(a \leqslant v \leqslant u \leqslant b)$, так что $f(x)$ на $[a, b]$ удовлетворяет условию Липшица (6.1) с постоянной $L = \max\{|f'(a+0)|; |f'(b-0)|\}$.

Доказательство. Из теоремы 1 имеем

$$(f(a+h)-f(a))/h \leqslant (f(v)-f(a))/(v-a) \leqslant$$
$$\leqslant (f(u)-f(v))/(u-v) \leqslant (f(b)-f(u))/(b-u) \leqslant (f(b)-f(b-h))/h$$

для всех $h>0$, $a+h<v<u<b-h$. Отсюда при $h \to +0$ получаем

$$f'(a+0) \leqslant (f(u)-f(v))/(u-v) \leqslant f'(b-0),$$

что равносильно (4) при любых u, v $(a<v<u<b)$. Неравенства (4) остаются верными также и при $v=a$ или $u=b$, так как при выполнении условий теоремы функция $f(x)$ непрерывна во всех точках отрезка $[a, b]$ и в (4) можно совершить предельный переход при $v \to a+0$ или $u \to b-0$. □

Пример 2 показывает, что конечность величин $f'(a+0)$, $f'(b-0)$ существенна для выполнения условия Липшица (6.1).

Теорема 4. *Пусть функция $f(x)$ выпукла на отрезке $[a, b]$, а $l(v)$ — любая функция, удовлетворяющая неравенствам $f'(v-0) \leqslant l(v) \leqslant f'(v+0)$ при $a<v<b$. Тогда $l(v)$ не убывает при $v \in (a, b)$ и справедливо неравенство*

$$f(u) \geqslant f(v)+l(v)(u-v), \qquad u \in [a, b]. \tag{5}$$

Если, кроме того, $f(x)$ дифференцируема во всех точках отрезка $[a, b]$, то

$$f(u) \geqslant f(v)+f'(v)(u-v), \qquad u \in [a, b], \tag{6}$$

при любом $v \in [a, b]$. Если неравенство (5) (или (6)) обращается в равенство при некотором $u=c \in [a, b]$ $(c \neq v)$, то $f(u) \equiv f(v)+l(v)(u-v)$ при всех u из отрезка $[c, v]$.

Доказательство. Перепишем неравенство (1) в виде $f(v+\alpha(u-v))-f(v) \leqslant \alpha(f(u)-f(v))$ $(0<\alpha<1)$. Разделив обе части этого неравенства на α и перейдя к пределу при $\alpha \to +0$, получим $f(u)-f(v) \geqslant f'(v+0)(u-v) \geqslant l(v)(u-v)$ при $u>v$ и $f(u)-f(v) \geqslant f'(v-0)(u-v) \geqslant l(v)(u-v)$ при $u<v$. Неравенство (5) доказано. Заметим, что при $a<u<b$ переменные u, v в (5) входят равноправно, поэтому, меняя их ролями, получаем $f(v) \geqslant f(u)+l(u)(v-u)$ при всех $v \in [a, b]$. Сложим последнее неравенство с (5) почленно. Будем иметь $(l(u)-l(v))(u-v) \geqslant 0$ при всех $u, v \in (a, b)$, что равносильно монотонному возрастанию $l(v)$.

Пусть теперь $f(x)$ дифференцируема во всех точках $x \in [a, b]$. Тогда $f'(u+0)=f'(u-0)=f'(u)$ при всех $u \in [a, b]$. Полагая в (5) $l(v)=f'(v)$, убеждаемся в справедливости неравенства (6) при всех $u \in (a, b)$. Из существования конечных функций $f'(a+0)$, $f'(b-0)$ и из (4) следует, что (6) верно и при $v=a$, $v=b$.

Наконец, пусть $f(c)=f(v)=l(v)(c-v)$ при некотором $c \in [a, b]$ $(c \neq v)$. Возьмем произвольную точку $u=\alpha c+(1-\alpha)v$ из отрезка $[c, v]$. Из выпуклости $f(x)$ тогда следует, что $f(u) \leqslant \alpha f(c)+(1-\alpha)f(v)=\alpha(f(v)+l(v)(c-v))+(1-\alpha)f(v)=f(v)+l(v)(u-v)$ $(u \in [c, v])$. Сравнивая это неравенство с (5), заключаем, что $f(u)=f(v)+l(v)(u-v)$ при всех $u \in [c, v]$. □

График линейной функции $f(v)+f'(v)(u-v)$ переменной $u \in [a, b]$ представляет собой касательную к графику $f=f(x)$ в точке v. Поэтому неравенство (6) означает, что график любой выпуклой дифференцируемой функции лежит не ниже любой касательной к нему. Обобщая понятие касательной на случай выпуклых недифференцируемых функций, прямую $g(u, v)=f(v)+l(v)(u-v)$, где $f'(v-0) \leqslant l(v) \leqslant f'(v+0)$, также будем называть *касательной* к графику $f=f(x)$ в точке v.

Следствие 1. *Пусть функция $f(x)$ выпукла на $[a, b]$. Тогда производные $f'(u+0)$, $f'(u-0)$ монотонно возрастают при $u \in (a, b)$ (если существуют конечные $f'(a+0)$, $f'(b-0)$, то утверждение справедливо на всем отрезке $[a, b]$).*

Доказательство этого утверждения непосредственно следует из теоремы 4, если в ней принять $l(v)=f'(v+0)$ или $l(v)=f'(v-0)$.

Теорема 5. *Пусть функция $f(x)$ выпукла на отрезке $[a, b]$ и $\lim\limits_{x \to a+0} f(x)=f(a)$, $\lim\limits_{x \to b-0} f(x)=f(b)$. Тогда множество X_* точек ее глобального минимума на $[a, b]$ непусто и все точки локального минимума $f(x)$ принадлежат X_*. Для того чтобы $x_* \in X_*$, необходимо и достаточно, чтобы*

$$f'(x_*+0) \geqslant 0, \quad f'(x_*-0) \leqslant 0 \tag{7}$$

(если $x_*=a$ или $x_*=b$, то (7) заменяется одним неравенством $f'(a+0) \geqslant 0$ или $f'(b-0) \leqslant 0$ соответственно).

Доказательство. Из условий на функцию $f(x)$ и теоремы 2 следует непрерывность $f(x)$ на $[a, b]$. Согласно теореме 1.1 тогда множество X_* непусто. Пусть x_* — какая-либо точка локального минимума $f(x)$. Тогда $f(x_*+h)-f(x_*) \geqslant 0$ при всех достаточно малых $|h|$, для которых $x_*+h \in [a, b]$. Разделив это неравенство на $h>0$ и $h<0$ и устремив h к нулю, получим условие (7). Заметим, что существование и конечность $f'(x_* \pm 0)$ при $a<x_*<b$ следует из теоремы 2. Если $x_*=a$, то существование и конечность $f'(a+0)$ следует из того, что $(f(a+h)-f(a))/h$ монотонно убывает при $h \to +0$ и ограничена снизу нулем. Аналогично доказывается существование и конечность $f'(b-0)$ при $x_*=b$. Таким образом, показано, что всякая точка локального минимума необходимо удовлетворяет условиям (7).

Пусть теперь некоторая точка $x_* \in (a, b)$ удовлетворяет условию (7). Положим в неравенстве (5) $v=x_*$, $l(v)=0$ и получим, что $f(u) \geqslant f(x_*)$ при всех $u \in [a, b]$. Это значит, что $x_* \in X_*$. Аналогично с использованием неравенств (4) рассматриваются случаи $x_*=a$ или $x_*=b$ и показывается, что $x_* \in X_*$. Отсюда следует, что всякая точка локального минимума выпуклой и непрерывной на $[a, b]$ функции является точкой ее глобального минимума на $[a, b]$. □

Теорема 6. *Пусть функция $f(x)$ выпукла на отрезке $[a, b]$ и $\lim\limits_{u \to a+0} f(u)=f(a)$, $\lim\limits_{u \to b-0} f(u)=f(b)$; пусть X_* — множество точек минимума $f(x)$ на $[a, b]$ и v — некоторая точка $(a<v<b)$. Тогда для того чтобы $X_* \cap [a, v]=\varnothing$ $(X_* \cap [v, b]=\varnothing)$, необходимо и достаточно выполнения неравенства $f'(v+0)<0$ $(f'(v-0)>0)$. Для того чтобы $X_* \cap [a, v] \neq \varnothing$ $(X_* \cap [v, b] \neq \varnothing)$, необходимо и достаточно, чтобы $f'(v+0) \geqslant 0$ $(f'(v-0) \leqslant 0)$.*

Доказательство. Достаточность. Пусть $f'(v+0)<0$. Тогда согласно следствию 1 $f'(u+0)<0$ при всех $u \in [a, v]$. Из теоремы 5 тогда имеем $X_* \cap [a, v]=\varnothing$. Если $f'(v-0)>0$, то аналогично получаем $f'(u-0)>0$ при всех $u \in [v, b]$, так что $X_* \cap [v, b]=\varnothing$.

Необходимость. Пусть $X_* \cap [a, v] = \varnothing$. Допустим, что $f'(v+0) \geqslant 0$. Тогда возможно, что $f'(v-0) \leqslant 0$ или $f'(v-0) > 0$. Если $f'(v-0) \leqslant 0$, то из (7) следует, что $v \in X_*$. Если же $f'(v-0) > 0$, то по доказанному выше $X_* \cap [v, b] = \varnothing$ и, следовательно, $X_* \cap [a, v] \neq \varnothing$. В обоих случаях приходим к противоречию с тем, что $X_* \cap [a, v] = \varnothing$. Это значит, что при $X_* \cap [a, v] = \varnothing$ необходимо, чтобы $f'(v+0) < 0$. Аналогично доказывается, что если $X_* \cap [v, b] = \varnothing$, то необходимо, чтобы $f'(v-0) > 0$.

В справедливости последнего утверждения теоремы 6 легко убедиться рассуждением от противного со ссылкой на уже доказанное первое утверждение. □

Теорема 7. *Если функция $f(x)$ выпукла на отрезке $[a, b]$ и $\lim\limits_{u \to a+0} f(u) = f(a)$, $\lim\limits_{u \to b-0} f(u) = f(b)$, то она унимодальна на $[a, b]$.*

Доказательство. Обозначим $u_* = \inf X_*$, $v_* = \sup X_*$. Из непрерывности $f(x)$ на $[a, b]$ и определения верхней и нижней грани множества X_* следует, что $u_*, v_* \in X_*$. Если $u_* = v_*$, то X_* состоит из одной точки u_*. Если $u_* < v_*$, то с учетом выпуклости $f(x)$ имеем $f_* = \inf\limits_{u \in [a, b]} f(u) \leqslant f(\alpha u_* + (1-\alpha)v_*) \leqslant \alpha f(u_*) + (1-\alpha) f(v_*) = f_*$. Это значит, что $f(\alpha u_* + (1-\alpha)v_*) = f_*$ при всех $\alpha \in [0, 1]$, т. е. $X_* = [u_*, v_*]$.

Далее, так как $X_* \cap [a, v] = \varnothing$ для любого v $(a \leqslant v < u_*)$, то по теореме 6 имеем $f'(v+0) < 0$ при $a \leqslant v < u_*$. А тогда $f'(v+0) \leqslant (f(v+h) - f(v))/h < 0$ при всех достаточно малых h, т. е. $f(x)$ строго монотонно убывает при $a \leqslant x \leqslant u_*$. Аналогично показывается, что при $v_* \leqslant x \leqslant b$ функция $f(x)$ строго монотонно возрастает. □

Как показывает пример 1, при нарушении условий теоремы 7 множество X_* может быть пустым. Приведем еще несколько примеров.

Пример 3. Функция $f(x) = x^2$ выпукла на отрезке $[-1, 1]$ и множество X_* состоит из единственной точки $x_* = 0$.

Пример 4. Функция $f(x) = |x| + |x-1|$ выпукла на отрезке $[-1, 2]$ и множество X_* представляет собой отрезок $X_* = [0, 1]$.

Пример 5. Пусть $f(x) = 0$ при $0 < x \leqslant 1$, $f(0) = 1$. Функция $f(x)$ выпукла на $[0, 1]$, но множество $X_* = \{x: 0 < x \leqslant 1\}$ не является отрезком. Здесь $\lim\limits_{u \to +0} f(u) \neq f(0)$ — нарушено одно из условий теоремы 7.

Критерий выпуклости функций, приведенный в теореме 1, не очень удобен для практической проверки. Приведем другие, часто более удобные критерии выпуклости функций.

Теорема 8. *Для того чтобы дифференцируемая функция $f(x)$ на отрезке $[a, b]$ была выпуклой, необходимо и достаточно, чтобы ее производная $f'(x)$ не убывала на $[a, b]$.*

Доказательство. Необходимость доказана в теореме 4, так как в рассматриваемом случае $l(v) = f'(v)$ $(v \in [a, b])$.

Достаточность Пусть $f'(x)$ не убывает на $[a, b]$. Пусть $a \leqslant v < u < w \leqslant b$. Применяя формулу Лагранжа, имеем

$$(f(u) - f(v))/(u - v) = f'(\xi_1), \quad v < \xi_1 < u,$$
$$(f(w) - f(u))/(w - u) = f'(\xi_2), \quad u < \xi_2 < w.$$

По условию $f'(\xi_1) \leqslant f'(\xi_2)$, поэтому из предыдущих равенств следует одно из неравенств (2), что согласно теореме 1 равносильно выпуклости $f(x)$ на $[a, b]$. □

Теорема 9. *Для того чтобы дважды дифференцируемая функция $f(x)$ на отрезке $[a, b]$ была выпуклой, необходимо и достаточно, чтобы $f''(x) \geqslant 0$ на $[a, b]$.*

Доказательство. Условие $f''(x) \geqslant 0$ является необходимым и достаточным для неубывания $f'(x)$ на $[a, b]$. Отсюда и из теоремы 8 следует требуемое. □

Используя теоремы 8, 9, легко проверить, что функции $f(x) = a^x$, $f(x) = -\ln x$, $f(x) = x \ln x$ выпуклы на любом отрезке из области своего определения; функции $f(x) = x^r$ при $r \leqslant 1$, $r \leqslant 0$ и $f(x) = -x^r$ при $0 < r < 1$ выпуклы на любых отрезках $[a, b]$ $(0 < a < b < \infty)$. Функция $f(x) = \sin x$ выпукла на отрезке $[-\pi, 0]$, но невыпукла на $[-\pi, \pi]$.

Упражнения

1. Доказать, что если функция $f(x)$ выпукла на отрезке $[a, b]$, то $f'(x+0) = \inf\limits_{h>0} (f(x+h) - f(x))/h$, $f'(x-0) = \sup\limits_{h>0} (f(x) - f(x-h))/h$ при всех $x \in [a, b]$.

2. Пусть функция $f(x)$ выпукла на отрезке $[a, b]$. Доказать, что тогда $f'(x+0)$ непрерывна слева, а $f'(x-0)$ непрерывна справа при всех x $(a < x < b)$. Указание: воспользоваться непрерывностью $f(x)$, следствием 1 и упражнением 1.

3. Пусть $f(x)$ выпукла на $[a, b]$. Доказать, что $f'(v-0) \leqslant f'(v+0) \leqslant f'(u-0) \leqslant f'(u+0)$ при всех u, v $(a < v < u < b)$. Пользуясь этими неравенствами, показать, что $f(x)$ дифференцируема в точке v $(a < v < b)$ тогда и только тогда, когда одна из функций $f'(x+0)$ или $f'(x-0)$ непрерывна в точке v.

4. Пусть $f(x)$ выпукла на $[a, b]$. Пользуясь упражнениями 2, 3, доказать, что множества точек непрерывности функции $f'(x+0)$ и $f'(x-0)$ совпадают. Вывести отсюда, что множество точек, в которых $f(x)$ недифференцируема, не более чем счетно.

5. Пусть функция $f(x)$ непрерывна на отрезке $[a, b]$, дифференцируема на отрезках $[a, a_1], [a_1, a_2], \ldots, [a_{n-1}, a_n], [a_n, b]$ $(a < a_1 < \ldots < a_n < b)$, причем на каждом таком отрезке производная $f'(x)$ суммируема, не убывает и $f'(a_i - 0) \leqslant f'(a_i + 0)$ $(i = 1, \ldots, n)$. Доказать, что тогда $f(x)$ выпукла на $[a, b]$.

6. Для выпуклости функции $f(x)$ на интервале (a, b) необходимо и достаточно, чтобы существовала функция $l(v)$ $(v \in (a, b))$ такая, что $f(x) \geqslant f(v) + l(v)(x - v)$ при всех $x \in (a, b)$. Необходимость доказана в теореме 4, докажите достаточность. Покажите, что $l(v) = f'(v)$ почти всюду на (a, b).

7. Пользуясь теоремой 3, доказать, что выпуклая на отрезке $[a, b]$ функция $f(x)$ абсолютно непрерывна на любом отрезке $[\alpha, \beta] \subset (a, b)$.

8. Если функция $g(t)$ возрастает на отрезке $[a, b]$ и суммируема на этом отрезке, то функция $f(x) = \int\limits_a^x g(t)\,dt$ выпукла на $[a, b]$. Доказать. Верно ли обратное утверждение?

9. Пусть $f(x)$ выпукла на $[a, b]$ и имеет обратную функцию. Можно ли утверждать, что обратная функция также будет выпуклой? Рассмотреть функции $f(x) = e^x$, $f(x) = e^{-x}$.

10. Пусть $f(x)$ выпукла на $[a, b]$ и $\lim\limits_{u \to a+0} f(u) = f(a)$, $\lim\limits_{u \to b-0} f(u) = f(b)$, а X_* — множество точек минимума $f(x)$ на $[a, b]$. Доказать, что $X_* \cap [\alpha, \beta] \neq \varnothing$, $(X_* \subset \operatorname{int} [\alpha, \beta])$ тогда и только тогда, когда $f'(\alpha - 0) \leqslant 0$, $f'(\beta + 0) \geqslant 0$, $(f'(\alpha - 0) < 0$, $f'(\beta + 0) > 0)$; здесь $a < \alpha < \beta < b$.

11. Доказать, что выпуклая ни отрезке $[a, b]$ функция $f(x)$, отличная от постоянной, не может достигать своей верхней грани внутри отрезка $[a, b]$.

12. Пусть функция $f(u)$ выпукла и монотонно возрастает на отрезке $[a, b]$, а функция $z(x)$ выпукла на $[c, d]$, причем $z(x) \in [a, b]$ при всех $x \in [c, d]$. Доказать, что тогда сложная функция $g(x) = f(z(x))$ выпукла на $[c, d]$.

13. Назовем функцию $f(x)$ выпуклой на отрезке $[a, b]$, если $f((u+v)/2) \leqslant (f(u)+f(v))/2$ при всех $u, v \in [a, b]$. Доказать, что для непрерывных функций это определение выпуклой функции равносильно определению 1. Если не требовать непрерывности $f(x)$, то новое определение выделяет более широкий класс функций — см. пример из [735, стр.119].

14. Пусть $f(x)$ — выпуклая функция при $x \geqslant 0$, $f(0) \leqslant 0$. Доказать, что тогда функция $\varphi(x) = f(x)/x$ монотонно возрастает при $x > 0$. На примере функции $f(x) = 1 + x^2$ убедиться, что при $f(0) > 0$ это утверждение неверно. Указание: воспользоваться равенством $f(u) = f\left(\frac{u}{u+h}(u+h) + \frac{h}{u+h} \cdot 0\right)$ $(h > 0)$.

15. Пусть функция $f(x)$ выпукла и дважды дифференцируема при $x \geqslant 0$, причем $\lim\limits_{x \to \infty} (xf'(x) - f(x)) \leqslant 0$. Доказать, что тогда $\varphi(x) = f(x)/x$ монотонно убывает при $x > 0$. Указание: вычислить производные функций $\varphi(x)$ и $xf'(x) - f(x)$.

16. Доказать, что $(a+b)^n \leqslant 2^{n-1}(a^n + b^n)$ при всех $n \geqslant 1$, $a \geqslant 0$, $b \geqslant 0$. Указание: воспользоваться выпуклостью функции $f(x) = x^n$ при $x \geqslant 0$, $n \geqslant 1$.

§ 9. Метод касательных

1. Пусть функция $f(x)$ выпукла и дифференцируема на отрезке $[a, b]$. Согласно теоремам 8.3, 8.7 такая функция удовлетворяет условию Липшица и унимодальна на $[a, b]$. Поэтому для минимизации $f(x)$ на $[a, b]$ применимы почти все описанные выше методы, в частности, метод ломаных из § 6. Однако, если значения функции $f(x)$ и ее производной $f'(x)$ вычисляются достаточно просто, то здесь можно предложить другой, вообще говоря, более эффективный вариант метода ломаных, когда в качестве звеньев ломаных берутся отрезки касательных к графику $f(x)$ в соответствующих точках.

Зафиксируем какую-либо точку $v \in [a, b]$ и определим функцию $g(x, v) = f(v) + f'(v)(x - v)$, $a \leqslant x \leqslant b$. Согласно теореме 8.4

$$g(x, v) \leqslant f(x) \qquad \forall x \in [a, b] \tag{1}$$

В качестве начального приближения возьмем любую точку $x_0 \in [a, b]$ (например, $x_0 = a$), составим функцию $p_0(x) = g(x, x_0)$ и определим точку $u_1 \in [a, b]$ из условия $p_0(x_1) = \min\limits_{u \in [a, b]} p_0(u)$ (ясно, что при $f'(x_0) \neq 0$ будет $x_1 = a$ или $x_1 = b$).

Далее, берем новую функцию $p_1(x) = \max\{p_0(x); g(x, x_1)\}$ и следующую точку $x_2 \in [a, b]$ найдем из условия $p_1(x_2) = \min\limits_{u \in [a, b]} p_1(u)$, и т. д. Если точки $x_0, x_1, \ldots, x_n$ $(n \geqslant 1)$ уже известны, то составляем функцию $p_n(x) = \max\{p_{n-1}(x); g(x, x_n)\} = \max\limits_{0 \leqslant i \leqslant n} g(x, x_i)$ и следующую точку x_{n+1} определяем из условий $p_n(x_{n+1}) = \min\limits_{u \in [a, b]} p_n(u)$ $(x_{n+1} \in [a, b])$. Если при каком-либо $n \geqslant 0$ окажется, что $f'(x_n + 0) \geqslant 0$, $f'(x_n - 0) \leqslant 0$ (если $a < x_n < b$, то это равносильно условию $f'(x_n) = 0$), то согласно теореме 8.5 $x_n \in X_*$ — в этом случае задача минимизации уже решена и итерации на этом заканчиваются.

Нетрудно видеть, что $p_n(x)$ — непрерывная кусочно линейная функция и ее график представляет собой ломаную, состоящую из отрезков касательных к графику функции $f(x)$ в точках $x_0, x_1, \ldots, x_n$ (рис. 1.5). Поэтому описанный метод естественно назвать *методом касательных*.

Теорема 1. *Пусть функция $f(x)$ на отрезке $[a, b]$ выпукла и дифференцируема, а последовательность $\{x_n\}$ получена описанным выше методом касательных, причем $x_n \notin X_*$ $(n = 0, 1, \ldots)$. Тогда:*

1) $\lim\limits_{n \to \infty} f(x_n) = \lim\limits_{n \to \infty} p_n(x_{n+1}) = f_*$ *и справедлива оценка*

$$0 \leqslant f(x_{n+1}) - f_* \leqslant f(x_{n+1}) - p_n(x_{n+1}), \qquad n = 1, 2, \ldots;$$

2) $\lim\limits_{n \to \infty} \rho(x_n, X_*) = 0$, *или точнее,* $\{x_n\}$ *имеет не более двух предельных точек, совпадающих с* $u_* = \inf X_*$ *или* $v_* = \sup X_*$.

Доказательство. Поскольку величины $f'(a+0)$, $f'(b-0)$ конечны по условию, то в силу теоремы 8.3 функция $f(x)$ удовлетворяет условию Липшица с постоянной $L = \max\{|f'(a)|, |f'(b)|\}$. Кроме того, согласно (1) и определению функции $p_n(x)$ имеем

$$p_{n-1}(x) \leqslant p_n(x) \leqslant f(x), \qquad x \in [a, b], \quad n = 1, 2, \ldots \tag{2}$$

Тогда $f(x_i) = g(x_i, x_i) \leqslant p_n(x_i) \leqslant f(x_i)$, т. е.

$$f(x_i) = p_n(x_i), \qquad i = 0, 1, \ldots, n. \tag{3}$$

Наконец, угловые коэффициенты касательных $g(x, x_i)$ равны $f'(x_i)$, причем $|f'(x_i)| \leqslant L$. Из теоремы 6.1 тогда следует, что $p_n(x)$ удовлетворяет условию Липшица с постоянной L. Отсюда с учетом (2), (3) с помощью тех же рассуждений, которые применялись при доказательстве теоремы 6.3, нетрудно убедиться в справедливости всех утверждений доказываемой теоремы. Остается лишь заметить, что из того, что функция $f(x)$ унимодальна и $x_n \notin X_* = [u_*, v_*]$ $(n \geqslant 0)$, равенство $\lim\limits_{n \to \infty} \rho(x_n, X_*) = 0$ возможно только в том случае, если предельными для $\{x_n\}$ будут лишь точки u_* или v_*. □

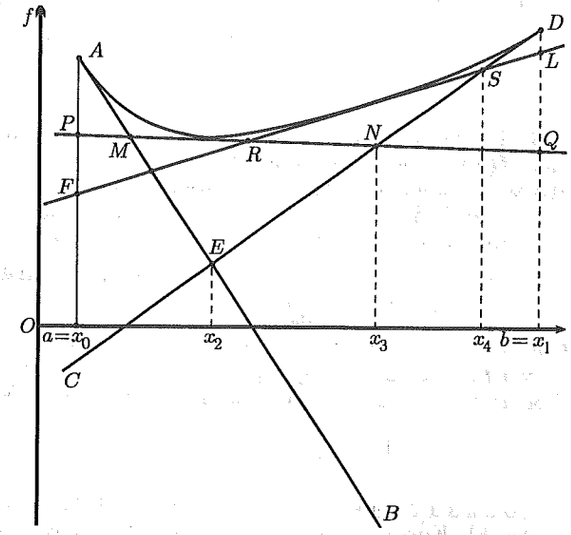

Рис. 1.5. AB — график $g(x, x_0)$, CD — график $g(x, x_1)$, AED — график $p_1(x)$, PQ — график $g(x, x_2)$, $AMND$ — график $p_2(x)$, FL — график $g(x, x_3)$, $AMRSD$ — график $p_3(x)$

2. Метод касательных обладает всеми достоинствами метода ломаных из § 6. Недостаток этого метода: он применим лишь в случае, когда минимизируемая функция выпукла и значения функции и ее производных вычисляются достаточно просто.

Можно предложить более удобную для использования на ЭВМ вычислительную схему метода касательных, которая не требует хранения в машинной памяти информации обо всей ломаной $p_n(x)$ при $x \in [a, b]$. А именно, возьмем $a_1 = a$, $b_1 = b$, вычислим $f'(a_1) = f'(a+0)$, $f'(b_1) = f'(b-0)$. Если $f'(a_1) \geqslant 0$ или $f'(b_1) \leqslant 0$, то по теореме 8.5 $a \in X_*$ или $b \in X_*$ — задача решена.

Поэтому, пусть $f'(a_1) < 0$, $f'(b_1) > 0$, что согласно теореме 8.6 означает $X_* \subset (a, b)$. Пусть отрезок $[a_{n-1}, b_{n-1}]$ $(n \geqslant 2)$ уже построен, причем $f'(a_{n-1}) < 0$, $f'(b_{n-1}) > 0$, $X_* \subset (a_{n-1}, b_{n-1})$. Обозначим через x_n точку пересечения касательных $g(x, a_{n-1})$ и $g(x, b_{n-1})$. Ясно, что $a_{n-1} < x_n < b_{n-1}$. Вычислим $f'(x_n)$. Если $f'(x_n) = 0$, то $x_n \in X_*$ — задача решена, итерации на

этом заканчиваются. Если $f'(x_n) \neq 0$, то положим

$$a_n = \begin{cases} a_{n-1}, & f'(x_n) > 0, \\ x_n, & f'(x_n) < 0, \end{cases} \qquad b_n = \begin{cases} x_n, & f'(x_n) > 0, \\ b_{n-1}, & f'(x_n) < 0. \end{cases} \tag{4}$$

По построению $f'(a_n) < 0$, $f'(b_n) > 0$, и согласно теореме 8.6 $X_* \subset (a, b)$. Индуктивное описание метода закончено.

Из геометрических построений нетрудно усмотреть (см. рис. 1.5), что этот метод совпадает с описанным выше методом касательных, в котором за начальную точку берется $x_0 = a$. В то же время приведенная схема метода более проста и удобна для реализации на ЭВМ, на каждом шаге метода здесь достаточно хранить в памяти ЭВМ величины a_n, b_n, $f(a_n)$, $f(b_n)$, $f'(a_n)$, $f'(b_n)$. Нетрудно выписать явное выражение для точки x_{n+1}, определяемой условием $g(x, a_n) = g(x, b_n)$ пересечения касательных в точках a_n, b_n при $f'(a_n) < 0$, $f'(b_n) > 0$:

$$x_{n+1} = \frac{f(a_n) - f(b_n) + b_n f'(b_n) - a_n f'(a_n)}{f'(b_n) - f'(a_n)}, \qquad n \geqslant 1. \tag{5}$$

3. Поскольку ломаная из отрезков касательных аппроксимирует функцию $f(x)$, вообще говоря, лучше, чем ломаные из § 6, то следует ожидать, что метод касательных для выпуклых функций сходится быстрее метода ломаных из § 6. Исследуем скорость сходимости метода касательных, считая минимизируемую функцию дважды дифференцируемой.

Теорема 2. *Пусть функция $f(x)$ дважды непрерывно дифференцируема на $[a, b]$, $\inf\limits_{u \in [a, b]} f''(x) > 0$, x_* — точка минимума $f(x)$ на $[a, b]$. Пусть последовательность $\{x_n\}$ получена методом касательных при $x_0 = a$ по схеме* (4), (5), *$a_1 = a$, $b_1 = b$, причем $x_n \neq x_*$ $(n = 0, 1, \ldots)$. Тогда, для любого числа $\varepsilon > 0$ существует номер $N = N(\varepsilon)$ такой, что*

$$|x_n - x_*| \leqslant ((1 + \varepsilon)/2)^{n-N}(b_N - a_N), \qquad n \geqslant N. \tag{6}$$

Доказательство. Из теоремы 8.9 следует выпуклость функции $f(x)$ на $[a, b]$. Кроме того, так как $f'(x)$ строго возрастает, то множество X_* состоит из единственной точки x_*. Тогда из теоремы 1 имеем $\lim\limits_{n \to \infty} x_n = x_*$. Поскольку $[a_{n+1}, b_{n+1}] \subset [a_n, b_n]$ $(n = 1, 2, \ldots)$, то последовательность $\{a_n\}$ монотонно возрастает, а $\{b_n\}$ — монотонно убывает, причем $a_n < x_* < b_n$ $(n = 1, 2, \ldots)$. Покажем, что $\lim\limits_{n \to \infty} a_n = \lim\limits_{n \to \infty} b_n = x_*$.

В силу (4) либо $\{a_n\}$, либо $\{b_n\}$ является подпоследовательностью последовательности $\{x_n\}$ и сходится к x_*. Пусть для определенности $\lim\limits_{n \to \infty} a_n = x_*$. Допустим, что $\{b_n\}$ не сходится к x_*. Тогда согласно (4) последовательность $\{b_n\}$ не может быть подпоследовательностью для $\{x_n\}$, т. е. найдется номер $n_0 \geqslant 1$ такой, что $b_n = b_{n_0}$ при всех $n \geqslant n_0$. При $n \to \infty$ из (5) получим $f(x_*) = f(b_{n_0}) + f'(b_{n_0})(x_* - b_{n_0})$. В силу теоремы 8.4 это возможно только тогда, когда $f(x) \equiv f(b_{n_0}) + f'(b_{n_0})(x - b_{n_0})$ $(x_* \leqslant x \leqslant b_{n_0})$. Отсюда $f'(b_{n_0}) = f'(x_*) = 0$, что противоречит условию $f'(b_{n_0}) > 0$. Тем самым доказано, что обе последовательности $\{a_n\}$, $\{b_n\}$ сходятся к x_*.

Из представления (5) для точки x_{n+1} с помощью формулы Тейлора имеем

$$x_{n+1} - a_n = \frac{f(a_n) - f(b_n) - f'(b_n)(a_n - b_n)}{f'(b_n) - f'(a_n)} = \frac{1}{2}\frac{f''(\xi_n)}{f''(\mu_n)}(b_n - a_n),$$

$$b_n - x_{n+1} = \frac{f(b_n) - f(a_n) - f'(a_n)(b_n - a_n)}{f'(b_n) - f'(a_n)} = \frac{1}{2}\frac{f''(\nu_n)}{f''(\mu_n)}(b_n - a_n),$$

где ξ_n, ν_n, μ_n — некоторые точки из отрезка $[a_n, b_n]$. Отсюда получаем, что $b_{n+1} - a_{n+1} \leqslant \max\{x_{n+1} - a_n; b_n - x_{n+1}\} \leqslant q_n(b_n - a_n)/2$, где $q_n = \max\{f''(\xi_n)/f''(\mu_n); f''(\nu_n)/f''(\mu_n)\}$. Поскольку последовательности $\{\xi_n\}$, $\{\nu_n\}$, $\{\mu_n\}$ вместе с $\{a_n\}$, $\{b_n\}$ стремятся к x_*, то в силу непрерывности функции $f''(x)$ и условия $\inf\limits_{u \in [a, b]} f(u) > 0$ имеем $\lim\limits_{n \to \infty} q_n = 1$.

Следовательно, для любого $\varepsilon > 0$ найдется номер $N = N(\varepsilon)$ такой, что $q_n \leqslant 1 + \varepsilon$ при всех $n \geqslant N$. Тогда $b_{n+1} - a_{n+1} \leqslant q(b_n - a_n)$ $(n \geqslant N)$, где $q = (1 + \varepsilon)/2$. Отсюда $b_n - a_n \leqslant q^{n-N}(b_N - a_N)$. Следовательно,

$$|x_n - x_*| \leqslant (b_n - a_n) \leqslant q^{n-N}(b_N - a_N), \qquad n \geqslant N. \qquad \square$$

4. Оценка (6) означает, что метод касательных сходится со скоростью, не меньшей скорости сходимости геометрической прогрессии со знаменателем $q = (1 + \varepsilon)/2 \approx 1/2$. Конечно, существуют выпуклые функции, для которых этот метод будет сходиться гораздо быстрее (возможно, например, что точка минимума найдется за конечное число шагов). Однако нетрудно привести пример, показывающий, что на классе дважды непрерывно дифференцируемых функций оценка (6) по порядку не может быть улучшена.

Пример 1. Пусть $f(x) = x^2$ $(-1 \leqslant x \leqslant 2)$. С помощью формулы (5) легко проверить, что касательные к параболе в точках a_n, b_n пересекаются в точке $x_{n+1} = (a_n + b_n)/2$. Возьмем $x_0 = a_1 = -1$, $x_1 = b_1 = 2$. Тогда $x_2 = 1/2$. С помощью индукции нетрудно показать, что $a_n = -1/2^{n-1}$, $b_n = -1/2^{n-2}$, $x_{n+1} = 1/2^n$ при нечетных n и $a_n = -1/2^{n-2}$, $b_n = 1/2^{n-1}$, $x_{n+1} = -1/2^n$ при четных n. Отсюда получается точная оценка $|x_n - x_*| = |x_n| = 1/2^{n-1}$, в то время как, используя методику вывода оценки (6), имеем $|x_n - x_*| \leqslant b_n - a_n = 3/2^{n-1}$ $(n \geqslant 1)$.

Таким образом, из примера 1 следует, что метод касательных на классе гладких выпуклых функций не лучше метода деления отрезка пополам. Более того, для этого класса функций нетрудно предложить вариант метода деления отрезка пополам, требующий лишь вычисления значений производных минимизируемой функции.

А именно, положим $x_0 = a_1 = a$, $x_1 = b_1 = b$, вычислим значения $f'(a_1) = f'(a + 0)$, $f'(b_1) = f'(b - 0)$. Если $f'(a_1) \geqslant 0$ или $f'(b_1) \leqslant 0$, то по теореме 8.5 имеем $a \in X_*$ или $b \in X_*$ — задача решена. Поэтому пусть $f'(a_1) < 0$, $f'(b_1) > 0$. Тогда $X_* \subset (a_1, b_1)$. Пусть отрезок $[a_{n-1}, b_{n-1}]$ $(n \geqslant 2)$ уже построен, причем $f'(a_{n-1}) < 0$, $f'(b_{n-1}) > 0$, так что $X_* \subset (a_{n-1}, b_{n-1})$. Положим $x_n = (a_{n-1} + b_{n-1})/2$ и вычислим $f'(x_n)$. Если $f'(x_n) = 0$, то $x_n \in X_*$ — задача решена. Если $f'(x_n) \neq 0$, то определим точки a_n, b_n по формулам (4), приняв в них $x_n = (a_{n-1} + b_{n-1})/2$. По построению $f'(a_n) < 0$, $f'(b_n) > 0$, и согласно теореме 8.6 $X_* \subset (a_n, b_n)$. Кроме того, ясно, что

$$b_n - a_n = (b_{n-1} - a_{n-1})/2 = (b_1 - a_1)/2^{n-1}, \qquad n = 1, 2, \ldots$$

Описанный метод деления отрезка пополам выгоднее применять при минимизации тех гладких выпуклых функций, у которых значения производ-

ных вычисляются проще, чем значения функции. Если же значения и функции, и ее производных вычисляются достаточно просто, то метод касательных может оказаться предпочтительнее — хотя, как мы выше убедились, метод касательных на классе гладких выпуклых функций в целом не лучше метода деления отрезка пополам, но в то же время нетрудно привести примеры таких выпуклых функций, для которых метод касательных сходится гораздо быстрее описанного метода деления отрезка пополам.

Заметим, что метод касательных можно описать и без требования дифференцируемости выпуклой функции, используя лишь односторонние производные во внутренних точках отрезка $[a, b]$ [148].

Упражнения

1. Применить метод касательных для минимизации функций $f(x)=x^2$, $f(x)=|x|$, $f(x)=|x|+(x-1)^2$ на отрезках $[-1, 1]$, $[-1, 2]$, $[0, 1]$, $[1, 2]$.

ГЛАВА 2

Классическая теория экстремума функций многих переменных

В этой главе собраны основные факты о задачах на безусловный и условный экстремум функций конечного числа переменных, обычно излагаемые в учебниках по математическому анализу [327; 350; 352; 534]. Кроме того, приводятся необходимые условия экстремума первого порядка для задач с ограничениями типа неравенств (см., например, [14; 286; 358; 374; 605; 617; 670]). Впервые в учебной литературе излагаются новые необходимые условия экстремума второго порядка, принадлежащие А. В. Арутюнову [44]. В последнем параграфе этой главы приведены некоторые вспомогательные формулы и оценки, необходимые для дальнейшего изложения.

§ 1. Постановка задачи. Теорема Вейерштрасса

1. Сначала введем обозначения и напомним некоторые определения из линейной алгебры и математического анализа. Через $\mathbb{R}^n$ будем обозначать n-мерное вещественное линейное пространство, состоящее из вектор-столбцов

$$x=\begin{bmatrix} x^1 \\ \dots \\ x^n \end{bmatrix}, \quad y=\begin{bmatrix} y^1 \\ \dots \\ y^n \end{bmatrix}, \quad z=\begin{bmatrix} z^1 \\ \dots \\ z^n \end{bmatrix}$$

с действительными координатами $x^i, y^i, z^i, \dots$ $(i=1, \dots, n)$; сумма $x+y$ двух вектор-столбцов и произведение αx вектор-столбца x на действительное число α в $\mathbb{R}^n$ определяется обычным образом:

$$x+y=\begin{bmatrix} x^1+y^1 \\ \dots \\ x^n+y^n \end{bmatrix}, \quad \alpha x=\begin{bmatrix} \alpha x^1 \\ \dots \\ \alpha x^n \end{bmatrix};$$

вектор-столбец

$$0=\begin{bmatrix} 0 \\ \dots \\ 0 \end{bmatrix}$$

называется *нулевым*. Вектор-строку, полученную транспонированием вектор-столбца x, обозначим через $x^\top=(x^1, \dots, x^n)$. Там, где не могут возникнуть недоразумения, вектор-столбец или вектор-строку из $\mathbb{R}^n$ для краткости мы часто будем называть просто вектором или точкой, а знак транспонирования «$\top$» будем опускать.

Если в $\mathbb{R}^n$ ввести скалярное произведение двух векторов $\langle x, y\rangle = \sum_{i=1}^{n} x^i y^i$, $x, y \in \mathbb{R}^n$, то $\mathbb{R}^n$ превращается в n-мерное евклидово пространство, которое будем обозначать через E^n. Длина вектора или *норма* вектора

в E^n определяется так $|x| = \langle x, x\rangle^{1/2} = \left(\sum_{i=1}^{n} |x^i|^2\right)^{1/2}$. Величину

$$\rho(x, y) = |x - y| = \left(\sum_{i=1}^{n} |x^i - y^i|^2\right)^{1/2}$$

называют евклидовым расстоянием между точками $x, y \in E^n$. Для любых точек $x, y, z \in E^n$ справедливо неравенство

$$|x - y| \leqslant |x - z| + |z - y|,$$

называемое *неравенством треугольника*. Когда важно подчеркнуть, что скалярное произведение, норма, расстояние взяты именно в E^n, мы будем писать $\langle x, y\rangle_{E^n}$, $|x|_{E^n}$, $|x - y|_{E^n}$. В E^n справедливо *неравенство Коши — Буняковского*

$$|\langle x, y\rangle| \leqslant |x| \cdot |y| \qquad \forall x, y \in E^n,$$

причем неравенство превращается в равенство тогда и только тогда, когда векторы x, y колинеарны, т. е. $x = \alpha y$ при некотором α.

Иногда мы будем пользоваться и другими *нормами* векторов из $\mathbb{R}^n$, такими, как $|x|_p = \left(\sum_{i=1}^{n} |x^i|^p\right)^{1/p}$, $1 \leqslant p < \infty$, $|x|_\infty = \max\limits_{1 \leqslant i \leqslant n} |x^i|$. Как известно [192], в конечномерных пространствах все нормы *эквивалентны*. Это значит, что если $|\cdot|_I$, $|\cdot|_{II}$ — две нормы в $\mathbb{R}^n$, то существуют числа $m_1 > 0$, $m_2 > 0$, что $m_1|\cdot|_I \leqslant |\cdot|_{II} \leqslant m_2|\cdot|_I$ $\forall x \in \mathbb{R}^n$. Отсюда следует, что если последовательность $\{x_k\}$ сходится к точке x в какой-либо норме $|\cdot|_I$, т. е. $|x_k - x|_I \to 0$ при $k \to \infty$, то $|x_k - x|_{II} \to 0$ в любой другой норме $|\cdot|_{II}$, в частности, $|x_k - x|_\infty = \max\limits_{1 \leqslant i \leqslant n} |x_k^i - x^i| \to 0$, что равносильно покоординатной сходимости.

В дальнейшем мы будем часто пользоваться различными геометрическими понятиями в E^n, такими, как прямая, луч, гиперплоскость, подпространство, шар, сфера, конус и т. п., обобщающие на многомерный случай соответствующие привычные понятия на плоскости E^2 и в пространстве E^3 [192; 213; 349; 351; 353]. Определения некоторых понятий будут даны ниже по ходу изложения; определения некоторых, наиболее широко используемых понятий напомним здесь.

Множество $X = \{x \in E^n\colon x = x(t) = x_0 + td,\ -\infty < t < +\infty\}$ называется *прямой*, проходящей через точку x_0 и имеющей направляющий вектор $d \in E^n$, $d \neq 0$. Множество $X = \{x \in E^n\colon x = x(t) = x_0 + td,\ t \geqslant 0\}$ — *луч* с направляющим вектором $d \in E^n$, $d \neq 0$, выходящий из точки x_0; $X = \{x \in E^n\colon x = x(t) = x_0 + td,\ t > 0\}$ — *открытый луч*. Множество $X = \{x \in E^n\colon \langle c, x\rangle = \gamma\}$, где γ — заданное число, $c \in E^n$, $c \neq 0$ — заданный вектор, называется *гиперплоскостью*. Множество $X = \{x \in E^n\colon \langle a_i, x\rangle = 0,\ i = 1, \ldots, s\} = \{x \in E^n\colon Ax = 0\}$ называется *подпространством*; здесь $a_1, \ldots, a_s$ — заданные векторы-строки, не все из которых равны нулю, A — матрица размера $s \times n$ со строками $a_1, \ldots, a_s$; число $\dim X = n - r$, где $r = \operatorname{rang} A$ — ранг матрицы A, называется *размерностью подпространства* X, а r — коразмерностью этого подпространства. Ясно, что $\dim X \geqslant \max\{n - s; 0\}$. Множество $X = \{x \in E^n\colon |x - x_0| \leqslant R\}$ — *шар* радиуса R с центром в точке x_0; $X = \{x \in E^n\colon |x - x_0| < R\}$ — *открытый шар*; $X = \{x \in E^n\colon |x| \leqslant 1\}$ — *единичный шар* с центром в точке $x_0 = 0$. Множество $X = \{x \in E^n\colon |x - x_0| = R\}$ — *сфера* радиуса R с центром в точке x_0; $X = \{x \in E^n\colon |x| = 1\}$ — *единичная сфера* с центром в точке $x_0 = 0$.

Конусом (с вершиной в нуле) называется множество, содержащее вместе с любой своей точкой x и точку λx при всех $\lambda > 0$. Примерами конуса являются прямая, проходящая через начало координат, луч, выходящий из начала координат, произвольное объединение прямых и лучей. Конусами являются множества $X = \{x \in E^n\colon \langle a_i, x\rangle \leqslant 0,\ i = 1, \ldots, m,\ \langle a_i, x\rangle = 0,\ i = m + 1, \ldots, s\}$, где вектора $a_i \in E^n$, $i = 1, \ldots, s$, заданы, $X = \{x \in E^n\colon \sum_{i=1}^{n} \alpha_i (x^i)^p \leqslant 0\}$, $X = \{x \in E^n\colon \sum_{i=1}^{n} \alpha_i (x^i)^p = 0\}$, где p, α_i, $i = 1, \ldots, n$ — заданные числа. Конус называется *острым*, если не содержит никакой прямой, и *неострым*, если содержит хотя бы одну прямую.

2. Перейдем к постановке задачи минимизации. Пусть X — некоторое непустое множество из E^n, а $f(x)$ — функция, определенная на этом множестве. Всюду ниже, если не оговорено противное, мы будем рассматривать лишь функции, принимающие во всех точках $x \in X$ конечные вещественные значения. Определения таких понятий, как точка минимума и максимума, наименьшее и наибольшее значение, ограниченность снизу и сверху, нижняя и верхняя грань функции $f(x)$ на множестве X, минимизирующая и максимизирующая последовательность, точка глобального (абсолютного) локального и строгого локального минимума и максимума функции, сходимость последовательности к заданному множеству в пространстве E^n получаются из определений 1.1.1–1.1.6, 1.1.8–1.1.10, нужно лишь под x понимать точку $x = (x^1, \ldots, x^n)$ из E^n, под $|x|$ — норму x в E^n. Поэтому здесь мы не будем воспроизводить определения перечисленных понятий. Примеры 1.1.1–1.1.5 могут служить иллюстрацией к этим понятиям и в E^n, так как функция одной переменной является частным случаем функции n переменных.

Нижнюю грань функции $f(x)$ на множестве X по-прежнему будем обозначать через

$$\inf_X f(x) = f_*,$$

а множество точек минимума $f(x)$ на X — через

$$X_* = \{x\colon x \in X,\ f(x) = f_*\}.$$

Для обозначения задачи минимизации функции $f(x)$ на множестве X часто будем пользоваться следующей краткой символической записью:

$$f(x) \to \inf; \quad x \in X.$$

В этом параграфе мы будем рассматривать лишь задачи поиска нижней грани f_* и точек, близких к X_*. Как и в § 1.1, будем различать задачи минимизации двух типов. В задачах первого типа ищется точное или приближенное значение величины f_*, и здесь неважно, будет ли множество X_* пустым или непустым. В задачах второго типа наряду с величиной f_* ищется точка $x \in X$, которая достаточно близка к множеству X_* или даже принадлежит X_*, — здесь естественно требовать, чтобы $f_* > -\infty$, $X_* \neq \varnothing$.

Для приближенного решения задач обоих типов на практике обычно строят какую-либо минимизирующую последовательность $\{x_k\}$:

$$x_k \in X, \quad k = 1, 2, \ldots, \quad \lim_{k \to \infty} f(x_k) = f_*$$

(при $X_* \neq \varnothing$ возможно, например, $x_k = x_* \in X_*$, $k = 1, 2, \ldots$). Тогда, как нетрудно видеть, в качестве приближения для f_* можно взять величину

$f(x_k)$ при достаточно большом k. В том случае, если $\{x_k\}$ сходится к множеству X_*, т. е. $\rho(x_k, X_*) = \inf\limits_{x \in X_*} |x_k - x| \to 0$ при $k \to \infty$, точку x_k и соответствующее значение функции $f(x_k)$ при достаточно большом k можно принять за приближенное решение задачи второго типа. Однако, как мы видели в примере 1.1.5, условие $\lim\limits_{k \to \infty} \rho(x_k, X_*) = 0$ имеет место не всегда. Поэтому в задачах второго типа построение минимизирующих последовательностей, сходящихся к X_*, в общем случае требует привлечения специальных методов (см. ниже, главу 9).

В то же время имеются классы задач второго типа, у которых любая минимизирующая последовательность $\{x_k\}$ сходится к X_*. Эти классы задач хороши тем, что для их приближенного решения достаточно построить произвольную минимизирующую последовательность $\{x_k\}$ и затем пару $(x_k, f(x_k))$ при достаточно большом k принять за приближенное решение. Один такой класс задач для функций одной переменной был сформулирован в теореме Вейерштрасса (см. теорему 1.1.1). Эта теорема остается верной и для функций многих переменных, нужно лишь уточнить ее формулировку, заменив $\mathbb{R}$ на E^n [327; 350; 352; 534].

3. Ниже, в теореме 1, которую также будем называть *теоремой Вейерштрасса*, приводится несколько более общее утверждение, тоньше учитывающее особенности задач минимизации. Для ее формулировки нам понадобятся понятия компактного множества, полунепрерывности снизу функции и некоторые другие понятия из математического анализа. Кратко напомним их определения.

Пусть $\{x_k\} = (x_1, x_2, \ldots)$ — некоторая последовательность, $\{x_k\} \in E^n$, т. е. $x_k \in E^n$ $(k = 1, 2, \ldots)$. Точка v называется *предельной точкой последовательности* $\{x_k\}$, если существует подпоследовательность $\{x_{k_m}\}$, сходящаяся к v. Последовательность $\{x_k\}$ называется *ограниченной*, если существует постоянная $M \geqslant 0$ такая, что $|x_k| \leqslant M$ для всех $k = 1, 2, \ldots$

Числовая последовательность $\{a_k\} = (a_1, a_2, \ldots)$ называется *ограниченной снизу* [*сверху*], если существует число A такое, что $a_k \geqslant A$ $[a_k \leqslant A]$ при всех $k = 1, 2, \ldots$ Если $\{a_k\}$ не ограничена снизу [сверху], то существует подпоследовательность $\{a_{k_m}\}$ такая, что $\lim\limits_{m \to \infty} a_{k_m} = -\infty$ $\left[\lim\limits_{m \to \infty} a_{k_m} = \infty\right]$.

Множество X из E^n называется *ограниченным*, если существует постоянная $M \geqslant 0$ такая, что $|x| \leqslant M$ для всех $x \in X$. Множество $O(v, \varepsilon) = \{x: x \in E^n, |x - v| < \varepsilon\}$, представляющее собой открытый шар с центром в точке v и радиусом $\varepsilon > 0$, называется *ε-окрестностью точки v*. Точка $v \in E^n$ называется *предельной точкой* множества $X \subset E^n$, если любая ее ε-окрестность содержит точки из X, отличные от v. Нетрудно видеть, что для любой предельной точки v множества X существует последовательность $\{x_k\} \in X$ $(x_k \neq v)$, сходящаяся к v, — для построения такой последовательности достаточно при каждом $k = 1, 2, \ldots$ взять точку $x_k \in O(v, 1/k)$ $(x_k \neq v)$. Верно и обратное: если в X существует последовательность $\{x_k\}$ $(x_k \neq v)$, сходящаяся к точке v, то v — предельная точка множества X. Множество X из E^n называется *замкнутым*, если оно содержит все свои предельные точки.

Определение 1. Множество X из E^n называется *компактным*, если любая последовательность $\{x_k\} \in X$ имеет хотя бы одну предельную точку v, причем $v \in X$.

Согласно теореме Больцано — Вейерштрасса [327; 350; 352; 534] всякая ограниченная последовательность имеет хотя бы одну предельную точку. Пользуясь этой теоремой, нетрудно доказать, что в E^n компактными являются все замкнутые ограниченные множества и только они.

Определение 2. Число a называется *нижним* [*верхним*] пределом ограниченной снизу [сверху] числовой последовательности $\{a_k\}$ и обозначается через $\lim\limits_{k \to \infty} a_k = a$ $\left[\overline{\lim\limits_{k \to \infty}} a_k = a\right]$, если:

1) существует хотя бы одна подпоследовательность $\{a_{k_m}\}$ сходящаяся к a,

2) все предельные точки последовательности $\{a_k\}$ не меньше [не больше] числа a, т. е. число a является наименьшей [наибольшей] предельной точкой последовательности $\{a_k\}$.

Иначе говоря, a — нижний [верхний] предел $\{a_k\}$, если для любого $\varepsilon > 0$: 1) существует номер N такой, что $a_k \geqslant a - \varepsilon$ $[a_k \leqslant a + \varepsilon]$ для всех $k \geqslant N$; 2) для любого номера m найдется номер $k_m > m$ такой, что $a_{k_m} \leqslant a + \varepsilon$ $[a_{k_m} \geqslant a - \varepsilon]$. В том случае, когда $\{a_k\}$ не ограничена снизу [сверху], то по определению принимают $\lim\limits_{k \to \infty} a_k = -\infty$ $\left[\overline{\lim\limits_{k \to \infty}} a_k = \infty\right]$; если $\lim\limits_{k \to \infty} a_k = -\infty$, то полагают $\overline{\lim\limits_{k \to \infty}} a_k = -\infty$; если $\lim\limits_{k \to \infty} a_k = \infty$, то $\lim\limits_{k \to \infty} a_k = \infty$.

Например, если $a_k = (-1)^k$ $(k = 1, 2, \ldots)$, то $\lim\limits_{k \to \infty} a_k = -1$, $\overline{\lim\limits_{k \to \infty}} a_k = 1$; если $a_k = (-1)^k k$ $(k = 1, 2, \ldots)$, то $\lim\limits_{k \to \infty} a_k = -\infty$, $\overline{\lim\limits_{k \to \infty}} a_k = \infty$; если $a_k = [1 + (-1)^k]k$ $(k = 1, 2, \ldots)$, то $\lim\limits_{k \to \infty} a_k = 0$, $\overline{\lim\limits_{k \to \infty}} a_k = \infty$; если $a_k = k^{-1}$ $(k = 1, 2, \ldots)$, то $\lim\limits_{k \to \infty} a_k = \overline{\lim\limits_{k \to \infty}} a_k = 0$.

У любой числовой последовательности $\{a_k\}$ существуют конечный или бесконечный нижний и верхний пределы. Для того чтобы последовательность $\{a_k\}$ имела предел, необходимо и достаточно, чтобы $\lim\limits_{k \to \infty} a_k = \overline{\lim\limits_{k \to \infty}} a_k = a$; тогда $\lim\limits_{k \to \infty} a_k = a$.

Определение 3. Пусть функция $f(x)$ определена на множестве $X \subseteq E^n$. Говорят, что функция $f(x)$ *полунепрерывна снизу* [*сверху*] в точке $x \in X$, если для любой последовательности $\{x_k\} \in X$, сходящейся к точке x, имеет место соотношение $\lim\limits_{k \to \infty} f(x_k) \geqslant f(x)$ $\left[\overline{\lim\limits_{k \to \infty}} f(x_k) \leqslant f(x)\right]$. Функцию $f(x)$ называют *полунепрерывной снизу* [*сверху*] на множестве X, если она полунепрерывна снизу [сверху] в каждой точке этого множества.

Предлагаем читателю доказать, что функция $f(x)$ полунепрерывна снизу [сверху] в точке $v \in X$ тогда и только тогда, если для любого $\varepsilon > 0$ существует $\delta > 0$ такое, что для всех $x \in \{x: x \in X, |x - v| < \delta\}$ справедливо неравенство $f(x) \geqslant f(v) - \varepsilon$ $[f(x) \leqslant f(v) + \varepsilon]$. Нетрудно убедиться, что функция непрерывна в точке v тогда и только тогда, когда она в этой точке полунепрерывна и снизу, и сверху.

Пример 1. Пусть $X = \{x: x \in E^n, |x| \leqslant 1\}$ — n-мерный единичный шар; пусть $f(x) = |x|$ при $0 < |x| \leqslant 1$ и $f(0) = a$. Тогда при $a \leqslant 0$ функция $f(x)$ будет полунепрерывна снизу на X; при $a \geqslant 0$ — полунепрерывна сверху на X; при $a = 0$ — непрерывна на X.

Пример 2. Пусть $X = \{x: x \in E^1, -1 \leqslant x \leqslant 1\}$; $f(x) = x$ при $0 < x \leqslant 1$; $f(x) = 1 - x$ $(-1 \leqslant x < 0)$; $f(0) = a$. Нетрудно видеть, что при $a \leqslant 0$ эта функция полунепрерывна снизу на X; при $a \geqslant 1$ — полунепрерывна сверху на X, а при $0 < a < 1$ в точке $x = 0$ она не будет полунепрерывной ни снизу, ни сверху.

Пример 3. Пусть $u=(x, y)\in E^2$; $f(u)=x^2+y^2$ при $x>0$, $y\geqslant 0$; $f(u)=0$ при $x\leqslant 0$, $y\geqslant 0$; $f(u)=1$ при $x\geqslant 0$, $y<0$; $f(u)=-1$ при $x<0$, $y<0$. Нетрудно показать, что эта функция на множестве $X_1=\{(x, y)$: $x>0$, $y\geqslant 0\}$ непрерывна; на $X_2=\{(x, y)$: $y\geqslant 0\}$ полунепрерывна снизу; на $X_3=\{(x, y)$: $x\leqslant 0\}$ полунепрерывна сверху; на $X_4=\{(x, y)$: $x\geqslant 0\}$ в некоторых точках полунепрерывна снизу, в некоторых — сверху.

Установим связь между свойством полунепрерывности снизу функции и замкнутостью множеств

$$M(c)=\{x\colon x\in X,\ f(x)\leqslant c\},\quad c=\text{const},$$

называемых *множествами Лебега* функция $f(x)$ на множестве X.

Лемма 1. *Пусть X — замкнутое множество из E^n. Тогда для того, чтобы функция $f(x)$ была полунепрерывна снизу на X, необходимо и достаточно, чтобы множество Лебега $M(c)$ было замкнутым при всех c (пустое множество считается замкнутым по определению). В частности, если $f(x)$ полунепрерывна снизу на X, то множество X_* точек минимума $f(x)$ на X замкнуто.*

Доказательство. Необходимость. Пусть $f(x)$ полунепрерывна снизу на X. Возьмем произвольное число c. Пусть $M(c)\neq\varnothing$. Возьмем какую-либо предельную точку w множества $M(c)$. Тогда существует последовательность $\{x_k\}\in M(c)$, сходящаяся к w. В силу замкнутости X точка $w\in X$. Из того, что $f(x_k)\leqslant c$ $(k=1, 2, \ldots)$, с учетом полунепрерывности снизу $f(x)$ в точке w имеем $f(w)\leqslant\varliminf f(x_k)\leqslant c$, т. е. $w\in M(c)$. Замкнутость $M(c)$ доказана. В частности, множество $X_*=\{x\colon x\in X,\ f(x)\leqslant f_*=\inf\limits_X f(x)\}$ замкнуто.

Достаточность. Пусть для некоторой функции $f(x)$ множество $M(c)$ замкнуто при любом c. Возьмем произвольные $\varepsilon>0$, $x\in X$ и последовательность $\{x_k\}\in X$, сходящуюся к точке x. Пусть $\lim\limits_{k\to\infty} f(x_k)=a=\lim\limits_{m\to\infty} f(x_{k_m})$. Тогда $f(x_{k_m})\leqslant a+\varepsilon$, т. е. $x_{k_m}\in M(a+\varepsilon)$ для всех достаточно больших номеров k_m. Но $M(a+\varepsilon)$ замкнуто по условию, а точка x является пределом для $\{x_{k_m}\}$. Следовательно, $x\in M(a+\varepsilon)$, т. е. $f(x)\leqslant a+\varepsilon$. В силу произвола $\varepsilon>0$ отсюда имеем $f(x)\leqslant a=\lim\limits_{k\to\infty} f(x_k)$. □

Установим одно интересное свойство расстояния от точки до множества.

Лемма 2. *Пусть X — произвольное непустое множество из E^n. Тогда расстояние $\rho(x, X)=\inf\limits_{w\in X}\rho(x, w)$ от точки x до множества X как функция переменной x непрерывна на E^n и, более того, удовлетворяет условию*

$$|\rho(x, X)-\rho(y, X)|\leqslant\rho(x, y)\quad\forall x, y\in E^n.$$

Доказательство. Прежде всего из $\rho(x, w)=|x-w|\geqslant 0$ и $\rho(x, X)\leqslant|x-w|$ $(w\in X)$ следует, что функция $\rho(x, X)$ неотрицательна и конечна во всех точках $x\in E^n$. Возьмем произвольное число $\varepsilon>0$. По определению нижней грани (см. определение 1.1.3) для любых $x, y\in E^n$ найдутся точки $x_\varepsilon, y_\varepsilon\in X$ такие, что

$$\rho(x, X)\leqslant\rho(x, x_\varepsilon)\leqslant\rho(x, X)+\varepsilon,\quad\rho(y, X)\leqslant\rho(y, y_\varepsilon)\leqslant\rho(y, X)+\varepsilon.$$

Поскольку $\rho(x, X)\leqslant\rho(x, y_\varepsilon)$, то с помощью неравенства треугольника $\rho(x, y_\varepsilon)\leqslant\rho(x, y)+\rho(y, y_\varepsilon)$ имеем $\rho(x, X)-\rho(y, X)\leqslant\rho(x, y_\varepsilon)-\rho(y, y_\varepsilon)+\varepsilon\leqslant\rho(x, y)+\varepsilon$. Аналогично получается неравенство $\rho(x, X)-\rho(y, X)\geqslant\rho(x, x_\varepsilon)-\varepsilon-\rho(y, x_\varepsilon)\geqslant-\rho(x, y)-\varepsilon$. Объединяя два последних неравенства, имеем $|\rho(x, X)-\rho(y, X)|\leqslant\rho(x, y)+\varepsilon$. Отсюда при $\varepsilon\to+0$ получим требуемое неравенство. □

4. Перейдем к формулировке теоремы Вейерштрасса.

Теорема 1. *Пусть X — компактное множество, а функция $f(x)$ определена, конечна и полунепрерывна снизу на X. Тогда $f_*=\inf\limits_X f(x)>-\infty$, множество $X_*=\{x\colon x\in X,\ f(x)=f_*\}$ непусто, компактно и любая минимизирующая последовательность сходится к X_*.*

Доказательство. Возьмем произвольную минимизирующую последовательность $\{x_k\}$: $x_k\in X$ $(k=1, 2, \ldots,\ \lim\limits_{k\to\infty} f(x_k)=f_*)$. Существование хотя бы одной такой последовательности следует из определения 1.1.3 нижней грани функции. Так как X — компактное множество, то $\{x_k\}$ имеет хотя бы одну предельную точку и все ее предельные точки принадлежат X. Возьмем любую предельную точку x_* этой последовательности. Тогда существует подпоследовательность $\{x_{k_m}\}$, сходящаяся к точке x_*. Пользуясь свойством нижней грани f_* и полунепрерывностью снизу функции $f(x)$ в точке x_*, имеем

$$f_*\leqslant f(x_*)\leqslant\varliminf_{m\to\infty} f(x_{k_m})=\lim_{k\to\infty} f(x_k)=f_*,$$

т. е. $f(x_*)=f_*$. Отсюда следует, что $f_*>-\infty$, $X_*\neq\varnothing$. Более того, показано, что любая предельная точка любой минимизирующей последовательности принадлежит X_*.

Покажем, что X_* компактно. Возьмем произвольную последовательность $\{v_k\}\in X_*$. Так как $\{v_k\}\in X$ — компактное множество, то существует подпоследовательность $\{v_{k_m}\}$, сходящаяся к некоторой точке $v_*\in X$. Но $\{v_k\}$ — минимизирующая последовательность, так как $f(v_k)=f_*$ $(k=1, 2, \ldots)$. По вышедоказанному тогда $v_*\in X_*$. Компактность X_* установлена.

Покажем, что любая минимизирующая последовательность $\{x_k\}$ сходится к X_*. Так как $\rho(x_k, X_*)=\inf\limits_{x\in X_*}\rho(x_k, x)\geqslant 0$ $(k=1, 2, \ldots)$, то ясно, что $\lim\limits_{m\to\infty}\rho(x_k, X_*)\geqslant 0$. Пусть $\overline{\lim\limits_{k\to\infty}}\rho(x_k, X_*)=\lim\limits_{m\to\infty}\rho(x_{k_m}, X_*)=a\leqslant\infty$. В силу компактности X из $\{x_{k_m}\}$ можно выбрать подпоследовательность, сходящуюся к некоторой точке x_*. Не умаляя общности, можем считать, что сама последовательность $\{x_{k_m}\}$ сходится к x_*. Согласно лемме 2 функция $\rho(x, X_*)$ непрерывна по переменной x, поэтому $\lim\limits_{m\to\infty}\rho(x_{k_m}, X_*)=\rho(x_*, X_*)=a$. Однако по доказанному $x_*\in X_*$. Тогда $a=\rho(x_*, X_*)=0$. Это значит, что $\lim\limits_{m\to\infty}\rho(x_k, X_*)=\overline{\lim\limits_{k\to\infty}}\rho(x_k, X_*)=0$. Следовательно, предел $\lim\limits_{k\to\infty}\rho(x_k, X_*)$ существует и равен нулю. Теорема 1 доказана. □

Предлагаем читателю рассмотреть функции, а также множества, из примеров 1–3 (см. также примеры 1.1.1–1.1.5) и проверить, в каких случаях условия теоремы 1 выполнены и, следовательно, нижняя грань достигается, в каких случаях она не достигается и в каких случаях нижняя грань достигается несмотря на то, что условия теоремы 1 нарушены.

5. Заметим, что в теореме 1 условие компактности множества X является довольно жестким. Например, такие часто встречающиеся на практике множества, как $X=E^n$ — все пространство или $X=\{x\colon x^1\geqslant 0, \ldots, x^n\geqslant 0\}$ —

неотрицательный ортант, не являются компактными. Приведем две теоремы, в которых компактность множества X не предполагается, но зато функция, кроме полунепрерывности снизу, удовлетворяет некоторым дополнительным требованиям.

Теорема 2. *Пусть X — непустое замкнутое множество из E^n, функция $f(x)$ конечна, полунепрерывна снизу на X и для некоторой фиксированной точки $v \in X$ множество Лебега*

$$M(v) = \{x:\ x \in X,\ f(x) \leqslant f(v)\}$$

ограничено. Тогда $f_* > -\infty$, множество X_* непусто, компактно и любая минимизирующая последовательность $\{x_k\}$, принадлежащая $M(v)$, сходится к X_*.

Доказательство. По определению множества $M(v)$ имеем: $f(x) > f(v)$ при всех $x \in X \setminus M(v)$ и $f(x) \leqslant f(v)$ при всех $x \in M(v)$. Это значит, что на $X \setminus M(v)$ функция $f(x)$ не может достигать своей нижней грани на X и для доказательства теоремы достаточно рассмотреть функцию $f(x)$ на множестве $M(v)$.

Замкнутость множества $M(v)$ вытекает из леммы 1. Из ограниченности и замкнутости $M(v)$ следует его компактность. Применяя теорему 1 к функции $f(x)$ на $M(v)$, получим все утверждения теоремы 2. Попутно установили, что $X_* \subseteq M(v)$. □

Заметим, что в теореме 2 утверждается сходимость к X_* только тех минимизирующих последовательностей x_k, которые принадлежат $M(v)$. Если $f(v) > f_*$, то условие $\{x_k\} \in M(v)$ можно не оговаривать, так как в этом случае для любой минимизирующей последовательности $\{x_k\}$ найдется номер k_0 такой, что $f(x_k) < f(v)$ для всех $k \geqslant k_0$, т. е. $x_k \in M(v)$ при $k \geqslant k_0$. Если же $f(v) = f_*$, то $X_* = M(v)$ и, как видно из примера 1.1.5, в этом случае могут существовать минимизирующие последовательности, которые не принадлежат $M(v)$ и не сходятся к X_*.

Теорема 3. *Пусть X — непустое замкнутое множество из E^n, функция $f(x)$ конечна, полунепрерывна снизу на X и для любой последовательности $\{x_k\} \in X$, $\lim\limits_{k\to\infty} |x_k| = \infty$ (если такие x_k существуют) имеет место соотношение*

$$\lim_{k\to\infty} f(x_k) = \infty.$$

Тогда $f_* > -\infty$, множество X_* непусто, компактно и любая минимизирующая последовательность $\{x_k\}$ сходится к X_*.

Доказательство. Если множество X ограничено, то все утверждения теоремы следуют из теоремы 1. Поэтому пусть X не ограничено, т. е. существует хотя бы одна последовательность $\{v_k\} \in X$ такая, что $\lim\limits_{k\to\infty} |v_k| = \infty$. Тогда согласно условию теоремы $\lim\limits_{k\to\infty} f(v_k) = \infty$. Возьмем какую-либо точку $v \in X$ такую, что $f(v) > f_*$ (например, можно принять $v = v_k$ при достаточно большом k), и рассмотрим множество Лебега $M(v) = \{x:\ x \in X,\ f(x) \leqslant f(v)\}$. Покажем, что $M(v)$ ограничено. Допустим противное: пусть существует последовательность $\{w_k\} \in M(v)$ такая, что $\lim\limits_{k\to\infty} |w_k| = \infty$. Тогда $\lim\limits_{k\to\infty} f(w_k) = \infty$, что противоречит неравенству $f(w_k) \leqslant f(v) < \infty$, вытекающему из включения $w_k \in M(v)$ $(k = 1, 2, \ldots)$. Таким образом, множество $M(v)$ ограничено. Отсюда и из теоремы 2 следуют все утверждения теоремы 3. □

Следствие 1. *Пусть X — непустое замкнутое множество из E^n. Тогда для любой точки $x \in E^n$ найдется точка $v = v(x) \in X$ такая, что $\rho(x, X) = \inf\limits_{w\in X} |x-w| = |x-v(x)|$, т. е. $v(x)$ — ближайшая к x точка из X.*

Доказательство. Пусть x — произвольная точка из E^n. Рассмотрим функцию $g(w) = |w - x|$ переменной $w \in E^n$. Ясно, что $g(w)$ непрерывна на E^n. Кроме того, $g(w) \geqslant |w| - |x|$, так что $\lim\limits_{|w|\to\infty} g(w) = \infty$. Таким образом, $g(w)$ удовлетворяет условиям теоремы 3 на любом непустом замкнутом множестве $X \subseteq E^n$. Существование искомой точки $v = v(x)$ теперь следует непосредственно из теоремы 3. Заметим, что такая точка $v(x)$, вообще говоря, неединственна. □

6. В заключение кратко остановимся на задаче максимизации функции $f(x)$ на множестве X. Для обозначения этой задачи также будем пользоваться символической записью

$$f(x) \to \sup, \quad x \in X.$$

Так как $\sup\limits_{x\in G} f(x) = -\inf\limits_{x\in G}(-f(x))$ для любого множества $G \subseteq X$, то ясно, что любая точка локального или глобального максимума, а также любая максимизирующая последовательность для функции $f(x)$ на множестве X будет соответственно точкой локального или глобального минимума или минимизирующей последовательностью для функции $(-f(x))$ на X.

Это значит, что любая задача максимизации функции $f(x)$ на X равносильна задаче минимизации функции $(-f(x))$ на том же множестве X. Поэтому можно ограничиться изучением лишь задач минимизации.

Предлагаем читателю, пользуясь указанной связью между задачами минимизации и максимизации, по аналогии с п. 2 сформулировать задачи максимизации первого и второго типов. Далее, учитывая, что полунепрерывность сверху функции $f(x)$ равносильна полунепрерывности снизу функции $-f(x)$, нетрудно сформулировать и доказать аналоги теорем 1–3 для задач максимизации. Для примера приведем формулировку теоремы Вейерштрасса, являющуюся аналогом теоремы 1.

Теорема 4. *Пусть X — компактное множество, а функция $f(x)$ определена, конечна и полунепрерывна сверху на X. Тогда $f^* = \sup\limits_X f(x) < +\infty$, множество $X^* = \{x:\ x \in X,\ f(x) = f^*\}$ непусто, компактно и любая максимизирующая последовательность сходится к X^*.*

Упражнения

1. Выяснить, будет ли произвольная минимизирующая последовательность сходиться к множеству точек минимума функции $f(u)$ на множестве X, если:

а) $X = \{u = (x, y) \in E^2,\ x \geqslant 0,\ y \geqslant 0,\ x + 2y \leqslant 1\}$, $f(u) = x + y$;

б) $X = E^n$, $f(x) = |x|(1 + |x|^2)^{-1}$;

в) $X = E^n$, $f(x) = |x|^2$.

2. Пусть X — замкнутое множество из E^n, функция $f(x)$ полунепрерывна снизу на X, $f_* = \inf\limits_{x\in X} f(x) > -\infty$, $M_\alpha = \{x \in X: f(x) < f_* + \alpha\}$. Доказать, что если множество M_α ограничено при некотором $\alpha > 0$, то $X_* = \{x \in X: f(x) = f_*\}$ непусто, компактно и любая минимизирующая последовательность сходится к X_* (ср. с теоремой 2). Если множество M_α неограничено при всех $\alpha > 0$, то возможны случаи, когда $X_* = \varnothing$ или $X_* \neq \varnothing$ и неограничено, или $X_* \neq \varnothing$ ограничено, но существует минимизирующая последовательность, которая не стремится к X_*.

Рассмотреть функцию $f(x)=x^2$ на множествах $X_1=\{x=(x^1, x^2)\in E^2\colon x^2\geqslant e^{-x^1}\}$, $X_2=\{x=(x^1, x^2)\in E^2\colon 0\leqslant x^2\leqslant 1\}$, $X_3=X_1\cup\{x=(0,0)\}$.

3. Доказать следующие свойства верхнего и нижнего пределов числовых последовательностей:

а) $\overline{\lim\limits_{n\to\infty}} ca_n = c\overline{\lim\limits_{n\to\infty}} a_n$, $\underline{\lim\limits_{n\to\infty}} ca_n = c\underline{\lim\limits_{n\to\infty}} a_n$, $\underline{\lim\limits_{n\to\infty}}(-ca_n) = -c\overline{\lim\limits_{n\to\infty}} a_n$, $\overline{\lim\limits_{n\to\infty}}(-ca_n) = -c\underline{\lim\limits_{n\to\infty}} a_n$ для любых $c=\text{const}>0$;

б) если $a_n\leqslant b_n$ $(n=1,2,\ldots)$, то $\underline{\lim\limits_{n\to\infty}} a_n\leqslant\underline{\lim\limits_{n\to\infty}} b_n$, $\overline{\lim\limits_{n\to\infty}} a_n\leqslant\overline{\lim\limits_{n\to\infty}} b_n$. Можно ли утверждать, что $\underline{\lim\limits_{n\to\infty}} a_n\geqslant\overline{\lim\limits_{n\to\infty}} b_n$? Рассмотреть пример: $a_n=(-1)^n$, $n=1,2,\ldots$, $b_n=0$ при $n=2k$, $b_n=-2$ при $n=2k-1$, $k=1,2,\ldots$;

в) $\underline{\lim\limits_{n\to\infty}} a_n+\overline{\lim\limits_{n\to\infty}} b_n\leqslant\overline{\lim\limits_{n\to\infty}}(a_n+b_n)\leqslant\overline{\lim\limits_{n\to\infty}} a_n+\overline{\lim\limits_{n\to\infty}} b_n$. Рассмотреть пример $a_n=(-1)^n$, $b_n=(-1)^n$ или $b_n=(-1)^{n+1}$ и убедиться, что здесь возможны строгие неравенства;

г) $\underline{\lim\limits_{n\to\infty}} a_n+\underline{\lim\limits_{n\to\infty}} b_n\leqslant\underline{\lim\limits_{n\to\infty}}(a_n+b_n)\leqslant\underline{\lim\limits_{n\to\infty}} a_n+\overline{\lim\limits_{n\to\infty}} b_n$. Привести примеры последовательностей, когда здесь возможны строгие неравенства;

д) если существует $\lim\limits_{n\to\infty} a_n$, то $\overline{\lim\limits_{n\to\infty}}(a_n+b_n)=\lim\limits_{n\to\infty} a_n+\overline{\lim\limits_{n\to\infty}} b_n$, $\underline{\lim\limits_{n\to\infty}}(a_n+b_n)=\lim\limits_{n\to\infty} a_n+\underline{\lim\limits_{n\to\infty}} b_n$;

е) если $a_n\geqslant 0$, $b_n\geqslant 0$ $(n=1,2,\ldots)$, то $\underline{\lim\limits_{n\to\infty}} a_n\cdot\underline{\lim\limits_{n\to\infty}} b_n\leqslant\underline{\lim\limits_{n\to\infty}} a_nb_n\leqslant\underline{\lim\limits_{n\to\infty}} a_n\cdot\overline{\lim\limits_{n\to\infty}} b_n$; $\underline{\lim\limits_{n\to\infty}} a_n\cdot\overline{\lim\limits_{n\to\infty}} b_n\leqslant\overline{\lim\limits_{n\to\infty}} a_nb_n\leqslant\overline{\lim\limits_{n\to\infty}} a_n\cdot\overline{\lim\limits_{n\to\infty}} b_n$. Привести примеры последовательностей, когда здесь возможны строгие неравенства;

ж) если $a_n\geqslant 0$, $b_n\geqslant 0$ $(n=1,2,\ldots)$ и существует $\lim\limits_{n\to\infty} a_n$, то $\overline{\lim\limits_{n\to\infty}} a_nb_n=\lim\limits_{n\to\infty} a_n\cdot\overline{\lim\limits_{n\to\infty}} b_n$, $\underline{\lim\limits_{n\to\infty}}(a_nb_n)=\lim\limits_{n\to\infty} a_n\cdot\underline{\lim\limits_{n\to\infty}} b_n$.

4. Найти верхний и нижний пределы последовательности $a_n=\sin n\alpha$, где α — фиксированное число.

5. Пусть $f(x)=(1-e^{-|x|})^{-1}$ $(x\neq 0)$. Как надо доопределить эту функцию при $x=0$, чтобы она стала полунепрерывной снизу или сверху на $E^1=\mathbb{R}$?

6. Пусть $f_1(x)$, $f_2(x)$ — две функции, полунепрерывные снизу на множестве X. Будут ли полунепрерывными снизу на X следующие функции:

а) $f(x)=\alpha_1 f_1(x)+\alpha_2 f_2(x)$ (рассмотреть случаи положительных и отрицательных α_1, α_2);

б) $f(x)=\min\{f_1(x); f_2(x)\}$;

в) $f(x)=\max\{f_1(x); f_2(x)\}$;

г) $f(x)=|f_1(x)|$?

7. Пусть функция $f(x)$ определена на множестве X. Говорят, что число A является *нижним [верхним] пределом* этой функции в точке v по множеству X и обозначают $\underline{\lim\limits_{x\to v}} f(x)=A$ $\left[\overline{\lim\limits_{x\to v}} f(x)=A\right]$, если:

а) для любой последовательности $\{x_k\}\in X$, сходящейся к v, имеет место неравенство $\underline{\lim\limits_{k\to\infty}} f(x_k)\geqslant A$ $\left[\overline{\lim\limits_{k\to\infty}} f(x_k)\leqslant A\right]$;

б) существует последовательность $\{x_k\}\in X$, сходящаяся к v и такая, что $\lim\limits_{k\to\infty} f(x_k)=A$.

Доказать, что функция $f(x)$ полунепрерывна снизу [сверху] в точке v, если $\underline{\lim\limits_{x\to v}} f(x)\geqslant f(v)$ $\left[\overline{\lim\limits_{x\to v}} f(x)\leqslant f(v)\right]$.

8. Пусть $X=E^n$, $f(x)=\sin(\pi|x|^{-1})$ при $x\neq 0$ и $f(0)=a$. При каких a эта функция будет полунепрерывна снизу или сверху на X? Найти $\underline{\lim\limits_{x\to 0}} f(x)$, $\overline{\lim\limits_{x\to 0}} f(x)$. Что изменится, если $X=\{x\colon |x|=n^{-1},\ n=1,2,\ldots\}$?

9. Показать, что понятия верхнего и нижнего предела функции в точке обладают свойствами, аналогичными свойствам верхнего и нижнего предела числовых последовательностей, приведенных в упражнении 3.

10. Пусть $X\subset E^n$ — произвольное множество, $\overline{X}$ — замыкание X (см. определение 4.1.6), функция $f(x)$ непрерывна на $\overline{X}$. Доказать, что

$$\inf_{x\in X} f(x)=\inf_{x\in\overline{X}} f(x),\qquad \sup_{x\in X} f(x)=\sup_{x\in\overline{X}} f(x).$$

11. Пусть $X\subset E^n$ — произвольное множество из E^n, функции $f(x)$, $g(x)$ определены на X. Доказать, что

а) $\inf\limits_{x\in X}\min\{f(x); g(x)\}=\min\{\inf\limits_{x\in X} f(x); \inf\limits_{x\in X} g(x)\}$;

б) $\inf\limits_{x\in X}\max\{f(x); g(x)\}\geqslant\max\{\inf\limits_{x\in X} f(x); \inf\limits_{x\in X} g(x)\}$.

12. Пусть $X\subset E^n$, $Y\subset E^m$ — произвольные множества, функция $f(x,y)$ определена на множестве $X\times Y=\{(x,y)\colon x\in X, y\in Y\}$. Доказать, что а) $\inf\limits_{X\times Y} f(x,y)=\inf\limits_{X}\{\inf\limits_{Y} f(x,y)\}=\inf\limits_{Y}\{\inf\limits_{X} f(x,y)\}$; б) $\sup\limits_{x\in X}\inf\limits_{y\in Y}\leqslant\inf\limits_{y\in Y}\sup\limits_{x\in X} f(x,y)$.

13. Пусть $X\subset E^n$, $Y\subset E^m$ — произвольные множества, функция $\varphi(x,y)$ определена на множестве $X\times Y$, полунепрерывна снизу на X при каждом фиксированном $y\in Y$ и ограничена сверху на Y при каждом фиксированном $x\in X$. Доказать, что функция $f(x)=\sup\limits_{y\in Y}\varphi(x,y)$ полунепрерывна снизу на X [54].

14. Пусть X — произвольное множество из E^n, Y — компактное множество из E^m, функция $\varphi(x,y)$ непрерывна [полунепрерывна снизу] на $X\times Y$. Доказать, что функция $f(x)=\inf\limits_{y\in Y}\varphi(x,y)$ непрерывна [полунепрерывна снизу] на X. Покажите, что условия на Y и $\varphi(x,y)$ в этом утверждении существенны [54]. Рассмотрите примеры: 1) $\varphi(x,y)=xy$, $X=Y=E^1$; 2) $\varphi(x,y)=x^2y^2+y$, $X=Y=E^1$; 3) $\varphi(x,y)=\dfrac{1}{1+x^2y^2}$, $X=Y=E^1$; 4) $\varphi(x,y)=\dfrac{xy}{x^2+y^2}$, $x^2+y^2\neq 0$; $\varphi(0,0)=0$, $X=Y=[-1,1]$; 5) $\varphi(x,y)=\max\{0; 1-xy\}$, $X=Y=[0,+\infty)$.

15. Пусть $X\subset E^n$, $Y\subset E^m$ — произвольные множества, функция $\varphi(x,y)$ непрерывна на $X\times Y$, пусть $\widetilde{Y}$ — плотное в Y подмножество, $\widetilde{X}$ — ограниченное подмножество из X, такие, что при каждом $y\in\widetilde{Y}$ функция $\varphi(x,y)$ своего максимума на X достигает в точке $x=x(y)\in\widetilde{X}$. Доказать, что функция $f(y)=\sup\limits_{x\in X}\varphi(x,y)$ непрерывна на Y.

§ 2. Классический метод решения задач на безусловный экстремум

Рассмотрим задачу поиска локального или глобального экстремума гладкой функции многих переменных на всем пространстве E^n. Такую задачу принято называть задачей на *безусловный экстремум*. В этом названии отражен тот факт, что на переменные $x=(x^1,\ldots,x^n)$ никакие дополнительные ограничения в такой задаче не накладываются. Кратко изложим *классический метод* поиска решения задач на безусловный экстремум, подразумевая под этим тот подход к ним, который основан на дифференциальном исчислении функций многих переменных и обычно излагается в учебниках по математическому анализу [327; 350; 352; 534; 768].

Сначала напомним некоторые понятия и факты.

Определение 1. Пусть функция $f(x)$ определена в некоторой ε-окрестности $O(x,\varepsilon)=\{v\in E^n\colon |v-x|<\varepsilon\}$ точки x. Говорят, что функция $f(x)$ *дифференцируема в точке* x, если существует вектор $f'(x)\in E^n$, такой, что приращение функции можно представить в виде:

$$\Delta f(x)=f(x+h)-f(x)=\langle f'(x), h\rangle+o(h,x),\quad |h|<\varepsilon, \tag{1}$$

где $o(h,x)$ — величина, бесконечно малая более высокого порядка, чем $|h|$, т. е. $\lim\limits_{|h|\to 0}\dfrac{o(h,x)}{|h|}=0$. Вектор $f'(x)$ называется *первой производной* или *градиентом* функции f в точке x.

Условие (1) однозначно определяет градиент $f'(x)$, причем

$$f'(x)=(f_{x^1}(x),\ldots,f_{x^n}(x)), \tag{2}$$

где

$$f_{x^i}(x)=\frac{\partial f(x)}{\partial x^i}=\lim_{t\to 0}\frac{f(x+te_i)-f(x)}{t}$$

есть *частная производная* функции f в точке x по переменной x^i, $e_i = (0, \ldots, 1, \ldots, 0)$ — единичный вектор, у которого i-я координата равна 1, остальные координаты равны нулю, $i = 1, \ldots, n$. Если функция дифференцируема в каждой точке множества X, то ее часто называют *гладкой на* X. Если производная $f'(x)$ существует и непрерывна в каждой точке $x \in X$, то функцию $f(x)$ называют *непрерывно дифференцируемой* на множестве X. Напомним, что функция, дифференцируемая в точке x, непрерывна в этой точке.

Определение 2. Пусть функция $f(x)$ определена и дифференцируема во всех точках некоторой ε-окрестности точки x. Говорят, что функция $f(x)$ *дважды дифференцируема в точке* x, если существует матрица $f''(x)$ размера $n \times n$, такая, что

$$f'(x+h) - f'(x) = f''(x)h + o_1(h, x), \quad |h| < \varepsilon, \tag{3}$$

где $\lim\limits_{|h|\to 0} \frac{|o_1(h, x)|}{|h|} = 0$. Матрица $f''(x)$ называется *второй производной* функции f в точке x.

Условие (3) однозначно определяет вторую производную $f''(x)$, причем

$$f''(x) = \begin{pmatrix} f_{x^1x^1}(x) & f_{x^1x^2}(x) & \ldots & f_{x^1x^n}(x) \\ f_{x^2x^1}(x) & f_{x^2x^2}(x) & \ldots & f_{x^2x^n}(x) \\ \ldots\ldots\ldots & \ldots\ldots\ldots & \ldots & \ldots\ldots\ldots \\ f_{x^nx^1}(x) & f_{x^nx^2}(x) & \ldots & f_{x^nx^n}(x) \end{pmatrix} = \{f_{x^ix^j}(x),\ i, j = 1, \ldots, n\}, \tag{4}$$

где $f_{x^ix^j}(x) = \frac{\partial^2 f(x)}{\partial x^i \partial x^j} = \frac{\partial}{\partial x^i}\left(\frac{\partial f(x)}{\partial x^j}\right) = \frac{\partial}{\partial x^j}\left(\frac{\partial f(x)}{\partial x^i}\right)$ — вторая частная производная функции $f(x)$ по переменным x^i, x^j. Как видим, $f''(x)$ — симметричная матрица. Если функция $f(x)$ дважды дифференцируема в некоторой ε-окрестности точки x, причем ее вторая производная непрерывна в точке x, т. е. $\lim\limits_{|h|\to 0} \|f''(x+h) - f''(x)\| = 0$, то

$$\Delta f(x) = f(x+h) - f(x) = \langle f'(x), h\rangle + \tfrac{1}{2}\langle f''(x)h, h\rangle + o_2(h, x), \quad |h| < \varepsilon, \tag{5}$$

где $\lim\limits_{|h|\to 0} \frac{|o_2(h, x)|}{|h|^2} = 0$ [327; 350; 352; 534; 768]. Квадратичную форму $d^2 f(x) = \langle f''(x)h, h\rangle = \sum\limits_{i,j=1}^{n} \frac{\partial^2 f(x)}{\partial x^i \partial x^j} h^i h^j$ переменной $h \in E^n$ называют *вторым дифференциалом* функции f в точке x. Если функция дважды дифференцируема в каждой точке множества X, то ее часто называют *дважды гладкой* на X. Если вторая производная $f''(x)$ существует и непрерывна в каждой точке $x \in X$, то функцию $f(x)$ называют *дважды непрерывно дифференцируемой* на множестве X.

Определение 3. Пусть $A = \{a_{ij},\ i, j = 1, \ldots, n\}$ симметричная матрица, $\langle Ah, h\rangle = \sum\limits_{i,j=1}^{n} a_{ij} h^i h^j$ — соответствующая ей квадратичная форма. Говорят, что матрица A *положительно* [*неотрицательно*] *определена* на E^n и обозначают $A > 0$ [$A \geqslant 0$], если $\langle Ah, h\rangle > 0\ \forall h \in E^n,\ h \neq 0$ [$\langle Ah, h\rangle \geqslant 0\ \forall h \in E^n$]. Аналогично, матрица A *отрицательно* [*неположительно*] *определена* на E^n, т. е. $A < 0$ [$A \leqslant 0$], если $\langle Ah, h\rangle < 0\ \forall h \in E^n,\ h \neq 0$ [$\langle Ah, h\rangle \leqslant 0\ \forall h \in E^n$] (см. [192; 213; 349; 353]).

Перейдем к изложению необходимых и достаточных условий оптимальности в задачах на безусловный экстремум.

Теорема 1 (Необходимое условие экстремума). *Пусть x_* — точка локального экстремума (минимума или максимума) функции $f(x)$ на E^n, пусть $f(x)$ дифференцируема в точке x_*. Тогда*

$$f'(x_*) = 0. \tag{6}$$

Если $f(x)$ дважды дифференцируема в некоторой ε-окрестности $O(x_*, \varepsilon)$ точки x_* и $f''(x)$ непрерывна в точке x_*, то $f''(x_*) \geqslant 0$ в точке локального минимума и $f''(x_*) \leqslant 0$ в точке локального максимума.

Доказательство. Пусть для определенности x_* — точка локального минимума $f(x)$ на E^n. Это значит, что существует ε-окрестность $O(x_*, \varepsilon)$ точки x_* такая, что $f(x) \geqslant f(x_*)\ \forall x \in O(x_*, \varepsilon)$. Отсюда и из (1) при $x = x_*$, $h = -tf'(x_*)$, $0 < t < t_0$, где число t_0 столь мало, что $t_0|f'(x_*)| < \varepsilon$, имеем $0 \leqslant \langle f'(x_*), -tf'(x_*)\rangle + o(t) = -t|f'(x_*)|^2 + o(t)$. Разделим это неравенство на $t > 0$ и затем устремим $t \to +0$. Получим $-t|f'(x_*)|^2 \geqslant 0$, что возможно только при выполнении равенства (6).

Далее, пусть $f(x)$ дважды дифференцируема в окрестности $O(x_*, \varepsilon)$ и $f''(x)$ непрерывна в точке x_*. Зафиксируем произвольное $h \in E^n$ и возьмем $t_0 > 0$ столь малым, что $t_0|h| < \varepsilon$. Тогда $x_* + th \in O(x_*, \varepsilon)$ и из (5) с учетом уже доказанного равенства (6) имеем

$$0 \leqslant f(x_* + th) - f(x_*) = \tfrac{1}{2}\langle f''(x_*)h, h\rangle t^2 + o(t^2)\ \forall t,\ 0 < t < t_0.$$

Разделим это неравенство на $t^2 > 0$ и устремим $t \to +0$. Получим $0 \leqslant \langle f''(x_*)h, h\rangle\ \forall h \in E^n$. Согласно определению 3 это значит, что $f''(x_*) \geqslant 0$. Так как точка локального максимума функции $f(x)$ является точкой локального минимума функции $(-f(x))$, то, применяя уже доказанные утверждения теоремы к функции $(-f(x))$, получим, что если x_* — точка локального максимума $f(x)$ на E^n, то $f'(x_*) = 0$, $f''(x_*) \leqslant 0$. □

Определение 4. Точка v, удовлетворяющая уравнению $f'(v) = 0$, называется *стационарной точкой* функции $f(x)$.

Из теоремы 1 следует, что только стационарные точки могут быть точками экстремума дифференцируемой на E^n функции. Однако стационарная точка необязательно является точкой экстремума. Более того, если в стационарной точке v еще выполняется условие $f''(v) \geqslant 0$ [$f''(v) \leqslant 0$], то это также не значит, что точка v непременно является точкой локального минимума [максимума]. Можно уверенно сказать лишь одно: стационарные точки являются подозрительными на экстремум.

Пример 1. Пусть $f(u) = x^4 - y^4$, $u = (x, y) \in E^2$. Очевидно, $v = (0, 0)$ — стационарная точка функции $f(x)$ и в ней $f''(v) = 0$. Однако в любой окрестности точки $v = 0$ существуют точки x, в которых $f(x) > f(v) = 0$ и $f(x) < 0$, т. е. $v = 0$ не является точкой экстремума.

Этот пример показывает, что условия экстремума, сформулированные в теореме 1, являются лишь необходимыми, но в общем случае этих условий

недостаточно для экстремума. Тем не менее, оказывается, несколько усилив условия теоремы 1, можно получить условия, достаточные для экстремума.

Теорема 2. *Пусть функция $f(x)$ дважды дифференцируема в окрестности $O(v, \varepsilon)$ стационарной точки v этой функции и $f''(x)$ непрерывна в точке v. Тогда если $f''(v) > 0$, то v — точка строгого локального минимума функции $f(x)$, а если $f''(v) < 0$, то v — точка строгого локального максимума.*

Доказательство. Пусть в точке v выполнены условия $f'(v) = 0$, $f''(v) > 0$, но v не является точкой строгого локального минимума. Тогда существует последовательность $\{x_k\}$, такая, что $x_k \neq v$, $\{x_k\} \to v$, $f(x_k) \leqslant f(v)$. Точки x_k можем представить в виде $x_k = v + t_k d_k$, где $d_k = \frac{x_k - v}{|x_k - v|}$, $t_k = |x_k - v| \to 0$ при $k \to \infty$. Так как $|d_k| = 1$, то, выбирая при необходимости подпоследовательность согласно теореме Больцано — Вейерштрасса [327; 350; 352; 534], можем считать, что $\{d_k\} \to d_0$, $|d_0| = 1$. Тогда полагая в (5) $x = v$, $h = t_k d_k$, имеем $0 \geqslant f(x_k) - f(v) = \frac{1}{2} t_k^2 \langle f''(v) d_k, d_k \rangle + o(t_k^2)$, $k = 1, 2, \ldots$. Разделим это неравенство на $t_k^2 > 0$ и устремим $k \to \infty$. Получим $\langle f''(v) d_0, d_0 \rangle \leqslant 0$, где $d_0 \neq 0$. Однако это противоречит условию $f''(v) > 0$. Следовательно, v — точка строгого локального минимума функции $f(x)$.

Аналогично доказывается, что если $f'(v) = 0$, $f''(v) < 0$, то v — точка строгого локального максимума. Теорема 2 доказана. □

Замечание 1. Для выяснения знакоопределенности квадратичных форм $< Ah, h >= \sum\limits_{i,j=1}^{n} a_{ij} h^i h^j$ существуют различные алгебраические критерии [192; 213; 349; 353].

Определение 5. *Главными минорами* матрицы A называются определители

$$\Delta_{i_1 i_2 \ldots i_k} = \det \begin{vmatrix} a_{i_1 i_1} & \ldots & a_{i_1 i_k} \\ \ldots\ldots\ldots\ldots \\ a_{i_k i_1} & \ldots & a_{i_k i_k} \end{vmatrix}, \quad 1 \leqslant i_1 < i_2 < \ldots < i_k \leqslant n, \quad k = 1, \ldots, n.$$

Главными угловыми минорами называются определители $\Delta_{1,2,\ldots,k}$, $k = 1, \ldots, n$.

Критерий Сильвестра: для того, чтобы $A \geqslant 0$, необходимо и достаточно, чтобы все главные миноры матрицы A были неотрицательны; для того, чтобы $A > 0$, необходимо и достаточно, чтобы все главные угловые миноры матрицы A были положительны.

Сформулируем критерии знакоопределенности матрицы A в терминах собственных чисел этой матрицы. Напомним, что *собственным числом* матрицы A называется решение λ уравнения $\det |A - \lambda I_n| = 0$, где I_n — единичная матрица размера $n \times n$. Так как у нас A — симметричная матрица, то у нее существует n действительных собственных чисел $\lambda_1, \lambda_2, \ldots, \lambda_n$ (с учетом их кратности). Для того, чтобы $A \geqslant 0$ [$A > 0$], необходимо и достаточно, чтобы все собственные числа матрицы A были неотрицательны [положительны]. Квадратичная форма $\langle Ah, h \rangle$ знакопеременна тогда и только тогда, когда у матрицы A имеется хотя бы одно положительное и хотя бы одно отрицательное собственное число.

В том случае, когда в стационарной точке v квадратичная форма $\langle f''(v)h, h \rangle$ не меняет знака при всех $h \in E^n$, но может равняться нулю при некоторых $h \neq 0$, то для выяснения поведения функции в окрестности точки v можно привлечь старшие производные и связанные с ними формы более высокого порядка:

$$d^m f(v) = \sum \frac{\partial^m f(v)}{(\partial x^1)^{r_1} \ldots (\partial x^n)^{r_n}} (h^1)^{r_1} \ldots (h^n)^{r_n},$$

где суммирование проводится по всем целым $r_1, \ldots, r_m$ таким, что $0 \leqslant r_i \leqslant m$, $i = 1, \ldots, n$, $r_1 + r_2 + \ldots + r_n = m$. Однако на практике исследование характера стационарных точек с помощью форм $d^m f(v)$, $m \geqslant 3$, почти не применяется из-за его громоздкости.

В тех случаях, когда описанным выше способом удается выявить все точки локального минимума [максимума] функции $f(x)$, то для определения глобального минимума [максимума] этой функции на всем пространстве E^n нужно перебрать все найденные точки и из них выбрать точку с наименьшим [наибольшим] значением функции, если такая точка существует.

Пример 2. Пусть в пространстве E^n даны p точек $x_i = (x_i^1, \ldots, x_i^n)$, $i = 1, \ldots, p$, и требуется найти точку $x \in E^n$, сумма квадратов расстояний от которой до этих данных точек минимальна.

Эта задача равносильна задаче минимизации функции $f(x) = \sum\limits_{i=1}^{p} |x - x_i|^2$ на E^n. Функцию $f(x)$ удобнее представить в виде: $f(x) = p|x|^2 - 2p\langle x, x_0 \rangle + \sum\limits_{i=1}^{p} |x_i|^2$, где $x_0 = \frac{1}{p} \sum\limits_{i=1}^{p} x_i$. Отсюда видно, что $f'(x) = 2p(x - x_0)$ и $v = x_0$ — стационарная точка. Матрица $f''(v) = 2pI_n$, где I_n — единичная матрица размера $n \times n$. Следовательно, $\langle f''(v)h, h \rangle = 2p|h|^2 > 0$ при всех $h \in E^n$, $h \neq 0$. Согласно теореме 2 это значит, что $v = x_0$ — точка строгого локального минимума. Однако здесь можно сказать больше: $v = x_0$ — точка глобального минимума $f(x)$ на E^n. В самом деле, рассматриваемая функция такова, что $\lim\limits_{|x| \to \infty} f(x) = \infty$. Тогда по теореме 1.3 множество X_* точек глобального минимума $f(x)$ на E^n непусто, а по теореме 1 $\forall x_* \in X_*$ является стационарной точкой. Поскольку здесь имеется единственная стационарная точка $v = x_0$, то $v \in X_*$. Следовательно, $X_* = \{x_0\}$, $f_* = f(x_0) = -p|x_0|^2 + \sum\limits_{i=1}^{p} |x_i|^2$. Заметим, что при исследовании этой несложной задачи можно было обойтись и без привлечения теоремы 1.3, поскольку здесь $f(x) - f(x_0) = p|x - x_0|^2 \geqslant 0$ $\forall x \in E^n$. Ясно также, что в этой задаче $f^* = \sup\limits_{x \in E^n} f(x) = +\infty$, т. е. задача максимизации $f(x)$ на E^n не имеет решения.

Упражнения

1. Найти экстремумы функций:

а) $f(u) = (x + y - 1)e^{-(x^2 - xy + y^2)}$, $u = (x, y) \in E^2$;

б) $f(u) = xy^2z^3(1 - x - 2y - 3z)$, $u = (x, y, z) \in E^3$;

в) $f(u) = \sin(x + y) - \sin x - \sin y$, $u = (x, y) \in E^2$.

2. Может ли функция двух переменных на плоскости иметь бесконечно много точек локального минимума и ни одной точки локального максимума? Рассмотрите функцию $f(u) = xe^x - (1 + e^x)\cos y$, $u = (x, y) \in E^2$.

В следующие упражнения 3–8 вошли задачи, которые автору сообщил Н. А. Бобылев.

3. В точке $u_0=(0,0)$ функция $f(u)$ переменных $u=(x,y)\in E^2$ имеет локальный минимум вдоль каждой прямой, проходящей через точку u_0. Можно ли утверждать, что в точке u_0 реализуется локальный минимум функции $f(x)$? Рассмотреть функции:

$$f_1(u)=(x-y^2)(2x-y^2),\quad f_2(u)=x^2-2xy^2+y^4-y^5.$$

4. Построить пример гладкой функции $f(u)$, $u=(x,y)\in E^2$, для которой точка $(0,0)$ является единственной стационарной точкой, причем $(0,0)$ — точка локального минимума функции f, не являющаяся точкой ее глобального минимума. Рассмотреть функцию

$$f(u)=\frac{3x^2-2x^3-1}{1+y^2}+(3x^2-2x^3)e^{-y}.$$

5. Пусть $x_0=0$ является единственной стационарной точкой гладкой функции $f(x)$ на E^n, в которой реализуется локальный минимум этой функции. Пусть для некоторых $\varepsilon>0$ и $R>0$ выполнено неравенство $|f'(x)|\geqslant\varepsilon\ \ \forall x,\ |x|\geqslant R$. Докажите, что $x_0=0$ — точка глобального минимума функции f на E^n.

6. Постройте пример (нарисуйте графики линий уровня) дифференцируемой функции $f(u)$, $u=(x,y)\in E^2$, у которой имеется ровно m стационарных точек и все они являются точками локального минимума этой функции.

7. Пусть заданы произвольные целые числа $m\geqslant 0$, $k\geqslant 0$, $l\geqslant 0$. Покажите, что существует гладкая функция $f(x)$, $u=(x,y)\in E^2$, у которой имеется ровно $m+k+l$ стационарных точек, причем m из них являются точками локального минимума, k — точками локального максимума, l — седловыми точками (определение 4. 9. 1).

8. Пусть $S^{n-1}=\{x\in E^n\colon |x|^2=1\}$ — единичная сфера в E^n, пусть гладкая функция $f(x)$, $x\in S^{n-1}$, имеет две точки локального максимума на S^{n-1}. Покажите, что функция f имеет стационарную точку, отличную от упомянутых двух и от точки глобального минимума этой функции на S^{n-1}.

§ 3. Задачи на условный экстремум. Необходимые условия первого порядка

В приложениях задачи на безусловный экстремум встречаются редко. Дело в том, что в практических задачах переменные, как правило, не могут быть совершенно произвольными и должны удовлетворять некоторым дополнительным условиям, выражающим, например, условия неотрицательности тех или иных переменных, условия ограниченности используемых ресурсов, ограничения на параметры конструкции системы, условия нормировки и т. п. Иначе говоря, переменные $x=(x^1,\ldots,x^n)$ должны принадлежать некоторому заданному множеству X из E^n. Тогда, чтобы подчеркнуть, что экстремум функции ищется при условии $x\in X\neq E^n$, часто говорят о задаче на *условный экстремум*.

В таких задачах мы также будем различать точки локального минимума и максимума. Напомним, что точка $v\in X$ называется *точкой локального минимума [максимума]* функции $f(x)$ на X, если существует такая ε-окрестность $O(v,\varepsilon)=\{x\in E^n\colon |x-v|<\varepsilon\}$ точки v, что $f(x)\geqslant f(v)$ $[f(x)\leqslant f(v)]$ для $\forall x\in X\cap O(v,\varepsilon)$. Если функция $f(x)$ дважды гладкая на X и точка v ее локального минимума [максимума] является внутренней точкой множества X, то необходимо $f'(v)=0$, $f''(v)\geqslant 0$ $[f''(v)\leqslant 0]$. Эти условия доказываются точно также, как в теореме 2.1. Однако, если v — граничная точка множества X, то такие условия, вообще говоря, не имеют места. Так, например,

функция $f(x)=-x^2$ на отрезке $X=\{x\in E^1\colon 1\leqslant x\leqslant 2\}$ имеет глобальный минимум в точке $v=2$, но $f'(2)=-4$, $f''(2)=-2$. Следовательно, условия экстремума в задачах на условный экстремум должны иметь другую форму, чем в теореме 2.1. В этом параграфе будут сформулированы необходимые условия экстремума, основанные на правиле множителей Лагранжа.

1. Начнем с классической задачи на условный экстремум, традиционно рассматриваемой в математическом анализе [327; 350; 352; 534; 768]: найти экстремумы функции $f(x)$ при условии, что

$$x\in X=\{x\in E^n\colon g_1(x)=0,\ldots,g_s(x)=0\}. \tag{1}$$

Здесь предполагается, что функции $f(x)$, $g_i(x)$ определены на всем пространстве E^n. Ограничения $g_i(x)=0$, $i=1,\ldots,s$, принято называть *ограничениями типа равенств*.

В тех случаях, когда систему уравнений (1) удается преобразовать к эквивалентному виду

$$x^1=\varphi_1(x^{p+1},\ldots,x^n),\ldots,\ x^p=\varphi_p(x^{p+1},\ldots,x^n), \tag{2}$$

выразив из (1) какие-то p переменных через остальные, рассматриваемую задачу на условный экстремум можно свести к задаче на безусловный экстремум функции $g(x^{p+1},\ldots,x^n)=f(\varphi_1(x^{p+1},\ldots,x^n),\ldots,\varphi_p(x^{p+1},\ldots,x^n),x^{p+1},\ldots,x^n)$ переменных $(x^{p+1},\ldots,x^n)\in E^{n-p}$, которую можно исследовать по описанной в § 2 схеме. Однако этот подход имеет ограниченное применение из-за того, что явное выражение вида (2) одной группы переменных через остальные удается получить лишь в редких случаях.

Более общий подход к исследованию задачи поиска экстремума функции $f(x)$ на множестве (1) дает правило множителей Лагранжа. Это правило заключается в следующем. Вводится функция

$$\mathcal{L}(x,\overline{\lambda})=\lambda_0 f(x)+\sum_{j=1}^{s}\lambda_j g_j(x) \tag{3}$$

переменных $x=(x^1,\ldots,x^n)\in E^n$, $\overline{\lambda}=(\lambda_0,\lambda_1,\ldots,\lambda_s)\in E^{s+1}$, называемая *функцией Лагранжа*. Имеет место

Теорема 1 (правило множителей Лагранжа). *Пусть x_* — точка локального минимума функции $f(x)$ на множестве* (1), *пусть функции $f(x)$, $g_i(x)$, $i=1,\ldots,s$, непрерывно дифференцируемы в окрестности точки x_*. Тогда существуют числа $\lambda_0^*,\ldots,\lambda_s^*$, называемые множителями Лагранжа, такие, что*

$$\overline{\lambda}^*=(\lambda_0^*,\ldots,\lambda_s^*)\neq 0,\quad \lambda_0^*\geqslant 0,$$
$$\left.\frac{\partial\mathcal{L}(x,\overline{\lambda}^*)}{\partial x^i}\right|_{x=x_*}=\lambda_0^*\frac{\partial f(x_*)}{\partial x^i}+\sum_{j=1}^{s}\lambda_j^*\frac{\partial g_j(x_*)}{\partial x^i}=0,\quad i=1,\ldots,n. \tag{4}$$

Таким образом, в теореме 1 утверждается, что всякая точка локального минимума x_* является стационарной точкой функции Лагранжа $\mathcal{L}(x,\overline{\lambda})$ при некотором, подходящим образом выбранном, нетривиальном наборе множителей $\overline{\lambda}$.

Доказательство. Условие (4) означает, что градиенты $f'(x_*),g_1'(x_*),\ldots,g_s'(x_*)$ линейно зависимы. Допустим противное: пусть это не так, пусть

эти векторы линейно независимы. Тогда $s+1 \leqslant n$. В случае $s+1 < n$ возьмем какие-либо векторы $e_{s+1}, \ldots, e_{n-1}$ так, чтобы система $f'(x_*), g_1'(x_*), \ldots, g_s'(x_*), e_{s+1}, \ldots, e_{n-1}$ образовала базис в E^n. Введем вектор-функцию $F(x, t) = (f_0(x, t), f_1(x, t), \ldots, f_{n-1}(x, t))$, где $f_0(x, t) = f(x) - f(x_*) + t$, $f_i(x, t) = g_i(x)$, $i = 1, \ldots, s$, $f_i(x, t) = \langle e_i, x - x_* \rangle$, $i = s+1, \ldots, n-1$, $(x, t) \in E^{n+1}$. Рассмотрим систему n уравнений

$$F(x, t) = (f_0(x, t), f_1(x, t), \ldots, f_{n-1}(x, t)) = 0 \tag{5}$$

относительно n неизвестных $x = (x^1, \ldots, x^n)$. Для ее исследования применим теорему о неявных функциях [327; 350; 352; 534; 768]. Заметим, что $F(x_*, 0) = 0$. Далее, функции $f_i(x, t)$ непрерывно дифференцируемы в окрестности точки $(x_*, 0) \in E^{n+1}$, причем $f_{0x}(x_*, 0) = f'(x_*)$, $f_{ix}(x_*, 0) = g_i'(x_*)$, $i = 1, \ldots, s$, $f_{ix}(x_*, 0) = e_i$, $i = s+1, \ldots, n-1$. Это значит, что в точке $(x_*, 0)$ якобиан системы функций $F(x, t) = (f_i(x, t), i = 0, \ldots, n-1)$, представляющий собой определитель квадратной матрицы со строками $f'(x_*), g_1'(x_*), \ldots, g_s'(x_*), e_{s+1}, \ldots, e_{n-1}$, образующими базис в E^n, отличен от нуля. Тогда по теореме о неявных функциях система (5) имеет решение при каждом t, $|t| \leqslant t_0$, где t_0 — достаточно малое положительное число, или, точнее, существует непрерывно-дифференцируемая вектор-функция $x = x(t) = (x^1(t), \ldots, x^n(t))$, такая, что при всех t, $|t| \leqslant t_0$

$$x(0) = x_*, \quad f(x(t)) = f(x_*) - t, \quad g_i(x(t)) = 0, \quad i = 1, \ldots, s,$$
$$\langle e_i, x(t) - x_* \rangle = 0, \quad i = s+1, \ldots, n-1.$$

Это значит, что $x(t) \in X$, $0 < t < t_0$. Однако $f(x(t)) = f(x_*) - t < f(x_*)$ $\forall t \in (0, t_0)$, что противоречит тому, что x_* — точка локального минимума. Тем самым доказано, что векторы $f'(x_*), g_1'(x_*), \ldots, g_s'(x_*)$ линейно зависимы, т. е. существует набор $\overline{\lambda}^* \neq 0$, что $\mathcal{L}_x(x_*, \overline{\lambda}^*) = 0$. Тогда для набора $(-\overline{\lambda}^*) \neq \neq 0$ также $\mathcal{L}_x(x_*, -\overline{\lambda}^*) = 0$. Поэтому можем считать, что $\lambda_0^* \geqslant 0$. Теорема 1 доказана. □

Условие (4), в котором используются лишь первые производные функций $f(x)$, $g_i(x)$, $i = 1, \ldots, s$, принято называть *необходимым условием первого порядка*. Разумеется, в (4) неравенство $\lambda_0^* \geqslant 0$ вполне можно было бы заменить на $\lambda_0^* \leqslant 0$ — это принципиального значения не имеет. Однако, следуя традициям, восходящим к Вейерштрассу, в литературе по экстремальным задачам для определенности обычно берут $\lambda_0^* \geqslant 0$, относя это неравенство к характеристике точки локального минимума.

Из теоремы 1 следует, что подозрительными на локальный минимум могут быть лишь те точки $x \in E^n$, для которых существуют множители Лагранжа $\overline{\lambda} = (\lambda_0, \ldots, \lambda_s)$ такие, что пара $(x, \overline{\lambda})$ является решением системы

$$\lambda_0 f'(x) + \sum_{j=1}^{s} \lambda_j g_j'(x) = 0, \quad g_i(x) = 0, \quad i = 1, \ldots, s, \quad \overline{\lambda} \neq 0, \quad \lambda_0 \geqslant 0. \tag{6}$$

Пусть v — какая-либо фиксированная точка локального минимума функции $f(x)$ на множестве (1). Множество всех $\overline{\lambda}$, для которых пара $(x = v, \overline{\lambda})$ является решением системы (6), будем называть *множителями Лагранжа, соответствующими точке* v, и обозначать через $\Lambda = \Lambda(v)$. Нетрудно видеть, что если $(v, \overline{\lambda})$ — решение системы (6), то $(v, \alpha\overline{\lambda})$ при любом $\alpha > 0$

также является решением этой системы. Отсюда следует, что множество $\Lambda(v)$ является конусом, будем его называть *конусом Лагранжа* точки v.

Замечание 1. Поскольку всякая точка $x^* \in X$ локального максимума функции $f(x)$ является точкой локального минимума функции $(-f(x))$, то, применяя теорему 1 к функции $(-f(x))$, получим, что для точки x^* необходимо существуют множители Лагранжа $\overline{\lambda}^* = (\lambda_0^*, \ldots, \lambda_s^*)$ такие, что $\mathcal{L}_x(x^*, \overline{\lambda}^*) = 0$, $\overline{\lambda}^* \neq 0$, $\lambda_0^* \leqslant 0$. Отсюда следует, что подозрительными на локальный максимум функции $f(x)$ на множестве (1) могут быть лишь те точки $x \in E^n$, для которых существуют множители $\overline{\lambda} = (\lambda_0^*, \ldots, \lambda_s^*)$ такие, что пара $(x, \overline{\lambda})$ является решением системы

$$\lambda_0 f'(x) + \sum_{j=1}^{s} \lambda_j g_j'(x) = 0, \quad g_i(x) = 0, \quad i = 1, \ldots, s, \quad \overline{\lambda} \neq 0, \quad \lambda_0 \leqslant 0. \tag{7}$$

Множество всех $\overline{\lambda}$, для которых пара $(x = v, \overline{\lambda})$ является решением системы (7), также является конусом, который мы будем обозначать через $\Lambda^-(v)$ и называть *конусом Лагранжа* точки v локального максимума. Отличие этого конуса от конуса, соответствующего точке локального минимума в том, что здесь у всех точек $\overline{\lambda} \in \Lambda^-(v)$ координата $\lambda_0 \leqslant 0$. Такое соглашение о знаке λ_0, несмотря на всю свою условность, ниже позволит нам единообразно формулировать необходимые и достаточные условия экстремума второго порядка, как-то упорядочит процедуру исследования точек экстремума.

Из теоремы 1 и замечания 1 следует, что в точке v экстремума функции $f(x)$ на множестве (1) конус Лагранжа не может быть пустым — если этот конус пуст, то такая точка v заведомо не может быть точкой экстремума. Из того, что множество множителей Лагранжа является конусом, условие $\overline{\lambda} \neq 0$ в (6), (7) можно заменить каким-либо условием нормировки, взяв, например, $|\overline{\lambda}|^2 = \sum_{i=0}^{s} \lambda_i^2 = 1$. Вместо отдельного исследования систем (6), (7) можно рассмотреть одну систему

$$\mathcal{L}_x(x, \overline{\lambda}) = \lambda_0 f'(x) + \sum_{j=1}^{s} \lambda_j g_j'(x) = 0, \quad g_i(x) = 0, \quad i = 1, \ldots, s,$$
$$(\lambda_0, \lambda_1, \ldots, \lambda_s) \neq 0, \tag{8}$$

последовательно полагая в ней $\lambda_0 = 1$, $\lambda_0 = -1$ и $\lambda_0 = 0$, $\sum_{i=1}^{s} \lambda_i^2 = 1$.

Система (8) с учетом условий нормировки представляет собой систему $n+s+1$ уравнений с $n+s+1$ неизвестными $(x, \overline{\lambda}) = (x^1, \ldots, x^n, \lambda_0, \lambda_1, \ldots, \lambda_s)$. Решив ее, мы найдем точки $x = v$ множества (1), подозрительные на экстремум, и соответствующие им множители Лагранжа $\overline{\lambda} = \overline{\lambda}(v)$. Для выяснения того, будет ли в этих точках в действительности реализовываться локальный минимум или максимум, нужно провести дополнительное изучение свойств функции $f(x)$ в окрестности точки v с учетом ограничений (1). Здесь могут быть привлечены геометрические, физические и т. п. соображения. При выполнении условий теорем Вейерштрасса из § 1 можно быть уверенным, что хотя бы одна из найденных точек v окажется точкой глобального минимума или максимума. Для выяснения характера экстремума точек v могут быть привлечены вторые производные функции Лагранжа по переменной x — об этом речь пойдет в § 5.

Изложенная схема поиска экстремума функции на множестве (1) составляет суть правила (метода) множителей Лагранжа. Для иллюстрации этого правила рассмотрим несколько примеров.

Пример 1. Пусть требуется на n-мерной единичной сфере $X = \{x \in E^n: |x|^2 = \langle x, x\rangle = 1\}$ найти точку, сумма квадратов расстояний от которой до p данных точек $x_1, \ldots, x_p$ была бы минимальной. Иначе говоря, нужно минимизировать функцию $f(x) = \sum_{i=1}^{p} |x - x_i|^2$ при условии $g(x) = \langle x, x\rangle - 1 = 0$. Как и в примере 2.2, здесь удобнее пользоваться следующим представлением функции $f(x) = p|x|^2 - 2p\langle x, x_0\rangle + \sum_{i=1}^{p} |x_i|^2$, где $x_0 = \frac{1}{p}\sum_{i=1}^{p} x_i$. Составим функцию Лагранжа этой задачи: $\mathcal{L}(x, \bar\lambda) = \lambda_0 f(x) + \lambda\,(\langle x, x\rangle - 1)$. Система (8) имеет вид

$$\mathcal{L}_x(x, \bar\lambda) = 2p\lambda_0(x - x_0) + 2\lambda x = 0, \quad \langle x, x\rangle = 1, \quad (\lambda_0, \lambda) \neq 0. \qquad (9)$$

При $\lambda_0 = 0$ эта система, очевидно, не имеет решения. Поэтому здесь можем принять $\lambda_0 = 1$ или $\lambda_0 = -1$. При $x_0 \neq 0$ из системы (9) получаем две точки: $v_1 = \frac{x_0}{|x_0|}$ и $v_2 = -\frac{x_0}{|x_0|}$, подозрительные на экстремум. Соответствующие этим точкам множители Лагранжа нетрудно выписать явно. Однако они нам ниже явно не понадобятся, важен лишь факт их существования. Поскольку X компактное множество, функция $f(x)$ непрерывна на X, то согласно теоремам 1.1, 1.4 эта функция достигает на X своего глобального минимума и максимума. Но точки глобального экстремума, конечно же, удовлетворяют системе (9). Но система (9) при $x_0 \neq 0$ имеет всего два решения v_1 и v_2. Следовательно, одна из этих точек является точкой глобального минимума, другая — точкой глобального максимума. Вычислив и сравнив значения $f(v_1), f(v_2)$, нетрудно убедиться, что $v_1 = \frac{x_0}{|x_0|}$ — точка глобального минимума со значением $f(v_1) = f_* = p - 2\left|\sum_{i=1}^{p} x_i\right| + \sum_{i=1}^{p} x_i^2$, $v_2 = -\frac{x_0}{|x_0|}$ — точка глобального максимума со значением $f(v_2) = f^* = p + 2\left|\sum_{i=1}^{p} x_i\right| + \sum_{i=1}^{p} x_i^2$. Поскольку при $x_0 \neq 0$ у функции $f(x)$ других точек экстремума на X нет, то $f(v_1) < f(x)$ $\forall x \in X$, $x \neq v_1$, и $f(x) < f(v_2)$ $\forall x \in X$, $x \neq v_2$, т. е. экстремумы строгие.

Рассмотрим случай $x_0 = 0$. Тогда системе (9) удовлетворяют все точки v, для которых $|v| = 1$. Это значит, что из необходимых условий экстремума (9) при $x_0 = 0$ нам не удалось извлечь никакой полезной информации — все точки единичной сферы как были, так и остались подозрительными на экстремум. Однако нетрудно убедиться, что при $x_0 = 0$ $f(x) = p + \sum_{i=1}^{p} x_i^2 = \text{const}$ $\forall x \in X$, и рассматриваемая задача стала тривиальной: можно сказать, что все точки $x \in X$ являются точкой абсолютного минимума (или максимума).

Пример 2. Определим точки экстремума функции $f(u) = x$ на множестве $X = \{u = (x, y) \in E^2: g(x) = x^3 - y^2 = 0\}$. Функция Лагранжа здесь равна $\mathcal{L}(u, \bar\lambda) = \lambda_0 x + \lambda(x^3 - y^2)$. Система (8) запишется в виде:

$$\lambda_0 + 3\lambda x^2 = 0, \quad -2\lambda y = 0, \quad x^3 - y^2 = 0, \quad (\lambda_0, \lambda) \neq 0.$$

Из этой системы находим единственную точку $v = (0, 0)$, подозрительную на экстремум. Ей соответствует конус Лагранжа $\Lambda(v) = \{\bar\lambda = (\lambda_0 = 0, \lambda): \forall \lambda \neq 0\}$. Нетрудно видеть, что точка $v = 0$ в этой задаче является точкой глобального минимума. В самом деле, из равенства $x^3 - y^2 = 0$ следует, что $f(u) = x = (y^2)^{1/3} \geqslant 0 = f(0) = f_*$ $\forall u \in X$. Здесь $f^* = +\infty$.

2. Изложим правило множителей Лагранжа для задачи поиска точек экстремума функции $f(x)$ на множестве, имеющем более общий вид:

$$X = \{x \in E^n: g_1(x) \leqslant 0, \ldots, g_m(x) \leqslant 0, g_{m+1}(x) = 0, \ldots, g_s(x) = 0\}, \qquad (10)$$

где предполагается, что функции $f(x), g_1(x), \ldots, g_s(x)$ определены на всем пространстве E^n. Ограничения $g_i(x) = 0$, $i = m+1, \ldots, s$, как и в (1), будем называть *ограничениями типа равенств*, а ограничения $g_i(x) \leqslant 0$, $i = 1, \ldots, m$ — *ограничениями типа неравенств*. В (10) не исключаются возможности, когда отсутствуют ограничения типа равенств ($s = m$) или типа равенств ($m = 0$); при $s = m = 0$ множество $X = E^n$ получаем задачу на безусловный экстремум из § 2. Для исследования задачи поиска экстремума функции $f(x)$ на множестве (10) введем функцию Лагранжа

$$\mathcal{L}(x, \bar\lambda) = \lambda_0 f(x) + \sum_{j=1}^{s} \lambda_j g_j(x)$$

переменных $x \in E^n$, $\bar\lambda \in E^{s+1}$, внешне ничем не отличающуюся от функции (3), но здесь, оказывается, достаточно ограничиться рассмотрением лишь неотрицательных множителей $\lambda_1, \ldots, \lambda_m$, соответствующих ограничениям типа неравенств.

Теорема 2. *Пусть x_* — точка локального минимума функции $f(x)$ на множестве* (10), *функции $f(x), g_1(x), \ldots, g_m(x)$ дифференцируемы в точке x_*, функции $g_{m+1}(x), \ldots, g_s(x)$ непрерывно дифференцируемы в некоторой окрестности точки x_*. Тогда существуют множители Лагранжа $\bar\lambda^* = (\lambda_0^*, \ldots, \lambda_s^*)$ такие, что*

$$\bar\lambda^* \neq 0, \quad \lambda_0^* \geqslant 0, \lambda_1^* \geqslant 0, \ldots, \lambda_m^* \geqslant 0, \quad \mathcal{L}_x(x_*, \bar\lambda^*) = 0, \quad \lambda_i^* g_i(x_*) = 0, \quad i = 1, \ldots, m.$$

Отметим, что теорему 2 в литературе иногда называют *теоремой Каруша — Джона* [234; 586]. Доказательство этой теоремы требует развития некоторого математического аппарата, и оно будет ниже проведено двумя способами для несколько более общих множеств, чем (10) (см. § 4.8, § 5.16). Из теоремы 2 следует, что точками локального минимума функции $f(x)$ на множестве (10) могут быть лишь те точки $x \in E^n$, для которых существуют множители Лагранжа $\bar\lambda = (\lambda_0, \ldots, \lambda_s)$, такие, что пара $(x, \bar\lambda)$ является решением системы

$$\lambda_0 f'(x) + \sum_{j=1}^{s} \lambda_j g_j'(x) = 0, \quad \lambda_i g_i(x) = 0, \quad g_i(x) \leqslant 0, \quad i = 1, \ldots, m,$$

$$g_i(x) = 0, \quad i = m+1, \ldots, s, \quad \bar\lambda \neq 0, \quad \lambda_0 \geqslant 0, \lambda_1 \geqslant 0, \ldots, \lambda_m \geqslant 0. \qquad (11)$$

Пусть v — какая-либо фиксированная точка локального минимума функции $f(x)$ на множестве (10). Множество всех точек $\bar\lambda$, для которых пара

$(x=v, \overline{\lambda})$ является решением системы (11), будем называть *множителями Лагранжа*, соответствующими точке v, и будем обозначать через $\Lambda=\Lambda(v)$. Нетрудно видеть, что если $(v, \overline{\lambda})$ — решение системы (11), то $(v, \alpha\overline{\lambda})$ при всех $\alpha>0$ также является решением этой системы, так что $\Lambda(v)$ — конус. Этот конус, как и в случае множества (1), будем называть *конусом Лагранжа* точки v.

Равенства $\lambda_i g_i(x)=0$, $i=1, \ldots, m$, из (11) принято называть *условиями дополняющей нежесткости*. Если $g_i(v)<0$ при некотором i, $1\leqslant i\leqslant m$, то из условия дополняющей нежесткости следует, что координата $\lambda_i=0$ у всех $\overline{\lambda}\in\Lambda(v)$; с другой стороны, если у некоторого набора $\overline{\lambda}\in\Lambda(v)$ оказалось, что $\lambda_i>0$ при некотором i, $1\leqslant i\leqslant m$, то соответствующее $g_i(v)=0$. Ограничение $g_i(x)\leqslant 0$ называется *активным в точке* v, если $g_i(v)=0$, и *пассивным (неактивным) в точке* v, если $g_i(v)<0$. Ограничения $g_i(x)=0$, $i=m+1, \ldots, s$, в любой точке $v\in X$, конечно, являются активными.

Замечание 2. Как и выше (см. замечание 1), нетрудно убедиться, что точки v локального максимума функции $f(x)$ на множестве (10) и соответствующие им множители Лагранжа $\overline{\lambda}$ являются решением системы

$$\lambda_0 f'(x)+\sum_{j=1}^{s}\lambda_j g_j'(x)=0,\quad \lambda_i g_i(x)=0,\quad g_i(x)\leqslant 0,$$
$$i=1, \ldots, m,\quad g_i(x)=0,\quad i=m+1, \ldots, s,$$
$$\overline{\lambda}=(\lambda_0, \lambda_1, \ldots, \lambda_s)\neq 0,\quad \lambda_0\leqslant 0,\ \lambda_1\geqslant 0,\ \ldots,\ \lambda_m\geqslant 0. \quad (12)$$

Множество всех $\overline{\lambda}$, для которых пара $(x=v, \overline{\lambda})$ является решением системы (12), будем обозначать через $\Lambda^-(v)$. Множество $\Lambda^-(v)$ и здесь будет конусом, и его также будем называть конусом Лагранжа. Отличие этого конуса от конуса, соответствующего точке локального минимума в том, что здесь у всех точек $\overline{\lambda}\in\Lambda^-(v)$ координата $\lambda_0\leqslant 0$. Такие соглашения о знаке λ_0, как уже отмечалось в замечании 1, несмотря на всю свою условность, позволят нам несколько унифицировать дальнейшее изложение.

Так как $\Lambda(v)$, $\Lambda^-(v)$ конусы, то в системах (11), (12) условие $\overline{\lambda}\neq 0$ можно заменить каким-либо условием нормировки, взяв, например, $|\overline{\lambda}|^2=\sum_{i=0}^{s}\lambda_i^2=1$. Вместо отдельного исследования систем (11), (12) можно рассмотреть одну систему

$$\mathcal{L}_x(x, \overline{\lambda})=\lambda_0 f'(x)+\sum_{j=1}^{s}\lambda_j g_j'(x)=0,\quad \lambda_i g_i(x)=0,\quad g_i(x)\leqslant 0,$$
$$i=1, \ldots, m,\quad g_i(x)=0,\quad i=m+1, \ldots, s,$$
$$\overline{\lambda}=(\lambda_0, \lambda_1, \ldots, \lambda_s)\neq 0,\quad \lambda_1\geqslant 0,\ \ldots,\ \lambda_m\geqslant 0, \quad (13)$$

полагая в ней последовательно $\lambda_0=1$, $\lambda_0=-1$ и $\lambda_0=0$, $\sum_{i=1}^{s}\lambda_i^2=1$. Нетрудно видеть, что в системе (13) с учетом условий нормировки содержится подсистема из $n+s+1$ уравнений с $n+s+1$ неизвестными $(x, \overline{\lambda})=(x^1, \ldots, x^n, \lambda_0, \ldots, \lambda_s)$. Определив решения этой подсистемы и отобрав из них те, которые удовлетворяют остальным условиям (13), получим множество точек v, подозрительных на экстремум и соответствующие им множители Лагранжа $\overline{\lambda}=\overline{\lambda}(v)$. Для дальнейшего выяснения того, будет ли в найденных точках v в самом деле реализовываться экстремум, как и в задачах

с ограничениями (1), нужно провести дополнительное исследование поведения функции $f(x)$ в окрестности точки v с учетом ограничений (10) или попытаться использовать достаточные условия экстремума из § 5.

Правило множителей Лагранжа для поиска точек экстремума функций на множествах вида (10) изложено. Проиллюстрируем его на примерах.

Пример 3. Пусть требуется найти точки экстремума функции $f(x)=\sum_{i=1}^{p}|x-x_i|^2$ из примера 1 на шаре $X=\{x\in E^n\colon |x|^2=\langle x, x\rangle\leqslant 1\}$.

Функция Лагранжа этой задачи

$$\mathcal{L}_x(x, \overline{\lambda})=\lambda_0 f'(x)+\lambda_1(\langle x, x\rangle-1),\quad x\in E^n,\quad \lambda_1\geqslant 0.$$

Ее производная равна $\mathcal{L}_x(x, \overline{\lambda})=2\lambda_0 p(x-x_0)+2\lambda_1 x$, где $x_0=\frac{1}{p}\sum_{i=1}^{p}x_i$. Система (13) имеет вид:

$$(\lambda_0 p+\lambda_1)x=\lambda_0 p x_0,\quad \lambda_1(|x|^2-1)=0,\quad |x|^2\leqslant 1,\quad \overline{\lambda}=(\lambda_0, \lambda_1)\neq 0,\quad \lambda_1\geqslant 0. \quad (14)$$

Нетрудно убедиться, что при $\lambda_0=0$ эта система не имеет решения. Анализируя систему (14) при $\lambda_0=1$ и $\lambda_0=-1$, получим следующие точки, подозрительные на экстремум: $v_1=x_0$ при $|x_0|\leqslant 1$ с соответствующими множителями Лагранжа $\overline{\lambda}_{1,1}=(1, 0)$ и $\overline{\lambda}_{1,2}=(-1, 0)$; $v_2=\frac{x_0}{|x_0|}$ при $|x_0|>1$ с $\overline{\lambda}_2=(\lambda_0=1, \lambda_1=p(|x_0|-1)>0)$; $v_3=\frac{x_0}{|x_0|}$ при $0<|x_0|<1$ с $\overline{\lambda}_3=(\lambda_0=-1, \lambda_1=p(1-|x_0|)>0)$; $v_4=-\frac{x_0}{|x_0|}$ при $|x_0|>0$ с $\overline{\lambda}_4=(\lambda_0=-1, \lambda_1=p(1+|x_0|)>0)$; наконец, при $x_0=0$ подозрительными на экстремум будут все точки v_5 на единичной сфере $|x|=1$ с множителем $\overline{\lambda}_5=(\lambda_0=-1, \lambda_1=p>0)$.

Выясним теперь, будет ли в отобранных точках действительно реализовываться экстремум, анализируя поведение функции $f(x)$ в окрестности этих точек с учетом ограничения $|x|\leqslant 1$. В данной конкретной задаче такой анализ удается провести до конца. Так как $f(x)-f(x_0)=p|x-x_0|^2$ $\forall x\in E^n$ и, тем более, $\forall x\in X$, то ясно, что $v_1=x_0$ является точкой глобального минимума $f(x)$ на X (ср. с примером 2.2). В точке v_2 согласно теореме 2 можно ожидать, что будет локальный минимум. Это ожидание оправдывается и, более того, v_2 — точка глобального минимума. В самом деле, с учетом неравенства $|x_0|>1$ имеем: $f(x)-f(v_2)=p(|x-x_0|+|x_0|-1)(|x-x_0|-|x_0|+1)\geqslant p|x-x_0|(1-|x|)\geqslant 0$ $\forall x\in X$.

В точке v_3, судя по знаку λ_0, может быть локальный максимум. Однако установленное в примере 1 неравенство $f(x)>f(v_3)$, справедливое при всех x, $|x|=1$, говорит о том, что это не так. Следовательно, точка v_3 при $0<|x_0|<1$ не может быть точкой экстремума функции $f(x)$ на X. Далее, для точки $v_4=-\frac{x_0}{|x_0|}$, $|x_0|>0$ имеем $f(x)-f(v_4)=p(|x-x_0|+|x_0|+1)(|x-x_0|-|x_0|-1)\leqslant p(|x-x_0|+|x_0|+1)(|x|-1)\leqslant 0$ $\forall x\in X$. Это значит, что v_4 — точка глобального максимума. Наконец, пусть $x_0=0$, пусть v_5 — произвольная точка такая, что $|v_5|=1$. Тогда $f(x)-f(v_5)=p(|x|^2-1)\leqslant 0$ $\forall x\in X$. Следовательно, при $x_0=0$ все точки единичной сферы являются точками глобального максимума (ср. с примером 1).

Пример 4. Найти точки экстремума функции $f(u)=x$ на множестве $X=\{u=(x, y)\in E^2\colon g_1(u)=-x\leqslant 0,\ g_2(u)=x^2-y\leqslant 0,\ g_3(u)=y-2x^2\leqslant 0\}$.

Здесь $\mathcal{L}(u, \bar{\lambda}) = \lambda_0 x + \lambda_1(-x) + \lambda_2(x^2 - y) + \lambda_3(y - 2x^2)$, $(x, y) \in E^2$, $\lambda_1 \geqslant 0$, $\lambda_2 \geqslant 0$, $\lambda_3 \geqslant 0$. Система (13) имеет вид

$$\lambda_0 - \lambda_1 + 2x(\lambda_2 - 2\lambda_3) = 0, \quad -\lambda_2 + \lambda_3 = 0, \quad \lambda_1(-x) = 0,$$
$$\lambda_2(x^2 - y) = 0, \quad \lambda_3(y - 2x^2) = 0, \quad -x \leqslant 0, \quad x^2 - y \leqslant 0, \quad y - 2x^2 \leqslant 0,$$
$$\lambda_1 \geqslant 0, \quad \lambda_2 \geqslant 0, \quad \lambda_3 \geqslant 0, \bar{\lambda} = (\lambda_0, \ldots, \lambda_3) \neq 0. \quad (15)$$

Допустим, что решением (15) является пара $(v, \bar{\lambda})$, где $v = (x, y)$, $x > 0$. Тогда $0 < x^2 \leqslant y \leqslant 2x^2$, причем хотя бы одно из неравенств $\leqslant$ строгое. Отсюда и из (15) следует, что $\lambda_1 = 0$, $\lambda_2 = \lambda_3 = 0$, $\lambda_0 = 0$, что противоречит условию $\bar{\lambda} \neq 0$. Остается одна возможность, что $v = (x = 0, y)$. Но тогда из предыдущих неравенств имеем $y = 0$. Таким образом, в рассматриваемой задаче подозрительной на экстремум является лишь точка $v = (0, 0)$. Ей соответствуют множители Лагранжа $\bar{\lambda}_1 = (1, 1, \lambda_2 \geqslant 0, \lambda_3 = \lambda_2 \geqslant 0)$, $\bar{\lambda}_2 = (0, 0, \lambda_2 \geqslant \geqslant 0, \lambda_3 = \lambda_2 \geqslant 0)$ (с учетом нормировки $\lambda_0 = 1$ или $\lambda_0 = 0$). Поскольку у всех точек $u = (x, y) \in X$ координата $x \geqslant 0$, то $f(u) = x \geqslant 0 = f(0)$ $\forall u \in X$. Следовательно, $v = 0$ — точка глобального минимума.

Пример 5. Найти точки экстремума функции $f(u) = x + \cos y$ на множестве $X = \{u = (x, y) \in E^2: g(u) = x \leqslant 0\}$.

Здесь $\mathcal{L}(u, \bar{\lambda}) = \lambda_0(x + \cos y) + \lambda_1 x$, $(x, y) \in E^2$, $\lambda_1 \geqslant 0$. Система (13) имеет вид:

$$\lambda_0 + \lambda_1 = 0, \quad \lambda_0 \sin y = 0, \quad \lambda_1 x = 0, \quad x \leqslant 0, \quad \lambda_1 \geqslant 0, \quad \bar{\lambda} = (\lambda_0, \lambda_1) \neq 0.$$

Из этой системы определяется бесконечно много подозрительных на экстремум точек $v_k = (x = 0, y = \pi k)$, $k = 0, \pm 1, \pm 2, \ldots$; всем им соответствует один и тот же набор множителей Лагранжа $\bar{\lambda} = (-1, 1)$ (с учетом нормировки). Отсюда видно, что в этой задаче точек локального минимума нет. В точках $v_{2m} = (0, 2\pi m)$, $m = 0, \pm 1, \ldots$, реализуется глобальный максимум, так как $f(u) = x + \cos y \leqslant 1 = f(v_{2m})$ $\forall x \in X$. Далее, $f(v_{2m+1}) = -1$, причем в любой сколь угодно малой ε-окрестности точки v_{2m+1} нетрудно указать точки $u \in X$, в которых $f(u) < -1$ (например, $u = (x, y = \pi(2m+1), -\varepsilon < x < 0)$ и $f(u) > -1$ (например, $u = (x = 0, y)$, $0 < |y - \pi(2m+1)| < \varepsilon$). Следовательно, в точках v_{2m+1}, $m = 0, \pm 1, \ldots$ экстремума нет.

Упражнения

1. Пользуясь правилом множителей Лагранжа, найти точки экстремума функции $f(u)$ на множестве X, если:

а) $f(u) = x^2 + y^2 + z^2$, $X = \{u = (x, y, z) \in E^3: x + y + z = 1\}$, или $X = \{u \in E^3: x \geqslant 0, y \geqslant \geqslant 0, z \geqslant 0, x + y + z = 1\}$, или $X = \{u \in E^3: x - 2y + 3z = 1\}$, или $X = \{u \in E^3: x + y + 2z = 2, x - 2y + z = -2\}$; рассмотреть множества, полученные из X заменой ограничений типа равенств на ограничения типа $\leqslant$ или $\geqslant$;

б) $f(u) = x$, $X = \{u = (x, y) \in E^2: x^2 + y^2 \leqslant 1, x^2 \leqslant y, x + y \leqslant 0\}$, или $X = \{u = (x, y) \in E^2: x^2 + y2 \leqslant 1, x^3 + y^3 = 1\}$, или $X = \{u = (x, y) \in E^2: x^2 + y^2 \leqslant 1, (-x)^3 \leqslant y \leqslant x^3\}$;

в) $f(u) = \sin(x + y) - \sin x - \sin y$, $X = \{u = (x, y) \in E^2: x \leqslant 0, y \geqslant o, x + y \leqslant 2\pi\}$.

2. Среди всех вписанных в данный круг радиуса R треугольников найти тот, площадь которого наибольшая [периметр которого наибольший].

3. Среди всех параллелепипедов, имеющих ребра данной длины, найти параллелепипед наибольшего объема.

4. Среди треугольных пирамид с данным основанием и высотой найти ту, которая имеет наименьшую боковую поверхность.

5. Пусть $\Delta = \Delta(x_1, \ldots, x_n)$ — определитель матрицы $(x_1, \ldots, x_n)$, столбцами которой являются вектор-столбцы x_i с координатами $x_i^1, \ldots, x_i^n$, $i = 1, \ldots, n$. Найти наибольшее и наименьшее значение величины определителя Δ при условии, что $|x_i| = a_i$, где a_i — заданные положительные числа, $i = 1, \ldots, n$. Доказать неравенство Адамара $|\Delta| \leqslant |x_1| \cdot |x_2| \cdot \ldots \cdot |x_n|$. Дать геометрическую интерпретацию задачи при $n = 2, 3$ (ср. с упражнением 3) [352, ч. I, с. 554–557].

6. Найти наименьшее и наибольшее значение квадратичной формы $f(x) = \langle Ax, x \rangle$ при условии $x \in X = \{x \in E^n: \langle x, x \rangle = 1\}$, где A — симметрическая матрица. Показать, что величины $f_* = \min\limits_X f(x)$ и $f^* = \max\limits_X f(x)$ представляют собой соответственно наименьшее и наибольшее собственное число матрицы A ([353, с. 209]).

7. Найти точки экстремума функций $f(u) = x + y$, $f(u) = |x| + |y - 1|$, $f(u) = x^2 + 2y^2$ на множествах X, где $X = \{u = (x, y) \in E^2: 0 \leqslant x \leqslant 1, 0 \leqslant y \leqslant 1\}$ или $X = \{u = (x, y) \in E^2: x - y^2 \geqslant 0, x^2 + y^2 \leqslant 1\}$, или $X = \{u = (x, y) \in E^2: x \geqslant 0, y \geqslant 0, ax + by = 1\}$, числа $a \geqslant 0$, $b \geqslant 0$. Указание: нарисовать пересечения графиков линий уровня $f(x) \equiv \text{const}$ со множеством X.

8. Пусть $X = \{x \in E^n: \langle c, x \rangle = 1\}$ или $X = \{x \in E^n: \langle c, x \rangle \leqslant 1\}$, где $c \in E^n$, $c \neq 0$. Найти точку $x \in X$, сумма квадратов расстояний от которой до p данных точек $x_1, x_2, \ldots, x_p \in E^n$ была бы минимальной [максимальной] (ср. с примерами 1,3).

9. Задачи из примеров 1, 3 и упражнения 8 исследовать геометрически (при $n = 2, 3$), используя тот факт, что поверхности уровня $f(x) = \text{const}$, где $f(x) = \sum\limits_{i=1}^{p} |x - x_i|^2$, являются сферами $|x - x_0| = R$ с центром $x_0 = \frac{1}{p} \sum\limits_{i=1}^{p} x_i$.

10. Пусть $f(u) = x$, $X = \{u = (x, y) \in E^2: g_1(u) = x^2 - y = 0, g_2(u) = x^2 + y = 0, g_3(x) = x = 0\}$. Дайте описание конуса Лагранжа точки $v = (0, 0)$, выясните характер экстремума в этой точке.

11. Приведите пример функции $f(x)$ и множества X, таких, что $f(x)$ не имеет ни одной точки экстремума на X, а система (13) имеет бесконечно много решений. Начните с функции $f(u) = x + y$ на множестве $X = \{u = (x, y) \in E^2: g(u) = x^2 - y^2 = 0\}$.

§ 4. Необходимые условия экстремума второго порядка

Для более тонкого анализа точек экстремума используются *необходимые условия второго порядка*. Так называются условия, в формулировке которых используются вторые производные функций, входящих в постановку задачи. С помощью этих условий проводят дополнительный отбор и суживают множество точек, подозрительных на экстремум, выделенных с помощью необходимых условий первого порядка. Для задач на безусловный экстремум необходимые условия второго порядка мы уже формулировали выше (теорема 2.1). Перейдем к формулировке таких условий для задач на условный экстремум.

1. Как в предыдущем параграфе, изложение начнем с классической задачи поиска экстремума на множестве

$$X = \{x \in E^n: g_1(x) = 0, \ldots, g_s(x) = 0\}, \quad (1)$$

задаваемом ограничения типа равенств. Известные в литературе необходимые условия второго порядка в задачах на условный экстремум обычно формулируются при дополнительном требовании нормальности точки, подозрительной на экстремум (см., например, [14; 670; 721]).

Определение 1. Точка v называется *нормальной точкой множества* (1), если $v \in X$ и векторы $g_1'(v), \ldots, g_s'(v)$ линейно независимы.

Требование нормальности точки в литературе часто называют *условием Люстерника*. Посмотрим, как устроен конус Лагранжа $\Lambda(v)$ в нормальной точке локального минимума. Перепишем уравнение $\mathcal{L}_x(v, \bar\lambda) = 0$ из системы (3.6) в виде:

$$\lambda_1 g_1'(v) + \ldots + \lambda_s g_s'(v) = -\lambda_0 f'(v). \qquad (2)$$

По определению точка $\bar\lambda = (\lambda_0, \ldots, \lambda_s)$ принадлежит конусу $\Lambda(v)$ тогда и только тогда, когда $\bar\lambda$ является решением уравнения (2) и $\bar\lambda \neq 0$, $\lambda_0 \geqslant 0$. При каждом фиксированном λ_0 (2) представляет собой систему линейных алгебраических уравнений относительно неизвестных $\lambda = (\lambda_1, \ldots, \lambda_s)$ с матрицей $\{g_1'(v), \ldots, g_s'(v)\}$, ранг которой в нормальной точке v равен s, причем $s \leqslant n$. Отсюда следует [192; 353], что при каждом λ_0 система (2) имеет единственное решение λ, и оно представимо в виде $\lambda = \lambda_0 \mu$, где μ — решение этой системы при $\lambda_0 = 1$. Это значит, что конус Лагранжа в нормальной точке v есть открытый луч $\Lambda(v) = \{\bar\lambda \in E^{s+1}: \bar\lambda = \lambda_0(1, \mu), \lambda_0 > 0\}$ с направляющим вектором $(1, \mu)$.

Теорема 1. *Пусть функции $f(x), g_1(x), \ldots, g_s(x)$ дважды непрерывно дифференцируемы в некоторой окрестности точки v локального минимума функции $f(x)$ на множестве (1), пусть v — нормальная точка этого множества, $\Lambda(v)$ — конус Лагранжа точки v. Тогда для любой точки $\bar\lambda \in \Lambda(v)$*

$$\langle \mathcal{L}_{xx}(v, \bar\lambda)h, h\rangle \geqslant 0 \quad \forall h \in K(v) = \{h \in E^n: \langle g_i'(v), h\rangle = 0, i = 1, \ldots, s\}. \qquad (3)$$

Конус $K(v)$, введенный в (3), называют *конусом критических направлений множества (1) в точке v*. Этот конус непуст, так как он всегда содержит точку $h = 0$. В нормальной точке v при $s = n$ конус $K(v)$ состоит из единственной точки $h = 0$, а при $0 < s < n$ он является подпространством размерности $n - s$. Так как функция $\mathcal{L}(x, \bar\lambda)$ однородна по переменной $\bar\lambda$, т. е. $\mathcal{L}(x, \alpha\bar\lambda) = \alpha\mathcal{L}(x, \bar\lambda)$ $\forall\alpha$, а $\Lambda(v)$ — луч с направляющим вектором $(1, \mu)$, то условие (3) достаточно проверить при $\bar\lambda = (1, \mu)$. Для нормальной точки локального максимума теорема 1 сохраняется, нужно лишь в ее формулировке конус $\Lambda(v)$ заменить на конус $\Lambda^-(v) = \{\bar\lambda \in E^{s+1}: \bar\lambda = \lambda_0(1, \mu), \lambda_0 < 0\}$.

Для иллюстрации теоремы 1 приведем пример.

Пример 1. Пусть $f(x) = (x^1)^2 - (x^2)^2$, $X = \{x = (x^1, x^2) \in E^2: g(x) = x^1 = 0\}$. Тогда функция Лагранжа $\mathcal{L}(x, \bar\lambda) = \lambda_0((x^1)^2 - (x^2)^2) + \lambda_1 x^1$, ее производные $\mathcal{L}_x(x, \bar\lambda) = (2\lambda_0 x^1 + \lambda_1, -2\lambda_0 x^2)$, $\mathcal{L}_{xx}(x, \bar\lambda) = \begin{pmatrix} 2\lambda_0 & 0 \\ 0 & -2\lambda_0 \end{pmatrix}$; квадратичная форма $\langle \mathcal{L}_{xx}(x, \bar\lambda)h, h\rangle = 2\lambda_0((h^1)^2 - (h^2)^2)$. Нетрудно видеть, что точка $v = (0, 0)$ нормальна и подозрительна на локальный минимум, ее конус Лагранжа $\Lambda(0) = \{\bar\lambda = \lambda_0(1, 0), \lambda_0 > 0\}$. Применим к ней теорему 1. Здесь конус критических направлений $K(0) = \{h = (h^1, h^2): h^1 = 0\}$. Тогда $\langle \mathcal{L}_{xx}(0, \bar\lambda)h, h\rangle = -2\lambda_0(h^2)^2 < 0$ $\forall h \in K(0)$, $h \neq 0$. Условие (3) не выполняется. Следовательно, точка $v = 0$ не может быть точкой локального минимума функции $f(x)$ на множестве X. Нетрудно убедиться, что точка $v = 0$ с конусом $\Lambda^-(0) = \{\bar\lambda = \lambda_0(1, 0), \lambda_0 < 0\}$ также удовлетворяет условию (3) и претендует на локальный максимум. В этом простом примере ясно, что $v = 0$ — точка глобального максимума $f(x)$ на X.

Теперь откажемся от априорного требования нормальности точки, подозрительной на экстремум.

Определение 2. Точка v называется *анормальной точкой множества* (1), если $v \in X$ и векторы $g_1'(v), \ldots, g_s'(v)$ линейно зависимы, т. е. существуют числа $\lambda_1, \ldots, \lambda_s$ такие, что

$$\lambda_1 g_1'(v) + \ldots + \lambda_s g_s'(v) = 0, \quad \lambda = (\lambda_1, \ldots, \lambda_s) \neq 0. \qquad (4)$$

Сразу отметим следующее интересное свойство анормальных точек. Условие (4) можно переписать в равносильном виде: $0 \cdot f'(v) + \lambda_1 g_1'(v) + \ldots + \lambda_s g_s'(v) = 0$, $\bar\lambda = (\lambda_0 = 0, \lambda) \neq 0$, для любой дифференцируемой функции $f(x)$. Это значит, что набор $\bar\lambda = (0, \lambda) \in \Lambda(v)$, т. е. конус $\Lambda(v) \neq \varnothing$. Следовательно, всякая анормальная точка v множества (1), по сути выражающая лишь некоторое специфическое свойство (4) этого множества, автоматически удовлетворяет необходимым условиям экстремума первого порядка и оказывается подозрительной на экстремум для любой дифференцируемой функции $f(x)$.

Посмотрим, как устроен конус Лагранжа в анормальной точке локального минимума. Оказывается, в такой точке v конус $\Lambda(v) \cup \{0\}$ содержит прямую. В самом деле, если набор $\lambda = (\lambda_1, \ldots, \lambda_s)$ удовлетворяет условию (4), то набор $(-\lambda)$ также удовлетворяет ему. Тогда точки $\bar\lambda_0 = (\lambda_0 = 0, \lambda) \neq 0$ и $(-\bar\lambda_0) = (\lambda_0 = 0, -\lambda) \neq 0$ являются решением системы (2) при $\lambda_0 = 0$ и, следовательно, $\bar\lambda_0, (-\bar\lambda_0) \in \Lambda(v)$. Так как $\Lambda(v)$ конус, то прямая $\bar\lambda(t) = t\bar\lambda_0$, $-\infty < t < +\infty$, с направляющим вектором $\bar\lambda_0 \neq 0$, принадлежит $\Lambda(v) \cup \{0\}$. Верно и обратное: если конус $\Lambda(v) \cup \{0\}$ содержит некоторую прямую $\bar\lambda(t) = t\bar\mu$, $-\infty < t < +\infty$, $\bar\mu = (\mu_0, \ldots, \mu_s) \neq 0$, то v — анормальная точка множества (1). Действительно, тогда $\bar\lambda(1) = \bar\mu$ и $\bar\lambda(-1) = -\bar\mu$ принадлежат $\Lambda(v)$. Поскольку у всех точек $\bar\lambda$ конуса $\Lambda(v)$ координата $\lambda_0 \geqslant 0$, то необходимо $\mu_0 \geqslant 0$, $-\mu_0 \geqslant 0$, так что $\mu_0 = 0$, $\lambda = (\mu_1, \ldots, \mu_s) \neq 0$. Это значит, что условия $\mathcal{L}_x(v, \bar\mu) = 0$, $\bar\mu \neq 0$ из системы (3.6) превращаются в условие (4). Следовательно, v — анормальная точка множества (1).

Из приведенных рассуждений следует, что точка v локального минимума анормальна тогда и только тогда, когда конус $\Lambda(v) \cup \{0\}$ неострый. Аналогично доказывается, что точка v локального максимума будет анормальной тогда и только тогда, когда конус $\Lambda^-(v) \cup \{0\}$ неострый.

Заметим, что наличие анормальных точек у множества (1) довольно частое явление. Если $s > n$, то всякая точка множества (1) анормальна и, следовательно, подозрительна на экстремум для любой дифференцируемой функции $f(x)$.

Типичным примером задачи, приводящей к необходимости исследования на экстремум анормальных точек, является следующая: при каких условиях на симметричные матрицы $Q_0, \ldots, Q_s$ квадратичная форма $\langle Q_0 x, x\rangle \geqslant 0$ $\forall x \in K = \{x \in E^n: \langle Q_i x, x\rangle = 0 [\leqslant 0], i = 1, \ldots, s\}$ (см. ниже упражнения 4, 5). Эта задача имеет различные приложения и показывает, что проблема изучения анормальных точек экстремума не является надуманной.

Для дальнейшего анализа на экстремум анормальных точек нужны необходимые условия второго порядка. Однако теоремой 1 мы здесь пользоваться не можем, так как она доказана в предположении, что v — нормальная точка. Более того, нетрудно привести примеры, показывающие, что для анормальных точек теорема 1 просто неверна.

Пример 2. Рассмотрим задачу: $f(x)=-(x^1)^2-(x^2)^2+(x^3)^2\to\inf$, $x\in X=\{x=(x^1,x^2,x^3)\in E^3:\ g_1(x)=x^1x^2=0,\ g_2(x)=(x^1)^2-(x^2)^2=0\}$. Нетрудно видеть, что множество X состоит из точек $x=(0,0,x^3)$ с $\forall x^3$. Отсюда ясно, что $v=(0,0,0)=0$ — точка глобального минимума функции $f(x)$ на X. Так как $g_1'(0)=0$, $g_2'(0)=0$, то условия (4) выполняются при всех $\lambda=(\lambda_1,\lambda_2)\neq 0$. Следовательно, $v=0$ — анормальная точка множества X. Кроме того, $f'(0)=0$, и, очевидно, конус Лагранжа $\Lambda(0)=\{\bar\lambda=(\lambda_0,\lambda_1,\lambda_2)$: $\lambda_0\geqslant 0,\ \forall\lambda_1,\ \forall\lambda_2,\ \bar\lambda\neq 0\}$. Убедимся, что условие (3) не выполняется ни для одного набора $\bar\lambda\in\Lambda(0)$. В самом деле, здесь

$$\mathcal{L}_{xx}(x,\bar\lambda)=2\lambda_0\begin{pmatrix}-1&0&0\\0&-1&0\\0&0&1\end{pmatrix}+\lambda_1\begin{pmatrix}0&1&0\\1&0&0\\0&0&0\end{pmatrix}+\lambda_2\begin{pmatrix}2&0&0\\0&-2&0\\0&0&0\end{pmatrix}\quad\forall x\in E^3,$$

конус $K(0)=\{h=(h^1,h^2,h^3)\in E^3:\ \langle g_1'(0),h\rangle=0,\langle g_2'(0),h\rangle=0\}=E^3$. По этому условие (3) означает, что $\mathcal{L}_{xx}(0,\bar\lambda)\geqslant 0$ (см. определение 2.2). Воспользуемся критерием неотрицательности матрицы, приведенном в замечании 2.2. Если $\bar\lambda=(\lambda_0>0,\lambda_1=0,\lambda_2=0)$, то неравенство

$$\mathcal{L}_{xx}(0,\bar\lambda)=2\lambda_0\begin{pmatrix}-1&0&0\\0&-1&0\\0&0&1\end{pmatrix}\geqslant 0$$

невозможно ни при каком $\lambda_0>0$. Если же $\bar\lambda=(\lambda_0=0,\forall\lambda_1,\forall\lambda_2)$, то неравенство

$$\mathcal{L}_{xx}(0,\bar\lambda)=\begin{pmatrix}2\lambda_2&\lambda_1&0\\\lambda_1&-2\lambda_2&0\\0&0&0\end{pmatrix}\geqslant 0$$

также невозможно, так как $\det\begin{vmatrix}2\lambda_2&\lambda_1\\\lambda_1&-2\lambda_2\end{vmatrix}=-4\lambda_2^2-\lambda_1^2<0\ \ \forall(\lambda_1,\lambda_2)\neq 0$. Аналогично исследуется случай $\lambda_0>0$, $(\lambda_1,\lambda_2)\neq 0$.

Таким образом, теорема 1 в анормальных точках перестает быть справедливой. Следовательно, необходимые условия экстремума второго порядка для анормальных точек должны формулироваться как-то иначе, чем в теореме 1. А как? Ниже приводится необходимое условие второго порядка, справедливое в общем случае, независимо от того, является ли точка v нормальной или анормальной. Этот интересный и изящный результат принадлежит А. В. Арутюнову [44] и в учебной литературе излагается впервые.

Определение 3. Пусть v — точка локального минимума функции $f(x)$ на множестве (1) и $\Lambda(v)$ — соответствующий ей конус Лагранжа. *Конусом Арутюнова* $\Lambda_a=\Lambda_a(v)$ будем называть конус, состоящий из таких наборов $\bar\lambda=(\lambda_0,\ldots,\lambda_s)\in\Lambda(v)$ для каждого из которых существует подпространство $\Pi=\Pi(\bar\lambda)$ пространства E^n, обладающее следующими тремя свойствами:

$$\dim\Pi(\bar\lambda)\geqslant\max\{n-s;0\},\ \text{где}\ \dim\Pi(\bar\lambda)\ —\ \text{размерность}\ \Pi(\bar\lambda);\qquad(5)$$

$$\Pi(\bar\lambda)\subseteq\ker G'(v)=\{h\in E^n:\ \langle g_i'(v),h\rangle=0,\ i=1,\ldots,s\},\ \ G=\begin{pmatrix}g_1\\\ldots\\g_s\end{pmatrix};\qquad(6)$$

$$\langle\mathcal{L}_{xx}(v,\bar\lambda)h,h\rangle\geqslant 0\quad\forall h\in\Pi(\bar\lambda).\qquad(7)$$

Подпространство $\Pi(\bar\lambda)$ со свойствами (5)–(7) будем называть *сопровождающим подпространством точки* $\bar\lambda\in\Lambda_a(v)$.

Убедимся, что множество $\Lambda_a(v)$ в самом деле является конусом. Возьмем произвольные точку $\bar\lambda\in\Lambda_a(v)$ и число $\alpha>0$. Для точки $\alpha\lambda$ в качестве сопровождающего подпространства можно взять то же подпространство $\Pi(\bar\lambda)$, которое является сопровождающим для точки $\bar\lambda$. Свойства (5), (6) не требуют доказательств, а свойство (7) вытекает из того, что $\langle\mathcal{L}_{xx}(v,\alpha\bar\lambda)h,h\rangle=\alpha\langle\mathcal{L}_{xx}(v,\bar\lambda)h,h\rangle\geqslant 0\ \ \forall h\in\Pi(\bar\lambda),\ \ \forall\alpha>0$. Это значит, что $\Lambda_a(v)$ — конус.

Заметим, что при $s\geqslant n$ каждая точка $\bar\lambda\in\Lambda(v)$ обладает сопровождающим подпространством $\Pi(\bar\lambda)=\{0\}$. Следовательно, $\Lambda_a(v)=\Lambda(v)$ при $s\geqslant n$.

Теорема 2 (Арутюнов [44]). *Пусть v — точка локального минимума функции $f(x)$ на множестве* (1), *пусть функции $f(x)$, $g_1(x),\ldots,g_s(x)$ дважды непрерывно дифференцируемы в некоторой окрестности точки v. Тогда*

$$\Lambda_a(v)\neq\varnothing;\qquad(8)$$

$$\max_{\bar\lambda\in\Lambda_a(v),\ |\bar\lambda|=1}\langle\mathcal{L}_{xx}(v,\bar\lambda)h,h\rangle\geqslant 0\qquad(9)$$

$$\forall h\in K(v)=\{h\in E^n:\ \langle g_i'(v),h\rangle=0,\ i=1,\ldots,s\}.$$

Доказательство этой теоремы будет дано ниже в § 5.16. Сейчас мы прокомментируем ее и проиллюстрируем примерами. Сразу же заметим, что в условиях (7) и (9) участвует одна и та же квадратичная форма $\langle\mathcal{L}_{xx}(v,\bar\lambda)h,h\rangle$, и у читателя может сложиться впечатление, что оба этих условия как-то связаны с неотрицательной определенностью указанной формы. Условие (7) в самом деле означает неотрицательность этой формы на сопровождающем подпространстве $\Pi(\bar\lambda)$. А условие (9) гарантирует лишь то, что для каждого фиксированного $h\in K(v)$ найдется своя точка $\bar\lambda=\bar\lambda(h)\in\Lambda_a(v)$, $|\bar\lambda(h)|=1$, для которой значение квадратичной формы $\langle\mathcal{L}_{xx}(v,\bar\lambda)h,h\rangle\geqslant 0$, что вовсе не исключает того, что в какой-либо другой точке $h\in K(v)$ значение той же формы будет <0. Как увидим ниже на примерах, бывают, конечно, случаи, когда максимум в (9) достигается в одной и той же точке $\bar\lambda^*$ при всех $h\in K(v)$, и для такого $\bar\lambda^*$ квадратичная форма $\langle\mathcal{L}_{xx}(v,\bar\lambda^*)h,h\rangle$ в самом деле будет неотрицательной на $K(v)$. На примерах мы также увидим, что такого универсального $\bar\lambda^*$, не зависящего от h, может и не существовать. Заметим также, что условие (8) непустоты конуса $\Lambda_a(v)$ вовсе не вытекает из условия $\Lambda(v)\neq\varnothing$ и весьма содержательно.

Посмотрим, во что превращается теорема 2, когда v — нормальная точка множества (1). Тогда, как было выяснено выше, конус Лагранжа $\Lambda(v)=\{\bar\lambda=\lambda_0(1,\mu),\ \lambda_0>0\}$ — открытый луч с направляющим вектором $(1,\mu)$ где μ — решение системы (2) при $\lambda_0=1$. Согласно теореме 2 конус $\Lambda_a(v)\neq\varnothing$. Отсюда и из включения $\Lambda_a(v)\subset\Lambda(v)$ следует, что $\Lambda_a(v)=\Lambda(v)$. Далее, в нормальной точке v конус $\ker G'(v)$ является подпространством в E^n размерности $n-s\geqslant 0$. С другой стороны, сопровождающее подпространство $\Pi(\bar\lambda)$ точки $\bar\lambda\in\Lambda_a(v)$ в силу (5), (6) имеет размерность $\dim\Pi(\bar\lambda)\geqslant n-s$, и $\Pi(\bar\lambda)\subseteq\ker G'(v)$. Следовательно, $\Pi(\bar\lambda)=\ker G'(v)\ \forall\bar\lambda\in\Lambda_a(v)$, т. е. все точки $\bar\lambda\in\Lambda_a(v)$ обладают одним и тем же сопровождающим подпространством, совпадающим с $\ker G'(v)$. Кроме того, в рассматриваемой задаче $\ker G'(v)=K(v)$, и условие (7) означает, что $\langle\mathcal{L}_{xx}(v,\bar\lambda)h,h\rangle\geqslant 0\ \ \forall h\in\Pi(\bar\lambda)=K(v)$

и $\forall \bar\lambda \in \Lambda_a(v) = \Lambda(v)$. Отсюда следует, что условие (9) в нормальной точке v совпадает с условием (7), так как множество $\{\bar\lambda \in \Lambda_a(v), |\bar\lambda| = 1\}$ состоит из единственной точки и знак максимума в (9) можно опустить. Таким образом, если точка локального минимума v функции $f(x)$ на множестве (1) является нормальной точкой этого множества, то теорема 2 превращается в теорему 1.

Теперь рассмотрим другую крайность, когда v — точка локального минимума, является анормальной точкой множества (1), и, более того, пусть $g_1'(v) = \ldots = g_s'(v) = 0$. Тогда конус $\ker G'(v) \equiv E^n$, и условие (6) тривиально выполняется для всех $\bar\lambda \in \Lambda_a(v)$. Для истолкования условий (5), (7) здесь удобно воспользоваться понятием индекса квадратичной формы. Как известно из линейной алгебры [192; 213; 349; 351; 353], всякая квадратичная форма $\langle Ax, x\rangle$ с помощью невырожденного линейного преобразования может быть приведена к каноническому (диагональному) виду, причем число положительных, число отрицательных и число нулевых членов в канонической форме не зависит от способа приведения (закон инерции квадратичных форм). Поэтому имеет смысл

Определение 4. Число положительных членов в каноническом виде квадратичной формы $\langle Ax, x\rangle$ называется *положительным индексом* этой формы, число отрицательных членов — *отрицательным индексом*.

Если при приведении квадратичной формы $\langle Ax, x\rangle$ к каноническому виду используются лишь ортогональные преобразования, то коэффициенты канонического вида квадратичной формы совпадают с собственными числами матрицы A, причем положительный индекс этой формы равен числу положительных собственных чисел матрицы A, отрицательный индекс — числу отрицательных собственных чисел этой матрицы (с учетом их кратности). Ниже нас будет интересовать лишь отрицательный индекс, и мы его для краткости будем называть просто *индексом* квадратичной формы. Можно доказать, что индекс квадратичной формы совпадает с размерностью максимального подпространства L_- пространства E^n, на котором форма $\langle Ax, x\rangle$ отрицательна, т. е. $\langle Ax, x\rangle < 0$ $\forall x \in L_-$, $x \neq 0$, или еще иначе, индекс формы равен коразмерности максимального подпространства L_+, на котором форма неотрицательно определена. Напомним, что коразмерность подпространства L_+, по определению, есть число $(n - \dim L_+)$ [192; 213; 349; 353].

Отсюда и из (5), (7) получаем содержательное утверждение: если $0 < s < n$ и $\ker G'(v) = E^n$ в анормальной точке v локального минимума функции $f(x)$ на множестве (1) индекс квадратичной формы $\langle \mathcal{L}_{xx}(v, \bar\lambda)h, h\rangle$ не превышает s при всех $\bar\lambda \in \Lambda_a(v)$, или, иначе говоря, количество отрицательных собственных чисел матрицы $\mathcal{L}_{xx}(v, \bar\lambda)$ не превышает s при $\forall \bar\lambda \in \Lambda_a(v)$. В общем случае, когда в анормальной точке подпространство $\ker G'(v) \neq E^n$, можно показать, что количество отрицательных собственных чисел матрицы $P(v)\mathcal{L}_{xx}(v, \bar\lambda)P(v)$ не превышает $s - \operatorname{rang}\{g_1'(v), \ldots, g_s'(v)\}$, где $P(v)$ — матрица оператора проектирования пространства E^n на подпространство $\ker G'(v)$, $\operatorname{rang}\{g_1'(v), \ldots, g_s'(v)\}$ обозначает ранг матрицы, столбцами которой являются векторы $g_1'(v), \ldots, g_s'(v)$. Заметим, что $P(v)\mathcal{L}_{xx}(v, \bar\lambda)P(v)$ — симметричная матрица, так как симметричны матрицы $P(v)$, $\mathcal{L}_{xx}(v, \bar\lambda)$.

Замечание 1. Учитывая, что всякая точка локального максимума функции $f(x)$ на множестве X является точкой локального минимума функции $(-f(x))$ на том же множестве, из теорем 1,2 нетрудно получить необходимые условия второго порядка для точки локального максимума. Предлагаем читателю убедиться, что все утверждения теорем 1, 2 полностью сохраняются, нужно лишь конусы $\Lambda(v)$, $\Lambda_a(v)$ заменить соответственно конусами $\Lambda^-(v)$, $\Lambda_a^-(v)$. Определение конуса Арутюнова $\Lambda_a^-(v)$ для точки v локального максимума функции $f(x)$ на множестве (1) получается из определения 3 заменой конуса $\Lambda(v)$ на конус $\Lambda^-(v)$, в котором, в отличие от $\Lambda(v)$, все наборы $\bar\lambda = (\lambda_0, \ldots, \lambda_s)$ имеют координату $\lambda_0 \leqslant 0$ (см. замечание 3.1).

Для иллюстрации теоремы 2 рассмотрим несколько примеров.

Пример 3. Задача: $f(x) = -(x^1)^2 - (x^2)^2 + (x^3)^2 \to \inf$, $x \in X = \{x = (x^1, x^2, x^3) \in E^3$: $g_1(x) = x^1x^2 = 0$, $g_2(x) = (x^1)^2 - (x^2)^2 = 0\}$. Эту задачу мы уже рассматривали в примере 2 и убедились, что точка $v = (0, 0, 0)$ глобального минимума является анормальной точкой множества X и теорема 1 к нет неприменима. Теорема 2 к этой точке, конечно, применима, и тем не менее интересно посмотреть, как конкретно устроен здесь конус $\Lambda_a(0)$, подпространства $\Pi(0)$ и т. п. Конус Лагранжа $\Lambda(0)$ для точки $v = 0$, как мы уже знаем, состоит из всех точек $\bar\lambda = (\lambda_0, \lambda_1, \lambda_2) \neq 0$ с произвольными $\lambda_0 \geqslant 0$, λ_1, λ_2, а конус $\ker G'(0) = E^3$. Покажем, что конус Арутюнова $\Lambda_a(0) = \Lambda(0)$. В самом деле, возьмем $\forall \bar\lambda \in \Lambda(0)$ и подпространство $\Pi = \{h = (h^1, h^2, h^3) \in E^3$: $h^1 = 0, h^2 = 0\}$. Здесь $n = 3$, $s = 2$, и $\dim \Pi = 1 = 3 - 2$, $\Pi \subset \ker G'(0) = E^3$. Кроме того, $\langle \mathcal{L}_{xx}(0, \bar\lambda)h, h\rangle = 2\lambda_0(h^3)^2 \geqslant 0$ $\forall h \in \Pi$. Таким образом, для всех $\bar\lambda \in \Lambda(0)$ нам удалось указать одно и то же сопровождающее подпространство Π. Это значит, что $\Lambda_a(0) = \Lambda(0)$. Заметим, что для некоторых $\bar\lambda \in \Lambda(0)$ можно указать и другие сопровождающие подпространства. Например, для $\bar\lambda_1 = (0, 1, 0)$ квадратичная форма $\langle \mathcal{L}_{xx}(0, \bar\lambda_1)h, h\rangle = \lambda_1 h^1 h^2 \geqslant 0$ для всех $h \in \Pi = \Pi(\bar\lambda_1) = \{h = (h^1, h^2, h^3)$: $h^1 = \gamma h^2\}$ с $\forall \gamma > 0$. Ясно, что $\dim \Pi(\bar\lambda_1) = 2 > 1$ и все свойства (5)–(7) выполнены. Нетрудно проверить, что индекс квадратичной формы $\langle \mathcal{L}_{xx}(0, \bar\lambda)h, h\rangle$ при всех $\bar\lambda \in \Lambda_a(0)$, как и положено по теореме 2, не превышает 2, причем для некоторых $\bar\lambda$, например, для $\bar\lambda_1 = (0, 1, 0)$ этот индекс равен 1, для $\bar\lambda_2 = (1, 0, 0)$ индекс равен 2.

Теперь обратимся к условию (9). Здесь

$$\langle \mathcal{L}_{xx}(0, \bar\lambda)h, h\rangle = 2\lambda_0(-(h^1)^2 - (h^2)^2 + (h^3)^2) + 2\lambda_1 h^1 h^2 + 2\lambda_2((h^1)^2 - (h^2)^2).$$

В примере 2 было показано, что условие $\langle \mathcal{L}_{xx}(0, \bar\lambda)h, h\rangle \geqslant 0$ $\forall h \in K(0) = E^3$ не может выполняться ни при каких $\bar\lambda \in \Lambda(0) = \Lambda_a(0)$. Это означает, что нет единого, «универсального» для всех h набора $\bar\lambda \in \Lambda_a(0)$, для которого $\langle \mathcal{L}_{xx}(0, \bar\lambda)h, h\rangle \geqslant 0$ $\forall h \in K(0) = E^3$. Однако для каждого отдельно взятого $h \in E^3$ можно указать $\bar\lambda = \bar\lambda(h) \in \Lambda_a(0)$, $|\lambda(h)| = 1$, что $\langle \mathcal{L}_{xx}(0, \bar\lambda(h))h, h\rangle \geqslant 0$. Например, можно взять $\bar\lambda = \bar\lambda(h) = (0, 1, 0)$ при $h^1h^2 > 0$, $\bar\lambda(h) = (0, -1, 0)$ при $h^1h^2 < 0$, $\bar\lambda(h) = (0, 0, 1)$ при $h^1h^2 = 0$, $(h^1)^2 - (h^2)^2 > 0$, $\bar\lambda(h) = (0, 0, -1)$ при $h^1h^2 = 0$, $(h^1)^2 - (h^2)^2 < 0$, $\bar\lambda(h) = (1, 0, 0)$ при $h^1h^2 = 0$, $(h^1)^2 - (h^2)^2 = 0$, т. е. $h^1 = h^2 = 0$. Тогда $\max\limits_{\bar\lambda \in \Lambda_a(0),\ |\bar\lambda| = 1} \langle \mathcal{L}_{xx}(0, \bar\lambda)h, h\rangle \geqslant \langle \mathcal{L}_{xx}(0, \bar\lambda(h))h, h\rangle \geqslant 0$. Из этого примера видно, что в условии (9) взятие максимума по множеству $\{\bar\lambda \in \Lambda_a(0)\ |\bar\lambda| = 1\}$ существенно.

Пример 4. Пусть $f(x)$ — произвольная дважды непрерывно дифференцируемая на E^4 функция, $f'(0) \neq 0$, пусть $X = \{x \in E^4$: $g(x) = (x^1)^2 + (x^2)^2 - (x^3)^2 - (x^4)^2 = 0\}$. Функция Лагранжа имеет вид: $\mathcal{L}(x, \bar\lambda) =$

$= \lambda_0 f(x) + \lambda_1 g(x)$, а ее производные:

$$\mathcal{L}_x(x, \bar\lambda) = \lambda_0 f'(x) + 2\lambda_1 \begin{pmatrix} x^1 \\ x^2 \\ -x^3 \\ -x^4 \end{pmatrix}, \quad \mathcal{L}_{xx}(x, \bar\lambda) = \lambda_0 f''(x) + 2\lambda_1 \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

Равенство $\mathcal{L}_x(0, \bar\lambda) = \lambda_0 f'(0) + \lambda g'(0) = \lambda_0 f'(0) = 0$ выполняется лишь при $\lambda_0 = 0$ и любых λ_1. Следовательно, точка $v = 0$ удовлетворяет необходимому условию экстремума первого порядка, ее конус Лагранжа $\Lambda(0) = \{\bar\lambda = (\lambda_0, \lambda_1)\colon \lambda_0 = 0, \forall \lambda_1 \neq 0\}$. Конусы $\ker G'(0)$, $K(0)$ здесь совпадают со всем пространством E^4. Ясно также, что $v = 0$ — анормальная точка множества X. Число отрицательных собственных чисел у матрицы

$$\mathcal{L}_{xx}(0, \bar\lambda) = 2\lambda_1 \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

при любых $\lambda_1 \neq 0$ равно 2. Однако по теореме 2 индекс квадратичной формы $\langle \mathcal{L}_{xx}(0, \bar\lambda)h, h\rangle$ не должен быть выше $s = 1$ для $\forall \lambda \in \Lambda_a(0)$. Это означает, что конус $\Lambda_a(0)$ пуст. Условие (8) нарушено. По теореме 2 и замечанию 1 точка $v = 0$ не может быть точкой экстремума функции $f(x)$ на множестве X.

Пример 5. Пусть $f(x)$ — произвольная дважды непрерывная дифференцируемая на E^3 функция, $f'(0) \neq 0$, пусть $X = \{x \in E^3\colon g(x) = (x^1)^2 + (x^2)^2 - (x^3)^2 = 0\}$. Тогда функция Лагранжа: $\mathcal{L}(x, \bar\lambda) = \lambda_0 f(x) + \lambda_1 g(x)$, ее производные:

$$\mathcal{L}_x(x, \bar\lambda) = \lambda_0 f'(x) + 2\lambda_1 \begin{pmatrix} x^1 \\ x^2 \\ -x^3 \end{pmatrix}, \quad \mathcal{L}_{xx}(x, \bar\lambda) = \lambda_0 f''(x) + 2\lambda_1 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

Нетрудно видеть, что точка $v = 0$ подозрительна на экстремум в силу теоремы 3.1 и ее конус Лагранжа $\Lambda(0) = \{\bar\lambda = (\lambda_0, \lambda_1)\colon \lambda_0 = 0, \forall \lambda_1 \neq 0\}$, конусы $\ker G'(0) = K(0) = E^3$. Кроме того, $v = 0$ — анормальная точка множества X. Применим к ней теорему 2. Заметим, что $\mathcal{L}_{xx}(0, \bar\lambda) = 2\lambda_1 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$ при всех $\bar\lambda \in \Lambda(0)$. Отсюда видно, что индекс квадратичной формы $\langle \mathcal{L}_{xx}(0, \bar\lambda)h, h\rangle$ равен 1 при $\lambda_1 > 0$ и равен 2 при $\lambda_1 < 0$. По теореме 2 индекс этой формы не может превышать $s = 1$. Это означает, что при $\lambda_1 < 0$ точка $\bar\lambda \in \Lambda(0)$ не может принадлежать конусу $\Lambda_a(0)$. Рассмотрим случай $\lambda > 0$. Тогда, оказывается, точка $\bar\lambda = (0, \lambda) \in \Lambda_a(0)$. В самом деле, для каждой точки $\bar\lambda = (\lambda_0 = 0, \lambda_1 > 0) \in \Lambda(0)$ можем указать сопровождающее подпространство $\Pi = \{h = (h^1, h^2, h^3), h^3 = 0, \forall h^1, h^2\}$; свойства (5)–(7) здесь легко проверяются: $\dim \Pi = 2 (= n - s)$, $\Pi \subset \ker G'(0) = E^3$, $\langle \mathcal{L}_{xx}(0, \bar\lambda)h, h\rangle = 2\lambda_1((h^1)^2 + (h^2)^2) \geqslant 0 \; \forall h \in \Pi$. Это и значит, что $\Lambda_a(0) = \{\bar\lambda = (0, \lambda)\colon \lambda \geq 0\}$. Далее воспользуемся утверждением (9) теоремы 2. Множество $\{\bar\lambda\colon \bar\lambda \in \Lambda_a(0), |\bar\lambda| = 1\}$ здесь состоит из единственной точки $\bar\lambda = (\lambda_0 = 0, \lambda_1 = 1)$ и для нее неравенство $\langle \mathcal{L}_{xx}(0, \bar\lambda)h, h\rangle = (h^1)^2 + (h^2)^2 - (h^3)^2 \geqslant 0$ не может выполняться при всех $h \in K(0) = E^3$, т. е. нарушено условие (9). В силу теоремы 2 и замечания 1 точка $v = 0$ не может быть точкой экстремума рассматриваемой функции $f(x)$ на X.

Пример 6. Пусть $f(x) = \langle c, x\rangle + \langle Q_0 x, x\rangle$, где $x \in E^{10}$, $c \in E^{10}$, $c \neq 0$, Q_0 — произвольная матрица размера 10×10. Пусть $X = \{x \in E^{10}\colon g_1(x) = -(x^1)^2 - (x^2)^2 + (x^3)^2 + (x^4)^2 + (x^5)^2 = 0, g_2(x) = -(x^6)^2 - (x^7)^2 + (x^8)^2 + (x^9)^2 + (x^{10})^2 = 0\}$. Так как $g_1'(0) = 0$, $g_2'(0) = 0$, то $v = 0$ — анормальная точка множества X, и, следовательно, она подозрительна на экстремум. Конус Лагранжа этой точки равен $\Lambda(0) = \{\bar\lambda = (\lambda_0, \lambda_1, \lambda_2)\colon \lambda_0 = 0, \forall \lambda_1, \forall \lambda_2, (\lambda_1, \lambda_2) \neq 0\}$, т. к. $f'(0) = c \neq 0$; $\ker G'(0) = K(0) = E^{10}$. Найдем конус Арутюнова $\Lambda_a(0)$. Для $\forall \bar\lambda \in \Lambda(0)$ квадратичная форма $\langle \mathcal{L}_{xx}(0, \bar\lambda)h, h\rangle = 2\lambda_1(-(h^1)^2 - (h^2)^2 + (h^3)^2 + (h^4)^2 + (h^5)^2) + 2\lambda_2(-(h^6)^2 - (h^7)^2 + (h^8)^2 + (h^9)^2 + (h^{10})^2)$. Если $\lambda_1 < 0$ или $\lambda_2 < 0$, то индекс этой формы $\geqslant 3$, а для $\lambda \in \Lambda_a(0)$ этот индекс не может превышать 2. Это значит, что точки $\bar\lambda = (\lambda_0 = 0, \lambda_1 < 0, \lambda_2 < 0)$ не могут принадлежать $\Lambda_a(0)$. Для точек $\bar\lambda = (\lambda_0 = 0, \lambda_1 > 0, \lambda_2 > 0)$ индекс формы $\langle \mathcal{L}_{xx}(0, \bar\lambda)h, h\rangle \geqslant 4$, поэтому такие точки также не могут принадлежать конусу $\Lambda_a(0)$. Следовательно, лишь лучи $\Lambda_1 = \{\bar\lambda = (\lambda_0 = 0, \lambda_1 > 0, \lambda_2 = 0)\}$ и $\Lambda_2 = \{\bar\lambda = (\lambda_0 = 0, \lambda_1 = 0, \lambda_2 > 0)\}$ могут принадлежать $\Lambda_a(0)$. В самом деле, для Λ_1 можем взять сопровождающее подпространство $\Pi_1 = \{h \in E^{10}\colon h^1 = h^2 = 0\}$, для Λ_2 — $\Pi_2 = \{h \in E^{10}\colon h^6 = h^7 = 0\}$, размерность которых равна 8. Так как здесь $\ker G'(0) = E^{10}$, то ясно, что все условия (5)–(7) выполняются. Таким образом, конус $\Lambda_a(0)$ представляет собой объединение двух лучей Λ_1 и Λ_2. Проверим условие (9). Множество $\{\lambda \in \Lambda_a(0)\colon |\bar\lambda| = 1\}$ состоит из двух точек $\bar\lambda_1 = (\lambda_0 = 0, \lambda_1 = 1, \lambda_2 = 0) \in \Lambda_1$, $\bar\lambda_2 = (\lambda_0 = 0, \lambda_1 = 0, \lambda_2 = 1) \in \Lambda_2$. Возьмем $h = h_0 \in E^{10} = K(0)$ с координатами $h^1 = 1$, $h^6 = 1$, все остальные координаты $h^i = 0$. Тогда $\langle \mathcal{L}_{xx}(0, \bar\lambda_i)h_0, h_0\rangle = -2$, $i = 1, 2$, и в точке h_0 условие (9) нарушается. В силу теоремы 2 и замечания 1 заключаем, что v не может быть точкой экстремума функции $f(x)$ на множестве X.

Следующий пример показывает, что теорема 2 не «всесильна» и не всегда «распознает» точки, которые подозрительны на экстремум в силу теоремы 3.1, но на самом деле в них нет экстремума.

Пример 7. Рассмотрим функцию $f(x) = -(x^1)^2 - (x^2)^2 + (x^3)^2$ на множестве $X = \{x = (x^1, x^2, x^3) \in E^3\colon g(x) = x^1 x^2 = 0\}$. Нетрудно проверить, что $v = 0$ — анормальная точка этого множества. Следовательно, она подозрительна на экстремум. Ее конус Лагранжа $\Lambda(0) = \{\bar\lambda = (\lambda_0, \lambda_1)\colon \lambda_0 \geqslant 0, \forall \lambda_1 \neq 0\}$, конусы $\ker G'(0) = K(0) = E^3$. Опишем конус Арутюнова $\Lambda_a(0)$. Согласно теореме 2 индекс квадратичной формы $\langle \mathcal{L}_{xx}(0, \bar\lambda)h, h\rangle = 2\lambda_0(-(h^1)^2 - (h^2)^2 + (h^3)^2) + 2\lambda_1 h^1 h^2$ не может превышать $s = 1$ для всех $\bar\lambda \in \Lambda_a(0)$. Матрица

$$\mathcal{L}_{xx}(0, \bar\lambda) = \begin{pmatrix} -2\lambda_0 & \lambda_1 & 0 \\ \lambda_1 & -2\lambda_0 & 0 \\ 0 & 0 & 2\lambda_0 \end{pmatrix}$$

имеет три собственных числа $\gamma_1 = -2\lambda_0 + \lambda_1$, $\gamma_2 = -2\lambda_0 - \lambda_1$, $\gamma_3 = 2\lambda_0 \geqslant 0$. Если $\bar\lambda \in \Lambda(0)$ и $|\lambda_1| < 2\lambda_0$, то $\gamma_1 < 0$, $\gamma_2 < 0$, т. е. индекс квадратичной формы $\langle \mathcal{L}_{xx}(0, \bar\lambda)h, h\rangle$ равен 2, и поэтому такие $\bar\lambda$ не могут принадлежать конусу $\Lambda_a(0)$. Если $\bar\lambda \in \Lambda(0)$ и $|\lambda_1| \geqslant 2\lambda_0 \geqslant 0$, то только одно из собственных чисел γ_1 или γ_2 будет отрицательным, поэтому индекс формы $\langle \mathcal{L}_{xx}(0, \bar\lambda)h, h\rangle$ будет равен 1. В качестве сопровождающего подпространства Π для всех $\bar\lambda \neq 0$, $|\lambda_1| \geqslant 2\lambda_0 \geqslant 0$ можно взять $\Pi = \{h \in E^3\colon h^1 = h^2 = 0, \forall h^3\}$. Тогда

$\langle \mathcal{L}_{xx}(0, \bar\lambda)h, h\rangle = 2\lambda_0(h^3)^2 \geqslant 0$ и, очевидно, все условия (5)–(7) выполнены. Следовательно, конус $\Lambda_a(0) = \{\bar\lambda = (\lambda_0, \lambda_1)\colon \bar\lambda \neq 0, |\lambda_1| \geqslant 2\lambda_0 \geqslant 0\}$. Проверим условие (9). Возьмем $\forall h = (h^1, h^2, h^3) \in K(0) = E^3$ и определим $\bar\lambda = \bar\lambda(h)$ так: $\lambda_0 = 0$, $\lambda_1 = 1$ при $h^1h^2 \geqslant 0$ и $\lambda_1 = -1$ при $h^1h^2 < 0$. Ясно, что $\lambda(h) \in \Lambda_a(0)$ и $|\bar\lambda(h)| = 1$. Тогда $\max\limits_{\bar\lambda \in \Lambda_a(0),\ |\bar\lambda|=1} \langle \mathcal{L}_{xx}(0, \bar\lambda)h, h\rangle \geqslant \langle \mathcal{L}_{xx}(0, \lambda(h))h, h\rangle = 2|\lambda_1||h^1h^2| \geqslant 0$ $\forall h \in E^3$. Как видим, точка $v = 0$ удовлетворяет всем необходимым условиям второго порядка теоремы 2. Но $v = 0$, очевидно, не является точкой экстремума функции $f(x)$ на X. Однако теорема 2 не сумела ее забраковать.

2. Сформулируем необходимое условие экстремума второго порядка для множеств более общего вида:

$$X = \{x \in E^n\colon g_1(x) \leqslant 0, \ldots, g_m(x) \leqslant 0, g_{m+1}(x) = 0, \ldots, g_s(x) = 0\}. \quad (10)$$

Пусть v — точка локального минимума функции $f(x)$ на множестве (10), пусть $\Lambda(v)$ — конус Лагранжа этой точки. Напомним, что конус Лагранжа точки v состоит из всех тех точек $\bar\lambda = (\lambda_0, \lambda_1, \ldots, \lambda_s) \neq 0$, $\lambda_0 \geqslant 0$, $\lambda_1 \geqslant 0, \ldots, \lambda_m \geqslant 0$, для которых пара $(v, \bar\lambda)$ является решением системы (3.13). Пусть $I(v) = \{i\colon 1 \leqslant i \leqslant m, g_i(v) = 0\} \cup \{i\colon m+1 \leqslant i \leqslant s\}$ — множество номеров активных ограничений точки v, $|I(v)|$ — количество элементов множества $I(v)$. *Конусом Арутюнова* точки v множества (10) будем называть подмножество $\Lambda_a(v)$ таких точек $\bar\lambda \in \Lambda(v)$, для которых существует подпространство $\Pi = \Pi(\bar\lambda)$ пространства E^n, обладающее следующими свойствами:

$$\dim \Pi(\bar\lambda) \geqslant \max\{n - |I(v)|; 0\}; \quad (11)$$

$$\Pi(\bar\lambda) \subseteq \ker G'(v) = \{h \in E^n\colon \langle g_i'(v), h\rangle = 0, i \in I(v)\}, \ G = \{g_i, i \in I(v)\}; \quad (12)$$

$$\langle \mathcal{L}_{xx}(v, \bar\lambda)h, h\rangle \geqslant 0 \quad \forall h \in \Pi(\bar\lambda). \quad (13)$$

Как и в определении 3, подпространство $\Pi(\bar\lambda)$ со свойствами (11)–(13) будем называть *сопровождающим подпространством* точки $\bar\lambda \in \Lambda_a(v)$. Рассуждая также, как в случае множества (1), нетрудно убедиться, что и здесь $\Lambda_a(v)$ действительно является конусом.

Заметим, что при $|I(v)| \geqslant n$ каждая точка $\bar\lambda \in \Lambda(v)$ обладает нулевым сопровождающим подпространством $\Pi(\bar\lambda) = \{0\}$. Поэтому конус $\Lambda_a(v) = \Lambda(v)$ при $|I(v)| \geqslant n$.

Теорема 3 (Арутюнов [44]). *Пусть v — точка локального минимума функции $f(x)$ на множестве* (10), *пусть функции $f(x), g_1(x), \ldots, g_s(x)$ дважды непрерывно дифференцируемы в некоторой окрестности точки v. Тогда*

$$\Lambda_a(v) \neq \varnothing, \quad (14)$$

$$\max_{\bar\lambda \in \Lambda_a(v),\ |\bar\lambda|=1} \langle \mathcal{L}_{xx}(v, \bar\lambda)h, h\rangle \geqslant 0 \quad \forall h \in K(v), \langle f'(v), h\rangle \leqslant 0, \quad (15)$$

где

$$K(v) = \{h \in E^n\colon \langle g_i'(v), h\rangle \leqslant 0, \ i \in I(v) \cap \{i\colon 1 \leqslant i \leqslant m\}, \ \langle g_i'(v), h\rangle = 0, i = m+1, \ldots, s\} \quad (16)$$

— конус критических направлений множества (10) *в точке v.*

Доказательство этой теоремы проводится с помощью метода штрафных функций в § 5.16. Сделаем несколько замечаний.

Замечание 2. Прежде всего убедимся, что при $m = 0$ теорема 3 превращается в теорему 2. Для этого покажем, что неравенство $\langle f'(v), h\rangle \leqslant 0$ из (15) в случае $m = 0$ может быть опущено без потерь. В самом деле, очевидно, для любого вектора $h \in E^n$ выполняется одно из двух неравенств: $\langle f'(v), h\rangle \leqslant 0$ или $\langle f'(v), h\rangle \geqslant 0$, и справедливо равенство $\langle \mathcal{L}_{xx}(v, \bar\lambda)h, h\rangle = \langle \mathcal{L}_{xx}(v, \bar\lambda)(-h), (-h)\rangle$. Кроме того из $\langle f'(v), h\rangle \leqslant 0$ следует, что $\langle f'(v), (-h)\rangle \geqslant 0$. Отсюда ясно, что если неравенство (9) выполняется для $\forall h \in K(v)$, $\langle f'(v), h\rangle \leqslant 0$, то оно выполняется и для $\forall h \in K(v)$, $\langle f'(v), h\rangle \geqslant 0$. Это значит, что при $m = 0$ в условии (15) требование $\langle f'(v), h\rangle \leqslant 0$ может быть опущено, а тогда теорема 3 превращается в теорему 2.

Замечание 3. Если v — точка локального максимума функции $f(x)$ на множестве (10), то утверждения (14), (15) сохраняют силу, нужно лишь конус $\Lambda_a(v)$ заменить на $\bar\Lambda_a(v)$, а в условии (15) неравенство $\langle f'(v), h\rangle \leqslant 0$ заменить на $\langle f'(v), h\rangle \geqslant 0$. Конус Арутюнова $\bar\Lambda_a(v)$ для точки v локального максимума функции $f(x)$ на множестве (10) определяется также, как конус $\Lambda_a(v)$, нужно лишь вместо $\Lambda(v)$ взять конус $\bar\Lambda(v)$ (см. замечание 3.2).

Обсудим понятия анормальной и нормальной точки для множества (10). Мы здесь будем придерживаться трактовки этих понятий, принятой в [44].

Определение 5. Точка v множества (10) называется *анормальной*, если $0 \leqslant m < s$ и градиенты $g_{m+1}'(v), \ldots, g_s'(v)$ линейно зависимы, т. е. существуют числа $\lambda_{m+1}, \ldots, \lambda_s$, такие, что

$$\lambda_{m+1}g_{m+1}'(v) + \ldots + \lambda_s g_s'(v) = 0, \quad (\lambda_{m+1}, \ldots, \lambda_s) \neq 0. \quad (18)$$

При $m = s$, когда в (10) ограничения типа равенств отсутствуют, во множестве (10) анормальных точек нет по определению.

Если $s - m > n$, то все точки множества (10) являются анормальными.

Нетрудно видеть, что всякая анормальная точка v множества (10) является подозрительной на экстремум для любой дифференцируемой функции $f(x)$, так как пара $(v, \bar\lambda)$, где $\bar\lambda$ имеет координаты $\lambda_0 = \lambda_1 = \ldots = \lambda_m = 0$, а $\lambda_{m+1}, \ldots, \lambda_s$ взяты из (18), является решением системы (3.13). Как и в случае множества (1) можно показать, что точка v локального минимума дифференцируемой функции $f(x)$ на множестве (10) будет анормальной тогда и только тогда, когда конус $\Lambda(v) \cup \{0\}$ является неострым, т. е. содержит прямую. В самом деле, если набор $(\lambda_{m+1}, \ldots, \lambda_s)$ удовлетворяет условию (18), то набор $(-\lambda_{m+1}, \ldots, -\lambda_s)$ также удовлетворяет этому условию, поэтому наборы $\bar\lambda_0 = (\lambda_0 = 0, \lambda_1 = 0, \ldots, \lambda_m = 0, \lambda_{m+1}, \ldots, \lambda_s)$ и $(-\bar\lambda_0) = (\lambda_0 = 0, \lambda_1 = 0, \ldots, \lambda_m = 0, -\lambda_{m+1}, \ldots, -\lambda_s)$ принадлежат $\Lambda(v)$. Поскольку $\Lambda(v)$ конус, то точки $t\bar\lambda_0$ и $t(-\bar\lambda_0) \in \Lambda(v)$ при всех $t > 0$. Следовательно, прямая $\lambda(t) = t\bar\lambda_0$, $-\infty < t < +\infty$, принадлежит $\Lambda(v) \cup \{0\}$, т. е. конус $\Lambda(v) \cup \{0\}$ неострый. Обратное, если конус $\Lambda(v) \cup \{0\}$ содержит некоторую прямую $\lambda(t) = t\bar\mu$, $-\infty < t < +\infty$, $\bar\mu = (\mu_0, \ldots, \mu_s) \neq 0$, то точки $\bar\lambda(1) = \bar\mu$ и $\bar\lambda(-1) = -\bar\mu$ принадлежат $\Lambda(v)$. Поскольку у всех точек $\bar\lambda$ конуса $\Lambda(v)$ координаты $\lambda_0 \geqslant 0, \ldots, \lambda_m \geqslant 0$, то $\mu_0 \geqslant 0, \ldots, \mu_m \geqslant 0$, $-\mu_0 \geqslant 0, \ldots, -\mu_m \geqslant 0$, так что $\mu_0 = 0, \mu_1 = 0, \ldots, \mu_m = 0$, и $(\mu_{m+1}, \ldots, \mu_s) \neq 0$. Это значит, что условия $\mathcal{L}_x(v, \bar\mu) = 0$, $\bar\mu \neq 0$ из системы (3.13) превращаются

в условие (18) с $(\lambda_{m+1}=\mu_{m+1}, \ldots, \lambda_s=\mu_s)\neq 0$. Тем самым показано, что если конус $\Lambda(v)\cup\{0\}$ неострый, то v — анормальная точка множества (10). Аналогично доказывается, что точка v локального максимума дифференцируемой функции $f(x)$ на множестве (10) анормальна тогда и только тогда, когда конус $\Lambda^-(0)\cup\{0\}$ неострый.

Заметим, что если в (10) $m=s$, то конус $\Lambda(v)\cup\{0\}$ или $\Lambda^-(0)\cup\{0\}$ не может содержать прямую. В противном случае нашлись бы точки $\bar\lambda$, $(-\bar\lambda)\in\Lambda(v)$ или $\in\Lambda^-(v)$. Это означало бы, что $\lambda_1\geqslant 0, \ldots, \lambda_m\geqslant 0$, $-\lambda_1\geqslant 0, \ldots, -\lambda_m\geqslant 0$, а $\lambda_0\geqslant 0$, $-\lambda_0\geqslant 0$, либо $\lambda_0\leqslant 0$, $-\lambda_0\leqslant 0$, отсюда следует, что $\bar\lambda=0$. Однако, $0\notin\Lambda(v)$ и $0\notin\Lambda^-(v)$. Противоречие. Приведенные соображения оправдывают определение 5, когда в (10) $m=s$.

Таким образом, понятие анормальной точки множества (10), выражая лишь специфическое свойство (18) этого множества, автоматически является подозрительной на экстремум для любой дифференцируемой функции $f(x)$. Поскольку при $0\leqslant m<s$ наличие анормальной точки у множества (10) не такое уж редкое явление, то возможность использования теоремы 2 для их анализа на экстремум, представляется весьма важным.

Определение 6. Точка v множества (10) называется *нормальной*, если система линейных уравнений и неравенств относительно неизвестных $\lambda_1, \ldots, \lambda_s$

$$\sum_{i=1}^{s}\lambda_i g_i'(v)=0, \quad \lambda_i g_i(v)=0, \quad \lambda_i\geqslant 0, \quad i=1, \ldots, m, \tag{19}$$

имеет лишь нулевое решение.

Как видим, при $m=0$ определение 6 совпадает с определением 1 нормальной точки для множества (1). Как и в случае множества (1) нетрудно заметить, что если точка v, подозрительная на экстремум функции $f(x)$ на множестве (10), является нормальной, то в конусе Лагранжа все точки $\bar\lambda=(\lambda_0, \ldots, \lambda_s)$ имеют координату $\lambda_0\neq 0$. В самом деле, если бы в такой точке существовал набор $\bar\lambda=(\lambda_0=0, \lambda_1, \ldots, \lambda_s)\neq 0$, то как видно из (3.13), система (19) имела бы решение. Верно и обратное: если в конусе Лагранжа у всех точек $\bar\lambda$ координата $\lambda_0\neq 0$, то v — нормальная точка множества (10).

Заметим, что если в нормальной точке множества (1) конус Лагранжа состоял из одного луча, то в случае $m>0$ этот конус, вообще говоря, богаче. Кроме того, не следует думать, что если точка множества (10) не является анормальной, то она непременно будет нормальной. Так, если $m=s$, то в множестве (10) по определению 5 нет анормальных точек. Однако если, например, при этом $g_i'(v)=0$, $i=1, \ldots, m$, и имеются активные ограничения, то система (19) будет иметь бесконечно много решений, и, следовательно, точка v не является нормальной. Таким образом, в множестве (10) могут быть точки, которые не являются ни нормальными, ни анормальными. Конечно, всякая нормальная точка множества (10) не может быть анормальной, так как конус $\Lambda(v)$ любой анормальной точки содержит $\bar\lambda$ с $\lambda_0=0$.

Покажем, что v будет нормальной точкой множества (10) тогда и только тогда, когда: 1) векторы $g_{m+1}'(v), \ldots, g_s'(v)$ линейно независимы, 2) существует вектор $d\in E^n$, для которого

$$\begin{aligned}&\langle g_i'(v), d\rangle=0, \quad i=m+1, \ldots, s,\\ &\langle g_i'(v), d\rangle<0, \quad \forall i\in I(v)\cap\{1\leqslant i\leqslant m\},\end{aligned} \tag{20}$$

где $I(v)$ — множество номеров активных ограничений точки v (возможность $m=0$ или $m=s$ здесь не исключается). Это условие в литературе по экстремальным задачам (особенно зарубежной) часто называют *условием Мангасариана — Фрамовица*. Убедимся, что при выполнении перечисленных условий точка v будет нормальной. Допустим, что это не так. Тогда система (19) имеет хотя бы одно решение λ, и для него $0=\langle\sum_{i=1}^{s}\lambda_i g_i'(v), d\rangle=\sum_{i\in I(v)}\lambda_i\langle g_i'(v), d\rangle\leqslant 0$, что возможно только при $\lambda_i=0$ $\forall i\in I(v)\cap\{1\leqslant i\leqslant m\}$, так что $\lambda_1=\ldots=\lambda_m=0$. Если при этом $m=s$, то это означает $\lambda=0$, что противоречит тому, что λ — решение системы (19). Если же $m<s$, то равенство $\sum_{i=1}^{s}\lambda_i g_i'(v)=0$ превращается в $\sum_{i=m+1}^{s}\lambda_i g_i'(v)=0$, что может быть только при $\lambda_{m+1}=\ldots=\lambda_s=0$. Таким образом, и в случае $m<s$ получаем $\lambda=0$, что невозможно для решения системы (19). Следовательно, система (19) не имеет решения, т. е. v — нормальная точка множества (10).

Докажем обратное: если v — нормальная точка множества (10), то условия Мангасариана — Фрамовица выполнены. Пусть это не так. Тогда либо векторы $g_{m+1}'(v), \ldots, g_s'(v)$ линейно зависимы, либо система (20) несовместна. В первом случае найдутся числа $(\alpha_{m+1}, \ldots, \alpha_s)\neq 0$, такие, что $\sum_{i=m+1}^{s}\alpha_i g_i'(v)=0$. Тогда $\lambda=(\lambda_1=0, \ldots, \lambda_m=0, \lambda_{m+1}=\alpha_{m+1}, \ldots, \lambda_s=\alpha_s)$ — решение системы (19), что противоречит нормальности точки v. Во втором случае, когда система (20) несовместна, из теоремы Моцкина (см. упражнение 3.5.14) следует, что совместна система

$$\sum_{i\in I(v)}p_i g_i'(v)=0, \quad p_i\geqslant 0, \quad i\in I(v)\cap\{i\colon 1\leqslant i\leqslant m\}, \quad (p_i, i\in I(v))\neq 0.$$

Тогда вектор $\lambda=(\lambda_i=p_i, i\in I(v), \lambda_i=0, i\notin I(v))$ будет решением системы (19), что опять-таки противоречит нормальности точки v. Таким образом, установлено, что условие Мангасариана — Фрамовица равносильно нормальности точки v множества (10).

Покажем, что в отличие от множества (1), в нормальных точках v множества (10), подозрительных на экстремум, в условии (15) операция взятия максимума, вообще говоря, не может быть опущена. С этой целью, как мы уже это делали неоднократно, позаимствуем пример из [44].

Пример 8. Рассмотрим задачу: $f(x)=-x^3\to\inf$ $x\in X=\{x=(x^1, x^2, x^3)\in E^3$: $g_1(x)=x^3+2x^1x^2-\varepsilon((x^1)^2+(x^2)^2)\leqslant 0$, $g_2(x)=x^3-2x^1x^2-\varepsilon((x^1)^2+(x^2)^2)\leqslant 0$, $g_3(x)=x^3+((x^1)^2-(x^2)^2)-\varepsilon((x^1)^2+(x^2)^2)\leqslant 0$, $g_4(x)=x^3-(x^1)^2+(x^2)^2-\varepsilon((x^1)^2+(x^2)^2)\leqslant 0\}$, $\varepsilon>0$. Сначала убедимся, что $v=(0, 0, 0)=0$ — решение этой задачи. На плоскости (x^1, x^2) введем полярную систему координат: $x^1=r\cos\varphi$, $x^2=r\sin\varphi$, $0\leqslant\varphi<2\pi$. Тогда $g_1(x)=x^3+r^2(\sin 2\varphi-\varepsilon)$, $g_2(x)=x^3-r^2(\sin 2\varphi+\varepsilon)$, $g_3(x)=x^3+r^2(\cos 2\varphi-\varepsilon)$, $g_4(x)=x^3-r^2(\cos 2\varphi+\varepsilon)$. Выберем $\varepsilon>0$ таким, чтобы

$$\min\{\sin 2\varphi, -\sin 2\varphi, \cos 2\varphi, -\cos 2\varphi\}-\varepsilon<0 \quad \forall\varphi, \quad 0\leqslant\varphi<2\pi,$$

например, можно взять $\varepsilon=1/4$. Тогда хотя бы в одном из неравенств $x^3\leqslant -r^2(\sin 2\varphi-\varepsilon)$, $x^3\leqslant r^2(\sin 2\varphi+\varepsilon)$, $x^3\leqslant -r^2(\cos 2\varphi-\varepsilon)$, $x^3\leqslant r^2(\cos 2\varphi+\varepsilon)$ правая часть отрицательна при каком-либо φ, $0\leqslant\varphi<2\pi$, и любых

$r>0$, и все правые части $=0$ при $r=0$. Это значит, что у всех точек $x=(x^1, x^2, x^3)\in X$ координата $x^3\leqslant 0$. Следовательно, $\inf\limits_{x\in X} f(x)=0=f(0)$ и $v=0$ — решение рассматриваемой задачи. В точке $v=0$ выполняется условие Мангасариана — Фрамовица с вектором $d=(0,0,-1)$. Следовательно $v=0$ — нормальная точка множества X. Конус Лагранжа точки $v=0$ равен $\Lambda(0)=\{\bar\lambda=(\lambda_0,\lambda_1,\lambda_2,\lambda_3,\lambda_4)$: $\lambda_0=\lambda_1+\lambda_2+\lambda_3+\lambda_4>0$, $\lambda_i\geqslant 0$, $i=1,\ldots,4\}$. Поскольку в точке $v=0$ все ограничения $g_i(x)\leqslant 0$, $i=1,\ldots,4$, активные, то $|I(0)|=4>n=3$ и каждая точка $\bar\lambda\in\Lambda(0)$ имеет тривиальное сопровождающее подпространство $\Pi(\bar\lambda)=\{0\}$. Следовательно, в рассматриваемой задаче конус Арутюнова $\Lambda_a(0)=\Lambda(0)$. Согласно теореме 3 в точке минимума $v=0$ условие (15) выполняется при всех $h\in K_1(0)=K(0)\cap\{h$: $\langle f'(0), h\rangle\leqslant 0\}=\{h\in E^3$: $\langle g_i'(0), h\rangle\leqslant 0$, $i=1,\ldots,4$, $\langle f'(0), h\rangle\leqslant 0\}=\{h=(h^1, h^2, h^3=0)\}$. Однако $\langle\mathcal{L}_{xx}(0,\bar\lambda)h, h\rangle=\langle\left(\sum\limits_{i=1}^{4}\lambda_i g_i''(0)\right)h, h\rangle=2|h|^2[A(\bar\lambda)\sin 2\varphi+B(\bar\lambda)\cos 2\varphi-\varepsilon]$ при каждом $h=(h^1=|h|\cos\varphi, h^2=|h|\sin\varphi, h^3=0)\in K_1(0)$, $\lambda\in\Lambda(0)$; явные выражения величин $A(\bar\lambda)$, $B(\bar\lambda)$ нам не понадобятся, поэтому мы их не будем здесь приводить. Отсюда видно, что для $\forall\bar\lambda\in\Lambda(0)=\Lambda_a(0)$, $|\bar\lambda|=1$ найдется φ, $0\leqslant\varphi<2\pi$, такой, что $A(\bar\lambda)\sin 2\varphi+B(\bar\lambda)\cos 2\varphi-\varepsilon=\sqrt{A^2(\bar\lambda)+B^2(\bar\lambda)}\sin(2\varphi+\psi_0)-\varepsilon<0$ (здесь ψ_0 вспомогательный угол, $0\leqslant\psi_0<2\pi$, определяемый равенствами $\cos\psi_0=\frac{A(\bar\lambda)}{\sqrt{A^2(\bar\lambda)+B^2(\bar\lambda)}}$, $\sin\psi_0=\frac{B(\bar\lambda)}{\sqrt{A^2(\bar\lambda)+B^2(\bar\lambda)}}$ при $A^2(\bar\lambda)+B^2(\bar\lambda)\neq 0$). Таким образом, при любом выборе $\bar\lambda\in\Lambda_a(0)$, $|\bar\lambda|=1$ найдется $h\in K(0)$, $h\neq 0$, что $\langle\mathcal{L}_{xx}(0,\bar\lambda)h, h\rangle<0$. В то же время условие (15) согласно теореме 3 выполняется. Это значит, что даже в нормальной точке экстремума знак максимума в (15) не может быть опущен.

3. Кратко обсудим один известный прием сведения задач с ограничениями типа неравенств к задачам с ограничениями типа равенств. Этот прием, по-видимому, впервые был предложен Н. Н. Гернет [221] для исследования задач вариационного исчисления с односторонними ограничениями. Опишем его для задачи поиска экстремума функции $f(x)$ на множестве

$$X=\{x\in E^n\colon g_1(x)\leqslant 0,\ldots, g_m(x)\leqslant 0\}. \tag{21}$$

Введем вспомогательные переменные $w=(w^1,\ldots,w^m)$ и в пространстве переменных $y=(x,w)=(x^1,\ldots,x^n,w^1,\ldots,w^m)$ рассмотрим задачу поиска экстремума функции $f(x)$ на множестве

$$Y=\{y=(x,w)\in E^{n+m}\colon q_i(y)=g_i(x)+(w^i)^2=0,\ i=1,\ldots,m\}. \tag{22}$$

Нетрудно видеть, что эта задача равносильна исходной задаче на множестве (21). В самом деле, если x_* — точка локального минимума [максимума] функции $f(x)$ на множестве (21), то точка $y_*=(x_*,w_*)$, где $w_*=(w_*^1,\ldots,w_*^m)$, $w_*^i=(-g_i(x_*))^{1/2}$, $i=1,\ldots,m$, будет точкой локального минимума [максимума] функции $f(x)$ на множестве (22) и наоборот, если $y_*=(x_*,w_*)$ — точка локального минимума [максимума] функции $f(x)$ на множестве (22), то x_* — точка локального минимума [максимума] функции $f(x)$ на множестве (21).

Допустим, что x_* — точка локального минимума функции $f(x)$ на множестве (21). Можем считать, что в (21) все ограничения в точке x_* активны, так как удаление неактивных ограничений из (21) не повлияет на свойство точки x_* быть локальным минимумом функции $f(x)$. Тогда точка $y_*=(x_*, w_*=0)\in Y$ будет точкой локального минимума функции $f(x)$ на множестве (22), и для нее будет справедливы приведенные в § 3, 4 теоремы для множеств, задаваемых ограничениями типа равенств. Применим их и посмотрим, что из этого получится для исходной задачи. Предположим, что градиенты $g_1'(x_*),\ldots,g_m'(x_*)$ линейно независимы. Тогда градиенты $q_i'(y_*)=\begin{pmatrix} g_i'(x_*)\\ 0\end{pmatrix}$, $i=1,\ldots,m$, также будут линейно независимы, т. е. $y_*=(x_*,0)$ — нормальная точка множества (22), и теперь можно будет воспользоваться теоремой 1. Составим функцию Лагранжа: $\mathcal{L}(y,\bar\lambda)=\lambda_0 f(x)+\sum\limits_{i=1}^{m}\lambda_i(g_i(x)+(w^i)^2)$, ее производными будут:

$$\mathcal{L}_x(y,\bar\lambda)=\lambda_0 f'(x)+\sum_{i=1}^{m}\lambda_i g_i'(x),\quad \mathcal{L}_w(y,\bar\lambda)=(2\lambda_1 w^1,\ldots,2\lambda_m w^m),$$

$$\mathcal{L}_y(y,\bar\lambda)=(\mathcal{L}_x(y,\bar\lambda),\mathcal{L}_w(y,\bar\lambda)),\qquad \mathcal{L}_{xx}(y,\bar\lambda)=\lambda_0 f''(x)+\sum_{i=1}^{m}\lambda_i g_i''(x),$$

$$\mathcal{L}_{ww}(y,\bar\lambda)=2\begin{pmatrix}\lambda_1 & & 0\\ & \ddots & \\ 0 & & \lambda_m\end{pmatrix},\quad \mathcal{L}_{yy}(y,\bar\lambda)=\begin{pmatrix}\mathcal{L}_{xx}(y,\bar\lambda) & 0\\ 0 & \mathcal{L}_{ww}(y,\bar\lambda)\end{pmatrix}.$$

Согласно теореме 3.1 существует такой набор $\bar\lambda=(\lambda_0,\ldots,\lambda_m)\neq 0$, $\lambda_0\geqslant 0$, $\mathcal{L}_y(y_*,\bar\lambda)=0$. Отсюда имеем $\mathcal{L}_x(y_*,\bar\lambda)=\lambda_0 f'(x_*)+\sum\limits_{i=1}^{m}\lambda_i g_i'(x_*)=0$, равенство $\mathcal{L}_w(y_*,\bar\lambda)=0$ выполняется автоматически и полезной информации не несет. Так как y_* — нормальная точка множества (22), то можем считать $\lambda_0=1$. Кроме того, у нас $g_i(x_*)=0$, $i=1,\ldots,m$, по предположению, и условия дополняющей нежесткости $\lambda_i g_i(x_*)=0$, $i=1,\ldots,m$, выполняются автоматически. А где же условия $\lambda_1\geqslant 0,\ldots,\lambda_m\geqslant 0$? Они, оказывается, при рассматриваемом подходе относятся к необходимым условиям второго порядка. В самом деле, применяя теорему 1, получим: $\langle\mathcal{L}_{yy}(y_*,\bar\lambda)h, h\rangle\geqslant 0$ $\forall h\in K_1(y_*)=\{h=(h_1,h_2)$: $h_1\in E^n$, $h_2\in E^m$, $\langle q_i'(y_*), h\rangle=0$, $i=1,\ldots,m\}=\{h=(h_1,h_2)$: $\langle g_i'(x_*), h_1\rangle+\langle 0, h_2\rangle=\langle g_i'(x_*), h_1\rangle=0$, $i=1,\ldots,m\}$. Отсюда следует, что $\langle\mathcal{L}_{xx}(y_*,\bar\lambda)h_1, h_1\rangle\geqslant 0$ $\forall h_1\in K_2(x_*)=\{h\in E^n$: $\langle g_i'(x_*), h\rangle=0$, $i=1,\ldots,m\}$, и $\langle\mathcal{L}_{ww}(y_*,\bar\lambda)h_2, h_2\rangle=2\lambda_1(h_2^1)^2+\ldots+2\lambda_m(h_2^m)^2\geqslant 0$ $\forall h_2\in E^m$, что возможно только при $\lambda_1\geqslant 0,\ldots,\lambda_m\geqslant 0$. Так как $\mathcal{L}_{xx}(y_*,\bar\lambda)=\mathcal{L}_{xx}(x_*,\bar\lambda)$, и, кроме того, рассуждая также, как в замечании 2, конус $K_2(x_*)$ без ущерба можем заменить на $K_3(x_*)=\{\langle f'(x_*), h\rangle\leqslant 0$, $\langle g_i'(x_*), h\rangle=0$, $i=1,\ldots,m\}$, то имеем $\langle\mathcal{L}_{xx}(x_*,\bar\lambda)h, h\rangle\geqslant 0$ $\forall h\in K_3(x_*)$. Конус $K_3(x_*)$, вообще говоря, беднее конуса из (15), поэтому полученные на этом пути необходимые условия второго порядка слабее, чем в теореме 3. Применение указанного приема ко множеству (10) в общем случае также приводит к более слабым необходимым условиям экстремума, чем это получено выше. Тем не менее простота этого приема привлекательна, и от него не стоит совсем отказываться при исследовании задач на экстремум при наличии ограничений типа неравенств.

Упражнения

1. Применить теоремы 2, 3 к исследованию задач из примеров, приведенных в § 3.

2. Исследовать задачи на экстремум, пользуясь правилом множителей Лагранжа и теоремами 2, 3, если:

а) $f(x)=\frac{1}{3}(x^1)^2+x^2$, $X=\{x=(x^1,x^2)\in E^2:\ g(x)=(x^1)^2+(x^2)^2-2\leqslant 0\}$;

б) $f(x)=x^1$, $X=\{x=(x^1,x^2)\in E^2:\ g_1(x)=(x^1)^2+(x^2)^2-1\leqslant 0,\ g_2(x)=(x^1)^3+(x^2)^3-1=0\}$;

в) $f_0(x)=f(x)-(x^{11})^2-(x^{12})^2-(x^{13})^2$, $X=\{x\in E^{13}:\ g_1(x)=0,\ g_2(x)=0\}$, где функции $f(x)$, $g_1(x)$, $g_2(x)$ взяты из примера 6; в определении функции $f(x)$ считать $c=0$.

3. Применить правило множителей Лагранжа и теоремы 2, 3 для поиска точек экстремума функций $f(u)=x$, $f(u)=x^2$, $f(x)=x^2+y^2$ на множестве $X=\{u=(x,y)\in E^2:\ g(u)=0[\leqslant 0, \geqslant 0]\}$, где $g(u)=x^p+y^q$ или $g(u)=x^p-y^q$, p, q — натуральные числа. Найти все нормальные и анормальные точки множества X.

4. Пусть $Q_0, Q_1, \ldots, Q_s$ — симметричные матрицы размера $n\times n$, пусть $K=\{x\in E^n:\ \langle Q_i x,x\rangle\leqslant 0,\ i=1,\ldots,m;\ \langle Q_i x,x\rangle=0,\ i=m+1,\ldots,s\}$. Доказать, что $\langle Q_0 x,x\rangle\geqslant 0$ на конусе K тогда и только тогда, когда точка $v=0$ является решением задачи:

$$f(x)=\langle Q_0 x,x\rangle\to\inf,\quad x\in K. \tag{23}$$

5. Пусть $\langle Q_0 x,x\rangle\geqslant 0\ \forall x\in K$ (обозначения см. в упражнении 4). Доказать, что для $\forall h\in E^n$ $\exists\bar\lambda=\bar\lambda(h)=(\lambda_0(h)\geqslant 0,\ldots,\lambda_m(h)\geqslant 0,\lambda_{m+1}(h),\ldots,\lambda_s(h))$, что $\left\langle\left(\sum_{i=0}^{s}\lambda_i(h)Q_i\right)h,h\right\rangle\geqslant 0$ и индекс квадратичной формы $\left\langle\left(\sum_{i=0}^{s}\lambda_i(h)Q_i\right)\xi,\xi\right\rangle$ не превышает $|I(0)|$, где $I(0)$ — множество активных ограничений из K в точке $v=0$, $|I(0)|$ — количество элементов множества $I(0)$. Указание: применить к задаче (23) теорему 3 в точке $v=0$.

6. Найти точки экстремума функций $f(x)=x$, $f(x)=x^2$, $f(x)=(x-3/2)^2$, $f(x)=(x-3)^2$, $f(x)=(x-1)(x-2)$ на множестве $X=\{x\in E^1:\ g_1(x)=-x\leqslant 0,\ g_2(x)=x^3-3x^2+2x\leqslant 0\}$. Обратить внимание, что $x=0$ является изолированной точкой множества X и ее можно считать точкой локального минимума [максимума] любой функции.

7. Пусть v — изолированное решение системы уравнений $g_i(x)=0$, $i=1,\ldots,s$; $n>s$. Доказать, что тогда для $\forall h\in K(v)=\{h\in E^n:\ \langle g_i'(v),h\rangle=0, i=1,\ldots,s\}$ существует $\lambda=\lambda(h)=(\lambda_1,\ldots,\lambda_s)$, что

$$\lambda(h)\neq 0,\quad \widetilde{\mathcal{L}}_x(v,\lambda(h))=\sum_{i=1}^{s}\lambda_i(h)g_i'(v)=0 \tag{24}$$

(т. е. v — анормальная точка множества (1)), $\widetilde\Lambda(v)\neq\varnothing$, $\max\limits_{\lambda\in\widetilde\Lambda_a(v),\,|\lambda|=1}\langle\widetilde{\mathcal{L}}_{xx}(v,\lambda)h,h\rangle\geqslant 0$ $\forall h\in K(v)$, где $\widetilde{\mathcal{L}}(x,\lambda)=\sum\limits_{i=1}^{s}\lambda_i g_i(v)$, $\widetilde\Lambda(v)$ — множество всех λ, удовлетворяющих условию (24), $\widetilde\Lambda_a(v)$ — подмножество тех $\lambda\in\widetilde\Lambda(v)$, для которых существует сопровождающее подпространство $\Pi=\Pi(\lambda)$ со свойствами (5)–(7) при замене в (7) $\mathcal{L}$ на $\widetilde{\mathcal{L}}$. Указание: убедиться, что точка v будет изолированной точкой множества (1) тогда и только тогда, когда v — точка локального минимума функции $f(x)=-|x-v|^2$, и затем к функции $f(x)$ на множестве (1) применить теорему 2.

8. Применяя к функции $f(x)=-|x-v|^2$ на множестве (10) теорему 3, получить необходимое условие изолированности точки v множества (10).

§ 5. Достаточные условия экстремума

Продолжим исследование задачи поиска экстремума функции $f(x)$ на множестве

$$X=\{x\in E^n:\ g_i(x)\leqslant 0,\ i=1,\ldots,m;\ g_i(x)=0,\ i=m+1,\ldots,s\}. \tag{1}$$

Приведенное в §§ 3, 4 условия первого и второго порядков являются лишь необходимыми условиями экстремума, и поэтому те точки, которые отбираются с их помощью, являются лишь подозрительными на экстремум и, как мы видели на примерах, в этих точках не всегда реализуется ожидаемый экстремум. Для выяснения характера экстремума в отобранных точках предназначены *достаточные условия*, в формулировке которых используются производные второго и более высокого порядков для функций $f(x)$, $g_i(x)$, $i=1,\ldots,s$. Здесь мы ограничимся достаточными условиями, которые формулируются с помощью второй производной функции Лагранжа.

Теорема 1. *Пусть функции $f(x)$, $g_i(x)$, $i=1,\ldots,s$, дважды непрерывно дифференцируемы в окрестности точки $v\in X$, пусть конус Лагранжа $\Lambda(v)$ этой точки непуст и*

$$\max_{\bar\lambda\in\Lambda(v),\,|\bar\lambda|=1}\langle\mathcal{L}_{xx}(v,\bar\lambda)h,h)\rangle>0\quad\forall h\neq 0,\quad h\in K(v),\quad\langle f'(v),h\rangle\leqslant 0, \tag{2}$$

где

$$K(v)=\{h\in E^n:\ \langle g_i'(v),h\rangle\leqslant 0,\ i\in I(v)\cap\{i:\ 1\leqslant i\leqslant m\},\ \langle g_i'(v),h\rangle=0,\ i=m+1,\ldots,s\},$$

$I(v)$ — множество индексов активных ограничений точки v. Тогда в этой точке реализуется строгий локальный минимум функции $f(x)$ на множестве (1).

Если $\Lambda(v)\neq\varnothing$ (см. замечания 3.1, 3.2) и, кроме того,

$$\max_{\bar\lambda\in\Lambda(v),\,|\bar\lambda|=1}\langle\mathcal{L}_{xx}(v,\bar\lambda)h,h\rangle>0\quad\forall h\neq 0,\quad h\in K(v),\quad\langle f'(v),h\rangle\geqslant 0, \tag{3}$$

то в точке v реализуется строгий локальный максимум функции $f(x)$ на множестве (1).

Замечание 1. Так как конус $\Lambda(v)\cup\{0\}$ замкнут, то множество $\{\bar\lambda\in\Lambda(v),|\bar\lambda|=1\}$ компактно и максимум в (2) при любом фиксированном h достигается хотя бы в одной точке $\bar\lambda=\bar\lambda(h)\in\Lambda(v)$, $|\bar\lambda(h)|=1$. Как мы видели в примерах 4.3, 4.8, одной «универсальной» точки $\bar\lambda$, в которой реализуется максимум в (2) одновременно для всех $h\in K(v)$, $\langle f'(v),h\rangle\leqslant 0$, может не быть. Аналогичное замечание справедливо и для условия (3).

Если в (1) ограничений типа неравенств нет ($m=0$), то в условиях (2), (3) ограничения $\langle f'(x),h\rangle\leqslant 0$ ($\geqslant 0$) могут быть опущены без ущерба — этот вопрос мы уже обсуждали в замечании 4.2. При $m=s=0$ теорема 1 превращается в теорему 2.2.

Доказательство теоремы 1. Если v — изолированная точка множества (1), то, по определению, v — точка строгого локального минимума [максимума]. Поэтому далее предполагается, что v не является изолированной точкой множества (1). Допустим, что условие (2) выполнено, но точка v не является точкой строгого локального минимума. Тогда найдется последовательность $\{x_k\}$, такая, что

$$x_k\neq v,\quad g_i(x_k)\leqslant 0,\quad i=1,\ldots,m,\quad g_i(x_k)=0,\quad i=m+1,\ldots,s$$
$$f(x_k)\leqslant f(v),\quad k=1,2,\ldots,\quad\{x_k\}\to v. \tag{4}$$

Точки x_k можем представить в виде $x_k=v+t_k d_k$, где $d_k=\frac{x_k-v}{|x_k-v|}$, $t_k=|x_k-v|\to 0$ при $k\to\infty$. Так как $|d_k|=1$, $k=1,2,\ldots$, то, выбирая при необходимости подпоследовательность согласно теореме Больцано — Вейерштрасса,

можем считать, что $\{d_k\} \to d_0$, $|d_0| = 1$. С учетом (4) и дифференцируемости функций $f(x)$, $g_i(x)$ в точке $x = v$ имеем

$$0 \geqslant f(x_k) - f(v) = \langle f'(v), d_k\rangle t_k + o(t_k),$$
$$0 \geqslant g_i(x_k) - g_i(v) = \langle g_i'(v), d_k\rangle t_k + o(t_k), \quad i \in I(v) \cap \{i\colon 1 \leqslant i \leqslant m\},$$
$$0 = g_i(x_k) - g_i(v) = \langle g_i'(v), d_k\rangle t_k + o(t_k), \quad i = m+1, \ldots, s, \quad k = 1, 2, \ldots$$

Разделив эти соотношения на $t_k > 0$ и устремив $k \to \infty$, получим:

$$\langle f'(v), d_0\rangle \leqslant 0, \quad \langle g_i'(v), d_0\rangle \leqslant 0, \quad i \in I(v) \cap \{i\colon 1 \leqslant i \leqslant m\},$$
$$\langle g_i'(v), d_0\rangle = 0, \quad i = m+1, \ldots, s, \quad |d_0| = 1.$$

Это означает, что d_0 принадлежит конусу $K_1(v) = \{v \in K(v), \langle f'(v), d_0\rangle \leqslant 0\}$ и $d_0 \neq 0$. Если конус $K_1(v)$ состоит лишь из точки 0, то уже получим противоречие. Пусть $K_1(v) \neq \{0\}$. Возьмем точку $\bar\lambda = \bar\lambda(d_0) \in \Lambda(v)$, $|\bar\lambda(d_0)| = 1$, в которой достигается максимум в (2) при $h = d_0$. Тогда $\langle \mathcal{L}_{xx}(v, \bar\lambda(d_0))d_0, d_0\rangle > 0$. С другой стороны, учитывая (4), неравенства $\lambda_i \geqslant 0$, $i = 0, \ldots, m$, и соотношения $g_i(v) = 0$ $\forall i \in I(v) \cap \{i\colon 1 \leqslant i \leqslant m\}$, $g_i(v) < 0$ и $\lambda_i = 0$ при $i \notin I(v)$, $\lambda_i g_i(v) = 0$, $i = 1, \ldots, m$, $g_i(v) = 0$, $i = m+1, \ldots, s$, имеем:

$$\mathcal{L}(x_k, \bar\lambda(d_0)) = \lambda_0 f(x_k) + \sum_{i=1}^{s} \lambda_i g_i(x_k) \leqslant \lambda_0 f(x_k) \leqslant \lambda_0 f(v) =$$
$$= \lambda_0 f(v) + \sum_{i=1}^{s} \lambda_i g_i(v) = \mathcal{L}(v, \bar\lambda(d_0)), \quad k = 1, 2, \ldots$$

Отсюда с помощью формулы Тейлора с учетом равенства $\mathcal{L}_{xx}(v, \lambda(d_0)) = 0$ получаем

$$0 \geqslant \mathcal{L}(x_k, \bar\lambda(d_0)) - \mathcal{L}(v, \bar\lambda(d_0)) = \tfrac{1}{2} t_k^2 \langle \mathcal{L}_{xx}(v, \bar\lambda(d_0))d_k, d_k\rangle + o(t_k^2), \quad k = 1, 2, \ldots$$

Разделив это неравенство на $t_k^2 > 0$ и устремив $k \to \infty$, будем иметь $\langle \mathcal{L}_{xx}(v, \bar\lambda(d_0))d_0, d_0\rangle \leqslant 0$, что противоречит (2) и определению $\bar\lambda(d_0)$. Следовательно, v — точка строгого локального минимума функции $f(x)$ на множестве (1).

Аналогично доказывается, что если для точки v выполнены условия (3), то v — точка строгого локального максимума функции $f(x)$ на множестве (1). Теорема 1 доказана. □

Для иллюстрации теоремы 1 приведем пример.

Пример 1. Задача: требуется найти точки экстремума функции $f(x) = \sum_{i=1}^{p} |x - x_i|^2$ на шаре $X = \{x \in E^n\colon |x|^2 \leqslant 1\}$; здесь $x_1, \ldots, x_p$ — заданные точки из E^n.

Эта задача была исследована в примере 4.3 с учетом конкретных особенностей задачи. Убедимся, что использование теоремы 1 упрощает анализ точек, подозрительных на экстремум. Поучительно также и то, что в некоторых точках теорема 1 «не работает» — с ее помощью не удается распознать характер экстремума точки. В этой задаче $\mathcal{L}_{xx}(x, \bar\lambda) = 2(\lambda_0 p + \lambda_1)I_n$, где I_n — единичная матрица $n \times n$. Переберем точки, подозрительные на экстремум, в том порядке, как они перечислены в примере 4.3.

1) точка $v_1 = x_0 = \frac{1}{p}\sum_{i=1}^{p} x_i$ при $|x_0| \leqslant 1$. Нам известны соответствующие v_1 два набора нормированных множителей Лагранжа: $\bar\lambda_{1,1} = (\lambda_0 = 1, \lambda_1 = 0)$ и $\bar\lambda_{1,2} = (\lambda_0 = -1, \lambda_1 = 0)$. Набор $\bar\lambda_{1,1} = (1, 0)$ принадлежит конусу $\Lambda(v_1)$; для него имеем $\langle \mathcal{L}_{xx}(v_1, \bar\lambda_{1,1})h, h\rangle = 2p|h|^2 > 0$ $\forall h \in E^n$, $h \neq 0$, в частности, для $\forall h \in K(v)$, $\langle f'(v), h\rangle \leqslant 0$. Отсюда следует, что условие (2) выполняется. Следовательно, v_1 — точка строгого локального минимума. А что мы получим, если точку v_1 аналогично проанализируем, используя набор $\bar\lambda_{1,2} = (-1, 0)$? Набор $\bar\lambda_{1,2}$ принадлежит конусу $\Lambda^-(v_1)$; для него $\langle \mathcal{L}_{xx}(v_1, \bar\lambda_{1,2})h, h\rangle = -2p|h|^2 < 0$ $\forall h \neq 0$. В этом случае конус $\Lambda^-(v_1) = \{\bar\lambda = t\bar\lambda_{1,2}, t > 0\}$. Отсюда видно, что условие (3) не будет выполняться на конусе $\Lambda^-(v_1)$. Это означает, что, используя набор $\bar\lambda_{1,2}$, с помощью теоремы 1 нам не удалось распознать характер экстремума точки v_1. Впрочем, нетрудно проверить, что здесь $\Lambda_a^-(v_1) = \varnothing$ и согласно теореме 4.3 и замечанию 4.3 точка v_1 не может быть точкой локального максимума.

2) $v_2 = \frac{x_0}{|x_0|}$ при $|x_0| > 1$, $\bar\lambda_2 = (\lambda_0 = 1, \lambda_1 = p(|x_0| - 1) > 0)$. Здесь $\bar\lambda_2 \in \Lambda(v_2)$ и $\langle \mathcal{L}_{xx}(v_2, \bar\lambda_2)h, h\rangle = 2p|x_0||h|^2 > 0$ $\forall h \neq 0$. Таким образом, здесь условие (2) выполнено, и v_2 — точка строгого локального минимума.

3) $v_3 = \frac{x_0}{|x_0|}$ при $0 < |x_0| < 1$, $\bar\lambda_3 = (-1, \lambda_1 = p(1 - |x_0|) > 0) \in \Lambda^-(v_3)$. Здесь $\langle \mathcal{L}_{xx}(v_3, \bar\lambda_3)h, h\rangle = -2p|x_0||h|^2 < 0$ $\forall h \neq 0$. Отсюда и из того, что здесь конус Лагранжа $\Lambda^-(v_3) = \{\bar\lambda = t\bar\lambda_3, t > 0\}$, заключаем, что условие (3) не выполняется. Таким образом, в точке v_3 теорема 1 «не работает». В примере 4.3 из других соображений было выяснено, что v_3 не является точкой экстремума. К такому же выводу мы придем, показав, что точка v_3 не удовлетворяет необходимым условиям второго порядка (теорема 4.3, замечание 4.3).

4) $v_4 = -\frac{x_0}{|x_0|}$ при $|x_0| > 0$, $\bar\lambda_4 = (\lambda_0 = -1, \lambda_1 = p(1 + |x_0|) > 0) \in \Lambda^-(v_4)$. Здесь $\langle \mathcal{L}_{xx}(v_4, \bar\lambda_4)h, h\rangle = 2p|h|^2 > 0$ $\forall h \neq 0$. Условие (3) заведомо выполняется. Следовательно, v_4 — точка строгого локального максимума.

5) при $x_0 = 0$ все точки v_5, $|v_5| = 1$, подозрительны на экстремум; им соответствует нормированный множитель Лагранжа $\bar\lambda_5 = (\lambda_0 = -1, \lambda_1 = p > 0)$. Здесь $\bar\lambda_5 \in \Lambda^-(v_5)$ и $\langle \mathcal{L}_{xx}(v_5, \bar\lambda_5)h, h\rangle = 0$ $\forall h \in E^n$ и $\forall \bar\lambda \in \Lambda^-(v_5)$. Отсюда ясно, что условие (3) не выполняется, и пользуясь лишь теоремой 1, мы не можем судить о характере экстремума в точках v_5. Здесь нам не могут помочь и необходимые условия второго порядка (теорема 4.3, замечание 4.3), которые, как нетрудно проверить, выполняются для всех точек v_5. В примере 4.3 из других соображений было выяснено, что при $x_0 = 0$ все точки единичной сферы $|x| = 1$ являются точкой глобального максимума и $f(x) \equiv f_*$.

При сравнении теорем 1 и 4.3 возникает интересный вопрос, насколько велик «зазор» между необходимыми и достаточными условиями второго порядка? Не вдаваясь в подробности, отметим, что исследования, проведенные в [43; 44], показывают, что для широких классов экстремальных задач этот «зазор» является минимально возможным при использовании вариаций, имеющих второй порядок малости.

В заключение этой главы заметим, что с помощью изложенных выше условий экстремума лишь в редких задачах удается найти и проанализировать все точки экстремума. Поэтому может создаться впечатление, что эти

условия имеют лишь теоретическое значение. Однако это не так. Как увидим ниже, многочисленные методы в той или иной степени представляют собой итерационные процессы, подсказанные условиями экстремума и предназначенные для решения систем уравнений и неравенств, составляющих суть этих условий. Нередко даже беглый теоретический анализ условий оптимальности позволяет получить немало информации о свойствах решений конкретной задачи, которая может быть использована при конструировании и реализации численных методов.

Упражнения

1. Применить теорему 1 для исследования задач из упражнений к §§ 2–4.

2. Пусть точка v принадлежит множеству (1), пусть функции $g_i(x)$, $i=1,\ldots,s$, дважды непрерывно дифференцируемы в окрестности точки v. Пусть конус $\tilde{\Lambda}(v)=\{\lambda=(\lambda_1,\ldots,\lambda_s)$: $\lambda\neq 0$, $\lambda_1\geqslant 0,\ldots,\lambda_m\geqslant 0$, $\sum_{i=1}^{s}\lambda_i g_i'(v)=0\}\neq\varnothing$ и $\max_{\lambda\in\tilde{\Lambda}(v),\,|\lambda|=1}\langle\left(\sum_{i=1}^{s}\lambda_i g_i'(v)\right)h,h\rangle>0$ $\forall h\neq 0$, $h\in K(v)=\{h\in E^n$: $\langle g_i'(v),h\rangle\leqslant 0$, $i\in I(v)\cap\{i$: $1\leqslant i\leqslant m\}$, $\langle g_i'(v),h\rangle=0$, $i=m+1,\ldots,s\}$. Доказать, что тогда v — изолированная точка множества (1). Указание: рассмотреть задачу минимизации функции $f(x)=-|x-v|^2$ на множестве (1) и применить к ней теорему 1 (см. упражнения 4.7, 4.8).

§ 6. Вспомогательные предложения

Ниже приводятся некоторые формулы и различные другие сведения, которые будут использованы в дальнейшем изложении.

1. Сначала напомним некоторые формулы для конечных приращений функций конечного числа переменных. Будем пользоваться обозначениями: $C^1(X)$ — множество всех функций, непрерывно дифференцируемых на множестве X, $C^2(X)$ — множество всех функций, дважды непрерывно дифференцируемых на множестве X.

Возьмем какую-либо функцию $f(x)$, определенную на множестве $X\subseteq E^n$. Пусть точки x, $x+h\in X$ таковы, что $h\neq 0$, $x+th\in X$ при всех t, $0\leqslant t\leqslant 1$. Тогда можно рассматривать функцию одной переменной $g(t)=f(x+th)$ при $t\in[0,1]$. Оказывается, если $f(x)\in C^p(X)$ при $p=1$ или $p=2$, то $g(t)\in C^p[0,1]$, причем

$$g'(t)=\langle f'(x+th),h\rangle,\quad g''(t)=\langle f''(x+th)h,h\rangle\quad 0\leqslant t\leqslant 1. \tag{1}$$

В самом деле, если, например, $f(x)\in C^2(X)$, то, заменив в формуле (2.5) x на $x+th$, h на Δth, получим

$$g(t+\Delta t)-g(t)=\Delta t\langle f'(x+th),h\rangle+\tfrac{1}{2}(\Delta t)^2\langle f''(x+th)h,h\rangle+o(|\Delta t|^2).$$

Такое разложение означает, что $g(t)\in C^2[0,1]$, и указывает на справедливость формул (1).

Для функции одной переменной имеют место формулы

$$g(t)-g(0)=g'(\theta_1 t)t=\int_0^t g'(\tau)d\tau=g'(0)t+\tfrac{1}{2}g''(\theta_2 t)t^2,$$

$$g'(t)-g'(0)=g''(\theta_3 t)t,\quad 0\leqslant\theta_1,\theta_2,\theta_3\leqslant 1.$$

Полагая в этих формулах $t=1$ и пользуясь равенствами (1), получаем различные формулы для конечных приращений функции многих переменных:

$$f(x+h)-f(x)=\langle f'(x+\theta_1 h),h\rangle=\int_0^1\langle f'(x+th),h\rangle dt, \tag{2}$$

$$f(x+h)-f(x)=\langle f'(x),h\rangle+\tfrac{1}{2}\langle f''(x+\theta_2 h)h,h\rangle, \tag{3}$$

$$\langle f'(x+h)-f'(x),h\rangle=\langle f''(x+\theta_3 h)h,h\rangle, \tag{4}$$

где $0\leqslant\theta_1,\theta_2,\theta_3\leqslant 1$. Далее, так как

$$\frac{d}{dt}(f'(x+th))=f''(x+th)h,\quad 0\leqslant t\leqslant 1,$$

то, интегрируя это равенство по t на отрезке $[0,1]$, получаем

$$f'(x+h)-f'(x)=\int_0^1 f''(x+th)hdt=\left(\int_0^1 f''(x+th)dt\right)h. \tag{5}$$

Подчеркнем еще раз, что в формулах (1)–(5) подразумевается, что точки x, $x+h$ принадлежат множеству X вместе с отрезком $x+th$, $0\leqslant t\leqslant 1$. В частности, эти формулы верны на любых выпуклых множествах — множествах, которые содержат вместе с любыми двумя своими точками u и v и отрезок $[u,v]=\{u_\alpha=\alpha u+(1-\alpha)v$, $0\leqslant\alpha\leqslant 1\}$, соединяющий эти точки (подробнее о выпуклых множествах см. § 4.1).

2. При описании и исследовании методов минимизации нам часто придется иметь дело с функциями, градиент которых удовлетворяет условию Липшица.

Определение 1. Пусть $f(x)\in C^1(X)$. Скажем, что градиент $f'(x)$ этой функции удовлетворяет *условию Липшица* на множестве X с постоянной $L\geqslant 0$, если

$$|f'(x)-f'(y)|\leqslant L|x-y|,\quad x,y\in X. \tag{6}$$

Класс таких функций будем обозначать через $C^{1,1}(X)$.

Лемма 1. *Пусть X — выпуклое множество, $f(x)\in C^{1,1}(X)$. Тогда*

$$|f(x)-f(y)-\langle f'(y),x-y\rangle|\leqslant L|x-y|^2/2 \tag{7}$$

при всех $x,y\in X$.

Доказательство. С помощью формулы (2) имеем

$$f(x)-f(y)-\langle f'(y),x-y\rangle=\int_0^1\langle f'(y+t(x-y))-f'(y),x-y\rangle dt.$$

Пользуясь неравенством Коши — Буняковского, с учетом условия (6) получим

$$|f(x)-f(y)-\langle f'(y),x-y\rangle|\leqslant$$

$$\leqslant\int_0^1|f'(y+t(x-y))-f'(y)||x-y|dt\leqslant\int_0^1 L|x-y|^2tdt=\frac{L|x-y|^2}{2}.\ \square$$

3. Приведем несколько лемм о числовых последовательностях, которые нам пригодятся при доказательстве сходимости методов минимизации, при оценке скорости их сходимости.

Лемма 2. *Пусть числовая последовательность $\{a_k\}$ такова, что*

$$a_{k+1}\leqslant a_k+\delta_k,\quad \delta_k\geqslant 0,\quad k=0,1,\ldots,\quad \sum_{k=0}^{\infty}\delta_k<\infty. \tag{8}$$

Тогда существует $\lim\limits_{k\to\infty} a_k < \infty$. *Если* $\{a_k\}$ *ограничена еще и снизу, то* $\lim\limits_{k\to\infty} a_k$ *конечен.*

Заметим, что если $\delta_k = 0$, $k = 0, 1, \ldots$, то последовательность $\{a_k\}$ не возрастает, и лемма 1 превращается в хорошо известное утверждение о пределе монотонной последовательности.

Доказательство. Суммируя первое из неравенств (8), имеем

$$a_m \leqslant a_k + \sum_{i=k}^{m-1} \delta_i \leqslant a_k + \sum_{i=k}^{\infty} \delta_i \tag{9}$$

при всех $m > k \geqslant 0$. Пусть $\lim\limits_{k\to\infty} a_k = \lim\limits_{n\to\infty} a_{k_n}$, $k_n < k_{n+1}$, $n = 0, 1, \ldots$; $\lim\limits_{n\to\infty} k_n = \infty$. Положим в (9) $k = k_n$. Получим $a_m \leqslant a_{k_n} + \sum\limits_{i=k_n}^{\infty} \delta_i$ $\forall m > k_n$. Следовательно, $\overline{\lim\limits_{m\to\infty}} a_m \leqslant a_{k_n} + \sum\limits_{i=k_n}^{\infty} \delta_i$ для всех $n = 1, 2, \ldots$ Отсюда при $n \to \infty$ имеем $\overline{\lim\limits_{m\to\infty}} a_m \leqslant \lim\limits_{n\to\infty} a_{k_n} = \lim\limits_{m\to\infty} a_m$. Но всегда $\lim\limits_{m\to\infty} a_m \leqslant \overline{\lim\limits_{m\to\infty}} a_m$, поэтому $\lim\limits_{m\to\infty} a_m = \overline{\lim\limits_{m\to\infty}} a_m$. Отсюда следует существование предела $\{a_k\}$. Далее, при $k = 0$ из (9) следует ограниченность $\{a_k\}$ сверху. Поэтому, если $\{a_k\}$ ограничена еще и снизу, то $\lim\limits_{k\to\infty} a_k$ конечен. □

Лемма 3. *Пусть числовая последовательность* $\{b_k\}$ *такова, что*

$$b_{k+1} \geqslant b_k - \delta_k, \quad \delta_k \geqslant 0 < \quad k = 0, 1, \ldots, \quad \sum_{k=0}^{\infty} \delta_k < \infty.$$

Тогда существует $\lim\limits_{k\to\infty} b_k > -\infty$. *Если* $\{b_k\}$ *ограничена еще и сверху, то* $\lim\limits_{k\to\infty} b_k$ *конечен.*

Эта лемма сводится к лемме 2, если принять $b_k = -a_k$, $k = 0, 1, \ldots$

Лемма 4. *Пусть числовая последовательность* $\{a_k\}$ *такова, что*

$$a_k \geqslant 0, \quad k = 0, 1, \ldots; \quad a_k - a_{k+1} \geqslant A a_k^2, \quad k \geqslant k_0 \geqslant 0. \tag{10}$$

Тогда $a_k = O(k^{-1})$, $k = 1, 2, \ldots$, *т. е. найдется постоянная* $B > 0$ *такая, что*

$$0 \leqslant a_k \leqslant Bk^{-1}, \quad k = 1, 2, \ldots \tag{11}$$

Доказательство. Если $a_m = 0$ при некотором $m \geqslant k_0$, то из (10) следует, что $a_k = 0$ при всех $k \geqslant m$, и оценка (11) становится тривиальной — в (11) достаточно взять $B = m \max\limits_{1\leqslant i\leqslant m} a_i$. Поэтому пусть $a_n > 0$ при всех $n \geqslant k_0$. Тогда из (10) имеем

$$\frac{1}{a_{n+1}} - \frac{1}{a_n} = \frac{a_n - a_{n+1}}{a_n a_{n+1}} \geqslant \frac{a_n}{a_{n+1}} A \geqslant A > 0, \quad n \geqslant k_0.$$

Суммируя эти неравенства по n от k_0 до некоторого $k - 1 \geqslant k_0$, получаем $a_k^{-1} - a_{k_0}^{-1} \geqslant A(k - k_0)$ или $a_k \leqslant A^{-1}(k - k_0)^{-1}$, $k > k_0$. Но $(k - k_0)^{-1} \leqslant (k_0 + 1)k^{-1}$ при $k > k_0$, поэтому $0 < a_k < (k_0 + 1)A^{-1}k^{-1}$. Если $1 \leqslant k \leqslant k_0$, то $0 \leqslant a_k = k a_k k^{-1} \leqslant k_0 (\max\limits_{1\leqslant k\leqslant k_0} a_k) k^{-1}$. Остается в (11) принять $B = \max\{(k_0 + 1)A^{-1}; k_0 \max\limits_{1\leqslant k\leqslant k_0} a_k\}$. □

Лемма 5. *Пусть числовая последовательность* $\{a_k\}$ *удовлетворяет условиям*

$$a_k \geqslant 0, \quad k \in N = \{1, 2, \ldots\}; \tag{12}$$

$$a_{k+1} \leqslant a_k - \frac{a_k^2}{A} + \frac{A}{k^{2\rho}}, \quad k \in I_0; \tag{13}$$

$$a_k \leqslant Bk^{-2\rho}, \quad k \in I_1; \tag{14}$$

$$a_{k+1} \leqslant a_k + Ck^{-2\rho}, \quad k \in I_1, \quad k+1 \in I_0, \tag{15}$$

где A, B, C, ρ *— положительные постоянные,* $\rho \leqslant 1$, *а множества индексов* I_0, I_1 *таковы, что* $I_0 \cup I_1 = N$, $I_0 \cap I_1 = \varnothing$ (*случаи* $I_0 = \varnothing$ *или* $I_1 = \varnothing$ *не исключаются*). *Тогда существует постоянная* $D > 0$ *такая, что*

$$0 \leqslant a_k \leqslant Dk^{-\rho}, \quad k = 1, 2, \ldots \tag{16}$$

Доказательство. Можем считать, что $A \geqslant B + C$, так как если неравенство (13) верно для некоторого $A = A_0 > 0$, то оно верно для всех $A > A_0$. Выберем натуральное число k_0 так, чтобы

$$4 \leqslant k_0^{\rho} < (k_0 + 1)^{\rho} \leqslant 6. \tag{17}$$

Убедимся в том, что такое число существует. Для этого перепишем (17) в равносильном виде: $4^{\varepsilon} \leqslant k_0 \leqslant 6^{\varepsilon} - 1$, где $\varepsilon = \rho^{-1} \geqslant 1$. Существование такого числа k_0 будет доказано, если покажем, что длина отрезка $[4^{\varepsilon}, 6^{\varepsilon} - 1]$ при любом $\varepsilon \geqslant 1$ не меньше 1, т. е. $6^{\varepsilon} - 4^{\varepsilon} - 1 \geqslant 1$ или $g(\varepsilon) = 6^{\varepsilon} - 4^{\varepsilon} \geqslant 2$ при всех $\varepsilon \geqslant 1$. Но $g'(\varepsilon) = 6^{\varepsilon} \ln 6 - 4^{\varepsilon} \ln 4 \geqslant \ln 6(6^{\varepsilon} - 4^{\varepsilon}) > 0$, $\varepsilon \geqslant 1$, так что $g(\varepsilon)$ строго монотонно возрастает при $\varepsilon \geqslant 1$. Следовательно, $g(\varepsilon) \geqslant g(1) = 2$ для всех $\varepsilon \geqslant 1$. Таким образом, при каждом ρ, $0 < \rho \leqslant 1$, число k_0, удовлетворяющее условиям (17), существует.

Покажем, что

$$a_{k_0+1} \leqslant 2A(k_0 + 1)^{-\rho}. \tag{18}$$

Может случиться, что $k_0 \in I_0$. Тогда воспользуемся неравенством (13). Заметим, что функция $f_k(a) = a - a^2 A^{-1} + A k^{-2\rho}$ достигает своего максимума на числовой оси при $a = A/2$, и поэтому $f_k(a) \leqslant f_k(A/2) = A/4 + A k^{-2\rho}$ для всех $a \geqslant 1$, $k = 1, 2, \ldots$ Тогда из (13) с учетом неравенств (17) имеем

$$a_{k_0+1} \leqslant f_k(a_{k_0}) \leqslant A/4 + A k_0^{-2\rho} \leqslant A/4 + A/16 \leqslant (5A/16)6(k_0+1)^{-\rho} < 2A(k_0+1)^{-\rho}.$$

Если же $k_0 \in I_1$, то возможно и $k_0 + 1 \in I_1$. Тогда из (14), (17) следует, что

$$a_{k_0+1} \leqslant B(k_0 + 1)^{-2\rho} \leqslant B(k_0 + 1)^{-\rho}/4 < 2A(k_0 + 1)^{-\rho}.$$

Если $k_0 \in I_1$, но $k_0 + 1 \in I_0$, то из (14), (15), (17) получим

$$a_{k_0+1} \leqslant Bk_0^{-2\rho} + Ck_0^{-2\rho} \leqslant A k_0^{-2\rho} < 2A(k_0 + 1)^{-\rho}.$$

Оценка (18) доказана. Далее, сделаем индуктивное предположение: пусть при некотором $k \geqslant k_0 + 1$ верна оценка $a_k \leqslant 2Ak^{-\rho}$. Возможно, что $k \in I_0$. Тогда с учетом (17) имеем $a_k \leqslant 2Ak^{-\rho} \leqslant 2Ak_0^{-\rho} \leqslant A/2$. Поскольку $f_k(a)$

монотонно возрастает на отрезке $[0, A/2]$, то из (13) следует, что $a_{k+1} \leqslant f_k(a_k) \leqslant f_k(2Ak^{-\rho}) = 2Ak^{-\rho} - 3Ak^{-2\rho} < 2A(k^{-\rho} - k^{-2\rho})$. Но при $0 < \rho \leqslant 1$ справедливы соотношения

$$k^{-\rho} - k^{-2\rho} < (k^\rho + 1)^{-1} < (k^\rho + k^{\rho-1})^{-1} = k^{-\rho+1}(k+1)^{-1} < (k+1)^{-\rho}, \qquad (19)$$

поэтому $a_{k+1} < 2A(k+1)^{-\rho}$. Если же $k \in I_1$ и $k+1 \in I_1$, то из (14), (17) получим

$$a_{k+1} \leqslant B(k+1)^{-2\rho} \leqslant B(k+1)^{-\rho}(k_0+1)^{-\rho} < 2A(k+1)^{-\rho}.$$

Если $k \in I_1$, но $k+1 \in I_0$, то из (14), (15), (17), (19) имеем

$$a_{k+1} \leqslant (B+C)k^{-2\rho} \leqslant Ak^{-2\rho} < A(k_0^\rho - 1)k^{-2\rho} < \\ < A(k^\rho - 1)k^{-2\rho} = A(k^{-\rho} - k^{-2\rho}) < 2A(k+1)^{-\rho}.$$

Этим показано, что $a_k \leqslant 2Ak^{-\rho}$ при всех $k \geqslant k_0 + 1$. Если $1 \leqslant k \leqslant k_0$, то $a_k = k^\rho a_k k^{-\rho} \leqslant k_0 k^{-\rho} \max\limits_{1 \leqslant k \leqslant k_0} a_k$. Остается в (16) принять $D = \max\{2A;\ k_0 \max\limits_{1 \leqslant k \leqslant k_0} a_k\}$. □

Лемма 6. *Пусть числовая последовательность* $\{w_k\}$ *такова, что*

$$0 \leqslant w_{k+1} \leqslant (1 - s_k)w_k + d_k, \quad k = 1, 2, \ldots, \quad w_1 \geqslant 0, \qquad (20)$$

где

$$0 < s_k \leqslant 1, \quad d_k \geqslant 0, \quad k = 1, 2, \ldots, \quad \sum_{k=1}^{\infty} s_k = \infty, \quad \lim_{k \to \infty} d_k/s_k = 0. \qquad (21)$$

Тогда $\lim\limits_{k \to \infty} w_k = 0$.

Доказательство. Поскольку $1 - x \leqslant e^{-x}$ при $0 \leqslant x \leqslant 1$, то $1 - s_k \leqslant e^{-s_k}$. Из неравенства (20) тогда имеем $0 \leqslant w_{k+1} \leqslant w_k e^{-s_k} + d_k$, $k = 1, 2, \ldots$ Отсюда с помощью индукции нетрудно получить, что

$$0 \leqslant w_{k+1} \leqslant \left(w_1 + \sum_{i=1}^{n} d_i \exp\left\{\sum_{j=1}^{i} s_j\right\}\right) \exp\left\{-\sum_{j=1}^{k} s_j\right\}, \quad k = 1, 2, \ldots \qquad (22)$$

Далее воспользуемся известной теоремой Штольца ([352, ч. 1, с. 88]), которая представляет собой разностный аналог правила Лопиталя и гласящей, что если последовательность $\{y_k\}$ монотонно возрастает, предел $\lim\limits_{k \to \infty} (x_k - x_{k-1})/(y_k - y_{k-1})$ существует, $\lim\limits_{k \to \infty} y_k = \infty$, то также существует и предел $\lim\limits_{k \to \infty} x_k/y_k$, причем

$$\lim_{k \to \infty} x_k/y_k = \lim_{k \to \infty} (x_k - x_{k-1})/(y_k - y_{k-1}).$$

Положим $y_k = \exp\left\{-\sum\limits_{j=1}^{k} s_j\right\}$, $x_k = w_1 + \sum\limits_{i=1}^{k} d_i \exp\left\{\sum\limits_{j=1}^{i} s_j\right\}$, $k = 1, 2, \ldots$ Из условий (21) следует, что $\{y_k\}$ монотонно возрастает и стремится к бесконечности. Кроме того,

$$\lim_{k \to \infty} \frac{x_k - x_{k-1}}{y_k - y_{k-1}} = \lim_{k \to \infty} d_k \exp\left\{+\sum_{j=1}^{k} s_j\right\}$$

$$\left(\exp\left\{-\sum_{j=1}^{k} s_j\right\} - \exp\left\{-\sum_{j=1}^{k-1} s_j\right\}\right)^{-1} = \lim_{k \to \infty} \frac{d_k}{1 - e^{-s_k}} = \lim_{k \to \infty} \left(\frac{d_k}{s_k}\right) \frac{s_k}{1 - e^{-s_k}} = 0,$$

так как функция $x/(1 - e^{-x})$ ограничена на множестве $0 < x \leqslant 1$. По теореме Штольца с учетом неравенства (22) получим

$$\lim_{k \to \infty} w_k = \lim_{k \to \infty} x_k/y_k = \lim_{k \to \infty} (x_k - x_{k-1})/(y_k - y_{k-1}) = 0.$$

Заметим, что неравенство (22) по существу представляет собой оценку скорости сходимости последовательности $\{w_k\}$. Однако правая часть оценки (22) трудно обозрима. Поэтому полезно иметь другие, быть может, более грубые, но более обозримые оценки. Здесь может быть полезна следующая простая

Лемма 7. *Пусть числовая последовательность* $\{w_k\}$ *такова, что*

$$0 \leqslant w_{k+1} \leqslant (1 - s_k)w_k + d_k, \quad k = 1, 2, \ldots, \quad w_1 \geqslant 0, \qquad (23)$$

где

$$0 \leqslant s_k \leqslant 1, \quad d_k \geqslant 0, \quad k = 1, 2, \ldots, \quad \sup_{k \geqslant 1} d_k/s_k = c < \infty. \qquad (24)$$

Тогда

$$0 \leqslant w_k \leqslant w_1 + c, \quad k = 1, 2, \ldots \qquad (25)$$

Доказательство легко проводится по индукции. При $k = 1$ оценка (25) очевидна. Если (25) верно для некоторого $k \geqslant 1$, то из (23), (24) следует $0 \leqslant w_{k+1} \leqslant (1 - s_k)(w_1 + c) + d_k \leqslant (1 - s_k)w_1 + (1 - s_k)c + cs_k \leqslant w_1 + c$, что и требовалось доказать. □

Покажем, как может быть применена лемма 7 для оценки конкретных последовательностей.

Лемма 8. *Пусть числовая последовательность* $\{a_k\}$ *такова, что*

$$0 \leqslant a_{k+1} \leqslant (1 - 1/k)a_k + c_1/k^2, \quad k = 1, 2, \ldots, \quad c_1 = \text{const} > 0, \quad a_1 \geqslant 0. \qquad (26)$$

Тогда справедлива оценка

$$0 \leqslant a_k \leqslant c_2 \ln(k+1)/k, \quad k = 1, 2, \ldots, \quad c_2 = \text{const} > 0. \qquad (27)$$

Доказательство. Сделаем замену $w_k = a_k k(\ln(k+1))^{-1}$ и, пользуясь леммой 7, докажем ограниченность $\{w_k\}$. Из (26) имеем

$$0 \leqslant w_{k+1} \leqslant \left(1 - \frac{1}{k}\right) \frac{k+1}{k} \frac{\ln(k+1)}{\ln(k+2)} w_k + c_1 \frac{k+1}{k^2 \ln(k+2)},$$

Таким образом, $\{w_k\}$ удовлетворяет условиям (23) при

$$s_k = 1 - \left(1 - \frac{1}{k}\right) \frac{\ln(k+1)}{\ln(k+2)}, \quad d_k = c_1 \frac{k+1}{k^2 \ln(k+2)}.$$

Нетрудно видеть, что $0 < s_k < 1$, $\lim\limits_{k \to \infty} d_k/s_k = c_1$, так что $\sup\limits_{k \geqslant 1} d_k/s_k = c_3 < \infty$. Из леммы 7 имеем $0 \leqslant w_k \leqslant w_1 + c_3$, $k = 1, 2, \ldots$, что равносильно оценке (27) с $c_2 = c_3 + a_1/\ln 2$. □

Лемма 9. *Пусть числовая последовательность* $\{a_k\}$ *такова, что*

$$0 \leqslant a_{k+1} \leqslant (1 - 1/k^\beta)a_k + c_1/k^{2\beta}, \quad k = 1, 2, \ldots, \\ c_1 = \text{const} > 0, \quad 0 < \beta < 1, \quad a_1 \geqslant 0. \qquad (28)$$

Тогда

$$0 \leqslant a_k \leqslant \left(a_1 + \frac{c_1}{1 - \beta}\right) \frac{1}{k^\beta}, \quad k = 1, 2, \ldots \qquad (29)$$

Доказательство. Сделаем замену $w_k = k^\beta a_k$. Тогда из (28) имеем

$$0 \leqslant w_{k+1} \leqslant \left(1 - \frac{1}{k^\beta}\right) \frac{(k+1)^\beta}{k^\beta} w_k + c_1 \frac{(k+1)^\beta}{k^{2\beta}}.$$

Это значит, что $\{w_k\}$ удовлетворяет условиям (23) при

$$s_k = 1 - \left(1 - \frac{1}{k^\beta}\right)\left(1 + \frac{1}{k}\right)^\beta < 1, \quad d_k = c_1\left(1 + \frac{1}{k}\right)^\beta \frac{1}{k^\beta}, \quad k = 1, 2, \ldots$$

Поскольку $(1 + 1/k)^{-\beta} \geqslant 1 - \beta/k$, $k = 1, 2, \ldots$, то

$$s_k = \left(1 + \frac{1}{k}\right)^\beta \left[\left(1 + \frac{1}{k}\right)^{-\beta} - 1 + \frac{1}{k^\beta}\right] \geqslant \left(1 + \frac{1}{k}\right)^\beta \left[1 - \frac{\beta}{k} - 1 + \frac{1}{k^\beta}\right] =$$
$$= \left(1 + \frac{1}{k}\right)^\beta \frac{1}{k^\beta}\left(1 - \frac{\beta}{k^{1-\beta}}\right) \geqslant d_k \frac{1}{c_1}(1 - \beta) > 0, \quad k = 1, 2, \ldots;$$

отсюда же следует, что $d_k/s_k \leqslant c_1/(1 - \beta)$, $k = 1, 2, \ldots$ По лемме 7 тогда $0 \leqslant w_k \leqslant w_1 + c_1/(1 - \beta)$, что равносильно оценке (29). □

Лемма 10. *Пусть $\{z_k\}$, $\{w_k\}$ — некоторые последовательности из евклидова пространства E^n, z_* — точка из E^n такая, что*

$$a|w_{k+1} - z_k|^2 + |w_{k+1} - z_*|^2 \leqslant |z_k - z_*|^2, \tag{30}$$

$$|z_{k+1} - w_{k+1}| \leqslant b\delta_k, \quad \delta_k \geqslant 0, \quad k = 0, 1, \ldots, \quad \sum_{k=0}^{\infty} \delta_k < \infty, \tag{31}$$

где a, b — положительные постоянные. Тогда существует конечный предел

$$\lim_{k \to \infty} |z_k - z_*| = \lim_{k \to \infty} |w_{k+1} - z_*| \tag{32}$$

и справедливо равенство

$$\lim_{k \to \infty} |w_{k+1} - z_k| = 0. \tag{33}$$

Если, кроме того, точка z_* из (30) является предельной для $\{z_k\}$, то обе последовательности $\{z_k\}$, $\{w_k\}$ сходятся к z_*.

Доказательство. Из (30) следует, что $|w_{k+1} - z_*| \leqslant |z_k - z_*|$. Тогда с помощью (31) имеем

$$|z_{k+1} - z_*| \leqslant |z_{k+1} - w_{k+1}| + |w_{k+1} - z_*| \leqslant b\delta_k + |z_k - z_*|, \tag{34}$$

или

$$|z_{k+1} - z_*| \leqslant |z_k - z_*| + b\delta_k, \quad k = 0, 1, \ldots$$

Отсюда и из леммы 2 при $a_k = |z_k - z_*|$ вытекает существование конечного предела $\lim\limits_{k \to \infty} |z_k - z_*|$. Из (34) следует (32), что в свою очередь гарантирует выполнение равенства (33). Наконец, пусть в (30) z_* — предельная точка $\{z_k\}$, пусть $|z_{k_m}| \to z_*$. Тогда $\lim\limits_{k \to \infty} |z_k - z_*| = \lim\limits_{m \to \infty} |z_{k_m} - z_*| = 0$, т. е. $\{z_k\} \to z_*$. Отсюда и из (32) получим, что $\{w_k\} \to z_*$. □

Лемма 11. *Пусть неотрицательное число z таково, что*

$$0 \leqslant z^p \leqslant bz + d, \tag{35}$$

где b, d — неотрицательные числа, $p > 1$. Тогда

$$0 \leqslant z \leqslant (b^q + qd)^{1/p}, \tag{36}$$

где q определяется равенством $p^{-1} + q^{-1} = 1$.

Доказательство. Если $d = 0$, то из (35) имеем $0 \leqslant z^{p-1} \leqslant b$, или $0 \leqslant z \leqslant b^{1/(p-1)} = b^{q/p}$, что совпадает с оценкой (36) при $d = 0$. Поэтому можем считать, что $d > 0$. Рассмотрим функцию

$$\varphi(x) = x^p - bx - d, \quad x \geqslant 0.$$

Нетрудно проверить, что эта функция имеет единственную точку минимума $x_* = (b/p)^{1/(p-1)} \geqslant 0$, $\varphi(x_*) \leqslant \varphi(0) = -d < 0$, $\lim\limits_{x \to \infty} \varphi(x) = \infty$, строго монотонно убывает при $0 \leqslant x \leqslant x_*$ (если $x_* > 0$), строго монотонно возрастает при $x \geqslant x_*$. Отсюда следует, что уравнение $\varphi(x) = 0$ имеет единственное решение $x = \gamma$, так что $\varphi(x) < 0$ при $0 \leqslant x < \gamma$ и $\varphi(x) > 0$ при $x > \gamma$. Согласно (35) $\varphi(z) \leqslant 0$, поэтому справедлива оценка $0 \leqslant z \leqslant \gamma$. Однако получить явное выражение для γ в общем случае не удается, поэтому в приложениях удобнее оценка (36). Для доказательства оценки (36) достаточно установить, что $\gamma \leqslant a = (b^q + qd)^{1/p}$. Пользуясь известным неравенством [327; 350; 393]

$$|ab| \leqslant |a|^p/p + |b|^q/q,$$

справедливым для всех действительных чисел a, b, $p > 1$, $q > 1$, $p^{-1} + q^{-1} = 1$, имеем $\varphi(a) = a^p - ab - d \geqslant a^p - a^p p^{-1} - b^q q^{-1} - d = a^p q^{-1} - (b^q + qd)q^{-1} = 0$. Следовательно, $a \geqslant \gamma$ и $0 \leqslant z \leqslant a = (b^q + qd)q^{1/p}$. □

ГЛАВА 3

Элементы линейного программирования

Изучение методов минимизации функций многих переменных начнем с методов решения сравнительно простых и достаточно хорошо изученных задач линейного программирования. Под *линейным программированием* понимается раздел теории экстремальных задач, в котором изучаются задачи минимизации (или максимизации) линейных функций на множествах, задаваемых системами линейных равенств и неравенств. Различные аспекты теории и методов линейного программирования, его приложения к технико-экономическим задачам изложены, например в [1; 13; 33; 48; 49; 52–54; 61; 76; 116; 135; 179; 203; 204; 214; 216; 231; 232; 243; 252; 259; 295; 297–299; 304; 317; 320; 330; 356; 361; 370; 373; 374; 398; 410; 422; 466; 470; 471; 487; 499; 506; 516; 517; 525; 541; 566; 584–586; 601; 612; 620; 636; 644; 652; 670; 676; 683; 685; 686; 688; 690; 719; 725; 736; 746; 747; 750–752; 775; 776; 796; 818].

§ 1. Постановка задачи

1. *Общая задача линейного программирования* может быть сформулирована следующим образом: минимизировать функцию

$$f(x)=c^1x^1+c^2x^2+\ldots+c^nx^n \tag{1}$$

при условиях

$$x^k\geqslant 0,\quad k\in I_+, \tag{2}$$

$$\left.\begin{aligned} &a_{11}x^1+a_{12}x^2+\ldots+a_{1n}x^n\leqslant b^1,\\ &\ldots\ldots\ldots\ldots\ldots\ldots\\ &a_{m1}x^1+a_{m2}x^2+\ldots+a_{mn}x^n\leqslant b^m, \end{aligned}\right\} \tag{3}$$

$$\left.\begin{aligned} &a_{m+11}x^1+a_{m+12}x^2+\ldots+a_{m+1n}x^n=b^{m+1},\\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\ &a_{s1}x^1+a_{s2}x^2+\ldots+a_{sn}x^n=b^s, \end{aligned}\right\} \tag{4}$$

где c^j, a_{ij}, b^i, $i=1,\ldots,s$, $j=1,\ldots,n$ заданные числа, причем не все из чисел c^j и не все из a_{ij} равны нулю, I_+ — заданное подмножество индексов из множества $\{1,2,\ldots,n\}$.

В частности, здесь возможно, что $I_+=\varnothing$ или $I_+=\{1,2,\ldots,n\}$; не исключаются случаи, когда отсутствуют ограничения типа равенств ($m=s$) или типа неравенств ($m=0$). Если ввести векторы $c=(c^1,\ldots,c^n)$, $a_i=(a_{i1},\ldots,a_{in})$, $x=(x^1,\ldots,x^n)$, то задачу (1)–(4) можно кратко записать так:

$$f(x)=\langle c,x\rangle\to\inf,\quad x\in X=\{x\in E^n\colon x^k\geqslant 0,\ k\in I_+,$$
$$\langle a_i,x\rangle\leqslant b^i,\ i=1,\ldots,m;\quad \langle a_i,x\rangle=b^i,\ i=m+1,\ldots,s\},$$

где $\langle c,x\rangle$, $\langle a_i,x\rangle$ — скалярное произведение соответствующих векторов.

Приведем еще одну форму матрично-векторной записи задачи (1)–(4). Предварительно договоримся о некоторых обозначениях. Если для каких-

либо двух векторов $x=(x^1,\ldots,x^p)$, $y=(y1,\ldots,y^p)$ справедливы неравенства $x^i\geqslant y^i$ при всех $i=1,\ldots,p$, то будем кратко писать: $x\geqslant y$. Тогда, например, неравенство $x\geqslant 0$ означает, что $x^i\geqslant 0$ для всех $i=1,\ldots,p$. Далее, не умаляя общности дальнейших рассмотрений, можем считать, что переменные $x^1,x^2,\ldots,x^n$ перенумерованы так, что $I_+=\{1,\ldots,n_1\}$, $0\leqslant n_1\leqslant n$ ($n_1=0$ соответствует случаю $I_+=\varnothing$). Отдельно выделяя неотрицательные координаты, вектор x можем представить так: $x=(x_1,x_2)$, $x_1=(x_1^1,x_1^2,\ldots,x_1^{n_1})\in E^{n_1}$, $x_2=(x_2^1,x_2^2,\ldots,x_2^{n_2})\in E^{n_2}$, $x_1\geqslant 0$, $n_1+n_2=n$. Используя принятые обозначения, задачу (1)–(4) можем записать в следующем виде:

$$f(x)=\langle c_1,x_1\rangle+\langle c_2,x_2\rangle\to\inf,\quad x=(x_1,x_2)\in X, \tag{5}$$

$$X=\{x=(x_1,x_2)\colon x_1\in E^{n_1},\ x_2\in E^{n_2},\ A_{11}x_1+A_{12}x_2\leqslant b_1,\ A_{21}x_1+A_{22}x_2=b_2,\ x_1\geqslant 0\}, \tag{6}$$

где A_{ij} — матрица размером $m_i\times n_j$, $b_i\in E^{m_i}$, $c^j\in E^{n_j}$, $i,j=1,2$; $m_1=m$, $m_1+m_2=s$,

$$A_{11}=\begin{bmatrix} a_{11},&\ldots,&a_{1n_1}\\ \ldots\ldots\ldots\ldots\\ a_{m1},&\ldots,&a_{mn_1}\end{bmatrix},\qquad A_{12}=\begin{bmatrix} a_{1n_1+1},&\ldots,&a_{1n}\\ \ldots\ldots\ldots\ldots\\ a_{mn_1+1},&\ldots,&a_{mn}\end{bmatrix},$$

$$A_{21}=\begin{bmatrix} a_{m+11},&\ldots,&a_{m+1n_1}\\ \ldots\ldots\ldots\ldots\\ a_{s1},&\ldots,&a_{sn_1}\end{bmatrix},\qquad A_{22}=\begin{bmatrix} a_{m+1n_1+1},&\ldots,&a_{m+1n}\\ \ldots\ldots\ldots\ldots\\ a_{sn_1+1},&\ldots,&a_{sn}\end{bmatrix},$$

$$b_1=\begin{bmatrix} b^1\\ \ldots\\ b^m\end{bmatrix},\quad b_2=\begin{bmatrix} b^{m+1}\\ \ldots\\ b^s\end{bmatrix},\quad c_1=\begin{bmatrix} c^1\\ \ldots\\ c^{n_1}\end{bmatrix},\quad c_2=\begin{bmatrix} c^{n_1+1}\\ \ldots\\ c^n\end{bmatrix}.$$

Подчеркнем, что в (6) и всюду ниже в произведениях вида $A_{i1}x_1$, $A_{i2}x_2$, Ax, $By,\ldots$ матриц A_{i1}, A_{i2}, A, $B\ldots$ на соответствующие вектора $x_1,x_2,x,y,\ldots$ будем подразумевать, что $x_1,x_2,x,y,\ldots$ — это векторы-столбцы подходящей размерности, хотя для экономии места, как мы уже делали выше, часто будем записывать эти векторы в виде строки.

Укажем еще на одну форму записи множества (6)

$$X=\{x=(x_1,x_2)\colon x_1\in E^{n_1},\ x_2\in E^{n_2},$$
$$\sum_{k=1}^{n_1}A_{11}^kx_1^k+\sum_{k=1}^{n_2}A_{12}^kx_2^k\leqslant b_1,\quad \sum_{k=1}^{n_1}A_{21}^kx_1^k+\sum_{k=1}^{n_2}A_{22}^kx_2^k=b_2,\ x_1\geqslant 0\},$$

где A_{ij}^k — k-й столбец матрицы A_{ij}.

Точку $x_*\in X$ назовем *точкой минимума* функции $\langle c,x\rangle$ на множестве X или, короче, *решением* задачи (5), (6) если $\langle c,x_*\rangle=\inf\limits_X\langle c,x\rangle$.

2. Приведем примеры прикладных задач, приводящих к задачам линейного программирования.

Задача оптимального планирования производства. Пусть на некотором предприятии изготовляются n видов продукции из s видов сырья. Известно, что на изготовление одной единицы продукции j-го вида нужно a_{ij} единиц сырья i-го вида. В распоряжении предприятия имеется b_i единиц сырья i-го вида. Известно также, что на каждой единице продукции j-го вида предприятие получает c_j единиц прибыли. Требуется определить, сколько единиц $x^1,\ldots,x^n$ каждого вида продукции должно изготовить предприятие, чтобы обеспечить себе максимальную прибыль.

Если предприятие наметит себе план производства $x=\{x^1,\ldots,x^n\}$, то оно израсходует $a_{i1}x^1+\ldots+a_{in}x^n$ единиц сырья i-го вида и получит $c_1x^1+\ldots+c_nx^n$ единиц прибыли. Ясно также, что все величины x^i, $i=1,\ldots,n$, неотрицательны. Поэтому мы приходим к следующей задаче линейного программирования: максимизировать функцию $f(x)=c_1x^1+\ldots+c_nx^n$ при ограничениях $x^1\geqslant 0,\ldots,x^n\geqslant 0$, $a_{i1}x^1+\ldots+a_{in}x^n\leqslant b^i$, $i=1,\ldots,s$. Поскольку задача максимизации функции $f(x)$ равносильна задаче минимизации функции $-f(x)$, то с учетом введенных выше обозначений сформулированную задачу линейного программирования можно кратко записать в виде

$$\langle -c,x\rangle\to\inf;\quad x\in X=\{x\in E^n\colon x\geqslant 0,\ Ax\leqslant b\}. \tag{7}$$

Ясно, что задача (7) является частным случаем задачи (5), (6).

Задача об оптимальном использовании посевной площади. Пусть под посев p культур отведено r земельных участков площадью соответственно в $b_1,\ldots,b_r$ гектаров. Известно, что средняя урожайность i-й культуры на j-м участке составляет a_{ij} центнеров с гектара, а прибыль за один центнер i-й культуры составляет c_i рублей. Требуется определить, какую площадь на каждом участке следует отвести под каждую из культур, чтобы получить максимальную прибыль, если по плану должно быть собрано не менее d_i центнеров i-й культуры.

Обозначим через u_{ij} площадь, которую планируется отвести под i-ю культуру на j-м участке. Тогда

$$u_{1j}+\ldots+u_{pj}=b_j,\qquad j=1,\ldots,r. \tag{8}$$

Ожидаемый средний урожай i-й культуры со всех участков равен $a_{i1}u_{i1}+\ldots+a_{in}u_{in}$ центнеров. Поскольку согласно плану должно быть произведено не менее d_i центнеров i-й культуры, то

$$a_{i1}u_{i1}+\ldots+a_{in}u_{in}\geqslant d_i\quad i=1,\ldots,p. \tag{9}$$

Ожидаемая прибыль за урожай i-й культуры равна $c_i(a_{i1}u_{i1}+\ldots+a_{in}u_{in})$, а за урожай всех культур —

$$\sum_{i=1}^{p}c_i(a_{i1}u_{i1}+\ldots+a_{in}u_{in})=f(u). \tag{10}$$

Таким образом, приходим к задаче максимизации функции (10) (или минимизации функции $-f(u)$) при условиях (8), (9) и естественных ограничениях

$$u_{ij}\geqslant 0,\quad i=1,\ldots,p,\ j=1,\ldots,r.$$

Если умножить соотношения (9) на -1 и переменные $\{u_{ij}\}$ переобозначить через $x^1,\ldots,x^n$, то придем к задаче вида (1)–(4).

Транспортная задача. Пусть имеется r карьеров, где добывается песок, и p потребителей песка (например, кирпичные заводы). В i-м карьере ежесуточно добывается a_i тонн песка, а j-му потребителю ежесуточно требуется b_j тонн песка. Пусть c_{ij} — стоимость перевозки одной тонны песка с i-го карьера j-му потребителю. Требуется составить план перевозок песка так, чтобы общая стоимость перевозок была минимальной.

Обозначим через u_{ij} планируемое количество тонн песка из i-го карьера j-му потребителю. Тогда с i-го карьера будет вывезено

$$u_{i1}+\ldots+u_{ip}=a_i,\quad i=1,\ldots,r, \tag{11}$$

тонн песка, j-му потребителю доставлено

$$u_{1j}+\ldots+u_{rj}=b_j,\quad j=1,\ldots,p, \tag{12}$$

тонн песка, а стоимость перевозок будет равна

$$f(u)=\sum_{j=1}^{p}\sum_{i=1}^{r}c_{ij}u_{ij}. \tag{13}$$

Естественно требовать, чтобы

$$u_{ij}\geqslant 0,\quad i=1,\ldots,r,\ j=1,\ldots,p. \tag{14}$$

В результате получили задачу минимизации функции (13) при условиях (11), (12), (14), которая, очевидно, является частным случаем общей задачи линейного программирования (1)–(4).

К задачам типа (1)–(4) сводятся также и многие другие прикладные задачи технико-экономического содержания.

Следует заметить, что приведенные выше примеры задач линейного программирования, вообще говоря, представляют лишь приближенную, упрощенную математическую модель реальных задач. Вполне может оказаться, что принятая математическая модель, обычно составляемая на основе приближенных данных о реальном моделируемом явлении (объекте, процессе), не охватывает какие-либо важные существенные стороны исследуемого явления и приводит к результатам, существенно расходящимся с реальностью. В этом случае математическая модель должна быть изменена, доработана с учетом вновь поступившей информации, а получаемые при анализе совершенствованной модели данные должны снова и снова критически сопоставляться с реальными данными и использоваться для выяснения границ применимости модели. Математическая модель лишь при высокой степени адекватности моделируемому явлению может быть использована для более глубокого анализа явления и проникновения в его сущность, для выработки целенаправленного управления.

Практика показала, что линейные модели, приводящие к задачам вида (1)–(4), вполне пригодны для исследования многих реальных явлений, или для их анализа могут быть использованы теория и методы линейного программирования. Линейное программирование является одним из наиболее изученных разделов теории экстремальных задач с достаточно богатым арсеналом методов. Ниже мы увидим, что задачи линейного программирования нередко используются в качестве вспомогательных во многих методах решения более сложных нелинейных задач минимизации.

3. Из общей задачи линейного программирования обычно выделяют так называемую *каноническую* задачу:

$$f(x)=\langle c,x\rangle\to\inf;\quad x\in X=\{x\in E^n\colon x\geqslant 0,\ Ax=b\}, \tag{15}$$

получающуюся из задачи (5), (6) при $m=0$, $I_+=\{1,2,\ldots,n\}$. Задача (15) привлекательна тем, что при ее исследовании, разработке методов ее решения можно пользоваться хорошо известной из линейной алгебры теорией систем линейных алгебраических уравнений. Замечательно также и то, что методы, созданные для решения канонической задачи (15), нетрудно модифицировать и применять для решения общей задачи линейного программирования (5), (6). Дело в том, что задача (5), (6) оказывается сама равносильна некоторой канонической задаче. Покажем это.

Для того, чтобы легче было понять последующие построения, прежде всего заметим, что любое действительное число a можно представить в виде разности двух неотрицательных чисел: $a = a^+ - a^-$, где $a^+ = \max\{0; a\} \geqslant 0$, $a^- = \max\{0; -a\} \geqslant 0$. Отсюда следует, что вектор $x_2 = (x_2^1, \ldots, x_2^{n_2})$ можно представить в виде разности неотрицательных векторов:

$$x_2 = z_1 - z_2, \quad z_1 = \max\{0; x_2\} \geqslant 0, \quad z_2 = \max\{0; -x_2\} \geqslant 0, \tag{16}$$

где операция взятия максимума проводится покоординатно: $z_1 = (z_1^1, \ldots, z_1^{n_2})$, $z_1^j = \max\{0; x_2^j\}$, $z_2 = (z_2^1, \ldots, z_2^{n_2})$, $z_2^j = \max\{0; -x_2^j\}$, $j = 1, \ldots, n_2$. Далее, заметим, что ограничения $Ax \leqslant b$ типа неравенств можно записать в виде ограничений типа равенств $Ax + v = b$, добавив сюда неравенство $v \geqslant 0$: ясно, что точка x будет решением неравенства $Ax \leqslant b$ тогда и только тогда, когда (x, v) — решение системы $Ax + v = b$, $v \geqslant 0$. Отсюда следует, что вводя переменную

$$v = b_1 - A_{11}x_1 - A_{12}x_2 \geqslant 0, \tag{17}$$

ограничение $A_{11}x_1 + A_{12}x_2 \leqslant b_1$ с учетом (16) можно представить в равносильном виде

$$A_{11}x_1 + A_{12}x_2 + v = A_{11}x_1 + A_{12}z_1 + (-A_{12})z_2 + v = b_1, \quad v \geqslant 0.$$

Ограничение $A_{21}x_1 + A_{22}x_2 = b_2$ с учетом (16) запишем в виде:

$$A_{21}x_1 + A_{22}z_2 + (-A_{22})z_2 + 0v = b_2.$$

Учитывая эти соображения, в пространстве переменных $w = (x_1, z_1, z_2, v)$, $x_1 \in E^{n_1}$, $z_1 \in E^{n_2}$, $z_2 \in E^{n_2}$, $v \in E^{m_1}$, рассмотрим следующую каноническую задачу:

$$g(w) = \langle c_1, x_1\rangle + \langle c_2, z_1\rangle + \langle -c_2, z_2\rangle + \langle 0, v\rangle \to \inf, \quad w \in W, \tag{18}$$

$$W = \{w = (x_1, z_1, z_2, v)\colon A_{11}x_1 + A_{12}z_1 + (-A_{12})z_2 + Iv = b_1,$$
$$A_{21}x_1 + A_{22}z_1 + (-A_{22})z_2 = b_2,\ w \geqslant 0\}, \tag{19}$$

где I — единичная матрица размера $m_1 \times m_1$. Оказывается, задачи (5), (6) и (18), (19) обе одновременно имеют или не имеют решение, причем, зная какое-либо решение одной из этих задач, нетрудно получить решение другой задачи. Точнее, справедлива следующая

Теорема 1. *Задачи* (5), (6) *и* (18), (19) *равносильны, т. е.:*
1) *множества X и W оба пусты или непусты одновременно;*
2) *если $X \neq \varnothing$, $W \neq \varnothing$, то $f_* = g_*$, где $f_* = \inf\limits_{u \in U} f(u)$, $g_* = \inf\limits_{w \in W} g(w)$;*
3) *множества решений $X_* = \{u \in X\colon f(x) = f_*\}$, $W_* = \{w \in W\colon g(w) = g_*\}$ этих задач оба пусты или непусты одновременно, причем если $x_* = (x_{1*}, x_{2*}) \in X_*$, то $w_* = (x_{1*}, z_{1*}, z_{2*}, v_*) \in W_*$, где $z_{1*} = \max\{0; x_{2*}\}$, $z_{2*} = \max\{0; -x_{2*}\}$, $v_* = b_1 - A_{11}x_{1*} - A_{12}x_{2*}$, и обратно, если $w_* = (x_{1*}, z_{1*}, z_{2*}, v_*) \in W_*$, то $x_* = (x_{1*}, x_{2*} = z_{1*} - z_{2*}) \in X_*$.*

Доказательство. Учитывая связи (16),(17) между переменными x_1, x_2, z_1, z_2, v и определения (6), (19) множеств X, W заключаем, что если точка $x = (x_1, x_2) \in X$, то $w = w(x) = (x_1, z_1 = \max\{0; x_2\}, z_2 = \max\{0; -x_2\}, v = b_1 - A_{11}x_1 - A_{12}x_2) \in W$. И обратно, если $w = (x_1, z_1, z_2, v) \in W$, то $x = x(w) = (x_1, x_2 = z_1 - z_2) \in X$. Отсюда ясно, что либо оба множества X и W пусты, либо оба непусты одновременно. Далее, из определений функций $f(x)$, $g(w)$, $w(x)$, $x(w)$ следуют тождества

$$f(x) \equiv g(w(x)), \quad g(w) \equiv f(x(w)), \quad \forall x, w. \tag{20}$$

Допустим, что $f_* = -\infty$. Тогда существует последовательность $\{x_k\}$: $x_k \in X$, $k = 1, 2, \ldots$, такая, что $\{f(x_k)\} \to f_* = -\infty$. Положим $w_k = w(x_k)$, $k = 1, 2, \ldots$. Из (20) тогда имеем $g(w_k) = g(w(x_k)) = f(x_k) \to -\infty$, откуда с учетом включения $w_k \in W$, $k = 1, 2, \ldots$, получим, что $g_* = -\infty$. Аналогично рассуждая, заключаем, что если $g_* = -\infty$, то $f_* = -\infty$.

Пусть далее $f_* > -\infty$, $g_* > -\infty$. Возьмем произвольное число $\varepsilon > 0$. По определению нижней грани найдется точка $x_\varepsilon \in X$, такая, что $f_* \leqslant f(x_\varepsilon) < f_* + \varepsilon$. Тогда $w_\varepsilon = w(x_\varepsilon) \in W$ и из (20) следует, что $g_* \leqslant g(w_\varepsilon) = f(x_\varepsilon) < f_* + \varepsilon$, т. е. $g_* < f_* + \varepsilon$. Аналогично, по определению g_* существует точка $\overline{w}_\varepsilon$ для которой $g_* \leqslant g(\overline{w}_\varepsilon) < g_* + \varepsilon$. Тогда $x_\varepsilon = x(\overline{w}_\varepsilon) \in X$ и $f_* \leqslant f(x_\varepsilon) = f(x(\overline{w}_\varepsilon)) = g(\overline{w}_\varepsilon) < g_* + \varepsilon$, т. е. $f_* < g_* + \varepsilon$. Следовательно, $f_* - \varepsilon < g_* < f_* + \varepsilon$. В силу произвольности $\varepsilon > 0$ отсюда вытекает, что $f_* = g_* > -\infty$.

Наконец, если $x_* \in X_*$, то $w_* = w(x_*) \in W$, и в силу вышедоказанного $g(w_*) = g(w(x_*)) = f(x_*) = f_* = g_*$. Это значит, что $w(x_*) \in W_*$ при любом $x_* \in X_*$. Аналогично доказывается, что если $w_* \in W_*$, то $x_* = x(w_*) \in X_*$. Отсюда следует, что либо оба множества X_* и W_* пусты или оба непусты одновременно. Теорема 1 доказана. □

4. В теории и методах линейного программирования наряду с канонической задачей принято выделять так называемую *основную* (или *стандартную*) задачу линейного программирования:

$$f(x) = \langle c, x\rangle \to \inf, \quad x \in X = \{x \geqslant 0\colon \overline{A}x \leqslant \overline{b}\}, \tag{21}$$

получающуюся из общей задачи (5), (6) при $m = s$, $I_+ = \{1, 2, \ldots, n\}$. Это объясняется тем, что в приложениях большое число линейных математических моделей изначально естественным образом записывается в виде задачи (21) (см., например, задачу (7)). Следует также отметить, что задача (21) весьма удобна для геометрических интерпретаций, делающих наглядными многие понятия и методы линейного программирования.

Если ввести дополнительные переменные $v = (v^1, \ldots, v^m)$ посредством соотношений

$$v = \overline{b} - \overline{A}x, \quad v \geqslant 0, \tag{22}$$

то задачу (21) в пространстве E^{n+m} переменных $z = (x, v)$ можно записать в канонической форме:

$$g(z) = \langle d, z\rangle \to \inf, \quad z \in Z,$$
$$Z = \{z = (x, v) \geqslant 0, \quad Cz = \overline{A}x + I_m v = \overline{b}\}, \tag{23}$$

где $d = (c, 0) \in E^{n+m}$, $C = (\overline{A}, I_m)$, I_m — единичная матрица размера $m \times m$. Из теоремы 1 следует, что задачи (21), (23) равносильны, и зная решение $x_* \in X_*$ задачи (21) по формуле (22) нетрудно получить решение задачи (23) $z_* = (x_*, v_* = \overline{b} - \overline{A}x_*)$, и обратно, если $z_* = (x_*, v_*) \in Z_*$, то $x_* \in X_*$.

С другой стороны, каноническую задачу (15) нетрудно записать в форме основной задачи. В самом деле, если ограничения типа равенств $Ax = b$ заменить на равносильную систему двух неравенств: $Ax \leqslant b$, $Ax \geqslant b$, то

тогда задачу (15) можно записать в следующем виде

$$f(x)=\langle c, x\rangle \to \inf, \quad x\in X=\{x\geqslant 0:\ Ax\leqslant b,\ (-A)x\leqslant -b\}=$$
$$=\{x\geqslant 0:\ Cx\leqslant f\},\quad C=\begin{bmatrix} A\\ -A\end{bmatrix},\quad f=\begin{bmatrix} b\\ -b\end{bmatrix}. \qquad (24)$$

Рассуждая также, как при доказательстве теоремы 1, нетрудно установить равносильность задач (15), (24).

Как видим, все три формы задач линейного программирования — общая задача (5),(6), каноническая (15), основная (21) — тесно связаны между собой и простыми преобразованиями от одной формы легко перейти к другой. Поэтому, если мы научимся решать одну из этих задач, то тем самым будем уметь решать задачу линейного программирования, записанную в любой другой форме. Заметим однако, что изложенные приемы сведения задач (5), (6), (15), (21) к канонической или основной задаче могут привести к чрезмерному увеличению размерности переменных или числа ограничений. Поэтому методы решения задач линейного программирования обычно разрабатывают для задач (15) или (21), как более простых для исследования, а затем, учитывая указанную связь между задачами (5), (6), (15), (21), модифицируют полученные методы применительно к другим классам задач линейного программирования, стараясь, по возможности, не увеличивать их размерность.

Упражнения

1. Требуется составить наиболее дешевую смесь, содержащую не менее b^i единиц i-го вещества, $i=1, 2, \ldots, m$, при условии, что для изготовления смеси имеется n видов продукции, причем в одной единице j-го продукта содержится a_{ij} единиц i-го вещества, а цена одной единицы j-го продукта равна c^j рублей (задача о смесях). Сформулировать эту задачу в виде основной задачи (21).

2. Задачу $f(x)=x^1+x^2+x^3+x^4-x^5\to \sup$, $x=(x^1, x^2, x^3, x^4, x^5)\in X=\{x^1\geqslant 0,\ x^3\geqslant 0,\ x^4\geqslant 0;\ x^1+x^2-x^3\leqslant 1,\ x^1+x^4+x^5=3,\ x^1-x^3+x^5\geqslant 1,\ -1\leqslant x^2\leqslant 1,\ x^5\geqslant 1\}$ записать в виде задач (5), (6), (15), (21).

3. Записать общую задачу (5), (6) линейного программирования в форме основной задачи, доказать их эквивалентность.

§ 2. Геометрическая интерпретация. Угловые точки

1. Кратко остановимся на геометрическом смысле задачи линейного программирования. Рассмотрим задачу (1.21) при $n=2$:

$$f(x)=c^1x^1+c^2x^2\to \inf$$
$$x\in X=\{x=(x^1, x^2):\ x^1\geqslant 0,\ x^2\geqslant 0,\ a_{i1}x^1+a_{i2}x^2\leqslant b^i,\ i=1, \ldots, m\} \qquad (1)$$

Введем множества: $X_0=\{x=(x^1, x^2):\ x^1\geqslant 0,\ x^2\geqslant 0\}=E^2_+$ — неотрицательный квадрант плоскости; (x^1, x^2), $X_i=\{x=(x^1, x^2):\ a_{i1}x^1+a_{i2}x^2\leqslant b^i\}$ — полуплоскость, образуемая прямой $a_{i1}x^1+a_{i2}x^2=b^i$, $i=1, \ldots, m$. Ясно, что множество X является пересечением множеств $X_0, X_1, \ldots, X_m$. Может случиться, что это пересечение пусто (рис. 3.1) — тогда задача (1) теряет смысл. Если множество X непусто, то оно, образованное пересечением конечного числа полуплоскостей, представляет собой выпуклое многоугольное множество, границей которого является ломаная, составленная из отрезков каких-либо координатных осей и прямых $a_{i1}x^1+a_{i2}x^2=b^i$, $i=1, \ldots, m$. Это многоугольное множество может быть как ограниченным (рис. 3.2) — тогда X представляет собой выпуклый многоугольник, так и неограниченным (рис. 3.3).

Пусть α — какое-либо значение функции $f(x)=\langle c, x\rangle=c^1x^1+c^2x^2$. Тогда уравнение

$$c^1x^1+c^2x^2=\alpha \qquad (2)$$

задает линию уровня функции $f(x)$, соответствующую ее значению α и на плоскости определяет прямую, перпендикулярную вектору $c=(c^1, c^2)\neq 0$. При изменении α от $-\infty$ до ∞ прямая (2), смещаясь параллельно самой себе, «зачертит» («заметет») всю плоскость. При этом вектор c — градиент функции $f(x)$ — указывает направление, в котором следует смещать прямую (2), чтобы увеличивать значение функции $f(x)=\langle c, x\rangle$. Может случиться, что при изменении α от $-\infty$ до ∞ прямая (2) при некотором значении $\alpha=f_*$ впервые коснется X и будет иметь с X общую точку x_* (на рис. 3.2–3.5 прямая (2) представлена при $\alpha=\alpha_1<f_*<\alpha_2<\alpha_3$). Ясно, что $\langle c, x_*\rangle=f_*=\inf\limits_X\langle c, x\rangle$, т. е. x_* — решение задачи (1). Возможен случай, когда прямая (2) при первом касании с многоугольником X будет иметь не одну

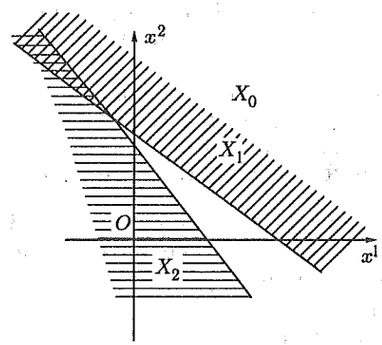

Рис. 3.1

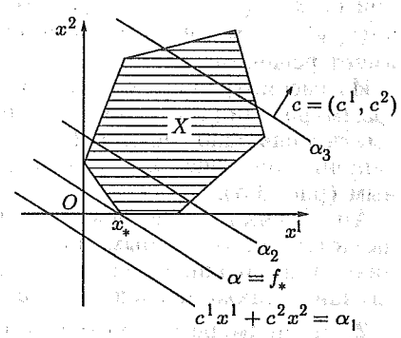

Рис. 3.2

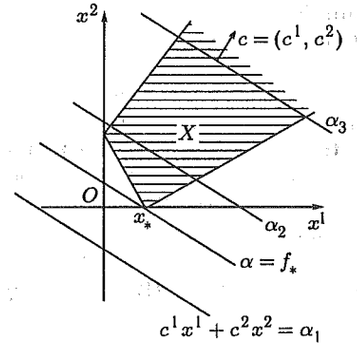

Рис. 3.3

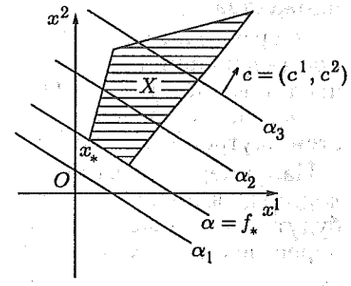

Рис. 3.4

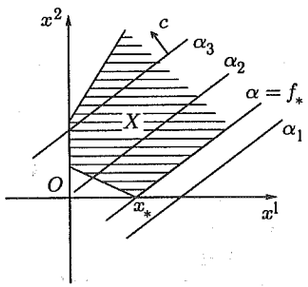

Рис. 3.5

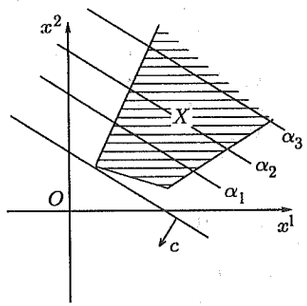

Рис. 3.6

общую точку с X, а целую сторону (рис. 3.4, 3.5) — это может случиться, если X имеет сторону, перпендикулярную вектору c.

Если многоугольное множество X не ограничено, то наряду со случаями, когда при первом касании прямая (2) будет иметь с X одну общую точку x_* (см. рис. 3.3) или сторону (рис. 3.5), возможна ситуация, когда прямая (2) при всех α $(-\infty < \alpha \leqslant \alpha_0 \leqslant \infty)$ имеет общую точку с X (рис. 3.6) — тогда $\inf\limits_X(c, x) = -\infty$ (первого касания прямой (2) с X нет), т. е. задача (1) не имеет решения.

Из рассмотренных случаев задачи (1) видно, что задача линейного программирования может не иметь ни одного решения (см. рис. 3.1, 3.6), может иметь лишь одно решение (см. рис. 3.2, 3.3), может иметь бесконечно много решений (см. рис. 3.4, 3.5), множество решений может быть неограниченным (рис. 3.5).

Аналогично можно показать, что множество X в задаче (1.21) при $n = 3$ является многогранным множеством, и дать геометрическую интерпретацию этой задачи. Предлагаем читателю самостоятельно рассмотреть этот случай, а также исследовать задачу (1.15) при $n = 2, 3$.

2. На примере рассмотренной выше задачи (1) нетрудно усмотреть, что если задача (1) имеет решение, то среди решений найдется хотя бы одна угловая точка (вершина) многоугольного множества X. Ниже мы увидим, что это не случайно: и в более общей задаче линейного программирования, оказывается, нижняя грань функции $\langle c, x\rangle$ на X достигается в угловой точке множества.

Определение 1. Точка v множества X называется *угловой точкой* (вершиной, крайней точкой, экстремальной точкой) множества X, если представление $v = \alpha v_1 + (1-\alpha)v_2$ при $v_1, v_2 \in X$ и $0 < \alpha < 1$ возможно лишь при $v_1 = v_2$. Иначе говоря, v — *угловая точка* множества X, если она не является внутренней точкой никакого отрезка, принадлежащего множеству X.

Например, угловыми точками многоугольника на плоскости или параллелепипеда в пространстве являются их вершины; все граничные точки шара будут его угловыми точками; при $n > 2$ замкнутое полупространство или пересечение двух замкнутых полупространств не имеют ни одной угловой точки.

В задачах линейного программирования понятие угловой точки играет фундаментальную роль и лежит в основе многих методов решения таких

задач. В дальнейшем мы будем подробно исследовать каноническую задачу (1.15). Поэтому начнем с изучения свойств угловых точек множества

$$X = \{x \in E^n:\ x \geqslant 0,\ Ax = b\}, \tag{3}$$

где A — матрица размера $m \times n$, $A \neq 0$, b — вектор из E^m. Ниже будет показано, что множество (3), если оно непусто, имеет хотя бы одну угловую точку (см. теорему 4.1). Возникает вопрос, как узнать, будет ли та или иная точка множества (3) угловой точкой? Приведем один достаточно простой алгебраический критерий угловой точки множества (3). Для этого вначале обозначим j-й столбец матрицы A через A_j и запишем систему уравнений $Ax = b$ в следующей эквивалентной форме:

$$A_1 x^1 + \ldots + A_n x^n = b. \tag{4}$$

Теорема 1. *Пусть множество X определено условиями* (3), $A \neq 0$, $r = \operatorname{rang} A$ *— ранг матрицы A. Для того чтобы точка $v = (v^1, \ldots, v^n) \in X$ была угловой точкой множества X, необходимо и достаточно, чтобы существовали номера $j_1, \ldots, j_r$, $1 \leqslant j_l \leqslant n$, $l = 1, \ldots, r$, такие, что*

$$A_{j_1} v^{j_1} + \ldots + A_{j_r} v^{j_r} = b; \quad v^j = 0, \quad j \neq j_l, \quad l = 1, \ldots, r, \tag{5}$$

причем столбцы $A_{j_1}, \ldots, A_{j_r}$ линейно независимы.

Доказательство. Необходимость. Пусть v — угловая точка множества X. Если $v = 0$, то из условия $0 \in X$ следует, что $b = 0$. Поскольку $A \neq 0$, то $r = \operatorname{rang} A \geqslant 1$ и существуют линейно независимые столбцы $A_{j_1}, \ldots, A_{j_r}$. Отсюда имеем $A_{j_1} \cdot 0 + \ldots + A_{j_r} \cdot 0 = 0$. Для случая $v = 0$ соотношения (5) доказаны.

Пусть теперь $v \neq 0$ и пусть $v^{j_1}, \ldots, v^{j_k}$ — все положительные координаты точки v. Отсюда и из условия $Av = b$ с учетом представления (4) имеем

$$A_{j_1} v^{j_1} + \ldots + A_{j_k} v^{j_k} = b; \quad v^j = 0, \quad j \neq j_l, \quad l = 1, \ldots, k. \tag{6}$$

Покажем, что столбцы $A_{j_1}, \ldots, A_{j_k}$ линейно независимы.

Пусть при некоторых $\alpha_1, \ldots, \alpha_k$ имеет место равенство

$$\alpha_1 A_{j_1} + \ldots + \alpha_k A_{j_k} = 0. \tag{7}$$

Возьмем точку $v_+ = (v_+^1, \ldots, v_+^n)$ с координатами $v_+^{j_p} = v^{j_p} + \varepsilon\alpha_p$, $v_+^j = 0$ при $j \neq j_p$, $p = 1, \ldots, k$, и точку $v_- = (v_-^1, \ldots, v_-^n)$ с координатами $v_-^{j_p} = v^{j_p} - \varepsilon\alpha_p$, $v_-^j = 0$ при $j \neq j_p$, $p = 1, \ldots, k$. Поскольку $v^{j_p} > 0$, $p = 1, \ldots, k$, то при достаточно малых $\varepsilon > 0$ будем иметь $v_+ \geqslant 0$, $v_- \geqslant 0$. Кроме того, умножая (7) на ε или $-\varepsilon$ и складывая с (6), приходим к равенствам $Av_+ = b$, $Av_- = b$. Таким образом, $v_+, v_- \in X$. Очевидно, $v = (v_+ + v_-)/2$, т. е. $v = \alpha v_+ + (1 - \alpha)v_-$ при $\alpha = 1/2$. По определению угловой точки это возможно лишь при $v_+ = v_- = v$, что в свою очередь означает, что $\alpha_1 = \ldots = \alpha_k = 0$. Таким образом, равенство (7) возможно только при $\alpha_1 = \ldots = \alpha_k = 0$. Линейная независимость столбцов $A_{j_1}, \ldots, A_{j_k}$ доказана. Отсюда следует, что $k \leqslant r$.

Если $k = r$, то соотношения (6) равносильны (5). Если $k < r$, то добавим к столбцам $A_{j_1}, \ldots, A_{j_k}$ новые столбцы $A_{j_{k+1}}, \ldots, A_{j_r}$ матрицы A так, чтобы система $A_{j_1}, \ldots, A_{j_k}, A_{j_{k+1}}, \ldots, A_{j_r}$ была линейно независимой, а при добавлении

любого другого столбца A_j эта система становилась линейно зависимой. Тогда система $A_{j_1}, \ldots, A_{j_r}$ образует некоторый базис линейной оболочки векторов $A_1, \ldots, A_n$. Размерность линейной оболочки векторов $A_1, \ldots, A_n$ равна рангу матрицы A, так что $s = r = \operatorname{rang} A$. Добавив к первому равенству (6) столбцы $A_{j_{k+1}}, \ldots, A_{j_r}$, умноженные соответственно на $v^{j_{k+1}} = 0, \ldots, v^{j_r} = 0$, из (6) получим соотношения (5). Тем самым необходимость доказана.

Достаточность. Пусть некоторая точка $v = (v^1, \ldots, v^n)$ удовлетворяет условиям (5), где $A_{j_1}, \ldots, A_{j_r}$ — линейно независимы, $r = \operatorname{rang} A$. Пусть $v = \alpha v_1 + (1-\alpha) v_2$ при некоторых $v_1, v_2 \in X$, $0 < \alpha < 1$. Покажем, что такое представление возможно только при $v_1 = v_2 = v$. Сразу же заметим, что если $v^j = 0$, то из этого представления с учетом неравенств $0 < \alpha < 1$, $v_1^j \geqslant 0$, $v_2^j \geqslant 0$ получим $0 \leqslant \alpha v_1^j + (1-\alpha) v_2^j = v^j = 0$, что возможно лишь при $v_1^j = v_2^j = v^j = 0$. Таким образом, для получения равенства $v = v_1 = v_2$ остается еще доказать, что $v_1^j = v_2^j = v^j$ и при тех j, для которых $v^j > 0$.

По условию (5) у точки v положительными могут быть лишь координаты $v^{j_1}, \ldots, v^{j_r}$. Произведя при необходимости перенумерацию переменных, можем считать, что $v^{j_1} > 0, \ldots, v^{j_k} > 0$, $v^{j_{k+1}} = 0, \ldots, v^{j_r} = 0$ (случаи $k = 0$ или $k = r$ здесь не исключаются). Тогда (4) можно переписать в виде $A_{j_1} v^{j_1} + \ldots + A_{j_k} v^{j_k} = b$. Кроме того, учитывая, что по доказанному $v_1^j = v_2^j = 0$ при всех $j \neq j_p$, $p = 1, \ldots, k$, равенства $Av_i = b$ также можно записать в виде $A_{j_1} v_i^{j_1} + \ldots + A_{j_k} v_i^{j_k} = b$, $i = 1, 2$. Вспомним, что векторы $A_{j_1}, \ldots, A_{j_k}$ линейно независимы. Поэтому вектор b может линейно выражаться через $A_{j_1}, \ldots, A_{j_k}$ единственным способом. Это значит, что $v^{j_p} = v_1^{j_p} = v_2^{j_p}$ для $p = 1, \ldots, k$. Тем самым установлено, что $v = v_1 = v_2$. Следовательно, v — угловая точка множества X. □

Определение 2. Систему векторов $A_{j_1}, \ldots, A_{j_r}$, входящих в первое из равенств (5), называют *базисом угловой точки* v, а соответствующие им переменные $v^{j_1}, \ldots, v^{j_r}$ — *базисными координатами* угловой точки v. Если все базисные координаты угловой точки положительны, то такую угловую точку называют *невырожденной*. Если же среди базисных координат $v^{j_1}, \ldots, v^{j_r}$ — хотя бы одна равна нулю, то такая угловая точка называется *вырожденной*. При фиксированном базисе $A_{j_1}, \ldots, A_{j_r}$ переменные $u^{j_1}, \ldots, u^{j_r}$ называются *базисными переменными* угловой точки, а остальные переменные u^j — *небазисными* (*свободными*) *переменными*.

Из теоремы 1 следует, что невырожденная угловая точка обладает единственным базисом — ее базис составляют столбцы с теми номерами, которым соответствуют положительные координаты угловой точки. Если угловая точка вырожденная, то она может обладать несколькими базисами. В самом деле, если $v^{j_1} > 0, \ldots, v^{j_k} > 0$, $k < r = \operatorname{rang} A$, а остальные координаты v^j угловой точки v равны нулю, то, как видно из доказательства теоремы 1, в базис такой точки обязательно войдут столбцы $A_{j_1}, \ldots, A_{j_k}$, а остальные базисные столбцы $A_{j_{k+1}}, \ldots, A_{j_r}$, входящие в представление (5), могут быть выбраны, вообще говоря, различными способами.

Поскольку из n столбцов матрицы A можно выбрать r линейно независимых столбцов не более чем C_n^r способами (C_n^r — число сочетаний из n элементов по r), то из теоремы 1 следует, что число угловых точек множества (3) конечно.

Пример 1. Пусть $X = \{x = (x^1, x^2, x^3, x^4) \in E^4\colon\ x^j \geqslant 0,\ j = 1, \ldots, 4,\ x^1 + x^2 + 3x^3 + x^4 = 3,\ x^1 - x^2 + x^3 + 2x^4 = 1\}$. Обозначим

$$A_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad A_3 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \quad A_4 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad b = \begin{bmatrix} 3 \\ 1 \end{bmatrix}.$$

Нетрудно видеть, что точки $x_1 = (2, 1, 0, 0)$ и $x_2 = (0, 5/3, 0, 4/3)$ являются невырожденными угловыми точками множества X, причем точке x_1 соответствует базис A_1, A_2, а точке x_2 — базис A_2, A_4; угловая точка $x_3 = (0, 0, 1, 0)$ вырожденная и ей соответствуют базисы A_1, A_3, или A_2, A_3, или A_3, A_4; точка $x_4 = (5, 0, 0, -2)$ не является угловой для множества X, так как $x_4 \notin X$.

Упражнения

1. При каких значениях параметра a задача: $f(x) = x^1 + ax^2 \to \inf$, $x \in X = \{x \in E^2\colon x \geqslant 0,\ x^1 - x^2 \geqslant 1,\ x^1 + 2x^2 \geqslant 4\}$ имеет решение? Не имеет решения? Имеет единственное решение? Нарисуйте график функции $f_* = f_*(a) = \inf\limits_{x \in X} f(x)$. Графически изобразите множество $X_*(a) = \{x \in X\colon f(x) = f_*(a)\}$ при различных a.

2. Найти все угловые точки и их базисы для множеств:

$X_1 = \{x \in E^4\colon x \geqslant 0,\ x^1 - 2x^2 - x^3 = 0,\ -x^1 + 3x^2 + x^4 = 1\}$,

$X_2 = \{x \in E^5\colon x \geqslant 0,\ x^1 + x^2 + x^3 + x^4 = 1,\ -x^1 + 2x^2 + x^3 + x^5 = 1\}$,

$X_3 = \{x \in E^5\colon x \geqslant 0,\ 2x^1 + 3x^3 + x^5 = 3,\ x^1 + x^2 + 2x^3 = 2,\ x^1 + x^3 + x^4 = 1\}$.

3. При каких значениях параметров a_i, b множество $X = \{x \in E^n\colon x \geqslant 0,\ a_1 x^1 + \ldots + a_n x^n = b\}$ непусто и имеет угловые точки? Какое максимальное и минимальное число угловых точек может иметь такое множество?

4. Пусть $X = \{x \in E^n\colon \langle a_i, x\rangle \leqslant b^i,\ i = 1, \ldots, m\}$, $m \geqslant n$. Показать, что точка $v \in X$ является угловой точкой множества X тогда и только тогда, если обращаются в точные равенства не менее, чем n из неравенств $\langle a_i, v\rangle \leqslant b^i$, среди которых есть n линейно независимых.

5. Вывести теорему 1 из утверждения упражнения 4. Указание: множество (3) записать в виде $X = \{x \in E^n\colon \langle a_i, x\rangle = b^i,\ i = 1, \ldots, m;\ \langle -e_i, x\rangle \leqslant 0,\ i = 1, \ldots, n\}$, где e_i — i-й столбец единичной матрицы размера $n \times n$.

6. Доказать, что всякая угловая точка множества (3) является угловой и для множества $X_1 = \{x \geqslant 0;\ Ax \leqslant b\}$.

7. Доказать, что множество $X = \{x \in E^n\colon \langle a_i, x\rangle \leqslant b^i,\ i = 1, \ldots, m\}$ при $m < n$ не имеет угловых точек.

§ 3. Симплекс-метод. Антициклин

1. Будем рассматривать каноническую задачу (1.15):

$$f(x) = \langle c, x\rangle \to \inf, \quad x \in X = \{x \in E^n\colon x \geqslant 0, Ax = b\}, \tag{1}$$

где A ненулевая матрица размера $m \times n$, $c \in E^n$, $b \in E^m$. Ниже будет показано, что всякое непустое множество X из (1) имеет хотя бы одну угловую точку и, кроме того, если $\inf\limits_{x \in X} \langle c, x\rangle = f_* > -\infty$, то эта нижняя грань достигается хотя бы в одной угловой точке множества X (см. теоремы 4.1, 4.2). Отсюда следует, что задачу (1) можно попытаться решить следующим

образом: сначала найти все угловые точки множества X, пользуясь, например, конструкциями теоремы 2.1, затем вычислить значение функции $f(x)=\langle c, x\rangle$ в каждой из угловых точек, число которых, как мы знаем, конечно, и определить наименьшее из них. Однако такой подход к решению задачи (1) практически не применяется, так как уже в задачах не очень большой размерности число угловых точек может быть столь большим, что простой перебор всех угловых точек множества X может оказаться невозможным за разумное время даже при использовании самых лучших современных компьютеров.

Тем не менее идея перебора угловых точек множества оказалась весьма плодотворной и послужила основой ряда методов решения канонической и других задач линейного программирования. Одним из таких методов является так называемый *симплекс-метод*. Название этого метода связано с тем, что он впервые разрабатывался применительно к задачам линейного программирования, в которых множество X представлял собой симплекс в E^n: $X=\{x=(x^1,\ldots,x^n);\ x\geqslant 0,\ \sum\limits_{i=1}^{n} x^i=1\}$, затем метод был обобщен на случай более общих множеств X, но первоначальное название за ним так и сохранилось; в литературе этот метод часто называют еще *методом последовательного улучшения плана*.

При реализации симплекс-метода осуществляется упорядоченный (направленный) перебор угловых точек множества X, при котором значение функции $\langle c, x\rangle$ убывает при переходе от одной угловой точки к другой, что позволит, перебрав, быть может, лишь относительно небольшое число угловых точек, выяснить, имеет ли задача (1) решение и, если имеет, то найти его. Такова общая идея симплекс-метода.

Перейдем к описанию симплекс-метода для решения канонической задачи (1). По условию $1\leqslant r=\operatorname{rang} A\leqslant\min\{m, n\}$. Предполагая, что из системы $(Ax)^i=b^i$, $i=1,\ldots,m$, исключены линейно зависимые уравнения, в этом параграфе будем считать, что $r=m$, и матрица A имеет размеры $r\times n$. Тогда $r\leqslant n$. Если $r=n$, то система $Ax=b$ будет иметь единственное решение x_0 и множество X будет либо пустым (если не соблюдается ограничение $x_0\geqslant 0$), либо X состоит из одной точки (если $x_0\geqslant 0$) — в этом случае задача (1) становится малосодержательной. Поэтому будем считать, что $r<n$. Тогда систему $Ax=b$ можем записать в виде

$$\begin{aligned} &a_{11}x^1+\ldots+a_{1n}x^n=b^1,\\ &\ldots\ldots\ldots\ldots\ldots\\ &a_{r1}x^1+\ldots+a_{rn}x^n=b^r,\quad r=m<n. \end{aligned}\tag{2}$$

Пусть известна некоторая угловая точка $v=(v^1, v^2,\ldots, v^n)$ множества X с базисом $A_{j_1}, A_{j_2},\ldots, A_{j_r}$ (о том, как найти такую точку v, см. § 4). Матрицу $B=(A_{j_1},\ldots, A_{j_r})$, столбцами которой являются базисные векторы, будем называть *базисной матрицей* или просто *базисом*. Через $I(v)=\{j_1,\ldots, j_r\}$ обозначим номера базисных переменных или, короче, базисных номеров. Перенумеровав переменные, можем считать, что $I(v)=\{1, 2,\ldots, r\}$; тогда столбцы $A_1, A_2,\ldots, A_r$ матрицы A составляют базис точки v, а $x^1, x^2,\ldots, x^r$ — ее базисные переменные. Обозначим

$$\bar{x}=\begin{bmatrix} x^1\\ \ldots\\ x^r\end{bmatrix},\quad \bar{v}=\begin{bmatrix} v^1\\ \ldots\\ v^r\end{bmatrix},\quad \bar{c}=\begin{bmatrix} c^1\\ \ldots\\ c^r\end{bmatrix},\quad A_j=\begin{bmatrix} a_{1j}\\ \ldots\\ a_{rj}\end{bmatrix},$$

$$B=\begin{bmatrix} a_{11}, & \ldots, & a_{1r}\\ \ldots\ldots\ldots\ldots\\ a_{r1}, & \ldots, & a_{rr}\end{bmatrix}=(A_1, A_2,\ldots, A_r).$$

Тогда систему (2) можно кратко переписать в виде

$$b=A_1x^1+\ldots+A_rx^r+A_{r+1}x^{r+1}+\ldots+A_nx^n=B\bar{x}+\sum_{k=r+1}^{n}A_kx^k.\tag{3}$$

Так как столбцы $A_1,\ldots, A_r$ линейно независимы, то $\det B\neq 0$ и, следовательно, существует обратная матрица B^{-1}. Кроме того, вспомним, что согласно теореме 2.1 небазисные координаты угловой точки v заведомо равны нулю, так что $v=\begin{bmatrix}\bar{v}\\ 0\end{bmatrix}$, где $\bar{v}\geqslant 0$. Отсюда и из (3) следует, что базисные координаты $\bar{v}$ удовлетворяют системе $B\bar{v}=b$, откуда имеем $\bar{v}=B^{-1}b$. Умножая систему (3) на B^{-1} слева, получим следующее соотношение между базисными переменными $\bar{x}$ и небазисными переменными $x^{r+1},\ldots, x^n$:

$$0\leqslant\bar{v}=B^{-1}b=\bar{x}+\sum_{k=r+1}^{n}B^{-1}A_kx^k.\tag{4}$$

Обозначим $(B^{-1}A_k)^i=\gamma_{ik}$ — i-я координата вектора столбца $r_k=B^{-1}A_k$. Тогда систему уравнений (4) можно записать в покоординатной форме:

$$\begin{aligned} v^1&=x^1 &&+\gamma_{1r+1}x^{r+1}+\ldots+\gamma_{1k}x^k+\ldots+\gamma_{1n}x^n,\\ v^i&=x^i &&+\gamma_{ir+1}x^{r+1}+\ldots+\gamma_{ik}x^k+\ldots+\gamma_{in}x^n,\\ v^s&=x^s &&+\gamma_{sr+1}x^{r+1}+\ldots+\gamma_{sk}x^k+\ldots+\gamma_{sn}x^n,\\ v^r&=x^r &&+\gamma_{rr+1}x^{r+1}+\ldots+\gamma_{rk}x^k+\ldots+\gamma_{rn}x^n. \end{aligned}\tag{5}$$

Систему $B^{-1}Ax=B^{-1}b$, полученную умножением исходной системы $Ax=b$ на матрицу B^{-1} слева, называют *приведенной системой угловой точки* v с базисной матрицей B. Системы (4) и (5), таким образом, представляют собой различные формы записи приведенной системы точки v с базисом $B=(A_1,\ldots, A_r)$. Подчеркнем, что из невырожденности матрицы B следует, что системы (4), (5) равносильны исходной системе (2) или (3).

Пользуясь равенством $\bar{x}=\bar{v}-\sum\limits_{j=r+1}^{n}B^{-1}A_jx^j$, вытекающим из (4), значения функции $f(x)$ выразим через небазисные переменные:

$$\begin{aligned} f(x)=\langle\bar{c},\bar{x}\rangle+\sum_{j=r+1}^{n}c_jx^j&=\langle\bar{c},\bar{v}-\sum_{j=r+1}^{n}B^{-1}A_jx^j\rangle+\\ &+\sum_{j=r+1}^{n}c_jx^j=\langle\bar{c},\bar{v}\rangle-\sum_{j=r+1}^{n}(\langle\bar{c},B^{-1}A_j\rangle-c_j)x^j, \end{aligned}$$

или, короче

$$f(x)=f(v)-\sum_{j=r+1}^{n}\Delta_jx^j,\tag{6}$$

где учтено, что $\langle\bar{c},\bar{v}\rangle=\langle c, v\rangle=f(v)$, и использованы обозначения

$$\Delta_j=\langle\bar{c},B^{-1}A_j\rangle-c_j=\sum_{i=1}^{r}c_i\gamma_{ij}-c_j,\quad j=1,\ldots, n.\tag{7}$$

Выражение (6) будем называть *приведенной формой целевой функции*, соответствующей угловой точке v с базисом B.

Входящие в (5), (6) величины γ_{ik}, v_i, Δ_j удобно записать в виде табл. 1, которую принято называть *симплекс-таблицей* угловой точки v с базисом $B=(A_1,\ldots,A_r)$. В столбце Б этой таблицы перечислены базисные переменные $x^1,\ldots,x^r$ точки v; в столбце V помещены значения базисных переменных $\bar v=B^{-1}b$ угловой точки v; в столбце x^k находятся координаты $\gamma_{ik}=(B^{-1}A_k)^i$, $i=1,\ldots,r$, вектора $\gamma_k=B^{-1}A_k$, $k=1,\ldots,n$; в столбцах базисных переменных $x^1,\ldots,x^r$ отражены равенства $B^{-1}A_j=e_j$, $j=1,\ldots,r$, вытекающие из определения обратной матрицы B^{-1}; здесь e_j — j-й столбец

Таблица 1

	Б	V	x^1	…	x^i	…	x^s	…	x^r	x^{r+1}	…	x^k	…	x^j	…	x^n
Γ_1	x^1	v^1	1	…	0	…	0	…	0	γ_{1r+1}	…	γ_{1k}	…	γ_{1j}	…	γ_{1n}
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
Γ_i	x^i	v^i	0	…	1	…	0	…	0	γ_{ir+1}	…	γ_{ik}	…	γ_{ij}	…	γ_{in}
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
Γ_s	x^s	v^s	0	…	0	…	1	…	0	γ_{sr+1}	…	γ_{sk}	…	γ_{sj}	…	γ_{sn}
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
Γ_r	x^r	v^r	0	…	0	…	0	…	1	γ_{rr+1}	…	γ_{rk}	…	γ_{rj}	…	γ_{rn}
Δ		$f(v)$	0	…	0	…	0	…	0	Δ_{r+1}	…	Δ_k	…	Δ_j	…	Δ_n

единичной матрицы размера $r\times r$. В крайнем левом столбце для удобства изложения приведены обозначения для строк симплекс-таблицы: $\Gamma_1,\Gamma_2,\ldots,\Gamma_n,\Delta$. Так, например, в строке $\Gamma_i=(v^i,0,\ldots,0,1,0,\ldots,0,\gamma_{ir+1},\ldots,\gamma_{in})$ записана вся информация, по которой удобно воспроизвести соответствующее i-е уравнение системы (5), и, наоборот, зная i-е уравнение этой системы легко можно восстановить строку Γ_i. В строке Δ помещены величины $\Delta_0=f(v)=\langle c,v\rangle$, $\Delta_1,\ldots,\Delta_n$, связанные с минимизируемой функцией $f(x)=\langle c,x\rangle$ формулами (6), (7); в этой строке отражено, что для базисных номеров $\Delta_j=\langle\bar c,e_j\rangle-c^j=c^j-c^j=0$, $j=1,\ldots,r$. По строке $\Delta=(f(v),0,\ldots,0,\Delta_{r+1},\ldots,\Delta_n)$ симплекс-таблицы легко воспроизвести формулу (6) и обратно, имея (6), несложно восстановить строку Δ. Из формул $\gamma_0=\bar v=B^{-1}b$, $\gamma_j=B^{-1}A_j$, $\Delta_0=\langle\bar c,B^{-1}b\rangle=f(v)$, $\Delta_j=\langle\bar c,B^{-1}A_j\rangle-c^j$, для величин, заполняющих симплекс-таблицу, следует, что эта таблица однозначно определяется заданием векторов c, b, матрицы A и базисной матрицы B угловой точки v.

После сделанных преобразований каноническую задачу (1) теперь можно сформулировать в следующей равносильной, так называемой, *приведенной форме*: минимизировать функцию (6) при условиях (5) и соблюдении неравенства $x\geqslant 0$. Конечно, от такой переформулировки задача (1) проще не стала, но тем не менее в новой ее формулировке с явным разделением базисных и небазисных переменных, оказывается, легче проследить за тем, как изменяется функция $f(x)$ при изменении небазисных переменных, и можно попытаться выбрать эти переменные так, чтобы в новой точке $w\in X$ было $f(w)<f(v)$. Однако, если мы начнем изменять все небазисные переменные сразу, то вряд ли сможем проследить и за изменением функции $f(x)$, и за соблюдением ограничений $x\geqslant 0$. Поэтому мы попробуем изменить лишь одну из небазисных переменных, скажем, переменную x^k, $r+1\leqslant k\leqslant n$, остальные небазисные переменные положим равными нулю, а базисные переменные будем определять из уравнений (5). Иначе говоря, новую точку $w=(w^1,\ldots,w^n)$ будем искать среди точек с координатами

$$w^1=v^1-\gamma_{1k}x^k,\ \ldots,\ w^i=v^i-\gamma_{ik}x^k,\ \ldots,\ w^r=v^r-\gamma_{rk}x^k,\quad w^{r+1}=0,\ \ldots,\ w^{k-1}=0,\ w^k=x^k\geqslant 0,\ w^{k+1}=0,\ \ldots,\ w^n=0. \tag{8}$$

В такой точке w согласно (6) значение функции $f(w)$ равно

$$f(w)=f(v)-\Delta_k x^k,\quad x^k\geqslant 0. \tag{9}$$

Наша ближайшая задача: выбрать номер k, $r+1\leqslant k\leqslant n$, и величину $x^k\geqslant 0$ так, чтобы новая точка (8) удовлетворяла требованиям: $Aw=b$, $w\geqslant 0$, $f(w)\leqslant f(v)$ (будет еще лучше, если удастся получить $f(w)<f(v)$). Что касается первого требования $Aw=b$, то здесь проблем нет: точка (8) при любом выборе номера k и величины x^k, очевидно, является решением системы (5) и равносильной ей системы (2). Анализируя знаки величин Δ_k, γ_{ik}, нетрудно выяснить, можно ли удовлетворить оставшимся двум требованиям: $w\geqslant 0$ и $f(w)\leqslant f(v)$, и указать правило выбора нужного номера k и нужной величины $x^k\geqslant 0$. Такой анализ приведет к рассмотрению следующих трех взаимоисключающих друг друга случаев I — III.

Случай I. Справедливы неравенства:

$$\Delta_j=\langle\bar c,B^{-1}A_j\rangle-c_j\leqslant 0,\quad j=r+1,\ldots,n, \tag{10}$$

т. е. в нижней строке симплекс-таблицы 1 все Δ_j, $1\leqslant j\leqslant n$, неположительны. Как видно из (8), (9), тогда невозможно добиться неравенства $f(w)<f(v)$ ни при каких k, $r+1\leqslant k\leqslant n$, и $x^k\geqslant 0$, в лучшем случае при $x^k=0$ получим $w=v$, $f(w)=f(v)$. Однако это обстоятельство не должно огорчать нас, так как оказывается, что при выполнении условий (10) рассматриваемая точка v является решением задачи (1). В самом деле, для любой точки $x\in X=\{x\geqslant 0\colon Ax=b\}$ с учетом представления (4) и неравенств (10) имеем

$$f(x)=\langle\bar c,\bar x\rangle+\sum_{j=r+1}^{n}c^jx^j\geqslant\langle\bar c,\bar x\rangle+\sum_{j=r+1}^{n}\langle\bar c,B^{-1}A_j\rangle x^j=\langle\bar c,\bar x+\sum_{j=r+1}^{n}B^{-1}A_jx^j\rangle=\langle\bar c,\bar v\rangle=f(v).$$

Таким образом, $f(x)\geqslant f(v)$ при всех $x\in X$, т. е. v — решение задачи (1).

Случай II. Существует номер k, $r+1\leqslant k\leqslant n$, такой, что

$$\Delta_k>0,\ \gamma_{ik}\leqslant 0,\ i=1,\ldots,r,\ \text{т. е. }\gamma_k=B^{-1}A_k\leqslant 0. \tag{11}$$

Это значит, что в k-м столбце симплекс-таблицы 1 над величиной $\Delta_k>0$ нет ни одного положительного числа γ_{ik}. В этом случае при всех $x^k\geqslant 0$ точка w, определяемая формулами (8), будет иметь неотрицательные координаты и, следовательно, будет принадлежать множеству X. Тогда как видно из (9), $f(w)=f(v)-\Delta_kx^k\to-\infty$ при $x^k\to+\infty$. Это значит, что $f_*=\inf\limits_{x\in X}f(x)=-\infty$, т. е. задача (1) не имеет решения.

Случай III. Существует номер k, $r+1\leqslant k\leqslant n$, для которого $\Delta_k>0$, причем для каждого такого номера k найдется номер i, $1\leqslant i\leqslant r$, что $\gamma_{ik}>0$, или, иначе говоря, в каждом k-м столбце симплекс-таблицы 1 над величиной

$\Delta_k > 0$ имеется хотя бы одно положительное число γ_{ik}. Используя известные кванторы $\forall$, $\exists$ всеобщности и существования, рассматриваемый случай кратко запишем в виде:

$$\forall \Delta_k > 0 \quad \exists \gamma_{ik} > 0. \tag{12}$$

Для точки w, определяемой формулами (8), согласно (9) здесь будем иметь $f(w) = f(v) - \Delta_k x^k \leqslant f(v)$ при любом $x^k \geqslant 0$. Остается лишь позаботиться о выполнении условия $w \geqslant 0$. В рассматриваемом случае множество номеров

$$I_k(v) = \{i\colon\ 1 \leqslant i \leqslant r,\ \gamma_{ik} > 0\} \neq \varnothing.$$

Если $i \notin I_k(v)$, т. е. $\gamma_{ik} \leqslant 0$, то как видно из формул (8), $w^i = v^i - \gamma_{ik} x^k \geqslant v^i \geqslant 0$ при любом выборе $x^k \geqslant 0$. Если же $\gamma_{ik} > 0$, то при слишком больших значениях x^k, а именно, при $x^k > \min\limits_{i \in I_k(v)} (v^i/\gamma_{ik})$, величина $w^i = v^i - \gamma_{ik} x^k$ станет отрицательной хотя бы для одного номера $i \in I_k(v)$. Таким образом, для обеспечения условия $w \geqslant 0$ для точек, определяемых формулами (8), здесь нужно x^k взять так, чтобы $0 \leqslant x^k \leqslant \min\limits_{i \in I_k(v)} (v^i/\gamma_{ik})$. Пусть

$$\min_{i \in I_k(v)} \frac{v^i}{\gamma_{ik}} = \frac{v^s}{\gamma_{sk}}, \qquad s \in I_k(v). \tag{13}$$

Так как множество $I_k(v)$ непусто и конечно, то хотя бы один такой номер s существует. Величину γ_{sk}, где номера k, s определяются условиями (12), (13), называют *разрешающим* (ведущим) элементом симплекс-таблицы 1.

Зафиксируем один из разрешающих элементов γ_{sk} таблицы 1 и в формулах (8), (9) положим $x^k = v_s/\gamma_{sk}$. Получим точку $w = (w^1, \ldots, w^n)$ с координатами

$$\begin{gathered} w^1 = v^1 - \gamma_{1k}\frac{v^s}{\gamma_{sk}}, \ \ldots, \ w^i = v^i - \gamma_{ik}\frac{v^s}{\gamma_{sk}}, \ \ldots, \ w^{s-1} = v^{s-1} - \gamma_{s-1\,k}\frac{v^s}{\gamma_{sk}}, \\ w^s = v^s - \gamma_{sk}\frac{v^s}{\gamma_{sk}} = 0, \ w^{s+1} = v^{s+1} - \gamma_{s+1\,k}\frac{v^s}{\gamma_{sk}}, \ \ldots, \ w^r = v^r - \gamma_{rk}\frac{v^s}{\gamma_{sk}}, \\ w^{r+1} = 0, \ \ldots, \ w^{k-1} = 0, \ w^k = \frac{v^s}{\gamma_{sk}}, \ w^{k+1} = 0, \ \ldots, \ w^n = 0, \end{gathered} \tag{14}$$

и значение функции $f(x)$ в этой точке

$$f(w) = \langle c, w\rangle = f(v) - \Delta_k v_s/\gamma_{sk}. \tag{15}$$

По построению точка w с координатами (14) принадлежит множеству X. Покажем, что w — угловая точка множества X с базисом

$$A_1, \ldots, A_{s-1}, A_k, A_{s+1}, \ldots, A_r, \tag{16}$$

получающимся из базиса точки v заменой столбца A_s на A_k. Учитывая, что $w^s = w^{r+1} = \ldots = w^{k-1} = w^{k+1} = \ldots = w^n = 0$, условие $Aw = b$ можно записать в виде $A_1 w^1 + \ldots + A_{s-1} w^{s-1} + A_{s+1} w^{s+1} + \ldots + A_r w^r + A_k w^k = b$. Согласно теореме 2.1 остается показать, что система векторов (16) линейно независима. Пусть для некоторых чисел $\alpha_1, \ldots, \alpha_{s-1}, \alpha_{s+1}, \ldots, \alpha_r, \alpha_k$ оказалось, что

$$\alpha_1 A_1 + \ldots + \alpha_{s-1} A_{s-1} + \alpha_{s+1} A_{s+1} + \ldots + \alpha_r A_r + \alpha_k A_k = 0. \tag{17}$$

Поскольку $A_k = BB^{-1}A_k = \sum\limits_{i=1}^{r} A_i (B^{-1}A_k)^i = \sum\limits_{i=1}^{r} \gamma_{ik} A_i$, то из (17) следует

$$\sum_{i=1,\, i\neq s}^{r} \alpha_i A_i + \alpha_k \sum_{i=1}^{r} \gamma_{ik} A_i = \sum_{i=1,\, i\neq s}^{r} (\alpha_i + \alpha_k \gamma_{ik}) A_i + \alpha_k \gamma_{sk} A_s = 0.$$

Но система $A_1, \ldots, A_s, \ldots, A_r$ является базисом точки v и, следовательно, линейно независима. Тогда последнее равенство возможно лишь при $\alpha_i + \alpha_k \gamma_{ik} = 0$, $i = 1, \ldots, r$, $i \neq s$; $\alpha_k \gamma_{sk} = 0$. Но $\gamma_{sk} > 0$, как разрешающий элемент, поэтому $\alpha_k = 0$. А тогда все остальные $\alpha_i = 0$, $i = 1, \ldots, r$, $i \neq s$. Таким образом, равенство (17) возможно лишь при $\alpha_1 = \ldots = \alpha_{s-1} = \alpha_{s+1} = \ldots = \alpha_r = \alpha_k = 0$. Это значит, что система (16) линейно независима. Тем самым показано, что точка w, определяемая формулами (14), является угловой точкой множества X с базисом (16), с базисными переменными $x^1, \ldots, x^{s-1}, x^k, x^{s+1}, \ldots, x^r$, причем $f(w) \leqslant f(v)$, так как в (15) $\Delta_k > 0$, $\gamma_{sk} > 0$, $v^s \geqslant 0$.

Замечание 1. Для дальнейшего полезно подчеркнуть, что при доказательстве того, что точка w является угловой точкой, мы нигде не пользовались тем, что $\Delta_k > 0$. Это означает, что независимо от знака Δ_k, формулы (13), (14) позволяют перейти от одной угловой точки v множества X к другой его угловой точке w, лишь бы $I_k(v) \neq \varnothing$, $v^s > 0$. Если $v^s = 0$, то формулы (13), (14) дают ту же угловую точку, т. е. $w = v$, но при этом происходит замена базиса $A_1, \ldots, A_r$ на базис (16).

Далее, познакомимся с правилами заполнения симплекс-таблицы точки w (см. табл. 2), постараемся понять, как связаны симплекс-таблицы точек v и w. Как и в таблице 1 в столбце Б укажем базисные переменные $x^1, \ldots, x^{s-1}, x^k, x^{s+1}, \ldots, x^r$ точки w, в столбце V — соответствующие значения $w^1, \ldots, w^{s-1}, w^k, w^{s+1}, \ldots, w^r$ ее базисных координат, вычисленных по формулам (14). В столбцах x^j нам нужно поместить координаты $\overline{\gamma}_{ij}$ вектора $\overline{\gamma}_j = \overline{B}^{-1} A_j$, где $\overline{B}^{-1}$ — матрица, обратная к матрице $\overline{B} = \{A_1, \ldots, A_{s-1}, A_k, A_{s+1}, \ldots, A_r\}$. Следует однако заметить, что обращение матриц, их умножение является довольно трудоемкими операциями, поэтому вычисление координат вектора $\overline{\gamma}_j$, опираясь на его определение, может потребовать большого объема вычислений. В связи с этим полезно вспомнить, что вектор $\overline{\gamma}_j$ совпадает со столбцом коэффициентов при переменной x^j в приведенной системе $\overline{B}^{-1} b = \overline{B}^{-1} Ax$, соответствующей угловой точке w с базисом (16). К счастью, имея приведенную систему (5) для угловой точки v, из нее нетрудно получить такую систему и для точки w. Покажем, как это делается. С этой целью разделим s-е уравнение системы (5) на разрешающий элемент $\gamma_{sk} > 0$; учитывая, что в силу (14) $w^k = v^s/\gamma_{sk}$, получим

$$w^k = \frac{v^s}{\gamma_{sk}} = \frac{1}{\gamma_{sk}} x^s + \sum_{j=r+1}^{k-1} \frac{\gamma_{sj}}{\gamma_{sk}} x^j + x^k + \sum_{j=r+1}^{n} \frac{\gamma_{sj}}{\gamma_{sk}} x^j. \tag{18}$$

Отсюда выразим переменную x^k через остальные переменные:

$$x^k = \frac{v^s}{\gamma_{sk}} - \frac{1}{\gamma_{sk}} x^s - \sum_{j=r+1}^{n}{}' \frac{\gamma_{sj}}{\gamma_{sk}} x^j, \tag{19}$$

и подставим ее в другие уравнения системы (5) (здесь и ниже в (20)–(24) знак $\sum'$ означает, что суммирование ведется по всем $j = r+1, \ldots, n$, исклю-

чая номер $j=k$). Будем иметь

$$v^i = x^i + \sum_{j=r+1}^{n}{}' \gamma_{ij}x^j + \gamma_{ik}\left(\frac{v^s}{\gamma_{sk}} - \frac{1}{\gamma_{sk}}x^s - \sum_{j=r+1}^{n}{}' \frac{\gamma_{sj}}{\gamma_{sk}}x^j\right) =$$

$$= \gamma_{ik}\frac{v^s}{\gamma_{sk}} + x^i + \left(-\frac{\gamma_{ik}}{\gamma_{sk}}\right)x^s + \sum_{j=r+1}^{n}{}' \left(\gamma_{ij} - \gamma_{ik}\frac{\gamma_{sj}}{\gamma_{sk}}\right)x^j,$$

откуда с учетом (14) получим

$$w^i = v^i - \gamma_{ik}\frac{v^s}{\gamma_{sk}} = x^i + \left(-\frac{\gamma_{ik}}{\gamma_{sk}}\right)x^s + \sum_{j=r+1}^{n}{}' \left(\gamma_{ij} - \gamma_{ik}\frac{\gamma_{sj}}{\gamma_{sk}}\right)x^j,$$

$$i = 1, \ldots, s-1, s+1, \ldots, r. \quad (20)$$

Система r уравнений (18), (20) относительно неизвестных $x^1, x^2 \ldots, x^n$ равносильна системам (5), (2) и представляет собой приведенную систему для угловой точки w (см. упражнение 7). Отсюда следует, что в строке Γ_s таблицы 2 согласно (18) мы должны записать величины $\bar\gamma_{sj}$, $j=0, 1, \ldots, n$, определяемые формулами

$$\bar\gamma_{s0} = w^k = \frac{v^s}{\gamma_{sk}}, \quad \bar\gamma_{ss} = \frac{1}{\gamma_{sk}}, \quad \bar\gamma_{sj} = \frac{\gamma_{sj}}{\gamma_{sk}}, \quad j = r+1, \ldots, k-1, k+1, \ldots, n;$$

$$\bar\gamma_{sk} = 1, \quad \bar\gamma_{sj} = 0, \quad j = 1, \ldots, s-1, s+1, \ldots, r. \quad (21)$$

В других строках Γ_i, $i \neq s$, таблицы 2 , согласно (14), (20) следует поместить величины $\bar\gamma_{ij}$, $j=0, 1, \ldots, n$, определяемые формулами

$$\bar\gamma_{i0} = w^i = v^i - \gamma_{ik}\frac{v^s}{\gamma_{sk}}, \quad \bar\gamma_{ii} = 1, \quad \bar\gamma_{is} = -\gamma_{ik}\frac{1}{\gamma_{sk}}, \quad \bar\gamma_{ij} = \gamma_{ij} - \gamma_{ik}\frac{\gamma_{sj}}{\gamma_{sk}},$$

$$j = r+1, \ldots, k-1, k+1, \ldots, n; \quad \bar\gamma_{ik} = 0, \quad \bar\gamma_{ij} = 0, \quad 1 \leqslant j \leqslant r, \quad j \neq i. \quad (22)$$

Наконец, заполним строку Δ таблицы 2. С этой целью подставим переменную x^k из (19) в (6), с учетом формулы (15) получим следующее выражение значения функции $f(x)$ через небазисные переменные точки w:

$$f(x) = f(v) - \sum_{j=r+1}^{n}{}' \Delta_j x^j - \Delta_k\left(\frac{v^s}{\gamma_{sk}} - \frac{1}{\gamma_{sk}}x^s - \sum_{j=r+1}^{n}{}' \frac{\gamma_{sj}}{\gamma_{sk}}x^j\right) =$$

$$= f(w) - \left(-\frac{\Delta_k}{\gamma_{sk}}\right)x^s - \sum_{j=r+1}^{n}{}' \left(\Delta_j - \Delta_k\frac{\gamma_{sj}}{\gamma_{sk}}\right)x^j. \quad (23)$$

Таблица 2

	Б	V	x^1	...	x^i	...	x^{s-1}	x^s	x^{s+1}	...	x^r
Γ_1	x^1	w^1	1	...	0	...	0	$\bar\gamma_{1s}$	0	...	0
...	...	...	...	...	...	...	...	...	...	...	...
Γ_i	x^i	w^i	0	...	1	...	0	$\bar\gamma_{is}$	0	...	0
...	...	...	...	...	...	...	...	...	...	...	...
Γ_{s-1}	x^s	w^{s-1}	0	...	0	...	1	$\bar\gamma_{s-1s}$	0	...	0
Γ_s	x^s	w^k	0	...	0	...	0	$\bar\gamma_{ss}$	0	...	0
Γ_{s+1}	x^s	w^{s+1}	0	...	0	...	0	$\bar\gamma_{s+1s}$	1	...	0
...	...	...	...	...	...	...	...	...	...	...	...
Γ_r	x^r	w^r	0	...	0	...	0	$\bar\gamma_{rs}$	0	...	1
Δ		$f(w)$	0	...	0	...	0	$\bar\Delta_s$	0	...	0

Из (23) следует, что в строке Δ симплекс-таблицы 2 точки w должны быть записаны величины $\bar\Delta_j$, $j=0, 1, \ldots, n$, определяемые формулами

$$\bar\Delta_0 = f(w) = f(v) - \Delta_k\frac{v^s}{\gamma_{sk}}, \quad \bar\Delta_s = -\Delta_k\frac{1}{\gamma_{sk}}, \quad \bar\Delta_j = \Delta_j - \Delta_k\frac{\gamma_{sj}}{\gamma_{sk}},$$

$$j = r+1, \ldots, k-1, k+1, \ldots, n; \quad \bar\Delta_k = 0; \quad \bar\Delta_j = 0,$$

$$j = 1, \ldots, s-1, s+1, \ldots, r. \quad (24)$$

Таким образом, симплекс-таблица 2 угловой точки w с базисом (16) полностью заполнена. Несложный анализ формул (21), (22), (24) с учетом конкретных числовых значений γ_{ij}, $\bar\gamma_{ij}$, Δ_j, $\bar\Delta_j$ в базисных столбцах таблиц 1,2 показывает, что элементы этих таблиц связаны следующими простыми соотношениями:

$$\bar\gamma_{sj} = \frac{\gamma_{sj}}{\gamma_{sk}}; \quad \bar\gamma_{ij} = \gamma_{ij} - \gamma_{ik}\frac{\gamma_{sj}}{\gamma_{sk}}, \quad i = 1, \ldots, r, \quad i \neq s, \quad j = 0, \ldots, n;$$

$$\bar\Delta_j = \Delta_j - \Delta_k\frac{\gamma_{sj}}{\gamma_{sk}}, \quad j = 0, \ldots, n. \quad (25)$$

Если элементы и строки таблицы 1 обозначить через $\gamma_{ij}(v)$, $\Delta_j(v)$, $\Gamma_i(v)$, $\Delta(v)$, а элементы и строки таблицы 2 через $\gamma_{ij}(w)$, $\Delta_j(w)$, $\Gamma_i(w)$, $\Delta(w)$, то соотношения (25) можно записать в векторной форме:

$$\Gamma_s(w) = \frac{\Gamma_s(v)}{\gamma_{sk}(v)}, \quad \Gamma_i(w) = \Gamma_i(v) - \gamma_{ik}(v)\frac{\Gamma_s(v)}{\gamma_{sk}(v)},$$

$$i = 1, \ldots, s-1, s+1, \ldots, r; \quad \Delta(w) = \Delta(v) - \Delta_k(v)\frac{\Gamma_s(v)}{\gamma_{sk}(v)}. \quad (26)$$

Соотношения (25) и (26) описывают один шаг известного метода Гаусса — Жордана, соответствующий исключению переменной x^k из всех строк симплекс-таблицы 1, кроме строки Γ_s, в которой переменная x^k остается с коэффициентом $\gamma_{sk}(w)=1$.

Таким образом, один шаг симплекс-метода, заключающийся в переходе от одной угловой точки v множества X к другой угловой точке w, описан. Этот шаг формально можно истолковать как переход от одной симплекс-таблицы 1 к другой симплекс-таблице 2 по формулам (25) или (26), где номера k, s и разрешающий элемент $\gamma_{sk}=\gamma_{sk}(v)$ выбираются из условий (12), (13).

Таблица 2

x^{r+1}	...	x^{k-1}	x^k	x^{k+1}	...	x^j	...	x^n
$\bar\gamma_{1r+1}$	...	$\bar\gamma_{1k-1}$	0	$\bar\gamma_{1k+1}$	...	$\bar\gamma_{1j}$	...	$\bar\gamma_{1n}$
...	...	...	...	...	...	...	...	...
$\bar\gamma_{ir+1}$	...	$\bar\gamma_{ik-1}$	0	$\bar\gamma_{ik+1}$	...	$\bar\gamma_{ij}$	...	$\bar\gamma_{in}$
...	...	...	...	...	...	...	...	...
$\bar\gamma_{s-1r+1}$	...	$\bar\gamma_{s-1k-1}$	0	$\bar\gamma_{s-1k+1}$	...	$\bar\gamma_{s-1j}$	...	$\bar\gamma_{s-1n}$
$\bar\gamma_{sr+1}$	...	$\bar\gamma_{sk-1}$	1	$\bar\gamma_{sk+1}$	...	$\bar\gamma_{sj}$	...	$\bar\gamma_{sn}$
$\bar\gamma_{s+1r+1}$	...	$\bar\gamma_{s+1k-1}$	0	$\bar\gamma_{s+1k+1}$	...	$\bar\gamma_{s+1j}$	...	$\bar\gamma_{s+1n}$
...	...	...	...	...	...	...	...	...
$\bar\gamma_{rr+1}$	...	$\bar\gamma_{rk-1}$	0	$\bar\gamma_{rk+1}$	...	$\bar\gamma_{rj}$	...	$\bar\gamma_{rn}$
$\bar\Delta_{r+1}$	...	$\bar\Delta_{k-1}$	0	$\bar\Delta_{k+1}$	...	$\bar\Delta_j$	...	$\bar\Delta_n$

2. Формулы перехода (25), (26) были получены в предположении, что множество номеров базисных переменных угловой точки v имеет специальный вид $I(v)=\{1, 2, \ldots, r\}$, что соответствует таблице 1. Конечно, путем перенумерации переменных всегда можно добиться, чтобы множество $I(v)$ имело указанный вид, но это связано с дополнительной обработкой числовых массивов, усложняет программную реализацию симплекс-метода на ЭВМ. Однако нетрудно убедиться, что можно обойтись без какой-либо перенумерации переменных, а формулы перехода (25), (26) остаются справедливыми для угловых точек с любым множеством базисных номеров.

В самом деле, пусть номера базисных переменных начальной точки v образуют множество $I(v)=\{j_1, j_2, \ldots, j_r\}$. Заметим, что в процессе применения симплекс-метода множество $I(v)$ обновляется на каждом шаге и нельзя ожидать, что номера $j_1, j_2, \ldots, j_r$ из этого множества будут упорядочены, скажем, в порядке монотонного возрастания или убывания (так, например, в таблице 2 в отличие от таблицы 1 монотонность номеров базисных переменных в столбце Б уже нарушена). Однако это обстоятельство нашим дальнейшим рассуждениям никак не помешает. Обозначим

$$\bar{x}=\begin{bmatrix} x^{j_1} \\ \ldots \\ x^{j_r} \end{bmatrix}, \quad \bar{v}=\begin{bmatrix} v^{j_1} \\ \ldots \\ v^{j_r} \end{bmatrix}, \quad \bar{c}=\begin{bmatrix} c^{j_1} \\ \ldots \\ c^{j_r} \end{bmatrix}, \quad A_j=\begin{bmatrix} a_{1j} \\ \ldots \\ a_{rj} \end{bmatrix},$$

$$B=\begin{bmatrix} a_{1j_1}, & \ldots, & a_{1j_r} \\ \ldots\ldots\ldots\ldots & & \\ a_{rj_1}, & \ldots, & a_{rj_r} \end{bmatrix}=(A_{j_1}, A_{j_2}, \ldots, A_{j_r}), \tag{27}$$

$$\gamma_k=\gamma_k(v)=B^{-1}A_k, \quad \gamma_0=\gamma_0(v)=B^{-1}b, \quad \gamma_{ik}=\gamma_{ik}(v)=(B^{-1}A_k)^i,$$
$$\gamma_{i0}=\gamma_{i0}(v)=(B^{-1}b)^i, \quad i=1, \ldots, r, \quad k=1, \ldots, n.$$

Так как $B^{-1}B=B^{-1}(A_{j_1}, \ldots, A_{j_r})=(B^{-1}A_{j_1}, \ldots, B^{-1}A_{j_r})=(e_1, \ldots, e_r)=I$ — единичная матрица размера $r\times r$, то $\gamma_{j_i}=B^{-1}A_{j_i}=e_i$ для всех $i=1, \ldots, r$. Кроме того, согласно теореме 2.1

$$B\bar{v}=A_{j_1}v^{j_1}+\ldots+A_{j_r}v^{j_r}=b; \quad v^j=0, \quad j\notin I(v),$$

поэтому $\bar{v}=B^{-1}b=\gamma_0$, $v^{j_i}=(B^{-1}b)^i=\gamma_{i0}$, $i=1, \ldots, r$; $v^j=0$, $j\notin I(v)$. Таким образом, умножая систему $Ax=\sum\limits_{i=1}^{n}A_ix^i=b$ слева на матрицу B^{-1} как и при выводе системы (4), (5), получим приведенную систему угловой точки v с базисом B в векторной форме

$$0\leqslant\bar{v}=B^{-1}b=\gamma_0=\bar{x}+\sum_{k\notin I(v)}(B^{-1}A_k)x^k=\sum_{k=1}^{n}\gamma_kx^k,$$

или в покоординатной форме

$$v^{j_i}=\gamma_{i0}=x^{j_i}+\sum_{k\notin I(v)}\gamma_{ik}x^k=\sum_{k=1}^{n}\gamma_{ik}x^k, \quad i=1, \ldots, r. \tag{28}$$

По аналогии с (6), (7) для целевой функции получим ее приведенную форму

$$f(x)=\langle\bar{c}, \bar{x}\rangle+\sum_{j\notin I(v)}c^jx^j=\langle\bar{c}, \bar{v}-\sum_{j\notin I(v)}(B^{-1}A_j)x^j\rangle+$$
$$+\sum_{j\notin I(v)}c^jx^j=\langle\bar{c}, \bar{v}\rangle-\sum_{j\notin I(v)}[\langle\bar{c}, B^{-1}A_j\rangle-c^j]x^j,$$

которую можно переписать в виде

$$\Delta_0=f(x)+\sum_{k\notin I(v)}\Delta_kx^k=f(x)+\sum_{k=1}^{n}\Delta_kx^k, \tag{29}$$

где приняты обозначения

$$\Delta_0=f(v)=\langle\bar{c}, \bar{v}\rangle, \quad \Delta_k=\langle\bar{c}, B^{-1}A_k\rangle-c^k=\langle\bar{c}, \gamma_k\rangle-c^k=\sum_{i=1}^{r}c^{j_i}\gamma_{ik}-c^k, \quad k=1, 2, \ldots, n, \tag{30}$$

причем учтено, что для всех $k=j_i\in I(v)$ величина $\Delta_{j_i}=\langle\bar{c}, B^{-1}A_{j_i}\rangle-c^{j_i}=\langle\bar{c}, e_i\rangle-c^{j_i}=c^{j_i}-c^{j_i}=0$, $i=1, \ldots, r$, т. е. $\Delta_k=0$, $\forall k\in I(v)$.

Информацию из (27)–(30) об угловой точке v с базисом $B=(A_{j_1}, \ldots, A_{j_r})$ удобно записать в виде симплекс-таблицы 3: строка Γ_i в ней соответствует i-у уравнению (28), строка Δ — представлению (29) для целевой функции.

Таблица 3

	Б	V	x^1	…	x^i	…	x^k	…	x^{j_i}	…	x^n
Γ_1	x^{j_1}	γ_{10}	γ_{11}	…	γ_{1i}	…	γ_{1k}	…	$\gamma_{1j_i}=0$	…	γ_{1n}
…	…	…	…	…	…	…	…	…	…	…	…
Γ_i	x^{j_i}	γ_{i0}	γ_{i1}	…	γ_{ii}	…	γ_{ik}	…	$\gamma_{ij_i}=1$	…	γ_{in}
…	…	…	…	…	…	…	…	…	…	…	…
Γ_s	x^{j_s}	γ_{s0}	γ_{s1}	…	γ_{si}	…	γ_{sk}	…	$\gamma_{sj_i}=0$	…	γ_{sn}
…	…	…	…	…	…	…	…	…	…	…	…
Γ_r	x^{j_r}	γ_{r0}	γ_{r1}	…	γ_{ri}	…	γ_{rk}	…	$\gamma_{rj_i}=0$	…	γ_{rn}
Δ		Δ_0	Δ_1	…	Δ_i	…	Δ_k	…	$\Delta_{j_i}=0$	…	Δ_n

Отметим, что в столбце базисной переменной x^{j_i} вектор $\gamma_{j_i}=e_i$, т. е. $\gamma_{lj_i}=0$ при всех $l\neq i$, $1\leqslant l\leqslant r$, $\gamma_{ij_i}=1$; в нижней строке этого столбца $\Delta_{j_i}=0$. Симплекс-таблицу 3 можно кратко записать в виде матрицы

$$S=(v, B)=\begin{pmatrix} \gamma_0, & \gamma_1, & \ldots, & \gamma_n \\ \Delta_0, & \Delta_1, & \ldots, & \Delta_n \end{pmatrix}=\begin{pmatrix} \Gamma \\ \Delta \end{pmatrix}$$

размера $(r+1)\times(n+1)$, где столбцы γ_k подматрицы $\Gamma=\begin{pmatrix}\Gamma_1\\ \ldots\\ \Gamma_r\end{pmatrix}$ и элементы строки Δ вычисляются по стандартным формулам:

$$\gamma_k=B^{-1}A_k, \quad \Delta_k=\langle\bar{c}, B^{-1}A_k\rangle-c^k=\langle\bar{c}, \gamma_k\rangle-c^k, \quad k=0, 1, \ldots, n; \tag{31}$$

здесь для единообразия формул принято $b=A_0$, предполагается, что $c^0=0$, остальные обозначения взяты из (27).

Опишем один шаг симплекс-метода в общем случае. По аналогии с (10)–(12) рассмотрим следующие три взаимоисключающие возможности.

Случай I. Справедливы неравенства

$$\Delta_j=\langle\bar{c}, B^{-1}A_j\rangle-c^j\leqslant 0, \quad j=1, \ldots, n, \tag{32}$$

т. е. в нижней строке симплекс-таблицы 3 все величины $\Delta_1, \ldots, \Delta_n$ непо-

ложительны. Тогда с учетом равносильности систем (2) и (28) для любой точки $x \in X$ имеем

$$f(x) = \langle \bar{c}, \bar{x} \rangle + \sum_{j \notin I(v)} c^j x^j \geqslant \langle \bar{c}, \bar{x} \rangle + \sum_{j \notin I(v)} \langle \bar{c}, B^{-1}A_j \rangle x^j = \langle \bar{c}, \bar{x} + \sum_{j \notin I(v)} B^{-1}A_j x^j \rangle = \langle \bar{c}, \bar{v} \rangle = f(v).$$

Это значит, что v — решение задачи (1).

Случай II. Существует номер $k > 0$, $k \notin I(v)$, такой, что

$$\Delta_k > 0, \quad \gamma_k = B^{-1}A_k \leqslant 0, \tag{33}$$

т. е. в k-м столбце симплекс-таблицы 3 над $\Delta_k > 0$ нет ни одной положительной величины γ_{ik}. Тогда точка $x = x(t) = (x^1, \ldots, x^n)$ с координатами $x^k = t$, $x^{j_i} = v^{j_i} - \gamma_{ik}t$, $i = 1, \ldots, r$; $x^j = 0$, $j \notin I(v)$, $j \neq k$, будет принадлежать множеству X при всех $t \geqslant 0$. Отсюда и из (29) следует, что $f(x(t)) = f(v) - \Delta_k t \to -\infty$ при $t \to +\infty$. Это значит, что $f_* = \inf\limits_{x \in X} f(x) = -\infty$, т. е. задача (1) не имеет решения.

Случай III. Существует номер $k > 0$, $k \notin I(v)$, для которого $\Delta_k > 0$, причем для каждого такого номера k найдется номер i, $1 \leqslant i \leqslant r$, что $\gamma_{ik} > 0$, или, короче,

$$\forall \Delta_k > 0 \quad \exists \gamma_{ik} = (B^{-1}A_k)^i > 0. \tag{34}$$

Это значит, что в каждом k-м столбце симплекс-таблицы 3 над величиной $\Delta_k > 0$ имеется хотя бы одно положительное число γ_{ik}. Тогда выберем номер s и разрешающий элемент $\gamma_{sk} > 0$ из условий:

$$\min_{i \in I_k(v)} \frac{\gamma_{i0}}{\gamma_{ik}} = \frac{\gamma_{s0}}{\gamma_{sk}}, \quad s \in I_k(v) = \{i: 1 \leqslant i \leqslant r, \ \gamma_{ik} > 0\}. \tag{35}$$

Далее, рассуждая также, как выше (см. формулы (14)–(16) и пояснения к ним), убеждаемся, что точка $w = (w^1, \ldots, w^n)$ с координатами

$$w^{j_i} = v^{j_i} - \gamma_{ik}\frac{v^{j_s}}{\gamma_{sk}} = \gamma_{i0} - \gamma_{ik}\frac{\gamma_{s0}}{\gamma_{sk}}, \quad i = 1, \ldots, r, \quad i \neq s; \quad w^{j_s} = 0,$$
$$w^k = \frac{v^{j_s}}{\gamma_{sk}} = \frac{\gamma_{s0}}{\gamma_{sk}}; \quad w^j = 0, \quad j \notin I(v), \quad j \neq k, \tag{36}$$

принадлежит множеству X, является угловой точкой этого множества с базисом

$$(A_{j_1}, \ldots, A_{j_{s-1}}, A_k, A_{j_{s+1}}, \ldots, A_{j_r}) = \overline{B}, \tag{37}$$

значение функции $f(x)$ в этой точке равно

$$f(w) = f(v) - \Delta_k \frac{v^{j_s}}{\gamma_{sk}} = \Delta_0 - \Delta_k \frac{\gamma_{s0}}{\gamma_{sk}}. \tag{38}$$

Замечание 1 с очевидными изменениями сохраняет силу в рассматриваемом общем случае. Приведенная система точки w выводится так же, как система (18), (20), и имеет вид:

$$w^k = \frac{1}{\gamma_{sk}}x^{j_s} + \sum_{j \notin I(v)}{}' \frac{\gamma_{sj}}{\gamma_{sk}}x^j + x^k,$$
$$w^{j_i} = x^{j_i} + \left(-\frac{\gamma_{ik}}{\gamma_{sk}}\right)x^{j_s} + \sum_{j \notin I(v)}{}' \left(\gamma_{ij} - \gamma_{ik}\frac{\gamma_{sj}}{\gamma_{sk}}\right)x^j; \tag{39}$$

аналогичное (23) выражение для функции $f(x)$ выглядит так

$$f(x) = f(w) - \left(-\frac{\Delta_k}{\gamma_{sk}}\right)x^{j_s} - \sum_{j \notin I(v)}{}' \left(\Delta_j - \Delta_k\frac{\gamma_{sj}}{\gamma_{sk}}\right)x^j; \tag{40}$$

в (39), (40) знак $\sum'$ означает, что суммирование ведется по всем $j \notin I(v)$, $j \neq k$. Нетрудно видеть, что если $I(v) = \{1, 2, \ldots, r\}$, то формулы (35)–(40) переходят в соответствующие формулы (13)–(16), (18), (20), (23). Анализируя коэффициенты при переменных $x^1, x^2, \ldots, x^n$ в выражениях (39), (40), получаем аналогичные (21), (22), (24) формулы для величин, которые должны находиться в строках $\Gamma_i(w)$, $i = 1, 2, \ldots, r$, $\Delta(w)$ симплекс-таблицы точки w, и убеждаемся в том, что переход от симплекс-таблицы точки v с базисом $B = (A_{j_1}, \ldots, A_{j_r})$ к симплекс-таблице точки w с базисом (37) совершается по тем же формулам (25), (26), где номера k, s определяются из условий (34), (35).

Для иллюстрации вышеизложенного приведем несколько примеров.

Пример 1. Будем минимизировать функцию $f(x) = 10x^2 - x^3 + 4x^4 + x^5$ на множестве $X = \{x = (x^1, x^2, \ldots, x^5) \geqslant 0: \ x^1 + 2x^3 + x^4 = 2, \ 2x^1 - x^3 + x^5 = 3, \ -x^1 + x^2 + x^3 = 1\}$. Уравнения, задающие это множество можно записать в виде

$$\begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} x^1 + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} x^2 + \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} x^3 + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} x^4 + \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} x^5.$$

Как и выше, столбец из коэффициентов при переменной x^i будем обозначать через A_i. Нетрудно видеть, что выписанная система уравнений является приведенной системой для угловой точки $v_0 = (0, 1, 0, 2, 3)$ с базисом A_2, A_4, A_5; здесь $I(v_0) = \{j_1 = 4, j_2 = 5, j_3 = 2\}$, $B = (A_4, A_5, A_2)$. Внесем коэффициенты этой системы в строки $\Gamma_1, \Gamma_2, \Gamma_3$ симплекс-таблицы 4 точки v_0. Пользуясь формулами (30), (31), вычислим значения величин Δ_j, $j = 0, 1, \ldots, 5$

Таблица 4

	Б	V	x^1	x^2	x^3	x^4	x^5
Γ_1	x^4	2	1	0	2	1	0
Γ_2	x^5	3	2	0	−1	0	1
Γ_3	x^2	1	−1	1	①	0	0
Δ		21	−4	0	18	0	0

и впишем их в строку Δ таблицы 4. В этой строке величина $\Delta_3 > 0$, а в столбце x^3 имеются положительные элементы $\gamma_{13} = 2$, $\gamma_{33} = 1$. Это значит, что в точке v_0 реализовались условия (34). Определим номер s из условия (35): $\min\{\gamma_{10}/\gamma_{13}, \gamma_{30}/\gamma_{33}\} = \min\{2/2, 1/1\} = 1$. Как видим, здесь минимум достигается сразу при двух значениях $s = 1$ и $s = 3$. Для определенности возьмем $s = 3$. Тогда разрешающий элемент равен $\gamma_{33} = 1$, $k = 3$, $s = 3$. В таблице 4 и в последующих таблицах разрешающий элемент будем обводить кружочком. В соответствии с выбранным разрешающим элементом переменную $x^{j_3} = x^2$ и столбец $A_{j_3} = A_2$ будем выводить из базиса и заменим их переменной x^3 и столбцом A_3 соответственно. Согласно формуле (26) разделим строку Γ_3 на $\gamma_{33} = 1$ и полученные величины внесем в строку Γ_3 таблицы 5. Затем будем

последовательно умножать строку Γ_3 таблицы 5 на величины $\gamma_{13}=2$, $\gamma_{23}=-1$, $\gamma_{23}=-1$, $\Delta_3=18$, получившиеся строки вычтем соответственно из строк Γ_1, Γ_2, Δ таблицы 4 и результат вычитания внесем в строки Γ_1, Γ_2, Δ таблицы 5. Таким образом, придем к симплекс-таблице 5 следующей угловой точки

Таблица 5

	Б	V	x^1	x^2	x^3	x^4	x^5
Γ_1	x^4	0	③	−2	0	1	0
Γ_2	x^5	4	1	1	0	0	1
Γ_3	x^3	1	−1	1	1	0	0
Δ		3	14	−18	0	0	0

$v_1=(0,0,1,0,4)$ с базисом $B_1=(A_4, A_5, A_3)$, со множеством базисных номеров $I(v_1)=\{j_1=4, j_2=5, j_3=3\}$ и со значением функции $f(v_1)=3<f(v_0)=21$.

В строке Δ таблицы 5 величина $\Delta_1=14>0$, в столбце x^1 имеются положительные элементы $\gamma_{11}=3$, $\gamma_{21}=1$, т. е. снова реализовались условия (34). По правилу (35): $\min\{0/3; 4/1\}=0$ однозначно определяем номер $s=1$ и разрешающий элемент $\gamma_{11}=3$. Это значит, что переменную $x^{j_1}=x^4$ и столбец $A_{j_1}=A_4$ выводим из базиса и заменяем их переменной x^1 и столбцом A_1 соответственно. По формулам (26) вычислим симплекс-таблицу 6 следующей

Таблица 6

	Б	V	x^1	x^2	x^3	x^4	x^5
Γ_1	x^4	0	1	−2/3	0	1/3	0
Γ_2	x^5	4	0	5/3	0	−1/3	1
Γ_3	x^3	1	0	1/3	1	1/3	0
Δ		3	0	−26/3	0	−14/3	0

угловой точки $v_2=(0,0,1,0,4)$ с базисом $B_2=(A_1, A_5, A_3)$, со множеством $I(v_2)=\{j_1=1, j_2=5, j_3=3\}$ и со значением функции $f(v_2)=3=f(v_1)$. В строке Δ этой таблицы среди величин $\Delta_1, \ldots, \Delta_5$ нет положительных. Это значит, что реализовался случай (32), точка $v_2=(0,0,1,0,4)$ является решением рассматриваемой задачи, $f_*=f(v_2)=3$.

Заметим, что точки v_1 и v_2 совпадают и различаются лишь базисами. Выясняется, что еще в таблице 5 мы, оказывается, уже получили решение задачи, но не смогли это распознать и вынуждены были сделать еще один шаг симплекс-метода.

Пример 2. Рассмотрим задачу: минимизировать функцию $f(x)=x^1+x^2-x^3-x^4+x^5$ при условиях $x=(x^1,\ldots,x^5)\geqslant 0$, $x^1-x^3+x^4+x^5=1$, $x^2+x^3-x^4+x^5=1$. Нетрудно видеть, что $v_0=(1,1,0,0,0)$ — угловая точка с базисом $B_0=(A_1, A_2)$, со множеством $I(v_0)=\{j_1=1, j_2=2\}$ и система уравнений, задающая множество, уже записана в приведенной форме. Табл. 7 представляет собой симплекс-таблицу точки v_0. В строке Δ имеется несколько положительных величин $\Delta_3=\Delta_4=\Delta_5=1$. В качестве разрешающего элемента выберем величину $\gamma_{23}=1$ из столбца x^3. По формулам (26)

при $s=2$, $k=3$ совершим переход к симплекс-таблице 8 угловой точки $v_1=(2,0,1,0,0)$ с базисом $B_1=(A_1, A_3)$, со множеством $I(v_1)=\{j_1=1, j_2=3\}$. В строке Δ таблицы 8 имеется величина $\Delta_4=2>0$, но столбец x^4 не содержит положительных элементов. Это значит, что реализовался случай (33). Следовательно, $f_*=-\infty$, рассматриваемая задача не имеет решения.

Таблица 7

Б	V	x^1	x^2	x^3	x^4	x^5
x^1	1	1	0	−1	1	1
x^2	1	0	1	①	−1	1
	2	0	0	1	1	1

Таблица 8

x^1	2	1	1	0	0	2
x^3	1	0	1	1	−1	1
	1	0	−1	0	2	0

Подобно тому, как это сделано в таблицах 7, 8, в последующих симплекс-таблицах мы иногда будем опускать обозначения строк или столбцов, полагая, что читатель уже привык к обозначениям.

3. Из вышеизложенного следует, что, имея какую-либо начальную угловую точку v_0 множества X, с помощью симплекс-метода последовательно переходя от одной угловой точки к другой, можно построить последовательность угловых точек $v_0, v_1, \ldots, v_p, \ldots$ Согласно (38) на каждом шаге имеем:

$$f(v_{p+1})=f(v_p)-\Delta_k(v_p)\frac{\gamma_{s0}(v_p)}{\gamma_{sk}(v_p)},$$

где $\Delta_k(v_p)>0$, $\gamma_{sk}(v_p)>0$, $\gamma_{s0}(v_p)=v_p^{j_s}\geqslant 0$. Отсюда следует, что

$$f(v_0)\geqslant f(v_1)\geqslant\ldots\geqslant f(v_p)\geqslant\ldots \tag{41}$$

Процесс получения последовательностей $\{v_p\}$, $\{f(v_p)\}$ в дальнейшем будем кратко называть *симплекс-процессом*.

Заметим, что в примерах 1, 2 симплекс-процесс завершился за конечное число шагов выполнением одного из условий (32) или (33). Однако всегда ли это будет так? Возможно, существуют канонические задачи, для которых симплекс-процесс может неограниченно продолжаться? Для ответа на этот принципиально важный вопрос, внимательнее проанализируем описанный симплекс-процесс. Прежде всего заметим, что поскольку варианты (32)–(34) изменения знаков величин $\Delta_k(v)$, $\gamma_{ik}(v)$ исчерпывают все возможности и взаимоисключают друг друга, то симплекс-процесс может быть бесконечным лишь в том случае, когда на каждом шаге этого процесса будут реализовываться условия (34). Каждая реализация условий (34) связана с переходом от одной угловой точки к другой угловой точке, от одной симплекс-таблицы к другой симплекс-таблице. Это значит, что всякий бесконечный симплекс-процесс порождает последовательности угловых точек

$\{v_p\}$, их базисов $\{B_p\}$, симплекс-таблиц $\{S_p\}$, где S_p является симплекс-таблицей точки v_p с базисом B_p, причем, как видно из (41), соответствующая последовательность $\{f(v_p)\}$ не возрастает. Поскольку угловых точек и их базисов в задаче (1) конечное число, то конечно и множество симплекс-таблиц этой задачи. Отсюда следует, что симплекс-процесс может быть бесконечным лишь в том случае, когда хотя бы одна из симплекс-таблиц S, соответствующая некоторой угловой точке v с базисом B, будет повторяться бесконечно много раз. Это значит, что найдется бесконечная подпоследовательность номеров $\{p_l\}$: $p_1 < p_2 < \ldots < p_l < \ldots$, такая, что $v_{p_l} = v$, $B_{p_l} = B$, $S_{p_l} = S$, $f(v_{p_l}) = f(v)$ при всех $l = 1, 2, \ldots$ В силу (41) это возможно лишь тогда, когда

$$f(v_p) = \text{const} \quad \forall p \geqslant p_1. \tag{42}$$

Таким образом, необходимым условием бесконечности симплекс-процесса является условие (42), которое должно выполняться, начиная с некоторого номера p_1. Посмотрим, когда это возможно. Начнем с выяснения того, когда $f(w) = f(v)$ и когда $f(w) < f(v)$, где угловая точка w получена из угловой точки v в результате одного шага симплекс-метода. В силу условий (34), (35) $\Delta_k(v) > 0$, $\gamma_{sk}(v) > 0$ и, кроме того, $\gamma_{s0}(v) = v^{j_s} \geqslant 0$ как базисная переменная угловой точки v. Отсюда и из формулы (38) следует, что $f(w) = f(v)$ тогда и только тогда, когда $v^{j_s} = 0$, т. е. v — вырожденная угловая точка. Такое явление мы наблюдали в примере 1 при переходе от таблицы 5 к таблице 6. Таким образом, среди канонических задач имеет смысл выделять задачи вырожденные и невырожденные.

Определение 1. Задачу (1) называют *вырожденной* или *невырожденной*, если множество X в этой задаче содержит хотя бы одну вырожденную угловую точку или не содержит таковые.

Покажем, что в невырожденных задачах (1) симплекс-процесс всегда конечен. В самом деле, в таких задачах все базисные координаты угловой точки v будут положительны. Поэтому какими ни были номера k, s, определяемые из условий (34), (35), всегда $v^{j_s} > 0$, и, согласно (38), тогда $f(w) < f(v)$.

Отсюда следует, что в невырожденных задачах симплекс-процесс порождает такую последовательность угловых точек $v_0, v_1, \ldots, v_p, \ldots$, для которых

$$f(v_0) > f(v_2) > \ldots > f(v_p) > \ldots \tag{43}$$

Поскольку угловых точек конечное число, и из-за строгих неравенств они повторяться в симплекс-процессе не могут, то этот процесс закончится на каком-то шаге выполнением либо условия (32), либо (33). Впрочем, конечность симплекс-процесса здесь вытекает и из несовместимости соотношений (42), (43). Таким образом, доказана

Теорема 1. *Пусть в канонической задаче* (1) *множество X непусто и невырождено*, $\operatorname{rang} A = r = m < n$, *пусть v_0 — произвольная угловая точка этого множества. Тогда симплекс-процесс, начинающийся с точки v_0 при выборе разрешающего элемента γ_{sk} из условий* (34), (35), *завершится за конечное число шагов определением некоторой угловой точки v_p множества X, в которой реализуются либо условия* (32), *либо* (33), *причем в случае* (32) *v_p — решение задачи (1), $f(v_p) = f_* > -\infty$, в случае* (33) *задача* (1) *не имеет решения, $f_* = -\infty$.*

Заметим, что хотя теорема 1 справедлива при любом выборе номеров k, s из условий (34), (35), но продолжительность симплекс-процесса и последняя точка v_p могут существенно зависеть от выбора этих номеров. Впрочем, интересно отметить, что если номер k из условий (34) как-то уже выбран и зафиксирован, то в невырожденных задачах номер s условием (35) определяется однозначно. В самом деле, для невырожденной угловой точки $w = (w^1, \ldots, w^n)$ с базисом (37) координаты $w^{j_i} > 0$ для $i \neq s$, $1 \leqslant i \leqslant r$. Из формул (36) тогда следует, что $v^{j_i} - \gamma_{ik} v^{j_s}/\gamma_{sk} > 0$ или $v^{j_i}/\gamma_{ik} > v^{j_s}/\gamma_{sk}$ для всех $i \in I_k(v)$, $i \neq s$, так что минимум в левой части (35) будет достигаться на единственном номере $s \in I_k(v)$.

Отсюда следует, что условие (35) может неоднозначно определять номер s лишь в вырожденных задачах. Кстати говоря, если в (35) минимум достигается по крайней мере на двух номерах $s, l \in I_k(v)$, $s \neq l$, то, согласно формулам (36), $w^{j_s} = w^{j_l} = 0$, т. е. угловая точка w непременно будет вырожденной (так случилось в таблицах 4, 5). Конечно, точка w может быть вырожденной и в том случае, когда условие (35) однозначно определяет номер $s \in I_k(v)$, для которого $v^{j_s} = 0$ (как видим, тогда сама точка v вырожденная); в этом случае, как видно из (36), у точки w базисная координата $w^k = 0$ (см. табл. 5, 6). Впрочем, если

$$s \in I_k(v), \quad v^{j_s} = 0, \tag{44}$$

то минимум в (35) равен нулю и будет достигаться именно на этом номере s, (и на всех других номерах $l \in I_k(v)$, для которых $v^{j_l} = 0$) и, согласно формулам (36), (38), тогда $w = v$, $f(w) = f(v)$. Это значит, что при выполнении условий (44) мы сделаем один шаг симплекс-метода и останемся в той же точке $w = v$, лишь заменив один ее базис $B = (A_{j_1}, \ldots, A_{j_r})$ на другой базис (37) (именно так случилось в таблицах 5, 6). Здесь возникает естественный вопрос: при выполнении условий (44) не может ли привести дальнейшее применение симплекс-метода к бесконечному перебору базисов угловой точки v, не может ли здесь реализоваться бесконечный симплекс-процесс? Оказывается, так вполне может быть. Приведем пример вырожденной задачи [775], в которой симплекс-процесс приводит к так называемому зацикливанию, заключающемуся в бесконечном циклическом переборе базисов одной и той же угловой точки (другой пример — см. упражнение 6).

Пример 3. Рассмотрим задачу минимизации функции

$$f(x) = x^1 - x^4 - x^5 + x^6 \tag{45}$$

при условиях

$$x = (x^1, x^2, \ldots, x^7) \geqslant 0, \quad x^1 + x^4 + x^5 + x^6 + x^7 = 1,$$
$$-2x^1 + x^2 + x^4 - 3x^5 + 4x^6 = 0, \quad 3x^1 + x^3 + 4x^4 - 2x^5 + x^6 = 0. \tag{46}$$

Нетрудно видеть, что точка $v_0 = (0, 0, 0, 0, 0, 0, 1)$ является угловой точкой с базисом $(A_7, A_2, A_3) = B_0$, система (46) представляет собой приведенную систему этой точки. Образуем симплекс-процесс, взяв в качестве начальной точку v_0 с указанным базисом. В таблицах 9–15 приведены результаты вычислений для первых точек $v_0, v_1, \ldots, v_7$; в кружочках указаны разрешающие элементы этих таблиц. В таблицах 9, 11, 13 разрешающий элемент условием (35) определяется неоднозначно, в таблицах 10, 12, 14 разрешающий элемент находится однозначно. Как видно, таблицы 9 и 15 совпадают

Таблица 9

	Б	V	x^1	x^2	x^3	x^4	x^5	x^6	x^7
Γ_1	x^7	1	1	0	0	1	1	1	1
Γ_2	x^2	0	−2	1	0	(1)	−3	4	0
Γ_3	x^3	0	−3	0	1	4	−2	1	0
Δ		0	−1	0	0	1	1	−1	0

Таблица 10

x^7	1	3	−1	0	0	4	−3	1
x^4	0	−2	1	0	1	−3	4	0
x^3	0	(5)	−4	1	0	10	−15	0
	0	1	−1	0	0	4	−5	0

Таблица 11

x^7	1	0	7/5	−3/5	0	−2	6	1
x^4	0	0	−3/5	2/5	1	(1)	−2	0
x^1	0	1	−4/5	1/5	0	2	−3	0
	0	0	−1/5	−1/5	0	2	−2	0

Таблица 12

x^7	1	0	1/5	1/5	2	0	2	1
x^5	0	0	−3/5	2/5	1	1	−2	0
x^1	0	1	2/5	−3/5	−2	0	(1)	0
	0	0	1	−1	−2	0	2	0

Таблица 13

Б	V	x^1	x^2	x^3	x^4	x^5	x^6	x^7
x^7	1	−2	−3/5	7/5	6	0	0	1
x^5	0	2	(1/5)	−4/5	−3	1	0	0
x^6	0	1	2/5	−3/5	−2	0	1	0
	0	−2	1/5	1/5	2	0	0	0

Таблица 14

x^7	1	4	0	−1	−3	3	0	1
x^2	0	10	1	−4	−15	5	0	0
x^6	0	−3	0	(1)	4	−2	1	0
	0	−4	0	1	5	−1	0	0

Таблица 15

Б	V	x^1	x^2	x^3	x^4	x^5	x^6	x^7
x^7	1	1	0	0	1	1	1	1
x^2	0	−2	1	0	(1)	−3	4	0
x^3	0	−3	0	1	4	−2	1	0
	0	−1	0	0	1	1	−1	0

и поэтому, если на следующих шагах продолжать выбор тех же разрешающих элементов в том же порядке, то придем к бесконечному симплекс-процессу, в котором будет осуществляться циклический перебор базисов точки v_0 в следующем порядке: $(A_7, A_2, A_3) \to (A_7, A_4, A_3) \to (A_7, A_4, A_1) \to (A_7, A_5, A_1) \to (A_7, A_5, A_6) \to (A_7, A_2, A_6) \to (A_7, A_2, A_3) \to \ldots$ Любопытно отметить, что длина цикла в задачах линейного программирования меньше шести не бывает [775].

Этот пример показывает, что описанный выше симплекс-метод действительно может привести к бесконечному симплекс-процессу и с его помощью может быть решена не всякая каноническая задача (1). Если функция $f(x) = \langle c, x \rangle$ принимает одинаковые значения в нескольких вырожденных угловых точках, то, по-видимому, возможны более сложные бесконечные симплекс-процессы, в частности, явления зацикливания с участием в цикле базисов различных таких точек.

4. Можно ли избежать зацикливания или, точнее, появления бесконечных симплекс-процессов? Нельзя ли уточнить правило (34), (35) выбора разрешающего элемента так, чтобы для любой задачи (1) симплекс-процесс, начинающийся с произвольной начальной угловой точки, завершался за конечное число шагов реализацией одного из условий (32) или (33)? Положительный ответ на эти вопросы имеет важное значение для обоснования симплекс-метода и означал бы, что можно, по крайней мере в принципе, решить любую задачу линейного программирования симплекс-методом.

Определение 2. Любое уточняющее (34), (35) правило выбора разрешающего элемента, с помощью которого можно избежать зацикливания или, точнее, появления бесконечного симплекс-процесса во всякой канонической задаче (1), назовем *антициклином*.

На практике правило (34), (35) нередко уточняют следующим образом: среди номеров k, удовлетворяющих условиям (34), выбирают тот, для которого Δ_k принимает максимальное значение, а если таких номеров несколько, то берут минимальный из них, и затем, после такой фиксации номера k, берут номер s минимально возможным из условий (35). Такое уточнение правил (34), (35) действительно гарантирует однозначность выбора разрешающего элемента γ_{sk}, выглядит вполне естественным и в примере 3, как легко проверить, на самом деле позволяет избежать зацикливания. Однако пример задачи из упражнения 6 (см. ниже) показывает, что в общем случае в классе канонических задач линейного программирования такое уточнение правила (34), (35) не спасает от зацикливания и, следовательно, не может служить антициклином. Это говорит о том, что построение антициклина — дело тонкое, и с первого взгляда неясно даже, существуют ли они. К счастью, антициклины существуют, и к настоящему времени уже созданы различные и не очень сложные антициклины (см., например, [52; 54; 116; 148; 259; 499; 517; 652; 685]).

Остановимся на одном из них [52]. Для описание этого антициклина нам понадобится понятие лексикографического упорядочения конечномерного пространства.

Определение 3. Говорят, что вектор $x=(x^1,\ldots,x^l)\in\mathbb{R}^l$ *лексикографически положителен (отрицателен)*, и обозначают $x\succ 0$ $[x\prec 0]$, если $x\neq 0$ и первая ненулевая координата вектора x положительна (отрицательна). Говорят, что вектор $x\in\mathbb{R}^l$ *лексикографически больше (меньше)* вектора $y\in\mathbb{R}^l$, и пишут $x\succ y$ $[x\prec y]$, если $x-y\succ 0$ $[x-y\prec 0]$.

Другими словами, запись $x\succ 0$ означает, что существует номер p, $1\leqslant p\leqslant l$, такой, что $x^1=\ldots=x^{p-1}=0$, $x^p>0$, остальные координаты $x^{p+1},\ldots,x^n$ могут быть любыми. Лексикографическое неравенство $x\succ y$ означает существование такого номера p, $1\leqslant p\leqslant l$, что $x^1=y^1,\ldots,x^{p-1}=y^{p-1}$, $x^p>y^p$.

Для любых $x,\ y\in\mathbb{R}^l$ выполнено одно и только одно из соотношений: $x\succ y$, $x\prec y$, или $x=y$. Ясно, что отношение $\succ$ транзитивно, т. е. если $x\succ y$, $y\succ z$, то $x\succ z$. Упорядочение векторов в их лексикографическом убывании (или возрастании) вполне аналогично упорядочению слов в словарях, что и объясняет присутствие слова «лексикографический» в определении 3.

Опираясь на определение 3, несложно вывести следующие, важные для дальнейшего, свойства отношения $\succ$:

1) если $x\succ 0$, то $\alpha x\succ 0$ для всех чисел $\alpha>0$;
2) если $x\succ y$, то $\alpha x\succ\alpha y$ для всех $\alpha>0$;
3) если $x\succ 0$, $y\succ 0$, то $x+\alpha y\succ 0$ для всех $\alpha\geqslant 0$;
4) если $x\succ 0$, то $y\succ y-\alpha x$ для всех $\alpha>0$ и $y\in\mathbb{R}^l$.

Определение 4. Пусть M_0 — некоторое множество целых чисел (номеров), пусть $G=\{y_i=(y_i^1,\ldots,y_i^l)\in\mathbb{R}^l,\ i\in M_0\}$. Вектор y_s, $s\in M_0$ называется *лексикографическим минимумом* множества G, если для всех $i\in M_0$ либо $y_i\succ y_s$, либо $y_i=x_s$. Обозначение: $y_s=\operatorname{lex}\min\limits_{i\in M_0} y_i$.

Лемма 1. *Пусть M_0 — конечное множество номеров и пусть во множестве $G=\{y_i\in\mathbb{R}^l,\ i\in M_0\}$ все векторы различны. Тогда лексикографический минимум множества G достигается на единственном векторе $y_s\in G$, т. е. $y_s\prec y_i\ \forall i\in M_0$ (случай $y_i=y_s$ при $i\neq s$ исключается). Для определения номера s нужно последовательно строить множества M_0, $M_1=\{s:\ s\in M_0,\ y_s^1=\min\limits_{i\in M_0} y_i^1\}$, $\ldots$, $M_p=\{s:\ s\in M_{p-1},\ y_s^p=\min\limits_{i\in M_{p-1}} y_i^p\}$ до тех пор, пока не будет обнаружено множество M_ν, $0\leqslant\nu\leqslant l$, состоящее из единственного номера s, который и будет искомым.*

Доказательство. В простейшем случае, когда множество M_0 состоит из единственного номера s, то, по определению, y_s — искомый вектор. Поэтому пусть M_0 содержит более одного номера. Тогда строим множество M_1. Если M_1 содержит лишь один номер s, то $y_s^1<y_i^1$ для всех $i\in M_0$, $i\neq s$, и ясно, что $y_s=\operatorname{lex}\min\limits_{i\in M_0} y_i$. Если M_1 содержит по крайней мере два номера, то строим множество $M_2=\{i\in M_1:\ y_s^2=\min\limits_{i\in M_1} y_i^2\}$ и т. д. Пусть уже построены множества $M_0\supset M_1\supset\ldots\supset M_p$, $p<l$, причем множества $M_0,\ldots,M_{p-1}$ содержат более одного номера. Если M_p состоит из одного номера s, то y_s — искомый вектор. Если M_p содержит более одного номера, то строим множество M_{p+1} и т. д. В крайнем случае, когда множества $M_0,\ldots,M_{l-1}$ окажутся состоящими более чем из одного номера, этот процесс закончится построением множества $M_l=\{s\in M_{l-1}:\ y_s^l=\min\limits_{i\in M_{l-1}} y_i^l\}$. Если бы M_l содержал два различных номера s, q, то у векторов y_s, y_q все координаты были бы одинаковыми, т. е. $y_s=y_q$. Однако по условию во множестве G нет двух одинаковых векторов. Следовательно, M_l состоит из единственного номера s, причем $y_s=\operatorname{lex}\min\limits_{i\in M_0} y_i$. Лемма доказана. □

Опираясь на отношение $\succ$ между векторами, введем отношения $\overset{\Gamma}{\succ}$, $\overset{\Delta}{\succ}$ на множестве симплекс-таблиц. Не стремясь к общности построений, мы можем ограничиться следующим определением, достаточным для дальнейших рассмотрений.

Определение 5. Симплекс-таблицу $S=S(v,B)$ угловой точки v с базисом B назовем *лексикографически положительной* и будем обозначать $S\overset{\Gamma}{\succ}0$, если все ее строки $\Gamma_i=(\gamma_{i0},\gamma_{i1},\ldots,\gamma_{in})\succ 0$, $i=1,\ldots,r$ (см. табл. 3). Скажем, что симплекс-таблица $S_1=S(v_1,B_1)$ лексикографически больше другой симплекс-таблицы $S_2=S(v_2,B_2)$ и будем обозначать $S_1\overset{\Delta}{\succ}S_2$, если строка $\Delta_1=(\Delta_{10},\ldots,\Delta_{1n})$ таблицы S_1 лексикографически больше строки $\Delta_2=(\Delta_{20},\ldots,\Delta_{2n})$ таблицы S_2.

Для примера укажем, что таблицы 4–8, 13 лексикографически положительны, таблицы 9–12, 14, 15 не являются таковыми; симплекс-таблица 12 лексикографически больше симплекс-таблицы 11.

Нетрудно видеть, что если угловая точка v с базисом $B=(A_{j_1},\ldots,A_{j_r})$ невырожденная, то ее симплекс-таблица $S\succ 0$, так как тогда (см. табл. 3) $\gamma_{i0}=v^{j_i}>0$ и, следовательно, $\Gamma_i\succ 0$ при всех $i=1,\ldots,r$. Если точка v вырожденная, то $\gamma_{i0}=0$ хотя бы для одного номера i, $1\leqslant i\leqslant r$, и первый отличный от нуля элемент в строке Γ_i может оказаться отрицательным и тогда $\Gamma_i\prec 0$. (см., например, строки Γ_2,Γ_3 таблицы 9). Впрочем, такой «недостаток» строки Γ_i легко исправить, если соответствующий базисный столбец x^{j_i} симплекс-таблицы переставить между столбцами V и x^1. Такая перестановка, равносильная перенумерации переменных, приведет к тому, что в строке Γ_i сразу после величины $\gamma_{i0}=0$ окажется величина $\gamma_{ij_i}=1$ и строка Γ_i станет $\succ 0$, а на лексикографической положительности или отрицательности других строк это не отразится, так как $\gamma_{sj_i}=0$ при всех $s\neq i$, $1\leqslant s\leqslant r$. Отсюда ясно, что последовательно переставляя указанным образом базисные столбцы x^{j_i} для всех строк $\Gamma_l\prec 0$, нетрудно добиться, чтобы симплекс-таблица стала лексикографически положительной. Так, например, в таблице 9 для этого достаточно переставить столбцы x^2, x^3 между столбцами V и x^1. После сказанного теперь неудивительно, что симплекс-таблица 1 лексикографически положительна. А вот симплекс-таблица 2 может потерять это свойство, если в строке Γ_s окажется $w^k=0$, $\overline{\gamma}_{ss}<0$.

Используя введенные лексикографические понятия, перейдем к описанию обещанного антициклина. Напомним, что применение симплекс-метода во всякой невырожденной задаче приводит к построению последовательности угловых точек $v_0,v_1,\ldots,v_p,\ldots$ со свойством (43). Так как $\Delta_{p0}=f(v_p)$, то согласно определению 4 свойство (43) будет означать, что соответствующие этим точкам симплекс-таблицы $S_0,S_1,\ldots,S_p,\ldots$ таковы, что

$$S_0\overset{\Delta}{\succ}S_1\overset{\Delta}{\succ}\ldots\overset{\Delta}{\succ}S_p\overset{\Delta}{\succ}\ldots \tag{47}$$

При написании цепочки лексикографических неравенств (47) мы учли транзитивность отношения $\overset{\Delta}{\succ}$ для симплекс-таблиц: если $S_1\overset{\Delta}{\succ}S_2$, $S_2\overset{\Delta}{\succ}S_3$, то

$S_1 \overset{\Delta}{\succ} S_3$. Так как симплекс-таблиц конечное число, а в (47) повторение таблиц невозможно, то еще раз убеждаемся в конечности симплекс-процесса в невырожденных задачах. А в вырожденных задачах последовательность $\{f(v_p)\}$ обладает, вообще говоря, лишь свойством (41), а свойство (47), как видно из примера 3 (см. табл. 9–15), может не выполняться. Возникает идея: нельзя ли как-то дополнить правило (34), (35) выбора разрешающего элемента так, чтобы и в вырожденных задачах получались последовательности симплекс-таблиц со свойством (47)? Реализация этой идеи приведет нас к антициклину.

Пусть v — какая-либо угловая точка множества X с базисом $B=(A_{j_1}, \ldots, A_{j_r})$ и с симплекс-таблицей $S=S(v, B)$, пусть это будет таблица 3. Предположим, что таблица S удовлетворяет условиям (34), и уже зафиксирован какой-либо номер $k \notin I(v)$, $k>0$, из (34). Выберем номер s и разрешающий элемент γ_{sk} из условия

$$\frac{\Gamma_s}{\gamma_{sk}} = \operatorname*{lex\,min}_{i \in I_k(v)} \frac{\Gamma_i}{\gamma_{ik}}, \quad s \in I_k(v) = \{i\colon 1 \leqslant i \leqslant r,\ \gamma_{ik} > 0\}. \tag{48}$$

С помощью леммы 1 убедимся, что условие (48) однозначно определяет номер s. Для этого нам надо показать, что множество $G=\{y_i=\frac{\Gamma_i}{\gamma_{ik}},\ i \in I_k(v)=M_0\}$ состоит из различных векторов. Допустим, что два вектора из этого множества оказались равными: $\Gamma_i/\gamma_{ik}=\Gamma_p/\gamma_{pk}$, $p, k \in I_k(v)$, $p \neq k$. Тогда $\Gamma_i=(\gamma_{ik}/\gamma_{pk})\Gamma_p$, т. е. строки Γ_i, Γ_p в матрице $\Gamma=(\gamma_0, \gamma_1, \ldots, \gamma_n)=(B^{-1}b, B^{-1}A_1, \ldots, B^{-1}A_n)=B^{-1}(b, A)$ пропорциональны. Однако по предположению множество X непусто, поэтому система (2), $Ax=b$, совместна, и тогда, согласно теореме Кронекера — Капелли [192; 353], $\operatorname{rang}(b, A)=\operatorname{rang}A=r$. Отсюда и из невырожденности матрицы B^{-1} следует, что $\operatorname{rang}\Gamma=\operatorname{rang}A=r$. Это значит, что строки $\Gamma_1, \ldots, \Gamma_r$ матрицы Γ образуют линейно независимую систему векторов, и никакие строки Γ_i, Γ_p в этой матрице пропорциональными не могут быть. Полученное противоречие показывает, что множество G состоит из различных векторов. Согласно лемме 1 условие (48) однозначно определяет номер $s \in I_k(v)$, причем для его практического нахождения можно воспользоваться конструкциями, указанными в этой лемме.

Важно заметить, что лексикографическое правило (48) выбора разрешающего элемента не отменяет ранее сформулированное правило (34), (35), а, наоборот, включает его в себя, дополняет и уточняет его. В самом деле, в соответствии с конструкциями леммы 1 для поиска номера s из условия (48) мы в первую очередь образуем множество $M_1=\{s \in M_0=I_k(v)\colon \gamma_{s0}/\gamma_{sk}=\min\limits_{i \in M_0}\gamma_{i0}/\gamma_{ik}\}$, которое в точности совпадает со множеством номеров s, определяемых условием (35). Отсюда, кстати, следует, что если множество M_1 состоит из единственного номера (так будет, например, в невырожденных задачах), то оба правила (34), (35) и (48) определяют один и тот же номер s, один и тот же разрешающий элемент γ_{sk}. Если же M_1 содержит более одного номера, то правило (48) устраняет возможную в вырожденных задачах неоднозначность в выборе разрешающего элемента при пользовании правилом (34), (35). Итак, выберем номера k, s и разрешающий элемент $\gamma_{sk}=\gamma_{sk}(v)$ из условий (34), (48) и по правилам (26) совершим переход от симплекс-таблицы $S(v, B)$ угловой точки v с базисом

B к симплекс-таблице $S(w, \overline{B})$ угловой точки w с координатами (36) и с базисом (37). Оказывается, если исходная симплекс-таблица $S(v, B) \overset{\Gamma}{\succ} 0$, то имеют место следующие лексикографические неравенства:

$$S(w, \overline{B}) \overset{\Gamma}{\succ} 0, \quad S(v, B) \overset{\Delta}{\succ} S(w, \overline{B}). \tag{49}$$

В самом деле, если $S(v, B) \overset{\Gamma}{\succ} 0$, то по определению 5 в этой симплекс-таблице строки $\Gamma_i=\Gamma_i(v) \succ 0$, $i=1, \ldots, r$. Тогда из (26), из неравенства $\gamma_{sk}>0$ и свойства 1) отношения $\succ$ следует, что $\Gamma_s(w)=\Gamma_s(v)/\gamma_{sk}(v) \succ 0$. Пусть теперь $i \neq s$. Тогда, возможно, либо $\gamma_{ik}>0$, либо $\gamma_{ik} \leqslant 0$. Если $\gamma_{ik}>0$, то $i \in I_k(v)$ и, согласно правилу (48), имеем $\Gamma_i(v)/\gamma_{ik}(v) \succ \Gamma_s(v)/\gamma_{sk}(v)$. В силу свойства 2) отношения $\succ$ тогда $\Gamma_i(v) \succ (\gamma_{ik}(v)/\gamma_{sk}(v))\Gamma_s(v)$, т. е. $\Gamma_i(w)=\Gamma_i(v)-(\gamma_{ik}(v)/\gamma_{sk}(v))\Gamma_s(v) \succ 0$. Если $\gamma_{ik} \leqslant 0$, то $\alpha=-(\gamma_{ik}(v)/\gamma_{sk}(v)) \geqslant 0$ и $\Gamma_i(w)=\Gamma_i(v)+\Gamma_s(w)(-\gamma_{ik}(v)/\gamma_{sk}(v)) \succ 0$ в силу свойства 3) отношения $\succ$. Таким образом, $\Gamma_i(w) \succ 0$ для всех $i=1, \ldots, r$, что означает, что симплекс-таблица $S(w, \overline{B}) \overset{\Gamma}{\succ} 0$. Наконец, из того, что $\Gamma_s(w) \succ 0$, $\Delta_k=\Delta_k(v)>0$, из формулы (26) для строки $\Delta(w)$ и свойства 4) отношения $\succ$ имеем: $\Delta(v) \succ \Delta(v)-\Delta_k(v)\Gamma_s(w)=\Delta(w)$. Это значит, что $S(v, B) \overset{\Delta}{\succ} S(w, \overline{B})$. Лексикографические неравенства (49) доказаны.

Теперь посмотрим, к какому симплекс-процессу приведет применение правил (34), (48). Пусть у нас имеется некоторая угловая точка v_0 с базисом B_0, с симплекс-таблицей $S_0 \overset{\Gamma}{\succ} 0$. Пользуясь правилами (34), (48), (26), организуем симплекс-процесс, начинающийся с точки v_0, и получим последовательности угловых точек $\{v_p\}$, их базисов $\{B_p\}$, симплекс-таблиц $\{S_p\}$, где S_p — симплекс-таблица точки v_p с базисом B_p. Оказывается, последовательность $\{S_p\}$ удовлетворяет лексикографическим неравенствам (47), в чем легко убедиться с помощью математической индукции, основываясь на неравенствах (49) и $S_0 \overset{\Gamma}{\succ} 0$. Так как в цепочке неравенств (47) повторение симплекс-таблиц невозможно в силу транзитивности отношения $\overset{\Delta}{\succ}$ и в задаче (1) множество симплекс-таблиц конечно, то такой симплекс-процесс закончится на каком-то шаге реализацией одного из условий (32) или (33). Это значит, что правило выбора разрешающего элемента по формулам (34), (48) является антициклином. Тем самым доказана

Теорема 2. *Пусть в канонической задаче* (1) *множество* $X \neq \varnothing$, $\operatorname{rang}A=r=m<n$, *пусть* v_0 — *какая-либо угловая точка этого множества с симплекс-таблицей* $S_0 \overset{\Gamma}{\succ} 0$. *Тогда симплекс-процесс, начинающийся с точки* v_0, *при выборе разрешающего элемента* γ_{sk} *из условий* (34), (48) *завершится за конечное число шагов определением некоторой угловой точки* v_p *множества* X, *в которой реализуются либо условия* (32), *либо* (33), *причем в случае* (32) v_p — *решение задачи* (1), $f(v_p)=f_*>-\infty$, *а в случае* (33) *задача* (1) *не имеет решения*, $f_*=-\infty$.

Напомним, что построенный антициклин (34), (48) обоснован при условии, что начальная симплекс-таблица $S_0 \overset{\Gamma}{\succ} 0$. Но это условие нельзя считать серьезным требованием к антициклину, так как выше мы показали, что переставив некоторые из базисных столбцов и соответствующим образом перенумеровав переменные, любую симплекс-таблицу S_0 легко сделать лексикографически положительной. К тому же такую операцию с перестановкой

столбцов и перенумерацией переменных нужно сделать самое большее один раз в самом начале симплекс-процесса. Впрочем, операцию с перестановкой столбцов и перенумерацией переменных можно явно и не делать, если эту операцию учесть при порядке формирования множеств $M_1, M_2, \ldots$, указанных в лемме 1 и используемых при поиске номера s из условия (48). Более того, можно доказать, что условия (34), (48) являются антициклином и без требования $S_0 \overset{\Gamma}{\succ} 0$ [148; 259].

Заметим также, что антициклин (34), (48) оставляет некоторый произвол в организации симплекс-процесса из-за того, что условие (34) определяет номер k, вообще говоря, неоднозначно; для устранения указанной неоднозначности к правилу (34), (48) можно сделать дополнение, руководствуясь какими-либо другими соображениями, например, можно выбирать минимальный или максимальный номер k, удовлетворяющий условиям (34).

Пример 4. Для иллюстрации изложенного антициклина (34), (48) рассмотрим задачу (45), (46), в которой, как обнаружилось выше, использование правила (34), (35) выбора разрешающего элемента может привести к зацикливанию. Сначала симплекс-таблицу 9 начальной угловой точки $v_0 = (0, 0, 0, 0, 0, 0, 1)$ сделаем лексикографически положительной, переставив базисные столбцы x^2, x^3 между столбцами V и x^1; в результате придем к таблице 16, в которой сохранена первоначальная нумерация переменных.

Таблица 16

	Б	V	x^2	x^3	x^1	x^4	x^5	x^6	x^7
Γ_1	x^7	1	0	0	1	1	1	1	1
Γ_2	x^2	0	1	0	−2	1	−3	4	0
Γ_3	x^3	0	0	1	−3	(4)	−2	1	0
Δ		0	0	0	−1	1	1	−1	0

Таблица 17

x^7	1	0	−1/4	7/4	0	(3/2)	3/4	1
x^2	0	1	−1/4	−5/4	0	−5/2	15/4	0
x^4	0	0	1/4	−3/4	1	−1/2	1/4	0
	0	0	−1/4	−1/4	0	3/2	−5/4	0

Таблица 18

x^5	2/3	0	−1/6	7/6	0	1	1/2	2/3
x^2	5/3	1	−2/3	5/3	0	0	5	5/3
x^4	1/3	0	1/6	−1/6	1	0	1/2	1/3
	−1	0	0	−1	0	0	−2	−1

В строке Δ таблицы 16 величина $\Delta_4 = 1 > 0$ и весь столбец x^4 заполнен положительными числами: $\gamma_{14} = 1$, $\gamma_{24} = 1$, $\gamma_{34} = 4$. Таким образом получаем, что условия (34) здесь выполнены и $I_4(v_0) = \{1, 2, 3\}$. Для применения

лексикографического правила (48) выпишем следующие строки:

$$\frac{\Gamma_1}{\gamma_{14}} = (1, 0, 0, 1, 1, 1, 1, 1), \quad \frac{\Gamma_2}{\gamma_{24}} = (0, 1, 0, -2, 1, -3, 4, 0),$$

$$\frac{\Gamma_3}{\gamma_{34}} = (0, 0, \tfrac{1}{4}, -\tfrac{3}{4}, 1, -\tfrac{1}{2}, \tfrac{1}{4}, 0).$$

Последовательно сравнивая по величине первые, вторые и т. д. координаты этих векторов-строк Γ_i/γ_{ik}, $i \in M_0 = I_4(v_0)$, легко находим указанные в лемме 1 множества $M_1 = \{2, 3\}$, $M_3 = \{3\}$, искомый номер $s = 3$, так что здесь $\operatorname{lex}\min\{\Gamma_1/\gamma_{14}, \Gamma_2/\gamma_{24}, \Gamma_3/\gamma_{34}\} = \Gamma_3/\gamma_{34}$. Понятно, что те же множества M_1, M_2 и номер $s = 3$ можно было получить непосредственно из таблицы 9, просматривая ее столбцы в таком порядке: V, x^2, x^3, x^1, x^4, x^5, x^6, x^7. Таким образом, с помощью правила (48) мы однозначно нашли разрешающий элемент $\gamma_{34} = 4$. Далее, по формулам (26) в базис вводим переменную x^4 и выводим из базиса x^3. В результате придем к симплекс-таблице 17 угловой точки v_1, совпадающей с v_0, но имеющей другой базис (A_7, A_2, A_4). В этой таблице разрешающим элементом может быть лишь величина $\gamma_{15} = 3/2$. Из базиса выведем переменную x^7, заменив ее x^5, по формулам (26) получим табл. 18. В строке Δ все величины Δ_i, $1 \leqslant i \leqslant 7$, неположительны, реализовались условия (32). Симплекс-процесс на этом заканчивается. Выясняется, что невырожденная угловая точка $v_3 = (0, 5/3, 0, 1/3, 2/3, 0, 0)$ является решением задачи (45), (46), $f(v_3) = f_* = -1$.

Заметим, что хотя среди задач линейного программирования вырожденные задачи встречаются довольно часто, но тем не менее на основе всего практического опыта применения симплекс-метода к таким задачам сложилось убеждение, что вероятность получения бесконечных симплекс-процессов ничтожно мала. Добавим также, что использование антициклина на каждом шаге симплекс-метода может привести к заметному увеличению машинного времени ЭВМ, требующегося для решения задачи. Поэтому на практике чаще всего пользуются упрощенным правилом выбора номеров k и s из условий (34), (35), беря, например, наименьшие или наибольшие номера, удовлетворяющие этим условиям.

Из сказанного следует, что наличие симплекс-метода, снабженного антициклином, для практики, видимо, не является слишком актуальным, но в теоретическом плане это принципиально важно и ставит симплекс-метод на надежный математический фундамент.

5. Кратко остановимся на применении симплекс-метода для решения канонической задачи максимизации:

$$f(x) = \langle c, x\rangle \to \sup, \qquad x \in X = \{x \in E^n\colon x \geqslant 0,\ Ax = b\}, \qquad (50)$$

где A — ненулевая матрица размера $r \times n$, $c \in E^n$, $b \in E^m$, $r = \operatorname{rang} A < n$. Конечно, эту задачу можно свести к равносильной задаче минимизации $g(x) = -f(x) \to \inf$, $x \in X$, и к ней применить описанный выше симплекс-метод без каких-либо изменений. В то же время нетрудно несколько видоизменить симплекс-метод и приспособить его для непосредственного применения к задаче (50). Легко понять, посмотрев на формулы (6),(9) и (29), что при решении задачи максимизации (50) нас прежде всего будут интересовать величины $\Delta_k < 0$, и мы естественно придем к рассмотрению следующих трех случаев, аналогичных случаям (32)–(34).

Случай I. В нижней строке симплекс-таблицы 3 все $\Delta_1, \ldots, \Delta_n$ неотрицательны. Исходная угловая точка v является решением задачи (50).

Случай II. В нижней строке симплекс-таблицы 3 найдется величина $\Delta_k < 0$, $1 \leqslant k \leqslant n$, и находящийся над ней столбец γ_k неположителен. Тогда $f^* = \sup\limits_{x \in X} \langle c, x\rangle = +\infty$, задача (50) не имеет решения.

Случай III. В нижней строке симплекс-таблицы 3 имеются величины $\Delta_k < 0$, $1 \leqslant k \leqslant n$, причем в каждом столбце над величиной $\Delta_k < 0$ найдется хотя бы одно число $\gamma_{ik} > 0$. Тогда фиксируем один из таких номеров k с $\Delta_k < 0$ и выбираем разрешающий элемент γ_{sk} по правилу (35) или, точнее, (48), а затем по формулам (36) совершаем переход от угловой точки v с базисом B к точке w, которая согласно замечанию 1 также будет угловой точкой множества X с базисом $\overline{B}$, имеющим вид (37), причем $f(w) \geqslant f(v)$. Один шаг симплекс-метода для задачи (50) описан. Как и выше, можем считать, что исходная симплекс-таблица $S(v, B) \overset{\Gamma}{\succ} 0$. Тогда будут справедливы следующие лексикографические неравенства, аналогичные неравенствам (49): $S(w, \overline{B}) \overset{\Gamma}{\succ} 0$, $S(v, B) \overset{\Delta}{\prec} S(w, \overline{B})$. Отсюда следует, что симплекс-процесс для задачи (50) будет конечным и закончится реализацией одного из случаев I или II.

Все высказанные здесь утверждения, касающиеся задачи (50), доказываются совершенно также, как и аналогичные утверждения, касающиеся задачи (1). Предлагаем читателю убедиться в этом самостоятельно.

Остается напомнить, что в течение всего этого параграфа задача (1) (а также задача (50)) рассматривалась в предположении, что $m = r = \operatorname{rang} A < n$, и нам известна хотя бы одна угловая точка множества X. В следующем параграфе покажем, как избавиться от этих ограничений.

Упражнения

1. С помощью симплекс-метода решить следующие канонические задачи:

а) $f(x) = x^1 + x^2 + x^3 + x^4 + x^5 \to \inf$; $x = (x^1, x^2, \ldots, x^5) \geqslant 0$, $x^1 + x^2 + 2x^3 + x^4 - 2x^5 = 2$, $2x^1 - x^2 + x^3 + 2x^4 + x^5 = 1$;

б) $f(x) = x^1 + 3x^2 + 2x^3 + x^4 - 3x^5 \to \inf$; $x = (x^1, \ldots, x^5) \geqslant 0$, $x^1 + 2x^2 - 42x^3 - x^5 = 0$, $-x^2 + x^4 - 2x^5 = 3$;

в) $f(x) = x^1 - 2x^2 + x^3 - x^4 \to \inf$; $x = (x^1, \ldots, x^4) \geqslant 0$, $x^1 - 2x^3 + x^4 = 1$, $x^2 + x^3 - 2x^4 = 1$;

г) $f(x) = x^1 + x^2 + x^3 + x^4 + x^5 \to \inf$; $x = (x^1, \ldots, x^6) \geqslant 0$, $x^1 + 2x^2 - x^4 + 2x^5 = 1$, $x^2 + x^3 + 2x^4 - x^5 = 0$, $-x^2 - 2x^4 + x^5 + x^6 = 1$.

2. Проверить, что точка v_0 является угловой, найти ее базис, приведенную систему и, взяв v_0 в качестве начальной точки, с помощью симплекс-метода решить следующие задачи:

а) $f(x) = x^1 + 2x^2 + x^4 + x^6 \to \inf[\sup]$; $x = (x^1, \ldots, x^6) \geqslant 0$, $x^1 + 2x^2 - x^3 + 2x^4 - x^5 + 2x^6 = 0$, $x^1 - 4x^2 + 2x^3 + 2x^4 - 4x^6 = 1$, $2x^1 - 2x^2 + x^3 + 4x^4 + x^5 - 2x^6 = 3$; $v_0 = (1, 0, 0, 0, 1, 0)$;

б) $f(x) = x^1 + x^2 + x^3 + 2x^4 + 3x^5 + 2x^6 \to \inf[\sup]$; $x = (x^1, \ldots, x^6) \geqslant 0$, $x^1 + x^2 + 4x^4 + 2x^6 = 5$, $x^2 + x^3 + x^4 - x^5 + 2x^6 = 0$, $x^1 + x^3 + x^4 - x^5 + 2x^6 = -1$; $v_0 = (0, 1, 0, 1, 2, 0)$;

в) $f(x) = x^1 + 2x^2 + 3x^3 + x^4 + x^5 \to \inf[\sup]$; $x = (x^1, \ldots, x^5) \geqslant 0$, $x^1 + 2x^2 + 4x^3 - x^4 + x^5 = 1$, $x^1 - x^2 - 2x^3 + x^4 + x^5 = 1$, $-x^1 + x^2 - 6x^3 + x^4 + x^5 = 1$; $v_0 = (1, 1, 0, 1, 0)$.

3. Найти начальную угловую точку v_0, ее базис, приведенную систему и решить следующие задачи:

а) $f(x) = -x^1 + 3x^2 + 5x^3 + x^4 \to \inf[\sup]$; $x = (x^1, \ldots, x^4) \geqslant 0$, $x^1 + 4x^2 + 4x^3 + x^4 = 5$, $x^1 + 7x^2 + 8x^3 + 2x^4 = 9$;

б) $f(x) = x^2 - x^3 + x^5 \to \inf[\sup]$; $x = (x^1, \ldots, x^5) \geqslant 0$, $x^1 + x^4 - x^5 = 1$, $x^1 + x^2 + 2x^4 = 3$; $x^3 + x^4 = 1$;

в) $f(x) = 4x^1 - x^2 - 3x^3 - 10x^4 \to \inf[\sup]$; $x = (x^1, \ldots, x^4) \geqslant 0$, $x^1 + x^2 - x^3 + x^4 = 0$, $14x^1 + x^2 + 10x^3 - 10x^4 = 11$.

4. С помощью приемов, описанных в § 1, записать задачи линейного программирования в каноническом виде и решить их с помощью симплекс-метода:

а) $f(x) = -x^1 + 4x^2 - 5x^3 \to \inf[\sup]$; $x = (x^1, x^2, x^3) \geqslant 0$, $2x^1 + x^2 + x^3 = 4$ [$\leqslant 4$; $\geqslant 4$], $x^1 - x^2 - x^3 \leqslant 2$ [$\leqslant 2$; $= 2$];

б) $f(x) = x^1 + x^2 + x^3 \to \inf[\sup]$; $x = (x^1, x^2, x^3) \geqslant 0$, $-1 \leqslant x^1 + x^2 + x^3 \leqslant 1$, $-x^1 + x^2 + x^3 \leqslant 1$, $x^1 - x^2 + x^3 \leqslant 1$;

в) $f(x) = x^1 + x^2 + x^3 + x^4 \to \inf[\sup]$; $x = (x^1, \ldots, x^5) \geqslant 0$, $x^1 - x^2 \geqslant 0$, $x^1 + x^2 - x^3 + x^4 - x^5 \geqslant 1$.

5. При каких значениях параметра a задача $f(x) = x^1 + 2x^2 + 3x^3 + 4x^4 \to \inf[\sup]$; $x = (x^1, \ldots, x^4) \geqslant 0$, $x^1 + x^2 + x^3 + ax^4 \leqslant 1$, [$\geqslant 1$; $= 1$] имеет решение? Не имеет решения? Единственное решение?

6. С помощью симплекс-метода с антициклином (48) решить задачу [216]:

$$f(x) = -\frac{3}{4}x^1 + 150x^2 - \frac{1}{50}x^3 + 6x^4 \to \inf; \quad x = (x^1, \ldots, x^7) \geqslant 0,$$

$$\frac{1}{4}x^1 - 60x^2 - \frac{1}{25}x^3 + 9x^4 + x^5 = 0$$

$$\frac{1}{2}x^1 - 90x^2 - \frac{1}{50}x^3 + 3x^4 + x^6 = 0$$

$$x^3 + x^7 = 1,$$

взяв в качестве начальной угловую точку $v_0 = (0, 0, 0, 0, 0, 0, 1)$ с базисом $B_0 = (A_5, A_6, A_7)$. Показать, что если разрешающий элемент γ_{sk} выбирать по правилу (34), (35) так, что k — наименьший из номеров, для которых $\Delta_k = \max\limits_{1 \leqslant i \leqslant 7} \Delta_i$, s — наименьший из номеров, удовлетворяющих условию (35), то придем к циклическому перебору базисов точки v_0 по схеме: $(A_5, A_6, A_7) \to (A_1, A_6, A_7) \to (A_1, A_2, A_7) \to (A_3, A_2, A_7) \to (A_3, A_4, A_7) \to (A_5, A_4, A_7) \to$ $\to (A_5, A_6, A_7) \to \ldots$

7. Пусть v — угловая точка множества $X = \{x \geqslant 0\colon Ax = b\}$ с базисом $B = (A_{j_1}, \ldots, A_{j_r})$, где A — матрица размера $m \times n$, $r = \operatorname{rang} A = m < n$. Пусть некоторая система уравнений

$$\mu^{j_i} = x^{j_i} + \sum_{k \notin I(v)} \mu_{ik} x^k, \quad i = 1, \ldots, r; \quad I(v) = \{j_1, \ldots, j_r\} \tag{51}$$

равносильна системе $Ax = b$. Доказать, что система (51) является приведенной системой для точки v с базисом B. Указание: сначала покажите, что системы (51) и (28) равносильны, затем подставьте в (51) следующие решения системы (28): $x_0 = v$; $x_p = (x_p^1, \ldots, x_p^n)$, $p \notin I(v)$, где $x_p^{j_i} = v^{j_i} - \gamma_{ip}$, $i = 1, \ldots, r$, $x_p^p = 1$, $x_p^j = 0$, $j \notin I(v)$, $j \neq p$, и убедитесь, что $\mu^{j_i} = v^{j_i}$, $\mu_{ik} = \gamma_{ik}$, $k \notin I(v)$, $i = 1, \ldots, r$.

8. Пусть задача (1) имеет своим решением угловую точку v с симплекс-таблицей 3. Можно ли утверждать, что тогда $\Delta_j \leqslant 0$ для всех $j = 1, 2, \ldots, n$? (см. табл. 3, 4).

9. Для того, чтобы задача (1) имела своим решением угловую точку v, необходимо и достаточно, чтобы для какого-либо базиса B точки v соответствующая симплекс-таблица удовлетворяла условиям (32). Доказать.

10. Пусть угловая точка v с базисом $B = (A_{j_1}, \ldots, A_{j_r})$ является решением задачи (1) и пусть ее симплекс-таблица (см. табл. 3) удовлетворяет условиям (32). Пусть при некотором $k \notin I(v) = \{j_1, \ldots, j_r\}$ величина $\Delta_k = 0$ и в столбце x^k имеется величина $\gamma_{ik} > 0$. Пусть разрешающий элемент γ_{sk} определен по правилам (35) или (48), и получена угловая точка w с координатами (36) и с базисом (37) (см. замечание 2). Доказать, что тогда w — решение задачи (1). Можно ли таким способом получить все угловые точки множества X, являющиеся решением задачи (1)?

11. Пусть v — угловая точка множества $X = \{x \geqslant 0\colon Ax = b\}$, $B = (A_{j_1}, \ldots, A_{j_r})$ — ее базис, A — матрица размера $r \times n$, $r = \operatorname{rang} A < n$; пусть в матрице $\Gamma = (\gamma_0, \gamma_1, \ldots, \gamma_n)$ (см. обозначения (27)) один из столбцов $\gamma_k \leqslant 0$, $k \notin I(v) = \{j_1, \ldots, j_r\}$. Доказать, что тогда множество X неограничено, и найдется такой вектор $c = (c_1, \ldots, c_n)$, для которого $\inf\limits_{x \in X} \langle c, x\rangle = -\infty$. Указание: взять $c_i = 0$, $i \in I(v)$, $c_k = -1$, остальные c_i — произвольные числа, составить симплекс-таблицу точки v для задачи (1) с выбранным c. Можно ли утверждать обратное: если множество X неограничено, то в матрице Γ найдется столбец $\gamma_k \leqslant 0$, $k \notin I(v)$?

§ 4. Поиск начальной угловой точки

1. Будем рассматривать каноническую задачу

$$f(x)=\langle c, x\rangle \to \inf, \quad x\in X=\{x=(x^1,\ldots,x^n)\geqslant 0,\ Ax=b\}, \qquad (1)$$

где A — матрица размера $m\times n$, $c\in E^n$, $b\in E^m$, в самом общем виде, отказавшись от ограничительных требований $\operatorname{rang}A=m<n$, принятых в предыдущем параграфе. Можем считать, что $A\neq 0$, так как при $A=0$ либо $X=E^n_+=\{x\geqslant 0\}$ (случай $b=0$), либо $X=\varnothing$ ($b\neq 0$), и задача (1) становится малосодержательной. Нас будут интересовать вопросы, как узнать, не пусто ли множество X, и если оно непусто, то имеет ли оно хотя бы одну угловую точку? Как найти эту угловую точку, ее базис, ее симплекс-таблицу? Ниже будут даны ответы на все эти вопросы. Здесь замечательно то, что для ответа на них может быть использован изложенный выше симплекс-метод с антициклином. Оказывается, по задаче (1) легко можно составить новую вспомогательную каноническую задачу, к которой очень удобно применять симплекс-метод, и в зависимости от того, чем закончится симплекс-процесс для вспомогательной задачи, можно будет сказать, пусто или непусто множество X, найти ее угловую точку. Этот метод поиска начальной угловой точки в литературе часто называют *методом искусственного базиса*.

Можем считать, что в (1) вектор $b\geqslant 0$, так как если $b^i<0$ при некотором i, $1\leqslant i\leqslant m$, то соответствующее i-е уравнение $(Ax)^i=b^i$ системы $Ax=b$ можно умножить на (-1). Наряду с основными переменными $x=(x^1,\ldots,x^n)$ введем вспомогательные (искусственные) переменные $u=(u^1,\ldots,u^m)$ и в пространстве E^{n+m} переменных $y=(u,x)=(u^1,\ldots,u^m,x^1,\ldots,x^n)$ рассмотрим следующую каноническую задачу линейного программирования:

$$\begin{gathered} g(y)=u^1+u^2+\ldots+u^m=\langle \mathbf{1}_m, u\rangle\to\inf; \quad y\in Y,\\ Y=\{y=[u,x]\in E^{n+m}:\ y\geqslant 0,\ Cy=u+Ax=b\}, \end{gathered} \qquad (2)$$

где $\mathbf{1}_m=(1,\ldots,1)\in E^m$, $C=(I_m,A)$, I_m — единичная матрица размера $m\times m$. Систему $Cy=u+Ax=b$ перепишем в покоординатной форме:

$$\begin{aligned} b^1&=u^1 &&+a_{11}x^1+\ldots+a_{1j}x^j+\ldots+a_{1n}x^n,\\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\ b^i&=u^i &&+a_{i1}x^1+\ldots+a_{ij}x^j+\ldots+a_{in}x^n,\\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\ b^m&=u^m+a_{m1}x^1+\ldots+a_{mj}x^j+\ldots+a_{mn}x^n. \end{aligned}$$

При $u^1=\ldots=u^m=0$ система (3) превращается в систему $Ax=b$. Множество Y непусто: оно содержит, например, точку $z_0=(b,0)\geqslant 0$. Более того, с помощью теоремы 2.1 легко убедиться, что точка z_0 является угловой точкой множества Y с базисом $(e_1,\ldots,e_m)=I_m$. Система (3) является приведенной системой этой точки, $\operatorname{rang}C=m$. Для целевой функции приведенная форма равна $g(y)=\langle \mathbf{1}_m, b-Ax\rangle=g(z_0)-\langle A^\top\mathbf{1}_m, x\rangle$, где $A^\top$ — транспонированная матрица A. Теперь симплекс-таблица точки z_0 составляется просто (см. табл. 19): в столбце Б — базисные координаты $u^1,\ldots,u^m$, в столбце $V-\gamma_0=b$, в столбцах $u^1,\ldots,u^m$ вспомогательных переменных $\gamma'_1=e_1,\ldots,\gamma'_m=e_m$, в столбцах $x^1,\ldots,x^n$ основных переменных

$\gamma_1=I_mA_1=A_1,\ldots,\gamma_n=I_mA_n=A_n$. Элементы строки Δ легко вычисляются по формулам (3.30), которые применительно к задаче (4) дадут: величины $\Delta_1,\ldots,\Delta_m$, соответствующие базисным переменным $u^1,\ldots,u^m$ равны нулю, а величины $\Delta_0,\Delta_1,\ldots,\Delta_n$ равны сумме элементов соответствующих столбцов $V, x^1,\ldots,x^n$: $\Delta_0=\sum_{i=1}^m b^i=g(z_0)$, $\Delta_j=\sum_{i=1}^m a_{ij}$. Привлекательно также и то, что симплекс-таблица 19 лексикографически положительна и поэтому здесь удобно применять симплекс-метод с антициклином (3.48).

Таблица 19

	Б	V	u^1	…	u^i	…	u^m	x^1	…	x^j	…	x^n
Γ_1	u^1	b^1	1	…	0	…	0	a_{11}	…	a_{1j}	…	a_{1n}
…	…	…	…	…	…	…	…	…	…	…	…	…
Γ_i	u^i	b^i	0	…	1	…	0	a_{i1}	…	a_{ij}	…	a_{in}
…	…	…	…	…	…	…	…	…	…	…	…	…
Γ_m	u^m	b^m	0	…	0	…	1	a_{m1}	…	a_{mj}	…	a_{mn}
Δ		Δ_0	0	…	0	…	0	Δ_1	…	Δ_j	…	Δ_n

Поскольку $g(y)\geqslant 0$ при всех $y\in Y$, то $g_*=\inf\limits_Y g(y)\geqslant 0$ и случай $g_*=-\infty$ здесь невозможен. Поэтому, взяв в качестве начальной точку z_0, с помощью симплекс-метода с антициклином за конечное число шагов найдем угловую точку $z_*=(u_*,v_*)$ множества Y, являющуюся решением задачи (2): $g(z_*)=g_*\geqslant 0$. Имеются две возможности: или $g(z_*)>0$ или $g(z_*)=0$. Если $g(z_*)=u^1_*+\ldots+u^m_*>0$, то $u_*\neq 0$ и, оказывается, множество X в (1) будет пустым. В самом деле, если бы существовала хотя бы одна точка $x_0\in X$, то точка $y_0=(0,x_0)$ принадлежала бы множеству Y и, кроме того, тогда $g(y_0)=0$, что противоречит неравенствам $g(y_0)\geqslant g(z_*)=g_*>0$. Таким образом, при $g(z_*)=g_*>0$ множество X пусто, и задача (1) не имеет смысла.

Пусть теперь $g(z_*)=u^1_*+\ldots+u^m_*=0$. Тогда $u_*=0$ и $z_*=(0,v_*)$. Кроме того, по построению $z_*=(0,v_*)$ угловая точка множества X. Покажем, что тогда v_* — угловая точка множества X. Прежде всего ясно, что из $z_*\geqslant 0$ следует $v_*\geqslant 0$, а из $Cz_*=b$ имеем $Av_*=b$. Это значит, что $v_*\in X$. Далее, рассмотрим представление

$$v_*=\alpha x_1+(1-\alpha)x_2, \quad 0<\alpha<1; \quad x_1,x_2\in X, \qquad (4)$$

и покажем, что оно возможно лишь при $v_*=x_1=x_2$. Точки $y_1=(0,x_1)$, $y_2=(0,x_2)$, очевидно, принадлежат Y. Тогда (4) можно переписать в виде $z_*=\alpha y_1+(1-\alpha)y_2$, $0<\alpha<1$. Но z_* — угловая точка множества Y. Поэтому последнее равенство для z_* возможно лишь при $z_*=y_1=y_2$. Отсюда следует, что $v_*=x_1=x_2$. Таким образом, v_* — угловая точка множества X. Тем самым доказана важная

Теорема 1. *Если множество $X=\{x\in E^n:\ x\geqslant 0,\ Ax=b\}$ непусто, то оно имеет хотя бы одну угловую точку.*

Таким образом, с помощью изложенного метода искусственного базиса, составной частью которого является симплекс-метод с антициклином, мы получили возможность узнать, пусто множество X задачи (1) или непусто, а в случае непустого X определили одну из его угловых точек v_*. Найденная точка v_* вполне может быть использована как начальная угловая точка для

организации симплекс-процесса для исходной канонической задачи (1). Зная координаты точки v_*, с помощью теоремы 2.1 можно определить номера базисных переменных и ранг матрицы A. В самом деле, положительные координаты $(v_*^{j_1}, \ldots, v_*^{j_k})$ точки v_* заведомо являются базисными (см. доказательство теоремы 2.1). Остается добавить к линейно независимым столбцам $A_{j_1}, \ldots, A_{j_k}$ матрицы A другие ее столбцы и получить базис $(A_{j_1}, \ldots, A_{j_r}) = B$ линейной оболочки векторов, натянутых на столбцы $A_1, \ldots, A_n$ матрицы A. Номера $j_1, \ldots, j_r$ будут базисными для точки v_* и $r = \operatorname{rang} A$. Далее, с помощью процесса Гаусса — Жордана можем выразить базисные переменные через небазисные и, попутно исключая линейно зависимые уравнения из системы $Ax = b$, придем к приведенной системе угловой точки v_* с базисом B. Остается применить симплекс-метод с антициклином и получить решение задачи (1) или узнать, что эта задача не имеет решения.

2. Таким образом, доказана принципиальная возможность использования симплекс-метода, оснащенного антициклином, для решения произвольной канонической задачи. Более того, с помощью симплекс-метода мы доказали важную для теории и методов линейного программирования теорему 1. Приведем еще две теоремы, касающиеся канонической задачи, в доказательстве которых симплекс-метод также играет существенную роль.

Теорема 2. *Если задача* (1) *разрешима, то среди ее решений найдется хотя бы одна угловая точка множества* X.

Доказательство. По условию теоремы $X \neq \varnothing$ и существует точка $v_* \in X$ такая, что $\langle c, v_* \rangle = f_* > -\infty$. По теореме 1 тогда множество X имеет хотя бы одну угловую точку. Отправляясь от одной из этих угловых точек, с помощью симплекс-метода с антициклином за конечное число шагов придем к угловой точке x_*, являющейся решением задачи (1)–(3). Теорема 2 доказана. □

Теорема 3. *Для того, чтобы каноническая задача* (1) *была разрешима, т. е. существовала точка* $x_* \in X$ *такая, что* $\langle c, x_* \rangle = \inf\limits_X \langle c, x \rangle = f_* > -\infty$, *необходимо и достаточно, чтобы:*

1) *множество* X *было непустым;*

2) *функция* $f(x) = \langle c, x \rangle$ *была ограничена снизу на* X.

Доказательство. Необходимость очевидна. Докажем достаточность. Из того, что $X \neq \varnothing$, по теореме 1 следует существование угловой точки множества X. Принимая эту точку за начальную, будем решать задачу (1) с помощью симплекс-метода, снабженного антициклином. Так как по условию $f_* > -\infty$, то случай (3.33) здесь невозможен, и симплекс-процесс завершится за конечное число шагов реализацией случая (3.32) и отысканием точки x_*, являющейся решением задачи (1). Теорема 3 доказана. □

Применение симплекс-метода для доказательства других важных теорем теории линейного программирования изложим в следующем параграфе.

3. На этом заканчиваем изложение симплекс-метода для канонической задачи (1). Учитывая возможность сведения общей задачи линейного программирования к канонической задаче (теорема 1.1), можно сказать, что симплекс-метод является универсальным методом решения задач линейного программирования. Конечно, компьютерная реализация описанной выше схемы симплекс-метода требует огромной дополнительной работы: надо выбрать подходящую модификацию метода, изучить влияние погрешности

на симплекс-процесс, организовать хранение исходной и текущей информации о задаче и т. п. — эти практические проблемы обсуждаются, например, в [116; 516; 586; 620].

Симплекс-метод относится к так называемым конечным методам, позволяющим найти решение задачи линейного программирования или обнаружить ее нерешаемость за конечное число арифметических действий. Это число, конечно, зависит от размерностей m, n задачи (1). Известен пример задачи линейного программирования с n переменными и $m = 2n$ ограничениями (этот пример приведен в [586], стр. 360), для решения которого требуется не менее $2^n - 1$ шагов симплекс-метода, и, следовательно, число арифметических операций, необходимых для получения решения, не меньше 2^n. Отсюда следует, что количество вычислений для решения «плохих» задач линейного программирования симплекс-методом оценивается экспоненциальной функцией параметров m, n размерности задачи, и уже при не очень больших m, n решение таких задач симплекс-методом невозможно за обозримое время даже на самых мощных компьютерах. Как принято говорить, на классе задач линейного программирования симплекс-метод имеет *экспоненциальную сложность*. Однако вопреки такому пессимистическому выводу в практических задачах симплекс-метод показывает высокую эффективность, причем в абсолютном большинстве реальных задач количество необходимых арифметических операций имеет порядок $n^2 m$ [52]. Причина этого удивительного явления пока еще не выяснена. В последнее время появились методы, имеющие *полиномиальную сложность*. Так называются конечные методы, для которых число элементарных операций, необходимых для получения решения задачи линейного программирования с нужной точностью, не превышает некоторого полинома от размерностей m, n задачи — более точные формулировки см. в [525; 676; 736]. Эти методы в самом деле эффективнее симплекс-метода на «плохих» искусственно придуманных задачах линейного программирования, но на реальных задачах пока не могут успешно конкурировать с ним. На практике симплекс-метод по-прежнему остается основным методом линейного программирования.

Кроме симплекс-метода имеется множество других (конечных, итерационных) методов решения задач линейного программирования [52; 76; 77; 116; 203; 259; 586; 620; 685; 719; 775; 776]. Для специальных классов задач линейного программирования таких, как, например, транспортная задача, существуют методы, лучше учитывающие конкретные особенности этих задач [52; 203; 232; 259; 471; 493; 620; 685; 725; 743]. Содержательный обзор многих существующих методов линейного программирования дан в [586].

В заключении подчеркнем, что всюду выше предполагалось, что исходные данные задачи линейного программирования — матрица A, векторы b, c — известны точно и, кроме того, все промежуточные вычисления в симплекс-методе проводятся без погрешностей. Такая идеализация позволила нам дать строгое обоснование симплекс-метода, доказать ряд важных теорем линейного программирования. Однако на практике исходные данные задаются, как правило, неточно, промежуточные вычисления проводятся с округлениями. Поэтому применение симплекс-метода или других методов в конкретных задачах линейного программирования может привести к большим погрешностям, неверным выводам из-за возможной неустойчивости решения по отношению к возмущениям исходных данных таких задач, и для получения их решения с нужной точностью могут понадобиться специальные методы регуляризации, которые будут рассмотрены в главе 9.

Упражнения

1. С помощью метода искусственного базиса найдите какую-нибудь угловую точку следующих множеств:

а) $x=(x^1,\ldots,x^4)\geqslant 0,\ x^1+x^2+x^3-x^4=2,\ x^1+x^3+2x^4=1,\ x^1+x^2+x^4=2;$

б) $x=(x^1,\ldots,x^4)\geqslant 0,\ x^1+x^2+2x^3-x^4=1,\ x^1-x^2+x^3+x^4=2,\ x^1+3x^2+3x^3-3x^4=0;$

в) $x=(x^1,\ldots,x^4)\geqslant 0,\ x^1+x^2+x^3+x^4=1,\ x^1+2x^2-2x^3+x^4=0,\ 2x^1+3x^2-x^3+2x^4=2;$

г) $x=(x^1,\ldots,x^5)\geqslant 0,\ x^1+2x^2+2x^3-x^4+x^5=1,\ 2x^1-x^2+x^3+x^4+2x^5=2,\ 5x^2+3x^3-3x^4=0,\ x^1-3x^2-x^3+2x^4+x^5=1;$

д) $x=(x^1,\ldots,x^5)\geqslant 0,\ x^1+2x^2+x^3+x^4+2x^5=5,\ x^1-3x^3-2x^4-x^5=2,\ 2x^1+x^2-x^3+x^4+x^5=1;$

е) $x=(x^1,\ldots,x^6)\geqslant 0,\ x^1+x^2+x^3+x^4+x^5+x^6=2,\ x^1+2x^2+x^3+2x^4-x^5+x^6=3,\ x^1+x^2+3x^3+x^4+x^5-x^6=2,\ x^1+2x^2+x^3+4x^4-x^5-x^6=3.$

2. С помощью симплекс-метода решите следующие канонические задачи:

а) $f(x)=x^1-x^2-x^3-x^4+2x^5\to\inf[\sup],\ x=(x^1,\ldots,x^5)\geqslant 0,\ x^1+3x^2+x^3+x^4-2x^5=10,\ 2x^1+6x^2+x^3+3x^4-4x^5=20,\ 3x^1+10x^2+x^3+6x^4-7x^5=30;$

б) $f(x)=x^1+2x^2+x^3+2x^4+x^5\to\inf[\sup],\ x=(x^1,\ldots,x^6)\geqslant 0,\ x^1-x^2+2x^3+x^4-3x^5-x^6=3,\ x^1+x^3+2x^4-x^5+2x^6=2,\ 2x^1+x^2+x^3-x^4+2x^5+x^6=3;$

в) $f(x)=x^1+2x^2-2x^3+5x^4\to\inf[\sup],\ x=(x^1,\ldots,x^4)\geqslant 0,\ x^1+2x^2-x^3-x^4=1,\ -x^1+2x^2+3x^3+x^4=2,\ x^1+5x^2+x^3-x^4=5;$

г) $f(x)=x^1+2x^2+3x^3+4x^4+5x^5\to\inf[\sup],\ x=(x^1,\ldots,x^5)\geqslant 0,\ x^1+x^3-2x^4-6x^5=2,\ x^2+x^3-2x^4+7x^5=2,\ x^1+x^2-2x^4+7x^5=2;$

д) $f(x)=x^1+x^3+x^5+x^7\to\inf[\sup],\ x=(x^1,\ldots,x^7)\geqslant 0,\ x^1+3x^2+x^3+2x^4+x^5+x^6=10,\ 2x^1+x^2-x^3+5x^4+3x^6-x^7=20,\ x^1+13x^2+7x^3+5x^5-x^6+2x^7=10.$

3. С помощью приемов, описанных в § 1, запишите задачу линейного программирования в каноническом виде и решите ее с помощью симплекс-метода:

а) $f(x)=2x^1+x^2-x^3+x^5\to\inf[\sup],\ x=(x^1,\ldots,x^5)\geqslant 0,\ 0\leqslant x^1-x^4\leqslant 2,\ x^1+x^2+x^3-x^4-x^5\geqslant 1;$

б) $f(x)=3x^1+10x^2+8x^3-6x^4\to\inf[\sup],\ x=(x^1,\ldots,x^4)\geqslant 0,\ 3x^1+2x^2+x^3-x^4\geqslant -1,\ x^1+3x^2+3x^3-2x^4=-1,\ x^1\leqslant 4;$

в) $f(x)=x^1+3x^2-x^3\to\inf[\sup],\ x^1\geqslant 0,\ x^3\geqslant 0,\ -1\leqslant x^1-x^2+x^3\geqslant 1,\ x^1+x^2+x^3\leqslant 4;$

г) $f(x)=6x^1-12x^2+5x^3+2x^4+3x^5\to\inf[\sup],\ x=(x^1,\ldots,x^5)\geqslant 0,\ 3x^1-6x^2+5x^3+4x^4\leqslant 3,\ x^1-2x^2+4x^4+x^5=2,\ -x^1+2x^2-3x^3-2x^4-x^5\geqslant -7;$

4. Суточный рацион группы животных включает не менее 10 кг продукта $П_1$, 25 кг продукта $П_2$, 15 кг продукта $П_3$, 30 кг продукта $П_4$, 5 кг продукта $П_5$. Эти продукты содержатся в концентратах трех видов k_1, k_2, k_3, причем концентрат k_1 содержит продукты $П_1$, $П_2$, $П_3$, $П_4$. $П_5$ в пропорциях 3 : 1 : 0 : 1 : 0, концентрат k_2 — в пропорциях 1 : 1 : 2 : 1 : 1, концентрат k_3 — 1 : 0 : 1 : 0 : 2, цены концентратов соответственно 0, 5; 0, 9, 0, 7 рублей за килограмм. Сколько нужно в сутки покупать концентратов, чтобы необходимый суточный рацион был наиболее дешевым?

5. Показать, что множество $X=\{x=(x^1,\ldots,x^6)\geqslant 0:\ x^1+4x^2-5x^3-5x^4-3x^5-x^6=2,\ -4x^1+4x^2-12x^3-2x^5+2x^6=2,\ x^1+2x^2-3x^3+3x^4+x^5+x^6=1\}$ состоит из единственной точки. Указание: применить метод искусственного базиса, выбрав в начальной симплекс-таблице разрешающий элемент из столбца x^2 с помощью лексикографического правила (3.48).

6. Множество X задано условиями: $x=(x^1,\ldots,x^6)\geqslant 0:\ x^1-2x^2+x^3=1,\ x^1-2x^2+2x^3+x^4=1,\ x^1-2x^2+2x^3+2x^4-x^5=1,\ x^1-2x^2+2x^3+2x^4-2x^5+x^6=1,\ x^1+x^2+x^3+x^4+x^5+x^6=1.$ Симплекс-методом решите задачу $f(x)=\langle c, x\rangle\to\inf,\ x\in X$ при различных $c\in E^6$ и убедитесь, что минимум достигается в одной и той же точке множества X. Объясните это явление.

7. Примените симплекс-метод к основной задаче (1.21) с вектором $b\geqslant 0$, сведя ее к канонической задаче (1.23). Указание: сравните системы (1.22) и (3) и найдите угловую точку множества Z задачи (1.23).

8. Обобщить симплекс-метод на задачу: $f(x)=\langle c, x\rangle\to\inf;\ x\in X=\{x\in E^n:\ x\geqslant 0,\ Ax=b,\ \alpha_i\leqslant x^i\leqslant\beta_i,\ i=1,\ldots,n\}$, где α_i, β_i заданные величины, $\alpha_i\leqslant\beta_i$ (возможно, некоторые $\alpha_i=-\infty$ и некоторые $\beta_j=+\infty$) [775].

9. Пользуясь упражнением 8, рассмотреть задачи из упражнений 1–3 при дополнительных ограничениях $0\leqslant x^i\leqslant 2$.

§ 5. Условие разрешимости задач линейного программирования. Теоремы двойственности

Будем рассматривать общую задачу линейного программирования:

$$f(x)=\langle c, x\rangle=\langle c_1, x_1\rangle+\langle c_2, x_2\rangle\to\inf,\quad x=(x_1, x_2)\in X, \tag{1}$$

$$X=\{x=(x_1, x_2):\ x_1\in E^{n_1},\ x_2\in E^{n_2},\ A_{11}x_1+A_{12}x_2\leqslant b_1,\ A_{21}x_1+A_{22}x_2=b_2,\ x_1\geqslant 0\}, \tag{2}$$

где A_{ij} — матрицы размера $m_i\times n_j$, $c_j\in E^{n_j}$, $b_i\in E^{m_i}$, $i, j=1, 2$. Как и выше, будем обозначать $f_*=\inf\limits_{x\in X} f(x)$, подразумевая при этом, что $X\neq\varnothing$. Для случая, когда $f_*>-\infty$, введем множество $X_*=\{x\in X:\ f(x)=f_*\}$. Напоминаем, что задача (1), (2) называется разрешимой, если $X_*\neq\varnothing$; каждую точку $x_*\in X_*$ называют решением этой задачи.

1. Приведем теорему существования решения задачи (1), (2), которая дополняет теоремы Вейерштрасса из § 2.1 и характеризует специфику задач линейного программирования.

Теорема 1. *Задача* (1), (2) *разрешима тогда и только тогда, когда* $X\neq\varnothing$ *и целевая функция* $f(x)$ *ограничена снизу на* X, *т. е.* $f_*>-\infty$.

Нетрудно видеть, что для нелинейных задач такая теорема неверна. Например, задача: $f(x)=e^{-x}\to\inf,\ x\in X=\{x\in E^1:\ x\geqslant 0\}$ не имеет решения, хотя и $f_*=0>-\infty$.

Доказательство. Необходимость очевидна, так как условие $X_*\neq\varnothing$ предполагает, что $X\neq\varnothing$ и $f_*>-\infty$. Докажем достаточность. Пусть $X\neq\varnothing$, $f_*>-\infty$. Покажем, что тогда $X_*\neq\varnothing$. Пользуясь конструкциями теоремы 1.1, задачу (1), (2) запишем в канонической форме:

$$f(x)=\langle c, x\rangle\to\inf,\quad x\in X=\{x\in E^n:\ Ax=b,\ x\geqslant 0\}, \tag{3}$$

при этом $c\in E^n$, $b\in E^m$, A — матрица размера $m\times n$. С этой целью положим (см. § 1)

$$x_2=z_1-z_2,\quad z_1=\max\{0; x_2\},\quad z_2=\max\{0; -x_2\},\quad v=b_1-A_{11}x_1-A_{12}x_2,$$

и в пространстве переменных $w=(x_1, z_1, z_2, v)\in E^q$, $q=n_1+2n_2+m_1$, рассмотрим задачу:

$$g(w)=\langle c_1, x_1\rangle+\langle c_2, z_1\rangle+\langle -c_2, z_2\rangle+\langle 0, v\rangle\to\inf,\ w\in W, \tag{4}$$

$$W=\{w\in E^q:\ w\geqslant 0,\ A_{11}x_1+A_{12}z_1+(-A_{12})z_2+I_{m_1}v=b_1,\ A_{21}x_1+A_{22}z_1+(-A_{22})z_2+0v=b_2\}, \tag{5}$$

где I_{m_1} — единичная матрица размера $m_1\times m_1$. Задача (4) ,(5) совпадает с задачей (3), если принять $c=(c_1, c_2, -c_2, 0)\in E^n$,

$$b=(b_1, b_2)\in E^m,\quad A=\begin{pmatrix} A_{11}, & A_{12}, & -A_{12}, & I_{m_1} \\ A_{21}, & A_{22}, & -A_{22}, & 0 \end{pmatrix}$$

— матрица размера $m\times n$, где $m=m_1+m_2$, $n=q=n_1+2n_2+m_1$. Согласно теореме 1.1 из $X\neq\varnothing$, $f_*>-\infty$ следует, что $W\neq\varnothing$, $g_*=\inf\limits_{w\in W} g(w)>-\infty$.

Тогда по теореме 4.3, примененной к канонической задаче (4), (5), множество $W_* = \{w \in W:\ g(w) = g_*\} \neq \varnothing$. Возьмем произвольную точку $w_* = (x_{1*}, z_{1*}, z_{2*}, v_*) \in W_*$. В силу теоремы 1.1 тогда точка $x_* = (x_{1*}, x_{2*} = z_{1*} - z_{2*})$ — решение задачи (1), (2), т. е. $X_* \neq \varnothing$. Теорема 1 доказана. □

Следствие 1. *Задача максимизации*

$$\langle d, x\rangle \to \sup,\ x \in X$$

имеет решение тогда и только тогда, когда $X \neq \varnothing$ *и функция* $\langle d, x\rangle$ *ограничена сверху на* X.

Для того, чтобы убедиться в справедливости этого утверждения, достаточно заметить, что такая задача максимизации равносильна задаче (1), (2) с $c = -d$, и воспользоваться теоремой 1.

2. Прежде чем переходить к изложению так называемых теорем двойственности, докажем несколько важных лемм.

Лемма 1. *Для того чтобы некоторая точка* x_* *из множества* X *была решением канонической задачи* (3), *т. е.* $x_* \in X_*$, *необходимо и достаточно существования точки* $\lambda^* = (\lambda_1^*, \ldots, \lambda_m^*) \in E^m$ *такой, что*

$$A^\top \lambda^* + c \geqslant 0,\quad \langle c, x_*\rangle = -\langle b, \lambda^*\rangle, \tag{6}$$

где $A^\top$ *— матрица, полученная транспонированием матрицы* A.

Доказательство. Необходимость. Возьмем произвольную точку $x_* \in X_*$. Покажем, что тогда необходимо существует точка $\lambda^* \in E^m$ со свойствами (6).

Сначала рассмотрим случай $m = r = \operatorname{rang} A$. Применим к задаче (3) симплекс-метод с антициклином. По условию $f(x_*) = f_* > -\infty$, поэтому симплекс-процесс закончится обнаружением некоторой угловой точки v_* множества X с базисом $B = (A_{j_1}, \ldots, A_{j_r})$, $f(v_*) = f_*$, причем будут выполняться неравенства (3.32):

$$\Delta_k = \langle \bar c, B^{-1}A_k\rangle - c^k \leqslant 0,\quad k = 1, 2, \ldots, n;\quad \bar c = (c^{j_1}, \ldots, c^{j_r}). \tag{7}$$

Положим $\lambda^* = -(B^{-1})^\top \bar c$. Пользуясь известным из линейной алгебры тождеством $\langle Mx, y\rangle = \langle x, M^\top y\rangle$, справедливым для любых $x \in R^n$, $y \in E^m$ и любых матриц M размера $m \times n$, из (7) имеем

$$0 \geqslant \Delta_k = \langle (B^{-1})^\top \bar c, A_k\rangle - c^k = -\langle \lambda^*, A_k\rangle - c_k^k,\quad k = 1, 2, \ldots, n.$$

В векторной форме эти неравенства можно записать так: $A^\top \lambda^* + c \geqslant 0$. Далее вспомним, что у угловой точки v_* базисные координаты $\bar v_* = (v_*^{j_1}, \ldots, v_*^{j_r}) = B^{-1}b$, а небазисные координаты равны нулю. Поэтому

$$\langle c, x_*\rangle = f_* = \langle c, v_*\rangle = \langle \bar c, \bar v_*\rangle = \langle \bar c, B^{-1}b\rangle = \langle (B^{-1})^\top \bar c, b\rangle = -\langle b, \lambda^*\rangle.$$

Таким образом, искомая точка λ^* со свойствами (6) найдена. Случай $m = r = \operatorname{rang} A$ рассмотрен. Пусть теперь $m > r = \operatorname{rang} A$. Тогда в системе уравнений $Ax = b$, которую можем записать в виде $\langle a_i, x\rangle = b^i$, $i = 1, \ldots, m$, где a_i — строки матрицы A, имеются ровно r линейно независимых уравнений. Перенумеровав уравнения, можем считать, что первые r уравнений этой системы линейно независимы, а остальные уравнения с номерами $i = r+1, \ldots, m$, линейно выражаются через первые, базисные уравнения $i = 1, \ldots, r$. Удаление линейно зависимых уравнений приведет к равносильной системе $\bar A x = \bar b$, где $\bar A$ — матрица, состоящая из строк $a_1, a_2, \ldots, a_r$ матрицы A, $\bar b = (b^1, \ldots, b^r)$, и задача (3) сведется к равносильной канонической задаче:

$$f(x) = \langle c, x\rangle \to \inf,\quad x \in X = \{x \geqslant 0;\ \bar A x = \bar b\}.$$

В этой задаче число уравнений равно $r = \operatorname{rang} \bar A$, и по доказанному существует точка $\bar\lambda^* = (\lambda_1^*, \ldots, \lambda_r^*) \in E^r$ такая, что

$$\bar A^\top \bar\lambda^* + c \geqslant 0,\quad \langle c, x_*\rangle = -\langle b, \bar\lambda^*\rangle. \tag{8}$$

Определим точку $\lambda^* = (\bar\lambda^*, 0) \in E^m$, полученную добавлением к координатам $\bar\lambda^*$ нулевых координат $\lambda_{r+1}^* = 0, \ldots, \lambda_m^* = 0$. Тогда из (8) следует, что $A^\top\lambda^* + c \geqslant 0$, $\langle c, x_*\rangle = -\langle b, \lambda^*\rangle$. Необходимость доказана.

Достаточность. Пусть для каких-либо точек $x_* \in X$, $\lambda^* \in E^m$ выполнены соотношения (6). Тогда для всех $x \in X$ имеем

$$0 \leqslant \langle x, A^\top\lambda^* + c\rangle = \langle c, x\rangle + \langle Ax, \lambda^*\rangle = \langle c, x\rangle + \langle b, \lambda^*\rangle = \langle c, x\rangle - \langle c, x_*\rangle.$$

Это значит, что $x_* \in X_*$. Лемма 1 доказана. □

Лемму 1 нетрудно обобщить на случай общей задачи линейного программирования (1), (2).

Лемма 2. *Для того чтобы некоторая точка* $x_* = (x_{1*}, x_{2*})$ *из множества* (2) *была решением задачи* (1), (2), *необходимо и достаточно, чтобы существовала точка* $\lambda^* = (\lambda_1^*, \lambda_2^*)$, $\lambda_1^* \in E^{m_1}$, $\lambda_2^* \in E^{m_2}$ *такая, что*

$$A_{11}^\top\lambda_1^* + A_{21}^\top\lambda_2^* + c_1 \geqslant 0,\quad A_{12}^\top\lambda_1^* + A_{22}^\top\lambda_2^* + c_2 = 0,\quad \lambda_1^* \geqslant 0, \tag{9}$$

$$\langle c_1, x_{1*}\rangle + \langle c_2, x_{2*}\rangle = -\langle b_1, \lambda_1^*\rangle - \langle b_2, \lambda_2^*\rangle, \tag{10}$$

где $A_{ij}^\top$ *— матрица, полученная транспонированием матрицы* A_{ij}.

Доказательство. Возьмем произвольную точку $x_* = (x_{1*}, x_{2*}) \in X_*$. Тогда согласно теореме 1.1 точка $w_* = (x_{1*}, z_{1*}, z_{2*}, v_*)$, где $z_{1*} = \max\{0; x_{2*}\}$, $z_{2*} = \max\{0; -x_{2*}\}$, $v_* = b_1 - A_{11}x_{1*} - A_{12}x_{2*}$, является решением задачи (4), (5), причем $g(w_*) = g_* = f_* = f(x_*)$. Применяя лемму 1 к канонической задаче (4), (5), заключаем, что это возможно тогда и только тогда, когда существует точка $\lambda^* = (\lambda_1^*, \lambda_2^*)$, $\lambda_1^* \in E^{m_1}$, $\lambda_2^* \in E^{m_2}$, такая, что

$$A^\top\lambda^* + c = \begin{bmatrix} A_{11}^\top & A_{21}^\top \\ A_{12}^\top & A_{22}^\top \\ -A_{12}^\top & -A_{22}^\top \\ I & 0 \end{bmatrix} \begin{bmatrix} \lambda_1^* \\ \lambda_2^* \end{bmatrix} + \begin{bmatrix} c_1 \\ c_2 \\ -c_2 \\ 0 \end{bmatrix} = \begin{bmatrix} A_{11}^\top\lambda_1^* + A_{21}^\top\lambda_2^* + c_1 \\ A_{12}^\top\lambda_1^* + A_{22}^\top\lambda_2^* + c_2 \\ -A_{12}^\top\lambda_1^* - A_{22}^\top\lambda_2^* - c_2 \\ I\lambda_1^* + 0\lambda_2^* \end{bmatrix} \geqslant 0,$$

$$f_* = g_* = g(w_*) = \langle c_1, x_{1*}\rangle + \langle c_2, z_{1*}\rangle + \langle -c_2, z_{2*}\rangle + \langle 0, v_*\rangle = -\langle b_1, \lambda_1^*\rangle - \langle b_2, \lambda_2^*\rangle.$$

Учитывая, что $x_{2*} = z_{1*} - z_{2*}$, эти соотношения нетрудно переписать в равносильном виде (9), (10). Лемма 2 доказана. □

С задачей (1), (2) тесно связана следующая задача линейного программирования:

$$\psi(\lambda) = -\langle b_1, \lambda_1\rangle - \langle b_2, \lambda_2\rangle \to \sup,\ \lambda = (\lambda_1, \lambda_2) \in \Lambda, \tag{11}$$

$$\Lambda = \{\lambda = (\lambda_1, \lambda_2):\ \lambda_1 \in E^{m_1},\ \lambda_2 \in E^{m_2},\ A_{11}^\top\lambda_1 + A_{21}^\top\lambda_2 + c_1 \geqslant 0,\ A_{12}^\top\lambda_1 + A_{22}^\top\lambda_2 + c_2 = 0,\ \lambda_1 \geqslant 0\}. \tag{12}$$

Задача (11), (12) называется *двойственной* задачей по отношению к исходной задаче (1), (2), переменные $\lambda = (\lambda_1, \lambda_2)$ называются *двойственными переменными* по отношению к исходным переменным $x = (x_1, x_2)$. Будем обозначать $\psi^* = \sup\limits_{\lambda \in \Lambda} \psi(\lambda)$, $\Lambda^* = \{\lambda \in \Lambda: \psi(\lambda) = \psi^*\}$. Как видим, двойственная задача (11), (12) однозначно определяется по элементам c_1, c_2, b_1, b_2, A_{11}, A_{12}, A_{21}, A_{22} исходной задачи (1), (2).

Лемма 3. *Если в задачах* (1), (2) *и* (11), (12) *множества* X *и* Λ *непусты, то величины* $f_* = \inf\limits_{x \in X} f(x)$, $\psi^* = \sup\limits_{\lambda \in \Lambda} \psi(\lambda)$ *конечны и*

$$\psi^* \leqslant f_*. \tag{13}$$

Доказательство. Возьмем произвольные $x \in X$, $\lambda \in \Lambda$. Тогда справедлива следующая цепочка неравенств, вытекающая из определений (2) и (12) множеств X и Λ:

$$\begin{aligned} f(x) - \psi(\lambda) &= \langle c_1, x_1\rangle + \langle c_2, x_2\rangle + \langle b_1, \lambda_1\rangle + \langle b_2, \lambda_2\rangle \geqslant \\ &\geqslant \langle c_1, x_1\rangle + \langle c_2, x_2\rangle + \langle A_{11}x_1 + A_{12}x_2, \lambda_1\rangle + \langle A_{21}x_1 + A_{22}x_2, \lambda_2\rangle = \\ &= \langle c_1 + A_{11}^\top \lambda_1 + A_{21}^\top \lambda_2, x_1\rangle + \langle c_2 + A_{12}^\top \lambda_1 + A_{22}^\top \lambda_2, x_2\rangle \geqslant 0. \end{aligned} \tag{14}$$

Таким образом,

$$f(x) \geqslant \psi(\lambda), \quad \forall x \in X, \quad \lambda \in \Lambda. \tag{15}$$

Последовательно переходя в неравенстве (15) сначала к нижней грани по $x \in X$, затем к верхней грани по $\lambda \in \Lambda$, убеждаемся, что величины f_*, ψ^* конечны и удовлетворяют неравенству (13). Лемма 3 доказана. □

Выясним, как выглядит задача, двойственная по отношению к двойственной задаче (11), (12). Замечательно, что эта задача, оказывается, с точностью до эквивалентной формы совпадает с исходной задачей (1), (2). Чтобы убедиться в этом, перепишем задачу (11), (12) в равносильном виде, как задачу минимизации:

$$\begin{gathered} -\psi(\lambda) = \langle b_1, \lambda_1\rangle + \langle b_2, \lambda_2\rangle \to \inf, \quad \lambda \in \Lambda, \\ \Lambda = \{\lambda = (\lambda_1, \lambda_2): \lambda_1 \in E^{m_1}, \ \lambda_2 \in E^{m_2}, \ (-A_{11}^\top)\lambda_1 + (-A_{21}^\top)\lambda_2 \leqslant c_1, \\ (-A_{12}^\top)\lambda_1 + (-A_{22}^\top)\lambda_2 = c_2, \ \lambda_1 \geqslant 0\}, \end{gathered} \tag{16}$$

по форме совпадающей с исходной задачей (1), (2), и затем, пользуясь тем же правилом, с помощью которого была сконструирована двойственная задача (11), (12) на основе исходной задачи (1), (2), составим двойственную к (13) задачу. Обозначив двойственные к $\lambda = (\lambda_1, \lambda_2)$ переменные через $x = (x_1, x_2)$, придем к следующей задаче:

$$\begin{gathered} -\langle c_1, x_1\rangle - \langle c_2, x_2\rangle \to \sup, \quad x = (x_1, x_2) \in M, \\ M = \{x = (x_1, x_2): x_1 \in E^{n_1}, \ x_2 \in E^{n_2}, \ (-A_{11}^\top)^\top x_1 + (-A_{12}^\top)^\top x_2 + b_1 \geqslant 0, \\ (-A_{21}^\top)^\top x_1 + (-A_{22}^\top)^\top x_2 + b_2 = 0, \ x_1 \geqslant 0\}, \end{gathered} \tag{17}$$

являющейся двойственной по отношению к задаче (16). Так как $(-A_{ij}^\top)^\top = -A_{ij}$, $i, j = 1, 2$, то нетрудно видеть, что $M = X$ и задача (17) равносильна задаче (1), (2). Таким образом, с учетом сделанных эквивалентных переходов от задачи (11), (12) к задаче (16), от (17) к (1), (2), можем сказать, что задача, двойственная по отношению к двойственной задаче (11), (12), совпадает с исходной задачей (1), (2), и, следовательно, задачи (1), (2) и (11), (12) образуют пару взаимодвойственных задач. Оказывается, параллельное изучение взаимодвойственных задач способствует более глубокому понимании природы этих задач, оказывается полезным при разработке методов их решения, обогащает теорию линейного программирования. Связь между взаимодвойственными задачами (1), (2) и (11), (12) отражена в следующих теоремах, называемых теоремами двойственности.

Теорема 2. *Задача* (1), (2) *имеет решение тогда и только тогда, когда имеет решение двойственная к ней задача* (11), (12). *Иначе говоря, взаимодвойственные задачи линейного программирования либо обе одновременно разрешимы, либо ни одна из них не имеет решения. Если задачи* (1), (2) *и* (11), (12) *разрешимы, то значения их экстремумов совпадают, т. е.*

$$f_* = \psi^*. \tag{18}$$

Доказательство. Пусть задача (1), (2) имеет решение, т. е. $X_* \neq \varnothing$. Возьмем произвольную точку $x_* \in X_*$. Согласно лемме 2 тогда существует точка $\lambda^* \in \Lambda$, для которой справедливо равенство (10). Таким образом, $\Lambda \neq \varnothing$, и, кроме того, $f_* = f(x_*) = \psi(\lambda^*) \leqslant \psi^*$. Отсюда и из (13) следует $f_* = f(x_*) = \psi(\lambda^*) = \psi^*$, т. е. $\lambda^* \in \Lambda^*$. Таким образом, из разрешимости задачи (1), (2) следует разрешимость двойственной к ней задачи (11), (12). Так как задача (1), (2) в свою очередь является двойственной к двойственной задаче (11), (12), то из разрешимости задачи (11), (12) следует разрешимость задачи (1), (2), причем $\psi^* = f_*$. Теорема 2 доказана. □

Теорема 3. *Взаимодвойственные задачи* (1), (2) *и* (11), (12) *имеют решение тогда и только тогда, когда существуют точки* $x_* = (x_{1*}, x_{2*})$, $\lambda^* = (\lambda_1^*, \lambda_2^*)$ *такие, что*

$$x_* \in X, \quad \lambda^* \in \Lambda, \quad f(x_*) = \psi(\lambda^*). \tag{19}$$

Соотношения (19) *справедливы для всех точек* $x_* \in X_*$, $\lambda^* \in \Lambda^*$ *и только для них.*

Доказательство. Необходимость. Пусть задачи (1), (2) и (11), (12) разрешимы, т. е. $X_* \neq \varnothing$, $\Lambda^* \neq \varnothing$. Возьмем любые точки $x_* \in X_*$, $\lambda^* \in \Lambda^*$. Это означает, что $x_* \in X_*$, $f(x_*) = f_*$, $\lambda^* \in \Lambda^*$, $\psi(\lambda^*) = \psi^*$. Но согласно теореме 2 тогда $f_* = \psi^*$, поэтому $f(x_*) = \psi(\lambda^*)$. Таким образом, в качестве точек x_*, λ^*, удовлетворяющих условиям (19), можно взять любые точки из множеств X_*, Λ^*.

Достаточность. Пусть для каких-то точек $x_* = (x_{1*}, x_{2*})$, $\lambda^* = (\lambda_1^*, \lambda_2^*)$ выполняются соотношения (19). Это значит, что множества X и Λ непусты и по лемме 3 тогда $f_* > -\infty$, $\psi^* < +\infty$. Отсюда, из теоремы 1 и следствия к ней следует, что задачи (1), (2) и (11), (12) разрешимы, т. е. $X_* \neq \varnothing$, $\Lambda^* \neq \varnothing$. Согласно теореме 2 тогда $f_* = \psi^*$. Отсюда и из (19) имеем $f_* \leqslant f(x_*) = \psi(\lambda^*) \leqslant \psi^* = f_*$. Это значит, что все неравенства здесь обращаются в равенства, т. е. $f(x_*) = f_*$, $\psi(\lambda^*) = \psi^*$ и, следовательно, $x_* \in X_*$, $\lambda^* \in \Lambda^*$. Теорема 3 доказана. □

Замечание. Условия (19) равносильны условиям

$$x_* \in X, \quad \lambda^* \in \Lambda, \quad f(x_*) \leqslant \psi(\lambda^*). \tag{20}$$

В самом деле, совмещая неравенство из (20) с неравенством (15) при $x = x_*$, $\lambda = \lambda^*$, приходим к равенству $f(x_*) = \psi(\lambda^*)$.

Теорема 4. *Взаимодвойственные задачи* (1), (2) *и* (11), (12) *имеют решение тогда и только тогда, когда существуют точки* $x_* = (x_{1*}, x_{2*})$, $\lambda^* = (\lambda_1^*, \lambda_2^*)$ *такие, что*

$$\begin{gathered} x_* \in X, \quad \lambda^* \in \Lambda, \quad x_{1*}^j (A_{11}^\top \lambda_1^* + A_{21}^\top \lambda_2^* + c_1)^j = 0, \quad j = 1, \ldots, n_1, \\ (\lambda_1^*)^i (b_1 - A_{11}x_{1*} - A_{12}x_{2*})^i = 0, \quad i = 1, \ldots, m_1. \end{gathered} \tag{21}$$

Соотношения (21) *справедливы для всех точек* $x_* \in X_*$, $\lambda^* \in \Lambda^*$ *и только для них.*

Равенства из (21) называются условиями дополняющей нежесткости (ср. с аналогичными условиями из § 2.3, теорема 2).

Доказательство. Необходимость. Пусть задачи (1), (2) и (11), (12) имеют решение. Согласно теореме 3 тогда условия (19) справедливы при всех $x_* \in X_*$, $\lambda^* \in \Lambda^*$. В частности, $f(x_*) - \psi(\lambda^*) = 0$. Отсюда и из (14) заключаем, что при $x = x_*$, $\lambda = \lambda^*$ все неравенства в (14) обращаются в равенство, что с учетом ограничений (2), (12) возможно только при

$$\langle A_{11}^\top \lambda_1^* + A_{21}^\top \lambda_2^* + c_1, x_{1*} \rangle = \sum_{j=1}^{n_1} (A_{11}^\top \lambda_1^* + A_{21}^\top \lambda_2^* + c_1)^j x_{1*}^j = 0. \qquad (22)$$

В силу (2), (12) каждое слагаемое в сумме (22) неотрицательно. Поэтому из (22) следуют первые равенства (21). Для доказательства остальных равенств (21) воспользуемся неравенствами

$$\begin{aligned} f(x) - \psi(\lambda) &= \langle c_1, x_1 \rangle + \langle c_2, x_2 \rangle + \langle b_1, \lambda_1 \rangle + \langle b_2, \lambda_2 \rangle \geqslant \\ &\geqslant \langle -A_{11}^\top \lambda_1 - A_{21}^\top \lambda_2, x_1 \rangle + \langle -A_{12}^\top \lambda_1 - A_{22}^\top \lambda_2, x_2 \rangle + \\ &+ \langle b_1, \lambda_1 \rangle + \langle b_2, \lambda_2 \rangle = \langle b_1 - A_{11} x_1 - A_{12} x_2, \lambda_1 \rangle + \\ &+ \langle b_2 - A_{21} x_1 - A_{22} x_2, \lambda_2 \rangle \geqslant 0 \quad \forall x \in X, \ \forall \lambda \in \Lambda, \end{aligned} \qquad (23)$$

аналогичными (14) и также вытекающими из определений (2), (12) множеств X, Λ. Из (23) при $x = x_*$, $\lambda = \lambda^*$ с учетом равенства (19) имеем

$$\langle b_1 - A_{11} x_{1*} - A_{12} x_{2*}, \lambda_1^* \rangle = \sum_{i=1}^{m_1} (b_1 - A_{11} x_{1*} - A_{12} x_{2*})^i (\lambda_1^*)^i = 0. \qquad (24)$$

Из неотрицательности каждого слагаемого в сумме (24) следует вторая группа равенств (21).

Достаточность. Пусть для каких-то точек $x_* = (x_{1*}, x_{2*})$, $\lambda^* = (\lambda_1^*, \lambda_2^*)$ выполнены условия (21). Тогда для них справедливы равенства (22), (24). Отсюда и из (2), (12) следует, что в (23) при $x = x_*$, $\lambda = \lambda^*$ все неравенства обращаются в равенства и, следовательно, $f(x_*) = \psi(\lambda^*)$. Таким образом, точки x_*, λ^* удовлетворяют условиям (19). Согласно теореме 3 тогда $x_* \in X_*$, $\lambda^* \in \Lambda^*$. Теорема 4 доказана. □

Покажем, что двойственные переменные в задачах линейного программирования можно истолковать как обобщение понятия множителей Лагранжа, используемых в классическом анализе при исследовании задач на условный экстремум, см. гл. 2. Введем функцию

$$L(x,\lambda) = \langle c_1, x_1 \rangle + \langle c_2, x_2 \rangle + \langle \lambda_1, A_{11} x_1 + A_{12} x_2 - b_1 \rangle + \langle \lambda_2, A_{21} x_1 + A_{22} x_2 - b_2 \rangle \qquad (25)$$

переменных $x = (x_1, x_2) \in X_0 = \{x = (x_1, x_2)\colon x_1 \in E^{n_1}, \ x_2 \in E^{n_2}, \ x_1 \geqslant 0\}$, $\lambda = (\lambda_1, \lambda_2) \in \Lambda_0 = \{\lambda = (\lambda_1, \lambda_2)\colon \lambda_1 \in E^{m_1}, \ \lambda_2 \in E^{m_2}, \ \lambda_1 \geqslant 0\}$. Эта функция называется функцией Лагранжа задачи (1), (2), переменные $\lambda = (\lambda_1, \lambda_2)$ называются множителями Лагранжа, причем $\lambda_1 \geqslant 0$ — множители, соответствующие ограничениям типа неравенств в определении множества (2), λ_2 — множители, соответствующие ограничениям типа равенств. Пользуясь тождеством $\langle A_{ij} x_j, \lambda_i \rangle = \langle x_j, A_{ij}^\top \lambda_i \rangle$, функцию (25) можно записать в виде

$$\begin{aligned} L(x,\lambda) = &\langle -b_1, \lambda_1 \rangle + \langle -b_2, \lambda_2 \rangle + \langle x_1, A_{11}^\top \lambda_1 + A_{21}^\top \lambda_2 + c_1 \rangle + \\ &+ \langle x_2, A_{12}^\top \lambda_1 + A_{22}^\top \lambda_2 + c_2 \rangle, \quad x \in X_0, \quad \lambda \in \Lambda_0. \end{aligned} \qquad (26)$$

Определение 1. Точка $(x_*, \lambda^*) \in X_0 \times \Lambda_0$ называется седловой точкой функции Лагранжа, если

$$L(x_*, \lambda) \leqslant L(x_*, \lambda^*) \leqslant L(x, \lambda^*) \quad \forall x \in X_0, \quad \forall \lambda \in \Lambda_0. \qquad (27)$$

Теорема 5. *Взаимодвойственные задачи* (1), (2) *и* (11), (12) *имеют решение тогда и только тогда, когда существуют точки* $x_* = (x_{1*}, x_{2*}) \in X_0$, $\lambda^* = (\lambda_1^*, \lambda_2^*) \in \Lambda_0$, *образующие седловую точку* (x_*, λ^*) *функции Лагранжа. Точка* $(x_*, \lambda^*) \in X_0 \times \Lambda_0$ *будет седловой точкой тогда и только тогда, когда* $x_* \in X_*$, $\lambda^* \in \Lambda^*$, *т. е. множество седловых точек функции Лагранжа совпадает со множеством* $X_* \times \Lambda^*$. *Справедливы равенства*

$$L(x_*, \lambda^*) = f_* = f(x_*) = \psi(\lambda^*) = \psi^* \quad \forall (x_*, \lambda^*) \in X_* \times \Lambda^*. \qquad (28)$$

Доказательство. Необходимость. Пусть задачи (1), (2) и (11), (12) имеют решение. Возьмем произвольную точку (x_*, λ^*), где $x_* \in X_*$, $\lambda^* \in \Lambda^*$. Согласно теоремам 2–4 тогда

$$\begin{gathered} f(x_*) = \psi(\lambda^*) = f_* = \psi^*, \quad \langle x_*, A_{11}^\top \lambda_1^* + A_{21}^\top \lambda_2^* + c_1 \rangle = 0, \\ \langle \lambda_1^*, b_1 - A_{11} x_{1*} - A_{12} x_{2*} \rangle = 0; \end{gathered}$$

кроме того $A_{21} x_{1*} + A_{22} x_{2*} = b_2$, $A_{12}^\top \lambda_1^* + A_{22}^\top \lambda_2^* + c_1 = 0$ по определению множеств X, Λ. С учетом перечисленных равенств из (25), (26) при $x = x_*$, $\lambda = \lambda^*$ получим равенства (28). Кроме того, из (26) при $\lambda = \lambda^*$ имеем: $L(x, \lambda^*) = \psi(\lambda^*) + \langle x_1, A_{11}^\top \lambda_1^* + A_{21}^\top \lambda_2^* + c_1 \rangle \ \forall x \in X_0$. Отсюда и из уже доказанных равенств (28) следует

$$L(x, \lambda^*) - L(x_*, \lambda^*) = \langle x_1, A_{11}^\top \lambda_1^* + A_{21}^\top \lambda_2^* + c_1 \rangle \geqslant 0 \quad \forall x \in X_0.$$

Правое неравенство (27) доказано. Далее, из (25) при $x = x_*$ имеем $L(x_*, \lambda) = f(x_*) + \langle \lambda_1, A_{11} x_{1*} + A_{12} x_{2*} - b_1 \rangle \ \forall \lambda \in \Lambda_0$. Отсюда и из (28) следует левое неравенство (27)

$$L(x_*, \lambda^*) - L(x_*, \lambda) = \langle \lambda_1, b_1 - A_{11} x_{1*} - A_{12} x_{2*} \rangle \geqslant 0 \quad \forall \lambda \in \Lambda_0.$$

Тем самым установлено, что любая точка $(x_*, \lambda^*) \in X_* \times \Lambda^*$ является седловой точкой функции Лагранжа.

Достаточность. Пусть $(x_*, \lambda^*) \in X_* \times \Lambda^*$ — какая-либо седловая точка функции (25). Покажем, что тогда $x_* \in X_*$, $\lambda^* \in \Lambda^*$, т. е. задачи (1), (2) и (11), (12) разрешимы. С учетом представлений (25), (26) функции Лагранжа перепишем неравенства (27) в развернутом виде

$$\begin{aligned} f(x_*) + \langle \lambda_1, A_{11} x_{1*} + A_{12} x_{2*} - b_1 \rangle + \langle \lambda_2, A_{21} x_{1*} + A_{22} x_{2*} - b_2 \rangle \leqslant \\ \leqslant L(x_*, \lambda^*) \leqslant \psi(\lambda^*) + \langle x_1, A_{11}^\top \lambda_1^* + A_{21}^\top \lambda_2^* + c_1 \rangle + \\ + \langle x_2, A_{12}^\top \lambda_1^* + A_{22}^\top \lambda_2^* + c_2 \rangle \quad \forall x \in X_0, \quad \forall \lambda \in \Lambda_0. \end{aligned} \qquad (29)$$

Точка $\lambda = (\lambda_1 = 0, \ \lambda_2 = t(A_{21} x_{1*} + A_{22} x_{2*} - b_2)) \in \Lambda_0 \ \forall t \in \mathbb{R}$. Подставив эту точку в (29), из левого неравенства имеем: $t |A_{21} x_{1*} + A_{22} x_{2*} - b_2|^2 \leqslant L(x_*, \lambda^*) - f(x_*) \ \forall t \in \mathbb{R}$. Разделим обе части этого неравенства на t, считая $t > 0$, и устремим $t \to +\infty$. Получим $|A_{21} x_{1*} + A_{22} x_{2*} - b_1|^2 \leqslant 0$, что возможно только при $A_{21} x_{1*} + A_{22} x_{2*} = b_2$. Далее, положим в (29) $\lambda = (\lambda_1 = (0, \ldots, 0, \lambda_1^i, 0, \ldots, 0), \lambda_2 = 0)$, считая $\lambda_1^i \geqslant 0$. Получим $\lambda_1^i (A_{11} x_{1*} + A_{12} x_{2*} - b_1)^i \leqslant L(x_*, \lambda^*) - f(x_*) \ \forall \lambda_1^i \geqslant 0$. Разделим это неравенство на $\lambda_1^i > 0$ и устремим $\lambda_1^i \to +\infty$.

Будем иметь $(A_{11}x_{1*} + A_{12}x_{2*} - b_1)^i \leqslant 0$ при каждом $i = 1, \ldots, n_1$, т. е. $A_{11}x_{1*} + A_{12}x_{2*} \leqslant b_1$. Следовательно, $x_* \in X$. Аналогичными рассуждениями, полагая в (29) $x = (x_1 = 0, x_2 = t(A_{12}^\top \lambda_1^* + A_{22}^\top \lambda_2^* + c_2))$, $\forall t \in \mathbb{R}$, и $x = (x_1 = (0, \ldots, 0, x_1^i, 0, \ldots, 0), x_2 = 0)$, $x_1^i \geqslant 0$, устанавливаем, что $\lambda^* \in \Lambda$. Таким образом, показано, что всякая седловая точка (x_*, λ^*) функции (25) принадлежит $X \times \Lambda$. Наконец, положив в (29) $x = (x_1 = 0, x_2 = 0)$, $\lambda = (\lambda_1 = 0, \lambda_2 = 0)$, получим $f(x_*) \leqslant L(x_*, \lambda^*) \leqslant \psi(\lambda^*)$. С другой стороны для точек $x_* \in X$, $\lambda^* \in \Lambda$ справедливо неравенство (17): $f(x_*) \geqslant \psi(\lambda^*)$. Следовательно, $f(x_*) = \psi(\lambda^*)$. Это значит, что точки x_*, λ^* удовлетворяют всем условиям (19). В силу теоремы 3 тогда $x_* \in X_*$, $\lambda^* \in \Lambda^*$. Тем самым показано, что все седловые точки функции Лагранжа принадлежат множеству $X_* \times \Lambda^*$. С другой стороны, выше было установлено, что каждая точка из $X_* \times \Lambda^*$ является седловой. Следовательно, множество седловых точек функции Лагранжа задачи (1), (2) совпадает со множеством $X_* \times \Lambda^*$. Теорема 5 доказана. □

В следующей теореме вопросы разрешимости и неразрешимости взаимодвойственных задач обсуждаются в терминах пустоты или непустоты множеств X, Λ. Предварительно отметим, что согласно теореме 1 и следствия к ней неразрешимость задачи (1), (2) означает, что либо $X = \varnothing$, либо $X \neq \varnothing$, но $f_* = -\infty$, а для двойственной задачи (11), (12) неразрешимость равносильна тому, что либо $\Lambda = \varnothing$, либо $\Lambda \neq \varnothing$, но $\psi^* = +\infty$.

Теорема 6. *Справедливы следующие утверждения* а)–г):

а) *взаимодвойственные задачи* (1), (2) *и* (11), (12) *разрешимы тогда и только тогда, когда множества* X *и* Λ *непусты одновременно;*

б) *в задаче* (1), (2) $X \neq \varnothing$, $f_* > -\infty$ *тогда и только тогда, когда в задаче* (11), (12) $\Lambda \neq \varnothing$, $\psi^* < +\infty$;

в) *если в задаче* (1), (2) $X \neq \varnothing$, $f_* = -\infty$, *то в двойственной задаче* (11), (12) $\Lambda = \varnothing$; *обратно: если* $\Lambda \neq \varnothing$, $\psi^* = +\infty$, *то* $X = \varnothing$;

г) *если в задаче* (1), (2) $X \neq \varnothing$, *а в задаче* (11), (12) $\Lambda = \varnothing$, *то* $f_* = -\infty$; *обратно: если* $X = \varnothing$, $\Lambda \neq \varnothing$, *то* $\psi^* = +\infty$.

Доказательство. а) если задачи (1), (2) и (11), (12) разрешимы, то, конечно, $X \neq \varnothing$, $\Lambda \neq \varnothing$. Обратно, если $X \neq \varnothing$, $\Lambda \neq \varnothing$, то из леммы 3 следует, что $f_* > -\infty$, $\psi^* < +\infty$, и разрешимость задач (1), (2) и (11), (12) вытекает из теоремы 1 и следствия к ней.

б) Пусть в задаче (1), (2) $X \neq \varnothing$, $f_* > -\infty$. Тогда согласно теореме 1 задача (1), (2) разрешима, а по теореме 2 разрешима и двойственная задача (11), (12), т. е. $\Lambda \neq \varnothing$, $\psi^* < +\infty$. Обратно: из $\Lambda \neq \varnothing$, $\psi^* < +\infty$ следует разрешимость задачи (11), (12), поэтому разрешима и двойственная к ней задача (1), (2), так что $X \neq \varnothing$, $f_* > -\infty$.

в) Это утверждение легко доказывается рассуждениями от противного. Пусть $X \neq \varnothing$, $f_* = -\infty$, но $\Lambda \neq \varnothing$. Согласно утверждению б) тогда обе задачи (1), (2) и (11), (12) имеют решение и $f_* > -\infty$, что противоречит условию. Аналогично доказывается, что если $\Lambda \neq \varnothing$, $\psi^* = +\infty$, то $X = \varnothing$.

г) Пусть $X \neq \varnothing$, $\Lambda = \varnothing$, но $f_* > -\infty$. Тогда в силу утверждения а) $\Lambda \neq \varnothing$, $\psi^* < +\infty$, что противоречит условию $\Lambda = \varnothing$. Аналогично убеждаемся, что если $X = \varnothing$, $\Lambda \neq \varnothing$, то $\psi^* = +\infty$. Теорема 6 доказана. □

Следующий пример показывает, что возможен случай, когда во взаимодвойственных задачах (1), (2) и (11), (12) оба множества X и Λ пусты.

Пример 1. Исходная задача: $f(x) = x^1 - 2x^2 \to \inf$, $x \in X = \{x = (x^1, x^2) \geqslant 0\colon\ x^1 - x^2 = 1,\ x^1 - x^2 = 2\}$. Двойственная задача: $\psi(\lambda) = -\lambda^1 - 2\lambda^2 \to \sup$, $\lambda \in \Lambda = \{\lambda = (\lambda^1, \lambda^2)\colon\ \lambda^1 + \lambda^2 \geqslant -1,\ \lambda^1 + \lambda^2 \leqslant -2\}$. Ясно, что $X = \varnothing$, $\Lambda = \varnothing$.

Задачи линейного программирования с противоречивыми условиями, когда $X = \varnothing$ или $\Lambda = \varnothing$, изучались в [49; 297; 298; 644]. Приведенные выше теоремы двойственности часто позволяют получить содержательную информацию о рассматриваемой задаче линейного программирования, иногда на этом пути удается провести полное исследование задачи и даже получить ее решение. Для иллюстрации рассмотрим задачу линейного программирования, не содержащую ограничения типа неравенств.

Пример 2. Рассмотрим задачу

$$f(x) = \langle c, x\rangle \to \inf, \quad x \in X = \{x \in E^n\colon\ Ax = b\},$$

где A — матрица размера $m \times n$, $c \in E^n$, $b \in E^m$. Эта задача является частным случаем задачи (1), (2), когда $n_1 = 0$, $n_2 = n$, $m_1 = 0$, $m_2 = m$, $A_{22} = A$, $b_2 = b$, матрицы A_{11}, A_{12}, A_{21}, b_1 отсутствуют. Двойственной к ней является задача:

$$\psi(\lambda) = -\langle b, \lambda\rangle \to \sup, \quad \lambda \in \Lambda = \{\lambda \in E^m\colon\ A^\top \lambda + c = 0\}.$$

Если $X \neq \varnothing$, $f_* > -\infty$, то, согласно теореме 6, $\Lambda \neq \varnothing$, $\psi^* < +\infty$ и, следовательно, вектор c представим в виде $c = -A^\top \lambda_0$, где $\lambda_0 \in \Lambda$. Но тогда

$$f(x) = \langle c, x\rangle = -\langle A^\top \lambda_0, x\rangle = -\langle \lambda_0, Ax\rangle = -\langle \lambda_0, b\rangle = \text{const}$$

при всех $x \in X$, так что $X_* = X$. Аналогично, если $x_0 \in X$, то $b = Ax_0$ и $\psi(\lambda) = -\langle b, \lambda\rangle = -\langle Ax_0, \lambda\rangle = -\langle x_0, A^\top \lambda\rangle = \langle x_0, c\rangle = \text{const}\ \forall \lambda \in \Lambda$, так что $\psi^* = \langle x_0, c\rangle = f_*$, $\Lambda^* = \Lambda$. Как видим, задачи линейного программирования без ограничений типа неравенств малосодержательны и большого интереса не представляют.

3. В заключение докажем еще одну теорему, известную в литературе под названием *теоремы Фаркаша*. Эта теорема имеет важные приложения в выпуклом анализе, в теории экстремальных задач и может быть легко доказана на основе приведенных выше теорем двойственности.

Теорема 7. *Пусть множества* X, Λ *определены согласно* (2), (12), $X \neq \varnothing$, *пусть* a *— заданное число. Тогда для того чтобы* $f(x) = \langle c_1, x_1\rangle + \langle c_2, x_2\rangle \geqslant a$ *для всех* $x = (x_1, x_2) \in X$, *необходимо и достаточно, чтобы* $\Lambda \neq \varnothing$ *и существовала точка* $\lambda^* = (\lambda_1^*, \lambda_2^*) \in \Lambda$ *такая, что* $\psi(\lambda^*) = -\langle b_1, \lambda_1^*\rangle - \langle b_2, \lambda_2^*\rangle \geqslant a$.

Доказательство. Необходимость. Пусть $X \neq \varnothing$ и $f(x) \geqslant a$ при всех $x \in X$. Тогда $f_* = \inf\limits_{x \in X} f(x) \geqslant a > -\infty$ и в силу теоремы 1 задача (1), (2) имеет решение. Согласно теореме 2 двойственная задача (11), (12) также будет разрешима, т. е. $\Lambda \neq \varnothing$ и найдется точка $\lambda^* \in \Lambda$, для которой $\psi(\lambda^*) = \sup\limits_{\lambda \in \Lambda} \psi(\lambda) = \psi^* = f_* \geqslant a$.

Достаточность. Пусть $X \neq \varnothing$, $\Lambda \neq \varnothing$ и точка $\lambda^* \in \Lambda$ такова, что $\psi(\lambda^*) \geqslant a$. Тогда с помощью неравенства (15) при $\lambda = \lambda^*$ имеем $f(x) \geqslant \psi(\lambda^*) \geqslant a$ при всех $x \in X$. Теорема 7 доказана. □

В приложениях чаще всего используется следующий частный вариант теоремы Фаркаша.

Теорема 8. *Пусть* A_1, A_2 *— матрицы размера* $m_1 \times n$, $m_2 \times n$, *вектор* $c \in E^n$. *Тогда для того чтобы* $\langle c, x\rangle \geqslant 0$ *при всех* x *таких, что* $A_1 x \leqslant 0$, $A_2 x = 0$, *необходимо и достаточно, чтобы существовала точка* $\lambda^* = (\lambda_1^*, \lambda_2^*)$, $\lambda_1^* \in E^{m_1}$, $\lambda_1^* \geqslant 0$, $\lambda_2^* \in E^{m_2}$ *такая, что*

$$c = -A_1^\top \lambda_1^* - A_2^\top \lambda_2^*. \tag{30}$$

Доказательство. Положим $X=\{x\in E^n\colon A_1x\leqslant 0,\ A_2x=0\}$, $\Lambda=\{\lambda=(\lambda_1,\lambda_2)\colon \lambda_1\in E^{m_1},\ \lambda_2\in E^{m_2},\ A_1^{\top}\lambda_1+A_2^{\top}\lambda_2+c=0\}$. Эти множества являются частными случаями множеств X, Λ из (2), (12) при $A_{11}=0$, $A_{12}=A_1$, $A_{21}=0$, $A_{22}=A_2$, $b_1=0$, $b_2=0$, $c_1=0$, $c_2=c$. Здесь $X\neq\varnothing$, так как $0\in X$. Отсюда и из теоремы 7 при $a=0$ следует утверждение теоремы 8, причем в качестве искомой точки $\lambda^*=(\lambda_1^*,\lambda_2^*)$ можно взять любую точку $\lambda^*\in\Lambda$. Попутно отметим, что здесь $\psi(\lambda)\equiv 0=\psi^*=f_*=f(0)\leqslant f(x)$ при всех $x\in X$, $\lambda\in\Lambda$, $\Lambda^*=\Lambda$, а равенство (30) вытекает из принадлежности точки λ^* множеству Λ. Теорема 8 доказана. □

Другие теоремы двойственности для задач линейного программирования, их обобщения на случай нелинейных экстремальных задач будут приведены ниже в § 4.9, там же будет установлена связь между двойственными переменными и множителями Лагранжа, объяснены некоторые тайны происхождения двойственных задач. Различные методы решения задач линейного программирования, основанные на теории двойственности, экономические и игровые интерпретации этой теории, ее приложения к теории линейных неравенств изложены, например, в [48; 49; 52; 76; 116; 203; 216; 231; 232; 259; 295; 297–299; 330; 356; 373; 470; 586; 612; 620; 670; 719; 746; 747; 750–752; 775; 776].

Упражнения

1. Напишите двойственные задачи к задачам из упражнения 1–4 к § § 3, 4 и решите их симплекс-методом, преобразовав их при необходимости к каноническому виду.

2. Напишите двойственные задачи для канонической и основной задач (1.15) и (1.21). Сформулируйте для них аналоги теорем 2–6.

3. Приведите примеры взаимодвойственных задач линейного программирования, в которых множества решений X_*, Λ^* непусты и, кроме того, 1) оба эти множества состоят из единственной точки; 2) оба содержат более одной точки и ограничены; 3) оба неограничены; 4) одно из них состоит из единственной точки, другое ограничено и содержит более одной точки; 5) одно из них состоит из единственной точки, другое неограничено; 6) оба множества содержат более одной точки, но одно из них ограничено, другое неограничено.

4. Приведите примеры взаимодвойственных задач линейного программирования, в которых реализуются случаи, описанные в утверждениях а)–г) теоремы 6.

5. В теореме 4 утверждается, что любые точки $x_*\in X_*$, $\lambda^*\in\Lambda^*$ удовлетворяют условию дополняющей нежесткости, т. е. в каждом из произведений (21) хотя бы один из сомножителей (возможно и оба) равны нулю. Доказать, что можно подобрать такие $x_*\in X_*$, $\lambda^*\in\Lambda^*$, когда в каждом из произведений (21) лишь один сомножитель обращается в нуль (условие строгой дополняющей нежесткости) (см. [670]).

6. Пусть в задаче (1), (2) $b_1=0$, $b_2=0$, и пусть эта задача разрешима. Доказать, что тогда $f(0)=f_*=\psi^*=0$. Указание: написать двойственную задачу и заметить, что $\psi(\lambda)\equiv 0$ $\forall\lambda\in\Lambda$, $0\in X$.

7. Доказать, что задача (1), (2) разрешима при любых $c=(c_1,c_2)$, $c_1\in E^{n_1}$, $c_2\in E^{n_2}$, тогда и только тогда, когда множество (2) ограничено.

8. Пусть множества X, Λ определены согласно (2), (12), пусть $X\neq\varnothing$. Доказать, что тогда совместна одна и только одна из систем: $x\in X$, $f(x)<a$ или $\lambda\in\Lambda$, $\psi(\lambda)\geqslant a$. Убедиться, что это утверждение равносильно теореме Фаркаша. Указание: пользуясь неравенством (17) и теоремой 7, показать, что эти системы не могут быть одновременно совместными и одновременно несовместными.

9. Доказать, что совместна одна и только одна из следующих систем: $A_1x\leqslant 0$, $A_2x=0$, $\langle c,x\rangle<0$ или $A_1^{\top}\lambda_1+A_2^{\top}\lambda_2+c=0$, $\lambda_1\geqslant 0$ (обозначения см. в теореме 8). Убедиться, что это утверждение равносильно теореме 8.

10. Доказать, что непустое множество (2) неограничено тогда и только тогда, когда существует вектор $c\in E^{n_1+n_2}$, для которого $\inf\limits_{x\in X}\langle c,x\rangle=-\infty$.

11. Доказать, что множество X, определенное согласно (2), непусто тогда и только тогда, когда $\langle b_1,\lambda_1\rangle+\langle b_2,\lambda_2\rangle\geqslant 0$ для всех $\lambda\in\Lambda=\{\lambda=(\lambda_1,\lambda_2)\colon \lambda_1\in E^{m_1},\ \lambda_2\in E^{m_2},\ \lambda_1\geqslant 0,\ A_{11}^{\top}\lambda_1+A_{21}^{\top}\lambda_2\geqslant 0,\ A_{12}^{\top}\lambda_1+A_{22}^{\top}\lambda_2=0\}$. Указание: для задачи $g(\lambda)=\langle b_1,\lambda_1\rangle+\langle b_2,\lambda_2\rangle\to\inf$, $\lambda\in\Lambda$ написать двойственную задачу и воспользоваться теоремой 7.

12. Пусть множества X,Λ взяты из упражнения 11. Доказать, что тогда совместна одна и только одна из двух систем: $x\in X$ или $\lambda\in\Lambda$, $\langle b_1,\lambda_1\rangle+\langle b_2,\lambda_2\rangle<0$. Убедиться, что это утверждение равносильно утверждению из упражнения 11.

В следующих упражнениях приведены формулировки ряда важных теорем теории линейных неравенств. Доказательства этих и других важных теорем из упомянутой теории читатель найдет, например, в [48; 54; 752].

13. Пусть A — матрица размера $m\times n$, $b\in E^m$. Доказать, что совместна одна и только одна из двух систем:

1) $Ax\leqslant b$ или $\langle b,\lambda\rangle<0$, $A^{\top}\lambda=0$, $\lambda\geqslant 0$ (теорема Александрова — Фана [54, стр. 75]);

2) $Ax=b$ или $A^{\top}\lambda=0$, $\langle b,\lambda\rangle<0$ [54, стр. 55, 74];

3) $Ax=b$, $x\geqslant 0$ или $A^{\top}\lambda\geqslant 0$, $\langle b,\lambda\rangle<0$ (теорема Минковского — Фаркаша, [54, стр. 55, 74]).

14. Пусть A_i — матрица размера $m_i\times n$, $b_i\in E^{m_i}$, $i=1,2,3$; пусть система $A_1x\leqslant b_1$, $A_2x=b_2$ совместна. Доказать, что тогда совместна одна и только одна из систем:

$$A_1x\leqslant b_1,\quad A_2x=b_2,\quad A_3x<b_3$$

или

$$A_1^{\top}\lambda_1+A_2^{\top}\lambda_2+A_3^{\top}\lambda_3=0,\quad \langle b_1,\lambda_1\rangle+\langle b_2,\lambda_2\rangle+\langle b_3,\lambda_3\rangle\geqslant 0,\quad \lambda_1\geqslant 0,\ \lambda_3\geqslant 0,\ \lambda_3\neq 0$$

(теорема Моцкина, [54, стр. 78]).

15. Пусть A — матрица размера $m\times n$. Доказать, что совместна одна и только одна из двух систем [54, стр. 79]:

1) $Ax=0$, $x\geqslant 0$, $x\neq 0$ или $A^{\top}\lambda>0$ (теорема Гордана);

2) $Ax=0$, $x>0$ или $A^{\top}\lambda\geqslant 0$, $A^{\top}\lambda\neq 0$ (теорема Штимке);

3) $Ax\leqslant 0$, $x\geqslant 0$, $x\neq 0$ или $A^{\top}\lambda>0$, $\lambda\geqslant 0$ (теорема Вилля);

4) $Ax\leqslant 0$, $x>0$ или $A^{\top}\lambda\geqslant 0$, $A^{\top}\lambda\neq 0$, $\lambda\geqslant 0$.

ГЛАВА 4

Элементы выпуклого анализа

Прежде чем переходить к изложению численных методов решения задач, более сложных, чем задачи линейного программирования, остановимся на элементах выпуклого анализа — области математики, в которой изучаются свойства выпуклых множеств и выпуклых функций, и которая играет фундаментальную роль в теории и методах решения экстремальных задач [5; 14; 15; 34; 48; 49; 54; 61; 83; 84; 106; 133; 186; 191; 204; 209; 225; 233; 264; 265; 278; 286; 295; 297; 314; 315; 358; 364; 366; 374; 446; 449; 471; 472; 499; 511; 543; 584; 587; 603–605; 613; 617; 636; 670; 683; 689; 735; 747; 767; 774; 785; 795; 798; 814].

§ 1. Выпуклые множества

1. Начнем с рассмотрения конкретных примеров выпуклых множеств. Сначала напомним

Определение 1. Множество X называется *выпуклым*, если для любых $u, v \in X$ точка $u_\alpha = v + \alpha(u-v) = \alpha u + (1-\alpha)v$ принадлежит X при всех α, $0 \leqslant \alpha \leqslant 1$. Иначе говоря, множество X выпукло, если отрезок $[v, u] = \{u_\alpha : u_\alpha = v\alpha(u-v), 0 \leqslant \alpha \leqslant 1\}$, соединяющий любые две точки u, v из X целиком лежит в X (рис. 4.1).

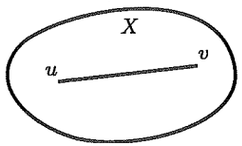

Рис. 4.1

Все пространство E^n, очевидно, образует выпуклое множество. Пустое множество и множество, состоящее из одной точки, удобно считать выпуклыми. Тогда из определения 1 непосредственно следует, что пересечение любого числа выпуклых множеств является выпуклым множеством. Приведем другие примеры выпуклых множеств.

Пример 1. Шар

$$S(u_0, R) = \{u : u \in E^n, |u - u_0| \leqslant R\} \tag{1}$$

радиуса $R > 0$ с центром в точке u_0 является выпуклым множеством. В самом деле, если $u, v \in S(u_0, R)$, то, пользуясь неравенством треугольника, имеем

$$|\alpha u + (1-\alpha)v - u_0| = |\alpha(u-u_0) + (1-\alpha)(v-u_0)| \leqslant$$
$$\leqslant \alpha|u-u_0| + (1-\alpha)|v-u_0| \leqslant \alpha R + (1-\alpha)R = R,$$

т. е. $u_\alpha = \alpha u + (1-\alpha)v \in S(u_0, R)$ для всех $\alpha \in [0, 1]$.

Пример 2. Напомним, что *гиперплоскостью* в E^n называется множество

$$\Gamma = \Gamma(c, \gamma) = \{u : u \in E^n, \langle c, u\rangle = \gamma\}, \tag{2}$$

где $c = (c_1, \ldots, c_n) \neq 0$ — вектор из E^n, γ — действительное число. Это множество всегда непусто: если, например, $c_i \neq 0$, то точка u_0 с координатами $u^i = \gamma/c_i$, $u^j = 0$ $(j = 1, \ldots, n,\ j \neq i)$ удовлетворяет равенству $\langle c, u_0\rangle = \gamma$, т. е.

$u_0 \in \Gamma$. Если u_0 — какая-либо точка из Γ, т. е. $\langle c, u_0\rangle = \gamma$, то гиперплоскость Γ можно представить в виде

$$\Gamma = \Gamma(c, u_0) = \{u : u \in E^n, \langle c, u - u_0\rangle = 0\}.$$

Напоминаем, что два вектора $a, b \in E^n$ называются *ортогональными*, если $\langle a, b\rangle = 0$. Предыдущее представление для Γ означает, что гиперплоскость состоит из тех и только тех точек u, для которых вектор $u - u_0$ ортогонален вектору c. Вектор c называют *нормальным* вектором гиперплоскости Γ.

Возьмем произвольные точки $u, v \in \Gamma$, т. е. $\langle c, u\rangle = \langle c, v\rangle = \gamma$. Тогда $\langle c, \alpha u + (1-\alpha)v\rangle = \alpha\langle c, u\rangle + (1-\alpha)\langle c, v\rangle = \gamma$ при всех $\alpha \in [0, 1]$. Следовательно, Γ — выпуклое множество.

Пример 3. Пусть $\Gamma = \{u : \langle c, u\rangle = \gamma\}$ — некоторая гиперплоскость. Тогда множества

$$\Gamma^+ = \{u : \langle c, u\rangle > \gamma\}, \quad \Gamma^- = \{u : \langle c, u\rangle < \gamma\}$$

называются *открытыми* полупространствами, а множества

$$\bar{\Gamma}^+ = \{u : \langle c, u\rangle \geqslant \gamma\}, \quad \bar{\Gamma}^- = \{u : \langle c, u\rangle \leqslant \gamma\}$$

называются *замкнутыми* полупространствами. Нетрудно видеть, что Γ^+, Γ^-, $\bar{\Gamma}^+$, $\bar{\Gamma}^-$ — выпуклые множества. Например, если $u, v \in \Gamma^+$, то $\langle c, \alpha u + (1-\alpha)v\rangle = \alpha\langle c, u\rangle + (1-\alpha)\langle c, v\rangle > \alpha\gamma + (1-\alpha)\gamma = \gamma$ для всех $\alpha \in [0, 1]$.

Пример 4. Прямая и луч в E^n (см. определение в § 2.1) — выпуклые множества.

Пример 5. Важным примером выпуклого множества являются аффинные множества (или линейные многообразия).

Определение 2. Множество M из E^n называется *аффинным*, если $\alpha u + (1-\alpha)v \in M$ при всех $u, v \in M$, $\alpha \in \mathbb{R}$, т. е. прямая, проходящая через любые две точки $u, v \in M$, целиком лежит в M.

Все пространство E^n является аффинным множеством. Пустое множество и множество, состоящее из одной точки, удобно считать аффинными. Любое подпространство пространства E^n представляет собой аффинное множество. Множество $M = L + u_0$, получаемое сдвигом подпространства L на произвольный фиксированный вектор u_0, также является аффинным.

Верно и обратное: всякое аффинное множество M может быть получено сдвигом некоторого подпространства L на некоторый вектор u_0. В самом деле, возьмем произвольную точку $u_0 \in M$ и положим $L = M - u_0$. Ясно, что L — аффинное множество, причем $0 \in L$. Тогда для каждого $u \in L$ имеем $\alpha u = \alpha u + (1-\alpha)\cdot 0 \in L$ при всех $\alpha \in \mathbb{R}$. Кроме того, если $u, v \in L$, то $(u+v)/2 = u/2 + (1-1/2)v \in L$ и, следовательно, $u + v = 2((u+v)/2) \in L$. Таким образом, сумма двух векторов из L и произведение вектора из L на любое число принадлежат L, т. е. L — подпространство.

Убедимся, что подпространство $L = M - u_0$ не зависит от выбора точки $u_0 \in M$. В самом деле, пусть $L_1 = M - u_1$, где $u_1 \in M$. Возьмем любую точку $u \in L$. Поскольку $u_1 - u_0 \in L$, то $u + (u_1 - u_0) \in L$ и, следовательно, $u \in L - (u_1 - u_0) = (L + u_0) - u_1 = M - u_1 = L_1$. Это значит, что $L \subseteq L_1$. Обратное включение $L_1 \subseteq L$ доказывается совершенно так же. Следовательно, $L_1 = L$.

Таким образом, всякое аффинное множество M из E^n представимо в виде

$$M = L + u_0, \tag{3}$$

где L — подпространство, однозначно определяемое множеством M, u_0 — произвольная точка из M. Подпространство из этого представления называют *параллельным* аффинному множеству M.

Опираясь на полученное представление (3), можно дать алгебраическое описание аффинных множеств на E^n. Поскольку всякое подпространство L из E^n представимо в виде $L=\{u\in E^n\colon\ Au=0\}=\{u\in E^n\colon\ \langle a_i, u\rangle=0, i=1,\ldots,m\}$, где A — некоторая матрица размера $m\times n$, a_i — i-я строка матрицы A (например, в качестве векторов a_i, $i=1,\ldots,m$, можно взять базис ортогонального дополнения к L в E^n). Отсюда и из (3) следует, что всякое аффинное множество из E^n может быть задано в виде

$$M=\{u\in E^n\colon Au=0\}+u_0=\{u\in E^n\colon u=u_0+v, Av=0\}=$$
$$=\{u\in E^n\colon A(u-u_0)=0\}=\{u\in E^n\colon Au=b\}=$$
$$=\{u\in E^n\colon \langle a_i, u\rangle=b_i, i=1,\ldots,m\}, \quad (4)$$

где $b=Au_0=(b^1,\ldots,b^m)$. Нетрудно проверить, что верно и обратное: всякое множество вида (4) является аффинным. В самом деле, если $u, v\in M$, т. е. $Au=b$, $Av=b$, то $A(\alpha u+(1-\alpha)v)=\alpha Au+(1-\alpha)Av=b$ или, иначе, $\alpha u+(1-\alpha)v\in M$ при всех $\alpha\in\mathbb{R}$. Таким образом, множества вида (4) и только они являются аффинными.

Согласно теореме Кронекера — Капелли [192; 351; 353] множество (4) непусто тогда и только тогда, когда матрица A и расширенная матрица $B=(A, b)$ имеют один и тот же ранг. Если $\operatorname{rang}A<\operatorname{rang}B$ (например, $A=0$, $b\neq 0$), то $M=\varnothing$. Если $A=0$, $b=0$, то $M=E^n$. Рассмотрим случай $A\neq 0$, $\operatorname{rang}A=\operatorname{rang}B=r$. Тогда множество (4) состоит из тех и только тех точек, которые представимы в виде [192; 351; 353]

$$u=u_0+\sum_{i=1}^{n-r} t_i u_i, \quad (5)$$

где u_0 — какое-либо частное решение неоднородной системы линейных алгебраических уравнений $Au=b$, а $u_1,\ldots,u_{n-r}$ — линейно независимые решения однородной системы $Au=0$; $t_1,\ldots,t_{n-r}$ — действительные числа. Векторы $u_1,\ldots,u_{n-r}$ образуют базис подпространства

$$L=\{u\in E^n\colon Au=0\}=\Big\{u\in E^n\colon u=\sum_{i=1}^{n-r} t_i u_i, t_i\in\mathbb{R}\Big\},$$

так что $\dim L$ — размерность L и равна $n-r$. С помощью введенного подпространства L равенство (5) можно переписать в виде $M=u_0+L$ — мы снова пришли к представлению (3).

Размерность аффинного множества M по определению принимается равной размерности подпространства L, параллельного M. Таким образом, размерность $\dim M$ аффинного множества (4) равна $n-$г, где $r=\operatorname{rang}A=\operatorname{rang}B$. Аффинное множество размерности p часто называют *гиперплоскостью* размерности p. В частности, если в (4), (5) $r=n$, то $L=\{0\}$ и $M=\{u_0\}$ состоят из одной точки и $\dim L=\dim M=0$. Если $r=n-1$, то $M_1=\{u\in E^n\colon u=u_0+tu_1, t\in\mathbb{R}\}$ — прямая (см. пример 4). Далее, гиперплоскость (2) также является аффинным множеством: в этом случае в (4) нужно принять $A=c$, $b=\gamma$. Поскольку $c\neq 0$, то $\operatorname{rang}A=\operatorname{rang}(A, b)=1$ и, следовательно, гиперплоскость имеет размерность $n-1$. Согласно (5) тогда $\Gamma=M_{n-1}=\Big\{u\in E^n\colon u=u_0+\sum_{i=1}^{n-1} t_i u_i, t_i\in\mathbb{R}\Big\}$, где $u_1,\ldots,u_{n-1}$ — базис параллельного подпространства $L_{n-1}=\{u\in E^n; \langle c, u\rangle=0\}$. Как видим, вектор c ортогонален к L_{n-1} и является базисом ортогонального дополнения к L_{n-1} до E^n, а векторы $u_1,\ldots,u_{n-1}, c$ образуют базис в E^n.

Заметим, что пересечение любого числа аффинных множеств само является аффинным множеством и, следовательно, представимо в виде (4).

Определение 3. Пересечение всех аффинных множеств, содержащих множество X из E^n, называется *аффинной оболочкой* множества X и обозначается через $\operatorname{aff}X$; подпространство L, параллельное $\operatorname{aff}X$, называется *несущим подпространством* множества X и обозначается через $\operatorname{Lin}X$.

Таким образом, $\operatorname{aff}X$ представляет собой минимальное аффинное множество, содержащее X. Пользуясь (3)–(5), нетрудно показать, что:

1) $\operatorname{aff}X=u_0+\operatorname{Lin}X$, где u_0 — произвольная точка из X;
2) если $0\in X$, то $\operatorname{aff}X=\operatorname{Lin}X$;
3) $u=v-w\in\operatorname{Lin}X$ для всех $v, w\in\operatorname{aff}X$, в частности, для $v, w\in X$;
4) если $d\in\operatorname{Lin}X$, $v\in\operatorname{aff}X$, то $u=v+\varepsilon d\in\operatorname{aff}X$ при всех $\varepsilon\in\mathbb{R}$.

Определение 4. *Размерностью произвольного множества* X из E^n называется размерность его аффинной оболочки; размерность множества X обозначают $\dim X$.

Согласно этому определению отрезок $[u, v]=\{u_\alpha=v+\alpha(u-v), 0\leqslant\alpha\leqslant 1\}$, соединяющий две точки $u, v\in E^n$ $(u\neq v)$, имеет размерность 1, так как его аффинной оболочкой является прямая $\Gamma=\{w\colon w=v+t(u-v), -\infty<t<\infty\}$. Размерность шара (1) равна n.

Пример 6. Предлагаем читателю самостоятельно доказать, что множество

$$X=\{x=(x_1, x_2)\colon x_1\in E^{n_1}, x_2\in E^{n_2}, A_{11}x_1+A_{12}x_2\leqslant b_1, A_{21}x_1+A_{22}x_2=b_2, x_1\geqslant 0\},$$

где A_{ij} — матрица размера $m_i\times n_j$, $c_j\in E^{n_j}$, $b_i\in E^{m_i}$, $i, j=1, 2$ (см. задачу (3.5.1), (3.5.2)), выпукло. Это множество называют *многогранным множеством* или *полиэдром*.

Пример 7. Множество $X=\{u=(u^1,\ldots,u^n)\colon \alpha_i\leqslant u^i\leqslant\beta_i, i=1,\ldots,n\}$, где α_i, β_i — заданные величины, $\alpha_i<\beta_i$ (возможно, что некоторые из $\alpha_i=\infty$ и некоторые из $\beta_j=\infty$), выпукло и имеет размерность n. В частности, *неотрицательный ортант* пространства E^n — это множество $E^n_+=\{u\colon u\in E^n, u\geqslant 0\}$ — выпукло, причем $\dim E^n_+=n$. Если в определении множества X величины α_i, β_i конечны при всех $i=1,\ldots,n$, то это множество называют *n-мерным параллелепипедом*.

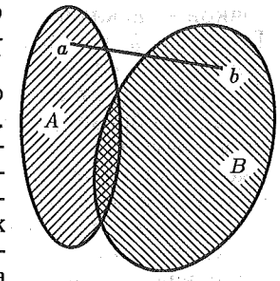

Рис. 4.2

2. Выше было отмечено, что пересечение любого числа выпуклых множеств выпукло. Нетрудно видеть, что объединение двух выпуклых множеств, вообще говоря, невыпукло (рис. 4.2). Посмотрим, как влияют на выпуклость другие операции над множествами: сложение, вычитание, умножение множества на число, замыкание и т. п.

Определение 5. *Суммой множеств* $A_1,\ldots,A_m$ называется множество $A=A_1+\ldots+A_m=\sum_{i=1}^{m} A_i$, состоящее из тех и только тех точек a, которые представимы в виде $a=\sum_{i=1}^{m} a_i$, где $a_i\in A_i$, $i=1,\ldots,m$. *Разностью двух множеств* A и B называется множество $C=A-B$, состоящее из тех и только тех точек c, которые представимы в виде $c=a-b$, $a\in A$,

$b \in B$. *Произведением множества A на действительное число λ* называется множество $B = \lambda A$, состоящее из всех точек вида $b = \lambda a$, $a \in A$.

Теорема 1. *Если $A_1, \ldots, A_m$, A, B — выпуклые множества, то множества $C = A_1 + \ldots + A_m$, $C = A - B$, $C = \lambda A$ выпуклы.*

Доказательство. Проведем его, например, для множества $C = A - B$. Пусть c_1, c_2 — произвольные точки из $C = A - B$. Это значит, что существуют $a_i \in A$, $b_i \in B$ такие, что $c_i = a_i - b_i$ $(i = 1, 2)$. Из выпуклости A и B следует, что $a_\alpha = \alpha a_1 + (1-\alpha)a_2 \in A$, $b_\alpha = \alpha b_1 + (1-\alpha)b_2 \in B$ при всех $\alpha \in [0, 1]$. Тогда $c_\alpha = \alpha c_1 + (1-\alpha)c_2 = \alpha(a_1 - b_1) + (1-\alpha)(a_2 - b_2) = a_\alpha - b_\alpha$, так что $c_\alpha \in C$ при всех $\alpha \in [0, 1]$. Выпуклость $C = A - B$ доказана. Аналогично доказывается выпуклость множеств $C = A_1 + \ldots + A_m$ и $C = \lambda A$. □

Определение 6. *Замыканием множества X* называется множество, являющееся объединением множества X и всех его предельных точек. Замыкание множества X будем обозначать через $\overline{X}$.

Для любой точки v и любого множества X из E^n имеет место одна и только одна из следующих трех возможностей.

1. Найдется ε-окрестность точки v, которая целиком принадлежит множеству X — тогда точка v называется *внутренней* точкой множества X. Совокупность всех внутренних точек множества X будем обозначать через $\operatorname{int} X$. Множество X, все точки которого являются внутренними, называют *открытым множеством*.

Примером открытого множества является открытое полупространство из примера 3.

2. Найдется ε-окрестность точки v, которая не содержит ни одной точки множества X — такая точка называется *внешней* по отношению ко множеству X.

3. Любая ε-окрестность точки v содержит как точки из X, так и точки из $E^n \setminus X$ — тогда точка v называется *граничной* точкой множества X. Совокупность всех граничных точек множества X будем обозначать через $\operatorname{Гр} X$.

Всякая внутренняя точка множества, очевидно, является его предельной точкой. Однако не всякая граничная точка множества будет его предельной точкой — исключение здесь составляют изолированные точки множества. Точку $v \in X$ называют *изолированной* точкой этого множества, если существует ε-окрестность этой точки, не содержащая ни одной точки множества X, отличной от v.

Таким образом, замыкание $\overline{X}$ множества X состоит, вообще говоря, из точек четырех типов: 1) внутренние точки множества X; 2) изолированные точки множества X; 3) предельные граничные точки множества X, принадлежащие X; 4) предельные граничные точки множества X, не принадлежащие X. Отсюда ясно, что замыкание любого множества является замкнутым множеством.

Очевидно, выпуклое множество, содержащее не менее двух точек, не может иметь изолированных точек.

Шар (1) замкнут и его замыкание состоит из внутренних точек $\operatorname{int} S(u_0, R) = \{u\colon |u - u_0| < R\}$ и граничных точек $\operatorname{Гр}(u_0, R) = \{u\colon |u - u_0| = R\}$. Множество $X = \{u = (x, y) \in E^2\colon 0 \leqslant x \leqslant 1, 0 \leqslant y \leqslant 1, x + y < 1\}$ выпукло, но не замкнуто — точки прямой $x + y = 1$ при $0 \leqslant x, y \leqslant 1$ являются предельными граничными для X, но не принадлежат X; $\overline{X} = \{u = (x, y)\colon 0 \leqslant x, y \leqslant 1, x + y \leqslant 1\}$. Множество $X = \{u = (x, y) \in E^2\colon -1 < x < 1, y = 0\}$ выпукло и не имеет внутренних точек; $\overline{X} = \{u = (x, y)\colon -1 \leqslant x \leqslant 1, y = 0\}$.

Для любого аффинного множества M из E^n имеем $\overline{M} = M$, так что M — замкнутое множество: это видно, например, из представления (4) аффинного множества; для множеств из примеров 6–8 также $\overline{X} = X$. В частности, $\operatorname{aff} X = \operatorname{aff} \overline{X}$, откуда будет следовать, что $\operatorname{aff} \overline{X} = \operatorname{aff} X$ для любого множества X из E^n.

Теорема 2. *Если A — выпуклое множество, то его замыкание тоже выпукло.*

Доказательство. Пусть a и b — произвольные точки множества $\overline{A}$. Поскольку выпуклое множество не имеет изолированных точек, то точки a и b будут предельными для A. Тогда существуют последовательности $\{a_k\}, \{b_k\} \in A$, сходящиеся соответственно к a, b. Возьмем произвольные $\alpha \in [0, 1]$. В силу выпуклости A тогда $c_k = \alpha a_k + (1 - \alpha)b_k \in A$. Отсюда при $k \to \infty$ получим $\lim\limits_{k \to \infty} c_k = c_\alpha = \alpha a + (1 - \alpha)b$. Таким образом, точка c_α является предельной для A и, следовательно, принадлежит $\overline{A}$ при любом $\alpha \in [0, 1]$. □

Теорема 3. *Пусть X — выпуклое множество и $\operatorname{int} X \neq \varnothing$. Пусть $u_0 \in \operatorname{int} X$, $v \in \overline{X}$. Тогда $v_\alpha = v + \alpha(u_0 - v) \in \operatorname{int} X$ при всех α, $0 < \alpha \leqslant 1$. Если $u \in \operatorname{int} X$, $y \notin \operatorname{int} X$, $y \in \overline{X}$, то $w_\lambda = u + \lambda(y - u) \notin \overline{X}$ при всех $\lambda > 1$.*

Доказательство. Поскольку точка $u_0 \in \operatorname{int} X$, то найдется ее δ-окрестность $O(u_0, \delta) = \{u\colon |u - u_0| < \delta\}$, целиком принадлежащая X. Сначала рассмотрим случай, когда $v \in X$. Возьмем произвольное α, $0 < \alpha < 1$.

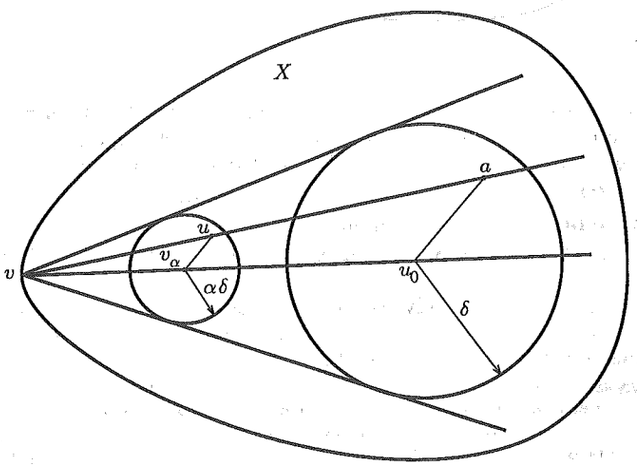

Рис. 4.3

Покажем, что окрестность $O(v_\alpha, \alpha\delta) = \{u\colon |u - u_\alpha| < \alpha\delta\}$ точки v_α принадлежит X. С этой целью возьмем произвольную точку $u \in O(v_\alpha, \alpha\delta)$ и положим $a = u_0 + (u - v_\alpha)/\alpha$ (рис. 4.3). Поскольку $|a - u_0| = |u - v_\alpha|/\alpha < \delta\alpha/\alpha = \delta$, то $a \in O(u_0, \delta) \subset X$. Из определения точки a имеем представление $u = v_\alpha + \alpha(a - u_0) = v + \alpha(u_0 - v) + \alpha(a - u_0) = \alpha a + (1 - \alpha)v$, где $a, v \in X$ и $0 < \alpha < 1$. Тогда $u \in X$ в силу выпуклости X. Тем самым показано, что произвольная точка u из $O(v_\alpha, \alpha\delta)$ принадлежит X. Следовательно, $O(v_\alpha, \alpha\delta) \subset X$, т. е. v_α — внутренняя точка X.

Пусть теперь $v \in \overline{X} \setminus X$. Поскольку v — предельная точка X, то найдется точка $w \in X$ такая, что $|v-w| < \alpha(1-\alpha)^{-1}\delta$. Возьмем точку $w_\alpha = w + \alpha(u_0 - w)$ (рис. 4.4). В силу только что доказанного точка w_α принадлежит множеству X вместе со своей окрестностью $O(w_\alpha, \alpha\delta)$. Но $|v_\alpha - w_\alpha| = |v + \alpha(u_0 - v) - w - \alpha(u_0 - w)| = (1-\alpha)|v-w| < (1-\alpha)\alpha(1-\alpha)^{-1}\delta = \alpha\delta$. Следовательно, $v_\alpha \in O(w_\alpha, \alpha\delta) \subset X$. Убедимся, что окрестность $O(v_\alpha, \beta)$ ($\beta = \delta\alpha - |v_\alpha - w_\alpha|$) точки v_α также принадлежит X. В самом деле, если $u \in O(v_\alpha, \beta)$, то $|u - w_\alpha| \leqslant |u - v_\alpha| + |v_\alpha - w_\alpha| < \beta + |v_\alpha - w_\alpha| = \delta\alpha$, т. е. $O(v_\alpha, \beta) \in O(w_\alpha, \alpha\delta) \in X$.

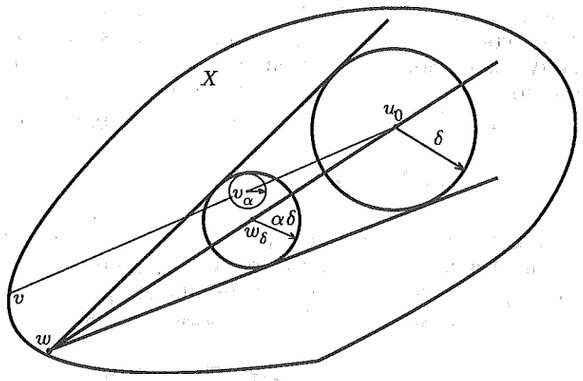

Рис. 4.4

Наконец, пусть $w_\lambda = u + \lambda(y-u)$ ($\lambda > 1$), где $u \in \operatorname{int} X$, $y \notin \operatorname{int} X$, $y \in \overline{X}$. Допустим, что $w_\lambda \in \overline{X}$ при каком-либо $\lambda > 1$. Из представления для w_λ имеем $y = u + (w_\lambda - u)/\lambda = w_\lambda + (1 - 1/\lambda)(u - w_\lambda) = w_\lambda + \alpha(u - w_\lambda)$, где $\alpha = 1 - 1/\lambda \in (0,1)$, $w_\lambda \in \overline{X}$, $u \in \operatorname{int} X$. По доказанному выше тогда $y \in \operatorname{int} X$, что противоречит условию. Следовательно, $w_\lambda \notin \overline{X}$ при всех $\lambda > 1$. □

Теорема 4. *Если X — выпуклое множество, то* int X *тоже выпукло.*

Доказательство. Пусть u, v — произвольные точки из int X. В теореме 3 было показано, что $u_\alpha = \alpha u + (1-\alpha)v \in \operatorname{int} X$ при всех α, $0 < \alpha < 1$. Это и означает выпуклость int X. □

3. В тех случаях, когда рассматриваемое множество X невыпукло, часто бывает полезно расширить его до выпуклого множества. Посмотрим, как это делается.

Определение 7. Точка u называется *выпуклой комбинацией* точек $u_1, \ldots, u_m$, если существуют числа $\alpha_1 \geqslant 0, \ldots, \alpha_m \geqslant 0$, $\alpha_1 + \ldots + \alpha_m = 1$ такие, что $u = \alpha_1 u_1 + \ldots + \alpha_m u_m$.

Теорема 5. *Множество выпукло тогда и только тогда, когда оно содержит все выпуклые комбинации любого конечного числа своих точек.*

Доказательство. Необходимость. Пусть X — выпуклое множество. Тогда по определению 1 множество X содержит выпуклые комбинации любых двух своих точек. Сделаем индуктивное предположение: пусть множество X содержит выпуклые комбинации любых $m-1$ своих точек. Рассмотрим выпуклую комбинацию $\alpha_1 u_1 + \ldots + \alpha_m u_m$ произвольных m точек из X. Можем считать, что $\alpha_i > 0$, $i = 1, \ldots, m$. Поскольку $\alpha_1 + \ldots + \alpha_m = 1$, то $0 < \alpha_i < 1$, $i = 1, \ldots, m$. Следовательно, точка $v = \sum_{i=1}^{m-1} \alpha_i (1-\alpha_m)^{-1} u_i$ является выпуклой комбинацией точек $u_1, \ldots, u_{m-1}$ и по предположению индукции принадлежит X. Однако $u = (1-\alpha_m) \sum_{i=1}^{m-1} \alpha_i (1-\alpha_m)^{-1} u_i + \alpha_m u_m = (1-\alpha_m)v + \alpha_m u_m \in X$ в силу выпуклости X.

Достаточность. Если множество X содержит все выпуклые комбинации любого конечного числа своих точек, то оно содержит, в частности, выпуклые комбинации любых двух своих точек и, следовательно, выпукло. □

Определение 8. Пересечение всех выпуклых множеств, содержащих множество X, называется *выпуклой оболочкой* множества X и обозначается через co X.

Ясно, что co X, как пересечение выпуклых множеств, является выпуклым множеством. Кроме того, co X содержится в любом выпуклом множестве, содержащем X. Так что co X — это минимальное выпуклое множество, содержащее X.

Теорема 6. *Выпуклая оболочка множества X состоит из тех и только тех точек, которые являются выпуклой комбинацией конечного числа точек из X. Если X — выпукло, то* co $X = X$.

Доказательство. Пусть W — множество всех точек, являющихся выпуклыми комбинациями любого конечного числа точек из X. Нам надо показать, что co $X = W$. Поскольку $X \subseteq$ co X и co X — выпуклое множество, то по теореме 5 co X содержит все выпуклые комбинации точек из co X и, в частности, точек из X. Следовательно, $W \subseteq$ co X.

Покажем, что W — выпуклое множество. В самом деле, пусть $u, v \in W$, т. е. $u = \alpha_1 u_1 + \ldots + \alpha_m u_m$, $u_i \in X$, $\alpha_i \geqslant 0$, $i = 1, \ldots, m$, $\alpha_1 + \ldots + \alpha_m = 1$, $v = \beta_1 v_1 + \ldots + \beta_p v_p$, $v_i \in X$, $\beta_i \geqslant 0$, $i = 1, \ldots, p$, $\beta_1 + \ldots + \beta_p = 1$. Тогда $u_\alpha = \alpha u + (1-\alpha)v = \sum_{i=1}^{m} \alpha\alpha_i u_i + \sum_{j=1}^{p} (1-\alpha)\beta_j v_j$, где $\alpha\alpha_i \geqslant 0$, $(1-\alpha)\beta_j \geqslant 0$, $\sum_{i=1}^{m} \alpha\alpha_i + \sum_{j=1}^{p} (1-\alpha)\beta_j = 1$ для каждого $\alpha \in [0,1]$. Следовательно, u_α является выпуклой комбинацией точек $u_1, \ldots, u_m$, $v_1, \ldots, v_p \in X$ и принадлежит W при всех $\alpha \in [0,1]$. Таким образом, W — выпуклое множество, содержащее X. Но co X по своему определению принадлежит всем выпуклым множествам, содержащим X, и поэтому co $X \subseteq W$. Сравнивая с ранее доказанным включением $W \subseteq$ co X, заключаем, что co $X = W$. Если X — выпуклое множество, то с учетом теоремы 5 имеем $X = W =$ co X. Теорема 6 доказана. □

Заметим, что выпуклая оболочка двух точек на плоскости представляет собой отрезок, выпуклая оболочка трех точек, не лежащих на одной прямой, — треугольник. В общем случае выпуклая оболочка конечного числа точек на плоскости образует выпуклый многоугольник, а в пространстве — выпуклый многогранник.

Определение 9. Выпуклая оболочка множества точек $u_0, u_1, \ldots, u_m$ из E^n таких, что система векторов $\{u_i - u_0, i = 1, \ldots, m\}$ линейно независима, называется *симплексом*, натянутым на эти точки, и обозначается через $S_m = S_m(u_0, u_1, \ldots, u_m)$. Точки $u_0, u_1, \ldots, u_m$ называются *вершинами симплекса*.

В случае $m = 0, 1, 2, 3$ симплекс представляет собой соответственно точку, отрезок, треугольник, тетраэдр.

Согласно теореме 6 симплекс S_m представим в виде

$$S_m = \Big\{u: u = \sum_{i=0}^{m} \alpha_i u_i, \alpha_i \geqslant 0, i = 0, \ldots, m, \sum_{i=0}^{m} \alpha_i = 1\Big\}.$$

По теореме 6 любая точка выпуклой оболочки множества X является выпуклой комбинацией конечного, но, быть может, довольно большого числа точек из X. Замечательно, однако, то, что в E^n для получения множества co X достаточно ограничиться рассмотрением выпуклых комбинаций не более чем $n+1$ точек из X. Точнее, верна

Теорема 7 (Каратеодори). *Пусть X — произвольное непустое множество из E^n. Тогда любая точка $u \in$ co X представима в виде выпуклой комбинации не более чем $n+1$ точек из X.*

Доказательство. Согласно теореме 6 любая точка $u \in$ co X представима в виде $u = \alpha_1 u_1 + \ldots + \alpha_m u_m$, где $u_i \in X$, $\alpha_i \geqslant 0$, $i = 1, \ldots, m$, $\alpha_1 + \ldots + \alpha_m = 1$. Пусть $m > n+1$ и все $\alpha_i > 0$ (если $\alpha_i = 0$, то число m может быть уменьшено). В $n+1$-мерном пространстве рассмотрим векторы $\bar{u}_i = (u_i, 1)$, $i = 1, \ldots, m$. Поскольку $m > n+1$, то эти векторы линейно зависимы, т. е. существуют числа $\gamma_1, \ldots, \gamma_m$, не все равные нулю и такие, что $\gamma_1 \bar{u}_1 + \ldots + \gamma_m \bar{u}_m = 0$. Это равенство эквивалентно следующим двум равенствам:

$$\gamma_1 u_1 + \ldots + \gamma_m u_m = 0, \quad \gamma_1 + \ldots + \gamma_m = 0.$$

Тогда точка u представима другими выпуклыми комбинациями тех же точек $u_1, \ldots, u_m$: $\sum_{i=1}^{m} (\alpha_i - t\gamma_i) u_i = \sum_{i=1}^{m} \alpha_i u_i - \sum_{i=1}^{m} \gamma_i u_i = u$. В самом деле, здесь $\alpha_i > 0$ и, следовательно, $\alpha_i - t\gamma_i \geqslant 0$, $i = 1, \ldots, m$ при всех достаточно малых t и, кроме того, $\sum_{i=1}^{m} (\alpha_i - t\gamma_i) = \sum_{i=1}^{m} \alpha_i - t \sum_{i=1}^{m} \gamma_i = 1$. Поскольку не все γ_i равны нулю, но $\gamma_1 + \ldots + \gamma_m = 0$, то среди $\{\gamma_i\}$ найдутся положительные.

Пусть $\alpha_s\gamma_s^{-1}=\min\limits_{\gamma_i>0}\alpha_i\gamma_i^{-1}$. Положим $t=\alpha_s\gamma_s^{-1}$. При таком выборе t все $\alpha_i-t\gamma_i$ останутся неотрицательными, причем $\alpha_s-t\gamma_s=0$. Это значит, что точку u удалось представить в виде выпуклой комбинации меньшего числа точек $u_1,\ldots,u_{s-1},u_{s+1},\ldots,u_m$. Ясно, что последовательно применяя описанный прием далее, число точек, участвующих в выпуклой комбинации, можно уменьшить до $n+1$. □

Теорема 8. *Если X — замкнутое ограниченное множество из E^n, то* co X *замкнуто и ограничено.*

Доказательство. По условию существует число $R>0$ такое, что $|u|\leqslant R$ для всех $u\in X$, т. е. $X\subseteq S(0,R)$ — шар радиуса R с центром в точке 0. Но шар — выпуклое множество. Согласно определению 8 тогда co $X\subseteq S(0,R)$, так что $|u|\leqslant R$ для всех $u\in\operatorname{co}X$. Ограниченность co X доказана.

Докажем замкнутость co X. Пусть u — предельная точка co X, $\{u_k\}\in\operatorname{co}X$ и $\lim\limits_{k\to\infty}u_k=u$. Согласно теореме 7 существуют точки $u_{ki}\in X$, числа $\alpha_{ki}\geqslant 0$, $i=1,\ldots,n+1$, $\alpha_{k1}+\ldots+\alpha_{k,n+1}=1$ такие, что $u_k=\alpha_{k1}u_{k1}+\ldots+\alpha_{k,n+1}u_{k,n+1}$. Заметим, что $|u_{ki}|\leqslant R$, $0\leqslant\alpha_{ki}\leqslant 1$ при всех $i=1,\ldots,n+1$ и всех $k=1,2,\ldots$ Пользуясь теоремой Больцано — Вейерштрасса, сначала из $\{u_{k1}\}$, $\{\alpha_{k1}\}$ выберем подпоследовательности $\{u_{k_1}\}$, $\{\alpha_{k_1}\}$, сходящиеся соответственно к некоторым u_1, α_1; затем из $\{u_{k_12}\}$, $\{\alpha_{k_12}\}$ — подпоследовательности $\{u_{k_22}\}\to u_2$, $\{\alpha_{k_22}\}\to\alpha_2$ и т. д., наконец, $\{u_{k_{n+1}n+1}\}\to u_{n+1}$, $\{\alpha_{k_{n+1}n+1}\}\to\alpha_{n+1}$. Тогда из $u_{k_{n+1}}=\sum\limits_{i=1}^{n+1}\alpha_{k_{n+1}}u_{k_{n+1}i}$, $u_{k_{n+1}i}\in X$, $\alpha_{k_{n+1}i}\geqslant 0$, $i=1,\ldots,n+1$, $\sum\limits_{i=1}^{n+1}\alpha_{k_{n+1}i}=1$, предельным переходом при $k_{n+1}\to\infty$ получим $u=\sum\limits_{i=1}^{n+1}\alpha_iu_i$, $\alpha_i\geqslant 0$, $\sum\limits_{i=1}^{n+1}\alpha_i=1$, где $u_i\in X$, $i=1,\ldots,n+1$, в силу замкнутости X. Следовательно, по теореме 6 $u\in\operatorname{co}X$, что и требовалось. □

Заметим, что требование ограниченности множества X в теореме 8 существенно. Например, множество $X=\{u=(x,y)\in E^2\colon x\geqslant 0, y=\sqrt{x}\}$ замкнуто, но co $X=\{u=(x,y)\in E^2\colon x>0, 0<y\leqslant\sqrt{x}\}\cup\{(0,0)\}$ незамкнуто. Если X — выпуклое, замкнутое множество из E^n, то co X будет замкнутым и без требования ограниченности X, поскольку в этом случае в силу теорем 5, 6 co $X=X=\overline{X}=\overline{\operatorname{co}X}$.

Теорема 9. *Пусть A — произвольное непустое множество из E^n. Тогда*

$$\sup_{a\in A}\langle c,a\rangle=\sup_{a\in\bar{A}}\langle c,a\rangle=\sup_{a\in\operatorname{co}A}\langle c,a\rangle=\sup_{a\in\overline{\operatorname{co}A}}\langle c,a\rangle\quad\forall c\in E^n.$$

Доказательство. Поскольку $A\subset\bar{A}$, то $\sup\limits_{A}\langle c,a\rangle\leqslant\sup\limits_{\bar{A}}\langle c,a\rangle$. С другой стороны, для любого $\bar{a}\in\bar{A}$ существуют $a_k\in A$, $\{a_k\}\to\bar{a}$. Поэтому из $\langle c,a_k\rangle\leqslant\sup\limits_{A}\langle c,a\rangle$ при $k\to\infty$ получим $\langle c,\bar{a}\rangle\leqslant\sup\limits_{A}\langle c,a\rangle$ для всех $\bar{a}\in\bar{A}$. Следовательно, $\sup\limits_{\bar{A}}\langle c,a\rangle\leqslant\sup\limits_{A}\langle c,a\rangle$, так что $\sup\limits_{\bar{A}}\langle c,a\rangle=\sup\limits_{A}\langle c,a\rangle$. Отсюда же имеем $\sup\limits_{\operatorname{co}A}=\sup\limits_{\overline{\operatorname{co}A}}\langle c,a\rangle$. Далее, так как $A\subset\operatorname{co}A$, то $\sup\limits_{A}\langle c,a\rangle\leqslant\sup\limits_{\operatorname{co}A}\langle c,a\rangle$. С другой стороны, для любого $\bar{a}\in\operatorname{co}A$ согласно теореме 6 найдутся $a_i\in A$, $\alpha_i\geqslant 0$, $i=1,\ldots,r$, $\alpha_1+\ldots+\alpha_r=1$ такие, что $\bar{a}=\sum\limits_{i=1}^{r}\alpha_ia_i$. Поэтому

$$\langle c,\bar{a}\rangle=\sum_{i=1}^{r}\alpha_i\langle c,a_i\rangle\leqslant\sum_{i=1}^{r}\alpha_i\sup_{a\in A}\langle c,a\rangle=\sup_{A}\langle c,a\rangle$$

для каждого $\bar{a}\in\operatorname{co}A$. Следовательно, $\sup\limits_{\operatorname{co}A}\langle c,a\rangle\leqslant\sup\limits_{A}\langle c,a\rangle$, так что $\sup\limits_{\operatorname{co}A}\langle c,a\rangle=\sup\limits_{A}\langle c,a\rangle$. □

4. Приведем условия существования внутренней точки выпуклого множества.

Теорема 10. *Пусть X — непустое выпуклое множество из E^n. Для того чтобы* int $X\neq\varnothing$, *необходимо и достаточно, чтобы* dim $X=n$.

Доказательство. Необходимость. Пусть int $X\neq\varnothing$. Тогда для любой внутренней точки v множества X существует ε-окрестность $O(v,\varepsilon)=\{u\colon|u-v|<\varepsilon\}$, также принадлежащая X. Отсюда вытекает, что минимальным аффинным множеством, содержащим множество X и, следовательно, шар $O(v,\varepsilon)$, является все пространство E^n. Это значит, что aff $X=E^n$, dim $X=n$.

Достаточность. Пусть dim $X=n$. Тогда aff $X=E^n$ и найдутся точки $u_0,u_1,\ldots,u_n\in X$ такие, что векторы $u_1-u_0,\ldots,u_n-u_0$ — линейно независимы. Натянем на эти точки симплекс

$$S_n=\Big\{u\in E^n\colon u=\sum_{i=0}^{n}\alpha_iu_i,\ \alpha_i\geqslant 0,\ i=0,\ldots,n,\ \sum_{i=0}^{n}\alpha_i=1\Big\}.$$

Согласно теореме 5 $S_n\subset X$, а по теореме 6 S_n выпукло.

Рассмотрим систему линейных алгебраических уравнений

$$\sum_{i=1}^{n}(u_i-u_0)x_i=u-u_0,\qquad u\in E^n.\tag{6}$$

Матрица этой системы $A=(u_1-u_0,\ldots,u_n-u_0)$, столбцами которой являются линейно независимые векторы u_i-u_0, $i=1,\ldots,n$, имеет размеры $n\times n$ и невырождена. Поэтому система (6) для каждого $u\in E^n$ имеет и притом единственное решение $x=x(u)=(x_1(u),\ldots,x_n(u))$. Из известных формул Крамера [192, 353] видно, что функции $x_i(u)$ непрерывно (и даже линейно) зависят от u.

Опираясь на это свойство решений системы (6), покажем, что любая точка $w=\sum\limits_{i=0}^{n}\alpha_iu_i$ симплекса S_n при $\alpha_i>0$, $i=0,\ldots,n$ является внутренней точкой S_n. В самом деле, для такой точки w имеем $w-u_0=\sum\limits_{i=0}^{n}\alpha_iu_i-\sum\limits_{i=0}^{n}\alpha_iu_0=\sum\limits_{i=0}^{n}\alpha_i(u_i-u_0)$. Сравнение с (6) показывает, что $\alpha_i=x_i(w)$, $i=1,\ldots,n$. В силу непрерывности $x_i(u)$ тогда $x_i(u)>0$, $i=1,\ldots,n$ для всех $u\in O(w,\varepsilon)=\{u\colon|u-w|<\varepsilon\}$, где $\varepsilon>0$ — достаточно малое число. Функция $x_0(u)=1-\sum\limits_{i=1}^{n}x_i(u)$ также непрерывна, причем $x_0(w)=1-\sum\limits_{i=1}^{n}\alpha_i=\alpha_0>0$. Взяв $\varepsilon>0$ достаточно малым, можем считать, что $x_0(u)>0$ для всех $u\in O(w,\varepsilon)$.

Таким образом, для каждой точки $u\in O(w,\varepsilon)$ в силу (6) имеем представление $u=u_0+\sum\limits_{i=1}^{n}x_i(u)(u_i-u_0)=\sum\limits_{i=0}^{n}x_i(u)u_i$, где $x_i(u)>0$, $\sum\limits_{i=0}^{n}x_i(u)=1$. Это означает, что $O(w,\varepsilon)\subset S_n$. Следовательно, $w\in\operatorname{int}S_n$. Отсюда и из включения $S_n\subset X$ следует, что $O(w,\varepsilon)\subset X$, т. е. $w\in\operatorname{int}X$ и int $X\neq\varnothing$. □

5. Выпуклое множество $X=\{u=(x,y,z)\in E^3\colon x^2+y^2\leqslant 1, z=0\}$, представляющее собой единичный круг в плоскости $\Gamma=\{u=(x,y,z)\in E^3\colon z=0\}$, не имеет внутренних точек. Кстати, здесь плоскость Γ представляет собой аффинную оболочку множества X. В то же время, если это множество X рассматривать лишь относительно плоскости Γ (т. е не «признавать» точки E^3, лежащие вне Γ), то X — единичный круг — конечно же, имеет внутренние точки. Приводимая ниже теорема 11 показывает, что это не случайно. Для ее формулировки нам понадобится

Определение 10. Точка $v\in X$ называется *относительно внутренней точкой* множества X, если существует ε-окрестность $O(v,\varepsilon)=\{u\in E^n\colon|u-v|<\varepsilon\}$ точки v такая, что пересечение $O(v,\varepsilon)\cap\operatorname{aff}X$ целиком принадлежит X. Множество всех относительно внутренних точек множества X обозначается через ri X (иногда обозначают relint X).

Например, если $X=\{u=(x,y,z)\in E^3\colon x^2+y^2\leqslant 1, z=0\}$, то ri $X=\{u=(x,y,z)\in E^3\colon x^2+y^2<1, z=0\}$.

Если множество $X\subseteq E^n$ имеет размерность n, т. е. dim aff $X=n$, то понятия внутренней и относительно внутренней точки для множества X совпадают и ri $X=$int X. Нетрудно указать множества X (например, множество, состоящие из двух различных точек E^n), у которых ri $X=\varnothing$. Однако для выпуклых множеств верна

Теорема 11. *Если X непустое выпуклое множество из E^n, то* ri X *непусто, выпукло. При этом если $u_0\in$ ri X, $v\in\overline{X}$, то $v_\alpha=v+\alpha(u_0-v)\in$ ri X при всех α, $0<\alpha\leqslant 1$. Если $u\in$ ri X, $y\notin$ ri X, $y\in\overline{X}$, то $w_\lambda=u+\lambda(y-u)\notin\overline{X}$ при всех $\lambda>1$.*

Доказательство. Можем считать, что точка $0\in X$, так как в противном случае вместо множества X мы рассмотрели бы множество $X-\{v\}=\{w\in E^n; w=u-v, u\in X\}$, где v — какая-либо точка из X. Тогда $0\in\operatorname{aff}X=\operatorname{Lin}X$ — подпространство в E^n. Пусть множество X имеет размерность dim $X=m$, $1\leqslant m<n$, (в случае $m=0$, когда X состоит из единственной точки, утверждения теоремы тривиальны; случай $m=n$ рассмотрен в теоремах 3, 4, 10). Тогда найдутся такие точки $u_0,u_1,\ldots,u_m\in X$, что векторы $e_1=u_1-u_0,\ldots,e_m=u_m-u_0$ линейно независимы и образуют базис aff X. Можно дополнить систему $e_1,\ldots,e_m$ векторами $e_{m+1},\ldots,e_n$ до базиса E^n, причем можно считать, что $\langle e_i,e_j\rangle=0$, $i=1,\ldots,m$, $j=m+1,\ldots,n$. В этом базисе aff $X=\{u=(u^1,\ldots,u^n)\in E^n\colon u^{m+1}=\langle e_{m+1},u\rangle=0,\ldots,u^n=\langle e_n,u\rangle=0\}=\{u=(x,u^{m+1}=0,\ldots,u^n=0)\colon x\in E^m\}$, так что aff X можно отождествить с пространством

E^n. Повторив в этом пространстве соответствующие рассуждения из доказательств теорем 3, 4, 10, убеждаемся в справедливости утверждений доказываемой теоремы. □

Теорема 12. *Пусть X — непустое выпуклое множество из E^n и $y \in \overline{X}$, $y \notin \operatorname{ri} X$. Тогда существует последовательность $\{y_k\}$, $y_k \notin \overline{X}$, $y_k \in \operatorname{aff} \overline{X}$, $k = 1, 2, \ldots$, сходящаяся к y.*

Доказательство. Возьмем какую-либо точку $u \in \operatorname{ri} X$. Согласно теореме 11 тогда $w_\lambda = u + \lambda(y-u) \notin \overline{X}$ при всех $\lambda > 1$. Кроме того, поскольку $\operatorname{aff} \overline{X} = \operatorname{aff} X$, то $u, y \in \operatorname{aff} X$ и, следовательно, $w_\lambda \in \operatorname{aff} X$. Тогда $y_k = w_{\lambda_k} = u + \lambda_k(y-u)$, где $\lambda_k > 1$, $\{\lambda_k\} \to 1$, — искомая последовательность. □

Заметим, что если множество X не является выпуклым, то утверждение теоремы 12 может оказаться неверным. Например, пусть X — множество точек на числовой оси E^1, имеющих рациональные координаты. Тогда $\operatorname{aff} X = E^1$, и любая точка $y \in E^1$ является граничной для X. Таким образом, $X = E^1$, и последовательности $\{y_k\} \notin \overline{X}$, сходящейся к y здесь не существует.

Теорема 13. *Пусть X — выпуклое множество из E^n. Тогда* $\operatorname{ri} X = \operatorname{ri} \overline{X}$, $\overline{X} = \overline{\operatorname{ri} X}$.

Доказательство. Возьмем любую точку $v \in \operatorname{ri} X$. Согласно определению 10 тогда существует такое $\varepsilon > 0$, что $O(v, \varepsilon) \cap \operatorname{aff} X = O(v, \varepsilon) \cap \operatorname{aff} \overline{X} \subset X \subset \overline{X}$. Это значит, что $v \in \operatorname{ri} \overline{X}$. Следовательно, $\operatorname{ri} X \subseteq \operatorname{ri} \overline{X}$. Докажем обратное включение. Возьмем $w \in \operatorname{ri} \overline{X}$. Тогда существует такое $\varepsilon > 0$, что $O(w, \varepsilon) \cap \operatorname{aff} \overline{X} \subset \overline{X}$.

Возьмем какую-либо точку $v \in \operatorname{ri} X$ и положим $w_\lambda = w + \lambda(w - v)$, $\lambda \in \mathbb{R}$. Поскольку $v, w \in \operatorname{aff} \overline{X} = \operatorname{aff} X$, то $w_\lambda \in \operatorname{aff} \overline{X}$ при всех $\lambda \in \mathbb{R}$. Кроме того, существует такое $\lambda_0 > 0$, что $w_\lambda \in O(w, \varepsilon)$ для всех λ, $|\lambda| \leqslant \lambda_0$. Следовательно, $w_\lambda \in O(w, \varepsilon) \cap \operatorname{aff} \overline{X} \subset \overline{X}$, $|\lambda| \leqslant \lambda_0$. Из выражения для w_λ следует, что $w = w_\lambda + \frac{\lambda}{1+\lambda}(v - w_\lambda)$. При $0 < \lambda < \lambda_0$ имеем $\alpha = \frac{\lambda}{1+\lambda} \in (0, 1)$. Согласно теореме 11 тогда $w \in \operatorname{ri} X$. Это значит, что $\operatorname{ri} \overline{X} \subseteq \operatorname{ri} X$. Тем самым установлено, что $\operatorname{ri} \overline{X} = \operatorname{ri} X$.

Далее, так как $\operatorname{ri} X \subset X$, то $\overline{\operatorname{ri} X} \subseteq \overline{X}$. Возьмем любую точку $u \in \overline{X}$ и $v \in \operatorname{ri} X$. По теореме 11 $u_\alpha = u + \alpha(v-u) \in \operatorname{ri} X$ при всех $\alpha \in (0, 1]$, причем $u_\alpha \to u$ при $\alpha \to 0$. Следовательно, $u \in \overline{\operatorname{ri} X}$. Это значит, что $\overline{X} \subseteq \overline{\operatorname{ri} X}$. Таким образом, показано, что $\overline{X} = \overline{\operatorname{ri} X}$. □

Некоторые другие свойства выпуклых множеств будут рассмотрены ниже.

Упражнения

1. Пусть X — некоторое множество из E^n, $\overline{X}$ — замыкание множества X. Если $\overline{X}$ выпукло, то можно ли утверждать, что X также выпукло?

2. Существует ли невыпуклое множество, удалив из которого одну точку (или несколько точек), можно получить выпуклое множество? Рассмотреть пример $X = \{u = (x, y) \in E^2 : x \geqslant 0, y \geqslant 0, x + y < 1\} \cup \{(0, 1)\} \cup \{(1, 0)\}$.

3. Показать, что равенство $\operatorname{ri} X = \operatorname{ri} \overline{X}$ для невыпуклых множеств, вообще говоря, неверно (рассмотреть круг с выколотым центром).

4. Доказать, что если $A \subset B$, то $\overline{A} \subset \overline{B}$, $\operatorname{int} A \subset \operatorname{int} B$, но, вообще говоря, не будет включения $\operatorname{ri} A \subset \operatorname{ri} B$ даже для выпуклых A и B. Рассмотреть пример B — куб в E^3, A — одна из его граней.

5. Если A — выпуклое множество из E^n, то $\operatorname{aff} A = \operatorname{aff}(\operatorname{ri} A)$. Доказать.

6. Доказать, что размерность выпуклого множества X совпадает с максимальной размерностью симплексов, содержащихся в X.

7. Если A, B — выпуклые множества из E^n, то $\overline{A} + \overline{B} \subset \overline{A + B}$, $\operatorname{ri} A + \operatorname{ri} B = \operatorname{ri}(A + B)$, $\operatorname{ri}(\lambda A) = \lambda \operatorname{ri} A$ для любых действительных чисел λ. Доказать.

8. Доказать, что если A, B — выпуклые множества из E^n, $\operatorname{ri} A \cap \operatorname{ri} B \neq \varnothing$, то $\operatorname{ri} A \cap \operatorname{ri} B = \operatorname{ri}(A \cap B)$. Существенно ли здесь требование $\operatorname{ri} A \cap \operatorname{ri} B \neq \varnothing$? Рассмотреть пример $A = \{u = (x, y) \in E^2 : x \geqslant 0\}$, $B = \{u = (x, y) \in E^2 : x \leqslant 0\}$.

9. Если X — открытое множество, то $\operatorname{co} X$ открыто. Доказать.

10. Доказать, что $\operatorname{co}(A + B) = \operatorname{co} A + \operatorname{co} B$.

11. Доказать, что $\overline{\operatorname{co}\, X} = \overline{\operatorname{co}} X$, где $\overline{\operatorname{co}} X$ — пересечение всех выпуклых замкнутых множеств, содержащих X.

12. Доказать, что вершины $u_0, u_1, \ldots, u_m$ m-мерного симплекса $S_m = S_m(u_0, u_1, \ldots, u_m)$ являются его угловыми точками (см. определения 9, 3.2.1).

13. Доказать, что аффинная оболочка множества X состоит из точек вида $\alpha_1 u_1 + \ldots + \alpha_m u_m$ при всевозможных $u_1, \ldots, u_m \in X$, α_i — действительные числа, $i = 1, \ldots, m$, $\alpha_1 + \ldots + \alpha_m = 1$, и только из них.

14. Пусть A, B — выпуклые замкнутые множества из E^n, причем хотя бы одно из них ограничено. Доказать, что тогда $A + B$ — выпуклое замкнутое множество. Будет ли $A + B$ замкнутым, если A, B не ограничены? Рассмотреть примеры:

а) $A = \{a = (x, y) \in E^2 : y = 0\}$, $B = \{b = (x, y) \in E^2 : y \geqslant e^x\}$;

б) множества A, B из рис. 4.14;

в) $A = \{a = (x, y, z) \in E^3 : x = z = 0, y \leqslant 0\}$, $B = \{b = (x, y, z) \in E^3 : x^2 + y^2 \leqslant 2yz, y \geqslant 0\}$.

§ 2. Выпуклые функции

1. В главе 1 были рассмотрены некоторые свойства выпуклых функций одной переменной. Здесь мы продолжим изучение свойств выпуклых функций многих переменных.

Определение 1. Функция $f(x)$, определенная на выпуклом множестве X, называется *выпуклой* на этом множестве, если

$$f(\alpha u + (1-\alpha)v) \leqslant \alpha f(u) + (1-\alpha) f(v) \tag{1}$$

при всех $u, v \in X$, всех α, $0 \leqslant \alpha \leqslant 1$. Если в (1) при $u \neq v$ равенство возможно только при $\alpha = 0$ и $\alpha = 1$, то функция $f(x)$ называется *строго выпуклой* на X. Функцию $f(x)$ называют *вогнутой* [*строго вогнутой*] на выпуклом множестве X, если $(-f(x))$ выпукла [строго выпукла] на X.

Если множество X пусто или состоит из одной точки, то функцию на таком множестве нам будет удобно считать выпуклой (или вогнутой) по определению. Подчеркнем также, что всюду, если не оговорено противное, будем рассматривать лишь функции, принимающие конечные значения во всех точках области определения.

Примерами выпуклой функции на всем пространстве E^n служат линейная функция $f(x) = \langle c, x \rangle$ и норма $f(x) = |x|$. Кстати, линейная функция $f(x) = \langle c, x \rangle$ одновременно является и вогнутой на E^n.

В теореме 1.5 было показано, что выпуклое множество X содержит выпуклые комбинации $\sum_{i=1}^m \alpha_i u_i$, $\alpha_i \geqslant 0$, $i = 1, \ldots, m$, $\sum_{i=1}^m \alpha_i = 1$, любых своих точек $u_1, \ldots, u_m$ при любых $m = 2, 3, \ldots$ Пользуясь индукцией по той же схеме, какая была использована при доказательстве теоремы 1.5, нетрудно показать, что для любой выпуклой функции $f(x)$ на выпуклом множестве имеет место *неравенство Йенсена*

$$f\left(\sum_{i=1}^m \alpha_i x_i\right) \leqslant \sum_{i=1}^m \alpha_i f(x_i) \tag{2}$$

для любых $m = 1, 2, \ldots$, любых $x_i \in X$, $\alpha_i \geqslant 0$, $i = 1, \ldots, m$, $\sum_{i=1}^m \alpha_i = 1$.

2. Как и в случае выпуклых функций одной переменной, выпуклые функции многих переменных на выпуклом множестве не могут иметь локальных минимумов. Точнее, верна

Теорема 1. *Пусть X — выпуклое множество, а функция $f(x)$ определена и выпукла на X. Тогда всякая точка локального минимума*

$f(x)$ одновременно является точкой ее глобального минимума на X, причем множество

$$X_* = \{x\colon x \in X,\ f(x) = f_* = \inf_{u \in X} f(u)\}$$

выпукло. Если $f(x)$ строго выпукла на X, то X_* содержит не более одной точки.

Доказательство. Пусть u_* — точка локального минимума функции $f(x)$ на множестве X. Это значит, что существует окрестность $O(u_*, \varepsilon) = \{u\colon |u - u_*| < \varepsilon\}$ точки u_* такая, что $f(u_*) \leqslant f(v)$ для всех $v \in O(u_*, \varepsilon) \cap X$. Возьмем произвольную точку $x \in X$ и число $\alpha > 0$, столь малое, что $\alpha|x - u_*| < \varepsilon$. Тогда $u_* + \alpha(x - u_*) \in O(u_*, \varepsilon) \cap X$, и с учетом выпуклости функции $f(x)$ имеем $f(u_*) \leqslant f(u_* + \alpha(x - u_*)) \leqslant f(u_*) + \alpha(f(x) - f(u_*))$ или $0 \leqslant \alpha(f(x) - f(u_*))$. Сокращая на $\alpha > 0$, отсюда получаем $f(x) \geqslant f(u_*)$ при любом $x \in X$. Следовательно, $u_* \in X_*$.

Пусть теперь $u, v \in X_*$, т. е. $f(u) = f(v) = f_*$, $u, v \in X$. Тогда

$$f_* \leqslant f(\alpha u + (1 - \alpha)v) \leqslant \alpha f(u) + (1 - \alpha)f(v) = f_*, \tag{3}$$

т. е. $f(\alpha u + (1 - \alpha)v) = f_*$ при всех α, $0 \leqslant \alpha \leqslant 1$. Следовательно, $\alpha u + (1 - \alpha)v \in X_*$, $0 \leqslant \alpha \leqslant 1$. Выпуклость X_* доказана.

Если $u \neq v$, то для строго выпуклых функций неравенства (3) не могут обратиться в равенства при $0 < \alpha < 1$. Следовательно, строго выпуклая функция может достигать своей нижней грани на выпуклом множестве не более чем в одной точке. □

Примеры 1.11.1, 1.11.3–1.11.5 показывают, что у выпуклых функций множество X_* может быть пустым, может содержать одну или бесконечно много точек.

3. Далее, остановимся на одном характеристическом свойстве гладких выпуклых функций.

Теорема 2. *Пусть X — выпуклое множество. Если функция $f(x)$ выпукла на X и дифференцируема в точке $v \in X$, то*

$$f(u) \geqslant f(v) + \langle f'(v), u - v\rangle \quad \forall u \in X. \tag{4}$$

Если $f(x) \in C^1(X)$, то $f(x)$ выпукла на X тогда и только тогда, когда неравенство (4) выполняется при всех $u, v \in X$.

Доказательство. Необходимость. Пусть $f(x)$ выпукла на X и дифференцируема в точке $v \in X$. Перепишем неравенство (1) в виде

$$f(v + \alpha(u - v)) - f(v) \leqslant \alpha(f(u) - f(v)) \quad \forall \alpha \in [0, 1],\ \forall u, v \in X.$$

С учетом дифференцируемости $f(x)$ в точке v отсюда имеем

$$\langle f'(v), \alpha(u - v)\rangle + o(\alpha|u - v|) \leqslant \alpha(f(u) - f(v))$$

Деля обе части этого неравенства на $\alpha > 0$ и устремляя $\alpha \to +0$, получим неравенство (4). Если $f(x) \in C^1(X)$, то неравенство (4) верно для всех $u, v \in X$.

Достаточность. Пусть $f(x) \in C^1(X)$, X — выпуклое множество и пусть неравенство (4) выполняется при всех $u, v \in X$. Покажем, что тогда $f(x)$ выпукла на X. Возьмем произвольные точки $u, v \in X$ и число α, $0 \leqslant \alpha \leqslant 1$. Положим $u_\alpha = \alpha u + (1 - \alpha)v$. Из (4) получим

$$f(u) - f(u_\alpha) \geqslant \langle f'(u_\alpha), u - u_\alpha\rangle, \quad f(v) - f(u_\alpha) \geqslant \langle f'(u_\alpha), v - u_\alpha\rangle.$$

Умножим первое из этих неравенств на α, а второе — на $1 - \alpha$ и сложим. Получим $\alpha f(u) + (1 - \alpha)f(v) - f(u_\alpha) \geqslant \langle f'(u_\alpha), u_\alpha - u_\alpha\rangle = 0$, что равносильно неравенству (1). □

Неравенство (4) имеет простой геометрический смысл. Как известно [327; 350; 352; 534], гиперплоскость $\Gamma = \{(u, \gamma) \in E^{n+1}\colon u \in E^n, \gamma = f(v) + \langle f'(v), u - v\rangle\}$ является касательной плоскостью к графику функции $\gamma = f(x)$ в точке v. Поэтому неравенство (4) означает, что график выпуклой функции лежит не ниже касательной плоскости к этому графику в любой точке $v \in X$, в которой существует производная $f'(v)$ (ср. с теоремой 1.8.4).

4. Следующая теорема, называемая *критерием оптимальности* для выпуклых функций, дает необходимое и достаточное условие минимума гладких функций на выпуклом множестве.

Теорема 3. *Пусть X — выпуклое множество, X_* — множество точек минимума функции $f(x)$ на X. Если в точке $x_* \in X_*$ функция $f(x)$ дифференцируема, то необходимо выполняется неравенство*

$$\langle f'(x_*), x - x_*\rangle \geqslant 0 \qquad \forall x \in X, \tag{5}$$

которое в случае $x_* \in \operatorname{int} X$ превращается в равенство: $f'(x_*) = 0$. Если, кроме того, функция $f(x)$ выпукла на X, то условие (5) является достаточным для того, чтобы $x_* \in X_*$.

Доказательство. Необходимость. Пусть $x_* \in X_*$. Тогда при любых $x \in X$ и $\alpha \in [0, 1]$ с учетом дифференцируемости функции $f(x)$ в точке x_* имеем

$$0 \leqslant f(x_* + \alpha(x - x_*)) - f(x_*) = \alpha\langle f'(x_*), x - x_*\rangle + o(\alpha)$$

или

$$0 \leqslant \langle f'(x_*), x - x_*\rangle + o(\alpha)/\alpha, \qquad \forall x \in X,\ \forall \alpha \in (0, 1].$$

отсюда при $\alpha \to +0$ получим условие (5). Если $x_* \in \operatorname{int} X$, то для $\forall e \in E^n$ найдется $\varepsilon_0 > 0$ такое, что $x = x_* + \varepsilon e \in X$ при всех ε, $|\varepsilon| \leqslant \varepsilon_0$. Полагая в (5) $x = x_* + \varepsilon e$, получим $\varepsilon\langle f'(x_*), e\rangle \geqslant 0$ при всех ε, $|\varepsilon| \leqslant \varepsilon_0$, что возможно только при $\langle f'(x_*), e\rangle = 0$. Пользуясь произволом в выборе e, здесь можем взять $e = f'(x_*)$, что приводит к равенству $f'(x_*) = 0$.

Замечание 1. Несколько подправив приведенное доказательство необходимости, нетрудно убедиться, что если x_* — точка локального минимума $f(x)$ на множестве X, то мы также придем к условию (5).

Достаточность. Пусть функция $f(x)$ выпукла на X, пусть для некоторой точки $x_* \in X$ выполнено условие (5). Тогда из неравенства (4) при $v = x_*$ получим $f(x) - f(x_*) \geqslant \langle f'(x_*), x - x_*\rangle \geqslant 0$ или $f(x) \geqslant f(x_*)$ $\forall x \in X$. Следовательно, $x_* \in X_*$. Теорема 3 доказана. □

Условие (5) имеет простой геометрический смысл и означает неотрицательность производной по направлению $e = \frac{x - x_*}{|x - x_*|}$ функции $f(x)$ в точке минимума x_* — об этом подробнее см. ниже в п. 8. Если $X = E^n$, то условие (5) эквивалентно равенству $f'(x_*) = 0$. Таким образом, условие (5) является естественным обобщением условия стационарности (2.2.6) для задач на условный экстремум с выпуклым множеством X. Заметим, что если x_* — граничная точка множества X, то равенство $f'(x_*) = 0$ может выполняться, может и не выполняться. Например, если $f(x) = x^2$, $X = \{x \in E^1\colon 1 \leqslant x \leqslant 2\}$, то $x_* = 1$, $f'(x_*) = 2$, и условие (5) в точке x_*, конечно, выполняется. Если ту же функцию $f(x) = x^2$ рассматривать на отрезке $X = \{x \in E^1\colon 0 \leqslant x \leqslant 2\}$, то $x_* = 0$, $f'(x_*) = 0$, хотя $x_* = 0$ — граничная точка X.

Неравенство (5), как и условие стационарности (2.2.6), вообще говоря, может быть записано в виде системы n уравнений с n неизвестными $x_* = (x_*^1, \ldots, x_*^n)$. Поясним это на примере.

Пример 1. Рассмотрим задачу: $f(x) \to \inf$, $x \in X = E_+^2 = \{x = (x^1, x^2) \in E^2: x^1 \geqslant 0, x^2 \geqslant 0\}$, предполагая, что $f(x)$ дифференцируема на E_+^2. Тогда условие (5) равносильно системе уравнений

$$x_*^i f_{x^i}(x_*) = 0, \quad i = 1, 2.$$

В этом нетрудно убедиться, последовательно анализируя возможности: 1) $x_* = (0, 0)$; 2) $x_* = (x_*^1, x_*^2) > 0$; тогда $x_* \in \operatorname{int} E_+^2$ и $f'(x_*) = 0$; 3) $x_*^1 = 0$, $x_*^2 > 0$; тогда $f_{x^1}'(x_*) \geqslant 0$, $f_{x^2}'(x_*) = 0$; 4) $x_*^1 > 0$, $x_*^2 = 0$; тогда $f_{x^1}'(x_*) = 0$, $f_{x^2}'(x_*) \geqslant 0$.

Если $X = E_+^n = \{x \in E^n: x = (x^1, \ldots, x^n) > 0\}$, то условие (5) аналогично сводится к системе уравнений

$$x_*^i f_{x^i}(x_*) = 0, \quad i = 1, \ldots, n.$$

В общем случае для множеств X с «хорошей» границей неравенство (5) также может быть записано в виде системы уравнений, составленной из условий принадлежности x_* границе множества X и условий равенства нулю производных функции $f(x)$ по некоторым касательным направлениям к границе в точке x_*; правда, получающаяся при этом система может оказаться громоздкой и сложной для исследования. Не вдаваясь в детали, заметим, что для некоторых классов задач минимизации на множествах вида (2.3.10) условие (5) тесно связано с правилом множителей Лагранжа (подтверждение этим соображениям читатель найдет ниже, в § 9).

Так как неравенство (5) при $x = x_*$ обращается в равенство, то условие (5) можно записать в равносильном виде

$$\min_{x \in X} \langle f'(x_*), x - x_* \rangle = 0.$$

Неравенства вида (5) называются *вариационными неравенствами*, они подробно рассмотрены в [31; 227; 378; 397; 407; 536; 640].

5. Сформулируем и докажем ниже два критерия выпуклости для гладких функций.

Теорема 4. *Пусть X — выпуклое множество, $f(x) \in C^1(X)$. Тогда для выпуклости функции $f(x)$ на X необходимо и достаточно, чтобы*

$$\langle f'(u) - f'(v), u - v \rangle \geqslant 0 \quad \forall u, v \in X. \tag{6}$$

Доказательство. Необходимость. Пусть $f(x)$ выпукла на X. Тогда для любых $u, v \in X$ имеет место неравенство (4). Поменяв в (4) переменные u и v ролями, получим

$$f(v) \geqslant f(u) + \langle f'(u), v - u \rangle.$$

Сложив это неравенство с (4), придем к условию (6).

Достаточность. Пусть для некоторой функции $f(u) \in C^1(X)$ выполнено условие (6). Тогда с помощью формулы конечных приращений (2.6.2) для любых $u, v \in X$ и $\alpha \in [0, 1]$ имеем

$$\begin{aligned} &\alpha f(u) + (1-\alpha) f(v) - f(\alpha u + (1-\alpha)v) = \\ &\quad = \alpha [f(u) - f(\alpha u + (1-\alpha)v)] + (1-\alpha)[f(v) - f(\alpha u + (1-\alpha)v)] = \end{aligned}$$

$$\begin{aligned} &= \alpha \int_0^1 \langle f'(\alpha u + (1-\alpha)v + t(u - \alpha u - (1-\alpha)v)), u - \alpha u - (1-\alpha)v \rangle dt + \\ &\quad + (1-\alpha) \int_0^1 \langle f'(\alpha u + (1-\alpha)v + t(v - \alpha u - (1-\alpha)v)), v - \alpha u - (1-\alpha)v \rangle dt = \\ &\quad = \alpha(1-\alpha) \int_0^1 \langle f'(\alpha u + (1-\alpha)v + t(1-\alpha)(u-v)) - \\ &\qquad - f'(\alpha u + (1-\alpha)v + t\alpha(u-v)), u - v \rangle dt, \end{aligned}$$

или

$$\begin{aligned} &\alpha f(u) + (1-\alpha) f(v) - f(\alpha u + (1-\alpha)v) = \\ &\qquad = \alpha(1-\alpha) \int_0^1 \langle f'(z_1) - f'(z_2), z_1 - z_2 \rangle \frac{1}{t} dt, \end{aligned} \tag{7}$$

где $z_1 = \alpha u + (1-\alpha)v + t(1-\alpha)(u-v)$, $z_2 = \alpha u + (1-\alpha)v + t\alpha(v-u)$. Поскольку $z_1 = \beta u + (1-\beta)v$, $\beta = t + \alpha(1-t) \in [0, 1]$, $z_2 = \gamma u + (1-\gamma)v$, $\gamma = \alpha(1-t) \in [0, 1]$ и множество X выпукло, то $z_1, z_2 \in X$. Отсюда и из условия (6) имеем $\langle f'(z_1) - f'(z_2), z_1 - z_2 \rangle \geqslant 0$ при всех t, $0 < t \leqslant 1$. Это значит, что правая часть (7) и, следовательно, левая часть (7) неотрицательна при любом выборе $u, v \in X$, $\alpha \in [0, 1]$, т. е. $f(x)$ выпукла на X. □

Заметим, что для функций одной переменной неравенство (6) равносильно неубыванию производной $f'(x)$. Это значит, что доказанная теорема 4 является естественным обобщением теоремы 1.8.8 на случай гладких функций многих переменных.

Следующий критерий выпуклости обобщает теорему 1.8.9.

Теорема 5. *Пусть X — выпуклое множество из E^n, $f(x) \in C^2(X)$. Тогда для выпуклости $f(x)$ на X необходимо и достаточно, чтобы*

$$\langle f''(u)\xi, \xi \rangle \geqslant 0 \tag{8}$$

при всех $u \in X$ и всех $\xi = (\xi^1, \ldots, \xi^n)$, принадлежащих подпространству $L = \operatorname{Lin} X$, параллельному аффинной оболочке множества X (в частности, если $\operatorname{int} X \neq \varnothing$, то (8) выполняется при всех $\xi \in E^n$).

Доказательство. Необходимость. Пусть $f(x)$ выпукла на X. Пусть $\operatorname{aff} X = \{u \in E^n: Au = b\}$, где A — некоторая матрица размера $m \times n$, а $b \in E^n$ (см. пример 1.5). Тогда подпространство L, параллельное $\operatorname{aff} X$, имеет вид $L = \{\xi \in E^n: A\xi = 0\}$. Далее, согласно теореме 1.11 $\operatorname{ri} X \neq \varnothing$. Возьмем произвольные $\xi \in L$, $u \in \operatorname{ri} X$. Тогда $A(u + \varepsilon\xi) = Au + \varepsilon A\xi = Au = b$, т. е. $u + \varepsilon\xi \in \operatorname{aff} X$ при всех ε.

По определению 1.10 относительно внутренней точки множества найдется такое число $\varepsilon_0 > 0$, что $u + \varepsilon\xi \in X$ при всех ε, $|\varepsilon| \leqslant \varepsilon_0$. Поскольку для гладкой выпуклой функции справедливо неравенство (6), то из него с учетом формулы (2.6.4) имеем $\langle f'(u + \varepsilon\xi) - f'(u), \xi \rangle \varepsilon = \langle f''(u + \theta\varepsilon\xi)\xi, \xi \rangle \varepsilon^2 \geqslant 0$, $0 \leqslant \theta \leqslant 1$, или $\langle f''(u + \theta\varepsilon\xi)\xi, \xi \rangle \geqslant 0$ для всех ε, $0 < |\varepsilon| \leqslant \varepsilon_0$. Отсюда, пользуясь непрерывностью $f''(x)$, при $\varepsilon \to 0$ получим условие (8) для всех $u \in \operatorname{ri} X$. Если $u \in X \setminus \operatorname{ri} X$, то существует последовательность $\{u_k\} \in \operatorname{ri} X$, сходящаяся к u. По доказанному $\langle f''(u_k)\xi, \xi \rangle \geqslant 0$ при всех $\xi \in L$. Отсюда при $k \to \infty$ получим неравенства (8) и для точек $u \in X \setminus \operatorname{ri} X$.

Достаточность. Пусть $f(x) \in C^2(X)$ и выполнено условие (8). Возьмем произвольные точки $u, v \in X$. Тогда $\xi = u - v \in L$. Пользуясь формулой (2.6.4) и неравенством (8) при $\xi = u - v$, получим

$$\langle f'(u) - f'(v), u - v \rangle = \langle f''(v + \theta(u-v))(u-v), u - v \rangle \geqslant 0 \quad \forall u, v \in X.$$

Таким образом, для функции $f(x)$ выполнено условие (6). Из теоремы 4 следует выпуклость $f(x)$ на X. □

Замечание 2. Следующий пример показывает, что при $\operatorname{int} X = \varnothing$ условие (8) может и не выполняться при каждом $\xi \in E^n$.

Пример 2. Пусть $f(u) = x^2 - y^2$, $X = \{u = (x, y) \in E^2 : y = 0\}$. Ясно, что $f(u)$ выпукла на X. Но условие $\langle f''(u)\xi, \xi\rangle = 2(\xi^1)^2 - 2(\xi^2)^2 \geqslant 0$ не выполняется, например, для $\xi = (0, 1)$. Здесь $\operatorname{int} X = \varnothing$, $\operatorname{aff} X = X = L$.

Замечание 3. Условие (8) при $\operatorname{int} X \neq \varnothing$ представляет собой условие неотрицательности квадратичной формы

$$\langle f''(u)\xi, \xi\rangle = \sum_{i,j=1}^{n} \frac{\partial^2 f(u)}{\partial u^i \partial u^j} \xi^i \xi^j$$

на E^n. Как было отмечено в замечании 2.2.1, для того чтобы $\langle f''(u)\xi, \xi\rangle \geqslant 0$ при всех $\xi = (\xi^1, \ldots, \xi^n)$, необходимо и достаточно, чтобы все главные миноры матрицы были неотрицательны.

Напомним также, что неотрицательность квадратичной формы $\langle f''(u)\xi, \xi\rangle$ равносильна тому, что собственные числа $\lambda_1(u), \ldots, \lambda_n(u)$ матрицы $f''(u)$ (т. е. решения уравнения $\det |f''(u) - \lambda I_n| = 0$, I_n — единичная матрица размера $n \times n$) неотрицательны при всех $u \in X$.

Пример 3. Определим, при каких a, b, c функция

$$f(u) = x^2 + 2axy + by^2 + cz^2, \quad u = (x, y, z),$$

будет выпуклой на E^n. Здесь

$$f''(u) = \begin{bmatrix} 2 & a & 0 \\ 2a & 2b & 0 \\ 0 & 0 & 2c \end{bmatrix}.$$

Условие неотрицательности всех главных миноров этой матрицы дает искомые условия на a, b, c: $b - a^2 \geqslant 0$, $c \geqslant 0$.

Пример 4. Пусть

$$f(u) = \tfrac{1}{2}\langle Au, u\rangle - \langle b, u\rangle, \qquad u \in E^n, \tag{9}$$

где A — симметричная неотрицательно определенная матрица размера $n \times n$, $b \in E^n$. В частности, если $A = 2I_n$ — единичная матрица, $b = 0$, то $f(u) = \langle u, u\rangle = |u|^2$.

Приращение функции (9) нетрудно записать в виде

$$f(u + h) - f(u) = \langle Au - b, h\rangle + \tfrac{1}{2}\langle Ah, h\rangle \tag{10}$$

при любых $u, h \in E^n$. Из (10) имеем

$$f'(u) = Au - b, \quad f''(u) = A$$

По условию $A \geqslant 0$. Отсюда и из теоремы 5 следует выпуклость функции $f(u)$ на E^n. Согласно теореме 3 для того, чтобы функция (9) достигала своей нижней грани на E^n в точке u_*, необходимо и достаточно, чтобы u_* являлась решением линейной алгебраической системы $Au = b$. Указанная связь между задачей минимизации функции (9) на E^n и системой $Au = b$ с матрицей $A \geqslant 0$ лежит в основе ряда численных методов линейной алгебры [74; 89; 481].

Пример 5. Пусть

$$f(u) = |Au - b|^2, \quad u \in E^n, \tag{11}$$

где A — матрица порядка $m \times n$, $b \in E^n$. Покажем, что такая функция выпукла на E^n. Для этого вычислим ее производные.

Пользуясь формулой $\langle Ax, y\rangle = \langle x, A^\top y\rangle$, $x \in E^n$, $y \in E^n$, где $A^\top$ — матрица, полученная транспонированием матрицы A, нетрудно представить приращение функции (11) в виде

$$f(u + h) - f(u) = 2\langle A^\top(Au - b), h\rangle + \tfrac{1}{2}\langle 2A^\top Ah, h\rangle$$

при всех $u, h \in E^n$. Отсюда имеем

$$f'(u) = 2A^\top(Au - b), \quad f''(u) = 2A^\top A.$$

Но $\langle f''(u)\xi, \xi\rangle = 2\langle A^\top A\xi, \xi\rangle = 2\langle A\xi, A\xi\rangle = 2|A\xi|^2 \geqslant 0$ при всех $\xi \in E^n$. В силу теоремы 5 функция (11) выпукла на E^n.

Согласно теореме 3 для того, чтобы функция (11) достигала своей нижней грани на E^n в точке u, необходимо и достаточно, чтобы u удовлетворяла системе линейных алгебраических уравнений

$$A^\top Au = A^\top b.$$

6. Посмотрим далее, как влияют на выпуклость сложение, умножение на число и некоторые другие операции над выпуклыми функциями. Легко доказывается

Теорема 6. *Если функции $f_i(u)$, $i = 1, \ldots, m$, выпуклы на выпуклом множестве, то функция*

$$f(u) = \alpha_1 f_1(u) + \ldots + \alpha_m f_m(u)$$

выпукла на этом множестве при любых $\alpha_i \geqslant 0$, $i = 1, \ldots, m$.

Теорема 7. *Пусть $f_i(u)$, $i \in I$, — произвольное семейство функций, конечных и выпуклых на выпуклом множестве X, пусть*

$$f(u) = \sup_{i \in I} f_i(u), \quad u \in X.$$

Тогда функция $f(u)$ выпукла на X.

Доказательство. Возьмем произвольные точки $u, v \in X$, число $\alpha \in [0, 1]$ и положим $u_\alpha = \alpha u + (1 - \alpha)v$. Для каждого фиксированного $i \in I$ функция $f_i(u)$ выпукла на X, поэтому $f_i(u_\alpha) \leqslant \alpha f_i(u) + (1 - \alpha)f_i(v) \leqslant \alpha f(u) + (1 - \alpha)f(v)$ $\forall i \in I$. Переходя в левой части этих неравенств к верхней грани по $i \in I$, получим: $f(u_\alpha) \leqslant \alpha f(u) + (1 - \alpha)f(v)$ $\forall i \in I$, $\forall \alpha \in [0, 1]$. Выпуклость функции $f(u)$ доказана. □

Следует заметить, что хотя каждая из функций $f_i(u)$, $i \in I$, принимает конечное значение в каждой точке $u \in X$, но тем не менее в каких-то точках $u \in X$ возможно $f(u) = +\infty$. Несмотря на указанное обстоятельство, приведенное доказательство, очевидно, сохраняет силу.

Следствие 1. *Пусть функция $g(u)$ выпукла на выпуклом множестве X. Тогда функция*

$$g^+(u) = \max\{g(u); 0\}$$

выпукла на X.

Теорема 8. *Пусть функция $\varphi(t)$ одной переменной выпукла и не убывает на отрезке $[a, b]$ (возможность $a = -\infty$ или $b = \infty$ здесь не*

исключается), пусть функция $g(u)$ выпукла на выпуклом множестве $X \subseteq E^n$, причем $g(u) \in [a, b]$ при всех $u \in X$. Тогда функция

$$f(u) = \varphi(g(u))$$

выпукла на X.

Доказательство. Возьмем произвольные $u, v \in X$ и $\alpha \in [0, 1]$. Тогда

$$f(\alpha u+(1-\alpha)v) = \varphi(g(\alpha u+(1-\alpha)v)) \leqslant \varphi(\alpha g(u)+(1-\alpha)g(v)) \leqslant$$
$$\leqslant \alpha\varphi(g(u)) + (1-\alpha)\varphi(g(u)) = \alpha f(u) + (1-\alpha)f(v),$$

что и требовалось. □

Иногда удобнее пользоваться другим вариантом этой теоремы: если функция $\varphi(t)$ выпукла и не возрастает на отрезке $[a, b]$, а $g(u)$ вогнута на выпуклом множестве $X \subseteq E^n$, $g(u) \in [a, b]$ при $u \in X$, то функция $f(u) = \varphi(g(u))$ выпукла на X.

Следствие 1. *Если функция $g(u)$ выпукла и неотрицательна на выпуклом множестве X, то функция*

$$f(u) = (g(u))^p$$

выпукла на X при всех $p \geqslant 1$.

Следствие 2. *Если функция $g(u)$ выпукла на выпуклом множестве X, то функция*

$$f(u) = (\max\{0; g(u)\})^p = (g^+(u))^p$$

выпукла на X при всех $p \geqslant 1$.

Следствие 3. *Если функция $g(u)$ выпукла на выпуклом множестве X, причем $g(u) < 0$ при всех $u \in X$, то функции*

$$f(u) = -1/g(u), \quad f(u) = \max\{-\ln(-g(u)); 0\}^p, \quad p \geqslant 1,$$

выпуклы на X.

Как увидим ниже, функции, указанные в следствиях к теоремам 7, 8, будут использованы при описании различных методов минимизации (например, в методах штрафных и барьерных функций и др.).

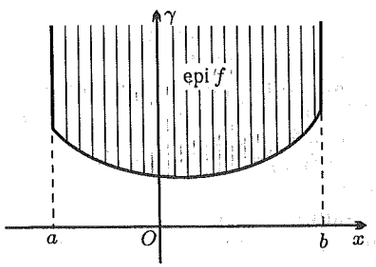

Рис. 4.5 (Надграфик)

7. Выпуклые функции являются удобным средством для задания выпуклых множеств. Это связано с тем, что надграфик всякой выпуклой функции является выпуклым множеством.

Определение 2. *Надграфиком* (или *эпиграфом*) всякой функции $f(x)$, определенной на множестве $X \subseteq E^n$, называется множество (рис. 4.5)

$$\operatorname{epi} f = \{(x, \gamma) \in E^{n+1}: x \in X, \gamma \geqslant f(x)\}.$$

Теорема 9. *Для того чтобы функция $f(x)$, определенная на выпуклом множестве X была выпуклой на X, необходимо и достаточно, чтобы ее надграфик был выпуклым множеством.*

Доказательство. Необходимость. Пусть функция $f(x)$ выпукла на выпуклом множестве X. Возьмем две произвольные точки $z_1 = (u_1, \gamma_1)$, $z_2 = (u_2, \gamma_2) \in \operatorname{epi} f$ и составим их выпуклую комбинацию $z_\alpha = \alpha z_1 + (1-\alpha)z_2 = (\alpha u_1 + (1-\alpha)u_2, \alpha\gamma_1 + (1-\alpha)\gamma_2)$, $0 \leqslant \alpha \leqslant 1$. Из выпуклости X следует, что $u_\alpha = \alpha u_1 + (1-\alpha)u_2 \in X$. Из выпуклости функции $f(x)$, учитывая, что $z_1, z_2 \in \operatorname{epi} f$, имеем $f(u_\alpha) \leqslant \alpha f(u_1) + (1-\alpha)f(u_2) \leqslant \alpha\gamma_1 + (1-\alpha)\gamma_2$. Следовательно, $z_\alpha \in \operatorname{epi} f$ при всех $\alpha \in [0, 1]$. Выпуклость $\operatorname{epi} f$ доказана.

Достаточность. Пусть $\operatorname{epi} f$ — выпуклое множество. Возьмем произвольные $u_1, u_2 \in X$ и $\alpha \in [0, 1]$. Тогда $z_1 = (u_1, f(u_1))$, $z_2 = (u_2, f(u_2)) \in \operatorname{epi} f$. В силу выпуклости $\operatorname{epi} f$ точка $z_\alpha = \alpha z_1 + (1-\alpha)z_2 \in \operatorname{epi} f$. Это значит, что $\alpha f(u_1) + (1-\alpha)f(u_2) \geqslant f(\alpha u_1 + (1-\alpha)u_2)$. Выпуклость $f(x)$ доказана. □

Теорема 10. *Пусть X — выпуклое множество, а функция $f(x)$ выпукла на X. Тогда множество $M(c) = \{u: u \in X, f(u) \leqslant c\}$ выпукло при любом c.*

Доказательство. Возьмем произвольные $u, v \in M(c)$, $\alpha \in [0, 1]$. Используя выпуклость множества X и функции $f(x)$, имеем $f(\alpha u+(1-\alpha)v) \leqslant \alpha f(u) + (1-\alpha)f(v) \leqslant c$, т. е. $\alpha u + (1-\alpha)v \in M(c)$, что и требовалось. □

Заметим, что обратное утверждение здесь неверно: из выпуклости множества $M(c)$ при любом c, вообще говоря, не следует выпуклость функции $f(x)$. Например, множество $M(c) = \{u: u \in E^1, u^3 \leqslant c\}$ выпукло при любом c, а функция $f(u) = u^3$ невыпукла на E^1 (см. упражнение 33).

Теорема 11. *Пусть X_0 — выпуклое множество, функции $g_i(x)$, $i = 1, \ldots, m$, выпуклы на X_0, а $g_i(x) = \langle a_i, x\rangle - b_i$, $i = m+1, \ldots, s$, где a_i — заданные векторы из E^n, b_i — заданные числа, $i = m+1, \ldots, s$. Тогда выпукло множество*

$$X = \{u: u \in X_0,\ g_i(u) \leqslant 0,\ i = 1, \ldots, m;\ g_i(u) = 0,\ i = m+1, \ldots, s\}. \qquad (12)$$

Доказательство. В силу теоремы 10 множество $X_i = \{u: u \in X_0, g_i(u) \leqslant 0\}$ выпукло при всех $i = 1, \ldots, m$. Выпукло также множество $M = \{u \in E^n: \langle a_i, u\rangle - b_i = 0, i = m+1, \ldots, s\}$ — см. пример 1.5. Тогда множество (12), являющееся пересечением выпуклых множеств $X_1, \ldots, X_m, M$ само будет выпуклым. □

8. Рассмотренное в теореме 3 условие оптимальности сформулировано для непрерывно-дифференцируемых функций. Однако аналогичное условие можно получить при гораздо меньших ограничениях на функцию, используя лишь существование производных по направлениям. Напоминаем, что *производной функции $f(x)$ в точке u по направлению e*, $|e| = 1$, называется число

$$\frac{df(u)}{de} = \lim_{t \to +0} \frac{f(u+te) - f(u)}{t}. \qquad (13)$$

Заметим, что для определения производной по направлению в точке u нужно, чтобы $u + te$ принадлежало области определения $f(x)$ при $0 \leqslant t \leqslant t_0$ хотя бы при малом $t_0 > 0$.

Определение 3. Пусть X — некоторое множество из E^n, пусть $u \in X$. Направление $e \neq 0$ называется *возможным* в точке u, если существует число $t_0 > 0$ такое, что $u + te \in X$ при всех t, $0 \leqslant t \leqslant t_0$. Иначе говоря, достаточно малое перемещение из точки u по возможному направлению не выводит за пределы множества X.

Очевидно, если $u \in \operatorname{int} X$, то любое направление $e \neq 0$ является возможным в этой точке. В граничных точках множества возможное направление может и не существовать.

Пример 6. Пусть $X = \{u = (x, y) \in E^2: x \geqslant 0, x^2 \leqslant y \leqslant 2x^2\}$. Нетрудно видеть, что в граничной точке $(0, 0)$ нет ни одного возможного направления.

Для выпуклых множеств X, содержащих не менее двух точек, приведенная в примере 6 ситуация невозможна: в любой точке u такого выпуклого множества X имеется хотя бы одно возможное направление, причем направление $e \neq 0$ будет возможным в точке u тогда и только тогда, когда существуют точка $v \in X$, $v \neq u$, и число $\gamma > 0$ такие, что $e = \gamma(v - u)$.

Таким образом, если функции $f(x)$ определена на множестве X, а направление e, $|e| = 1$, является возможным в точке $u \in X$, то функция $g(t) = f(u + te)$ определена на отрезке $[0, t_0]$, где $t_0 > 0$, и $df(u)/de = g'(+0)$ — правая производная $g(t)$ в точке $t = 0$.

исключается), пусть функция $g(u)$ выпукла на выпуклом множестве $X \subseteq E^n$, причем $g(u) \in [a, b]$ при всех $u \in X$. Тогда функция

$$f(u) = \varphi(g(u))$$

выпукла на X.

Доказательство. Возьмем произвольные $u, v \in X$ и $\alpha \in [0, 1]$. Тогда

$$f(\alpha u+(1-\alpha)v) = \varphi(g(\alpha u+(1-\alpha)v)) \leqslant \varphi(\alpha g(u)+(1-\alpha)g(v)) \leqslant$$
$$\leqslant \alpha\varphi(g(u)) + (1-\alpha)\varphi(g(u)) = \alpha f(u) + (1-\alpha)f(v),$$

что и требовалось. □

Иногда удобнее пользоваться другим вариантом этой теоремы: если функция $\varphi(t)$ выпукла и не возрастает на отрезке $[a, b]$, а $g(u)$ вогнута на выпуклом множестве $X \subseteq E^n$, $g(u) \in [a, b]$ при $u \in X$, то функция $f(u) = \varphi(g(u))$ выпукла на X.

Следствие 1. *Если функция $g(u)$ выпукла и неотрицательна на выпуклом множестве X, то функция*

$$f(u) = (g(u))^p$$

выпукла на X при всех $p \geqslant 1$.

Следствие 2. *Если функция $g(u)$ выпукла на выпуклом множестве X, то функция*

$$f(u) = (\max\{0; g(u)\})^p = (g^+(u))^p$$

выпукла на X при всех $p \geqslant 1$.

Следствие 3. *Если функция $g(u)$ выпукла на выпуклом множестве X, причем $g(u) < 0$ при всех $u \in X$, то функции*

$$f(u) = -1/g(u), \quad f(u) = \max\{-\ln(-g(u)); 0\}^p, \quad p \geqslant 1,$$

выпуклы на X.

Как увидим ниже, функции, указанные в следствиях к теоремам 7, 8, будут использованы при описании различных методов минимизации (например, в методах штрафных и барьерных функций и др.).

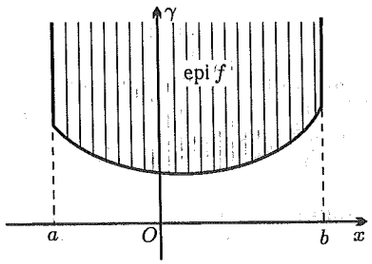

Рис. 4.5 (Надграфик)

7. Выпуклые функции являются удобным средством для задания выпуклых множеств. Это связано с тем, что надграфик всякой выпуклой функции является выпуклым множеством.

Определение 2. *Надграфиком* (или *эпиграфом*) всякой функции $f(x)$, определенной на множестве $X \subseteq E^n$, называется множество (рис. 4.5)

$$\operatorname{epi} f = \{(x, \gamma) \in E^{n+1}: x \in X, \gamma \geqslant f(x)\}.$$

Теорема 9. *Для того чтобы функция $f(x)$, определенная на выпуклом множестве X была выпуклой на X, необходимо и достаточно, чтобы ее надграфик был выпуклым множеством.*

Доказательство. Необходимость. Пусть функция $f(x)$ выпукла на выпуклом множестве X. Возьмем две произвольные точки

$z_1 = (u_1, \gamma_1)$, $z_2 = (u_2, \gamma_2) \in \operatorname{epi} f$ и составим их выпуклую комбинацию $z_\alpha = \alpha z_1 + (1-\alpha)z_2 = (\alpha u_1 + (1-\alpha)u_2, \alpha\gamma_1 + (1-\alpha)\gamma_2)$, $0 \leqslant \alpha \leqslant 1$. Из выпуклости X следует, что $u_\alpha = \alpha u_1 + (1-\alpha)u_2 \in X$. Из выпуклости функции $f(x)$, учитывая, что $z_1, z_2 \in \operatorname{epi} f$, имеем $f(u_\alpha) \leqslant \alpha f(u_1) + (1-\alpha)f(u_2) \leqslant \alpha\gamma_1 + (1-\alpha)\gamma_2$. Следовательно, $z_\alpha \in \operatorname{epi} f$ при всех $\alpha \in [0, 1]$. Выпуклость $\operatorname{epi} f$ доказана.

Достаточность. Пусть $\operatorname{epi} f$ — выпуклое множество. Возьмем произвольные $u_1, u_2 \in X$ и $\alpha \in [0, 1]$. Тогда $z_1 = (u_1, f(u_1))$, $z_2 = (u_2, f(u_2)) \in \operatorname{epi} f$. В силу выпуклости $\operatorname{epi} f$ точка $z_\alpha = \alpha z_1 + (1-\alpha)z_2 \in \operatorname{epi} f$. Это значит, что $\alpha f(u_1) + (1-\alpha)f(u_2) \geqslant f(\alpha u_1 + (1-\alpha)u_2)$. Выпуклость $f(x)$ доказана. □

Теорема 10. *Пусть X — выпуклое множество, а функция $f(x)$ выпукла на X. Тогда множество $M(c) = \{u: u \in X, f(u) \leqslant c\}$ выпукло при любом c.*

Доказательство. Возьмем произвольные $u, v \in M(c)$, $\alpha \in [0, 1]$. Используя выпуклость множества X и функции $f(x)$, имеем $f(\alpha u+(1-\alpha)v) \leqslant \alpha f(u) + (1-\alpha)f(v) \leqslant c$, т. е. $\alpha u + (1-\alpha)v \in M(c)$, что и требовалось. □

Заметим, что обратное утверждение здесь неверно: из выпуклости множества $M(c)$ при любом c, вообще говоря, не следует выпуклость функции $f(x)$. Например, множество $M(c) = \{u: u \in E^1, u^3 \leqslant c\}$ выпукло при любом c, а функция $f(u) = u^3$ невыпукла на E^1 (см. упражнение 33).

Теорема 11. *Пусть X_0 — выпуклое множество, функции $g_i(x)$, $i = 1, \ldots, m$, выпуклы на X_0, а $g_i(x) = \langle a_i, x\rangle - b_i$, $i = m+1, \ldots, s$, где a_i — заданные векторы из E^n, b_i — заданные числа, $i = m+1, \ldots, s$. Тогда выпукло множество*

$$X = \{u: u \in X_0,\ g_i(u) \leqslant 0,\ i = 1, \ldots, m;\ g_i(u) = 0,\ i = m+1, \ldots, s\}. \qquad (12)$$

Доказательство. В силу теоремы 10 множество $X_i = \{u: u \in X_0, g_i(u) \leqslant 0\}$ выпукло при всех $i = 1, \ldots, m$. Выпукло также множество $M = \{u \in E^n: \langle a_i, u\rangle - b_i = 0, i = m+1, \ldots, s\}$ — см. пример 1.5. Тогда множество (12), являющееся пересечением выпуклых множеств $X_1, \ldots, X_m, M$ само будет выпуклым. □

8. Рассмотренное в теореме 3 условие оптимальности сформулировано для непрерывно-дифференцируемых функций. Однако аналогичное условие можно получить при гораздо меньших ограничениях на функцию, используя лишь существование производных по направлениям. Напоминаем, что *производной функции $f(x)$ в точке u по направлению e*, $|e| = 1$, называется число

$$\frac{df(u)}{de} = \lim_{t \to +0} \frac{f(u+te) - f(u)}{t}. \qquad (13)$$

Заметим, что для определения производной по направлению в точке u нужно, чтобы $u + te$ принадлежало области определения $f(x)$ при $0 \leqslant t \leqslant t_0$ хотя бы при малом $t_0 > 0$.

Определение 3. Пусть X — некоторое множество из E^n, пусть $u \in X$. Направление $e \neq 0$ называется *возможным* в точке u, если существует число $t_0 > 0$ такое, что $u + te \in X$ при всех t, $0 \leqslant t \leqslant t_0$. Иначе говоря, достаточно малое перемещение из точки u по возможному направлению не выводит за пределы множества X.

Очевидно, если $u \in \operatorname{int} X$, то любое направление $e \neq 0$ является возможным в этой точке. В граничных точках множества возможное направление может и не существовать.

Пример 6. Пусть $X = \{u = (x, y) \in E^2: x \geqslant 0, x^2 \leqslant y \leqslant 2x^2\}$. Нетрудно видеть, что в граничной точке $(0, 0)$ нет ни одного возможного направления.

Для выпуклых множеств X, содержащих не менее двух точек, приведенная в примере 6 ситуация невозможна: в любой точке u такого выпуклого множества X имеется хотя бы одно возможное направление, причем направление $e \neq 0$ будет возможным в точке u тогда и только тогда, когда существуют точка $v \in X$, $v \neq u$, и число $\gamma > 0$ такие, что $e = \gamma(v - u)$.

Таким образом, если функции $f(x)$ определена на множестве X, а направление e, $|e| = 1$, является возможным в точке $u \in X$, то функция $g(t) = f(u + te)$ определена на отрезке $[0, t_0]$, где $t_0 > 0$, и $df(u)/de = g'(+0)$ — правая производная $g(t)$ в точке $t = 0$.

Заметим, что если функция $f(x)$ определена в некоторой ε-окрестности точки u и дифференцируема в этой точке, то $f(x)$ имеет производные по всем направлениям, причем

$$\frac{df(u)}{de} = \langle f'(u), e\rangle, \quad |e| = 1 \tag{14}$$

(ср. с (2.6.1) при $t \to +0$). Однако обратное неверно: из того, что функция в некоторой точке имеет производные по всем направлениям, вообще говоря, не следует ее дифференцируемость в этой точке, и более того, нельзя гарантировать даже ее непрерывность.

Пример 7. Пусть

$$f(u) = f(x, y) = \begin{cases} \dfrac{x^2 y}{x^4 + y^2}, & u = (x, y) \neq 0, \\ 0, & u = (0, 0) = 0. \end{cases}$$

Возьмем произвольное направление $e = (\cos\alpha, \sin\alpha)$. Тогда

$$\frac{f(0 + te) - f(0)}{t} \equiv \frac{1}{t} f(t\cos\alpha, t\sin\alpha) = \frac{\cos^2\alpha \sin\alpha}{t^2\cos^4\alpha + \sin^2\alpha}, \quad t > 0.$$

Отсюда имеем

$$\frac{df(0)}{de} = \begin{cases} \cos^2\alpha/\sin\alpha, & \sin\alpha \neq 0, \\ 0, & \sin\alpha = 0. \end{cases}$$

Однако эта функция разрывна в точке $u = 0$. В самом деле, устремим точку $u = (x, y)$ к нулю по параболе $y = x^2$. Тогда $f(x, x^2) \equiv 1/2 \not\to f(0, 0) = 0$.

Таким образом, требование существования производных по направлению существенно менее жесткое, чем требование дифференцируемости. В связи с этим представляет интерес получить условия оптимальности в терминах производных по направлению.

Теорема 12. *Пусть X — выпуклое множество, X_* — множество точек минимума функции $f(x)$ на X, пусть в точке $u_* \in X_*$ функция $f(x)$ имеет производные по всем возможным направлениям. Тогда необходимо выполняется условие*

$$\frac{df(u_*)}{de} \geqslant 0 \tag{15}$$

для всех возможных направлений e, $|e| = 1$, в точке u_*. Если, кроме того, функция $f(x)$ выпукла на X, то условие (15) *достаточно для того, чтобы $u_* \in X_*$.*

Доказательство. Необходимость. Пусть $u_* \in X_*$ и e, $|e| = 1$ — возможное направление в точке u_*. Тогда $f(u_* + te) \geqslant f(u_*)$ или $(f(u_* + te) - f(u_*))/t \geqslant 0$ при всех достаточно малых $t > 0$. Отсюда при $t \to +0$ получим условие (15).

Достаточность. Пусть $f(x)$ — выпуклая функция на X, пусть в некоторой точке $u_* \in x$ выполняется условие (15). Возьмем любую точку $u \in X$, $u \neq u_*$, и положим $e = (u - u_*)/|u - u_*|$. Направление e — возможное в точке u_*, так как $u_* + te \in X$ при всех t, $0 \leqslant t \leqslant t_0 = |u - u_*|$, $t_0 > 0$. Из условия (15) тогда имеем $g'(+0) \geqslant 0$, где $g(t) = f(u_* + te)$.

Ниже в теореме 13 будет показано, что $g(t)$ выпукла на $[0, t_0]$. Из неравенства (1.8.6) тогда следует, что $g(t) - g(0) \geqslant g'(+0)t$ или $g(t) \geqslant g(0)$ при всех $t \in [0, t_0]$. В частности, при $t = t_0 = |u - u_*|$ отсюда имеем $f(u) \geqslant f(u_*)$, что и требовалось. □

В частности, если в точке u_* существует градиент $f'(u_*)$, то для $e = (u - u_*)/|u - u_*|$, $u \in X$, $u \neq u_*$, согласно формуле (14) имеем $df(u_*)/de = \langle f'(u_*), u - u_*\rangle/|u - u_*|$, и в этом случае условие (15) превращается в условие (5). Таким образом, теорема 12 является обобщением теоремы 3 на существенно более широкий класс функций. Более того, условие (15) является наиболее естественным для класса выпуклых функций. Дело в том, что, оказывается, всякая выпуклая функция в любой внутренней точке множества имеет производные по всем направлениям. Это вытекает из следующих двух теорем.

Теорема 13. *Пусть X — выпуклое множество, функция $f(x)$ определена на X. Для того чтобы $f(x)$ была выпуклой на X, необходимо и достаточно, чтобы для любой точки $u \in X$ и любого возможного направления e в точке u функция $g(t) = f(u + te)$ одной переменной t была выпукла на отрезке $[a, b]$, где $a = \inf\{t: u + te \in X\}$, $b = \sup\{t: u + te \in X\}$ (ясно, что $a \leqslant 0 < b$; если $u + ae \notin X$ или $u + be \notin X$, то функцию $g(t)$ не следует рассматривать соответственно при $t = a$ или $t = b$).*

Доказательство. Необходимость. Пусть $f(x)$ выпукла на X. Возьмем произвольную точку $u \in X$, какое-либо возможное направление e в этой точке и составим функцию $g(t) = f(u + te)$, $a \leqslant t \leqslant b$. Пусть t_1, t_2 — произвольные точки из $[a, b]$ и $\alpha \in [0, 1]$. Тогда $g(\alpha t_1 + (1-\alpha)t_2) = f(\alpha(u + t_1 e) + (1-\alpha)(u + t_2 e)) \leqslant \alpha f(u + t_1 e) + (1-\alpha) f(u + t_2 e) = \alpha g(t_1) + (1 - \alpha) g(t_2)$, что и требовалось.

Достаточность. Пусть для всех $u \in X$ и всех возможных направлений e в точке u функция $g(t) = f(u + te)$ выпукла на соответствующем отрезке $[a, b]$. Возьмем любые точки $u, v \in X$, $u \neq v$, положим $e = v - u$ — это возможное направление в точке u, так как $u + t(v - u) \in X$ при $0 \leqslant t \leqslant 1$. Из выпуклости $g(t) = f(u + te)$ получим $f(\alpha v + (1 - \alpha)u) = g(\alpha) = g(\alpha \cdot 1 + (1 - \alpha) \cdot 0) \leqslant \alpha g(1) + (1 - \alpha) g(0) = \alpha f(v) + (1 - \alpha) f(u)$ при всех $\alpha \in [0, 1]$. □

Теорема 14. *Пусть X — выпуклое множество, функция $f(x)$ выпукла на X. Тогда в любой точке $u \in \operatorname{ri} X$ функция $f(x)$ имеет производные по всем направлениям $e \in \operatorname{Lin} X$. В частности, если $\operatorname{int} X \neq 0$, то в точке $u \in \operatorname{int} X$ существуют производные функции $f(x)$ по всем направлениям $e \in E^n$, $|e| = 1$.*

Доказательство. Зафиксируем какое-либо направление $e \in \operatorname{Lin} X$, $|e| = 1$, и точку $u \in \operatorname{ri} X$. Согласно определению 1.10 существует ε-окрестность $O(u, \varepsilon) = \{v \in E^n: |v - u| < \varepsilon\}$ точки u, такая, что пересечение $O(u, \varepsilon) \cap \operatorname{aff} X$ целиком принадлежит X. Учитывая, что $-e$ также принадлежит $\operatorname{Lin} X$, можем сказать, что $u + te \in X$ для всех t, $|t| \leqslant t_0$, $0 < t_0 < \varepsilon$. Это значит, что функция $g(t) = f(u + te)$ определена на отрезке $[-t_0, t_0]$ и согласно теореме 13 она выпукла на этом отрезке. Поскольку $t = 0$ — внутренняя точка отрезка $[-t_0, t_0]$, то по теореме 1.8.2 существует

$$g'(+0) = \lim_{t \to +0} \frac{g(t) - g(0)}{t} = \lim_{t \to +0} \frac{f(u + te) - f(u)}{t} = \frac{df(u)}{de}.$$

Если $u \in X \setminus \operatorname{ri} X$, то в такой точке у выпуклой функции производные по возможным направлениям могут и не существовать — об этом свидетельствует пример 1.8.2. □

9. Приведенный выше пример 7 показывает, что существование производных по всем направлениям не гарантирует непрерывности функции. Но для выпуклых функций такая ситуация, оказывается, невозможна.

Теорема 15. *Пусть множество X выпукло и $\operatorname{int} X \neq \varnothing$. Тогда выпуклая функция $f(x)$ во всех внутренних точках множества X непрерывна. В частности, функция, выпуклая на всем пространстве E^n, непрерывна во всех точках.*

Доказательство. Возьмем произвольные $u \in \operatorname{int} X$ и $\varepsilon > 0$. По определению внутренней точки существует число $\delta > 0$ такое, что $u + h \in X$, $u + nh^i e_i \in X$ для всех $h = (h^1, \ldots, h^n)$, $|h| < \delta/n$; здесь $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$, $i = 1, \ldots, n$ — базис в E^n. Поскольку по теореме 14 функция $f(x)$ в точке u имеет производные по направлениям e_i, то она непрерывна в этой точке по направлениям e_i, $i = 1, \ldots, n$. Поэтому можно взять число δ столь малым, чтобы $|f(u + nh^i e_i) - f(u)| < \varepsilon$ при всех h, $|h| < \delta/n$, $i = 1, \ldots, n$. Тогда, пользуясь неравенством (2), получаем

$$f(u + h) - f(u) = f\Big(\frac{1}{n}\sum_{i=1}^{n}(u + nh^i e_i)\Big) - f(u) \leqslant \frac{1}{n}\sum_{i=1}^{n}(f(u + nh^i e_i) - f(u)) < \varepsilon \tag{16}$$

для всех h, $|h| < \delta/n$. В частности, для $-h$, удовлетворяющих неравенству $|-h| < \delta/n$, из (16) следует $f(u - h) - f(u) < \varepsilon$. Но в силу выпуклости $f(x)$ имеем $f(u) = f((u + h)/2 + (u - h)/2) \leqslant (f(u + h) + f(u - h))/2$, поэтому $f(u) - f(u + h) \leqslant f(u - h) - f(u) < \varepsilon$. Отсюда и из (16) следует $|f(u + h) - f(u)| < \varepsilon$ при всех h, $|h| < \delta/n$. □

Заметим, что если $\operatorname{int} X = \varnothing$, то, рассматривая лишь точки из $\operatorname{aff} X$, аналогично можно доказать непрерывность выпуклой на X функции во всех точках $u \in \operatorname{ri} X$. В качестве базиса $\{e_i\}$, участвующего в доказательстве, в этом случае нужно взять базис подпространства $\operatorname{Lin} X$. В точках $u \in X \setminus \operatorname{ri} X$ выпуклая функция может терпеть разрыв — об этом говорит пример 1.8.1. Теорему 15 дополняет доказываемая ниже теорема 6.6.

10. Рассмотрим выпуклые функции на выпуклом множестве X, принадлежащие классу $C^{1,1}(X)$ (см. определение 2.6.3), т. е. гладкие выпуклые функции, градиент которых удовлетворяет условию

$$|f'(u) - f'(v)| \leqslant L|u - v| \quad \forall u, v \in X, \ L = \text{const} \geqslant 0. \tag{17}$$

Для таких функций имеют место неравенства

$$0 \leqslant \langle f'(u) - f'(v), u - v\rangle \leqslant L|u - v|^2 \quad \forall u, v \in X. \tag{18}$$

В самом деле, левое неравенство следует из теоремы 4, а правое — из условия (17). Оказывается, при $\operatorname{int} X \neq \varnothing$ эти два неравенства можно записать в виде одного равносильного неравенства [24; 234; 525], полностью характеризующего класс выпуклых функций из $C^{1,1}(X)$ с данной постоянной $L \geqslant 0$.

Теорема 16. *Пусть X — выпуклое множество из E^n, $\operatorname{int} X \neq \varnothing$. Для того чтобы функция $f(x)$ из класса $C^{1,1}(X)$ была выпуклой и удовлетворяла условию* (17) *с постоянной L, необходимо и достаточно, чтобы*

$$|f'(u) - f'(v)|^2 \leqslant L\langle f'(u) - f'(v), u - v\rangle \quad \forall u, v \in X. \tag{19}$$

Из (19) *следует неравенство* [24]

$$\langle f'(u) - f'(v), v - w\rangle \leqslant \frac{1}{4}L|u - w|^2 \quad \forall u, v, w \in X. \tag{20}$$

Доказательство. Достаточность. Если выполняется неравенство (19), то из него, во-первых, следует, что $\langle f'(u) - f'(v), u - v\rangle \geqslant 0$, $\forall u, v \in X$, и выпуклость $f(x)$ гарантируется теоремой 4, и, во-вторых, применяя к правой части (19) неравенство Коши — Буняковского и деля на $|f'(u) - f'(v)|$, получаем условие (17). Из выпуклости $f(x)$ и условия (17) имеем неравенства (18). Таким образом, из (19) следует (18).

Кроме того, из (19) имеем

$$\langle f'(u) - f'(v), v - w\rangle = \langle f'(u) - f'(v), u - w\rangle - \langle f'(u) - f'(v), u - v\rangle \leqslant$$
$$\leqslant \langle f'(u) - f'(v), u - w\rangle - \frac{1}{L}|f'(u) - f'(v)|^2 = -|L^{-1/2}(f'(u) - f'(v)) -$$
$$- \frac{1}{2}L^{1/2}(u - w)|^2 + \frac{1}{4}L|u - w|^2 \leqslant \frac{1}{4}L|u - w|^2 \quad \forall u, v, w \in X.$$

неравенство (20) установлено. Заметим, что при доказательстве достаточности условие $\operatorname{int} X \neq \varnothing$ не использовано.

Необходимость. Пусть функция $f(x)$ выпукла и удовлетворяет условию (17). Тогда, как было показано выше, справедливы неравенства (18). Остается из (18) получить (19). Сначала рассмотрим случай, когда $f(x) \in C^2(X)$. Тогда

$$0 \leqslant \langle f''(u)\xi, \xi\rangle \leqslant L|\xi|^2 \quad \forall u \in X \tag{21}$$

при всех $\xi \in E^n$. В самом деле, из неравенств (18) с помощью формулы (2.6.4) в случае $u \in \operatorname{int} X$ имеем

$$0 \leqslant \langle f'(u + \varepsilon\xi) - f'(u), \varepsilon\xi\rangle = \varepsilon^2\langle f''(u + \theta\varepsilon\xi)\xi, \xi\rangle \leqslant L|\xi|^2\varepsilon^2$$

или $0 \leqslant \langle f''(u+\theta\varepsilon\xi)\xi, \xi\rangle \leqslant L|\xi|^2$, $0 \leqslant \theta \leqslant 1$, для всех ε, $|\varepsilon| \leqslant \varepsilon_0$, $\varepsilon_0 > 0$. Отсюда при $\varepsilon \to +0$ получим (21) для точек $u \in \operatorname{int} X$. Если $u \in \operatorname{Гр} X$, то оценка (21) доказывается с помощью предельного перехода от внутренних точек так же, как это делалось при доказательстве теоремы 5.

Далее, пользуясь формулой (2.6.5), имеем

$$f'(u + h) - f'(u) = Ah, \quad A = \int_0^1 f''(u + th)dt, \quad h = v - u. \tag{22}$$

Разумеется, матрица A зависит от u, v, но эту зависимость мы для краткости не будем явно указывать. Согласно (21) $0 \leqslant \langle f''(u + th)\xi, \xi\rangle \leqslant L|\xi|^2$, $0 \leqslant t \leqslant 1$, откуда, интегрируя по t, получаем

$$0 \leqslant \langle A\xi, \xi\rangle \leqslant L|\xi|^2, \quad \xi \in E^n. \tag{23}$$

Таким образом, симметричная матрица A неотрицательно определена. Тогда существует симметричная неотрицательно определенная матрица $A^{1/2}$ такая, что $(A^{1/2})^2 = A$ [192, 353]. Пользуясь оценкой (23) при $\xi = A^{1/2}h$, с помощью формул (22) имеем

$$|f'(u) - f'(v)|^2 = \langle Ah, Ah\rangle = \langle AA^{1/2}h, A^{1/2}h\rangle \leqslant L|A^{1/2}h|^2 = L\langle Ah, h\rangle = L\langle f'(u) - f'(v), u - v\rangle.$$

Неравенство (19) доказано при дополнительном предположении $f(x) \in C^2(X)$.

Наметим схему доказательства для случая, когда $f(x) \in C^1(X)$. Построим последовательность функций $\{f_k(x)\}$, $x \in X$, и последовательность $\{X_k\}$ строго внутренних и выпуклых подмножеств множества X таких, что $X = \bigcup_{k \geqslant 1} X_k$, $X_k \subset X_{k+1}$, $k = 1, 2, \ldots$, для всех $k \geqslant 1$ и всех $m \geqslant k$ функция $f_m(x)$ выпукла на X_k, $f_m(x) \in C^2(X_k)$, $|f_m'(u) - f_m'(v)| \leqslant L|u - v|$ для любых $u, v \in X_k$, $\lim_{m\to\infty} f_m(u) = f(u)$, $\lim_{m\to\infty} f_m'(u) = f'(u)$ при всех $u \in X_k$. В силу доказанного тогда $|f_m'(u) - f_m'(v)|^2 \leqslant L\langle f_m'(u) - f_m'(v), u - v\rangle$ при всех $u, v \in X_k$ и всех $m \geqslant k$. Отсюда при $m \to \infty$ получим неравенство (19) на множестве X_k. Далее, при $k \to \infty$ убеждаемся в справедливости (19) для всех $u, v \in \operatorname{int} X$. Наконец, для граничных точек множества X (неравенство (19) доказывается с помощью предельного перехода от внутренних точек).

Как построить упомянутые последовательности $\{X_k\}$ и $\{f_k(x)\}$? В качестве $\{X_k\}$ может быть взята последовательность подмножеств всех внутренних точек множества X, удаленных от границы X на расстояние не менее чем $3\delta_k$, где $\lim \delta_k = 0$, $\delta_k > \delta_{k+1} > 0$, $k = 1, 2, \ldots$. В качестве $f_k(x)$ могут быть взяты средние функции Стеклова — Соболева [534], например,

$$f_k(x) = \int_{E^n} f(w)\omega_k(w - x)dw, \quad \omega_k(x) = \delta_k^{-n}\varkappa^{-1}\omega(|x|\delta_k^{-1}),$$

где $\omega(r) = \exp\{-1/(1 - r^2)\}$ при $|r| < 1$, $\omega(r) = 0$ при $|r| \geqslant 1$, $\varkappa = \int_{-1}^{1} \omega(r)dr$ и функция $f(x)$ вне X доопределена тождественным нулем. □

Приведем пример, показывающий, что условие $\operatorname{int} X \neq \varnothing$ в теореме 16 существенно [798].

Пример 8. Рассмотрим функцию $f(u) = xy$ на множестве $X = \{u = (x, y) \in E^2 : y = 0\}$. Так как $f(u) \equiv 0$ $\forall u \in X$, то ясно, что $f(u)$ выпукла на X. Далее, $f'(u) = (y, x)$, так что $f'(u) = (0, x)$ $\forall u \in X$. Возьмем произвольные точки $u = (x, 0)$, $v = (a, 0) \in X$, $u \neq v$. Тогда $|f'(u) - f'(v)| = |x - a|$, т. е. условие (17) выполнено с $L = 1$. Наконец, здесь $|f'(u) - f'(v)|^2 = |x - a|^2 > L\langle f'(u) - f'(v), u - v\rangle = 0$ $\forall u, v \in X$, $u \neq v$. Как видим, неравенство (19) не выполняется. Остается заметить, что здесь $\operatorname{int} X = \varnothing$.

11. Остановимся на одном замечательном свойстве выпуклых множеств, задаваемых ограничениями $g(x) \leqslant c$, где $g(x)$ — выпуклая функция.

Теорема 17. *Пусть X_0 — непустое выпуклое замкнутое множество из E^n, функция $g(x)$ выпукла и полунепрерывна снизу на X_0, пусть*

$$M(c) = \{u\colon u \in X_0,\ g(u) \leqslant c\}.$$

Тогда для ограниченности множества $M(c)$ при каждом c необходимо и достаточно, чтобы при некотором a множество $M(a)$ было непустым и ограниченным.

Доказательство. Необходимость. Она очевидна. Достаточность. Пусть $M(a) \neq \varnothing$ и это множество ограничено. Поскольку $M(c) \subset M(a)$ при всех $c \leqslant a$, то $M(c)$ ограничено при всех $c \leqslant a$ (пустое множество ограничено по определению). Остается рассмотреть случай $c > a$. Предположим, что при некотором $c > a$ множество $M(c)$ не ограничено. Заметим, что $M(c)$ выпукло и замкнуто, что следует из леммы 2.1.1 и теоремы 10.

Покажем, что существует вектор $e \neq 0$ такой, что $u + te \in M(c)$ при всех $t \geqslant 0$ и всех $u \in M(c)$ (направление, задаваемое таким вектором e, принято называть *рецессивным направлением* неограниченного выпуклого множества).

Поскольку множество $M(c)$ не ограничено по предположению, то существует последовательность $\{u_k\} \in M(c)$ такая, что $|u_k| \to \infty$ при $k \to \infty$. Возьмем какую-либо точку $\bar{u} \in M(c)$ и построим вектор $e_k = (u_k - \bar{u}) \times |u_k - \bar{u}|^{-1}$, $k = 1, 2, \ldots$. По теореме Больцано — Вейерштрасса из последовательности $\{e_k\}$ можно выбрать подпоследовательность $\{e_{k_m}\}$, сходящуюся к некоторому вектору e, $|e| = 1$.

Возьмем произвольное $t > 0$. Поскольку $0 < t/|u_k - \bar{u}| < 1$ при всех $k \geqslant k_0$, то в силу выпуклости $M(c)$ имеем

$$\bar{u} + te_k = \frac{t}{|u_k - \bar{u}|}u_k + \left(1 - \frac{t}{|u_k - \bar{u}|}\right)\bar{u} \in M(c), \quad k \geqslant k_0.$$

Отсюда при $k_m \to \infty$ получим $\bar{u} + te \in M(c)$, так как $M(c)$ замкнуто. В силу произвольности $t > 0$ заключаем, что $\bar{u} + te \subset M(c)$ при всех $t \geqslant 0$.

Теперь возьмем любую точку $u \in M(c)$ и покажем, что $u + te \in M(c)$ при каждом $t \geqslant 0$. По доказанному $\bar{u} + \mu e \in M(c)$, $\mu \geqslant 0$. В силу выпуклости $M(c)$ тогда

$$\frac{t}{\mu}(\bar{u} + \mu e) + \left(1 - \frac{t}{\mu}\right)u = u + te + \frac{t(\bar{u} - u)}{\mu} \in M(c)$$

при всех $\mu > t$. Отсюда при $\mu \to \infty$ с учетом замкнутости $M(c)$ получим $u + te \in M(c)$ при каждом $t \geqslant 0$.

Зафиксируем какую-либо точку $v \in M(a) \subset M(c)$. В силу построения вектора e тогда $v + te \in M(c)$, $t \geqslant 0$. По условию множество $M(a)$ ограничено, поэтому луч $\{v + te, t \geqslant 0\}$ пересекает границу выпуклого множества $M(a)$ в некоторой точке, или, точнее говоря, найдется число $t_0 = \sup\{t\colon v + te \in M(a)\}$ такое, что $v + te \notin M(a)$ при всех $t > t_0$. Это значит, что $v + te \in X_0$, но $c \geqslant g(v + te) > a \geqslant g(v)$ для всех $t > t_0$.

Зафиксируем какое-либо $t > t_0$. Тогда, пользуясь представлением

$$v + te = \lambda\left(v + \frac{t}{\lambda}e\right) + (1 - \lambda)v, \quad 0 < \lambda < 1,$$

и выпуклостью функции $g(x)$, имеем $g(v+te) \leqslant \lambda g(v+(t/\lambda)e)+(1-\lambda)g(v)$, или $g(v+(t/\lambda)e) \geqslant (g(v + te) - g(v))/\lambda + g(v)$, $0 < \lambda < 1$. Поскольку $g(v + te) > g(v)$, то при $\lambda \to +0$ отсюда

получим $g(v+(t/\lambda)e)\to\infty$. Тогда найдется число $\lambda_0>0$ такое, что $g(v+(t/\lambda)e)>c$ при всех λ, $0<\lambda<\lambda_0$. С другой стороны, по построению вектора e имеем $v+(t/\lambda)e\in M(c)$ или $g(v+(t/\lambda)e)\leqslant c$ при всех λ, $0<\lambda<1$. Полученное противоречие доказывает теорему. □

Отметим, что требование полунепрерывности снизу функции в теореме 17 существенно (см. ниже упражнение 21).

12. Наконец, получим оценку снизу скорости роста выпуклой функции для случая, когда множество точек ее минимума ограничено.

Теорема 18. *Пусть $f(x)$ — выпуклая функция на $X=E^n$, $f_*=\inf\limits_{E^n} f(x)>-\infty$, $X_*=\{u\in E^n: f(u)=f_*\}\neq\varnothing$, причем X_* — ограниченное множество, т. е. существует такое число $R>0$, что $X_*\subset S_*=\{u\in E^n: |u-u_*|<R\}$, где u_* — какая-либо фиксированная точка из X_*. Тогда $\lim\limits_{|u|\to\infty} f(u)=\infty$ и, более того, верна оценка*

$$f(u)\geqslant |u-u_*|\frac{f_{*R}-f_*}{R}+f_* \quad \forall u\notin S_*, \tag{24}$$

где $f_{*R}=\inf\limits_{u\in \operatorname{Гр} S_*} f(u)>f_*$, $\operatorname{Гр} S_*=\{u\in E^n: |u-u_*|=R\}$.

Доказательство. Возьмем любую точку $u\notin S_*$. Поскольку

$$v=u_*+R\frac{u-u_*}{|u-u_*|}=\frac{R}{|u-u_*|}u+\Big(1-\frac{R}{|u-u_*|}\Big)u_*\in \operatorname{Гр} S_*,$$

то с учетом выпуклости $f(u)$ имеем

$$f_{*R}\leqslant f(v)\leqslant \frac{R}{|u-u_*|}f(u)+\Big(1-\frac{R}{|u-u_*|}\Big)f(u_*).$$

Отсюда получаем требуемое неравенство (24). Согласно теореме 15 функция $f(x)$ непрерывна на E^n, и в силу теоремы 2.1.1 на замкнутом ограниченном множестве $\operatorname{Гр} S_*$ она достигает своей нижней грани хотя бы в одной точке. Отсюда и из того, что замкнутые множества X_* и $\operatorname{Гр} S_*$ не пересекаются, следует, что $f_{*R}>f_*$. Тогда из оценки (24) имеем $\lim\limits_{|u|\to\infty} f(u)=\infty$. □

Отметим, что для функции $f(u)=|u|$, $u\in E^n$, неравенство (24) превращается в тождественное равенство. Это значит, что оценка (24) на классе выпуклых функций является точной.

Некоторые другие свойства выпуклых функций и множеств будут рассмотрены ниже.

Существуют более широкие, чем выпуклые, классы функций, которые наследуют некоторые важные свойства выпуклых функций [774, 806]. Некоторые такие классы приведены ниже в упражнениях 33, 34.

Упражнения

1. При каких a, b, c функция $f(u)=ax^2+2bxy+cy^2$ переменных $u=(x,y)\in E^2$ будет выпуклой на E^2? Вогнутой на E^2?

2. Найти области выпуклости и вогнутости функций $f(u)=\sin(x+y+z)$, $f(u)=\sin(x^2+y^2+z^2)$.

3. При каких p, q функция $f(u)=x^p y^q$ будет выпуклой (или вогнутой) на множестве $X=\{u=(x,y)\in E^2: x>0, y>0\}$? Аналогичное исследование провести для функции $f(u)=x^p y^q z^r$ на $X=\{u=(x,y,z): x>0, y>0, z>0\}$.

4. Если функция $f(u)$ выпукла, то будет ли выпуклой функция $|f(u)|$?

5. Если функция $f(x)$ выпукла на E^m, а A — матрица размера $m\times n$, то функция $g(u)=f(Au)$ выпукла на E^m. Доказать.

6. Если $f_1(u)$, $f_2(u)$ выпуклы, то будет ли их произведение выпуклой функцией? Рассмотреть пример $f_1(u)=u$, $f_2(u)=u^2$. Что изменится, если от функций $f_1(u)$, $f_2(u)$ потребовать неотрицательности? Или монотонности?

7. Пусть функции $f_k(u)$ выпуклы на выпуклом множестве X при всех $k=0,1,\ldots$ и пусть существует предел $\lim\limits_{k\to\infty} f_k(u)=f(u)$ или сходится ряд $\sum\limits_{k=0}^{\infty} f_k(u)=f(u)$ при всех $u\in X$. Доказать, что функция $f(u)$ выпукла на X.

8. Выяснить, когда в неравенстве (2) возможно равенство, если $f(u)$ — строго выпуклая функция.

9. Доказать неравенства:

а) $\sqrt[m]{x_1\ldots x_m}\leqslant\frac{1}{m}(x_1+\ldots+x_m)$, $x_1\geqslant 0,\ldots, x_m\geqslant 0$;

б) $(x_1+\ldots+x_m)^n\leqslant m^{n-1}(x_1^n+\ldots+x_m^n)$, $x_1\geqslant 0,\ldots, x_m\geqslant 0$, $n\geqslant 1$;

в) $(x_1+\ldots+x_m)(x_1^{-1}+\ldots+x_m^{-1})\geqslant m^2$, $x_1>0,\ldots, x_m>0$;

г) $\sum\limits_{i=1}^m |a_i||b_i|\leqslant\Big(\sum\limits_{i=1}^m |a_i|^p\Big)^{1/p}\Big(\sum\limits_{i=1}^m |b_i|^q\Big)^{1/q}$ $\forall a_i, b_i\in\mathbb{R}$, $\frac{1}{p}+\frac{1}{q}=1$, $p>1$, $q>1$;

д) $(a_1\ldots a_m)^{1/m}+(b_1\ldots b_m)^{1/m}\leqslant((a_1+b_1)\ldots(a_m+b_m))^{1/m}$ $\forall a_i\geqslant 0$, $b_i\geqslant 0$.

Указание: воспользоваться неравенством (2) для выпуклых функций $f(u)=-\ln u$, $u>0$; $f(u)=u^n$, $u\geqslant 0$, $n\geqslant 1$; $f(u)=u^{-1}$, $u>0$, $f(u)=u^p$, $u\geqslant 0$, $p>1$; $f(u)=\ln(1+e^u)$ при подходящим образом выбранных α_i, x_i, $i=1,\ldots,m$. Выяснить, при каких условиях в этих неравенствах возможно равенство.

10. Пусть функция $f(u)$, $u\in E^n$, такова, что $f(\alpha u+(1-\alpha)v)\leqslant\alpha f(u)+(1-\alpha)f(v)$ при всех $u, v\in E^n$, $\alpha\in\mathbb{R}$. Проверить, что аффинная функция $f(u)=<a,u>+b$, $a\in E^n$, $b\in E^1$, удовлетворяет этому неравенству. Существуют ли другие функции, обладающие этим свойством?

11. Для того чтобы функция $f(x)$ была строго выпуклой на выпуклом множестве X, необходимо и достаточно выполнения неравенства (4) (а в случае $f(x)\in C^1(X)$ — неравенства (6)), которое может обратиться в равенство лишь при $u=v$. Доказать.

12. Доказать, что если X — выпуклое множество, $f(x)\in C^2(X)$ и неравенство (8) является строгим при всех $\xi\in\operatorname{Lin} X$, $\xi\neq 0$, то функция $f(x)$ строго выпукла на X. Верно ли обратное утверждение? Рассмотреть пример $f(x)=x^4$, $x\in E^1$.

13. Пусть X — выпуклое множество, $f(x)$ выпукла на X и $f(x)\in C^1(X)$. Доказать, что тогда критерий оптимальности (5) равносилен неравенству $\langle f'(u), u-u_*\rangle\geqslant 0$ при всех $u\in X$.

14. Пусть $f(x)$ — выпуклая функция на выпуклом множестве X, $f(x)\in C^1(X)$ и $f_*=\inf\limits_X f(x)>-\infty$. Доказать, что для того чтобы некоторая последовательность $\{u_k\}\in X$ была минимизирующей, т. е. $\lim\limits_{k\to\infty} f(u_k)=f_*$, необходимо и достаточно выполнения условия $\lim\limits_{k\to\infty}\langle f'(u_k), u-u_k\rangle\geqslant 0$ при всех $u\in X$ (ср. с теоремой 3).

15. Пусть строго выпуклая функция $f(x)$ достигает на E^n своей нижней грани. Доказать, что тогда $\lim\limits_{|u|\to\infty} f(u)=\infty$.

Указание: воспользоваться теоремами 1, 18.

16. Пусть функция $f(x)$ выпукла и полунепрерывна снизу на выпуклом замкнутом множестве X из E^n, $f_*>-\infty$, $X_*\neq\varnothing$, причем $X_*\subset S_*=\{u\in X: |u-u_*|<R\}$, где u_* — какая-либо фиксированная точка из X_*. Тогда

$$f(u)\geqslant |u-u_*|\frac{f_{*R}-f_*}{R}+f_* \qquad \forall u\in X,\ u\notin S_*,$$

где $f_{*R}=\inf\limits_{\operatorname{Гр} S_*} f(u)>f_*$, $\operatorname{Гр} S_*=\{u\in X: |u-u_*|=R\}$. Доказать, пользуясь схемой доказательства теоремы 18.

17. Выпуклая функция, отличная от постоянной, может достигать своей верхней грани на выпуклом множестве лишь в его граничных точках. Доказать.

18. Для того чтобы функция $\rho(u,X)=\inf\limits_{v\in X}|u-v|$ была выпуклой на E^n, необходимо и достаточно, чтобы замыкание множества X было выпуклым. Доказать.

19. Пусть X — ограниченное множество из E^n. Доказать, что функция $\delta(c,X)=\sup\limits_{u\in X}\langle c,u\rangle$ переменной $c\in E^n$, называемая *опорной функцией* множества X, выпукла на E^n.

20. Пусть X — выпуклое множество из E^n, $0\in\operatorname{int} X$. Доказать, что функция $\mu(u,X)=\inf\limits_{\alpha\in A_u}\alpha$, $A_u=\{\alpha: \alpha>0, u/\alpha\in X\}$, называемая *функцией Минковского*, выпукла на E^n.

21. Пусть $X=\{u=(x,y): x\geqslant 0, y\geqslant 0\}=E^2_+$. Показать, что функция

$$g(u)=\begin{cases} y, & x\geqslant 0,\ y>0 \text{ или } 0\leqslant x\leqslant 1,\ y=0,\\ x-1, & x>1,\ y=0,\end{cases}$$

выпукла и полунепрерывна сверху, но не является полунепрерывной снизу на X. Убедиться,

что множество $M(c)=\{u\in X\colon g(u)\leqslant c\}$ ограничено при $c=0$ и не ограничено при всех $c>0$ (ср. с теоремой 17). Показать, что $M(c)$ не замкнуто при каждом $c>0$.

22. Множество $X_c=\{u\in E^n\colon g_i(u)\leqslant c,\ i=1,\ldots,m\}$, где $g_i(u)$ — выпуклая функция на E^n, будет ограничено при любых c тогда и только тогда, когда X_c ограничено хотя бы при одном значении $c=c_0$. Доказать.

23. Пусть X — неограниченное замкнутое выпуклое множество из E^n. Доказать, что

1) для любой точки $v\in X$ существует ненулевой вектор e такой, что луч $\{u=v+te,\ t\geqslant 0\}\in X$;

2) если луч $\{u=v+te,\ t\geqslant 0\}\in X$ при некотором $v\in X$, то луч $\{u=w+te,\ t\geqslant 0\}\in X$ при всех $w\in X$. Показать, что требование замкнутости X существенно для обоих утверждений, рассмотрев множество $X=\{u=(x,y)\colon 0<x<1\}\cup\{(0,0)\}$.

Указание: воспользоваться рассуждениями из доказательства теоремы 17.

24. Доказать, что функция

$$f(u)=\begin{cases} x^2/y, & y\neq 0,\\ 0, & y=0.\end{cases}$$

выпукла на множестве $X=\{u=(x,y)\colon y>0\}\cup\{(0,0)\}$ и полунепрерывна снизу на X. Убедиться, что $f(u)$ не является полунепрерывной сверху в точке $u_0=(0,0)$, и, более того, показать, что для любого числа $A\geqslant 0$ существует такая последовательность $\{u_k\}\in X$, $\{u_k\}\to 0$, что $\lim\limits_{k\to\infty} f(u_k)=A$.

25. Пусть $X=\{u\in E^2\colon Au\leqslant b\}$ — многогранное множество, функция $f(u)$ выпукла на X. Доказать, что $f(u)$ полунепрерывна сверху на X [617, стр. 101].

26. Пусть функция $f(x)$ выпукла и ограничена сверху на $E^n_+=\{u=(u^1,\ldots,u^n)\in E^n\colon u^1\geqslant 0,\ldots,u^n\geqslant 0\}$. Доказать, что $f(x)$ монотонна и не возрастает на E^n_+ по каждой переменной.

27. Доказать, что если выпуклая функция $f(x)$ на E^n ограничена сверху, то $f(x)$ постоянна.

28. Пусть $f(u)$ — выпуклая дифференцируемая функция на отрытом выпуклом множестве W из E^n. Доказать, что тогда градиент $f'(u)=(\partial f(u)/\partial u^1,\ldots,\partial f(u)/\partial u^n)$ непрерывен на W [617, стр. 263] (см. теорему 6.7).

29. Пусть $f(x)$ — выпуклая функция на выпуклом множестве X из E^n. Доказать, что $f(x)$ удовлетворяет условию Липшица на каждом ограниченном множестве V, замыкание которого принадлежит ri X. [617, стр. 103].

30. Пусть $f(x)$ — выпуклая функция на открытом выпуклом множестве W из E^n. Доказать, что $f(x)$ почти всюду на W дифференцируема [617, стр. 262].

31. Пусть X_0 — выпуклое замкнутое множество из E^n, $g(u)$ — выпуклая функция на X_0, $X=\{u\in X_0\colon g(u)\leqslant 0\}$. Пусть $\{u_k\}\in X_0$, $\{g(u_k)\}\to 0$. Можно ли утверждать, что $\{\rho(u_k,X)\}\to 0$? Рассмотреть пример $X_0=\{u=(x,y)\in E^2\colon x\geqslant 1\}$, $g(u)=y^2/x$, $u_k=(k,\sqrt[4]{k})$, $k=1,2,\ldots$.

32. Пусть X — выпуклое замкнутое множество из E^n, $f(u)$ — выпуклая непрерывная функция на X, $f_*>-\infty$, $X_*\neq\varnothing$. Пусть $\{u_k\}\in X$, $\{f(u_k)\}\to f_*$. Можно ли ожидать, что $\{\rho(u_k,X_*)\}\to 0$? Рассмотреть пример $X=\{u=(x,y)\in E^2\colon x\geqslant 1\}$, $f(u)=y^2/x$, $u_k=(k,\sqrt[4]{k})$, $k=1,2,\ldots$.

33. Пусть X — выпуклое множество. Функция $f(x)$ называется *квазивыпуклой* на множестве X, если

$$f(\alpha x+(1-\alpha)y)\leqslant\max\{f(x);f(y)\}\quad\forall x,y\in X\forall\alpha\in[0,1].$$

Доказать, что [774; 806]:

1) если $f(x)$ выпукла на X, то $f(x)$ квазивыпукла на X. Показать, что функция $f(x)=x^3$ квазивыпукла на любом отрезке $a\leqslant x\leqslant b$, где $a<0$, но невыпукла на $[a,b]$;

2) функция $f(x)$ квазивыпукла на X тогда и только тогда, когда множество Лебега $M(x)=\{y\in X\colon f(y)\leqslant f(x)\}$ выпукло при всех $x\in X$ (ср. с теоремой 10);

3) унимодальная функция на отрезке $[a,b]$ (определение 1.1.7) квазивыпукла на $[a,b]$;

4) привести пример квазивыпуклой функции, имеющей разрывы во внутренних точках множества X (ср. с теоремой 15);

5) будет ли сумма двух квазивыпуклых функций квазивыпуклой? Рассмотреть пример: $f_1(x)=x^3$, $f_2(x)=-3x$, $x\in X=E^1$;

6) если $f(x)\in C^1(X)$, то $f(x)$ квазивыпукла на X тогда и только тогда, когда для всех $x,y\in X$, для которых $f(x)\leqslant f(y)$, справедливо неравенство $\langle f'(y),y-x\rangle\geqslant 0$;

7) привести пример квазивыпуклой функции, для которой выполнение неравенства (5) не гарантирует, что $x_*\in X_*$;

8) можно ли утверждать, что точка локального минимума квазивыпуклой функции $f(x)$, $x\in X$, является точкой ее глобального минимума на X? Рассмотреть пример: $f(x)=0$ при $|x|\leqslant 1$, $f(x)=-(x+1)^2$ при $x<-1$, $f(x)=(x-1)^2$ при $x>1$, $X=E^1$.

34. Пусть X — выпуклое множество. Функция $f(x)\in C^1(X)$ называется *псевдовыпуклой* на X, если для всех $x,y\in X$, для которых $\langle f'(x),y-x\rangle\geqslant 0$, справедливо неравенство $f(y)\geqslant f(x)$ (ср. с п. 6 упражнения 33).

Доказать, что [315; 774; 806]:

1) если $f(x)$ выпукла на X, $f(x)\in C^1(X)$, то $f(x)$ псевдовыпукла на X. Указание: воспользоваться неравенством (4): $\langle f'(x),y-x\rangle\leqslant f(y)-f(x)\ \forall x,y\in X$. Показать, что функция $f(x)=-x^2$, $x\in X=\{x\in E^1\colon x<0\}$ псевдовыпукла, но невыпукла на X;

2) если $f(x)$ псевдовыпукла на X, то $f(x)$ квазивыпукла на X. Показать, что функция $f(x)=-x^2$, $x\in X=\{x\in E^1\colon x\leqslant 0\}$ квазивыпукла, но не является псевдовыпуклой на X;

3) если $f(x)$ псевдовыпукла на X, то для того чтобы $x_*\in X_*$, необходимо и достаточно, чтобы $\langle f'(x_*),x-x_*\rangle\geqslant 0\ \forall x\in X$ (ср. с теоремой 3 и с п. 7 упражнения 33);

4) всякая точка локального минимума псевдовыпуклой функции $f(x)$ на X является точкой ее глобального минимума на X (ср. с теоремой 1 и с п. 8 упражнения 33);

5) будет ли сумма двух псевдовыпуклых функций псевдовыпуклой? Рассмотреть пример: $f_1(x)=x^3+x$, $f_2(x)=-x$, $x\in X=E^1$;

6) функция $f(x)$ псевдовыпукла на X тогда и только тогда, когда для всех $x,y\in X$, для которых $f(y)<f(x)$, справедливо неравенство $\langle f'(x),y-x\rangle<0$;

7) если $f(x)$ псевдовыпукла на $X\subseteq E^1$, то $f(x)$ — унимодальная функция (определение 1.1.7). Верно ли обратное утверждение?

8) дробно-рациональная функция $f(x)=\dfrac{ax+b}{cx+d}$, $ad-bc\neq 0$, псевдовыпукла на любом отрезке $[a,b]$, не содержащем точку $x=-\dfrac{d}{c}$;

9) для гладких выпуклых функций из (4) следует двойное неравенство

$$\langle f'(x),y-x\rangle\leqslant f(y)-f(x)\leqslant\langle f'(y),y-x\rangle\quad\forall x,y\in X.$$

Автор полагает, что левое из этих неравенств подсказывает определение псевдовыпуклой функции, а правое — определение квазивыпуклой функции (в форме п. 6 упражнения 33). Подумайте над этим эвристическим соображением.

35. Пусть $f(x)\in C^{1,1}(E^n)$ (см. определение 2.6.3). Доказать, что [234; 525]:

$$f_*=\inf_{x\in X} f(x)\leqslant f(x)-\frac{1}{2L}|f'(x)|^2\quad\forall x\in E^n. \tag{25}$$

Указание: в неравенстве (2.6.7) принять $x=y-\dfrac{1}{L}f'(y)$.

36. Пусть функция $f(x)\in C^{1,1}(E^n)$ и выпукла на E^n. Доказать, что [234; 525]:

$$\frac{1}{2L}|f'(x)-f'(y)|^2\leqslant f(x)-f(y)-\langle f'(y),x-y\rangle\quad\forall x,y\in E^n. \tag{26}$$

Указание: к функции $g(x)=f(x)-f(y)-\langle f'(y),x-y\rangle$ применить неравенство (25); убедиться, что $g_*=\inf\limits_{x\in E^n} g(x)=g(y)=0$. Сравните неравенство (26) с (4).

37. Для выпуклых функций $f(x)\in C^{1,1}(E^n)$ доказать неравенство (19), опираясь на неравенство (26) [234; 525]. Указание: в (26) поменять ролями x,y и сложить получившееся неравенство с (26).

38. Остается ли верной теорема 5, если функция $f(x)$ в некоторых точках множества X не имеет второй производной, но непрерывна в них? Рассмотреть примеры: $f(x)=(|x|-1)^2$, $f(x)=-|x|$, $x\in E^1$.

39. Доказать, что каждая дважды непрерывно дифференцируемая функция $f(x)$ на компактном множестве представима в виде разности двух выпуклых функций. Указание: рассмотреть функции $f_1(x)=f(x)+ax^2$, $f_2(x)=ax^2$, где a достаточно большое положительное число.

40. Доказать, что всякая непрерывная функция на компактном множестве является пределом равномерно сходящейся последовательности функций, представимых в виде разности двух выпуклых функций.

41. Пусть функция $f(x)$ конечна и выпукла на E^n, множество X непусто и выпукло, $\{x_k\}$ — произвольная последовательность из E^n, для которой $\lim\limits_{k\to\infty}\rho(x_k, X)=0$. Доказать, что тогда $\varliminf\limits_{k\to\infty} f(x_k)\geqslant f_*=\inf\limits_{x\in X} f(x)$ [85].

42. Пусть X — выпуклое замкнутое множество из E^n, функция $f(x)$ выпукла и полунепрерывна снизу на X, пусть $f_*>-\infty$, $X_*\neq\varnothing$, X_* — ограничено. Тогда для $\forall\alpha>0$ множество $M(f_*+\alpha)=\{x\in X: f(x)<f_*+\alpha\}$ ограничено и каждая минимизирующая последовательность $\{x_k\}$ сходится к X_*. Доказать. Указание: воспользоваться теоремой 17.

§ 3. Сильно выпуклые функции

1. Непрерывная выпуклая функция на выпуклом замкнутом множестве может не достигать своей нижней грани на этом множестве. Например, если $f(x)=1/x$, $X=\{x\in E^1: x\geqslant 1\}$, то $f_*=\inf\limits_X fx)=0$, но $f(x)>0$ при всех $x\in X$. Однако можно выделить подкласс выпуклых функций, для которых подобная ситуация невозможна.

Определение 1. Функция $f(x)$, определенная на выпуклом множестве X, называется *сильно выпуклой* на X, если существует постоянная $\varkappa>0$ такая, что

$$f(\alpha u+(1-\alpha)v)\leqslant\alpha f(u)+(1-\alpha)f(v)-\tfrac{1}{2}\alpha(1-\alpha)\varkappa|u-v|^2 \qquad (1)$$

при всех $u, v\in X$ и всех α, $0\leqslant\alpha\leqslant 1$. Постоянную $\varkappa$ называют *постоянной сильной выпуклости* функции $f(x)$ на множестве X.

Очевидно, сильно выпуклая на X функция будет выпуклой и даже строго выпуклой на X. Примером сильно выпуклой функции на всем пространстве E^n может служить функция

$$\Omega(x)=\langle x, x\rangle=|x|^2, \quad x\in E^n.$$

Для этой функции неравенство (1) превращается в тождественное равенство с постоянной $\varkappa=2$:

$$|\alpha u+(1-\alpha)v|^2=\alpha|u|^2+(1-\alpha)|v|^2-\alpha(1-\alpha)|u-v|^2 \qquad (2)$$

при всех $u, v\in E^n$, $\alpha\in[0,1]$. Линейная функция $f(x)=\langle c, x\rangle$ выпукла на E^n, но не сильно выпукла. Упомянутая выше функция $f(x)=1/x$ при $x\geqslant 1$ выпукла, но не сильно выпукла.

Нетрудно видеть, что сумма выпуклой функции на выпуклом множестве X и сильно выпуклой функции на том же множестве с постоянной $\varkappa$ будет сильно выпуклой функцией на X с той же постоянной $\varkappa$. Если $f(x)$ сильно выпукла на X с постоянной $\varkappa$, то $g(x)=cf(x)$ при любом $c=\mathrm{const}>0$ будет сильно выпуклой на X с постоянной $c\varkappa$.

Теорема 1. *Пусть множество X выпукло и замкнуто, а функция $f(x)$ сильно выпукла и полунепрерывна снизу на X. Тогда:*

1) *множество Лебега*

$$M(v)=\{u: u\in X,\ f(u)\leqslant f(v)\}$$

выпукло, замкнуто и ограничено при всех $v\in X$;

2) $f_*=\inf\limits_X f(u)>-\infty$, *множество* $X_*=\{u: u\in X,\ f(u)=f_*\}$ *непусто и, более того, состоит из единственной точки u_*;*

3) *имеет место неравенство*

$$\tfrac{1}{2}\varkappa|u-u_*|^2\leqslant f(u)-f(u_*) \qquad \forall u\in X; \qquad (3)$$

4) *любая минимизирующая последовательность* $\{u_k\}$: $u_k\in X$, $k=1,2,\ldots$, $\lim\limits_{k\to\infty} f(u_k)=f_*$, *сходится к точке u_*.*

Сформулированная теорема является обобщением теоремы Вейерштрасса 2.1.1. В отличие от теоремы 2.1.1, здесь на функцию накладывается более жесткое ограничение, но зато от множества X не требуется ограниченности. В частности, в теореме 1 возможно $X=E^n$.

Доказательство. Если множество X ограничено, замкнуто, т. е. компактно, то все утверждения теоремы 1, кроме неравенства (3), следуют из теоремы 2.1.1. Поэтому пусть X — неограниченное множество. Возьмем произвольную точку $v\in X$ и рассмотрим множество

$$S=S(v,1)=\{u: u\in X,\ |u-v|\leqslant 1\}.$$

Из теоремы 2.1.1 следует, что $\inf\limits_S f(u)=f_{*S}>-\infty$, так что

$$f(u)\geqslant f_{*S}=f(v)-\nu \quad \forall u\in S, \quad \nu=f(v)-f_{*S}\geqslant 0. \qquad (4)$$

Возьмем произвольную точку $u\in X\setminus S$, т. е. $u\in X$, $|u-v|>1$. Тогда

$$0<\alpha_0=1/|u-v|<1. \qquad (5)$$

При $\alpha=\alpha_0$ из (1) получаем

$$\alpha_0 f(u)\geqslant f(v+\alpha_0(u-v))-(1-\alpha_0)f(v)+\alpha_0(1-\alpha_0)\varkappa|u-v|^2/2. \qquad (6)$$

В силу (5) $\alpha_0|u-v|=1$, поэтому $v+\alpha_0(u-v)\in S$. Согласно (4) тогда $f(v+\alpha_0(u-v))\geqslant f(v)-\nu$. Учитывая эту оценку, из (6) получаем

$$\alpha_0 f(u)\geqslant\alpha_0 f(v)-\nu+\alpha_0(1-\alpha_0)\varkappa|u-v|^2/2.$$

Отсюда, сокращая на $\alpha_0>0$ и вспоминая определение (5) величины α_0, получаем

$$f(u)\geqslant f(v)+(1-\alpha_0)\varkappa|u-v|^2/2-\nu/\alpha_0=$$
$$=f(v)+\varkappa|u-v|^2/2-\left(|u-v|\sqrt{\tfrac{\varkappa}{2}}\right)\left(\sqrt{\tfrac{\varkappa}{2}}+\nu\sqrt{\tfrac{2}{\varkappa}}\right)$$

Применяя к последнему слагаемому неравенство $ab\leqslant(a^2+b^2)/2$, будем иметь

$$f(u)\geqslant f(v)+\varkappa|u-v|^2/4-\left(\sqrt{\tfrac{\varkappa}{2}}+\nu\sqrt{\tfrac{2}{\varkappa}}\right)^2/2 \qquad (7)$$

для всех $u\in X\setminus S$. Нетрудно видеть, что неравенство (7) справедливо и при $u\in S$. В самом деле, если $u\in S$, т. е. $|u-v|\leqslant 1$, то

$$\nu\leqslant\nu+\frac{\nu^2}{\varkappa}=\tfrac{1}{2}\left(\sqrt{\tfrac{\varkappa}{2}}+\nu\sqrt{\tfrac{2}{\varkappa}}\right)^2-\tfrac{\varkappa}{4}\leqslant\tfrac{1}{2}\left(\sqrt{\tfrac{\varkappa}{2}}+\nu\sqrt{\tfrac{2}{\varkappa}}\right)^2-\tfrac{\varkappa}{4}|u-v|^2 \quad \forall u\in S.$$

Отсюда и из (4) следует справедливость (7) и для $u\in S$.

Таким образом, неравенство (7) имеет место для всех $u\in X$. Для любых $u\in M(v)$ из (7) следует

$$\varkappa|u-v|^2/4-\left(\sqrt{\tfrac{\varkappa}{2}}+\nu\sqrt{\tfrac{2}{\varkappa}}\right)^2/2\leqslant f(u)-f(v)\leqslant 0$$

или

$$|u-v|\leqslant 1+2\nu/\varkappa \quad \forall u\in M(v).$$

Ограниченность $M(v)$ доказана. Выпуклость $M(v)$ следует из теоремы 2.10, а замкнутость $M(v)$ — из леммы 2.1.1. Из теоремы 2.1.2 имеем $f_* > -\infty$, $X_* \neq \varnothing$. Поскольку сильно выпуклая функция строго выпукла, то в силу теоремы 2.1 множество X_* состоит из единственной точки u_*.

Докажем неравенство (3). Учитывая, что $f(u_*) \leqslant f(u)$ при всех $u \in X$, из (1) имеем $0 \leqslant f(\alpha u + (1-\alpha)u_*) - f(u_*) \leqslant \alpha(f(u) - f(u_*)) - \alpha(1-\alpha)\varkappa|u - u_*|^2/2$, или $\frac{1}{2}\alpha(1-\alpha)\varkappa|u-u_*|^2 \leqslant \alpha(f(u)-f(u_*))$ при всех $\alpha \in [0, 1]$ и $u \in X$. Деля на $\alpha > 0$ и устремляя $\alpha \to +0$, отсюда получаем неравенство (3).

Наконец, пусть $\{u_k\} \in X$, $\lim\limits_{k\to\infty} f(u_k) = f_*$. Полагая в (3) $u = u_k$ получаем $\frac{1}{2}\varkappa|u_k - u_*|^2 \leqslant f(u_k) - f_*$, $k = 1, 2, \ldots$. Отсюда при $k \to \infty$ следует, что $|u_k - u_*| \to 0$. Теорема 1 доказана. □

Замечание 1. При выполнении условий теоремы 1 утверждение $f_* > -\infty$, $X_* \neq \varnothing$ остается верным для любого замкнутого (необязательно выпуклого) подмножества $W \subseteq X$ — это следует из замкнутости и ограниченности $M(v) = \{u\colon u \in W,\ f(u) \leqslant f(v)\}$ и теоремы 2.1.2.

2. Укажем теперь критерии сильной выпуклости для гладких функций, аналогичные теоремам 2.2. 2.4, 2.5. Сначала докажем одно вспомогательное утверждение.

Лемма 1. *Пусть X — выпуклое множество. Функция $f(x)$ сильно выпукла на X с постоянной сильной выпуклости $\varkappa > 0$ тогда и только тогда, когда функция $g(u) = f(u) - \varkappa|u|^2/2$ выпукла на X.*

Доказательство. Необходимость. Пусть функция $f(x)$ сильно выпукла на X, т. е. выполнено неравенство (1). Умножим равенство (2) на $\varkappa/2 > 0$ и почленно вычтем получившееся равенство из (1). Будем иметь неравенство, которое с помощью функции $g(u)$ можно написать в виде

$$g(\alpha u + (1-\alpha)v) \leqslant \alpha g(u) + (1-\alpha)g(v) \quad \forall u, v \in X, \quad \alpha \in [0, 1]. \tag{8}$$

Это значит, что $g(u)$ выпукла на X.

Достаточность. Пусть функция $g(u)$ выпукла на X, т. е. выполняется (8). Сложив (8) с равенством (2), умноженным на $\varkappa/2 > 0$, придем к неравенству (1). Это значит, что $f(x)$ сильно выпукла на X. □

Теорема 2. *Пусть X — выпуклое множество. Если функция $f(x)$ сильно выпукла на X и дифференцируема в точке $v \in X$, то*

$$f(u) \geqslant f(v) + \langle f'(v), u - v\rangle + \varkappa|u - v|^2/2 \quad \forall u, v \in X. \tag{9}$$

Если $f(x) \in C^1(X)$, то она сильно выпукла на X тогда и только тогда, когда $f(x)$ удовлетворяет неравенству (9).

Доказательство. Заметим, что для сильно выпуклой функции $\Omega(x) = |x|^2$ неравенство (9) превращается в легко проверяемое тождественное равенство с постоянной $\varkappa = 2$.

$$|u|^2 = |v|^2 + \langle 2v, u - v\rangle + |u - v|^2 \quad \forall u, v \in X. \tag{10}$$

Теперь нетрудно доказать, что условие (9) равносильно неравенству

$$g(u) \geqslant g(v) + \langle g'(v), u - v\rangle \quad \forall u, v \in X, \tag{11}$$

где $g(u) = f(u) - \varkappa|u|^2/2$, $g'(u) = f'(u) - \varkappa u$. В самом деле, если умножить равенство (10) на $\varkappa/2$ и сложить с (11), то получим неравенство (9). И обратно: вычитая из (9) равенство (10), умноженное на $\varkappa/2$, приходим к (11).

Из равносильности неравенств (9) и (11), из леммы 1 и теоремы 2.2 следует утверждение теоремы. □

Теорема 3. *Пусть X — выпуклое множество, $f(x) \in C^1(X)$. Тогда для сильной выпуклости функции $f(x)$ на X необходимо и достаточно существования такой постоянной $\mu > 0$, что*

$$\langle f'(u) - f'(v), u - v\rangle \geqslant \mu|u - v|^2, \quad \forall u, v \in X. \tag{12}$$

Доказательство. Легко проверить, что неравенство (12) равносильно неравенству

$$\langle g'(u) - g'(v), u - v\rangle \geqslant 0 \qquad \forall u, v \in X$$

для функции $g(u) = f(u) - \varkappa|u|^2/2$, где $\varkappa = \mu$. Отсюда и из леммы 1, теоремы 2.4 следует утверждение теоремы. □

Теорема 4. *Пусть X выпуклое множество из E^n, $f(x) \in C^2(X)$. Тогда для сильной выпуклости функции $f(x)$ на X необходимо и достаточно существования такой постоянной $\mu > 0$, что*

$$\langle f''(u), \xi, \xi\rangle \geqslant \mu|\xi|^2 \tag{13}$$

при всех $u \in X$ и всех $\xi = (\xi^1, \ldots, \xi^n)$, принадлежащих подпространству $L = \operatorname{Lin} X$, параллельному аффинной оболочке множества X (в частности, если $\operatorname{int} X \neq \varnothing$, то (13) *выполняется при всех $\xi \in E^n$).*

Доказательство. Для функции $g(u) = f(u) - \varkappa|u|^2/2$ ее вторая производная равна $g''(u) = f''(u) - \varkappa I_n$, где I_n — единичная матрица. Отсюда ясно, что неравенство (13) с $\mu = \varkappa$ равносильно неравенству

$$\langle g''(u)\xi, \xi\rangle \geqslant 0 \quad \forall u \in X, \quad \forall \xi \in L.$$

Отсюда и из леммы 1, теоремы 2.5 следует справедливость теоремы. □

Замечание 1. Пример 2.1 показывает, что при $\operatorname{int} X = \varnothing$, условие (13) может и не выполняться при каждом $\xi \in E^n$.

Замечание 2. Для функций одной переменной неравенство (13) имеет вид $f''(u) \geqslant \mu > 0$ при всех $u \in X$. Отсюда нетрудно вывести, что функция $f(u) = u^p$ при любом $p > 1$ будет сильно выпукла на множестве $X_\varepsilon = \{u \in E^1\colon u \geqslant \varepsilon\}$ при всех $\varepsilon > 0$; если здесь $\varepsilon \leqslant 0$, то такая функция сильно выпукла лишь при $p = 2$. Функция $f(u) = \sin u$ сильно выпукла на $X_\varepsilon = [-\pi + \varepsilon, -\varepsilon]$ при всех ε, $0 < \varepsilon < \pi/2$, но не сильно выпукла на $[-\pi, 0]$.

Замечание 3. Условие (13) в теореме 4 не может быть заменено условием

$$\langle f''(u)\xi, \xi\rangle > 0 \qquad \forall u \in X,\ \forall \xi \in L,\ \xi \neq 0. \tag{14}$$

Например, для функции $f(u) = e^u$, $u \in X = E^1 = L$ имеем $\langle f''(u)\xi, \xi\rangle = e^u\xi^2 > 0$ при всех $u \in X$, $\xi \in L$, $\xi \neq 0$, но эта функция не является сильно выпуклой. Однако если $\operatorname{int} X \neq \varnothing$, $f''(u) = A$ — постоянная матрица, то условие (14), означающее положительную определенность квадратичной формы $\langle A\xi, \xi\rangle$, влечет за собой условие $\langle A\xi, \xi\rangle \geqslant \mu|\xi|^2$, $\xi \in E^n$, где $\mu = \inf\limits_{|\xi|=1} \langle A\xi, \xi\rangle > 0$. Согласно критерию Сильвестра для положительной определенности квадратичной формы $\langle A\xi, \xi\rangle$ необходимо и достаточно, чтобы все главные угловые миноры матрицы A были положительны (см. замечание 2.2.1). Пользуясь этим условием, для функции $f(u) = x^2 + 2axy + by^2 + cz^2$

из примера 2.2 находим, что $f(u)$ будет сильно выпукла на E^3 тогда и только тогда, когда $b-a^2>0$, $c>0$.

Далее, для функции

$$f(u)=\langle Au,u\rangle/2-\langle b,u\rangle,\quad u\in E^n,$$

из примера 2.3 сильная выпуклость на E^n будет тогда и только тогда, когда A — положительно определенная матрица. Аналогично для функции

$$f(u)=|Au-b|^2$$

из примера 2.4 сильная выпуклость на E^n будет тогда и только тогда, когда матрица A^TA — невырожденная.

3. Рассмотрим теперь сильно выпуклые функции $f(x)$ из класса $C^{1,1}(X)$, т. е. гладкие сильно выпуклые функции, градиент которых удовлетворяет условию

$$|f'(u)-f'(v)|\leqslant L|u-v|,\quad u,v\in X. \tag{15}$$

Полезно установить связь между постоянными $\varkappa$, μ, L из (1), (9), (12), (13), (15). Из условия (12) с помощью неравенства Коши — Буняковского и условия (15) имеем $\mu\leqslant L$. При доказательстве теорем 3, 4 было установлено, что $\mu=\varkappa$. Поэтому $\varkappa\leqslant L$.

Из определения 1 видно, что если неравенство (1) имеет место при некотором $\varkappa$, то оно будет иметь место и для всех меньших положительных значений $\varkappa$. Можно поставить вопрос об определении самой большой, точной постоянной $\varkappa$ в (1). Очевидно, такой постоянной в (1) будет

$$\varkappa_0=\inf_{0<\alpha<1}\inf_{u,v\in X}\frac{\alpha f(u)+(1-\alpha)f(v)-f(\alpha u+(1-\alpha)v)}{\alpha(1-\alpha)|u-v|^2/2}.$$

Аналогично самые большие постоянные в (12), (13) соответственно имеют вид

$$\mu_0=\inf_{u,v\in X}\frac{<f'(u)-f'(v),u-v>}{|u-v|^2},\quad \mu_1=\inf_{u\in X}\inf_{\xi\in \operatorname{Lin}X,\,\xi\neq 0}\frac{<f''(u)\xi,\xi>}{|\xi|^2}.$$

Из доказательства теорем 3, 4 следует, что $\mu_0=\mu_1=\varkappa_0$, причем для функций из $C^{1,1}(X)$ все эти постоянные не больше L. Заметим, что для функции $\Omega(x)=|x|^2$ на E^n имеем $\mu_0=\mu_1=\varkappa_0=L=2$.

4. Продолжим рассмотрение сильно выпуклых функций $f(x)\in C^{1,1}(X)$. Для таких функций из (12) и (15) имеем неравенства

$$\mu|u-v|^2\leqslant\langle f'(u)-f'(v),u-v\rangle\leqslant L|u-v|^2,\quad u,v\in X. \tag{16}$$

Оказывается, как в теореме 2.16, два неравенства (16) можно записать в виде одного равносильного (16) неравенства [24; 234; 525], полностью характеризующего класс сильно выпуклых функций из $C^{1,1}(X)$ при $\operatorname{int}X\neq\varnothing$ с данными постоянными L, μ, $L\geqslant\mu>0$.

Теорема 5. *Пусть X — выпуклое множество из E^n, $\operatorname{int}X\neq\varnothing$ и $f(x)\in C^1(X)$. Тогда для того чтобы функция $f(x)$ была сильно выпуклой с постоянной $\varkappa=\mu>0$ и удовлетворяла условию (15) с постоянной $L>0$, необходимо и достаточно, чтобы*

$$|f'(u)-f'(v)|^2+L\mu|u-v|^2\leqslant(L+\mu)\langle f'(u)-f'(v),u-v\rangle \tag{17}$$

при всех $u,v\in X$.

Доказательство. Необходимость. Пусть функция $f(x)\in C^{1,1}(X)$ и сильно выпукла на X. Тогда справедливы неравенства (16). Введем функцию

$$g(u)=f(u)-\mu|u|^2/2,\quad u\in X.$$

Имеем $g'(u)=f'(u)-\mu u$. Тогда из левого неравенства (16) следует

$$\langle g'(u)-g'(v),u-v\rangle=\langle f'(u)-f'(v),u-v\rangle-\mu|u-v|^2\geqslant 0,\quad \forall u,v\in X,$$

а из правого неравенства (16) получим

$$\langle g'(u)-g'(v),u-v\rangle\leqslant(L-\mu)|u-v|^2\quad\forall u,v\in X.$$

Объединяя оба полученных неравенства, имеем

$$0\leqslant\langle g'(u)-g'(v),u-v\rangle\leqslant(L-\mu)|u-v|^2\quad\forall u,v\in X.$$

Таким образом, функция $g(u)$ удовлетворяет неравенствам вида (2.18). Согласно теореме 2.16 эти два неравенства равносильны одному неравенству

$$|g'(u)-g'(v)|^2\leqslant(L-\mu)\langle g'(u)-g'(v),u-v\rangle\quad\forall u,v\in X.$$

Подставляя сюда $g'(u)=f'(u)-\mu u$, после несложных тождественных преобразований получаем неравенство (17).

Достаточность. Пусть некоторая функция $f(x)\in C^1(X)$ и удовлетворяет неравенству (17). Покажем, что тогда функция $f(x)$ сильно выпукла с постоянной $\varkappa=\mu$ и удовлетворяет условию (15) с L, где μ, L взяты из (17). С помощью неравенства Коши — Буняковского из (17) имеем

$$|f'(u)-f'(v)|^2+L\mu|u-v|^2\leqslant(L+\mu)|f'(u)-f'(v)|\cdot|u-v|.$$

Приняв $x=|f'(u)-f'(v)|$, последнее неравенство можно переписать в виде $x^2-(L+\mu)|u-v|x+L\mu|u-v|^2\leqslant 0$. Квадратный трехчлен в левой части этого неравенства имеет корни $x_1=\mu|u-v|$, $x_2=L|u-v|$. Поэтому $x_1\leqslant x\leqslant x_2$, т. е.

$$\mu|u-v|\leqslant|f'(u)-f'(v)|\leqslant L|u-v|\quad\forall u,v\in X. \tag{18}$$

Тогда $\langle f'(u)-f'(v),u-v\rangle\leqslant L|u-v|^2$ — правое неравенство (16) получено.

Используя левое неравенство (18), из (17) имеем $\mu^2|u-v|^2+L\mu|u-v|^2\leqslant(L+\mu)\langle f'(u)-f'(v),u-v\rangle$. Поделив обе части этого неравенства на $L+\mu>0$ придем к левому неравенству (16). Таким образом, из (17) получили неравенства (18) и (16). Левое неравенство (16) согласно теореме 3 означает сильную выпуклость $f(u)$ с постоянной $\varkappa=\mu$, а правое неравенство (16) (или (18)) дает условие (15). □

Из (17) вытекает неравенство

$$\langle f'(u)-f'(v),v-w\rangle\leqslant\frac{1}{4}(L+\mu)|u-w|^2-\frac{L\mu}{L+\mu}|u-v|^2\quad\forall u,v,w\in X.$$

Оно доказывается так же, как и подобное неравенство (2.20).

Упражнения

1. При каких a, b, c функция $f(u)=ax^2+2bxy+cy^2$ переменных $u=(x,y)\in E^3$ будет сильно выпукла на E^2?

2. Найти области сильной выпуклости функций $f(u)=\sin(x+y+z)$, $f(u)=\sin(x^2+y^2z^2)$.

3. Рассмотреть функцию одной переменной

$$f(x)=\begin{cases}x^2(1+\nu\sin(\ln|x|)), & x\neq 0\\ 0, & x=0.\end{cases}$$

При каких значениях параметра ν и на каких отрезках $a\leqslant x\leqslant b$ эта функция выпукла? Строго выпукла? Сильно выпукла? Нарисуйте график этой функции.

4. Доказать, что функция $f(x)$ сильно выпукла на выпуклом множестве X с постоянной сильной выпуклости $\varkappa>0$ тогда и только тогда, когда функция $g(t)=f(v+t(u-v))$ переменной t, $0\leqslant t\leqslant 1$, при любых $u,v\in X$ сильно выпукла с постоянной сильной выпуклости $\varkappa|u-v|^2$ [364].

5. Пусть функция $f(x)$ сильно выпукла и дифференцируема на выпуклом множестве X. Пользуясь теоремами 1–5, доказать, что:

а) $f'(u)\neq f'(v)$ $\forall u,v\in X$, $u\neq v$;

б) $|u-v| \leqslant \frac{1}{\varkappa}|f'(v)|^2$ при всех $u \in M(v) = \{u \in X: f(u) \leqslant f(v)\}$, $v \in X$;

в) $0 \leqslant f(u) - f_* \leqslant \frac{1}{4\varkappa}|f'(v)|^2$, $|u-u_*| \leqslant \frac{1}{2\varkappa}|f'(v)|$ $\forall u \in X$.

6. Для того чтобы симметричная матрица A порядка $n \times n$ была положительно определенной, необходимо и достаточно существования постоянных L, μ, $0 < \mu \leqslant L$, таких, что

$$|Ae|^2 + L\mu|e|^2 \leqslant (L+\mu)\langle Ae, e\rangle \quad \forall e \in E^n.$$

Доказать. Убедиться, что в приведенном неравенстве в качестве μ можно взять минимальное собственное число матрицы A, в качестве L — максимальное собственное число. Указание: к функции $f(x) = \langle Ax, x\rangle/2$ применить (17).

7. Пусть X — выпуклое множество, $f(x) \in C^1(X)$. Показать, что для того, чтобы функция $f(x)$ была сильно выпуклой и удовлетворяла условию (15), необходимо и достаточно выполнения неравенств (18) при каких-нибудь постоянных L, μ, $0 < \mu \leqslant L$.

8. Можно ли утверждать, что сильно выпуклая функция обладает более лучшими дифференциальными свойствами по сравнению с выпуклыми функциями? Рассмотреть функцию $f(x) = \langle x, x\rangle + g(x)$, где $g(x)$ — выпуклая функция.

9. Пусть X — выпуклое множество, $f(x, \alpha)$ при каждом значении параметра $\alpha \in A$ сильно выпукла на множестве X с постоянной сильной выпуклости $\varkappa(\alpha)$, $\inf\limits_{\alpha \in A} \varkappa(\alpha) \geqslant \varkappa_0 > 0$. Доказать, что функция $f(x) = \sup\limits_{\alpha \in A} f(x, \alpha)$ сильно выпукла на X с постоянной $\varkappa_0$. Указание: воспользоваться схемой доказательства теоремы 4.2.7.

10. Функция $f(x)$, определенная на выпуклом множестве X, называется *сильно квазивыпуклой* на X, если существует постоянная $\varkappa > 0$, такая, что

$$f(\alpha u + (1-\alpha)v) \leqslant \max\{f(u); f(v)\} - \frac{1}{2}\alpha(1-\alpha)\varkappa|u-v|^2 \quad \forall u, v \in X, \quad \forall \alpha \in [0,1]$$

(ср. с упражнением 2.33). Доказать, что:

а) функция $g(t) = |t+a|$ сильно квазивыпукла на любом отрезке $[0, c]$ с $\varkappa = \frac{1}{c}$, но не является сильно выпуклой (a — параметр, $a \in \mathbb{R}$);

б) функция $g(t) = \sqrt{(t+a)^2 + b^2} - a^2$ сильно выпукла на $[0, c]$ с постоянной $\varkappa = (b^2 - a^2)(\max\{g(0), g(c)\})^{-3}$ и сильно квазивыпукла на $[0, c]$ с постоянной $\varkappa = (\max\{g(0), g(c)\})^{-1}$ (a, b — параметры, a, $b \in \mathbb{R}$);

в) функция $f(x) = |x|$, $x \in E^n$, сильно квазивыпукла на любом ограниченном выпуклом множестве X с постоянной $\varkappa = \frac{1}{R}$, где R — радиус шара, содержащего X, и не является сильно квазивыпуклой на неограниченном множестве X. Функция $f(x) = |x|$ не является сильно выпуклой ни на каком выпуклом множестве X с $\operatorname{int} X \neq \varnothing$ [364].

§ 4. Проекция точки на множество

1. При описании и исследовании некоторых методов минимизации ниже нам понадобится понятие проекции точки на множество.

Определение 1. Пусть X — некоторое множество из E^n. *Проекцией точки* u из E^n называется ближайшая к u точка w множества X, т. е. точка $w \in X$, удовлетворяющая условию

$$|u-w| = \inf_{v \in X}|u-v|.$$

Проекцию точки u на множество X будем обозначать через $\mathcal{P}_X(u) = w$.

Поскольку $\rho(u, X) = \inf\limits_{v \in X}|u-v|$ — расстояние от точки u до множества X, то из определения 1 следует, что

$$\rho(u, X) = |u - \mathcal{P}_X(u)| \leqslant |u-v| \quad \forall v \in X, \quad \forall u \in E^n.$$

Если $u \in X$, то, очевидно, всегда $\mathcal{P}_X(u) = u$. Однако проекция на множество существует не всегда. Например, если $X = \{u \in E^n: |u| < 1\}$ — открытый единичный шар в E^n, то ни одна точка $u \notin X$ не будет иметь проекции на это множество. Однако если множество X замкнуто, то любая точка $u \in E^n$ имеет проекцию на X — это было доказано в следствии 1 к теореме 2.1.3.

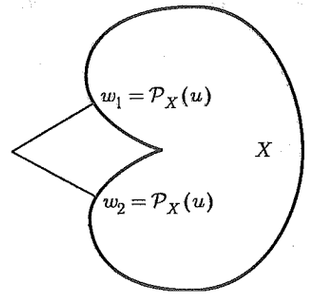

Рис. 4.6

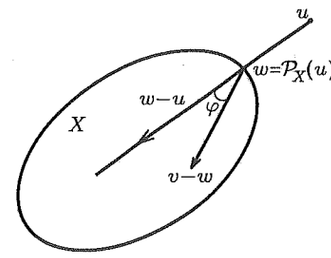

Рис. 4.7

Проекция точки на множество может определяться неоднозначно (рис. 4.6). Однако, как показывает следующая теорема, для выпуклых множеств такая ситуация невозможна (рис. 4.7).

Теорема 1. *Пусть X — выпуклое замкнутое множество из E^n. Тогда:*

1) всякая точка $u \in E^n$ имеет, и притом единственную, проекцию на это множество;

2) для того чтобы точка $w \in X$ была проекцией точки u на множество X, необходимо и достаточно выполнения неравенства (см. рис. 4.7)

$$\langle w-u, v-w\rangle \geqslant 0 \quad \forall v \in X. \tag{1}$$

При этом если X — аффинное множество (см. пример 1.4), то вместо (1) можно писать

$$\langle w-u, v-w\rangle = 0 \quad \forall v \in X. \tag{2}$$

Доказательство. Рассмотрим функцию $g(v) = |v-u|^2$ переменной $v \in E^n$ при произвольной фиксированной $u \in E^n$. Поскольку $g(v)$ сильно выпукла на E^n, то по теореме 3.1 эта функция достигает своей нижней грани на X в единственной точке $w \in X$. Это означает, что $|v-u|^2 \geqslant |w-u|^2$ или $|v-u| \geqslant |w-u|$ при всех $v \in X$, причем равенство здесь возможно только при $v = w$. Остается принять $\mathcal{P}_X(u) = w$.

Докажем второе утверждение теоремы. Согласно теореме 2.3 для того, чтобы функция $g(v)$ достигала минимума на X в точке w, необходимо и достаточно, чтобы $\langle g'(w), v-w\rangle = 2\langle w-u, v-w\rangle \geqslant 0$ при всех $v \in X$, что равносильно неравенству (1).

Наконец, пусть $X = \{u \in E^n: Au = b\}$ — аффинное множество. Поскольку это множество выпукло и замкнуто, то неравенство (1) сохраняет силу и здесь. Аффинное множество обладает следующим замечательным свойством: если $v, v_0 \in X$, $v \neq v_0$, то и $2v_0 - v \in X$, что проверяется непосредственно. Поэтому если здесь взять $v_0 = \mathcal{P}_X(u) = w \in X$, то $2w - v \in X$ при любом

выборе $v \in X$. Подставим в (1) вместо v точку $2w - v$. Получим $\langle w-u, 2w-v-w\rangle = \langle w-u, w-v\rangle \geqslant 0$ при всех $v \in X$. Сравнивая полученное неравенство с (1), приходим к равенству (2). □

Покажем, что оператор проектирования на выпуклое множество обладает сжимающим свойством.

Теорема 2. *Если X — выпуклое замкнутое множество из E^n, то*

$$|\mathcal{P}_X(u) - \mathcal{P}_X(v)| \leqslant |u - v| \quad \forall u, v \in E^n. \tag{3}$$

Доказательство. Из неравенства (1) имеем

$$\langle \mathcal{P}_X(u) - u, \mathcal{P}_X(v) - \mathcal{P}_X(u)\rangle \geqslant 0.$$

Поменяв ролями точки u и v в последнем неравенстве, получим

$$\langle \mathcal{P}_X(v) - v, \mathcal{P}_X(u) - \mathcal{P}_X(v)\rangle \geqslant 0.$$

Сложим эти два неравенства. Имеем

$$\langle \mathcal{P}_X(u) - u - \mathcal{P}_X(v) + v, \mathcal{P}_X(v) - \mathcal{P}_X(u)\rangle \geqslant 0.$$

Отсюда следует

$$|\mathcal{P}_X(u) - \mathcal{P}_X(v)|^2 \leqslant \langle \mathcal{P}_X(u) - \mathcal{P}_X(v), u - v\rangle \quad \forall u, v \in E^n. \tag{4}$$

Применим к правой части (4) неравенство Коши — Буняковского:

$$|\mathcal{P}_X(u) - \mathcal{P}_X(v)|^2 \leqslant |\mathcal{P}_X(u) - \mathcal{P}_X(v)| \cdot |u - v| \quad u, v \in E^n.$$

Разделив на $|\mathcal{P}_X(u) - \mathcal{P}_X(v)| \neq 0$, получим требуемое неравенство (3). Если $|\mathcal{P}_X(u) - \mathcal{P}_X(v)| = 0$, то (3) очевидно. □

Теорема 3. *Пусть X — выпуклое замкнутое множество из E^n, пусть $\mathcal{P}_X(M) = \{u \in X: \exists x \in M,$ что $u = \mathcal{P}_X(x)\}$ — множество значений оператора проектирования на множестве $M \subset E^n$. Если M — компактное множество, то $\mathcal{P}_X(M)$ также компактно.*

Доказательство. Возьмем $\forall x \in M$. Так как $\inf\limits_{v \in X} |v - x|^2 = |\mathcal{P}_X(x) - x|^2$, то, применяя к сильно выпуклой функции $g(v) = |v - x|^2$ неравенство (3.3), получим: $|v - \mathcal{P}_X(x)|^2 \leqslant \varphi(v) - \varphi(\mathcal{P}_X(x)) \leqslant \varphi(v) = |v - x|^2$ или $|v - \mathcal{P}_X(x)| \leqslant |v - x| \quad \forall x \in M, \ \forall v \in X$. Отсюда, фиксируя $v \in X$ имеем: $|\mathcal{P}_X(x)| \leqslant |v| + |v - x| \leqslant 2|v| + |x| \leqslant 2|v| + \sup\limits_{z \in M} |z| \quad \forall x \in M$. Ограниченность множества $\mathcal{P}_X(v)$ доказана. Докажем замкнутость $\mathcal{P}_X(M)$. Возьмем произвольную последовательность $\{x_k\} \in \mathcal{P}_X(M)$, $\{x_k\} \to x$. По определению множества $\mathcal{P}_X(M)$ найдется точка $u_k \in M$, такая, что $x_k = \mathcal{P}_X(u_k)$, $k = 1, 2, \ldots$ Из ограниченности M следует ограниченность $\{u_k\}$. Применяя теорему Больцано — Вейерштрасса, можем считать, что $\{u_k\} \to u_0$. Так как M замкнуто, то $u_0 \in M$. По теореме 2 оператор проектирования непрерывен, поэтому $\{x_k = \mathcal{P}_X(u_k)\} \to x = \mathcal{P}_X(u_0)$. Это значит, что $x \in \mathcal{P}_X(M)$, т. е. множество $\mathcal{P}_X(M)$ замкнуто. Следовательно, $\mathcal{P}_X(M)$ компактное множество. Теорема 3 доказана. □

2. Приведем примеры множеств, проекция на которые может быть выписана явно.

Пример 1. Пусть $X = S(u_0, R) = \{u \in E^n: |u - u_0| \leqslant R\}$ — шар радиуса $R > 0$ с центром в точке u_0. Из геометрических соображений (рис. 4.8) ясно, что проекцией точки $u \notin X$ является точка

$$w = u_0 + R(u - u_0)/|u - u_0|.$$

Для строгого доказательства этого факта достаточно проверить выполнение неравенства (1). Имеем

$$\langle w - u, v - w\rangle = (R/|u - u_0| - 1)(\langle u - u_0, v - u_0\rangle - R|u - u_0|) \geqslant 0,$$

так как $|u - u_0| > R$, а $\langle u - u_0, v - u_0\rangle \leqslant |u - u_0| \cdot |v - u_0| \leqslant |u - u_0|R$ в силу неравенства Коши — Буняковского для всех $v \in X$.

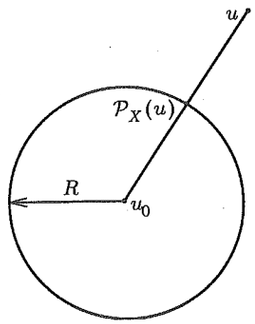

Рис. 4.8

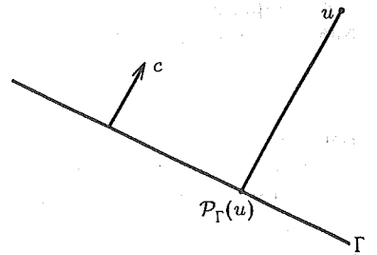

Рис. 4.9

Пример 2. Пусть $X = \Gamma = \{u \in E^n: \langle c, u\rangle = \gamma\}$ — гиперплоскость; здесь $c \in E^n$, $c \neq 0$, $\gamma = \text{const}$. Пользуясь геометрическими соображениями (рис. 4.9), проекцию точки $u \notin X$ на X будем искать в виде $w = u + \alpha c$. Определяя число α из условия $w \in X$, имеем

$$w = u + (\gamma - \langle c, u\rangle)c/|c|^2.$$

Поскольку $\langle w - u, v - w\rangle = (\gamma - \langle c, u\rangle)c/|c|^2 \cdot \langle c, v - w\rangle = 0$ при всех $v \in X$, то согласно теореме 1 найденная точка w представляет собой проекцию точки u на X.

Пример 3. Пусть $X = \{u \in E^n: \langle a_i, u\rangle = b^i, \ i = 1, \ldots, m\}$ — аффинное множество; здесь $a_i \in E^n$, $b^i = \text{const}$, $i = 1, \ldots, m$. Можем считать, что векторы $a_1, \ldots, a_m$ линейно независимы и $m < n$ (если $m = n$, то X будет состоять из одной точки). Проекцию точки u на множество X будем искать в виде

$$w = u + \sum_{j=1}^{m} \alpha_j a_j. \tag{5}$$

Из требования $w \in X$ имеем систему линейных алгебраических уравнений

$$\sum_{j=1}^{m} \alpha_j \langle a_i, a_j\rangle = b^i - \langle a_i, u\rangle, \quad i = 1, \ldots, m, \tag{6}$$

для определения коэффициентов $\alpha_1, \ldots, \alpha_m$. Определителем этой системы является определитель Грама [89; 192; 353], который для линейно независи-

мых $a_1, \ldots, a_m$ будет отличным от нуля. Поэтому искомые $\alpha_1, \ldots, \alpha_m$ существуют и однозначно определяются из системы (6). Для точки w из (5) будем иметь

$$\langle w-u, v-w\rangle = \sum_{j=1}^{m} \alpha_j \langle a_j, v-u\rangle - \sum_{i=1}^{m} \alpha_i \Big(\sum_{j=1}^{m} \alpha_j \langle a_i, a_j\rangle \Big) = 0$$

для всех $v \in X$. Следовательно, по теореме 1 найденная из (5), (6) точка w будет проекцией точки u на множество X. Если ввести матрицу A, строками которой являются векторы a_i, $i = 1, \ldots, m$, то точку (5) можно записать в виде

$$w = u - A^{\top}(AA^{\top})^{-1}(Au - b).$$

Предлагаем читателю провести проверку того, что такая точка w принадлежит X, т. е. $Aw = b$, и выполняется условие (1) (см. пример 9.3).

Пример 4. Пусть $X = \{u \in E^n: \langle c, u\rangle \leqslant \gamma\}$ — замкнутое полупространство, определяемое гиперплоскостью $\langle c, u\rangle = \gamma$. Пусть $u \notin X$, т. е. $\langle c, u\rangle > \gamma$. Как и в примере 2, попробуем представить проекцию точки u на X в виде

$$w = u + (\gamma - \langle c, u\rangle)c/|c|^2.$$

Имеем $\langle w-u, v-w\rangle = (\gamma - \langle c, u\rangle)c/|c|^{-2}(\langle c, v\rangle - \gamma) \geqslant 0$ при всех $v \in X$. Следовательно, точка w — искомая проекция.

Пример 5. Пусть $X = \{u = (u^1, \ldots, u^n) \in E^n: \alpha_i \leqslant u^i \leqslant \beta_i,\ i = 1, \ldots, n\}$ — n-мерный параллелепипед, где $\alpha_i, \beta_i,\ \alpha_i < \beta_i$ — заданные числа, $i = 1, \ldots, n$. Пусть $u \notin X$. Положим $w = (w^1, \ldots, w^n)$, где

$$w^i = \begin{cases} \alpha_i, & u^i < \alpha_i, \\ \beta_i, & u^i > \beta_i, \\ u^i, & \alpha_i \leqslant u^i \leqslant \beta_i, \quad i = 1, \ldots, n. \end{cases}$$

Тогда $(w^i - u^i)(v^i - w^i) \geqslant 0$ для всех $v^i,\ \alpha_i \leqslant v^i \leqslant \beta_i,\ i = 1, \ldots, n$. Отсюда, суммируя по i от 1 до n, получаем $\langle w-u, v-w\rangle \geqslant 0$ для всех $v \in X$. Следовательно, построенная точка w является проекцией точки u на множество X.

Пример 6. Пусть $X = E^n_+ = \{u = (u^1, \ldots, u^n) \in E^n: u^i \geqslant 0,\ i = 1, \ldots, n\}$ — неотрицательный ортант пространства E^n. Легко проверить, что проекцией точки u на X является точка $u^+ = ((u^1)^+, \ldots, (u^n)^+)$, где $(u^i)^+ = \max\{0; u^i\},\ i = 1, \ldots, n$.

3. Критерий оптимальности, сформулированный ранее в теореме 2.3, с помощью оператора проектирования может быть переформулирован следующим образом.

Теорема 4. *Пусть X — выпуклое замкнутое множество, X_* — множество точек минимума функции $f(x)$ на X. Если $u_* \in X_*$ и $f(x)$ дифференцируема в точке u_*, то необходимо выполняется равенство*

$$u_* = \mathcal{P}_X(u_* - \alpha f'(u_*)) \quad \forall \alpha > 0. \tag{7}$$

Если, кроме того, $f(x)$ выпукла на X, то всякая точка u_*, удовлетворяющая уравнению (7), принадлежит X_*.

Доказательство. Согласно теореме 1 равенство (7) эквивалентно неравенству $\langle u_* - (u_* - \alpha f'(u_*)), v - u_*\rangle \geqslant 0\ \forall v \in X$, откуда имеем $\alpha\langle f'(u_*), v - u_*\rangle \geqslant 0\ \forall v \in X$. Так как $\alpha > 0$, отсюда получим неравенство $\langle f'(u_*), v - u_*\rangle \geqslant 0\ \forall v \in X$. Таким образом, условия (7) и (2.5) эквивалентны. Отсюда и из теоремы 2.3 следует утверждение теоремы 3. □

Таким образом, если ввести отображение A из E^n в E^n по формуле

$$Au = \mathcal{P}_X(u - \alpha f'(u)), \quad \alpha > 0,$$

то условие (7) перепишется в виде $u_* = Au_*$, т. е. u_* — неподвижная точка отображения A. Ниже мы увидим, что при некоторых условиях на функцию $f(x)$ отображение будет сжимающим и для определения точки u_* могут быть использованы свойства сжимающих отображений [89; 393].

Упражнения

1. Найти проекцию точки $u \in E^n$ на множество

$$X = \{u \in E^n: \langle a_1, u\rangle \leqslant b^1,\ \langle a_2, u\rangle \leqslant b^2\}.$$

2. Найти проекцию точки $u \in E^n$ на множество

$$X = \{u = (u^1, \ldots, u^n): \alpha_i \leqslant u^i \leqslant \beta_i,\ i = 1, \ldots, n\}$$

(здесь $\alpha_i \leqslant \beta_i$, причем возможно, что $\alpha_i = \beta_i$, или $\alpha_j = -\infty$, или $\beta_k = \infty$ при некоторых i, j, k).

3. Выяснить геометрический смысл равенства (2).

4. Будут ли верными неравенство (1) или равенство (2), если X — невыпуклое множество?

5. Охарактеризовать все множества X из E^n, для которых существует точка $u \notin X$ такая, что $\mathcal{P}_X(u) = v$ для всех $v \in X$.

6. Для того чтобы точка $w \in X$ была проекцией точки u на выпуклое множество X, необходимо и достаточно, чтобы $\langle v-u, v-w\rangle \geqslant 0$ при всех $v \in X$. Доказать. Выяснить геометрический смысл этого условия.

7. Доказать, что для любого замкнутого множества X имеет место неравенство $\|u - \mathcal{P}_X(u)| - |v - \mathcal{P}_X(v)\| \leqslant |u - v|$ для всех $u, v \in X$ (Ср. с леммой 2.1.2).

8. Пусть X выпуклое замкнутое множество из E^n. Доказать, что тогда

$$|v - \mathcal{P}_X(u)|^2 \leqslant \langle v - u, v - \mathcal{P}_X(u)\rangle \quad \forall v \in X,\ \forall u \in E^n,$$
$$|v - \mathcal{P}_X(u)|^2 + |u - \mathcal{P}_X(u)|^2 \leqslant |v - u|^2 \quad \forall v \in X,\ \forall u \in E^n.$$

9. Пусть $f(u) = |Au - b|^2$, где A — матрица порядка $m \times n$, $b \in E^m$ (см. пример 2.4). Доказать, что

$$X_* = \Big\{u \in E^n: f(u) = \inf_{E^n} f(u) = f_*\Big\} \neq \varnothing.$$

Указание: взять проекцию точки b на множество $X = \{v \in E^m: v = Au, u \in E^n\}$ и показать, что $f(u) = |Au - \mathcal{P}_X(b)|^2 + |b - \mathcal{P}_X(b)|^2$, $f_* = |b - \mathcal{P}_X(b)|^2$; доказать замкнутость X.

10. Убедиться, что если $X = L$ — подпространство из E^n, то условие (2) можно заменить равенством

$$\langle w - u, g\rangle = 0 \quad \forall g \in L. \tag{8}$$

11. Пользуясь (8), доказать, что оператор $\mathcal{P}$ проектирования на подпространство $L \subset E^n$ является линейным, самосопряженным оператором и, кроме того, $\|\mathcal{P}\| = 1$, $\mathcal{P}^2 = \mathcal{P}$.

§ 5. Отделимость выпуклых множеств

1. В теории экстремальных задач важную роль играют теоремы, называемые *теоремами отделимости*. Основное содержание этих теорем сводится к тому, что для некоторых двух множеств A а B утверждается существование гиперплоскости такой, что множество A находится в одном из открытых или замкнутых полупространств, определяемых этой гиперплоскостью, а множество B — в другом открытом или замкнутом полупространстве (см. пример 1.3), т. е. гиперплоскости, которая отделяет эти два множества.

Определение 1. Пусть A и B — два множества из E^n. Говорят, что *гиперплоскость* $\langle c, x\rangle = \gamma$ *с нормальным вектором* $c \neq 0$ *отделяет (разделяет) множества* A и B, если $\langle c, a\rangle \geqslant \gamma$ при всех $a \in A$ и $\langle c, b\rangle \leqslant \gamma$ при всех $b \in B$, или, иначе говоря, выполняются неравенства

$$\sup_{b\in B}\langle c, b\rangle \leqslant \gamma \leqslant \inf_{a\in A}\langle c, a\rangle. \tag{1}$$

Если $\sup\limits_{b\in B}\langle c, b\rangle < \inf\limits_{a\in A}\langle c, a\rangle$, то говорят, что множества A и B *сильно отделены*. Если $\langle c, b\rangle < \langle c, a\rangle$ при всех $a \in A$, $b \in B$, то говорят о *строгом отделении* этих множеств. Если выполнено (1), причем существуют такие точки $a_0 \in A$, $b_0 \in B$, что $\langle c, b_0\rangle < \langle c, a_0\rangle$, то говорят, что множества A, B *собственно отделимы*.

Понятие собственной отделимости введено для того, чтобы выделить из (1) вырожденный случай, когда оба множества A, B лежат в разделяющей гиперплоскости и, возможно, даже имеют общие относительно внутренние точки.

Заметим, что в определение 1 множества A и B входят несколько несимметрично. Однако симметрию здесь нетрудно восстановить: нужно лишь взять вектор $-c$ вместо c, постоянную $-\gamma$ вместо γ и записать уравнение отделяющей гиперплоскости в виде $\langle -c, x\rangle = -\gamma$. Очевидно, если гиперплоскость $\langle c, x\rangle = \gamma$ отделяет множества A и B, то гиперплоскость $\langle \mu c, x\rangle = \mu\gamma$ при $\mu \neq 0$ также отделяет эти множества. Поэтому при необходимости можно считать, что $|c| = 1$.

На рис. 4.10–4.14 изображены случаи, когда два множества собственно отделимы, на рис. 4.13 — сильно отделимы, на рис. 4.14 — строго отделимы. Однако ясно, что не всякие два множества можно отделить гиперплоскостью (рис. 4.15). Ниже приводятся теоремы об отделимости выпуклых множеств.

Теорема 1. *Пусть* X *— непустое выпуклое множество из* E^n. *Тогда для любой точки* $y \notin \operatorname{ri} X$ *существует гиперплоскость* $\langle c, x\rangle = \gamma$, *собственно отделяющая множество* X *и точку* y, *или, точнее,*

$$\begin{aligned} &\langle c, x\rangle \geqslant \gamma \quad \forall x \in X, \quad \gamma \geqslant \langle c, y\rangle, \\ &\langle c, x\rangle > \gamma \quad \forall x \in \operatorname{ri} X. \end{aligned} \tag{2}$$

Если точка y *не принадлежит* $\overline{X}$ *— замыканию* X, *то множество* X *(а также и* $\overline{X}$*) сильно отделимо от* y.

Доказательство. Сначала рассмотрим случай $y \notin \overline{X}$. Напомним, что если X — выпуклое множество, то $\overline{X}$ также выпукло (см. теорему 1.2).

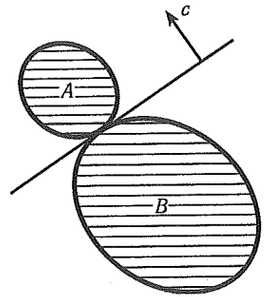

Рис. 4.10

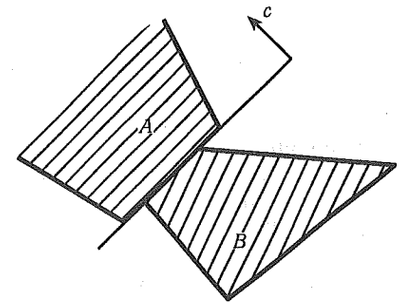

Рис. 4.11

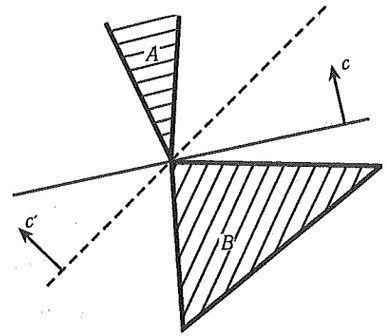

Рис. 4.12

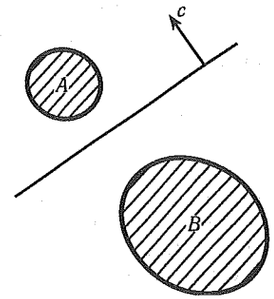

Рис. 4.13

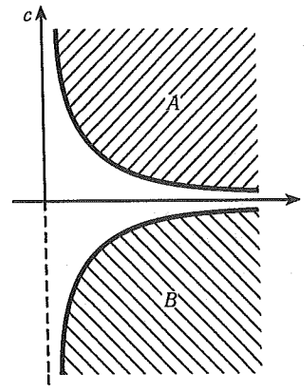

Рис. 4.14

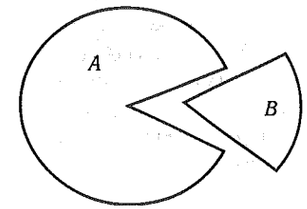

Рис. 4.15

Согласно теореме 4.1 тогда существует проекция $z=\mathcal{P}_{\overline{X}}(y)$ точки y на множество $\overline{X}$, причем $\langle z-y, x-z\rangle \geqslant 0$ для всех $x \in \overline{X}$. Положим $c=z-y$. С учетом предыдущего неравенства будем иметь $\langle c, x-y\rangle = \langle z-y, x-z\rangle + \langle z-y, z-y\rangle \geqslant |c|^2 > 0$ или $\langle c, x\rangle \geqslant \langle c, y\rangle + |c|^2 > \langle c, y\rangle$, $x \in \overline{X}$. Это значит, что гиперплоскость $\langle c, x\rangle = \langle c, y\rangle = \gamma$ сильно отделяет точку y от множества $\overline{X}$ и, тем более, множества X. Нетрудно видеть, что такая гиперплоскость не единственна. Например, гиперплоскость $\langle c, x\rangle = \langle c, z\rangle = \gamma_0$ также сильно отделяет $\overline{X}$ и y, так как $\langle c, x-z\rangle \geqslant 0$ при всех $x \in \overline{X}$, $\langle c, y-z\rangle = \langle z-y, y-z\rangle = -|c|^2 < 0$.

Теперь пусть $y \in \overline{X}$, $y \notin \operatorname{ri} X$. Согласно теореме 1.12 тогда существует последовательность $\{y_k\} \to y$: $y_k \notin \overline{X}$, $y_k \in \operatorname{aff} \overline{X}$, $k=1,2\ldots$ По доказанному гиперплоскость $\langle c_k, x\rangle = \langle c_k, y_k\rangle$, где $c_k = (z_k - y_k)/|z_k - y_k|$, $z_k = \mathcal{P}_{\overline{X}}(y_k)$, сильно отделяет $\overline{X}$ и y_k, так что $\langle c_k, x\rangle > \langle c_k, y_k\rangle$ при всех $x \in \overline{X}$. По теореме Больцано — Вейерштрасса из последовательности $\{c_k\}$, $|c_k|=1$, можно выбрать подпоследовательность, сходящуюся к некоторому вектору c, $|c|=1$. Без умаления общности можем считать, что вся последовательность $\{c_k\} \to c$. Переходя к пределу при $k \to \infty$ в неравенстве $\langle c_k, x\rangle > \langle c_k, y_k\rangle$, $x \in \overline{X}$, получаем $\langle c, x\rangle \geqslant \langle c, y\rangle = \gamma$ для всех $x \in \overline{X}$. Отсюда следует первое из неравенств (2).

Далее, возьмем любую точку $x \in \operatorname{ri} X$. Согласно определению 1.10 существует такое $\varepsilon > 0$, что $O(x, 2\varepsilon) \cap \operatorname{aff} X \in X$. Тогда $u_k = x - \varepsilon c_k \in O(x, 2\varepsilon) \cap \operatorname{aff} X \in X$, $k=1,2\ldots$ Ясно, что $\{u_k\} \to u = x - \varepsilon c$. Покажем, что $u = x - \varepsilon c \in \operatorname{ri} X$. Поскольку $z_k = \mathcal{P}_{\overline{X}}(y_k) \in \overline{X} \subset \operatorname{aff} \overline{X} = \operatorname{aff} X$, $y_k \in \operatorname{aff} X$, то $z_k - y_k \in \operatorname{Lin} X$ — подпространство, параллельное $\operatorname{aff} X$. Тогда $c_k = (z_k - y_k)/|z_k - y_k| \in \operatorname{Lin} X$. Поскольку $\operatorname{Lin} X$ замкнуто в E^n, $\{c_k\} \to c$, то $c \in \operatorname{Lin} X$. Отсюда следует, что $u = x - \varepsilon c \in \operatorname{aff} X$. Кроме того, так как $|u_k - x| = \varepsilon$, $k=1,2,\ldots$, то при $k \to \infty$ имеем $|u-x| = \varepsilon$, так что $u = x - \varepsilon c \in O(x, 2\varepsilon)$. Следовательно, $u = x - \varepsilon c \in \operatorname{ri} X \subset X \subset \overline{X}$. Тогда $\gamma = \langle c, y\rangle \leqslant \langle c, u\rangle = \langle c, x - \varepsilon c\rangle = \langle c, x\rangle - \varepsilon < \langle c, x\rangle$, или $\gamma < \langle c, x\rangle$ для каждой точки $x \in \operatorname{ri} X$, т. е. $\operatorname{ri} X$ и y строго отделимы. □

Теорема 2. *Пусть A и B — непустые выпуклые множества из E^n, $\operatorname{ri} A \cap \operatorname{ri} B = \varnothing$ (например, $A \cap B = \varnothing$). Тогда существует гиперплоскость $\langle c, u\rangle = \gamma$, строго отделяющая множества $\operatorname{ri} A$ и $\operatorname{ri} B$, собственно отделяющая множества A и B, а также их замыкания $\overline{A}$, $\overline{B}$, причем если $\overline{A}$ и $\overline{B}$ имеют общую граничную точку y, то $\gamma = \langle c, y\rangle$. Верно и обратное: если два выпуклых множества A и B собственно отделимы, то* $\operatorname{ri} A \cap \operatorname{ri} B = \varnothing$.

Доказательство. Введем множество $X = \operatorname{ri} A - \operatorname{ri} B = \{x \in E^n: x = a - b,\ a \in \operatorname{ri} A,\ b \in \operatorname{ri} B\}$. Согласно теоремам 1.1, 1.11 множество X выпукло. Поскольку $\operatorname{ri} A \cap \operatorname{ri} B = \varnothing$, то $0 \notin X$. Возможно два случая: $0 \notin \overline{X}$ или $0 \in \overline{X}$. Если $0 \notin \overline{X}$, то согласно теореме 1 точка 0 и множество X сильно отделимы, т. е. существуют такие $c \neq 0$, $\varepsilon > 0$, что $\langle c, x\rangle \geqslant \langle c, 0\rangle + \varepsilon = \varepsilon$ при всех $x \in X$. Если $0 \in \overline{X}$, то по той же теореме 1 найдется такой вектор $c \neq 0$, что $\langle c, x\rangle \geqslant 0$ для всех $x \in X$, причем $\langle c, x\rangle > 0$ при $x \in \operatorname{ri} X$. Таким образом, в обоих случаях существует такой вектор $c \neq 0$, что

$$\langle c, x\rangle \geqslant 0 \quad \forall x \in X, \quad \langle c, x\rangle > 0 \quad \forall x \in \operatorname{ri} X. \tag{3}$$

Из первого неравенства (3) с учетом определения множества X имеем

$$\langle c, a\rangle \geqslant \langle c, b\rangle \quad \forall a \in \operatorname{ri} A,\ b \in \operatorname{ri} B. \tag{4}$$

Далее, по теореме 1.11 $\operatorname{ri} X \neq \varnothing$. Это значит, что существуют $a_0 \in \operatorname{ri} A$, $b_0 \in \operatorname{ri} B$ такие, что $x_0 = a_0 - b_0 \in \operatorname{ri} X$. Из второго неравенства (3) тогда имеем $\langle c, x_0\rangle = \langle c, a_0 - b_0\rangle > 0$, или

$$\langle c, a_0\rangle > \langle c, b_0\rangle, \quad a_0 \in \operatorname{ri} A, \quad b_0 \in \operatorname{ri} B. \tag{5}$$

Неравенство (4) остается справедливым для всех предельных точек множеств $\operatorname{ri} A$, $\operatorname{ri} B$, т. е. для всех $a \in \overline{\operatorname{ri} A}$, $b \in \overline{\operatorname{ri} B}$. Но по теореме 1.13 $\overline{\operatorname{ri} A} = \overline{A}$, $\overline{\operatorname{ri} B} = \overline{B}$, так что $\langle c, a\rangle \geqslant \langle c, b\rangle$ при любых $a \in \overline{A}$, $b \in \overline{B}$. Отсюда $\inf\limits_{a \in \overline{A}} \langle c, a\rangle \geqslant \sup\limits_{b \in \overline{B}} \langle c, b\rangle$. Возьмем гиперплоскость $\langle c, x\rangle = \gamma$, где γ — произвольное число, удовлетворяющее неравенству $\inf\limits_{a \in \overline{A}} \langle c, a\rangle \geqslant \sup\limits_{b \in \overline{B}} \langle c, b\rangle$. Тогда

$$\langle c, a\rangle \geqslant \gamma \geqslant \langle c, b\rangle \quad \forall a \in \overline{A}, \quad b \in \overline{B}. \tag{6}$$

Из (5), (6) следует собственно отделимость множеств $\overline{A}$ и $\overline{B}$ и, тем более, множеств A и B. Если $y \in \overline{A} \cap \overline{B}$, то $\gamma = \langle c, y\rangle$.

Покажем, что построенная гиперплоскость $\langle c, x\rangle = \gamma$ строго отделяет множества $\operatorname{ri} A$ и $\operatorname{ri} B$. Из (5), (6) следует, что либо $\langle c, a_0\rangle > \gamma$, либо $\langle c, b_0\rangle < \gamma$. Пусть $\langle c, a_0\rangle > \gamma$ (случай $\langle c, b_0\rangle < \gamma$ рассматривается аналогично). Возьмем произвольную точку $a \in \operatorname{ri} A$. Тогда $a - a_0 \in \operatorname{Lin} A$, $a + \varepsilon(a - a_0) \in \operatorname{aff} A$ при всех $\varepsilon \in \mathbb{R}$ и по определению относительно внутренней точки найдется такое $\varepsilon > 0$, что $d = a + \varepsilon(a - a_0) \in \operatorname{ri} A$. Отсюда $a = \alpha a_0 + (1-\alpha)d$, $\alpha = \varepsilon/(1+\varepsilon) \in (0,1)$. Умножим неравенство $\langle c, a_0\rangle > \gamma$ на α, $\langle c, d\rangle \geqslant \gamma$ на $1-\alpha$ и сложим. Получим $\alpha\langle c, a_0\rangle + (1-\alpha)\langle c, d\rangle = \langle c, a\rangle > \gamma$. Таким образом, $\langle c, a\rangle > \gamma$ при всех $a \in \operatorname{ri} A$. Отсюда и из (6) следует, что множества $\operatorname{ri} A$ и $\operatorname{ri} B$ строго отделимы.

Докажем вторую часть теоремы. Пусть множества A и B собственно отделимы, но пусть тем не менее $\operatorname{ri} A \cap \operatorname{ri} B \neq \varnothing$. Возьмем какую-либо точку $u \in \operatorname{ri} A \cap \operatorname{ri} B$. Тогда при достаточно малом $\varepsilon > 0$ имеем $a_\varepsilon = u - \varepsilon(a_0 - u) \in \operatorname{ri} A$, $b_\varepsilon = u - \varepsilon(b_0 - u) \in \operatorname{ri} B$, где a_0, b_0 взяты из (5). В силу (4) $\langle c, a_\varepsilon\rangle = \langle c, u\rangle(1+\varepsilon) - \varepsilon\langle c, a_0\rangle \geqslant \langle c, b_\varepsilon\rangle = \langle c, u\rangle(1+\varepsilon) - \varepsilon\langle c, b_0\rangle$. Отсюда получаем $\langle c, a_0\rangle \leqslant \langle c, b_0\rangle$, что противоречит (5). Следовательно, $\operatorname{ri} A \cap \operatorname{ri} B = \varnothing$. □

Приведем одну теорему о сильной отделимости двух выпуклых множеств.

Теорема 3. *Пусть A и B — два выпуклых замкнутых множества, не имеющие общих точек, причем хотя бы одно из этих множеств ограничено. Тогда множества A и B сильно отделимы.*

Доказательство. Введем множество $X = A - B$. Покажем, что оно замкнуто. Пусть x — некоторая предельная точка множества X, пусть последовательность $\{x_k\} \in X$ сходится к x. Поскольку $x_k \in X$, то найдутся $a_k \in A$, $b_k \in B$ такие, что $x_k = a_k - b_k$, $k=1,2,\ldots$. По условию одно из множеств A или B ограничено. Пусть для определенности ограничено множество A. Тогда последовательность $\{a_k\} \in A$ ограничена. По теореме Больцано — Вейерштрасса найдется подпоследовательность $\{a_{k_n}\}$ сходящаяся к некоторой точке a. В силу замкнутости A точка a принадлежит A. Тогда $b_{k_n} = a_{k_n} - x_{k_n} \to b = a - x$ при $k_n \to \infty$, причем $b \in B$ в силу замкнутости B. Таким образом, для точки x получили представление $x = a - b$, где $a \in A$, $b \in B$. Это значит, что $x \in X$. Замкнутость X доказана.

Далее, по условию множества A и B не имеют общих точек. Поэтому $0 \notin X = \overline{X}$. По теореме 1 тогда существует гиперплоскость $\langle c, x\rangle = 0$ такая, что $\langle c, x\rangle \geqslant |c|^2 > 0$ для всех $x \in X$. Отсюда имеем $\langle c, a-b\rangle \geqslant |c|^2$, или $\langle c, a\rangle \geqslant$

$\geqslant \langle c, b\rangle + |c|^2$ для всех $a \in A$, $b \in B$. Следовательно, $\inf\limits_{a\in A}\langle c, a\rangle \geqslant \sup\limits_{b\in B}\langle c, b\rangle + |c|^2$. Любая гиперплоскость $\langle c, x\rangle = \gamma$, где $\sup\limits_{b\in B}\langle c, b\rangle \leqslant \gamma \leqslant \inf\limits_{a\in A}\langle c, a\rangle$ будет сильно отделять множества A и B, что и требовалось. □

Заметим, что требование ограниченности хотя бы одного из множеств в теореме 3 не может быть ослаблено (см. рис. 4.14).

2. Теоремы отделимости являются одним из важных инструментов исследования свойств выпуклых функции и множеств, экстремальных задач. Ряд приложений этих теорем будут даны в последующих параграфах. Здесь же мы воспользуемся ими для получения представления любого выпуклого замкнутого множества из E^n в виде пересечения некоторого семейства полупространств.

Определение 2. Гиперплоскость $\Gamma = \{u \in E^n: \langle c, u\rangle = \gamma\}$ называют *опорной* к множеству X, если $\langle c, x\rangle \geqslant \gamma$ при всех $x \in X$ и $\langle c, y\rangle = \gamma$ для некоторой точки $y \in \overline{X}$. Опорную к X гиперплоскость Γ называют *собственно опорной* к X, если X не содержится в Γ, т. е. $\langle c, x_0\rangle > \gamma$ при некотором $x_0 \in X$. Вектор c, являющийся нормальным вектором опорной [собственно опорной] к X гиперплоскости, проходящей через точку $y \in \overline{X}$, называют *опорным* [*собственно опорным*] вектором множества X в точке y (рис. 4.16).

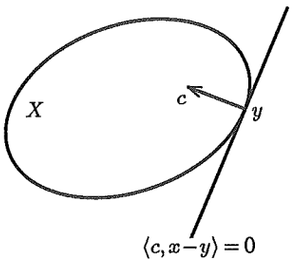

Рис. 4.16

Отметим, что через любую граничную точку y выпуклого множества X из E^n может быть проведена хотя бы одна опорная к X гиперплоскость. В самом деле, если $\operatorname{int} X \neq \varnothing$, то граничными для X будут только точки $y \in \overline{X}$, $y \notin \operatorname{int} X$, и согласно теореме 1 через каждую такую точку y можно провести собственно опорную к X гиперплоскость. Если $\operatorname{int} X = \varnothing$, то $\operatorname{aff} X \neq E^n$, $\operatorname{Гр} X = \overline{X}$, и через каждую точку $y \in \overline{X}$ можно провести гиперплоскость Γ: $\langle c, x - y\rangle = 0$, где c — любой ненулевой вектор из ортогонального дополнения к $\operatorname{Lin} X$. Тогда $X \subset \operatorname{aff} X \subset \Gamma$, так что Γ — опорная к X гиперплоскость, не являющаяся собственно опорной. Теорема 1 уточняет, что если $y \in \overline{X}$, $y \notin \operatorname{ri} X$, то среди опорных к X гиперплоскостей, проходящих через точку y, можно найти собственно опорную.

Заметим, что выпуклое множество с непустой внутренностью не может иметь опорных гиперплоскостей, не являющихся собственно опорными. Это вытекает из следующего несколько более общего утверждения.

Теорема 4. *Пусть X — выпуклое множество из E^n, $\operatorname{int} X \neq \varnothing$. Пусть вектор $c \neq 0$ и число γ таковы, что $\langle c, x\rangle \geqslant \gamma$ при всех $x \in X$. Тогда*

$$\langle c, x\rangle > \gamma \quad \forall x \in \operatorname{int} X.$$

Доказательство. Допустим противное: пусть существует такая точка $x_0 \in \operatorname{int} X$, что $\langle c, x_0\rangle = \gamma$. По определению внутренней точки найдется такое $\varepsilon_0 > 0$, что $z = x_0 - \varepsilon c/|c| \in X$ при всех ε, $0 < \varepsilon < \varepsilon_0$. Тогда $\gamma \leqslant \langle c, z\rangle = \langle c, x_0\rangle - \varepsilon|c| = \gamma - \varepsilon|c| < \gamma$. Получилось противоречивое неравенство, из которого и следует утверждение теоремы. □

Следствие 1. *Пусть X — выпуклое множество из E^n, $\operatorname{int} X \neq \varnothing$. Тогда любая гиперплоскость $\langle c, x - y\rangle = 0$, опорная к множеству X в какой-либо точке $y \in \operatorname{Гр} X$, является собственно опорной к X, или, точнее,*

$$\langle c, x - y\rangle > 0 \quad \forall x \in \operatorname{int} X.$$

Доказательство. Из определения опорной гиперплоскости к X в точке y следует, что $\langle c, x\rangle \geqslant \langle c, y\rangle = \inf\limits_{x\in X}\langle c, x\rangle = \inf\limits_{x\in\overline{X}}\langle c, x\rangle = \gamma$ при всех $x \in X$. В силу теоремы 4 тогда $\langle c, x\rangle > \gamma = \langle c, y\rangle$ для любой точки $x \in \operatorname{int} X$. □

В следующей теореме показывается, что выпуклое замкнутое множество полностью характеризуется своими опорными гиперплоскостями.

Теорема 5. *Всякое непустое выпуклое замкнутое множество X из E^n ($X \neq E^n$) является пересечением замкнутых полупространств, образованных всевозможными опорными гиперплоскостями к множеству X, содержащими X.*

Доказательство. Поскольку $X \neq E^n$, то $\operatorname{Гр} X \neq \varnothing$. Возьмем любую точку $y \in \operatorname{Гр} X$. Множество всех опорных векторов множества X в точке y обозначим через C_y. Выше было замечено, что $C_y \neq \varnothing$ при всех $y \in \operatorname{Гр} X$. Обозначим $A = \bigcap\limits_{y\in\operatorname{Гр} X}\ \bigcap\limits_{c\in C_y}\{x: \langle c, x - y\rangle \geqslant 0\}$. Нам

надо показать, что $A = X$. Если $x \in X$, то для всех $y \in \operatorname{Гр} X$ и всех $c \in C_y$ имеем $\langle c, x - y\rangle \geqslant 0$, т. е. $x \in A$. Следовательно, $X \subset A$.

Докажем обратное включение $A \subset X$. Допустим противное: пусть существует точка $a \in A$, $a \notin X$. Поскольку X — замкнутое множество, то по теореме 1 множество X и точка a сильно отделимы. Точнее, при доказательстве теоремы 1 было показано, что гиперплоскость $\langle c_z, u - z\rangle = 0$, где $z = \mathcal{P}_X(a)$, $c_z = z - a$, такова, что $\langle c_z, x - z\rangle \geqslant 0$ при всех $x \in X = \overline{X}$, а $\langle c_z, a - z\rangle < 0$. Это значит, что $c_z \in C_z$, $z \in \operatorname{Гр} X$, и поэтому для точки $a \in A$ должно бы быть $\langle c_z, a - z\rangle \geqslant 0$ в силу определения A. Полученное противоречие показывает, что $A \subset X$. Требуемое равенство $A = X$ доказано. □

Согласно теореме 5 выпуклое замкнутое множество характеризуется системой неравенств $\langle c, x\rangle \geqslant \langle c, y\rangle$, $x \in X$, которые можно записать в виде $\inf\limits_X\langle c, x\rangle = \langle c, y\rangle$, $c \in C_y$, $y \in \operatorname{Гр} X$. Если заменить c на $e = -c$, то эти условия приводят к равенствам $\sup\limits_X\langle e, x\rangle = \langle e, y\rangle$, $e \in -C_y$, $y \in \operatorname{Гр} X$. Таким образом, всякое выпуклое замкнутое множество X характеризуется значениями функции $\delta(e, X) = \sup\limits_X\langle e, x\rangle$, называемой *опорной функцией* множества X (здесь возможны значения $\delta(e, X) = \infty$ для некоторых $e \in E^n$). Это обстоятельство отражено также и в следующей теореме.

Теорема 6. *Пусть для двух множеств A, B из E^n известно, что*

$$\sup_{a\in A}\langle e, a\rangle \leqslant \sup_{b\in B}\langle e, b\rangle \quad \forall e \in E^n, \quad |e| = 1.$$

Тогда $\overline{\operatorname{co} A} \subseteq \overline{\operatorname{co} B}$; в частности, если A, B выпуклы и замкнуты, то $A \subseteq B$. Если

$$\sup_{a\in A}\langle e, a\rangle = \sup_{b\in B}\langle e, b\rangle \quad \forall e \in E^n, \quad |e| = 1,$$

то $\overline{\operatorname{co} A} = \overline{\operatorname{co} B}$; в частности, если A, B выпуклы и замкнуты, то $A = B$.

Доказательство. Допустим, что $\overline{\operatorname{co} A} \not\subseteq \overline{\operatorname{co} B}$. Тогда существует точка $a_0 \in \overline{\operatorname{co} A}$, но $a_0 \notin \overline{\operatorname{co} B}$. По теореме 1 множество $\overline{\operatorname{co} B}$ и точка a_0 сильно отделимы, т. е. существуют такое $e_0 \in E^n$, $|e_0| = 1$, $\varepsilon_0 > 0$, что $\langle e_0, b\rangle \leqslant \langle e_0, a_0\rangle - \varepsilon_0$ при всех $b \in \overline{\operatorname{co} B}$. Отсюда, пользуясь теоремой 1.9, имеем $\langle e_0, b\rangle \leqslant \sup\limits_{a\in\overline{\operatorname{co} A}}\langle e_0, a\rangle - \varepsilon_0 = \sup\limits_{a\in A}\langle e_0, a\rangle - \varepsilon_0$ при всех $b \in \overline{\operatorname{co} B}$, так что $\sup\limits_{b\in B}\langle e_0, b\rangle = \sup\limits_{b\in\overline{\operatorname{co} B}}\langle e_0, b\rangle \leqslant \sup\limits_{a\in A}\langle e_0, a\rangle - \varepsilon_0 < \sup\limits_{a\in A}\langle e_0, a\rangle$. Пришли к противоречию с условием теоремы. Следовательно, $\overline{\operatorname{co} A} \subseteq \overline{\operatorname{co} B}$. Если A, B выпуклы и замкнуты, то в силу теорем 1.2, 1.6 $A = \operatorname{co} A = \overline{A} = \overline{\operatorname{co} A}$, $\overline{\operatorname{co} B} = B$, и поэтому $A \subseteq B$. Справедливость последнего утверждения теоремы следует из того, что равенство $\delta(e, A) = \delta(e, B)$ эквивалентно двум неравенствам $\delta(e, A) \leqslant \delta(e, B)$, $\delta(e, B) \leqslant \delta(e, A)$, $e \in E^n$. □

3. Теорему 2 можно истолковать как необходимое условие пустоты пересечения двух выпуклых множеств A и B: если $A \cap B = \varnothing$, A и B выпуклы (тогда $\operatorname{ri} A \cap \operatorname{ri} B = \varnothing$), то необходимо существуют вектор $c \in E^n$, $c \neq 0$, и число γ такие, что $\langle c, a\rangle \geqslant \gamma$ при всех $a \in A$ и $\langle c, b\rangle \leqslant \gamma$ при всех $b \in B$. Положим $c_1 = c$, $c_2 = -c$, $\gamma_1 = \gamma$, $\gamma_2 = -\gamma$. Тогда приведенное необходимое условие пустоты пересечения двух выпуклых множеств A и B может быть записано в следующей симметричной форме:

$$\langle c_1, u\rangle \geqslant \gamma_1 \quad \forall u \in A,$$
$$\langle c_2, u\rangle \geqslant \gamma_2 \quad \forall u \in B,$$
$$c_1 + c_2 = 0, \quad \gamma_1 + \gamma_2 = 0,$$

где хотя бы один из векторов c_1 или c_2 не равен нулю.

Следующая теорема обобщает это утверждение и дает необходимое условие пустоты пересечения любого конечного числа выпуклых множеств [83; 225; 278]. Она называется теоремой Дубовицкого А. Я., Милютина А. А., которые впервые ее доказали для конусов.

Теорема 7 [465]. *Пусть непустые множества $A_0, A_1, \ldots, A_m$ из E^n выпуклы и $A_0 \cap A_1 \cap \ldots \cap A_m = \varnothing$. Тогда необходимо существуют векторы $c_0, c_1, \ldots, c_m \in E^n$, не все равные нулю, и числа $\gamma_0, \gamma_1, \ldots, \gamma_m$ такие, что*

$$\langle c_i, u\rangle \geqslant \gamma_i \quad \forall u \in A_i, \quad i = 0, \ldots, m, \tag{7}$$
$$c_0 + c_1 + \ldots + c_m = 0, \tag{8}$$
$$\gamma_0 + \gamma_1 + \ldots + \gamma_m = 0. \tag{9}$$

Для доказательства этой теоремы нам понадобится прямое (декартово) произведение конечного числа множеств, а также прямое произведение евклидовых пространств. Напомним соответствующие определения.

Определение 3. Пусть $A_1, \ldots, A_m$ какие-либо множества. Множество A, состоящее из всевозможных упорядоченных наборов (точек) $a=(a_1, \ldots, a_m)$, где $a_i = A_i$, $i=1, \ldots, m$, называется *прямым произведением множеств* $A_1, \ldots, A_m$ и обозначается через $A_1 \times \ldots \times A_m = A$.

Пусть $L^{n_1}, \ldots, L^{n_m}$ — вещественные линейные пространства. Положим $L = L^{n_1} \times \ldots \times L^{n_m}$. Для элементов (точек) $a=(a_1, \ldots, a_m)$, $b=(b_1, \ldots, b_m) \in L$ определим сумму $a+b=(a_1+b_1, \ldots, a_m+b_m)$ и произведение на вещественное число $\alpha a=(\alpha a_1, \ldots, \alpha a_m)$, где под a_i+b_i и αa_i понимаются соответствующие операции в L^{n_i}, $i=1, \ldots, m$. В результате получим вещественное линейное пространство L, называемое *прямым произведением линейных пространств* $L^{n_1}, \ldots, L^{n_m}$. Если $L^{n_i} = E^{n_i}$ — евклидовы пространства размерности n_i, $i=1, \ldots, m$, то в прямом произведении $E = E^{n_1} \times \ldots \times E^{n_m}$ также можно ввести скалярное произведение $\langle a, b\rangle = \langle a_1, b_1\rangle + \ldots + \langle a_m, b_m\rangle$ и норму $|a| = \langle a, a\rangle^{1/2} = (|a_1|^2 + \ldots + |a_m|^2)^{1/2}$, где $\langle a_i, b_i\rangle$ и $|a_i|$ — соответственно скалярное произведение и норма в E^{n_i}. Полученное евклидово пространство E называют *прямым произведением евклидовых пространств* $E^{n_1}, \ldots, E^{n_m}$; размерность пространства E равна $n_1 + \ldots + n_m$. Например, само евклидово пространство E^n является прямым произведением n одномерных евклидовых пространств: $E^n = E^1 \times \ldots \times E^1$. Параллелепипед $\{u=(u^1, \ldots, u^n)\colon \alpha_1 \leqslant u^i \leqslant \beta_i,\ i=1, \ldots, n\}$ представляет собой прямое произведение отрезков $[\alpha_i, \beta_i]$, $i=1, \ldots, n$.

Прямое произведение выпуклых множеств, очевидно, само выпукло.

Доказательство теоремы 7. Пусть $E = E^n \times \ldots \times E^n$ — прямое произведение m n-мерных евклидовых пространств. Тогда $A = A_1 \times \ldots \times A_m \in E$. Введем в E «диагональное» множество $B = \{b=(b_1, \ldots, b_m)\colon b_1 = \ldots = b_m = a_0,\ a_0 \in A_0\}$. Нетрудно видеть, что пересечение $\bigcap_{i=1}^{m} A_i$ пусто тогда и только тогда, когда $A \cap B$ пусто. Далее, A и B — выпуклые множества. По теореме 2 множества A и B отделимы, т. е. существует $c = \{c_1, \ldots, c_m\} \in E$, не все c_i равны нулю и $\langle c, a\rangle \geqslant \langle c, b\rangle$ при всех $a \in A$, $b \in B$, или

$$\sum_{i=1}^{m} \langle c_i, a_i\rangle \geqslant \sum_{i=1}^{m} \langle c_i, b_i\rangle = \Big\langle \sum_{i=1}^{m} c_i, a_0\Big\rangle \qquad \forall a_i \in A_i,\ i=0, \ldots, m. \tag{10}$$

Положим $c_0 = -(c_1 + \ldots + c_m)$, так что равенство (8) будет выполнено. Тогда неравенство (10) принимает вид

$$\sum_{i=0}^{m} \langle c_i, a_i\rangle \geqslant 0 \qquad \forall a_i \in A_i,\ i=0, \ldots, m. \tag{11}$$

Если в этом неравенстве зафиксируем какие-либо $a_i = \bar{a}_i \in A_i$ при всех $i=0, \ldots, m$, кроме $i=k$, то получим $\langle c_k, a_k\rangle \geqslant \sum_{i \neq k} \langle c_i, \bar{a}_i\rangle = \text{const}$ для всех $a_k \in A_k$. Следовательно,

$$\langle c_k, u\rangle \geqslant \gamma_k = \inf_{a \in A_k} \langle c_k, a\rangle > -\infty \qquad \forall u \in A_k,\ k=1, \ldots, m. \tag{12}$$

Положим

$$\gamma_0 = -(\gamma_1 + \ldots + \gamma_m). \tag{13}$$

Тогда, переходя в (11) к нижней грани по всем $a_i \in A_i$, $i=1, \ldots, m$, получаем $\langle c_0, a_0\rangle + \sum_{i=1}^{m} \gamma_i = \langle c_0, a_0\rangle - \gamma_0 \geqslant 0$ для каждого $a_0 \in A_0$ или

$$\langle c_0, u\rangle \geqslant \gamma_0 \quad \forall u \in A_0.$$

Все соотношения (7)–(9) получены.

При некоторых дополнительных ограничениях на множества $A_0, A_1, \ldots, A_m$ теорема 7 обратима. А именно, верна

Теорема 8. *Пусть $A_0, A_1, \ldots, A_m$ — непустые выпуклые множества из E^n, пусть все эти множества, кроме, быть может, одного, открыты. Тогда для того чтобы $A_0 \cap A_1 \cap \ldots \cap A_m = \varnothing$, необходимо и достаточно, чтобы существовали векторы $c_0, c_1, \ldots, c_m \in E^n$, не все равные нулю, и числа $\gamma_0, \gamma_1, \ldots, \gamma_m$, для которых выполнены соотношения (7)–(9).*

Доказательство. Необходимость доказана в теореме 7. Достаточность докажем, рассуждая от противного. Допустим, что условия (7)–(9) выполнены, но тем не менее существует точка $v \in \bigcap_{i=0}^{m} A_i$. Поскольку не все c_i равны нулю, то из (8) вытекает существование по крайней мере двух векторов c_i, c_j, $i \neq j$, отличных от нуля. По условию все множества $A_0, A_1, \ldots, A_m$, кроме, быть может, одного, открыты. Поэтому можем считать $c_i \neq 0$, A_i — открытое множество, т. е. $A_i = \operatorname{int} A_i$.

Согласно условию (7) $\langle c_i, u\rangle \geqslant \gamma_i$ при всех $u \in A_i$. В силу теоремы 4 тогда $\langle c_i, u\rangle > \gamma_i$ для всех $u \in A_i = \operatorname{int} A_i$. В частности, для точки $v \in \bigcap_{j=0}^{m} A_j \subset A_i$ также имеем $\langle c_i, v\rangle \geqslant \gamma_i$. Кроме того, для всех остальных номеров $j \neq i$ также $v \in A_j$ и в силу (7) $\langle c_j, v\rangle \geqslant \gamma_j$. Сложим все эти неравенства. С учетом равенства (9) получим $\langle c_0, v\rangle + \langle c_1, v\rangle + \ldots + \langle c_m, v\rangle > \gamma_0 + \gamma_1 + \ldots + \gamma_m = 0$, т. е. $\langle c_0 + c_1 + \ldots + c_m, v\rangle > 0$. Однако это невозможно в силу равенства (8). Полученное противоречие показывает, что $\bigcap_{i=0}^{m} A_i = \varnothing$. □

Приведенное выше доказательство теоремы 7 принадлежит В. И. Плотникову. Оно привлекает своей простотой и тем, что позволяет убедиться в справедливости теоремы 7 и в бесконечномерных гильбертовых (и более общих) пространствах — ее доказательство при этом остается неизменным, нужно лишь уточнить ссылки на соответствующие теоремы отделимости в бесконечномерных пространствах.

4. Переформулируем теоремы 7, 8 для случая, когда $A_0, A_1, \ldots, A_m$ являются выпуклыми конусами в E^n. Напомним

Определение 4. *Конусом* (с вершиной в нуле) называется множество K, содержащее вместе с любой своей точкой u и точки λu при всех $\lambda > 0$. Если множество K выпукло, то K называют *выпуклым конусом*, если K замкнуто — *замкнутым конусом*, если K открыто — *открытым конусом*.

Рассмотрим множество

$$K^* = \{c \in E^n\colon \langle c, u\rangle \geqslant 0 \quad \forall u \in K\}. \tag{14}$$

Это множество всегда непусто, так как $0 \in K^*$. Далее, если $c \in K^*$, то для λc при любом $\lambda > 0$ имеем $\langle \lambda c, u\rangle = \lambda \langle c, u\rangle \geqslant 0$ для всех $u \in K$, т. е. $\lambda c \in K^*$. Следовательно, K^* — конус.

Определение 5. Конус K^*, определенный посредством (14), называется *двойственным (сопряженным) конусом* к конусу K (рис. 4.17).

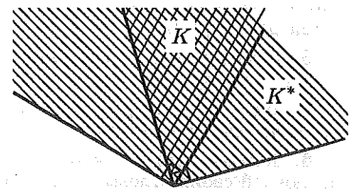

Рис. 4.17

Например, если $K = \{u \in E^n\colon < a, u >= 0\}$ — гиперплоскость, то $K^* = \{c \in E^n\colon c = \lambda a,\ \lambda \in \mathbb{R}\}$; если $K = \{u \in E^n\colon \langle a, u\rangle \leqslant 0\}$ — замкнутое полупространство или $K = \{u \in E^n\colon \langle a, u\rangle < 0\}$ — открытое полупространство, то $K^* = \{c \in E^n\colon c = -\lambda a,\ \lambda \geqslant 0\}$; если $K = E^n$, то $K^* = \{0\}$; если $K = \{0\}$, то $K^* = E^n$; если $K = \{u \in E^n\colon u \geqslant 0\}$, то $K^* = \{c \in E^n\colon c \geqslant 0\}$.

С помощью двойственных конусов удобно переформулировать теорему 7 для случая, когда множества $A_0, A_1, \ldots, A_m$ являются конусами.

Теорема 9. Пусть $K_0, K_1, \ldots, K_m$ — непустые выпуклые конусы из E^n (с вершиной в нуле), пусть $K_0 \cap K_1 \cap \ldots \cap K_m = \varnothing$. Тогда необходимо существуют векторы $c_0, c_1, \ldots, c_m$, не все равные нулю, $c_i \in K_i^*$, $i=0, \ldots, m$, и такие, что

$$c_0 + c_1 + \ldots + c_m = 0. \tag{15}$$

Доказательство. Согласно теореме 7 существуют векторы $c_0, c_1, \ldots, c_m$, не все равные нулю, и числа $\gamma_0, \gamma_1, \ldots, \gamma_m$, удовлетворяющие условиям (7)–(9). Воспользуемся тем, что рассматриваемые множества $K_0, K_1, \ldots, K_m$ являются именно конусами, и покажем, что тогда $\gamma_0 = \gamma_1 = \ldots = \gamma_m = 0$. В самом деле, если $\langle c_i, u\rangle \geqslant \gamma_i$ при всех $u \in K_i$, то $\langle c_i, \lambda u\rangle \geqslant \gamma_i$ или $\langle c_i, u\rangle \geqslant \gamma_i/\lambda$ для любых $\lambda > 0$ и $u \in K_i$. Отсюда при $\lambda \to +\infty$ получим $\langle c_i, u\rangle \geqslant 0$ при всех $u \in K_i$, т. е. $c_i \in K_i^*$, $i=0, \ldots, m$.

Кроме того, если $u \in K_i$, то, взяв в неравенстве $\langle c_i, u\rangle \geqslant 0$ вместо u точку λu при малых $\lambda > 0$, получим сколь угодно малые значения функции $\langle c_i, u\rangle$ на K_i и придем к равенству $\inf_{u \in K_i} \langle c_i, u\rangle = 0$. Согласно (12) это означает, что все величины γ_i, $i=1, \ldots, m$, участвующие в неравенствах (7), равны нулю. Из (13) тогда имеем $\gamma_0 = 0$, Таким образом, если в теореме 7 множества $A_0, A_1, \ldots, A_m$ являются выпуклыми конусами, то условие (9) выполняется тривиально, так как все $\gamma_i = 0$, $i=0, \ldots, m$, условия (7) означают, что $c_i \in K_i^*$, $i=0, \ldots, m$, а из (8) следует (15). □

При некоторых дополнительных ограничениях на конусы $K_0, K_1, \ldots, K_m$ теорема 9 обратима. А именно верна

Теорема 10. *Пусть $K_0, K_1, \ldots, K_m$ — непустые выпуклые конусы из E^n (с вершиной в нуле), пусть все эти конусы, кроме, быть может, одного, открыты. Тогда для того чтобы $K_0 \cap K_1 \cap \ldots \cap K_m = \varnothing$, необходимо и достаточно, чтобы существовали векторы $c_0, c_1, \ldots, c_m$, не все равные нулю, $c_i \in K_i^*$, $i = 0, \ldots, m$ и удовлетворяющие равенству* (15).

Доказательство. Необходимость доказана в теореме 9. Достаточность вытекает из теоремы 8, если заметить, что условие $c_i \in K_i^*$ равносильно неравенству $\langle c_i, u\rangle \geqslant 0 \ \forall u \in K_i$, то отсюда и из (15) следуют условия (7)–(9) при $\gamma_0 = \gamma_1 = \ldots = \gamma_m = 0$. □

Приведенные в этом параграфе теоремы отделимости и их различные обобщения играют важную роль в выпуклом анализе, в теории и методах математического программирования, оптимального управления, в теории уравнений и неравенств и т. д. (см., например, [48–50; 54; 83; 192; 225; 278; 279; 613; 617; 752]).

Упражнения

1. Пусть A и B — выпуклые множества, не имеющие общих внутренних точек. Можно ли утверждать, что A и B отделимы? Рассмотреть пример $A = \{u = (x, y)\colon y = 0, |x| \leqslant 1\}$, $B = \{u = (x, y)\colon x = 0, |y| \leqslant 1\}$ в E^2.

2. Пусть X — выпуклое множество из E^n, $\operatorname{int} X = \varnothing$. Доказать, что любая гиперплоскость, опорная к X и проходящая через точку $y \in \operatorname{ri} X$, содержит X, т. е. не является собственно опорной.

3. Пусть A — выпуклое множество из E^n, причем $A \cap \operatorname{int} E_+^n = \varnothing$. Доказать, что существует такой вектор $c = (c_1, \ldots, c_n) \neq 0$, $c_1 \geqslant 0, \ldots, c_n \geqslant 0$, что $< c, a > \leqslant 0$ при всех $a \in A$.

4. Пусть A — выпуклое множество из E^n, M — аффинное или многогранное множество из E^n. Для того чтобы A и M были собственно отделимы и разделяющая гиперплоскость не содержала A, необходимо и достаточно, чтобы $M \cap \operatorname{ri} A = \varnothing$. Доказать.

5. Пусть $\rho(A, B) = \inf\limits_{a \in A} \inf\limits_{b \in B} |a - b|$ — расстояние между множествами A и B. Доказать, что два непустых выпуклых множества A, B из E^n сильно отделимы тогда и только тогда, когда $\rho(A, B) > 0$.

6. Доказать, что всякое выпуклое замкнутое ограниченное множество из E^n имеет хотя бы одну угловую точку (см. определение 3.2.1).

7. Пусть X — выпуклое замкнутое множество из E^n. Доказать, что X имеет хотя бы одну угловую точку тогда и только тогда, когда X не содержит прямых.

8. Доказать, что выпуклое замкнутое ограниченное множество A из E^n является выпуклой оболочкой своих угловых точек. Показать, что без требования ограниченности множества A это утверждение неверно. Рассмотреть пример $A = \{a = (x, y) \in E^2\colon y \geqslant |x|\}$.

9. Если $A_1, \ldots, A_m$ — выпуклые множества из E^n, причем $\bigcap\limits_{i=1}^m \operatorname{ri} A_i \neq \varnothing$, то $\overline{\bigcap\limits_{i=1}^m A_i} = \bigcap\limits_{i=1}^m \overline{A}_i$; $\operatorname{ri}\left(\bigcap\limits_{i=1}^m A_i\right) = \bigcap\limits_{i=1}^m (\operatorname{ri} A_i)$; $\operatorname{aff}\left(\bigcap\limits_{i=1}^m A_i\right) = \bigcap\limits_{i=1}^m (\operatorname{aff} A_i)$. Доказать.

10. Пусть K — произвольный конус из E^n. Доказать, что тогда конус K^* будет замкнутым и выпуклым.

11. Доказать, что если K — выпуклый конус, то конусы $\overline{K}$, $\operatorname{ri} K$ также выпуклы и $K^* = (\overline{K})^* = (\operatorname{ri} K)^*$.

12. Доказать, что если K — замкнутый выпуклый конус, то $(K^*)^* = K$.

13. Пусть K — выпуклый конус. Доказать, что для ограниченности снизу линейной функции $\langle c, u\rangle$ на K необходимо и достаточно, чтобы $c \in K^*$.

14. Пусть K — выпуклый конус, $\operatorname{int} K \neq \varnothing$. Тогда $\langle c, u\rangle > 0$ для всех $u \in \operatorname{int} K$ при любом выборе $c \in K^*$ $(c \neq 0)$. Доказать.

15. Доказать, что если $K_1, \ldots, K_m$ выпуклые конусы, то $K = K_1 + \ldots + K_m$ — выпуклый конус, причем $K = \operatorname{co}(K_1 \cup K_2 \cup \ldots \cup K_m)$. Можно ли утверждать, что если $K_1, \ldots, K_m$ — выпуклые замкнутые конусы, то K — выпуклый замкнутый конус? (см. упражнение 1.14, в)).

16. Доказать, что $(K_1 + \ldots + K_m)^* = K_1^* \cap \ldots \cap K_m^*$, где $K_1, \ldots, K_m$ — конусы из E^n.

17. Пусть K_1, K_2 — выпуклые замкнутые конусы. Доказать, что $(K_1 \cap K_2)^* = \overline{K_1^* + K_2^*}$.

18. Пусть $K_0, K_1, \ldots, K_m$ — выпуклые конусы, пусть $K_0 \cap \operatorname{int} K_1 \cap \ldots \cap \operatorname{int} K_m \neq \varnothing$. Тогда $\left(\bigcap\limits_{i=0}^m K_i\right)^* = K_0^* + K_1^* + \ldots + K_m^*$. Доказать.

19. Пусть $K_0, K_1, \ldots, K_m$ — выпуклые конусы. Тогда либо $\left(\bigcap\limits_{i=0}^m K_i\right)^* = K_0^* + K_1^* + \ldots + K_m^*$, либо существуют не все равные нулю векторы $c_i \in K_i^*$, $i = 0, \ldots, m$, такие, что $c_0 + c_1 + \ldots + c_m = 0$. Доказать.

20. Для того чтобы выпуклые конусы K_0, K_1 были неотделимы, необходимо и достаточно, чтобы $0 \in \operatorname{int}(K_0 - K_1)$. Доказать.

21. Доказать, что два выпуклых конуса K_0, K_1 неотделимы тогда и только тогда, когда одновременно выполнены два условия: $\operatorname{ri} K_0 \cap \operatorname{ri} K_1 \neq \varnothing$, $\operatorname{Lin} K_0 + \operatorname{Lin} K_1 = E^n$.

22. Для того, чтобы два непустых выпуклых множества A, B из E^n были сильно отделимы, необходимо и достаточно, чтобы $0 \notin \overline{A - B}$. Доказать.

23. Доказать, что два непересекающихся многогранных множества сильно отделимы.

24. Множество $A^* = \{c \in E^n\colon \langle c, u\rangle \leqslant 1 \ \forall u \in A\}$ называется *полярой* множества A. Найти поляры множеств A, если $A = \{0\}$; $A = [a, b] \subset E^1$; $A = \{u \in E^n\colon u = te,\ 0 < t < \infty,\ e \neq 0\}$; $A = \{u \in E^n\colon \langle c, u\rangle \leqslant \gamma\}$; A — шар; A — конус с вершиной в нуле. Выяснить связь между полярой конуса и двойственным конусом.

§ 6. Субградиент. Субдифференциал

1. Для выпуклых дифференцируемых функций на выпуклом множестве выше было доказано неравенство (см. теорему 2.2)

$$f(u) \geqslant f(v) + \langle f'(v), u - v\rangle \qquad \forall u \in X. \tag{1}$$

К сожалению, выпуклая функция может не быть дифференцируемой даже во внутренних точках множества, и в этом случае полезное во многих случаях неравенство (1) не будет иметь смысла. Тем не менее, оказывается, для выпуклых функций это неравенство можно сохранить, если надлежащим образом обобщить понятие градиента.

Определение 1. Пусть функция $f(x)$ определена на множестве X из E^n. Вектор $c = c(v) \in E^n$ называется *субградиентом* функции $f(x)$ в точке $v \in X$, если

$$f(u) \geqslant f(v) + \langle c(v), u - v\rangle \quad \forall u \in X. \tag{2}$$

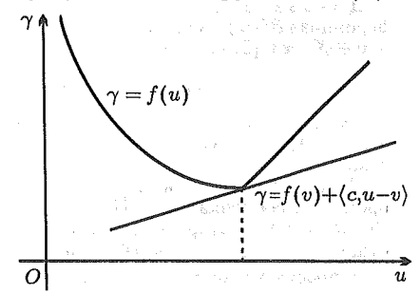

Рис. 4.18

Множество всех субградиентов функции $f(x)$ точке v называют *субдифференциалом* этой функции в точке v и обозначают через $\partial f(v)$.

Неравенство (2) имеет простой геометрический смысл и означает, что график функции $\gamma = f(u)$, $u \in X$ в пространстве переменных (u, γ) лежит не ниже графика линейной функции $\gamma = f(v) + \langle c(v), u - v\rangle$, $u \in X$, причем в точке $u = v$ оба графика пересекаются (рис. 4.18).

Для гладких выпуклых функций, как показывает неравенство (1), субдифференциал непуст и градиенты этих функций являются их субградиен-

тами. Во внутренних точках множества гладкая функция, оказывается, других субградиентов, кроме градиента, иметь не может. В самом деле, пусть $v \in \operatorname{int} X$, $c(v) \in \partial f(v)$. Поскольку $f(u) \in C^1(X)$, то $f(u) = f(v) + \langle f'(v), u - v\rangle + o(|u-v|)$, $u \in X$. Отсюда и из (2) следует, что $\langle f'(v) - c(v), u - v\rangle \geqslant o(|u-v|)$, $u \in X$. Поскольку $v \in \operatorname{int} X$, то $u = v - \varepsilon(f'(v) - c(v)) \in X$ при всех ε, $0 < \varepsilon < \varepsilon_0$. Подставив эту точку в предыдущее неравенство получим $-\varepsilon|f'(v) - c(v)|^2 \geqslant o(\varepsilon)$, $0 < \varepsilon < \varepsilon_0$. Деля на $\varepsilon > 0$ и устремляя $\varepsilon \to +0$, отсюда будем иметь $-|f'(v) - c(v)|^2 \geqslant 0$, т. е. $c(v) = f'(v)$.

Тем самым показано, что для гладкой выпуклой функции $\partial f(v) = \{f'(v)\}$ при всех $v \in \operatorname{int} X$. Существуют функции, которые недифференцируемы в точке, но тем не менее субдифференциал в этой точке непуст.

Пример 1. Функция $f(u) = |g(u)|$, $u \in X$ в точке v, где $g(v) = 0$, всегда имеет субградиент $c(v) = 0$, так как $|g(u)| - |g(v)| = |g(u)| \geqslant 0 = \langle 0, u - v\rangle$ для всех $u \in X$. В то же время в точках v, где $g(v) \neq 0$, эта функция может быть недифференцируемой и не имеющей субградиента.

Пример 2. Пусть $f(u) = |u|$, $u \in E^n$. В точке $v = 0$ эта функция недифференцируема, но для нее верно соотношение

$$f(u) - f(0) = |u| \geqslant \langle c, u - 0\rangle = \langle c, u\rangle, \quad u \in E^n,$$

для всех c, $|c| \leqslant 1$. Это значит, что $\partial(|u|)\Big|_{u=0} = \partial f(0) = \{c \in E^n\colon |c| \geqslant 1\}$ — единичный шар с центром в нуле. Если $v \neq 0$, то $\partial f(v) = \{v/|v| = f'(v)\}$.

Заметим, что в примере 2 функция $f(u) = |u|$ выпукла на E^n. Оказывается, если совсем отказаться от выпуклости функции, то даже гладкая функция может не иметь субградиента ни в одной точке. Например, для функции $f(u) = u^3$ на E^1 субдифференциал пуст во всех точках. В то же время эта функция $f(u) = u^3$ на множестве $X = \{u \in E^1\colon u \geqslant 0\}$ выпукла и во всех точках $v \in X$ имеет на X непустой субдифференциал. Ниже увидим, что это не случайно.

2. Следующая теорема показывает, что понятия субградиента и субдифференциала являются естественными для выпуклых функций.

Теорема 1. *Пусть X открытое выпуклое множество на E^n (например, возможно, $X \equiv E^n$). Тогда для того чтобы функция $f(x)$, определенная на X, имела непустой субдифференциал во всех точках X, необходимо и достаточно, чтобы $f(x)$ была выпукла на множестве X.*

Доказательство. Необходимость. Пусть для некоторой функции $f(x)$ субдифференциал $\partial f(u) \neq \varnothing$ при всех $u \in X$. Покажем, что $f(x)$ выпукла на X. Возьмем произвольные $u, v \in X$, $\alpha \in [0, 1]$ и положим $u_\alpha = \alpha u + (1-\alpha)v$. Пусть $c = c(u_\alpha) \in \partial f(u_\alpha)$. Тогда

$$f(u) - f(u_\alpha) \geqslant \langle c, u - u_\alpha\rangle, \quad f(v) - f(u_\alpha) \geqslant \langle c, v - u_\alpha\rangle.$$

Умножим первое из этих неравенств на α, второе на $1-\alpha$ и сложим. Получим $\alpha f(u) + (1-\alpha)f(v) - f(u_\alpha) \geqslant \langle c, u_\alpha - u_\alpha\rangle = 0$ при всех $u, v \in X$, $\alpha \in [0, 1]$. Выпуклость $f(x)$ на X доказана.

Достаточность. Пусть $f(x)$ выпукла на открытом выпуклом множестве X. Пусть v — произвольная точка на X. Покажем, что $\partial f(v) \neq \varnothing$. Возьмем некоторый единичный вектор e. Поскольку X — открытое множество, то $v + te \in X$ при всех t, $0 \leqslant t \leqslant t_0$, $t_0 > 0$. По теореме 2.14 существует производная $df(v)/de$ по направлению e.

В пространстве E^{n+1} переменных (u, γ) введем два множества

$$A = \{(u, \gamma) \in E^{n+1}\colon u \in X,\ \gamma > f(u)\},$$

$$B = \{(u, \gamma) \in E^{n+1}\colon u = v + te,\ \gamma = f(v) + t\frac{df(v)}{de},\ 0 \leqslant t \leqslant t_0\}.$$

Нетрудно показать, что множество A выпукло — это делается так же, как доказывалась выпуклость надграфика, выпуклой функции в теореме 2.9 (кстати, в данном случае $A = \operatorname{int}(\operatorname{epi} f)$).

Множество B является отрезком прямой в E^{n+1} и тоже выпукло. Покажем, что множества A и B не имеют общих точек.

В самом деле, пусть $(u, \gamma) \in A$. Имеются две возможности: 1) $u \neq v + te$ при всех t, $0 \leqslant t \leqslant t_0$ — тогда заведомо $(u, \gamma) \notin B$; 2) при некотором t, $0 \leqslant t \leqslant t_0$, оказалось, что $u = v + te$ — тогда с учетом неравенства (1.8.4) и $\gamma > f(u) = f(v + te)$ имеем $\gamma - f(v) > f(v + te) - f(v) \geqslant tdf(v)/de$, т. е. $\gamma > f(v) + tdf(v)/de$, и снова $(u, \gamma) \notin B$.

Итак, множества A и B выпуклы, $A \cap B = \varnothing$. По теореме 5.2 тогда существует гиперплоскость с нормальным вектором $(d, \nu) \neq 0$, отделяющая A и B, т. е.

$$\langle d, u\rangle + \nu\gamma \geqslant \langle d, v + te\rangle + \nu\left(f(v) + t\frac{df(v)}{de}\right) \tag{3}$$

при всех $\gamma > f(u)$, $u \in X$, $0 \leqslant t \leqslant t_0$. В частности, при $u = v$, $t = 0$ из (3) имеем $\nu(\gamma - f(v)) \geqslant 0$ для всех $\gamma > f(v)$. Отсюда следует, что $\nu \geqslant 0$.

Допустим, что $\nu = 0$. Тогда из (3) имеем $\langle d, u\rangle \geqslant \langle d, v + te\rangle$ для всех $u \in X$, $0 \leqslant t \leqslant t_0$. Положим здесь $u = v + \varepsilon d$, $t = 0$ — так можно делать, ибо $v + \varepsilon d \in X$ при всех ε, $0 < |\varepsilon| < \varepsilon_0$ в силу открытости X. Получим $\langle d, v + \varepsilon d\rangle \geqslant \langle d, v\rangle$, или $\varepsilon|d|^2 \geqslant 0$ при всех ε, $0 < |\varepsilon| \leqslant \varepsilon_0$, что возможно только при $d = 0$. Однако $(d, \nu) \neq 0$ по построению. Полученное противоречие показывает, что $\nu = 0$ не может быть.

Итак, $\nu > 0$. Поделим (3) на $\nu > 0$. Обозначая $c = -d/\nu$ и устремляя $\gamma \to f(u) + 0$, из (3) получаем

$$f(u) - f(v) + t\langle c, e\rangle \geqslant \langle c, u - v\rangle + t\frac{df(v)}{de} \tag{4}$$

при всех $u \in X$ и всех $0 \leqslant t \leqslant t_0$. Полагая здесь $t = 0$, будем иметь

$$f(u) - f(v) \geqslant \langle c, u - v\rangle \quad \forall u \in X.$$

Это означает, что $c \in \partial f(v)$, т. е. $\partial f(v) \neq \varnothing$. □

В следующей теореме изучаются некоторые свойства субдифференциала выпуклой функции.

Теорема 2. *Пусть X — открытое выпуклое множество из E^n (например, $X = E^n$), $f(x)$ — выпуклая функция на X. Тогда субдифференциал $\partial f(v)$ при всех $v \in X$ является непустым выпуклым, замкнутым и ограниченным множеством.*

Доказательство. Непустота субдифференциала доказана в теореме 1. Покажем выпуклость $\partial f(v)$. Пусть $c_1, c_2 \in \partial f(v)$, т. е.

$$f(u) - f(v) \geqslant \langle c_1, u - v\rangle, \quad f(u) - f(v) \geqslant \langle c_2, u - v\rangle, \quad u \in X.$$

Возьмем $\alpha \in [0, 1]$. Умножая первое неравенство на α, второе на $1-\alpha$ и, складывая, получаем $f(u) - f(v) \geqslant \langle \alpha c_1 + (1-\alpha)c_2, u - v\rangle$ при всех $u \in X$. Это значит, что $\alpha c_1 + (1-\alpha)c_2 \in \partial f(v)$ для любых $\alpha \in [0, 1]$. Выпуклость $\partial f(v)$ доказана.

Пусть c — предельная точка множества $\partial f(v)$, пусть $\{c_k\} \in \partial f(v)$ и $c_k \to c$ при $k \to \infty$. Из $c_k \in \partial f(v)$ следует, что $f(u) - f(v) \geqslant \langle c_k, u - v\rangle$ $\forall u \in X$. При $k \to \infty$ отсюда получим $c \in \partial f(v)$. Замкнутость $\partial f(v)$ доказана.

Покажем ограниченность $\partial f(v)$. Возьмем любой вектор $c \in \partial f(v)$. Поскольку X — открытое множество, то $S(v, \varepsilon) = \{u \in E^n\colon |u - v| \leqslant \varepsilon\} \subset X$ при достаточно малом $\varepsilon > 0$. Далее, в силу теорем 2.15 и 2.1.4 $\sup\limits_{S(v, \varepsilon)} f(u) = F^*(S) < \infty$. Положим в неравенстве (2) $u = v + \varepsilon c/|c| \in S(v, \varepsilon)$. Получим $|c| \leqslant (f(v + \varepsilon c/|c|) - f(v))/\varepsilon < (f^*(S) - f(v))/\varepsilon < \infty$ при всех $c \in \partial f(v)$. □

3. Теоремы 1, 2 не дают конструктивного описания субдифференциала выпуклой функции. Такое описание удается получить лишь в немногих случаях.

Пример 3. Пусть $f(u) = \max\limits_{i \in I} u^i$, $u = (u^1, \ldots, u^n) \in E^n$, $I = \{1, \ldots, n\}$. Согласно теореме 2.7 функция $f(u)$ выпукла на E^n. Покажем, что

$$\partial f(v) = \{c = (c_1, \ldots, c_n)\colon c_i \geqslant 0,\ i \in I(v),\ c_i = 0,\ i \notin I(v);\ c_1 + \ldots + c_n = 1\}, \tag{5}$$

где $I(v) = \{i \in I\colon \max\limits_{j \in I} v^j = v^i\}$. Множество, определяемое правой частью формулы (5), обозначим через $A(v)$. Пусть $c \in A(v)$. Умножим неравенство $f(u) - f(v) = \max\limits_{j \in I} u^j - v^i \geqslant u^i - v^i$, верное при всех $i \in I(v)$, $u \in E^n$ на $c_i \geqslant 0$ и сложим по всем $i \in I(v)$. С учетом равенств $c_i = 0$ при $i \notin I(v)$, $c_1 + \ldots + c_n = 1$ получим $f(u) - f(v) \geqslant \langle c, u - v\rangle$ $\forall u \in E^n$, т. е. $c \in \partial f(v)$. Это значит, что $A(v) \subseteq \partial f(v)$.

Докажем включение $\partial f(v) \subseteq A(v)$. Пусть $c \in \partial f(v)$, т. е.

$$f(u) - f(v) = \max_{i \in I} u^i - \max_{i \in I} v^i \geqslant \langle c, u - v\rangle \quad \forall u \in E^n. \tag{6}$$

Возьмем в (6) $u = u_\pm = (v^1 \pm 1, \ldots, v^n \pm 1)$. Тогда $f(u_\pm) - f(v) = \pm 1$ и из (6) получим $\pm 1 \geqslant \langle c, u_\pm - v\rangle = \pm \sum_{i=1}^{n} c_i$, что возможно лишь при $\sum_{i=1}^{n} c_i = 1$. Далее, в (6) возьмем $u_\varepsilon = (u^1, \ldots, u^n)$, где $u^i = v^i - \varepsilon$ при некотором $i \in I$, $u^j = v^j$ при $j \neq i$, $\varepsilon > 0$. Тогда $f(u_\varepsilon) \leqslant f(v)$ и из (6) получим $0 \geqslant \langle c, u_\varepsilon - v\rangle = c_i(-\varepsilon)$, так что $c_i \geqslant 0$, $i = 1, \ldots, n$.

Далее, пусть $i \notin I(v)$. Тогда $v^i < f(v)$ и можно выбрать $\varepsilon > 0$ так, что $v^i + \varepsilon < f(v)$. Положим $u_\varepsilon = (u^1, \ldots, u^n)$, где $u^i = v^i + \varepsilon$, $u^j = v^j$ при $j \neq i$. Тогда $f(v) = f(u_\varepsilon)$, и из (6) получим $0 \geqslant \langle c, u_\varepsilon - v\rangle = \varepsilon c_i$, т. е. $c_i \leqslant 0$, $i \notin I(v)$. Сравнивая это неравенство с уже доказанным $c_i \geqslant 0$, получаем $c_i = 0$, $i \notin I(v)$. Это значит, что $c \in A(v)$, так что $\partial f(v) \subseteq A(v)$. Равенство (5) установлено.

4. Установим связь между производными по направлению и субдифференциалом выпуклой функции.

Теорема 3. *Пусть X — открытое выпуклое множество из E^n, $f(x)$ — выпуклая функция на X. Тогда во всех точках $v \in X$ производная функции $f(x)$ по любому направлению e, $|e| = 1$, существует, причем*

$$\frac{df(v)}{de} = \max_{c \in \partial f(v)} \langle c, e\rangle. \tag{7}$$

Доказательство. Существование производной $df(v)/de$ установлено в теореме 2.14. Докажем формулу (7). Из (2) имеем $(f(v + te) - f(v))/t \geqslant \langle c, e\rangle$ при всех $c \in \partial f(v)$ и всех достаточно малых $t > 0$. Отсюда при $t \to +0$ получим $df(v)/de \geqslant \langle c, e\rangle$ для любого $c \in \partial f(v)$, так что $df(v)/de \geqslant \sup_{c \in \partial f(v)} \langle c, e\rangle$. С другой стороны, при доказательстве теоремы 1 был построен специальный субградиент c, для которого выполняется неравенство (4). Полагая в (4) $u = v$, будем иметь $df(v)/de \leqslant \langle c, e\rangle$, $c \in \partial f(v)$. Сравнивая это неравенство с предыдущим, приходим к формуле (7). Попутно показали, что максимум в правой части (7) достигается именно на том субградиенте, который был построен в теореме 1. □

Формула (7) обобщает известную формулу $df(v)/de = \langle f'(v), e\rangle$ для гладких функций.

5. С помощью субдифференциала можно сформулировать критерий оптимальности, обобщающий теорему 2.3 на случай негладких выпуклых функций.

Теорема 4. *Пусть $f(x)$ — выпуклая функция на открытом выпуклом множестве W из E^n, X — выпуклое подмножество множества W. Тогда для того чтобы функция $f(x)$ достигала своей нижней грани на множестве X в точке $u_* \in X$, необходимо и достаточно, чтобы существовал субградиент $c_* = c(u_*) \in \partial f(u_*)$ такой, что*

$$\langle c_*, u - u_*\rangle \geqslant 0 \quad \forall u \in X. \tag{8}$$

Если $u_* \in \operatorname{int} X$, то в (8) $c_* = 0$.

Необходимость. Пусть $u_* \in X_* = \{u \in X\colon f(u) = \inf_X f(u) = f_* > -\infty\}$. Так как $f(x)$ выпукла на открытом выпуклом множестве W, то по теореме 2.14 существует производная $\frac{df(u_*)}{de}$ по всем направлениям $e = (u - u_*)|u - u_*|^{-1}$, $u \in X$, $u \neq u_*$. Согласно теореме 2.12 $\frac{df(u_*)}{de} \geqslant 0$. Отсюда и из формулы (7) следует: $\frac{df(u_*)}{de} = \max_{c \in \partial f(u_*)} \langle c, e\rangle \geqslant 0$. Возьмем $\forall c_* \in \partial f(u_*)$, для которого $\max_{c \in \partial f(u_*)} \langle c, e\rangle = \langle c_*, e\rangle$. Тогда $\langle c_*, e\rangle = \langle c_*, u - u_*\rangle |u - u_*|^{-1} \geqslant 0$ $\forall u \in X$, $u \neq u_*$, откуда вытекает неравенство (8).

Если $u_* \in \operatorname{int} X$, то $u = u_* + \varepsilon c_* \in X$ при всех ε, $0 < |\varepsilon| < \varepsilon_0$, и из (8) получим $\langle c_*, \varepsilon c_*\rangle = \varepsilon |c_*|^2 \geqslant 0$, $0 < |\varepsilon| < \varepsilon_0$. Это возможно лишь при $c_* = 0$.

Достаточность. Пусть для некоторой точки $u_* \in X$ выполнено неравенство (8) при каком-либо $c_* \in \partial f(u_*)$. По определению субградиента тогда $f(u) - f(u_*) \geqslant \langle c_*, u - u_*\rangle \geqslant 0$ при всех $u \in X$, т. е. $u_* \in X_*$. □

Замечание 1. Вариационное неравенство (8) можно записать в эквивалентном виде:

$$u_* = \mathcal{P}_X(u_* - \alpha c_*) \quad \forall \alpha > 0.$$

Доказательство этого равенства проводится совершенно также, как и теоремы 4.4.4, и предоставляется читателю.

Следующие примеры показывают, что субградиент c_* из (8) в общем случае определяется неоднозначно.

Пример 4. Пусть $f(u) = |u|$, $u \in W = E^1$. Если $X = E^1$, то $X_* = \{0\}$, $\partial f(0) = [-1, 1]$ и неравенство (8) выполняется лишь при $c_* = 0$. Если $X = \{u \in E^1\colon u \geqslant 0\}$, то $X_* = \{0\}$ и (8) выполняется для всех $c_* \in [0, 1] \subset \partial f(0)$.

Пример 5. Пусть $f(u) = \max\{u; 0\}$, $u \in W = E^1$. Если $X = E^1$, то $X_* = \{0\}$, $\partial f(0) = [0, 1]$ и (8) имеет место лишь для $c_* = 0$. Если $X = \{u \in E^1\colon u \geqslant 0\}$, то по-прежнему $X_* = \{0\}$, но неравенство (8) здесь выполняется для всех $c_* \in \partial f(0) = [0, 1]$.

6. Определение 2. Пусть E^n, E^m — евклидовы пространства, $W \subset E^n$, $\Pi(E^m)$ — множество всех непустых множеств из E^m. Говорят, что на W задано *многозначное отображение* $F\colon W \to \Pi(E^m)$, если каждой точке $u \in W$ поставлено в соответствие некоторое множество $F(u) \subset \Pi(E^m)$.

Определение 3. Многозначное отображение, которое каждой точке u из открытого выпуклого множества $W \subset E^n$ ставит в соответствие субдифференциал $\partial f(u)$ некоторой выпуклой на W функции $f(u)$, называется *субдифференциальным отображением* и обозначается через ∂f (здесь $m = n$).

Субдифференциальное отображение обладает рядом замечательных свойств [264; 604; 605; 617; 670]; на некоторых из них мы здесь кратко остановимся.

Определение 4. Пусть W — множество из E^n. Многозначное отображение $F\colon W \to \Pi(E^m)$ называется:

1) *компактным*, если для любого компактного множества $U \subset W$ множество $F(U) = \bigcup_{u \in U} F(u)$ компактно;

2) *монотонным*, если $\langle c(u) - c(v), u - v\rangle \geqslant 0$ при всех $u, v \in W$, $c(u) \in F(u)$, $c(v) \in F(v)$ (здесь подразумевается, что $n = m$);

3) *выпуклозначным*, если $F(u)$ — выпуклое множество при каждом $u \in W$;

4) *замкнутым (полунепрерывным сверху)* в точке $v \in W$, если из того, что $\{v_k\} \to v$, $v_k \in W$, и $\{c_k\} \to c$, $c_k \in F(v_k)$, $k = 1, 2, \ldots$ следует $c \in F(v)$.

Теорема 5. *Пусть $f(x)$ — выпуклая функция на открытом выпуклом множестве W из E^n. Тогда субдифференциальное отображение $\partial f\colon W \to \Pi(E^n)$ выпуклозначно, монотонно, замкнуто, компактно.*

Доказательство. Выпуклозначность отображения ∂f следует из теоремы 2. Возьмем произвольные $u, v \in W$, $c(u) \in \partial f(u)$, $c(v) \in \partial f(v)$. Тогда согласно (2) $f(u) - f(v) \geqslant \langle c(v), u - v\rangle$, $f(v) - f(u) \geqslant \langle c(u), v - u\rangle$. Сложив эти два неравенства, получим $\langle c(u) - c(v), u - v\rangle \geqslant 0$. Монотонность ∂f установлена. Далее, пусть $v \in W$, $\{v_k\} \to v$, $v_k \in W$, пусть $\{c_k\} \to c$, $c_k \in \partial f(v_k)$. Это значит, что $f(u) - f(v_k) \geqslant \langle c_k, u - v_k\rangle$ при всех $u \in W$. Поскольку функция $f(u)$ непрерывна на W (см. теорему 2.15), то, переходя к пределу в этом неравенстве при $k \to \infty$, приходим к неравенству (2). Это значит, что $c \in \partial f(v)$. Замкнутость доказана.

Наконец, возьмем произвольное ограниченное замкнутое множество $U \subset W$. Поскольку W — открытое множество, то все точки U являются внутренними для W и найдется такое число $\delta > 0$, что ограниченное замкнутое множество $U_\delta = \{u \in E^n\colon |u - v| \leqslant \delta, v \in U\}$, представляющее собой δ-раздутие множества U, принадлежит W. В самом деле, если $W = E^n$, то $U_\delta \subset E^n$ при любом $\delta > 0$. Если же $W \neq E^n$, то граница Гр W выпуклого множества непуста и $\rho(v, \text{Гр } W) = \inf_{w \in \text{Гр } W} |v - w| > 0$ при всех $v \in U$. В силу леммы 2.1.2 функция $\rho(v, \text{Гр } W)$ непрерывна на компактном множестве U и согласно теореме 2.1.1 найдется такая точка $v_* \in U$, что $\inf_{v \in U} \rho(v, \text{Гр } W) = \rho(v_*, \text{Гр } W) = 2\delta > 0$. Это значит, что $U_\delta \subset W$. Функция $f(u)$ непрерывна на компактном множестве U_δ, поэтому $\sup_{U_\delta} f(u) = f_\delta^* < \infty$ (теорема 2.1.4).

Возьмем любые $v \in U$, $c \in \partial f(v)$. В неравенстве (2) положим $u = v + \delta c/|c| \subset U_\delta \subset W$. Получим $|c| \leqslant [f(v + \delta c/|c|) - f(v)]/\delta \leqslant 2f_\delta^*/\delta = M < \infty$ для всех $c \in \partial f(v)$, $v \in U$. Таким образом, $\sup_{v \in V} \sup_{c \in \partial f(v)} |c| = \sup_{c \in \partial f(U)} |c| \leqslant M < \infty$, т. е. множество $\partial f(U)$ ограничено. Докажем замкнутость $\partial f(U)$. Пусть $\{c_k\} \to c$, $c_k \in \partial f(U)$. Это значит, что существует такая точка $v_k \in U$, что $c_k \in \partial f(v_k)$. Поскольку U — компактное множество, то без ограничения общности считаем, что $\{v_k\} \to v \subset U$. В силу замкнутости отображения ∂f тогда $c \in \partial f(v) \subset \partial f(U)$. Следовательно, $\partial f(U)$ — замкнутое множество. Компактность отображения ∂f установлена. □

Опираясь на теорему 5, можем уточнить свойства непрерывности и дифференцируемости выпуклых функций на открытом множестве, в частности, обобщить теорему 1.8.3 на многомерный случай.

Теорема 6. *Пусть $f(x)$ — выпуклая функция на открытом выпуклом множестве W из E^n. Тогда на любом компактном множестве $G \subset W$ функция $f(x)$ удовлетворяет*

условию Липшица, т. е. существует такая постоянная $L = L(G) \geqslant 0$, *что* $|f(u) - f(v)| \leqslant L|u - v|$, $u, v \in G$.

Доказательство. Возьмем $U = \operatorname{co} G$ — это выпуклая оболочка компактного множества G. В силу теоремы 1.8 U — выпуклое компактное множество, $U \subset W$. Тогда $f(u) - f(v) \geqslant \langle c(v), u - v\rangle \geqslant -L|u - v|$, $u, v \in U$, где $L = \sup\limits_{\partial f(U)} |c| < \infty$ в силу теоремы 5. Поменяв здесь u, v местами, имеем $f(v) - f(u) \geqslant -L|u - v|$, $u, v \in U$. Отсюда следует утверждение теоремы (см. упражнение 2.29). □

Теорема 7. *Если функция* $f(x)$ *выпукла и дифференцируема в каждой точке открытого выпуклого множества* $W \subseteq E^n$, *то ее градиент* $f'(x)$ *непрерывен на* W.

Доказательство. В начале параграфа мы установили, что субградиент выпуклой дифференцируемой функции совпадает с градиентом, так что $\partial f(u) = \{f'(u)\}$ $\forall u \in W$. Возьмем произвольную точку $v \in W$ и последовательность $\{v_k\} \in W$, $\{v_k\} \to v$. Множество $U = \{u = v_k,\ k = 1, 2, \ldots\} \cup \{v\}$ компактно и $U \subset W$. Тогда в силу теоремы 5 множество $\partial F(U) = \{c_k = f'(v_k),\ k = 1, 2, \ldots\}$ компактно. Пусть c — произвольная предельная точка множества $\partial F(U)$. Тогда существует подпоследовательность $c_{k_r} = f'(v_{k_r}) \to c$. Из замкнутости субдифференциального отображения следует, что $c \in \partial f(v) = \{f'(v)\}$, т. е. $c = f'(v)$. Это значит, что последовательность $\{f'(v_k)\}$ имеет единственную предельную точку, совпадающую с $f'(v)$. Отсюда и из произвольности точки v и последовательности $\{v_k\} \to v$ следует непрерывность градиента $f'(x)$ на W. □

Для получения интересных экстремальных свойств субдифференциального отображения нам понадобится

Лемма 1. *Пусть* W — *открытое множество из* E^n, *многозначные отображения* A *и* B: $W \to \Pi(E^n)$ *таковы, что* A *замкнуто и компактно,* B *монотонно, причем* $A(u) \cap B(u) \neq \varnothing$ *при всех* $u \in W$. *Тогда* $\overline{\operatorname{co}}\, B(u) \subseteq \overline{\operatorname{co}}\, A(u)$ $\forall u \in W$.

Доказательство. Зафиксируем любые $u \in W$ и $e \in E^n$, $|e| = 1$. Поскольку W — открытое множество, то $S = \{v \in E^n\colon |v - u| \leqslant \varepsilon_0\} \subset W$ при некотором $\varepsilon_0 > 0$. Возьмем последовательность $\{\varepsilon_k\} \to 0$, $0 < \varepsilon_k < \varepsilon_0$. Тогда $v_k = u + \varepsilon_k e \subset S \subset W$, $\{v_k\} \to u$. По условию существуют $a_k \in A(v_k) \cap B(v_k)$. В силу монотонности B для всех $b \in B(u)$ имеем $\langle a_k - b, v_k - u\rangle = \langle a_k - b, \varepsilon_k e\rangle \geqslant 0$ или $\langle a_k - b, e\rangle \geqslant 0$. Поскольку отображение A компактно, то множество $A(S)$ является компактным. Поэтому, учитывая включение $a_k \in A(v_k) \subset A(S)$, можем считать, что $\{a_k\} \to a_0$. В силу замкнутости отображения A тогда $a_0 \in A(u)$. Поэтому, переходя к пределу в неравенстве $\langle a_k - b, e\rangle \geqslant 0$, получаем $\langle a_0 - b, e\rangle \geqslant 0$, или $\langle e, b\rangle \leqslant \langle e, a_0\rangle \leqslant \sup\limits_{a \in A(U)} \langle e, a\rangle$ при всех $b \in B(u)$. Отсюда $\sup\limits_{b \in B(U)} \langle e, b\rangle \leqslant \sup\limits_{a \in A(U)} \langle e, a\rangle$, и требуемое утверждение следует из теоремы 5.6. □

Теорема 8. *Пусть* $f(x)$ — *выпуклая функция на открытом выпуклом множестве* W *из* E^n, *пусть* F: $W \to \Pi(E^n)$ — *какое-либо многозначное отображение. Тогда:*

а) если F *монотонно,* $\partial f(u) \subseteq F(u)$ *при всех* $u \in W$, *то* $\partial f(u) = F(u)$ $\forall u \in W$, *т. е. субдифференциальное отображение максимально в классе монотонных отображений;*

б) если F *замкнуто, выпуклозначно и* $F(u) \subseteq \partial f(u)$ *при всех* $u \in W$, *то* $\partial f(u) = F(u)$, *т. е. субдифференциальное отображение минимально в классе замкнутых выпуклозначных отображений.*

Доказательство. В случае а), пользуясь леммой 1 при $A(u) = \partial f(u)$, $B(u) = F(u)$, получаем $F(u) \subseteq \overline{\operatorname{co}}\, F(u) \subseteq \overline{\operatorname{co}}\, \partial f(u) = \partial f(u)$, так что $\partial f(u) = F(u)$ $\forall u \in W$. В случае б) в лемме 1 возьмем $A(u) = F(u)$, $B(u) = \partial f(u)$. Нужно проверить компактность отображения F. Пусть U — какой-либо компакт из W. Из включения $F(u) \subseteq \partial f(u)$ $\forall u \in W$ следует $F(U) \subseteq \partial f(U)$. В силу теоремы 5 множество $\partial f(U)$ компактно. Это значит, что его подмножество $F(U)$ ограничено.

Далее, пусть $\{c_k\} \to c$, $c_k \in F(U)$. Тогда найдутся такие точки $u_k \in U$, что $c_k \in F(u_k)$, $k = 1, 2, \ldots$. В силу компактности U можем считать, что $\{u_k\} \to u \in U$. Из замкнутости отображения F следует, что $c \in F(u) \subset F(U)$, т. е. $F(U)$ замкнуто. Компактность F установлена. В частности, взяв здесь одноточечный компакт $U = \{u\}$, заключаем, что $F(u)$ — компактное, т. е. ограниченное и замкнутое множество при каждом $u \in W$. Отсюда и из теоремы 1.6 с учетом выпуклости $F(u)$ имеем равенство $\operatorname{co} F(u) = F(u)$. Из леммы 1 теперь получаем $\operatorname{co} \partial f(u) = \partial f(u) \subseteq \operatorname{co} F(u) = F(u)$ $\forall u \in U$. Отсюда и из включения $F(u) \subseteq \partial f(u)$ $\forall u \in W$ следует утверждение б) теоремы. □

7. Субдифференциал для выпуклых функций играет роль, аналогичную той, какую играет градиент для дифференцируемых функций. Как для работы с градиентами полезно иметь некоторый набор правил дифференцирования, так и для работы с субдифференциалами нужно иметь некоторые правила субдифференцирования. Предлагаем читателю самостоятельно доказать следующие правила субдифференцирования 1–4:

1. Если $g(u) = f(u + u_0)$, то $\partial g(u) = \partial f(u + u_0)$.
2. Если $g(u) = \lambda f(u)$, $\lambda > 0$, то $\partial g(u) = \lambda \partial f(u)$.
3. Если $g(u) = f(\lambda u)$, то $\partial g(u) = \lambda \partial f(\lambda u)$.
4. Если функция $f(u)$ выпукла на E^n, а A — матрица порядков $n \times n$, $b \in E^n$, то функция

$$g(u) = f(Au + b), \qquad u \in E^n,$$

выпукла на E^n, причем

$$\partial g(u) = A^{\top} \partial f(v)\Big|_{v = Au + b}.$$

5. Справедлива следующая теорема [670; 234], обобщающая известную теорему о производной сложной функции.

Теорема 9. *Пусть* $f_1(u), \ldots, f_m(u)$ — *выпуклые функции, определенные на открытом выпуклом множестве* W *из* E^n, *функция* $\varphi(x) = \varphi(x^1, \ldots, x^m)$ — *выпуклая функция на открытом выпуклом множестве* X *из* E^m, *причем* $f(u) = (f_1(u), \ldots, f_m(u)) \in X$ *при всех* $u \in W$, $\varphi(x)$ *монотонно возрастает на* X, *т. е.* $\varphi(x) \geqslant \varphi(y)$ *для всех* $x = (x^1, \ldots, x^m)$, $y = (y^1, \ldots, y^m) \in X$, $x^i \geqslant y^i$, $i = 1, \ldots, m$. *Тогда функция* $\Phi(u) = \varphi(f(u))$ *выпукла на* W *и ее субдифференциал имеет вид*

$$\partial \Phi(u) = \bigcup_{p = (p_1, \ldots, p_m) \in \partial \varphi(f(u))} \Big\{ \sum_{i=1}^{m} p_i \partial f_i(u) \Big\}, \quad u \in W. \tag{9}$$

Для доказательства этой теоремы нам понадобится

Лемма 2. *Пусть* $A_1, \ldots, A_m$ — *выпуклые множества из* E^n, P — *выпуклое множество из* E^n_+, *тогда множество* $A = \bigcup\limits_{p = (p_1, \ldots, p_m) \in P} \Big\{ \sum\limits_{i=1}^{m} p_i A_i \Big\}$ *выпукло.*

Доказательство. Возьмем произвольные $c_1, c_2 \in A$, $\alpha \in (0, 1)$. По определению A существуют такие $p_i = (p_{i1}, \ldots, p_{im}) \in P \subset E^n_+$, $a_{ij} \in A_j$, $j = 1, \ldots, m$, что $c_i = \sum\limits_{j=1}^{m} p_{ij} a_{ij}$, $i = 1, 2$. Тогда $\alpha c_1 + (1 - \alpha) c_2 = \sum\limits_{j=1}^{m} (\alpha p_{1j} a_{1j} + (1 - \alpha) p_{2j} a_{2j})$. По условию $p_{ij} \geqslant 0$. Обозначим через I множество всех номеров $j = 1, \ldots, m$, для которых $p_{1j} > 0$ или $p_{2j} > 0$. Тогда $\alpha p_{1j} + (1 - \alpha) p_{2j} > 0$, $\gamma_{\alpha j} = \dfrac{\alpha p_{1j}}{\alpha p_{1j} + (1 - \alpha) p_{2j}} \in (0, 1)$, $j \in I$.

Положим $a^{\alpha}_j = \gamma_{\alpha j} a_{1j} + (1 - \gamma_{\alpha j}) a_{2j}$ при $j \in I$, $a^{\alpha}_j = a_{1j}$ при $j \neq I$. В силу выпуклости A_j точки a^{α}_j принадлежат A_j, $j = 1, \ldots, m$. Кроме того, $p^{\alpha} = (p^{\alpha}_1, \ldots, p^{\alpha}_m) = \alpha p_1 + (1 - \alpha) p_2 \in P$ из-за выпуклости P, причем здесь $p^{\alpha}_j = 0$ при $j \notin I$. Тогда $\alpha c_1 + (1 - \alpha) c_2 = \sum\limits_{j \in I} (\alpha p_{1j} a_{1j} + (1 - \alpha) p_{2j} a_{2j}) = \sum\limits_{j \in I} (\alpha p_{1j} + (1 - \alpha) p_{2j})(\gamma_{\alpha j} a_{1j} + (1 - \gamma_{\alpha j}) a_{2j}) = \sum\limits_{j \in I} p^{\alpha}_j a^{\alpha}_j = \sum\limits_{j=1}^{m} p^{\alpha}_j a^{\alpha}_j$, где $a^{\alpha}_j \in A_j$, $j = 1, \ldots, m$, $p^{\alpha} \in P$. Значит, $\alpha c_1 + (1 - \alpha) c_2 \in A$ при всех $\alpha \in (0, 1)$, т. е. A — выпуклое множество. □

Доказательство теоремы 9. Из выпуклости функций $f_i(u)$, $\varphi(x)$ и монотонности $\varphi(x)$ следует выпуклость сложной функции $\Phi(u) = \varphi(f(u))$ на открытом выпуклом множестве W — это доказывается так же, как и теорема 2.8. Согласно теореме 2 тогда субдифференциал $\partial \Phi(u)$ при каждом $u \in W$ представляет собой непустое выпуклое компактное множество.

Докажем формулу (11). Обозначим $F(u) = \bigcup\limits_{p \in \partial \varphi(f(u))} \Big\{ \sum\limits_{i=1}^{m} p_i \partial f_i(u) \Big\}$. По теореме 2 субдифференциалы $\partial f_i(u)$, $\partial \varphi(x)$ также непусты, выпуклы, компактны и поэтому $F(u) \neq \varnothing$ $\forall u \in W$. Отметим, что $\partial \varphi(x) \in E^m_+$ при всех $x \in X$. В самом деле, возьмем любые $x = (x^1, \ldots, x^m) \in X$, $p = (p_1, \ldots, p_m) \in \partial \varphi(x)$. Поскольку множество X открыто, то при достаточно малом $\varepsilon > 0$ точка $y = (y^1, \ldots, y^m)$, где $y^i = x^i - \varepsilon$, $y^j = x^j$ при $j \neq i$, принадлежит X. С учетом монотонности $\varphi(x)$ тогда $0 \geqslant \varphi(y) - \varphi(x) \geqslant \langle p, y - x\rangle = p_i(-\varepsilon)$, так что $p_i \geqslant 0$, $i = 1, \ldots, m$. Следовательно, $\partial \varphi(x) \in E^m_+$. По лемме 2 тогда множество $F(u)$ выпукло при каждом $u \in W$.

Покажем, что $F = F(u)$, как многозначное отображение $W \to \Pi(E^n)$, замкнуто. Пусть $u \in W$, $\{u_k\} \to u$, $\{c_k\} \to c$, $c_k \in F(u_k)$. Тогда найдутся $p_k \in \partial \varphi(f(u_k))$, $c_{ik} \in \partial f_i(u_k)$ такие, что $c_k = \sum\limits_{i=1}^{m} p_{ik} c_{ik}$. Поскольку сходящаяся последовательность $\{u_k\}$ ограничена, то найдется

компактное множество $G \subset W$, содержащее все точки $u, u_1, u_2, \ldots$ Аналогично, поскольку в силу непрерывности выпуклых функций $f_i(u)$ последовательность $\{x_k = f(u_k)\} \to f(u) \in X$, то существует компактное множество $Y \subset X$, содержащее все точки $f(u), f(u_1), f(u_2), \ldots$ (можно взять $Y = f(G) = f_1(G) \times \ldots \times f_m(G)$). По теореме 5 множества $\partial f_i(G)$, $\partial\varphi(Y)$ компактны. Поскольку $c_{ik} \subset \partial f_i(u_k) \subset \partial f_i(G)$, $p_k \in \partial\varphi(f(u_k)) \in \partial\varphi(Y)$, $k = 1, 2, \ldots$, то не теряя общности можем считать, что $\{c_{ik}\} \to c_i$, $\{p_k\} \to p$. Из замкнутости отображений $\partial f_i(u)$, $\partial\varphi(x)$ имеем $c_i \in \partial f_i(u)$, $p \in \partial\varphi(f(u))$. Переходя к пределу в равенстве $c_k = \sum_{i=1}^{m} p_{ik} c_{ik}$, получаем $c = \sum_{i=1}^{m} p_i c_i$, т. е. $c \in F(u)$. Это значит, что отображение F замкнуто.

Возьмем любые $u \in W$ и $c \in F(u)$. Тогда $c = \sum_{i=1}^{m} p_i c_i$ при некоторых $c_i \in \partial f_i(u)$, $p = (p_1, \ldots, p_m) \in \partial\varphi(f(u))$. Учитывая определение субградиента и неотрицательность p_i, получим

$$\Phi(v) - \Phi(u) = \varphi(f(v)) - \varphi(f(u)) \geqslant \langle p, f(v) - f(u)\rangle = \sum_{i=1}^{m} p_i (f_i(v) - f_i(u)) \geqslant \sum_{i=1}^{m} p_i \langle c_i, v - u\rangle = \langle \sum_{i=1}^{m} p_i c_i, v - u\rangle = \langle c, v - u\rangle \quad \forall v \in W.$$

Это значит, что $c \in \partial\Phi(u)$ и, следовательно, $F(u) \subseteq \partial\Phi(u)$ при всех $u \in W$. Отсюда, пользуясь утверждением б) теоремы 8, заключаем, что $\partial\Phi(u) = F(u)$ $\forall u \in W$. Формула (9) доказана. □

С помощью теоремы 9 можно получить более сложные правила субдифференцирования, дополняющие приведенные выше правила 1–4. Ниже при ссылках на формулу (9) предполагается, что выполнены условия теоремы 9.

6. Если $\varphi(x)$ — дифференцируемая функция, то $\partial\varphi(x) = \{\varphi'(x)\} = \{(\partial\varphi/\partial x^1, \ldots, \partial\varphi/\partial x^m)\}$, $\partial\varphi(f(u)) = \{\varphi'(f(u))\}$, и из формулы (9) имеем

$$\partial\Phi(u) = \sum_{i=1}^{m} \frac{\partial\varphi(f(u))}{\partial x^i} \partial f_i(u), \quad u \in W.$$

В частности, если $f_i(u)$ дифференцируема и $\partial f_i(u) = \{f_i'(u)\}$, отсюда получаем классическое правило дифференцирования сложной функции.

7. Если $\varphi(x) = \sum_{i=1}^{m} \alpha_i x^i$, $\alpha_i \geqslant 0$, то $\partial\varphi(x) = \{(\alpha_1, \ldots, \alpha_m)\} \in E_+^m$ и для функции $\Phi(u) = \sum_{i=1}^{m} \alpha_i f_i(u)$, $u \in W$, из (11) имеем $\partial\Phi(u) = \sum_{i=1}^{m} \alpha_i \partial f_i(u)$ $\forall u \in W$.

8. Если $\varphi(x) = \max\limits_{1 \leqslant i \leqslant m} x^i$, то согласно формуле (5) $\partial\varphi(x) = \{p = (p_1, \ldots, p_m)\colon p_i \geqslant 0, i \in I(x); p_i = 0, i \notin I(x), p_1 + \ldots + p_m = 1\}$, где $I(x) = \{i\colon 1 \leqslant i \leqslant m, \max\limits_{1 \leqslant j \leqslant m} x^j = x^i\}$, $x \in E^m$. Отсюда и из (9) для функции $\Phi(u) = \max\limits_{1 \leqslant i \leqslant m} f_i(u)$ имеем

$$\partial\Phi(u) = \{c\colon c = \sum_{i \in I(f(u))} p_i c_i,\ c_i \in \partial f_i(u),\ p_i \geqslant 0,\ i \in I(f(u)),\ \sum_{i \in I(f(u))} p_i = 1\} = \operatorname{co}\Big(\bigcup_{i \in I(f(u))} \partial f_i(u)\Big), \quad I(f(u)) = \{i\colon 1 \leqslant i \leqslant m, \max_{1 \leqslant j \leqslant m} f_j(u) = f_i(u)\}, \quad u \in W. \quad (10)$$

9. Если $\varphi(x) = \max\{0; x\}$, $x \in E^1$, то согласно (10) $\partial\varphi(0) = \{c\colon c = p_1 \cdot 0 + p_2 \cdot 1 = p_2, p_1 + p_2 = 1, p_1 \geqslant 0, p_2 \geqslant 0\} = [0, 1]$, $\partial\varphi(x) = \{1\}$ при $x > 0$, $\partial\varphi(x = \{0\}$ при $x < 0$, и для функции $\Phi(u) = \max\{0; f(u)\}$, $u \in W$ из (9) имеем $\partial\Phi(u) = p\partial f(u)$, $0 \leqslant p \leqslant 1$, при $f(u) = 0$, $\partial\Phi(u) = \partial f(u)$ при $f(u) > 0$, $\partial\Phi(u) = 0$ при $f(u) < 0$.

10. Если $\varphi(x) = (\max\{0; x\})^p$, $p > 1$, $x \in E^1$, то $\partial\varphi(x) = \{\varphi'(x) = p(\max\{0; x\})^{p-1}\}$ и для функции $\Phi(u) = (\max\{0; f(u)\})^p$, $u \in W$, имеем

$$\partial\Phi(u) = p(\max\{0; f(u)\})^{p-1} \partial f(u), \quad u \in W, \quad p > 1.$$

11. Приведем еще одну теорему, в которой дается обобщение формулы (10).

Теорема 10. *Пусть A — компактное множество из E^m, W — открытое выпуклое множество из E^n, функция $G(u, a)$ определена на $W \times A$, полунепрерывна сверху по a при каждом $u \in W$, выпукла по переменной $u \in W$ при каждом $a \in A$. Тогда функция $\Phi(u) = \max\limits_{a \in A} G(u, a)$ выпукла на W и ее субдифференциал имеет вид*

$$\partial\Phi(u) = \operatorname{co}\Big(\overline{\bigcup_{a \in R(u)} \partial G(u, a)}\Big), \quad R(u) = \{a\colon a \in A, G(u, a) = \Phi(u)\}, \quad u \in W. \quad (11)$$

Доказательство может быть проведено по той же схеме, как и теорема 9, и представляется читателю.

12. Если A — выпуклое замкнутое ограниченное множество из E^n, то функция $\Phi(u) = \max\limits_{a \in A} \langle a, u\rangle$, $u \in E^n$, выпукла на E^n, причем, как следует из (11) при $G(u, a) = \langle a, u\rangle$, имеем $\partial\Phi(u) = \{a\colon a \in A, \langle a, u\rangle = \Phi(u)\}$.

Более подробно о перечисленных и других правилах субдифференцирования, о различных свойствах субдифференциала, о различных обобщениях понятий субградиента и субдифференциала, о применении этих понятий для исследования и приближенного решения экстремальных задач см., например [234; 251; 263; 264; 302; 314; 358; 434; 495; 499; 502; 542; 543; 604; 605; 613; 617; 670; 720; 795; 814].

Упражнения

1. Найти субдифференциалы функций:
а) $f(u) = |u - 1|$, $u \in E^1$;
б) $f(u) = |u - 1| + |u + 1|$, $u \in E^1$;
в) $f(u) = |x + y| + |x - y|$, $u = (x, y) \in E^2$;
г) $f(u) = \max\{u^2, u + 2\}$, $u \in E^1$;
д) $f(u) = \max\{|u|; |u - 1|\}$, $u \in E^1$;
е) $f(u) = |\langle a, u\rangle - b|$, $u \in E^n$;

2. Пусть функции $f_1(u), \ldots, f_m(u)$, $u \in E^n$, непрерывно дифференцируемы в некоторой окрестности точки v. Доказать, что тогда функция $f(u) = \max\limits_{1 \leqslant i \leqslant m} f_i(u)$ в точке v имеет производные по любому направлению e, $|e| = 1$, причем

$$\frac{df(v)}{de} = \max_{i \in I(v)} \langle f_i'(v), e\rangle, \quad I(v) = \{i\colon 1 \leqslant i \leqslant m,\ f_i(v) = f(v)\}.$$

Установить связь между этой формулой и формулами (7), (10).

3. Найти субдифференциалы функций $f(u) = \max\limits_{|t| \leqslant 1} |t^2 + xt + y|$, $f(u) = \max\limits_{|t| \leqslant 1} |xt^2 + yt|$, $f(u) = \max\limits_{0 \leqslant t \leqslant 1} |x + ty|$, $u = (x, y) \in E^2$.

4. Пусть A — замкнутое ограниченное множество из E^m, функция $g(u, a)$ непрерывна по совокупности переменных (u, a) на $E^n \times A$ вместе с производной $\partial g(u, a)/\partial u$. Доказать, что тогда функция $f(u) = \max\limits_{a \in A} g(u, a)$ во всех точках $v \in E^n$ имеет производную по любому направлению e, $|e| = 1$, причем

$$\frac{df(v)}{de} = \max_{a \in A_0(v)} \langle \frac{\partial g(v, a)}{\partial u}, e\rangle, \quad A_0(v) = \{a\colon a \in A, g(v, a) = f(v)\}.$$

Установить связь между этой формулой и формулами (7), (11).

5. Пусть $f(u)$ — выпуклая функция одной переменной на отрезке $[a, b]$. Доказать, что $\partial f(u) = [f'(u - 0), f'(u + 0)]$ при всех $u \in (a, b)$, где $f'(u - 0)$, $f'(u + 0)$ — левая и правая производные в точке u. Показать, что в точках $u = a$ или $u = b$ субдифференциал может быть пустым (рассмотреть пример $f(u) = -\sqrt{1 - u^2}$, $|u| \leqslant 1$).

6. Пусть $f(u)$ — выпуклая функция на выпуклом множестве X из E^n. Доказать, что при всех $u \in \operatorname{ri} X$ множество $\partial f(u)$ непусто, выпукло, компактно, причем $\frac{df(u)}{de} = \max\limits_{c \in \partial f(u)} \langle c, e\rangle$ для всех $e \in \operatorname{Lin} X$.

7. Пусть X — выпуклое множество, функция $f(x)$ выпукла на X. Доказать, что $\partial f(x) \cap \operatorname{Lin} X \neq \varnothing$ при $\forall x \in \operatorname{ri} X$, т. е. субградиент всегда можно выбрать в $\operatorname{Lin} X$.

8. Пусть функция $f(u)$ определена на открытом выпуклом множестве $W \subset E^n$ и такова, что функция $\Phi(u) = |f(u)|$ выпукла на W. Описать множество $\partial\Phi(u)$, $u \in W$.

9. Описать субдифференциалы функций $\rho(u, X)$, $\delta(c, X)$, $\mu(u, X)$ из упражнений 18–20 к § 4.2.

10. Пусть $f(u)$ — выпуклая функция на открытом выпуклом множестве W из E^n, пусть субдифференциал $\partial f(u)$ в некоторой точке $u \in W$ состоит из единственного элемента c. Доказать, что $f(u)$ дифференцируема в точке u, причем $f'(u) = c$.

11. Пусть $f(u)$, $G(u)$ — выпуклые функции на открытом выпуклом множестве W из E^n, причем $\partial f(u) = \partial G(u)$ при всех $u \in W$. Доказать, что тогда $f(u) = G(u) + \text{const}$, $u \in W$.

12. Пусть функция $f(u)$ выпукла на открытом выпуклом множестве W из E^n. Доказать, что для того чтобы $f(u)$ была сильно выпуклой на W, необходимо и достаточно, чтобы для каждой точки $v \in W$ существовал субградиент $c(v) \in \partial f(v)$ такой, что

$$f(u) - f(v) \geqslant \langle c(v), u - v\rangle + \frac{1}{2}\varkappa|u - v|^2 \quad \forall u \in W, \quad \varkappa = \text{const} > 0.$$

13. Пусть функция $f(u)$ сильно выпукла на открытом выпуклом множестве W из E^n с постоянной сильной выпуклости $\varkappa$. Доказать:

а) $\langle c(u) - c(v), u - v\rangle \geqslant \varkappa|u - v|^2$ для всех $\forall u, v \in W$, $c(u) \in \partial f(u)$, $c(v) \in \partial f(v)$;

б) $\partial f(u) \cap \partial f(v) = \varnothing$ для всех $u, v \in W$, $u \neq v$;

в) для любой точки $v \in W$ справедливо неравенство $|u - v| \leqslant \frac{1}{\varkappa} \min\limits_{c \in \partial f(v)} |c|$ для всех $u \in M(v) = \{u \in W: f(u) \leqslant f(v)\}$.

Опираясь на это утверждение, доказать теорему 3.1 для любого выпуклого замкнутого множества $X \subseteq W$.

14. Пусть функция $f(u)$ выпукла на открытом выпуклом множестве $W \subset E^n$ и сильно выпукла на выпуклом замкнутом подмножестве $X \subset W$. Доказать, что тогда

$$0 \leqslant f(u) - f_* \leqslant \frac{1}{4\varkappa} \min_{c \in \partial f(v)} |c|^2, \quad |u - u_*| \leqslant \frac{1}{2\varkappa} \min_{c \in \partial f(v)} |c|,$$

где u_* — точка минимума $f(u)$ на X, $f_* = f(u_*)$.

15. Пусть функция $f(u)$ сильно выпукла на E^n. Доказать, что для любого $c \in E^n$ существует такая единственная точка $u(c) \in E^n$, что $c \in \partial f(u(c))$.

Указание: рассмотреть точку минимума функции $g(u) = f(u) - \langle c, u\rangle$ на E^n.

16. Доказать, что оператор проектирования на выпуклое замкнутое множество является монотонным, замкнутым, компактным отображением. Указание: воспользоваться теоремами 4.4.2, 4.4.3, неравенством (4.4.4).

§ 7. Равномерно выпуклые функции

1. Рассмотренный в § 3 класс сильно выпуклых функций обладает замечательным свойством — для функций этого класса имеет место теорема 3.1. Однако этот подкласс выпуклых функций недостаточно широк и не содержит, например, такую выпуклую функцию, как $f(x) = x^4$, $x \in E^1$, которая, между прочим, достигает своей нижней грани на любом выпуклом замкнутом множестве из E^1, причем в единственной точке. Хотелось бы выделить такой подкласс выпуклых функций, для которого была бы верна теорема типа теоремы 3.1 и который был бы шире класса сильно выпуклых функций. Оказывается, таким классом является класс равномерно выпуклых функций.

Определение 1. Функцию $f(x)$, определенную на выпуклом множестве X, называют *равномерно выпуклой* на X, если существует неотрицательная функция $\delta(t)$, определенная при всех t, $0 \leqslant t \leqslant \operatorname{diam} X = \sup\limits_{u, v \in X} |u - v|$, $\delta(0) = 0$, $\delta(t_0) > 0$ при некотором t_0, $0 < t_0 < \operatorname{diam} X$, и такая, что

$$f(\alpha u + (1 - \alpha)v) \leqslant \alpha f(u) + (1 - \alpha)f(v) - \alpha(1 - \alpha)\delta(|u - v|) \qquad (1)$$

при всех $u, v \in X$, $\alpha \in [0, 1]$. Функцию $\delta(t)$ называют *модулем выпуклости* функции $f(u)$ на X, а функцию

$$\mu(t) = \inf_{0 < \alpha < 1} \inf_{|u - v| = t,\, u, v \in X} \frac{\alpha f(u) + (1 - \alpha)f(v) - f(\alpha u + (1 - \alpha)v)}{\alpha(1 - \alpha)}$$

точным модулем выпуклости $f(x)$ на X. Если равномерно выпуклая функция имеет модуль выпуклости $\delta(t) > 0$ при всех t, $0 < t < \operatorname{diam} X$, то такую функцию называют *строго равномерно выпуклой* на X [191].

Очевидно, всякая сильно выпуклая функция является строго равномерно выпуклой с модулем $\delta(t) = \frac{1}{2}\varkappa t^2$. Сумма равномерно выпуклой функции с модулем $\delta(t)$ и выпуклой функции будет равномерно выпуклой с модулем $\delta(t)$. Если $f(x)$ равномерно выпукла с модулем $\delta(t)$, то функция $g(x) = cf(x)$ при любом $c = \text{const} > 0$ также будет равномерно выпуклой с модулем $c\delta(t)$. Если $\mu(t)$ — точный модуль выпуклости равномерно выпуклой функции $f(x)$ на X, то любая функция $\delta(t) \leqslant \mu(t)$, $0 \leqslant t \leqslant \operatorname{diam} X$, неотрицательная, нетождественно равная нулю, $\delta(t) = 0$, будет модулем выпуклости функции $f(x)$ на X.

Следующая теорема является обобщением теоремы 3.1.

Теорема 1. *Пусть X — выпуклое замкнутое множество из E^n (например, $X = E^n$), а функция $f(x)$ равномерно выпукла и полунепрерывна снизу на X. Тогда:*

1) множество Лебега $M(v) = \{u: u \in X, f(u) \leqslant f(v)\}$ выпукло, замкнуто и ограничено при всех $v \in X$;

2) $f_* = \inf\limits_X f(u) > -\infty$, $X_* = \{u: u \in X, f(u) = f_*\} \neq \varnothing$;

3) имеет место неравенство

$$\delta(|u - u_*|) \leqslant f(u) - f(u_*) \qquad (2)$$

при всех $u \in X$, $u_* \in X_*$;

4) если, кроме того, $f(x)$ строго равномерно выпукла на X, то X_* состоит из единственной точки u_* и всякая минимизирующая последовательность $\{u_k\}$: $\{u_k\} \in X$, $\lim\limits_{k \to \infty} f(u_k) = f_*$, сходится к точке u_*.

Для доказательства этой теоремы нам понадобятся следующие две леммы о свойствах точного модуля выпуклости.

Лемма 1. *Пусть $\mu(t)$ — точный модуль выпуклости равномерно выпуклой функции $f(x)$ на выпуклом множестве X. Тогда*

$$\mu(ct) \geqslant c^2\mu(t) \qquad (3)$$

для всех $c \geqslant 1$, $t \geqslant 0$, $0 \leqslant ct \leqslant \operatorname{diam} X$.

Доказательство. Сначала рассмотрим случай $1 \leqslant c < 2$. По определению $\mu(ct)$ для любого $\varepsilon > 0$ существуют точки $u_1, u_2 \in X$ и число α, $0 < \alpha < 1$, такие, что $|u_1 - u_2| = ct$ и

$$\mu(ct) \leqslant \frac{\alpha f(u_1) + (1 - \alpha)f(u_2) - f(u_\alpha)}{\alpha(1 - \alpha)} \leqslant \mu(ct) + \varepsilon,$$

где $u_\alpha = \alpha u_1 + (1 - \alpha)u_2$. Отсюда имеем

$$\alpha f(u_1) + (1 - \alpha)f(u_2) - f(u_\alpha) \leqslant \alpha(1 - \alpha)\mu(ct) + \alpha(1 - \alpha)\varepsilon. \qquad (4)$$

Можем считать, что $0 < \alpha \leqslant 1/2$, так как в противном случае в (4) точки u_1 и u_2 можно поменять ролями. Тогда с учетом $1 \leqslant c < 2$ можем сказать, что $0 < \alpha \leqslant \alpha c < 1$. Кроме того, $1/2 < 1/c \leqslant 1$, поэтому $u_3 = (1/c)u_1 + (1 - 1/c)u_2 \in X$, причем $|u_3 - u_2| = |u_1 - u_2|/c = t$. Заметим также, что $u_\alpha = \alpha c u_3 + (1 - \alpha c)u_2$. Тогда

$$f(u_3) \leqslant (1/c)f(u_1) + (1 - 1/c)f(u_2) - (1/c)(1 - 1/c)\mu(ct),$$
$$f(u_\alpha) \leqslant \alpha c f(u_3) + (1 - \alpha c)f(u_2) - \alpha c(1 - \alpha c)\mu(t).$$

Умножим первое из этих неравенств на αc и сложим со вторым. Учитывая неравенство (4), получаем

$$\alpha c(1/c)(1 - 1/c)\mu(ct) + \alpha c(1 - \alpha c)\mu(t) \leqslant$$
$$\leqslant \alpha f(u_1) + (\alpha c - \alpha)f(u_2) - (1 - \alpha c)f(u_2) - f(u_\alpha) =$$
$$= \alpha f(u_1) + (1 - \alpha)f(u_2) - f(u_\alpha) \leqslant \alpha(1 - \alpha)\mu(ct) + \alpha(1 - \alpha)\varepsilon$$

или

$$\alpha(1-1/c)\mu(ct)+\alpha c(1-\alpha c)\mu(t)\leqslant\alpha(1-\alpha)\mu(ct)+\alpha(1-\alpha)\varepsilon.$$

Поскольку здесь $\varepsilon>0$ — произвольное число, то можем ε устремить к $+0$. Будем иметь.

$$\alpha c(1-\alpha c)\mu(t)\leqslant(\alpha/c)(1-\alpha c)\mu(ct) \quad \text{или} \quad c^2\mu(t)\leqslant\mu(ct).$$

Неравенство (3) для случая $1\leqslant c<2$ доказано. Пусть теперь $c\geqslant 1$ — произвольное число, $0\leqslant ct\leqslant\operatorname{diam}X$. Поскольку $\lim\limits_{n\to\infty}\sqrt[n]{c}=1$, то найдется b $(1\leqslant b<2)$ такое, что $c=b^n$ при некотором натуральном $n\geqslant 1$. Учитывая, что по доказанному $\mu(bt)\geqslant b^2\mu(t)$, получаем $\mu(ct)=$ $=\mu(b^nt)=\mu(b\cdot b^{n-1}t)\geqslant b^2\mu(b^{n-1}t)\geqslant b^4\mu(b^{n-2}t)\geqslant\ldots\geqslant b^{2n}\mu(t)=c^2\mu(t)$. □

Лемма 2. *Пусть $\mu(t)$ — точный модуль выпуклости равномерно выпуклой функции $f(x)$ на выпуклом множестве X. Тогда:*

1) $\mu(t)=O(t^2)$ при $t\to+0$;

2) $\mu(t)\equiv 0$ при $0\leqslant t<\tau=\inf\{t\colon\mu(t)>0\}$, $\mu(t)$ строго монотонно растет при $\tau<t\leqslant$ $\leqslant\operatorname{diam}X$;

3) если $\operatorname{diam}X=\infty$, то $\lim\limits_{t\to\infty}\mu(t)=\infty$.

Доказательство. Из определения 1 следует, что $\mu(t_0)>0$ при некотором t_0, $0<t_0<$ $<\operatorname{diam}X$. Поэтому $0\leqslant\tau<\operatorname{diam}X$. Если $\tau>0$, то $\mu(t)\equiv 0$ при $0\leqslant t<\tau$ по определению τ. Пусть $\tau<t<a<\operatorname{diam}X$. Тогда с помощью неравенства (3) имеем $\mu(a)=\mu((a/t)t)\geqslant$ $\geqslant(a/t)^2\mu(t)>\mu(t)>0$, т. е. $\mu(t)$ строго монотонна при $\tau<t\leqslant\operatorname{diam}X$.

Далее, если $\tau>0$, то условие $\mu(t)=O(t^2)$ при $t\to+0$ выполняется тривиально. Поэтому пусть $\tau=0$. Тогда, фиксируя какое-либо t_0, $0<t_0\leqslant\operatorname{diam}X$, для всех $0<t<t_0$ имеем $\mu(t_0)=$ $=\mu((t_0/t)t)\geqslant(t_0/t)^2\mu(t)$ или $\mu(t)\leqslant\mu(t_0)t^2/t_0^2=\text{const}\cdot t^2$. Это и означает, что $\mu(t)=O(t^2)$ при $t\to+0$.

Наконец, пусть $\operatorname{diam}X=\infty$. Тогда $\mu(t)$ определена при всех $t\geqslant 0$. Пусть $t\geqslant t_0>\tau$. Тогда $\mu(t)=\mu((t/t_0)t_0)\geqslant(t/t_0)^2\mu(t_0)=\text{const}\cdot t^2$. Это значит, что $\mu(t)\to\infty$ при $t\to\infty$ со скоростью не медленнее, чем t^2. □

Заметим, что из неравенства $0\leqslant\delta(t)\leqslant\mu(t)$, справедливого для любого модуля выпуклости равномерно выпуклой функции, и из леммы 2 следует, что условие $\delta(t)=O(t^2)$ при $t\to$ $\to+0$ является необходимым для того, чтобы некоторая функция $\delta(t)$ могла служить модулем выпуклости для какой-либо равномерно выпуклой функции.

Доказательство теоремы 1. Если множество X ограничено, замкнуто, т. е. X компактно, то утверждения 1), 2) теоремы следуют из теорем 2.1.1, 2.10, леммы 2.1.1. Остается рассмотреть случай, когда X — неограниченное множество. Тогда $\operatorname{diam}X=\infty$ и точный модуль выпуклости $\mu(t)$ функции $f(x)$ будет определен при всех $t\geqslant 0$.

Пусть $t_0>0$ и $\mu(t_0)>0$. Возьмем произвольную точку $v\in X$ и рассмотрим шар

$$S=S(v,t_0)=\{u\colon u\in X,\ |u-v|\leqslant t_0\}.$$

Из теоремы 2.1.1 следует, что $\inf\limits_S f(x)=f_S^*>-\infty$, так что

$$f(u)\geqslant f_S^*=f(v)-\nu,\quad \nu=f(v)-f_S^*\geqslant 0, \tag{5}$$

при всех $u\in S$. Возьмем произвольную точку $u\in X\setminus S$, т. е. $|u-v|>t_0$. Тогда, учитывая доказанную в лемме 2 строгую монотонность $\mu(t)$ при $t>\tau$, имеем

$$0<\alpha_0=(\mu(t_0)/\mu(|u-v|))^{1/2}<1. \tag{6}$$

При $\alpha=\alpha_0$ из (1) получаем

$$\alpha_0 f(u)\geqslant f(v+\alpha_0(u-v))-(1-\alpha_0)f(v)+\alpha_0(1-\alpha_0)\mu(|u-v|). \tag{7}$$

Из (6) и леммы 1 следует $\mu(t_0)=\alpha_0^2\mu(|u-v|)=\alpha_0^2\mu((1/\alpha_0)\alpha_0|u-v|)\geqslant\mu(\alpha_0|u-v|)$ или $\mu(t_0)\geqslant\mu(\alpha_0|u-v|)$. В силу монотонности $\mu(t)$ это означает, что $\alpha_0|u-v|\leqslant t_0$. Тогда $v+$ $+\alpha_0(u-v)\in S$ и согласно (5) $f(v+\alpha_0(u-v))\geqslant f(v)-\nu$. Учитывая эту оценку, из (7) имеем

$$\alpha_0 f(u)\geqslant\alpha_0 f(v)-\nu+\alpha_0(1-\alpha_0)\mu(|u-v|).$$

Отсюда, сокращая на $\alpha_0>0$ и вспоминая определение (6) величины α_0, получаем

$$f(u)\geqslant f(v)+(1-\alpha_0)\mu(|u-v|)-\nu/\alpha_0=f(v)+\mu(|u-v|)-\sqrt{\mu(|u-v|)}\left(\sqrt{\mu(t_0)}+\nu/\sqrt{\mu(t_0)}\right).$$

Применяя к последнему слагаемому неравенство $ab\leqslant(a^2+b^2)/2$, будем иметь

$$f(u)\geqslant f(v)+\mu(|u-v|)/2-\left(\sqrt{\mu(t_0)}+\nu/\sqrt{\mu(t_0)}\right)^2/2 \tag{8}$$

для всех $u\in X\setminus S$. На самом деле, неравенство (8) имеет место для всех $u\in X$. Действительно, если $u\in S$, то $\mu(|u-v|)\leqslant\mu(t_0)$, а тогда $\nu\leqslant\left(\sqrt{\mu(t_0)}+\nu/\sqrt{\mu(t_0)}\right)^2/2-\mu(|u-v|)/2$. Отсюда и из (5) следует справедливость (8) и для $u\in S$.

Для всех $u\in M(v)$ из (8) имеем $\mu(|u-v|)/2-\left(\sqrt{\mu(t_0)}+\nu/\sqrt{\mu(t_0)}\right)^2/2\leqslant f(u)-f(v)\leqslant 0$, т. е. $\mu(|u-v|)\leqslant\left(\sqrt{\mu(t_0)}+\nu/\sqrt{\mu(t_0)}\right)^2$ при любом $u\in M(v)$. Поскольку $\mu(t)\to\infty$ при $t\to\infty$ и только в этом случае, то из последнего неравенства следует ограниченность множества $M(v)$. Выпуклость $M(v)$ следует из теоремы 2.10, а замкнутость $M(v)$ — из леммы 2.1.1. Из теоремы 2.1.2 имеем, что $f_*>-\infty$, $X_*\neq\varnothing$.

Докажем неравенство (2). Поскольку любой модуль выпуклости $\delta(t)\leqslant\mu(t)$, то неравенство (2) достаточно доказать для $\mu(t)$. Возьмем любую точку $u_*\in X_*$. Тогда $0\leqslant f(\alpha u+(1-$ $-\alpha)u_*)-f(u_*)\leqslant\alpha(f(u)-f(u_*))-\alpha(1-\alpha)\mu(|u-v|)$ или $\alpha(1-\alpha)\mu(|u-u_*|)\leqslant\alpha(f(u)-f(u_*))$, $0<\alpha<1$, $u\in X$. Деля на $\alpha>0$ и устремляя $\alpha\to+0$, отсюда получаем неравенство (2).

Наконец, пусть функция $f(x)$ строго равномерно выпукла на X. Тогда она строго выпукла на X и согласно теореме 2.1 множество X_* будет состоять из единственной точки u_*. Возьмем произвольную минимизирующую последовательность $\{u_k\}$. Полагая в (2) $u=u_k$ и устремляя $k\to\infty$, получаем $\delta(|u_k-u_*|)\to 0$. Это возможно только при $|u_k-u_*|\to 0$, так как $f(x)$ строго равномерно выпукла.

2. Остановимся на некоторых необходимых, а также достаточных условиях равномерной выпуклости функции.

Теорема 2. *Пусть X — открытое выпуклое множество из E^n, пусть функция $f(x)$ равномерно выпукла на X с модулем выпуклости $\delta(t)$. Тогда необходимо выполняются неравенства*

$$f(u)\geqslant f(v)+\langle c(v),u-v\rangle+\delta(|u-v|), \tag{9}$$

$$\langle c(u)-c(v),u-v\rangle\geqslant 2\delta(|u-v|) \tag{10}$$

при всех $c(v)\in\partial f(v)$, $c(u)\in\partial f(u)$ и всех $u,v\in X$.

Доказательство. Поскольку равномерно выпуклая функция является и просто выпуклой, то из теоремы 6.1 следует, что $\partial f(u)\neq\varnothing$ при всех $u\in X$. Возьмем произвольные $u,v\in X$, $c(v)\in\partial f(v)$. Из определения субградиента и из (1) при всех α, $0<\alpha<1$, имеем

$$\alpha\langle c(v),u-v\rangle+f(v)\leqslant f(\alpha u+(1-\alpha)v)\leqslant\alpha f(u)+(1-\alpha)f(v)-\alpha(1-\alpha)\delta(|u-v|)$$

или $(1-\alpha)\delta(|u-v|)+\langle c(v),u-v\rangle\leqslant f(u)-f(v)$. Отсюда при $\alpha\to+0$ получим неравенство (9). Меняя в (9) переменные u и v ролями, будем иметь

$$f(v)\geqslant f(u)+\langle c(u),v-u\rangle+\delta(|u-v|),\quad c(u)\in\partial f(u).$$

Складывая это неравенство с (9), приходим к (10). □

Приведем одно достаточное условие равномерной выпуклости функции.

Теорема 3. *Пусть X — выпуклое множество, $f(x)\in C^1(X)$, и пусть для некоторой непрерывной неотрицательной функции $\xi(t)$, $0\leqslant t\leqslant\operatorname{diam}X$, $\xi(t)=O(t^2)$ при $t\to+0$, $\xi(t)\not\equiv 0$, выполняется неравенство*

$$\langle f'(u)-f'(v),u-v\rangle\geqslant\xi(|u-v|) \tag{11}$$

при всех $u,v\in X$. Тогда функция $f(x)$ равномерно выпукла на X с модулем выпуклости $\delta(t)=\int\limits_0^1(\xi(\tau t)/\tau)d\tau$.

Доказательство. Из формулы (2.7) и условия (11) имеем

$$\alpha f(u)+(1-\alpha)f(v)-f(\alpha u+(1-\alpha)v)=$$

$$=\alpha(1-\alpha)\int_0^1\langle f'(z_1)-f'(z_2),z_1-z_2\rangle\frac{d\tau}{\tau}\geqslant\alpha(1-\alpha)\int_0^1\xi(|z_1-z_2|)\frac{d\tau}{\tau}=$$

$$=\alpha(1-\alpha)\int_0^1\xi(\tau|u-v|)\frac{d\tau}{\tau}=\alpha(1-\alpha)\delta(|u-v|),\quad u,v\in X,\quad\alpha\in[0,1],$$

что и требовалось. □

Доказанная теорема 3 может быть использована для установления равномерной выпуклости конкретных функций.

Теорема 4. *Функция $f(x)=|x|^p$ строго равномерно выпукла на E^n при всех $p\geqslant 2$.*

Доказательство. Покажем, что

$$\langle f'(u)-f'(v),\, u-v\rangle \geqslant \frac{p}{2}\min\{1;2^{3-p}\}|u-v|^p, \quad u,v\in E^n. \tag{12}$$

Здесь $f'(x)=p|x|^{p-2}x$. Тогда

$$\begin{aligned}\langle f'(u)-f'(v),\, u-v\rangle &= \langle p|u|^{p-2}u-p|v|^{p-2}v,\, u-v\rangle = \\ &= p[|u|^p+|v|^p-\langle u,v\rangle(|u|^{p-2}+|v|^{p-2})] = \\ &= p\Big[|u|^p+|v|^p-\frac{|u|^2+|v|^2-|u-v|^2}{2}(|u|^{p-2}+|v|^{p-2})\Big] = \\ &= \frac{p}{2}[(|u|^{p-2}-|v|^{p-2})(|u|^2-|v|^2)+|u-v|^2(|u|^{p-2}+|v|^{p-2})] \geqslant \\ &\geqslant \frac{p}{2}|u-v|^2(|u|^{p-2}+|v|^{p-2}), \quad u,v\in E^n. \end{aligned}\tag{13}$$

Покажем, что

$$|u|^{p-2}+|v|^{p-2}\geqslant |u-v|^{p-2}\min\{1;2^{3-p}\}, \quad u,v\in E^n. \tag{14}$$

Рассмотрим функцию $\varphi(x)=(x^\alpha+1)/(x+1)^\alpha$ при $x\geqslant 1$, $\alpha>0$. Имеем $\varphi'(x)=\alpha(x^{\alpha-1}-1)(x+1)^{-\alpha-1}$. Если $\alpha\geqslant 1$, то $\varphi'(x)\geqslant 0$, и $\varphi(x)\geqslant\varphi(1)=2^{1-\alpha}$ для всех $x\geqslant 1$. Если $0<\alpha<1$, то $\varphi'(x)<0$ и $\varphi(x)\geqslant\varphi(\infty)=1$ при всех $x\geqslant 1$. Следовательно, $\varphi(x)\geqslant A_\alpha=\min\{1;2^{1-\alpha}\}$ $\forall x\geqslant 1$, или

$$A_\alpha(x+1)^\alpha\leqslant x^\alpha+1, \quad x\geqslant 1, \quad \alpha>0. \tag{15}$$

Далее имеем $|u-v|^{p-2}\leqslant(|u|+|v|)^{p-2}$. Без ограничения общности можем считать $|u|\geqslant|v|$. Тогда с помощью неравенства (15) получим

$$|u-v|^{p-2}\leqslant|v|^{p-2}(|u|/|v|+1)^{p-2}\leqslant A_{p-2}^{-1}((|u|/|v|)^{p-2}+1)|v|^{p-2},$$

что равносильно (14). Из (13) и (14) следует неравенство (12). С помощью теоремы 3 отсюда заключаем, что функция $f(x)=|x|^p$ при $p\geqslant 2$ равномерно выпукла на E^n с модулем $\delta(t)=t^p\min\{1;2^{3-p}\}/2$. □

Более тонкие оценки показывают, что функция $f(x)=|x|^p$ при $p\geqslant 2$ имеет точный модуль выпуклости $\mu(t)=t^p/2^{p-2}$, $t\geqslant 0$.

Будет ли функция $f(x)=|x|^p$ равномерно выпуклой на E^n при $1<p<2$? Оказывается, не будет. Чтобы убедиться в этом, достаточно показать, что функция $\varphi(x)=x^p$ одной переменной при $1<p<2$ не будет равномерно выпуклой на полуоси $x\geqslant 0$, поскольку функция $f(x)=|x|^p$ вдоль лучей $x=te$, $|e|=1$, ведет себя как функция t^p одной переменной.

Если бы функция $\varphi(x)=x^p$, $1<p<2$, была равномерно выпуклой при $x\geqslant 1$ с некоторым модулем выпуклости $\delta(t)$, то согласно теореме 2 необходимо выполнялось бы неравенство (10). В данном случае неравенство (10) имеет вид

$$2\delta(t)\leqslant tp[(x+t)^{p-1}-x^{p-1}], \quad x\geqslant 0, \quad t\geqslant 0.$$

С помощью формулы конечных превращений отсюда имеем

$$2\delta(t)\leqslant t^2p(p-1)(x+\theta t)^{p-2}\leqslant t^2p(p-1)x^{p-2}, \quad x\geqslant 0, \quad t\geqslant 0.$$

Зафиксируем здесь произвольное $t>0$, а x устремим к ∞. Получим $\delta(t)\equiv 0$ при всех $t>0$.

Таким образом, функция $f(x)=|x|^p$ при $1<p<2$ не является равномерно выпуклой на E^n. Можно, однако, показать, что эта функция строго равномерно выпукла на любом выпуклом ограниченном множестве из E^n [191].

§ 8. Обоснование правила множителей Лагранжа

Пользуясь теоремами отделимости выпуклых множеств и теорией неявных функций из классического анализа, мы уже в состоянии дать строгое обоснование правила множителей Лагранжа, изложенного в § 2.3. Более того, мы получим необходимые условия экстремума первого порядка для несколько более общей задачи:

$$f(x)\to\inf, \quad x\in X, \tag{1}$$

$$X=\{x\in X_0:\ g_i(x)\leqslant 0,\ i=1,\ldots,m;\ g_i(x)=0,\ i=m+1,\ldots,s\}. \tag{2}$$

где X_0 — заданное множество из E^n, функции $f(x)$, $g_i(x)$, $i=1,\ldots,s$, определены на X_0. Здесь не исключаются возможности, когда отсутствуют либо ограничения $g_i(x)\leqslant 0$ типа неравенств ($m=0$), либо ограничения $g_i(x)=0$ типа равенств ($s=m$), либо оба вида ограничений ($m=s=0$, $X=X_0$). Если $X_0=E^n$, то получим задачи, рассмотренные в главе 2.

Разумеется, и само множество X_0 в (2) может задаваться ограничениями типа равенств и неравенств. При выделении множества X_0 обычно руководствуются тем, чтобы X_0 имело простую структуру, чтобы легко (без трудоемкой вычислительной работы) можно было проверить включение $x\in X_0$, указать какую-либо конкретную точку из X_0, чтобы легко было проектировать точку на X_0. В задачах линейного программирования (глава 3) роль X_0 играл неотрицательный ортант E^n_+. Часто множество X_0 представляет собой параллелепипед

$$X_0=\{x=(x^1,\ldots,x^n)\in E^n:\ \alpha_i\leqslant x^i\leqslant\beta_i,\ i=1,\ldots,n\},$$

где α_i, β_i — заданные числа, $\alpha_i\leqslant\beta_i$ (возможно, некоторые $\alpha_i=\infty$, $\beta_j=+\infty$).

Функция Лагранжа для задачи (1), (2) определяется также, как в главе 2:

$$\mathcal{L}(x,\bar\lambda)=\lambda_0f(x)+\lambda_1g_1(x)+\ldots+\lambda_sg_s(x), \tag{3}$$

где $x=(x^1,\ldots,x^n)\in X_0$, $\bar\lambda=(\lambda_0,\ldots,\lambda_m)\in E^{s+1}$, $\lambda_1\geqslant 0,\ldots,\lambda_m\geqslant 0$.

Теорема 1. *Пусть множество X задается условиями* (2), *где X_0 — выпуклое множество из E^n, функции $f(x)$, $g_i(x)$, $i=1,\ldots,s$, определены на X_0. Пусть $x_*\in X$ — точка локального минимума в задаче* (1), (2), *пусть функции $f(x)$, $g_1(x),\ldots,g_m(x)$ дифференцируемы в точке x_*, а функции $g_{m+1}(x),\ldots,g_s(x)$ непрерывно дифференцируемы в некоторой окрестности $O(x_*,\varepsilon)\cap X_0$ точки x_*. Тогда существуют числа $\bar\lambda=(\lambda_0,\ldots,\lambda_s)$ такие, что*

$$\bar\lambda=(\lambda_0,\ldots,\lambda_s)\neq 0,\ \lambda_0\geqslant 0,\lambda_1\geqslant 0,\ldots,\lambda_m\geqslant 0, \tag{4}$$

$$\langle\mathcal{L}_x(x_*,\bar\lambda),\,x-x_*\rangle\geqslant 0 \qquad \forall x\in X_0, \tag{5}$$

$$\lambda_ig_i(x_*)=0, \quad i=1,\ldots,m. \tag{6}$$

Сразу же заметим, что при $X_0=E^n$ условие (5) эквивалентно равенству $\mathcal{L}_x(x_*,\bar\lambda)=0$ — это легко доказывается с помощью тех же рассуждений, использованных в теореме 2.3 в аналогичной ситуации. Отсюда следует, что при $X_0=E^n$ теорема 1 превращается в теорему 2.3.2. Поэтому, доказав теорему 1, мы получим также и доказательство теорем 2.3.1, 2.3.2. Как и в главе 2, числа $\bar\lambda=(\lambda_0,\ldots,\lambda_m)$ из (3)–(6) будем называть *множителями*

Лагранжа, соответствующими точке x_*; равенства (6) — *условиями дополняющей нежесткости*. Будем также придерживаться прежних определений *активных* и *пассивных* ограничений: ограничение $g_i(x) \leqslant 0$ активно в точке x_*, если $g_i(x_*) = 0$, и пассивно в точке x_*, если $g_i(x_*) < 0$.

Из теоремы 1 следует, что точками локального минимума в задаче (1), (2) могут быть лишь те точки $x = v$, для которых существуют множители Лагранжа $\bar\lambda = (\lambda_0, \ldots, \lambda_s)$, такие, что пара $(v, \bar\lambda) \in E^{n+s+1}$ является решением системы

$$\langle \lambda_0 f'(v) + \lambda_1 g_1'(v) + \ldots + \lambda_s g_s'(v), x - v\rangle \geqslant 0 \quad \forall x \in X_0, \qquad (7)$$
$$\lambda_i g_i(v) = 0, \quad i = 1, \ldots, m;$$
$$v \in X_0, \quad g_1(v) \leqslant 0, \ \ldots, \ g_m(v) \leqslant 0, \ g_{m+1}(v) = 0, \ \ldots, \ g_s(v) = 0, \qquad (8)$$
$$\bar\lambda \neq 0, \ \lambda_0 \geqslant 0, \ \ldots, \ \lambda_m \geqslant 0. \qquad (9)$$

Множество всех тех $\bar\lambda$, для которых пара $(v, \bar\lambda)$ — решение системы (7)–(9), будем обозначать через $\Lambda = \Lambda(v)$. Так как если $(v, \bar\lambda)$ — решение системы (7)–(9), то пара $(v, \mu\bar\lambda)$ при $\forall \mu > 0$ также решение этой системы. Следовательно, $\Lambda(v)$ — конус, который как и в главе 1, будем называть *конусом Лагранжа*. Предлагаем читателю доказать, что $\Lambda(v)$ — выпуклый конус, а конус $\Lambda(v) \cup \{0\}$ замкнут.

Замечание 1. Если v — точка локального максимума функции $f(x)$ на множестве (2), то учитывая, что v — точка локального минимума функции $(-f(x))$, и применяя к $(-f(x))$ теорему 1, получим, что для точки локального максимума существуют множители Лагранжа $\bar\lambda = (\lambda_0, \ldots, \lambda_s)$, такие, что пара (v, λ) удовлетворяет соотношениям (7), (8), а условие (9) заменяется на

$$\bar\lambda \neq 0, \ \lambda_0 \leqslant 0, \ \lambda_1 \geqslant 0, \ \ldots, \ \lambda_m \geqslant 0. \qquad (10)$$

множество таких λ образует выпуклый конус, который также будем обозначать через $\Lambda(v)$ и называть конусом Лагранжа точки локального максимума.

В системах (7)–(9) и (7), (8), (10) условие $\bar\lambda \neq 0$ можно заменить каким-либо условием нормировки, например, $|\bar\lambda|^2 = \sum\limits_{i=0}^{s} \lambda_i^2 = 1$. Для выявления точек, подозрительных на локальный экстремум (минимум или максимум) достаточно рассмотреть систему (7), (8) с требованием $\lambda_1 \geqslant 0, \ldots, \lambda_m \geqslant 0$, последовательно полагая в ней $\lambda_0 = 1$, $\lambda_0 = -1$ и $\lambda_0 = 0$, $\sum\limits_{i=0}^{s} \lambda_i^2 = 1$. При обсуждении теоремы 2.3 было замечено, что вариационные неравенства $\langle \mathcal{L}_x(v, \bar\lambda), x - v\rangle \geqslant 0 \ \ \forall x \in X_0$, вообще говоря, могут быть записаны в виде системы n уравнений. Поэтому можно сказать, что система (7), (8) с учетом условий нормировки, содержит подсистему из $n+s+1$ уравнений с $n+s+1$ неизвестными $(v, \bar\lambda)$. Определив решения этой подсистемы и отобрав из них те, которые удовлетворяют остальным условиям (7)–(9) или (10), получим множество точек v, подозрительных на локальный экстремум, и соответствующий множитель $\bar\lambda$ из конуса $\Lambda(v)$. Описанный подход к поиску точек экстремума функции $f(x)$ на множестве (2), как и в главе 2, будем называть *правилом множителей Лагранжа*.

Доказательство теоремы 1 проведем, используя методику книги [670] (гл. 4, § 2). Пусть $I_* = I(x_*) = \{i\colon 1 \leqslant i \leqslant m, g_i(x_*) = 0\}$ — множество номеров активных ограничений в точке x_* (возможность $I_* = \varnothing$ здесь не исключается), пусть $|I_*|$ — количество номеров I_*. Определим множество A,

состоящее из точек $a = (a_0, a_i, i \in I_*, a_{m+1}, \ldots, a_s) \in E^{s-m+|I_*|+1}$, представимых в виде $a_0 = \langle f'(x_*), x - x_*\rangle$, $a_i = \langle g_i'(x_*), x - x_*\rangle$ для $i \in I_*$ и $i = m+1, \ldots, s$ при некотором $x \in \operatorname{ri} X_0$. Так как $\operatorname{ri} X_0 \neq \varnothing$ (теорема 1.11), то $A \neq \varnothing$. Кроме того, введем множество $B = \{b = (b_0, b_i, i \in I_*, b_{m+1}, \ldots, b_s) \in E^{s-m+|I_*|+1}\colon b_0 < 0, b_i < 0, i \in I_*, b_{m+1} = 0, \ldots, b_s = 0\}$. Очевидно, B — непустое выпуклое множество. Нетрудно доказать, что множество A также выпукло. В самом деле, пусть $d_1, d_2 \in A$. По определению множества A тогда существуют точки $x_1, x_2 \in \operatorname{ri} X_0$, такие, что $d_j = (a_{0j} = \langle f'(x_*), x_j - x_*\rangle, a_{ij} = \langle g_i'(x_*), x_j - x_*\rangle, i \in I_*, i = m+1, \ldots, s)$, $j = 1, 2$. Возьмем $\alpha \in [0, 1]$ и положим $d_\alpha = \alpha d_1 + (1-\alpha) d_2$, $x_\alpha = \alpha x_1 + (1-\alpha) x_2$. Из выпуклости $\operatorname{ri} X_0$ (теорема 1.11) следует, что $x_\alpha \in \operatorname{ri} X_0$. Далее, $\alpha a_{01} + (1-\alpha) a_{22} = \alpha \langle f'(x_*), x_1 - x_*\rangle + (1-\alpha)\langle f'(x_*), x_2 - x_*\rangle = \langle f'(x_*), x_\alpha - x_*\rangle$, $\alpha a_{i1} + (1-\alpha) a_{i2} = \alpha \langle g'(x_*), x_1 - x_*\rangle + (1-\alpha)\langle g'(x_*), x_2 - x_*\rangle = \langle g'(x_*), x_\alpha - x_*\rangle \ \ \forall i \in I_*$ и $\forall i = m+1, \ldots, s$. Следовательно, $d_\alpha \in A \ \ \forall \alpha \in [0, 1]$. Выпуклость A доказана.

Далее, возьмем несущее подпространство $L = \operatorname{Lin} X_0$ множества X_0 (определение 1.3). Пусть $e_1, \ldots, e_p$ — базис подпространства $L^\perp$, являющегося ортогональным дополнением L до E^n. Тогда $L = \{h \in E^n\colon \langle e_i, h\rangle = 0, \ i = 1, \ldots, p\}$. Заметим, что если система векторов $\{g_{m+1}'(x_*), \ldots, g_s'(x_*), e_1, \ldots, e_p\}$ линейно зависима, то теорема 1 доказывается просто. В самом деле, в этом случае существуют числа $\lambda_{m+1}, \ldots, \lambda_s, \alpha_1, \ldots, \alpha_p$, такие, что

$$\sum_{i=m+1}^{s} \lambda_i g_i'(x_*) + \sum_{i=1}^{p} \alpha_i e_i = 0, \quad (\lambda_{m+1}, \ldots, \lambda_s, \alpha_1, \ldots, \alpha_p) \neq 0. \qquad (11)$$

Тогда среди чисел $(\lambda_{m+1}, \ldots, \lambda_s)$ найдутся отличные от нуля числа, так как в противном случае из (11) следовала бы линейная зависимость векторов $e_1, \ldots, e_p$, представляющих базис подпространства $L^\perp$. Кроме того, из (11) следует, что

$$\sum_{i=m+1}^{s} \lambda_i \langle g_i'(x_*), h\rangle = -\sum_{i=1}^{p} \alpha_i \langle e_i, h\rangle = 0, \quad \forall h \in \operatorname{Lin} X_0.$$

Но $\operatorname{Lin} X_0 = \operatorname{aff} X_0 - x_*$, поэтому полагая в этом равенстве $h = x - x_*$, $x \in \operatorname{aff} X_0$, имеем: $\sum\limits_{i=m+1}^{s} \lambda_i \langle g_i'(x_*), x - x_*\rangle = 0 \ \ \forall x \in X_0 \subset \operatorname{aff} X_0$. Отсюда следует, что набор чисел $\bar\lambda = (\lambda_0 = 0, \lambda_1 = 0, \ldots, \lambda_m = 0, \lambda_{m+1}, \ldots, \lambda_s)$, где $(\lambda_{m+1}, \ldots, \lambda_s) \neq 0$ взяты из предыдущего равенства, удовлетворяют всем условиям (4)–(6).

Теорему 1 остается доказать для случая, когда система векторов $\{g_{m+1}'(x_*), \ldots, g_s'(x_*), e_1, \ldots, e_p\}$ линейно независима. Покажем, что тогда введенные выше множества A и B не пересекаются. Допустим, что $A \cap B \neq \varnothing$. Тогда найдется точка $\bar x \in \operatorname{ri} X_0$, такая, что

$$a_0 = \langle f'(x_*), \bar x - x_*\rangle < 0, \quad a_i = \langle g_i'(x_*), \bar x - x_*\rangle < 0, \quad \forall i \in I_*,$$
$$a_i = \langle g_i'(x_*), \bar x - x_*\rangle = 0, \quad \forall i = m+1, \ldots, s. \qquad (12)$$

Обозначим $h = \bar x - x_*$. Линейно независимую систему $\{g_{m+1}'(x_*), \ldots, g_s'(x_*), e_1, \ldots, e_p\}$ дополним до базиса пространства E^n любыми подходящим образом выбранными векторами $e_{p+1}, \ldots, e_{n-s+m}$ и введем функции

$$f_i(r, t) = g_{m+i}(x_* + th + r), \quad i = 1, \ldots, s-m;$$
$$f_i(r, t) = \langle e_{i-s+m}, r\rangle, \quad i = s-m+1, \ldots, n. \qquad (13)$$

Рассмотрим систему n уравнений

$$f(r, t) = (f_1(r, t), \ldots, f_n(r, t)) = 0 \tag{14}$$

относительно n неизвестных $r = (r_1, \ldots, r_n)$. Для доказательства разрешимости системы (14) воспользуемся известной из математического анализа теоремой о неявных функциях [327; 350; 352; 534]. С этой целью прежде всего заметим, что $f(0, 0) = 0$. Далее, функции $f_i(r, t)$ непрерывно дифференцируемы в окрестности точки $(0, 0)$, причем с учетом (12), (13):

$$\frac{\partial f_i(0,0)}{\partial r} = g_{m+1}'(x_*), \quad \frac{\partial f_i(0,0)}{\partial t} = \langle g_{m+1}'(x_*), h\rangle = 0, \quad i = 1, \ldots, s-m,$$
$$\frac{\partial f_i(0,0)}{\partial r} = e_{i-s+m}, \quad \frac{\partial f_i(0,0)}{\partial t} = 0, \quad i = s-m+1, \ldots, n.$$

Таким образом, якобиан $\frac{\partial(f_1, \ldots, f_n)}{\partial(r_1, \ldots, r_n)}$ системы (14) в точке $(0, 0)$, представляющий собой определитель квадратной матрицы $\frac{\partial f(0,0)}{\partial r}$ со строками $g_{m+1}'(x_*), \ldots, g_s'(x_*), e_1, \ldots, e_{n-s+m}$, образующими базис в E^n, отличен от нуля. Все условия теоремы о неявных функциях выполнены. Согласно этой теореме существуют непрерывно дифференцируемые функции $r = r(t) = (r_1(t), \ldots, r_n(t))$, определенные при всех t, $|t| \leqslant t_0$, где t_0 — достаточно малое положительное число, и такие, что

$$r(0) = 0, \quad f(r(t), t) \equiv 0 \quad \forall t, \quad |t| \leqslant t_0.$$

Дифференцируя последнее тождество по t, получаем

$$\frac{\partial f(r(t), t)}{\partial r} r'(t) + \frac{\partial f(r(t), t)}{\partial t} \equiv 0, \quad |t| \leqslant t_0.$$

Отсюда при $t = 0$ с учетом равенства $\frac{\partial f(0,0)}{\partial t} = 0$ будем иметь: $\frac{\partial f(0,0)}{\partial r} r'(0) = 0$. Однако матрица $\frac{\partial f(0,0)}{\partial r}$ невырожденная, поэтому $r'(0) = 0$. Это значит, что $r(t) = r(0) + tr'(0) + o(t) = o(t)$, т. е. $\lim\limits_{t \to 0} r(t)/t = 0$. Таким образом, найдена вектор-функция $r(t) = (r_1(t), \ldots, r_n(t))$, для которой

$$g_i(x_* + t(\bar{x} - x_*) + r(t)) = 0, \quad i = m+1, \ldots, s,$$
$$\langle e_i, r(t)\rangle = 0, \quad i = 1, \ldots, n-s+m, \quad |t| \leqslant t_0, \quad \lim_{t \to 0} r(t)/t = 0. \tag{15}$$

Покажем, что по кривой $x = x(t) = x_* + t(\bar{x} - x_*) + r(t)$ можно двигаться, оставаясь в множестве X при всех t, $0 < t < t_1$, где t_1 — достаточно малое число, $0 < t_1 \leqslant \min\{t_0; 1\}$. В самом деле, равенства $\langle e_i, r(t)\rangle = 0$, $i = 1, \ldots, p$, означают, что $r(t) \in \operatorname{Lin} X_0$. Кроме того, $\bar{x} \in \operatorname{ri} X_0$, $\lim\limits_{t \to 0} r(t)/t = 0$, поэтому $\bar{x} + r(t)/t \in X_0$ при всех малых t. Тогда, учитывая выпуклость X_0, имеем $x(t) = t(\bar{x} + r(t)/t) + (1 - t)x_* \in X_0$ $\forall t$, $0 < t \leqslant t_1$. Далее, первые равенства (15) означают, что $g_i(x(t)) = 0$, $0 \leqslant t \leqslant t_1$, $\forall i = m+1, \ldots, s$. Покажем, что $g_i(x(t)) \leqslant 0$ $\forall t$, $0 \leqslant t \leqslant t_1$, и $\forall i = 1, \ldots, m$. Если $i \in I_*$, то $g_i(x_*) = 0$, и с учетом (12) имеем

$$g_i(x(t)) = g_i(x_*) + \langle g_i'(x_*), t(\bar{x} - x_*) + r(t)\rangle + o(t) =$$
$$= t[\langle g_i'(x_*), \bar{x} - x_*\rangle + \langle g_i'(x_*), r(t)/t\rangle + o(t)/t] < 0$$

при всех малых $t > 0$. Если $i \notin I_*$, $1 \leqslant i \leqslant m$, то $g_i(x_*) < 0$ и в силу непрерывности $g_i(x)$ неравенство $g_i(x(t)) = g_i(x_* + t(\bar{x} - x_*) + r(t)) < 0$ также сохранится при всех малых $t > 0$. Таким образом, существует достаточно малое число $t_1 > 0$, такое, что $x(t) \in X$ при всех t, $0 \leqslant t \leqslant t_1$. Беря при необходимости t_1 еще меньшим, с учетом (12) имеем

$$f(x(t)) - f(x_*) = t\left[\langle f_i'(x_*), \bar{x} - x_*\rangle + \langle f_i'(x_*), r(t)/t\rangle + \frac{o(t)}{t}\right] < 0 \quad \forall t,\ 0 < t \leqslant t_1.$$

Однако $x(t) \to x_*$ при $t \to 0$ и $x(t) \in X$, $0 < t < t_1$, и последнее неравенство противоречит тому, что x_* — точка локального минимума в задаче (1), (2). Полученное противоречие доказывает, что $A \cap B = \varnothing$.

Итак, A и B — выпуклые множества, $A \cap B = \varnothing$. По теореме 5.2 тогда существует гиперплоскость $\langle c, a\rangle = \gamma$ с нормальным вектором $c = (\lambda_0; \lambda_i, i \in I_*; \lambda_{m+1}, \ldots, \lambda_s) \in E^{s-m+|I_*|+1}$, $c \neq 0$, отделяющая множества A и B, а также A и $\overline{B} = \{b = (b_0; b_i, i \in I_*; b_{m+1}, \ldots, b_s)$: $b_0 \leqslant 0$; $b_i \leqslant 0$, $i \in I_*$, $b_{m+1} = 0, \ldots, b_s = 0\}$. Это значит, что

$$\langle c, b\rangle = \lambda_0 b_0 + \sum_{i \in I_*} \lambda_i b_i + \sum_{i=m+1}^{s} \lambda_i b_i \leqslant \gamma \leqslant \langle c, a\rangle = \lambda_0 a_0 + \sum_{i \in I_*} \lambda_i a_i + \sum_{i=m+1}^{s} \lambda_i a_i \tag{16}$$

при всех $a \in A$, $b \in \overline{B}$. Разделив (16) почленно на $b_j < 0$, где $j = 0$ или $j \in I_*$, и устремляя затем $b_j \to -\infty$ при фиксированных остальных b_i, a, получим $\lambda_j \geqslant 0$ при $j = 0$ или $\forall j \in I_*$. Далее полагая в (16) $a_0 = \langle f'(x_*), x - x_*\rangle$, $a_i = \langle g'(x_*), x - x_*\rangle$, $i \in I_*$ и $i = m+1, \ldots, s$, где $x \in \operatorname{ri} X$, $b = 0 \in \overline{B}$, будем иметь

$$\lambda_0\langle f'(x_*), x - x_*\rangle + \sum_{i \in I_*} \lambda_i \langle g'(x_*), x - x_*\rangle + \sum_{i=m+1}^{s} \lambda_i \langle g'(x_*), x - x_*\rangle \geqslant 0 \quad \forall x \in \operatorname{ri} X_0.$$

Отсюда, доопределив $\lambda_i = 0$ при $i \notin I_*$, $1 \leqslant i \leqslant m$, получим

$$\langle \lambda_0 f'(x_*) + \sum_{i=1}^{s} \lambda_i g_i'(x_*), x - x_*\rangle \geqslant 0 \qquad \forall x \in \operatorname{ri} X_0.$$

Совершая в этом неравенстве предельные переходы с учетом того, что $X_0 \subset \overline{X_0} = \overline{\operatorname{ri} X_0}$ (теорема 1.13), придем к неравенству (5). Справедливость условий (4), (6) вытекает из определения множества I_* и построения $\bar{\lambda} = (\lambda_0, \ldots, \lambda_s)$. Теорема 1 доказана. □

Другое доказательство теоремы 1, а также теорем 2.3.1, 2.3.2, не использующее теоремы отделимости и теорию неявных функций, будет приведено ниже в § 5.16 с помощью штрафных функций при несколько более жестких требованиях на дифференциальные свойства функций $f(x)$, $g_i(x)$, $i = 1, \ldots, m$. Различные доказательства, обобщения и модификации правила множителей Лагранжа см., например, в [5–7; 14; 15; 24; 34; 44; 106; 209; 225; 233; 234; 278; 279; 286; 297; 314; 347; 358; 366; 386; 434; 465; 502; 587; 602; 604; 605; 613; 617; 660; 670; 673; 683; 724; 759; 816].

Замечание 2. Если в конусе Лагранжа $\Lambda(v)$ точки $v \in X$ существуют наборы $\bar{\lambda} = (\lambda_0, \ldots, \lambda_s)$ с $\lambda_0 = 0$, то в условиях (7)–(9) целевая функция «исчезает» и эти условия превращаются в некоторую специфическую характеристику множества (2) в точке v. Как и в главе 2, выделим класс задач на экстремум функции $f(x)$ на множестве (2), у которых любой элемент $\bar{\lambda}$ конуса $\Lambda(v)$ имеет координату $\lambda_0 \neq 0$.

Определение 1. Точку v множества (2) назовем *нормальной* точкой этого множества, если система

$$\langle \sum_{i=1}^{s} \lambda_i g_i'(v), x-v\rangle \geqslant 0 \quad \forall x \in X_0, \quad \lambda_i g_i(v)=0, \quad i=1,\ldots,m, \qquad (17)$$
$$\lambda=(\lambda_1,\ldots,\lambda_m)\neq 0, \quad \lambda_1 \geqslant,\ldots,\lambda_m \geqslant 0,$$

относительно переменных λ не имеет решения.

Система (17) получена из систем (7)–(9) и (7), (8), (10) при $\lambda_0=0$. Поэтому рассуждениями от противного нетрудно доказать, что в нормальной точке v множества X все точки $\bar\lambda$ из конуса Лагранжа $\Lambda(v)$ имеют координату $\lambda_0\neq 0$.

По аналогии с § 2.4 можно сформулировать условие Мангасариана — Фрамовица, гарантирующее нормальность точки v из множества (2). А именно, пусть в (2) множество X_0 имеет непустую внутренность и $v\in \operatorname{int} X_0$, векторы $g_{m+1}'(v),\ldots,g_s'(v)$ линейно независимы, и существует вектор $d\in E^n$, для которого $\langle g_i'(v), d\rangle=0$, $i=m+1,\ldots,s$, $\langle g_i'(v), d\rangle<0$ $\forall i\in I(v)$, где $I(v)$ множество номеров активных ограничений точки v. При $v\in \operatorname{int} X_0$ неравенство $\langle \mathcal{L}_x(v,\bar\lambda), x-v\rangle\geqslant 0$ $\forall x\in X_0$ эквивалентно равенству $\mathcal{L}_x(v,\bar\lambda)=0$, и доказательство нормальности точки v при выполнении перечисленных условий проводится также, как в § 2.4.

Если в (2) ограничения типа равенств отсутствуют ($m=s$), то условие Мангасариана — Фрамовица можно модифицировать следующим образом: существует вектор $d\in E^n$ такой, что

$$\exists t_0>0, \text{ что } v+t_0 d\in X_0, \quad \langle g_i'(v), d\rangle<0, \quad \forall i\in I(v). \qquad (18)$$

Покажем, что при выполнении условия (18) точка v нормальна. Допустим противное: пусть условие (18) выполнено, но система (17) при $m=s$ имеет хотя бы одно решение λ. В неравенстве $\langle \sum_{i=1}^{m} \lambda_i g_i'(v), x-v\rangle\geqslant 0$ $\forall x\in X_0$ положим $x=v+t_0 d$. С учетом (18) получим: $0\leqslant \langle \sum_{i=1}^{m} \lambda_i g_i'(v), t_0 d\rangle = t_0 \sum_{i\in I(v)} \lambda_i \langle g_i'(v), d\rangle \leqslant 0$, что возможно только при $\lambda=0$. Однако $\lambda=0$ не может быть решением системы (17) при $m=s$. Полученное противоречие доказывает, что v — нормальная точка множества (2).

Приведем еще одно достаточное условие нормальности точки v из множества (2) при $m=s$. Пусть существует точка $\bar x\in X_0$, для которой $g_i(\bar x)<0$, $i=1,\ldots,m$. Это условие в литературе принято называть *условием Слейтера*. Если выполнено условие Слейтера и, кроме того X_0 — выпуклое множество и функции $g_i(x)$, $i=1,\ldots,m$, выпуклы на X_0, то, оказывается, выполняется условие (18). В самом деле, положим $d=\bar x-v$. Тогда точка $x=v+td\in X_0$ при $t=t_0=1$. Кроме того, из выпуклости функций $g_i(x)$ на X_0 и теоремы 2.2 следует, что $0>g_i(\bar x)=g_i(\bar x)-g_i(v)\geqslant\langle g_i'(v), \bar x-v\rangle=\langle g_i'(v), d\rangle$ $\forall i\in I(v)$. Как видим, условие (18) выполнено. Следовательно, v — нормальная точка множества (2) при $m=s$.

Кратко скажем, что понятие анормальной точки для множеств (2) можно ввести точно также, как в § 2.4 (определение 2.4.5); тогда в конусе Лагранжа $\Lambda(v)$ существует точка $\bar\lambda$ с координатой $\lambda_0=0$, конус $\Lambda(v)\cup\{0\}$ неострый.

Нетрудно привести примеры задач, когда в конусе Лагранжа все наборы $\bar\lambda$ имеют координату $\lambda_0=0$.

Пример 1. Задача: $f(x)=-x\to\inf$, $x\in X=\{x\in X_0:\ g(x)=x^2\leqslant 0\}$, где $X_0=\{x\in E^1:\ 0\leqslant x\leqslant a\}$, $a>0$ (возможно, $a=+\infty$). Тогда $X=\{0\}=X_*$, $f_*=0$, функция Лагранжа $\mathcal{L}(x,\bar\lambda)=-\lambda_0 x+\lambda x^2$, $x\in X_0$, $\lambda\geqslant 0$. Система (7)–(9) имеет вид:

$$(-\lambda_0+2\lambda v)(x-v)\geqslant 0 \quad \forall x\in X_0, \quad \lambda v^2=0, \quad v^2\leqslant 0,$$
$$\bar\lambda=(\lambda_0,\lambda)\neq 0, \quad \lambda_0\geqslant 0, \quad \lambda\geqslant 0.$$

Отсюда видно, что $v=0$. Тогда первое неравенство этой системы дает: $-\lambda_0 x\geqslant 0$ $\forall x\in[0,a]$, что возможно только при $-\lambda_0\geqslant 0$. С другой стороны $\lambda_0\geqslant 0$. Следовательно, $\lambda_0=0$, и конус $\Lambda(0)=\{\bar\lambda=(0,\lambda):\ \lambda>0\}$.

Упражнения

1. С помощью правила множителей Лагранжа исследовать задачи на экстремум, если:

а) $f(u)=x^2+y^2+z^2$, $X=\{u=(x,y,z)\in X_0,\ x+y+z=1[\leqslant 1;\geqslant 1]\}$, $X_0=\{u=(x,y,z)\in E^3:\ x\geqslant 0\}$, или $X_0=\{u=(x,y,z)\in E^3:\ x\geqslant 0, y\geqslant 0\}$, или $X_0=E^3_+$;

б) $f(u)=\sin(x+y)-\sin x-\sin y$, $X=\{u=(x,y)\in X_0:\ x+y\leqslant 2\pi\}$, $X_0=\{u=(x,y)\in E^2:\ x\geqslant 0\}$ или $X_0=E^2_+$.

2. Найти решения задач:

а) $f(u)=2x^{-2}+4x^5y^{-2}\inf$, $u\in X=\{u=(x,y)\in E^2:\ x>0, y>0, x^4y^2\leqslant 1\}$;

б) $f(u)=x+y^{-1}z^{-1/2}\to\inf$, $u\in X=\{u=(x,y,z)\in E^3:\ x>0, y>0, z>0, x^{-1}y+x^{-1}z\leqslant 1\}$.

3. Найти точки экстремума функции $f(u)=|u-a|^2$, где $a=(a_1,a_2)\in E^2$ — заданная точка, на множествах $X_1=\{u\in E^2_+:\ x^2+y^2\leqslant 1\}$, $X_2=\{u\in E^2_+:\ x^2+y^2\geqslant 1\}$, $X_3=\{u\in E^2_+:\ x^2+y^2=1\}$.

Указание: изобразить на плоскости E^2 множества X и линии уровня функции $f(u)$.

§ 9. Теорема Куна — Таккера. Двойственная задача

1. Перейдем к рассмотрению условий оптимальности для задач выпуклого программирования. Под *выпуклым программированием* понимается раздел теории экстремальных задач, в котором изучаются задачи минимизации [или максимизации] выпуклых [вогнутых] функций на выпуклых множествах. В частности, задача

$$f(x)\to\inf, \quad x\in X \qquad (1)$$
$$X=\{x\in X_0:\ g_i(x)\leqslant 0,\ i=1,\ldots,m;\ g_i(x)=0,\ i=m+1,\ldots,s\}, \qquad (2)$$

исследованная в § 8, превращается в задачу выпуклого программирования, если X_0 — выпуклое множество, функции $f(x), g_1(x),\ldots,g_m(x)$ выпуклы на X_0, а $g_i(x)=\langle a_i, x\rangle-b_i$, $i=m+1,\ldots,s$ — линейные функции с известными $a_i\in E^n$, $b_i\in\mathbb{R}$ (теорема 2.11). Такую задачу кратко принято называть *выпуклой задачей*. Ряд характерных свойств выпуклых задач были отмечены выше (см., например, теоремы 2.1, 2.3, 2.12, 6.4).

Важное место в теории выпуклого программирования занимает теорема о седловой точке функции Лагранжа, известная в литературе под названием теоремы Куна — Таккера. Эта теорема дает необходимое и достаточное условие оптимальности в задаче (1), (2), т. е. условие принадлежности той

или иной точки множеству $X_* = \{x \in X: f(x) = \inf\limits_{v \in X} f(v) = f_*\}$. Для формулировки теоремы Куна — Таккера введем функцию

$$L(x, \lambda) = f(x) + \sum_{i=1}^{s} \lambda_i g_i(x), \tag{3}$$

называемую в отличие от (8.3) *нормальной функцией Лагранжа* задачи (1), (2), где $x \in X_0$, а переменные $\lambda = (\lambda_1, \ldots, \lambda_s)$ принадлежат множеству

$$\Lambda_0 = \{\lambda = (\lambda_1, \ldots, \lambda_s) \in E^s: \lambda_1 \geqslant 0, \ldots, \lambda_m \geqslant 0\}. \tag{4}$$

Определение 1. Точку $(x_*, \lambda^*) \in X_0 \times \Lambda_0$ называют *седловой точкой* функции Лагранжа (3), если

$$L(x_*, \lambda) \leqslant L(x_*, \lambda^*) \leqslant L(x, \lambda^*) \quad \forall x \in X_0, \quad \lambda \in \Lambda_0. \tag{5}$$

Прежде чем переходить к выяснению связи между седловой точкой функции Лагранжа и решением задачи (1), (2), дадим другую равносильную (5) формулировку определения седловой точки.

Лемма 1. *Для того чтобы точка $(x_*, \lambda^*) \in X_0 \times \Lambda_0$ была седловой точкой функции Лагранжа, необходимо и достаточно, чтобы выполнялись следующие условия:*

$$L(x_*, \lambda^*) \leqslant L(x, \lambda^*) \qquad \forall x \in X_0, \tag{6}$$

$$\lambda_i^* g_i(x_*) = 0, \qquad i = 1, \ldots, m, \quad x_* \in X. \tag{7}$$

Подчеркнем, что в лемме 1 от множеств X_0 и функций $f(x)$, $g_i(x)$, $i = 1, \ldots, s$, не требуется ни выпуклость, ни какая-либо гладкость — здесь важно то, что $X_0 \neq \varnothing$ и функции $f(x)$, $g_i(x)$, $i = 1, \ldots, s$, определены и конечны на X_0.

Доказательство. Необходимость. Пусть $(x_*, \lambda^*) \in X_0 \times \Lambda_0$ — седловая точка. Тогда условие (6) представляет собой правое неравенство (5). Остается получить условия (7). Для этого перепишем левое неравенство (5) с учетом конкретного вида (3) функции Лагранжа:

$$f(x_*) + \sum_{i=1}^{s} \lambda_i g_i(x_*) \leqslant f(x_*) + \sum_{i=1}^{s} \lambda_i^* g_i(x_*) \quad \forall \lambda \in \lambda_0. \tag{8}$$

Отсюда имеем

$$\sum_{i=1}^{s} (\lambda_i^* - \lambda_i) g_i(x_*) \geqslant 0 \quad \forall \lambda \in \Lambda_0. \tag{9}$$

Покажем, что $x_* \in X$. Возьмем точку $\lambda = (\lambda_1, \ldots, \lambda_s)$, где $\lambda_j = \lambda_j^* + 1$ при некотором j, $1 \leqslant j \leqslant m$, и $\lambda_i = \lambda_i^*$ при всех остальных $i = 1, \ldots, s$ $(i \neq j)$. Из определения (4) множества Λ_0 и из того, что $\lambda^* \in \lambda_0$, следует, что выбранная точка $\lambda \in \Lambda_0$. Из (9) при таком λ получим $(-1) g_j(x_*) \geqslant 0$, т. е. $g_j(x_*) \leqslant 0$ при всех $j = 1, \ldots, m$.

Далее, пусть $\lambda = (\lambda_1, \ldots, \lambda_s)$ — точка с координатами $\lambda_j = \lambda_j^* + g_j(x_*)$ при некотором j, $m+1 \leqslant j \leqslant s$, и $\lambda_i = \lambda_i^*$ при всех $i = 1, \ldots, s$, $i \neq j$. Ясно, что $\lambda \in \Lambda_0$. Поэтому из (9) имеем $-|g_j(x_*)|^2 \geqslant 0$, т. е. $g_j(x_*) = 0$ при всех $j = m+1, \ldots, s$. Таким образом, доказано, что $x_* \in X$.

Возьмем точку $\lambda = (\lambda_1, \ldots, \lambda_s)$ с координатами $\lambda_j = 0$ при некотором j, $1 \leqslant j \leqslant m$, и $\lambda_i = \lambda_i^*$ при всех остальных $i = 1, \ldots, s$, $i \neq j$. Такая точка принадлежит Λ_0, поэтому из (9) получим $0 \leqslant \lambda_j^* g_j(x_*)$. Но $\lambda_j^* \geqslant 0$, $g_j(x_*) \leqslant 0$ при $j = 1, \ldots, m$, поэтому последнее неравенство возможно лишь при $\lambda_j^* g_j(x_*) = 0$, $j = 1, \ldots, m$. Все соотношения (6), (7) получены.

Достаточность. Пусть для некоторой точки $(x_*, \lambda^*) \in X_0 \times \Lambda_0$ выполнены соотношения (6), (7). Покажем, что тогда (x_*, λ^*) — седловая точка. Из (6) следует правое неравенство (5). Остается доказать левое неравенство (5). По условию (7) $x_* \in X$, т. е. $g_i(x_*) \leqslant 0$, $i = 1, \ldots, m$, $g_i(x_*) = 0$, $i = m+1, \ldots, s$. Тогда

$$(\lambda_i^* - \lambda_i) g_i(x_*) = 0 \tag{10}$$

при всех $i = m+1, \ldots, s$ и всех тех i, $1 \leqslant i \leqslant m$, для которых $g_i(x_*) = 0$. Если $g_i(x_*) < 0$ при некотором i, $1 \leqslant i \leqslant m$, то из равенства (7) следует, что $\lambda_i^* = 0$. Поэтому $(\lambda_i^* - \lambda_i) g_i(x_*) = -\lambda_i g_i(x_*) \geqslant 0$ для всех $\lambda_i \geqslant 0$, $1 \leqslant i \leqslant m$, для которых $g_i(x_*) < 0$. Складывая полученные неравенства с (10), будем иметь $\sum\limits_{i=1}^{s} (\lambda_i^* - \lambda_i) g_i(x_*) \geqslant 0$ для всех $\lambda \in \Lambda_0$. Отсюда $\sum\limits_{i=1}^{s} \lambda_i g_i(x_*) \leqslant \sum\limits_{i=1}^{s} \lambda_i^* g_i(x_*)$ при всех $\lambda \in \Lambda_0$. Добавляя к обеим частям этого неравенства $f(x_*)$, придем к неравенству (8), представляющему собой левое неравенство (5). □

Если сделать дополнительные предположения о выпуклости и гладкости задачи (1), (2), то лемму 1 можно переформулировать в следующей так называемой дифференциальной форме.

Лемма 2. *Пусть (1), (2) представляет собой задачу выпуклого программирования и функции $f(x), g_1(x), \ldots, g_m(x)$ дифференцируемы в точке $x_* \in X_0$. Тогда для того чтобы точка $(x_*, \lambda^*) \in X_0 \times \Lambda_0$ была седловой точкой функции Лагранжа, необходимо и достаточно, чтобы*

$$\langle \mathcal{L}_x(x_*, \lambda^*), x - x_* \rangle = \langle f'(x_*) + \sum_{i=1}^{s} \lambda_i^* g_i'(x_*), x - x_* \rangle \geqslant 0 \quad \forall x \in X_0, \tag{11}$$

$$\lambda_i^* g_i(x_*) = 0, \quad i = 1, \ldots, m, \quad x_* \in X. \tag{12}$$

Доказательство. При сделанных предположениях функция Лагранжа (3) выпукла и дифференцируема в точке $x_* \in X_0$ при каждом $\lambda \in \Lambda_0$. Поэтому условие (6) согласно теореме 2.3 равносильно условию (11). Условия (7) и (12) совпадают. □

Теперь выясним, как связаны между собой седловая точка функции Лагранжа и решение задачи (1), (2).

Теорема 1. *Пусть $(x_*, \lambda^*) \in X_0 \times \Lambda_0$ — седловая точка функции Лагранжа. Тогда $x_* \in X$, $f_* = L(x_*, \lambda^*) = f(x_*)$, т. е. x_* является решением задачи* (1), (2).

Доказательство. Из условия (7) имеем $x_* \in X$ и $L(x_*, \lambda^*) = f(x_*)$. Тогда неравенство (6) перепишется в виде

$$f(x_*) \leqslant L(x, \lambda^*) = f(x) + \sum_{i=1}^{s} \lambda_i^* g_i(x), \quad x \in X_0. \tag{13}$$

В частности, (13) верно и для всех $x \in X$. Но $\sum\limits_{i=1}^{s} \lambda_i^* g_i(x) \leqslant 0$ при $x \in X$, так как тогда $g_i(x) \leqslant 0$ и $\lambda_i^* \geqslant 0$ при $i = 1, \ldots, m$ и $g_i(x) = 0$ при $i = m+1, \ldots, s$. Поэтому из (13) следует, что $f(x_*) \leqslant L(x, \lambda^*) \leqslant f(x)$ при всех $x \in X$, т. е. $x_* \in X_*$. □

Заметим, что теорема 1, как и лемма 1, доказаны без каких-либо ограничений на функции $f(x)$, $g_i(x)$, $i = 1, \ldots, s$, и на множество X_0; в частности, никакие предположения о выпуклости, сделанные выше при формулировке задачи (1), (2), мы пока не использовали.

2. Возникает вопрос: во всякой ли задаче вида (1), (2) функция Лагранжа имеет седловую точку? Ответ здесь, конечно, отрицательный: если в зада-

че (1), (2) $X_* = \varnothing$, то, как следует из теоремы 1, функция Лагранжа такой задачи не может иметь седловую точку. Более того, даже в выпуклых задачах с $X_* \neq \varnothing$ в общем случае нельзя ожидать, что функция Лагранжа будет иметь седловую точку.

Пример 1. Рассмотрим задачу из примера 8.1: $f(x) = -x \to \inf$, $x \in X = \{x \in X_0\colon\ g(x) = x^2 \leqslant 0\}$, где $X_0 = \{x \in E^1\colon\ 0 \leqslant x \leqslant a\}$, $0 < a \leqslant \infty$. Здесь множество X_0 выпукло, функции $f(x)$, $g(x)$ выпуклы на X_0. Множество X состоит из одной точки $x = 0$, так что $f_* = f(0) = 0$, $X_* = \{0\}$. Функция Лагранжа

$$L(x, \lambda) = -x + \lambda x^2, \quad 0 \leqslant x \leqslant a, \quad \lambda \geqslant 0,$$

рассматриваемой задачи не имеет седловой точки.

Таким образом, для существования седловой точки на задачу (1), (2), кроме условий выпуклости, должны быть наложены какие-то дополнительные ограничения. Начнем с рассмотрения случая, когда в (2) ограничения типа равенств отсутствуют ($m = s$), т. е. множество X имеет вид

$$X = \{x \in X_0\colon\ g_i(x) \leqslant 0,\ i = 1, \ldots, m\}. \tag{14}$$

Предположим, что выполнено условие Слейтера, т. е. существует точка $\bar{x} \in X$ такая, что

$$g_1(\bar{x}) < 0, \ \ldots, \ g_m(\bar{x}) < 0. \tag{15}$$

Напомним, что условием (15) мы уже пользовались в § 8.

Если X_0 — выпуклое множество, функции $g_i(x)$ выпуклы на X_0, то вместо (15) достаточно потребовать для каждого i существования точки $\bar{x}_i \in X$ такой, что $g_i(\bar{x}_i) < 0$, $i = 1, \ldots, m$. Тогда в качестве $\bar{x}$ из (15) можно взять $\bar{x} = \sum\limits_{i=1}^{m} \alpha_i \bar{x}_i$, $\alpha_i > 0$, $\alpha_1 + \alpha_2 + \ldots + \alpha_m = 1$, поскольку $\bar{x} \in X_0$ и в силу неравенства (2.2) $g_j(\bar{x}) \leqslant \sum\limits_{i=1}^{m} \alpha_i g_j(\bar{x}_i) \leqslant \alpha_i g_j(\bar{x}_j) < 0$, $j = 1, \ldots, m$.

Не следует думать, что если множество (14) выпукло и имеет внутренние точки, то условие Слейтера непременно выполняется.

Пример 2. Задача: $f(x) = x \to \inf$, $x \in X = \{x \in E^1\colon\ g(x) \leqslant 0\}$, где

$$g(x) = \begin{cases} x^2 & \text{при } x < 0, \\ 0 & \text{при } x \geqslant 0. \end{cases}$$

Очевидно, функции $f(x)$, $g(x)$ выпуклы (и даже дифференцируемы) на $X_0 = E^1$, так что задача выпукла. Здесь $X = E^1_+$, $X_* = \{0\}$, $f_* = 0$. Как видим, все точки $x > 0$ являются внутренними для множества X, но $g(x) \equiv 0$ $\forall x > 0$, и условие (15) заведомо не может выполняться. Убедимся, что функция Лагранжа $L(x, \lambda) = x + \lambda g(x)$, $x \in X_0 = E^1$, $\lambda \in \Lambda_0 = E^1_+$ этой задачи не имеет седловой точки. Согласно теореме 1 седловыми могут быть лишь точки вида ($x_* = 0$, $\lambda \geqslant 0$). Однако неравенство $L(0, \lambda) = 0 = f_* \leqslant L(x, \lambda)$ не может выполняться при всех $x \in E^1$ ни при каком $\lambda \geqslant 0$. В самом деле, если $\lambda = 0$, то $L(x, \lambda) = x < 0$ $\forall x < 0$; если $\lambda > 0$, то $L(x, \lambda) = x + \lambda x^2 < 0$ при всех x, $-\lambda < x < 0$.

Теорема 2 (Кун — Таккер). *Пусть множество X_0 выпукло, функции $f(x)$, $g_i(x)$, $i = 1, m$, выпуклы на X_0 и выполнено условие* (15). *Пусть множество X_* точек минимума функции $f(x)$ на множестве* (14) *непусто. Тогда для каждой точки $x_* \in X_*$ необходимо существуют множители Лагранжа $\lambda^* = (\lambda_1^*, \ldots, \lambda_m^*) \in \Lambda_0 = \{\lambda \in E^m\colon\ \lambda_1 \geqslant 0, \ldots, \lambda_m \geqslant 0\}$ такие, что*

пара (x_*, λ^*) образует седловую точку функции Лагранжа на множестве $X_0 \times \Lambda_0$.

Доказательство. В пространстве E^{m+1} переменных $a = (a_0, a_1, \ldots, a_m)$ введем множества

$$A = \{a = (a_0, a_1, \ldots, a_m) \in E^{m+1}\colon\ a_0 \geqslant f(x),\ a_1 \geqslant g_1(x), \ldots, a_m \geqslant g_m(x),\ x \in X_0\},$$
$$B = \{b = (b_0, b_1, \ldots, b_m) \in E^{m+1}\colon\ b_0 < f_*,\ b_1 < 0, \ldots, b_m < 0\}.$$

Покажем, что A и B не имеют общих точек. В самом деле, пусть $a \in A$. Тогда найдется точка $x \in X_0$ такая, что $a_0 \geqslant f(x)$, $a_1 \geqslant g_1(x), \ldots, a_m \geqslant g_m(x)$. Возможно, что $x \in X$. Тогда $a_0 \geqslant f(x) \geqslant f_*$ и заведомо $a \notin B$. Если же $x \in X_0 \setminus X$, то найдется номер i, $1 \leqslant i \leqslant m$, такой, что $g_i(x) > 0$. Тогда $a_i \geqslant g_i(x) > 0$ и снова $a \notin B$. Итак, $A \cap B = \varnothing$.

Далее, нетрудно видеть, что A и B — выпуклые множества. Покажем, например, что A выпукло. Пусть a, c — две произвольные точки из A. Тогда существуют точки u, $v \in X_0$ такие, что $a_0 \geqslant f(u)$, $c_0 \geqslant f(v)$, $a_i \geqslant g_i(u)$, $c_i \geqslant g_i(v)$, $i = 1, \ldots, m$. Возьмем произвольное $\alpha \in [0, 1]$ и положим $a_\alpha = \alpha a + (1 - \alpha)c$, $u_\alpha = \alpha u + (1 - \alpha)v$. Из выпуклости X_0 следует $u_\alpha \in X_0$. Далее, из выпуклости функций $f(u)$, $g_i(u)$ имеем

$$f(u_\alpha) \leqslant \alpha f(u) + (1 - \alpha) f(v) \leqslant \alpha a_0 + (1 - \alpha) c_0,$$
$$g_i(u_\alpha) \leqslant \alpha g_i(u) + (1 - \alpha) g_i(v) \leqslant \alpha a_i + (1 - \alpha) c_i, \quad i = 1, \ldots, m.$$

Это означает, что $a_\alpha \in A$. Выпуклость A доказана. Аналогично доказывается выпуклость B.

В силу теоремы 5.2 тогда существует гиперплоскость $\langle c, a\rangle = \gamma$ с нормальным вектором $c = (\lambda_0^*, \lambda_1^*, \ldots, \lambda_m^*) \neq 0$, отделяющая A и B, а также A и $\overline{B} = \{b = (b_0, b_1, \ldots, b_m) \in E^{m+1}\colon\ b_0 \leqslant f_*, b_1 \leqslant 0, \ldots, b_m \leqslant 0\}$. Это значит, что

$$\langle c, b\rangle = \sum_{i=0}^{m} \lambda_i^* b_i \leqslant \gamma \leqslant \langle c, a\rangle = \sum_{i=0}^{m} \lambda_i^* a_i \quad \forall a \in A, \quad b \in \overline{B}. \tag{16}$$

Заметим, что $y = (f_*, 0, \ldots, 0) \in A \cap \overline{B}$. В самом деле, возьмем какую-либо точку $x_* \in X_*$. Тогда $f(x_*) = f_*$, $g_i(x_*) \leqslant 0$, $i = 1, \ldots, m$, что означает $y \in A$. Включение $y \in \overline{B}$ очевидно. Тогда по теореме 5.2 величина γ из (16) равна $\gamma = \langle c, y\rangle = \lambda_0^* f_*$, и (16) можно переписать в виде

$$\lambda_0^* b_0 + \sum_{i=1}^{m} \lambda_i^* b_i \leqslant \lambda_0^* f_* \leqslant \lambda_0^* a_0 + \sum_{i=1}^{m} \lambda_i^* a_i \quad \forall a \in A, \quad b \in \overline{B}. \tag{17}$$

Возьмем точку $b = (f_* - 1, 0, \ldots, 0) \in \overline{B}$. Из левого неравенства (17) получим $\lambda_0^*(f_* - 1) \leqslant \lambda_0^* f_*$, откуда $\lambda_0^* \geqslant 0$. Далее, беря $b = (f_*, 0, \ldots, 0, -1, 0, \ldots, 0)$, из левого неравенства (17) имеем $\lambda_0^* f_* - \lambda_i^* \leqslant \lambda_0^* f_*$, т. е. $\lambda_i^* \geqslant 0$, $i = 1, \ldots, m$. Таким образом, показано, что $\lambda^* = (\lambda_1^*, \ldots, \lambda_m^*) \geqslant 0$, $\lambda_0^* \geqslant 0$.

Далее, возьмем произвольную точку $x_* \in X_*$. Тогда $a = (f(x_*) = f_*, 0, \ldots, 0, g_i(x_*), 0, \ldots, 0) \in A \cap \overline{B}$. Подставляя эту точку в левое и правое неравенства (17), получаем $\lambda_0^* f_* + \lambda_i^* g_i(x_*) \leqslant \lambda_0^* f_* \leqslant \lambda_0^* f_* + \lambda_i^* g_i(x_*)$, откуда $\lambda_i^* g_i(x_*) \leqslant 0 \leqslant \lambda_i^* g_i(x_*)$ или $\lambda_i^* g_i(x_*) = 0$, $i = 1, \ldots, m$. Равенства (7) доказаны.

Покажем, что $\lambda_0^* > 0$. В самом деле, в (17) подставим $a = (f(\bar{x}), g_1(\bar{x}), \ldots, g_m(\bar{x})) \in A$, где $\bar{x}$ взято из (15). Получим $\lambda_0^* f_* \leqslant \lambda_0^* f(\bar{x}) + \sum\limits_{i=0}^{m} \lambda_i^* g_i(\bar{x})$. Допустим, что $\lambda_0^* = 0$. Тогда $\lambda^* \neq 0$ и из предыдущего неравенства при $\lambda_0^* = 0$ с

учетом условия (15) имеем $0 \leqslant \sum_{i=1}^{m} \lambda_i^* g_i(\bar{x}) < 0$. Полученное противоречие показывает, что $\lambda_0^* > 0$. Неравенства (17) сохраняют силу, если (17) разделить на $\lambda_0^* > 0$. Поэтому в (17) можем считать $\lambda_0^* = 1$.

Наконец, возьмем произвольную точку $x \in X_0$. Тогда $a = (f(x), g_1(x), \ldots, g_m(x)) \in A$. Подставим эту точку в правое неравенство (17). С учетом того, что $\lambda_0^* = 1$, получим $f_* \leqslant f(x) + \sum_{i=1}^{m} \lambda_i^* g_i(x) = L(x, \lambda^*)$, $x \in X_0$. Но в силу (7) $f_* = L(x_*, \lambda^*)$ при любом выборе $x_* \in X_*$. Отсюда и из предыдущего неравенства следует условие (6). Согласно лемме 1 тогда (x_*, λ^*) — седловая точка. Теорема 2 доказана. □

Другую форму теоремы 2 читатель найдет в § 5.5.

3. Приведенные выше примеры 1, 2 показывают, что без дополнительных условий вида (15) теорема 2, вообще говоря, неверна. Однако, если (2) — многогранное множество (пример 1.6), то, оказывается, существование седловой точки функции Лагранжа выпуклой задачи (1), (2) можно доказать без каких-либо дополнительных условий. Это мы уже показали для общей задачи линейного программирования — см. теорему 3.5.5. Теперь рассмотрим более общий случай, не предполагая линейности целевой функции. Будем пользоваться следующим представлением многогранного множества

$$X = \{x \in E^n:\ x \in X_0,\ g_i(x) = \langle a_i, x\rangle - b_i \leqslant 0,\ i = 1, \ldots, m;\ g_i(x) = \langle a_i, x\rangle - b_i = 0,\ i = m+1, \ldots, s\}, \quad (18)$$

где X_0, в свою очередь, является многогранным множеством и задается в виде:

$$X_0 = \{x \in E^n:\ \langle d_i, x\rangle \leqslant r_i,\ i = 1, \ldots, p;\ \langle d_i, x\rangle = r_i,\ i = p+1, \ldots, q\},$$

$a_i, d_i \in E^n$ — заданные векторы, b_i, r_i — заданные числа. В частности, здесь возможно $X_0 = E^n$, $X_0 = E_+^n$, $X_0 = \{x = (x^1, \ldots, x^n):\ x^i \geqslant 0,\ i \in I\}$, I — некоторое подмножество номеров $\{1, \ldots, n\}$; $X_0 = \{x = (x^1, \ldots, x^n):\ \alpha_i \leqslant x^i \leqslant \beta_i,\ i = 1, \ldots, n\}$, α_i, β_i — заданные величины, $\alpha_i \leqslant \beta_i$, причем некоторые $\alpha_i = -\infty$, или $\beta_j = +\infty$.

Для многогранного множества несложно дать полное описание всех возможных направлений в любой его точке (определение 2.3).

Лемма 3. *Множество возможных направлений множества* (18) *в любой его точке x_* совпадает с конусом:*

$$K = K(x_*) = \{e \in E^n:\ e \neq 0,\ \langle a_i, e\rangle \leqslant 0,\ i \in I_1^*,\ \langle a_i, e\rangle = 0,\ i = m+1, \ldots, s,\ \langle d_i, e\rangle \leqslant 0,\ i \in I_2^*,\ \langle d_i, e\rangle = 0,\ i = p+1, \ldots, s\}, \quad (19)$$

где $I_1^* = \{i:\ 1 \leqslant i \leqslant m,\ \langle a_i, x_*\rangle = b_i\}$, $I_2^* = \{i:\ 1 \leqslant i \leqslant p,\ \langle d_i, x_*\rangle = r_i\}$.

Доказательство. Пусть $e = (e^1, \ldots, e^n) \neq 0$ — произвольное возможное направление множества (18) в точке x_*. Согласно определению 2.3 тогда существует такое число $t_0 > 0$, что $x = x_* + te \in X$ или

$$\begin{aligned} &\langle a_i, x_* + te\rangle \leqslant b_i,\ i = 1, \ldots, m;\ \langle a_i, x_* + te\rangle = b_i,\ i = m+1, \ldots, s;\\ &\langle d_i, x_* + te\rangle \leqslant r_i,\ i = 1, \ldots, p;\ \langle d_i, x_* + te\rangle = r_i,\ i = p+1, \ldots, q; \end{aligned} \quad (20)$$

при всех t, $0 < t \leqslant t_0$. С учетом того, что $x_* \in X$ и определения множеств I_1^*, I_2^* активных индексов точки x_* из (20), сразу получаем $e \in K$. Верно и

обратное: если $e \in K$, то e — возможное направление в точке x_*. В самом деле, пусть $e \in K$. Тогда для $i \in I_1^*$ имеем $\langle a_i, x_* + te\rangle = b_i + t\langle a_i, e\rangle \leqslant b_i$ при всех $t \geqslant 0$, а если $i \notin I_1^*$, $1 \leqslant i \leqslant m$, то $\langle a_i, x_*\rangle < b_i$ и найдется такое $t_0 > 0$, что $\langle a_i, x_* + te\rangle \leqslant b_i$ при $0 \leqslant t \leqslant t_0$. Если $m+1 \leqslant i \leqslant s$, то $\langle a_i, x_* + te\rangle = b_i$ при всех t. Аналогично, взяв при необходимости $t_0 > 0$ еще меньшим, убедимся, что выполняются и остальные соотношения (18), так что $x_* + te \in X$, $0 \leqslant t \leqslant t_0$. Лемма 3 доказана. □

Теорема 3. *Пусть множество решений X_* задачи* (1), (18) *непусто. Пусть $f(x)$ выпукла на X_0 и дифференцируема в точке $x_* \in X_*$. Тогда существуют множители Лагранжа $\lambda^* = (\lambda_1^*, \ldots, \lambda_s^*) \in \Lambda_0 = \{\lambda = (\lambda_1, \ldots, \lambda_s) \in E^s:\ \lambda_1 \geqslant 0, \ldots, \lambda_m \geqslant 0\}$, такие, что пара $(x*, \lambda^*)$ образует седловую точку функции Лагранжа задачи* (1), (18).

Доказательство. Согласно теореме 2.3 для того, чтобы $x_* \in X_*$, необходимо и достаточно выполнения неравенства

$$\langle f'(x_*), x - x_*\rangle \geqslant 0 \quad \forall x \in X. \quad (21)$$

Возьмем любое $e \in K$. Тогда по лемме 3 $x = x_* + te \in X$, $0 \leqslant t \leqslant t_0$, $t_0 > 0$. Подставим такую точку x в (21). Получим: $\langle f'(x_*), e\rangle t \geqslant 0$ или $\langle f'(x_*), e\rangle \geqslant 0$ при всех $e \in K$. По теореме Фаркаша 3.5.8 тогда найдутся числа $\lambda_i^* \geqslant 0$, $i \in I_1^*$, $\lambda_{m+1}^*, \ldots, \lambda_s^*$, $\mu_i^* \geqslant 0$, $i \in I_2*$, $\mu_{p+1}^*, \ldots, \mu_q^*$ такие, что

$$f'(x_*) = -\sum_{i \in I_1^*} \lambda_i^* a_i - \sum_{i=m+1}^{s} \lambda_i^* a_i - \sum_{i \in I_2^*} \mu_i^* d_i - \sum_{i=p+1}^{q} \mu_i^* d_i \quad (22)$$

Если доопределим $\lambda_i^* = 0$ при $i \in \{1, \ldots, m\} \setminus I_1^*$, то получим точку $\lambda^* = (\lambda_1^*, \ldots, \lambda_s^*) \in \Lambda_0$. Отсюда, учитывая определение множества I_1^* и условие $x_* \in X_* \subseteq X$, имеем

$$\lambda_i^*(\langle a_i, x_*\rangle - b_i) = \lambda_i^* g_i(x_*) = 0, \quad i = 1, \ldots, s, \quad (23)$$

а равенство (22) можем переписать в виде

$$f'(x_*) + \sum_{i=1}^{s} \lambda_i^* a_i = -\sum_{i \in I_2^*} \mu_i^* d_i - \sum_{i=p+1}^{q} \mu_i^* d_i. \quad (24)$$

Функция Лагранжа в рассматриваемой задаче (1), (18) такая:

$$L(x, \lambda) = f(x) + \sum_{i=1}^{s} \lambda_i(\langle a_i, x\rangle - b_i), \quad x \in X_0, \quad \lambda \in \Lambda_0.$$

Тогда, используя неравенство $f(x) - f(x_*) \geqslant \langle f'(x_*), x - x_*\rangle$, $x \in X$ (теорема 2.2), определение множества I_2^*, условие $\mu_i^* \geqslant 0$, $i \in I_2^*$, и равенство (24), для каждого $x \in X_0$ получаем

$$\begin{aligned} L(x, \lambda^*) - L(x_*, \lambda^*) &= f(x) - f(x_*) + \sum_{i=1}^{s} \lambda_i^* \langle a_i, x - x_*\rangle \geqslant\\ &\geqslant \langle f'(x_*) + \sum_{i=1}^{s} \lambda_i^* a_i, x - x_*\rangle = -\sum_{i \in I_2^*} \mu_i^* \langle d_i, x - x_*\rangle -\\ &\quad - \sum_{i=p+1}^{q} \mu_i^* \langle d_i, x - x_*\rangle = -\sum_{i \in I_2^*} \mu_i^*(\langle d_i, x\rangle - r_i) \geqslant 0, \end{aligned}$$

или

$$L(x_*, \lambda^*) \leqslant L(x, \lambda^*) \qquad \forall x \in X_0.$$

Отсюда и из (23) с помощью леммы 1 заключаем, что (x_*, λ^*) — седловая точка функции Лагранжа. Теорема 3 доказана. □

Замечание 1. Если $f(x) = \langle c, x\rangle$, то из теоремы 3 вытекает теорема 3.5.5 для задач линейного программирования. Однако принятая здесь схема изложения не позволяет считать теорему 3.5.5 следствием теоремы 3, так как при доказательстве теоремы 3 была существенно использована теорема Фаркаша 3.5.8, которая в свою очередь (как, впрочем, и сама теорема 3.5.5) получена как следствие доказанных в § 3.5 утверждений.

Замечание 2. Условие дифференцируемости функции $f(x)$ в точке $x_* \in X_*$ в теореме 3 можно заменить условием непустоты субдифференциала $\partial f(x_*)$, считая, например, функцию $f(x)$ выпуклой на открытом выпуклом множестве W, $X_0 \subset W$ (теорема 6.1). Доказательство теоремы 3 в этом случае полностью сохраняет силу, если в нем вектор $f'(x_*)$ заменить на субградиент $c_* \in \partial f(x_*)$, взятый из условия (6.8), а ссылки на теоремы 2.3, 2.2 заменить соответственно ссылками на теорему 6.4 и определение 6.1 субградиента.

Для иллюстрации теоремы 3 рассмотрим задачу определения проекции точки на множество (18) при $X_0 = E^n$, $m = 0$.

Пример 3. Задача: $f(x) = \frac{1}{2}|x - z|^2 \to \inf$, $x \in X = \{x \in E^n\colon\ Ax = b\}$, где A — матрица размера $m \times n$, $b \in E^m$, z — произвольная точка из E^n. Согласно теореме 4.1 эта задача имеет и притом единственное решение $x_* = \mathcal{P}_X(z)\ \ \forall z \in E^n$. Функция Лагранжа этой задачи $L(x, \lambda) = \frac{1}{2}|x - z|^2 + \langle \lambda, Ax - b\rangle = \frac{1}{2}|x - z|^2 + \langle x, A^\top\lambda\rangle - \langle b, \lambda\rangle$. По теореме 3 функция $L(x, \lambda)$ имеет седловую точку $(x_*, \lambda^*) \in E^n \times E^m$ и условия (6), (7) приводят к системе

$$L_x(x_*, \lambda) = (x_* - z) + A^\top\lambda = 0, \quad Ax_* = b.$$

Отсюда для проекции $x_* = \mathcal{P}_X(z)$ точки z на множество X получаем представление $x_* = z - A^\top\lambda$, где λ — произвольное решение системы линейных алгебраических уравнений $AA^\top\lambda = Az - b$. Если квадратная матрица $AA^\top$ невырожденная, то получаем уже известную нам формулу для проекции из примера 4.3.

4. Наконец, приведем еще один вариант теоремы Куна — Таккера.

Теорема 4. *Пусть в задаче* (1), (2) X_0 *— многогранное множество из* E^n, *функции* $f(x)$, $g_i(x)$, $i = 1, \ldots, m$, *выпуклы на открытом выпуклом множестве* W, *содержащем* X_0, $g_i(x) = \langle a_i, x\rangle - b^i$, $i = m+1, \ldots, s$; *существует точка* $\bar{x} \in X$ *такая, что* $g_i(\bar{x}) < 0$, $i = 1, \ldots, m$; *множество* X_* *решений задачи* (1), (2) *непусто. Тогда для каждой точки* $x_* \in X_*$ *существуют множители Лагранжа* $\lambda^* = (\lambda_1^*, \ldots, \lambda_s^*) \in \Lambda_0 = \{\lambda \in E^s\colon \lambda_1 \geqslant 0, \ldots, \lambda_m \geqslant 0\}$ *такие, что пара* (x_*, λ^*) *образует седловую точку функции Лагранжа* (3).

Доказательство проведем, пользуясь приемом, изложенным в [239]. Заметим, что каждая точка $x_* \in X_*$ является решением задачи

$$f(x) \to \inf, \quad x \in X = \{x \in \Gamma_0\colon\ g_i(x) \leqslant 0,\ i = 1, \ldots, m\},$$

где $\Gamma_0 = \{x \in X_0\colon\ g_i(x) = 0,\ i = m+1, \ldots, s\}$. Согласно теореме 2 в этой задаче функция Лагранжа $L_1(x, u) = f(x) + \sum\limits_{i=1}^{m} u_i g_i(x)$, $x \in \Gamma_0$, $u = (u^1, \ldots, u^m) \geqslant 0$ имеет седловую точку (x_*, u^*), т. е.

$$L_1(x_*, u^*) \leqslant L_1(x, u^*)\ \ \forall x \in \Gamma_0; \quad u_i^* g_i(x_*) = 0, \quad i = 1, \ldots, m; \quad u^* \geqslant 0.$$

Отсюда следует, что x_* является решением задачи

$$L_1(x, u^*) \to \inf, \quad x \in \Gamma_0.$$

Функция Лагранжа $L_2(x, v) = L_1(x, u^*) + \sum\limits_{i=m+1}^{s} v_i g_i(x)$, $x \in X_0$, $v \in E^{s-m}$, последней задачи в силу теоремы 3 и замечания 2 к ней имеет седловую точку (x_*, v^*), что, согласно лемме 1, означает:

$$L_2(x_*, v^*) \leqslant L_2(x, v^*)\ \ \forall x \in X_0; \quad v_i^* g_i(x_*) = 0, \quad i = m+1, \ldots, s.$$

Положим $\lambda = (u, v)$, $\lambda^* = (u^*, v^*)$. Функция Лагранжа задачи (1), (2) тогда представима в виде $L(x, \lambda) = f(x) + \sum\limits_{i=1}^{s} \lambda_i g_i(x) = L_1(x, u) + \sum\limits_{i=m+1}^{s} v_i g_i(x)$, $x \in X_0$, $\lambda \in \Lambda_0$. Из предыдущих рассуждений следует, что

$$L(x_*, \lambda^*) = L_1(x_*, u^*) + \sum_{i=m+1}^{s} v_i^* g_i(x_*) = L_2(x_*, v^*) \leqslant L_2(x, v^*) =$$
$$= L_1(x, u^*) + \sum_{i=m+1}^{s} v_i^* g_i(x) = L(x, \lambda^*)\ \ \forall x \in X_0,$$
$$u_i^* g_i(x_*) = 0, \quad i = 1, \ldots, m; \quad v_i^* g_i(x_*) = 0, \quad i = m+1, \ldots, s; \quad x_* \in X.$$

Отсюда и из леммы 1 вытекает утверждение теоремы 4. □

Пример 4. Пусть $X_0 = \{u = (x, y) \in E^2\colon\ x \geqslant 0,\ y \geqslant 0\} = E^2_+$, $f(u) = -\sqrt{xy}$, $g(u) = x$, $X = \{u \in X_0\colon\ g(u) \leqslant 0\}$. Здесь X_0 выпукло, $f(u)$, $g(u)$ выпуклы на X_0, $f_* = 0$, $X_* = X = \{u_* = (0, y),\ y \geqslant 0\}$. Функция Лагранжа $L(u, \lambda) = -\sqrt{xy} + \lambda x$, $x \geqslant 0$, $y \geqslant 0$, $\lambda \geqslant 0$, не имеет седловой точки. Нарушены условия теоремы 4: $f(u)$ выпукла лишь на X_0, требуемой точки $\bar{x}$ нет.

Теоремы 2, 3, 4 дают достаточные условия существования седловой точки в задачах выпуклого программирования. Однако существуют и невыпуклые задачи, в которых функция Лагранжа имеет седловую точку.

Пример 5. Пусть $X_0 = \{u \in E^1\colon\ u \leqslant 1\}$, $f(u) = u^3$, $g(u) = -u^3 - 1$, $X = \{u\colon\ u \in X_0,\ g(u) \leqslant 0\}$. Здесь X_0 выпукло, но функции $f(u)$, $g(u)$ не являются выпуклыми на X_0. Множество X представляет собой отрезок $-1 \leqslant u \leqslant 1$, так что $f_* = -1$, $u_* = -1$. Функция Лагранжа $L(u, \lambda) = u^3 + \lambda(-u^3 - 1)$ имеет единственную седловую точку $(u_* = -1,\ \lambda^* = 1)$ на множестве $X_0 \times \Lambda_0$, $\Lambda_0 = \{\lambda \in E^1\colon\ \lambda \geqslant 0\}$.

5. С помощью функции Лагранжа $L(x, \lambda)$, $x \in X_0$, $\lambda \in \Lambda_0$, задачу (1), (2) можно переформулировать следующим образом. Введем функцию

$$\nu(x) = \sup_{\lambda \in \Lambda_0} L(x, \lambda), \quad x \in X_0. \tag{25}$$

Вычислим верхнюю грань в правой части этой формулы. Если $x \in X$, то $\sum\limits_{i=1}^{s} \lambda_i g_i(x) \leqslant 0\ \ \forall \lambda \in \Lambda_0$, причем равенство здесь реализуется при $\lambda = 0 \in \Lambda_0$. Если же $x_0 \in X_0 \setminus X$, то найдется номер i такой, что либо $1 \leqslant i \leqslant m$ и $g_i(x) > 0$, либо $m+1 \leqslant i \leqslant s$ и $g_i(x) \neq 0$, так что подходящим выбором $\lambda \in \Lambda_0$ сумму $\sum\limits_{i=1}^{s} \lambda_i g_i(x)$ можно сделать сколь угодно большой. Следовательно, функция (25) равна

$$\nu(x) = \begin{cases} f(x) & \forall x \in X, \\ +\infty & \forall x \in X_0 \setminus X. \end{cases}$$

Отсюда ясно, что $\inf\limits_{x\in X_0}\nu(x)=\inf\limits_{x\in X}f(x)=f_*$, и задачу (1), (2) можно переписать в равносильном виде

$$\nu(x)\to\inf,\quad x\in X_0. \tag{26}$$

Как и выше, в задаче (1), (2) будем предполагать, что $f_*>-\infty$, $X_*\neq\varnothing$. Тогда задача (26) будет иметь то же множество решений X_* с тем же минимальным значением $f_*>-\infty$, т. е.

$$\inf_{x\in X_0}\nu(x)=\nu_*=f_*>-\infty,\quad X_*=\{x\colon x\in X_0,\ \nu(x)=\nu_*=f_*\}\neq\varnothing. \tag{27}$$

Читатель, конечно, заметил, что переменные $(x,\lambda)\in X_0\times\Lambda_0$ в определении 1 седловой точки входят симметрично и, надо полагать, его уже озарила мысль о том, что переменные λ в неравенствах (5), по-видимому, относятся к какой-то неизвестной экстремальной задаче, а точка λ^* является ее решением. Действительно, это так и есть. Более того, переход от исходной задачи (1), (2) к задаче (26), осуществленный с помощью функции Лагранжа, наводит на симметричные действия по поиску формулировки этой таинственной задачи. А именно, по аналогии с (25) введем функцию

$$\psi(\lambda)=\inf_{x\in X_0}L(x,\lambda),\quad \lambda\in\Lambda_0, \tag{28}$$

и рассмотрим задачу

$$\psi(\lambda)\to\sup,\quad \lambda\in\Lambda_0. \tag{29}$$

Задачу (29) называют *двойственной задачей* к задаче (26) или к исходной основной задаче (1), (2), а переменные $\lambda=(\lambda_1,\ldots,\lambda_s)$ называют *двойственными переменными* в отличие от исходных, основных переменных $x=(x^1,\ldots,x^n)$.

При формулировке задач вида (1), (2) мы обычно подразумеваем, что функции $f(x)$, $g_i(x)$, $i=1,\ldots,s$, принимают конечные значения во всех точках множества X_0. Поэтому $\psi(\lambda)<+\infty$ при любых $\lambda\in\Lambda_0$. Но формула (28) не исключает возможность появления значений $\psi(\lambda)=-\infty$ при некоторых $\lambda\in\Lambda_0$. Имея в виду это обстоятельство, задачу (29) нетрудно записать в привычной форме

$$\psi(\lambda)\to\sup,\quad \lambda\in\Lambda=\{\lambda\in E^s\colon \lambda\in\Lambda_0,\ \psi(\lambda)>-\infty\}, \tag{30}$$

используя лишь конечные значения функции $\psi(\lambda)$. Обозначим

$$\sup_{\lambda\in\Lambda_0}\psi(\lambda)=\sup_{\lambda\in\Lambda}\psi(\lambda)=\psi^*,\quad \Lambda^*=\{\lambda\in\Lambda_0\colon \psi(\lambda)=\psi^*\}=\{\lambda\in\Lambda\colon \psi(\lambda)=\psi^*\}. \tag{31}$$

Важно заметить, что двойственная задача (29) или (30) равносильна задаче выпуклого программирования независимо от того, является ли исходная задача (1), (2) выпуклой или нет. В самом деле, функция $(-L(x,\lambda))$ линейна и, следовательно, выпукла по λ на выпуклом множестве Λ_0, и по теореме 2.7 функция $(-\psi(\lambda))=\sup\limits_{x\in X_0}(-L(x,\lambda))$ выпукла на Λ_0. Иначе говоря, функция $\psi(\lambda)$ вогнута на Λ_0, т. е. $\psi(\alpha\lambda_1+(1-\alpha)\lambda_2)\geqslant\alpha\psi(\lambda_1)+(1-\alpha)\psi(\lambda_2)$ $\forall\alpha\in[0,1]$ $\forall\lambda_1,\lambda_2\in\Lambda_0$. Отсюда видно, что если $\lambda_1,\lambda_2\in\Lambda_0$ и $\psi(\lambda_1)>-\infty$, $\psi(\lambda_2)>-\infty$, то и $\psi(\alpha\lambda_1+(1-\alpha)\lambda_2)>-\infty$ $\forall\alpha\in[0,1]$. Это значит, что множество Λ в (30) также выпукло. Следовательно, задачи (29), (30), записанные в равносильном виде

$$-\psi(\lambda)\to\sup,\quad \lambda\in\Lambda_0\quad\text{или}\quad \lambda\in\Lambda, \tag{32}$$

представляют собой задачи выпуклого программирования. Благодаря этому обстоятельству исследовать двойственную задачу нередко бывает проще, чем исходную. Возникает вопрос, зачем это делать? Какую информацию мы можем получить об исходной задаче (1), (2), изучая двойственную задачу? Оказывается, задачи (26) и (29) и, следовательно, задачи (1), (2) и (30) тесно связаны между собой и параллельное изучение этих задач зачастую бывает плодотворным, позволяет полнее изучить каждую из них, наметить новые подходы к их решению.

Прежде всего можно отметить неравенства:

$$\psi(\lambda)\leqslant\psi^*\leqslant f_*\leqslant\nu(x)\quad\forall x\in X_0,\quad\forall\lambda\in\Lambda_0. \tag{33}$$

В самом деле, из (25), (28) имеем $\psi(\lambda)=\inf\limits_{x\in X_0}L(x,\lambda)\leqslant L(x,\lambda)$ $\forall\lambda\in\Lambda_0$, $\forall x\in X_0$, поэтому $\psi^*=\sup\limits_{\lambda\in\Lambda_0}\psi(\lambda)\leqslant\sup\limits_{\lambda\in\Lambda_0}L(x,\lambda)=\nu(x)$ $\forall x\in X_0$. Переходя к нижней грани по $x\in X_0$ в этом неравенстве, получаем $\psi^*\leqslant f_*$, откуда следуют неравенства (33).

Интересно выяснить, когда $\psi^*=f_*$, и обе задачи (26) и (29) имеют решение, т. е.

$$X_*\neq\varnothing,\quad \Lambda^*\neq\varnothing,\quad f_*=\psi^*. \tag{34}$$

Оказывается, соотношения (34) тесно связаны с седловой точкой функции Лагранжа.

Теорема 5. *Для того чтобы имели место соотношения* (34), *необходимо и достаточно, чтобы функция* $L(x,\lambda)$, $x\in X_0$, $\lambda\in\Lambda_0$, *имела седловую точку на* $X_0\times\Lambda_0$ *в смысле определения 1. Множество седловых точек функции* $L(x,\lambda)$ *на* $X_0\times\Lambda_0$ *совпадает с множеством* $X_*\times\Lambda^*$.

Доказательство. Необходимость. Пусть выполнены соотношения (34). Возьмем произвольные $x_*\in X_*$ и $\lambda^*\in\Lambda^*$ и покажем, что (x_*,λ^*) — седловая точка. Имеем

$$\psi^*=\psi(\lambda^*)=\inf_{u\in X_0}L(u,\lambda^*)\leqslant L(x_*,\lambda^*)\leqslant\sup_{\lambda\in\Lambda_0}L(x_*,\lambda)=\nu(x_*)=f_*.$$

По условию $\psi^*=f_*$. Поэтому предыдущие неравенства превращаются в равенства:

$$L(x_*,\lambda^*)=\inf_{u\in X_0}L(u,\lambda^*)=\sup_{\lambda\in\Lambda_0}L(x_*,\lambda)=f_*.$$

Отсюда имеем неравенства (5), т. е. (x_*,λ^*) — седловая точка. Тем самым показано, что $X_*\times\Lambda^*$ принадлежит множеству седловых точек функции $L(x,\lambda)$ на $X_0\times\Lambda_0$.

Достаточность. Пусть $(x_*,\lambda^*)\in X_0\times\Lambda_0$ — седловая точка функции $L(x,\lambda)$ на $X_0\times\Lambda_0$. Согласно (5) это значит, что $L(x_*,\lambda)\leqslant L(x_*,\lambda^*)$, $\lambda\in\Lambda_0$. Отсюда имеем

$$\sup_{\lambda\in\Lambda_0}L(x_*,\lambda)=\nu(x_*)=L(x_*,\lambda^*).$$

Кроме того, $L(x_*,\lambda^*)\leqslant L(x,\lambda^*)$, $x\in X_0$, так что

$$L(x_*,\lambda^*)=\inf_{u\in X_0}L(u,\lambda^*)=\psi(\lambda^*),$$

откуда и из неравенств (33) следует

$$L(x_*,\lambda^*)=\psi(\lambda^*)\leqslant\psi^*\leqslant f_*\leqslant\nu(x_*)=L(x_*,\lambda^*),$$

т. е. $\psi(\lambda^*)=\psi^*=f_*=\nu(x_*)$. Это значит, что $\psi^*=f_*$, $\lambda^*\in\Lambda^*$, $x_*\in X_*$. Тем самым установлено, что множество седловых точек функции $L(x,\lambda)$ на $X_0\times\Lambda_0$ принадлежит множеству $X_*\times\Lambda^*$. Теорема 5 доказана. □

Следствие 1. *Следующие четыре утверждения равносильны:*
1) $(x_*,\lambda^*)\in X_0\times\Lambda_0$ *— седловая точка функции* $L(x,\lambda)$ *на* $X_0\times\Lambda_0$;
2) *выполняются соотношения* (34);
3) *существуют точки* $x_*\in X_0$, $\lambda^*\in\Lambda_0$ *такие, что*

$$\nu(x_*)=\psi(\lambda^*);$$

4) *справедливо равенство*

$$\max_{\lambda\in\Lambda_0}\inf_{u\in X_0}L(u,\lambda)=\min_{u\in X_0}\sup_{\lambda\in\Lambda_0}L(u,\lambda)$$

(напоминаем, что когда пишут max *или* min, *то достижение соответствующей верхней или нижней грани предполагается).*

Следствие 2. *Если* (x_*,λ^*) *и* $(a_*,b^*)\in X_0\times\Lambda_0$ *— седловые точки функции* $L(x,\lambda)$ *на* $X_0\times\Lambda_0$, *то* (x_*,b^*), (a_*,λ^*) *также являются седловыми точками этой функции на* $X_0\times\Lambda_0$, *причем*

$$L(x_*,b^*)=L(a_*,\lambda^*)=L(x_*,\lambda^*)=L(a_*,b^*)=f_*=\psi^*.$$

Отсюда и из теоремы 1 вытекает, что в теоремах 2, 3, 4 можно выбрать одни и те же множители Лагранжа λ^* для всех $x_*\in X_*$ сразу.

Полезно заметить, что в доказательстве теоремы 5 нигде не использовано то, что $L(x,\lambda)$ является функцией Лагранжа какой-либо задачи вида (1), (2), а множества X_0, Λ_0 выпуклы — там были важны лишь функции (25), (28), задачи (26), (29) и множества X_*, Λ^* из (27), (31), которые могут быть введены для любой функции $L(x,\lambda)$ на любых множествах X_0, Λ_0. Это значит, что теорема 5 и следствие 1, 2 к ней верны для произвольных функций $L(x,\lambda)$ и множеств X_0, Λ_0.

В приводимых ниже примерах иллюстрируются различные свойства двойственной задачи.

Пример 6. Задача: $f(x)=-x\to\inf$, $x\in X=\{x\in E^1\colon\ x\geqslant 0,\ g(x)=x^2\leqslant 0\}$. Здесь $f_*=0$, $X_*=X=\{0\}$. Задача выпуклая. В примере 1 было замечено, что функция Лагранжа $L(x,\lambda)=-x+\lambda x^2$, $x\in X_0=\{x\geqslant 0\}$, $\lambda\in\Lambda_0=\{\lambda\geqslant 0\}$ не имеет седловой точки. Функция $\psi(\lambda)=\inf\limits_{x\geqslant 0}L(x,\lambda)=-\frac{1}{4\lambda}$ при $\lambda>0$ и $\psi(0)=-\infty$. Двойственная задача (30) имеет вид: $\psi(\lambda)=-\frac{1}{4\lambda}\to\sup$, $\lambda\in\Lambda=\{\lambda>0\}$. Множество Λ — открытое, $\psi^*=f^*=0$, но $\Lambda^*=\varnothing$, т. е. двойственная задача не имеет решения. Множество X не имеет внутренних точек.

Пример 7. В задаче из примера 2 функция $\psi(\lambda)=\inf\limits_{x\in E^1}L(x,\lambda)=-\frac{1}{4\lambda}$ при $\lambda>0$, $\psi(0)=-\infty$. Как видим, двойственная задача здесь полностью совпадает с такой же задачей из примера 6, хотя исходные задачи разные. Здесь $\operatorname{int}X\neq\varnothing$, $\psi^*=f^*=0$, $X_*=\{0\}$, $\Lambda^*=\varnothing$.

Пример 8. Рассмотрим задачу из примера 4. Здесь $L(u,\lambda)=-\sqrt{xy}+\lambda x$, $u\in X_0=E^2_+$, $\lambda\in\Lambda_0=E^1_+$. Функция $\psi(\lambda)=-\infty$ при всех $\lambda\in\Lambda_0$, так что в двойственной задаче (30) множество $\Lambda=\varnothing$.

Пример 9. Задача: $f(x)=e^{-x}\to\inf$, $x\in X=\{x\in E^1,\ g(x)=xe^{-x}=0\}$. Множество X состоит из единственной точки $x=0$, так что $f_*=f(0)=1$, $X_*=\{0\}$. Здесь функция Лагранжа $L(x,\lambda)=e^{-x}+\lambda xe^{-x}$, $x\in X_0=E^1$, $\lambda\in\Lambda_0=E^1$; $\psi(\lambda)=-\infty$ при $\lambda>0$, $\psi(\lambda)=0$ при $\lambda=0$, $\psi(\lambda)=\lambda\exp\left(-1+\frac{1}{\lambda}\right)$ при $\lambda<0$. Множество $\Lambda=\{\lambda\leqslant 0\}$ замкнуто, функция $\psi(\lambda)$ непрерывна на Λ, $\psi^*=0$, $\Lambda^*=\{0\}$. Таким образом, здесь $X_*\neq\varnothing$, $\Lambda^*\neq\varnothing$, но $\psi^*<f_*$. Согласно теореме 5 функция $L(x,\lambda)$ не имеет седловой точки.

Не следует думать, что если $(x_*,\lambda^*)\in X_0\times\Lambda_0$ — седловая точка функции $L(x,\lambda)$, то и точки $(a,b)\in X_0\times\Lambda_0$, для которых $L(a,b)=L(x_*,\lambda^*)$, также будут седловыми точками. В общем случае можно лишь утверждать, что

$$X_*\subseteq X(\lambda^*)=\{x\in X_0\colon L(x,\lambda^*)=L(x_*,\lambda^*)\},$$
$$\Lambda^*\subseteq\Lambda(x_*)=\{\lambda\in\Lambda_0\colon L(x_*,\lambda)=L(x_*,\lambda^*)\}\quad(35)$$

где множества X_*, Λ^* взяты из (27), (31).

Пример 10. Функция $L(x,\lambda)=\lambda x$, $x\in X_0=E^1$, $\lambda\in\Lambda_0=E^1$, имеет единственную седловую точку ($x_*=0$, $\lambda^*=0$), $L(0,0)=0$. Здесь $X(\lambda^*)=E^1$, $\Lambda(x_*)=E^1$, и, как видим, включения (35) являются строгими. Далее, функции $\nu(x)$, $\psi(\lambda)$ из (25), (28) соответственно равны $\nu(x)=+\infty$ при $x\neq 0$, $\nu(0)=0$, и $\psi(\lambda)=-\infty$ при $\lambda\neq 0$, $\psi(0)=0$, так что оба множества $X=X_*=\{0\}$, $\Lambda=\Lambda^*=\{0\}$ являются одноточечными.

6. Напоминаем, что в главе 3 мы уже рассматривали двойственную задачу для задачи линейного программирования, причем двойственная задача была введена по определению, без объяснения, откуда она появилась. Убедимся, что введенная в § 3.5 двойственная задача является частным случаем задачи (30).

Рассмотрим общую задачу линейного программирования:

$$f(x)=\langle c_1,x_1\rangle+\langle c_2,x_2\rangle\inf,\qquad x\in X,\qquad(36)$$

$$X=\{x=(x_1,x_2)\in E^{n_1}\times E^{n_2}\colon x_1\geqslant 0,$$
$$A_{11}x_1+A_{12}x_2-b_1\leqslant 0,\ A_{21}x_1+A_{22}x_2-b_2=0\},\qquad(37)$$

где $c_1\in E^{n_1}$, $c_2\in E^{n_2}$, $b_1\in E^{m_1}$, $b_2\in E^{m_2}$ — заданные векторы, матрицы A_{ij} также заданы и имеют размерность $m_i\times n_j$, $i,j=1,2$. Функция Лагранжа этой задачи:

$$L(x,\lambda)=\langle c_1,x_1\rangle+\langle c_2,x_2\rangle+\langle\lambda_1,A_{11}x_1+A_{12}x_2-b_1\rangle+$$
$$+\langle\lambda_2,A_{21}x_1+A_{22}x_2-b_2\rangle=\langle c_1,x_1\rangle+\langle c_2,x_2\rangle+$$
$$+\sum_{i=1}^{m_1}\lambda_1^i(A_{11}x_1+A_{12}x_2-b_1)^i+\sum_{i=1}^{m_2}\lambda_2^i(A_{21}x_1+A_{22}x_2-b_2)^i,$$
$$x=(x_1,x_2)\in X_0=E^{n_1}_+\times E^{n_2},\quad\lambda=(\lambda_1,\lambda_2)\in\Lambda_0=E^{m_1}_+\times E^{m_2}.$$

Отсюда нетрудно видеть, что функция $\nu(x)=\sup\limits_{\lambda\in\Lambda_0}L(x,\lambda)$, $x\in X_0$, определяемая согласно (25), в случае задачи (36), (37) имеет вид:

$$\nu(x)=\begin{cases}<c_1,x_1>+<c_2,x_2> & \text{при}\quad x\in X,\\ +\infty & \text{при}\quad x\in X_0\setminus X.\end{cases}\qquad(38)$$

Для вычисления функции $\psi(\lambda)$, определяемой по формуле (28), удобнее представить функцию Лагранжа $L(x, \lambda)$ в следующем виде

$$L(x,\lambda)=\langle -b_1,\lambda_1\rangle+\langle -b_2,\lambda_2\rangle+\langle x_1, A_{11}^{\top}\lambda_1+A_{21}^{\top}\lambda_2+c_1\rangle+\langle x_2, A_{12}^{\top}\lambda_1+A_{22}^{\top}\lambda_2+c_2\rangle=$$
$$=-\langle b_1,\lambda_1\rangle-\langle b_2,\lambda_2\rangle+\sum_{j=1}^{n_1}x_1^j(A_{11}^{\top}\lambda_1+A_{21}^{\top}\lambda_2+c_1)^j+$$
$$+\sum_{j=1}^{n_2}x_2^j(A_{12}^{\top}\lambda_1+A_{22}^{\top}\lambda_2+c_2)^j,\quad x\in X_0,\quad \lambda\in\Lambda_0.$$

Отсюда следует, что

$$\psi(\lambda)=\begin{cases}-<b_1,\lambda_1>-<b_2,\lambda_2> & \text{при}\quad \lambda\in\Lambda,\\ -\infty & \text{при}\quad \lambda\in\Lambda_0\setminus\Lambda,\end{cases}\tag{39}$$

где $\Lambda=\{\lambda=(\lambda_1,\lambda_2)\in E^{m_1}\times E^{m_2}\colon \lambda_1\geqslant 0,\ A_{11}^{\top}\lambda_1+A_{21}^{\top}\lambda_2+c_1\geqslant 0,\ A_{12}^{\top}\lambda_1+A_{22}^{\top}\lambda_2+c_2=0\}$. Из полученных выражений (38), (39) для функций $\nu(x)$, $\psi(\lambda)$ следует, что задача (26): $\nu(x)\to\inf$, $x\in X_0$, равносильна исходной задаче (36), (37), а двойственная к ней задача (29): $\psi(\lambda)\to\sup$, $\lambda\in\Lambda_0$, или (30) равносильна задаче

$$\psi(\lambda)=-\langle b_1,\lambda_1\rangle-\langle b_2,\lambda_2\rangle\to\sup\quad(\text{или}\quad(-\psi(\lambda))\to\inf),\quad\lambda\in\Lambda.\tag{40}$$

Как видим, именно задача (40) в § 3.5 была по определению названа двойственной к (36), (37) задачей. Сравнивая утверждения, доказанные в этом параграфе, с теоремами из § 3.5, можем сделать вывод, что развитые здесь элементы теории двойственности являются прямым обобщением теории, изложенной в § 3.5, на случай нелинейных задач. Можно также заметить, что не все утверждения, справедливые для задач линейного программирования, допускают обобщения на нелинейные выпуклые задачи. Так, например, нельзя утверждать, что задача, двойственная к двойственной задаче (29), в нелинейных задачах также может быть приведена к виду, совпадающему с исходной задачей (1), (2). Для невыпуклых задач это очевидно, так как двойственная задача всегда равносильна задаче выпуклого программирования (32), и потому задача двойственная к двойственной, могла бы совпасть с исходной лишь тогда, если бы она была выпуклой. Однако требование выпуклости задачи здесь также не спасает положение, что видно из примеров 6, 7, в которых двойственные задачи совпадают, а двойственная к последней не может совпасть с исходной задачей, так как исходные задачи в этих примерах разные. Ничего не меняет здесь и требование существования седловой точки функции Лагранжа, о чем свидетельствует следующий пример.

Пример 11. Задача: $f(x)=|x|^2\to\inf$, $x\in X=\{x\in E^n\colon g(x)=|x|^2-1\leqslant 0\}$. Здесь $X_0=E^n$, $f_*=0$, $x_*=0$, $\Lambda_0=\{\lambda\in E^1\colon\lambda\geqslant 0\}$. Функция Лагранжа $L(x,\lambda)=|x|^2+\lambda(|x|^2-1)=(1+\lambda)|x|^2-\lambda$, $x\in X_0$, $\lambda\in\Lambda_0$. Ясно, что $\psi(\lambda)=\inf\limits_{x\in E^n}L(x,\lambda)=-\lambda$ при всех $\lambda\in\Lambda_0$. Таким образом, двойственная задача имеет вид $\psi(\lambda)=-\lambda\to\sup$, $\lambda\in\Lambda=\Lambda_0$. Эта задача линейного программирования. Двойственная к ней задача также будет задачей линейного программирования и не может совпасть с исходной. Остается заметить, рассматриваемая задача выпукла, и функция Лагранжа в ней имеет седловую точку ($x_*=0$, $\lambda^*=0$).

Заметим, что теорема 3.5.2 об одновременной разрешимости исходной и двойственной задач также специфична для задач линейного программирования и не может быть обобщена на нелинейные задачи даже при условии их выпуклости — см. примеры 6–8.

7. Кратко остановимся еще на одном интересном классе задач, называемых задачами *геометрического программирования*, в которых переход к двойственной задаче весьма плодотворен. Речь идет о задачах минимизации следующего вида:

$$g(x)=\sum_{i=1}^{n}c_i x_1^{a_{i1}}\dots x_r^{a_{ir}}\to\inf;\quad x\in X=\operatorname{int}E_+^r,\tag{40}$$

где $c_i>0$, a_{ij} — заданные числа, $\operatorname{int}E_+^r=\{x=(x_1,\dots,x_r)\colon x_1>0,\dots,x_r>0\}$. Функция $g(x)$ из (40) называется *позиномом*.

Для исследования задачи (40) удобнее перейти к новым переменным $u=(u_1,\dots,u_{r+n})$ по формулам

$$u_i=\ln x_i,\quad i=1,\dots,r;\quad u_{r+1}=-b_i+\sum_{j=1}^{r}a_{ij}u_j,\quad b_i=-\ln c_i,\quad i=1,\dots,n,\tag{41}$$

и переписать ее в эквивалентном виде:

$$f(u)=\sum_{i=1}^{n}e^{u_{r+i}}\to\inf;$$
$$u\in U=\Big\{u\in E^{r+n}\colon\sum_{j=1}^{r}a_{ij}u_j-u_{r+i}-b_i=0,\ i=1,\dots,n\Big\}.\tag{42}$$

Отметим, что функция $f(u)$ выпукла на E^{r+n}, U — многогранное множество, и поэтому к задаче (42) применима теорема 3. Составим функцию Лагранжа (3) для этой задачи:

$$L(u,\lambda)=\sum_{i=1}^{n}e^{u_{r+i}}+\sum_{i=1}^{n}\lambda_i\Big(\sum_{j=1}^{r}a_{ij}u_j-u_{r+i}-b_i\Big)=$$
$$=\sum_{i=1}^{n}\Big(e^{u_{r+i}}-\lambda_i u_{r+i}-\lambda_i b_i\Big)+\sum_{j=1}^{r}\Big(\sum_{i=1}^{n}a_{ij}\lambda_i\Big)u_j,\quad u\in E^{r+n}=U_0,\quad\lambda\in E^n=\Lambda_0.$$

С помощью классического метода (§ 1.2) нетрудно показать, что нижняя грань функции $\varphi(z)=e^z-\lambda_i z-\lambda_i b_i$ переменной z на числовой оси равна $\varphi_*=\lambda_i-\lambda_i\ln\lambda_i-\lambda_i b_i$, причем при $\lambda_i>0$ она достигается в точке $z_*=-\ln\lambda_i$; функция $\lambda_i\ln\lambda_i$ при $\lambda_i=0$ здесь считается доопределенной по непрерывности нулем. Отсюда, опираясь на линейность функции $L(u,\lambda)$ по переменным $u_1,\dots,u_r$, получаем

$$\psi(\lambda)=\inf_{u\in E^{r+n}}L(u,\lambda)=$$
$$=\begin{cases}\sum\limits_{i=1}^{n}(\lambda_i-\lambda_i\ln\lambda_i-\lambda_i b_i), & \lambda\in E_+^n,\quad\sum\limits_{i=1}^{n}a_{ij}\lambda_i=0,\quad j=1,\dots,r,\\ -\infty & \text{при других}\quad\lambda.\end{cases}$$

Поэтому двойственная задача (30) здесь будет иметь вид

$$\psi(\lambda)=\sum_{i=1}^{n}(\lambda_i-\lambda_i\ln\lambda_i-\lambda_i b_i)\to\sup;\quad\lambda\in\Lambda,$$
$$\Lambda=\Big\{\lambda=(\lambda_1,\dots,\lambda_n)\in E_+^n\colon\sum_{i=1}^{n}a_{ij}\lambda_i=0,\ j=1,\dots,r\Big\}.\tag{43}$$

Если здесь верхняя грань достигается в точке $\lambda^*\neq 0$, то задачу (43) можно записать в более простой форме. А именно, учитывая, что любую точку $\lambda=$

$= (\lambda_1, \ldots, \lambda_n) \neq 0$ можно представить в виде $\lambda = \alpha\nu$, где $\alpha = \sum_{i=1}^{n} \lambda_i$, $\nu = (\nu_1, \ldots, \nu_n)$, $\nu_i = \lambda_i/\alpha$, $\nu_1 + \ldots + \nu_n = 1$, задачу (43) перепишем сначала в терминах переменных (α, ν):

$$\psi_1(\alpha, \nu) = \psi(\alpha\nu) = \sum_{i=1}^{n} \alpha(\nu_i - \nu_i \ln \alpha\nu_i - \nu_i b_i) = \alpha\left[1 - \ln \alpha + \sum_{i=1}^{n} (\nu_i \ln c_i - \nu_i \ln \nu_i)\right] \to \sup;$$

$$(\alpha, \nu) \in \Lambda_1 = \left\{(\alpha, \nu):\ \alpha > 0,\ \nu \in E_+^n,\ \sum_{i=1}^{n} \nu_i = 1,\ \sum_{i=1}^{n} a_{ij}\nu_i = 0,\ j = 1, \ldots, r\right\}.$$

Далее, пользуясь классическим методом (§ 1.2), убеждаемся, что точка $\alpha^* = \prod_{i=1}^{n} \left(\frac{c_i^{\nu_i}}{\nu_i^{\nu_i}}\right) > 0$ (здесь принято $0^0 = 1$) доставляет функции $\psi_1(\alpha, \nu)$ максимальное значение по $\alpha > 0$ при фиксированном $\nu \in E_+^n$, причем $\psi_1(\alpha^*, \nu) = \prod_{i=1}^{n} \left(\frac{c_i^{\nu_i}}{\nu_i^{\nu_i}}\right)$.

Тогда двойственная задача (43) перепишется в следующем виде:

$$\psi_2(\nu) = \frac{c_1^{\nu_1} \ldots c_n^{\nu_n}}{\nu_1^{\nu_1} \ldots \nu_n^{\nu_n}} \to \sup; \quad \nu \in \Lambda_2, \tag{44}$$

$\Lambda_2 = \left\{\nu = (\nu_1, \ldots, \nu_n) \in E_+^n:\ \sum_{i=1}^{n} \nu_i = 1,\ \sum_{i=1}^{n} a_{ij}\nu_i = 0,\ j = 1, \ldots, r\right\}$. Если $\nu^* = (\nu_1^*, \ldots, \nu_n^*) \in \operatorname{int} E_+^n$ — решение задачи (44), то, полагая $\lambda^* = \alpha^*\nu^*$, где $\alpha^* = \prod_{i=1}^{n} \left(\frac{c_i}{\nu_i^*}\right)^{\nu_i^*}$, $u_{r+i,*} = \lambda_i^*$, $i = 1, \ldots, n$, из системы линейных алгебраических уравнений (41) можно получить $u_{1*}, \ldots, u_{r*}$, откуда имеем решение $x_* = (x_{1*} = e^{u_{1*}}, \ldots, x_{r*} = e^{u_{r*}})$ исходной задачи (40). Задача (44) часто бывает проще задачи (40). Переход к двойственной задаче особенно эффективным оказывается тогда, когда множество Λ_2 в задаче (44) состоит из единственной точки ν^*, которая и будет решением этой задачи.

Аналогично может быть исследована и более общая задача геометрического программирования

$$g_0(x) \to \inf; \quad x \in X = \{x \in \operatorname{int} E_+^r:\ g_1(x) \leqslant 1, \ldots, g_m(x) \leqslant 1\},$$

где $g_0(x), \ldots, g_m(r)$ — позиномы. Подробнее о геометрическом программировании, его приложениях см., например, в [204; 260; 541].

Читателей, желающих подробнее ознакомиться с красивой и богатой результатами теорией двойственности, с различными ее приложениями, отсылаем к [6; 7; 14; 40; 44; 49; 52; 83; 84; 209; 225; 233; 234; 278; 297; 314; 358; 366; 373; 434; 465; 584; 604; 605; 613; 617; 670; 683; 687].

Заметим также, что в последнее время растет интерес к задачам, в которых нарушены соотношения двойственности (34), такие задачи возникают при исследовании объектов, описываемых противоречивыми системами ограничений, и имеют интересные приложения [297; 298; 644].

Упражнения

1. Сформулировать аналоги теорем Куна — Таккера для задачи максимизации: $g(x) \to \sup$, $x \in X$, где множество X определено посредством (2). Указание: рассмотреть задачу: $f(x) = -g(x) \to \inf$, $x \in X$, и воспользоваться теоремами 2–4.

2. С помощью теорем Куна — Таккера исследовать задачу: $f(x) = \sum_{i=1}^{n} |x^i - a_i| \to \inf$, $x \in X$, где $X = \{x \in E^n:\ |x| \leqslant 1\}$ или $X = \{x \in E^n:\ x^1 + \ldots + x^n = 0\}$, или $X = E_+^n$; $a_1, \ldots, a_n$ — заданные числа.

3. Применить теорему Куна — Таккера к задаче квадратичного программирования: $f(x) = (x^1)^2 + \ldots + n(x^n)^2 \to \inf$, $x \in X$, где $X = \{x \in E^n:\ x^1 + \ldots + x^n = 1\}$, или $X = \{x \in E^n:\ x^1 + \ldots + x^n \leqslant 1\}$, или $X = \{x \in E^n:\ -1 \leqslant x^1 + \ldots + x^n \leqslant 1\}$, или X является пересечением предыдущих множеств с E_+^n.

4. Решить задачи геометрического программирования:

а) $g(x) = c_1 x^{a_1} + c_2 x^{-a_2} \to \inf$ при $x > 0$, где $c_i > 0$, $a_i > 0$ — заданные числа;

б) $g(x, y) = x^{-1}y + 2x^2y + 3x^{-1}y^{-2} \to \inf$ при $x > 0$, $y > 0$;

в) $g(x, y, z) = 4x^{-1}y^{-1}z^{-1} + xy + 4xz + 2yz \to \inf$ при $x > 0$, $y > 0$, $z > 0$;

г) $g(x, y) = y \to \inf$ при $x \geqslant 0$, $y > 0$, $x^4y^{-4} + x^{-1}y^{1/2} \leqslant 1$;

д) $g(x, y, z) = x + y^{-1}z^{-1/2} \to \inf$ при $x > 0$, $y > 0$, $z > 0$, $x^{-1}y + x^{-1}z \leqslant 1$.

5. Найти решение задачи: $f(u) = -x - y \to \inf$, $u = (x, y) \in X = \{u \in E^2:\ g(u) = x^2 + y^2 - 2 \leqslant 0\}$ и двойственной к ней задачи. Убедиться, что здесь множество Λ из (30) выпукло и открыто.

6. Показать, что двойственная задача к задаче из примера 5 является задачей линейного программирования, и решить ее.

7. Доказать, что выпуклая квадратичная функции $f(x) = \frac{1}{2}\langle Cx, x\rangle + \langle c, x\rangle$ либо достигает своей нижней грани на E^n, либо не ограничена снизу [84].

8. Пусть U_0 — выпуклое множество из E^n, функции $g_1(u), \ldots, g_m(u)$ выпуклы на U_0, $g_i(u) = \langle a_i, u\rangle - b_i$, $i = m+1, \ldots, s$. Доказать, что если система неравенств $g_i(u) < 0$, $i = 1, \ldots, m$, $g_i(u) = 0$, $i = m+1, \ldots, s$, не имеет решения на U_0, то существуют числа $\lambda_1 \geqslant 0, \ldots, \lambda_m \geqslant 0$, $\lambda_{m+1}, \ldots, \lambda_s$ такие, что $\lambda_1 g_1(u) + \ldots + \lambda_s g_s(u) \geqslant 0$ при всех $u \in U_0$.

Указание: построить множества, аналогичные множествам A и B из доказательства теоремы 2, и применить к ним теорему отделимости 5.2.

9. Пользуясь теоремой Фаркаша, доказать, что для несовместности системы линейных неравенств $\langle e_i, x\rangle \leqslant \mu_i$, $i = 0, \ldots, m$, необходимо и достаточно, чтобы существовали такие числа $\lambda_0 \geqslant 0$, $\lambda_1 \geqslant 0, \ldots, \lambda_m \geqslant 0$, что $\lambda_0 e_0 + \lambda_1 e_1 + \ldots + \lambda_m e_m = 0$, $\lambda_0\mu_0 + \lambda_1\mu_1 + \ldots + \lambda_m\mu_m < 0$ [752].

10. Доказать, что два непустых многогранных множества $A = \{x \in E^n:\ \langle e_i, x\rangle \leqslant \mu_i,\ i = 0, \ldots, k\}$ и $B = \{x \in E^n:\ \langle e_i, x\rangle \leqslant \mu_i,\ i = k+1, \ldots, m\}$, не имеющие общих точек, сильно отделимы.

Указание: рассмотреть гиперплоскость $\langle c, x\rangle = \gamma$, где $c = \sum_{i=0}^{k} \lambda_i e_i$, $\gamma = \sum_{i=0}^{k} \lambda_i\mu_i$, числа $\lambda_0, \ldots, \lambda_m$ взяты из упражнения 9.

11. Доказать, что если система линейных неравенств $\langle e_0, x\rangle < 0$, $\langle e_1, x\rangle \leqslant 0, \ldots, \langle e_m, x\rangle \leqslant 0$ несовместна, то существуют такие числа $\lambda_1 \geqslant 0, \ldots, \lambda_m \geqslant 0$, что $e_0 = -\lambda_1 e_1 - \ldots - \lambda_m e_m$.

Указание: воспользоваться теоремой Фаркаша.

12. Пусть система $\langle e_0, x\rangle < \mu_0$, $\langle e_i, x\rangle \leqslant \mu_i$, $i = 1, \ldots, m$, несовместна, а подсистема $\langle e_i, x\rangle \leqslant \mu_i$, $i = 1, \ldots, m$, совместна. Доказать, что тогда существуют числа $\lambda_1 \geqslant 0, \ldots, \lambda_m \geqslant 0$ такие, что $e_0 = -\lambda_1 e_1 - \ldots - \lambda_m e_m$, $\mu_0 + \lambda_1\mu_1 + \ldots + \lambda_m\mu_m \leqslant 0$ [54; 670].

Указание: в пространстве переменных $(x, t) \in E^{n+1}$ рассмотреть систему $\langle e_0, x\rangle - t\mu_0 < 0$, $\langle e_i, x\rangle - t\mu_i \leqslant 0$, $i = 1, \ldots, m$, $\langle 0, x\rangle - t \leqslant 0$ и воспользоваться утверждением из упражнения 11.

13. Рассмотрите задачу (1), (2) при $X_0 = \{x \in E^1:\ x \geqslant 0\}$, $f(x) = x$, $X = \{x \geqslant 0:\ g(x) = x^2 \leqslant 0\}$ и убедитесь, что функция Лагранжа $L(x, \lambda) = x + \lambda x^2$, $x \geqslant 0$, $\lambda \geqslant 0$, этой задачи имеет бесконечно много седловых точек (ср. с примером 1). Сформулируйте двойственную задачу.

ГЛАВА 5

Методы минимизации функций многих переменных

Выше в гл. 3 был рассмотрен симплекс-метод для решения задач линейного программирования. Перейдем к изложению других методов минимизации функций конечного числа переменных, не предполагая линейности рассматриваемых задач.

К настоящему времени разработано и исследовано большое число методов минимизации функций многих переменных. Мы ниже остановимся на некоторых наиболее известных и часто используемых на практике методах минимизации, будет дано краткое описание каждого из рассматриваемых методов, исследованы вопросы сходимости, обсуждены некоторые вычислительные аспекты этих методов. При этом мы ограничимся рассмотрением лишь одного-двух основных вариантов излагаемых методов, чтобы ознакомить читателя с основами этих методов, полагая, что знание основ методов облегчит читателю изучение литературы, позволит ему без особого труда понять суть того или иного метода и выбрать подходящий вариант метода или самому разработать более удобные его модификации, лучше приспособленные для решения интересующего читателя класса задач. В конце главы будут высказаны некоторые общие замечания по методам минимизации.

Из обширной литературы, посвященной методам минимизации функций конечного числа переменных и их приложениям, мы можем здесь упомянуть лишь весьма незначительную ее часть [1; 13; 16; 18–20; 24–30; 32; 33; 52; 53; 56; 61; 62; 66; 70; 71; 73; 74; 76–78; 83–85; 89–92; 94; 102; 109; 116; 117; 128; 131; 134; 135; 140; 143; 148; 154; 161; 179; 183; 184; 194; 203–205; 214; 216; 218; 222; 226; 227; 231; 232; 234; 243; 250–252; 255–257; 259–266; 272; 273; 281; 286; 292; 294–299; 301; 302; 304–309; 314; 316–320; 326; 330; 341–345; 347; 356; 358; 361; 368–370; 372–375; 377; 386–388; 390; 396; 398; 401; 410; 422; 423; 426; 437; 442; 447; 448; 465; 466; 470; 471; 481; 491; 493–495; 499; 506; 511; 516–518; 520; 521; 523; 525; 527; 539; 541; 542; 548–550; 561–566; 572; 580–582; 586; 590; 591; 601; 603; 606; 608; 610; 612–614; 620; 623; 624; 635; 639; 652; 657; 661; 662; 670; 671; 675; 676; 678; 681; 684; 686–688; 695; 697; 704; 709–711; 713; 718–721; 725; 732; 737; 738; 742; 743–747; 750–752; 759; 760; 765; 769; 774–777; 786; 792; 793; 795; 799; 803; 807; 811; 813; 818].

§ 1. Градиентный метод

1. Будем рассматривать задачу

$$f(x) \to \inf; \quad x \in X \equiv E^n, \tag{1}$$

предполагая, что функция $f(x)$ непрерывно дифференцируема на E^n, т. е. $f(x) \in C^1(E^n)$. Согласно определению 2.2.1 дифференцируемой функции

$$f(x+h) - f(x) = \langle f'(x), h\rangle + o(h; x), \tag{2}$$

где $\lim\limits_{|h|\to 0} o(h; x)|h|^{-1} = 0$. Если $f'(x) \neq 0$, то при достаточно малых $|h|$ главная часть приращения (2) будет определяться величиной $\langle f'(x), h\rangle$. В силу неравенства Коши — Буняковского

$$-|f'(x)| \cdot |h| \leqslant \langle f'(x), h\rangle \leqslant |f'(x)| \cdot |h|,$$

причем, если $f'(x) \neq 0$, то правое неравенство превращается в равенство только при $h = \alpha f'(x)$, а левое неравенство — только при $h = -\alpha f'(x)$, где $\alpha = \text{const} \geqslant 0$. Отсюда ясно, что при $f'(x) \neq 0$ направление наибыстрейшего возрастания функции $f(x)$ в точке x совпадает с направлением градиента $f'(x)$, а направление наибыстрейшего убывания — с направлением антиградиента $(-f'(x))$.

Это замечательное свойство градиента лежит в основе ряда итерационных методов минимизации функций. Одним из таких методов является градиентный метод, к описанию которого мы переходим. Этот метод, как и все итерационные методы, предполагает выбор начального приближения — некоторой точки x_0. Общих правил выбора точки x_0 в градиентном методе, как, впрочем, и в других методах, к сожалению, нет. В тех случаях, когда из геометрических, физических или каких-либо других соображений может быть получена априорная информация об области расположения точки (или точек) минимума, то начальное приближение x_0 стараются выбрать поближе к этой области.

Будем считать, что некоторая начальная точка x_0 уже выбрана. Тогда градиентный метод заключается в построении последовательности $\{x_k\}$ по правилу

$$x_{k+1} = x_k - \alpha_k f'(x_k), \quad \alpha_k > 0, \quad k = 0, 1, \ldots \tag{3}$$

Число α_k из (3) часто называют *длиной шага* или просто *шагом градиентного метода*. Если $f'(x_k) \neq 0$, то шаг $\alpha_k > 0$ можно выбрать так, чтобы $f(x_{k+1}) < f(x_k)$. В самом деле, из равенства (2) имеем

$$f(x_{k+1}) - f(x_k) = \alpha_k[-|f'(x_k)|^2 + o(\alpha_k)\alpha_k^{-1}] < 0$$

при всех достаточно малых $\alpha_k > 0$. Если $f'(x_k) = 0$, то x_k — стационарная точка. В этом случае процесс (3) прекращается, и при необходимости проводится дополнительное исследование поведения функции в окрестности точки x_k для выяснения того, достигается ли в точке x_k минимум функции $f(x)$ или не достигается. В частности, если $f(x)$ — выпуклая функция, то согласно теореме 4.2.3 в стационарной точке всегда достигается минимум.

Существуют различные способы выбора величины α_k в методе (3). В зависимости от способа выбора α_k можно получить различные варианты градиентного метода. Укажем несколько наиболее употребительных на практике способов выбора α_k.

1) На луче $x = x_k - \alpha f'(x_k)$, $\alpha \geqslant 0$, направленном по антиградиенту, введем функцию одной переменной

$$g_k(\alpha) = f(x_k - \alpha f'(x_k)), \quad \alpha \geqslant 0,$$

и определим α_k из условий

$$g_k(\alpha_k) = \inf_{\alpha \geqslant 0} g_k(\alpha) = g_{k*}, \quad \alpha_k > 0. \tag{4}$$

Метод (3), (4) принято называть *методом скорейшего спуска*. При $f'(x_k) \neq 0$ согласно формуле (2.6.1) имеем $g_k'(0) = -|f'(x_k)|^2 < 0$, поэтому нижняя

грань в (4) может достигаться только при $\alpha_k > 0$. Приведем пример, когда величина α_k, определяемая условием (4), существует и может быть выписана в явном виде.

Пример 1. Пусть дана квадратичная функция

$$f(x) = \tfrac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle, \tag{5}$$

где A — симметричная положительно определенная матрица порядка $n \times n$, b — вектор из E^n. Выше было доказано, что эта функция сильно выпукла и ее производные вычисляются по формулам

$$f'(x) = Ax - b; \quad f''(x) = A.$$

Поэтому метод (3) в данном случае будет выглядеть так:

$$x_{k+1} = x_k - \alpha_k(Ax_k - b), \quad k = 0, 1, \ldots$$

Таким образом, градиентный метод для функции (5) представляет собой хорошо известный итерационный метод решения системы линейных алгебраических уравнений $Ax = b$. Определим α_k из условий (4). Пользуясь формулой (4.2.10), имеем

$$g_k(\alpha) = f(x_k) - \alpha|f'(x_k)|^2 + (\alpha^2/2)\langle Af'(x_k), f'(x_k)\rangle, \quad \alpha \geqslant 0.$$

При $f'(x_k) \neq 0$ условие $g_k'(\alpha) = -|f'(x_k)|^2 + \alpha\langle Af'(x_k), f'(x_k)\rangle = 0$ дает

$$\alpha_k = \frac{|f'(x_k)|^2}{\langle Af'(x_k), f'(x_k)\rangle} = \frac{|Ax_k - b|^2}{\langle A(Ax_k - b), Ax_k - b\rangle} > 0.$$

Поскольку функция $g_k(\alpha)$ выпукла, то в найденной точке α_k эта функция достигает своей нижней грани при $\alpha \geqslant 0$. Метод скорейшего спуска для функции (5) описан.

Однако точное определение величины α_k из условий (4) не всегда возможно. Кроме того, нижняя грань в (4) при некоторых k может и не достигаться. Поэтому на практике ограничиваются нахождением величины α_k, приближенно удовлетворяющей условиям (4). Здесь возможен, например, выбор α_k из условий

$$g_{k*} \leqslant g_k(\alpha_k) \leqslant g_{k*} + \delta_k, \quad \delta_k \geqslant 0, \quad \sum_{k=0}^{\infty} \delta_k = \delta < \infty, \tag{6}$$

или из условий

$$g_{k*} \leqslant g_k(\alpha_k) \leqslant (1 - \lambda_k)g_k(0) + \lambda_k g_{k*}, \quad 0 < \bar{\lambda} \leqslant \lambda_k \leqslant 1. \tag{7}$$

Величины δ_k, λ_k из (6), (7) характеризуют погрешность выполнения условия (4): чем ближе δ_k к нулю или λ_k к единице, тем точнее выполняется условие (4). При поиске α_k из условий (6), (7) можно пользоваться различными методами минимизации функций одной переменной (например, методами гл. 1).

Следует также заметить, что антиградиент $(-f'(x_k))$ указывает направление быстрейшего спуска лишь в достаточно малой окрестности точки x_k. Это означает, что если функция $f(x)$ меняется быстро, то в следующей точке x_{k+1} направление антиградиента $(-f'(x_{k+1}))$ может сильно отличаться

от направления $(-f'(x_k))$. Поэтому слишком точное определение величины α_k из условий (4) не всегда целесообразно.

2) На практике нередко довольствуются нахождением какого-либо $\alpha_k > 0$, обеспечивающего условие монотонности: $f(x_{k+1}) < f(x_k)$. С этой целью задаются какой-либо постоянной $\alpha > 0$ и в методе (3) на каждой итерации берут $\alpha_k = \alpha$. При этом для каждого $k \geqslant 0$ проверяют условие монотонности, и в случае его нарушения $\alpha_k = \alpha$ дробят до тех пор, пока не восстановится монотонность метода. Время от времени полезно менять α, пробовать увеличить α с сохранением условия монотонности.

3) Если функция $f(x) \in C^{1,1}(E^n)$, т. е. $f(x) \in C^1(E^n)$, и градиент $f'(x)$ удовлетворяет условию

$$|f'(u) - f'(v)| \leqslant L|u - v|, \quad u, v \in E^n,$$

причем константа L известна, то в (3) в качестве α_k может быть взято любое число, удовлетворяющее условиям

$$0 < \varepsilon \leqslant \alpha_k \leqslant 2/(L + 2\varepsilon), \tag{8}$$

где ε_0, ε — положительные числа, являющиеся параметрами метода. В частности, при $\varepsilon = L/2$, $\varepsilon_0 = 1/L$ получим метод (3) с постоянным шагом $\alpha_k = 1/L$. Отсюда ясно, что если константа L большая или получена с помощью слишком грубых оценок, то шаг α_k в (3) будет маленьким. Метод (3), (8) подробнее рассмотрим в следующем параграфе.

4) Возможен выбор α_k из условия [94; 374; 603]:

$$f(x_k) - f(x_k - \alpha_k f'(x_k)) \geqslant \varepsilon\alpha_k|f'(x_k)|^2, \quad \varepsilon > 0. \tag{9}$$

Для удовлетворения условия (9) сначала обычно берут некоторое число $\alpha_k = \alpha > 0$ (одно и то же на всех итерациях; например, $\alpha_k = 1$), а затем при необходимости дробят его, т. е. изменяют по закону $\alpha_k = \lambda^i\alpha$, $i = 0, 1\ldots$, $0 < \lambda < 1$, до тех пор, пока впервые не выполнится условие (9). Такой способ определения α_k в литературе часто называют выбором шага по Армихо [94].

5) Возможно априорное задание величин α_k из условий

$$\alpha_k > 0, \quad k = 0, 1, \ldots; \quad \sum_{k=0}^{\infty} \alpha_k = \infty, \quad \sum_{k=0}^{\infty} \alpha_k^2 < \infty. \tag{10}$$

Например, в качестве α_k можно взять $\alpha_k = c(k+1)^{-\alpha}$, где $c = \text{const} > 0$, а число α таково, что $1/2 < \alpha \leqslant 1$. В частности, если $\alpha = 1$, $c = 1$, то получим $\alpha_k = (k+1)^{-1}$, $k = 0, 1, \ldots$ Такой выбор $\{\alpha_k\}$ в (3) очень прост для реализации, но не гарантирует выполнения условия монотонности $f(x_{k+1}) < f(x_k)$ и, вообще говоря, сходится медленно. Более подробно о методе (3), (10) см. ниже в § 3.

6) В тех случаях, когда заранее известна величина $f_* = \inf\limits_{E^n}(x) > -\infty$, в (3) можно принять

$$\alpha_k = (f(x_k) - f_*)|f'(x_k)|^{-2}$$

— это абсцисса точки пересечения прямой $f = f_*$ с касательной к кривой $f = g_k(\alpha) = f(x_k - \alpha f'(x_k))$ в точке $(0, g_k(0))$.

Допустим, что какой-либо способ выбора α_k в (3) (например, один из перечисленных выше способов) уже выбран. Тогда на практике итерации (3)

продолжают до тех пор, пока не выполнится некоторый критерий окончания счета. Здесь часто используются следующие критерии:

$$|x_k - x_{k+1}| \leqslant \varepsilon, \quad \text{или} \quad |f(x_k) - f(x_{k+1})| \leqslant \varepsilon, \quad \text{или} \quad |f'(x_k)| \leqslant \varepsilon,$$

$$\text{или} \ \frac{|f(x_{k+1}) - f(x_k)|}{|x_{k+1} - x_k|} < \varepsilon, \ \text{или} \ |f(x_k) - f(x_{k+1})| + |x_k - x_{k+1}| \leqslant \varepsilon,$$

где $\varepsilon > 0$ — заданное число. Иногда заранее задают число итераций; возможны различные сочетания этих и других критериев. Разумеется, к этим критериям окончания счета надо относиться критически, поскольку они могут выполняться и вдали от искомой точки минимума. К сожалению, надежных критериев окончания счета, которые гарантировали бы получение решения задачи (1) с требуемой точностью, и применимых к широкому классу задач, пока нет. Сделанное замечание о критериях окончания счета относится и к другим излагаемым ниже методам.

В теоретических вопросах, когда исследуется сходимость метода, предполагается, что процесс (3) продолжается неограниченно и приводит к последовательности $\{x_k\}$. Здесь возникают вопросы, будет ли полученная последовательность $\{x_k\}$ минимизирующей для задачи (1), будет ли она сходиться к множеству точек минимума

$$X_* = \left\{ x \in E^n, f(x) = f_* = \inf_{E^n} f(x) \right\}$$

или, иначе говоря, выполняются ли соотношения

$$\lim_{k \to \infty} f(x_k) = f_*, \quad \lim_{k \to \infty} \rho(x_k, X_*) = 0. \tag{11}$$

Для положительного ответа на эти вопросы на функцию $f(x)$, кроме условия $f(x) \in C^1(E^n)$, приходится накладывать дополнительные более жесткие ограничения.

2. Подробнее рассмотрим эти вопросы для метода скорейшего спуска, когда в (3) величина α_k выбирается из условия (6).

Теорема 1. *Пусть* $f_* = \inf_{E^n} f(x) > -\infty$, $f(x) \in C^{1,1}(E^n)$. *Тогда последовательность* $\{x_k\}$, *полученная методом* (3), (6), *при произвольном начальном приближении* x_0 *такова, что* $\lim_{k \to \infty} f'(x_k) = 0$. *Если при этом множество* $M_0(x_0) = \{x \in E^n: f(x) \leqslant f(x_0) + \delta\}$, *где* δ *взято из* (6), *ограничено, то* $\lim_{k \to \infty} \rho(x_k, S_*) = 0$, *где* $S_* = \{x \in M_\delta(x_0): f'(x) = 0\}$ — *множество стационарных точек функции* $f(x)$ *на* $M_\delta(x_0)$.

Доказательство. Если при некотором $k \geqslant 0$ окажется, что $f'(x_k) = 0$, то из (3), (6) формально получаем $x_k = x_{k+1} = \ldots$ и утверждения теоремы становятся тривиальными. Поэтому будем считать, что $f'(x_k) \neq 0$ при всех $k = 0, 1, \ldots$ Так как $f(x_{k+1}) = g_k(\alpha_k) \leqslant \inf_{\alpha \geqslant 0} g_k(\alpha) + \delta_k \leqslant f(x_k - \alpha f'(x_k)) + \delta_k$ при всех $\alpha \geqslant 0$, то из неравенства (2.6.7) при $y = x_k$, $x = x_k - \alpha f'(x_k)$ имеем

$$f(x_k) - f(x_{k+1}) \geqslant f(x_k) - f(x_k - \alpha f'(x_k)) - \delta_k \geqslant$$
$$\geqslant \alpha |f'(x_k)|^2 - L\alpha^2 |f'(x_k)|^2/2 - \delta_k \geqslant \alpha(1 - \alpha L/2)|f'(x_k)|^2 - \delta_k$$

при всех $\alpha \geqslant 0$ и $k = 0, 1, \ldots$ Следовательно,

$$f(x_k) - f(x_{k+1}) \geqslant \max_{\alpha \geqslant 0} \alpha(1 - \alpha L/2)|f'(x_k)|^2 - \delta_k =$$
$$= (1/(2L))|f'(x_k)|^2 - \delta_k, \quad k = 0, 1, \ldots \tag{12}$$

Отсюда получаем

$$f(x_{k+1}) \leqslant f(x_k) + \delta_k, \quad k = 0, 1, \ldots \tag{13}$$

Так как $f(x_k) \geqslant f_* > -\infty$, $k = 0, 1, \ldots$, то из леммы 2.6.2 и (13) следует существование предела $\lim_{k \to \infty} f(x_k) \geqslant f_*$. Тогда $\lim_{k \to \infty} (f(x_k) - f(x_{k+1})) = 0$ и из (12) будем иметь $\lim_{k \to \infty} f'(x_k) = 0$.

Наконец, пусть множество $M_\delta(x_0)$ ограничено. Суммируя неравенства (13) по k от 0 до $m - 1$, получим

$$f(x_m) \leqslant f(x_0) + \sum_{k=0}^{m-1} \delta_k \leqslant f(x_0) + \delta, \quad m = 1, 2, \ldots,$$

т. е. $\{x_k\} \in M_\delta(x_0)$. По теореме Больцано — Вейерштрасса ограниченная последовательность $\{x_k\}$ имеет хотя бы одну предельную точку. Пусть x_* — произвольная предельная точка $\{x_k\}$ и $\{x_{k_m}\} \to x_*$. Пользуясь непрерывностью $f'(x)$, отсюда имеем $\lim_{m \to \infty} f'(x_{k_m}) = f'(x_*) = 0$, т. е. $x_* \in S_*$. Так как расстояние $\rho(x, S_*)$ непрерывно (см. лемму 2.1.2), то $\lim_{m \to \infty} \rho(x_{k_m}, S_*) = \rho(x_*, S_*) = 0$. Отсюда следует, что числовая последовательность $\{\rho(x_k, S_*)\}$ имеет единственную предельную точку, равную нулю, т. е. $\lim_{k \to \infty} \rho(x_k, S_*) = 0$. Теорема 1 доказана. □

Теорема 2. *Пусть выполнены все условия теоремы* 1 *и, кроме того, функция* $f(x)$ *выпукла на* E^n. *Тогда, для последовательности* $\{x_k\}$, *определяемой условиями* (3), (6), *имеют место соотношения* (11). *Если, кроме того, в* (6) $\{\delta_k\} = O(k^{-2})$, *то справедлива оценка*

$$0 \leqslant f(x_k) - f_* \leqslant c_0 k^{-1}, \quad c_0 = \text{const} > 0. \tag{14}$$

Доказательство. Из ограниченности $M_\delta(x_0)$, непрерывности $f(x)$, согласно теореме 2.1.2, имеем $f_* > -\infty$, $X_* \neq \varnothing$, $X_* \subset M_\delta(x_0)$. Тогда для любой точки $x_* \in X_*$ с помощью неравенства (4.2.4) получаем

$$0 \leqslant f(x_k) - f_* = f(x_k) - f(x_*) \leqslant \langle f'(x_k), x_k - x_* \rangle \leqslant$$
$$\leqslant |f'(x_k)| \cdot |x_k - x_*| \leqslant d|f'(x_k)|, \quad k = 0, 1, \ldots, \tag{15}$$

где $d \geqslant \operatorname{diam} M_\delta(x_0) = \sup_{u, v \in M_\delta(x_0)} |u - v|$ — диаметр множества $M_\delta(x_0)$. В теореме 1 было доказано, что $\lim_{k \to \infty} f'(x_k) = 0$. Отсюда и из (15) следует, что $\lim_{k \to \infty} f(x_k) = f_*$. Учитывая включение $\{x_k\} \in M_\delta(x_0)$, тогда с помощью теоремы 2.1.2 получаем второе из равенств (11).

Докажем оценку (14). Обозначим $a_k = f(x_k) - f_*$. Из неравенств (12), (15) имеем $a_k - a_{k+1} = f(x_k) - f(x_{k+1}) \geqslant (1/(2L))\,|f'(x_k)|^2 - \delta_k \geqslant a_k^2/(2Ld^2) - \delta_k$. По условию $\delta_k = O(k^{-2})$, т. е. $0 \leqslant \delta_k \leqslant c_1 k^{-2}$, $k = 1, 2, \ldots$, $c_1 = \text{const} > 0$. Полагая $A = \max\{c_1; 2Ld^2\}$, получим

$$a_{k+1} \leqslant a_k - a_k^2/A + Ak^{-2}, \quad k = 1, 2, \ldots$$

Отсюда и из леммы 2.6.5 при $I_0 = \{1, 2, \ldots\}$, $I_1 = \varnothing$ следует оценка (14). Если $\delta_k = 0$, $k = 0, 1, \ldots$, то оценка (14) вытекает из неравенств (12), (15) и леммы 2.6.4. Теорема 2 доказана. □

Теорема 3. *Пусть $f(x)\in C^{1,1}(E^n)$, $f(x)$ сильно выпукла на E^n. Тогда для последовательности $\{x_k\}$, получаемой из* (3), (6) *при любом начальном приближении x_0, справедливы соотношения* (11). *Если при этом $\delta_k = O(k^{-2})$, то имеет место оценка* (14). *Если $\delta_k = 0$, $k = 0, 1, \ldots$, то верна более сильная, чем* (14), *оценка*

$$0 \leqslant f(x_k) - f_* \leqslant (f(x_0) - f_*)q^k, \tag{16}$$

$$|x_k - x_*|^2 \leqslant (2/\mu)(f(x_0) - f_*)q^k, \qquad k = 0, 1, \ldots, \tag{17}$$

где x_* — точка минимума $f(x)$ на E^n, $q = 1 - \mu/L$, $0 \leqslant q < 1$, μ — постоянная из теоремы 4.3.3.

Доказательство. Согласно теореме 4.3.1 множество $M_\delta(x_0)$ ограничено, $f_* > -\infty$, X_* состоит из единственной точки x_*. Поэтому равенства (11) и оценка (14) следуют из теорем 1, 2. Докажем оценки (16), (17). Из (4.3.9) при $v = x_k$, $u = x_*$ имеем

$$a_k = f(x_k) - f(x_*) \leqslant \langle f'(x_k), x_k - x_*\rangle - \varkappa|x_k - x_*|^2/2 \leqslant |f'(x_k)||x_k - x_*| -$$
$$- \varkappa|x_k - x_*|^2/2 \leqslant \sup_{z\geqslant 0}(|f'(x_k)|z - \varkappa z^2/2) = |f'(x_k)|^2/(2\varkappa),$$

т. е.

$$a_k = f(x_k) - f(x_*) \leqslant |f'(x_k)|^2/(2\varkappa), \quad k = 0, 1, \ldots \tag{18}$$

Подставив неравенство (18) в правую часть (12) при $\delta_k = 0$, получим

$$a_k - a_{k+1} \geqslant \frac{2\varkappa}{2L}a_k = \frac{\varkappa}{L}a_k, \quad k = 0, 1, \ldots$$

В § 4.3 было установлено, что $\varkappa = \mu \leqslant L$. Поэтому $0 \leqslant q = 1 - (\mu/L) < 1$, и предыдущее неравенство можно переписать в виде $0 \leqslant a_{k+1} \leqslant a_k(1 - \mu/L) = qa_k$. Отсюда имеем $a_k \leqslant qa_{k-1} \leqslant q^2 a_{k-2} \leqslant \ldots \leqslant q^k a_0$, что равносильно оценке (16). Наконец, из неравенства (4.3.3) следует

$$\varkappa|x_k - x_*|^2 \leqslant f(x_k) - f(x_*) = a_k, \qquad k = 0, 1, \ldots$$

Отсюда и из (16) получим оценку (17). Теорема 3 доказана. □

Метод скорейшего спуска имеет простой геометрический смысл: оказывается, точка x_{k+1}, определяемая условиями (3), (4), лежит на луче $L_k = \{x\colon x = x_k - \alpha f'(x_k),\ \alpha \geqslant 0\}$ в точке его касания поверхности уровня $\Gamma_{k+1} = \{x \in E^n\colon f(x) = f(x_{k+1})\}$, а сам луч L_k перпендикулярен к поверхности уровня $\Gamma_k = \{x \in E^n\colon f(x) = f(x_k)\}$ — см. рис. 5.1 и 5.2. В самом деле, пусть $x = x(t)$, $a \leqslant t \leqslant b$ — некоторое параметрическое уравнение кривой, принадлежащей Γ_k, т. е. $f(x(t)) = f(x_k) = \text{const}$, $a \leqslant t \leqslant b$, причем $x(t_0) = x_k$. Тогда $\frac{d}{dt}f(x(t)) = \langle f'(x(t)), \dot{x}(t)\rangle = 0$, $a \leqslant t \leqslant b$. В частности, при $t = t_0$ имеем $\langle f'(x_k), \dot{x}(t_0)\rangle = 0$. Это означает, что градиент (или антиградиент) $f'(x_k)$ перпендикулярен к касательному направлению поверхности уровня Γ_k в точке x_k, или, иначе говоря, луч L_k перпендикулярен к Γ_k. Далее, из условия (4) при $\alpha_k > 0$ получаем $g_k'(\alpha_k) = -\langle f'(x_k - \alpha_k f'(x_k)), f'(x_k)\rangle = -\langle f'(x_{k+1}), f'(x_k)\rangle = 0$. Но вектор $f'(x_{k+1})$ перпендикулярен к Γ_{k+1} в точке x_{k+1}, поэтому последнее равенство означает, что направление $f'(x_k)$ и, следовательно, луч L_k являются касательными к поверхности уровня Γ_{k+1} в точке x_{k+1}.

3. Из рис. 5.1 и 5.2 можно понять, что чем ближе поверхность уровня $f(x) = \text{const}$ к сфере, тем лучше сходится метод скорейшего спуска. Это же явление можно усмотреть и из оценок (16), (17) — чем ближе μ/L к единице (для функции $f(x) = |x|^2$, у которой поверхностями уровня являются сферы, как раз имеем $\mu/L = 1$), тем ближе q к нулю и тем лучше сходимость.

Те же рис. 5.1 и 5.2 показывают, а теоретические исследования и численные эксперименты подтверждают, что метод скорейшего спуска и другие

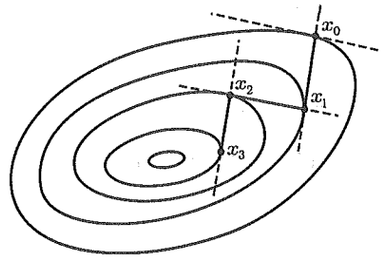

Рис. 5.1

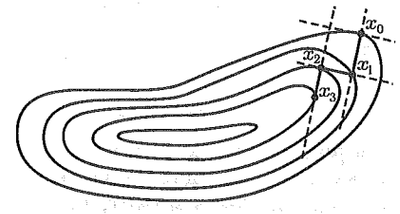

Рис. 5.2

варианты градиентного метода медленно сходятся в тех случаях, когда поверхности уровня функции $f(x)$ сильно вытянуты и функция имеет так называемый «овражный» характер. Это означает, что небольшое изменение некоторых переменных приводит к резкому изменению значений функции — эта группа переменных характеризует «склон оврага», а по остальным переменным, задающим направление «дна оврага», функция меняется незначительно (на рис. 5.2 и 5.3 изображены линии уровня «овражной»

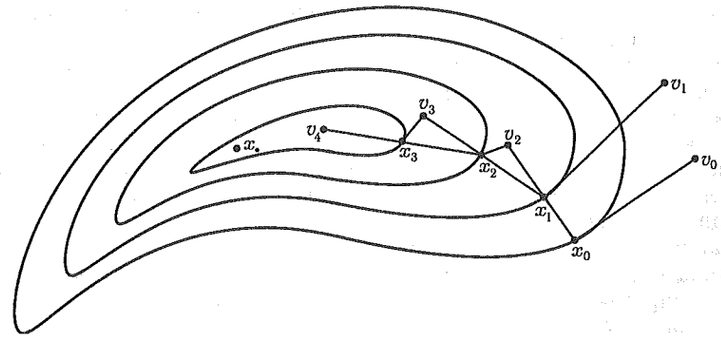

Рис. 5.3

функции двух переменных). Если точка лежит на «склоне оврага», то направление спуска из этой точки будет почти перпендикулярным к направлению «дна оврага», и в результате приближения $\{x_k\}$, получаемые градиентным методом, будут поочередно находиться то на одном, то на другом «склоне оврага». Если «склоны оврага» достаточно круты, то такие скачки «со склона на склон» точек x_k могут сильно замедлить сходимость градиентного метода.

Для ускорения сходимости этого метода при поиске минимума «овражной» функции можно предложить следующий эвристический прием, называемый *овражным методом*. Сначала опишем простейший вариант этого метода. В начале поиска задаются две точки v_0, v_1, из которых производят спуск с помощью какого-либо варианта градиентного метода, и получают две точки x_0, x_1 на «дне оврага». Затем полагают

$$v_2 = x_1 - (x_1 - x_0)|x_1 - x_0|^{-1} h \operatorname{sign}(f(x_1) - f(x_0)),$$

где h — положительная постоянная, называемая овражным шагом. Из точки v_2, которая, вообще говоря, находится на «склоне оврага», производят спуск с помощью градиентного метода и определяют следующую точку x_2 на «дне оврага».

Если уже известны точки $x_0, x_1, \ldots, x_k$, $k \geqslant 2$, то из точки

$$v_{k+1} = x_k - (x_k - x_{k-1})|x_k - x_{k-1}|^{-1} h \operatorname{sign}[f(x_k) - f(x_{k-1})] \qquad (19)$$

совершают спуск с помощью градиентного метода и находят следующую точку x_{k+1} на «дне оврага» (см. рис. 5.3; спуск из точки v_k в точку x_k, состоящий, быть может, из нескольких итерационных шагов градиентного метода, условно изображен отрезком прямой, соединяющей точки v_k, x_k, $k = 0, 1, \ldots$).

Величина овражного шага h подбирается эмпирически с учетом информации о минимизируемой функции, получаемой в ходе поиска минимума. От правильного выбора h существенно зависит скорость сходимости метода. Если шаг h велик, то на крутых поворотах «оврага» точки v_k могут слишком удаляться от «дна оврага» и спуск из точки v_k в точку x_k, может потребовать большого объема вычислений. Кроме того, при больших h на крутых поворотах может произойти выброс точки v_k из «оврага», и правильное направление поиска точки минимума будет потеряно. Если шаг h слишком мал, то поиск может очень замедлиться и эффект от применения овражного метода может стать незначительным.

Эффективность овражного метода может существенно возрасти, если величину овражного шага выбирать переменной, реагирующей на повороты «оврага» с тем, чтобы: 1) по возможности быстрее проходить прямолинейные участки на «дне оврага» за счет увеличения овражного шага; 2) на крутых поворотах «оврага» избежать выброса из «оврага» за счет уменьшения овражного шага; 3) добиться по возможности меньшего отклонения точек v_k от «дна оврага» и тем самым сократить объем вычислений, требуемый для градиентного спуска из точки v_k в точку x_k, $k = 0, 1, \ldots$ Интуитивно ясно, что для правильной реакции на поворот «оврага» надо учитывать «кривизну дна оврага», причем информацию о «кривизне» желательно получить, опираясь на результаты предыдущих итераций овражного метода.

В работе [657] предлагается следующий способ выбора овражного шага:

$$h_{k+1} = h_k \cdot c^{\cos\alpha_k - \cos\alpha_{k-1}}, \qquad k = 2, 3, \ldots, \qquad (20)$$

где α_k — угол между векторами $v_k - x_{k-1}$, $x_k - x_{k-1}$, определяемый условием

$$\cos\alpha_k = \langle v_k - x_{k-1}, x_k - x_{k-1}\rangle |v_k - x_{k-1}|^{-1} |x_k - x_{k-1}|^{-1},$$

а постоянная $c > 1$ является параметром алгоритма. Точка v_{k+1} определяется из (19) при $h = h_{k+1}$. Разность $\cos\alpha_k - \cos\alpha_{k-1}$ в равенстве (20)

связана с «кривизной дна оврага» и, кроме того, обладает важным свойством указывать направление изменения «кривизны». А именно, при переходе с участков «дна оврага» с малой «кривизной» на участки с большей «кривизной» будем иметь $\cos\alpha_k - \cos\alpha_{k-1} < 0$ (см. рис. 5.4). Тогда, в силу (19) имеем $h_{k+1} < h_k$, т. е. овражный шаг уменьшается, приспосабливаясь к повороту «дна оврага», что в свою очередь приводит к уменьшению выбросов точки v_{k+1} на «склоны оврага». При переходе с участков «дна оврага» с большой «кривизной» на участки с меньшей «кривизной», наоборот, $\cos\alpha_k - \cos\alpha_{k-1} > 0$, поэтому овражный шаг увеличится и появится возможность сравнительно быстро пройти участки с малой «кривизной», в частности, прямолинейные участки на «дне оврага». Если «кривизна дна оврага» на некоторых участках остается постоянной, то разность $\cos\alpha_k - \cos\alpha_{k-1}$ будет близка к нулю, и поиск минимума на таких участках будет проводиться с почти постоянным шагом, сформированным с учетом величины «кривизны» при выходе на рассматриваемый участок.

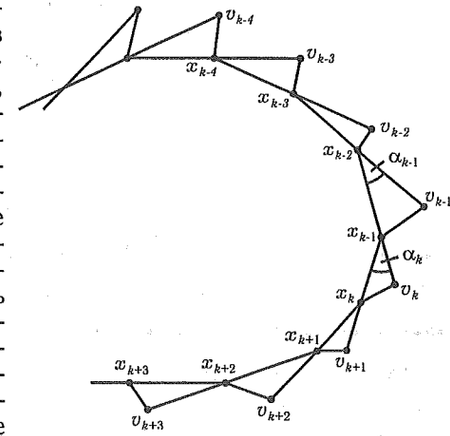

Рис. 5.4

Параметр c в равенстве (20) регулирует «чувствительность» метода к изменению «кривизны дна оврага», и правильный выбор этого параметра во многом определяет скорость движения по «оврагу». Некоторые эвристические соображения по поводу выбора c и другие аспекты применения овражного метода обсуждены в [657]. Выражение (20) для овражного шага удобнее преобразовать так

$$h_{k+1} = h_k c^{\cos\alpha_k - \cos\alpha_{k-1}} = h_{k-1} c^{\cos\alpha_k - \cos\alpha_{k-2}} = \ldots = h_2 c^{\cos\alpha_k - \cos\alpha_1},$$

откуда имеем

$$h_{k+1} = A c^{\cos\alpha_k}, \quad A = h_2 c^{-\cos\alpha_1} = \text{const} > 0, \quad k = 2, 3, \ldots$$

Другой способ ускорения сходимости градиентного метода заключается в выборе подходящей замены переменных $x = g(\xi) = (g_1(\xi), \ldots, g_n(\xi))$ с тем, чтобы поверхности уровня функции $f(g(\xi)) = G(\xi)$ в пространстве переменных $\xi = (\xi^1, \ldots, \xi^n)$ были близки к сферам. Заметим, что $G'(\xi) = (g'(\xi))^{\top} f'(g(\xi))$, где $g'(\xi) = \{g_{i\xi^j}(\xi)\}$ — матрица, i-я строка которой представляет собой $g_i'(\xi) = (g_{i\xi^1}(\xi), \ldots, g_{i\xi^n}(\xi))$, а $(g'(\xi))^{\top}$ — матрица, полученная из $g'(\xi)$ транспонированием. В пространстве переменных ξ градиентный метод выглядит так:

$$\xi_{k+1} = \xi_k - \beta_k (g'(\xi_k))^{\top} f'(g(\xi_k)), \quad \beta_k > 0, \quad k = 0, 1, \ldots$$

В пространстве исходных переменных $x = (x^1, \ldots, x^n)$ этот подход можно трактовать как итерационный процесс вида

$$x_{k+1} = x_k - \alpha_k A_k f'(x_k), \quad \alpha_k > 0, \quad k = 0, 1, \ldots,$$

— где A_k — некоторая невырожденная матрица порядка $n \times n$, представляющая собой параметр метода. То, что на этом пути можно добиться существенного ускорения скорости сходимости итераций, подтверждается, например, излагаемым ниже методом Ньютона, в котором полагается $A_k = (f''(x_k))^{-1}$, $k = 0, 1, \ldots$ О методах минимизации овражных функций и различных приемах ускорения сходимости итерационных методов см. [74; 84; 89; 222; 442; 525; 550; 586; 603; 657; 721; 738; 769].

4. Исследуем сходимость другого варианта градиентного метода (3), в котором параметр α_k определяется из условия (9) с помощью дробления. А именно, пусть $1 < \varepsilon < 2$, $\alpha > 0$ — фиксированные числа, а $i \geqslant 0$ — наименьший номер, для которого выполняется неравенство [374; 603]

$$f(x_k) - f(x_k - 2^{-i}\alpha f'(x_k)) \geqslant 2^{-i-1}\alpha\varepsilon|f'(x_k)|^2, \tag{21}$$

и пусть

$$\alpha_k = \alpha/2^i. \tag{22}$$

Теорема 4. *Пусть в задаче* (1) $f_* > -\infty$, $X_* \neq \varnothing$, *функция* $f(x)$ *выпукла на* E^n, $f(x) \in C^{1,1}(E^n)$. *Тогда для последовательности* $\{x_k\}$, *определяемой методом* (3), (21), (22), *имеют место соотношения* (11) *и, более того, существует точка* $v_* \in X_*$ *такая, что* $\{x_k\} \to v_*$,

$$|x_{k+1} - v_*| \leqslant |x_k - v_*|, \quad \rho(x_{k+1}, X_*) \leqslant \rho(x_k, X_*), \quad k = 0, 1, \ldots, \tag{23}$$

причем равенство в (23) *возможно лишь при* $x_k = x_{k+1} = \ldots = v_*$; *справедлива оценка*

$$0 \leqslant f(x_k) - f_* \leqslant \{\min\{(2-\varepsilon)/(2L); \alpha\}\}^{-1}(2/\varepsilon)|x_0 - v_*|^2 k^{-1} = O(1/k), \quad k = 1, 2, \ldots, \tag{24}$$

и если X_* — *аффинное множество, то* $v_* = \mathcal{P}_{X_*}(x_0)$, *т. е.* v_* — *ближайшая к* x_0 *точка из* X_*.

Доказательство. Сначала покажем возможность выбора α_k из условий (21), (22). Пусть $j \geqslant 0$ — наименьший номер, для которого

$$L \cdot 2^{-j}\alpha \leqslant 2 - \varepsilon; \tag{25}$$

здесь $L > 0$ — константа Липшица для $f'(x)$. Из неравенства (2.6.7) при $y = x_k$, $x = x_k - 2^{-j}\alpha f'(x_k)$ с учетом (25) имеем

$$f(x_k) - f(x_k - 2^{-j}\alpha f'(x_k)) \geqslant \langle f'(x_k), 2^{-j}\alpha f'(x_k)\rangle - L \cdot 2^{-2j-1}\alpha^2|f'(x_k)|^2 = 2^{-j-1}\alpha(2 - 2^{-j}\alpha L)|f'(x_k)|^2 \geqslant 2^{-j-1}\alpha\varepsilon|f'(x_k)|^2. \tag{26}$$

Это значит, что при $i = j$ неравенство (21) выполняется, и, следовательно, минимальный номер $i \geqslant 0$, при котором справедливо (21), существует и не превышает номера j из (25). Покажем, что для α_k из (21), (22) справедлива оценка

$$\alpha_k \geqslant \min\{(2 - \varepsilon)/(2L); \alpha\}, \quad k = 0, 1, \ldots \tag{27}$$

Сначала рассмотрим случай $\alpha > (2 - \varepsilon)/(2L)$. Тогда оказывается, $\alpha_k > (2 - \varepsilon)/(2L)$ при всех $k = 0, 1, \ldots$ В самом деле, для номера j из (25) в этом случае имеем $2^{-j}\alpha \leqslant (2 - \varepsilon)/L < 2^{-j+1}\alpha$, $j \geqslant 0$. Поэтому с учетом правила выбора номера i, определения α_k из (22) и неравенства $i \leqslant j$ получим $\alpha_k = \alpha/2^i \geqslant \alpha/2^j > (2 - \varepsilon)/(2L)$. Пусть теперь $\alpha \leqslant (2 - \varepsilon)/(2L)$. Тогда неравенство (25) и, следовательно, (26) выполняется при $j = 0$. Отсюда и из (21) следует, что $i = 0$. Согласно (22) тогда $\alpha_k = \alpha/2^0 = \alpha$, $k = 0, 1, \ldots$ Объединяя оба рассмотренных случая, приходим к оценке (27).

Далее, возьмем любую точку $x_* \in X_*$. Из (3), (21), (22) и теоремы 4.2.2 имеем

$$(\varepsilon/2)\alpha_k|f'(x_k)|^2 \leqslant f(x_k) - f(x_{k+1}) \leqslant f(x_k) - f(x_*) \leqslant \langle f'(x_k), x_k - x_*\rangle. \tag{28}$$

Кроме того, из (3) следует $|x_{k+1} - x_*|^2 = |x_k - \alpha_k f'(x_k) - x_*|^2 = |x_k - x_*|^2 - 2\alpha_k\langle f'(x_k), x_k - x_*\rangle + \alpha_k^2|f'(x_k)|^2$. Отсюда с учетом оценки (28) получаем

$$|x_{k+1} - x_*|^2 \leqslant |x_k - x_*|^2 - (\varepsilon - 1)\alpha_k^2|f'(x_k)|^2, \quad 1 < \varepsilon < 2. \tag{29}$$

Следовательно,

$$|x_{k+1} - x_*|^2 \leqslant |x_k - x_*|^2 \leqslant \ldots \leqslant |x_0 - x_*|^2 \quad \forall x_* \in X_*. \tag{30}$$

Из (30) вытекает существование предела $\lim\limits_{k\to\infty}|x_k - x_*|^2$ и ограниченность последовательности $\{x_k\}$. Тогда найдется подпоследовательность $\{x_{k_m}\}$, сходящаяся к некоторой точке v_*. Из (27), (29) следует, что $\{f'(x_{k_m})\} \to f'(v_*) = 0$. По теореме 4.2.3 тогда $v_* \in X_*$. Приняв $x_* = v_*$, из (30) получаем $\lim\limits_{k\to\infty}|x_k - v_*| = \lim\limits_{m\to\infty}|x_{k_m} - v_*| = 0$, т. е. вся последовательность $\{x_k\}$ сходится к точке v_*. Отсюда и из (29), (30) следуют неравенства (23). Как видно из (29), равенство в (23) возможно лишь при $f'(x_k) = 0$. Тогда в силу теоремы 4.2.3 $x_k = v_* \in X_*$, и процесс (3), (21), (22) на этом заканчивается.

Докажем оценку (24). Обозначим $a_k = f(x_k) - f_*$. Из (28), (30) при $x_* = v_*$ имеем

$$(\varepsilon/2)\alpha_k|f'(x_k)|^2 \leqslant a_k - a_{k+1}, \quad a_k \leqslant |f'(x_k)||x_0 - v_*|, \quad k = 0, 1, \ldots$$

Отсюда с учетом (27) получаем $a_k - a_{k+1} \geqslant (\varepsilon/2)\min\{(2-\varepsilon)/(2L); \alpha\}|x_0 - v_*|^{-2}a_k^2$, $k = 0, 1, \ldots$ Из леммы 2.6.4 тогда следует оценка (24).

Наконец, пусть X_* — аффинное множество. При $k \to \infty$ из (30) имеем $|v_* - x_*|^2 \leqslant |x_0 - x_*|^2$ при любом $x_* \in X_*$. В частности, в этом неравенстве можно взять $x_* = v_* + \alpha(\mathcal{P}_{X_*}(x_0) - v_*) = v_\alpha \in X_*$, $\alpha > 0$. Получим $|x_0 - v_\alpha|^2 \geqslant |v_* - v_\alpha|^2 = |(v_* - x_0) - (v_\alpha - x_0)|^2 = |v_* - x_0|^2 + |v_\alpha - x_0|^2 - 2\langle v_* - x_0, v_\alpha - x_0\rangle = |v_\alpha - x_0|^2 - |v_* - x_0|^2 - 2\alpha\langle v_* - x_0, \mathcal{P}_{X_*}(x_0) - v_*\rangle$ или

$$|v_* - x_0|^2 \geqslant 2\alpha|v_* - \mathcal{P}_{X_*}(x_0)|^2 + 2\alpha\langle\mathcal{P}_{X_*}(x_0) - x_0, v_* - \mathcal{P}_{X_*}(x_0)\rangle.$$

Отсюда с учетом равенства (4.4.2) имеем $|v_* - x_0|^2 \geqslant 2\alpha|v_* - \mathcal{P}_{X_*}(x_0)|^2$ при всех $\alpha > 0$. Разделив это неравенство на $\alpha > 0$ и устремив $\alpha \to \infty$, получим $v_* = \mathcal{P}_{X_*}(x_0)$. Теорема 4 доказана. □

5. Следуя [525], рассмотрим метод, представляющий собой комбинацию несколько модифицированного метода (3), (21), (22) и овражного метода.

Возьмем начальные приближения: $v_0 \in E^n$, $b_0 = 1$, $\alpha_{-1} > 0$, положим $x_{-1} = v_0$. Пусть для некоторого $k \geqslant 0$ уже известны $v_k \in E^n$, $b_k \geqslant 1$, $\alpha_{k-1} > 0$, $x_{k-1} \in E^n$. Определим наименьший номер $i \geqslant 0$, для которого выполняется неравенство

$$f(v_k) - f(v_k - 2^{-i}\alpha_{k-1}f'(v_k)) \geqslant 2^{-i-1}\alpha_{k-1}|f'(v_k)|^2. \tag{31}$$

Далее, положим

$$\alpha_k = \alpha_{k-1}/2^i, \quad x_k = v_k - \alpha_k f'(v_k), \tag{32}$$

$$b_{k+1} = \frac{1}{2}\left(1 + \sqrt{4b_k^2 + 1}\right), \quad v_{k+1} = x_k + \frac{b_k - 1}{b_{k+1}}(x_k - x_{k-1}). \tag{33}$$

Таким образом, в описанном методе (31)–(33) спуск из точки v_k на «дно оврага» осуществляется по формулам (31), (32) с помощью одного шага градиентного метода (3) с правилом выбора параметра α_k, близким к (21), (22). Здесь возможно использование некоторых других вариантов градиентного метода: по аналогии с (8) в (32) можно взять $\alpha_k = 1/L$. Как видно из (33), пересчет точки v_k осуществляется с помощью овражного метода по формуле, близкой к (19). Первое из равенств (33) представляет собой правило пересчета длины овражного шага; величина b_{k+1} является положительным корнем квадратного уравнения $x^2 - x - b_k^2 = 0$, так что

$$b_{k+1}^2 - b_{k+1} = b_k^2, \quad b_0 = 1, \quad b_k > 0, \quad k = 0, 1, \ldots \tag{34}$$

С помощью индукции нетрудно получить оценку

$$b_k > k, \quad k = 1, 2, \ldots \tag{35}$$

Теорема 5. *Пусть функция* $f(x)$ *выпукла на* E^n, $f(x) \in C^{1,1}(E^n)$, $f_* > -\infty$, $X_* \neq \varnothing$, *последовательность* $\{x_k\}$ *определена методом* (31)–(33). *Тогда*

$$0 \leqslant f(x_k) - f_* \leqslant (\min\{1/(2L); \alpha_{-1}\})^{-1}(2\alpha_0(f(x_0) - f_*) + \inf_{x \in X_*}|x - x_0|^2)/(2k^2) = O(1/k^2), \quad k = 1, 2, \ldots \tag{36}$$

Доказательство. Пусть $j \geqslant 0$ — наименьший номер, для которого

$$2^{-j}\alpha_{k-1} \leqslant 1/L. \tag{37}$$

Нетрудно видеть, что тогда

$$f(v_k) - f(v_k - 2^{-j}\alpha_{k-1}f'(v_k)) \geqslant 2^{-j-1}\alpha_{k-1}|f'(v_k)|^2; \tag{38}$$

это неравенство доказывается так же, как (26) при $\varepsilon = 1$. Отсюда следует существование номера $i \leqslant j$, удовлетворяющего неравенству (31). Рассуждая так же, как при доказательстве оценки (27), из (31), (32), (37), (38) с помощью индукции получаем

$$\alpha_k \geqslant \min\{1/(2L);\ \alpha_{-1}\}, \quad k = 0, 1, \ldots \tag{39}$$

Обозначим $p_k = (b_k - 1)(x_{k-1} - x_k)$. Тогда из (33) следует

$$v_{k+1} = x_k - p_k/b_{k+1}, \quad p_k = b_{k+1}(x_k - v_{k+1}). \tag{40}$$

Далее, с учетом (32), (40) имеем

$$p_{k+1} - x_{k+1} = (b_{k+1} - 1)(x_k - x_{k+1}) - x_{k+1} = b_{k+1}(x_k - x_{k+1}) - x_k =$$
$$= b_{k+1}(x_k - v_{k+1} + \alpha_{k+1} f'(v_{k+1})) - x_k = p_k - x_k + \alpha_{k+1} b_{k+1} f'(v_{k+1}).$$

Тогда для любого $x_* \in X_*$ получаем

$$|p_{k+1} - x_{k+1} + x_*|^2 = |p_k - x_k + x_*|^2 + 2\alpha_{k+1} b_{k+1} \langle f'(v_{k+1}), p_k - x_k + x_* \rangle + \alpha_{k+1}^2 b_{k+1}^2 |f'(v_{k+1})|^2,$$

или с учетом (40)

$$|p_{k+1} - x_{k+1} + x_*|^2 - |p_k - x_k + x_*|^2 = 2\alpha_{k+1} \langle f'(v_{k+1}), (b_{k+1} p_k - p_k) + (p_k - b_{k+1} x_k) + b_{k+1} x_* \rangle +$$
$$+ \alpha_{k+1}^2 b_{k+1}^2 |f'(v_{k+1})|^2 = 2\alpha_{k+1}(b_{k+1} - 1)\langle f'(v_{k+1}), p_k \rangle +$$
$$+ 2\alpha_{k+1} b_{k+1} \langle f'(v_{k+1}), x_* - v_{k+1} \rangle + \alpha_{k+1}^2 b_{k+1}^2 |f'(v_{k+1})|^2, \quad k = 0, 1, \ldots \tag{41}$$

Заметим, что (31) с учетом (32) можно переписать в виде $f(v_k) - f(x_k) \geqslant (\alpha_k/2)|f'(v_k)|^2$. Для $k+1$-й итерации это неравенство имеет вид

$$f(v_{k+1}) \geqslant f(x_{k+1}) + \frac{1}{2}\alpha_{k+1} |f'(v_{k+1})|^2. \tag{42}$$

Из теоремы 4.2.2 с учетом (42) получаем

$$\langle f'(v_{k+1}), x_* - v_{k+1} \rangle \leqslant f(x_*) - f(v_{k+1}) \leqslant f_* - f(x_{k+1}) - \frac{1}{2}\alpha_{k+1} |f'(v_{k+1})|^2. \tag{43}$$

Далее, из теоремы 4.2.2 и из (40), (42) следует

$$(\alpha_{k+1}/2)|f'(v_{k+1})|^2 \leqslant f(v_{k+1}) - f(x_{k+1}) \leqslant f(x_k) - \langle f'(v_{k+1}), x_k - v_{k+1} \rangle - f(x_{k+1}) =$$
$$= f(x_k) - f(x_{k+1}) - \langle f'(v_{k+1}), p_k \rangle / b_{k+1},$$

откуда

$$\langle f'(v_{k+1}), p_k \rangle \leqslant b_{k+1}[(f(x_k) - f(x_{k+1})) - (\alpha_{k+1}/2)|f'(v_{k+1})|^2]. \tag{44}$$

Обозначим $a_k = f(x_k) - f_*$. Подставим оценки (43), (44) в (41). С учетом (32), (34) получим

$$|p_{k+1} - x_{k+1} + x_*|^2 - |p_k - x_k + x_*|^2 \leqslant$$
$$\leqslant 2\alpha_{k+1}(b_{k+1} - 1) b_{k+1}(a_k - a_{k+1} - (\alpha_{k+1}/2)|f'(v_{k+1})|^2) +$$
$$+ 2\alpha_{k+1} b_{k+1}(-a_{k+1} - (\alpha_{k+1}/2)|f'(v_{k+1})|^2) + \alpha_{k+1}^2 b_{k+1}^2 |f'(v_{k+1})|^2 =$$
$$= 2\alpha_{k+1} b_k^2 a_k - 2\alpha_{k+1} a_{k+1}(b_k^2 + b_{k+1}) \leqslant 2\alpha_k b_k^2 a_k - 2\alpha_{k+1} b_{k+1}^2 a_{k+1}.$$

Таким образом,

$$|p_{m+1} - x_{m+1} + x_*|^2 - |p_m - x_m + x_*|^2 \leqslant 2\alpha_m b_m^2 a_m - 2\alpha_{m+1} b_{m+1}^2 a_{m+1}, \quad m = 0, 1, \ldots$$

Суммируя эти неравенства по m от 0 до некоторого $m = k - 1$, получим

$$|p_k - x_k + x_*|^2 + 2\alpha_k b_k^2 a_k \leqslant 2\alpha_0 b_0^2 a_0 + |p_0 - x_0 + x_*|^2.$$

Отсюда с учетом равенств $b_0 = 1$, $p_0 = 0$, оценок (35), (39), произвольности выбора точки x_* из X_* приходим к оценке (36). Теорема 5 доказана. □

Отметим, что метод (31)–(33) не обеспечивает монотонное убывание функции $f(x)$ на последовательностях $\{x_k\}$, $\{v_k\}$. Сравнение оценок (24) и (36) показывает, что для выпуклых гладких задач овражный метод имеет более высокую скорость сходимости, чем градиентный метод (3), (9). В [523] показано, что оценка $f(x_k) - f_* = O(1/k^2)$ является неулучшаемой на этом классе функции среди всех методов, использующих лишь значения $f(x)$, $f'(x)$.

6. Остановимся на непрерывном варианте градиентного метода. В этом методе вместо итерационного процесса (3), порождающего траекторию $\{x_k\}$, которая зависит от дискретного времени $k = 0, 1, \ldots$, за основу берется система дифференциальных уравнений:

$$\dot{x}(t) = -\alpha(t) f'(x(t)), \quad t \geqslant 0, \tag{45}$$

где $\alpha(t) > 0$ заданная функция (параметр метода). Эта система описывает движение материальной точки, движущейся в силовом поле, задаваемом антиградиентом $(-f'(x))$, со скоростью $\dot{x}(t)$, пропорциональной антиградиенту в точке $x(t)$. Сразу заметим, что итерационный процесс (3) представляет собой известный метод ломаных Эйлера для приближенного определения траектории системы (45), выходящей из точки $x(0) = x_0$. По аналогии с теоремами 1–3 можно надеяться, что при некоторых ограничениях на функцию $f(x)$, $\alpha(t)$ траектории $x(t)$, $t \geqslant 0$, системы (45) при больших t притягиваются ко множеству $S_* = \{x \in E^n : f'(x_*) = 0\}$ стационарных точек задачи (1) или, в лучшем случае, ко множеству X_* решений задачи (1). Очевидно, все точки множеств S_*, X_* являются точками равновесия (стационарными решениями) системы (45). Приведем две теоремы о сходимости метода (45).

Теорема 6. *Пусть функция $f(x) \in C^{1,1}(E^n)$, выпукла на E^n, $f_* > -\infty$, $X_* \neq \varnothing$, а функция $\alpha(t)$ непрерывно дифференцируема при $t \geqslant 0$, $\alpha(t) \geqslant \alpha_0 > 0$, $\alpha'(t) \leqslant 0$ $\forall t \geqslant 0$. Тогда траектория $x(t)$ системы (45) с любым начальным условием $x(0) = x_0$ определена при всех $t \geqslant 0$ и существует точка $v_* \in X_*$ такая, что*

$$\lim_{t \to +\infty} x(t) = v_*, \qquad \lim_{t \to +\infty} \dot{x}(t) = 0,$$
$$\lim_{t \to +\infty} f(x(t)) = f(v_*) = f_*, \qquad \lim_{t \to +\infty} f'(x(t)) = f'(v_*) = 0.$$

Доказательство. При выполнении условий теоремы $0 < \alpha_0 \leqslant \alpha(t) \leqslant \alpha(0)$ и правая часть $(-\alpha(t) f'(x))$ дифференциального уравнения (45) удовлетворяет условию Липшица по x, непрерывна по совокупности аргументов (t, x). Тогда задача Коши для уравнения (45) с начальным условием $x(0) = x_0$ имеет решение $x = x(t)$, определенное при всех $t \geqslant 0$ (см. ниже теорему 6.1.1). Возьмем $\forall x_* \in X_*$ и умножим (45) скалярно на $x(t) - x_*$:

$$\langle \dot{x}(t), x(t) - x_* \rangle = \frac{1}{2}\frac{d}{dt}|x(t) - x_*|^2 = -\alpha(t)\langle f'(x(t)), x(t) - x_* \rangle.$$

Отсюда с учетом равенства $f'(x_*) = 0$, условия $\alpha(t) > 0$ и теоремы 4.2.4 имеем: $\frac{d}{dt}|x(t) - x_*|^2 = -2\alpha(t)\langle f'(x(t)) - f'(x_*), x(t) - x_* \rangle \leqslant 0$ $\forall t \geqslant 0$. Таким образом, функция $|x(t) - x_*|^2$ не возрастает при $t \geqslant 0$, т. е.

$$|x(t) - x_*|^2 \leqslant |x(\tau) - x_*|^2 \quad \forall t > \tau \geqslant 0, \quad \forall x_* \in X_*. \tag{46}$$

В частности, при $\tau = 0$: $|x(t) - x_*| \leqslant |x_0 - x_*|$, т. е. траектория $x(t)$ ограничена равномерно на $t \geqslant 0$. Далее, умножим уравнение (45) скалярно на $\dot{x}(t)$:

$$|\dot{x}(t)|^2 = -\alpha(t)\langle f'(x(t)), \dot{x}(t) \rangle = -\alpha(t)\frac{d}{dt}(f(x(t) - f(x_*)), \quad t \geqslant 0.$$

Интегрируя это равенство и преобразуя по частям, получим:

$$\int_0^t |\dot{x}(\tau)|^2 d\tau = -\alpha(\tau)(f(x(\tau)) - f(x_*))\Big|_{\tau=0}^{\tau=t} + \int_0^t \alpha'(\tau)(f(x(\tau)) - f(x_*))d\tau.$$

Так как $0 < \alpha_0 \leqslant \alpha(t) \leqslant \alpha(0)$, $\alpha'(t) \leqslant 0$, $f(x(t)) - f(x_*) \geqslant 0$ $\forall t \geqslant 0$, то

$$\int_0^t |\dot{x}(\tau)|^2 d\tau \leqslant \alpha(0)(f(x_0) - f(x_*)) \quad \forall t \geqslant 0.$$

Это значит, что $\int_0^\infty |\dot{x}(\tau)|^2 d\tau < \infty$. Тогда найдется последовательность $\{t_i\} \to +\infty$, что $\{\dot{x}(t_i)\} \to 0$. Так как $|x(t)|$ ограничено при $t \geqslant 0$, то, пользуясь теоремой Больцано — Вейерштрасса, можем считать, что $x(t_i) \to v_*$. Из (45) при $t = t_i \to \infty$ с учетом $\lim_{t\to\infty} \alpha(t) \geqslant \alpha_0 > 0$ получим $f'(v_*) = 0$. Отсюда и из выпуклости $f(x)$ следует, что $v_* \in X_*$. Из (46) при $\tau = t_i$, $x_* = v_*$, имеем: $|x(t) - v_*|^2 \leqslant |x(t_i) - v_*|^2$ $\forall t > t_i$. Переходя к пределу сначала при $t \to +\infty$, затем $i \to \infty$, отсюда получим $\lim_{t\to\infty} x(t) = v_*$. Тогда $\lim_{t\to\infty} f(x(t)) = f(v_*) = f_*$, $\lim_{t\to\infty} f'(x(t)) = f'(v_*) = 0$, а из (45) следует: $\lim_{t\to\infty} \dot{x}(t) = 0$. Теорема 6 доказана. □

Для сильно выпуклых функций несложно получить оценку скорости метода (45).

Теорема 7. *Пусть функция $f(x) \in C^{1,1}(E^n)$ и сильно выпукла на E^n, а функция $\alpha(t)$, $\forall t \geqslant 0$, $\int_0^\infty \alpha(t)dt = +\infty$. Тогда для траектории $x(t)$ системы* (45) *с любым начальным условием $x(0) = x_0$ справедлива оценка:*

$$|x(t) - x_*| \leqslant |x_0 - x_*| \exp\left(-\mu \int_0^t \alpha(\tau)d\tau\right) \quad \forall t \geqslant 0, \tag{47}$$

где постоянная $\mu > 0$ взята из теоремы 4.3.3.

Доказательство. Прежде всего заметим, что по теореме 4.3.1 точка минимума x_* функции $f(x)$ на E^n существует и единственна, а по теореме 4.2.3 $f'(x_*) = 0$. Положим

$$V(t) = \frac{1}{2}|x(t) - x_*|^2, \quad t \geqslant 0. \tag{48}$$

Тогда с учетом (45) и теоремы 4.3.3 имеем:

$$\dot{V}(t) = \langle x(t) - x_*, \dot{x}(t)\rangle = -\alpha(t)\langle f'(x(t)) - f'(x_*), x(t) - x_*\rangle \leqslant$$
$$\leqslant -\mu\alpha(t)|x(t) - x_*|^2 = -2\mu\alpha(t)V(t), \quad \forall t \geqslant 0; \quad V(0) = |x_0 - x_*|^2/2.$$

Отсюда следует: $\frac{d}{dt}\left(V(t)\exp\left(2\mu\int_0^t \alpha(\tau)d\tau\right)\right) \leqslant 0$ $\forall t \geqslant 0$. Интегрируя это неравенство, получим

$$0 \leqslant V(t) \leqslant V(0)\exp\left(-2\mu\int_0^t \alpha(\tau)d\tau\right) = |x_0 - x_*|^2 \exp\left(-2\mu\int_0^t \alpha(\tau)d\tau\right)/2,$$

что равносильно оценке (47). Теорема 7 доказана. □

Пользуясь терминологией, принятой в теории устойчивости обыкновенных дифференциальных уравнений [328; 376; 588; 694], можно сказать, что в теореме 7 доказана асимптотическая устойчивость системы (45) относительно точки равновесия x_* этой системы. Для доказательства этого факта использован второй метод Ляпунова, в качестве функции Ляпунова была взята функция (48). В связи с этим полезно заметить, что при исследовании многих методов минимизации явно или неявно используется второй метод Ляпунова или его дискретный аналог; в качестве функции Ляпунова наряду с (48) часто используются также функции $V(t) = f(x(t)) - f_*$, $V(t) = |f'(x(t))|^2$ и др. Систематическое исследование сходимости методов минимизации с помощью метода Ляпунова проведено в [77].

Существуют и другие дифференциальные уравнения, траектории которых являются минимизирующими. Например, так называемый метод тяжелого шарика [74] заключается в рассмотрении системы дифференциальных уравнений вида:

$$\ddot{x}(t) + \dot{x}(t) + \alpha(t)f'(x(t)) = 0, \quad t \geqslant 0, \quad \alpha(t) > 0. \tag{49}$$

Оказывается, траектории системы (49) при довольно широких предположениях сходятся к точке минимума функции $f(x)$ на E^n, причем скорость сходимости, вообще говоря, выше, чем у траекторий системы (45).

Следует заметить, что непрерывные методы минимизации привлекательны тем, что для приближенного решения возникающих здесь задач Коши могут быть использованы не только метод ломаных Эйлера, но и другие известные методы [59; 74; 89; 481], которые, возможно, будут сходиться быстрее и лучше приспособлены для минимизации овражных функций, приводящих к так называемым *жестким системам* дифференциальных уравнений. На этом пути можно получить различные классы дискретных методов минимизации, которые подчас трудно обнаружить, оставаясь в рамках привычных представлений, навязанных итеративными схемами. Перечисленные обстоятельства стимулируют развитие непрерывных методов решения экстремальных задач (см., например, [25; 26; 28–30; 732]). Непрерывные аналоги некоторых методов изложены ниже в §§ 2, 6, 11.

7. В заключение отметим, что градиентный метод, вообще говоря, хорошо работает лишь на первых этапах поиска минимума, когда точки x_k из (3) не слишком близки к точке минимума x_*, а вблизи точки x_* расстояние $|x_k - x_*|$ часто перестает уменьшаться, сходимость метода ухудшается. Это связано с тем, что в окрестности точки минимума градиент $f'(x_k)$ близок к нулю, главная линейная часть приращения $f(x_k) - f(x_*)$, на базе которой выбирается направление спуска в методе (3), становится малой, усиливается влияние квадратичной части приращения, метод (3) становится слишком чувствительным к неизбежным погрешностям вычислений. Поэтому вблизи точки минимума при необходимости пользуются более точными и, вообще говоря, более трудоемкими методами, лучше учитывающими не только линейные, но и квадратичные части приращения.

Упражнения

1. Описать различные варианты градиентного метода для задачи из примера 2.2.2.

2. Установить сходимость, метода скорейшего спуска для функции (5); описать другие варианты градиентного метода для этой функции.

3. Рассмотреть метод скорейшего спуска и другие варианты градиентного метода для задачи минимизации функции $f(x) = |Ax - b|^2$, $x \in E^n$, где A — матрица порядка $m \times n$, $b \in E^n$; исследовать их сходимость.

4. Рассмотреть метод скорейшего спуска для минимизации функций $f(u) = x^2 + ay^2$, $u = (x, y) \in E^2$, и $f(u) = x^2 + y^2 + az^2$, $u = (x, y, z) \in E^3$, при различном начальном приближении u_0, считая коэффициент a намного больше единицы.

5. Доказать теоремы 1, 2 для метода (3), (7).

§ 2. Метод проекции градиента

1. Будем рассматривать задачу

$$f(x) \to \inf; \quad x \in X \subseteq E^n, \tag{1}$$

где множество X необязательно совпадает со всем пространством E^n, а функция $f(x) \in C^1(X)$. Непосредственное применение описанного выше градиентного метода в случае $X \neq E^n$ может привести к затруднениям, так как точка x_{k+1} из (1.3) при каком-то k может не принадлежать X. Однако эту трудность можно преодолеть, если полученную с помощью формулы (1.3) точку $x_k - \alpha_k f'(x_k)$ при каждом k проектировать на множество X (см. определение 4.4.1). В результате мы придем к так называемому *методу проекции градиента*.

А именно, пусть $x_0 \in X$ — некоторое начальное приближение. Далее будем строить последовательность $\{x_k\}$ по правилу

$$x_{k+1} = \mathcal{P}_X(x_k - \alpha_k f'(x_k)), \quad k = 0, 1, \ldots, \tag{2}$$

где α_k — положительная величина. Если X — выпуклое замкнутое множество и способ выбора $\{\alpha_k\}$ в (2) задан, то в силу теоремы 4.4.1 последовательность $\{x_k\}$ будет однозначно определяться условием (2). В частности, при $X = E^n$ метод (2) превратится в градиентный метод.

Если в (2) на некоторой итерации оказалось $x_{k+1} = x_k$ (например, это случится при $f'(x_k) = 0$), то процесс (2) прекращают. В этом случае точка x_k удовлетворяет необходимому условию оптимальности $x_k = \mathcal{P}_X(x_k - \alpha_k f'(x_k))$ (см. теорему (4.4.3), и для выяснения того, является ли в действительности x_k решением задачи (1) или нет, при необходимости нужно провести дополнительное исследование поведения функции $f(x)$ в окрестности точки x_k. В частности, если $f(x)$ — выпуклая функция, то такая точка x_k является решением задачи (1).

В зависимости от способа выбора α_k в (2) можно получить различные варианты метода проекции градиента. Укажем несколько наиболее употребительных на практике способов выбора α_k.

1) Введем функцию одной переменной $g_k(\alpha) = f(\mathcal{P}_X(x_k - \alpha f'(x_k)))$, $\alpha \geqslant 0$, и определим α_k из условий

$$g_k(\alpha_k) = \inf_{\alpha \geqslant 0} g_k(\alpha) = g_{k*}, \quad \alpha_k > 0. \tag{3}$$

Очевидно, при $X = E^n$ метод (2), (3) превратится в метод скорейшего спуска. Поскольку величину α_k из условий (3) удается найти точно лишь в редких случаях (возможно также, что нижняя грань в (3) не всегда достигается), то α_k на практике определяют приближенно из условий типа (1.6) или (1.7).

2) Иногда приходится довольствоваться нахождением какого-либо $\alpha_k > 0$, обеспечивающего условие монотонности: $f(x_{k+1}) < f(x_k)$. Для этого обычно выбирают какую-либо постоянную $\alpha > 0$ и в методе (2) на каждой итерации берут $\alpha_k = \alpha$, а затем проверяют условие монотонности и при необходимости дробят величину $\alpha_k = \alpha$, добиваясь выполнения условия монотонности.

3) Если функция $f(x)$ принадлежит $C^{1,1}(X)$ и константа Липшица L для градиента $f'(x)$ известна, то в (2) в качестве α_k можно взять любое число, удовлетворяющее условиям

$$0 < \varepsilon_0 \leqslant \alpha_k \leqslant 2/(L + 2\varepsilon), \tag{4}$$

где ε_0, ε — положительные числа, являющиеся параметрами метода.

4) Возможен выбор α_k из условия

$$f(x_k) - f(\mathcal{P}_X(x_k - \alpha_k f'(x_k))) \geqslant \varepsilon |x_k - \mathcal{P}_X(x_k - \alpha_k f'(x_k))|^2, \tag{5}$$

где $\varepsilon > 0$ — параметр метода. Для определения такого α_k можно взять какое-либо число $\alpha_k = \alpha$ (например, $\alpha = 1$) и затем дробить его до тех пор, пока не выполнится условие (5). Если $f(x) \in C^{1,1}(X)$, то нетрудно показать, что выполнения условия (5) можно добиться за конечное число дроблений.

5) Возможно априорное задание величин α_k из условий

$$\alpha_k > 0, \quad k = 0, 1, \ldots; \quad \sum_{k=0}^{\infty} \alpha_k = \infty, \quad \sum_{k=0}^{\infty} \alpha_k^2 < \infty, \tag{6}$$

например, $\alpha_k = (k+1)^{-1}$, $k = 0, 1, \ldots$ Сходимость метода (2), (6) будет исследована в § 3.

Заметим, что описанные здесь варианты метода (2) при $X = E^n$ переходят в соответствующие варианты градиентного метода.

На практике для ускорения сходимости вместо (2) часто пользуются более общим вариантом метода проекции градиента

$$x_{k+1} = x_k + \beta_k(\mathcal{P}_X(x_k - \alpha_k f'(x_k)) - x_k) = \\ = \beta_k \mathcal{P}_X(x_k - \alpha_k f'(x_k)) + (1 - \beta_k)x_k, \quad 0 < \beta_k \geqslant 1, \ \alpha_k > 0, \tag{2'}$$

где параметры α_k, β_k могут выбираться различными способами.

Заметим, что в методах (2) или (2') на каждой итерации, кроме выбора параметров α_k, β_k, нужно еще проектировать точку на множество X или, иначе говоря, решить задачу минимизации

$$\Phi_k(x) = |x - (x_k - \alpha_k f'(x_k))|^2 \to \inf, \quad x \in X; \tag{7}$$

здесь возможно использование функции $\Phi_k(x) = |x - x_k|^2 + 2\alpha_k \langle f'(x_k), x - x_k \rangle$, отличающейся от предыдущей функции постоянным слагаемым. Задачу (7) можно решать приближенно и вместо точки $x_{k+1} \in X$, $\Phi_k(x_{k+1}) = \inf_X \Phi_k(x) = \Phi_{k*}$ определить ее приближение z_{k+1} из условий

$$z_{k+1} \in X: \ \Phi_k(z_{k+1}) \leqslant \Phi_{k*} + \delta_k^2. \tag{8}$$

Предполагая, что X — выпуклое замкнутое множество, из (8) с помощью неравенства (4.3.3) имеем $|z_{k+1} - x_{k+1}|^2 \leqslant \Phi_k(z_{k+1}) - \Phi_k(x_{k+1}) \leqslant \delta_k^2$ или

$$z_{k+1} \in X: \ |z_{k+1} - x_{k+1}| \leqslant \delta_k.$$

Конечно, задачи (7), (8) далеко не всегда просто решаются. Поэтому методом проекции градиента обычно пользуются лишь в тех случаях, когда проекция точки на множество легко определяется. Например, когда множество X представляет собой шар в E^n, параллелепипед, гиперплоскость, полупространство или положительный ортант (см. примеры 4.4.1–4.4.6), задача проектирования точки решается просто и в явном виде, и реализация каждой итерации метода проекции градиента в этом случае не вызывает особых затруднений. Если же задача проектирования для своего решения в свою очередь требует применения тех или иных итерационных методов, то эффективность метода проекции градиента, вообще говоря, значительно снижается.

2. Остановимся на вопросах сходимости метода (2), (4).

Теорема 1. *Пусть X — выпуклое замкнутое множество из E^n, функция $f(x) \in C^{1,1}(X)$, $\inf_X f(x) = f_* > -\infty$. Тогда для последовательности $\{x_k\}$, полученной методом (2), (4) при любом начальном приближении x_0, имеет место соотношение $\lim_{k\to\infty} |x_{k+1} - x_k| = 0$. Если при этом множество $M(x_0) = \{x: x \in X, f(x) \leqslant f(x_0)\}$ ограничено, то $\lim_{k\to\infty} \rho(x_k, S_*) = 0$, где $S_* = \{x: x \in M(x_0), \langle f'(x), v - x \rangle \geqslant 0$ при всех $v \in X\}$.*

Доказательство. Из неравенства (2.6.7) при $y = x_k$, $x = x_{k+1}$ имеем

$$f(x_k) - f(x_{k+1}) \geqslant \langle f'(x_k), x_k - x_{k+1} \rangle - (L/2)|x_k - x_{k+1}|^2, \quad k = 0, 1, \ldots \tag{9}$$

Из (2) и теоремы 4.4.1 следует, что

$$\langle x_{k+1} - [x_k - \alpha_k f'(x_k)], x - x_{k+1} \rangle \geqslant 0 \qquad \forall x \in X.$$

Перепишем это неравенство в виде

$$\langle f'(x_k), x - x_{k+1} \rangle \geqslant \langle x_k - x_{k+1}, x - x_{k+1} \rangle / \alpha_k, \quad k = 0, 1, \ldots \tag{10}$$

Отсюда при $x = x_k$ с учетом условия (4) получим

$$\langle f'(x_k), x_k - x_{k+1}\rangle \geqslant |x_k - x_{k+1}|^2/\alpha_k \geqslant (L/2 + \varepsilon)|x_k - x_{k+1}|^2.$$

Подставим эту оценку в (9):

$$f(x_k) - f(x_{k+1}) \geqslant \varepsilon |x_k - x_{k+1}|^2, \qquad k = 0, 1, \ldots \tag{11}$$

Так как $f(x_k) \geqslant f_* > -\infty$ и последовательность $\{f(x_k)\}$ — убывающая, то существует конечный предел $\lim\limits_{k\to\infty} f(x_k) \geqslant f_*$ и, следовательно, $\lim\limits_{k\to\infty}(f(x_k) - f(x_{k+1})) = 0$. Отсюда и из (11) сразу получим $\lim\limits_{k\to\infty} |x_k - x_{k+1}| = 0$.

Пусть теперь множество $M(x_0)$ ограничено. Так как согласно (11) $f(x_{k+1}) \leqslant f(x_k) \leqslant \ldots \leqslant f(x_0)$, то $\{x_k\} \in M(x_0)$. По теореме Больцано — Вейерштрасса ограниченная последовательность $\{x_k\}$ имеет хотя бы одну предельную точку. Пусть x_* — произвольная предельная точка $\{x_k\}$ и $\{x_{k_m}\} \to x_*$. По доказанному $\lim\limits_{k\to\infty} |x_{k+1} - x_k| = 0$, поэтому $\{x_{k_m+1}\} \to x_*$. Переходя в (10) к пределу при $k = k_m \to \infty$, с учетом условий (4) и непрерывности $f'(x)$ получим $\langle f'(x_*), x - x_*\rangle \geqslant 0$ при любом $x \in X$, т. е. $x_* \in S_*$. По лемме 2.1.2 расстояние $\rho(x, S_*)$ непрерывно по x, поэтому $\lim\limits_{m\to\infty} \rho(x_{k_m}, S_*) = \rho(x_*, S_*) = 0$. Отсюда следует, что $\{\rho(x_k, S_*)\}$ имеет единственную предельную точку, равную нулю, т. е. $\lim\limits_{k\to\infty} \rho(x_k, S_*) = 0$. Теорема 1 доказана. □

Теорема 2. *Пусть выполнены все условия теоремы* 1 *и, кроме того, функция* $f(x)$ *выпукла на* X. *Тогда для последовательности* $\{x_k\}$ *из* (2), (4) *имеем*

$$\lim_{k\to\infty} f(x_k) = f_*, \quad \lim_{k\to\infty} \rho(x_k, X_*) = 0, \tag{12}$$

причем справедлива оценка

$$0 \leqslant f(x_k) - f_* \leqslant C_0 k^{-1}, \quad C_0 = \text{const} > 0, \quad k = 1, 2, \ldots \tag{13}$$

Доказательство. Из ограниченности $M(x_0)$, непрерывности $f(x)$, согласно теореме 2.1.2, следует $f_* > -\infty$, $X_* = \{x\colon x \in X,\ f(x) = f_*\} \neq \varnothing$, $X_* \subset M(x_0)$. Возьмем произвольную точку $x_* \in X_*$. Из неравенства (4.2.4) тогда имеем

$$0 \leqslant a_k = f(x_k) - f(x_*) \leqslant \langle f'(x_k), x_k - x_*\rangle =$$
$$= \langle f'(x_k), x_k - x_{k+1}\rangle - \langle f'(x_k), x_* - x_{k+1}\rangle, \quad k = 0, 1, \ldots$$

Пользуясь неравенством (10) при $x = x_*$ и условием (4) выбора α_k, отсюда получим

$$0 \leqslant a_k \leqslant \langle f'(x_k), x_k - x_{k+1}\rangle - \langle x_k - x_{k+1}, x_* - x_{k+1}\rangle/\alpha_k \leqslant$$
$$\leqslant |x_k - x_{k+1}|(\sup_{M(x_0)} |f'(x)| + D/\varepsilon_0) = C_1 |x_k - x_{k+1}|, \quad k = 0, 1, \ldots \tag{14}$$

Здесь мы учли ограниченность множества $M(x_0)$, поэтому $D = \sup\limits_{u,\, v \in M(x_0)} |u - v| < \infty$ и, кроме того, $|f'(x)| \leqslant |f'(x) - f'(x_0)| + |f'(x_0)| \leqslant L|x - x_0| + |f'(x_0)| \leqslant LD + |f'(x_0)|$ при любом $x \in M(x_0)$, так что $\sup\limits_{M(x_0)} |f'(x)| < \infty$. Из (11), (14) следует $a_k - a_{k+1} \geqslant \varepsilon C_1^{-2} a_k^2 = A a_k^2$, $k = 0, 1, \ldots$ Отсюда с помощью леммы 2.6.4 придем к оценке (13), из которой также следует первое из равенств (12). Второе равенство (12) является следствием теоремы 2.1.2. □

Рассмотрим случай сильно выпуклой функции, предполагая, что в методе (2) величина α_k выбирается постоянной.

Теорема 3. *Пусть* X *— выпуклое замкнутое множество, функция* $f(x) \in C^{1,1}(X)$ *и сильно выпукла на* X. *Пусть* $0 < \alpha < 2\mu L^{-2}$, *где постоянные* μ, L, $\mu \leqslant L$, *взяты из* (2.6.6), (4.3.12). *Тогда последовательность* $\{x_k\}$, *получаемая из* (2) *при* $\alpha_k = \alpha$, $k = 0, 1, \ldots$, *сходится к точке минимума* x_*, *причем справедлива оценка*

$$|x_k - x_*| \leqslant |x_0 - x_*|(q(\alpha))^k, \quad k = 0, 1, \ldots, \tag{15}$$

где $q(\alpha) = (1 - 2\mu\alpha + \alpha^2 L^2)^{1/2}$, $0 < q(\alpha) < 1$.

Доказательство. Введем отображение

$$Ax = \mathcal{P}_X(x - \alpha f'(x)),$$

действующее из X в X. Покажем его сжимаемость при $0 < \alpha < 2\mu L^{-2}$. С помощью теоремы 4.4.2 имеем

$$|Au - Av|^2 = |\mathcal{P}_X(u - \alpha f'(u)) - \mathcal{P}_X(v - \alpha f'(v))|^2 \leqslant$$
$$\leqslant |u - \alpha f'(u) - v + \alpha f'(v)|^2 = |u - v|^2 + \alpha^2 |f'(u) - f'(v)|^2 -$$
$$- 2\alpha\langle f'(u) - f'(v), u - v\rangle \leqslant |u - v|^2(1 + \alpha^2 L^2 - 2\mu\alpha) = q^2(\alpha)|u - v|^2,$$

т. е.

$$|Au - Av| \leqslant q(\alpha)|u - v|, \qquad u, v \in X. \tag{16}$$

Так как $0 < \alpha < 2\mu L^{-2}$, то $0 < q(\alpha) < 1$. Это значит, что отображение A — сжимающее. Заметим также, что замкнутое множество $X \subseteq E^n$ представляет собой полное метрическое пространство с метрикой $\rho(u, v) = |u - v|$. Следовательно, можно пользоваться принципом сжимающих отображений [393]. Метод (2) при $\alpha_k = \alpha$, записанный в виде $x_{k+1} = Ax_k$, представляет собой известный процесс поиска неподвижной точки x_* сжимающего отображения A, т. е. точки x_*, для которой $x_* = Ax_*$. Известно [393], что такая точка x_* существует, единственна и $\lim\limits_{k\to\infty} |x_k - x_*| = 0$. Из (16) следует, что

$$|x_k - x_m| \leqslant (q(\alpha))^k |x_0 - x_{m-k}| \quad \forall m \geqslant k.$$

Отсюда при $m \to \infty$ получим оценку (15). Так как $x_* = \mathcal{P}_X(x_* - \alpha f'(x_*))$, то из теоремы 4.4.4 следует, что x_* — точка минимума функции $f(x)$ на множестве X. Теорема 3 доказана. □

Заметим, что наименьшее значение $q(\alpha)$ из (15) при $0 < \alpha < 2\mu L^{-2}$ достигается при $\alpha_* = \mu L^{-2}$ и равно $q(\alpha_*) = (1 - (\mu/L)^2)^{1/2}$.

3. Следуя [24], рассмотрим сходимость метода (2), (4), не требуя, в отличие от теорем 1, 2, ограниченности множества $M(x_0)$. Кроме того, будем считать, что вычисление градиента функции и проектирование на множество на каждой итерации проводятся с погрешностями.

Теорема 4. *Пусть* X *— выпуклое замкнутое множество из* E^n, $\operatorname{int} X \neq \varnothing$, *функция* $f(x)$ *выпукла на* X, $f(x) \in C^{1,1}(X)$, $f_* > -\infty$, $X_* \neq \varnothing$. *Пусть вместо точного значения градиента* $f'(x)$ *и проекции* $\mathcal{P}_X(x) \equiv \mathcal{P}(x)$ *известны их приближения* $f_k'(x)$ *и соответственно* $\mathcal{P}_k(x)$ *с погрешностью*

$$|f'(x) - f_k'(x)| \leqslant \delta_k, \quad x \in X; \quad |\mathcal{P}(x) - \mathcal{P}_k(x)| \leqslant C_0\delta_k, \quad x \in E^n,$$
$$C_0 = \text{const} > 0, \quad \delta_k \geqslant 0, \quad k = 0, 1, \ldots; \quad \sum_{k=0}^{\infty} \delta_k = \delta < \infty. \tag{17}$$

Наконец, пусть последовательность $\{x_k\}$ определяется условиями

$$x_{k+1} = \mathcal{P}_k(x_k - \alpha_k f_k'(x_k)), \quad x_0 \in X, \quad k = 0, 1, \ldots, \tag{18}$$

где α_k выбирается так:

$$0 < \varepsilon_0 \leqslant \alpha_k \leqslant 2(1-\varepsilon)/L, \quad k = 0, 1, \ldots, \quad 0 < \varepsilon < 1. \tag{19}$$

Тогда $\{x_k\}$ сходится к некоторой точке $v_* \in X_*$.

Доказательство. Наряду с $\{x_k\}$ введем вспомогательную последовательность $\{v_k\}$, определяемую следующим образом:

$$v_k = \mathcal{P}(x_k - \alpha_k f'(x_k)), \quad k = 0, 1, \ldots; \quad v_0 = x_0. \tag{20}$$

Отсюда и из (18) с помощью теоремы 4.4.2 и условий (17) получаем

$$\begin{aligned} |x_{k+1} - v_k| &\leqslant |\mathcal{P}_k(x_k - \alpha_k f_k'(x_k)) - \mathcal{P}(x_k - \alpha_k f_k'(x_k))| + \\ &+ |\mathcal{P}(x_k - \alpha_k f_k'(x_k)) - \mathcal{P}(x_k - \alpha_k f'(x_k))| \leqslant C_0\delta_k + \alpha_k |f_k'(x_k) - f'(x_k)| \leqslant \\ &\leqslant C_0\delta_k + (2(1-\varepsilon)/L)\delta_k = C_1\delta_k, \quad k = 0, 1, \ldots \end{aligned} \tag{21}$$

Возьмем произвольную точку $x_* \in X_*$. Согласно теореме 4.2.3 тогда

$$\langle f'(x_*), x - x_* \rangle \geqslant 0, \quad x \in X. \tag{22}$$

Из (20) и неравенства (4.4.1) получаем

$$\langle v_k - x_k + \alpha_k f'(x_k), x - v_k \rangle \geqslant 0 \qquad \forall x \in X. \tag{23}$$

Положим в (23) $x = x_*$, а (22) умножим на $\alpha_k > 0$ и примем $x = v_k$. Сложим получившиеся неравенства

$$\langle v_k - x_k, x_* - v_k \rangle + \alpha_k \langle f'(x_k) - f'(x_*), x_* - v_k \rangle \geqslant 0, \quad k = 0, 1, \ldots. \tag{24}$$

Преобразуем каждое слагаемое в правой части (24). Прежде всего имеем

$$2\langle v_k - x_k, x_* - v_k \rangle = |x_k - x_*|^2 - |x_k - v_k|^2 - |v_k - x_*|^2. \tag{25}$$

Далее, воспользуемся неравенством (4.2.20) при $u = x_k$, $v = x_*$, $w = v_k$; получим

$$\langle f'(x_k) - f'(x_*), x_* - v_k \rangle \leqslant (L/4)|x_k - v_k|^2, \quad k = 0, 1, \ldots \tag{26}$$

Подставив (25), (26) в (24), получим

$$|x_k - x_*|^2 - |v_k - x_*|^2 - (1 - \alpha_k L/2)|x_k - v_k|^2 \geqslant 0.$$

Отсюда, учитывая условие (19), имеем

$$|x_k - x_*|^2 \geqslant |v_k - x_*|^2 + \varepsilon |x_k - v_k|^2, \quad k = 0, 1, \ldots \tag{27}$$

Далее, воспользуемся леммой 2.6.10 при $z_k = x_k$, $z_* = x_*$, $w_{k+1} = v_k$; из (17), (21), (27) получим

$$\lim_{k\to\infty} |x_k - x_*| = \lim_{k\to\infty} |v_k - x_*| < \infty, \quad \lim_{k\to\infty} |v_k - x_k| = 0. \tag{28}$$

Отсюда следует, что последовательность $\{x_k\}$ ограничена. Тогда существует хотя бы одна предельная точка v_* этой последовательности и подпоследовательность $\{x_{k_m}\}$, сходящаяся к v_*. Из (28) имеем $\lim\limits_{m\to\infty} v_{k_m} = v_*$.

Согласно (23) с учетом (19) получаем

$$\langle f'(x_k), x - v_k \rangle \geqslant -\langle v_k - x_k, x - v_k \rangle \alpha_k^{-1} \geqslant -|v_k - x_k||x - v_k|\varepsilon_0^{-1}.$$

Отсюда при $k = k_m \to \infty$ будем иметь $\langle f'(v_*), x - v_* \rangle \geqslant 0$ при всех $x \in X$. По теореме 4.2.3 тогда $v_* \in X_*$. Вспомним, что неравенство (27) было получено для любой точки $x_* \in X_*$. В частности, (27) верно и для $x_* = v_*$. Но v_* — предельная точка последовательности $\{x_k\}$. Из леммы 2.6.10 тогда следует, что $\{x_k\}$ сходится к v_*. Теорема 4 доказана. □

Замечание 1. Если в (17) $\delta_k = 0$, $k = 0, 1, \ldots$, то из (18)–(20) следует, что $x_{k+1} = v_k$, $k = 0, 1, \ldots$. Тогда из (27) имеем

$$|x_k - x_*|^2 \geqslant |x_{k+1} - x_*|^2 + \varepsilon |x_k - x_{k+1}|^2, \quad k = 0, 1, \ldots, \quad \forall x_* \in X_*.$$

Пользуясь произволом в выборе $x_* \in X_*$, отсюда получаем

$$|x_k - v_*| \geqslant |x_{k+1} - v_*|, \quad \rho(x_k, X_*) \geqslant \rho(x_{k+1}, X_*), \quad k = 0, 1, \ldots,$$

причем равенство здесь возможно лишь при $x_{k+1} = x_k$, что в силу теоремы 4.4.3 означает $x_k \in X_*$. Таким образом, при точной реализации метода (17)–(19) расстояние от точки x_k до множества X_* или до точки v_* монотонно убывает. Как мы видели, таким же свойством обладает градиентный метод (1.3), (1.21), (1.22).

4. Опираясь на неравенства, полученные при доказательстве теоремы 4, можно оценить скорость сходимости метода (2), (4) для сильно выпуклых функций, причем, в отличие от теоремы 3, новая оценка оказывается неулучшаемой на классе сильно выпуклых функций, принадлежащих $C^{1,1}(X)$.

Теорема 5. *Пусть X — выпуклое замкнутое множество из E^n, $\operatorname{int} X \neq \varnothing$, а функция $f(x)$ сильно выпукла на X и принадлежит $C^{1,1}(X)$. Пусть последовательность $\{x_k\}$ построена методом (2) при $\alpha_k = \alpha$, $k = 0, 1, \ldots$, $0 < \alpha < 2/L$. Тогда*

$$|x_k - x_*| \leqslant (q(\alpha))^k |x_0 - x_*|, \quad k = 0, 1, \ldots, \tag{29}$$

где

$$q(\alpha) = \begin{cases} 1 - \mu\alpha, & 0 < \alpha < 2(L+\mu)^{-1}, \\ L\alpha - 1, & 2(L+\mu)^{-1} \leqslant \alpha < 2L^{-1}, \end{cases}$$

$0 < q(\alpha) < 1$, постоянные μ, L взяты из (2.6.6), (4.3.12), а x_* — точка минимума $f(x)$ на X. Наименьшее значение $q(\alpha)$ при $0 < \alpha < 2L^{-1}$ достигается при $\alpha = \alpha_* = 2(L+\mu)^{-1}$ и равно $q(\alpha_*) = (L-\mu)(L+\mu)^{-1}$.

Доказательство. Из теоремы 4.3.1 следует, что $f_* > -\infty$, X_* состоит из единственной точки x_*. Тогда из теоремы 4 имеем $\lim\limits_{k\to\infty} |x_k - x_*| = 0$. Здесь мы предполагаем, что в (17) $\delta_k = 0$, $k = 0, 1, \ldots$, поэтому из (18), (20), (21) следует $v_k = x_{k+1}$, $k = 0, 1, \ldots$. Учитывая последнее равенство и условие $\alpha_k = \alpha$, подставим (25) в (24). Получим

$$\begin{aligned} |x_{k+1} - x_*|^2 &\leqslant |x_k - x_*|^2 - |x_{k+1} - x_k|^2 + 2\alpha\langle f'(x_k) - f'(x_*), x_* - x_{k+1} \rangle = \\ &= |x_k - x_*|^2 - |x_k - x_{k+1} - \alpha(f'(x_k) - f'(x_*))|^2 + \\ &+ \alpha^2 |f'(x_k) - f'(x_*)|^2 - 2\alpha\langle f'(x_k) - f'(x_*), x_k - x_* \rangle, \quad k = 0, 1, \ldots \end{aligned} \tag{30}$$

Вспомним неравенства (4.3.17), (4.3.18), из которых имеем

$$|f'(x_k) - f'(x_*)|^2 + L\mu |x_k - x_*|^2 \leqslant (L+\mu)\langle f'(x_k) - f'(x_*), x_k - x_* \rangle, \quad k = 0, 1, \ldots \tag{31}$$

$$\mu |x_k - x_*| \leqslant |f'(x_k) - f'(x_*)| \leqslant L |x_k - x_*|, \quad k = 0, 1, \ldots \tag{32}$$

Из (30), (31) следует

$$\begin{aligned} |x_{k+1} - x_*|^2 &\leqslant |x_k - x_*|^2 + \alpha^2 |f'(x_k) - f'(x_*)|^2 - \\ &- 2\alpha(L+\mu)^{-1} |f'(x_k) - f'(x_*)|^2 - 2\alpha L\mu(L+\mu)^{-1} |x_k - x_*|^2 = \\ &= \alpha[\alpha - 2(L+\mu)^{-1}] |f'(x_k) - f'(x_*)|^2 + [1 - 2\alpha L\mu(L+\mu)^{-1}] |x_k - x_*|^2, \quad k = 0, 1, \ldots \end{aligned} \tag{33}$$

Рассмотрим два случая:

1) если $0 < \alpha < 2(L+\mu)^{-1} \leqslant \mu^{-1}$, то из (33) и левого неравенства (32) имеем $|x_{k+1} - x_*|^2 \leqslant |x_k - x_*|^2 [\alpha(\alpha - 2(L+\mu)^{-1})\mu^2 + 1 - 2\alpha L\mu(L+\mu)^{-1}] = (1 - \alpha\mu)^2 |x_k - x_*|^2$;

2) если $L^{-1} \leqslant 2(L+\mu)^{-1} \leqslant \alpha < 2L^{-1}$, то из (33) и правого неравенства (32) получим $|x_{k+1} - x_*|^2 \leqslant |x_k - x_*|^2 [\alpha(\alpha - 2(L+\mu)^{-1})L^2 + 1 - 2\alpha L\mu(L+\mu)^{-1}] = (L\alpha - 1)^2 |x_k - x_*|^2$. Объединяя оба случая, имеем $|x_{k+1} - x_*| \leqslant q(\alpha)|x_k - x_*|$, $k = 0, 1, \ldots$, откуда следует оценка (29). Из графика

функции $q(\alpha)$ (рис. 5.5) видно, что функция $q(\alpha)$ достигает минимума при $0<\alpha<2L^{-1}$ в точке $\alpha_*=2(L+\mu)^{-1}$, причем $q(\alpha_*)=(L-\mu)(L+\mu)$. Теорема 5 доказана. □

Приведем пример, показывающий, что оценка (29) неулучшаема на классе сильно выпуклых функций из $C^{1,1}(X)$.

Пример 1. Пусть $u=(x,y)\in X=E^2$, $f(u)=(Lx^2+\mu y^2)/2$, $0<\mu\leqslant L$. Ясно, что эта функция сильно выпукла с константой $\varkappa=\mu$, принадлежит $C^{1,1}(E^2)$ с константой L и достигает минимума на E^2 в точке $x_*=(0,0)$. Процесс (2) при $\alpha_k=\alpha$, $0<\alpha<2L^{-1}$, имеет вид

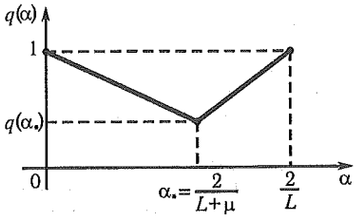

Рис. 5.5

$$x_{k+1}=x_k-\alpha L x_k=(1-\alpha L)x_k,$$
$$y_{k+1}=y_k-\alpha\mu y_k=(1-\alpha\mu)y_k,$$
$$k=0,1,\ldots$$

Положим здесь $\alpha=2(L+\mu)^{-1}$, $q=(L-\mu)(L+\mu)^{-1}$. Тогда $x_{k+1}=-qx_k$, $y_{k+1}=qy_k$. Следовательно, $|u_{k+1}-x_*|=q|u_k|=q^{k+1}|u_0|$, т. е. оценка (29) неулучшаема.

Заметим, что если в теоремах 1–5, в частности, $X=E^n$, то мы получим сходимость соответствующих вариантов градиентного метода (1.3).

5. Опираясь на подходы к непрерывным методам минимизации, развитые в работе А. С. Антипина [25], можно предложить следующий непрерывный вариант метода проекции градиента, основанный на системе дифференциальных уравнений:

$$\dot x(t)=\mathcal{P}_X(x(t)-\alpha(t)f'(x(t)))-x(t),\quad t\geqslant 0,\tag{34}$$

где $\alpha(t)>0$ — заданная функция (параметр метода), X — выпуклое замкнутое множество из E^n. В частности, если $X=E^n$, то (34) превращается в систему (1.45). Согласно теореме 4.4.4 решение x_* задачи (1) удовлетворяет уравнению $\mathcal{P}_X(x_*-\alpha(t)f'(x_*))-x_*=0$ при $\forall t\geqslant 0$. Это значит, что каждая точка $x_*\in X_*$ является точкой равновесия (стационарным решением) системы (34). Можно ожидать, что при некоторых ограничениях на функцию $f(x)$, $\alpha(t)$ траектории $x(t)$ системы (34) при больших t приближаются ко множеству X_*. Справедлива

Теорема 6 (Антипин). *Пусть функция $f(x)\in C^{1,1}(E^n)$, выпукла на E^n, $f_*>-\infty$, $X_*\neq\varnothing$, а функция $\alpha(t)$ непрерывно дифференцируема при $t\geqslant 0$, $\alpha(t)\geqslant\alpha_0>0$, $\alpha'(t)\leqslant 0$ $\forall t\geqslant 0$. Тогда траектория $x(t)$ системы* (34) *с любым начальным условием $x(0)=x_0\in E^n$ определена при всех $t\geqslant 0$ и существует точка $v_*\in X_*$ такая, что*

$$\lim_{t\to+\infty}x(t)=v_*,\quad \lim_{t\to+\infty}f(x(t))=f_*,\quad \lim_{t\to+\infty}\dot x(t)=0.$$

Доказательство. При выполнении условий теоремы $0<\alpha_0\leqslant\alpha(t)\leqslant\alpha(0)$, и с учетом свойства сжимаемости оператора проектирования (теорема 4.4.2) нетрудно доказать, что правая часть $\mathcal{P}_X(x-\alpha(t)f'(x))-x$ дифференциального уравнения (34) удовлетворяет условию Липшица по x, непрерывна по совокупности (t,x). Тогда задача Коши для уравнения (34) с начальным условием $x(0)=x_0$ имеет решение $x=x(t)$, определенное при всех $t\geqslant 0$ (см. ниже теорему 6.1.1). Используя характеристическое свойство проекции (неравенство (4.4.1)) уравнение (34) можно записать в эквивалентном виде

$$\langle(\dot x(t)+x(t))-(x(t)-\alpha(t)f'(x(t))),v-(\dot x(t)+x(t))\rangle\geqslant 0\quad\forall v\in X,\ \forall t\geqslant 0.\tag{35}$$

Кроме того, в силу теоремы 4.2.3 имеет неравенство

$$\langle f'(x_*),v-x_*\rangle\geqslant 0\quad\text{или}\quad\alpha(t)\langle f'(x_*),v-x_*\rangle\geqslant 0\quad\forall v\in X,\ \forall t\geqslant 0,\ \forall x_*\in X_*.\tag{36}$$

Согласно (34) тогда $\dot x(t)+x(t)\in X$ $\forall t\geqslant 0$ (сама траектория $x(t)$ может и не принадлежать X при каких-то $t\geqslant 0$). Положим в (35) $v=x_*$, в (36) $v=\dot x(t)+x(t)$ и сложим получившиеся неравенства. Будем иметь

$$\langle\dot x(t)+\alpha(t)(f'(x(t))-f'(x_*)),x_*-\dot x(t)-x(t)\rangle\geqslant 0$$

или

$$|\dot x(t)|^2+\langle\dot x(t),x(t)-x_*\rangle+\alpha(t)\langle f'(x(t))-f'(x_*)),x(t)-x_*\rangle+\alpha(t)\langle f'(x(t))-f'(x_*)),\dot x(t)\rangle\leqslant 0\quad\forall t\geqslant 0,\quad\forall x_*\in X_*.$$

Отсюда с учетом неравенств $\alpha(t)>0$, $\langle f'(x(t))-f'(x_*)),x(t)-x_*\rangle\geqslant 0$ $\forall t\geqslant 0$ (теорема 4.2.4) имеем:

$$|\dot x(t)|^2+\tfrac12\tfrac{d}{dt}|x(t)-x_*|^2+\alpha(t)\tfrac{d}{dt}(f(x(t))-f(x_*)-\langle f'(x_*),x(t)-x_*\rangle)\leqslant 0\ \ \forall t\geqslant 0.$$

Интегрируя это неравенство на произвольном отрезке $[\tau,t]$, $t>\tau\geqslant 0$, и, преобразуя интеграл от третьего слагаемого по частям, получим

$$\int_\tau^t|\dot x(s)|^2ds+\tfrac12|x(s)-x_*|^2\Big|_{s=\tau}^{s=t}+\alpha(s)(f(x(s))-f(x_*)-\langle f'(x_*),x(t)-x_*\rangle)\Big|_{s=\tau}^{s=t}-\int_\tau^t\alpha'(s)(f(x(s))-f(x_*)-\langle f'(x_*),x(s)-x_*\rangle)ds\leqslant 0\quad\forall t>\tau\geqslant 0,\quad\forall x_*\in X_*.$$

Отсюда с учетом неравенств $\alpha'(t)\leqslant 0$, $\alpha(t)>0$, $\langle f'(x_*),x(t)-x_*\rangle\geqslant 0$ (теорема 4.2.3), $f(x(t))-f(x_*)-\langle f'(x_*),x(t)-x_*\rangle\geqslant 0$ (теорема 4.2.2) имеем

$$\int_\tau^t|\dot x(s)|^2ds+\tfrac12|x(t)-x_*|^2\leqslant\tfrac12|x(\tau)-x_*|^2+\alpha(\tau)(f(x(\tau))-f(x_*))\quad\forall t>\tau\geqslant 0,\quad\forall x_*\in X_*.\tag{37}$$

Из (37) при $\tau=0$ следует

$$\int_0^t|\dot x(s)|^2ds+\tfrac12|x(t)-x_*|^2\leqslant\tfrac12|x_0-x_*|^2+\alpha(0)(f(x_0)-f(x_*))\quad\forall t\geqslant 0.$$

Отсюда заключаем, что траектория $x(t)$ равномерно на $t\geqslant 0$ ограничена. Кроме того, $\int_0^\infty|\dot x(s)|^2ds<\infty$, и поэтому найдется последовательность $\{t_i\}\to+\infty$, для которой $\dot x(t_i)\to 0$. Кроме того, пользуясь теоремой Больцано — Вейерштрасса, можем считать, что последовательность $\{x(t_i)\}$ сходится к некоторой точке v_*. Так как множество X замкнуто, а $\dot x(t)+x(t)\in X$ $\forall t\geqslant 0$, то $\lim\limits_{i\to\infty}(\dot x(t_i)+x(t_i))=v_*\in X$. Далее, переходя в (34) к пределу при $t=t_i\to\infty$ с учетом непрерывности оператора проектирования (теорема 4.4.2), $\lim\limits_{i\to\infty}\alpha(t_i)=\alpha(\infty)\geqslant\alpha_0>0$, получим

$$0=\mathcal{P}_X(v_*-\alpha(\infty)f'(v_*))-v_*.\tag{38}$$

Согласно теореме 4.4.4 это значит, что $v_* \in X_*$. Полагая в (37) $\tau = t_i$ и $x_* = v_*$, имеем

$$\frac{1}{2}|x(t) - v_*|^2 \leqslant \frac{1}{2}|x(t_i) - v_*|^2 + \alpha(t_i)(f(x(t_i)) - f(v_*)) \quad \forall t > t_i.$$

Отсюда, переходя к пределу сначала при $t \to +\infty$, затем $t_i \to +\infty$ с учетом равенства $\lim\limits_{i\to\infty} x(t_i) = v_*$, получим: $\lim\limits_{t\to\infty} x(t) = v_*$. Тогда $\lim\limits_{t\to\infty} f(x(t)) = f(v_*) = f_*$. Наконец, из уравнения (34) при $t \to +\infty$ с учетом равенства (38) имеем: $\lim\limits_{t\to\infty} \dot{x}(t) = 0$. Теорема 6 доказана. □

Для сильно выпуклых функций можно доказать следующую оценку скорости сходимости метода (34) при $\alpha(t) \equiv \alpha_0 = \text{const}$ [25]:

$$|x(t) - x_*| \leqslant |x_0 - x_*| \exp\Big(-\int_0^t b(\tau)d\tau\Big),$$

где $b(\tau) = \alpha\mu\left(1 - \frac{\alpha\mu}{4}\right)$ при $0 < \alpha < \frac{4}{L+\mu}$, $b(\tau) = \alpha L\left(1 - \frac{\alpha L}{4}\right)$ при $\alpha \geqslant \frac{4}{L+\mu}$.

При построении непрерывных методов проекции градиента могут быть использованы дифференциальные уравнения второго и более высокого порядков. Так, например, для выпуклых задач (1) в [25] исследована сходимость процесса

$$\beta(t)\ddot{x}(t) + \dot{x}(t) + x(t) = \mathcal{P}_X(x(t) - \alpha(t)f'(x(t))), \quad t \geqslant 0, \quad \beta(t) \geqslant \beta_0 > 0, \tag{39}$$

и его двухшагового дискретного аналога.

При $X = E^n$ метод (39) превращается в метод тяжелого шарика (1.49). Непрерывный и дискретный варианты методов более высокого порядка см., например, в [520–521].

Упражнения

1. Вычислить несколько итераций метода проекции градиента при различных способах выбора α_k в (2) для функции

$$f(u) = (x-1)^2 + (y+1)^2,$$
$$u \in X = E_+^2 = \{u = (x, y) \in E^2 : x \geqslant 0,\ y \geqslant 0\}.$$

Рассмотреть начальные приближения $u_0 = (0, 0)$, $u_0 = (0, 1)$, $u_0 = (1, 0)$, $u_0 = (1, 1)$.

2. Описать одну итерацию метода проекции градиента для функции (1.5), считая, что множество X представляет собой шар, гиперплоскость, параллелепипед, полупространство или положительный ортант (см. примеры 4.4.1–4.4.6). Исследовать сходимость метода.

3. Рассмотреть метод проекции градиента для функции $f(x) = |Ax - b|^2$, где A — матрица порядка $m \times n$, $b \in E^m$, считая, что множество X имеет вид, описанный в примерах 4.4.1–4.4.6. Исследовать сходимость.

§ 3. Метод проекции субградиента

1. В рассмотренных выше градиентном методе и методе проекции градиента требовалась дифференцируемость минимизируемой функции. Однако для выпуклых функций указанные методы более естественно описать на языке субградиентов (см. § 4.6). А именно, пусть X — выпуклое замкнутое множество из E^n, функция $f(x)$ выпукла на X и ее субдифференциал $\partial f(x)$ непуст при всех $x \in X$. Тогда для приближенного решения задачи

$$f(x) \to \inf; \quad x \in X, \tag{1}$$

можно предложить следующий итерационный метод:

$$x_{k+1} = \mathcal{P}_X(x_k - \alpha_k c_k), \quad \alpha_k > 0, \quad c_k \in \partial f(x_k), \quad k = 0, 1, \ldots, \tag{2}$$

где x_0 — некоторая точка из X, а субградиент c_k выбирается из $\partial f(x_k)$ произвольным образом. Если при некотором k окажется, что $x_{k+1} = x_k$, то процесс (2) прекращается, так как в этом

случае x_k — решение задачи (1). В самом деле, при $x_k = \mathcal{P}_X(x_k - \alpha_k c_k)$ согласно теореме 4.4.1 $\langle x_k - (x_k - \alpha_k c_k), x - x_k\rangle = \langle c_k, x - x_k\rangle \alpha_k \geqslant 0$ или $\langle c_k, x - x_k\rangle \geqslant 0$ при всех $x \in X$. Отсюда из теоремы 4.6.4 следует, что $x_k \in X_*$.

Согласно замечанию 1 к теореме 4.6.4 точка $x_* \in X_*$ тогда и только тогда, когда $x_* = \mathcal{P}_X(x_* - \alpha c_*)$, $\alpha > 0$, где c_* — некоторый элемент субдифференциала $\partial f(x_*)$. Отсюда ясно, что итерационный процесс (2) представляет собой метод поиска неподвижной точки оператора $\mathcal{P}_X(x - \alpha c)$, $c \in \partial f(x)$.

В том случае, когда функция $f(x)$ дифференцируема во всех точках $x \in X$, метод (2) превращается в метод проекции градиента, а при $X = E^n$ — в градиентный метод. При выборе длины шага α_k в (2) можно руководствоваться теми же соображениями, которые были описаны выше в § 1, 2. Мы здесь ограничимся рассмотрением случая, когда α_k в (2) выбирается из условий

$$\alpha_k > 0, \quad k = 0, 1, \ldots; \quad \sum_{k=0}^{\infty} \alpha_k = \infty, \quad \sum_{k=0}^{\infty} \alpha_k^2 < \infty. \tag{3}$$

Как уже отмечалось в § 1, в качестве α_k можно взять $\alpha_k = C(k+1)^{-\alpha}$, где $C = \text{const} > 0$, $1/2 < \alpha \leqslant 1$; например, $\alpha_k = (k+1)^{-1}$, $k = 0, 1, \ldots$

Метод (2), (3) не гарантирует выполнения условия монотонности $f(x_k) > f(x_{k+1})$ на каждой итерации и сходится, вообще говоря, медленно, но если проекция точки на множестве X и субградиент $c_k \in \partial f(x_k)$ находятся несложно, то этот метод очень прост для реализации на ЭВМ. Докажем его сходимость.

Теорема 1. *Пусть X — выпуклое замкнутое множество из E^n, функция $f(x)$ определена и выпукла на некотором открытом выпуклом множестве W, содержащем X (например, $W = E^n$). Пусть $f_* > -\infty$, множество X_* точек минимума $f(x)$ на X непусто и ограничено, и пусть, кроме того,*

$$\sup_{x\in X} \sup_{c\in\partial f(x)} |c| = A < \infty. \tag{4}$$

Тогда последовательность $\{x_k\}$, определяемая условиями (2), (3), *такова, что*

$$\lim_{k\to\infty} f(x_k) = f_*, \quad \lim_{k\to\infty} \rho(x_k, X_*) = 0. \tag{5}$$

Доказательство. При сделанных предположениях функция $f(x)$ непрерывна на X, субдифференциалы $\partial f(x)$ непусты, выпуклы, замкнуты и ограничены при всех $x \in X$ (см. теоремы 4.2.15, 4.6.1, 4.6.2). Из ограниченности множества X_* и теоремы 4.2.17 следует, что множество $M(C) = \{x \in X : f(x) \leqslant C\}$ ограничено при любом $C \geqslant f_*$. Множество X_* выпукло и замкнуто в силу теоремы 4.2.1 и леммы 2.1.1.

Согласно определению 4.4.1 проекции точки на множество и теореме 4.4.2 имеем

$$\rho^2(x_{k+1}, X_*) = |x_{k+1} - \mathcal{P}_{X_*}(x_{k+1})|^2 \leqslant |x_{k+1} - \mathcal{P}_{X_*}(x_k)|^2 = |\mathcal{P}_X(x_k - \alpha_k c_k) - \mathcal{P}_X(\mathcal{P}_{X_*}(x_k))|^2 \leqslant$$
$$\leqslant |x_k - \alpha_k c_k - \mathcal{P}_{X_*}(x_k)|^2 = \rho^2(x_k, X_*) + \alpha_k^2|c_k|^2 - 2\alpha_k\langle c_k, x_k - \mathcal{P}_{X_*}(x_k)\rangle$$

или

$$2\alpha_k\langle c_k, x_k - \mathcal{P}_{X_*}(x_k)\rangle \leqslant \alpha_k^2|c_k|^2 + \rho^2(x_k, X_*) - \rho^2(x_{k+1}, X_*), \quad k = 0, 1, \ldots \tag{6}$$

Суммируя неравенства (6) по k от 0 до некоторого $s \geqslant 1$, с учетом условий (3), (4) получим

$$\sum_{k=0}^{s} 2\alpha_k\langle c_k, x_k - \mathcal{P}_{X_*}(x_k)\rangle \leqslant A^2 \sum_{k=0}^{m} \alpha_k^2 + \rho^2(x_0, X_*) - \rho^2(x_{k+1}, X_*) \leqslant$$
$$\leqslant A^2 \sum_{k=0}^{\infty} \alpha_k^2 + \rho^2(x_0, X_*) = B < \infty, \quad s = 1, 2, \ldots \tag{7}$$

Далее, по определению субградиента имеем

$$0 \leqslant f(x_k) - f_* = f(x_k) - f(\mathcal{P}_{X_*}(x_k)) \leqslant \langle c_k, x_k - \mathcal{P}_{X_*}(x_k)\rangle, \quad k = 0, 1, \ldots \tag{8}$$

Из (7), (8) следует, что числовой ряд

$$\sum_{k=0}^{\infty} \alpha_k\langle c_k, x_k - \mathcal{P}_{X_*}(x_k)\rangle$$

с неотрицательными членами сходится. Но согласно (3) имеем $\sum\limits_{k=0}^{\infty} \alpha_k = \infty$. Поэтому сходи-

мость предыдущего ряда возможна лишь при $\lim\limits_{k\to\infty}\langle c_k, x_k - \mathcal{P}_{X_*}(x_k)\rangle = 0$. Это значит, что существуют номера $k_1 < k_2 < \ldots < k_p < \ldots$ такие, что

$$\lim_{p\to\infty}\langle c_{k_p}, x_{k_p} - \mathcal{P}_{X_*}(x_{k_p})\rangle = 0. \tag{9}$$

Тогда из (8) при $k = k_p \to \infty$ получим $\lim\limits_{p\to\infty} f(x_{k_p}) = f_*$. Кроме того, из (8), (9) следует, что $f(x_{k_p}) \leqslant f_* + \sup\limits_p \langle c_{k_p}, x_{k_p} - \mathcal{P}_{X_*}(x_{k_p})\rangle = C < \infty$, т. е. $\{x_{k_p}\} \in M(C)$. Но $M(C)$ ограничено, а $\{x_{k_p}\}$ — минимизирующая последовательность, поэтому из теоремы 2.1.2 имеем $\lim\limits_{p\to\infty}\rho(x_{k_p}, X_*) = 0$.

Тем самым показано, что для подпоследовательности $\{x_{k_p}\}$, удовлетворяющей условию (9), справедливы равенства (5). Опираясь на это, покажем, что равенства (5) имеют место для всей последовательности $\{x_k\}$. Из (3), (4), (6), (8) получаем, что

$$\rho^2(x_{k+1}, X_*) \leqslant \rho^2(x_k, X_*) + \alpha_k^2|c_k|^2 \leqslant \rho^2(x_k, X_*) + \alpha_k^2 A^2, \quad k = 0, 1, \ldots$$

Это значит, что числовая последовательность $a_k = \rho^2(x_k, X_*)$, $k = 0, 1, \ldots$, удовлетворяет условиям леммы 2.6.2, из которой следует существование предела $\lim\limits_{k\to\infty}\rho^2(x_k, X_*)$. Так как подпоследовательность $\{\rho^2(x_{k_p}, X_*)\}$ сходится к нулю, то этот предел может равняться лишь нулю. Следовательно, $\lim\limits_{k\to\infty}\rho(x_k, X_*) = 0$. Покажем, что тогда $\lim\limits_{k\to\infty} f(x_k) = f_*$. По условию множество X_* ограничено. Тогда последовательность $\{x_k\}$, сходящаяся к X_*, также ограничена. Возьмем любую предельную точку x_* этой последовательности. Пусть $\{x_{k_r}\} \to x_*$. Так как X_* — замкнутое множество и $\lim\limits_{r\to\infty}\rho(x_{k_r}, X_*) = \rho(x_*, X_*) = 0$, то $x_* \in X_*$. А тогда $\lim\limits_{r\to\infty} f(x_{k_r}) = f(x_*) = f_*$. Это означает, что числовая последовательность $\{f(x_k)\}$ имеет единственную предельную точку, равную f_*, т. е. $\lim\limits_{k\to\infty} f(x_k) = f_*$. Теорема 1 доказана. □

Замечание 1. В условии теоремы 1 предполагается выполнение условия (4). В том случае, когда X ограничено, то, как следует из теоремы 4.6.5, условие (4) всегда выполняется. Заметим также, что в теореме 1 вместо (4) можно потребовать сходимость ряда $\sum\limits_{k=0}^{\infty}\alpha_k^2|c_k|^2$.

Замечание 2. Описанный выше метод проекции субградиента после некоторой модификации можно использовать для решения следующей задачи выпуклого программирования:

$$f(x) \to \inf; \quad x \in X = \{x \in E^n: x \in X_0,\ g_i(x) \leqslant 0,\ i = 1, \ldots, m\}. \tag{10}$$

Заметим, что система неравенств $g_i(x) \leqslant 0$, $i = 1, \ldots, m$, равносильна одному неравенству $g(x) \leqslant 0$, где $g(x) = \max\limits_{1\leqslant i\leqslant m} g_i(x)$, $x \in X$. Кроме того, из выпуклости функций $g_i(x)$, $i = 1, \ldots, m$, на X_0 следует выпуклость $g(x)$ на X_0 (см. теорему 4.2.7). Поэтому задачу (10) можно переформулировать в виде эквивалентной задачи

$$f(x) \to \inf; \quad x \in X = \{x \in X_0,\ g(x) \leqslant 0\}, \tag{11}$$

также являющейся задачей выпуклого программирования.

Предположим, что субдифференциалы $\partial f(x)$, $\partial g(x)$ непусты при всех $x \in X_0$. Следуя [766], рассмотрим метод

$$x_{k+1} = \mathcal{P}_{X_0}(x_k - \alpha_k c_k), \quad k = 0, 1, \ldots; \quad x_0 \in X_0, \tag{12}$$

где $\{\alpha_k\}$ выбирается из условий

$$\alpha_k > 0, \quad k = 0, 1, \ldots, \quad \sum_{k=0}^{\infty}\alpha_k^{1+\gamma} = \infty, \quad \sum_{k=0}^{\infty}\alpha_k^2 < \infty, \quad 0 < \gamma < 1, \tag{13}$$

а субградиенты $\{c_k\}$ таковы, что

$$c_k \in \partial f(x_k) \ \text{ при } \ g(x_k) \leqslant \alpha_k^{\gamma} \ \text{ и } \ c_k \in \partial g(x_k) \ \text{ при } \ g(x_k) > \alpha_k^{\gamma}. \tag{14}$$

Таким образом, метод (12)–(14) работает так: если ограничение $g(x) \leqslant 0$ при $x = x_k$ не нарушено или нарушено немного, то минимизируем функцию $f(x)$, а если нарушение этого

ограничения велико, то минимизируем функцию $g(x)$. Если функции $f(x)$, $g(x)$ дифференцируемы на X_0, то в (12), (14) вместо c_k нужно брать соответствующие градиенты $f'(x_k)$ или $g'(x_k)$. В качестве последовательности $\{\alpha_k\}$, удовлетворяющей условиям (13), можно взять, например, $\alpha_k = C(k+1)^{-\alpha}$, где $C = \text{const} > 0$, а число α таково, что $1/2 < \alpha < (1+\gamma)^{-1}$. В частности, при $\alpha = 3/5$, $\gamma = 1/2$, $C = 1$ получим $\alpha_k = (k+1)^{-3/5}$, $k = 0, 1, \ldots$

Теорема 2. *Пусть X_0 — выпуклое замкнутое множество из E^n, функции $f(x)$, $g(x)$ определены и выпуклы на некотором открытом выпуклом множестве W, содержащем X_0 (например, $W = E^n$). Пусть $f_* = \inf\limits_X f(x) > -\infty$, множество X_* точек минимума задачи* (11) *непусто, ограничено и, кроме того,*

$$\sup_{x\in X_0}\ \sup_{c\in\partial f(x)\cup\partial g(x)}|c| = A < \infty.$$

Тогда для последовательности $\{x_k\}$, определяемой условиями (12)–(14), *справедливы равенства* (5).

Доказательство. При выполнении условий теоремы функции $f(x)$, $g(x)$ непрерывны на X_0, субдифференциалы $\partial f(x)$, $\partial g(x)$ непусты, выпуклы, замкнуты и ограничены при всех $x \in X_0$ (см. теоремы 4.2.15, 4.6.1, 4.6.2), а множество X_* выпукло и замкнуто (см. теорему 4.2.1 и лемму 2.1.1). Покажем, что множество

$$M(C_1, C_2) = \{x \in X_0: f(x) \leqslant C_1,\ g(x) \leqslant C_2\}$$

ограничено при всех $C_1 > \inf\limits_{X_0} f(x)$, $C_2 > \inf\limits_{X_0} g(x)$. В самом деле, $M(f_*, 0) = X_*$ ограничено по условию. Тогда по теореме 4.2.17 множество $M(C_1, 0) = \{x: x \in X_0,\ g(x) \leqslant 0,\ f(x) \leqslant C_1\}$ ограничено при всех $C_1 > \inf\limits_{X_0} f(x)$. Теперь, фиксируя любое C_1, по той же теореме 4.2.17 получаем ограниченность $M(C_1, C_2)$ при каждом $C_2 > \inf\limits_{X_0} g(x)$.

Нетрудно видеть, что неравенства (6), (7) сохраняют силу и для метода (12)–(14). Из (7) имеем

$$\sum_{k=0}^{s}\alpha_k^{1+\gamma}\alpha_k^{-\gamma}\langle c_k, x_k - \mathcal{P}_{X_*}(x_k)\rangle \leqslant B < \infty, \quad s = 1, 2, \ldots \tag{15}$$

Отсюда следует существование номеров $k_1 < k_2 < \ldots < k_p < \ldots$ таких, что

$$\langle c_{k_p}, x_{k_p} - \mathcal{P}_{X_*}(x_{k_p})\rangle \leqslant \alpha_{k_p}^{\gamma}, \quad p = 1, 2, \ldots \tag{16}$$

В самом деле, допустим, что (16) не имеет места. Тогда $\langle c_k, x_k - \mathcal{P}_{X_*}(x_k)\rangle > \alpha_k^{\gamma}$ при всех $k = 0, 1, \ldots$ Отсюда и из (15) имеем

$$\sum_{k=0}^{s}\alpha_k^{1+\gamma} \leqslant \sum_{k=0}^{s}\alpha_k\langle c_k, x_k - \mathcal{P}_{X_*}(x_k)\rangle \leqslant B < \infty, \quad s = 1, 2, \ldots,$$

что противоречит расходимости $\sum\limits_{k=0}^{\infty}\alpha_k^{1+\gamma}$.

Тем самым показано существование подпоследовательности $\{x_{k_p}\}$, удовлетворяющей условию (16). Докажем, что

$$\lim_{p\to\infty} f(x_{k_p}) = f_*, \quad \lim_{p\to\infty}\rho(x_{k_p}, X_*) = 0. \tag{17}$$

Сначала убедимся в том, что

$$c_{k_p} \in \partial f(x_{k_p}), \quad p = 1, 2, \ldots \tag{18}$$

Для этого достаточно показать, что $g(x_{k_p}) \leqslant \alpha_{k_p}^{\gamma}$, $p = 1, 2, \ldots$, и вспомнить условия (14). Допустим, что $g(x_{k_p}) > \alpha_{k_p}^{\gamma}$ при некотором $p \geqslant 1$. Учитывая, что тогда $c_{k_p} \in \partial g(x_{k_p})$ и, кроме того, $\mathcal{P}_{X_*}(x_{k_p}) \in X_* \subset X$, т. е. $g(\mathcal{P}_{X_*}(x_{k_p})) \leqslant 0$, из (16) имеем

$$\alpha_{k_p}^{\gamma} < g(x_{k_p}) \leqslant g(x_{k_p}) - g(\mathcal{P}_{X_*}(x_{k_p})) \leqslant \langle c_{k_p}, x_{k_p} - \mathcal{P}_{X_*}(x_{k_p})\rangle \leqslant \alpha_{k_p}^{\gamma}.$$

Получили противоречивое неравенство. Включения (18) доказаны, и попутно установлено, что

$$g(x_{k_p}) \leqslant \alpha_{k_p}^{\gamma}, \quad p = 1, 2, \ldots \tag{19}$$

Множество номеров $\{k_p\}$, удовлетворяющих условиям (16), представимо в виде объединения непересекающихся множеств $I_1=\{k_p\colon f(x_{k_p})\geqslant f_*\}$ и $I_2=\{k_p\colon f(x_{k_p})<f_*\}$.

Сначала рассмотрим случаи, когда множество I_1 бесконечно. Из (16), (18) имеем

$$0\leqslant f(x_{k_p})-f_*=f(x_{k_p})-f(\mathcal{P}_{X_*}(x_{k_p}))\leqslant\langle c_{k_p},\, x_{k_p}-\mathcal{P}_{X_*}(x_{k_p})\rangle\leqslant\alpha_{k_p}^{\gamma},\quad k_p\in I_1.$$

Отсюда следует, что $f(x_{k_p})\to f_*$ при $p\to\infty$, $k_p\in I_1$. Тогда $f(x_{k_p})\leqslant C_1<\infty$, и, кроме того, согласно (13), (19) имеем $g(x_{k_p})\leqslant\alpha_{k_p}^{\gamma}\leqslant\sup\limits_{k\geqslant 0}\alpha_k^{\gamma}=C_2<\infty$, т. е. $\{x_{k_p}\}\in M(C_1, C_2)$, $k_m\in I_1$. Так как $M(C_1, C_2)$ ограничено, то $\{x_{k_p},\, k_p\in I_1\}$ имеет хотя бы одну предельную точку. Не умаляя общности, можем считать, что $\{x_{k_p}\}\to x_*$, $k_p\in I_1$. Из замкнутости X_0, неравенства (19) и непрерывности $g(x)$ следует, что $x_*\in X$. Но $f(x_{k_p})\to f(x_*)=f_*$ при $k_p\to\infty$, $k_p\in I_1$, так что $x_*\in X_*$. Отсюда и из непрерывности $\rho(x, X_*)$ имеем $\rho(x_{k_p}, X_*)\to\rho(x_*, X_*)=0$ при $k_p\to\infty$, $k_p\in I_1$.

Теперь рассмотрим случай, когда множество I_2 бесконечно. Если $k_p\in I_2$, то $f(x_{k_p})<f_*$, а $g(x_{k_p})\leqslant\alpha_{k_p}^{\gamma}<\sup\limits_{k\geqslant 0}\alpha_k^{\gamma}=C_2$ в силу (19). Это значит, что $\{x_{k_p}\}\in M(f_*, C_2)$, $k_p\in I_2$. Поскольку множество $M(f_*, C_2)$ ограничено, то $\{x_{k_p},\, k_p\in I_2\}$ имеет предельную точку. Не умаляя общности, можем считать, что $\{x_{k_p}\}\to x_*$, $k_p\in I_2$. Из (19) и замкнутости X_0 следует, что $x_*\in X$. Поэтому $f(x_*)\geqslant f_*$. С другой стороны, $f(x_{k_p})<f_*$, $k_p\in I_2$, так что $\lim\limits_{k_p\in I_2,\, p\to\infty}f(x_{k_p})=f(x_*)\leqslant f_*$. Следовательно, $f(x_*)=f_*$, т. е. $x_*\in X_*$. Отсюда и из непрерывности $\rho(x, X_*)$ получаем $\rho(x_{k_p}, X_*)\to\rho(x_*, X_*)=0$ при $k_p\to\infty$, $k_p\in I_2$.

Объединяя оба рассмотренных случая, заключаем, что для подпоследовательности $\{x_{k_p}\}$, удовлетворяющей условию (16), справедливы равенства (17). Отсюда, повторив заключительные рассуждения из доказательства теоремы 1, убеждаемся в справедливости равенств (5) и для метода (12)–(14). Замечание 1 сохраняет силу и здесь.

На этом закончим рассмотрение методов минимизации негладких выпуклых функций. Отметим, что негладкие задачи в последние годы интенсивно исследуются, продолжается разработка различных методов их решения [73; 226; 251; 256; 263–266; 302; 314; 318; 361; 386; 396; 426; 434; 495; 502; 542; 572; 586; 613; 718; 720; 769; 777; 795].

Упражнения

1. Рассмотреть возможность применения метода проекции субградиента к задачам из упражнений 4.6.1 и 4.6.3.

2. Описать метод (12)–(14) применительно к задаче

$$f(u)=|x+y|+|x-y|\to\inf,\quad u\in X=\{u=(x, y)\in E^2\colon u\in X_0,\ g(u)=u^2-1\leqslant 0\},$$
$$X_0=\{u=(x, y)\colon x\geqslant 0,\ y\leqslant 0\}.$$

3. Проверить условия теоремы 2 для задачи

$$f(u)=|\langle c, u\rangle|\to\inf;$$
$$u\in X=\{u\in E^n\colon u\geqslant 0,\ g(u)=|\langle a, u\rangle|-1\leqslant 0\},$$

где $a, c\in E^n$. Описать метод (12)–(14) применительно к этой задаче.

4. Пользуясь формулой (4.6.10), модифицировать метод (12)–(14) так, чтобы его можно было применять к задаче (10) непосредственно, не сводя ее к задаче (11).

5. Пусть W — открытое выпуклое множество, $f(x)$ — выпуклая функция на W. Показать, что вектор c_*, удовлетворяющий условиям

$$|c_*|=\inf_{c\in\partial f(x)}|c|>0,\quad c_*\in\partial f(x),$$

является направлением убывания функции $f(x)$ в точке x.

§ 4. Метод условного градиента

1. Этот метод приспособлен для решения задачи

$$f(x)\to\inf;\quad x\in X,\tag{1}$$

где X — выпуклое замкнутое ограниченное множество из E^n, функция $f(x)\in C^1(X)$. Опишем его. Пусть $x_0\in X$ — некоторое начальное приближение. Если известно k-е приближение $x_k\in X$, $k\geqslant 0$, то приращение функции $f(x)$ в точке x_k можем представить в виде

$$f(x)-f(x_k)=\langle f'(x_k),\, x-x_k\rangle+o(|x-x_k|).$$

Возьмем главную линейную часть этого приращения

$$f_k(x)=\langle f'(x_k),\, x-x_k\rangle,\tag{2}$$

и определим вспомогательное приближение $\bar{x}_k$ из условий

$$\bar{x}_k\in X,\quad \inf_X f_k(x)=f_k(\bar{x}_k)=\langle f'(x_k),\, \bar{x}_k-x_k\rangle.\tag{3}$$

Так как множество X замкнуто и ограничено, а линейная функция $f_k(x)$ непрерывна, то точка $\bar{x}_k$ из (3) всегда существует. Если функция $f_k(x)$ достигает своей нижней грани на X более чем в одной точке, то в качестве точки $\bar{x}_k$ возьмем любую из них.

Заметим, что если

$$X=\{x\in E^n\colon x\geqslant 0,\ \langle a_i, x\rangle\leqslant b^i,\ i=1, \ldots, m;\ \langle a_i, x\rangle=b^i,\ i=m+1, \ldots, s\},$$

то задача (3) превратится в задачу линейного программирования, которая может быть решена известными методами (например, описанным в гл. 3 симплекс-методом).

Укажем случаи, когда решение задачи (3) будет выписываться в явном виде. Если

$$X=\{x=(x^1, \ldots, x^n)\colon \alpha_i\leqslant x^i\leqslant\beta_i,\ i=1, \ldots, n\}$$

— n-мерный параллелепипед, то функция $f_k(x)=\sum\limits_{i=1}^{n}f_{x^i}(x_k)(x^i-x_k^i)$ или $\sum\limits_{i=1}^{n}f_{x^i}(x_k)x^i$, очевидно, достигает своей нижней грани на X в точке $\bar{x}_k=(\bar{x}_k^1, \ldots, \bar{x}_k^n)$, где

$$\bar{x}_k^i=\begin{cases}\alpha_i, & f_{x^i}(x_k)>0,\\ \beta_i, & f_{x^i}(x_k)<0;\end{cases}$$

в случае $f_{x^i}(x_k)=0$ здесь возникает неопределенность и в качестве $\bar{x}_k^i$ можно взять любое число из отрезка $[\alpha_i, \beta_i]$ (обычно берут $\bar{x}_k^i=\alpha_i$, или $\bar{x}_k^i=\beta_i$, или $\bar{x}_k^i=(\alpha_i+\beta_i)/2$).

Если

$$X=\{x\in E^n\colon |x-v_0|\leqslant r\}$$

— шар радиуса r с центром в точке v_0, то с помощью неравенства Коши — Буняковского $\langle f'(x_k), x\rangle=\langle f'(x_k), x-v_0\rangle+\langle f'(x_k), v_0\rangle\geqslant-|f'(x_k)|r+\langle f'(x_k), v_0\rangle$, $x\in X$, получаем, что

$$\bar{x}_k=v_0-rf'(x_k)|f'(x_k)|^{-1}.$$

Разумеется, так просто получить вспомогательное приближение $\bar{x}_k$ удается далеко не всегда, и вместо точного решения задачи (3) часто приходится довольствоваться определением какого-либо приближенного решения. А именно, будем предполагать, что оно определяется из следующих условий:

$$\bar{x}_k \in X, \quad f_k(\bar{x}_k) \leqslant \min_X f_k(x) + \varepsilon_k, \quad \varepsilon_k \geqslant 0, \quad \lim_{k\to\infty} \varepsilon_k = 0. \tag{4}$$

Допустим, что точка $\bar{x}_k$, удовлетворяющая условиям (4) (или (3)), уже найдена. Тогда следующее $(k+1)$-е приближение будем искать в виде

$$x_{k+1} = x_k + \alpha_k(\bar{x}_k - x_k), \quad 0 \leqslant \alpha_k \leqslant 1. \tag{5}$$

В силу выпуклости множества X всегда $x_{k+1} \in X$.

Заметим, что при $\bar{x}_k = x_k$ (это может случиться, например, когда $f'(x_k) = 0$) имеем $x_{k+1} = x_k$ независимо от способа выбора α_k в (5). Если при этом $\bar{x}_k$ было определено точно из условий (3), то имеем

$$f_k(\bar{x}_k) = f_k(x_k) = 0 = \min_X f_k(x) \quad \text{и} \quad \langle f'(x_k), x - x_k\rangle \geqslant 0$$

при всех $x \in X$. Согласно теореме 4.2.3 это означает, что точка x_k удовлетворяет необходимому условию минимума в задаче (1). В этом случае итерации прекращаются, и для выяснения того, будет ли $x_k \in X_*$, при необходимости проводится дополнительное исследование поведения функции $f(x)$ в окрестности точки x_k. В частности, если $f(x)$ выпукла, то согласно теореме 4.2.3 имеем $x_k \in X_*$, т. е. задача (1) решена. Если случай $\bar{x}_k = x_k$ реализовался при определении $\bar{x}_k$ из условия (4), то будем иметь $-\varepsilon_k \leqslant \min_X f_k(x) \leqslant f_k(\bar{x}_k) = f_k(x_k) = 0$, и при $\varepsilon_k > 0$ здесь теорему 4.2.3 применять нельзя. В этом случае согласно (5) полагаем $x_{k+1} = x_k$ и переходим к проверке условия (4) для номера $k+1$ и т. д.

В зависимости от способа выбора величины α_k в (5) можно получить различные варианты описанного метода, часто именуемого в литературе *методом условного градиента*. Укажем некоторые наиболее употребительные способы выбора α_k в (5).

1) Величина α_k может выбираться из условий

$$0 \leqslant \alpha_k \leqslant 1, \quad g_k(\alpha_k) = \min_{0\leqslant\alpha\leqslant 1} g_k(\alpha) = g_{k*}, \quad g_k(\alpha) = f(x_k + \alpha(\bar{x}_k - x_k)). \tag{6}$$

Для некоторых классов задач можно получить из (6) явное выражение для α_k.

Пример 1. Пусть

$$f(x) = \tfrac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle,$$

где A — симметричная положительно определенная матрица размера $n \times n$, $b \in E^n$. Тогда $f'(x_k) = Ax_k - b$. Пользуясь формулой (4.2.10), имеем

$$g_k(\alpha) = f(x_k) + \alpha\langle f'(x_k), \bar{x}_k - x_k\rangle + (\alpha^2/2)\langle A(\bar{x}_k - x_k), \bar{x}_k - x_k\rangle. \tag{7}$$

Если $\langle A(\bar{x}_k - x_k), \bar{x}_k - x_k\rangle = 0$, то $x_k = \bar{x}_k$ и, как было указано выше, тогда $x_k \in X_*$. Поэтому пусть $\langle A(\bar{x}_k - x_k), \bar{x}_k - x_k\rangle > 0$. Тогда функция (7) представляет собой квадратный трехчлен, достигающий своего наименьшего значения на числовой оси $-\infty < \alpha < +\infty$ при

$$\alpha_k^* = -\langle f'(x_k), \bar{x}_k - x_k\rangle(\langle A(\bar{x}_k - x_k), \bar{x}_k - x_k\rangle)^{-1}.$$

Рассматривая возможные случаи $\alpha_k^* < 0$, $0 \leqslant \alpha_k^* \leqslant 1$, $\alpha_k^* > 1$, из условий (6) тогда получаем

$$\alpha_k = \begin{cases} 0, & \alpha_k^* < 0, \\ \alpha_k^*, & 0 \leqslant \alpha_k^* \leqslant 1, \\ 1, & \alpha_k^* > 1. \end{cases} \tag{8}$$

Кстати, если точка $\bar{x}_k$ в (7) найдена из условий (3), то $f_k(\bar{x}_k) \leqslant f_k(x_k) = 0$ и, следовательно, $\alpha_k^* \geqslant 0$ — в этом случае формула (8) для α_k запишется в виде $\alpha_k = \min\{1; \alpha_k^*\}$.

Однако точное определение α_k из условия (6) возможно далеко не всегда. Поэтому вместо (6) можно ограничиться определением величины α_k из условий

$$0 \leqslant \alpha_k \leqslant 1, \quad g_k(\alpha_k) \leqslant g_{k*} + \delta_k, \quad \delta_k \geqslant 0, \quad \sum_{k=0}^{\infty} \delta_k = \delta < \infty \tag{9}$$

или

$$0 \leqslant \alpha_k \leqslant 1, \quad g_k(\alpha_k) \leqslant (1 - \lambda_k)g_k(0) + \lambda_k g_{k*}, \quad 0 < \bar{\lambda} \leqslant \lambda_k \leqslant 1.$$

Здесь могут быть использованы известные методы минимизации функций одной переменной (например, методы из гл. 1).

2) Если $f(x) \in C^{1,1}(X)$ и константа Липшица L для $f'(x)$ известна, то возможен выбор α_k в (5) из условий

$$\alpha_k = \begin{cases} \min\{1; \rho_k|f_k(\bar{x}_k)||\bar{x}_k - x_k|^{-2}\}, & f_k(\bar{x}_k) \leqslant 0, \\ 0, & f_k(\bar{x}_k) > 0, \end{cases} \tag{10}$$

где

$$0 < \varepsilon_0 \leqslant \rho_k \leqslant 2(1 - \varepsilon)/L, \tag{11}$$

$\varepsilon_0, \varepsilon$ — параметры метода, $0 < \varepsilon_0 < 1$.

3) Другой способ выбора α_k: при $f_k(\bar{x}_k) > 0$ полагают $\alpha_k = 0$, а если $f_k(\bar{x}_k) \leqslant 0$, то $\alpha_k = \lambda^{i_0}$, где i_0 — минимальный номер среди номеров $i \geqslant 0$, удовлетворяющих условию

$$f(x_k) - f(x_k + \lambda^i(\bar{x}_k - x_k)) \geqslant \lambda^i\varepsilon|f_k(\bar{x}_k)|,$$

где λ, ε — параметры метода, $0 < \lambda;\ \varepsilon < 1$.

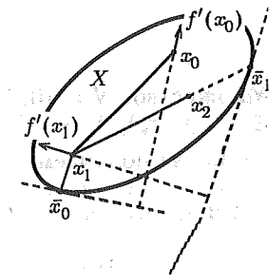

Рис. 5.6

4) Величины α_k в (5) можно априорно задавать из условий [783]

$$0 < \alpha_k \leqslant 1, \quad \lim_{k\to\infty} \alpha_k = 0, \quad \sum_{k=0}^{\infty} \alpha_k = \infty, \tag{12}$$

например, $\alpha_k = (k+1)^{-1}$, $k = 0, 1, \ldots$ Такой выбор α_k очень прост для реализации на ЭВМ, но, вообще говоря, не гарантирует выполнение условия монотонности $f(x_{k+1}) < f(x_k)$.

5) Возможны и другие способы выбора α_k в (5). Например, можно задавать $\alpha_k = 1$ и проверять условие монотонности $f(x_{k+1}) < f(x_k)$, а затем при необходимости дробить α_k до тех пор, пока не выполнится условие монотонности.

На рис. 5.6 поясняется геометрический смысл метода (3), (5) в двумерном случае.

2. Рассмотрим теперь сходимость метода (4), (5), (9).

Теорема 1. *Пусть X — выпуклое замкнутое ограниченное множество из E^n, функция $f(x)\in C^{1,1}(X)$. Тогда при любом выборе $x_0\in X$ для последовательности $\{x_k\}$, определяемой условиями* (4), (5), (9), *справедливы равенства*

$$\lim_{k\to\infty}\langle f'(x_k),\bar{x}_k-x_k\rangle=0,\quad \lim_{k\to\infty}\rho(x_k,S_*)=0, \tag{13}$$

где $S_*=\{x\colon x\in X,\ \langle f'(x),v-x\rangle\geqslant 0$ *при всех* $v\in X\}$.

Если, кроме перечисленных условий, $f(x)$ *выпукла на* X *и*

$$\varepsilon_k+\delta_k\leqslant C_0k^{-2\rho},\quad C_0=\mathrm{const}>0,\quad 1/2<\rho\leqslant 1, \tag{14}$$

то

$$\lim_{k\to\infty}f(x_k)=f_*,\quad \lim_{k\to\infty}\rho(x_k,X_*)=0, \tag{15}$$

и справедлива оценка

$$0\leqslant f(x_k)-f_*\leqslant C_1k^{-\rho},\quad k=1,2,\ldots;\quad C_1=\mathrm{const}\geqslant 0. \tag{16}$$

Наконец, если, кроме того, $f(x)$ *сильно выпукла на* X, *то*

$$|x_k-x_*|^2\leqslant(2C_1/\varkappa)k^{-\rho},\quad k=1,2,\ldots \tag{17}$$

Доказательство. При сделанных предположениях $f_*>-\infty$, $X_*\neq\varnothing$. Так как множество X ограничено, то $\sup\limits_{u,v\in X}|u-v|\leqslant d<\infty$. Из условия (9) следует $f(x_{k+1})=g_k(\alpha_k)\leqslant g_{k*}+\delta_k\leqslant f(x_k+\alpha(\bar{x}_k-x_k))+\delta_k$ при всех α, $0\leqslant\alpha\leqslant 1$. Поэтому пользуясь неравенством (2.6.7), имеем

$$\begin{aligned}f(x_k)-f(x_{k+1})+\delta_k&\geqslant f(x_k)-f(x_k+\alpha(\bar{x}_k-x_k))\geqslant-\alpha\langle f'(x_k),\bar{x}_k-x_k\rangle-\\&-\alpha^2L|\bar{x}_k-x_k|^2/2\geqslant\\&-\alpha f'(\bar{x}_k)-\alpha^2Ld^2/2,\quad 0\leqslant\alpha\leqslant 1,\quad k=0,1,\ldots\end{aligned} \tag{18}$$

Множество $N=\{0,1,2,\ldots\}$ разобьем на два множества $N^+=\{k\colon k\in N,\ \langle f'(x_k),\bar{x}_k-x_k\rangle>0\}$ и $N^-=N\setminus N^+$. Так как $\inf\limits_X f_k(x)\leqslant f_k(x_k)=0$, то из (4) получаем $0\leqslant f_k(\bar{x}_k)\leqslant\varepsilon_k$ при всех $k\in N^+$. Поэтому если N^+ — бесконечное множество, то $f_k(\bar{x}_k)\to 0$ при $k\to\infty$, $k\in N^+$.

Теперь пусть $k\in N^-$. Тогда из (18) имеем

$$0\leqslant-f_k(\bar{x}_k)\leqslant(f(x_k)-f(x_{k+1})+\delta_k)/\alpha+\alpha Ld^2/2 \tag{19}$$

при всех α, $0<\alpha<1$, $k\in N^-$. Далее, из (9) следует, что $f(x_{k+1})\leqslant f(x_k)+\delta_k$, $k=0,1,\ldots$ Так как $f(x_k)\geqslant f_*>-\infty$, $k=0,1,\ldots$, то из леммы 2.6.2 вытекает существование конечного предела $\lim\limits_{k\to\infty}f(x_k)\geqslant f_*$. Следовательно, $\lim\limits_{k\to\infty}(f(x_k)-f(x_{k+1}))=0$. Если N^- — бесконечное множество, то при $k\to\infty$, $k\in N^-$, из (19) имеем

$$0\leqslant\varlimsup_{k\to\infty}|f_k(\bar{x}_k)|\leqslant\varlimsup_{k\to\infty}|f_k(\bar{x}_k)|\leqslant\alpha Ld^2/2$$

при всех α, $0<\alpha<1$. Устремляя $\alpha\to+0$, отсюда получим $f_k(\bar{x}_k)\to 0$ при $k\to\infty$, $k\in N^-$. Объединяя оба случая $k\in N^+$ и $k\in N^-$, приходим к первому равенству (13). Так как X ограничено и $\{x_k\}\in X$, то последовательность $\{x_k\}$ имеет хотя бы одну предельную точку. Пусть x_* — произвольная предельная точка $\{x_k\}$, пусть $\{x_{k_m}\}\to x_*$. Согласно (4) имеем $f_k(\bar{x}_k)-\varepsilon_k\leqslant\inf\limits_X f_k(x)\leqslant\langle f'(x_k),x-x_k\rangle$ при всех $x\in X$ и $k=0,1,\ldots$ Отсюда при $k=k_m\to\infty$ с учетом первого равенства (13) получим, что $\langle f'(x_*),x-x_*\rangle\geqslant 0$ при всех $x\in X$. Тем самым показано, что любая предельная точка последовательности $\{x_k\}$ принадлежит S_*. Отсюда следуют второе равенство (13).

Пусть теперь $f(x)$ выпукла на X и x_* — произвольная точка из X_*. Тогда из теоремы 4.2.2 и условия (4) имеем

$$\begin{aligned}0\leqslant a_k=f(x_k)-f(x_*)&\leqslant\langle f'(x_k),x_k-x_*\rangle=\\&=-f_k(x_*)\leqslant-\min_X f_k(x)\leqslant-f_k(\bar{x}_k)+\varepsilon_k,\quad k=0,1,\ldots\end{aligned} \tag{20}$$

Отсюда и из первого равенства (13) следует $\lim\limits_{k\to\infty}f(x_k)=f_*$, т. е. $\{x_k\}$ — минимизирующая последовательность. Из теоремы 2.1.1 тогда получаем $\lim\limits_{k\to\infty}\rho(x_k,X_*)=0$. Равенства (15) доказаны. Заметим, что неравенство (20) может служить полезной апостериорной оценкой при практическом использовании метода (4), (5), (9).

Остается получить оценку (16). Для этого множество $N=\{0,1,2,\ldots\}$ разобьем на два множества $I_0=\{k\colon k\in N,\ a_k\geqslant\varepsilon_k\}$, $I_1=\{k\colon k\in N,\ 0\leqslant a_k<\varepsilon_k\}$. Из оценки

$$0\leqslant a_k-\varepsilon_k\leqslant-f_k(\bar{x}_k),\quad k\in I_0, \tag{21}$$

являющейся следствием неравенства (20), следует, что $I_0\subseteq N^-$. Поэтому (18) можно переписать в виде

$$a_k-a_{k+1}\geqslant\alpha|f_k(\bar{x}_k)|-\alpha^2Ld^2/2-\delta_k,\quad 0\leqslant\alpha\leqslant 1,\quad k\in I_0. \tag{22}$$

Так как в силу (13) $\{|f_k(\bar{x}_k)|\}$ ограничена, то, взяв при необходимости d еще большим, можем сделать $0\leqslant\bar{\alpha}_k=|f_k(\bar{x}_k)|d^{-2}L^{-1}\leqslant 1$ при всех $k=0,1,\ldots$ Принимая в (22) $\alpha=\bar{\alpha}_k$, получим

$$a_k-a_{k+1}\geqslant 1/(2Ld^2)|f_k(\bar{x}_k)|^2-\delta_k,\quad k\in I_0.$$

Отсюда и из (21) с учетом условия (14) имеем

$$\begin{aligned}a_{k+1}&\leqslant a_k-(a_k-\varepsilon_k)^2/(2Ld^2)+\delta_k\leqslant a_k-a_k^2/(2Ld^2)+\\&+(\sup_{k\geqslant 0}a_k)L^{-1}d^{-2}\varepsilon_k+\delta_k\leqslant a_k-a_k^2/A+Ak^{-2\rho},\quad k\in I_0,\end{aligned} \tag{23}$$

где $A=\max\{2Ld^2;(\sup\limits_{k\geqslant 0}a_k)L^{-1}d^{-2}C_0;C_0\}$.

Если $k\in I_1$, то $0\leqslant a_k<\varepsilon_k\leqslant C_0k^{-2\rho}$. Кроме того, из (18) при $\alpha\to+0$ получим $a_k-a_{k+1}+\delta_k\geqslant 0$ или $a_{k+1}\leqslant a_k+\delta_k\leqslant a_k+C_0k^{-2\rho}$ для всех $k=0,1,\ldots$ Таким образом, последовательность $\{a_k\}$ удовлетворяет условиям леммы 2.6.5, из которой следует оценка (16).

Наконец, оценка (17) вытекает из неравенства (4.3.3) и оценки (16). Теорема 1 доказана. □

3. Исследуем сходимость метода (4), (5), (10).

Теорема 2. *Пусть X — выпуклое замкнутое ограниченное множество из E^n, функция $f(x)$ принадлежит $C^{1,1}(X)$. Тогда при любом $x_0 \in X$ для последовательности $\{x_k\}$, определяемой условиями* (4), (5), (10), *справедливы равенства* (13). *Если, кроме того, $f(x)$ выпукла на X, то имеют место равенства* (15), *а при $\varepsilon_k \leqslant C_0 k^{-2\rho}$, $C_0 = \text{const} > 0$, $0 < \rho \leqslant 1$, верна оценка* (16). *Для сильно выпуклой функции справедлива оценка* (17).

Доказательство. Так же, как неравенство (18), нетрудно показать, что

$$f(x_k) - f(x_{k+1}) \geqslant -\alpha_k f_k(\bar{x}_k) - \alpha_k^2 L |\bar{x}_k - x_k|^2/2, \quad k = 0, 1, \ldots \tag{24}$$

В соответствии с формулой (10), определяющей величину α_k рассмотрим три возможных случая:

1) Если $f_k(\bar{x}_k) \leqslant 0$, $\alpha_k = 1 \leqslant \rho_k |f_k(\bar{x}_k)| |\bar{x}_k - x_k|^{-2}$, то из (24) с учетом (11) имеем

$$f(x_k) - f(x_{k+1}) \geqslant |f_k(\bar{x}_k)| - L\rho_k |f_k(\bar{x}_k)|/2 \geqslant \varepsilon |f_k(\bar{x}_k)|. \tag{25}$$

2) Если $f_k(\bar{x}_k) \leqslant 0$, $\alpha_k = \rho_k |f_k(\bar{x}_k)| |\bar{x}_k - x_k|^{-2} < 1$, то из (24) с учетом (11) получаем

$$f(x_k) - f(x_{k+1}) \geqslant \rho_k |f_k(\bar{x}_k)|^2 |\bar{x}_k - x_k|^{-2} - L\rho_k^2 |f_k(\bar{x}_k)|^2 |\bar{x}_k - x_k|^{-2}/2 =$$

$$= |f_k(\bar{x}_k)|^2 |\bar{x}_k - x_k|^{-2} \rho_k (1 - L\rho_k/2) \geqslant |f_k(\bar{x}_k)|^2 d^{-2} \varepsilon_0 \varepsilon, \quad d \geqslant \sup_{u, v \in X} |u - v|. \tag{26}$$

3) Наконец, если $f_k(\bar{x}_k) > 0$, то согласно (10) и из (24) имеем

$$f(x_k) - f(x_{k+1}) \geqslant 0, \tag{27}$$

а из (4) следует

$$0 < f_k(\bar{x}_k) \leqslant \varepsilon_k. \tag{28}$$

Из (25)–(27) вытекает, что последовательность $\{f(x_k)\}$ не возрастает. Так как $f(x_k) \geqslant f_* > -\infty$, то существует $\lim\limits_{k \to \infty} f(x_k) \geqslant f_*$ и, следовательно, $\lim\limits_{k \to \infty} (f(x_k) - f(x_{k+1})) = 0$. Отсюда и из (25), (26), (28) имеем $0 \leqslant |f_k(\bar{x}_k)| \leqslant \max\{\varepsilon_k; \text{const} \cdot (f(x_k) - f(x_{k+1}))^{1/2}\} \to 0$ при всех $k \to \infty$. Первое из равенств (13) доказано. Второе равенство (13) устанавливается так же, как в теореме 1.

Пусть теперь функция $f(x)$ выпукла на X. Тогда справедлива цепочка неравенств (20), из которой следуют равенства (15). Предполагая, что $\varepsilon_k \leqslant C_0 k^{-2\rho}$, $0 < \rho \leqslant 1$, докажем оценку (16). Предварительно заметим, что $0 \leqslant a_k = f(x_k) - f_* \leqslant \sup\limits_X f(x) - f_* = C_2 < \infty$, поэтому

$$a_k^2 \leqslant \sup_{k \geqslant 0} a_k \cdot a_k \leqslant C_2 a_k, \quad k = 0, 1, \ldots \tag{29}$$

Еще раз переберем рассмотренные выше три возможности.

1) Если $f_k(\bar{x}_k) \leqslant 0$, $\alpha_k = 1 \leqslant \rho_k |f_k(\bar{x}_k)| |\bar{x}_k - x_k|^{-2}$, то из (20), (25), (29) имеем $a_k - a_{k+1} \geqslant \varepsilon a_k - \varepsilon \varepsilon_k$ или

$$a_{k+1} \leqslant a_k - \varepsilon a_k + \varepsilon \varepsilon_k \leqslant a_k - a_k^2 (\varepsilon/C_2) + \varepsilon C_0 k^{-2\rho}. \tag{30}$$

2) Пусть $f_k(\bar{x}_k) \leqslant 0$, $\alpha_k = \rho_k |f_k(\bar{x}_k)| |\bar{x}_k - x_k|^{-2} < 1$. Здесь, в свою очередь, имеются две возможности: $a_k \geqslant \varepsilon_k$ или $0 \leqslant a_k < \varepsilon_k$. Если $a_k \geqslant \varepsilon_k$, то из (20), (26), $0 \leqslant a_k \leqslant C_2$ получим $a_k - a_{k+1} \geqslant (a_k - \varepsilon_k)^2 d^{-2} \varepsilon_0 \varepsilon \geqslant a_k^2 d^{-2} \varepsilon_0 \varepsilon - 2C_2 \varepsilon_k d^{-2} \varepsilon_0 \varepsilon$ или

$$a_{k+1} \leqslant a_k - a_k^2 (\varepsilon_0 \varepsilon / d^2) + 2C_2 \varepsilon_0 \varepsilon d^{-2} C_0 k^{-2\rho}. \tag{31}$$

Если же $0 \leqslant a_k < \varepsilon_k$, то достаточно воспользоваться более простым следствием (26): $a_{k+1} \leqslant a_k$. Последние два неравенства можно переписать в виде

$$0 \leqslant a_k \leqslant C_0 k^{-2\rho}, \quad a_{k+1} \leqslant a_k \leqslant a_k + C_0 k^{-2\rho}. \tag{32}$$

3) Наконец, пусть $f_k(\bar{x}_k) > 0$, $\alpha_k = 0$. Тогда из (20), (27) получим $0 \leqslant a_k \leqslant \varepsilon_k$, $a_{k+1} \leqslant a_k$, что снова приведет к неравенствам (32).

Из (30)–(32) следует, что последовательность $\{a_k\}$ удовлетворяет условиям леммы 2.6.5, на которой получаем оценку (16). Теорема 2 доказана. □

4. Наконец, рассмотрим вариант метода условного градиента (4), (5), (12) [783].

Теорема 3. *Пусть X — выпуклое замкнутое ограниченное множество из E^n, функция $f(x) \in C^{1,1}(X)$ и выпукла на X. Тогда при любом $x_0 \in X$ для последовательности $\{x_k\}$, определяемой условиями* (4), (5), (12), *справедливы равенства* (15). *Если при этом $\alpha_k = (k+1)^{-1}$, $\varepsilon_k = C_3 (k+1)^{-1}$, $k = 0, 1, \ldots$, то*

$$0 \leqslant f(x_k) - f_* \leqslant C_4 \ln(k+1)/k, \quad k = 1, 2, \ldots, \tag{33}$$

а если $\alpha_k = (k+1)^{-\beta}$, $\varepsilon_k = C_3 (k+1)^{-\beta}$, $k = 0, 1, \ldots$, $0 < \beta < 1$, то

$$0 \leqslant f(x_k) - f_* \leqslant C_4 k^{-\beta}, \quad k = 1, 2, \ldots; \tag{34}$$

здесь C_3, C_4 — некоторые положительные постоянные.

Доказательство. Заметим, что неравенства (20), (24) не зависят от способа выбора α_k, $0 \leqslant \alpha_k \leqslant 1$, в (5), поэтому сохраняют силу и в рассматриваемом случае. Из них имеем $a_k - a_{k+1} \geqslant \alpha_k (a_k - \varepsilon_k) - \alpha_k^2 L d^2/2$ или

$$a_{k+1} \leqslant (1 - \alpha_k) a_k + \alpha_k^2 L d^2/2 + \alpha_k \varepsilon_k, \quad k = 0, 1, \ldots$$

Отсюда с учетом свойств последовательностей $\{\alpha_k\}$, $\{\varepsilon_k\}$ из (4), (12) заключаем, что $\{a_k\}$ удовлетворяет условиям леммы 2.6.6. Поэтому $\lim\limits_{k \to \infty} a_k = 0$ или $\lim\limits_{k \to \infty} f(x_k) = f_*$. Отсюда и из теоремы 2.1.1 получаем равенства (15). Оценки (33), (34) следуют из лемм 2.6.8, 2.6.9.

Упражнения

1. Вычислить несколько итераций метода (3), (5), (6) для функции $f(u) = x^2 + xy + y^2$ при $u \in X = \{u = (x, y) \in E^2 : 0 \leqslant x \leqslant 1, -1 \leqslant y \leqslant 0\}$, выбирая $u_0 = (1, -1), (-1, 0), (1, 0)$ или $(0, 0)$.

2. Для функции из примера 1 проверить выполнение условий теорем 1–3 и сформулировать условия сходимости соответствующих вариантов метода условного градиента.

3. Дать описание различных вариантов метода условного градиента для функции $f(x) = |Ax - b|^2$, где A — матрица $m \times n$, $b \in E^m$, а множество X является шаром или параллелепипедом. Опираясь на теоремы 1–3, доказать сходимость метода.

§ 5. Метод возможных направлений

1. Продолжим рассмотрение задачи минимизации гладкой функции $f(x)$ на заданном множестве $X \subseteq E^n$. Напомним, что направление $e \neq 0$ называется *возможным* в точке $x \in X$, если $x + te \in X$ при всех t, $0 \leqslant t \leqslant t_0$, где t_0 — положительное число, зависящее от точки x, направления e и от структуры множества X (см. определение 4.2.3).

Определение 1. Направление $e \neq 0$ назовем *возможным направлением убывания* функции $f(x)$ точке x на множестве X, если e — возможное направление в точке x и $f(x + \alpha e) < f(x)$ при всех α, $0 < \alpha < \beta$, где $0 < \beta \leqslant t_0$.

Метод возможных направлений основан на следующей естественной и прозрачной идее: на каждой итерации этого метода определяется возможное направление убывания функции, и по этому направлению осуществляется спуск с некоторым положительным шагом. Собственно говоря, эта идея для нас уже не новая — именно на ней были основаны многие варианты изложенных в § 1, 2, 4 методов. В самом деле, если $X = E^n$, $f'(x) \neq 0$,

то возможное направление убывания функции легко находится — это направление антиградиента $e = -f'(x)$. Более трудным был выбор возможного направления убывания в методах § 2, 4: в методе проекция градиента (см. формулы (2.2) и (2.2')) для этого нужно было проектировать точку на исходное множество X, а в методе условного градиента — решать задачу минимизации линейной функции на множестве X (см. задачу (4.3)).

Понятно, что если задача выбора возможного направления убывания на каждой итерации слишком сложна и требует решения вспомогательных задач минимизации, сравнимых по трудности, быть может, с исходной задачей, то такой метод минимизации будет малоэффективным. Возникает вопрос: нельзя ли указать простые и достаточно удобные для реализации на ЭВМ способы выбора возможных направлений убывания? Оказывается, для достаточно широких классов гладких задач такие способы существуют. Покажем это на примере следующей задачи:

$$f(x) \to \inf; \quad x \in X = \{x \in E^n: g_i(x) \leqslant 0,\ i = 1, \ldots, m\}, \tag{1}$$

где функции $f(x)$, $g_i(x)$, $i = 1, \ldots, m$, определены на всем пространстве E^n и $f(x)$, $g_i(x) \in C^1(X)$.

Чтобы проще было пояснить суть метода возможных направлений для задачи (1), сначала опишем более простой вариант этого метода. Пусть $x_0 \in X$ — некоторое начальное приближение. Пусть известно k-е приближение $x_k \in X$, $k \geqslant 0$. Введем множество номеров

$$I_k = \{i: 1 \leqslant i \leqslant m,\ g_i(x_k) = 0\}.$$

Возможно, что $I_k = \varnothing$, — это будет означать, что $g_i(x_k) < 0$ при всех $i = 1, \ldots, m$, т. е. $x_k \in \operatorname{int} X$ — такая возможность ниже не исключается. В пространстве переменных

$$z = (e, \sigma) = (e^1, \ldots, e^n, \sigma) \in E^{n+1}$$

рассмотрим вспомогательную задачу

$$\sigma \to \inf, \quad z = (e, \sigma) \in W_k = \{(e, \sigma): \langle f'(x_k), e\rangle \leqslant \sigma,$$
$$\langle g_i'(x_k), e\rangle \leqslant \sigma,\ i \in I_k;\ |e^j| \leqslant 1,\ j = 1, \ldots, n\}. \tag{2}$$

Заметим, что задача (2) является задачей линейного программирования, причем минимизируемая функция $\langle c, z\rangle = \langle 0, e\rangle + 1 \cdot \sigma$, $c = (0, 1) \in E^{n+1}$, явно не зависит от переменных $e = (e^1, \ldots, e^n)$. Далее, ясно, что точка $z = (e = 0, \sigma = 0) = (0, 0) \in W_k$, так что $W_k \neq \varnothing$ и $\inf\limits_{W_k} \sigma = \sigma_k \leqslant 0$ при всех $k = 0, 1, \ldots$. Очевидно, множество W_k замкнуто. Наконец, условия $|e^j| \leqslant 1$, $j = 1, \ldots, n$, называемые условиями нормировки, гарантируют ограниченность множества W_k. Тогда из теоремы 2.1.1 следует, что задача (2) имеет хотя бы одно решение. Для получения решения задачи (2) могут быть использованы известные конечные методы линейного программирования (например, симплекс-метод, описанный в гл. 3).

Предположим, что задача (2) решена и найдены $(e_k, \sigma_k) \in W_k$ такие, что $\sigma_k = \inf\limits_{W_k} \sigma$. Выше было замечено, что $\sigma_k \leqslant 0$.

Сначала рассмотрим случай $\sigma_k < 0$. Оказывается, в этом случае направление e_k, полученное из задачи (2), является возможным направлением

убывания функции $f(x)$ в точке x_k. В самом деле, из условия $(e_k, \sigma_k) \in W_k$ следует, что

$$\langle f'(x_k), e_k\rangle \leqslant \sigma_k < 0, \quad \langle g_i'(x_k), e_k\rangle \leqslant \sigma_k < 0, \quad i \in I_k.$$

Отсюда ясно, что $e_k \neq 0$. Кроме того, для любого номера $i \in I_k$ имеем

$$g_i(x_k + \alpha e_k) = g_i(x_k + \alpha e_k) - g_i(x_k) = \langle g_i'(x_k), e_k\rangle \alpha + o(\alpha) \leqslant$$
$$\leqslant \alpha[\sigma_k + o(\alpha)/\alpha] < 0 \quad \text{при всех} \quad \alpha,\ 0 < \alpha < \alpha_i,\ \alpha_i > 0.$$

Если $i \notin I_k$, т. е. $g_i(x_k) < 0$, то в силу непрерывности функции $g_i(x)$ неравенство $g_i(x_k + \alpha e_k) < 0$ сохранится при всех α, $0 < \alpha < \alpha_i$, где $\alpha_i > 0$ — достаточно малое число. Положим $\alpha_0 = \min\{\alpha_1, \ldots, \alpha_m\} > 0$. Тогда

$$g_i(x_k + \alpha e_k) < 0, \quad i = 1, \ldots, m; \quad 0 < \alpha < \alpha_0,$$

т. е. e_k — возможное направление множества X в точке x_k.

Далее, взяв при необходимости число $\alpha_0 > 0$ еще меньшим, можно добиться выполнения неравенства

$$f(x_k + \alpha e_k) - f(x_k) = \langle f'(x_k), e_k\rangle \alpha + o(\alpha) \leqslant$$
$$\leqslant \alpha[\sigma_k + o(\alpha)/\alpha] < 0 \quad \text{при всех} \quad \alpha,\ 0 < \alpha < \alpha_0.$$

Тем самым показано, что если (e_k, σ_k) — решение задачи (2), причем $\sigma_k < 0$, то e_k — возможное направление убывания функции $f(x)$ в точке x_k на множестве X.

Используя найденное таким образом направление e_k, следующее $(k+1)$-е приближение определим так:

$$x_{k+1} = x_k + \alpha_k e_k, \quad 0 < \alpha_k \leqslant \beta_k, \tag{3}$$

где

$$\beta_k = \sup\{\alpha; x_k + te_k \in X,\ 0 \leqslant t \leqslant \alpha\} > 0. \tag{4}$$

Выбирая α_k в (3) различными способами, будем получать различные варианты метода возможных направлений. Перечислим некоторые способы выбора α_k.

1) Величина α_k может выбираться из условий

$$0 < \alpha_k \leqslant \beta_k, \quad \bar{g}_k(\alpha_k) = \inf_{0 < \alpha \leqslant \beta_k} \bar{g}_k(\alpha) = \bar{g}_{k*}; \quad \bar{g}_k(\alpha) = f(x_k + \alpha e_k). \tag{5}$$

Для минимизации функции $\bar{g}_k(\alpha)$ могут быть использованы известные методы (см., например, гл. 1). Точное решение задачи (5) удается найти лишь в редких случаях; возможно также, что на некоторых направлениях e_k величина $\beta_k = \infty$ и нижняя грань функции $\bar{g}_k(\alpha)$ при $\alpha > 0$ не достигается. Поэтому вместо (5) на практике целесообразно употреблять такой способ выбора α_k:

$$0 < \alpha_k \leqslant \beta_k, \quad \bar{g}_k(\alpha_k) \leqslant \bar{g}_{k*} + \delta_k, \quad \delta_k \geqslant 0, \quad \sum_{k=0}^{\infty} \delta_k = \delta < \infty \tag{6}$$

или

$$f(x_k + \alpha_k e_k) \leqslant (1 - \lambda_k) f(x_k) + \lambda_k \bar{g}_{k*}, \quad 0 < \lambda \leqslant \lambda_k \leqslant 1.$$

2) Если функция $f(x) \in C^{1,1}(X)$ и константа Липшица L для градиента $f'(x)$ известна, то в (3) в качестве α_k можно принять

$$\alpha_k = \min\{\beta_k; |\sigma_k| L^{-1}\}.$$

3) Возможен выбор величин α_k из следующих условий:

$$f(x_k) - f(x_k + \alpha_k e_k) \geqslant \varepsilon\alpha_k|\sigma_k|, \quad 0 < \alpha_k \leqslant \beta_k, \quad 0 < \varepsilon < 1/2.$$

Для определения такого α_k сначала можно положить $\alpha_k = \beta_k$, а затем при необходимости дробить эту величину.

4) В тех случаях, когда трудно оценить величину β_k из (4), приходится довольствоваться нахождением какого-либо $\alpha_k > 0$, обеспечивающего включение $x_k + \alpha_k e_k \in X$ и условие монотонности $f(x_k + \alpha_k e_k) < f(x_k)$. Для этого обычно выбирают какую-либо постоянную $\alpha > 0$, полагают $\alpha_k = \alpha$ и проверяют условие монотонности и принадлежность точки x_{k+1} множеству X; при необходимости дробят величину $\alpha_k = \alpha$, добиваясь выполнения упомянутых условий.

Один шаг метода возможных направлений для задачи (1) в случае $\sigma_k < 0$ описан. Попутно выяснен смысл вспомогательной задачи (2): минимизируя σ, мы добиваемся того, чтобы направление e_k было как можно ближе к направлению антиградиента (это обеспечивается условием $\langle f'(x_k), e_k\rangle \leqslant \sigma$) и в то же время оставалось возможным направлением для множества X в точке x_k (это обеспечивается условиями $\langle g_i'(x_k), e_k\rangle \leqslant \sigma$, $i \in I_k$), причем, чем меньше σ, тем ярче выражены указанные свойства направления e_k. Кстати, если $I_k = \varnothing$, т. е. $x_k \in \operatorname{int} X$, то $e_k = -\alpha f'(x_k)$, $\alpha = (\max\limits_{1 \leqslant j \leqslant n} |f_{x^j}(x_k)|)^{-1} > 0$ — направление антиградиента.

Теперь рассмотрим случай, когда в решении (e_k, σ_k) задачи (2) координата $\sigma_k = 0$. Как видно из (2), это может случиться, например, при $f'(x_k) = 0$ или $g_i'(x_k) = 0$ для некоторого номера $i \in I_k$. При $\sigma_k = 0$ уже нельзя гарантировать, что e_k будет возможным направлением убывания. В этом случае итерационный процесс (2)–(4) прекращается. Оказывается, при $\sigma_k = 0$ в точке x_k выполняются необходимые условия минимума, выраженные в теореме 4.8.1. Для выпуклой задачи (1) со свойством (4.9.15) условие $\sigma_k = 0$ гарантирует, что $x_k \in X_*$. А именно справедлива

Теорема 1. *Пусть функции $f(x)$, $g_i(x)$, $i = 1, \ldots, m$, определены на E^n, $f(x)$, $g_i(x) \in C^1(X)$, где множество X задано условиями* (1), *и пусть задача* (1) *имеет решение, т. е. $f_* > -\infty$, $X_* \neq \varnothing$. Тогда для любой точки $x_* \in X_*$ задача*

$$\sigma \to \inf; \quad z = (e, \sigma) \in W_* = \{(e, \sigma): \langle f'(x_*), e\rangle \leqslant \sigma,$$
$$\langle g_i'(x_*), e\rangle \leqslant \sigma,\ i \in I_*,\ |e^j| \leqslant 1,\ j = 1, \ldots, n\}, \quad (7)$$

где

$$I_* = \{i: 1 \leqslant i \leqslant m,\ g_i(x_*) = 0\},$$

необходимо имеет решение (e_*, σ_*) с $\sigma_* = \min\limits_{W_*} \sigma = 0$. Если, кроме того, $f(x)$, $g_i(x)$ выпуклы на E^n, и выполнено условие Слейтера (4.9.15), *то всякая точка $x_* \in X$, для которой задача* (7) *определяет величину $\sigma_* = \inf\limits_{W_*} \sigma = 0$, является решением задачи* (1).

Доказательство. Необходимость. Пусть $x_* \in X_*$. По теореме 2.3.2 или 4.8.1 тогда существуют множители Лагранжа $\lambda_0^*, \ldots, \lambda_m^*$, неотрицательные и не все равные нулю, такие, что

$$\lambda_0^* f'(x_*) + \sum_{i=1}^m \lambda_i^* g_i'(x_*) = 0, \quad \lambda_i^* g_i(x_*) = 0, \quad i = 1, \ldots, m. \quad (8)$$

Если $i \notin I_*$, то из второго равенства (8) следует $\lambda_i^* = 0$, поэтому первое равенство (8) можно переписать в виде

$$\lambda_0^* f'(x_*) + \sum_{i \in I_*} \lambda_i^* g_i'(x_*) = 0. \quad (9)$$

Возьмем любую точку $(e, \sigma) \in W_*$. Тогда $\langle f'(x_*), e\rangle \leqslant \sigma$, $\langle g_i'(x_*), e\rangle \leqslant \sigma$, $i \in I_*$. Умножим первое из этих неравенств на $\lambda_0^* \geqslant 0$, остальные — на соответствующие $\lambda_i^* \geqslant 0$ и сложим. С учетом равенства (9) получим

$$\langle \lambda_0^* f'(x_*) + \sum_{i \in I_*} \lambda_i^* g_i'(x_*), e\rangle = 0 \leqslant \sigma(\lambda_0^* + \lambda_1^* + \ldots + \lambda_m^*).$$

Следовательно, $\sigma \geqslant 0$ $\forall (e, \sigma) \in W_*$, и $\sigma_* = 0$.

Достаточность. Пусть теперь $f(x)$, $g_i(x)$ выпуклы на X, выполнено условие Слейтера (4.9.15), и пусть для некоторой точки $x_* \in X$ задача (7) определяет величину $\sigma_* = \inf\limits_{W_*} \sigma = 0$. Покажем, что тогда $x_* \in X_*$. С этой целью в пространстве E^{n+1} введем конус

$$K = \{z = (e, \sigma) \in E^{n+1}: \langle f'(x_*), e\rangle + (-1)\sigma \leqslant 0,\ \langle g_i'(x_*), e\rangle + (-1)\sigma \leqslant 0,\ i \in I_*\}$$

и вектор $d = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Из (7) с учетом $\inf\limits_{W_*} \sigma = \sigma_* = 0$ имеем

$$\langle d, z\rangle = \langle 0, e\rangle + 1 \cdot \sigma = \sigma \geqslant \sigma_* = 0 \quad (10)$$

для всех $z = (e, \sigma) \in K$, для которых $|e^j| \leqslant 1$, $j = 1, \ldots, n$. Однако условие $|e^j| \leqslant 1$, $j = 1, \ldots, n$ здесь можно отбросить, и неравенство (10) на самом деле верно для всех $z \in K$. В самом деле, пусть $z = (e, \sigma) \in K$ и $|e^j| > 1$ для некоторого номера j, $1 \leqslant j \leqslant n$. Тогда $\|e\| = \max\limits_{1 \leqslant j \leqslant n} |e^j| > 1$. Положим

$$\bar{z} = (\bar{e}, \bar{\sigma}), \quad \bar{e} = e/\|e\|, \quad \bar{\sigma} = \sigma/\|e\|.$$

Ясно, что $\bar{z} \in W_*$. Следовательно, $\langle d, \bar{z}\rangle = \langle d, z\rangle/\|e\| = \bar{\sigma} = \sigma/\|e\| \geqslant 0$, так что $\langle d, z\rangle = \sigma \geqslant 0$. Тем самым показано, что неравенство (10) верно для всех $z \in K$.

По теореме Фаркаша 3.5.8 тогда существуют неотрицательные числа $\lambda_0^*, \ldots, \lambda_m^*$ такие, что

$$d = \begin{pmatrix} 0 \\ 1 \end{pmatrix} = -\lambda_0^* \begin{pmatrix} f'(x_*) \\ -1 \end{pmatrix} - \sum_{i \in I_*} \lambda_i^* \begin{pmatrix} g_i'(x_*) \\ -1 \end{pmatrix}$$

или

$$0 = -\lambda_0^* f'(x_*) - \sum_{i \in I_*} \lambda_i^* g_i'(x_*), \quad 1 = \lambda_0^* + \sum_{i \in I_*} \lambda_i^*. \quad (11)$$

Кроме того, из определения множества I_* следует, что $g_i(x_*) = 0$, поэтому $\lambda_i^* g_i(x_*) = 0$, $i \in I_*$. Доопределим $\lambda_i^* = 0$ при всех $i \notin I_*$. В результате с учетом первого равенства (11) получим

$$\lambda_0^* f'(x_*) + \sum_{i=1}^m \lambda_i^* g_i'(x_*) = 0, \quad \lambda_i^* g_i(x_*) = 0, \quad i = 1, \ldots, m, \quad (12)$$

а из второго равенства (11) следует, что $\bar{\lambda} = (\lambda_0^*, \lambda_1^*, \ldots, \lambda_m^*) \neq 0$.

Покажем, что $\lambda_0^* > 0$. Если $I_* = \varnothing$, то из (11) сразу имеем $\lambda_0^* = 1$. Допустим, что $I_* \neq \varnothing$, но тем не менее $\lambda_0^* = 0$. Тогда среди неотрицательных чисел λ_i^*,

$i \in I_*$, найдется хотя бы одно положительное число. Пусть $\lambda_p^* > 0$, $p \in I_*$. По условию существует точка $\bar{x} \in X$ такая, что $g_i(\bar{x}) < 0$ для всех $i = 1, \ldots, m$. Поскольку $I_* \neq \varnothing$, то $\bar{x} \neq x_*$. В силу выпуклости множества X тогда $\alpha\bar{x} + (1-\alpha)x_* = x_* + \alpha(\bar{x} - x_*) \in X$ при всех α, $0 \leqslant \alpha \leqslant 1$. Это значит, что направление $e = \bar{x} - x_* \neq 0$ является возможным для множества X в точке x_*. Из выпуклости функций $g_i(x)$ для всех $i \in I_*$ имеем $0 > g_i(\bar{x}) = g_i(\bar{x}) - g_i(x_*) \geqslant \langle g_i'(x_*), \bar{x} - x_* \rangle = \langle g_i'(x_*), e \rangle$. Поэтому $\sum_{i=1}^{m} \lambda_i^* \langle g_i'(x_*), e \rangle \leqslant \lambda_p^* \langle g_p'(x_*), e \rangle < 0$. Но с другой стороны, из первого равенства (12) при $\lambda_0^* = 0$ получим $\sum_{i=1}^{m} \lambda_i^* \langle g_i'(x_*), e \rangle = 0$. Полученное противоречие показывает, что $\lambda_0^* > 0$. Разделив первое равенство (12) на $\lambda_0^* > 0$ и сделав очевидные переобозначения, придем к равенству $f'(x_*) + \sum_{i=1}^{m} \lambda_i^* g_i(x_*) = 0$. Отсюда и из второго равенства (12) с помощью леммы 4.9.2 и теоремы 4.9.1 получим, что $x_* \in X_*$. Теорема 1 доказана. □

В невыпуклых задачах условие $\sigma_* = 0$ не является достаточным для оптимальности точки x_*. Это показывает следующий

Пример 1. Пусть $f(u) = x + \cos y$, $u \in X = \{u = (x, y) \in E^2 : g(u) = -x \leqslant 0\}$. Возьмем точку $u_* = (0, 0)$. Тогда $f'(u_*) = (1, 0)$, $g'(u_*) = (-1, 0)$, $W_* = \{(e, \sigma) = (e^1, e^2, \sigma) : e^1 \leqslant \sigma, -e^1 \leqslant \sigma, |e^1| \leqslant 1, |e^2| \leqslant 1\}$. Отсюда $|e^1| \leqslant \sigma$ при всех $(e, \sigma) \in W_*$. Это значит, что $\inf_{W_*} \sigma = \sigma_* = 0$, причем нижняя грань достигается при $e_* = (0, 1)$ или $e_* = (0, -1)$, $\sigma_* = 0$. Но здесь $u_* = (0, 0)$ не является точкой минимума $f(u)$ на X. Любопытно заметить, что векторы $e_* = (0, 1)$ или $(0, -1)$ в данном случае являются возможными направлениями убывания.

2. Описанный выше вариант метода возможных направлений (2)–(4) на практике применяют редко. Дело и том, что когда в решении (e_k, σ_k) задачи (2) координата $\sigma_k < 0$ мала по абсолютной величине, направление e_k теоретически являясь возможным направлением убывания в точке x_k, практически может обладать указанными свойствами в весьма слабой форме. Это означает, что либо $\langle g_i'(x_k), e_k \rangle \approx \sigma_k \approx 0$ при некотором $i \in I_k$ и направление e_k почти «касается» множества X, не ведет «вглубь» X, а величина β_k из (4) может оказаться очень малой, либо $\langle f'(x_k), e_k \rangle \approx \sigma_k \approx 0$, т. е. вдоль e_k функция $f(x)$ в точке x_k убывает слишком медленно. В результате длина шага α_k в (3) может получиться очень малой даже вдали от стационарной точки, и сходимость метода может оказаться очень медленной.

Чтобы избежать таких неприятных явлений, можно попытаться варьировать множество номеров I_k в (2) и осуществлять спуск из точки x_k только в том случае, когда получаемое из (2) направление e_k обладает достаточно ярко выраженными свойствами возможного направления убывания.

Опишем один из таких подходов. Пусть $x_0 \in X$, $\varepsilon_0 > 0$ — некоторое начальное приближение. Допустим, что k-е приближение (x_k, ε_k), $x_k \in X$, $\varepsilon > 0$, при каком-то $k \geqslant 0$ уже известно. Определим множество номеров

$$I_k = \{i : 1 \leqslant i \leqslant m, \ -\varepsilon_k \leqslant g_i(x_k) \leqslant 0\} \tag{13}$$

и решим вспомогательную задачу (2) при таком I_k. Задача (2) по-прежнему будет задачей линейного программирования и будет обладать хотя бы одним решением (e_k, σ_k) с $\sigma_k = \inf_{W_k} \sigma \leqslant 0$. Имеются две возможности:

1) $\sigma_k \leqslant -\varepsilon_k$. В этом случае считаем, что e_k является достаточно хорошим возможным направлением убывания в точке x_k, и полагаем

$$x_{k+1} = x_k + \alpha_k e_k, \quad 0 < \alpha_k \leqslant \beta_k, \quad \varepsilon_{k+1} = \varepsilon_k, \tag{14}$$

где β_k определяется из (4), а выбор α_k может быть осуществлен одним из описанных выше способов.

2) $-\varepsilon_k < \sigma_k \leqslant 0$. В этом случае считаем, что направление e_k не обладает ясно выраженным свойством возможного направления убывания в точке x_k, полагаем

$$x_{k+1} = x_k, \quad \varepsilon_{k+1} = \varepsilon_k / 2 \tag{15}$$

и снова переходим к рассмотрению задачи (2) с заменой множества I_k на множество $I_{k+1} = \{i : 1 \leqslant i \leqslant m, /, -\varepsilon_{k+1} = -\varepsilon_k/2 \leqslant g_i(x_k) \leqslant 0\}$, надеясь на то, что на более широком множестве (при сужении I_k множество W_k, вообще говоря, расширяется) удастся найти лучшее возможное направление убывания и т. д.

Описание одной итерации метода возможных направлений для задачи (1) закончено. В методе (2), (13)–(15) имеются параметры $\varepsilon_0, \varepsilon_1, \ldots$, удачным выбором которых, вообще говоря, можно улучшить выбор направлений e_k на каждой итерации, ускорить сходимость метода. Кстати, в (15) вместо деления пополам можно принять иной способ дробления ε_k, например, $\varepsilon_{k+1} = 0,9\varepsilon_k$.

3. Следуя [374], изучим сходимость метода (2), (3), (6), (13)–(15). Предварительно докажем несколько лемм.

Лемма 1. *Пусть $f(x), g_i(x) \in C^{1,1}(X)$, $i = 1, \ldots, m$, и I — некоторое фиксированное множество номеров, взятых из $\{1, 2, \ldots, m\}$ (возможности $I = \varnothing$ или $I = \{1, \ldots, m\}$ не исключаются). Для каждого $x \in X$ положим $\sigma(x) = \min_{G(x)} \sigma$, где $G(x) = \{(e, \sigma) = (e^1, \ldots, e^n, \sigma) \in E^{n+1} : \langle f'(x), e \rangle \leqslant \sigma, \langle g_i'(x), e \rangle \leqslant \sigma, i \in I; |e^j| \leqslant 1, j = 1, \ldots, n\}$. Тогда*

$$|\sigma(u) - \sigma(v)| \leqslant L\sqrt{n}|u - v|, \quad u, v \in X, \tag{16}$$

где L — константа Липшица для градиентов $f'(x)$, $g_i'(x)$, $i = 1, \ldots, m$.

Доказательство. Возьмем произвольные точки $u, v \in X$. Пусть $(e, \sigma) \in G(v)$, т. е.

$$\langle f'(v), e \rangle \leqslant \sigma, \quad \langle g_i'(v), e \rangle \leqslant \sigma, \quad i \in I, \quad |e^j| \leqslant 1, \quad j = 1, \ldots, n.$$

Тогда

$$\langle f'(u), e \rangle = \langle f'(v), e \rangle + \langle f'(u) - f'(v), e \rangle \leqslant \sigma + L|u - v||e| \leqslant \sigma + L|u - v|\sqrt{n}$$

и, аналогично,

$$\langle g_i'(u), e \rangle \leqslant \sigma + L|u - v|\sqrt{n}, \quad i \in I.$$

Это значит, что при каждом $(e, \sigma) \in G(v)$ точка $(e, \sigma + L\sqrt{n}|u - v|)$ принадлежит множеству $G(u)$. Тогда $\sigma(u) = \min_{G(u)} \sigma \leqslant \sigma + L\sqrt{n}|u - v|$ для любых $(e, \sigma) \in G(v)$. Следовательно, $\sigma(u) \leqslant \sigma(v) + L\sqrt{n}|u - v|$. Поменяв в этих рассуждениях точки u, v ролями, получим $\sigma(v) \leqslant \sigma(u) + L\sqrt{n}|u - v|$. Из последних двух неравенств следует неравенство (16). □

Лемма 2. *Пусть*

$$f(x), g_1(x), \ldots, g_m(x) \in C^{1,1}(X), \quad A_0 = \max_{1 \leqslant i \leqslant m} \sup_{x \in X} |g_i'(x)| < \infty,$$

а последовательности $\{x_k\}$, $\{e_k\}$, $\{\alpha_k\}$, $\{\beta_k\}$, $\{\varepsilon_k\}$, $\{\sigma_k\}$ определены условиями (2), (3), (6), (13)–(15). *Тогда*

$$\beta_k \geqslant A_1 \min\{\varepsilon_k, |\sigma_k|\}, \quad k = 0, 1, \ldots, \tag{17}$$

где $A_1 = \min\{1/(A_0\sqrt{n}); 1/(nL)\} > 0$, L — константа Липшица для градиентов $f'(x)$, $g_1'(x), \ldots, g_m'(x)$.

Доказательство. Если $\beta_k = \infty$, то неравенство (17) верно. Поэтому пусть $\beta_k < \infty$. Из определения (4) величины β_k и замкнутости X следует, что $x_k + \beta_k e_k \in X$ и $g_i(x_k + \beta_k e_k) = 0$ хотя бы для одного номера i. Зафиксируем один из таких номеров i. Может оказаться, что $g_i(x_k) < -\varepsilon_k$. Тогда $\varepsilon_k < -g_i(x_k) = g_i(x_k + \beta_k e_k) - g_i(x_k) = \langle g_i'(x_k + \theta\beta_k e_k), \beta_k e_k\rangle \leqslant A_0\beta_k|e_k| \leqslant A_0\sqrt{n}\beta_k$, т. е. $\beta_k \geqslant \varepsilon_k/(A_0\sqrt{n})$.

Если же оказалось, что $-\varepsilon_k \leqslant g_i(x_k) \leqslant 0$, то $i \in I_k$ и $\langle g_i'(x_k), e_k\rangle \leqslant \sigma_k \leqslant 0$. Допустим, что $\sigma_k < 0$. Тогда направление e_k является возможным в точке x_k и заведомо $\beta_k > 0$. По определению β_k имеем $g_i(x_k + \alpha e_k) \leqslant 0$ при всех α, $0 < \alpha < \beta_k$. Кроме того, $g_i(x_k + \beta_k e_k) = 0$ по выбору номера i. Тогда $0 \geqslant g_i(x_k + \alpha e_k) - g_i(x_k + \beta_k e_k) = \langle g_i'(x_k + \beta_k e_k), e_k\rangle(\alpha - \beta_k) + o(|\alpha - \beta_k|)$ или $\langle g_i'(x_k + \beta_k e_k), e_k\rangle \geqslant o(|\alpha - \beta_k|)(\beta_k - \alpha)^{-1}$ при всех α, $0 < \alpha < \beta_k$. Отсюда при $\alpha \to \beta_k - 0$ получим $\langle g_i'(x_k + \beta_k e_k), e_k\rangle \geqslant 0$. Тогда $|\sigma_k| = -\sigma_k \leqslant \langle -g_i'(x_k), e_k\rangle \leqslant \langle g_i'(x_k + \beta_k e_k) - g_i'(x_k), e_k\rangle \leqslant L\beta_k|e_k|^2 \leqslant Ln\beta_k$, т. е. $\beta_k \geqslant |\sigma_k|/(nL)$. Если $\sigma_k = 0$, то последнее неравенство также остается верным, так как согласно (4) всегда $\beta_k \geqslant 0$. Объединяя обе полученные оценки для β_k, приходим к оценке (17). Лемма 2 доказана. □

Лемма 3. *Пусть $f(x)$, $g_i(x) \in C^{1,1}(X)$, $i = 1, \ldots, m$. Пусть, кроме того, в процессе* (2), (3), (6), (13)–(15) *на некоторой k-й итерации оказалось $\sigma_k \leqslant -\varepsilon_k$. Тогда*

$$f(x_k) - f(x_{k+1}) \geqslant A_2 \min\{\beta_k|\sigma_k|; \sigma_k^2\} - \delta_k, \tag{18}$$

где $A_2 = \min\{1/2; 1/(2nL)\} > 0$.

Доказательство. Из неравенства $\sigma_k \leqslant -\varepsilon_k$ и определения e_k, σ_k следует, что $\langle f'(x_k), e_k\rangle \leqslant \sigma_k \leqslant -\varepsilon_k < 0$. Кроме того, e_k является возможным направлением в точке x_k и, следовательно, $\beta_k > 0$. Из (6) и леммы 2.6.1 имеем

$$f(x_k) - f(x_{k+1}) \geqslant f(x_k) - \inf_{0 \leqslant \alpha \leqslant \beta_k} g_k(\alpha) - \delta_k \geqslant f(x_k) - f(x_k + \alpha e_k) - \delta_k \geqslant$$
$$\geqslant -\alpha\langle f'(x_k), e_k\rangle - \alpha^2 L|e_k|^2/2 - \delta_k \geqslant -\alpha\sigma_k - \alpha^2 n/2 - \delta_k \tag{19}$$

при всех α, $0 \leqslant \alpha \leqslant \beta_k$. Положим здесь $\alpha = \min\{\beta_k; |\sigma_k|/(nL)\}$.

Может случиться, что $\alpha = \beta_k \leqslant |\sigma_k|/(nL)$. Тогда из (19) получаем

$$f(x_k) - f(x_{k+1}) \geqslant \alpha|\sigma_k| - \alpha \cdot \alpha \cdot nL/2 - \delta_k \geqslant \beta_k|\sigma_k| - \beta_k(|\sigma_k|/(nL))(nL/2) - \delta_k = \beta_k|\sigma_k|/2 - \delta_k.$$

Если же $\alpha = |\sigma_k|/(nL) < \beta_k$, то из (19) следует

$$f(x_k) - f(x_{k+1}) \geqslant \sigma_k^2/(nL) - (\sigma_k^2/(nL)^2)(nL/2) - \delta_k = \sigma_k^2/(2nL) - \delta_k.$$

Объединяя оба рассмотренных случая, приходим к оценке (18). □

Теорема 2. *Пусть функции $f(x)$, $g_i(x)$, $i = 1, \ldots, m$ определены и выпуклы на E^n, выполнено условие Слейтера* (4.9.15); *$f(x)$, $g_i(x) \in C^{1,1}(X)$, $A_0 = \max\limits_{1 \leqslant i \leqslant m} \sup\limits_X |g_i'(x)| < \infty$. Пусть задача* (1) *имеет решение, т. е. $f_* > -\infty$, $X_* \neq \varnothing$, и начальная точка $x_0 \in X$ такова, что множество $M_\delta(x_0) = \{x: x \in X, f(x) \leqslant f(x_0) + \delta\}$ ограничено. Тогда при любом выборе $\varepsilon_0 > 0$ для последовательности $\{x_k\}$, определяемой условиями* (2), (3), (6), (13)–(15), *справедливы равенства*

$$\lim_{k \to \infty} f(x_k) = f_*, \quad \lim_{k \to \infty} \rho(x_k, X_*) = 0. \tag{20}$$

Доказательство. Сначала установим, что

$$\lim_{k \to \infty} (f(x_k) - f(x_{k+1})) = 0. \tag{21}$$

Если $\sigma_k \leqslant -\varepsilon_k$, то из (14), (6) имеем $f(x_{k+1}) = \bar{g}_k(\alpha_k) \leqslant \bar{g}_k(0) + \delta_k = f(x_k) + \delta_k$. Если же $-\varepsilon_k < \sigma_k \leqslant 0$, то из (15) следует $f(x_{k+1}) = f(x_k) \leqslant f(x_k) + \delta_k$. Таким образом,

$$f(x_{k+1}) \leqslant f(x_k) + \delta_k, \quad k = 0, 1, \ldots; \quad \sum_{k=0}^{\infty} \delta_k = \delta < \infty. \tag{22}$$

и, кроме того, $f(x_k) \geqslant f_* > -\infty$, $k = 0, 1, \ldots$

Согласно лемме 2.6.2 тогда существует $\lim\limits_{k \to \infty} f(x_k) \geqslant f_*$. Отсюда следует равенство (21).

Далее, покажем, что $\lim\limits_{k \to \infty} \varepsilon_k = 0$. Согласно (14), (15) последовательность $\{\varepsilon_k\}$ получается дроблением и не возрастает. Допустим, что $\lim\limits_{k \to \infty} \varepsilon_k = \varepsilon > 0$. Это значит, что в процессе построения $\{x_k\}$ было конечное число дроблений и $\varepsilon_k = \varepsilon > 0$ при всех $k \geqslant k_0$. Из (14) тогда имеем $\sigma_k \leqslant -\varepsilon_k = -\varepsilon$, т. е. $|\sigma_k| \geqslant \varepsilon$, $k \geqslant k_0$. В этом случае согласно лемме 2 имеем $\beta_k \geqslant A_1\varepsilon$, $k \geqslant k_0$. Поэтому из леммы 3 получим $f(x_k) - f(x_{k+1}) \geqslant A_2 \min\{A_1\varepsilon^2; \varepsilon^2\} - \delta_k$, $k \geqslant k_0$, что противоречит равенству (21). Итак, показано, что $\lim\limits_{k \to \infty} \varepsilon_k = 0$.

Пусть $k_1 < k_2 < \ldots < k_r < \ldots$ — номера тех итераций, когда происходит дробление ε_k. Согласно (14), (15) тогда $-\varepsilon_{k_r} \leqslant \sigma_{k_r} \leqslant 0$, $r = 1, 2, \ldots$ Следовательно, $\lim\limits_{r \to \infty} \sigma_{k_r} = 0$. Тем самым установлено, что существует хотя бы одна подпоследовательность $\{\sigma_{k_r}\}$, сходящаяся к нулю.

Возьмем произвольную подпоследовательность $\{\sigma_{k_r}\} \to 0$. Покажем, что тогда любая предельная точка соответствующей подпоследовательности $\{x_{k_r}\}$ принадлежит множеству X_*. Из (22) следует, что $f(x_{k+1}) \leqslant f(x_0) + \delta$, $k = 0, 1, \ldots$, т. е. $\{x_k\} \in M_\delta(x_0)$. По условию множество $M_\delta(x_0)$ ограничено. Поэтому можем считать, что взятая выше подпоследовательность $\{x_{k_r}\}$ сходится к некоторой точке x_*. Далее, множество номеров I_k, определяемое согласно (13), представляет собой подмножество конечного числа номеров $\{1, 2, \ldots, m\}$, поэтому число различных множеств I_k конечно. Это значит, что среди $\{I_{k_r}, r = 1, 2, \ldots\}$ найдется хотя бы одно множество $I_{k_r} = I$, которое повторяется бесконечно много раз. Выбирая при необходимости подпоследовательности, можем, таким образом, считать, что

$$\{\sigma_{k_r}\} \to 0, \quad \{x_{k_r}\} \to x_*, \quad I_{k_r} = I, \quad r = 1, 2, \ldots$$

Согласно лемме 1 при $G(x_{k_r}) = W_{k_r}$ имеем

$$|\sigma_{k_r} - \sigma(x_*)| = |\sigma(x_{k_r}) - \sigma(x_*)| \leqslant L\sqrt{n}|x_{k_r} - x_*| \to 0,$$

т. е. $\lim\limits_{r \to \infty} \sigma(x_{k_r}) = \lim\limits_{r \to \infty} \sigma_{k_r} = 0 = \sigma(x_*)$, где $\sigma(x_*) = \inf\limits_{G(x_*)} \sigma$, $G(x_*) = \{(e, \sigma) \in E^{n+1}: \langle f'(x_*), e\rangle \leqslant \sigma, \langle g_i'(x_*), e\rangle \leqslant \sigma, i \in I; |e^j| \leqslant 1, j = 1, \ldots, n\}$.

Рассмотрим задачу (7), соответствующую точке $x_* = \lim\limits_{r \to \infty} x_{k_r}$. Покажем, что $W_* \subseteq G(x_*)$. По определению множеств $I = I_{k_r} = \{i: 1 \leqslant i \leqslant r, -\varepsilon_{k_r} \leqslant g_i(x_{k_r}) \leqslant 0\}$, $r = 1, 2, \ldots$ Отсюда при $r \to \infty$ получим $g_i(x_*) = 0$ для всех $i \in I$. Это значит, что $I \subseteq I_*$, т. е. в определении множества W_* число ограничений типа неравенств не меньше, чем число таких ограничений в определении $G(x_*)$. Тем самым установлено, что $W_* \subseteq G(x_*)$. А тогда, замечая, что одна и та же функция на более широком множестве имеет меньшую нижнюю грань, получаем $\sigma_* = \inf\limits_{W_*} \sigma \geqslant \inf\limits_{G(x_*)} \sigma = \sigma(x_*) = 0$. С другой стороны, $(0, 0) \in W_*$, поэтому $\sigma_* \leqslant 0$. Следовательно, $\sigma_* = 0$ и согласно теореме 1 имеем $x_* \in X_*$.

Выше было доказано существование предела $\lim\limits_{k \to \infty} f(x_k)$. Теперь можем сказать, чему равен этот предел: $\lim\limits_{k \to \infty} f(x_k) = \lim\limits_{k \to \infty} f(x_{k_r}) = f(x_*) = f_*$.

Таким образом, построенная последовательность $\{x_k\}$ минимизирует функцию $f(x)$ на множестве X. Поскольку $\{x_k\} \in M_\delta(x_0)$ — ограниченное множество, то из теоремы 2.1.2 следует, что $\lim\limits_{k \to \infty} \rho(x_k, X_*) = 0$. Равенства (20) и, тем самым, теорема 2 доказаны. □

4. Для задачи

$$f(x) \to \inf; \quad x \in X = \{x \in E^n: g_i(x) \leqslant 0, \ i = 1, \ldots, m, \ g_i(x) = \langle a_i, x\rangle - b^i = 0, \ i = m+1, \ldots, s\},$$

содержащей линейные ограничения типа равенств, метод возможных направлений описывается так же, как выше, лишь в задаче (2) нужно добавить еще ограничения $\langle a_i, e\rangle = 0$, $i = m+1, \ldots, s$.

Можно заметить, что описанный в гл. 3 симплекс-метод для решения канонической задачи линейного программирования по существу является вариантом метода возможных направлений. Более того, опираясь на идеи метода возможных направлений, можно получить симплекс-метод непосредственно для основной задачи линейного программирования (без ее сведения к канонической задаче).

Выше во вспомогательной задаче (2) было принято условие нормировки $|e^j| \leqslant 1$, $j = 1, \ldots, n$. Возможны и другие условия нормировки, например, $|e|^2 \leqslant 1$ или $|B_k e| \leqslant 1$, где B_k — специально выбираемая матрица. Заметим, что при такой нормировке задача (2) уже не будет задачей

линейного программирования. Тем не менее удачный выбор B_k может облегчить выбор возможного направления убывания, ускорить сходимость метода. О других способах нормировки, о сходимости различных вариантов метода возможных направлений и других аспектах этого метода см., например, [319; 326; 374; 774].

Упражнения

1. Сделать несколько итераций метода возможных направлений для задачи минимизации $f(u)=x+y$ на множестве $X=\{u=(x, y)\colon g_1(u)=x^2-y\leqslant 0,\ g_2(u)=y-1\leqslant 0\}$ при различном выборе начальной точки u_0.

2. Вычислить несколько приближений по методу возможных направлений для задачи из примера 1 при различном начальном приближении u_0.

§ 6. Проксимальный метод

1. Этот метод используется для решения выпуклых задач минимизации

$$f(x)\to\inf,\quad x\in X, \tag{1}$$

когда X — выпуклое замкнутое множество из E^n, функция $f(x)$ выпукла на X. В его основе лежит понятие проксимального оператора, который определяется следующим образом. Фиксируются точка $x\in E^n$ и число $\alpha>0$, определяется функция

$$\varphi(z, x, \alpha)=\tfrac{1}{2}|z-x|^2+\alpha f(z) \tag{2}$$

переменной $z\in X$, и рассматривается задача минимизации

$$\varphi(z, x, \alpha)\to\inf,\quad z\in X. \tag{3}$$

Так как функция $\varphi(z, x, \alpha)$ сильно выпукла на X с постоянной сильной выпуклости $\varkappa=1$, то согласно теореме 4.3.1 задача (3) имеет, притом единственное решение $z_*=z_*(x, \alpha)$. Тем самым определен оператор, который каждой точке $x\in E^n$ и числу $\alpha>0$ ставит в соответствие решение z_* задачи (3). Этот оператор называется *проксимальным*, его обозначают символом pr. Таким образом, $\operatorname{pr}(x, \alpha)=z_*\in X$ — решение задачи (3). Изучим некоторые свойства проксимального оператора. Из (3) и неравенства (4.3.3) следует, что

$$\tfrac{1}{2}|z-\operatorname{pr}(x, \alpha)|^2\leqslant\varphi(z, x, \alpha)-\varphi(\operatorname{pr}(x, \alpha), x, \alpha)\quad\forall z\in X,\ x\in E^n,\ \alpha>0. \tag{4}$$

Если функция $f(x)$ дифференцируема на X, то $\varphi_z(z, x, \alpha)=z-x+\alpha f'(z)$ и для решения $\operatorname{pr}(x, \alpha)$ задачи (3) по теореме 4.2.3 имеем

$$\langle\operatorname{pr}(x, \alpha)-x+\alpha f'(\operatorname{pr}(x, \alpha)), z-\operatorname{pr}(x, \alpha)\rangle\geqslant 0\quad\forall z\in X.$$

Отсюда, вспоминая характеристическое свойство проекции точки на множество (неравенство (4.4.1)), обнаруживаем следующую связь между проксимальным оператором и оператором проектирования:

$$\operatorname{pr}(x, \alpha)=\mathcal{P}_X(x-\alpha f'(\operatorname{pr}(x, \alpha)))\quad\forall x\in E^n,\quad\alpha>0. \tag{5}$$

Если функция $f(x)$ не является дифференцируемой, то эту связь можно выразить с помощью субградиентов. Всюду ниже в этом параграфе мы будем предполагать, что функция $f(x)$ определена и выпукла на открытом выпуклом множестве W, содержащем множество X. Тогда при каждом $z\in X$ субдифференциал $\partial f(z)$ является непустым выпуклым замкнутым ограниченным множеством (теорема 4.6.2). По правилу 7 субдифференцирования из § 4.6 для функции (2) имеем

$$\partial\varphi(z, x, \alpha)=z-x+\alpha\partial f(z)\quad\forall z\in X. \tag{6}$$

Согласно теореме 4.6.4 для решения $\operatorname{pr}(x, \alpha)$ задачи (3) найдется субградиент $c_1(\operatorname{pr}(x, \alpha))\in\partial\varphi(\operatorname{pr}(x, \alpha), x, \alpha)$, такой, что

$$\langle c_1(\operatorname{pr}(x, \alpha)), z-\operatorname{pr}(x, \alpha)\rangle\geqslant 0\quad\forall z\in X.$$

Из формулы (6) следует существование $c(\operatorname{pr}(x, \alpha))\in\partial f(\operatorname{pr}(x, \alpha))$, для которого $c_1(\operatorname{pr}(x, \alpha))=\operatorname{pr}(x, \alpha)-x+\alpha c(\operatorname{pr}(x, \alpha))$, и поэтому предыдущее неравенство можно записать в виде

$$\langle\operatorname{pr}(x, \alpha)-x+\alpha c(\operatorname{pr}(x, \alpha)), z-\operatorname{pr}(x, \alpha)\rangle\geqslant 0\quad\forall z\in X. \tag{7}$$

Отсюда вместо формулы (5) получим

$$\operatorname{pr}(x, \alpha)=\mathcal{P}_X(x-\alpha c(\operatorname{pr}(x, \alpha)))\quad\forall x\in E^n,\quad\alpha>0, \tag{8}$$

для некоторого субградиента $c(\operatorname{pr}(x, \alpha))\in\partial f(\operatorname{pr}(x, \alpha))$.

С помощью установленной связи (8) между проксимальным оператором и оператором проектирования нетрудно доказать следующий критерий оптимальности для задачи (1).

Теорема 1. *Для того, чтобы $x_*\in X_*$, необходимо и достаточно, чтобы*

$$x_*=\operatorname{pr}(x_*, \alpha)\quad\forall\alpha>0.$$

Доказательство. Необходимость. Пусть $x_*\in X_*$. По теореме 4.6.4 найдется субградиент $c(x_*)\in\partial f(x_*)$ такой, что $\langle c(x_*), z-x_*\rangle\geqslant 0$ $\forall z\in X$. Тогда, как следует из формулы (6) $c_1(x_*)=x_*-x_*+\alpha c(x_*)=\alpha c(x_*)\in\partial\varphi(x_*, x_*, \alpha)$. Умножая на $\alpha>0$ предыдущее вариационное неравенство, получаем: $\langle c_1(x_*), z-x_*\rangle\geqslant 0$ $\forall z\in X$. Согласно теореме 4.6.4 это означает, что x_* — решение задачи (3) при $x=x_*$, т. е. $x_*=\operatorname{pr}(x_*, \alpha)$.

Достаточность. Пусть $x_*=\operatorname{pr}(x_*, \alpha)$. Согласно формуле (8) найдется $c(\operatorname{pr}(x, \alpha))=c(x_*)\in\partial f(\operatorname{pr}(x_*, \alpha))=\partial f(x_*)$, что $x_*=\mathcal{P}_X(x_*-\alpha c(x_*))$. Отсюда и из замечания 1 к теореме 4.6.4 имеем: $x_*\in X_*$. Теорема 1 доказана. □

Теорема 2. *Проксимальный оператор $\operatorname{pr}(x, \alpha)$ непрерывен по переменной x равномерно относительно $\alpha>0$ и, более того,*

$$|\operatorname{pr}(x, \alpha)-\operatorname{pr}(y, \alpha)|\leqslant|x-y|\quad\forall x, y\in E^n,\quad\alpha>0. \tag{9}$$

Доказательство. Из неравенства (4) при $z=\operatorname{pr}(y, \alpha)$ имеем: $\frac{1}{2}|\operatorname{pr}(y, \alpha)-\operatorname{pr}(x, \alpha)|^2\leqslant\varphi(\operatorname{pr}(y, \alpha), x, \alpha)-\varphi(\operatorname{pr}(x, \alpha), x, \alpha)$. Меняя здесь x и y ролями, получим $\frac{1}{2}|\operatorname{pr}(x, \alpha)-\operatorname{pr}(y, \alpha)|^2\leqslant\varphi(\operatorname{pr}(x, \alpha), y, \alpha)-\varphi(\operatorname{pr}(y, \alpha), y, \alpha)$. Сложим эти два неравенства: $|\operatorname{pr}(x, \alpha)-\operatorname{pr}(y, \alpha)|^2\leqslant\varphi(\operatorname{pr}(y, \alpha), x, \alpha)-\varphi(\operatorname{pr}(x, \alpha), x, \alpha)+\varphi(\operatorname{pr}(x, \alpha), y, \alpha)-\varphi(\operatorname{pr}(y, \alpha), y, \alpha)$.

В правую часть этого неравенства подставим соответствующие значения функции $\varphi(z, x, \alpha)$ согласно ее определению (2). После простых преобразований получим

$$|\mathrm{pr}\,(x,\alpha)-\mathrm{pr}\,(y,\alpha)|^2\leqslant\langle\mathrm{pr}\,(x,\alpha)-\mathrm{pr}\,(y,\alpha),x-y\rangle\quad\forall x,\,y\in E^n,\quad\alpha>0. \tag{10}$$

К правой части неравенства (10) применим неравенство Коши — Буняковского: $|\mathrm{pr}\,(x,\alpha)-\mathrm{pr}\,(y,\alpha)|^2\leqslant|\mathrm{pr}\,(x,\alpha)-\mathrm{pr}\,(y,\alpha)|\cdot|x-y|$. Отсюда следует неравенство (9). □

В следующей теореме приводятся условия, обеспечивающие непрерывность $\mathrm{pr}\,(x,\alpha)$ по совокупности аргументов (x,α) во всех точках $x\in E^n$, $\alpha>0$.

Теорема 3. *Пусть в дополнение к сделанным выше предположениям функция $f(x)$ удовлетворяет условию Липшица: $|f(x)-f(y)|\leqslant L|x-y|$ $\forall x,y\in X$, пусть $x_k\in E^n$, $\alpha_k>0$, $k=1,2,\ldots$, $\{x_k\}\to x$, $\{\alpha_k\}\to\alpha>0$. Тогда $\lim\limits_{k\to\infty}\mathrm{pr}\,(x_k,\alpha_k)=\mathrm{pr}\,(x,\alpha)$.*

Доказательство. По неравенству треугольника имеем:

$$|\mathrm{pr}\,(x_k,\alpha_k)-\mathrm{pr}\,(x,\alpha)|\leqslant|\mathrm{pr}\,(x_k,\alpha_k)-\mathrm{pr}\,(x,\alpha_k)|+|\mathrm{pr}\,(x,\alpha_k)-\mathrm{pr}\,(x,\alpha)|,\quad k=1,2,\ldots \tag{11}$$

Первое слагаемое в правой части неравенства (11) в силу (9): $|\mathrm{pr}\,(x_k,\alpha_k)-\mathrm{pr}\,(x,\alpha_k)|\leqslant|x_k-x|\to0$ при $k\to\infty$. Докажем, что второе слагаемое также стремится к нулю. Из неравенства (4) при $\alpha=\alpha_k$, $z=\mathrm{pr}\,(x,\alpha)$ следует

$$\tfrac12|\mathrm{pr}\,(x,\alpha)-\mathrm{pr}\,(x,\alpha_k)|^2\leqslant\tfrac12|\mathrm{pr}\,(x,\alpha)-x|^2-\tfrac12|\mathrm{pr}\,(x,\alpha_k)-x|^2+\alpha_k(f(\mathrm{pr}\,(x,\alpha))-f(\mathrm{pr}\,(x,\alpha_k))),\quad k=1,2,\ldots$$

По условию функция $f(x)$ удовлетворяет условию Липшица, поэтому $|f(\mathrm{pr}\,(x,\alpha))-f(\mathrm{pr}\,(x,\alpha_k))|\leqslant L|\mathrm{pr}\,(x,\alpha)-\mathrm{pr}\,(x,\alpha_k)|$. Отсюда и из предыдущего неравенства для величины $a_k=|\mathrm{pr}\,(x,\alpha)-\mathrm{pr}\,(x,\alpha_k)|\geqslant0$ имеем: $a_k^2-2\alpha_kLa_k-|\mathrm{pr}\,(x,\alpha)-x|^2\leqslant0$, $k=1,2,\ldots$. Замечая, что левая часть этого неравенства представляет собой квадратный трехчлен относительно переменной a_k, получаем оценку для a_k: $0\leqslant a_k\leqslant\alpha_kL+\sqrt{|\mathrm{pr}\,(x,\alpha)-x|^2+\alpha_k^2L^2}\leqslant2L\sup\limits_{k\geqslant0}\alpha_k+|\mathrm{pr}\,(x,\alpha)-x|$. Тогда последовательность $\{\mathrm{pr}\,(x,\alpha_k)\}$ также ограничена и согласно теореме Больцано — Вейерштрасса имеет хотя бы одну предельную точку w. Выбирая при необходимости подпоследовательность, можем считать, что $\{\mathrm{pr}\,(x,\alpha_k)\}\to w$. Тогда множество Y, состоящее из точек $\mathrm{pr}\,(x,\alpha_k)$, $k=1,2,\ldots$ и точки w компактно. В силу компактности субдифференциального отображения (теорема 4.6.5) тогда компактно и множество $\bigcup\limits_{y\in Y}\partial f(y)$, которое содержит субградиенты $c(\mathrm{pr}\,(x,\alpha_k))$, входящие в неравенства (7) при $\alpha=\alpha_k$, $k=1,2,\ldots$. Поэтому можем считать, что $\{c(\mathrm{pr}\,(x,\alpha_k))\}\to c$. Из замкнутости субдифференциального отображения (теорема 4.6.5) следует, что $c\in\partial f(w)$. Теперь мы можем совершить предельный переход при $k\to\infty$ в вариационных неравенствах $\langle\mathrm{pr}\,(x,\alpha_k)-x+\alpha_kc(\mathrm{pr}\,(x,\alpha_k)),z-\mathrm{pr}\,(x,\alpha_k)\rangle\geqslant0$ $\forall z\in X$, полученных из (7) при $\alpha=\alpha_k$. Будем иметь $\langle w-x+\alpha c,z-w\rangle\geqslant0$ $\forall z\in X$. Здесь $c\in\partial f(w)$ и согласно формуле (6) имеем $c_1=w-x+\alpha c\in\partial\varphi(w,x,\alpha)$. Отсюда и из предыдущего неравенства, записанного в виде $\langle c_1,z-w\rangle\geqslant0$ $\forall z\in X$ с помощью

теоремы 4.6.4 получим, что w решение задачи (3), т. е. $w=\mathrm{pr}\,(x,\alpha)$. Это значит, что последовательность $\{\mathrm{pr}\,(x,\alpha_k)\}$ имеет единственную предельную точку $\mathrm{pr}\,(x,\alpha)$, т. е. $\lim\limits_{k\to\infty}\mathrm{pr}\,(x,\alpha_k)=\mathrm{pr}\,(x,\alpha)$. Тем самым мы доказали, что второе слагаемое из правой части неравенства (11) также стремится к нулю. Непрерывность проксимального отображения $\mathrm{pr}\,(x,\alpha)$ по совокупности аргументов (x,α) установлена. □

Замечание 1. В силу теоремы 4.6.6 выраженное в теореме 3 требование условия Липшица от функции $f(x)$ можно заменить условием ограниченности множества X.

Проксимальный метод заключается в построении последовательности $\{x_k\}$ по следующему правилу:

$$x_{k+1}=\mathrm{pr}\,(x_k,\alpha_k),\quad k=0,1,\ldots, \tag{12}$$

где начальная точка x_0 и последовательность $\{\alpha_k\}>0$ предполагаются заданными. Итерационный процесс (12) основан на известном из анализа [393] методе поиска неподвижных точек сжимающих операторов и идейно близок к рассмотренным выше методам проекции градиента и субградиента. В соответствии с определением проксимального оператора на k-м шаге процесса (12) для определения очередного приближения x_{k+1} нужно решить задачу минимизации (3) при $x=x_k$, $\alpha=\alpha_k$:

$$\varphi(z,x_k,\alpha_k)=\tfrac12|z-x_k|^2+\alpha_kf(z)\to\inf,\quad z\in X,\quad k=0,1,\ldots \tag{13}$$

Если функция $f(x)$ дифференцируема на X, то согласно теореме 4.2.3 точка x_{k+1} будет решением задачи (13) тогда и только тогда, когда

$$\langle\varphi_z'(x_{k+1},x_k,\alpha_k),z-x_{k+1}\rangle=\langle x_{k+1}-x_k+\alpha_kf'(x_{k+1}),z-x_{k+1}\rangle\geqslant0\quad\forall z\in X.$$

Отсюда, используя характеристическое свойство проекции (неравенство (4.4.1)), получаем, что

$$x_{k+1}=\mathcal{P}_X(x_k-\alpha_kf'(x_{k+1})),\quad k=0,1,\ldots \tag{14}$$

Таким образом, для дифференцируемых функций проксимальный метод (12) равносилен методу проекции градиента в так называемой неявной форме, когда x_{k+1} явно не выражено через предыдущее приближение x_k и должно определяться как решение уравнения (14).

Теорема 4. *Пусть X — выпуклое замкнутое множество из E^n, функция $f(x)$ определена и выпукла на открытом выпуклом множестве W, содержащем множество X, и полунепрерывна снизу на X, $f_*=\inf\limits_{x\in X}f(x)>-\infty$, $X_*=\{x\in X\colon f(x)=f_*\}\neq\varnothing$, пусть $0<\gamma_0\leqslant\alpha_k\leqslant\gamma_1$, $k=0,1,\ldots$ Тогда последовательность $\{x_k\}$, определяемая методом (12) при любом начальном приближении $x_0\in E^n$, сходится к некоторой точке $v_*=v_*(x_0)\in X_*$.*

Доказательство. Применяя неравенство (4.3.3) к задаче (13) имеем

$$\tfrac12|z-x_{k+1}|^2\leqslant\varphi(z,x_k,\alpha_k)-\varphi(x_{k+1},x_k,\alpha_k)=\tfrac12|z-x_k|^2-\tfrac12|x_{k+1}-x_k|^2+\alpha_k(f(z)-f(x_{k+1}))\quad\forall z\in X,\quad k=0,1,\ldots \tag{15}$$

Возьмем произвольную точку $x_* \in X_*$ и в (15) положим $z = x_*$. С учетом неравенства $f(x_*) - f(x_{k+1}) \leqslant 0$ получим

$$|x_* - x_{k+1}|^2 \leqslant |x_* - x_k|^2 - |x_{k+1} - x_k|^2 \leqslant |x_* - x_k|^2, \quad k = 0, 1, \ldots, \quad \forall x_* \in X_*. \tag{16}$$

Отсюда видно, что последовательность $\{|x_k - x_*|^2\}$ не возрастает, а $\{x_k\}$ ограничена. Тогда существует подпоследовательность $\{x_{k_r}\}$, сходящаяся к некоторой точке v_*, причем $v_* \in X$ в силу замкнутости X. Можем считать, что $\{\alpha_{k_r}\} \to \alpha_\infty$, $0 < \gamma_0 \leqslant \alpha_\infty \leqslant \gamma_1$. Далее, суммируя неравенство (16), имеем

$$\sum_{k=0}^{N} |x_{k+1} - x_k|^2 \leqslant |x_* - x_0|^2 - |x_* - x_{N+1}|^2 \leqslant |x_* - x_0|^2 \quad \forall N = 1, 2, \ldots$$

Это значит, что ряд $\sum\limits_{k=0}^{\infty} |x_{k+1} - x_k|^2$ сходится и, следовательно, $\lim\limits_{k \to \infty} |x_{k+1} - x_k| = 0$. Тогда подпоследовательность $\{x_{k_r+1}\}$ также сходится к v_*. Переходя в (15) к пределу при $k = k_r \to \infty$ с учетом неравенства $\underline{\lim} f(x_{k+1}) \geqslant f(v_*)$, получим: $0 \leqslant \alpha_\infty (f(z) - f(v_*))$ $\forall z \in X$. Это значит, что $v_* \in X_*$. Поскольку последовательность $\{|x_k - x_*|^2\}$ не возрастает при $\forall x_* \in X_*$, то в частности при $x_* = v_*$ мы также получим невозрастающую последовательность $\{|x_k - v_*|^2\}$. Тогда $\lim\limits_{k \to \infty} |x_k - v_*|^2 = \lim\limits_{r \to \infty} |x_{k_r} - v_*|^2 = 0$. Теорема 4 доказана. □

Для сильно выпуклых функций $f(x)$ несложно получить оценку скорости сходимости метода (12) при $\alpha_k = \alpha > 0$, $k = 0, 1, \ldots$

Теорема 5. *Пусть выполнены условия теоремы* 4 *и функция* $f(x)$ *сильно выпукла на множестве* X *с постоянной сильной выпуклости* $\varkappa > 0$. *Пусть последовательность* $\{x_k\}$ *определена методом* (12) *при* $\alpha_k = \alpha > 0$, $k = 0, 1, \ldots$ *Тогда*

$$|x_k - x_*|^2 \leqslant |x_0 - x_*|^2 q^k, \quad k = 0, 1, \ldots; \quad q = \frac{1}{1 + 2\alpha\varkappa}. \tag{17}$$

Доказательство. Из теоремы 4.3.1 следует, что при сделанных предположениях задача (1) имеет, притом единственное, решение x_*, и

$$\frac{\varkappa}{2} |z - x_*|^2 \leqslant f(z) - f(x_*) \quad \forall z \in X. \tag{18}$$

Кроме того, функция $\varphi(z, x, \alpha)$ из (2) сильно выпукла с постоянной сильной выпуклости $1 + \alpha\varkappa$ и из (13) имеем

$$\frac{1 + \alpha\varkappa}{2} |z - x_{k+1}|^2 \leqslant \varphi(z, x_k, \alpha) - \varphi(x_{k+1}, x_k, \alpha), \quad \forall z \in X, \; k = 0, 1, \ldots \tag{19}$$

В неравенстве (18) положим $z = x_{k+1}$ и умножим его на $\alpha > 0$ и сложим с (19) при $z = x_*$. Получим

$$\left(\frac{1}{2} + \alpha\varkappa\right) |x_* - x_{k+1}|^2 \leqslant \frac{1}{2} |x_* - x_k|^2 - \frac{1}{2} |x_{k+1} - x_k|^2 \leqslant \frac{1}{2} |x_* - x_k|^2, \quad k = 0, 1, \ldots$$

Отсюда следует, что $|x_{k+1} - x_*|^2 \leqslant |x_k - x_*|^2 q \leqslant |x_{k-1} - x_*|^2 q^2 \leqslant \ldots \leqslant |x_0 - x_*|^2 q^{k+1}$, $k = 0, 1, \ldots$ Оценка (17) доказана. □

Отметим, что если функция $f(x)$ дифференцируема на X, то теоремы 4, 5 сохраняют силу и для метода (14).

2. Опишем непрерывный вариант проксимального метода, следуя [25]. Рассмотрим задачу

$$\varphi(z, x, \alpha(t)) = \frac{1}{2} |z - x|^2 + \alpha(t) f(z) \to \inf, \quad z \in X, \tag{20}$$

где $\alpha(t) > 0$ $\forall t \geqslant 0$, X — выпуклое замкнутое множество, $f(x)$ выпукла на X. Тогда $\varphi(z, x, \alpha(t))$ сильно выпукла с постоянной сильной выпуклости $\varkappa = 1$ и по теореме 4.3.1 задача (20) имеет, притом единственное, решение $z = \operatorname{pr}(x, \alpha(t))$. Введем систему дифференциальных уравнений

$$\dot{x}(t) = \operatorname{pr}(x(t), \alpha(t)) - x(t), \quad t \geqslant 0. \tag{21}$$

Согласно теореме 1 решение x_* задачи (1) удовлетворяет уравнению $\operatorname{pr}(x_*, \alpha(t)) - x_* = 0$ при $\forall t \geqslant 0$. Это значит, что каждая точка $x_* \in X_*$ является точкой равновесия (стационарным решением) системы (21). Можно ожидать, что при некоторых требованиях на функции $f(x)$, $\alpha(t)$ траектории $x(t)$ системы (21) при больших t приближаются ко множеству X_*. Справедлива

Теорема 6 (Антипин [25]). *Пусть множество* X *и функция* $f(x)$ *удовлетворяют условиям теорем* 3, 4, *функция* $\alpha(t)$ *непрерывна и* $0 \leqslant \gamma_0 \leqslant \alpha(t) \leqslant \gamma_1$ $\forall t \geqslant 0$. *Тогда траектория* $x(t)$ *системы* (21), *выходящая из любой точки* $x(0) = x_0$, *определена при всех* $t \geqslant 0$ *и сходится при* $t \to +\infty$ *к некоторой точке* $v_* = v_*(x_0) \in X_*$; *кроме того,* $\lim\limits_{t \to \infty} \dot{x}(t) = 0$.

Доказательство. В силу теорем 2, 3 правая часть $\operatorname{pr}(x, \alpha(t)) - x$ уравнения (21) удовлетворяет условию Липшица по x и непрерывна по t, потому все решения системы (21) определены при всех $t \geqslant 0$ (см. ниже теорему 6.1.1). В силу уравнения (21) имеем $\operatorname{pr}(x(t), \alpha(t)) = \dot{x}(t) + x(t) \in X$ и по теореме 6.3.1

$$\frac{1}{2} |z - \operatorname{pr}(x(t), \alpha(t))|^2 = \frac{1}{2} |z - (\dot{x}(t) + x(t))|^2 \leqslant \varphi(z, x(t), \alpha(t)) -$$

$$- \varphi(\operatorname{pr}(x(t), \alpha(t)), x(t), \alpha(t)) = \frac{1}{2} |z - x(t)|^2 - \frac{1}{2} |(\dot{x}(t) + x(t)) - x(t)|^2 +$$

$$+ \alpha(t)(f(z) - f(\dot{x}(t) + x(t))) \qquad \forall z \in X, \quad t \geqslant 0. \tag{22}$$

Полагая в (22) $z = x_* \in X_*$, с учетом неравенств $f(x_*) - f(\dot{x}(t) + x(t)) \leqslant 0$, $\alpha(t) > 0$ получим:

$$|\dot{x}(t) + x(t) - x_*|^2 = |\dot{x}(t)|^2 + |x(t) - x_*|^2 + 2\langle \dot{x}(t), x(t) - x_* \rangle \leqslant |x(t) - x_*|^2 - |\dot{x}(t)|^2$$

или

$$|\dot{x}(t)|^2 + \frac{1}{2} \frac{d}{dt} |x(t) - x_*|^2 \leqslant 0 \quad \forall t \geqslant 0, \quad \forall x_* \in X_*. \tag{23}$$

Интегрируем это неравенство на отрезке $[\tau, t]$

$$2 \int_\tau^t |\dot{x}(s)|^2 ds + |x(t) - x_*|^2 \leqslant |x(\tau) - x_*|^2 \quad \forall t > \tau \geqslant 0, \quad \forall x_* \in X_*.$$

Это означает, что функция $|x(t) - x_*|^2$ не возрастает при всех $x_* \in X_*$. В частности, при $\tau = 0$ отсюда имеем: $|x(t) - x_*|^2 \leqslant |x_0 - x_*|^2$, так что траектория $\{x(t), t \geqslant 0\}$ ограничена и, кроме того, $\int\limits_0^\infty |\dot{x}(t)|^2 dt < \infty$. Отсюда следует существование последовательности $\{t_i\} \to +\infty$ такой, что $\{\dot{x}(t_i)\} \to 0$, $\{x(t_i)\} \to v_*$, $\{\alpha(t_i)\} \to \alpha_\infty$, $0 < \gamma_0 \leqslant \alpha_\infty \leqslant \gamma_1$. Так как X — замкнутое множество, $\dot{x}(t) + x(t) \in X$, то $\dot{x}(t_i) + x(t_i) \to v_* \in X$. Далее в (22) перейдем к пределу при $t = t_i \to \infty$. Учитывая, что $\lim\limits_{i \to \infty} f(\dot{x}(t_i) + x(t_i)) \geqslant f(v_*)$, получим $0 \leqslant \alpha_\infty (f(z) - f(v_*))$ $\forall z \in X$. Это значит, что $v_* \in X_*$. Тогда функция $|x(t) - v_*|^2$ не возрастает. Потому $\lim\limits_{t \to \infty} |x(t) - v_*|^2 = \lim\limits_{i \to \infty} |x(t_i) - v_*|^2 = 0$, т. е. $\lim\limits_{t \to \infty} x(t) = v_*$. Наконец, из неравенства (23) при $x_* = v_*$ имеем $|\dot{x}(t)|^2 \leqslant -\langle \dot{x}(t), x(t) - v_* \rangle \leqslant |\dot{x}(t)| \cdot |x(t) - v_*|$ или $|\dot{x}(t)| \leqslant |x(t) - v_*|$ $\forall t \geqslant 0$. Отсюда следует, что $\lim\limits_{t \to \infty} \dot{x}(t) = 0$. Теорема 6 доказана. □

Для сильно выпуклых функций $f(x)$ можно получить следующую оценку скорости сходимости метода (21).

Теорема 7. *Пусть множество* X *и функция* $f(x)$ *удовлетворяют условиям теорем* 3, 4, *функция* $f(x)$ *сильно выпукла с постоянной сильной выпуклости* $\varkappa > 0$, *функция* $\alpha(t) > 0$ *непрерывна и* $\int\limits_0^\infty \frac{\alpha(\tau)}{1 + 2\alpha(\tau)\varkappa} d\tau = +\infty$, *пусть* $x(t)$, $t \geqslant 0$, — *любая траектория системы* (21). *Тогда*

$$|x(t) - x_*| \leqslant |x(0) - x_*| \exp\left(-\int_0^t \frac{2\alpha(\tau)\varkappa}{1 + 2\alpha(\tau)\varkappa} d\tau\right) \quad \forall t \geqslant 0. \tag{24}$$

Доказательство. Из теоремы 4.3.1 следует, что задача (1) имеет, притом единственное, решение x_* и

$$\frac{\varkappa}{2} |z - x_*|^2 \leqslant f(z) - f(x_*) \qquad \forall z \in X. \tag{25}$$

Функция $\varphi(z, x, \alpha(t))$ сильно выпукла на X с постоянной сильной выпуклости $1 + \alpha(t)\varkappa$ и из (20), (21) имеем

$$\frac{1}{2}(1 + \alpha(t)\varkappa) |z - (\dot{x}(t) + x(t))|^2 \leqslant \varphi(z, x(t), \alpha(t)) - \varphi(\dot{x}(t) + x(t), x(t), \alpha(t)), \quad t \geqslant 0. \tag{26}$$

В (25) положим $z = \dot{x}(t) + x(t)$, умножим на $\alpha(t) > 0$ и сложим с (26) при $x = x_*$. Получим

$$\tfrac{1}{2}(1+2\alpha(t)\varkappa)(|\dot{x}(t)|^2+|x(t)-x_*|^2+2\langle\dot{x}(t),x(t)-x_*\rangle) \leqslant \tfrac{1}{2}|x(t)-x_*|^2-\tfrac{1}{2}|\dot{x}(t)|^2, \quad t \geqslant 0$$

или

$$\frac{d}{dt}|x(t)-x_*|^2 + \frac{2\alpha(t)\varkappa}{1+2\alpha(t)\varkappa}|x(t)-x_*|^2 \leqslant 0, \quad t \geqslant 0,$$

или

$$\frac{d}{dt}\left(|x(t)-x_*|\exp\left(\int_0^t \frac{2\alpha(\tau)\varkappa}{1+2\alpha(\tau)\varkappa}d\tau\right)\right) \leqslant 0, \quad t \geqslant 0.$$

Интегрируя это неравенство на отрезке $[0, t]$, придем к оценке (24). □

Замечание 2. При $\alpha(t) \equiv \alpha = \text{const} > 0$ требование условия Липшица от функции $f(x)$ в теоремах 6, 7 излишне. Это требование нужно было в теореме 3 при доказательстве непрерывности $\text{pr}(x, \alpha)$ по совокупности (x, α), что в свою очередь обеспечивало непрерывность правой части системы (21) по t и продолжимость траекторий $x(t)$ на всю полуось $t \geqslant 0$. Однако при $\alpha(t) \equiv \alpha$ правая часть (21) от t не зависит и для продолжимости траектории достаточно условия (9).

Замечание 3. При доказательстве теорем 4–7 для нас не было существенно, каким методом решаются вспомогательные задачи (3), (13), (20) для определения значений проксимального оператора в требуемых точках (x, α). Однако ясно, что проксимальный метод имеет смысл применять лишь тогда, когда имеется удобный быстро сходящийся метод для решения упомянутых вспомогательных задач. В этих задачах, в отличие от исходной задачи (1), минимизируемая функция сильно выпукла благодаря слагаемому $\frac{1}{2}|z-x|^2$, что обеспечивает их однозначную разрешимость и, можно надеяться, улучшает сходимость используемых методов их решения, повышает устойчивость этих методов к погрешностям вычислений. Различные варианты проксимального метода, вычислительные аспекты этого метода исследованы, например, в [25; 26; 30; 799; 803; 813].

В следующем параграфе рассматривается метод, который можно истолковать как некоторое развитие проксимального метода.

Упражнения

1. Реализовать один шаг проксимального метода для задач из упражнений 4.6.1, 4.6.3, 5.2, 5.3 при различных начальных приближениях.

2. Опираясь на теоремы 1–3, доказать, что проксимальный оператор является монотонным, замкнутым, компактным отображением.

§ 7. Метод линеаризации

Этот метод на каждой итерации использует линейные аппроксимации минимизируемой функции и функций, задающих ограничения. Опишем его для задачи

$$f(x) \to \inf, \quad x \in X = \{x \in X_0: g_1(x) \leqslant 0, \ldots, g_m(x) \leqslant 0\}, \tag{1}$$

предполагая, что X_0 — выпуклое замкнутое множество из E^n и функции $f(x)$, $g_i(x) \in C^1(X_0)$. Пусть x_0 — начальное приближение, $x_0 \in X_0$. Предположим, что k-е приближение $x_k \in X_0$ при некотором $k \geqslant 0$ уже известно. Введем функцию

$$\Phi_k(x) = \tfrac{1}{2}|x-x_k|^2 + \alpha_k\langle f'(x_k), x-x_k\rangle, \quad \alpha_k > 0, \tag{2}$$

и множество

$$W_k = \{x \in X_0: g_i(x_k) + \langle g_i'(x_k), x-x_k\rangle \leqslant 0,\ i = 1, \ldots, m\}. \tag{3}$$

Пусть $W_k \neq \varnothing$. В качестве $k+1$-го приближения x_{k+1} возьмем решение следующей задачи минимизации:

$$\Phi_k(x) \to \inf, \quad x \in W_k. \tag{4}$$

Поскольку функция (2) сильно выпукла, множество (3) выпукло и замкнуто, то согласно теореме 4.3.1 задача (4) имеет, притом единственное, решение. Задачу (4) необязательно решать точно: достаточно найти точку x_{k+1} из условий

$$x_{k+1} \in W_k: \Phi_k(x_{k+1}) \leqslant \inf_{W_k}\Phi_k(x) + \varepsilon_k, \quad \varepsilon_k \geqslant 0. \tag{5}$$

Если X_0 многогранное множество, то задача (4) представляет собой задачу квадратичного программирования и может быть решена конечношаговым методом (см. ниже § 7). Если W_k — ограниченное множество, то для решения задачи (4) может быть использован, например, метод условного градиента, который будет сходиться и при $\varepsilon_k > 0$ позволит определить точку x_{k+1} из (5) за конечное число шагов. В общем случае задача (5), конечно, не всегда просто решается. Метод линеаризации (5) обычно используют лишь в тех случаях, когда определение точки x_{k+1} из (5) не требует большого объема вычислений. Полезно заметить, что задача (4) равносильна задаче

$$\varphi_k(x) = \tfrac{1}{2}|x-(x_k-\alpha_k f'(x_k))|^2 \to \inf, \quad x \in W_k,$$

так как $\varphi_k(x) - \Phi_k(x) = \alpha_k^2|f'(x_k)|^2 = \text{const}$, $x \in E^n$. Это значит, что точное решение v_k задачи (4) представляет собой проекцию точки $x_k - \alpha_k f'(x_k)$ на множество W_k, а точка x_{k+1} из (5) является приближением для v_k. Отсюда следует, что если в (1) ограничения $g_i(x) \leqslant 0$ отсутствуют $(m = 0)$, то $X = X_0 = W_k$ и метод линеаризации превратится в метод проекции градиента.

Теорема 1. *Пусть X_0 — выпуклое замкнутое множество из E^n, $\text{int}\, X_0 \neq \varnothing$ (в частности, возможно $X_0 = E^n$); функции $f(x)$, $g_i(x) \in C^1(X_0)$, выпуклы на X_0 и*

$$\max\{|f'(u)-f'(v)|; \max_{1 \leqslant i \leqslant m}|g_i'(u)-g_i'(v)|\} \leqslant L|u-v| \quad \forall u, v \in X_0;$$

выполнено условие Слейтера, т. е. существует такая точка $\bar{x} \in X$, что

$$g_1(\bar{x}) < 0, \ldots, g_m(\bar{x}) < 0; \tag{6}$$

$f_* > -\infty$, $X_* \neq \varnothing$; числа α_k, ε_k в (2), (5) таковы, что

$$\varepsilon_k \geqslant 0, \quad \sum_{k=0}^{\infty}\sqrt{\varepsilon_k} < \infty, \quad 0 < \gamma_0 \leqslant \alpha_k \leqslant \alpha, \tag{7}$$

где α определяется ниже формулой (22). Тогда множество (3) непусто при всех $k \geqslant 0$, последовательность $\{x_k\}$, определяемая методом (5), сходится к некоторой точке $v_* \in X_*$.

Доказательство. Согласно теореме 4.2.2

$$g_i(x_k) + \langle g_i'(x_k), x-x_k\rangle \leqslant g_i(x) \quad \forall x \in X_0. \tag{8}$$

Отсюда следует, что если $x \in X$, то $x \in W_k$, так что $X \subset W_k$. По условию $X \neq \varnothing$, поэтому $W_k \neq \varnothing$, $k = 0, 1, \ldots$. Таким образом, при каждом $k \geqslant 0$, $\varepsilon_k \geqslant 0$ существует точка x_{k+1}, удовлетворяющая условиям (5); например, можно взять $x_{k+1} = v_k$, где v_k — точное решение задачи (4). Применяя теорему 4.3.1 к задаче (4) с учетом (5) имеем $|x_{k+1}-v_k|^2/2 \leqslant \Phi_k(x_{k+1}) - \Phi_k(v_k) \leqslant \varepsilon_k$, так что

$$|x_{k+1}-v_k| \leqslant \sqrt{2\varepsilon_k}. \tag{9}$$

Возьмем произвольную точку $x_* \in X_*$. При сделанных предположениях по теореме 4.9.2 и лемме 4.9.2 найдутся такие числа $\lambda_1^* \geqslant 0, \ldots, \lambda_m^* \geqslant 0$, что

$$\langle f'(x_*) + \sum_{i=1}^{m} \lambda_i^* g_i'(x_*), x - x_* \rangle \geqslant 0 \qquad \forall x \in X_0, \tag{10}$$

$$\lambda_i^* g_i(x_*) = 0, \quad i = 1, \ldots, m, \quad x_* \in X_*. \tag{11}$$

Подчеркнем, что в силу замечания, сделанного после формулировки следствия 2 к теореме 4.9.5, числа $\lambda_1^*, \ldots, \lambda_m^*$ в (10), (11) могут быть выбраны одни и те же для всех $x_* \in X_*$.

Далее, из условия (6) и неравенств (8) следует

$$g_i(x_k) + \langle g_i'(x_*), \bar{x} - x_* \rangle \leqslant g_i(\bar{x}) < 0, \quad i = 1, \ldots, m. \tag{12}$$

Это значит, что множество (3) также удовлетворяет условию Слейтера, и к задаче (4) также применима теорема 4.9.2, из которой следует, что функция Лагранжа этой задачи $L_1(x, \xi) = \frac{1}{2}|x - x_k|^2 + \alpha_k \langle f'(x_k), x - x_k \rangle + \sum_{i=1}^{m} \xi_i (g_i(x_k) + \langle g_i'(x_k), x - x_k \rangle)$, $x \in X_0$, $\xi = (\xi_1, \ldots, \xi_m) \in \Lambda_0 = E_+^m$, имеет седловую точку (v_k, ξ^k), v_k — решение задачи (4), $\xi^k = (\xi_{1k}, \ldots, \xi_{mk}) \in E_+^m$. В силу леммы 4.9.2

$$\langle v_k - x_k + \alpha_k f'(x_k) + \sum_{i=1}^{m} \xi_{ik} g_i'(x_k), x - v_k \rangle \geqslant 0 \qquad \forall x \in X_0, \tag{13}$$

$$\xi_{ik} (g_i(x_k) + \langle g_i'(x_k), v_k - x_k \rangle) = 0, \quad i = 1, \ldots, m, \tag{14}$$

$$g_i(x_k) + \langle g_i'(x_k), v_k - x_k \rangle \leqslant 0, \quad i = 1, \ldots, m, \quad v_k \in X_0. \tag{15}$$

Возьмем в (10) $x = v_k$, умножим на $\alpha_k > 0$ и сложим с (13) при $x = x_*$. Получим

$$0 \leqslant \langle v_k - x_k, x_* - v_k \rangle + \alpha_k \langle f'(x_*) - f'(x_k), v_k - x_* \rangle + \alpha_k \sum_{i=1}^{m} \lambda_i^* \langle g_i'(x_*), v_k - x_* \rangle + \sum_{i=1}^{m} \xi_{ik} \langle g_i'(x_k), x_* - v_k \rangle. \tag{16}$$

Преобразуем и оценим каждое слагаемое в левой части (16). Для первого слагаемого имеем

$$\langle v_k - x_k, x_* - v_k \rangle = \frac{1}{2}|x_k - x_*|^2 - \frac{1}{2}|v_k - x_k|^2 - \frac{1}{2}|x_* - v_k|^2. \tag{17}$$

Пользуясь неравенством (4.2.20) при $u = x_k$, $w = v_k$, $v = x_*$, получаем оценку для второго слагаемого

$$\alpha_k \langle f'(x_*) - f'(x_k), v_k - x_* \rangle \leqslant \alpha_k L |x_k - v_k|^2 / 4. \tag{18}$$

Далее из леммы 2.6.1 при $f(x) = g_i(x)$, $x = v_k$, $y = x_k$ с учетом неравенств (15) имеем $g_i(v_k) \leqslant g_i(x_k) + \langle g_i'(x_k), v_k - x_k \rangle + L|x_k - v_k|^2/2 \leqslant L|x_k - v_k|^2/2$, $i = 1, \ldots, m$. Отсюда для третьего слагаемого из (16) с помощью равенств (11), неравенств (8) при $x = v_k$ и $\lambda_k^* \geqslant 0$ получаем

$$\alpha_k \sum_{i=1}^{m} \lambda_i^* \langle g_i'(x_*), v_k - x_* \rangle = \alpha_k \sum_{i=1}^{m} \lambda_i^* (g_i(x_*) + \langle g_i'(x_*), v_k - x_* \rangle) \leqslant \alpha_k \sum_{i=1}^{m} \lambda_i^* g_i(v_k) \leqslant \alpha_k |\lambda^*|_1 L |x_k - v_k|^2/2, \quad |\lambda^*|_1 = \sum_{i=1}^{m} |\lambda_i^*|. \tag{19}$$

Наконец, для четвертого слагаемого из (16) с учетом равенств (14), неравенств (8) при $x = x_*$, включений $x_* \in X$, $\xi^k \in E_+^m$ имеем

$$\sum_{i=1}^{m} \xi_{ik} \langle g_i'(x_k), x_* - v_k \rangle = \sum_{i=1}^{m} [\xi_{ik} (g_i(x_k) + \langle g_i'(x_k), x_* - x_k \rangle) - \xi_{ik} (g_i(x_k) + \langle g_i'(x_k), v_k - x_k \rangle)] \leqslant \sum_{i=1}^{m} \xi_{ik} g_i(x_*) \leqslant 0. \tag{20}$$

Сложим оценки (17)–(20); с помощью (16) получим

$$0 \leqslant \frac{1}{2}|x_k - x_*|^2 - \frac{1}{2}|v_k - x_*|^2 - \frac{1}{2}|v_k - x_k|^2 \left(1 - \frac{1}{2} L \alpha_k - L |\lambda^*|_1 \alpha_k\right). \tag{21}$$

Выберем α_k из условий

$$0 < \gamma_0 \leqslant \alpha_k \leqslant \frac{2(1 - \gamma)}{L(1 + 2|\lambda^*|_1)} = \alpha, \tag{22}$$

где γ_0, γ такие малые положительные числа, что $\gamma_0 \leqslant 2(1 - \gamma)/(L(1 + 2|\lambda^*|_1))$. Из (21), (22) тогда имеем

$$|v_k - x_*|^2 + \gamma |v_k - x_k|^2 \leqslant |x_k - x_*|^2, \quad k = 0, 1, \ldots \tag{23}$$

Из неравенств (7), (9), (23) и леммы 2.6.10 следует существование конечных пределов

$$\lim_{k \to \infty} |x_k - x_*| = \lim_{k \to \infty} |v_k - x_*|, \quad \lim_{k \to \infty} |x_k - v_k| = 0. \tag{24}$$

Это значит, что последовательность $\{x_k\}$, $\{v_k\}$ ограничены. Покажем ограниченность последовательности $\{\xi^k\}$ из (13)–(15). С помощью (12), (14) из (13) при $x = \bar{x}$ имеем

$$\langle v_k - x_k + \alpha_k f'(x_k), \bar{x} - v_k \rangle \geqslant -\sum_{i=1}^{m} \xi_{ik} \langle g_i'(x_k), \bar{x} - v_k \rangle = \sum_{i=1}^{m} [\xi_{ik} (-g_i(x_k) - \langle g_i'(x_k), \bar{x} - x_k \rangle) + \xi_{ik} (g_i(x_k) + \langle g_i'(x_k), v_k - x_k \rangle)] \geqslant \sum_{i=1}^{m} \xi_{ik} (-g_i(\bar{x})) \geqslant \xi_{jk} \min_{1 \leqslant i \leqslant m} |g_i(\bar{x})|, \quad j = 1, \ldots, m.$$

Отсюда и из неравенств (22), ограниченности $\{x_k\}$ и $\{v_k\}$ получаем

$$0 \leqslant \xi_{jk} \leqslant \frac{1}{\min\limits_{1 \leqslant i \leqslant m} |g_i(\bar{x})|} [\langle v_k - x_k + \alpha_k f'(x_k), \bar{x} - v_k \rangle] \leqslant \text{const} < \infty, \quad j = 1, \ldots, m.$$

Таким образом, последовательность $\{\xi^k\}$ ограничена. Отсюда и из (22) следует ограниченность $\{\xi^k/\alpha_k\}$. Перепишем (13), (14) в виде

$$\left\langle \frac{v_k - x_k}{\alpha_k} + f'(x_k) + \sum_{i=1}^{m} \frac{\xi_{ik}}{\alpha_k} g_i'(x_k), x - v_k \right\rangle \geqslant 0 \qquad \forall x \in X_0,$$
$$\frac{\xi_{ik}}{\alpha_k} (g_i(x_k) + \langle g_i'(x_k), v_k - x_k \rangle) = 0, \quad i = 1, \ldots, m. \tag{25}$$

Выбирая при необходимости подпоследовательности из ограниченных последовательностей $\{x_k\}$, $\{v_k\}$, $\{\xi^k/\alpha_k\}$, можем считать, что эти последовательности сходятся. С учетом (24) тогда

$$\lim_{k \to \infty} x_k = \lim_{k \to \infty} v_k = v_*, \quad \lim_{k \to \infty} \frac{\xi^k}{\alpha_k} = \mu_i^* \geqslant 0, \quad i = 1, \ldots, m.$$

Из замкнутости X_0 следует, что $v_* \in X_0$, а из (15) при $k \to \infty$ получим $g_i(v_*) \leqslant 0$, $i = 1, \ldots, m$. Следовательно, $v_* \in X$. Далее, из (25) при $k \to \infty$ с учетом (22) имеем

$$\langle f'(v_*) + \sum_{i=1}^{m} \mu_i^* g_i'(v_*), x - v_* \rangle \geqslant 0 \qquad \forall x \in X_0, \quad \mu_i^* g_i(v_*) = 0, \quad i = 1, \ldots, m. \tag{26}$$

Как следует из леммы 4.9.2 и теоремы 4.9.2, соотношения (26) означают, что $v_* \in X_*$. Вспомним, что неравенство (23) было получено при любых $x_* \in X_*$. В частности, (23) верно и при $x_* = v_*$. Но v_* — предельная точка последовательности $\{x_k\}$. Согласно лемме 2.6.10 тогда $\{x_k\} \to v_*$. Теорема доказана. □

Замечание 1. Если в (5) $\varepsilon_k = 0$, то согласно (9) тогда $x_{k+1} = v_k$, $k = 0, 1, \ldots$, и неравенство (23) можно записать в виде

$$|x_{k+1} - x_*|^2 + \gamma |x_{k+1} - x_k|^2 \leqslant |x_k - x_*|^2, \quad k = 0, 1, \ldots, \quad \forall x_* \in X_*.$$

Пользуясь произволом в выборе $x_* \in X_*$, отсюда имеем

$$|x_k - v_*| \geqslant |x_{k+1} - v_*|, \quad \rho(x_k, X_*) \geqslant \rho(x_{k+1}, X_*), \quad k = 0, 1, \ldots,$$

причем равенство здесь возможно лишь при $x_{k+1} = x_k = v_* \in X_*$. Таким образом, при точной реализации описанного метода линеаризации расстояние от точки x_k до множества X_* или до точки v_* монотонно убывает. В то же время можно отметить, что хотя и $\{f(x_k)\} \to f(v_*) = f_*$, но $\{f(x_k)\}$ не обязательно монотонно убывает и не обязательно $x_k \in X$.

Непрерывные и другие различные варианты метода линеаризации описаны и исследованы, например, в [24; 27; 29; 286; 304; 603; 606; 670; 738; 774].

Упражнения

1. Доказать, что если в (4) окажется $v_k = x_k$ при некотором $k \geqslant 0$, то точка x_k удовлетворяет необходимым условиям оптимальности. Указание: применить теорему 4.8.1 к задаче (4), затем принять $x_k = v_k$.

2. Доказать, что если выполнены условия теоремы 1, $\varepsilon_k = 0$, $k = 0, 1, \ldots$, и в (4) $v_k = x_k$ при некотором $k \geqslant 0$, то $x_k \in X_*$. Указание: положить в (25), (15) $v_k = x_k$ и воспользоваться леммой 4.9.2 и теоремой 4.9.2.

3. Рассмотреть метод линеаризации для задачи (1) при $X_0 = E^n_+$, $m = 0$.

4. Описать метод линеаризации для задачи (1) с дополнительными линейными ограничениями $\langle a_i, x\rangle = b^i$, $i = m+1, \ldots, s$.

§ 8. Квадратичное программирование

1. Рассмотрим задачу

$$f(x) = \tfrac{1}{2}\langle Cx, x\rangle + \langle c, x\rangle \to \inf, \quad x \in X, \tag{1}$$

$$X = \{x \in E^n\colon \langle a_i, x\rangle \leqslant b^i,\ i = 1, \ldots, m;\ \langle a_i, x\rangle = b^i,\ i = m+1, \ldots, s\}, \tag{2}$$

где C — симметричная неотрицательно определенная матрица размера $n \times n$, т. е. $C \geqslant 0$; $c, a_i \in E^n$, $b^i \in \mathbb{R}$, $i = 1, \ldots, s$, (возможности $m = 0$, или $s = m$, или $s = m = 0$ не исключаются). Задачу (1), (2) принято называть *задачей квадратичного программирования*: в ней квадратичная выпуклая функция минимизируется на многогранном множестве. Такие задачи возникают в различных приложениях. Задачи определения расстояния от точки до многогранного множества, проектирования на такое множество также представляют примеры задачи квадратичного программирования, когда в (1) $C = I$ — единичная матрица. Задачи вида (1), (2) часто возникают как вспомогательные при описании различных методов минимизации (см., например, § 6). Поэтому важно иметь достаточно простые методы решения задачи квадратичного программирования. Оказывается, для задачи (1), (2), как и для задачи линейного программирования, существуют конечные (конечношаговые) методы их решения. Для построения таких методов сначала нужно выявить некоторые специфические особенности этой задачи. В частности, здесь полезно рассмотреть двойственную к (1), (2) задачу.

Введем функцию Лагранжа задачи (1), (2):

$$L(x, \lambda) = \tfrac{1}{2}\langle Cx, x\rangle + \langle c, x\rangle + \langle \lambda, Ax - b\rangle = \tfrac{1}{2}\langle Cx, x\rangle + \langle c + A^\top\lambda, x\rangle - \langle \lambda, b\rangle,$$
$$x \in X_0 = E^n, \quad \lambda \in \Lambda_0 = \{\lambda = (\lambda_1, \ldots, \lambda_s) \in E^s\colon \lambda_1 \geqslant 0, \ldots, \lambda_m \geqslant 0\},$$

где A — матрица размера $s \times n$ со строками $a_1, \ldots, a_s$, $b = (b^1, \ldots, b^s)$. Если $f_* > -\infty$, $X_* \neq \varnothing$, то согласно теореме 4.9.3 функция $L(x, \lambda)$ имеет седловую точку (x_*, λ^*), причем в силу леммы 4.9.2

$$L_x(x_*, \lambda^*) = Cx_* + c + A^\top\lambda^* = 0, \tag{3}$$

$$\lambda^*_i(\langle a_i, x_*\rangle - b^i) = 0, \quad i = 1, \ldots, m, \quad x_* \in X_*, \quad \lambda^* \in \Lambda_0. \tag{4}$$

Тогда двойственная к (1), (2) задача

$$\psi(\lambda) = \inf_{x \in E^n} L(x, \lambda) \to \sup, \quad \lambda \in \Lambda_0, \tag{5}$$

согласно теореме 4.9.5 также имеет решение, причем

$$f_* = \psi^* = \sup_{\Lambda_0} \psi(\lambda) = \psi(\lambda^*) = f(x_*), \quad x_* \in X_*, \quad \lambda^* \in \Lambda^* = \{\lambda \in \Lambda_0\colon \psi(\lambda) = \psi^*\}.$$

При дополнительном предположении положительной определенности матрицы C, т. е. $C > 0$, функция $\psi(\lambda)$ может быть выписана явно. В самом деле, тогда C невырождена и точка минимума $x = x(\lambda)$ функции $L(x, \lambda)$ по $x \in E^n$ однозначно определяется из системы $Cx + c + A^\top\lambda = 0$, так что $x(\lambda) = -C^{-1}(c + A^\top\lambda)$. Поэтому

$$\psi(\lambda) = L(x(\lambda), \lambda) = -\tfrac{1}{2}\langle c + A^\top\lambda, C^{-1}(c + A^\top\lambda)\rangle -$$
$$- \langle \lambda, b\rangle = -\tfrac{1}{2}\langle (AC^{-1}A^\top)\lambda, \lambda\rangle - \langle AC^{-1}c + b, \lambda\rangle - \tfrac{1}{2}\langle C^{-1}c, c\rangle, \quad \lambda \in \Lambda_0,$$

где $AC^{-1}A^\top \geqslant 0$. Таким образом, при $C > 0$ двойственная задача (5), записанная в виде

$$-\psi(\lambda) \to \inf, \quad \lambda \in \Lambda_0, \tag{6}$$

также является задачей квадратичного программирования вида (1), (2), но множество Λ_0 по сравнению с (2) имеет более простую структуру. Зная какое-либо решение λ^* задачи (6), можно записать решение исходной задачи (1), (2) в виде

$$x_* = -C^{-1}(c + A^\top\lambda^*). \tag{7}$$

В самом деле, при $C > 0$ функция $f(x)$ сильно выпукла и согласно теореме 4.3.1 задача (1), (2) имеет единственное решение x_*, которое обязательно будет решением системы (3), (4), где λ^* — решение задачи (6). И поскольку система (3) при фиксированном λ^* однозначно определяет точку x_*, то необходимо приходим к формуле (7).

Особенно проста задача (6) в том случае, когда в исходной задаче (1), (2) отсутствуют ограничения типа неравенств ($m = 0$) и множество X имеет вид

$$X = \{x \in E^n\colon \langle a_i, x\rangle = b^i,\ i = 1, \ldots, s\}. \tag{8}$$

Тогда $\Lambda_0 = E^s$, и задача (6) запишется в форме

$$-\psi(\lambda) \to \inf, \quad \lambda \in E^s. \tag{9}$$

Множество Λ^* решений задачи (9) в силу теоремы 4.2.3 совпадает со множеством решений системы

$$-\psi'(\lambda^*) = AC^{-1}A^\top\lambda^* + AC^{-1}c + b = 0. \tag{10}$$

В общем случае система (10) может иметь более одного решения. Если матрица A невырожденная, т. е. векторы $a_1, \ldots, a_s$ в (8) линейно независимы, то из $C > 0$ следует $AC^{-1}A^\top > 0$, и тогда задача (9) и, следовательно, система (10) будут иметь единственное решение. Таким образом, при $C > 0$ для решения задачи (1), (8) достаточно решить две системы линейных алгебраических уравнений (10), (3). Здесь могут быть использованы известные методы линейной алгебры [59; 74; 89; 192; 353]. Поскольку для линейных систем имеется принципиальная возможность получить решение за конечное число арифметических операций (например, методом исключения Гаусса), то такая возможность имеется и для задачи (1), (8) при $C > 0$.

2. Следуя [670], покажем, что исходная задача (1), (2) при $C > 0$ может быть сведена к решению конечного числа задач вида (1), (8). Здесь важную роль играет понятие особой точки задачи (1), (2).

Определение 1. Точка v называется *особой точкой задачи* (1), (2), если $v \in X$ и v является решением задачи

$$f(x) \to \inf, \quad x \in V = \{x \in E^n: \langle a_i, x\rangle = b^i,\ i \in I \cup \{m+1, \ldots, s\}\}, \quad (11)$$

где I — какое-либо подмножество индексов $\{1, \ldots, m\}$ (возможность $I = \varnothing$ не исключается).

Лемма 1. *Пусть в задаче* (1), (2) $C \geqslant 0$, $f_* > -\infty$, $X_* \neq \varnothing$. *Тогда каждое решение x_* задачи* (1), (2) *является особой точкой этой задачи.*

Доказательство. Положим $I(x_*) = \{i: 1 \leqslant i \leqslant m, \langle a_i, x_*\rangle = b^i\}$ и рассмотрим задачу (11) с $I = I(x_*)$. Заметим, что $x_* \in V$. Так как $\langle a_i, x_*\rangle < b^i$ при $i \notin I(x_*)$, $1 \leqslant i \leqslant m$, и функция $\langle a_i, x\rangle$ непрерывна, то существует такая окрестность $S(x_*, \varepsilon) = \{x \in E^n: |x - x_*| < \varepsilon\}$, $\varepsilon > 0$, точки x_*, что $\langle a_i, x\rangle < b^i$ при всех $i \notin I(x_*)$, $1 \leqslant i \leqslant m$, и $x \in S(x_*, \varepsilon)$. Это значит, что $V \cap S(x_*, \varepsilon) \subset X$. Тогда $f(x) \geqslant f(x_*) = f_*$ при всех $x \in V \cap S(x_*, \varepsilon)$. Таким образом, x_* — точка локального минимума выпуклой задачи (11) с $I = I(x_*)$. По теореме 4.2.1 тогда x_* является точкой глобального минимума функции $f(x)$ на множестве V. Следовательно, x_* — решение задачи (11) с $I = I(x_*)$, так что x_* — особая точка задачи (1), (2). □

Теорема 1. *Пусть $C > 0$, множество* (2) *непусто. Тогда существует конечный метод решения задачи* (1), (2).

Доказательство. Так как множество $\{1, \ldots, m\}$ имеет конечное число подмножеств I, а всякая задача (11) при $C > 0$ имеет одно решение (теорема 4.3.1), то и число особых точек задачи (1), (2) конечно. Согласно лемме 1 для отыскания решения задачи (1), (2) достаточно перебрать все ее особые точки и найти ту из них, в которой функция (1) принимает меньшее значение. Так как задача (11) имеет вид (1), (8), то каждая особая точка может быть найдена за конечное число арифметических операций. Значит, поиск решения задачи (1), (2) закончится за конечное число шагов. □

3. Установленная в теореме 1 принципиальная возможность получения решения задачи (1), (2) за конечное число шагов имеет лишь теоретический интерес. Дело в том, что полный перебор особых точек задачи (1), (2) на практике требует слишком большого объема вычислений уже при не очень больших значениях n, s. Опишем один из методов упорядоченного перебора особых точек задачи (1), (2), более экономичного по сравнению с полным перебором [670]. При описании этого метода можно выделить три этапа.

На 1-м начальном этапе определяется, будет ли множество (2) непустым, и, если $X \neq \varnothing$, то находится какая-либо точка $v \in X$. Здесь может быть использован, например, симплекс-метод, описанный в гл. 3.

2-й этап состоит в переходе от какой-либо точки $v \in X$ к особой точке $w \in X$ со значением $f(w) \leqslant f(v)$. Для построения такой точки w можно воспользоваться следующим итерационным процессом. В качестве начального приближения берется $u_0 = v$. Пусть известно k-е приближение $u_k \in X$, $f(u_k) \leqslant f(v)$. Определим вспомогательное приближение $\bar{u}_k$ как решение задачи (11) при

$$I = I(u_k) = \{i: 1 \leqslant i \leqslant m, \langle a_i, u_k\rangle = b^i\}.$$

Поскольку $u_k \in V$, то $f(\bar{u}_k) \leqslant f(u_k) \leqslant f(v)$. Поэтому, если $\bar{u}_k \in X$, то в качестве требуемой особой точки можем взять $w = \bar{u}_k$. Допустим, что $\bar{u}_k \notin X$. Тогда $\langle a_j, \bar{u}_k\rangle > b^j$ хотя бы для одного $j \notin I(u_k)$, $1 \leqslant j \leqslant m$, так что множество индексов $I_1 = \{j: \langle a_j, \bar{u}_k\rangle > b^j,\ j \notin I(u_k),\ 1 \leqslant j \leqslant m\} \neq \varnothing$. Положим

$$u_{k+1} = u_k + \alpha_k(\bar{u}_k - u_k), \quad \alpha_k = \min_{j \in I_1}[(b^j - \langle a_j, u_k\rangle)/\langle a_j, \bar{u}_k - u_k\rangle]. \quad (12)$$

Из определения I_1 и включения $u_k \in X$ следует, что

$$0 < \frac{b^j - \langle a_j, u_k\rangle}{\langle a_j, \bar{u}_k - u_k\rangle} = \frac{b^j - \langle a_j, u_k\rangle}{(b^j - \langle a_j, u_k\rangle) + (\langle a_j, \bar{u}_k\rangle - b^j)} < 1, \quad j \in I_1,$$

поэтому $0 < \alpha_k < 1$. Покажем, что $u_{k+1} \in X$. Из выпуклости множества V и из $u_k \in V$, $\bar{u}_k \in V$ следует, что $u_{k+1} \in V$, где V взято из (11) при $I = I(u_k)$. Остается доказать, что $\langle a_j, u_{k+1}\rangle \leqslant b^j$ при всех $j \notin I(u_k)$, $1 \leqslant j \leqslant m$. Если $j \in I_1$, то с учетом определения (12) величины α_k имеем

$$\langle a_j, u_{k+1}\rangle = \langle a_j, u_k\rangle + \alpha_k\langle a_j, \bar{u}_k - u_k\rangle \leqslant b^j, \quad j \in I_1. \quad (13)$$

Если $j \notin I_1 \cup I(u_k)$, $1 \leqslant j \leqslant m$, то $\langle a_j, u_{k+1}\rangle = \alpha_k\langle a_j, \bar{u}_k\rangle + (1-\alpha_k)\langle a_j, u_k\rangle \leqslant b^j$. Таким образом, показано, что $u_{k+1} \in X$. Далее, поскольку $f(\bar{u}_k) \leqslant f(u_k) \leqslant f(v)$ и функция $f(x)$ выпукла, то $f(u_{k+1}) \leqslant \alpha_k f(\bar{u}_k) + (1-\alpha_k)f(u_k) \leqslant f(v)$. Далее, из включения $u_{k+1} \in V$, где V взято из (11) при $I = I(u_k)$, следует, что $I(u_k) \subset I(u_{k+1}) = \{i: \langle a_i, u_{k+1}\rangle = b^i,\ 1 \leqslant i \leqslant m\}$. В то же время для тех $j_0 \in I_1$, для которых в (12) реализовывается минимальное значение при определении α_k, неравенство (13) превращается в равенство, так что $j_0 \in I(u_{k+1})$, но $j_0 \notin I(u_k)$. Следовательно, множество $I(u_{k+1})$ содержит по крайней мере на один элемент больше, чем $I(u_k)$. Таким образом, следующее приближение $u_{k+1} \in X$ со значением $f(u_{k+1}) \leqslant f(v)$ построено, причем множество $I(u_{k+1})$ существенно шире $I(u_k)$. Однако множества $I(u_k) \in \{1, \ldots, m\}$ не могут бесконечно расширяться, и поэтому описанный процесс закончится на какой-то k-й итерации, когда $\bar{u}_k \in X$, причем $w \equiv \bar{u}_k$ — особая точка задачи (1), (2) со значением $f(w) \leqslant f(v)$. Поскольку решаемая на каждой итерации задача (11) с $I = I(u_k)$ имеет вид (1), (8) и для ее решения существует конечный метод, то и весь переход от точки v к точке w осуществим за конечное число шагов.

На 3-м этапе выясняется, не будет ли особая точка w, построенная на 2-м этапе, решением задачи (1), (2), и в том случае, если $w \notin X_*$, осуществляется переход к следующей точке $z \in X$, для которой $f(z) < f(w)$. Для этих целей достаточно совершить один шаг несколько модифицированного метода условного градиента, приняв в качестве начальной точку w, полученную на 2-м этапе. А именно, сначала можно решить следующую задачу линейного программирования

$$\langle f'(w), e\rangle = \langle Cw + c, e\rangle \to \inf, \quad e \in \mathcal{E} = \{e = (e^1, \ldots, e^n) \in E^n:$$
$$\langle a_i, e\rangle \leqslant 0,\ i \in I(w) = \{i: 1 \leqslant i \leqslant m, \langle a_i, w\rangle = b^i\},$$
$$\langle a_i, e\rangle = 0,\ i = m+1, \ldots, s,\ -1 \leqslant e^j \leqslant 1,\ j = 1, \ldots, n\}. \quad (14)$$

Пусть $e = e_*$ — решение задачи (14), которое может быть получено, например, симплекс-методом. Так как $e = 0 \in \mathcal{E}$, то $\beta = \langle f'(w), e_*\rangle = \min_{e \in \mathcal{E}}\langle f'(w), e\rangle \leqslant \langle f'(w), 0\rangle = 0$. Поэтому имеются две возможности: либо $\beta = 0$, либо $\beta < 0$. Покажем, что в случае $\beta = 0$ точка w — решение задачи (1), (2). С этой целью возьмем произвольную точку $u \in X$, $u \neq w$, и положим $e = t(u - w)$, где $t > 0$ столь мало, что $|e^j| = t|u^j - w^j| \leqslant 1$, $j = 1, \ldots, n$. Если $i \in I(w)$, то $\langle a_i, e\rangle = \langle a_i, u - w\rangle t = (\langle a_i, u\rangle - b^i)t \leqslant 0$. Если $m+1 \leqslant i \leqslant s$, то $\langle a_i, e\rangle = t(\langle a_i, u\rangle - b^i) = 0$. Следовательно, $e = t(u - w) \in \mathcal{E}$. Поэтому $\beta = 0 = \langle f'(w), e_*\rangle \leqslant \langle f'(w), e\rangle$. Пользуясь теоремой 4.2.2 тогда имеем

$f(u)-f(w)\geqslant\langle f'(w), u-w\rangle=\langle f'(w), e\rangle t^{-1}\geqslant 0$ при любом $u\in X$. Это значит, что $w\in X_*$, т. е. задача (1), (2) решена.

Рассмотрим вторую возможность: $\beta<0$. Тогда e_* — возможное направление убывания функции $f(u)$ в точке w. В самом деле, при достаточно малых $\alpha>0$ с учетом того, что e_* — решение задачи (13), имеем $f(w+\alpha e_*)-f(w)=\langle f'(w), e_*\rangle\alpha+o(\alpha)=\alpha(\beta+o(\alpha)/\alpha)<0$; если $i\in I(w)$ или $m+1\leqslant i\leqslant s$, то $\langle a_i, w+\alpha e_*\rangle=b^i+\alpha\langle a_i, e_*\rangle\leqslant b^i$; если $i\notin I(w)$, $1\leqslant i\leqslant m$, то $\langle a_i, w\rangle<b^i$ и $\langle a_i, w+\alpha e_*\rangle<b^i$. Тогда в качестве искомой точки $z=X$, $f(z)<f(w)$, можно взять $z=w+\alpha_0 e_*$, где $\alpha_0>0$ — достаточно малое число, которое может быть найдено за конечное число шагов, например, перебором чисел $\alpha_0=2^{-p}$, $p=0, 1, \ldots$. Описание 3-го этапа закончено.

Отправляясь от точки z, полученной на 3-м этапе, можно снова перейти ко 2-му этапу при $v=z$, затем к 3-му этапу и т. д. В результате будет построена последовательность особых точек, на которой функция $f(x)$ строго убывает. Так как при $C>0$ число особых точек конечно, то на каком-то шаге процесс, состоящий в последовательном применении 2-го и 3-го этапов, закончится нахождением решения задачи (1), (2). Таким образом, описанный метод позволяет за конечное число шагов найти решение задачи (1), (2) при $C>0$.

Существуют конечные методы решения задачи квадратичного программирования (1), (2) и при $C\geqslant 0$; об этих методах читатель сможет прочесть, например, в [1; 33; 48; 61; 203; 204; 222; 273; 319; 422; 471; 481; 586; 603; 606; 608; 612; 670; 738; 746; 759]. О задачах кубического программирования и, в общем случае, полиномиального программирования, когда минимизируемая функция является многочленом, см. [70; 83; 84; 527].

Упражнения

1. Уточните описание каждого этапа приведенного выше метода для задачи (1), (2) при $C=I$ — единичная матрица, а также при $X=E^n_+$ или $X=\{u=(u^1, \ldots, u^n)\in E^n\colon \alpha_i\leqslant u^i\leqslant\beta_i,\ i=1, \ldots, n\}$.

2. Примените описанный выше метод к задачам квадратичного программирования из упражнения 3 к § 4.9.

3. Пусть $X=\{u=(x, y)\in E^2\colon -1\leqslant x, y\leqslant 1\}$ или $X=\{u=(x, y)\in E^2\colon 0\leqslant x, y\leqslant 1\}$. Найдите особые точки задачи минимизации функций $f(u)=(x-a)^2+(y-b)^2$, $f(u)=(ax+by)^2$ на этих множествах при различных a, b.

4. Доказать, что множество точек минимума квадратичной функции $f(x)=|Ax-b|^2$ на E^n совпадает со множеством решений системы $A^\top Ax=A^\top b$ (см. пример 4.2.4).

5. Доказать, что квадратичная (или кубическая) функция достигает своей нижней грани на множестве (2), либо неограничена снизу [84].

§ 9. Метод сопряженных направлений

В описанных выше методах, использующих градиент функции, на каждой итерации учитывается информация лишь о линейной части приращения минимизируемой функции в окрестности полученной точки. С помощью этих методов точку минимума квадратичной функции удается найти лишь за бесконечное число итераций. Возникает вопрос: нельзя ли придумать метод,

использующий лишь градиент функции, который позволяет найти точку минимума квадратичной функции на всем пространстве за конечное число шагов? Если бы такой метод существовал, то можно было бы ожидать, что он сходится к точке минимума гладких функций быстрее градиентного метода, поскольку в окрестности точки минимума гладкая функция достаточно хорошо аппроксимируется квадратичной функцией.

Оказывается, методы с упомянутыми свойствами существуют. Одним из таких методов является *метод сопряженных направлений*. Опишем один из вариантов этого метода.

1. Сначала рассмотрим квадратичную задачу:

$$f(x)=\tfrac{1}{2}\langle Ax, x\rangle-\langle b, x\rangle\to\inf;\quad x\in X=E^n, \tag{1}$$

где A — симметричная положительная матрица, $b\in E^n$. Тогда, как было показано выше, справедливы формулы

$$f'(x)=Ax-b,\quad f''(x)=A,$$

и, кроме того, $f(x)$ сильно выпукла на E^n и достигает своей нижней грани на E^n в единственной точке x_* такой, что

$$f'(x_*)=Ax_*-b=0\quad\text{или}\quad x_*=A^{-1}b. \tag{2}$$

Возьмем произвольную начальную точку $x_0\in E^n$ и вычислим $p_0=f'(x_0)$. Если $f'(x_0)=0$, то $x_0=x_*$ — задача (1) решена. Поэтому пусть $f'(x_0)\neq 0$. Тогда положим

$$x_1=x_0-\alpha_0 p_0,\quad \alpha_0\geqslant 0,$$

где величина α_0 определяется условием

$$g_0(\alpha_0)=\min_{\alpha\geqslant 0} g_0(\alpha),\quad g_0(\alpha)=f(x_0-\alpha p_0).$$

Таким образом, первая итерация метода сопряженных направлений совпадает с итерацией метода скорейшего спуска. Заметим, что $g_0(\alpha)$ сильно выпукла, поэтому величина α_0 существует и определяется однозначно (см. ниже формулу (14)). Поскольку $g_0'(0)=-\langle f'(x_0), p_0\rangle=-|f'(x_0)|^2<0$, то $\alpha_0>0$. Следовательно,

$$g_0'(\alpha_0)=0=-\langle f'(x_0-\alpha_0 p_0), p_0\rangle=-\langle f'(x_1), p_0\rangle=-\langle f'(x_1), f'(x_0)\rangle. \tag{3}$$

Можем считать, что $f'(x_1)\neq 0$, иначе $x_1=x_*$ и задача (1) решена.

Так как $p_0\neq 0$, то $Ap_0\neq 0$ и множество

$$\Gamma_1=\{x\in E^n\colon \langle Ap_0, x-x_1\rangle=0\}$$

представляет собой гиперплоскость размерности $n-1$, проходящую через точку x_1. Важно заметить, что искомая точка x_* также принадлежит Γ_1. В самом деле, поскольку матрицы A, A^{-1} симметричны, то с учетом равенств (2), (3) имеем $\langle Ap_0, x_*-x_1\rangle=\langle Ap_0, A^{-1}b-x_1\rangle=\langle p_0, b-Ax_1\rangle=-\langle p_0, f'(x_1)\rangle=0$. Поэтому дальнейший поиск точки x_* имеет смысл проводить в гиперплоскости Γ_1. Для этого нужно найти какое-либо направление p_1, параллельное гиперплоскости Γ_1. Можно искать p_1, например, в виде

$$p_1=f'(x_1)-\beta_0 p_0,\quad \beta_0=\text{const}.$$

Условие параллельности p_1 гиперплоскости Γ_1 дает равенство $\langle Ap_0, p_1\rangle = \langle Ap_0, f'(x_1) - \beta_0 p_0\rangle = 0$, т. е.

$$\beta_0 = \langle Ap_0, f'(x_1)\rangle / \langle Ap_0, p_0\rangle.$$

Поскольку $f'(x_1) \neq 0$, то $p_1 \neq 0$. В самом деле, если бы $p_1 = 0$, то $f'(x_1) = \beta_0 p_0$ и согласно (3) тогда $|f'(x_1)|^2 = \beta_0 \langle p_0, f'(x_1)\rangle = 0$ — противоречие с условием $f'(x_1) \neq 0$. Из $p_1 \neq 0$ следует, что $Ap_1 \neq 0$. Положим

$$x_2 = x_1 - \alpha_1 p_1, \quad \alpha_1 \geqslant 0,$$

где величину α_1 будем определять из условия

$$g_1(\alpha_1) = \min_{\alpha \geqslant 0} g_1(\alpha), \quad g_1(\alpha) = f(x_1 - \alpha p_1).$$

С учетом равенств (3) имеем $g_1'(0) = \langle f'(x_1), -p_1\rangle = \langle f'(x_1), -f'(x_1) + \beta_0 p_0\rangle = -|f'(x_1)|^2 < 0$, поэтому $\alpha_1 > 0$. Тогда

$$g_1'(\alpha_1) = 0 = \langle f'(x_1 - \alpha_1 p_1), -p_1\rangle = -\langle f'(x_2), p_1\rangle = 0.$$

Заметим, что

$$f'(x_1) - f'(x_2) = Ax_1 - b - Ax_2 + b = \alpha_1 Ap_1.$$

Тогда в силу (3) и выбора p_1 получаем

$$\langle f'(x_2), f'(x_0)\rangle = \langle f'(x_2), p_0\rangle = \langle f'(x_1) - \alpha_1 Ap_1, p_0\rangle = 0.$$

Отсюда следует равенство

$$\langle f'(x_2), f'(x_0)\rangle = \langle f'(x_2), p_1 + \beta_0 p_0\rangle = 0.$$

Таким образом, первые две итерации метода сопряженных направлений для задачи (1) описаны. Показано, что

$$\langle f'(x_1), p_0\rangle = \langle f'(x_1), f'(x_0)\rangle = 0, \quad \langle Ap_0, p_1\rangle = \langle Ap_1, p_0\rangle = 0,$$
$$\langle f'(x_2), p_1\rangle = \langle f'(x_2), p_0\rangle = \langle f'(x_2), f'(x_1)\rangle = \langle f'(x_2), f'(x_0)\rangle = 0.$$

Кроме того, заметим, что векторы Ap_0, Ap_1 линейно независимы. В самом деле, если $\gamma_0 Ap_0 + \gamma_1 Ap_1 = 0$, то, умножая это равенство скалярно сначала на p_0, затем на p_1, получим $\gamma_0 = \gamma_1 = 0$. Можем считать, что $f'(x_2) \neq 0$, иначе $x_2 = x_*$ — задача (1) решена.

Теперь у нас есть основание сделать следующее индуктивное предположение: пусть при некотором $k \geqslant 2$ уже найдены точки $x_0, x_1, \ldots, x_k$, $x_{i+1} = x_i - \alpha_i p_i$, $i = 0, 1, \ldots, k-1$, где

$$p_i = f'(x_i) - \beta_i p_{i-1} \neq 0, \quad \beta_i = \langle f'(x_i), Ap_{i-1}\rangle / \langle Ap_{i-1}, p_{i-1}\rangle,$$

а величины $\alpha_i > 0$ определены из условий

$$g_i(\alpha_i) = \min_{\alpha \geqslant 0} g_i(\alpha), \quad g_i(\alpha) = f(x_i - \alpha p_i), \quad i = 0, 1, \ldots, k-1;$$

пусть

$$\langle Ap_i, p_j\rangle = 0, \quad i \neq j, \quad 0 \leqslant i, \quad j \leqslant k-1, \tag{4}$$
$$\langle f'(x_i), p_j\rangle = 0, \quad 0 \leqslant j < i \leqslant k, \tag{5}$$
$$\langle f'(x_i), f'(x_j)\rangle = 0, \quad i \neq j, \quad 0 \leqslant i, \quad j \leqslant k; \tag{6}$$

и, кроме того, пусть $f'(x_i) \neq 0$, $i = 0, 1, \ldots, k$, и система векторов $\{Ap_0, Ap_1, \ldots, Ap_{k-1}\}$ линейно независима.

Тогда множество

$$\Gamma_k = \{x \in E^n: \langle Ap_i, x - x_{k+1}\rangle = 0, \ i = 0, 1, \ldots, k-1\}$$

представляет собой гиперплоскость (аффинное множество — см. пример 4.1.5) размерности $n - k$. Поскольку из (5) следует

$$\langle Ap_i, x_k - x_{i+1}\rangle = \langle p_i, Ax_k - Ax_{i+1}\rangle = \langle p_i, f'(x_k) - f'(x_{i+1})\rangle = 0$$

для всех $i = 0, 1, \ldots, k-1$, то $x_k \in \Gamma_k$. Замечательно то, что $x_* \in \Gamma_k$, так как согласно (2), (5)

$$\langle Ap_i, x_* - x_{i+1}\rangle = \langle Ap_i, A^{-1}b - x_{i+1}\rangle = \langle p_i, b - Ax_{i+1}\rangle = \langle p_i, -f'(x_{i+1})\rangle = 0, \quad i = 0, 1, \ldots, k-1.$$

Поэтому дальнейший поиск точки x_* целесообразно продолжать в гиперплоскости Γ_k. Для этого нужно найти направление p_k, параллельное Γ_k, т. е. удовлетворяющее условиям $\langle Ap_i, p_k\rangle = 0$, $i = 0, 1, \ldots, k-1$. Будем искать p_k в виде

$$p_k = f'(x_k) - \beta_k p_{k-1}. \tag{7}$$

Заметим, что

$$f'(x_i) - f'(x_{i+1}) = Ax_i - Ax_{i+1} = \alpha_i Ap_i, \quad i = 0, 1, \ldots, k-1. \tag{8}$$

Из (4), (6), (8) следует

$$\langle Ap_i, p_k\rangle = \langle Ap_i, f'(x_k) - \beta_k p_{k-1}\rangle = \langle Ap_i, f'(x_k)\rangle - \beta_k \langle Ap_i, p_{k-1}\rangle = \langle f'(x_i) - f'(x_{i+1}), f'(x_k)\rangle \alpha_i^{-1} = 0$$

для всех $i = 0, 1, \ldots, k-2$ при любом выборе β_k в (7). Поэтому для параллельности направления p_k гиперплоскости Γ_k остается удовлетворить равенству $\langle Ap_{k-1}, p_k\rangle = 0$. Отсюда имеем $\langle Ap_{k-1}, f'(x_k) - \beta_k p_{k-1}\rangle = \langle Ap_{k-1}, f'(x_k)\rangle - \beta_k \langle Ap_{k-1}, p_{k-1}\rangle = 0$ или

$$\beta_k = \langle Ap_{k-1}, f'(x_k)\rangle / \langle Ap_{k-1}, p_{k-1}\rangle. \tag{9}$$

Заметим, что $p_k \neq 0$, ибо в противном случае $f'(x_k) = \beta_k p_{k-1}$, и тогда в силу (5) имеем $|f'(x_k)|^2 = \beta_k \langle f'(x_k), p_{k-1}\rangle = 0$, что противоречит индуктивному предположению.

Итак, учитывая выбор направления p_k и равенства (4), имеем

$$\langle Ap_i, p_j\rangle = 0, \quad i \neq j, \quad 0 \leqslant i, \quad j \leqslant k. \tag{10}$$

Следующее $(k+1)$-е приближение будем искать в виде

$$x_{k+1} = x_k - \alpha_k p_k, \quad \alpha_k \geqslant 0, \tag{11}$$

где α_k определяется из условия

$$g_k(\alpha_k) = \min_{\alpha \geqslant 0} g_k(\alpha), \quad g_k(\alpha) = f(x_k - \alpha p_k). \tag{12}$$

Поскольку $g_k(\alpha)$ — сильно выпуклая функция, то величина α_k существует и единственна. С учетом предположений индукции и формулы (7) имеем

$$g_k'(0) = \langle f'(x_k), -p_k\rangle = \langle f'(x_k), -f'(x_k) + \beta_k p_{k-1}\rangle = -|f'(x_k)|^2 < 0.$$

Это значит, что $\alpha_k > 0$ и $g_k'(\alpha_k) = 0 = \langle f'(x_{k+1}), -p_k\rangle$ или

$$\langle f'(x_{k+1}), p_k\rangle = 0. \tag{13}$$

Отсюда нетрудно получить явное выражение для α_k. В самом деле,

$$0 = \langle f'(x_{k+1}), p_k\rangle = \langle Ax_{k+1} - b, p_k\rangle = \\ = \langle Ax_k - \alpha_k Ap_k - b, p_k\rangle = \langle f'(x_k), p_k\rangle - \alpha_k\langle Ap_k, p_k\rangle.$$

Так как $p_k \neq 0$, то $\langle Ap_k, p_k\rangle \neq 0$ и из последнего равенства вытекает

$$\alpha_k = \frac{\langle f'(x_k), p_k\rangle}{\langle Ap_k, p_k\rangle} = \frac{\langle f'(x_k), f'(x_k) - \beta_k p_{k-1}\rangle}{\langle Ap_k, p_k\rangle} = \frac{|f'(x_k)|^2}{\langle Ap_k, p_k\rangle}. \tag{14}$$

Далее, заметим

$$f'(x_k) - f'(x_{k+1}) = Ax_k - Ax_{k+1} = \alpha_k Ap_k.$$

Отсюда и из (4), (5) имеем

$$\langle f'(x_{k+1}), p_i\rangle = \langle f'(x_k) - \alpha_k Ap_k, p_i\rangle = 0, \quad i = 0, 1, \ldots, k-1. \tag{15}$$

Собрав все равенства (5), (13), (15), получим

$$\langle f'(x_i), p_j\rangle = 0, \quad 0 \leqslant j < i \leqslant k+1. \tag{16}$$

Из предположения индукции и равенств (16) следует

$$\langle f'(x_{k+1}), f'(x_i)\rangle = \langle f'(x_{k+1}), p_i + \beta_i p_{i-1}\rangle = 0, \quad i = 1, \ldots, k,$$
$$\langle f'(x_{k+1}), f'(x_0)\rangle = \langle f'(x_{k+1}), p_0\rangle = 0.$$

Отсюда и из (6) имеем

$$\langle f'(x_i), f'(x_j)\rangle = 0, \quad i \neq j, \quad 0 \leqslant i, \quad j \leqslant k+1.$$

Наконец, покажем, что система $\{Ap_0, \ldots, Ap_k\}$ линейно независима. В самом деле, если $\gamma_0 Ap_0 + \gamma_1 Ap_1 + \ldots + \gamma_k Ap_k = 0$, то, умножая это равенство на p_j скалярно, с учетом (10) получим $\gamma_j\langle Ap_j, p_j\rangle = 0$, $j = 0, 1, \ldots, k$. Так как $p_j \neq 0$, то $\langle Ap_j, p_j\rangle > 0$ и последние равенства возможны лишь при $\gamma_j = 0$, $j = 0, 1, \ldots, k$.

Тем самым все этапы индукции проведены, следующее $(k+1)$-е приближение x_{k+1} построено. Если $f'(x_{k+1}) = 0$, то $x_{k+1} = x_*$ — решение задачи (1) найдено. Если же $f'(x_{k+1}) \neq 0$, то согласно индукции процесс можно продолжать дальше.

Метод сопряженных направлений для задачи (1), заключающийся в построении последовательности $\{x_k\}$ по правилу (11), где α_k, p_k определяются из (7), (9), (12) (или (14)), $p_0 = f'(x_0)$, описан. Название этого метода объясняет следующее

Определение 1. Векторы $p_0, p_1, \ldots, p_k$ называются *сопряженными относительно матрицы* A или A*-ортогональными*, если $\langle p_i, p_j\rangle = 0$ при всех $i \neq j$, $0 \leqslant i, j \leqslant k$.

Нетрудно видеть, что для квадратичной задачи (1) метод сопряженных направлений закончится за конечное число итераций нахождением точки x_*. В самом деле, векторы $f'(x_0), f'(x_1), \ldots, f'(x_k), \ldots$, получаемые этим методом, образуют ортогональную систему: $\langle f'(x_i), f'(x_j)\rangle = 0$, $i \neq j$. Однако в n-мерном пространстве не может быть более n ненулевых взаимно ортогональных векторов. Следовательно, найдется номер k, $0 \leqslant k < n$ такой, что $f'(x_k) = 0$. Тогда $x_k = x_*$ — решение задачи (1).

2. Перейдем к рассмотрению задачи

$$f(x) \to \inf; \quad x \in X \equiv E^n, \tag{17}$$

где функция $f(x) \in C^1(E^n)$, причем в отличие от задачи (1) здесь $f(x)$ не предполагается квадратичной. Так как формула (9) содержит матрицу A, характеризующую квадратичную функцию (1), то описанный выше метод сопряженных направлений (7), (9), (11), (12) не может быть непосредственно применен для решения задачи (17). Поэтому сначала формулу (9) приведем к виду, не содержащему матрицу A. С учетом равенств (6), (8) числитель и знаменатель дроби (9) можно преобразовать так:

$$\langle Ap_{k-1}, f'(x_k)\rangle = \langle f'(x_{k-1}) - f'(x_k), f'(x_k)\rangle\alpha_{k-1}^{-1} = -|f'(x_k)|^2\alpha_{k-1}^{-1},$$
$$\langle Ap_{k-1}, p_{k-1}\rangle = \langle f'(x_{k-1}) - f'(x_k), p_{k-1}\rangle\alpha_{k-1}^{-1} = \langle f'(x_{k-1}), p_{k-1}\rangle\alpha_{k-1}^{-1} = \\ = \langle f'(x_{k-1}), f'(x_{k-1}) - \beta_{k-1}p_{k-2}\rangle\alpha_{k-1}^{-1} = |f'(x_{k-1})|^2\alpha_{k-1}^{-1}.$$

Тогда формула (9) запишется в виде

$$\beta_k = \frac{\langle f'(x_k), f'(x_{k-1}) - f'(x_k)\rangle}{|f'(x_{k-1})|^2}, \tag{18}$$

или

$$\beta_k = -\frac{|f'(x_k)|^2}{|f'(x_{k-1})|^2}. \tag{19}$$

Кроме того, вспоминая, что для функции (1) $A = f''(x_k)$, формулу (9) можно представить еще и в такой форме:

$$\beta_k = \frac{\langle f''(x_k)p_{k-1}, f'(x_k)\rangle}{\langle f''(x_k)p_{k-1}, p_{k-1}\rangle}. \tag{20}$$

Для квадратичной функции (1) все три формулы (18)–(20) дают одну и ту же величину β_k. Но если функция $f(x)$ отлична от квадратичной, то из этих формул будут получаться, вообще говоря, различные значения β_k.

В результате, отправляясь от соотношений (7), (11), (12), (18)–(20), придем к следующему описанию метода сопряженных направлений для задачи (17). Пусть x_0 — некоторое начальное приближение. Будем строить последовательность $\{x_k\}$ по правилам

$$x_{k+1} = x_k - \alpha_k p_k, \quad k = 0, 1, \ldots, \tag{21}$$

где

$$p_0 = f'(x_0), \quad p_k = f'(x_k) - \beta_k p_{k-1}, \quad k = 1, 2, \ldots, \tag{22}$$

величина α_k определяется условиями

$$\alpha_k \geqslant 0, \quad g_k(\alpha_k) = \min_{\alpha \geqslant 0} g_k(\alpha), \quad g_k(\alpha) = f(x_k - \alpha p_k), \tag{23}$$

а β_k в (22) вычисляется по одной из формул (18), (19) или (20). Отметим, что в варианте (20)–(23) метода сопряженных направлений требуется, чтобы $f(x) \in C^2(E^n)$, и поэтому на практике он применяется очень редко и лишь в тех случаях, когда матрица $f''(x)$ вычисляется достаточно просто.

Так как в задаче (17) квадратичность функции не предполагается, то нельзя ожидать, что описанный метод сопряженных направлений за конечное число итераций приведет к точке минимума функции $f(x)$ на E^n. Далее, точное определение величины α_k из условий (23) возможно лишь в редких случаях, поэтому реализация каждой итерации метода будет сопровождаться неизбежными погрешностями. Как показывает практика, эти погрешности, накапливаясь, могут привести к тому, что векторы $\{p_k\}$ перестают указывать направление убывания функции, и сходимость метода может нарушиться. Чтобы бороться с этим явлением, метод сопряженных направлений время от времени обновляют, полагая в (22) $\beta_k = 0$. Обозначим множество тех номеров $k \geqslant 1$, при которых принимается $\beta_k = 0$, через I_0. Номера $k \in I_0$ называются *моментами обновления метода*. Если метод используется без обновления, то $I_0 = \varnothing$. На практике часто берут $I_0 = \{n, 2n, 3n, \ldots\}$, где n-размерность рассматриваемого пространства. Возможны и другие правила выбора моментов обновления. Кстати, если $I_0 = \{1, 2, 3, \ldots\}$, то метод (21)–(23) превратится в метод скорейшего спуска.

Если функция $f(x)$ не является квадратичной, то для описанного метода сопряженных направлений равенства (5), (6), вообще говоря, не выполняются. Однако, тем не менее, и в общем случае при любом выборе моментов обновления справедливы равенства

$$\langle f'(x_{k+1}), p_k\rangle = 0, \quad \langle f'(x_k), p_k\rangle = |f'(x_k)|^2, \quad k = 0, 1, \ldots \tag{24}$$

В самом деле, при $k = 0$ имеем $p_0 = f'(x_0)$, поэтому $\langle f'(x_0), p_0\rangle = |f'(x_0)|^2$. Из условий (23) при $k = 0$ в случае $\alpha_0 > 0$ следует $g_0'(\alpha_0) = \langle f'(x_0 - \alpha_0 p_0), -p_0\rangle = -\langle f'(x_1), p_0\rangle = 0$. Если же $\alpha_0 = 0$, то $x_1 = x_0$ и $0 \leqslant g_0'(0) = -\langle f'(x_0), p_0\rangle = -|f'(x_0)|^2 \leqslant 0$, так что $f'(x_0) = f'(x_1) = 0$, $\langle f'(x_1), p_0\rangle = 0$. Таким образом, равенства (24) при $k = 0$ верны. Сделаем индуктивное предположение: пусть для некоторого $k \geqslant 1$ имеют место равенства $\langle f'(x_k), p_{k-1}\rangle = 0$, $\langle f'(x_{k-1}), p_{k-1}\rangle = |f'(x_{k-1})|^2$. Тогда из (23) при $\alpha_k > 0$ получим $g_k'(\alpha_k) = -\langle f'(x_{k+1}), p_k\rangle = 0$. Если же $\alpha_k = 0$, то $x_{k+1} = x_k$ и $0 \leqslant g_k'(0) = -\langle f'(x_k), p_k\rangle = -\langle f'(x_k), f'(x_k) - \beta_k p_{k-1}\rangle = -|f'(x_k)|^2 = -|f'(x_{k+1})|^2 \leqslant 0$, поэтому $f'(x_{k+1}) = 0$ и $\langle f'(x_{k+1}), p_k\rangle = 0$. Наконец,

$$\langle f'(x_k), p_k\rangle = \langle f'(x_k), f'(x_k) - \beta_k p_{k-1}\rangle = |f'(x_k)|^2.$$

Равенства (24) доказаны. Из первого равенства (24) и определения (22) вектора p_k следует

$$|p_k|^2 = |f'(x_k) - \beta_k p_{k-1}|^2 = |f'(x_k)|^2 + \beta_k^2 |p_{k-1}|^2, \quad k = 1, 2, \ldots \tag{25}$$

3. Пользуясь соотношениями (24), (25), установим сходимость метода сопряженных направлений (21)–(23), (18).

Теорема 1. *Пусть функция $f(x)$ сильно выпукла на E^n, $f(x) \in C^{1,1}(E^n)$. Тогда при любом выборе множества I_0 моментов обновления и любом начальном приближении x_0 последовательность $\{x_k\}$, определяемая условиями* (21)–(23), (18), *сходится к точке x_* минимума функции $f(x)$ на E^n, причем справедливы оценки*

$$0 \leqslant a_k = f(x_k) - f_* \leqslant q^k a_0, \quad |x_k - x_*|^2 \leqslant \frac{2}{\mu} q^k a_0, \quad k = 0, 1, \ldots, \tag{26}$$

где $q = 1 - \dfrac{\mu^3}{L(\mu^2 + L^2)}$, $0 < q < 1$, μ — постоянная из теоремы 4.3.3, *L — константа Липшица для градиента $f'(x)$ на E^n.*

Доказательство. Из теоремы 4.3.1 следует существование и единственность точки x_*, в которой $f(x_*) = f_* = \inf\limits_{E^n} f(x)$. Функция $g_k(\alpha) = f(x_k - \alpha p_k)$ при $p_k \neq 0$ также сильно выпукла, и условия (23) однозначно определяют величину $\alpha_k > 0$. Будем считать, что $p_k \neq 0$, $f'(x_k) \neq 0$, $\alpha_k > 0$ при всех $k = 0, 1, \ldots$, ибо в противном случае из (24) при $p_k = 0$ получим $f'(x_k) = 0$ и $x_k = x_*$ — решение задачи (17).

В силу выбора α_k при всех $\alpha \geqslant 0$ имеем $f(x_{k+1}) \leqslant f(x_k - \alpha p_k)$. Отсюда и из леммы 2.6.1 с учетом второго равенства (24) получим

$$f(x_k) - f(x_{k+1}) \geqslant f(x_k) - f(x_k - \alpha p_k) \geqslant \alpha\langle f'(x_k), p_k\rangle - \frac{\alpha^2 L}{2}|p_k|^2 =$$
$$= \alpha|f'(x_k)|^2 - \frac{\alpha^2 L}{2}|p_k|^2, \quad \alpha \geqslant 0, \quad k = 0, 1, \ldots \tag{27}$$

Докажем теперь неравенство

$$\gamma L |p_k|^2 \leqslant |f'(x_k)|^2, \quad \gamma = \mu^2 L^{-1}(\mu^2 + L^2)^{-1}, \quad k = 0, 1, \ldots \tag{28}$$

Согласно теореме 4.3.3 имеем $\mu|x_k - x_{k-1}|^2 = \mu\alpha_{k-1}^2|p_{k-1}|^2 \leqslant \langle f'(x_k) - f'(x_{k-1}), x_k - x_{k-1}\rangle = \langle f'(x_k) - f'(x_{k-1}), p_{k-1}\rangle(-\alpha_{k-1})$. Отсюда с учетом равенств (24) получаем

$$\mu\alpha_{k-1}|p_{k-1}|^2 \leqslant \langle f'(x_{k-1}), p_{k-1}\rangle = |f'(x_{k-1})|^2.$$

Тогда из (18) следует

$$|\beta_k| \leqslant |f'(x_k)|L|x_k - x_{k-1}||f'(x_{k-1})|^{-2} \leqslant$$
$$|f'(x_k)|L\alpha_{k-1}|p_{k-1}|(\mu\alpha_{k-1}|p_{k-1}|^2)^{-1} = L\mu^{-1}|f'(x_k)||p_{k-1}|^{-1},$$

т. е.

$$\beta_k|p_{k-1}| \leqslant L\mu^{-1}|f'(x_k)|, \quad k = 0, 1, \ldots$$

Отсюда и из (25) получим $|p_k|^2 \leqslant |f'(x_k)|^2(1 + L^2\mu^{-2})$, что равносильно неравенству (28).

Теперь нетрудно доказать оценки (26). Из (27) с учетом (28) имеем

$$a_k - a_{k+1} \geqslant \alpha\left(1 - \frac{\alpha}{2\gamma}\right)|f'(x_k)|^2, \quad \forall\alpha \geqslant 0, \quad k = 0, 1, \ldots$$

Следовательно,

$$a_k - a_{k+1} \geqslant \max_{\alpha \geqslant 0} \alpha\left(1 - \frac{\alpha}{2\gamma}\right)|f'(x_k)|^2 = \frac{\gamma}{2}|f'(x_k)|^2, \quad k = 0, 1, \ldots$$

Но $2\mu a_k \leqslant |f'(x_k)|^2$ (см. неравенство (1.18)), поэтому $a_k - a_{k+1} \geqslant \gamma\mu a_k$ или $a_{k+1} \leqslant (1 - \gamma\mu)a_k = qa_k$, $k = 0, 1, \ldots$ Отсюда следует первая из оценок (26). Вторая оценка (26) вытекает из первой оценки и неравенства (4.3.3). Остается заметить, что $0 < q < 1$, ибо $\mu \leqslant L$. Теорема доказана. □

Отметим, что оценки (26) являются довольно грубыми. Более тонкие исследования показывают, что метод сопряженных направлений на самом деле имеет более высокую скорость сходимости, чем это следует из оценок (20). В то же время этот метод ненамного сложнее метода скорейшего спуска. Недостатком метода сопряженных направлений является его чувствительность к погрешностям при определении величины α_k из условия (23) — недостаточно точное определение α_k может привести к ухудшению сходимости метода.

4. В методе (7), (9), (11), (12) направления $p_0, p_1, \ldots, p_k$ строятся с помощью процесса A-ортогонализации последовательно вычисляемых градиентов $f'(x_0), f'(x_1), \ldots, f'(x_k)$, и поэтому этот метод для задачи (1) и полученный на его основе метод (21)–(23) для задачи (17) в литературе часто называют *методом сопряженных градиентов*. В общем случае в методе сопряженных направлений могут быть использованы и другие способы построения векторов p_k, отличные от (22). А именно, пусть направления $p_0, p_1, \ldots, p_k$, удовлетворяющие условиям (10), уже известны и с их помощью последовательно построены точки $x_1, \ldots, x_{k+1}$ по формулам (21), (23). Следующий вектор p_{k+1} будем определять из условий $\langle p_{k+1}, Ap_i\rangle = 0$, $i = 0, 1, \ldots, k$. В случае квадратичных функций (1) формула (8) остается справедливой при любом выборе векторов $p_0, p_1, \ldots, p_k$ в (21), (23), поэтому условие ортогональности вектора p_{k+1} к векторам $Ap_0, Ap_1, \ldots, Ap_k$ здесь приводит к равенствам

$$\langle p_{k+1}, q_i\rangle = 0, \quad q_i = f'(x_i) - f'(x_{i+1}), \quad i = 0, 1, \ldots, k. \tag{29}$$

Условия (29) имеют смысл и для неквадратичных функций, и ими пользуются для определения p_{k+1} в общем случае. Обычно вектор p_{k+1} ищут в виде [61; 71; 76; 222; 374; 586; 759]

$$p_{k+1} = H_{k+1}f'(x_{k+1}), \quad H_{k+1} = H_k + \Delta H_k, \tag{30}$$

где матрица ΔH_k определяется из условий (29). Нетрудно видеть, что перечисленные условия (29), (30) матрицу ΔH_k определяют неоднозначно и в зависимости от того, как распорядиться этим произволом, можно получить различные варианты метода сопряженных направлений. Если на каком-либо шаге $H_k = 0$, то метод (21), (23), (29), (30) обновляют, полагая $A_k = I$ — единичная матрица. Приведем один из вариантов этого метода, в котором матрицы H_k определяются по правилу

$$H_{k+1} = H_k - \frac{(Hq_k)(H_k q_k)^\top}{\langle Hq_k, q_k\rangle}, \quad q_k = f'(x_{k+1}) - f'(x_k), \quad H_0 = I.$$

В [603] предлагается и исследуется метод сопряженных направлений, позволяющий за конечное число итераций найти точку минимума квадратичной функции (1) на множестве, задаваемом линейными ограничениями типа равенств и неравенств.

Различные варианты метода сопряженных направлений, более тонкие оценки скорости сходимости читатель может найти в [74; 374; 603].

Упражнения

1. Показать, что точка x_k, полученная методом сопряженных направлений для квадратичной функции (1) при $I_0 = \varnothing$, есть точка минимума этой функции на гиперплоскости, проходящей через точку x_0 и натянутой на векторы $f'(x_0), f'(x_1), \ldots, f'(x_{k-1})$.

2. Описать метод сопряженных направлений для функции $f(x) = |Ax - b|^2$, $x \in E^n$, где A — матрица порядка $m \times n$, $b \in E^m$.

§ 10. Метод Ньютона

До сих пор мы рассматривали методы первого порядка — так называются методы минимизации, использующие лишь первые производные минимизируемой функции. В этих методах для определения направления убывания функции используется лишь линейная часть разложения функции в ряд Тейлора. Если минимизируемая функция $f(x)$ дважды непрерывно дифференцируема и производные $f'(x)$, $f''(x)$ вычисляются достаточно просто, то возможно применение методов минимизации второго порядка, которые используют квадратичную часть разложения этой функции в ряд Тейлора. Поскольку квадратичная часть разложения аппроксимирует функцию гораздо точнее, чем линейная, то естественно ожидать, что методы второго порядка сходятся быстрее, чем методы первого порядка.

Ниже будет рассмотрен метод Ньютона, имеющий квадратичную скорость сходимости на классе сильно выпуклых функций. Здесь мы пользуемся следующей терминологией, принятой в литературе: говорят, что последовательность $\{x_k\}$ сходится к точке x_* с *линейной скоростью* или со скоростью геометрической прогрессии (со знаменателем q), если, начиная с некоторого номера, выполняется неравенство $|x_{k+1} - x_*| \leqslant q|x_k - x_*|$, $0 < q < 1$; при выполнении неравенства $|x_{k+1} - x_*| \leqslant q_k|x_k - x_*|$, где $\{q_k\} \to 0$, говорят о *сверхлинейной скорости* сходимости последовательности $\{x_k\}$ к x_*, а если здесь $q_k = C|x_k - x_*|^{s-1}$, т. е. $|x_{k+1} - x_*| \leqslant C|x_k - x_*|^s$, то говорят о *скорости сходимости порядка* s (при $s=2$ получим *квадратичную скорость* сходимости). Для некоторых методов выше была установлена линейная скорость сходимости на классе сильно выпуклых функций; в тех случаях, когда $|x_k - x_*| = O(1/k)$, скорость сходимости ниже линейной; для метода сопряженных направлений можно показать сверхлинейную скорость сходимости [603].

1. Опишем метод Ньютона для задачи

$$f(x) \to \inf; \quad x \in X, \tag{1}$$

где $f(x) \in C^2(X)$, X — выпуклое замкнутое множество из E^n (например, $X = E^n$). Пусть $x_0 \in X$ — некоторое начальное приближение. Если известно k-е приближение x_k, то приращение функции $f(x) \in C^2(X)$ в точке x_k можно представить в виде

$$f(x) - f(x_k) = \langle f'(x_k), x - x_k\rangle + \tfrac{1}{2}\langle f''(x_k)(x - x_k), x - x_k\rangle + o(|x - x_k|^2).$$

Возьмем квадратичную часть этого приращения

$$f_k(x) \equiv \langle f'(x_k), x - x_k\rangle + \tfrac{1}{2}\langle f''(x_k)(x - x_k), x - x_k\rangle \tag{2}$$

и определим вспомогательное приближение $\bar{x}_k$ из условий

$$\bar{x}_k \in X, \quad f_k(\bar{x}_k) = \inf_X f_k(x). \tag{3}$$

Следующее $(k+1)$-е приближение будем искать в виде

$$x_{k+1} = x_k + \alpha_k(\bar{x}_k - x_k), \quad 0 \leqslant \alpha_k \leqslant 1. \tag{4}$$

В зависимости от способа выбора величины α_k в (4) можно получить различные варианты метода (2)–(4), называемого методом Ньютона. Укажем несколько наиболее употребительных способов выбора α_k.

1) В (4) можно принять

$$\alpha_k = 1, \quad k = 0, 1, \ldots \tag{5}$$

В этом случае, как следует из (4), $x_{k+1} = \bar{x}_k$, $k = 0, 1, \ldots$, т. е. условие (3) сразу определяет следующее $(k+1)$-е приближение. Иначе говоря,

$$x_{k+1} \in X, \quad f_k(x_{k+1}) = \inf_X f_k(x), \quad k = 0, 1, \ldots \tag{6}$$

В частности, когда $X = E^n$, в точке минимума функции $f_k(x)$ ее производная $f_k'(x)$ обращается в нуль, т. е.

$$f_k'(x_{k+1}) = f'(x_k) + f''(x_k)(x_{k+1} - x_k) = 0. \tag{7}$$

Это значит, что на каждой итерации метода (2)–(5) или (6) нужно решать линейную алгебраическую систему уравнений (7) относительно неизвестной разности $x_{k+1} - x_k$. Если матрица этой системы $f''(x_k)$ — невырожденная, то из (7) имеем

$$x_{k+1} = x_k - (f''(x_k))^{-1} f'(x_k), \quad k = 0, 1, \ldots \tag{8}$$

Широко известный метод Ньютона для решения системы уравнений

$$F(x) = \{F_1(x), \ldots, F_n(x)\} = 0, \quad x \in E^n,$$

представляет собой итерационный процесс [74; 89; 550; 635]

$$x_{k+1} = x_k - (F'(x_k))^{-1} F(x_k), \quad k = 0, 1, \ldots, \tag{9}$$

где $F'(x)$ — матрица, i-я строка которой равна $F_i'(x) = (F_{ix^1}, \ldots, F_{ix^n})$. Сравнение формул (8) и (9) показывает, что метод (8) решения задачи (1) в

случае $X = E^n$ представляет собой известный метод Ньютона для решения уравнения $f'(x) = 0$, определяющего стационарные точки функции $f(x)$. Отсюда происходит название метода (2)–(4) и в общем случае.

2) В качестве α_k в (4) можно принять $\alpha_k = \lambda^{i_0}$, где i_0 — минимальный среди $i \geqslant 0$ номер, для которых выполняется неравенство [603]

$$f(x_k) - f(x_k + \lambda^i(\bar{x}_k - x_k)) \geqslant \varepsilon\lambda^i |f_k(\bar{x}_k)|, \tag{10}$$

где λ, ε — параметры метода, $0 < \lambda$; $\varepsilon < 1$.

3) Возможен выбор α_k в (4) из условий [603]

$$0 \leqslant \alpha_k \leqslant 1, \quad g_k(\alpha_k) = \min_{0 \leqslant \alpha \leqslant 1} g_k(\alpha), \quad g_k(\alpha) = f(x_k + \alpha(\bar{x}_k - x_k)). \tag{11}$$

Заметим, что метод (2)–(4) с выбором длины шага α_k по правилам (10), (11) аналогичен соответствующим вариантам метода условного градиента, где для определения $\bar{x}_k$ использовалась линейная часть приращений, а в методе Ньютона — квадратичная часть (2).

Если $f_k(x)$ из (2) сильно выпукла, а $X = E^n$ или X задается линейными ограничениями типа равенств или неравенств, то для определения $\bar{x}_k$ из (3) могут быть использованы методы из § 8, 9. Следует заметить, что задача (3) в общем случае может оказаться весьма сложной и сравнимой по объему требуемой для своего решения вычислительной работы с исходной задачей (1). Метод Ньютона для решения задачи (1) обычно применяют в тех случаях, когда вычисление производных $f'(x)$, $f''(x)$ не представляет особых трудностей и вспомогательная задача (3) решается достаточно просто. Достоинством метода Ньютона является высокая скорость сходимости. Поэтому, хотя трудоемкость каждой итерации этого метода, вообще говоря, выше, чем в методах первого порядка, но общий объем вычислительной работы, необходимой для решения задачи (1) с требуемой точностью, при применении метода Ньютона может оказаться меньше, чем при применении других более простых методов.

2. Сначала исследуем сходимость метода Ньютона (2)–(4) с выбором шага α_k из условия (5) при условии $X = E^n$ или, проще говоря, метода (8).

Теорема 1. *Пусть функция $f(x)$ сильно выпукла на E^n, $f(x) \in C^2(E^n)$ и, кроме того,*

$$\|f''(x) - f''(y)\| \leqslant L|x - y|, \quad x, y \in E^n, \quad L = \text{const} > 0. \tag{12}$$

Пусть начальное приближение x_0 выбрано таким, что

$$L|f'(x_0)| \leqslant 2\mu^2 q, \tag{13}$$

где $\mu > 0$ — постоянная из теоремы 4.3.4, а q — некоторая константа, $0 < q < 1$. Тогда последовательность $\{x_k\}$, определяемая условиями (8), существует, сходится к точке x_* минимума $f(x)$ на E^n, причем справедлива оценка

$$|x_k - x_*| \leqslant 2\mu L^{-1} q^{2^k}, \quad k = 0, 1, \ldots \tag{14}$$

Доказательство. Существование и единственность точки x_* установлена в теореме 4.3.1. Согласно теореме 4.3.4

$$\langle f''(x)\xi, \xi\rangle \geqslant \mu|\xi|^2, \quad x \in E^n, \quad \xi \in E^n. \tag{15}$$

Отсюда следует, что система уравнений $f''(x)\xi = 0$ имеет единственное решение $\xi = 0$ и, следовательно, матрица $f''(x)$ невырожденная при всех $x \in E^n$. Это значит, что система (7) при каждом $k = 0, 1, \ldots$ имеет, и притом единственное, решение, т. е. последовательность $\{x_k\}$ однозначно определяется условиями (8). Кроме того, полагая в (15) $\xi = (f''(x))^{-1}z$, получим $\mu|(f''(x))^{-1}z|^2 \leqslant \langle z, (f''(x))^{-1}z\rangle \leqslant |z||(f''(x))^{-1}z|$ или $|(f''(x))^{-1}z| \leqslant |z|\mu^{-1}$ при всех $z \in E^n$. Это значит, что

$$\|(f''(x))^{-1}\| \leqslant \mu^{-1}, \quad x \in E^n. \tag{16}$$

Введем числовую последовательность $a_k = |f'(x_k)|$ и покажем, что

$$a_k \leqslant 2\mu^2 L^{-1} q^{2^k}, \quad k = 0, 1, \ldots \tag{17}$$

При $k = 0$ неравенство (17) следует из условия (13). Пусть (17) справедливо при некотором $k \geqslant 0$. Из условия (8) и формулы (2.6.5) имеем

$$f'(x_{k+1}) = f'(x_k) + \int_0^1 f''(x_k + t(x_{k+1} - x_k))(x_{k+1} - x_k)\,dt = \int_0^1 [f''(x_k) - f''(x_k + t(x_{k+1} - x_k))]\,dt\,(f''(x_k))^{-1}f'(x_k).$$

Отсюда и из (8), (12), (16) с помощью предположения индукции получим

$$a_{k+1} \leqslant (L/2)|x_{k+1} - x_k|\mu^{-1}a_k \leqslant (L/(2\mu^2))a_k^2 \leqslant (L/(2\mu^2))(2\mu^2/L)^2(q^{2^k})^2 = (2\mu^2/L)q^{2^{k+1}}.$$

Неравенства (17) доказаны. Тогда из теоремы 4.3.3 с учетом равенства $f'(x_*) = 0$ имеем $\mu|x_k - x_*|^2 \leqslant \langle f'(x_k) - f'(x_*), x_k - x_*\rangle \leqslant |f'(x_k)||x_k - x_*|$ или $|x_k - x_*| \leqslant a_k\mu^{-1}$. Отсюда и из неравенства (17) следует оценка (14).

Теорема 1 доказана. □

Как видно из оценки (14) и как показывает практика, метод Ньютона (8) сходится очень быстро. Однако у него есть один существенный недостаток: для его сходимости начальная точка x_0 должна выбираться достаточно близкой к искомой точке x_*. Это требование в теореме 1 выражено условием (13), означающим, что $|x_0 - x_*| \leqslant a_0\mu^{-1} \leqslant (2\mu/L)q$. Приведем пример, показывающий, что при отсутствии хорошего начального приближения метод (8) может расходиться.

Пример 1. Пусть

$$f(x) = \begin{cases} -\frac{1}{4\delta^3}x^4 + \frac{1}{2}\left(1 + \frac{3}{\delta}\right)x^2, & |x| \leqslant \delta, \\ \frac{x^2}{2} + 2|x| - \frac{3}{4}\delta, & |x| > \delta, \end{cases}$$

где $x \in E^1$, а δ — сколь угодно малое фиксированное положительное число, $0 < \delta < 1$. Нетрудно видеть, что $f(x) \in C^2(E^1)$ и, кроме того, $f''(x) \geqslant 1$ при всех $x \in E^1$, так что $f(x)$ сильно выпукла на E^1. Далее, ясно, что $f_* = 0$, $x_* = 0$. В качестве начального приближения возьмем $x_0 = \delta$. Из (8) получим последовательность $x_k = (-1)^k \cdot 2$, $k = 1, 2, \ldots$, которая расходится, хотя начальное приближение x_0 отличается от $x_* = 0$ на малое число δ.

Метод (8) часто применяют на завершающем этапе поиска минимума, когда с помощью более грубых, менее трудоемких методов уже найдена некоторая точка, достаточно близкая к точке минимума.

3. Исследуем сходимость метода (2)–(5) без предположения, что $X = E^n$.

Теорема 2. *Пусть X — выпуклое замкнутое множество из E^n, функция $f(x)$ сильно выпукла и принадлежит классу $C^2(X)$ и*

$$|f''(x) - f''(y)| \leqslant L|x - y|, \quad x, y \in X, \quad L = \text{const}. \tag{18}$$

Тогда последовательность $\{x_k\}$ однозначно определяется условиями (6), *при любом выборе начального приближения x_0. Если*

$$q = (L/(2\mu))|x_1 - x_0| < 1, \tag{19}$$

то последовательность $\{x_k\}$, определяемая условиями (6), *сходится к точке x_* — решению задачи* (1), *причем справедлива оценка*

$$|x_k - x_*| \leqslant \frac{2\mu}{L} \sum_{m=k}^{\infty} q^{2^m} \leqslant \frac{2\mu}{L} q^{2^k} (1 - q^{2^k})^{-1}, \quad k = 0, 1, \ldots; \tag{20}$$

здесь $\mu > 0$ — постоянная из теоремы 4.3.4.

Доказательство. В силу теоремы 4.3.1 функция $f(x)$ ограничена снизу и достигает своей нижней грани на X в единственной точке x_*. Из теоремы 4.3.4 следует

$$\langle f''(x)\xi, \xi\rangle \geqslant \mu|\xi|^2, \quad x \in X, \quad \xi \in L_X, \tag{21}$$

где L_X — подпространство, параллельное аффинной оболочке множества X. Так как $f_k''(x) = f''(x_k)$, то из предыдущего неравенства и теоремы 4.3.4 вытекает сильная выпуклость функции $f_k(x)$ на множестве X при всех $k = 0, 1, \ldots$ Снова обращаясь к теореме 4.3.1, заключаем, что условия (6) однозначно определяют точку x_{k+1}. Таким образом, существование последовательности $\{x_k\}$ из (6) доказано. Применив теорему 4.2.3 к функции $f_k(x)$ на X, получим

$$\langle f_k'(x_{k+1}), x - x_{k+1}\rangle \geqslant 0, \quad x \in X, \quad k = 0, 1, \ldots \tag{22}$$

Так как $f_k'(x) \equiv f'(x_k) + f''(x_k)(x - x_k)$, то неравенство (22) перепишется в виде

$$\langle f'(x_k) + f''(x_k)(x_{k+1} - x_k), x - x_{k+1}\rangle \geqslant 0, \quad x \in X, \quad k = 0, 1, \ldots \tag{23}$$

Может случиться, что $x_{k+1} = x_k$. Тогда на (23) имеем $\langle f'(x_k), x - x_k\rangle \geqslant 0$ при всех $x \in X$. Согласно теореме 4.2.3 в этом случае $x_k = x_*$ — задача (1) решена. Поэтому можем считать, что $x_k \neq x_{k+1}$ при всех $k = 0, 1, \ldots$ Положим в (23) $x = x_k$. Получим

$$\langle f'(x_k) + f''(x_k)(x_{k+1} - x_k), x_k - x_{k+1}\rangle \geqslant 0.$$

Отсюда и из (21) имеем

$$\mu|x_{k+1} - x_k|^2 \leqslant \langle f''(x_k)(x_{k+1} - x_k), x_{k+1} - x_k\rangle \leqslant \langle f'(x_k), x_k - x_{k+1}\rangle, \quad k = 0, 1, \ldots \tag{24}$$

Оценим правую часть (24) сверху. Для этого в (22) заменим k на $k-1$. Получим $\langle f_{k-1}'(x_k), x - x_k\rangle \geqslant 0$, $x \in X$. Полагая здесь $x = x_{k+1}$, имеем

$$\langle f_{k-1}'(x_k), x_k - x_{k+1}\rangle \leqslant 0, \quad k = 1, 2, \ldots$$

Отсюда, из формулы (2.6.5) и условия (18) следует

$$\begin{aligned}\langle f'(x_k), x_k - x_{k+1}\rangle &\leqslant \langle f'(x_k) - f_{k-1}'(x_k), x_k - x_{k+1}\rangle = \\ &= \langle f'(x_k) - f'(x_{k-1}) - f''(x_{k-1})(x_k - x_{k-1}), x_k - x_{k+1}\rangle = \\ &= \left\langle \int_0^1 [f''(x_{k-1} + t(x_k - x_{k-1})) - f''(x_{k-1})]\,dt\,(x_k - x_{k-1}), x_k - x_{k+1}\right\rangle \leqslant \\ &\leqslant \frac{L}{2}|x_k - x_{k-1}|^2|x_k - x_{k+1}|, \quad k = 1, 2, \ldots\end{aligned}$$

Подставив полученную оценку в (24), имеем

$$|x_{k+1} - x_k| \leqslant (L/(2\mu))|x_k - x_{k-1}|^2, \quad k = 1, 2, \ldots \tag{25}$$

Докажем оценку

$$|x_{k+1} - x_k| \leqslant (2\mu/L)q^{2^k}, \quad k = 0, 1, \ldots \tag{26}$$

При $k = 0$ эта оценка следует из условия (19). Сделаем индуктивное предположение: пусть $|x_k - x_{k-1}| \leqslant (2\mu/L)q^{2^{k-1}}$ при некотором $k \geqslant 1$. Отсюда и из (25) имеем $|x_{k+1} - x_k| \leqslant (L/(2\mu))(2\mu/L)^2(q^{2^{k-1}})^2 = (2\mu/L)q^{2^k}$. Оценка (26) доказана. Из (26) следует

$$|x_k - x_p| \leqslant \sum_{m=k}^{p-1} |x_{m+1} - x_m| \leqslant \sum_{m=k}^{p-1} \frac{2\mu}{L} q^{2^m} \leqslant \sum_{m=k}^{\infty} \frac{2\mu}{L} q^{2^m} \leqslant \frac{2\mu}{L} q^{2^k}(1 - q^{2^k})^{-1} \tag{27}$$

для всех p, k, $p > k \geqslant 0$. Так как $0 < q < 1$, то правая часть (27) стремится к нулю при $k \to \infty$. Это значит, что последовательность $\{x_k\}$ фундаментальна и сходится к некоторой точке x_*. В силу замкнутости множества X точка $x_* \in X$. Переходя к пределу при $p \to \infty$, из (27) получим оценку (20). Остается убедиться в том, что x_* — точка минимума $f(x)$ на X. Так как $f(x) \in C^2(X)$, то при $k \to \infty$ из (23) имеем $\langle f'(x_*), x - x_*\rangle \geqslant 0$ при всех $x \in X$. Учитывая выпуклость $f(x)$, отсюда и из теоремы 4.2.3 заключаем, что x_* — решение задачи (1). Теорема 2 доказана. □

Из (20) при $k = 0$ имеем $|x_0 - x_*| \leqslant (2\mu/L)q(1-q)^{-1}$. Это неравенство означает, что метод (6) при $X \neq E^n$, так же как и метод (8), который получен из (6) при $X = E^n$, сходится, вообще говоря, лишь при выборе достаточно хорошего начального приближения.

4. Перейдем к рассмотрению метода (2)–(4) с выбором шага $\alpha_k = \lambda^{i_0}$, где i_0 — минимальный номер, для которого выполняется неравенство (10). Этот вариант метода Ньютона будем называть методом (2)–(4), (10). Покажем, что метод (2)–(4), (10) сходится при любом выборе начального приближения и этим выгодно отличается от метода (2)–(4), (5).

Теорема 3. *Пусть X — замкнутое выпуклое множество из E^n, $f(x) \in C^2(X)$ и*

$$\mu|\xi|^2 \leqslant \langle f''(x)\xi, \xi\rangle \leqslant M|\xi|^2, \quad x \in X, \quad \xi \in L_X, \tag{28}$$

где L_X — подпространство, параллельное аффинной оболочке множества X, а μ, M — постоянные, $0 < \mu \leqslant M$. Тогда последовательность $\{x_k\}$, определяемая методом (2)–(4), (10), *при любом начальном приближении $x_0 \in X$ существует и сходится к точке x_* — решению задачи* (1). *Если, кроме того, $f''(x)$ удовлетворяет условию Липшица* (18), *то найдется номер k_0 такой, что в* (4) $\alpha_k = 1$ *при всех $k \geqslant k_0$ и справедлива оценка*

$$|x_k - x_*| \leqslant \frac{2\mu}{L} \sum_{m=k}^{\infty} q^{2^m} \leqslant \frac{2\mu}{L} q^{2^k}(1 - q^{2^k})^{-1}, \quad k \geqslant k_0. \tag{29}$$

Доказательство. Согласно теореме 4.3.4 функция $f(x)$ сильно выпукла. Тогда из теоремы 4.3.1 следует существование и единственность точки $\bar{x}_k$, удовлетворяющей условиям (3). Согласно теореме 4.2.3 тогда $\langle f_k'(\bar{x}_k, x - \bar{x}_k\rangle \geqslant 0$ или

$$\langle f'(x_k) + f''(x_k)(\bar{x}_k - x_k), x - \bar{x}_k\rangle \geqslant 0 \quad \text{при всех} \quad x \in X. \tag{30}$$

Если оказалось, что $\bar{x}_k = x_k$, то из (30) имеем $\langle f'(x_k), x - x_k\rangle \geqslant 0$, $x \in X$. В силу теоремы 4.2.3 и выпуклости $f(x)$ отсюда следует $\bar{x}_k = x_k = x_*$ — задача (1) решена. Поэтому можем считать, что $\bar{x}_k \neq x_k$. Тогда $f_k(\bar{x}_k) < f_k(x_k) = 0$. Покажем, что тогда существует хотя бы один номер $i \geqslant 0$, для которого выполняется условие (10). С этой целью возьмем произвольное число α, $0 \leqslant \alpha \leqslant 1$, и положим $x_\alpha = x_k + \alpha(\bar{x}_k - x_k)$. Отсюда и из выпуклости $f_k(x)$ следует

$$f_k(x_\alpha) \leqslant \alpha f_k(\bar{x}_k) + (1 - \alpha)f_k(x_k) = \alpha f_k(\bar{x}_k) < 0.$$

Тогда из формулы

$$\begin{aligned}f(x_\alpha) - f(x_k) = f_k(x_\alpha) + (\alpha^2/2)\langle (f''(x_k + \theta\alpha(\bar{x}_k - x_k)) - \\ - f''(x_k))(\bar{x}_k - x_k), \bar{x}_k - x_k\rangle, \quad 0 \leqslant \alpha \leqslant 1,\end{aligned} \tag{31}$$

с учетом условий (28) получим

$$\begin{aligned}f(x_\alpha) - f(x_k) &\leqslant f_k(x_\alpha) + (\alpha^2/2)(M - \mu)|\bar{x}_k - x_k|^2 \leqslant \\ &\leqslant \alpha f_k(\bar{x}_k) + (\alpha^2/2)M|\bar{x}_k - x_k|^2, \quad 0 \leqslant \alpha \leqslant 1.\end{aligned} \tag{32}$$

Так как $\bar{x}_k$ — точка минимума сильно выпуклой функции $f_k(x)$ на X, то согласно теореме 4.3.1

$$|x_k - \bar{x}_k|^2 \leqslant (2/\mu)[f_k(x_k) - f_k(\bar{x}_k)] = (2/\mu)|f_k(\bar{x}_k)|. \tag{33}$$

Подставив эту оценку в (32), получим

$$f(x_\alpha) - f(x_k) \leqslant -\alpha|f_k(\bar{x}_k)| + \alpha^2(M/\mu)|f_k(\bar{x}_k)|, \quad 0 \leqslant \alpha \leqslant 1.$$

Возьмем произвольное α, удовлетворяющее условиям

$$0 < \varepsilon_0 = \lambda(1-\varepsilon)\mu/M \leqslant \alpha \leqslant (1-\varepsilon)\mu/M < 1. \tag{34}$$

Отсюда и из предыдущего неравенства будем иметь

$$f(x_k) - f(x_k + \alpha(\bar{x}_k - x_k)) \geqslant \alpha(1 - \alpha(M/\mu))|f_k(\bar{x}_k)| \geqslant \varepsilon\alpha|f_k(\bar{x}_k)| \tag{35}$$

при всех α, удовлетворяющих условиям (34). Возьмем такой номер $m \geqslant 1$, для которого $\lambda^m \leqslant$ $\leqslant (1-\varepsilon)\mu/M < \lambda^{m-1}$. Отсюда следует, что

$$0 < \varepsilon_0 = \lambda(1-\varepsilon)\mu/M < \lambda^m \leqslant (1-\varepsilon)\mu/M. \tag{36}$$

Таким образом, $\alpha = \lambda^m$ удовлетворяет условиям (34) и, следовательно, при $\alpha = \lambda^m$ будет справедливо неравенство (35). Это значит, при $i = m$ выполняется условие (10). Тогда найдется наименьший номер $i = i_0$, $0 \leqslant i_0 \leqslant m$, удовлетворяющий неравенству (10). Приняв в (4) $\alpha_k = \lambda^{i_0}$, получим следующее приближение x_{k+1}.

Тем самым показано, что последовательность $\{x_k\}$ из метода (2)–(4), (10) при любом начальном приближении существует. Из (10) при $i = i_0$ имеем

$$f(x_k) - f(x_{k+1}) \geqslant \varepsilon\alpha_k|f_k(\bar{x}_k)|, \quad k = 0, 1, \ldots$$

Учитывая, что согласно (36) $\alpha_k = \lambda^{i_0} \geqslant \lambda^m > \varepsilon_0$, отсюда получим

$$f(x_k) - f(x_{k+1}) \geqslant \varepsilon\varepsilon_0|f_k(\bar{x}_k)|, \quad k = 0, 1, \ldots \tag{37}$$

Таким образом, $f(x_k) \geqslant f(x_{k+1}) \geqslant f_*$, $k = 0, 1, \ldots$ Тогда существует $\lim\limits_{k\to\infty} f(x_k) \geqslant f_*$ и $\lim\limits_{k\to\infty}(f(x_k) - f(x_{k+1})) = 0$. Из (37) теперь имеем $\lim\limits_{k\to\infty} f(\bar{x}_k) = 0$, а из (33) следует

$$\lim_{k\to\infty} |x_k - \bar{x}_k| = 0. \tag{38}$$

Далее заметим, что согласно (37) последовательность $\{x_k\} \in M(x_0) = \{x\colon x \in X,\ f(x) \leqslant f(x_0)\}$. Для сильно выпуклых непрерывных функций множество $M(x_0)$ выпукло, замкнуто и ограничено. Тогда последовательность $\{x_k\}$ имеет хотя бы одну предельную точку. Пусть v_* — произвольная предельная точка $\{x_k\}$ и пусть $\{x_{k_m}\} \to v_*$. С учетом (38) и условия $f(x) \in C^2(X)$ из (30) при $k = k_m \to \infty$ получим $\langle f'(v_*), x - v_*\rangle \geqslant 0$ для всех $x \in X$. Согласно теореме 4.2.3 тогда $v_* = x_*$ — точка минимума $f(x)$ на X. Следовательно, $\lim\limits_{k\to\infty} f(x_k) = \lim\limits_{m\to\infty} f(x_{k_m}) = f(x_*) = f_*$, т. е. $\{x_k\}$ — минимизирующая последовательность. Отсюда и из теоремы 4.3.1 следует $\{x_k\} \to$ $\to x_*$.

Пусть теперь выполнено условие (18). В силу (38) существует номер k_0 такой, что $(L/\mu)|\bar{x}_k - x_k| \leqslant 1 - \varepsilon$ при всех $k \geqslant k_0$. Из (31) с учетом условия (18) и оценки (33) тогда имеем

$$f(x_\alpha) - f(x_k) \leqslant f_k(x_\alpha) + (\alpha^3/2)L|\bar{x}_k - x_k|^3 \leqslant -\alpha|f_k(\bar{x}_k)| + \alpha^2(L/\mu)|f_k(\bar{x}_k)||\bar{x}_k - x_k|,$$

т. е.

$$f(x_k) - f(x_\alpha) \geqslant |f_k(\bar{x}_k)|\alpha(1 - \alpha(L/\mu)|\bar{x}_k - x_k|) \geqslant \alpha\varepsilon|f_k(\bar{x}_k)|$$

при всех α, $0 \leqslant \alpha \leqslant 1$, $k \geqslant k_0$. В частности, при $\alpha = 1$ отсюда заключаем, что условие (10) выполнено при $i = i_0 = 0$, и, следовательно, $\alpha_k = \lambda^0 = 1$, $x_{k+1} = \bar{x}_k$ при каждом $k \geqslant k_0$. Таким образом, начиная с номера $k = k_0$, метод (2)–(4), (10) превращается в метод (2)–(5) с начальным приближением x_{k_0}, удовлетворяющим условию

$$q = (L/(2\mu))|x_{k_0+1} - x_{k_0}| = (L/(2\mu))|\bar{x}_{k_0} - x_{k_0}| \leqslant (1-\varepsilon)/2 < 1.$$

Отсюда и из теоремы 2 следует оценка (29), что и требовалось. □

Таким образом, метод (2)–(4), (10) ненамного сложнее метода (2)–(5), (10), по скорости сходимости не уступает ему и в то же время не столь чувствителен к выбору начального приближения, как метод (2)–(5). При наличии эффективных методов минимизации квадратичной функции $f_k(x)$ на множестве X метод (2)–(4), (10) можно с успехом применять для минимизации достаточно гладких функций.

Другие теоремы о сходимости описанных выше вариантов метода Ньютона читатель может найти в [603].

5. Существуют и другие модификации метода Ньютона, широко используемые в вычислительной практике. Так, например, в случае $X = E^n$ вместо (8) часто применяют метод Ньютона с переменным шагом:

$$x_{k+1} = x_k - \alpha_k(f''(x_k))^{-1}f'(x_k), \quad k = 0, 1, \ldots \tag{39}$$

Параметр $\alpha_k > 0$ в (39) выбирается из тех же соображений, что и в основном варианте метода Ньютона. Заметим, что приближение x_{k+1} в (39) может быть получено, как решение задачи минимизации:

$$f_k(x, \alpha) \equiv \alpha\langle f'(x_k), x - x_k\rangle + \frac{1}{2}\langle f''(x_k)(x - x_k), x - x_k\rangle \to \inf, \quad x \in E^n \tag{40}$$

при $\alpha = \alpha_k$. Задача (40) подсказывает, как обобщить метод (39) на случай $X \neq E^n$ — тогда приближение x_{k+1} следует определять как решение задачи:

$$f_k(x, \alpha) \to \inf, \quad x \in X \tag{41}$$

при $\alpha = \alpha_k$, или как решение вариационного неравенства

$$\langle f''(x_k)(x_{k+1} - x_k) + \alpha_k f'(x_k), x - x_{k+1}\rangle \geqslant 0 \qquad \forall x \in X. \tag{42}$$

Практика показывает, что, умело выбирая параметр α_k в (39)–(42), можно сделать эти методы менее чувствительными к выбору начального приближения x_0.

Еще раз подчеркнем, что все выше перечисленные варианты метода Ньютона могут быть эффективно использованы лишь тогда, когда матрица вторых производных $f''(x_k)$ легко вычисляется и все последующие вспомогательные задачи решаются достаточно просто. Желание преодолеть возникающие здесь трудности привело к появлению так называемых *квазиньютоновских методов*

$$x_{k+1} = x_k - \alpha_k A_k f'(x_k), \quad \alpha_k > 0, \quad k = 0, 1, \ldots, \tag{43}$$

предназначенных для решения задачи (1) при $X = E^n$. В (43) матрица A_k выбирается из условия

$$\lim_{k\to\infty} \|A_k - (f''(x_k))^{-1}\| = 0. \tag{44}$$

Оказывается, при таком выборе A_k метод (44) также сохраняет высокую скорость сходимости, присущую методу Ньютона. Другое достоинство методов (43) — существуют конструктивные способы построения матриц A_k со свойством (44) на основе достаточно простых рекуррентных соотношений, использующих информацию с предыдущей итерации, обходясь без вычисления и обращения матрицы $f''(x)$. Примером квазиньютоновского метода является *метод Давидона — Флетчера — Пауэлла*, в котором матрицы A_k определяются соотношениями

$$A_{k+1} = A_k + \frac{r_k r_k^\top}{\langle r_k, q_k\rangle} - \frac{(A_k q_k)(A_k q_k)^\top}{\langle A_k q_k, q_k\rangle}, \quad k = 0, 1, \ldots; \quad A_0 = I, \tag{45}$$

где $q_k = f'(x_{k+1}) - f'(x_k)$, $r_k = x_{k+1} - x_k$, а величина α_k находится из условия

$$g_k(\alpha_k) = \min_{\alpha \geqslant 0} g_k(\alpha), \quad g_k(\alpha) = f(x_k - \alpha A_k f'(x_k)). \tag{46}$$

Отметим, что векторы $p_k = A_k f'(x_k)$ удовлетворяют равенствам (9.29), так что метод (39), (45), (46) одновременно является методом сопряженных направлений.

С квазиньютоновскими методами читатель может подробнее познакомиться в [76; 222; 586; 603; 721; 738; 759; 769]. Непрерывные варианты метода Ньютона и его обобщений рассматриваются в следующем параграфе.

§ 11. Непрерывные методы с переменной метрикой

Рассмотрим задачу минимизации

$$f(x) \to \inf, \quad x \in X, \tag{1}$$

где X — выпуклое замкнутое множество, функция $f(x)$ дифференцируема на X. Пусть при каждом $x \in E^n$ определена квадратная матрица $G(x)$ n-го порядка, симметричная, положительно определенная. Всякая такая матрица $G(x)$ в E^n задает новое скалярное произведение $\langle y, z\rangle_G = \langle G(x)y, z\rangle$ $\forall y, z \in E^n$, соответствующую норму $|y|_G = \sqrt{\langle G(x)y, y\rangle}$ и порождает переменную (зависящую от x) метрику $\rho(y, z) = |y-z|_G = \sqrt{\langle G(x)(y-z), y-z\rangle}$, которую мы кратко будем называть *G-метрикой*.

Определение 1. Точку $w \in X$ будем называть *G-проекцией* точки $z \in E^n$, если

$$|w - z|_G = \inf_{y \in X} |y - z|_G.$$

и будем обозначать через $w = \mathcal{P}_X^{G(x)}(z)$.

Иначе, G-проекция точки z является решением задачи минимизации

$$\varphi(y) = \tfrac{1}{2}\langle G(x)(y-z), y-z\rangle \to \inf, \quad y \in X. \tag{2}$$

Так как $\varphi''(y) = G(x) > 0$, то функция $\varphi(y)$ сильно выпукла на E^n и на выпуклом замкнутом множестве X достигает своей нижней грани, притом в единственной точке w (теорема 4.3.1). Как вытекает из (2) и теоремы 4.2.3 точка w будет G-проекцией точки $z \in E^n$ тогда и только тогда, когда

$$\langle \varphi'(w), y - w\rangle = \langle G(x)(w-z), y-w\rangle \geqslant 0 \quad \forall y \in X. \tag{3}$$

Если $G(x) = I_n$ — единичная матрица, то G-проекция превращается в обычную проекцию точки на множество (см. § 4.4). По аналогии с теоремой 4.4.4 с помощью G-проекции можно сформулировать критерий оптимальности для выпуклых задач (1). Справедлива

Теорема 1. *Пусть X — выпуклое замкнутое множество, X_* — множество точек минимума функции $f(x)$ на X. Если $x_* \in X_*$ и функция $f(x)$ дифференцируема в точке x_*, то необходимо выполняется равенство*

$$x_* = \mathcal{P}_X^{G(x_*)}(x_* - \alpha(G(x_*))^{-1} f'(x_*)) \quad \forall \alpha > 0. \tag{4}$$

Если, кроме того, $f(x)$ выпукла на X, то всякая точка x_*, удовлетворяющая уравнению (4), принадлежит X_*.

Доказательство. В силу (3) равенство (4) эквивалентно неравенству

$$\langle G(x_*)[x_* - (x_* - \alpha(G(x_*))^{-1} f'(x_*)], y - x_*\rangle \geqslant 0 \quad \forall y \in X,$$

или

$$\langle \alpha f'(x_*), y - x_*\rangle \geqslant 0 \quad \forall y \in X.$$

Поскольку $\alpha > 0$, то отсюда имеем:

$$\langle f'(x_*), y - x_*\rangle \geqslant 0 \quad \forall y \in X \tag{5}$$

Так как проведенные выкладки обратимы, то вариационные неравенства (4) и (5) равносильны. Отсюда и из теоремы 4.2.3 следует утверждение теоремы 1. □

Рассмотрим систему дифференциальных уравнений [28]:

$$\dot{x}(t) = \mathcal{P}_X^{G(x(t))}(x(t) - \alpha(t)(G(x(t)))^{-1} f'(x(t))) - x(t), \quad t \geqslant 0, \tag{6}$$

где $\alpha(t) > 0$ — заданная функция. Согласно теореме 1 решение x_* задачи (1) удовлетворяет уравнению (4) при $\alpha = \alpha(t) > 0$ $\forall t \geqslant 0$. Это значит, что каждая точка $x_* \in X_*$ является точкой равновесия (стационарным решением) системы (6). Можно надеяться, что при некоторых ограничениях на функции $f(x)$, $\alpha(t)$, матрицу $G(x)$ траектория $x(t)$ системы (6) при больших t приближается ко множеству X_*. Непрерывный метод с переменной метрикой описан.

Если $X = E^n$, то $\mathcal{P}_X^{G(x)}(z) = z$ $\forall z \in E^n$, и (6) превращается в систему

$$\dot{x}(t) = -\alpha(t)(G(x(t)))^{-1} f'(x(t)), \quad t \geqslant 0. \tag{7}$$

Если $X \neq E^n$, то уравнение (6) эквивалентно вариационному неравенству, которое вытекает из (3):

$$\langle G(x(t))[\dot{x}(t) + x(t) - (x(t) - \alpha(t)(G(x(t)))^{-1} f'(x(t)))], y - (\dot{x}(t) + x(t))\rangle \geqslant 0$$

или

$$\langle G(x(t))\dot{x}(t) + \alpha(t) f'(x(t)), y - \dot{x}(t) - x(t)\rangle \geqslant 0, \quad \forall y \in X, \quad \forall t \geqslant 0. \tag{8}$$

Из (6), (7) видно, что при $G(x) \equiv I_n$ метод (6) превращается в непрерывный вариант градиентного метода (1.45) или метода проекции градиента (2.34). Посмотрим, что будет, если $G(x) \equiv f''(x)$, предполагая, что $f(x) \in C^2(E^n)$ и сильно выпукла на E^n. В случае $X = E^n$ из (7) имеем:

$$\dot{x}(t) = -\alpha(t)(f''(x(t)))^{-1} f'(x(t)), \quad t \geqslant 0. \tag{9}$$

Нетрудно видеть, что метод (10.39) является разностным аналогом (схема Эйлера) метода (9), а классический метод Ньютона (10.8) — это разностный аналог метода (9) при $\alpha(t) = \alpha$ [25]. Если $X \neq E^n$, то из (8) при $G(x) \equiv f''(x)$ получим вариационное неравенство

$$\langle f''(x(t))\dot{x}(t) + \alpha(t) f'(x(t)), y - \dot{x}(t) - x(t)\rangle \geqslant 0, \quad \forall y \in X, \quad \forall t \geqslant 0. \tag{10}$$

Неравенство (10.42) можно истолковать как разностный аналог неравенства (10). Как видим, метод (6) при $G(x) = f''(x)$ является непрерывным аналогом метода Ньютона. Поэтому можно ожидать, что если в (6) матрицу $G(x)$ выбирать близкой к $f''(x)$, то на этом пути удастся получить непрерывные аналоги квазиньютоновских методов, также имеющих высокую скорость сходимости, хорошо приспособленных для минимизации овражных функций. Следует сказать, что проблема конструктивного выбора матрицы $G(x)$ в методе (6) пока еще мало изучена. Приведем теорему сходимости метода (6).

Теорема 2. *Пусть X — выпуклое замкнутое множество, функция $f(x) \in C^1(E^n)$ и выпукла на E^n, $f_* > -\infty$, $X_* \neq \varnothing$; функция $\alpha(t) \geqslant \alpha_0 > 0$, непрерывно дифференцируема и $\alpha'(t) \leqslant 0$ $\forall t \geqslant 0$; матрица $G(x)$ симметрична, и существует сильно выпуклая функция $\psi(x) \in C^2(E^n)$ такая, что $\psi''(x) = G(x)$ $\forall x \in E^n$. Пусть траектория системы (6) с начальным условием $x(0) = x_0$ определена при всех $t \geqslant 0$. Тогда существует точка $v_* = v_*(x_0) \in X_*$ такая, что*

$$\lim_{t\to\infty} x(t) = v_*, \quad \lim_{t\to\infty} f(x(t)) = f_*, \quad \lim_{t\to\infty} \dot{x}(t) = 0.$$

Доказательство. Примем в (5) $y = \dot{x}(t) + x(t) \in X$, умножим на $\alpha(t) > 0$ и сложим с (8) при $y = x_*$:

$$\langle G(x(t))\dot{x}(t) + \alpha(t)(f'(x(t)) - f'(x_*)), x_* - \dot{x}(t) - x(t)\rangle \geqslant 0$$

или

$$\langle G(x(t))\dot{x}(t), \dot{x}(t)\rangle + \langle G(x(t))\dot{x}(t), x(t) - x_*\rangle + \alpha(t)\langle f'(x(t)) - f'(x_*), \dot{x}(t)\rangle + \\ + \alpha(t)\langle f'(x(t)) - f'(x_*), x(t) - x_*\rangle \leqslant 0 \quad \forall t \geqslant 0, \ \forall x_* \in X_*. \tag{11}$$

По условию теоремы существует функция $\psi(x)$, для которой $\psi''(x) = G(x)$ $\forall x \in E^n$, поэтому

$$\frac{d}{dt}(\psi(x_*) - \psi(x(t)) + \langle \psi'(x(t)), x(t) - x_*\rangle) = \langle \psi''(x(t))\dot{x}(t), x(t) - x_*\rangle = \langle G(x(t))\dot{x}(t), x(t) - x_*\rangle.$$

Кроме того

$$\frac{d}{dt}(f(x(t)) - f(x_*) - \langle f'(x_*), x(t) - x_*\rangle) = \langle f'(x(t)) - f'(x_*), \dot{x}(t)\rangle, \quad \alpha(t) > 0,$$

и

$$\langle f'(x(t)) - f'(x_*), x(t) - x_*\rangle \geqslant 0 \tag{12}$$

в силу выпуклости $f(x)$ (теорема 4.2.4). С учетом этих соотношений из (11) имеем

$$\langle G(x(t))\dot{x}(t),\dot{x}(t)\rangle+\frac{d}{dt}(\psi(x_*)-\psi(x(t))+\langle\psi'(x(t)),x(t)-x_*\rangle)+$$
$$+\alpha(t)\frac{d}{dt}(f(x(t))-f(x_*)-\langle f'(x_*),x(t)-x_*\rangle)\leqslant 0\quad\forall t\geqslant 0,\quad\forall x_*\in X_*.$$

Интегрируя это неравенство на произвольном отрезке $[\tau,t]$, получим

$$\int_\tau^t\langle G(x(s))\dot{x}(s),\dot{x}(s)\rangle ds+(\psi(x_*)-\psi(x(s))+\langle\psi'(x(s)),x(s)-x_*\rangle)\Big|_{s=\tau}^{s=t}+$$
$$+\alpha(t)(f(x(s))-f(x_*)-\langle f'(x_*),x(s)-x_*\rangle)\Big|_{s=\tau}^{s=t}-$$
$$-\int_\tau^t\alpha'(s)[f(x(s))-f(x_*)-\langle f'(x_*),x(s)-x_*\rangle]ds\leqslant 0\quad\forall t>\tau\geqslant 0,\quad\forall x_*\in X_*.\qquad(13)$$

Из выпуклости $f(x)$ (теорема 4.2.2) и сильной выпуклости $\psi(x)$ (теоремы 4.3.2 и 4.3.4) следует

$$f(x(t))-f(x_*)-\langle f'(x_*),x(t)-x_*\rangle\geqslant 0\quad\forall t\geqslant 0,$$
$$\psi(x_*)-\psi(x(t))+\langle\psi'(x(t)),x(t)-x_*\rangle\geqslant\varkappa|x(t)-x_*|^2,\qquad(14)$$
$$\langle G(x(t))\dot{x}(t),\dot{x}(t)\rangle\geqslant\varkappa|\dot{x}(t)|^2\quad\forall t\geqslant 0.$$

Отсюда и из (13) с учетом $\alpha(t)>0$, $\alpha'(t)\leqslant 0$ имеем

$$\varkappa\int_\tau^t|\dot{x}(s)|^2ds+\varkappa|x(t)-x_*|^2\leqslant\psi(x_*)-\psi(x(\tau))+\langle\psi'(x(\tau)),x(\tau)-x_*\rangle+\alpha(\tau)f(x(\tau))-$$
$$-f(x_*)-\langle f'(x_*),x(\tau)-x_*\rangle\equiv\nu(\tau,x_*)\quad\forall t>\tau\geqslant 0,\quad\forall x_*\in X_*.\qquad(15)$$

Это означает, что $|x(t)-x_*|^2\leqslant\nu(0,x_*)/\varkappa\ \forall t\geqslant 0$ и $\int_0^\infty|\dot{x}(t)|^2dt\leqslant\nu(0,x_*)/\varkappa$. Поэтому существует последовательность $\{t_i\}\to\infty$ такая, что $\{x(t_i)\}\to v_*$, $\{\dot{x}(t_i)\}\to 0$. Так как множество X замкнуто, $\dot{x}(t)+x(t)\in X\ \forall t\geqslant 0$, то $\lim\limits_{i\to\infty}(\dot{x}(t_i)+x(t_i))=v_*\in X$. Положим в (8) $t=t_i$; при $i\to\infty$ с учетом $\lim\limits_{i\to\infty}\alpha(t_i)=\alpha(\infty)\geqslant\alpha_0>0$ получим $\alpha(\infty)\langle f'(v_*),y-v_*\rangle\geqslant 0\ \forall y\in X$. Согласно теореме 4.2.3 тогда $v_*\in X_*$. Из (15) при $\tau=t_i$, $x_*=v_*$ следует: $|x(t)-v_*|^2\leqslant\nu(t_i,v_*)/\varkappa\ \forall t\geqslant 0$. Переходя здесь к пределу сначала при $t\to+\infty$, затем при $t_i\to\infty$, имеем $\lim\limits_{t\to\infty}x(t)=v_*$. Тогда $\lim\limits_{t\to\infty}f(x(t))=f(v_*)=f_*$, $\lim\limits_{t\to\infty}f'(x(t))=f'(v_*)$. Наконец, из (11) при $x_*=v_*$ с учетом (12), (14) получим: $\varkappa|\dot{x}(t)|^2\leqslant\|G(x(t))\|\|\dot{x}(t)\|x(t)-v_*|+\alpha(t)|f'(x(t))-f'(v_*)\|\dot{x}(t)|$ или $\varkappa|\dot{x}(t)|\leqslant\|G(x(t))\|\|x(t)-v_*|+\alpha(t)|f'(x(t))-f'(v_*)|\ \forall t\geqslant 0$. Отсюда при $t\to\infty$ следует, что $\lim\limits_{t\to\infty}\dot{x}(t)=0$. Теорема 2 доказана. □

В заключение заметим, что метод (6), основанный на изменяющейся вдоль траектории $x(t)$ G-метрике, можно рассматривать как непрерывный аналог большой группы итерационных методов, которые в литературе принято называть методами «с растяжением пространства» или методами с переменной метрикой [76; 273; 586; 738; 769].

§ 12. Метод покоординатного спуска

В предыдущих параграфах мы рассмотрели методы, которые для своей реализации требуют вычисления первых или вторых производных минимизируемой функции. Однако в практических задачах нередко встречаются случаи, когда минимизируемая функция либо не обладает нужной гладкостью, либо является гладкой, но вычисление ее производных с нужной точностью требует слишком большого объема работ, много машинного времени. В таких случаях желательно иметь методы минимизации, которые требуют лишь вычисления значения функции. Одним из таких методов является метод покоординатного спуска [74; 374; 753].

1. Сначала опишем этот метод для задачи

$$f(x)\to\inf;\quad x\in X=E^n.\qquad(1)$$

Обозначим $e_i=(0,\ldots,0,1,0,\ldots,0)$ — единичный координатный вектор, у которого i-я координата равна 1, остальные равны нулю, $i=1,\ldots,n$. Пусть x_0 — некоторое начальное приближение, а α_0 — некоторое положительное число, являющееся параметром метода. Допустим, что нам уже известны точка $x_k\in E^n$ и число $\alpha_k>0$ при каком-либо $k\geqslant 0$. Положим:

$$p_k=e_{i_k},\quad i_k=k-n\left[\frac{k}{n}\right]+1,\qquad(2)$$

где $\left[\frac{k}{n}\right]$ означает целую часть числа k/n. Условие (2) обеспечивает циклический перебор координатных векторов $e_1,e_2,\ldots,e_n$, т. е.

$$p_0=e_1,\ \ldots,\ p_{n-1}=e_n,\ p_n=e_1,\ \ldots,\ p_{2n-1}=e_n,\ p_{2n}=e_1,\ \ldots$$

Вычислим значение функции $f(x)$ в точке $x=x_k+\alpha_kp_k$ и проверим неравенство

$$f(x_k+\alpha_kp_k)<f(x_k).\qquad(3)$$

Если (3) выполняется, то примем

$$x_{k+1}=x_k+\alpha_kp_k,\quad\alpha_{k+1}=\alpha_k.\qquad(4)$$

В том случае, если (3) не выполняется, вычисляем значение функции $f(x)$ в точке $x=x_k-\alpha_kp_k$ и проверяем неравенство

$$f(x_k-\alpha_kp_k)<f(x_k).\qquad(5)$$

В случае выполнения (5) положим

$$x_{k+1}=x_k-\alpha_kp_k,\quad\alpha_{k+1}=\alpha_k.\qquad(6)$$

Назовем $(k+1)$-ю итерацию удачной, если справедливо хотя бы одно из неравенств (3) или (5). Если $(k+1)$-я итерация неудачная, т. е. не выполняются оба неравенства (3) и (5), то полагаем

$$x_{k+1}=x_k,\quad\alpha_{k+1}=\begin{cases}\lambda\alpha_k, & i_k=n,\ x_k=x_{k-n+1},\\ \alpha_k, & i_k\neq n\text{ или }x_k\neq x_{k-n+1},\\ & \text{или }0\leqslant k\leqslant n-1;\end{cases}\qquad(7)$$

здесь λ, $0<\lambda<1$ — фиксированное число, являющееся параметром метода. Условия (7) означают, что если за один цикл из n итераций при переборе направлений всех координатных осей $e_1,\ldots,e_n$ с шагом α_k реализовалась хотя бы одна удачная итерация, то длина шага α_k не дробится и сохраняется на протяжении по крайней мере следующего цикла из n итераций. Если же среди последних n итераций не оказалось ни одной удачной итерации, то шаг α_k дробится. Таким образом, если на итерации с номером $k=k_m$ произошло дробление α_k, то

$$f(x_{k_m}+\alpha_{k_m}e_i)\geqslant f(x_{k_m}),\quad f(x_{k_m}-\alpha_{k_m}e_i)\geqslant f(x_{k_m})\qquad(8)$$

при всех $i=1,\ldots,n$. Метод покоординатного спуска для задачи (1) описан. Справедлива

Теорема 1. *Пусть функция $f(x)$ выпукла на E^n и принадлежит классу $C^1(E^n)$, а начальное приближение x_0 таково, что множество $M(x_0)=\{x\in E^n\colon f(x)\leqslant f(x_0)\}$ ограничено. Тогда последовательность $\{x_k\}$, получаемая описанным методом (2)–(7), минимизирует функцию $f(x)$ на E^n и сходится ко множеству X_*.*

Доказательство. Согласно теореме 2.1.2 имеем $f_*>-\infty$, $X_*\neq\varnothing$. Из описания метода (2)–(7) следует, что $f(x_{k+1})\leqslant f(x_k)$, $k=0,1,\ldots$, так что $\{x_k\}\in M(x_0)$ и существует $\lim\limits_{k\to\infty} f(x_k)\geqslant f_*$. Покажем, что найдется бесконечно много номеров $k_1,\ldots,k_m,\ldots$ итераций, на которых шаг α_k дробится, и поэтому $\lim\limits_{k\to\infty}\alpha_k=0$. Допустим противное: пусть процесс дробления конечен, т. е. $\alpha_k=\alpha>0$ при всех $k\geqslant N$. Обозначим $M_\alpha=\{u\colon u=x_N+\alpha re_i\in M(x_0),\ i=1,\ldots,n,\ r=0,\pm1,\pm2,\ldots\}$ — сетку (решетку) с шагом α. Из описания метода покоординатного спуска при $\alpha_k=\alpha$, $k\geqslant N$, следует, что начиная с номера N все последующие циклы из n итераций будут содержать хотя бы одну удачную итерацию, и на каждой удачной итерации будет происходить переход от одной точки сетки M_α к другой соседней точке этой сетки. По определению удачной итерации переход от точки к точке сопровождается строгим уменьшением значения функции $f(x)$, поэтому каждая точка сетки M_α будет просматриваться не более одного раза. Но множество $M(x_0)$ по условию ограничено, и поэтому сетка M_α состоит из конечного числа точек. Следовательно, процесс перебора точек этой сетки закончится через конечное число итераций определением точки x_{k_m}, $k_m>N$, для которой выполняются неравенства (8) при всех $i=1,\ldots,n$. А тогда вопреки допущению придется дробить число $\alpha_k=\alpha$. Полученное противоречие показывает, что процесс дробления α_k бесконечен и $\lim\limits_{k\to\infty}\alpha_k=0$.

Пусть $k_1<k_2<\ldots<k_m<\ldots$ — номера тех итераций, на которых длина шага α_k дробится и выполняются неравенства (8). Так как последовательность $\{x_k\}$ принадлежит ограниченному множеству $M(x_0)$, то из $\{x_{k_m}\}$ можно выбрать сходящуюся подпоследовательность. Без умаления общности можем считать, что сама последовательность $\{x_{k_m}\}$ сходится к некоторой точке x_*. С помощью формулы конечных приращений из (8) имеем

$$\langle f'(x_{k_m}+\theta_m\alpha_{k_m}e_i),e_i\rangle\alpha_{k_m}\geqslant 0,\quad \langle f'(x_{k_m}-\bar\theta_m\alpha_{k_m}e_i),e_i\rangle(-\alpha_{k_m})\geqslant 0,$$

откуда

$$f_{x^i}(x_{k_m}+\theta_m\alpha_{k_m}e_i)\geqslant 0,\quad f_{x^i}(x_{k_m}-\bar\theta_m\alpha_{k_m}e_i)\leqslant 0,$$

$0\leqslant\theta_m,\bar\theta_m\leqslant 1$ при всех $i=1,\ldots,n$ и $m=1,2,\ldots$ Пользуясь тем, что $f(x)\in C^1(E^n)$ и $\lim\limits_{m\to\infty}\alpha_{k_m}=0$, отсюда получим $f_{x^i}(x_*)=0$, $i=1,\ldots,n$, т. е. $f'(x_*)=0$. В силу выпуклости $f(x)$ тогда $x_*\in X_*$. Следовательно, $\lim\limits_{k\to\infty} f(x_k)=\lim\limits_{m\to\infty} f(x_{k_m})=f(x_*)=f_*$. Таким образом, последовательность $\{x_k\}$ является минимизирующей. Отсюда и из теоремы 2.1.2 следует, что $\rho(x_k,X_*)\to 0$ при $k\to\infty$. Теорема доказана. □

Заметим, что хотя метод (2)–(7) для своей реализации не требует знания градиента минимизируемой функции, однако в условии теоремы 1 содержится требование гладкости этой функции. Оказывается, если функция $f(x)$ не является гладкой, то метод покоординатного спуска может не сходиться ко множеству решений задачи (1). Об этом говорит следующий

Пример 1. Пусть

$$f(u)=x^2+y^2-2(x+y)+2|x-y|,\quad u=(x,y)\in E^2.$$

Нетрудно проверить, что $f(u)$ сильно выпукла на E^2 и, следовательно, ограничена снизу и достигает своей нижней грани на E^2 в единственной точке. Возьмем в качестве начального приближения точку $u_0=(0,0)$. Тогда имеем $f(u_0+\alpha e_1)=f(\alpha e_1)=\alpha^2-2\alpha+2|\alpha|\geqslant 0=f(0)$, $f(u_0+\alpha e_2)=f(\alpha e_2)=\alpha^2-2\alpha+2|\alpha|\geqslant 0=f(0)$ при всех действительных α. Отсюда следует, что все итерации метода (2)–(7) при начальной точке $u_0=(0,0)$ и любом выборе начального параметра $\alpha=\alpha_0>0$ будут неудачными, т. е. $u_k=u_0$ при всех $k=0,1,\ldots$ Однако в точке $u_0=(0,0)$ функция $f(u)$ не достигает своей нижней грани на E^2: например, в точке $v=(1,1)$ имеем $f(v)=-2<f(u_0)=0$.

2. Описанный выше метод покоординатного спуска нетрудно модифицировать применительно к задаче минимизации функции на параллелепипеде:

$$f(x)\to\inf;\quad x\in X=\{(x^1,\ldots,x^n)\colon a_i\leqslant x^i\leqslant b_i,\ i=1,\ldots,n\},$$

где a_i, b_i — заданные числа, $a_i<b_i$, $i=1,\ldots,n$. А именно, пусть k-е приближение $x_k\in X$ и число $\alpha_k>0$ при некотором $k\geqslant 0$ уже найдены. Выберем вектор $p_k=e_{i_k}$ согласно формуле (2), составим точку $x_k+\alpha_k p_k$ и проверим условия

$$x_k+\alpha_k p_k\in X,\quad f(x_k+\alpha_k p_k)<f(x_k). \tag{10}$$

Если оба условия (10) выполняются, то следующее приближение x_{k+1}, α_{k+1} определяем по формулам (4). Если же хотя бы одно условие (10) не выполняется, то составляем точку $x_k-\alpha_k p_k$ и проверяем условия

$$x_k-\alpha_k p_k\in X,\quad f(x_k-\alpha_k p_k)<f(x_k). \tag{11}$$

В случае выполнения обоих условий (11) следующее приближение определяем по формулам (6), а если хотя бы одно из условий (11) не выполняется, то следующее приближение находится по формулам (7).

Теорема 2. *Пусть функция $f(x)$ выпукла на X и $f(x)\in C^1(X)$. Тогда при любом выборе начальных $x_0\in X$ и $\alpha_0>0$ последовательность $\{x_k\}$, получаемая методом (10), (4), (11), (6), (7), минимизирует функцию $f(x)$ на X и сходится ко множеству решений задачи (9).*

Доказательство. Так как X — параллелепипед, то множество $M(x_0)=\{x\colon x\in X,\ f(x)\leqslant f(x_0)\}$ ограничено. Так как $f(x_{k+1})\leqslant f(x_k)$, $k=0,1,\ldots$, то $\{x_k\}\in X$ и существует $\lim\limits_{k\to\infty} f(x_k)\geqslant f_*$. Так же, как в теореме 1, доказывается существование бесконечного числа номеров $k_1<\ldots<k_m<\ldots$ итераций, на которых длина шага α_k дробится, и поэтому $\lim\limits_{k\to\infty}\alpha_k=0$. В силу ограниченности $M(x_0)$ из $\{x_{k_m}\}$ можно выбрать сходящуюся подпоследовательность. Не умаляя общности, можем считать, что $\{x_{k_m}\}\to x_*=(x_*^1,\ldots,x_*^n)$. При каждом $i=1,\ldots,n$ возможны следующие три случая.

1) $a_i<x_*^i<b_i$. Так как $\lim\limits_{k\to\infty}\alpha_k=0$, то найдется номер N такой, что $x_{k_m}+\alpha_{k_m}e_i\in X$ и $x_{k_m}-\alpha_{k_m}e_i\in X$ при всех $m\geqslant N$. Поскольку α_k при $k=k_m$ дробится, то

$$f(x_{k_m}+\alpha_{k_m}e_i)\geqslant f(x_{k_m}),\quad f(x_{k_m}-\alpha_{k_m}e_i)\geqslant f(x_{k_m})$$

для всех $m \geqslant N$. Отсюда, как и в теореме 1, получаем $f_{x^i}(x_*) = 0$, так что

$$f_{x^i}(x_*)(x^i - x_*^i) = 0, \quad a_i \leqslant x^i \leqslant b_i.$$

2) $x_*^i = a_i$. Тогда $x_{k_m} + \alpha_{k_m} e_i \in X$ и $f(x_{k_m} + \alpha_{k_m} e_i) \geqslant f(x_{k_m})$ при всех $m \geqslant N$. Следовательно, $\langle f'(x_{k_m} + \theta_m \alpha_{k_m} e_i), e_i \rangle \alpha_{k_m} \geqslant 0$ или $f_{x^i}(x_{k_m} + \theta_m \alpha_{k_m} e_i) \geqslant 0$ для каждого $m \geqslant N$. Отсюда при $m \to \infty$ получим $f_{x^i}(x_*) \geqslant 0$ или $f_{x^i}(x_*)(x^i - a_i) = f_{x^i}(x_*)(x^i - x_*^i) \geqslant 0$, $a_i \leqslant x^i \leqslant b_i$.

3) $x_*^i = b_i$. Тогда $x_{k_m} - \alpha_{k_m} e_i \in X$ и $f(x_{k_m} - \alpha_{k_m} e_i) \geqslant f(x_{k_m})$ при всех $m \geqslant N$. Поэтому $\langle f(x_{k_m} - \bar{\theta}_m \alpha_{k_m} e_i), e_i \rangle (-\alpha_{k_m}) \geqslant 0$ или $f_{x^i}(x_{k_m} - \bar{\theta}_m \alpha_{k_m} e_i) \leqslant 0$, $m \geqslant N$. Отсюда при $m \to \infty$ получим $f_{x^i}(x_*) \leqslant 0$, следовательно,

$$f_{x^i}(x_*)(x^i - b_i) = f_{x^i}(x_*)(x^i - x_*^i) \geqslant 0, \quad a_i \leqslant x^i \leqslant b_i.$$

Объединяя все три рассмотренных случая, заключаем, что

$$f_{x^i}(x_*)(x^i - x_*^i) \geqslant 0, \quad a_i \leqslant x^i \leqslant b_i, \quad i = 1, \ldots, n.$$

Суммируя эти неравенства по всем $i = 1, \ldots, n$, получим

$$\langle f'(x_*), x - x_* \rangle \geqslant 0 \quad \text{для всех} \quad x \in X.$$

Согласно теореме 4.2.3 тогда $x_* \in X_*$. Следовательно, $\lim\limits_{k \to \infty} f(x_k) = \lim\limits_{m \to \infty} f(x_{k_m}) = f(x_*) = f_*$, т. е. $\{x_k\}$ — минимизирующая последовательность. Отсюда и из теоремы 2.1.2 следует, что $\lim\limits_{k \to \infty} \rho(x_k, X_*) = 0$. Теорема 2 доказана. □

3. Существуют и другие варианты метода покоординатного спуска. Можно, например, строить последовательность $\{x_k\}$ по правилу

$$x_{k+1} = x_k + \alpha_k p_k, \tag{12}$$

где p_k определяется согласно (2), а α_k — условиями

$$\alpha_k \geqslant 0, \quad g_k(\alpha_k) = \min_{-\infty < \alpha < +\infty} g_k(\alpha), \quad g_k(\alpha) = f(x_k + \alpha p_k). \tag{13}$$

Метод (12), (13) имеет смысл применять в том случае, когда величина α_k из (13) находится в явном виде. Так будет, если функция $f(x)$ — квадратичная, т. е.

$$f(x) = \tfrac{1}{2}\langle Ax, x \rangle - \langle b, x \rangle, \quad x \in E^n, \tag{14}$$

где A — симметричная положительно определенная матрица, $b \in E^n$. Нетрудно убедиться, что для функции (14) метод (12), (13) приводит к хорошо известному методу Зейделя из линейной алгебры [74; 89].

Хотя и скорость сходимости метода покоординатного спуска, вообще говоря, невысокая, благодаря простоте каждой итерации, скромным требованиям к гладкости минимизируемой функции этот метод довольно широко применяется на практике.

Существуют и другие методы минимизации, использующие лишь значения функции и не требующие для своей реализации вычисления производных. Например, используя вместо производных их разностные аппроксимации, можно построить модификации рассматривавшихся в предыдущих параграфах методов, требующие вычисления лишь значений функции в подходящим образом выбранных точках.

Другой подход для минимизации негладких функций, основанный лишь на вычислении значений функции, дает метод случайного поиска, который будет рассмотрен ниже в § 19. Метод поиска глобального минимума, излагаемый в следующем параграфе, также относится к методам, не требующим вычисления производных минимизируемой функции.

Упражнения

1. Нарисуйте линии уровня $f(x) = C = \text{const}$ функции из примера 1 и поясните причину расходимости метода покоординатного спуска для этой функции при выборе $u_0 = (0, 0)$.

2. Опишите метод покоординатного спуска и докажите его сходимость для случая, когда в задаче (9) $a_i = -\infty$ или $b_j = \infty$ для каких-либо i, j, $1 \leqslant i, j \leqslant n$.

3. Докажите сходимость метода (12), (13) для функции (14) [74].

§ 13. Метод покрытия в многомерных задачах

Опишем еще один метод минимизации, основанный лишь на вычислении значений целевой функции без привлечения значений каких-либо ее производных. Речь пойдет о методах покрытий, одномерный вариант которых был изложен в § 1.7. Эти методы служат для минимизации функций, удовлетворяющих условию Липшица. Заметим, что такие задачи в общем случае являются *многоэкстремальными*, т. е. в них могут существовать точки локального минимума, отличные от глобального минимума. Большинство методов, описанных выше в этой главе, при их применении к многоэкстремальным задачам скорее всего нам помогут найти лишь какую-либо точку локального минимума, расположенную поблизости от начальной точки. Поэтому эти методы часто называют *локальными методами*. На практике для решения многоэкстремальных задач локальные методы обычно используются следующим образом: на множестве задают некоторую сетку точек и, выбирая в качестве начального приближения точки этой сетки, с помощью того или иного локального метода находят локальные минимумы функции, а затем, сравнивая полученные результаты, определяют ее глобальный минимум. Однако ясно, что такой подход к решению многоэкстремальных задач весьма трудоемок и не всегда приводит к цели. Поэтому представляют большой интерес методы поиска глобального минимума в многоэкстремальных задачах.

Перейдем к изложению одного из методов покрытий, которые служат для решения многоэкстремальных задач с целевой функцией, удовлетворяющей условию Липшица. Ограничимся рассмотрением задачи минимизации на параллелепипеде:

$$f(x) \to \inf, \quad x \in \Pi = \{x = (x^1, \ldots, x^n) : a_i \leqslant x^i \leqslant b_i, \ i = 1, \ldots, n\}, \tag{1}$$

где a_i, b_i — заданные числа, $a_i < b_i$, а функция $f(x)$ удовлетворяет условию

$$|f(x) - f(y)| \leqslant L|x - y|_\infty \quad \forall x, y \in \Pi, \tag{2}$$

где $L = \text{const} > 0$, $|x - y|_\infty = \max\limits_{1 \leqslant i \leqslant n} |x^i - y^i|$. В правой части (2) можно было поставить любую другую норму $|x - y|_p$, $1 \leqslant p < \infty$, например, евклидову норму $|x - y|$, как это мы неоднократно делали выше, когда требовали условие Липшица от функции или ее производных. В силу эквивалентности норм в

E^n условие Липшица в любой норме может быть сведено к виду (2). А норма $|x-y|_\infty$ здесь привлекает нас тем, что такие множества, как параллелепипед, куб удобно описывать с помощью именно такой нормы. Так, например, множество $\{x \in E^n: |x-x_0|_\infty \leqslant h/2\} = \{x \in E^n: |x^i - x_0^i| \leqslant h/2,\ i=1,\ldots,n\}$ представляет собой куб с центром в точке x_0, ребром длины h и с гранями, параллельными осям координат. Именно такими кубами мы будем покрывать параллелепипед Π. Кроме того, использование нормы $|\cdot|_\infty$ позволит нам изложить многомерный вариант метода покрытий для решения задачи (1) также просто, как в одномерном случае (см. § 1.7, п. 4).

На параллелепипеде Π введем сетку Π_h, состоящую из точек $x_{i_1\ldots i_n} = (x_{i_1}^1, x_{i_2}^2, \ldots, x_{i_j}^j, \ldots, x_{i_n}^n)$, j-я координата $x_{i_j}^j$ которых при каждом $j=1,\ldots,n$ образована по правилу (ср. с § 1.7):

$$x_1^j = a_j + \frac{h}{2},\ x_2^j = x_1^j + h,\ \ldots,\ x_{i+1}^j = x_i^j + h,\ \ldots,$$
$$x_{m_j-1}^j = x_1 + (m_j-2)h,\ x_{m_j} = \min\{x_1 + (m_j-1)h;\ b_j\},$$

где $h = \frac{2\varepsilon}{L}$ — шаг метода, а натуральное число m_j определяется условием $x_{m_j-1} < b_j - \frac{h}{2} \leqslant x_1^j + (m_j-1)h$. В качестве приближения нижней грани f_* в задаче (1), можно взять величину $\min\limits_{\Pi_h} f(x_{i_1\ldots i_n}) = F_h$, которую можно найти с помощью простого перебора всех значений функции $f(x)$ по точкам сетки Π_h. Справедлива

Теорема 1. *Для любой функции $f(x)$, удовлетворяющей условию* (2), *справедлива оценка*

$$f_* \leqslant F_h \leqslant f_* + \varepsilon. \tag{3}$$

Доказательство. Кубы $\Pi_{i_1\ldots i_n} = \{x \in E^n: |x - x_{i_1\ldots i_n}|_\infty \leqslant h/2\}$ с центрами $x_{i_1\ldots i_n} \in \Pi_h$ покрывают весь параллелепипед Π. Это означает, что для любой точки $x \in \Pi$ найдется куб $\Pi_{i_1\ldots i_n}$, содержащий эту точку. Отсюда и из (2) имеем: $f(x) \geqslant f(x_{i_1\ldots i_n}) - L|x - x_{i_1\ldots i_n}|_\infty \geqslant F_h - L\frac{h}{2} = F_h - \varepsilon\ \ \forall x \in \Pi$. Переходя здесь к нижней грани по $x \in \Pi$, приходим к оценке (3).

Метод простого перебора предполагает, что в каждой точке сетки Π_h вычислены значения функции $f(x)$, которые в определенном порядке перебираются с целью определения величины F_h. Однако, как и в одномерном случае, нетрудно указать более эффективные способы определения величины F_h, которые, вообще говоря, не предполагают вычисления значений функции $f(x)$ во всех точках сетки Π_h и перебора всех точек этой сетки. Опишем один из таких методов последовательного перебора. На первом шаге выбирается произвольная точка $v_1 \in \Pi_h$, вычисляется значение $f(v_1)$ и полагается $F_1 = f(v_1)$. Допустим, что в точках $v_1, v_2, \ldots, v_k$ сетки Π_h уже вычислены значения функции $f(v_1), \ldots, f(v_k)$ и найдена величина $F_k = \min\limits_{1 \leqslant i \leqslant k} f(v_i) = \min\{F_{k-1}; f(v_k)\}$, $k \geqslant 2$. Через v_{j_k} обозначим ту из точек $v_1, \ldots, v_k$, в которой $F_k = f(v_{j_k})$. Далее, возьмем любую из точек $v_{k+1} \in \Pi_h$, которая в предыдущих шагах не исключалась из рассмотрения и в которой еще не вычислялось значение функции $f(x)$. Вычислим значение $f(v_{k+1})$ и величину $F_{k+1} = \min\{F_k; f(v_{k+1})\} = \min\limits_{1 \leqslant i \leqslant k+1} f(v_i)$. Имеются две возможности: либо $F_{k+1} = f(v_{k+1}) < F_k$, либо $F_{k+1} = F_k \leqslant f(v_{k+1})$. В первом случае,

когда $F_{k+1} < F_k$, полагаем $v_{j_{k+1}} = v_{k+1}$ и из дальнейшего перебора исключаем точку v_{j_k} и вместе с нею все точки $x_{i_1\ldots i_n} \in \Pi_h$, для которых

$$|x_{i_1\ldots i_n} - v_{j_k}| \leqslant \frac{F_k - F_{k+1}}{L}. \tag{4}$$

Заметим, что некоторые из этих точек могли оказаться исключенными из перебора уже на предыдущих шагах. Для нас важно лишь то, что среди исключенных точек заведомо нет таких, в которых значение функции $f(x)$ было бы меньше, чем F_{k+1}. В самом деле, $f(v_{j_k}) = F_k > F_{k+1}$. Для остальных исключенных точек $x_{i_1\ldots i_n}$, не зная значения $f(x_{i_1\ldots i_n})$, можем сказать, что $f(x_{i_1\ldots i_n}) - F_{k+1} = f(x_{i_1\ldots i_n}) - f(v_{j_k}) + F_k - F_{k+1} \geqslant -L|x_{i_1\ldots i_n} - v_{j_k}| + F_k - F_{k+1} \geqslant 0$ в силу (2) и (4). Рассмотрим вторую возможность: $F_{k+1} = F_k \leqslant f(v_{j_k})$. Тогда полагаем $v_{j_{k+1}} = v_{j_k}$ и из дальнейшего перебора исключаем точку v_{k+1} вместе с точками $x_{i_1\ldots i_n} \in \Pi_h$, для которых

$$|x_{i_1\ldots i_n} - v_{k+1}| \leqslant \frac{f(v_{k+1}) - F_k}{L}. \tag{5}$$

Нетрудно убедиться, что и в этом случае в исключенных точках значения функции не могут быть меньше F_{k+1}. В самом деле, здесь $f(x_{i_1\ldots i_n}) - F_{k+1} = f(x_{i_1\ldots i_n}) - F_k = f(x_{i_1\ldots i_n}) - f(v_{k+1}) + f(v_{k+1}) - F_k \geqslant -L|x_{i_1\ldots i_n} - v_{k+1}| + f(v_{k+1}) - F_k \geqslant 0$ в силу (2) и (5). Общий шаг метода описан. Так как на каждом шаге метода берется новая точка сетки Π_h, которая еще не исключалась из перебора и в которой значение функции $f(x)$ еще не вычислялось, то ясно, что на каком-то шаге описанного процесса перебора такая новая точка не найдется и процесс закончится за N шагов, $N \leqslant m_1 \cdot m_2 \cdot \ldots \cdot m_n$, перебором точек $v_1, \ldots, v_N$ сетки Π_h и определением величины $F_N = \min\limits_{1 \leqslant i \leqslant N} f(v_i) = \min\limits_{\Pi_h} f(x_{i_1\ldots i_n}) = F_h$. В силу теоремы 1 величина F_h удовлетворяет неравенствам (3). □

Как и в одномерном случае, нетрудно привести примеры, когда изложенный метод покрытий может превратиться в метод простого перебора точек сетки Π_h. В то же время ясно, что если величины $F_k - F_{k+1}$, $f(v_{k+1}) - F_k$ в (4), (5) достаточно большие, то многие точки сетки Π_h будут исключены из перебора без вычисления в них значения функции.

Различные модификации метода покрытий на классе функций (2), обобщения этого метода на более сложные области, чем параллелепипед, на многокритериальные задачи, а также другие методы поиска глобального минимума читатель может найти, например, в [286; 309; 493; 526; 590; 591; 661; 662; 671].

§ 14. Метод модифицированных функций Лагранжа

1. Рассмотрим задачу

$$f(x) \to \inf;\quad x \in X = \{x \in E^n: x \in X_0,$$
$$g_i(x) \leqslant 0,\ i=1,\ldots,m,\ g_i(x) = 0,\ i = m+1,\ldots,s\}, \tag{1}$$

где $f(x), g_1(x), \ldots, g_s(x)$ — заданные функции на множестве X_0. Пусть $f_* > -\infty$, $X_* \neq \varnothing$. Для выпуклой задачи (1) при различных дополнительных

предположениях в § 4.9 было установлено, что найдутся множители Лагранжа $\lambda^* = (\lambda_1^*, \ldots, \lambda_s^*)$:

$$\lambda^* \in \Lambda_0 = \{\lambda \in E^s: \lambda_1 \geqslant 0, \ldots, \lambda_m \geqslant 0\}$$

такие, что пара (x_*, λ^*), где $x_* \in X_*$, образует седловую точку функции Лагранжа

$$L(x, \lambda) = f(x) + \sum_{i=1}^{s} \lambda_i g_i(x), \quad x \in X_0, \quad \lambda \in \Lambda_0, \tag{2}$$

т. е.

$$L(x_*, \lambda) \leqslant L(x_*, \lambda^*) = f_* \leqslant L(x, \lambda^*), \quad x \in X_0, \quad \lambda \in \Lambda_0. \tag{3}$$

Была также доказана справедливость обратного утверждения: если $(x_*, \lambda^*) \in X_0 \times \Lambda_0$ является седловой точкой функции (2), то $x_* \in X_*$.

Основываясь на этих фактах, можно предложить различные методы решения задачи (1), сводящиеся к поиску седловой точки функции Лагранжа. Например, здесь естественным образом напрашивается итерационный процесс, представляющий собой метод проекции градиента по каждой из переменных x и λ (спуск по переменной x и подъем — по λ):

$$x_{k+1} = P_{X_0}(x_k - \alpha_k L_x(x_k, \lambda_k)), \tag{4}$$

$$\lambda_{k+1} = P_{\Lambda_0}(\lambda_k + \alpha_k L_\lambda(x_k, \lambda_k)) = P_{\Lambda_0}(\lambda_k + \alpha_k g(x_k)), \quad k = 0, 1, \ldots \tag{5}$$

где

$$L_x(x_k, \lambda_k) = (L_{x^1}(x, \lambda), \ldots, L_{x^n}(x, \lambda)),$$

$$L_\lambda(x, \lambda = (L_{\lambda_1}(x, \lambda), \ldots, L_{\lambda_s}(x, \lambda)) = (g_1(x), \ldots, g_s(x)) = g(x);$$

длину шага α_k в (4), (5) можно выбирать из тех же соображений, как это делалось выше в § 2. Заметим, что проекция любой точки $\lambda \in E^s$ на множество Λ_0 вычисляется просто по формулам $P_{\Lambda_0}(\lambda) = (\mu_1, \ldots, \mu_s)$, где

$$\mu_i = \max\{\lambda_i, 0\}, \quad i = 1, \ldots, m, \quad \mu_i = \lambda_i, \quad i = m+1, \ldots, s.$$

Вместо (4) возможно использование других итерационных процессов, таких, как метод Ньютона и др. В тех случаях, когда задача минимизации функции (x, λ) по переменной $x \in X_0$ при каждом фиксированном $\lambda \in \Lambda_0$ решается достаточно просто, можно предложить следующий итерационный метод:

$$L(x_{k+1}, \lambda) = \inf_{x \in X_0} L(x, \lambda_k), \quad \lambda_{k+1} = P_{\Lambda_0}(\lambda_k + \alpha_k g(x_{k+1})), \quad k = 0, 1 \ldots$$

Однако, как оказалось, сходимость перечисленных методов удается доказать лишь при довольно жестких ограничениях на данные задачи (1).

Приведем простейший пример выпуклой задачи, когда метод (4), (5) не сходится к седловой точке функции Лагранжа.

Пример 1. Пусть $f(x) \equiv 0$, $X = \{x \in E^1: g(x) \equiv x = 0\}$. Тогда $f_* = 0$, $X_* = \{0\}$. Функция Лагранжа $L(x, \lambda) = f(x) + \lambda g(x) \equiv \lambda x$ на $X_0 \times \Lambda_0 = E^1 \times E^1$ имеет седловую точку $(0, 0)$. Процесс (4), (5) здесь имеет вид

$$x_{k+1} = x_k - \alpha_k \lambda_k, \quad \lambda_{k+1} = \lambda_k + \alpha_k x_k, \quad k = 0, 1 \ldots$$

Поскольку $x_{k+1}^2 + \lambda_{k+1}^2 = (x_k^2 + \lambda_k^2)(1 + \alpha_k^2) \geqslant x_k^2 + \lambda_k^2$, $k = 0, 1, \ldots$, то ясно, что при любых $(x_0, \lambda_0) \neq 0$ и любом выборе длины шага $\alpha_k \geqslant 0$ этот процесс расходится.

Анализ перечисленных методов показывает, что причина их расходимости заключается в том, что функция Лагранжа (2) по переменной λ не очень хорошо «устроена». Чтобы преодолеть возникающие здесь трудности, можно попытаться видоизменить функцию Лагранжа, строить так называемые модифицированные функции Лагранжа, которые имеют то же множество седловых точек, что и функция (2), и которые обладают лучшими свойствами, чем функция (2). Такие функции, оказывается, существуют и могут быть использованы для поиска седловой точки функции (2) и для решения задачи (1). Следуя [24], мы рассмотрим один из возможных здесь подходов.

2. Будем рассматривать задачу

$$f(x) \to \inf, \quad x \in X = \{x \in E^n: x \in X_0, \ g(x) \leqslant 0\}, \tag{6}$$

где $f(x)$, $g(x) = (g_1(x), \ldots, g_m(x))$ — заданные функции из $C^1(X_0)$. Как и в гл. 3, векторное неравенство $g = (g_1, \ldots, g_m) \geqslant 0$ $[g \leqslant 0]$ здесь и ниже означает, что $g_i \geqslant 0$ $[g_i \leqslant 0]$ при всех $i = 1, \ldots, m$, а неравенство $a \geqslant b$ для $a, b \in E^m$ эквивалентно неравенству $a - b \geqslant 0$.

Наряду с классической функцией Лагранжа задачи (6)

$$L(x, \lambda) = f(x) + \langle g(x), \lambda \rangle, \quad x \in X_0, \quad \lambda \in \Lambda_0 = \{\lambda \in E^m: \lambda \geqslant 0\} = E_+^m \tag{7}$$

еще рассмотрим следующую модифицированную функцию Лагранжа:

$$M(x, \lambda) = f(x) + \frac{1}{2A}[(\lambda + Ag(x))^+]^2 - \frac{1}{2A}|\lambda|^2 \tag{8}$$

переменных $x \in X_0$, $\lambda \in \Lambda_0$, где A — произвольная фиксированная положительная константа; в (8) принято обозначение

$$a^+ = \mathcal{P}_{E_+^m}(a) = (a_1^+, \ldots, a_m^+), \quad a_i^+ = \max\{a_i; 0\}, \quad i = 1, \ldots, m, \tag{9}$$

— проекция точки $a \in E^m$ на положительный ортант E_+^m.

Нетрудно видеть, что функция $\varphi(z) = (\max\{z; 0\})^2 = (z^+)^2$ одной переменной непрерывно дифференцируема на всей числовой оси E^1, причем

$$\varphi'(z) = 2 \max\{z; 0\} = 2z^+.$$

Отсюда следует, что при $f(x), g(x) \in C^1(X_0)$ функция (8) непрерывно дифференцируема по x и λ, причем

$$\begin{aligned} \frac{\partial M}{\partial x} &= M_x(x, \lambda) = f'(x) + (g'(x))^\top (\lambda + Ag(x))^+, \\ \frac{\partial M}{\partial \lambda} &= M_\lambda(x, \lambda) = \frac{1}{A}[(\lambda + Ag(x))^+ - \lambda], \quad x \in X_0, \quad \lambda \in E^m, \end{aligned} \tag{10}$$

где $g'(x)$ — матрица порядка $m \times n$, у которой в i-й строке, j-м столбце $g_{ix^j}(x) = \frac{\partial g_i(x)}{\partial x^j}$, $i = 1, \ldots, m$, $j = 1, \ldots, n$, а матрица $(g'(x))^\top$ получена транспонированием $g'(x)$. Далее, пользуясь теоремами 4.2.7, 4.2.8 и следствиями из них, нетрудно показать, что если X_0 — выпуклое множество и функции $f(x)$, $g_i(x)$ выпуклы на X_0, то функция $M(x, \lambda)$ выпукла по переменной x на множестве X_0 при любом фиксированном $\lambda \in E^m$. Отметим также, хотя это ниже явно не будет использовано, что $M(x, \lambda)$ является вогнутой по переменной λ на множестве Λ_0 при любом фиксированном $x \in X_0$ — в этом проще всего убедиться, доказав неравенство $\langle M_\lambda(x, \lambda) - M_\lambda(x, \mu), \lambda - \mu \rangle \leqslant 0$ для всех $\lambda, \mu \in \Lambda_0$ и затем обратившись к теореме 4.2.4.

Перейдем к описанию метода решения задачи (6), использующего функцию $M(x, \lambda)$. В качестве начального приближения возьмем любые точки $x_0 \in X_0$, $\lambda_0 \in \Lambda_0$. Пусть k-е приближение $x_k \in X_0$, $\lambda_k \in \Lambda_0$ уже известно. Составим функцию (ср. с § 6)

$$\Phi_k(x) = \tfrac{1}{2}|x - x_k|^2 + \alpha M(x, \lambda_k), \quad x \in X_0, \tag{11}$$

где α — некоторое положительное число, являющееся параметром метода. Предположим, что существует точка v_k, удовлетворяющая условиям

$$v_k \in X_0, \quad \Phi_k(v_k) = \inf_{X_0} \Phi_k(x). \tag{12}$$

В качестве следующего $(k+1)$-го приближения возьмем точку x_{k+1} такую, что

$$x_{k+1} \in X_0, \quad \Phi_k(x_{k+1}) \leqslant \inf_{X_0} \Phi_k(x) + \delta_k^2/2, \quad |g(x_{k+1}) - g(v_k)| \leqslant \delta_k, \tag{13}$$

где $\delta_k \geqslant 0$, $\lim\limits_{k\to\infty} \delta_k = 0$. В частности, если точка v_k из (12) известна, то можно принять $x_{k+1} = v_k$; в общем случае для определения x_{k+1} из условий (13) нужно решать задачу (12) с помощью какого-либо сходящегося метода минимизации. Дальнейшее изложение не зависит от того, каким методом решается задача (12), поэтому здесь мы можем ограничиться предположением, что имеется какой-либо достаточно эффективный метод решения задачи (12), позволяющий за конечное число итераций найти точку x_{k+1}, которая удовлетворяет условиям (13). После определения x_{k+1} точка λ_{k+1} находится по формуле

$$\lambda_{k+1} = (\lambda_k + Ag(x_{k+1}))^+. \tag{14}$$

Правила получения $(k+1)$-го приближения $x_{k+1} \in X_0$, $\lambda_{k+1} \in \Lambda_0$ изложены.

Описанный метод кратко будем называть методом (13), (14). Для исследования сходимости метода (13), (14) нам понадобятся некоторые свойства функции a^+, определенной равенствами (9). Из теоремы 4.4.2 следует, что

$$|a^+ - b^+| \leqslant |a - b| \qquad \forall a, b \in E^m. \tag{15}$$

Далее, система соотношений

$$g \leqslant 0 < \quad \lambda \geqslant 0, \quad \lambda_i g_i = 0, \quad i = 1, \ldots, m, \tag{16}$$

эквивалентна равенству

$$\lambda = (\lambda + Ag)^+ \tag{17}$$

при любых постоянных $A > 0$. В самом деле, если выполняются соотношения (16), то либо $g_i = 0$, $\lambda_i \geqslant 0$, либо $\lambda_i = 0$, $g_i \leqslant 0$. В каждом из этих случаев, очевидно, равенство (17) верно. Таким образом, из (16) следует (17). Докажем обратное. Пусть имеет место равенство (17). Распишем это равенство в координатной форме

$$\lambda_i = (\lambda_i + Ag_i)^+ = \max\{\lambda_i + Ag_i; 0\}, \quad i = 1, \ldots, m. \tag{17'}$$

Отсюда ясно, что $\lambda_i \geqslant 0$ при всех $i = 1, \ldots, m$, т. е. $\lambda_i \geqslant 0$. Если $\lambda_i = 0$, то $\lambda_i g_i = 0$ и, кроме того, из (17') получим $0 = (0 + Ag_i)^+ = \lambda_i + Ag_i$, т. е. $g_i \leqslant 0$. Если же $\lambda_i > 0$, то из (17') следует $0 < \lambda_i = (\lambda_i + Ag_i)^+ = \lambda_i + Ag_i$, что возможно лишь при $g_i = 0$ и $\lambda_i g_i = 0$. Эквивалентность (16) и (17) доказана.

Далее, пользуясь определением (9) функции a^+, нетрудно получить, что

$$\langle a^+, a\rangle = \langle a^+, a^+\rangle, \quad \langle a^+, b\rangle \leqslant \langle a^+, b^+\rangle \quad \forall a, b \in E^n.$$

Отсюда имеем

$$\langle a^+ - b^+, a - b\rangle = \langle a^+, a\rangle + \langle b^+, b\rangle - \langle a^+, b\rangle - \langle b^+, a\rangle \geqslant$$
$$\geqslant \langle a^+, a^+\rangle + \langle b^+, b^+\rangle - \langle a^+, b^+\rangle - \langle b^+, a^+\rangle = \langle a^+ - b^+, a^+ - b^+\rangle,$$

т. е.

$$\langle a^+ - b^+, a - b\rangle \geqslant \langle a^+ - b^+, a^+ - b^+\rangle. \tag{18}$$

Теорема 1. *Пусть X_0 — выпуклое замкнутое множество из E^n (например, $X_0 = E^n$), функции $f(x), g_1 x), \ldots, g_m(x)$ выпуклы на X_0 и принадлежат классу $C^1(X_0)$, $f_* > -\infty$, $X_* \neq \varnothing$, функция Лагранжа (7) имеет хотя бы одну седловую точку $(x_*, \lambda^*) \in X_0 \times \Lambda_0$ в смысле неравенств (3). Пусть, кроме того, последовательность $\{\delta_k\}$ из (13) неотрицательна и $\sum\limits_{k=0}^{\infty} \delta_k < \infty$. Тогда последовательность $\{(x_k, \lambda_k)\}$, удовлетворяющая условиям (13), (14), при любом выборе начальных $(x_0, \lambda_0) \in X_0 \times \Lambda_0$ и любых фиксированных параметрах $\alpha > 0$, $A > 0$ существует и сходится к некоторой седловой точке функции Лагранжа (7).*

Доказательство. При сделанных предположениях функция $M(x, \lambda)$ выпукла по переменной $x \in X_0$ при всех $\lambda \in \Lambda_0$, $A > 0$, поэтому при любых $x_k \in X_0$, $\lambda_k \in \Lambda_0$, $\alpha > 0$, $A > 0$ функция $\Phi_k(x)$, определяемая формулой (11), сильно выпукла на X_0 с константой сильной выпуклости $\varkappa = 1$. Отсюда и из теоремы 4.3.1 следует, что точка v_k, удовлетворяющая условиям (12), существует и определяется однозначно. Тогда существует и точка x_{k+1}, удовлетворяющая условиям (13): например, в (13) можно взять $x_{k+1} = v_k$. Здесь важно заметить, что многие из описанных выше методов минимизации для задачи (12) сходятся и при любом $\delta_k > 0$ позволяют получить точку x_{k+1} из (13) за конечное число итераций. Таким образом, при выполнении условий теоремы последовательность $\{(x_k, \lambda_k)\}$ существует и имеются достаточно эффективные способы реализации каждой итерации метода (13), (14).

Наряду с точкой λ_{k+1}, определяемой по формуле (14), введем еще точку

$$\mu_k = (\lambda_k + Ag(v_k))^+, \quad k = 0, 1, \ldots \tag{19}$$

Покажем, что для любой седловой точки (x_*, λ^*) функции Лагранжа (7) справедливо неравенство

$$|x_k - x_*|^2 + \frac{\alpha}{A}|\lambda_k - \lambda^*|^2 \geqslant |v_k - x_*|^2 + \frac{\alpha}{A}|\mu_k - \lambda^*|^2 + |v_k - x_k|^2 + \frac{\alpha}{A}|\mu_k - \lambda_k|^2, \quad k = 0, 1, \ldots \tag{20}$$

Согласно лемме 4.9.2 существование седловой точки (x_*, λ^*) в задаче (6) эквивалентно соотношениям

$$\langle f'(x_*) + (g'(x_*))^\top \lambda^*, x - x_*\rangle \geqslant 0 \qquad \forall x \in X_0, \tag{21}$$

$$g(x_*) \leqslant 0, \quad \lambda^* \geqslant 0, \quad \lambda_i^* g_i(x_*) = 0, \quad i = 1, \ldots, m. \tag{22}$$

В силу эквивалентности соотношений (16), (17) условия (22) можно переписать в следующей равносильной форме:

$$\lambda^* = (\lambda^* + Ag(x_*))^+. \tag{23}$$

Из (21) с учетом равенства (23) имеем

$$\langle f'(x_*) + (g'(x_*))^\top (\lambda^* + Ag(x_*))^+, x - x_*\rangle \geqslant 0, \quad \forall x \in X_0. \tag{24}$$

Далее, из условия (12) и теоремы 4.2.3 следует

$$\langle \Phi_k'(v_k), x - v_k\rangle \geqslant 0, \quad \forall x \in X_0.$$

Отсюда с учетом формулы (10) получим

$$\langle v_k - x_k + \alpha f'(v_k) + \alpha (g'(v_k))^\top (\lambda_k + Ag(v_k))^+, x - v_k\rangle \geqslant 0, \quad \forall x \in X_0. \tag{25}$$

Примем в (24) $x = v_k$, умножим это неравенство на $\alpha > 0$ и сложим с неравенством (25) при $x = x_*$. Получим

$$\langle v_k - x_k + \alpha(f'(v_k) - f'(x_*)) + \alpha(g'(v_k))^\top (\lambda_k + Ag(v_k))^+ -$$
$$- \alpha (g'(x_*))^\top (\lambda^* + Ag(x_*))^+, x_* - v_k\rangle \geqslant 0, \quad k = 0, 1, \ldots$$

Отсюда имеем

$$\langle v_k - x_k, x_* - v_k\rangle \geqslant \alpha\langle f'(v_k) - f'(x_*), v_k - x_*\rangle + \alpha\langle (\lambda_k + Ag(v_k))^+, g'(v_k)(v_k - x_*)\rangle -$$
$$- \alpha\langle (\lambda^* + Ag(x_*))^+, g'(x_*)(v_k - x_*)\rangle, \quad k = 0, 1, \ldots \tag{26}$$

Так как функции $f(x)$, $g_i(x)$ выпуклы, то согласно теореме 4.2.4

$$\langle f'(v_k) - f'(x_*), v_k - x_*\rangle \geqslant 0, \quad g'(v_k)(v_k - x_*) \geqslant g(v_k) - g(x_*) \geqslant g'(x_*)(v_k - x_*).$$

Отсюда и из (26) следует

$$\langle v_k - x_k, x_* - v_k\rangle \geqslant \alpha\langle (\lambda_k + Ag(v_k))^+ - (\lambda^* + Ag(x_*))^+, g(v_k) - g(x_*)\rangle =$$
$$= \frac{\alpha}{A}\langle (\lambda_k + Ag(v_k))^+ - (\lambda^* + Ag(x_*))^+, [(\lambda_k + Ag(v_k)) - \lambda_k] - [\lambda^* + Ag(x_*)) - \lambda^*]\rangle.$$

К правой части этой оценки применим неравенство (18). С учетом формулы (19), определяющей точку μ_k, и равенства (23) получим

$$\langle v_k - x_k, x_* - v_k\rangle \geqslant \frac{\alpha}{A}\langle(\lambda_k + Ag(v_k))^+ - (\lambda^* + Ag(x_*))^+,$$

$$[(\lambda_k + Ag(v_k))^+ - \lambda_k] - [(\lambda^* + Ag(x_*))^* - \lambda^*]\rangle = \frac{\alpha}{A}\langle\mu_k - \lambda^*, \mu_k - \lambda_k\rangle,$$

т. е.

$$\langle v_k - x_k, x_* - v_k\rangle + \frac{\alpha}{A}\langle\mu_k - \lambda_k, \lambda^* - \mu_k\rangle \geqslant 0, \quad k = 0, 1, \ldots \tag{27}$$

Справедливы тождества

$$|x_k - x_*|^2 = |(x_k - v_k) + (v_k - x_*)|^2 = |v_k - x_k|^2 + |x_* - v_k|^2 + 2\langle v_k - x_k, x_* - v_k\rangle,$$

$$|\lambda_k - \lambda^*|^2 = |\mu_k - \lambda_k|^2 + |\lambda^* - \mu_k|^2 + 2\langle\mu_k - \lambda_k, \lambda^* - \mu_k\rangle.$$

Умножим второе из этих тождеств на α/A и сложим с первым. Отсюда с учетом оценки (27) получим обещанное неравенство (20).

Далее, покажем, что

$$|x_{k+1} - v_k| \leqslant \delta_k, \quad |\lambda_{k+1} - \mu_k| \leqslant A\delta_k, \quad k = 0, 1, \ldots. \tag{28}$$

Поскольку функция $\Phi_k(x)$ сильно выпукла на X_0, то с помощью теоремы 4.3.1 и первого неравенства (13) получим

$$|x_{k+1} - v_k|^2/2 \leqslant \Phi_k(x_{k+1}) - \Phi_k(v_k) \leqslant \delta_k^2/2,$$

что равносильно первой оценке (28). Из формул (14), (19), определяющих точки λ_{k+1}, μ_{k+1}, неравенства (15) и условий (13) следует

$$|\lambda_{k+1} - \mu_k| \leqslant A|g(x_{k+1}) - g(v_k)| \leqslant A\delta_k.$$

Оценки (28) доказаны.

В $(n+m)$-мерном пространстве E^{n+m} переменных $z = (x, \lambda) = (x^1, \ldots, x^n, \lambda_1, \ldots, \lambda_m)$ введем скалярное произведение $\langle z_1, z_2\rangle = \langle x_1, x_2\rangle + (\alpha/A)\langle\lambda^1, \lambda^2\rangle$ и соответствующую ему норму

$$\|z\| = (|x|^2 + (\alpha/A)|\lambda|^2)^{1/2}. \tag{29}$$

Тогда, обозначив $z_k = (x_k, \lambda_k)$, $w_k = (v_k, \mu_k)$, $z^* = (x_*, \lambda^*)$, неравенства (20) и (28) можем записать и виде

$$\|z_k - z^*\|^2 \geqslant \|w_k - z^*\|^2 + \|w_k - z_k\|^2, \quad k = 0, 1, \ldots, \tag{30}$$

$$\|z_{k+1} - w_k\| \leqslant (A\alpha + 1)\delta_k, \quad k = 0, 1, \ldots \tag{31}$$

Напомним, что по условию $\sum_{k=0}^{\infty} \delta_k < \infty$. Таким образом, последовательности $\{z_k\}$, $\{w_k\}$, $\{\delta_k\}$ удовлетворяют условиям леммы 2.6.10. Для полной строгости, конечно, нужно заметить, что в неравенствах (2.6.30), (2.6.31) использована евклидова норма пространства E^{n+m}, а в только что полученных неравенствах (30), (31) — норма (29). Тем не менее, рассуждая так же, как при доказательстве леммы 2.6.10, нетрудно показать, что существует конечный предел $\lim_{k\to\infty} \|z_k - z^*\|$ и, кроме того,

$$\lim_{k\to\infty} \|w_k - z_k\| = 0. \tag{32}$$

Заметим, что

$$\min\{1; \alpha/A\}|z|^2 \leqslant \|z\|^2 \leqslant \max\{1; \alpha/A\}|z|^2,$$

т. е. нормы $|z|$ и $\|z\|$ эквивалентны. Отсюда и из существования конечного предела $\lim_{k\to\infty} \|z_k - z^*\|$ следует, что последовательность $\{z_k = (x_k, \lambda_k)\} \in X_0 \times \Lambda_0$ ограничена в E^{n+m} и из нее можно выбрать подпоследовательность $\{z_{k_r} = (x_{k_r}, \lambda_{k_r})\}$, которая сходится в E^{n+m} к некоторой точке $c^* = (a_*, b^*)$, причем $a_* \in X_0$, $b^* \in \Lambda_0$ в силу замкнутости X_0 и Λ_0. Покажем, что $c^* = (a_*, b^*)$ — седловая точка функции Лагранжа (7). Из $\{z_{k_r}\} \to c^*$ и (31), (32) следует, что $\{w_{k_r}\} \to c^*$, $\{z_{k_r+1}\} \to c^*$. Тогда из (14) при $k = k_r \to \infty$ получим

$$b^* = (b^* + Ag(a_*))^+. \tag{33}$$

В силу эквивалентности соотношений (16) и (17) из (33) следует

$$g(a_*) \leqslant 0, \quad b^* \geqslant 0, \quad b_i^* g_i(a_*) = 0, \quad i = 1, \ldots, m. \tag{34}$$

Далее, переходя в (25) к пределу при $k = k_r \to \infty$ будем иметь

$$\langle f'(a_*) + (g'(a_*))^\top (b^* + Ag(a_*))^+, x - a_*\rangle \geqslant 0, \quad x \in X_0,$$

или с учетом (33)

$$\langle f'(a_*) + (g'(a_*))^\top b^*, x - a_*\rangle \geqslant 0 \qquad \forall x \in X_0. \tag{35}$$

Из соотношений (34), (35) и леммы 4.9.2 следует, что $c^* = (a_*, b^*)$ — седловая точка функции $L(x, \lambda)$ в смысле неравенств (3), а тогда согласно теореме 4.9.1 получаем, что a_* — решение задачи (6).

Заметим, что неравенство (30) верно для любой седловой точки, в частности, оно верно и для найденной точки $c^* = (a_*, b^*)$. Поэтому существует конечный предел $\lim_{k\to\infty} \|z_k - c^*\|$, причем в силу определения точки c^* имеем $\lim_{k\to\infty} \|z_k - c^*\| = \lim_{r\to\infty} \|z_{k_r} - c^*\| = 0$. Это значит, что вся последовательность $\{z_k = (x_k, \lambda_k)\}$ сходится к точке $c^* = (a_*, b^*)$, и, в частности, $\{x_k\}$ сходится к a_* — решению задачи (6). Теорема 1 доказана. □

Другие методы поиска седловой точки функции Лагранжа, другие методы решения задачи (1) или (6), основанные на связи между двойственными задачами (см. теорему 4.9.6), а также библиографию по таким методам читатель найдет в [24–26; 222; 234; 286; 344; 759].

§ 15. Метод штрафных функций

1. Метод штрафных функций является одним из наиболее простых и широко применяемых методов решения задач минимизации. Основная идея метода заключается в сведении исходной задачи

$$f(x) \to \inf; \quad x \in X \tag{1}$$

к последовательности задач минимизации

$$\Phi_k(x) \to \inf; \quad x \in X_0, \quad k = 1, 2, \ldots, \tag{2}$$

где $\Phi_k(x)$ — некоторая вспомогательная функция, а множество X_0 содержит X. При этом функция $\Phi_k(x)$ подбирается так, чтобы она с ростом номера k мало отличалась от исходной функции $f(x)$ на множестве X и быстро возрастала на множестве $X_0 \setminus X$. Можно ожидать, что быстрый рост функции $\Phi_k(x)$ вне X приведет к тому, что при больших k нижняя грань этой функции на X_0 будет достигаться в точках, близких ко множеству X, и решение задачи (2) будет приближаться к решению задачи (1). Кроме того, как увидим ниже, имеется достаточно широкий произвол в выборе функций $\Phi_k(x)$ и множества X_0 для задач (2), и можно надеяться на то, что задачи (2) удастся составить более простыми по сравнению с задачей (1) и допускающими применение несложных методов минимизации.

Определение 1. Последовательность функций $\{P_k(x), k = 1, 2, \ldots\}$, определенных и неотрицательных на множестве X_0, содержащем множество X, называют *штрафом* или *штрафной функцией* множества X на множестве X_0, если

$$\lim_{k\to\infty} P_k(x) = \begin{cases} 0, & \forall x \in X, \\ \infty, & \forall x \in X_0 \setminus X. \end{cases}$$

Из этого определения видно, что при больших номерах k за нарушение условия $x \in X$ приходится «платить» большой штраф, в то время как при $x \in X$ штрафная функция представляет собой бесконечно малую величину при $k \to \infty$.

Для любого множества $X \subset E^n$ можно указать сколько угодно различных штрафных функций. Например, если $\{A_k\}$ — какая-либо положительная последовательность, $\lim\limits_{k\to\infty} A_k = \infty$, то можно взять

$$P_k(x) = A_k \rho(x, X), \quad x \in E^n = X_0, \quad k = 1, 2, \ldots$$

(здесь X предполагается замкнутым) или

$$P_k(x) = \begin{cases} 0, & x \in X, \\ A_k |x - \bar{x}|, & x \notin X, \quad k = 1, 2, \ldots; \end{cases}$$

где $\rho(x, X) = \inf\limits_{y \in X} |x - y|$ — расстояние от точки x до множества X, а $\bar{x}$ — какая-либо точка из X. Другие примеры штрафных функций будут приведены ниже.

Допустим, что некоторое множество X_0, содержащее X, а также штрафная функция $\{P_k(x)\}$ множества X на X_0 уже выбраны. Предполагая, что функция $f(x)$ определена на X_0, введем функции

$$\Phi_k(x) = f(x) + P_k(x), \quad x \in X_0, \quad k = 1, 2, \ldots \tag{3}$$

и рассмотрим последовательность задач (2) с функциями (3). Будем считать, что

$$\Phi_{k*} = \inf_{X_0} \Phi_k(x) > -\infty, \quad k = 1, 2, \ldots \tag{4}$$

Если здесь при каждом $k = 1, 2, \ldots$ нижняя грань достигается, то условия

$$\Phi_k(x_k) = \Phi_{k*}, \quad x_k \in X_0, \tag{5}$$

определяют последовательность $\{x_k\}$. Однако точно определить x_k из (5) удается лишь в редких случаях. Кроме того, нижняя грань в (4) при некоторых или даже всех $k = 0, 1, \ldots$ может и не достигаться. Поэтому будем считать, что при каждом $k = 1, 2, \ldots$ с помощью какого-либо метода минимизации найдена точка x_k, определяемая условиями

$$x_k \in X_0, \quad \Phi_k(x_k) \leqslant \Phi_{k*} + \varepsilon_k, \tag{6}$$

где $\{\varepsilon_k\}$ — некоторая заданная последовательность, $\varepsilon_k > 0$, $k = 1, 2, \ldots$, $\lim\limits_{k\to\infty} \varepsilon_k = 0$ (если x_k удовлетворяет условиям (5), то в (6) допускается возможность $\varepsilon_k = 0$). Отметим, что, вообще говоря, $x_k \notin X$. Метод штрафных функций описан.

Подчеркнем, что дальнейшее изложение не зависит от того, каким конкретным методом будет найдена точка x_k из (6). Поэтому мы здесь можем ограничиться предположением, что имеется достаточно эффективный метод определения такой точки.

2. Перейдем теперь к исследованию сходимости метода штрафных функций. Так как

$$\lim_{k\to\infty} P_k(x) = \infty \quad \text{при} \quad x \in X_0 \setminus X,$$

то можно ожидать, что для широкого класса задача (1), последовательность $\{x_k\}$, определяемая условиями (6), будет приближаться ко множеству X и будут справедливы равенства

$$\lim_{k\to\infty} f(x_k) = f_*, \quad \lim_{k\to\infty} \rho(x_k, X_*) = 0. \tag{7}$$

Мы здесь ограничимся рассмотрением задачи (1) для случая, когда множество X имеет вид

$$X = \{x \in E^n: x \in X_0, \ g_i(x) \leqslant 0, \ i = 1, \ldots, m; \ g_i(x) = 0, \ i = m+1, \ldots, s\}, \tag{8}$$

где X_0 — заданное множество из E^n (например, $X_0 = E^n$), функции $f(x)$, $g_i(x)$, $i = 1, \ldots, s$, определены на X_0. В качестве штрафной функции множества (8) возьмем

$$\begin{aligned} &P_k(x) = A_k P(x), \\ &P(x) = \sum_{i=1}^{m} (\max\{g_i(x); 0\})^p + \sum_{i=m+1}^{s} |g_i(x)|^p, \quad x \in X_0, \end{aligned} \tag{9}$$

где $A_k > 0$, $k = 1, 2, \ldots$, $\lim\limits_{k\to\infty} A_k = \infty$, а $p \geqslant 1$ — фиксированное число.

Очевидно, если функции $g_i(x)$ будут r раз непрерывно дифференцируемы на множестве X_0, то при любом $p > r$ функция (9) также будет r раз непрерывно дифференцируема на X_0. Если в (9) $p = 1$, то из непрерывности $g_i(x)$, $i = 1, \ldots, s$, следует непрерывность $P_k(x)$ на X_0, но гладкости $P_k(x)$ в этом случае ожидать не приходится. Полезно также заметить, что если X_0 — выпуклое множество, функции $g_i(x)$ при $i = 1, \ldots, m$ выпуклы на X_0, $g_i(x) = \langle a_i, x\rangle - b^i$ — линейные функции при $i = m+1, \ldots, s$, то функция (9) выпукла на X_0 — это вытекает из следствий к теореме 4.2.8.

Если для краткости ввести обозначения

$$g_i^+(x) = \begin{cases} \max\{g_i(x); 0\}, & i = 1, \ldots, m, \\ |g_i(x)|, & i = m+1, \ldots, s, \end{cases} \tag{10}$$

то функцию (9) можно записать в виде

$$P_k(x) = A_k P(x), \quad P(x) = \sum_{i=1}^{s} (g_i^+(x))^p, \quad x \in X_0.$$

Функцию $P(x)$ мы также будем называть штрафной функцией множества (8), подразумевая при этом, что после умножения на $A_k > 0$, $\lim\limits_{k\to\infty} A_k = \infty$, она превратится в штрафную функцию в смысле определения 1. Величины A_k из (9) будем называть *штрафными коэффициентами*.

Заметим, что существуют и другие штрафные функции множества (8). Например, вместо (9) можно взять

$$P_k(x) = \sum_{i=1}^{s} A_{ki} (g_i^+(x))^{p_i}, \quad x \in X_0, \quad k = 1, 2, \ldots, \tag{9'}$$

где $p_i \geqslant 1$, $A_{ki} > 0$, $\lim\limits_{k\to\infty} A_{ki} = \infty$, $i = 1, \ldots, s$; здесь каждое ограничение из (8) имеет свой штрафной коэффициент. Весьма широкий класс штрафных функций множества (8) дает следующая конструкция:

$$P_k(x) = \sum_{i=1}^{s} A_{ki} \varphi_i(g_i^+(x)), \quad x \in X_0, \quad k = 1, 2, \ldots,$$

где $\varphi_i(g)$ — произвольная функция, определенная при $g \geqslant 0$ такая, что $\varphi_i(0) = 0$, $\varphi_i(g) > 0$ при $g > 0$, $i = 1, \ldots, s$. При необходимости можно выбрать функции $\varphi_i(g)$ так, чтобы штрафная функция $P_k(x)$ обладала различными полезными свойствами, такими, как, например, непрерывность, гладкость,

выпуклость, простота вычисления значений функции и нужных производных и т. п. Возможны и другие конструкции штрафных функций множества (8). Приведем еще два конкретных примера штрафной функции

$$P_k(x)=\Big(1+\sum_{i=1}^{s}(g_i^+(x))^{p_i}\Big)^{A_k}-1,\quad p_i\geqslant 1,$$

$$P_k(x)=A_k^{-1}\Big(\sum_{i=1}^{m}\exp\{A_k g_i(x)\}+\sum_{i=m+1}^{s}\exp\{A_k g_i^2(x)\}\Big),\quad x\in X_0,$$

где $A_k>0$, $k=1,2,\ldots$, $\lim\limits_{k\to\infty}A_k=\infty$.

Прежде чем переходить к строгим формулировкам теорем сходимости метода штрафных функций, рассмотрим несколько примеров.

Пример 1. Пусть требуется решить задачу

$$f(u)=x^2+xy+y^2\to\inf,\quad u\in X=\{u=(x,y)\in E^2:\ x+y-2=0\}.$$

В качестве штрафной функции возьмем $P_k(u)=k(x+y-2)^2$ и положим

$$\Phi_k(u)=x^2+xy+y^2+k(x+y-2)^2,\quad u\in X_0=E^2;\quad k=1,2,\ldots$$

Функция $\Phi_k(u)$ при каждом фиксированном $k=1,2,\ldots$ сильно выпукла на E^2 и достигает своей нижней грани на E^2 в точке $u_k=(x_k,y_k)$, которая определяется уравнениями

$$\frac{\partial\Phi_k(u_k)}{\partial x}=2x_k+y_k+2k(x_k+y_k-2)=0,$$
$$\frac{\partial\Phi_k(u_k)}{\partial y}=x_k+2y_k+2k(x_k+y_k-2)=0.$$

Отсюда получаем

$$u_k=\Big(\frac{4k}{3+4k},\frac{4k}{3+4k}\Big),\quad \Phi_k(u_k)=\frac{12k}{4k+3}=\inf_{E^2}\Phi_k(u).$$

При $k\to\infty$ будем иметь $u_k\to u_*=(1,1)$, $\Phi_k(u_k)\to 3$. Нетрудно видеть, что u_* — решение исходной задачи. В самом деле, $f'(u_*)=(3;3)$, $\langle f'(u_*),u-u_*\rangle=3(x-1)+3(y-1)=0$ для всех $u\in X$. В силу выпуклости множества X и функции $f(u)$, согласно теореме 4.2.3, тогда u_* — точка минимума $f(u)$ на X, причем $f(u_*)=f_*=3=\lim\limits_{k\to\infty}\Phi_k(u_k)$. Таким образом, в рассмотренном примере метод штрафных функций сходится.

Пример 2. Пусть

$$f(x)=e^{-x}\to\inf;\quad x\in X=\{x\in E^1:\ g(x)=xe^{-x}=0\}.$$

Здесь $X=\{0\}=X_*$, $f_*=1$. Возьмем штрафную функцию $P_k(x)=kg^2(x)=kx^2e^{-2x}$ и положим $\Phi_k(x)=e^{-x}+kx^2e^{-2x}$, $x\in X_0=E^1$. Так как $\Phi_k(x)>0$ при всех $x\in E^1$, $\lim\limits_{x\to\infty}\Phi_k(x)=0$, то $\Phi_{k*}=\inf\limits_{E^1}\Phi_k(x)=0$. В качестве точки x_k, удовлетворяющей условиям (6) при $\varepsilon_k=e^{-k}+k^2e^{-2k}$, здесь можно взять $x_k=k$, $k=1,2,\ldots$ Получим $\lim\limits_{k\to\infty}f(x_k)=0<f_*=1$, $\lim\limits_{k\to\infty}\rho(x_k,X_*)=\infty$. Таким образом, выясняется, что метод штрафных функций не всегда сходится.

Пример 3. Задача: $f(u)=(x-1)^2-y\to\inf$, $u\in X=\{u=(x,y,z)\in X_0=E^3:\ g_1(u)=y^2\leqslant 0,\ g_2(u)=-z\leqslant 0,\ g_3(u)=x^2-yz\leqslant 0\}$. Здесь $f_*=1$, $X_*=X=\{u=(0,0,z)\ \forall z\geqslant 0\}$. Возьмем штрафную функцию такую: $\Phi_k(x)=$

$=(x-1)^2-y+ky^2+k(\max\{-z;0\})^2+k(\max\{x^2-yz;0\})^2$, $u\in X_0=E^3$, $k=1,2,\ldots$ Очевидно, $\Phi_k(u)\geqslant\min\limits_{y\in E^1}(-y+ky^2)=-\frac{1}{4k}$ $\forall u\in E^3$, причем в точке $u_k=(1,\frac{1}{2k},2k)$ значение $\Phi_k(u_k)=-\frac{1}{4k}$. Следовательно, $\Phi_{k*}=-\frac{1}{4k}>-\infty$, $k=1,2,\ldots$, и точка u_k удовлетворяет условию (6) при $\varepsilon_k=0$. Однако, $\lim\limits_{k\to\infty}f(u_k)=\lim\limits_{k\to\infty}\frac{-1}{4k}=0<f_*=1$, $\rho(u_k,X_*)=\inf\limits_{u\in X_*}\Big(1+\Big(\frac{1}{2k}\Big)^2+(2k-z)^2\Big)^{1/2}\geqslant 1$, $k=1,2,\ldots$, и $\lim\limits_{k\to\infty}\rho(u_k,X_*)=1$, т. е. метод штрафов не сходится. В этой задаче функции $f(u)$, $g_1(u)$, $g_2(u)$, $g_3(u)$ являются полиномами, множество X выпукло и имеет внутренние точки, нижняя грань Φ_{k*} достигается.

Приведем пример задачи, в которой $\Phi_{k*}=-\infty$, $k=1,2,\ldots$

Пример 4. Задача: $f(x)=-x^2\to\inf$, $x\in X=\{x\in X_0=E^1:\ g(x)=|x|\leqslant 0\}$. Здесь $f_*=0$, $X_*=X=\{0\}$. Возьмем штрафную функцию $\Phi_k(x)=-x^2+k|x|$. Ясно, что $\Phi_{k*}=\inf\limits_{x\in E^1}\Phi_k(x)=-\infty$, $k=1,2,\ldots$, и условие (6) теряет смысл. В то же время, если в этой задаче мы выберем другую штрафную функцию, как, например, $\Phi_k(x)=-x^2+(k+1)x^2$ или $\Phi_k(x)=-x^2+kx^4$, то получим $\Phi_{k*}=0>-\infty$, причем нижняя грань достигается в точке $x_k=0$, $k=1,2,\ldots$

Замечание 1. Напомним, что неравенство (6) написано в предположении, что выполнено условие (4). Если $f_{**}=\inf\limits_{x\in X_0}f(x)>-\infty$, то из (3) и из $P_k(x)\geqslant 0$ $\forall x\in X_0$ следует, что $\Phi_{k*}>-\infty$, $k=1,2,\ldots$, при любом выборе штрафной функции.

Отметим, что в примерах 1, 2 выполняется условие $f_{**}>-\infty$, в примерах 3, 4 — $f_{**}=-\infty$. Для конкретных классов штрафных функций можно указать другие достаточные условия для выполнения (4). Так, например, если функция $P(x)$ взята из (9) или (9′), то $\Phi(x,A)=f(x)+AP(x)\leqslant\Phi(x,B)=f(x)+BP(x)$ $\forall A\leqslant B$, $\forall x\in X_0$. Поэтому $\Phi_*(A)=\inf\limits_{x\in X_0}\Phi(x,A)\leqslant\Phi_*(B)$ $\forall A\leqslant B$, т. е. функция $\Phi_*(A)$ монотонно растет (точнее, не убывает). Отсюда следует, что если $\Phi_*(A)>-\infty$ при некотором A, то $\Phi_{k*}=\Phi_*(A)>-\infty$ для всех k, для которых $A_k\geqslant A$.

Перейдем к исследованию вопросов сходимости метода штрафных функций для задачи (1), (8). Для определенности все формулировки и доказательства теорем проведем для штрафной функции (9), хотя некоторые из нижеследующих утверждений будут справедливы и для более широкого класса штрафных функций.

Теорема 1. *Пусть функции $f(x)$, $g_i(x)$, $i=1,\ldots,s$, определены на множестве X_0, а последовательность $\{x_k\}$ определена условиями* (3), (4), (6), (9). *Тогда*

$$\overline{\lim_{k\to\infty}}f(x_k)\leqslant\overline{\lim_{k\to\infty}}\Phi_k(x_k)=\overline{\lim_{k\to\infty}}\Phi_{k*}\leqslant f_*.\qquad(11)$$

Если, кроме того, $f_{**}=\inf\limits_{X_0}f(x)>-\infty$, то

$$P(x_k)=\sum_{i=1}^{s}(g_i^+(x_k))^p=O(A_k^{-1}),\quad k=1,2,\ldots,\qquad(12)$$

$$\overline{\lim_{k\to\infty}}g_i(x_k)\leqslant 0,\quad i=1,\ldots,m;\quad \lim_{k\to\infty}g_i(x_k)=0,\quad i=m+1,\ldots,s.\qquad(13)$$

Доказательство. Так как $P(x) \geqslant 0$, то из (3), (6), (9) имеем

$$f(x_k) \leqslant f(x_k) + A_k P(x_k) = \Phi_k(x_k) \leqslant \Phi_{k*} + \varepsilon_k \leqslant$$
$$\leqslant \Phi_k(x) + \varepsilon_k = f(x) + A_k P(x) + \varepsilon_k \quad \forall x \in X_0, \quad k = 1, 2, \ldots$$

Отсюда, переходя к нижней грани по $x \in X$ и учитывая, что $P(x) = 0$, $x \in X$, получим

$$f(x_k) \leqslant \Phi_k(x_k) \leqslant \Phi_{k*} + \varepsilon_k \leqslant f_* + \varepsilon_k, \quad k = 1, 2, \ldots \tag{14}$$

При $k \to \infty$ из (14) вытекает (11).

Пусть теперь $f_{**} > -\infty$. Так как $f_* \geqslant f_{**}$, то $f_* > -\infty$, а из (3), $P_k(x) \geqslant 0$, следует (4) (см. замечание 1). С учетом (14) имеем

$$0 \leqslant A_k P(x_k) = \Phi_k(x_k) - f(x_k) \leqslant f_* + \varepsilon_k - f_{**}, \quad k = 1, 2, \ldots$$

или

$$0 \leqslant P(x_k) \leqslant (f_* + \sup_{k \geqslant 0} \varepsilon_k - f_{**}) A_k^{-1}, \quad k = 1, 2, \ldots$$

Оценка (12) доказана. Из нее следует, что $\lim\limits_{k \to \infty} P(x_k) = 0$ или $\lim\limits_{k \to \infty} g_i^+(x_k) = 0$, $i = 1, \ldots, s$. Вспоминая определение (10) для $g_i^+(x)$, отсюда получим соотношения (13). □

Примеры 2, 3 показывают, что в общем случае неравенства в (11) могут быть строгими. Приведем достаточные условия, когда справедливы равенства (7).

Теорема 2. *Пусть X_0 — замкнутое множество из E^n, функции $f(x), g_1(x), \ldots, g_m(x), |g_{m+1}(x)|, \ldots, |g_s(x)|$ полунепрерывны снизу на X_0, $f_{**} = \inf\limits_{X_0} f(x) > -\infty$. Пусть последовательность $\{x_k\}$, определяемая условиями* (3), (6), (9), *имеет хотя бы одну предельную точку. Тогда все предельные точки $\{x_k\}$ принадлежат множеству X_* точек минимума задачи* (1), (8). *Если, кроме того, множество*

$$X_\delta = \{x: x \in X_0, \ g_i^+(x) \leqslant \delta, \ i = 1, \ldots, s\} \tag{15}$$

ограничено хотя бы при одном значении $\delta > 0$, то для последовательности $\{x_k\}$ выполняются равенства (7).

Доказательство. При сделанных предположениях для последовательности $\{x_k\}$ соотношения (11)–(13) сохраняют силу. Пусть v_* — какая-либо предельная точка последовательности $\{x_k\}$, пусть $\{x_{k_r}\} \to v_*$. Заметим, что $v_* \in X_0$ в силу замкнутости X_0. Тогда с учетом полунепрерывности снизу указанных в условии теоремы функций из соотношений (13) получим

$$g_i(v_*) \leqslant \varliminf_{r \to \infty} g_i(x_{k_r}) \leqslant \varlimsup_{k \to \infty} g_i(x_k) \leqslant 0, \quad i = 1, \ldots, m,$$
$$|g_i(v_*)| \leqslant \varliminf_{r \to \infty} |g_i(x_{k_r})| = \varlimsup_{k \to \infty} |g_i(x_k)| = 0, \quad i = m+1, \ldots, s.$$

Следовательно, $v_* \in X$. Тогда с учетом (11) имеем $f_* \leqslant f(v_*) \leqslant \varliminf\limits_{r \to \infty} f(x_{k_r}) \leqslant \varlimsup\limits_{k \to \infty} f(x_k) \leqslant f_*$, т. е. $\lim\limits_{r \to \infty} f(x_{k_r}) = f(v_*) = f_*$ или $v_* \in X_*$.

Наконец, пусть множество (15) ограничено при некоторых $\delta > 0$. Из соотношений (13) следует, что $\{x_k\} \in X_\delta$ для всех $k \geqslant k_0$. Это означает, что $\{x_k\}$ имеет хотя бы одну предельную точку. Тогда, как было выше показано,

все предельные точки $\{x_k\}$ принадлежат X_*. Следовательно, $\lim\limits_{k \to \infty} \rho(x_k, X_*) = 0$. Из тех же рассуждений и неравенств (11) вытекает первое равенство (7). Теорема 2 доказана. □

Для иллюстрации теоремы 2 рассмотрим

Пример 5. Пусть

$$f(x) = e^{-x} \to \inf; \quad x \in X = \{x \in E^1: g(x) = x = 0\}.$$

Здесь $f_* = 1$, $X_* = \{0\}$. Функции $f(x), g(x)$ непрерывны на замкнутом множестве $X_0 = E^1$, $f_{**} = \inf\limits_{E^1} e^{-x} = 0$, множество $X_\delta = \{x \in E^1: |x| \leqslant \delta\}$ ограничено при любом $\delta > 0$. Таким образом, все условия теоремы 2 выполнены. Возьмем штрафную функцию $P(x) = (g(x))^2 = x^2$ и положим

$$\Phi_k(x) = e^{-x} + kx^2, \quad x \in E^1, \quad k = 1, 2, \ldots$$

Нетрудно видеть, что $\Phi_k(x)$ сильно выпукла на E^1, поэтому $\Phi_{k*} = \inf\limits_{E^1} \Phi_k(x) > -\infty$. Пусть $\{\varepsilon_k\}$ — произвольная неотрицательная последовательность, стремящаяся к нулю. Определим точку x_k из условия $\Phi_k(x_k) \leqslant \Phi_{k*} + \varepsilon_k$, $k = 1, 2, \ldots$. Для получаемой таким образом последовательности $\{x_k\}$ согласно теореме 2 имеют место равенства (7).

3. Нетрудно видеть, что рассмотренные в примерах 2 и 5 задачи по существу одинаковые: минимизируется одна и та же функция e^{-x} на одном и том же множестве $X = \{0\}$, и отличие лишь в том, что в примере 2 множество X задается ограничениями $g(x) = xe^{-x} = 0$, а в примере 5 — $g(x) = x = 0$. Тем не менее, в примере 2 метод штрафных функций расходится, в примере 5 сходится.

Отсюда заключаем, что для сходимости метода штрафной функции важное значение имеет способ задания множества X: ограничения, задающие множество X и штрафные функции этого множества должны быть как-то согласованы с минимизируемой функцией $f(x)$.

Определение 2. Скажем, что задача (1), (8) имеет *согласованную постановку* на множестве X_0, если для любой последовательности $\{x_k\} \in X_0$, для которой

$$\lim_{k \to \infty} g_i^+(x_k) = 0, \quad i = 1, \ldots, s, \tag{16}$$

имеет место соотношение

$$\varliminf_{k \to \infty} f(x_k) \geqslant f_* = \inf_X f(x). \tag{17}$$

Отметим, что в примере 5 задача имеет согласованную постановку на E^1, а в примере 2 такой согласованности нет.

Теорема 3. *Пусть $\Phi_k(x) = f(x) + A_k P(x)$, где $P(x)$ определена формулой* (9), *пусть $\Phi_{k*} = \inf\limits_{X_0} \Phi_k(x)$, $k = 1, 2, \ldots$ Тогда для того чтобы*

$$\lim_{k \to \infty} \Phi_{k*} = f_*, \tag{18}$$

необходимо, чтобы задача (1), (8) *имела согласованную постановку на множестве X_0. Если $f_{**} = \inf\limits_{X_0} f(x) > -\infty$, то согласованной постановки задачи* (1), (8) *на X_0 достаточно для справедливости равенства* (18).

Доказательство. Необходимость. Пусть имеет место равенство (18). Возьмем произвольную последовательность $\{x_r\} \in X_0$, удовлетворяющую условиям (16). Тогда $\lim\limits_{r\to\infty} P(x_r) = 0$. Справедливы неравенства $\Phi_{k*} \leqslant \Phi_k(x_r) \leqslant f(x_r) + A_k P(x_r)$, $r = 1, 2, \ldots$ Отсюда при $r \to \infty$ получим $\Phi_{k*} \leqslant \lim\limits_{r\to\infty} f(x_r)$ при всех $k = 1, 2, \ldots$ Переходя здесь к пределу при $k \to \infty$, с учетом (18) будем иметь $\lim\limits_{r\to\infty} f(x_r) \geqslant \lim\limits_{k\to\infty} \Phi_{k*} = f_*$, что и требовалось.

Достаточность. Пусть $f_{**} > -\infty$, задача (1), (8) имеет согласованную постановку на множестве X_0. Поскольку $\Phi_k(x) \geqslant f(x)$ при всех $x \in X_0$, то $\Phi_{k*} \geqslant f_{**} > -\infty$, и имеет смысл говорить о последовательностях, удовлетворяющих условиям (6). Возьмем одну из таких последовательностей $\{x_k\}$. Согласно теореме 1 тогда справедливы соотношения (11)–(13). Заметим, что (13) равносильно (16), откуда следует (17). Из (14), (17) получим $\lim\limits_{k\to\infty} f(x_k) = \lim\limits_{k\to\infty} \Phi_{k*} = f_*$. Теорема 3 доказана. □

Утверждения, уточняющие и дополняющие теорему 3 и определение 2, приведены ниже в упражнениях 14–16.

Класс задач (1), (8), имеющих согласованную постановку на X_0, указан в теореме 2. Другой такой класс задач выделяется в следующей лемме.

Лемма 1. *Пусть в задаче* (1), (8) *функция Лагранжа* $L(x, \lambda) = f(x) + \sum\limits_{i=1}^{s} \lambda_i g_i(x)$, $x \in X_0$, $\lambda \in \Lambda_0 = \{\lambda = (\lambda_1, \ldots, \lambda_s) \in E^s : \lambda_1 \geqslant 0, \ldots, \lambda_m \geqslant 0\}$ *имеет седловую точку на* $X_0 \times \Lambda_0$. *Тогда задача* (1), (8) *имеет согласованную постановку на* X_0.

Доказательство. Пусть $(x_*, \lambda^*) \in X_0 \times \Lambda_0$ — седловая точка функции $L(x, \lambda)$, т. е.

$$L(x_*, \lambda) \leqslant L(x_*, \lambda^*) \leqslant L(x, \lambda^*) \quad \forall x \in X_0, \quad \lambda \in \Lambda_0. \tag{19}$$

Согласно теореме 4.9.1 тогда $x_* \in X_*$, $f(x_*) = f_* = L(x_*, \lambda^*)$. Из определения (10) функции $g_i^+(x)$ с учетом условия $\lambda \in \Lambda_0$ имеем

$$\lambda_i^* g_i(x) \leqslant |\lambda_i^*| g_i^+(x) \quad \forall x \in X_0, \quad i = 1, \ldots, s.$$

Отсюда и из (19) получим

$$f_* \leqslant f(x) + \sum_{i=1}^{s} |\lambda_i^*| g_i^+(x) \quad \forall x \in X_0. \tag{20}$$

Возьмем любую последовательность $\{x_k\} \in X_0$, которая удовлетворяет условиям (16). Тогда из (20) при $x = x_k$ получим $\lim\limits_{k\to\infty} f(x_k) \geqslant f_*$, т. е. задача (1), (8) имеет согласованную постановку на X_0. □

Из теорем 1, 3, леммы 1 следует

Теорема 4. *Пусть функция Лагранжа задачи* (1), (8) *имеет седловую точку и* $f_{**} = \inf\limits_{X_0} f(x) > -\infty$, *пусть последовательность* $\{x_k\}$ *определена условиями* (3), (6), (9). *Тогда* $\lim\limits_{k\to\infty} f(x_k) = \lim\limits_{k\to\infty} \Phi_k(x_k) = \lim\limits_{k\to\infty} \Phi_{k*} = f_*$ *и справедливы соотношения* (11), (12).

4. Покажем, что теорема 4 сохраняет силу и без требования $f_{**} > -\infty$. Более того, для задач, у которых функция Лагранжа имеет седловую точку и даже для несколько более общего класса задач (1), (8), можно получить оценку скорости сходимости метода штрафных функций.

Определение 3. Скажем, что задача (1), (8) имеет *сильно согласованную постановку*, если найдутся такие числа $c_1 \geqslant 0, \ldots, c_s \geqslant 0$, $\nu > 0$, что

$$f_* \leqslant f(x) + \sum_{i=1}^{s} c_i (g_i^+(x))^\nu \quad \forall x \in X_0. \tag{21}$$

Как видно из неравенства (20), задачи (1), (8), функция Лагранжа которых обладает седловой точкой, имеют сильно согласованную постановку, причем в (21) можно взять $c_i = |\lambda_i^*|$, $\nu = 1$. Другой важный класс задач с сильно согласованной постановкой будет приведем ниже в лемме 5. Заметим, что неравенство (21) обобщает очевидное неравенство $f_* \leqslant f(x)$ $\forall x \in X$ на более широкое множество X_0.

Теорема 5. *Пусть задача* (1), (8) *имеет сильно согласованную постановку в смысле определения* 3, $f_* > -\infty$, *последовательность* $\{x_k\}$ *определена условиями* (3), (6), (9), *где* $p > \nu$. *Тогда*

$$0 \leqslant (g_i^+(x_k))^p \leqslant P(x_k) \leqslant \rho_k, \quad k = 1, 2, \ldots, \tag{22}$$

$$-|c|(\rho_k)^{\nu/p} \leqslant f(x_k) - f_* \leqslant \varepsilon_k, \quad k = 1, 2, \ldots, \tag{23}$$

$$-BA_k^{-\nu/(p-\nu)} \leqslant \Phi_k(x_k) - f_* \leqslant \varepsilon_k, \quad -BA_k^{-\nu/(p-\nu)} \leqslant \Phi_{k*} - f_* \leqslant \varepsilon_k, \quad k = 1, 2, \ldots, \tag{24}$$

где

$$\rho_k = \left(\frac{|c|}{A_k}\right)^{p/(p-\nu)} + \frac{p}{p-\nu}\frac{\varepsilon_k}{A_k}, \quad k = 1, 2, \ldots$$

$$|c| = \left(\sum_{i=1}^{s} |c_i|^{p/(p-\nu)}\right)^{p/(p-\nu)}, \quad B = (p-\nu)\nu^{\nu/(p-\nu)} p^{-p/(p-\nu)} |c|^{p/(p-\nu)}.$$

Если, кроме того, X_0 *замкнутое множество, функции* $f(x)$, $g_i^+(x)$ *полунепрерывны снизу на* X_0, $\{A_k\} \to \infty$, $\{\varepsilon_k\} \to 0$, *и* v_* *— предельная точка последовательности* $\{x_k\}$, *то* $v_* \in X_*$.

Доказательство. Прежде всего покажем, что $\Phi_{k*} > -\infty$. Из (21) следует

$$\Phi_k(x) - f_* = f(x) - f_* + A_k P(x) \geqslant -\sum_{i=1}^{s} c_i (g_i^+(x))^\nu + A_k \sum_{i=1}^{s} (g_i^+(x))^p \geqslant \sum_{i=1}^{s} \min_{z \geqslant 0} (-c_i z^\nu + A_k z^p) \quad \forall x \in X_0. \tag{25}$$

Нетрудно видеть, что функция $\varphi(z) = -c_i z^\nu + A_k z^p$, где $p > \nu$, достигает своей нижней грани при $z \geqslant 0$ в точке $z_* = \left(\frac{\nu c_i}{p A_k}\right)^{1/(p-\nu)}$, причем $\varphi(z_*) = \min\limits_{z \geqslant 0} \varphi(z) = -\nu^{\nu/(p-\nu)} p^{-p/(p-\nu)} c_i^{p/(p-\nu)} (p-\nu) A_k^{-\nu/(p-\nu)}$. Отсюда и из (25) следует, что

$$\Phi_k(x) - f_* \geqslant -BA_k^{-\nu/(p-\nu)} \quad \forall x \in X_0. \tag{26}$$

Переходя к нижней грани по $x \in X_0$, из (26) имеем

$$\Phi_{k*} - f_* \geqslant -BA_k^{-\nu/(p-\nu)}. \tag{27}$$

Отсюда вытекает, что $\Phi_{k*} > -\infty$. Это значит, что при $\varepsilon_k > 0$ точка x_k, удовлетворяющая условиям (6), существует по определению нижней грани (при $\varepsilon_k = 0$ существование такой точки предполагается).

Далее, из (14), (21) имеем

$$f(x_k) + A_k P(x_k) \leqslant f_* + \varepsilon_k \leqslant f(x_k) + \sum_{i=1}^{s} c_i (g_i^+(x_k))^\nu + \varepsilon_k, \tag{28}$$

так что

$$0 \leqslant A_k P(x_k) \leqslant \sum_{i=1}^{s} c_i (g_i^+(x_k))^\nu + \varepsilon_k, \quad k = 1, 2, \ldots \tag{29}$$

Пользуясь неравенством Гельдера

$$\left|\sum_{i=1}^{s} a_i b_i\right| \leqslant \left(\sum_{i=1}^{s} a_i^m\right)^{1/m} \left(\sum_{i=1}^{s} b_i^r\right)^{1/r}, \quad \frac{1}{m} + \frac{1}{r} = 1$$

при $b_i = c_i$, $a_i = (g_i^+(x_k))^\nu$, $m = p/\nu$, $r = p/(p-\nu)$, получаем

$$0 \leqslant \sum_{i=1}^{s} c_i (g_i^+(x_k))^\nu \leqslant |c| (P(x_k))^{\nu/p}. \tag{30}$$

Отсюда и из (29) следует $0 \leqslant A_k P(x_k) \leqslant |c| (P(x_k))^{\nu/p} + \varepsilon_k$ или $0 \leqslant z^{p/\nu} \leqslant |c| A_k^{-\nu/p} z + \varepsilon_k$, $k = 1, 2, \ldots$, где $z = (A_k P(x_k))^{\nu/p}$. С помощью леммы 2.6.11 тогда получаем

$$0 \leqslant (A_k P(x_k))^{\nu/p} \leqslant \left((|c| A_k^{-\nu/p})^{p/(p-\nu)} + \frac{p}{p-\nu}\varepsilon_k\right)^{\nu/p},$$

что равносильно оценке (22). Далее, из (28) с учетом (30) имеем

$$-|c| (P(x_k))^{\nu/p} \leqslant f(x_k) - f_* \leqslant \varepsilon_k.$$

Отсюда и из уже доказанной оценки (22) следует оценка (23). Левые неравенства (24) получаются из (26) при $x = x_k$ и из (27), правые неравенства (24) вытекают из (14). Наконец, утверждение о том, что предельная точка v_* последовательности $\{x_k\}$ принадлежит X_*, вытекает из оценок (22)–(24) и доказывается так же, как аналогичное утверждение в теореме 2.

Теорема 6. *Пусть задача* (1), (8) *имеет сильно согласованную постановку в смысле определения* 3, $f_* > -\infty$, *последовательность* $\{x_k\}$ *определена условиями* (3), (6), (9), *где* $p = \nu$, $A_k > |c| = \max\limits_{1 \leqslant i \leqslant s} |c_i|$. *Тогда*

$$0 \leqslant (g_i^+(x_k))^p \leqslant P(x_k) \leqslant \frac{\varepsilon_k}{A_k - |c|}, \tag{31}$$

$$-|c|\frac{\varepsilon_k}{A_k - |c|} \leqslant f(x_k) - f_* \leqslant \varepsilon_k, \tag{32}$$

$$0 \leqslant \Phi_k(x_k) - f_* \leqslant \varepsilon_k, \quad \Phi_{k*} = f_*. \tag{33}$$

Если $X_* \neq \varnothing$, *то*

$$X_* = X_{k*} = \{x \in X_0 \colon \Phi_k(x) = \Phi_{k*}\}. \tag{34}$$

Если, кроме того, X_0 *— замкнутое множество, функции* $f(x)$, $g_i^+(x)$ *полунепрерывны снизу на* X_0, $\{\varepsilon_k\} \to 0$, *и* v_* *— предельная точка* $\{x_k\}$, *то* $v_* \in X_*$.

Доказательство. Из (25) при $p = \nu$, $A_k > |c|$ имеем $\Phi_k(x) - f_* \geqslant 0$, $x \in X_0$, так что $\Phi_{k*} \geqslant f_* > -\infty$, и последовательность $\{x_k\}$, удовлетворяющая условиям (6), при $\varepsilon_k > 0$ существует. Из (28) при $p = \nu$, $A_k > |c|$ сразу получаем оценку (31). Из нее и из (29) при $p = \nu$ следует оценка (32). Из (14) и (25) при $p = \nu$, $A_k > |c|$ с учетом того, что $\min\limits_{z \geqslant 0}(-c_i z^\nu + A_k z^\nu) = 0$, приходим к соотношениям (33). Докажем равенство (34). Возьмем произвольную точку $x_* \in X_*$. Тогда $P(x_*) = 0$ и $\Phi_k(x_*) = f(x_*) = f_* = \Phi_{k*}$, так что $x_* \in X_{k*}$. Следовательно, $X_* \subset X_{k*}$. Пусть теперь $x_{k*} \in X_{k*}$, т. е. $\Phi_k(x_{k*}) = \Phi_{k*}$. Это значит, что условие (6) при $x_k = x_{k*}$ выполняется с $\varepsilon_k = 0$. Тогда из оценки (31) при $\varepsilon_k = 0$ получаем $P(x_{k*}) = 0$, т. е. $x_{k*} \in X$. Отсюда, из (33) и из того, что $f(x_{k*}) = \Phi_k(x_{k*}) = \Phi_{k*} = f_*$, следует, что $x_{k*} \in X_*$. Следовательно, $X_{k*} \subset X_*$. Равенство (34) доказано. Последнее утверждение теоремы вытекает из оценок (31)–(33) и доказывается так же, как аналогичное утверждение в теореме 2. □

Из теоремы 6 следует, что случай $p = \nu$ интересен тем, что при точной реализации метода штрафных функции (3), (6), (9) решение исходной задачи (1), (8) может быть получено при конечных значениях штрафного коэффициента A_k.

Определение 4. Пусть A — некоторый класс задач (1). Говорят, что штрафная функция $P_k(x)$ является *точной на классе* A, если существует номер k_0 такой, что множество решений задачи

$$\Phi_k(x) = f(x) + P_k(x) \to \inf, \quad x \in X_0$$

при всех $k \geqslant k_0$ совпадает со множеством решений произвольной задачи (1) из класса A.

Типичным примером точной штрафной функции на классе задач (1), (8), у которых функция Лагранжа имеет седловую точку, является функция $P_k(x) = A_k \sum\limits_{i=1}^{s} g_i^+(x)$, $A_k > |\lambda^*|_\infty$ [294]. Это следует из теоремы 6 при $p = \nu = 1$, $c = \lambda^*$ и леммы 1. О точных штрафных функциях см., например, [85; 266; 294; 562]

Рассмотрим примеры, которые показывают, что оценки, полученные в теоремах 5, 6, не могут быть существенно улучшены на классе задач (1), (8), имеющих сильно согласованную постановку.

Пример 6. Рассмотрим задачу

$$f(x) = -x \to \inf, \quad x \in X = \{x \in E^1 \colon g(x) = x \leqslant 0\}.$$

Здесь $f_* = 0$, $X_* = \{0\}$. Функция Лагранжа $L(x, \lambda) = -x + \lambda x$, $x \in X_0 = E^1$, $\lambda \in \Lambda_0 = \{\lambda \in E^1 \colon \lambda \geqslant 0\}$ имеет седловую точку $(x_* = 0, \lambda^* = 1)$, так что согласно (20) неравенство (21) выполнено при $\nu = 1$, $c_i = 1$, $s = 1$. Возьмем штрафную функцию $P_k(x) = A_k(g^+(x))^p = A_k(\max\{x; 0\})^p$, $p \geqslant 1$, $A_k > 1$, $\{A_k\} \to \infty$. Тогда функция (3) будет иметь вид

$$\Phi_k(x) = \begin{cases} -x + A_k x^p, & x \geqslant 0, \\ -x, & x < 0. \end{cases}$$

Нетрудно показать, что

$$\Phi_{k*} = \inf_{E^1} \Phi_k(x) = \begin{cases} 0, & p = \nu = 1, \\ -\dfrac{p-1}{p}(pA_k)^{-1/(p-1)}, & p > 1, \end{cases}$$

причем нижняя грань достигается в точке $x_{k*} = 0$ при $p = 1$ и $x_{k*} = (pA_k)^{-1/(p-1)}$ при $p > 1$, $k = 1, 2, \ldots$ Последовательность $\{x_k\}$ удовлетворяет условиям (5) или (6) с $\varepsilon_k = 0$, причем

$$P(x_{k*}) = \begin{cases} 0, & p = 1, \\ (pA_k)^{-p/(p-1)}, & p > 1, \end{cases} \quad f(x_{k*}) - f_* = \begin{cases} 0, & p = 1, \\ -(pA_k)^{-1/(p-1)}, & p > 1. \end{cases}$$

Сравнение этих точных равенств с оценками теорем 5, 6 при $\varepsilon_k = 0$ показывает, что в случае $p = 1$ оценки (31), (32) точны, а в случае $p > 1$ оценки (22)–(24) точны по порядку и отличаются от точных оценок лишь константами при степени A_k. Если $\varepsilon_k > 0$, то при $p = 1$ в качестве точки x_k, удовлетворяющей условиям (6) и наиболее удаленной от X_*, здесь можно взять $x_k = \varepsilon_k/(A_k - 1)$, $k = 1, 2, \ldots$. Тогда $P(x_k) = x_k = \varepsilon_k(A_k - |c|)^{-1}$, $f(x_k) - f_* = -x_k = -|c|\varepsilon_k(A_k - |c|)^{-1}$, $\Phi_k(x_k) - f_* = (A_k - 1)x_k = \varepsilon_k$, $k = 1, 2, \ldots$, что совпадает с оценками (31)–(33). Если $\varepsilon_k > 0$, $p = 2$, то точка $x_k = (1/(2A_k)) + (\varepsilon_k/A_k)^{1/2}$ удовлетворяет условиям (6), причем $A_k P(x_k) = A_k x_k^2 = (1/(4A_k)) + \varepsilon_k + (\varepsilon/A_k)^{1/2}$, $f(x_k) - f_* = -x_k$, $\Phi_k(x_k) - f_* = \varepsilon_k - (1/(4A_k))$, что также свидетельствует о том, что оценки (22)–(24) на классе задач с сильно согласованной постановкой не являются грубыми.

Этот же пример показывает, что в теореме 6 требование $A_k > |c|$ не может быть ослаблено. В самом деле, если $A_k < 1 = |c|$, $p = 1$, то $\Phi_{k*} = -\infty$, если же $A_k = 1 = |c|$, $p = 1$, то $\Phi_k(x) \equiv 0$, $\Phi_{k*} = 0$, $X_{k*} = E^1$ и нарушено равенство (34).

Пример 7. Рассмотрим задачу

$$f(x) = -x \to \inf, \quad x \in X = \{x \in E^1 \colon g(x) = x^2 \leqslant 0\}.$$

Здесь $f_* = 0$, $X_* = \{0\}$. Функция Лагранжа $L(x, \lambda) = -x + \lambda x^2$, $x \in E^1$, $\lambda \in E^1_+$, седловой точки не имеет, но тем не менее задача имеет сильно согласованную постановку. В самом деле, справедливо неравенство $f_* = 0 \leqslant -x + |x| = -x + (g(x))^{1/2}$ при всех $x \in E^1$, так что неравенство (21) выполняется при $c = 1$, $\nu = 1/2$. Возьмем штрафную функцию $P(x) = (\max\{x^2; 0\})^p = (x^2)^p$. Если $p > 1/2 = \nu$, то функция $\Phi_k(x) = -x + A_k x^{2p}$, $A_k > 0$, $\{A_k\} \to \infty$, достигает нижней грани на $X_0 = E^1$ при $x_{k*} = (2pA_k)^{-1/(2p-1)}$, $k = 1, 2, \ldots$, причем $P(x_{k*}) = (2pA_k)^{-2p/(2p-1)}$, $f(x_{k*}) - f_* = -x_{k*}$, $\Phi_{k*} - f_* = -(2p-1)((2p)^{2p}A_k)^{-1/(2p-1)}$, $k = 1, 2, \ldots$ Как видим эти оценки лишь константами при степенях A_k отличаются от оценок (22)–(24). Интересно заметить, что с увеличением p оценки ухудшаются. Если $p = \nu = 1/2$, $A_k > 1 = |c|$, то $\Phi_k(x) = -x + A_k|x|$, $\Phi_{k*} = 0$. Точка $x_k = \varepsilon_k(A_k - 1)^{-1}$ удовлетворяет условиям (6) и наиболее удалена от $X_* = \{0\}$. Тогда $P(x_k) = |x_k| = \varepsilon_k(A_k - 1)^{-1}$, $f(x_k) - f_* = -x_k$, $\Phi_k(x_k) - f_* = \varepsilon_k$, что совпадает с оценками (31)–(33).

5. Выполнение соотношений (13) или (16), как показывает пример 2, еще не гарантирует сходимость последовательности $\{x_k\}$ из (6) ко множеству X. Для такой сходимости множество должно удовлетворять некоторым дополнительным условиям.

Определение 5. Скажем, что множество (8) задано *корректными ограничениями* на X_0, если всякая последовательность $\{x_k\} \in X_0$, удовлетворяющая условиям (16), сходится ко множеству X.

Примеры 2, 5 показывают, что одно и то же множество может быть задано как корректными, так и некорректными ограничениями. Как следует из доказательства теоремы 2, ограничения из (8) будут корректными на X_0, если функции $g_i^+(x)$, $i = 1, \ldots, s$, полунепрерывны снизу на замкнутом множестве X_0, а множество $X(\delta)$, определяемое согласно (15), ограничено при некотором $\delta > 0$. Корректными будут также ограничения, для которых удается доказать неравенство

$$\rho(x, X) \leqslant h(g_1^+(x), \ldots, g_s^+(x)) \quad \forall x \in X_0, \tag{35}$$

где функция $h(t) = h(t_1, \ldots, t_s) > 0$ при всех $t \in E^s_+$, $t \neq 0$, $h(0) = 0$, $\lim\limits_{t \to \infty} h(t) = 0$. Приведем важные классы множеств (8), задаваемых корректными ограничениями, для которых неравенство (35) имеет вид

$$\rho(x, X) \leqslant M\Big(\max_{1 \leqslant i \leqslant s} g_i^+(x)\Big)^\gamma \quad \forall x \in X_0; \quad M > 0, \quad \gamma > 0. \tag{36}$$

Лемма 2. *Пусть* X_0 *— выпуклое замкнутое множество, функции* $g_1(x), \ldots, g_m(x)$ *выпуклы и непрерывны на* X_0, *пусть существует такая точка* $\bar{x} \in X_0$, *что* $g_1(\bar{x}) < 0, \ldots, g_m(\bar{x}) < 0$; *пусть множество* $X = \{x \in X_0 \colon g_1(x) \leqslant 0, \ldots, g_m(x) \leqslant 0\}$ *ограничено. Тогда неравенство* (36) *выполняется с* $\gamma = 1$, $M = \operatorname{diam} X \Big(\min\limits_{1 \leqslant i \leqslant m} |g_i(\bar{x})|\Big)^{-1}$, $\operatorname{diam} X = \sup\limits_{u, v \in X} |u - v|$.

Доказательство. Введем функцию $g(x)=\max\limits_{1\leqslant i\leqslant m} g_i(x)$. В силу теоремы 4.2.7 функция $g(x)$ выпукла на X_0. Возьмем произвольную точку $x\in X_0\setminus X$. Тогда $g(x)>0$. Функция $f(t)=g(x+t(\bar{x}-x))$ переменной t непрерывна на отрезке $[0,1]$, $f(0)=g(x)>0$, $f(1)=g(\bar{x})<0$. Следовательно, существует точка $t_0\in(0,1)$ такая, что $f(t_0)=0$. Положим $v=x+t_0(\bar{x}-x)$; тогда $t_0=|v-x||\bar{x}-x|^{-1}$, $1-t_0=|v-\bar{x}||\bar{x}-x|^{-1}$. Пользуясь выпуклостью функции $g(x)$, имеем $g(v)=f(t_0)=0\leqslant t_0 g(\bar{x})+(1-t_0)g(x)$ или $-t_0 g(\bar{x})\leqslant(1-t_0)g(x)$ или $|v-x||g(\bar{x})|\leqslant|v-\bar{x}|g^+(x)$. Отсюда с учетом $\bar{x}, v\in X$ получаем, что $\rho(x,X)\leqslant|x-v|\leqslant g^+(x)|v-\bar{x}|(g(\bar{x}))^{-1}=Mg^+(x)$, что и требовалось. □

Лемма 3 (Хоффман [796]). *Пусть* $X=\{x\in E^n\colon g_i(x)=\langle a_i,x\rangle-b^i\leqslant 0,\ i=1,\ldots,m\}\neq\varnothing$, *где* $a_i\in E^n$, $b^i\in\mathbb{R}$. *Тогда*

$$\rho(x,X)\leqslant M_1\max_{1\leqslant i\leqslant s} g_i^+(x)\quad\forall x\in E^n,\quad M_1=\mathrm{const}>0,$$

т. е. неравенство (36) *выполняется с* $\gamma=1$, $X_0=E^n$.

Доказательство. Возьмем произвольную точку $x\notin X$. Так как X — выпуклое замкнутое множество, то согласно теореме 4.4.1 однозначно определяется проекция $w=P_X(x)$ точки x на X. К задаче определения проекции: $g(y)=|y-x|\to\inf$, $y\in X$, применимы теорема 4.9.3 и лемма 4.9.2, которые гарантируют существование таких чисел $\lambda_1\geqslant 0,\ldots,\lambda_m\geqslant 0$, что

$$g'(w)+\sum_{i=1}^m\lambda_i g_i'(w)=\frac{w-x}{|w-x|}+\sum_{i=1}^s\lambda_i a_i=0,\quad\lambda_i(\langle a_i,w\rangle-b^i)=0,\quad i=1,\ldots,m.$$

Отсюда, учитывая, что $|w-x|=\rho(x,X)>0$ имеем

$$x-w=\rho(x,X)\sum_{i\in I(x)}\lambda_i a_i,\quad\Big|\sum_{i\in I(x)}\lambda_i a_i\Big|=1,\quad I(x)=\{i\colon 1\leqslant i\leqslant m,\ \lambda_i>0,\ \langle a_i,w\rangle-b^i=0\}.\quad(37)$$

Можно считать, что система векторов $\{a_i, i\in I(x)\}$ линейно независима. В самом деле, если существуют числа γ_i, $i\in I(x)$, не все равные нулю, $\sum\limits_{i\in I(x)}\gamma_i a_i=0$, то $x-w=\rho(x,X)\sum\limits_{i\in I(x)}(\lambda_i-t\gamma_i)a_i$, где $\nu_i=\lambda_i-t\gamma_i\geqslant 0$, $i\in I(x)$, при всех t, $0<t<t_0$, t_0 — достаточно малое число. Можно считать, что среди γ_i, $i\in I(x)$, есть положительные числа, иначе изменим знаки всех γ_i, $i\in I(x)$. Положим $t=\alpha_s/\gamma_s=\min\limits_{\gamma_i>0,\ i\in I(x)}\alpha_i/\gamma_i$. Тогда $\nu_i=\lambda_i-t\gamma_i\geqslant 0$, $i\in I(x)$, причем по крайней мере одно число $\nu_s=\lambda_s-t\gamma_s=0$. Таким образом, заменив в (37) λ_i на ν_i и исключив из $I(x)$ те номера, для которых $\nu_i=0$, снова придем к равенству вида (37) с меньшим числом слагаемых. Последовательно применяя этот прием далее, за конечное число шагов придем к представлению (37), в котором система $\{a_i, i\in I(x)\}$ линейно независима. Из (37) следует

$$\max_{1\leqslant i\leqslant m} g_i^+(x)\geqslant\max_{i\in I(x)} g_i^+(x)\geqslant\max_{i\in I(x)} g_i(x)=\max_{i\in I(x)}(\langle a_i,x\rangle-b^i)=$$
$$=\max_{i\in I(x)}\langle a_i,x-w\rangle=\rho(x,X)\max_{i\in I(x)}\langle a_i,\sum_{j\in I(x)}\lambda_j a_j\rangle.\quad(38)$$

Покажем, что величину $\max\limits_{i\in I(x)}\langle a_i,\sum\limits_{j\in I(x)}\lambda_j a_j\rangle$, где система $\{a_i, i\in I(x)\}$ линейно независима, можно оценить снизу положительной величиной, не зависящей от x. С этой целью возьмем любое множество индексов $I\subset\{1,\ldots,m\}$ таких, что векторы $\{a_i, i\in I\}$ линейно независимы, и введем множество

$$\Lambda_I=\Big\{(\lambda_i,\ i\in I)\colon\lambda_i\geqslant 0,\ \Big|\sum_{i\in I}\lambda_i a_i\Big|=1\Big\}.\quad(39)$$

Заметим, что Λ_I — замкнутое ограниченное множество. В самом деле, если $\lambda^k=(\lambda_i^k,\ i\in I)\in\Lambda_I$, $\lambda^k\to\lambda$, то предельным переходом в (39) легко убедиться, что $\lambda\in\Lambda_I$. Следовательно, Λ_I замкнуто. Покажем ограниченность Λ_I. Допустим противное: пусть найдутся $\lambda^k\in\Lambda_I$, $k=1,2,\ldots$, $|\lambda^k|\to\infty$. Тогда последовательность $\mu^k=\lambda^k/|\lambda^k|$, $k=1,2,\ldots$, ограничена: $|\mu^k|=1$. Выбирая при необходимости подпоследовательность, можем считать, что $\{\mu^k\}\to\mu$, $|\mu|=1$. Поскольку $\Big|\sum\limits_{i\in I}\lambda_i^k a_i\Big|=1$, то $\Big|\sum\limits_{i\in I}\mu_i^k a_i\Big|=1/|\lambda^k|\to 0=\sum\limits_{i\in I}\mu_i^0 a_i$, где $\mu=(\mu_i^0,\ i\in I)\neq 0$. Однако это противоречит линейной независимости $\{a_i, i\in I\}$. Следовательно, Λ_I ограничено.

На множестве Λ_I рассмотрим функцию $d(\lambda,I)=\max\limits_{i\in I}\langle a_i,\sum\limits_{j\in I}\lambda_j a_j\rangle$. Убедимся в том, что $d(\lambda,I)>0$ при всех $\lambda\in\Lambda_I$. В самом деле, если существует $\lambda^0=(\lambda_j^0,\ j\in I)\in\Lambda_I$, что $d(\lambda^0,I)\leqslant 0$, то $\langle a_i,\sum\limits_{j\in I}\lambda_j^0 a_j\rangle\leqslant 0$ при всех $i\in I$. Умножим эти неравенства на λ_i^0, $i\in I$, и сложим; получим равенство $\Big|\sum\limits_{i\in I}\lambda_i^0 a_i\Big|=0$, противоречащее определению Λ_I. Таким образом, $d(\lambda,I)>0$ при всех $\lambda\in\Lambda_I$. Функция $d(\lambda,I)$ полунепрерывна снизу на Λ_I. В самом деле, пусть $\lambda\in\Lambda_I$, $\{\lambda^k\}\to\lambda$, $\lambda^k\in\Lambda_I$, $k=1,2,\ldots$, пусть $i_0\in I$ и $d(\lambda,I)=\langle a_{i_0},\sum\limits_{j\in I}\lambda_j a_j\rangle$. Тогда $d(\lambda^k,I)=\max\limits_{i\in I}\langle a_i,\sum\limits_{j\in I}\lambda_j^k a_j\rangle\geqslant\langle a_{i_0},\sum\limits_{j\in I}\lambda_j^k a_j\rangle$. Отсюда при $k\to\infty$ получим $\varliminf\limits_{k\to\infty} d(\lambda^k,I)\geqslant\langle a_{i_0},\sum\limits_{j\in I}\lambda_j a_j\rangle=d(\lambda,I)$ $\forall\lambda\in\Lambda_I$. Согласно теореме 2.1.1 полунепрерывная снизу функция $d(\lambda,I)$ на компактном множестве Λ_I достигает своей нижней грани в некоторой точке $\lambda_*\in\Lambda_I$, причем $d_*(I)=\inf\limits_{\lambda\in\Lambda_I} d(\lambda,I)=d(\lambda^*,I)>0$. Поскольку множество $\{I\}$ различных подмножеств I множества $\{1,\ldots,m\}$, для которых векторы $\{a_i,\ i\in I\}$ линейно независимы, конечно, то $d_*=\inf\limits_{\{I\}} d_*(I)>0$ Отсюда и из (37), (38) имеем $\max\limits_{1\leqslant i\leqslant m} g_i^+(x)\geqslant\rho(x,X)d(\lambda,I(x))\geqslant\rho(x,X)d_*$, или $\rho(x,X)\leqslant(1/d_*)\max\limits_{1\leqslant i\leqslant s} g_i^+(x)$, $x\in E^n$. Таким образом, неравенство (36) справедливо с $\gamma=1$, $M=1/d_*$. Лемма 3 доказана. □

Лемма 4. *Пусть множество*

$$X=\{x\in E^n\colon g_i(x)=\langle a_i,x\rangle-b^i\leqslant 0,\ i=1,\ldots,m;\ gi(x)=\langle a_i,x\rangle-b^i=0,\ i=m+1,\ldots,s\},\quad(40)$$

где $a_i\in E^n$, $b^i\in\mathbb{R}$ *непусто. Тогда*

$$\rho(x,X)\leqslant M\max_{1\leqslant i\leqslant s} g_i^+(x)\quad\forall x\in E^n,\quad M=\mathrm{const}.\quad(41)$$

Доказательство. Каждое ограничение $g_i(x)=\langle a_i,x\rangle-b^i=0$ заменим равносильными ограничениями $h_{1i}(x)=g_i(x)\leqslant 0$, $h_{2i}(x)=-g_i(x)\leqslant 0$ и воспользуемся леммой 3. Получим

$$\rho(x,X)\leqslant M_1\max\{g_1^+(x),\ldots,g_m^+(x);\ h_{1m+1}^+(x),\ldots,h_{1s}^+(x),h_{2m+1}^+(x),\ldots,h_{2s}^+(x)\}.$$

Отсюда и из

$$h_{1s}^+(x)=\max\{g_i(x);0\}\leqslant|g_i(x)|=g_i^+(x),$$
$$h_{2s}^+(x)=\max\{-g_i(x);0\}\leqslant|g_i(x)|=g_i^+(x),\quad i=m+1,\ldots,s,$$

приходим к неравенству (41). Лемма 4 доказана. □

Другие классы множеств (8), заданных корректными ограничениями, читатель найдет в [84; 527; 670].

6. В лемме 1 был выделен класс задач (1), (8), имеющих сильно согласованную постановку (см. неравенства (20), (21)). Следуя [670], приведем еще один содержательный класс таких задач.

Лемма 5. *Пусть функция* $f(x)$ *на множестве* X_0 *удовлетворяет условию Гельдера*

$$|f(x)-f(y)|\leqslant L|x-y|^\alpha\quad\forall x,y\in X_0,\quad L>0,\quad 0<\alpha\leqslant 1;\quad(42)$$

ограничения, задающие множество (8), *корректны на* X_0 *и удовлетворяют неравенству* (36). *Тогда задача* (1), (8) *имеет сильно согласованную постановку, причем неравенство* (21) *выполняется при* $c_1=\ldots=c_s=LM^\alpha$, $\nu=\alpha\gamma$.

Доказательство. Возьмем произвольную точку $x\in X_0$. По определению $\rho(x,X)=\inf\limits_{v\in X}|x-v|$ для любого $\varepsilon>0$ найдется такая точка $x_\varepsilon\in X$, что $|x-x_\varepsilon|\leqslant\rho(x,X)+\varepsilon$. Тогда с учетом условий (36), (42) имеем

$$LM^\alpha\sum_{i=1}^s(g_i^+(x))^{\alpha\gamma}+f(x)-f_*\geqslant LM^\alpha\Big(\max_{1\leqslant i\leqslant s} g_i^+(x)\Big)^{\alpha\gamma}+f(x)-f(x_\varepsilon)\geqslant$$
$$\geqslant L(\rho(x,X))^\alpha-L|x-x_\varepsilon|^\alpha\geqslant L(\rho(x,X))^\alpha-L(\rho(x,X)+\varepsilon)^\alpha.$$

Пользуясь произволом $\varepsilon>0$, отсюда при $\varepsilon\to 0$ получим

$$LM^\alpha\sum_{i=1}^s(g_i^+(x))^{\alpha\gamma}+f(x)-f_*\geqslant 0\quad\forall x\in X_0.$$

Лемма 5 доказана. □

Заметим, что в общем случае из выполнения условий леммы 5 не следует существование седловой точки функции Лагранжа задачи (1), (8) и, наоборот, существование седловой точки не гарантирует выполнение условий леммы 5. Это означает, что выделенные в леммах 1, 5 два класса задач (1), (8), имеющих сильно согласованную постановку, взаимно дополняют друг друга. Отметим также, что этими двумя классами не исчерпываются задачи (1), (8) с сильно согласованной постановкой. Поясним это на примере.

Пример 8. Рассмотрим задачу

$$f(x)=-x^{\alpha}\to\inf,\quad x\in X=\{x\geqslant 0:\ g(x)=x^{\beta}\leqslant 0\},\tag{43}$$

где $\alpha>0$, $\beta>0$, $X_0=\{x\in E^1:\ x\geqslant 0\}=E^1_+$. Ясно, что $X=X_*=\{0\}$, $f_*=0$.

Далее, имеем

$$f_*=0=-x^{\alpha}+(x^{\beta})^{\alpha/\beta}=f(x)+(g^+(x))^{\alpha/\beta}\quad\forall x\in X_0,$$

так что неравенство (21) выполняется при $s=m=1$, $c_i=1$, $\nu=\alpha/\beta$. Следовательно, задача (43) имеет сильно согласованную постановку и к ней применимы теоремы 5, 6. Отметим, что здесь $\rho(x,X)=|x-0|=|x|=(g^+(x))^{1/\beta}$, $x\in X_0$, т. е. условие (36) выполняется с $M=1$, $\gamma=1/\beta$. Далее, при $0<\alpha\leqslant 1$ функция $f(x)=-x^{\alpha}$ удовлетворяет условию Гельдера: $|x^{\alpha}-y^{\alpha}|\leqslant|x-y|^{\alpha}$, $x, y\in X_0$, так что в этом случае применима лемма 5. При $\alpha>1$ условие Гельдера на $X_0=E^1_+$ не выполняется и лемма 5 неприменима. Далее, функция Лагранжа $L(x,\lambda)=-x^{\alpha}+\lambda x^{\beta}$, $x\geqslant 0$, $\lambda\geqslant 0$, задачи (43) при $\alpha=\beta$ имеет седловую точку $(x_*=0,\lambda^*=1)$. Кстати, седловая точка здесь не единственная: любая точка $(0,\lambda^*)$, $\lambda^*\geqslant 1$, также является седловой. Заметим также, что функция $f(x)=-x^{\alpha}$, $x\geqslant 0$, выпукла лишь при $0<\alpha\leqslant 1$, а $g(x)=x^{\beta}$, $x\geqslant 0$, выпукла лишь при $\beta\geqslant 1$. Если $\alpha\neq\beta$, то функция Лагранжа не имеет седловой точки. Таким образом, при $\alpha>1$, $\beta>0$, $\alpha\neq\beta$ задача (43) не охватывается леммами 1, 5.

7. Рассмотренный выше метод штрафных функций дает простую и универсальную схему решения задач минимизации на множествах, не совпадающих со всем пространством, и часто применяется на практике. Поскольку имеется достаточно богатый выбор штрафных функций, то при составлении функции $\Phi_k(x)$ можно постараться обеспечить нужную гладкость этой функции, выпуклость, подумать об удобствах вычисления значений функции и требуемых ее производных и т. п. Кроме того, имеется определенная свобода в выборе множества X_0 для задачи (2): в задании множества (8) всегда можно отнести ко множеству X_0 наиболее простые ограничения (например, X_0 может быть шаром или параллелепипедом в E^n, совпадать с полупространством или со всем пространством E^n и т. д.), а остальные ограничения оформить в виде $g_i(x)\leqslant 0$ или $g_j(x)=0$ и учесть их с помощью штрафной функции. Поэтому можно надеяться на то, что вспомогательные задачи (2), (3) удастся сформировать более простыми, более удобными для применения известных и сложных методов минимизации, чем исходная задача (1).

Следует заметить, что хотя сама схема метода штрафных функций довольно проста, но при практическом использовании этого метода для решения конкретных задач минимизации могут встретиться серьезные трудности. Дело в том, что для получения хорошего приближения решения задачи (1) номер k в (2), (3) (или штрафной коэффициент A_k в (9)) приходится брать достаточно большим. А с увеличением номера k свойства функции $\varphi_k(x)=f(x)+P_k(x)$, $x\in X_0$, оказывается, во многих случаях начинают ухудшаться: эта функция может стать более овражной, некоторые координаты градиента $\Phi_k'(x)$ могут быть слишком большими, могут появиться дополнительные локальные минимумы и т. п. Это все может привести к тому, что при больших k методы минимизации, используемые для решения задачи (2), будут плохо сходиться и определение точки x_k, удовлетворяющей условиям (6), с возрастанием k может потребовать все большего и большего объема вычислительной работы.

Поэтому при практическом применении метода штрафных функций вспомогательные задачи (2) обычно решают лишь для таких номеров k (возможно больших), для которых удается обеспечить достаточно быстрое убывание функции $f(x)$ и достаточную близость получаемых точек ко множеству X при небольшом объеме вычислительной работы. Если полученное на этом пути приближение к решению задачи (1) недостаточно хорошее, то привлекают более тонкие и, вообще говоря, более трудоемкие методы минимизации, стараясь при этом получше использовать ту информацию, которая получена с помощью метода штрафных функций.

Заметим, что если выполнены условия теоремы 6, то штрафная функция (9) множества (8) при $p=\nu$ будет точной и нет необходимости неограниченно увеличивать штрафной коэффициент A_k, и в этом случае упомянутый недостаток метода штрафных функций, вообще говоря, не будет проявляться. Правда, штрафная функция (9) при $p=\nu$ не всегда будет обладать достаточной гладкостью, но появившиеся в последнее время методы минимизации, не требующие гладкости минимизируемой функции (см., например, [264; 265; 361; 386; 396; 426; 572; 586; 718; 769; 777]), позволяют надеяться, что численное решение задачи (2) в рассматриваемом случае не будет слишком трудным.

Отметим, что при описании и исследовании метода штрафных функций выше мы предполагали, что функция $f(x)$ и множество (8) известны точно. Если же указанные исходные данные известны лишь приближенно, то метод штрафных функций полезно регуляризовать — об этом подробнее см. гл. 9.

Различные прикладные и теоретические аспекты метода штрафных функций исследованы в [18–20; 84; 85; 151; 218; 222; 250; 266; 286; 294; 295; 302; 319; 343; 374; 377; 379; 471; 562; 603; 613; 670; 720; 721; 738; 759; 774; 785; 786].

Упражнения

1. Применить метод штрафных функций к задачам

а) $f(u)=x^2+y^2\to\inf$; $u\in X=\{u=(x,y)\in E^2:\ g(u)=-x-y+1\leqslant 0\}$ или $u\in X=\{u=(x,y)\in E^2:\ g(u)=-x-y+1=0\}$;

б) $f(u)=xy\to\inf$; $u\in X=\{u=(x,y)\in E^2:\ x^2+y^2\leqslant 25\}$ или $u\in X=\{u=(x,y)\in E^2:\ x^2+y^2=25\}$;

в) $f(u)=x^2+y^2+z^2\to\inf$; $u\in X=\{u=(x,y,z)\in E^2:\ x+y+z+1\leqslant 0\}$.

2. Применить метод штрафных функций к задачам из примеров § 2.3.

3. Пусть $f(x)=e^{-2x}$, а множество $X=\{x\in E^1:\ 0\leqslant x\leqslant 1\}$ задано либо ограничениями $g_1(x)=-x\leqslant 0$, $g_2(x)=x-1\leqslant 0$, либо $g(x)=|x|+|x-1|-1=0$, либо $g(x)=e^{-x}(|x|+|x-1|-1)=0$. Выяснить, в каких случаях задача $f(x)\to\inf$, $x\in X$, имеет согласованную или сильно согласованную постановку на E^1.

4. Пусть $\{P_k(x)\}$ — штрафная функция некоторого множества X. Пусть функция $\varphi(t)$ определена при $t\geqslant 0$, $\varphi(0)=0$, причем $\varphi(t)\to 0$ при $t\to 0$, $\varphi(t)\to\infty$ при $t\to\infty$. Показать, что тогда $\{\varphi(P_k(x))\}$ является штрафной функцией множества X.

5. Применить метод (3), (6), (9) к задаче $f(x)=x^2-x\to\inf$ и $x\in X=\{x\in E^1:\ g(x=x\leqslant 0\}$, взяв в качестве штрафной функции $P(x)=(\max\{0;x\})^2$. Получить точную оценку погрешности, сравнить ее с оценками из теоремы 5.

6. Пусть множество X задано либо ограничениями $g_1(x)=x-1\leqslant 0$, $g_2(x)=-x-1\leqslant 0$, либо $g(x)=e^{-x^2}(x^2-1)\leqslant 0$, либо $g(x)=x^2-1\leqslant 0$. Выяснить, какие из этих ограничений являются корректными на E^1 или $X_0=\{x\in E^1:\ -1\leqslant x\leqslant 1\}$.

7. Пусть $X=\{x\in E^n\colon g(x)\leqslant 0\}$, где $g(x)$ — непрерывная функция на E^n. Доказать, что для того, чтобы множество X было ограниченным и ограничение $g(x)\leqslant 0$ было корректным на E^n, необходимо и достаточно, чтобы множество $X_\delta=\{x\in E^n\colon g(x)\leqslant\delta\}$ было ограниченным хотя бы при одном $\delta>0$.

8. Пусть X_0 — выпуклое замкнутое множество из E^n, функция $g(x)$ выпукла и полунепрерывна снизу на X_0, и пусть множество $M(C)=\{x\in X_0\colon g(x)\leqslant C\}$ непусто и ограничено при некотором C. Доказать, что тогда ограничение $g(x)\leqslant 0$ корректно на X_0 (см. теорему 4.2.17).

9. Рассмотреть задачу: $f(x)=x\to\inf$; $x\in X=\{x\in E^1\colon g(x)=x^2+\varepsilon|x|\leqslant 0\}$, $\varepsilon>0$. Доказать, что здесь выполняется неравенство (36) с $M=1/\varepsilon$, $\gamma=1$. Установить существование седловой точки функции Лагранжа. Выполняются ли здесь условия теорем 4.9.2, 4.9.4?

10. Применить метод штрафных функций к задаче (43), получив оценки скорости сходимости метода, и сравнить их с оценками из теорем 5, 6.

11. Применить метод штрафных функций к задаче: $f(u)=x^2+(1-xy)^2\to\inf$; $u\in X=\{u=(x,y)\in E^2\colon g(u)=x-a=0\}$, исследовать его сходимость при различных значениях параметра a.

12. Доказать, что множество $X=\{x\in E^n\colon g_i(x)=\langle a_i,x\rangle-b^i=0,\ i=1,\ldots,s\}$, где $a_1,\ldots,a_s$ — линейно независимые векторы из E^n, $b^i\in\mathbb{R}$, является корректным на E^n и неравенство (36) выполняется с $\gamma=1$, $M=s\|A^\top(AA^\top)^{-1}\|$, A — матрица размера $s\times n$, строками которой являются векторы $a_1,\ldots,a_s$. Указание: воспользоваться результатами примера 4.4.3.

13. Пусть задача (1), (8) удовлетворяет условиям теоремы 4.9.2, причем $\bar x\notin X_*$. Доказать, что тогда $X_*=\{x\in X_0\colon \Phi(x)=\Phi_*\}$, где

$$\Phi(x)=\ln\frac{1}{f(\bar x)-f(x)}+\sum_{i=1}^{m}\max\Big\{0;\frac{g_i(x)}{g_i(\bar x)}\Big\},\quad \Phi_*=\inf_{X_0}\Phi(x).$$

14. Пусть в задаче (1), (8) $f_*>-\infty$, $X_*\neq\varnothing$, пусть функция $\Phi(x,A)=f(x)+A\Phi(x)$, где штрафная функция $P(x)$ определена формулой (9) при $p>0$, пусть $\Phi_*(A)=\inf\limits_{x\in X_0}\Phi(x,A)$, $A>0$. Для того чтобы $\lim\limits_{A\to+\infty}\Phi_*(A)=f_*$, необходимо и достаточно, чтобы задача (1), (8) имела согласованную постановку на X_0 (определение 2) и существовало число $A_0>0$, что $\Phi_*(A_0)>-\infty$. Доказать [18].

15. Пусть в задаче (1), (8) X_0 — выпуклое замкнутое множество, функции $f(x)$, $g_1(x),\ldots,g_m(x)$, $|g_{m+1}(x)|,\ldots,|g_s(x)|$ выпуклы на X_0, $f_*>-\infty$, пусть функции $\Phi(x,A)$, $\Phi_*(A)$ взяты из упражнения 14 при $p\geqslant 1$. Для того чтобы $\lim\limits_{A\to+\infty}\Phi_*(A)=f_*$, необходимо и достаточно, чтобы задача (1), (8) имела согласованную постановку на множестве X_0. Доказать [18].

16. Доказать, что согласованная постановка задачи (1), (8) на множестве X_0 равносильна тому, что функция $f_*(c)=\inf\limits_{x\in X(c)}f(x)$, где $X(c)=\{x\in X_0\colon g_i^+(x)\leqslant c_i,\ i=1,\ldots,s\}$, полунепрерывна снизу при $c_i\to+0$, $i=1,\ldots,s$.

17. Привести пример задачи минимизации, для которой существует точная штрафная функция (определение 4), дифференцируемая любое конечное число раз. Указание: рассмотреть задачу: $f(x)\equiv 0\to\inf$, $x\in X=\{x\in X_0=E^1\colon g(x)=-x\leqslant 0\}$, взять $P(x)=(\max\{-x;0\})^p$ $\forall p>0$.

18. Показать, что если в лемме 5 условие (42) выполняется не на всем множестве X_0, а лишь на множестве X_δ из (15), то метод штрафных функций может не сходиться. Указание: рассмотреть задачу из примера 4; взять $P(x)=\max\{|x|;0\}=|x|$.

19. Пусть W — открытое выпуклое множество, функции $f(x)$, $g_i(x)$, $i=1,\ldots,m$, выпуклы на W, $g_i(x)=\langle a_i,x\rangle-b^i$, $i=m+1,\ldots,s$; X_0 — выпукло, замкнуто и $X_0\subset W$, пусть в задаче (1), (8) $f_*>-\infty$, $X_*\neq\varnothing$. Доказать, что для того чтобы функция Лагранжа задачи (1), (8) имела седловую точку, необходимо и достаточно, чтобы неравенство (21) выполнялось с показателем $\nu=1$. Указание: необходимость см. в лемме 1; для доказательства достаточности применить теоремы 6, 4.6.4 к задаче $\Phi(x,A)=f(x)+A\sum\limits_{i=1}^{s}g_i^+(x)\to\inf$, $x\in X_0$. [83, 2-е издание].

20. Пусть каноническая задача линейного программирования

$$f(x)=\langle c,x\rangle\to\inf,\quad x\in X=\{x\in E^n\colon x\geqslant 0,\ Ax=b\},\quad b\in E^m,\quad b\geqslant 0,\qquad (44)$$

имеет решение. Докажите, что задача (44) равносильна следующей канонической задаче

$$g(z)=\langle c,x\rangle+M(u^1+\ldots+u^m)\to\inf,$$
$$z\in Z=\{z=(u,x)\in E^m\times E^n\colon u\geqslant 0,\ x\geqslant 0,\ u+Ax=b\}\qquad (45)$$

при всех достаточно больших $M>0$ (M-метод [179; 259; 374; 471; 775]). Убедитесь, что $z_0=(b,0)$ — угловая точка множества Z. Указание: рассмотрите задачу: $f_1(z)=\langle c,x\rangle\to\inf$, $z\in Z_1=\{z=(u,x)\geqslant 0\colon u+Ax=b,\ u=0\}$, равносильную (44), ограничение $u=0$ учтите с помощью штрафной функции $P(z)=|u|_1$, $u\geqslant 0$ и, пользуясь леммой 1 и теоремой 6 при $p=\nu=1$, $\varepsilon_k=0$, $A_k=M>|\lambda^*|_\infty$, $c=\lambda^*$ — решение двойственной к (44) задачи, установите, что $P(z)$ — точная штрафная функция.

21. Пусть X_* — множество решений задачи (1), (8), $X_*(A)$ — множество решений задачи $\Phi(x,A)=f(x)+AP(x)\to\inf$, $x\in X_0$, где функция $P(x)$ взята из (9), $\Phi_*(A)=\inf\limits_{x\in X_0}\Phi(x,A)$, $\lim\limits_{A\to\infty}\Phi_*(A)=f_*$. Можно ли тогда утверждать, что $\inf\limits_{x\in X_*,\,y\in X_*(A)}|x-y|\to 0$ при $A\to+\infty$. Указание: рассмотреть задачу [18]: $f(x)=\max\{\omega(x^1,x^2);\omega(1-x^1,x^2)\}-x^2\to\inf$, $x\in X=\{x=(x^1,x^2,x^3,x^4)\in E^4\colon g_1(x)=x^2\leqslant 0,\ g_2(x)=\omega(x^3,x^1)-x^2\leqslant 0,\ g_3(x)=\omega(x^4,x^3)-x^2\leqslant 0\}$, где $\omega(u,v)=\sqrt{u^2+v^2}-u$ — функция Белоусова — Андронова [84]; штрафную функцию взять равной $P(x)=\sum\limits_{i=1}^{3}(\max\{g_i(x);0\})^2$; показать, что задача выпукла, $f_*=0$, $X_*=\{x\in E^4\colon x^1=x^2=x^3=0,\ x^4\geqslant 0\}$, $\Phi_*(A)=\frac12-\sqrt{\frac14+t^2}+t-At^2$, $X_*(A)=\{x\in E^4\colon x^1=\frac12,\ x^2=t,\ x^3\geqslant-\frac{t}{2}+\frac{1}{8t},\ x^4\geqslant-\frac{t}{2}+\frac{(x^3)^2}{2t},\ 0<t<\frac{1}{2A}\}$; убедиться, что $\lim\limits_{A\to\infty}\Phi_*(A)=0$, $\lim\limits_{A\to\infty}\inf\limits_{x\in X_*,\,y\in X_*(A)}|x-y|=+\infty$. Другие примеры, другие типы сходимости $X_*(A)$ к X_* исследованы в [18–20].

22. Применить метод штрафных функций к задаче: $J(u)=u\to\inf$, $u\in U=\{u\in U_0\colon g(u)=\frac{u^2}{1+u^4}\leqslant 0\}$, где $U_0=\{u\in E^1\colon u\geqslant -a\}$, $0<a\leqslant+\infty$. Показать, что при $0<a<+\infty$ задача имеет сильно согласованную постановку (в (21) взять $s=m=1$, $c_1=\sqrt{1+a^4}$, $\gamma=\frac12$). Что будет при $a=+\infty$? Проверить, что условия лемм 5.15.1, 5.15.4 не выполняются при всех a, $0<a\leqslant+\infty$.

§ 16. Доказательство необходимых условий экстремума первого и второго порядков с помощью штрафных функций

1. Начнем с необходимых условий первого порядка — дадим другое и, по-видимому, более простое доказательство правила множителей Лагранжа, отличное от изложенного в § 4.8 и не опирающееся на теорему отделимости и теорему о неявных функциях, для задачи

$$f(x)\to\inf,\quad x\in X,\qquad (1)$$
$$X=\{x\in X_0\colon g_i(x)\leqslant 0,\ i=1,\ldots,m;\ g_i(x)=0,\ i=m+1,\ldots,s\}.\qquad (2)$$

Как и в § 4.8, введем функцию Лагранжа задачи (1), (2)

$$\mathcal{L}(x,\bar\lambda)=\lambda_0 f(x)+\sum_{i=1}^{s}\lambda_i g_i(x),\quad x\in X_0,$$
$$\bar\lambda=(\lambda_0,\ldots,\lambda_s),\quad \lambda_0\geqslant 0,\ldots,\lambda_m\geqslant 0.\qquad (3)$$

Теорема 1. *Пусть X_0 — выпуклое замкнутое множество из E^n, функции $f(x)$, $g_i(x)$, $i=1,\ldots,s$, определены на X_0. Пусть v — точка локального минимума в задаче* (1), (2), *пусть функции $f(x)$, $g_i(x)$, $i=1,\ldots$*

..., s, непрерывно дифференцируемы в некоторой окрестности $O(v,\varepsilon)\cap$ $\cap X_0$ точки v. Тогда существуют числа $\lambda_0, \lambda_1, \ldots, \lambda_s$, такие, что

$$\bar\lambda = (\lambda_0, \ldots, \lambda_s) \neq 0, \quad \lambda_0 \geqslant 0, \lambda_1 \geqslant 0, \ldots, \lambda_m \geqslant 0, \tag{4}$$

$$\langle \mathcal{L}_x(v, \bar\lambda), x - v\rangle \geqslant 0 \qquad \forall x \in X_0, \tag{5}$$

$$\lambda_i g_i(v) = 0, \quad i = 1, \ldots, m. \tag{6}$$

Как видим, в этой теореме на задачу (1), (2) наложены более жесткие ограничения, чем в теореме 4.8.1, что связано со способом доказательства, использующим штрафные функции. Множество всех $\bar\lambda$, удовлетворяющих условиям (4)–(6), как и выше, будем обозначать через $\Lambda(v)$ — это конус Лагранжа. В § 4.8 было замечено, что конус $\Lambda(v)$ выпуклый, а конус $\Lambda(v)\cup$ $\cup\{0\}$ замкнутый.

Доказательство теоремы 1. Введем функцию $g_0(x) = f(x) + |x-v|^2$, $x \in X_0$, множество $W_0 = X_0 \cap S(v,\gamma)$, где $S(v,\gamma) = \{x \in E^n: |x-v| \leqslant \gamma\}$, $\gamma > 0$. Так как v — точка локального минимума функции $f(x)$ на X, то можем считать число γ столь малым, что $\gamma < 1$ и $f(x) \geqslant f(v)$ $\forall x \in X \cap S(v,\gamma)$. Рассмотрим вспомогательную задачу минимизации

$$g_0(x) \to \inf, \quad x \in W = \{x \in W_0: g_i(x) \leqslant 0, \quad i = 1, \ldots, m; \quad g_i(x) = 0, \quad i = m+1, \ldots, s\} \tag{7}$$

Так как $g_0(x) > f(x) \geqslant f(v)$ при всех $x \in W$, $x \neq v$, причем $g_0(v) = f(v)$, то ясно, что v — единственное решение задачи (7) и $g_{0*} = \inf\limits_{x\in W} g_0(x) = f(v)$. Применим к задаче (7) метод штрафных функций. Введем функцию $\Phi_k(x) = g_0(x) + k\sum\limits_{i=1}^{m}(\max\{g_i(x);0\})^2 + k\sum\limits_{i=m+1}^{s} g_i^2(x)$, $x \in X_0$. Так как W_0 компактное множество, функции $g_0(x), \Phi_k(x)$ непрерывны на W_0, то $g_{0*} = \inf\limits_{x\in W_0} g_0(x) > -\infty$, $\Phi_{k*} = \inf\limits_{x\in W_0} \Phi_k(x) > -\infty$ и существует точка $x_k \in W_0$, для которой $\Phi_k(x_k) = \Phi_{k*}$. Далее, множество $W(\delta) = \{x \in W_0: g_i^+(x) \leqslant \delta,\ i = 1, \ldots, s\}$ ограничено при всех $\delta > 0$, так как W_0 ограничено. По теореме 15.2 тогда

$$\lim_{k\to\infty} |x_k - v| = 0, \quad \lim_{k\to\infty} g_0(x_k) = g_{0*} = f(v) = \lim_{k\to\infty} f(x_k). \tag{8}$$

Применяя теорему 4.2.3 к задаче: $\Phi_k(x) \to \inf$, $x \in W_0$, имеем

$$\langle \Phi_k'(x_k), x - x_k\rangle \geqslant 0 \qquad \forall x \in W_0. \tag{9}$$

Покажем, что неравенство (9) на самом деле верно для всех $x \in X_0$ при всех достаточно больших номерах k. Возьмем произвольную точку $x \in X_0$ и положим $x_{k\alpha} = x_k + \alpha(x - x_k)$, $0 \leqslant \alpha \leqslant 1$. Так как X_0 выпуклое множество, то $x_{k\alpha} \in X_0$. Далее, $|x_{k\alpha} - v| \leqslant |x_{k\alpha} - x_k| + |x_k - v| = \alpha|x - x_k| + |x_k - v|$. С учетом (8) имеем: $|x_k - v| < \frac{\gamma}{2}$ $\forall k \geqslant k_0$, $\alpha|x - x_k| < \frac{\gamma}{2}$ $\forall \alpha$, $0 < \alpha < \alpha_0 = \alpha_0(x) < 1$. Поэтому $|x_{k\alpha} - v| < \gamma$ или $x_{k\alpha} \in W_0$ $\forall k \geqslant k_0$, $\forall \alpha$, $0 < \alpha < \alpha_0$, и в (9) можем положить $x = x_{k\alpha}$. Получим $\langle \Phi_k'(x_k), \alpha(x - x_k)\rangle \geqslant 0$ при всех α, $0 < \alpha < \alpha_0 = \alpha_0(x)$, $k \geqslant k_0$. Следовательно,

$$\langle \Phi_k'(x_k), x - x_k\rangle \geqslant 0 \quad \forall x \in X_0, \quad \forall k \geqslant k_0. \tag{10}$$

Подставим в (10) явное выражение для производной $\Phi_k'(x_k) = f'(x_k) + 2(x_k - v) + \sum\limits_{i=1}^{m} 2k \max\{g_i(x_k);0\} g_i'(x_k) + \sum\limits_{i=m+1}^{s} 2k g_i(x_k) g_i'(x_k)$. Будем иметь

$$\langle f'(x_k) + 2(x_k - v) + \sum_{i=1}^{s} \mu_{ik} g_i'(x_k), x - x_k\rangle \geqslant 0 \quad \forall x \in X_0, \quad \forall k \geqslant k_0, \tag{11}$$

$$\mu_{ik} = \begin{cases} 2k \max\{g_i(x_k);0\} \geqslant 0, & i = 1, \ldots, m; \\ 2k g_i(x_k), & i = m+1, \ldots, s. \end{cases} \tag{12}$$

Разделим неравенство (11) на $\left(1 + \sum\limits_{i=1}^{s} \mu_{ik}^2\right)^{1/2} \geqslant 1$. Получим

$$\langle \lambda_{0k} f'(x_k) + 2\lambda_{0k}(x_k - v) + \sum_{i=1}^{s} \lambda_{ik} g_i'(x_k), x - x_k\rangle \geqslant 0 \quad \forall x \in X_0, \quad k \geqslant k_0, \tag{13}$$

где $\lambda_{0k} = \left(1 + \sum\limits_{i=1}^{s} \mu_{ik}^2\right)^{-1/2} > 0$, $\lambda_{ik} = \mu_{ik}\lambda_{0k}$, $i = 1, \ldots, s$, причем в силу (12) $\lambda_{ik} \geqslant 0$, $i = 1, \ldots, m$, $\forall k \geqslant k_0$. Последовательность $\{\bar\lambda_k = (\lambda_{0k}, \ldots, \lambda_{sk})\}$ ограничена, так как $|\bar\lambda_k| = 1$. Пользуясь теоремой Больцано — Вейерштрасса и выбирая при необходимости подпоследовательность, можем считать, что $\{\bar\lambda_k\} \to \bar\lambda = (\lambda_0, \ldots, \lambda_s)$, где $\lambda_i \geqslant 0$, $i = 0, \ldots, m$, $|\bar\lambda| = 1$. Как видим, условия (4) для полученного $\bar\lambda$ выполнены. Так как $f'(x)$, $g_i'(x)$ непрерывны и $\{x_k\} \to v$ в силу (8), то из (13) при $k \to \infty$ получим неравенство (5). Наконец, если $g_i(v) = 0$, то $\lambda_i g_i(v) = 0$. Если же $g_i(v) < 0$ при некотором i, $1 \leqslant i \leqslant m$, то $g_i(x_k) < 0$ $\forall k \geqslant k_1 \geqslant k_0$. Тогда из (12) видно, что $\mu_{ik} = 0$ и поэтому $\lambda_{ik} = 0$ $\forall k \geqslant k_1$. При $k \to \infty$ отсюда имеем $\lambda_i = 0$ при $g_i(v) < 0$, так что снова $\lambda_i g_i(v) = 0$. Равенства (6) получены. Теорема 1 доказана. □

Читатель, конечно, обратил внимание, что в условиях близких теорем 1 и 4.8.1 (см. также теоремы 2.3.1, 2.3.2) к функциям $f(x)$, $g_i(x)$, $i = 1, \ldots, s$, предъявляются несколько разные требования гладкости. Ясно, что это связано с методами доказательства упомянутых теорем. В связи с этим возникает интересный вопрос, каковы минимальные требования к гладкости функций $f(x)$, $g_i(x)$, $i = 1, \ldots, s$, для того, чтобы правило множителей Лагранжа оставалось справедливым? Более тонкие исследования показывают [40], что для этого достаточно, чтобы функции $f(x)$, $g_i(x)$, $i = 1, \ldots, s$, были непрерывны в окрестности точки v и дифференцируемы в точке v.

2. Перейдем к доказательству необходимых условий экстремума второго порядка (см. теоремы 2.4.2, 2.4.3) для задачи

$$f(x) \to \inf, \quad x \in X = \{x \in E^n: g_i(x) \leqslant 0,\ i = 1, \ldots, m,\ g_i(x) = 0,\ i = m+1, \ldots, s\}, \tag{13.A}$$

получающейся из задачи (1), (2) при $X_0 = E^n$. Пусть v — точка локального минимума задачи (13.A). Согласно теореме 1 конус Лагранжа $\Lambda(v)$, состоящий из точек $\bar\lambda = (\lambda_0, \ldots, \lambda_s)$, которые удовлетворяют условиям

$$\bar\lambda \neq 0, \quad \lambda_0 \geqslant 0, \ldots, \lambda_m \geqslant 0, \quad \mathcal{L}_x(v, \bar\lambda) = 0, \quad \lambda_i g_i(v) = 0, \quad i = 1, \ldots, m,$$

непуст. Напомним, что конусом Арутюнова точки v множества (2) называется подмножество $\Lambda_a(v)$ таких точек $\bar\lambda \in \Lambda(v)$, для каждой из которых существует сопровождающее подпространство $\Pi = \Pi(\bar\lambda) \subseteq E^n$ со свойствами:

$$\dim \Pi(\bar\lambda) \geqslant \max\{n - |I(v)|; 0\}, \tag{14}$$

$$\Pi(\bar\lambda) \subseteq \ker G'(v) = \{h \in E^n: \langle g_i'(v), h\rangle = 0,\ i \in I(v)\}, \tag{15}$$

$$\langle \mathcal{L}_{xx}(v, \bar\lambda)h, h\rangle \geqslant 0 \qquad \forall h \in \Pi(\bar\lambda), \tag{16}$$

где $I(v) = \{i: 1 \leqslant i \leqslant m,\ g_i(v) = 0\} \cup \{i: m+1 \leqslant i \leqslant s\}$ — множество индексов (номеров) активных ограничений точки v, $|I(v)|$ — количество элементов множества $I(v)$, $G(x) = \{g_i(x),\ i \in I(v)\}$.

В § 2.4 было показано, что так введенное множество $\Lambda_a(v)$ в самом деле является конусом. Как мы видели на примерах, конус Арутюнова, в отличие от конуса Лагранжа $\Lambda(v)$, необязательно является выпуклым и может иметь весьма сложную структуру (пример 2.4.6). Покажем, что конус $\Lambda_a(v)\cup\{0\}$ замкнут. Для этого сначала докажем одну важную лемму, обобщающую известную в математическом анализе теорему Больцано — Вейерштрасса.

Определение 1. Пусть задана последовательность множеств $\{\Pi_k\}\subseteq E^n$. *Верхним пределом* последовательности $\{\Pi_k\}$ называется множество, обозначаемое $\mathop{\mathcal{L}s}\limits_{k\to\infty}\Pi_k$ и состоящее из таких точек x_0, для каждой из которых существует своя последовательность $\{x_k\}$, $x_k\in\Pi_k$, $k=1,2,\ldots$, для которой x_0 является предельной точкой [428, стр. 344].

Лемма 1 (Арутюнов [44]). *Пусть r — целое число, $0\leqslant r\leqslant n$, а $\{\Pi_k\}$ — последовательность таких подпространств из E^n, что $\dim\Pi_k\geqslant r$, $k=1,2,\ldots$ Тогда существует подпространство Π из E^n, такое, что $\dim\Pi\geqslant r$ и $\Pi\subseteq\mathop{\mathcal{L}s}\limits_{k\to\infty}\Pi_k$.*

Пример 1. Пусть $\Pi_1=\{u=(x,y,z)\in E^3\colon x=0,\ y=0\}$, $\Pi_2=\{u=(x,y,z)\in E^3\colon z=0\}$. Очевидно, Π_1, Π_2 — подпространства из E^n, $\dim\Pi_1=1$, $\dim\Pi_2=2$. Рассмотрим последовательность $\{\Pi_k\}$ такую, что $\Pi_{2i-1}=\Pi_1$, $\Pi_{2i}=\Pi_2$, $i=1,2,\ldots$. Здесь $\mathop{\mathcal{L}s}\limits_{k\to\infty}\Pi_k=\Pi_1\cup\Pi_2$. Очевидно, $\dim\Pi_k\geqslant r=1$, так что условие леммы 1 выполнено. В качестве подпространства Π из утверждения леммы можем взять $\Pi=\Pi_1$ или $\Pi=\Pi_2$. Отметим, что здесь $\mathop{\mathcal{L}s}\limits_{k\to\infty}\Pi_k$ не является подпространством и включение $\Pi\subset\mathop{\mathcal{L}s}\limits_{k\to\infty}\Pi_k$ строгое.

Пример 2. Пусть $\Pi_k=\{x\in E^n\colon\langle c_k,x\rangle=0\}$, где c_k — заданный вектор из E^n, $|c_k|=1$, $k=1,2,\ldots$ Здесь $\dim\Pi_k=n-1=r$. В качестве Π из утверждения леммы можем взять подпространство $\Pi=\{x\in E^n\colon\langle c,x\rangle=0\}$, где c — произвольная предельная точка последовательности $\{c_k\}$. Верхний предел последовательности $\{\Pi_k\}$ является объединением всех таких подпространств Π.

Доказательство леммы 1. Пусть $r_k=\dim\Pi_k$. Тогда $\dim\Pi_k^\perp=n-r_k$, где $\Pi_k^\perp$ — ортогональное дополнение к Π в E^n. В $\Pi_k^\perp$ возьмем ортонормированный базис $e_{k1},\ldots,e_{kn-r_k}$, т. е. $|e_{kj}|=1$, $j=1,\ldots,n-r_k$, $\langle e_{kj},e_{kp}\rangle=0$ $\forall j\neq p$, $1\leqslant j,p\leqslant n-r_k$. Тогда $\Pi_k=\{x\in E^n\colon\langle e_{kj},x\rangle=0,\ j=1,\ldots,n-r_k\}$. Добавим к базисным векторам $\{e_{kj}\}$ нулевые вектора $e_{kn-r_k+1}=0,\ldots,e_{kn-r}=0$. Тогда для Π_k получаем представление $\Pi_k=\{x\in E^n\colon\langle e_{kj},x\rangle=0,\ j=1,\ldots,n-r\}$, в котором количество «порождающих» подпространство Π_k векторов $e_{k1},\ldots,e_{kn-r}$ уже не зависит от k. Эти вектора, очевидно, образуют ортогональную систему: $\langle e_{kj},e_{kp}\rangle=0$ $\forall j\neq p$, $1\leqslant j,p\leqslant n-r$. Вместо свойства $|e_{kj}|=1$ теперь можем гарантировать лишь ограниченность: $|e_{kj}|\leqslant 1$, $j=1,\ldots,n-r$. Пользуясь классической теоремой Больцано — Вейерштрасса из анализа применительно к последовательностям векторов $\{e_{k1}\},\ldots,\{e_{kn-r}\}$, можем утверждать, что существует подпоследовательность $\{e_{k_\nu j}\}\to e_j$, где $k_\nu\to\infty$ при $\nu\to\infty$, $j=1,\ldots,n-r$, причем $\langle e_j,e_p\rangle=0$ $\forall j\neq p$, $1\leqslant j,p\leqslant n-r$, $|e_j|\leqslant 1$, $\forall j=1,\ldots,n-r$, и часть векторов e_j может оказаться равной нулю. Пусть для определенности $|e_j|=1$ при $j=1,\ldots,p$, и $e_j=0$ при $j=p+1,\ldots,n-r$ (возможность $p=0$ здесь не исключается). Положим $\Pi=\{x\in E^n\colon\langle e_j,x\rangle=0,\ j=1,\ldots,n-r\}=\{x\in E^n\colon\langle e_j,x\rangle=0,\ j=1,\ldots,p\}$. Поскольку $e_1,\ldots,e_p$ линейно независимы, то ясно, что $\dim\Pi^\perp=p$, $\dim\Pi=n-p\geqslant n-(n-r)=r$. Покажем, что $\Pi\subseteq\mathop{\mathcal{L}s}\limits_{k\to\infty}\Pi_k$. Возьмем произвольную точку $x_0\in\Pi$. Убедимся, что x_0 — предельная точка последовательности $\{x_k=\mathcal{P}_{\Pi_k}(x_0)\}$ проекций точки x_0 на Π_k, $k=1,2,\ldots$ Воспользуемся неравенством Хоффмана (15.41): $\rho(x_0,\Pi_{k_\nu})=|x_0-\mathcal{P}_{\Pi_{k_\nu}}(x_0)|=|x_0-x_{k_\nu}|\leqslant M\max\limits_{1\leqslant j\leqslant n-r}|\langle e_{k_\nu j},x_0\rangle|=M\max|\langle e_{k_\nu j}-e_j,x_0\rangle|\to 0$ при $\nu\to\infty$. Таким образом, $\Pi\subseteq\mathop{\mathcal{L}s}\limits_{k\to\infty}\Pi_k$. Лемма 1 доказана. □

Теперь можем доказать замкнутость конуса $\Lambda_a(v)\cup\{0\}$. Пусть последовательность $\{\bar\lambda_k\}\in\Lambda_a(v)$, $\{\bar\lambda_k\}\to\bar\lambda\neq 0$. Покажем, что $\bar\lambda\in\Lambda_a(v)$. По определению конуса $\Lambda_a(v)$ для каждого $\bar\lambda_k$ найдется сопровождающее подпространство $\Pi(\bar\lambda_k)$ со свойствами $\dim\Pi(\bar\lambda_k)\geqslant\max\{n-|I(v)|;0\}=r$, $\Pi(\bar\lambda_k)\subseteq\ker G'(v)$, $\langle\mathcal{L}_{xx}(v,\bar\lambda_k)h,h\rangle\geqslant 0$ $\forall h\in\Pi(\bar\lambda_k)$, $k=1,2,\ldots$ По лемме 1 существует подпространство Π такое, что $\dim\Pi\geqslant r=\max\{n-|I(v)|;0\}$, $\Pi\subset\mathop{\mathcal{L}s}\limits_{k\to\infty}\Pi(\bar\lambda_k)$. Так как конус $\ker G'(v)$ замкнут, $\Pi(\bar\lambda_k)\subseteq\ker G'(v)$, то нетрудно убедиться, что $\mathop{\mathcal{L}s}\limits_{k\to\infty}\Pi(\bar\lambda_k)\subset\ker G'(v)$. Следовательно, $\Pi\subset\ker G'(v)$. Покажем, что $\langle\mathcal{L}_{xx}(v,\bar\lambda)h,h\rangle\geqslant 0$ $\forall h\in\Pi$. Возьмем произвольную точку $h_0\in\Pi\subseteq\mathop{\mathcal{L}s}\limits_{k\to\infty}\Pi(\bar\lambda_k)$. По определению $\mathop{\mathcal{L}s}\limits_{k\to\infty}\Pi(\bar\lambda_k)$ найдутся $h_k\in\Pi(\bar\lambda_k)$ такие, что h_0 будет предельной точкой последовательности $\{h_k\}$, т. е. $\exists\{h_{k_\nu}\}\to h_0$. А по определению $\Pi(\bar\lambda_{k_\nu})$ имеем $\langle\mathcal{L}_{xx}(v,\bar\lambda_{k_\nu})h_{k_\nu},h_{k_\nu}\rangle\geqslant 0$, $\nu=1,2,\ldots$ Отсюда при $\nu\to\infty$ получим $\langle\mathcal{L}_{xx}(v,\bar\lambda)h_0,h_0\rangle\geqslant 0$ $\forall h_0\in\Pi$. Это значит, что подпространство Π обладает свойствами (14)–(16) и, следовательно, является сопровождающим подпространством точки $\bar\lambda$. Включение $\bar\lambda\in\Lambda_a(v)$ доказано в случае $\bar\lambda\neq 0$. Отсюда, учитывая, что последовательность $\{\bar\lambda_k\}\in\Lambda_a(v)$ может также сходиться к точке $\bar\lambda=0$, заключаем, что конус $\Lambda_a(v)\cup\{0\}$ замкнут.

Теперь можем приступить к доказательству основной теоремы настоящего параграфа.

Теорема 2 (Арутюнов [44]). *Пусть v — точка локального минимума задачи (13.A), пусть функции $f(x)$, $g_1(x),\ldots,g_s(x)$ дважды непрерывно дифференцируемы в некоторой окрестности точки v. Тогда*

$$\Lambda_a(v)\neq\varnothing, \tag{17}$$

$$\max_{\bar\lambda\in\Lambda_a(v),\,|\bar\lambda|=1}\langle\mathcal{L}_{xx}(v,\bar\lambda)h,h\rangle\geqslant 0\quad\forall h\in K(v)\cap\{h\in E^n\colon\langle f'(v),h\rangle\leqslant 0\}, \tag{18}$$

где

$$K(v)=\{h\in E^n\colon\langle g_i'(v),h\rangle\leqslant 0\ \ \forall i\in I(v)\cap\{i\colon 1\leqslant i\leqslant m\},\ \ \langle g_i'(v),h\rangle=0,\ i=m+1,\ldots,s\}. \tag{19}$$

Доказательство. Рассмотрим вспомогательную задачу минимизации

$$\begin{gathered}g_0(x)=f(x)+|x-v|^4\to\inf,\\ x\in W=\{x\in E^n\colon|x-v|\leqslant\gamma,\ g_i(x)\leqslant 0,\ i=1,\ldots,m,\ g_i(x)=0,\ i=m+1,\ldots,s\},\end{gathered} \tag{19.A}$$

аналогичную задаче (7), предполагая число $\gamma>0$ столь малым, что $\gamma<1$, $f(x)\geqslant f(v)$ $\forall x\in X\cap S(v,\gamma)=W$, $S(v,\gamma)=\{x\in E^n\colon|x-v|\leqslant\gamma\}$. Тогда, как и в (7), нетрудно доказать, что v — точка строгого локального минимума задачи (19.A). К задаче (19.A) применим метод штрафных функций. Имея в виду, что при выводе необходимых условий второго порядка нам придется иметь дело со вторыми производными рассматриваемых функций, воспользуемся более гладкими, чем в (7), штрафными функциями и рассмотрим задачу

$$\Phi_k(x)=g_0(x)+k\sum_{i=1}^m(\max\{g_i(x);0\})^4+k\sum_{i=m+1}^s g_i^2(x)\to\inf,\quad x\in W_0=S(v,\gamma). \tag{20}$$

Рассуждая также, как при доказательстве теоремы 1, нетрудно установить, что задача (20) имеет решение, т. е. существуют точки $x_k\in W_0$, $\Phi_k(x_k)=\Phi_{k*}=\inf\limits_{x\in W_0}\Phi_k(x)$, и справедливы равенства (8), означающие сходимость метода штрафных функций (20). В частности, из $\{x_k\}\to v$ следует, что $|x_k-v|<\gamma$, т. е. $x_k\in\operatorname{int}W_0$ $\forall k\geqslant k_0$. Во внутренних точках локального минимума необходимо $\Phi_k'(x_k)=0$, $\Phi_k''(x_k)\geqslant 0$ $\forall k\geqslant k_0$ (теорема 2.2.1). Пользуясь выражениями производных функции $\Phi(x)$ из (20), имеем

$$\Phi_k'(x_k)=g_0'(x_k)+\sum_{i=1}^m 4k(\max\{g_i(x_k);0\})^3g_i'(x_k)+\sum_{i=m+1}^s 2kg_i(x_k)g_i'(x_k)=0,$$

$$\begin{aligned}\Phi_k''(x_k)=g_0''(x_k)+\sum_{i=1}^m[&4k(\max\{g_i(x_k);0\})^3g_i''(x_k)+12k(\max\{g_i(x_k);0\})^2(g_i'(x_k))^\top g_i'(x_k)]+\\&+\sum_{i=m+1}^s[2kg_i(x_k)g_i''(x_k)+2k(g_i'(x_k))^\top g_i'(x_k)]\geqslant 0,\quad\forall k\geqslant k_0.\end{aligned}$$

Обозначим

$$\mu_{ik}=\begin{cases}4k(\max\{g_i(x_k);0\})^3\geqslant 0, & i=1,\ldots,m,\\ 2kg_i(x_k), & i=m+1,\ldots,s.\end{cases} \tag{21}$$

Разделив предыдущие соотношения на $\left(1+\sum\limits_{i=1}^s\mu_{ik}^2\right)^{1/2}\geqslant 1$, получим

$$\lambda_{0k}g_0'(x_k)+\sum_{i=1}^s\lambda_{ik}g_i'(x_k)=0,\quad\forall k\geqslant k_0, \tag{22}$$

$$\begin{aligned}\lambda_{0k}g_0''(x_k)+\sum_{i=1}^s\lambda_{ik}g_i''(x_k)+\sum_{i=1}^s 12k\lambda_{0k}(\max\{g_i(x_k);0\})^2(g_i'(x_k))^\top g_i'(x_k)+\\+\sum_{i=m+1}^s 2k\lambda_{0k}(g_i'(x_k))^\top g_i'(x_k)\geqslant 0,\quad\forall k\geqslant k_0,\end{aligned} \tag{23}$$

где $\lambda_{0k}=\left(1+\sum\limits_{i=1}^s\mu_{ik}^2\right)^{-1/2}$, $\lambda_{ik}=\mu_{ik}\lambda_{0k}$, $i=1,\ldots,s$.

По определению множества $I(v)$ имеем $g_i(v)<0\ \forall i\notin I(v)$. Поэтому с учетом непрерывности $g_i(x)$ и первого равенства (8), беря при необходимости k_0 еще большим, можем считать, что $g_i(x_k)<0\ \ \forall k\geqslant k_0$. Отсюда с учетом (21) имеем

$$\max\{g_i(x_k);0\}=0,\quad \mu_{ik}=0\quad \forall i\notin I(v)\quad \forall k\geqslant k_0,$$

так что $\lambda_{ik}=\mu_{ik}\lambda_{0k}\geqslant 0$, $i=1,\ldots,m$, $\lambda_{ik}=0\ \forall i\notin I(v)$, $\forall k\geqslant k_0$. Так как $|\bar\lambda_k|=1$, где $\bar\lambda_k=(\lambda_{0k},\ldots,\lambda_{sk})$, то, выбирая при необходимости подпоследовательность, можем считать, что $\{\bar\lambda_k\}\to\bar\lambda$. Кроме того, заметим, что в силу (8) $g_0'(x_k)=f'(x_k)+4|x_k-v|^2(x_k-v)\to f'(v)$ при $k\to\infty$. Переходя в (22) к пределу при $k\to\infty$, получим $\lambda_0 f'(x_0)+\sum\limits_{i=1}^{s}\lambda_i g_i'(v)=0$. Таким образом, $\mathcal{L}_x(v,\bar\lambda)=0$, $\lambda_i g_i(v)=0$, $i=1,\ldots,m$, $|\bar\lambda|=1$, $\lambda_0\geqslant 0,\ldots,\lambda_m\geqslant 0$, т. е. $\bar\lambda\in\Lambda(v)$.

Теперь совершим предельный переход при $k\to\infty$ в неравенстве (23). Введем последовательность подпространств $\Pi_k=\{h\in E^n\colon \langle g_i'(x_k),h\rangle=0,\ i\in I(v)\}$, $k=1,2,\ldots$ Размерность ортогонального подпространства $\Pi_k^{\perp}$ равна числу линейно независимых векторов системы $\{g_i'(x_k),\ i\in I(v)\}$, т. е. не превышает числа $|I(v)|$. Тогда $\dim\Pi_k\geqslant\max\{n-|I(v)|;0\}=r$. По лемме 1 найдется подпространство Π такое, что $\dim\Pi\geqslant\max\{n-|I(v)|;0\}$ и $\Pi\subset\operatorname*{\mathcal{L}s}\limits_{k\to\infty}\Pi_k$. По определению $\operatorname*{\mathcal{L}s}\limits_{k\to\infty}\Pi_k$ каждая точка $h\in\operatorname*{\mathcal{L}s}\limits_{k\to\infty}\Pi_k$ является пределом какой-либо подпоследовательности $\{h_{k_\nu}\}$, $h_{k_\nu}\in\Pi_{k_\nu}$, т. е. $\langle g_i'(x_{k_\nu}),h_{k_\nu}\rangle=0$, $i\in I(v)$. Однако $\{x_{k_\nu}\}\to v$, поэтому при $\nu\to\infty$ из последних равенств получим $\langle g'(v),h\rangle=0$, $i\in I(v)$, т. е. $h\in\ker G'(v)$. Это значит, что $\operatorname*{\mathcal{L}s}\limits_{k\to\infty}\Pi_k\subset\ker G'(v)$ и подавно $\Pi\subset\ker G'(v)$. Поскольку $\max\{g_i(x_k);0\}=0$ при $\forall i\notin I(v)$, $k\geqslant k_0$, а для $i\in I(v)$ имеет место равенство $\langle(g_i'(x_k))^{\top}g_i'(x_k)h_k,h_k\rangle=|g_i'(x_k)h_k|^2=0$ для $\forall h_k\in\Pi_k$, то из (23) при $k=k_\nu$ имеем

$$\langle(\lambda_{0k_\nu}g_0''(x_{k_\nu})+\sum_{i=1}^{s}\lambda_{ik_\nu}g_i''(x_{k_\nu}))h_{k_\nu},h_{k_\nu}\rangle=\langle\mathcal{L}_{xx}(x_{k_\nu},\bar\lambda_{k_\nu})h_{k_\nu},h_{k_\nu}\rangle\geqslant 0,$$

где $h_{k_\nu}\in\Pi_{k_\nu}$, $\nu=1,2,\ldots$, взяты такими, что $\{h_{k_\nu}\}\to h\in\Pi$. Так как $\{x_k\}\to v$, $\{\bar\lambda_{k_\nu}\}\to\bar\lambda$, $g_0''(x_k)=f''(x_k)+4|x_k-v|^2I_n+8(x_k-v)(x_k-v)^{\top}\to f''(v)$, где I_n — единичная матрица $n\times n$, то из последнего неравенства при $\nu\to\infty$ получим $\langle\mathcal{L}_{xx}(v,\bar\lambda)h,h\rangle\geqslant 0\ \forall h\in\Pi$. Таким образом, доказано, что подпространство Π обладает свойствами (14)–(16) и, следовательно, является сопровождающим подпространством точки $\bar\lambda\in\Lambda(v)$. Это значит, что $\bar\lambda\in\Lambda_a(v)$, т. е. $\Lambda_a(v)\neq\varnothing$. Утверждение (17) доказано.

Остается доказать неравенство (18). Заметим, что поскольку конус $\Lambda_a(v)\cup\{0\}$ замкнут, то множество $\{\bar\lambda\in\Lambda_a(v),\ |\bar\lambda|=1\}$ замкнуто и ограничено, т. е. компактно. Отсюда и из непрерывности $\mathcal{L}_{xx}(v,\bar\lambda)$ по $\bar\lambda$ следует, что в правой части (18) максимум достигается хотя бы в одной точке $\bar\lambda=\bar\lambda(h)$ при каждом $h\in K(v)$, $\langle f'(v),h\rangle\leqslant 0$. Нам надо доказать, что этот максимум неотрицателен при всех $h\in K(v)$, $\langle f'(v),h\rangle\leqslant 0$. Зафиксируем произвольную точку h из $K(v)$, $\langle f'(v),h\rangle\leqslant 0$. Можем считать, что $h\neq 0$, так как при $h=0$ неравенство (18) выполняется тривиально при всех $\bar\lambda$. Для сокращения записей примем, что $v=0$, $f(0)=0$, так что $f(x)\geqslant f(0)=0\ \forall x\in X$, $|x|\leqslant\gamma$; к этому случаю легко прийти, переходя к переменным $x-v$, $f(x)-f(v)$. Неравенство (18) сначала докажем при дополнительном предположении, что в задаче (13.А) все ограничения $g_i(x)\leqslant 0$, $i=1,\ldots,m$, в точке $x=v=0$ являются активными, т. е. $I(0)=\{i\colon i=1,2,\ldots,s\}$. Рассмотрим следующую вспомогательную задачу минимизации в пространстве переменных $z=(x,\chi)\in E^n\times E^1$:

$$g_{0\varepsilon}(z)=g_0(x)-\chi g_0(\varepsilon h)+|x-\varepsilon h|^4+\psi(\chi)\to\inf,\quad z\in Z(\varepsilon),\tag{24}$$

$$Z(\varepsilon)=\{z\in Z_0\colon g_{i\varepsilon}(z)=g_i(x)-\chi g_i(\varepsilon h)\leqslant 0,\ i=1,\ldots,m,$$
$$g_{i\varepsilon}(z)=g_i(x)-\chi g_i(\varepsilon h)=0,\ i=m+1,\ldots,s\},\tag{25}$$

где $g_0(x)=f(x)+|x|^4$, $Z_0=\{z=(x,\chi)\in E^{n+1}\colon |x|\leqslant\gamma,\ \chi\geqslant 0\}$; параметр ε столь мал, что $0<\varepsilon|h|<\gamma$;

$$\psi(\chi)=\begin{cases}0 & \text{при } 0\leqslant\chi\leqslant 1,\\ (\chi-1)^4 & \text{при } \chi\geqslant 1.\end{cases}$$

Нетрудно видеть, что точка $\bar z=(x=\varepsilon h,\chi=1)\in Z(\varepsilon)$, так что $Z(\varepsilon)\neq\varnothing$ и $g_{0\varepsilon}(\bar z)=0$. Введем множество Лебега $M_\varepsilon(\bar z)=\{z\in Z(\varepsilon)\colon g_{0\varepsilon}(z)\leqslant g_{0\varepsilon}(\bar z)\}$. Так как функции $g_{i\varepsilon}(z)$, $i=0,\ldots,s$, непрерывны при всех $z\in Z_0$, то $M_\varepsilon(\bar z)$ замкнутое множество. Убедимся, что оно ограничено, причем равномерно по ε, $0<\varepsilon<\gamma|h|^{-1}$. Заметим, что $M_\varepsilon(\bar z)=M_{\varepsilon1}(\bar z)\cup M_{\varepsilon2}(\bar z)$, где $M_{\varepsilon1}(\bar z)=M_\varepsilon(\bar z)\cap\{z\colon 0\leqslant\chi\leqslant 1\}$, $M_{\varepsilon2}(\bar z)=M_\varepsilon(\bar z)\cap\{z\colon\chi>1\}$. Множество $M_{\varepsilon1}(\bar z)\subset\{z=(x,\chi)\colon |x|\leqslant\gamma,\ 0\leqslant\chi\leqslant 1\}$ и, очевидно, ограничено равномерно по ε, $0<\varepsilon<\gamma|h|^{-1}$. Рассмотрим множество $M_{\varepsilon2}(\bar z)$. Пусть $z\in M_{\varepsilon2}(\bar z)$. Тогда имеем $0=g_{0\varepsilon}(\bar z)\geqslant\inf\limits_{|x|\leqslant\gamma}g_0(x)-\chi\sup\limits_{|x|\leqslant\gamma}g_0(x)-(|x|+|\varepsilon h|)^4+(\chi-1)^4$ или $(\chi-1)^4-\chi A\leqslant(2\gamma)^4-B\ \ \forall z=(x,\chi)\in M_{\varepsilon2}(\bar z)$, где $A=\sup\limits_{|x|\leqslant\gamma}g_0(x)\geqslant 0$, $B=\inf\limits_{|x|\leqslant\gamma}g_0(x)\leqslant 0$. Проведя элементарное исследование графика функции $\varphi(\chi)=(\chi-1)^4-\chi A$, $\chi\in E^1$, нетрудно показать, что уравнение $\varphi(\chi)=(2\gamma)^4-B$ имеет два решения χ_1,χ_2, причем $\chi_2>1$. Отсюда заключаем, что $M_{\varepsilon2}(\bar z)\subset\{z=(x,\chi)\colon |x|\leqslant\gamma,\ 1\leqslant\chi\leqslant\chi_2\}$, т. е. множество $M_{\varepsilon2}(\bar z)$ также ограничено, причем равномерно по ε, $0<\varepsilon<\gamma|h|^{-1}$. Таким образом, множество $M_\varepsilon(\bar z)$ компактно, и непрерывная функция $g_{0\varepsilon}(z)$ достигает на этом множестве своей нижней грани хотя бы в одной точке $z_\varepsilon=(x_\varepsilon,\chi_\varepsilon)\in M_\varepsilon(\bar z)$ (теорема 2.1.1). Однако, $\inf\limits_{z\in Z(\varepsilon)}g_{0\varepsilon}(z)=\inf\limits_{z\in M_\varepsilon(\bar z)}g_{0\varepsilon}(z)=g_{0\varepsilon}(z_\varepsilon)\leqslant g_{0\varepsilon}(\bar z)\leqslant 0$. Так как множество $M_\varepsilon(\bar z)$ ограничено равномерно по ε, $0<\varepsilon<\gamma|h|^{-1}$, то семейство $\{z_\varepsilon\}$ решений задач (24), (25) при $0<\varepsilon<\gamma|h|^{-1}$ равномерно ограничено: $|x_\varepsilon|\leqslant\gamma$, $0\leqslant\chi_\varepsilon\leqslant\chi_2$.

Покажем, что $z_\varepsilon\in\operatorname{int}Z_0$ при всех достаточно малых $\varepsilon>0$. Сначала убедимся, что $\chi_\varepsilon>0$. Заметим, что если $x\in W$ (см. множество задачи (19.А)), то $z=(x,\chi=0)\in Z(\varepsilon)$. Для всех таких точек $z=(x,\chi=0)\in Z(\varepsilon)$ имеем $g_{0\varepsilon}(z)=g_0(x)+|x-\varepsilon h|^4=f(x)+|x|^4+|x-\varepsilon h|^4>0\ \forall\varepsilon$, $0<\varepsilon<\gamma|h|^{-1}$. Однако $g_{0\varepsilon}(z_\varepsilon)\leqslant g_{0\varepsilon}(\bar z)\leqslant 0$. Следовательно, $z_\varepsilon\neq(x,\chi=0)$, т. е. $\chi_\varepsilon>0$. Далее, покажем, что $\lim\limits_{\varepsilon\to0}x_\varepsilon=0=v$. Пусть a — произвольная предельная точка семейства $\{x_\varepsilon\}$ при $\varepsilon\to 0$, пусть $a=\lim\limits_{k\to\infty}x_{\varepsilon_k}$. Так как $z_\varepsilon=(x_\varepsilon,\chi_\varepsilon)\in Z(\varepsilon)$, $0<\chi_\varepsilon\leqslant\chi_2$, то из (25) имеем: $|x_\varepsilon|\leqslant\gamma$, $g_i(x_\varepsilon)\leqslant\chi_\varepsilon g_i(\varepsilon h)\leqslant\chi_2|g_i(\varepsilon h)|$, $i=1,\ldots,m$, $|g_i(x_\varepsilon)|\leqslant\chi_\varepsilon|g_i(\varepsilon h)|\leqslant\chi_2|g_i(\varepsilon h)|$, $i=m+1,\ldots,s$. Отсюда, учитывая, что $\lim\limits_{\varepsilon\to0}g_i(\varepsilon h)=g_i(0)=0\ \ \forall i\in I(0)=\{i\colon i=1,\ldots,s\}$, при $\varepsilon=\varepsilon_k\to 0$ получим: $|a|\leqslant\gamma$, $g_i(a)\leqslant 0$, $i=1,\ldots,m$, $g_i(a)=0$, $i=m+1,\ldots,s$, т. е. $a\in W$. Из (24) следует: $g_0(x_\varepsilon)=g_{0\varepsilon}(z_\varepsilon)+\chi_\varepsilon g_0(\varepsilon h)-|x_\varepsilon-\varepsilon h|^4-\psi(\chi_\varepsilon)\leqslant\chi_\varepsilon g_0(\varepsilon h)\leqslant\chi_2|g_0(\varepsilon h)|$. Полагая здесь $\varepsilon=\varepsilon_k$ и устремляя $k\to\infty$ с учетом равенства $\lim\limits_{\varepsilon\to0}g_0(\varepsilon h)=g_0(0)=0$, имеем $g_0(a)\leqslant 0$. Однако $v=0$ — точка минимума $g_0(x)$ на W, $a\in W$, поэтому $0=g_0(0)\leqslant g_0(a)$. Таким образом, $g_0(a)=0$, т. е. $x=a$ — решение задачи (19.А). Но задача (19.А) имеет единственное решение, поэтому $a=0$. Это означает, что семейство $\{x_\varepsilon\}$ при $\varepsilon\to 0$ имеет единственную предельную точку $a=0$. Следовательно, $\lim\limits_{\varepsilon\to\infty}x_\varepsilon=0$ и $|x_\varepsilon|<\gamma$ при всех ε, $0<\varepsilon<\varepsilon_0\leqslant\gamma|h|^{-1}$. Таким образом, $z_\varepsilon\in\operatorname{int}Z_0$ при $\forall\varepsilon$, $0<\varepsilon<\varepsilon_0$.

Применим к задаче (24). (25) уже доказанное утверждение (17), согласно которому конус Арутюнова $\Lambda_{a\varepsilon}(z_\varepsilon)$ точки z_ε непуст при всех ε, $0<\varepsilon<\gamma|h|^{-1}$. Покажем, что все отличные от нуля предельные точки семейства конусов $\{\Lambda_{a\varepsilon}(z_\varepsilon),\ \varepsilon>0\}$ при $\varepsilon\to 0$ принадлежат конусу $\Lambda_a(0)$. Пусть $\bar\lambda\neq 0$ — одна из таких предельных точек. Это значит, что существуют $\{\varepsilon_k\}\to 0$ $\varepsilon_k>0$, и точки $\bar\lambda_{\varepsilon_k}\in\Lambda_{a\varepsilon_k}(z_{\varepsilon_k})$ такие, что $\lim\limits_{k\to\infty}\bar\lambda_{\varepsilon_k}=\bar\lambda$. Нам нужно показать, что $\bar\lambda\in\Lambda_a(0)$. Сначала установим, что $\bar\lambda\in\Lambda(0)$. Пусть $\Lambda_\varepsilon(z_\varepsilon)$ — конус Лагранжа задачи (24), (25), соответствующий точке $z_\varepsilon=(x_\varepsilon,\chi_\varepsilon)$. Так как $\bar\lambda_\varepsilon\subset\Lambda_{a\varepsilon}(z_\varepsilon)\subset\Lambda_\varepsilon(z_\varepsilon)$ и $z_\varepsilon\in\operatorname{int}Z_0\ \forall\varepsilon$, $0<\varepsilon<\varepsilon_0$, то по определению множителей Лагранжа задачи (24), (25)

$$\bar\lambda_\varepsilon=(\lambda_{0\varepsilon},\lambda_{1\varepsilon},\ldots,\lambda_{s\varepsilon}),\quad |\bar\lambda_\varepsilon|=1,\quad \lambda_{0\varepsilon}\geqslant 0,\lambda_{1\varepsilon}\geqslant 0,\ldots,\lambda_{m\varepsilon}\geqslant 0,\tag{26}$$

$$\frac{\partial\mathcal{L}_\varepsilon(z_\varepsilon,\bar\lambda_\varepsilon)}{\partial z}=\Big(\frac{\partial\mathcal{L}_\varepsilon(z_\varepsilon,\bar\lambda_\varepsilon)}{\partial x},\frac{\partial\mathcal{L}_\varepsilon(z_\varepsilon,\bar\lambda_\varepsilon)}{\partial\chi}\Big)=0,\tag{27}$$

$$\lambda_{i\varepsilon}(g_i(x_\varepsilon)-\chi_\varepsilon g_i(\varepsilon h))=0,\quad i=1,\ldots,m,\tag{28}$$

где $\mathcal{L}_\varepsilon(z,\bar\lambda)=\lambda_0g_{0\varepsilon}(z)+\sum\limits_{i=1}^{s}\lambda_i(g_i(x)-\chi g_i(\varepsilon h))$, $z=(x,\chi)\in Z_0$, $\lambda_i\geqslant 0$, $i=0,\ldots,m$, $\varepsilon=\varepsilon_k$, $k\geqslant k_0$. Подробнее распишем равенство (27):

$$\frac{\partial \mathcal{L}_\varepsilon(z_\varepsilon, \bar\lambda_\varepsilon)}{\partial x} = \lambda_{0\varepsilon}(f'(x_\varepsilon) + 4|x_\varepsilon|^2 x_\varepsilon + 4|x_\varepsilon - \varepsilon h|^2(x_\varepsilon - \varepsilon h)) + \sum_{i=1}^{s} \lambda_{i\varepsilon} g_i'(x_\varepsilon) = 0, \quad (29)$$

$$\frac{\partial \mathcal{L}_\varepsilon(z_\varepsilon, \bar\lambda_\varepsilon)}{\partial \chi} = \lambda_{0\varepsilon}(-g_0(\varepsilon h) + \psi'(\chi_\varepsilon)) + \sum_{i=1}^{s} \lambda_{i\varepsilon}(-g_i(\varepsilon h)) = 0. \quad (30)$$

Учитывая, что $x_\varepsilon \to v = 0$, $g_i(\varepsilon h) \to g_i(0)$ при $\varepsilon \to 0$, $0 < \chi_\varepsilon \leqslant \chi_2$, $0 < \varepsilon < \varepsilon_0$, $\lim\limits_{k\to\infty} \bar\lambda_{\varepsilon_k} = \bar\lambda$, из (26), (28), (29) при $\varepsilon = \varepsilon_k \to 0$ имеем

$$\bar\lambda = (\lambda_0, \ldots, \lambda_s) \neq 0, \quad \lambda_i \geqslant 0, \quad i = 0, \ldots, m, \quad \mathcal{L}_x(0, \bar\lambda) = 0, \quad \lambda_i g_i(0) = 0, \quad i = 1, \ldots, m.$$

Это означает, что предельная точка $\bar\lambda \in \Lambda(0)$.

Далее, покажем, что $\bar\lambda \in \Lambda_a(0)$. По определению конуса $\Lambda_{a\varepsilon}(z_\varepsilon)$ для каждого набора $\bar\lambda_\varepsilon \in \Lambda_{a\varepsilon}(z_\varepsilon)$ существует сопровождающее подпространство $\Pi_\varepsilon \subset E^{n+1}$, которое обладает свойствами:

$$\dim \Pi_\varepsilon \geqslant n + 1 - |I_\varepsilon(z_\varepsilon)|, \quad (31)$$

$$\Pi_\varepsilon \subset \ker G_\varepsilon'(z_\varepsilon) = \{\mu = (h, \eta) \in E^n \times E^1 : \langle \frac{\partial g_{i\varepsilon}(z_\varepsilon)}{\partial z}, \mu\rangle = 0, \quad i \in I_\varepsilon(z_\varepsilon)\}, \quad (32)$$

$$\langle \mathcal{L}_{\varepsilon zz}(z_\varepsilon, \bar\lambda_\varepsilon)\mu, \mu\rangle \geqslant 0 \qquad \forall \mu \in \Pi_\varepsilon, \quad (33)$$

где $I_\varepsilon(z_\varepsilon)$ — множество активных индексов точки z_ε, G_ε — вектор-функция с координатами $g_{i\varepsilon}$, $i \in I_\varepsilon(z_\varepsilon)$. Так как в рассматриваемом случае $I(0) = \{i: i = 1, 2, \ldots, s\}$, то $I_\varepsilon(z_\varepsilon) \subset I(0)$ и $|I_\varepsilon(z_\varepsilon)| \leqslant |I(0)|$, $0 < \varepsilon < \varepsilon_0$.

Введем подпространство

$$\Pi_{1\varepsilon} = \Pi_\varepsilon \cap \{\mu = (h, \eta) \in E^n \times E^1 : \langle \frac{\partial g_{i\varepsilon}(z_\varepsilon)}{\partial z}, \mu\rangle = 0, \quad i \in I(0) \setminus I_\varepsilon(z_\varepsilon)\}.$$

С учетом оценки (31) имеем

$$\dim \Pi_{1\varepsilon} \geqslant n + 1 - I(0), \quad \forall \varepsilon, \quad 0 < \varepsilon < \varepsilon_0. \quad (34)$$

Далее, из включения (32) следует

$$\Pi_{1\varepsilon} \subset \ker G_\varepsilon'(z_\varepsilon) \cap \{\mu : \langle \frac{\partial g_{i\varepsilon}(z_\varepsilon)}{\partial z}, \mu\rangle = 0, \ i \in I(0) \setminus I_\varepsilon(z_\varepsilon)\} =$$

$$= \{\mu = (h, \eta) \in E^n \times E^1 : \langle \frac{\partial g_{i\varepsilon}(z_\varepsilon)}{\partial z}, \mu\rangle = \langle g_i'(x_\varepsilon), h\rangle - g_{i\varepsilon}(\varepsilon h)\eta = 0, \ i \in I(0)\}. \quad (35)$$

Заметим, что

$$\mathcal{L}_{\varepsilon zz}(z, \bar\lambda) = \begin{pmatrix} \mathcal{L}_{\varepsilon xx}(z, \bar\lambda) & (\mathcal{L}_{\varepsilon x\chi}(z, \bar\lambda))^{\mathrm T} \\ \mathcal{L}_{\varepsilon x\chi}(z, \bar\lambda) & \mathcal{L}_{\varepsilon\chi\chi}(z, \lambda) \end{pmatrix}$$

— матрица размера $(n+1)\times(n+1)$, где $\mathcal{L}_{\varepsilon xx}(z, \bar\lambda) = \lambda_0 g_{0\varepsilon}''(z) + \sum\limits_{i=1}^{s} \lambda_i g_{i\varepsilon}''(x)$, $\mathcal{L}_{\varepsilon x\chi}(z, \bar\lambda) = 0$, $\mathcal{L}_{\varepsilon\chi\chi}(z, \lambda) = \lambda_0 \psi''(\chi_\varepsilon)$. Отсюда с учетом (33) получаем

$$\langle \mathcal{L}_{\varepsilon zz}(z_\varepsilon, \bar\lambda_\varepsilon)\mu, \mu\rangle = \langle \mathcal{L}_{\varepsilon xx}(z_\varepsilon, \bar\lambda_\varepsilon)h, h\rangle + \mathcal{L}_{\varepsilon\chi\chi}(z_\varepsilon, \bar\lambda_\varepsilon)\eta^2 =$$

$$= \langle (\lambda_{0\varepsilon} g_{0\varepsilon}''(z_\varepsilon) + \sum_{i=1}^{s} \lambda_{i\varepsilon} g_{i\varepsilon}''(x_\varepsilon))h, h\rangle + \lambda_0 \psi''(\chi_\varepsilon)\eta^2 \geqslant 0 \quad \forall \mu \in \Pi_{1\varepsilon} \subset \Pi_\varepsilon. \quad (36)$$

В E^n введем подпространство $\Pi_{2\varepsilon}$, состоящее из тех $h \in E^n$, для которых $\mu = (h, \eta = 0) \in \Pi_{1\varepsilon}$. Для таких μ из (34)–(36) имеем:

$$\dim \Pi_{2\varepsilon} \geqslant \max\{n - |I(0)|; 0\}, \quad 0 < \varepsilon < \varepsilon_0, \quad (37)$$

$$\Pi_{2\varepsilon} \subset \ker G'(z_\varepsilon) = \{h \in E^n : \langle g_i'(x_\varepsilon), h\rangle = 0, \ i \in I(0)\}, \quad 0 < \varepsilon < \varepsilon_0, \quad (38)$$

$$\langle \mathcal{L}_{\varepsilon xx}(z_\varepsilon, \bar\lambda_\varepsilon)h, h\rangle = \langle (\lambda_{0\varepsilon} g_{0\varepsilon}''(z_\varepsilon) + \sum_{i=1}^{s} \lambda_{i\varepsilon} g_{i\varepsilon}''(x_\varepsilon))h, h\rangle \geqslant 0 \quad \forall h \in \Pi_{2\varepsilon}, \quad 0 < \varepsilon < \varepsilon_0. \quad (39)$$

В (37)–(39) возьмем $\varepsilon = \varepsilon_k$ и устремим $k \to \infty$. Вспомним, что $\{x_{\varepsilon_k}\} \to 0$, $\{\bar\lambda_{\varepsilon_k}\} \to \bar\lambda \in \Lambda(0)$. К последовательности подпространств $\Pi_{2\varepsilon_k}$ применим лемму 1. Согласно этой лемме существует подпространство $\Pi \subset \underset{k\to\infty}{\mathcal{L}s}\, \Pi_{2\varepsilon_k}$, $\dim \Pi \geqslant \max\{n - |I(0)|; 0\}$. Возьмем произвольную точку $h_0 \in \Pi$.

По определению $\underset{k\to\infty}{\mathcal{L}s}\, \Pi_{2\varepsilon_k}$ точка h_0 является предельной для некоторой последовательности $\{h_k\}$, $h_k \in \Pi_{2\varepsilon_k}$. Без умаления общности можем считать, что сама последовательность $\{h_k\}$ сходится к h_0. Тогда из (38) при $\varepsilon = \varepsilon_k$, $h = h_k$ и $k \to \infty$ имеем $\langle g_i'(0), h_0\rangle = 0$, $\forall i \in I(0)$, т. е. $h_0 \in \ker G'(0)$. Следовательно, $\Pi \subset \ker G'(0)$. Наконец, учитывая, что $g_{0\varepsilon}''(x) = f''(x) + 4|x|^2 I_n + 8xx^{\mathrm T} + 4|x - \varepsilon h|^2 I_n + 8(x - \varepsilon h)(x - \varepsilon h)^{\mathrm T}$ из (39) при $\varepsilon = \varepsilon_k$, $h = h_k \in \Pi_{2\varepsilon_k}$ и $k \to \infty$ получим $\langle (\lambda_0 f''(0) + \sum\limits_{i=1}^{s} \lambda_i g_i''(0)h_0, h_0\rangle \geqslant 0$, т. е. $\langle \mathcal{L}_{xx}(0, \bar\lambda)h, h\rangle \geqslant 0 \ \forall h \in \Pi$. Это значит, что подпространство Π обладает свойствами (14)–(16) и является сопровождающим подпространством для точки $\bar\lambda$. Следовательно, $\bar\lambda = \bar\lambda(h) \in \Lambda_a(0)$. Тем самым показано, что любая отличная от нуля предельная точка семейства конусов $\{\Lambda_{a\varepsilon}(z_\varepsilon), \ \varepsilon > 0\}$ при $\varepsilon \to 0$ принадлежит $\Lambda_a(0)$.

Теперь мы можем доказать утверждение (18). С этой целью воспользуемся равенством (30). Напомним, что задачу (24), (25) мы рассматриваем при произвольном фиксированном $h \neq 0$, $h \in K(v) \cap \{h \in E^n : \langle f'(v), h\rangle \leqslant 0\}$, где конус $K(v)$ определен согласно (19); $v = 0$, $I(0) = \{i: i = 1, \ldots, s\}$. Тогда

$$g_0(\varepsilon h) = g_0(0) + \langle g_0'(0), \varepsilon h\rangle + \frac{1}{2}\langle g_0''(0)\varepsilon h, \varepsilon h\rangle + o(\varepsilon^2) =$$

$$= f(0) + \varepsilon\langle f'(0), h\rangle + \frac{1}{2}\varepsilon^2\langle f''(0)h, h\rangle + o(\varepsilon^2) \leqslant \frac{1}{2}\varepsilon^2\langle f''(0)h, h\rangle + o(\varepsilon^2),$$

$$g_i(\varepsilon h) = g_i(0) + \langle g_i'(0), \varepsilon h\rangle + \frac{1}{2}\varepsilon^2\langle g_i''(0)\varepsilon h, \varepsilon h\rangle + o(\varepsilon^2) \leqslant \frac{1}{2}\varepsilon^2\langle g_i''(0)h, h\rangle + o(\varepsilon^2), \quad i \in I(0),$$

Подставим эти неравенства в (30). Учитывая, что $\lambda_{i\varepsilon} \geqslant 0$, $i = 0, \ldots, m$, $\psi'(\chi_\varepsilon) \geqslant 0$, получим

$$\lambda_{0\varepsilon}[\frac{1}{2}\varepsilon^2\langle f''(0)h, h\rangle + o(\varepsilon^2)] + \sum_{i=1}^{s} \lambda_{i\varepsilon}[\frac{1}{2}\varepsilon^2\langle g_i''(0)h, h\rangle + o(\varepsilon^2)] \geqslant$$

$$\geqslant \lambda_{0\varepsilon} g_0(\varepsilon h) + \sum_{i=1}^{s} \lambda_{i\varepsilon} g_i(\varepsilon h) = \lambda_{0\varepsilon}\psi'(\chi_\varepsilon) \geqslant 0.$$

Отсюда, разделив на $\varepsilon^2 > 0$, имеем

$$\lambda_{0\varepsilon}\langle f''(0)h, h\rangle + \sum_{i=1}^{s} \lambda_{i\varepsilon}\langle g_i''(0)h, h\rangle + 2\lambda_{0\varepsilon}\frac{o(\varepsilon^2)}{\varepsilon^2} + \sum_{i=1}^{s} 2\lambda_{i\varepsilon}\frac{o(\varepsilon^2)}{\varepsilon^2} \geqslant 0.$$

Перейдем в этом неравенстве к пределу при $\varepsilon = \varepsilon_k \to 0$. Учитывая, что $\bar\lambda_{\varepsilon_k} \to \bar\lambda = \bar\lambda(h) \in \Lambda_a(0)$, получим

$$\langle (\lambda_0 f''(0) + \sum_{i=1}^{s} \lambda_i g_i''(0))h, h\rangle = \langle \mathcal{L}_{xx}(0, \bar\lambda(h))h, h\rangle \geqslant 0.$$

Так как $\Lambda_a(0)$ конус, то $\frac{\lambda(h)}{|\lambda(h)|} \in \Lambda_a(0)$. Следовательно, $\max\limits_{\bar\lambda \in \Lambda_a(0), |\bar\lambda|=1} \langle \mathcal{L}_{xx}(v, \lambda)h, h\rangle \geqslant \langle \mathcal{L}_{xx}\left(v, \frac{\lambda(h)}{|\lambda(h)|}\right)h, h\rangle \geqslant 0$. Неравенство (18) и, следовательно, теорема 2 доказаны в предположении, что в точке v локального минимума задачи (13.A) все ограничения $g_i(x) \leqslant 0$, $i = 1, \ldots, m$, активны.

Рассмотрим общий случай, когда среди этих ограничений имеются неактивные, т. е. $g_i(v) < 0$ для некоторых номеров i, $1 \leqslant i \leqslant m$ (возможность, когда все такие ограничения неактивны, здесь не исключается). Так как функции $g_i(x)$ непрерывны, то число γ в задаче (19.A) можно считать столь малым, что $g_i(x) < 0$ $\forall x$, $|x - v| \leqslant \gamma$, $i \notin I(0)$. Рассмотрим задачу:

$$g_0(x) \to \inf, \quad x \in W_1 = \{x \in E^n : |x - v| \leqslant \gamma, \ g_i(x) \leqslant 0,$$

$$i \in I(v) \cap \{1 \leqslant i \leqslant m\}, \ g_i(x) = 0, \ i = m+1, \ldots, s\}, \quad (40)$$

полученную из задачи (19.A) исключением неактивных ограничений. Заметим, что $W_1 \subseteq W$, так как если $x \in W_1$, то $g_i(x) < 0$ $\forall i \notin I(v)$ в силу выбора γ и поэтому $x \in W$. Отсюда, учитывая, что $v \in W_1$, имеем $\inf\limits_{x\in W_1} f(x) \geqslant \inf\limits_{x\in W} f(x) = f(v) \geqslant \inf\limits_{x\in W_1} f(x)$. Это значит, что $\inf\limits_{x\in W_1} f(x) = \inf\limits_{x\in W} f(x) = f(v)$, т. е. точка v — решение задачи (40). Однако в задаче (40) все ограничения $g_i(x) \leqslant 0$ в точке v активны, и, следовательно, к этой задаче применима уже доказанная часть теоремы. Поэтому конус Лагранжа $\tilde\Lambda(v)$ и конус Арутюнова $\tilde\Lambda_a(v)$ задачи (40) непусты, и для функции Лагранжа $\tilde{\mathcal{L}}(x, \tilde\lambda) = \sum\limits_{i\in I(0)} \tilde\lambda_i g_i(x)$ этой задачи справедливо неравенство

$$\max_{\tilde\lambda \in \tilde\Lambda_a(v), |\tilde\lambda|=1} \langle \tilde{\mathcal{L}}_{xx}(v, \tilde\lambda)h, h\rangle \geqslant 0 \quad \forall h \in K(v) \cap \{f'(v), h \leqslant 0\}. \quad (41)$$

Каждой точке $\tilde\lambda=(\tilde\lambda_0, \tilde\lambda_i, i\in I(v))$ поставим в соответствие точку $\bar\lambda=(\lambda_0,\dots,\lambda_s)$ по правилу $\lambda_0=\tilde\lambda_0$, $\lambda_i=\tilde\lambda_i$, $i\in I(v)$, $\lambda_i=0$, $i\notin I(v)$. Образуем множества $\Lambda(v)$, $\Lambda_a(v)$, состоящие из всех тех точек $\bar\lambda$, которые получены с помощью указанного правила из точек $\tilde\lambda$ множеств $\tilde\Lambda(v)$, $\tilde\Lambda_a(v)$ соответственно. Нетрудно убедиться, что так построенное множество $\Lambda(v)$ является конусом Лагранжа задачи (13.A) в точке v, а множество $\Lambda_a(v)$ — конусом Арутюнова, причем для каждой точки $\bar\lambda\in\Lambda_a(v)$ в качестве сопровождающего подпространства $\Pi(\bar\lambda)$ можно взять подпространство $\Pi(\tilde\lambda)$ для соответствующей точки $\tilde\lambda\in\tilde\Lambda_a(v)$. Неравенство (18) является следствием неравенства (41) и равенства $\mathcal{L}_{xx}(v,\bar\lambda)=\tilde{\mathcal{L}}_{xx}(v,\tilde\lambda)$ для любых соответствующих точек $\bar\lambda\in\Lambda(v)$ и $\tilde\lambda\in\tilde\Lambda(v)$. Теорема 2 доказана. □

Тем самым доказаны и теоремы 2.4.2 и 2.4.3.

Упражнения

1. Пусть $Q=(Q_0,\dots,Q_s)$, где Q_i — симметричная матрица $n\times n$, $i=0,\dots,s$, $K=\{x\in E^n: \langle Q_i x,x\rangle\leqslant 0,\ i=1,\dots,m;\ \langle Q_i x,x\rangle=0,\ i=m+1,\dots,s\}$. Пусть $\Lambda_a(0,Q)$ — конус Арутюнова задачи $f(x)=\langle Q_0 x,x\rangle\to\inf$, $x\in K$ в точке $v=0$. Доказать, что многозначное отображение Λ_a: $Q\to\Pi(E^{s+1})$ замкнуто (полунепрерывно сверху).

2. Пусть v — точка локального минимума задачи: $f(x)\to\inf$, $x\in X=\{x\in E^n: g_i(x)\leqslant 0,\ i=1,\dots,m;\ g_i(x)=0,\ i=m+1,\dots,s\}$, где $g_i(x)=\langle a_i,F(x)\rangle$, F: $E^n\to E^r$ — заданное отображение, $a_1,\dots,a_s$ — заданные векторы из E^r. Доказать, что тогда конус $\Lambda_a(v)$ можно заменить на другой конус $\tilde\Lambda_a(v)$, в котором коразмерность сопровождающих подпространств Π не превышает r, т. е. $\dim\Pi\geqslant\max\{n-r;0\}$. Указание: заметить, что $g_i'(x)=(F'(x))^{\top}a_i$, и, следовательно, $\langle g_i'(x),h\rangle=0\ \forall h\in\ker F'(x)$ (подробности см. [44]).

§ 17. Метод барьерных функций

1. Идеи метода штрафных функций могут быть использованы для построения минимизирующих последовательностей задачи

$$f(x)\to\inf,\quad x\in X, \tag{1}$$

которые обладают какими-либо дополнительными свойствами. Скажем, можно строить последовательность $\{x_k\}$, каждый член которой принадлежит множеству X, но находится вне некоторого заданного «запрещенного» подмножества $\gamma\subset X$. В качестве «запрещенного» множества γ может служить, например, граница Гр X множества X или какая-либо часть границы. Дело в том, что при применении того или иного метода решения задачи (1) при $X\neq E^n$ может случиться, что каждое получаемое приближение x_k будет принадлежать Гр X. Однако если структура границы множества слишком сложна, то реализация такого метода может потребовать большого объема вычислительной работы и, кроме того, сходимость метода может оказаться очень медленной. В таких случаях можно попробовать как-то построить «барьер» вблизи всей границы $\gamma=$ Гр X или какой-либо ее части γ (или какого-либо другого заданного подмножества $\gamma\subset X$), который исключал бы возможность попадания очередного приближения x_k на γ.

Определение 1. Пусть γ — некоторое подмножество множества X. Функцию $B(x)$ назовем *барьером* или *барьерной функцией* подмножества γ, если $B(x)$ определена, конечна и неотрицательна во всех точках $x\in X\setminus\gamma$, причем $\lim\limits_{r\to\infty}B(v_r)=\infty$ для всех последовательностей $\{v_r\}\in X\setminus\gamma$, которые сходятся к какой-либо точке $v\in\gamma$.

Заметим, что в определении 1 подразумевается, что $X\setminus\gamma\neq\varnothing$. Это значит, что если $\gamma=$ Гр X, то int $X=X\setminus\gamma\neq\varnothing$. Заметим также, что в точках $x\in\gamma$ барьерная функция $B(x)$ не определена (можно принять $B(x)=\infty$, $x\in\gamma$).

Пользуясь теми же конструкциями, которые использовались при построении штрафных функций, нетрудно выписать барьерные функции для множеств γ, задаваемых ограничениями типа равенств или неравенств. Например, если $\gamma=\{x\in E^n: x\in X, g(x)=0\}$, где $g(x)$ непрерывна на X, $X\setminus\gamma\neq\varnothing$, то в качестве барьерной функции здесь можно взять $B(x)=|g(x)|^{-1}$, или $B(x)=|g(x)|^{-2}$, или $B(x)=\max\{-\ln|g(x)|;0\}$. Если же $\gamma=\{x\in E^n: x\in X,\ g(x)\leqslant 0\}$, где $X\setminus\gamma\neq\varnothing$, $g(x)$ непрерывна на X, то можно принять $B(x)=(g(x))^{-p}$, $p>0$, или $B(x)=|\ln g(x)|$, $x\in X\setminus\gamma$ и т. п.

Перейдем к описанию метода барьерных функций для решения задачи (1), предполагая, что подмножество $\gamma\subseteq X$ и некоторая его барьерная функция уже заданы. Введем функции

$$F_k(x)=f(x)+a_kB(x),\quad x\in X\setminus\gamma,\quad k=1,2,\dots, \tag{2}$$

где $\{a_k\}$ — положительная последовательность, сходящаяся к нулю. Величины $\{a_k\}$ из (2) называются *барьерными коэффициентами*. Рассмотрим последовательность задач

$$F_k(x)\to\inf;\quad x\in X\setminus\gamma,\quad k=1,2,\dots \tag{3}$$

Обозначим $F_{k*}=\inf\limits_{X\setminus\gamma}F_K(x)$, $k=1,2,\dots$. Будем предполагать, что в исходной задаче (1) $f_*=\inf\limits_X f(x)>-\infty$. Так как $F_k(x)\geqslant f(x)$ при всех $x\in X\setminus\gamma$, то $F_{k*}\geqslant f_*>-\infty$. Тогда условия

$$x_k\in X\setminus\gamma,\quad F_k(x_k)\leqslant F_{k*}+\varepsilon_k,\quad k=1,2,\dots \tag{4}$$

определяют последовательность $\{x_k\}$, где $\varepsilon_k>0$, $\lim\limits_{k\to\infty}\varepsilon_k=0$; если окажется, что $F_k(x_k)=F_{k*}$, то в (4) допускается $\varepsilon_k=0$.

Поскольку, как обычно, мы подразумеваем, что функция $f(x)$ конечна во всех точках $x\in X$, то согласно определению 1 для любой последовательности $\{v_r\}\in X\setminus\gamma$, $\{v_r\}\to v\in\gamma$ справедливо равенство $\lim\limits_{r\to\infty}F_k(v_r)=\infty$ при каждом фиксированном $k=1,2,\dots$. Таким образом, функция $F_k(x)$ неограниченно возрастает вблизи γ. Поэтому следует ожидать, что при фиксированном k функция $F_k(x)$ вблизи γ не может принимать значения, близкие к F_{k*}, и точка x_k, определяемая условиями (4), не будет расположена на слишком близком расстоянии от γ. В то же время благодаря тому, что барьерные коэффициенты $\{a_k\}\to 0$, не исключается возможность того, что с увеличением номера k точки x_k, постепенно «преодолевая барьер», будут приближаться к γ.

Для приближенного решения задачи (3) при фиксированном k и определения точки x_k, удовлетворяющей условиям (4), могут быть использованы различные методы минимизации. В частности, если $\gamma=$ Гр X и $X\setminus\gamma=$ int $X\neq\varnothing$, то для решения задачи (3) может быть применен, например, градиентный метод (см. § 1):

$$x_{k,r+1}=x_{kr}-\alpha_rF_k'(x_{kr}),\quad x_{k0}=x_{k-1},\quad r=0,1\dots$$

Поскольку $x_{kr}\in$ int X, то при достаточно малых $\alpha_r>0$ точка $x_{k,r+1}$ также будет принадлежать int X, и мы избавлены от неудобств, связанных с

учетом границы X, — нужно лишь на каждой итерации следить за соблюдением включения $x_k \in \operatorname{int} X$, а при его нарушении уменьшать длину шага α_r. Правда, для этого величину α_{r1}, быть может, придется брать слишком малой, и сходимость градиентного метода, возможно, замедлится, но это уже будет «платой» за выполнение условия $x_k \in \operatorname{int} X$.

Дальнейшее изложение не зависит от того, с помощью какого конкретного метода минимизации будет найдена точка x_k, удовлетворяющая условиям (4). Поэтому мы здесь можем ограничиться предположением, что имеется достаточно удобный метод определения точки x_k из (4).

Метод барьерных функций описан. Отметим, что в литературе этот метод иногда называют *методом внутренних штрафов* (или методом внутренней точки), а метод штрафных функций из § 15 — *методом внешних штрафов* (или методом внешней точки) [721]. Для иллюстрации метода барьерных функций приведем пример.

Пример 1. Пусть требуется решить задачу

$$f(x) = -x \to \inf; \quad x \in X = \{x \in E^1: g(x) = x \leqslant 0\}.$$

Очевидно, здесь $f_* = 0$, $X_* = \{0\}$. Границей множества X является $\gamma = \operatorname{Гр} X = \{x \in E^1: g(x) = x = 0\} = \{0\}$, а $X \setminus \gamma = \{x \in E^1: g(x) = x < 0\} = \operatorname{int} X$. В качестве барьерной функции для γ возьмем $B(x) = -1/x$ $(x < 0)$. Пусть $a_k = k^{-1}$, $k = 1, 2, \ldots$ Тогда функция (2) будет иметь вид $F_k(x) = -x - (kx)^{-1}$, $x < 0$. Нетрудно видеть, что здесь $F_{k*} = \inf\limits_{x<0} F_k(x) = 2/\sqrt{k}$ и точка $x_k = -1/\sqrt{k}$ удовлетворяет условиям (4) при $\varepsilon_k = 0$, $k = 1, 2, \ldots$ Ясно также, что $\lim\limits_{k\to\infty} F_{k*} = \lim\limits_{k\to\infty} f(x_k) = 0 = f_*$, $\lim\limits_{k\to\infty} x_k = 0 = x_*$.

В качестве барьерной функции здесь можно также взять и $B(x) = |\ln(-x)|$. В этом случае $F_k(x) = -x + |\ln(-x)|k^{-1}$, $x < 0$, $k = 1, 2, \ldots$, $F_{k*} = (1 + \ln k)k^{-1}$, а точка $x_k = -k^{-1}$ удовлетворяет условиям (4) при $\varepsilon_k = 0$. И здесь $\{f(x_k)\} \to f_* = 0$, $\{x_k\} \to x_* = 0$.

Перейдем к исследованию сходимости метода барьерных функций.

Теорема 1. *Пусть γ — некоторое подмножество из X, $X \setminus \gamma \neq \varnothing$, и*

$$f_* = f_{**}, \quad \text{где} \quad f_* = \inf_X f(x), \quad f_{**} = \inf_{X\setminus\gamma} f(x) > -\infty. \tag{5}$$

Пусть $B(x)$ — какая-либо барьерная функция подмножества γ, а последовательность $\{x_k\}$ определена условиями (4). *Тогда*

$$\lim_{k\to\infty} F_{k*} = \lim_{k\to\infty} F_k(x_k) = \lim_{k\to\infty} f(x_k) = f_*, \quad \lim_{k\to\infty} a_k B(x_k) = 0. \tag{6}$$

Кроме того, если множество X ограничено и замкнуто, а $f(x)$ полунепрерывна снизу на X, то $\{x_k\}$ сходится к X_*.

Доказательство. Из определения f_{**}, F_{k*}, неотрицательности барьерной функции и условий (4) следует

$$-\infty < f_{**} \leqslant f(x_k) \leqslant F_k(x_k) \leqslant F_{k*} + \varepsilon_k \leqslant F_k(x) + \varepsilon_k = f(x) + a_k B(x) + \varepsilon_k \tag{7}$$

при всех $x \in X \setminus \gamma$, $k = 1, 2, \ldots$ Так как $B(x)$ конечна в любой точке $x \in X \setminus \gamma$, $\{a_k\} \to 0$, то из (7) при $k \to \infty$ получим

$$f_{**} < \varliminf_{k\to\infty} F_{k*} \leqslant \varlimsup_{k\to\infty} F_{k*} \leqslant f(x), \quad x \in X \setminus \gamma.$$

Переходя в этих неравенствах к нижней грани по $x \in X \setminus \gamma$, будем иметь $f_{**} \leqslant \varliminf\limits_{k\to\infty} F_{k*} \leqslant \varlimsup\limits_{k\to\infty} F_{k*} \leqslant f_{**}$, т. е. $\lim\limits_{k\to\infty} F_{k*} = f_{**}$. Отсюда и из (7) вытекает

$\lim\limits_{k\to\infty} F_k(x_k) = \lim\limits_{k\to\infty} f(x_k) = f_{**}$. Так как $f_* = f_{**}$, то первые соотношения (6) доказаны. А тогда из $0 \leqslant a_k B(x_k) = F_k(x_k) - f(x_k) \to 0$ при $k \to \infty$ получим и второе из соотношений (6). Последнее утверждение о сходимости минимизирующей последовательности $\{x_k\}$ к X_* следует из теоремы 2.1.1. □

Полезно заметить, что при доказательстве теоремы 1 были использованы не все свойства барьерных функций: соотношение $\lim\limits_{r\to\infty} B(v_r) = \infty$, где $\{v_r\} \in X \setminus \gamma$, $\{v_r\} \to v \in \gamma$, нам не понадобилось. Поэтому теорему 1 и ряд доказываемых ниже теорем можно использовать не только как теоремы о сходимости метода барьерных функций, но и как утверждения, выражающие собой достаточные условия устойчивости нижней грани относительно возмущений (погрешностей) минимизируемой функции и некоторых типов возмущений множества, на котором ищется минимум.

2. Рассмотрим возможности построения барьерных функций для задачи (1) в случае, когда

$$X = \{x \in X_0,\ g_i(x) \leqslant 0,\ i = 1, \ldots, m\}; \tag{8}$$

здесь X_0 — заданное множество из E^n, функции $g_1(x), \ldots, g_m(x)$ определены и полунепрерывны снизу на X_0. Положим

$$\gamma = \{x \in X:\ g_i(x) = 0 \quad \text{хотя бы для одного } i,\ 1 \leqslant i \leqslant m\}. \tag{9}$$

Будем предполагать, что множество

$$X(-0) = \{x \in X_0:\ g_i(x) < 0, \quad i = 1, \ldots, m\} \tag{10}$$

непусто. Тогда $X \setminus \gamma = X(-0) \neq \varnothing$. Довольно широкий класс барьерных функций для множества (9) дает следующая конструкция:

$$B(x) = \sum_{i=1}^{m} \varphi_i(-g_i(x)), \quad x \in X(-0), \tag{11}$$

где $\varphi_i(t)$ — неотрицательная функция переменной $t > 0$ такая, что $\lim\limits_{t\to\infty} \varphi_i(t) = \infty$ при всех $i = 1, \ldots, m$. В самом деле, возьмем произвольную последовательность $\{v_r\} \in X \setminus \gamma$, сходящуюся к некоторой точке $v \in \gamma$. Согласно (9) тогда найдется номер j, $1 \leqslant j \leqslant m$, для которого $g_j(v) = 0$. Так как $g_j(x)$ полунепрерывна снизу, а $g_j(v_r) < 0$, $r = 1, 2, \ldots$, то $0 = g_j(v) \leqslant \varliminf\limits_{r\to\infty} g_j(v_r) \leqslant \varlimsup\limits_{r\to\infty} g_j(v_r) \leqslant 0$, т. е. $\lim\limits_{r\to\infty} g_j(v_r) = 0$. Это значит, что $B(v_r) \geqslant \varphi_j(-g_j(v_r)) \to \infty$ при $r \to \infty$, так что функция (11) является барьерной для множества (9).

При необходимости в (11) функции $\varphi_i(t)$ нетрудно выбрать так, чтобы барьерная функция $B(x)$ обладала различными полезными свойствами, такими, как непрерывность, гладкость, выпуклость, простота вычисления значения функции и нужных ее производных и т. п., если, конечно, исходные данные в задаче (1), (8) обладают такими свойствами. Например, взяв в (11) $\varphi_i(t) = 1/t$ или $\varphi_i(t) = (\max\{-\ln t; 0\})^p$, $p \geqslant 1$, получим соответственно

$$\begin{aligned} B(x) &= -\sum_{i=1}^{m} \frac{1}{g_i(x)}, \\ B(x) &= \sum_{i=1}^{m} (\max\{-\ln(-g_i(x)); 0\})^p, \quad x \in X(-0). \end{aligned} \tag{12}$$

Если X_0 выпукло, функции $g_i(x)$, $i=1,\ldots,m$, выпуклы на X_0, то множество $X(-0)=X\setminus\gamma$ выпукло и функции (12) также будут выпуклыми на $X(-0)$ — это следует из следствий к теореме 4.2.8. Далее, функции (12) будут обладать той же гладкостью, какой обладают функции $g_i(x)$, $i=1,\ldots,m$ — у второй функции (12) для этого нужно взять параметр p достаточно большим.

Может сложиться впечатление, что если функции $g_1(x),\ldots,g_m(x)$ непрерывны на X_0, то множество γ, определяемое условиями (9), будет состоять только лишь из граничных точек множества (8). Однако это не всегда так — множество γ может содержать и внутренние точки X.

Пример 2. Пусть $g(x)=|x|-1$ при $|x|\leqslant 1$, $g(x)=0$ при $1<|x|<2$, $g(x)=|x|-2$ при $|x|\geqslant 2$. Тогда множество $X=\{x\in E^1=X_0\colon g(x)\leqslant 0\}$ представляет собой отрезок $-2\leqslant x\leqslant 2$ на числовой оси, а $\operatorname{int} X=\{x\in E^1\colon -2<x<2\}$. В то же время множество $X(-0)=\{x\in E^1\colon g(x)<0\}=\{x\colon -1<x<1\}\subset\operatorname{int} X$, но $X(-0)\neq\operatorname{int} X$, а $\gamma=\{x\in X\colon g(x)=0\}=\{x\colon 1\leqslant|x|\leqslant 2\}$ наряду с граничными точками $x=2$ и $x=-2$ содержит и внутренние точки множества X (см. также множество X из примера 4.9.2).

Таким образом, для множества (8) не всегда выполняется равенство

$$\Gamma p\ X=\Gamma p\ X_0\cup\gamma, \tag{13}$$

где γ определяется условиями (9), а функции (11), (12), являющиеся барьерными функциями для подмножества γ, могут и не быть таковыми хотя бы для части границы X.

3. Отдельно остановимся на условии (5), которое было существенно использовано в теореме 1 при доказательстве сходимости метода барьерных функций.

Нетрудно привести примеры задач (1), (8), в которых функции $f(x), g_1(x),\ldots,g_m(x)$ непрерывны, множество X замкнуто и ограничено, но условие (5) не имеет места. Например, если $f(x)=x$, а множество X взято из примера 2, то $f_{**}=-1>f_*=-2$.

Однако даже выполнение условия (13), при котором функции (11), (12) будут барьерными функциями γ — части границы X, еще не гарантирует справедливость равенства (5).

Пример 3. Пусть $f(u_k)=e^{-x}$, $X=\{u=(x,y)\in E^2=X_0\colon g(u)=(x^2+y^2-1)(y-1)^2\leqslant 0\}$. Тогда $\gamma=\Gamma p\ X=\{u\in X\colon g(u)=0\}=\{u\in E^2\colon x^2+y^2=1$ или $y=1\}$, $X_0=E^2$, $\Gamma p\ X_0=\varnothing$, так что условие (13) выполнено. Далее, $X\setminus\gamma=X(-0)=\{u\in E^2\colon g(u)<0\}=\{u\colon x^2+y^2<1\}=\operatorname{int} X$, поэтому $f_{**}=\inf\limits_{X\setminus\gamma} f(u)=e^{-1}>0$. В то же время $f_*=\lim\limits_{k\to\infty} f(u_k)=0$, где $u_k=(k,1)\in X$ при $k=1,2,\ldots$

Заметим, что в этом примере $\overline{X(-0)}=\{u\colon x^2+y^2\leqslant 1\}\subset X$, но $\overline{X(-0)}\neq X$. (Напоминаем, что через $\overline{Z}$ мы обозначаем замыкание множества Z.)

Приведем две теоремы, дающие достаточные условия для выполнения равенства (5). Имея в виду дальнейшие применения, утверждения сформулируем для множества

$$X(C)=\{x\in E^n\colon x\in X_0,\ g_i(x)\leqslant C,\ i=1,\ldots,m\}, \tag{14}$$

где C — некоторая постоянная. Обозначим

$$f_*(C)=\inf_{X(C)} f(x),\quad f_*(C-0)=\lim_{\varepsilon\to+0} f_*(C-\varepsilon),\quad f_*(C+0)=\lim_{\varepsilon\to+0} f_*(C+\varepsilon). \tag{15}$$

Теорема 2. *Пусть для некоторых C, $\varepsilon_0>0$ множество $X(C-\varepsilon_0)$ непусто и*

$$\overline{X(C-0)}=X(C), \tag{16}$$

где

$$X(C-0)=\{u\in X_0\colon g_i(u)<C,\ i=1,\ldots,m\}.$$

Пусть, кроме того, функция $f(x)$ полунепрерывна сверху на множестве $X(C)$. Тогда

$$f_*(C-0)=f_*(C)=\inf_{X(C-0)} f(x). \tag{17}$$

Доказательство. Прежде всего, заметим, что

$$X(C-\varepsilon)\subseteq X(C-\delta)\subseteq X(C-0)\subseteq X(C)$$

при всех $0<\delta<\varepsilon\leqslant\varepsilon_0$, поэтому

$$f_*(C)\leqslant\inf_{X(C-0)} f(x)\leqslant f_*(C-\delta)\leqslant f_*(C-\varepsilon).$$

Это значит, что функция $f_*(C)$ переменной C не возрастает и существует предел

$$\lim_{\varepsilon\to+0} f_*(C-\varepsilon)=f_*(C-0)\geqslant\inf_{X(C-0)} f(x)\geqslant f_*(C). \tag{18}$$

Возьмем произвольную точку $x\in X(C)$. В силу условия (16) найдется последовательность $\{x_k\}\in X(C-0)$, сходящаяся к точке x. Это значит, что $x_k\in X_0$, $g_i(x_k)\leqslant C-\varepsilon_{ik}<C$, $\varepsilon_{ik}>0$, $k=1,2,\ldots$, где $\lim\limits_{k\to\infty}\varepsilon_{ik}=0$, $i=1,\ldots,m$. Таким образом, $x_k\in X(C-\varepsilon_k)$, где $\varepsilon_k=\min\limits_{1\leqslant i\leqslant m}\varepsilon_{ik}>0$, $\{\varepsilon_k\}\to 0$, и $f_*(C-\varepsilon_k)\leqslant f(x_k)$, $k=1,2,\ldots$ Отсюда при $k\to\infty$, учитывая полунепрерывность сверху функции $f(x)$, получим $\lim\limits_{\varepsilon\to+0} f_*(C-\varepsilon)=\lim\limits_{k\to\infty} f_*(C-\varepsilon_k)\leqslant f(x)$. В силу произвольности $x\in X(C)$ тогда $f_*(C-0)\leqslant f_*(C)$. Сравнивая это неравенство с (18), приходим к равенству (17). Теорема 2 доказана. □

Таким образом, если условия теоремы 2 выполнены при $C=0$, то методом барьерных функций (2), (4), (8)–(11) для задачи (1), (8) можно получить последовательность $\{x_k\}$, обладающую свойствами (6).

Аналогичное утверждение справедливо для выпуклых задач (1), (8).

Теорема 3. *Пусть X_0 — выпуклое множество на E^n, функции $f(x)$, $g_1(x),\ldots,g_m(x)$ выпуклы на X_0. Тогда равенства* (17) *справедливы при всех $C>C_*=\max\limits_{1\leqslant i\leqslant m}\inf\limits_{X_0} g_i(x)$.*

Доказательство. Как было установлено в теореме 2, функция $f_*(C)$ переменной C не возрастает. Возьмем произвольные C, $\varepsilon>0$, $C>C-\varepsilon>C_*$. Пусть $x\in X(C)$, $y\in X(C-\varepsilon)$. В силу выпуклости X_0 тогда $x_\alpha=\alpha y+(1-\alpha)x\in X_0$ при всех α, $0\leqslant\alpha\leqslant 1$. Кроме того, из выпуклости $g_i(x)$ имеем $g_i(x_\alpha)\leqslant\alpha g_i(y)+(1-\alpha)g_i(x)\leqslant\alpha(C-\varepsilon)+(1-\alpha)C=C-\alpha\varepsilon$, $0<\alpha\leqslant 1$. Это значит, что $x_\alpha\in X(C-\alpha\varepsilon)$. Тогда с учетом выпуклости функции $f(x)$ получим $f_*(C)\leqslant\inf\limits_{X(C-0)} f(x)\leqslant f_*(C-\alpha\varepsilon)\leqslant f(x_\alpha)\leqslant\alpha f(y)+(1-\alpha)f(x)$ для всех $x\in X(C)$, $y\in X(C-\varepsilon)$. Следовательно, $f_*(C)\leqslant\inf\limits_{X(C-0)} f(x)\leqslant f_*(C-\alpha\varepsilon)\leqslant\alpha f_*(C-\varepsilon)+(1-\alpha)f_*(C)\leqslant f_*(C)+\alpha[f_*(C-\varepsilon_0)-f_*(C)]$ для всех α, $0<\alpha\leqslant 1$, $0<\varepsilon\leqslant\varepsilon_0<C-C_*$. Отсюда при $\alpha\to+0$ с учетом монотонности $f_*(C)$ получаем равенства (17), что и требовалось. □

4. Пусть множество X задается условиями

$$X=\{x\in E^n\colon x\in X_0,\quad g_i(x)\leqslant 0,\quad i=1,\ldots,m;\quad g_i(x)=0,\quad i=m+1,\ldots,s\}. \tag{19}$$

Если это множество не имеет внутренних точек, то реализация ряда методов минимизации (например, методов из § 3–5, 11 и др.) на X может стать затруднительной или даже невозможной. В то же время при применении методов § 14, 15 к задаче (1), (19) могут получиться такие последовательности $\{x_k\}$, которые не принадлежат множеству X и нарушают какие-либо из ограничений $g_i(x)\leqslant 0$, $g_j(x)=0$ на недопустимо большую величину. В таких случаях может оказаться целесообразным использование метода барьерных функций.

Заметим, что этот метод выше изложен для задачи (1), (8) в предположении, что множество $X(-0)$, определяемое условиями (10), непусто. Однако такое предположение для множества (19) при $m<s$ не имеет смысла. Поэтому описанный выше метод барьерных функций к задаче (1), (19) непосредственно неприменим и требует модификации, обобщения. Опишем один из возможных здесь подходов [390].

Введем теперь последовательность расширенных множеств

$$V_k = \{x \in X_0\colon g_i(x) \leqslant \theta_k,\ i = 1, \ldots, m;\ |g_i(x)| \leqslant \theta_k,\ i = m+1, \ldots, s\}, \tag{20}$$

где $\theta_k > 0$, $k = 1, 2, \ldots$, $\lim\limits_{k\to\infty} \theta_k = 0$. Так как $X \subset V_k$, $k = 1, 2, \ldots$, то из $X \neq \varnothing$ следует $V_k \neq \varnothing$, $k = 1, 2, \ldots$ Предполагая, что функция $f(x)$ определена на множестве $\bigcup\limits_{k=1}^{\infty} V_k$, рассмотрим последовательность задач

$$f(x) \to \inf; \quad x \in V_k, \quad k = 1, 2, \ldots \tag{21}$$

Для решения задач (21) могут быть использованы различные методы минимизации. Мы здесь остановимся лишь на методе барьерных функций. Обозначим

$$\gamma_k = \{x \in V_k \text{ и выполняется хотя бы одно из равенств } g_i(x) = \theta_k,\ i = 1, \ldots, m;$$
$$g_j(x) = \theta_k,\ g_j(x) = -\theta_k,\ j = m+1, \ldots, s\}. \tag{22}$$

Поскольку $X \subset V_k$, $X \cap \gamma_k = \varnothing$, то $X \subset V_k \setminus \gamma_k \neq \varnothing$, $k = 1, 2, \ldots$ В качестве барьерной функции $B_k(x)$ подмножества γ_k возьмем

$$B_k(x) = \sum_{i=1}^{s} \varphi_i(\theta_k - g_i(x)) + \sum_{i=m+1}^{s} \varphi_i(\theta_k + g_i(x)), \quad x \in V_k \setminus \gamma_k, \tag{23}$$

где функция $\varphi_i(t)$ определена, конечна, неотрицательна и не возрастает при $t > 0$, $\lim\limits_{t\to+0} \varphi_i(t) = \infty$, $i = 1, \ldots, s$. Например, в качестве $\varphi_i(t)$ можно взять $\varphi_i(t) = t^{-1}$, $\varphi(t) = (\max\{-\ln t; 0\})^p$, $p \geqslant 1$.

Далее, составим функцию

$$F_k(x) = f(x) + a_k B_k(x), \quad x \in V_k \setminus \gamma_k, \tag{24}$$

где $\{a_k\}$ — барьерные коэффициенты: $a_k > 0$, $k = 1, 2, \ldots$, $\{a_k\} \to 0$. В отличие от рассмотренного выше варианта метода барьерных функций, здесь мы будем требовать, чтобы барьерные коэффициенты $\{a_k\}$ и параметры $\{\theta_k\}$ стремились к нулю согласованно в следующем смысле:

$$\lim_{k\to\infty} a_k \varphi_i(\theta_k) = 0, \quad i = 1, \ldots, s. \tag{25}$$

Предположим, что $f_{k*} = \inf\limits_{V_k} f(x) > -\infty$, $k = 1, 2, \ldots$ Так как $B_k(x) \geqslant 0$, $a_k > 0$, то $F_k(x) \geqslant f(x)$ при всех $x \in V_k \setminus \gamma_k$, и поэтому $F_{k*} = \inf\limits_{V_k \setminus \gamma_k} F_k(x) \geqslant f_{k*} > -\infty$, $k = 1, 2, \ldots$ С помощью какого-либо метода минимизации определим точку x_k, удовлетворяющую условиям

$$x_k \in V_k \setminus \gamma_k, \quad F_{k*} \leqslant F_k(x_k) \leqslant F_{k*} + \varepsilon_k, \quad k = 1, 2, \ldots, \tag{26}$$

где $\{\varepsilon_k\}$ — некоторая положительная последовательность, сходящаяся к нулю; если $F_k(x_k) = F_{k*}$, то в (26) допускается $\varepsilon_k = 0$. Метод барьерных функций для задачи (1), (19) описан.

Теорема 4. *Пусть функции $F_k(x)$, $B_k(x)$, множество V_k, γ_k определены соотношениями* (20), (22)–(24), *выполняются равенства* (25) *и, кроме того,*

$$\lim_{k\to\infty} f_{k*} = f_* > -\infty, \quad f_{k*} = \inf_{V_k} f(x), \quad f_* = \inf_{X} f(x). \tag{27}$$

Тогда для последовательности $\{x_k\}$, определяемой условиями (20), *справедливы соотношения*

$$\lim_{k\to\infty} F_{k*} = \lim_{k\to\infty} F_k(x_k) = \lim_{k\to\infty} f(x_k) = f_*, \quad \lim_{k\to\infty} a_k B_k(x_k) = 0. \tag{28}$$

Если, кроме того, множество

$$X(\delta) = \{x \in X_0\colon g_i(x) \leqslant \delta,\ i = 1, \ldots, m;\ |g_i(x)| \leqslant \delta,\ i = m+1, \ldots, s\} \tag{29}$$

компактно при некотором $\delta > 0$, множество X_0 замкнуто, а функции $g_1(x), \ldots, g_m(x)$, $|g_{m+1}(x)|, \ldots, |g_s(x)|$ полунепрерывны снизу на $X(\delta)$, то $\{x_k\} \to X_*$ — множество решений задачи (1), (19).

Доказательство. Из определения f_{k*}, F_{k*}, неотрицательности $B_k(x)$ и условий (26) имеем

$$-\infty < f_{k*} \leqslant f(x_k) \leqslant F_{k*} + \varepsilon_k \leqslant F_k(x) + \varepsilon_k = f(x) + a_k B_k(x) + \varepsilon_k, \quad x \in V_k \setminus \gamma_k, \quad k = 1, 2, \ldots \tag{30}$$

Так как функции $\varphi_i(t)$ из (23) не возрастают при $t > 0$, то $\varphi_i(\theta_k - g_i(x)) \leqslant \varphi_i(\theta_k)$, $i = 1, \ldots, m$; $\varphi_i(\theta_k \pm g_i(x)) = \varphi_i(\theta_k)$, $i = m+1, \ldots, s$, для всех $x \in X$. Поэтому в силу условия (25)

$$0 \leqslant a_k B_k(x) \leqslant 2a_k \sum_{i=1}^{s} \varphi_i(\theta_k) \to 0, \quad k \to \infty, \quad \forall x \in X. \tag{31}$$

Тогда при $k \to \infty$ из (30) с учетом условия (27) получим

$$f_* \leqslant \varliminf_{k\to\infty} F_{k*} \leqslant \varlimsup_{k\to\infty} F_{k*} \leqslant f(x) \quad \text{при всех} \quad x \in X.$$

Переходя к нижней грани по $x \in X$, отсюда имеем $\lim\limits_{k\to\infty} F_{k*} = f_*$. Тогда из (30) следует $\lim\limits_{k\to\infty} F_k(x_k) = \lim\limits_{k\to\infty} f(x_k) = f_*$. Наконец, $0 \leqslant a_k B_k(x_k) = F_k(x_k) - f(x_k) \to 0$ при $k \to \infty$. Равенства (28) доказаны.

Пусть теперь выполнены все условия теоремы. Так как $\{\theta_k\} \to 0$, то $V_k \subset X(\delta)$ при всех $k \geqslant k_0$. Тогда $x_k \in X(\delta)$, $k \geqslant k_0$. В силу компактности $X(\delta)$ последовательность $\{x_k\}$ имеет хотя бы одну предельную точку. Пусть x_* — произвольная предельная точка $\{x_k\}$, пусть подпоследовательность $\{x_{k_r}\} \to x_*$. В силу замкнутости X_0 тогда $x_* \in X_0$. Из полунепрерывности снизу функций $g_1(x), \ldots, g_m(x)$, $|g_{m+1}(x)|, \ldots, |g_s(x)|$ и условия $x_k \in V_k$ следует, что

$$g_i(x_*) \leqslant \varliminf_{r\to\infty} g_i(x_{k_r}) \leqslant \lim_{k\to\infty} \theta_k = 0, \quad i = 1, \ldots, m, \quad |g_i(x_*)| \leqslant \varliminf_{r\to\infty} |g_i(x_{k_r})| \leqslant \lim_{k\to\infty} \theta_k = 0$$

или

$$g_i(x_*) = 0, \quad i = m+1, \ldots, s.$$

Таким образом, $x_* \in X$. Отсюда с учетом полунепрерывности снизу $f(x)$ на $X(\delta)$ получим $f_* \leqslant f(x_*) \leqslant \varliminf\limits_{r\to\infty} f(x_{k_r}) = \lim\limits_{k\to\infty} f(x_k) = f_*$, т. е. $f(x_*) = f_*$ или $x_* \in X_*$. Тем самым показано, что любая предельная точка последовательности $\{x_k\}$ принадлежит X_*. Отсюда следует, что $\{x_k\} \to X_*$. Теорема 4 доказана. □

При некоторых более жестких ограничениях на данные задачи (1), (19) можно получить оценки погрешности метода (20)–(26).

Теорема 5. *Пусть для задачи* (1), (19) *справедливо неравенство*

$$-\infty < f_* \leqslant f(x) + \sum_{i=1}^{s} c_i (g_i^+(x))^\nu \quad \forall x \in X_0, \quad c_i \geqslant 0, \quad \nu > 0 \tag{32}$$

(*см. определение* 15.3 *и леммы* 15.1, 15.5). *Тогда последовательность $\{x_k\}$, определяемая, условиями* (20)–(26), *существует и справедливы оценки*

$$-|c|_1 \theta_k^\nu \leqslant f(x_k) - f_* \leqslant F_k(x_k) - f_* \leqslant 2a_k \sum_{i=1}^{s} \varphi_i(\theta_k) + \varepsilon_k, \tag{33}$$

$$0 \leqslant a_k B_k(x_k) \leqslant 2a_k \sum_{i=1}^{s} \varphi_i(\theta_k) + \varepsilon_k + |c|_1 \theta_k^\nu, \quad k = 1, 2, \ldots, \tag{34}$$

где $|c|_1 = \sum\limits_{i=1}^{s} |c_i|$. Если, кроме того, множество (29) *компактно при некотором $\delta > 0$, X_0 замкнуто, а функции $f(x), g_1(x), \ldots, g_m(x)$, $|g_{m+1}(x)|, \ldots, |g_s(x)|$ полунепрерывны снизу на $X(\delta)$, то $\{x_k\} \to X_*$.*

Доказательство. Из определения (20) множества V_k и условия (32) следует

$$-\infty < f_* \leqslant f(x) + \sum_{i=1}^{s} c_i (g_i^+(x))^\nu \leqslant f(x) + |c|_1 \theta_k^\nu \leqslant F_k(x) + |c|_1 \theta_k^\nu \tag{35}$$

при всех $x \in V_k \setminus \gamma_k$. Отсюда имеем $F_k(x) \geqslant f_* - |c|_1 \theta_k^\nu > -\infty$, $x \in V_k \setminus \gamma_k$, или $F_{k*} \geqslant f_* - |c|_1 \theta_k^\nu > -\infty$, $k = 1, 2, \ldots$ Таким образом, последовательность $\{x_k\}$, удовлетворяющая условиям (26), существует. Далее из (31) следует

$$0 \leqslant a_k B_k(x_*) \leqslant 2a_k \sum_{i=1}^{s} \varphi_i(\theta_k), \quad x_* \in X_* \subset X \subset X_k \setminus \gamma_k, \quad k = 1, 2, \ldots$$

Поэтому с учетом неравенств (26), (35) имеем

$$f_* \leqslant f(x_k) + |c|_1 \theta_k^\nu \leqslant F_k(x_k) + |c|_1 \theta_k^\nu \leqslant F_{k*} + \varepsilon_k + |c|_1 \theta_k^\nu \leqslant$$
$$\leqslant F_k(x_*) + \varepsilon_k + |c|_1 \theta_k^\nu \leqslant f_* + 2a_k \sum_{i=1}^{s} \varphi_i(\theta_k) + \varepsilon_k + |c|_1 \theta_k^\nu, \quad k = 1, 2, \ldots$$

Отсюда получаем оценку (33).

Далее, из соотношений $0 \leqslant a_k B_k(x_k) = (F_k(x_k) - f_*) - (f(x_k) - f_*)$ и уже доказанной оценки (33) вытекает оценка (34). Последнее утверждение доказывается так же, как аналогичное утверждение теоремы 4. □

5. Отдельно остановимся на условии (27), которое существенно использовалось при доказательстве равенств (28). Нетрудно привести примеры задач (1), (19), когда это условие не выполняется.

Пример 4. Пусть $f(x) = e^{-x}$, $X = \{x \in E^1 = X_0\colon g(x) = (x^2-1)(1+x^4)^{-1} \leqslant 0\}$. Ясно, что $X = \{x \in E^1\colon |x| \leqslant 1\}$, $f_* = \inf\limits_X f(x) = e^{-1}$. Возьмем $V_k = \{x \in E^1\colon g(x) \leqslant \theta_k = 1/k^2\}$. Так как $x_r = r \in V_k$ при $r \geqslant k$, то $\lim\limits_{r\to\infty} f(x_r) = 0 = f_{k*}$, $k = 1, 2, \ldots$ Таким образом, здесь $\lim\limits_{k\to\infty} f_{k*} = 0 < e^{-1} = f_*$ — условие (27) не выполняется. Заметим, что в рассмотренном примере множество $X(\delta) = \{x \in E^1\colon g(x) \leqslant \delta\}$ не является компактным ни при каком $\delta > 0$.

Приведем теорему, дающую достаточные условия для выполнения условия (27).

Теорема 6. *Пусть множество X_0 замкнуто, функции $f(x), g_1(x), \ldots, g_m(x)$, $|g_{m+1}(x)|, \ldots, |g_s(x)|$ определены и полунепрерывны снизу на X_0. Кроме того, пусть множество*

$$X(C) = \{x \in E^n\colon x \in X_0,\ g_i^+(x) \leqslant C,\ i = 1, s\}$$

непусто, а множество $X(C + \varepsilon_0)$ ограничено и замкнуто при некотором $\varepsilon_0 > 0$. Тогда (см. обозначения (15))

$$\lim_{\varepsilon\to+0} f_*(C+\varepsilon) = f_*(C+0) = f_*(C). \tag{36}$$

Доказательство. Так как $X(C) \subseteq X(C+\delta) \subseteq X(C+\varepsilon)$ при любых $0 < \delta < \varepsilon \leqslant \varepsilon_0$, то $f_*(C+\varepsilon) \leqslant f_*(C+\delta) \leqslant f_*(C)$. Таким образом, функция $f_*(C)$ переменной C не возрастает и существует предел $\lim\limits_{\varepsilon\to+0} f_*(C+\varepsilon) = f_*(C+0) \leqslant f_*(C)$. Возьмем произвольную последовательность $\{\varepsilon_k\}$, $0 < \varepsilon_k \leqslant \varepsilon_0$, сходящуюся к нулю. При сделанных предположениях множества $X(C+\varepsilon_k) \subseteq X(C+\varepsilon_0)$ при каждом $k = 1, 2, \ldots$ ограничены и замкнуты. Согласно теореме 2.1.1 тогда существует точка $w_k \in X(C+\varepsilon_k)$ такая, что $f(w_k) = f_*(C+\varepsilon_k)$, $k = 1, 2, \ldots$ Поскольку $X(C+\varepsilon_0)$ — компактное множество и $w_k \in X(C+\varepsilon_k) \subseteq X(C+\varepsilon_0)$, то последовательность $\{w_k\}$ имеет хотя бы одну предельную точку. Пусть w_* — какая-либо предельная точка $\{w_k\}$. Не умаляя общности, можем считать, что сама последовательность $\{w_k\} \to w_*$. По построению $w_k \in X(C+\varepsilon_k)$, т. е. $w_k \in X_0$, $g_i^+(w_k) \leqslant C + \varepsilon_k$, $i = 1, \ldots, s$. Используя замкнутость множества X_0, полунепрерывность рассматриваемых функций, отсюда при $k \to \infty$ получаем $w_* \in X(C)$. А тогда $f_*(C) \leqslant f(w_*) \leqslant \lim\limits_{k\to\infty} f(w_k) = \lim\limits_{k\to\infty} f_*(C+\varepsilon_k) = f_*(C+0)$. Сравнивая с ранее установленным неравенством $f_*(C+0) \leqslant f_*(C)$, получаем равенство (36).

Нетрудно видеть, что при $C = 0$ из (36) вытекает условие (27).

Различные аспекты метода барьерных функций исследованы в [222; 286; 319; 390; 613; 721; 759; 774].

Упражнения

1. Применить метод барьерных функций к задачам:

а) $f(u) = x + y \to \inf$; $u \in X = \{u = (x, y) \in E^2\colon g_1(u) = x^2 - y^2 \leqslant 0,\ g_2(u) = -x \leqslant 0\}$;

б) $f(u) = y \to \inf$; $u \in X = \{u = (x, y) \in E^2\colon g(u) = \sin x + x - y \leqslant 0\}$;

в) $f(u) = (u-1)^3 \to \inf$; $u \in X = \{u = (x, y) \in E^1\colon g(u) = -u - 1 \leqslant 0\}$;

г) к задачам из упражнений 15.1;

д) к задачам из примеров § 2.3, если считать, что множество γ совпадает с границей множества X.

§ 18. Метод нагруженных функций

1. Методы, рассмотренные в § 15, 17, объединяет общая идея — в ней исходная задача минимизации заменяется свойством вспомогательных задач минимизации, в которых множество имеет более «простую» структуру, а целевая функция становится более «сложной» и содержит штрафные или

барьерные слагаемые, учитывающие ограничения, задающие множество исходной задачи. К методам такого типа относится также метод модифицированных функций Лагранжа из § 14, в котором вместо исходной задачи минимизации решается задача поиска седловой точки функции Лагранжа на «простом» множестве $X_0 \times \Lambda_0$. К упомянутым методам идейно примыкает и излагаемый ниже метод нагруженных функций. Как и выше будем рассматривать задачу

$$f(x) \to \inf; \quad X = \{x \in E^n\colon x \in X_0,\ g_i(x) \leqslant 0,\ i = 1, \ldots, m;\ g_i(x) = 0,\ i = m+1, \ldots, s\}. \tag{1}$$

В методе нагруженных функций задача (1) сводится к задачам минимизации некоторых вспомогательных функций на множестве X_0 и к поиску минимального решения (корня) некоторого уравнения. Для учета ограничений типа равенств и неравенств в этом методе также используется идея штрафов, но, в отличие от метода штрафных функций, в нем нет неограниченно возрастающих коэффициентов, аналогичных штрафным коэффициентам. Заметим также, что метод нагруженных функций применим к более широкому классу задач, чем метод модифицированных функций Лагранжа. Введем семейство функций

$$\Phi(x, t) = L\,|\max\{f(x) - t; 0\}|^{p_0} + MP(x), \quad x \in X_0, \tag{2}$$

зависящее от скалярного параметра t, $-\infty < t < \infty$, где $P(x)$ — уже знакомая нам штрафная функция множества X:

$$P(x) = \sum_{i=1}^{m} (\max\{g_i(x); 0\})^{p_i} + \sum_{i=m+1}^{s} |g_i(x)|^{p_i}, \tag{3}$$

величины $p_i \geqslant 1$, $i = 0, \ldots, s$, $L > 0$, $M > 0$ фиксированы и являются параметрами метода. Положим

$$\rho(t) = \inf_{x \in X_0} \Phi(x, t). \tag{4}$$

Поскольку $\Phi(x, t) \geqslant 0$ при всех t и $x \in X_0$, то $\rho(t) \geqslant 0$ при любом t. Предположим, что в задаче (1) $f_* > -\infty$, $X_* \neq \varnothing$. Возьмем произвольную точку $x_* \in X_*$. Тогда $P(x_*) = 0$ и $\Phi(x_*, f_*) = 0$. Следовательно, $\rho(f_*) = 0$, т. е. f_* является корнем уравнения

$$\rho(t) = 0. \tag{5}$$

С другой стороны, $\Phi(x, t) > 0$ при всех $t < f_*$ и $x \in X_0$, и поэтому можно ожидать, что для широкого класса задач будет выполняться неравенство $\rho(t) > 0$ при всех $t < f_*$. Если это в самом деле так, то задача поиска f_* сведется к поиску минимального корня уравнения (5). Такое сведение задачи минимизации привлекательно тем, что для поиска минимального корня уравнения (5) с одной неизвестной могут быть использованы такие широко известные методы решения уравнений, как методы деления отрезка пополам, простой итерации и т. п. [59; 74; 89]. Основная идея метода нагруженных функций описана.

Заметим, что в этом методе могут быть использованы и другие конструкции функции $\Phi(x, t)$, отличные от (2). Например, можно принять

$$\Phi(x, t) = L\,|f(x) - t|^{p_0} + MP(x), \quad x \in X_0, \quad -\infty < t < \infty, \tag{6}$$

где функция $P(x)$ взята из (3), $L > 0$, $M > 0$, $p_i \geqslant 1$. Повторив предыдущие рассуждения для функции $\rho(t)$, определяемой из условий (4), (6), можно

показать, что здесь также $\rho(f_*)=0$, и высказать гипотезу о том, что для широкого класса задач (1) число f_*, по-видимому, будет минимальным корнем уравнения (5).

2. Прежде чем переходить к формулировке условий, при которых высказанная гипотеза в самом деле будет справедлива, рассмотрим примеры. Во всех примерах ограничимся рассмотрением функций $\Phi(x, t)$ из (2) и (6) при $L=M=p_0=\ldots=p_s=1$.

Пример 1. Задача: $f(x)=-x\to\inf$, $x\in X=\{x\in E^1\colon g(x)=x\leqslant 0\}$. Здесь $f_*=0$, $x_*=0$. Функция (2) имеет вид

$$\Phi(x, t)=\max\{-x-t;0\}+\max\{x;0\},\quad x\in X_0=E^1.$$

Если $t\geqslant 0$, то $\Phi(0, t)=0=\inf\limits_{E^1}\Phi(x, t)=\rho(t)$. Если же $t<0$, то $\Phi(x, t)=x$ при $x\geqslant -t$, $\Phi(x, t)=-t$ при $0\leqslant x\leqslant -t$; $\Phi(x, t)=-x-t$ при $x\leqslant 0$ (нарисуйте график функции $\Phi(x, t)$ при различных t). Поэтому $\inf\limits_{E^1}\Phi(x, t)=\rho(t)=-t$ при $t<0$. Таким образом, $\rho(t)=\max\{-t;0\}$. Очевидно, минимальный корень уравнения (5) здесь совпадает с $f_*=0$.

Функция (6) будет иметь вид

$$\Phi(x, t)=|-x-t|+\max\{x;0\},\quad x\in E^1.$$

Если $t\geqslant 0$, то, взяв $x=-t$, получим $\Phi(-t; t)=0=\rho(t)$. Если же $t<0$, то $\Phi(x, t)=2x+t$ при $x\geqslant -t$; $\Phi(x, t)=-x-t$ при $x\leqslant 0$, и, следовательно, $\rho(t)=-t$ при $t<0$. В рассматриваемой задаче функции $\rho(t)$, построенные на основе функций (2) и (6), совпали.

Пример 2. Пусть $f(x)=x$, $X=\{x\in E^1\colon g(x)=x^2-1\leqslant 0\}$. Ясно, что здесь $X=\{x\in E^1\colon -1\leqslant x\leqslant 1\}$, $f_*=-1$, $x_*=-1$. Если согласно (2) принять

$$\Phi(x, t)=\max\{x-t;0\}+\max\{x^2-1;0\},\quad x\in X_0=E^1,$$

то нетрудно показать, что $\rho(t)=\inf\limits_{E^1}\Phi(x, t)=\max\{-t-1;0\}$. Если же за основу взять функцию (6), то

$$\Phi(x,t)=|x-t|+\max\{x^2-1;0\},\quad x\in E^1,\quad \rho(t)=\inf_{E^1}\Phi(x, t)=\max\{|t|-1;0\}.$$

В рассматриваемой задаче функции $\rho(t)$, построенные с помощью функции (2) и (6), оказались разными, но минимальный корень уравнения (5) в обоих случаях совпадает с $f_*=-1$.

Пример 3. Пусть $f(x)=x$, $X=\{x\in E^1\colon g(x)=x^2=0\}$. Тогда $X=\{0\}$, $f_*=0$, $x_*=0$. Для функции (2), $\Phi(x, t)=\max\{x-t;0\}+x^2$, $x\in X_0=E^1$, получим

$$\rho(t)=\begin{cases}0, & t\geqslant 0,\\ t^2, & -1/2\leqslant t<0,\\ -t-1/4, & t<-1/2.\end{cases}$$

Если взять функцию (6), то $\Phi(x, t)=|x-t|+x^2$, $x\in E^1$, и

$$\rho(t)=\begin{cases}t^2, & |t|\leqslant 1/2,\\ |t|-1/4, & |t|>1/2.\end{cases}$$

Здесь также минимальный корень уравнения (5) совпадает с $f_*=0$.

Однако нетрудно привести примеры задач (1), когда минимальный корень уравнения (5) строго меньше f_*.

Пример 4. Пусть $f(x)=x$, $X=\{x\in E^1\colon g(x)=(x^2-1)(x^4+1)^{-1}\leqslant 0\}$. Здесь $X=\{x\in E^1\colon -1\leqslant x\leqslant 1\}$ и, очевидно, $f_*=-1$, $x_*=-1$. Если согласно (2) примем

$$\Phi(x, t)=\max\{x-t;0\}+\max\{g(x);0\},\quad x\in X_0=E^1,$$

то при $t\geqslant -1$ получим $\Phi(-1, t)=0=\inf\limits_{E^1}\Phi(x, t)=\rho(t)$. При $t<-1$, взяв $x_k=-k\leqslant t$, также будем иметь $\lim\limits_{k\to\infty}\Phi(-k, t)=\lim\limits_{k\to\infty}g(-k)=0=\inf\limits_{E^1}\Phi(x, t)=\rho(t)$. Таким образом, в рассматриваемом случае $\rho(t)\equiv 0$ при всех t. Если в качестве минимального решения уравнения (5) здесь взять $t_*=-\infty$, то получим $t_*<f_*=-1$.

Рассмотрим функцию (6)

$$\Phi(x, t)=|x-t|+\max\{g(x);0\},\quad x\in X_0=E^1.$$

Если $|t|\leqslant 1$, то при $x=t$ получим $\Phi(t, t)=0=\rho(t)$. Пусть $|t|>1$. Введем множества $A_1=\left\{x\in E^1\colon |x-t|\leqslant\frac{|t|-1}{2}\right\}$, $A_2=\left\{x\in E^2\colon |x-t|>\frac{|t|-1}{2}\right\}$. Так как $A_1\cup A_2=E^1$, $A_1\cap A_2=\varnothing$, то тогда $\rho(t)=\inf\limits_{E^1}\Phi(x, t)=\min\left\{\inf\limits_{A^1}\Phi(x, t);\ \inf\limits_{A^2}\Phi(x, t)\right\}\geqslant\min\left\{\min\limits_{A^1}g(x);\ (|t|-1)/2\right\}>0$ при всех t, $|t|>1$. На первый взгляд создается впечатление, что здесь $f_*=-1$ — минимальный корень уравнения (5). Однако $0<\rho(t)\leqslant\Phi(t, t)=g(t)$ при $|t|>1$ и $\lim\limits_{t\to-\infty}\rho(t)=\lim\limits_{t\to-\infty}g(t)=0$. Поэтому есть основание считать, что минимальный корень уравнения (5) и в этом случае равен $t_*=-\infty<f_*$.

Любопытно сравнить задачи из примеров 2 и 4. В них функции $f(x)$ и множества X совпадают. Но эти задачи отличаются способом задания множества X. Это различие приводит к тому, что в примере 2 минимальный корень t_* уравнения (5) совпадает с f_*, а в примере 4 получим $t_*<f_*$. Отсюда можно сделать вывод: для выполнения равенства $t_*=f_*$, лежащего в основе метода нагруженных функций, исходные данные задачи (1) должны удовлетворять некоторым дополнительным условиям, они должны быть как-то согласованы. Для формулировки этих условий нам прежде всего нужно уточнить, что понимать под минимальным корнем уравнения (5).

Определение 1. Число t_* назовем *минимальным корнем* уравнения (5), если $\rho(t_*)=0$, $\rho(t)>0$ при всех $t<t_*$ и $\lim\limits_{t\to-\infty}\rho(t)>0$. Если же $\lim\limits_{t\to-\infty}\rho(t)=0$, то примем $t_*=-\infty$. Если $\rho(t)>0$ при всех $t>0$ и $\lim\limits_{t\to-\infty}\rho(t)>0$, то по определению положим $t_*=\infty$.

Чтобы показать, что все указанные в определении 1 возможности в самом деле могут реализоваться, рассмотрим еще несколько примеров.

Пример 5. Пусть

$$f(x)=\begin{cases}x, & x>-2,\\ -k^2, & -(k+1)<x\leqslant -k,\quad k=2, 3,\ldots;\end{cases}\qquad g(x)=\begin{cases}x^2-1, & |x|\leqslant 2,\\ 6/|x|, & |x|>2,\end{cases}\qquad (7)$$

$X=\{x\in E^1=X_0\colon g(x)\leqslant 0\}$. Тогда $f_*=-1$, $x_*=-1$. Рассмотрим функцию (6)

$$\Phi(x,t)=|f(x)-t|+\max\{g(x);0\},\quad x\in E^1.$$

Покажем, что $\rho(-k^2-k)\geqslant k$, $k=2,3,\ldots$ В самом деле, если $x\geqslant -2$, то $\Phi(x,-k^2-k)\geqslant|x+k^2+k|=k^2+k+x\geqslant k^2\geqslant k$ при всех $k=2,3,\ldots$ Если $-(i+1)<x\leqslant -i$, $2\leqslant i\leqslant k$, то $\Phi(x,-k^2-k)\geqslant|-i^2+k^2+k|=k^2-i^2+k\geqslant k$, а если $-(i+1)<x\leqslant -i$, $i\geqslant k+1$, то $\Phi(x,-k^2-k)\geqslant|-i^2+k^2+k|=i^2-(k+1)^2+k+1\geqslant k+1>k$. Таким образом, $\Phi(x,-k^2-k)\geqslant k$ для всех $x\in E^1$, поэтому $\rho(-k^2-k)\geqslant k$, $k=2,3,\ldots$ Следовательно, $\varliminf\limits_{t\to-\infty}\rho(t)=\lim\limits_{k\to\infty}\rho(-k^2-k)=\infty$. С другой стороны, $0\leqslant\rho(-k^2)\leqslant\Phi(-k,-k^2)=g(-k)=6k^{-1}$, $k=2,3,\ldots$, так что $\lim\limits_{t\to-\infty}\rho(t)=\lim\limits_{k\to\infty}\rho(-k^2)=0$. Согласно определению 1 тогда $t_*=-\infty<f_*=-1$.

Остановимся также на функции (2)

$$\Phi(x,t)=\max\{f(x)-t;0\}+\max\{g(x);0\},\quad x\in E^1.$$

Нетрудно видеть, что если $t\geqslant -1$, то $\Phi(-1,t)=0=\rho(t)$. Если же $t<-1$, то при $k\sqrt{-t}$ получим $\Phi(-k,t)=\max\{-k^2-t;0\}+g(-k)=g(-k)\to 0=\rho(t)$. Таким образом, здесь $\rho(t)\equiv 0$ при всех t и согласно определению 1 имеем $t^*=-\infty$.

Пример 6. Переопределим функцию $f(x)$ из (7) в точках $x=k^{-1}$ так: $f(k^{-1})=-k^2$, $k=1,2,\ldots$ Функцию $g(x)$ и множество X оставим такими же, как и примере 5. Повторив прежние рассуждения, нетрудно убедиться, что минимальный корень уравнения (5) здесь будет равным $t_*=f_*$ как при использовании функции (2), так и функции (6). В отличие от примера 5, здесь $f_*=-\infty$, поэтому справедливо $t_*=f_*$.

Любопытно посмотреть, что будет, если множество X из (1) пусто, но X_0 непусто. В этом случае задача (1), конечно, перестает быть содержательной, но тем не менее функции $\Phi(x,t)$, $\rho(t)$ из (2), (4), (6) будут иметь смысл.

Пример 7. Пусть $f(x)\equiv 1$, $X=\{x\in E^1\colon g(x)=e^{-x^2}\leqslant 0\}$. Здесь $X_0=E^1$, $X=\varnothing$. Согласно формуле (2) имеем $\Phi(x,t)=\max\{1-t;0\}+e^{-x^2}$, $x\in E^1$, поэтому $\rho(t)=\max\{1-t;0\}$ и $t_*=1$. Если же воспользуемся функцией (6) $\Phi(x,t)=|1-t|+e^{-x^2}$, то $\rho(t)=|1-t|$ и $t_*=1$.

Если здесь взять $g(x)=e^{-x^2}+1$, то получим $\rho(t)=\max\{1-t;0\}+1$ для функции (2) и $\rho(t)=|1-t|+1$ для функции (6), так что минимальный корень уравнения (5) согласно определению 1 будет равен $t_*=\infty$.

Пример 8. Пусть $f(x)=x$, $X=\{x\in E^1\colon g(x)=e^{-x^2}\leqslant 0\}$. Здесь $X_0=E^1$, $X=\varnothing$. Согласно (2) имеем $\Phi(x,t)=\max\{x-t;0\}+e^{-x^2}$. Так как при $x=-k\leqslant t$ функция $\Phi(-k,t)=e^{-k^2}\to 0$ при $k\to\infty$, то $\rho(t)\equiv 0$ при всех t, и $t_*=-\infty$. В случае функции (6) $\Phi(x,t)=|x-t|+e^{-x^2}=\min\{\inf\limits_{|x-t|\leqslant 1}\Phi(x,t);\inf\limits_{|x-t|>1}\Phi(x,t)\}\geqslant\min\{\inf\limits_{|x-t|\leqslant 1}e^{-x^2};1\}=c(t)>0$ при всех $x\in E^1$, поэтому $\rho(t)>0$ при всех t. Но $0<\rho(t)<\Phi(t,t)\to 0$ при $t\to+\infty$ или $t\to-\infty$, так что $\lim\limits_{t\to+\infty}\rho(t)=\lim\limits_{t\to-\infty}\rho(t)=0$ и $t_*=-\infty$.

Если же здесь взять $g(x)=e^{-x^2}+1$, то $\Phi(x,t)\geqslant 1$, $x\in E^1$ и $\rho(t)\geqslant 1$ при всех t, и поэтому $t_*=+\infty$.

3. Примеры 7, 8 подсказывают, что для того чтобы единообразно охватить возможность, когда в задаче (1) $X=\varnothing$, целесообразно принять

$$f_*=\begin{cases}\inf\limits_X f(x), & X\neq\varnothing,\\ +\infty, & X=\varnothing,\quad X_0\neq\varnothing.\end{cases}\tag{8}$$

Тогда справедлива следующая

Теорема 1. *Пусть функция $\rho(t)$ определена формулой* (4), *где функция $\Phi(x,t)$ взята из* (2) *или* (6). *Пусть t_* — минимальный корень уравнения* (5) *в смысле определения* 1, *а величина f_* определена согласно* (8). *Тогда $t_*\leqslant f_*$.*

Доказательство. Если $X=\varnothing$, то $f_*=+\infty$ и утверждение теоремы тривиально. Поэтому пусть $X\neq\varnothing$. Так как мы условились рассматривать функции, принимающие лишь конечные значения в области своего определения, то $f_*<\infty$. По определению f_*, существует последовательность $\{x_k\}\in X$ такая, что $\lim\limits_{k\to\infty}f(x_k)=f_*\geqslant-\infty$. Если $f_*>-\infty$, то $\lim\limits_{k\to\infty}\Phi(x_k,f_*)=0=\rho(f_*)$ и поэтому $t_*\leqslant f_*$. Если же $f_*=-\infty$, то, взяв $t_k=f(x_k)$, получим $\rho(t_k)=\Phi(x_k,t_k)=0$, $k=1,2,\ldots$ Поскольку $\{t_k\}\to-\infty$, то отсюда следует $\varliminf\limits_{t\to-\infty}\rho(t)=\lim\limits_{k\to\infty}\rho(t_k)=0$, так что $t_*=f_*=-\infty$. Теорема доказана. □

Рассмотренные выше примеры показывают, что для выполнения равенства $t_*=f_*$ важное значение имеет способ задания множества X: ограничения, задающие множество X должны быть как-то согласованы с минимизируемой функцией $f(x)$. Напоминаем, что в § 15 было введено понятие согласованной постановки задачи (1) на X_0 (см. определение 15.2), означающее, что для любой последовательности $\{x_k\}\in X_0$, которая удовлетворяет условиям

$$\lim_{k\to\infty}g_i^+(x_k)=0,\quad i=1,\ldots,s,\tag{9}$$

имеет место соотношение

$$\varliminf_{k\to\infty}f(x_k)\geqslant f_*.\tag{10}$$

Распространим это понятие на случай, когда $X_0\neq\varnothing$, $X=\varnothing$ и согласно (8) $f_*=+\infty$. Здесь следует различать две возможности: $\inf\limits_{X_0}P(x)=0$ и $\inf\limits_{X_0}P(x)>0$. Если $\inf\limits_{X_0}P(x)=0$, то существует хотя бы одна последовательность $\{x_k\}\in X_0$, удовлетворяющая условиям (9), — в этом случае скажем, что задача (1) имеет согласованную постановку на X_0, если для любой последовательности $\{x_k\}\in X_0$, для которой справедливы соотношения (9), имеет место равенство $\lim\limits_{k\to\infty}f(x_k)=+\infty=f_*$. Кстати, это же равенство получается и из (10) при $f_*=+\infty$. Наконец, если $X_0\neq\varnothing$, $X=\varnothing$, $\inf\limits_{X_0}P(x)>0$, то по определению будем считать, что задача (1) имеет согласованную постановку на множестве X_0.

Оказывается, введенное понятие согласованной постановки задачи (1) играет важную роль при выяснении того, будет ли $t_*=f_*$ или $t_*<f_*$.

Теорема 2. *Пусть функция $\rho(t)$ определена формулой* (4), *где функция $\Phi(x,t)$ взята из* (2) *или* (6), *пусть t_* — минимальный корень уравнения* (5), *а величина f_* определена формулой* (8). *Тогда для выпол-*

нения равенства $t_* = f_*$ необходимо и достаточно, чтобы задача (1) *имела согласованную постановку на множестве X_0.*

Доказательство. Необходимость. Пусть $t_* = f_*$. Если $f_* = -\infty$, то постановка задачи (1) согласована, так как $\varliminf\limits_{k\to\infty} f(x_k) \geqslant -\infty = f_*$ для любой последовательности $\{x_k\} \in X_0$. Поэтому пусть $f_* > -\infty$. Возьмем произвольную последовательность $\{x_k\} \in X_0$, удовлетворяющую условиям (9). Согласно определению (3) функции $P(x)$ тогда $\lim\limits_{k\to\infty} P(x_k) = 0$. Отсюда и из неравенств $\Phi(x_k, t) \geqslant \rho(t) > 0$, справедливых для всех $t < t_* = f_*$ и $k = 1, 2, \ldots$, при $k \to \infty$ получим

$$\varliminf_{k\to\infty} \max\{f(x_k) - t; 0\} \geqslant \rho(t) > 0, \quad t < t_*, \tag{11}$$

в случае использования функции (2) и

$$\varliminf_{k\to\infty} |f(x_k) - t| \geqslant \rho(t) > 0, \quad t < t_{**}, \tag{12}$$

в случае использования функции (6).

Покажем, что из (11), (12) следует неравенство $\varliminf\limits_{k\to\infty} f(x_k) \geqslant t_* = f_*$. В самом деле, при выполнении (11) для каждого $t < t_*$ найдется номер $k_0 = k_0(t)$ такой, что $\max\{f(x_k) - t; 0\} \geqslant \rho(t)/2 > 0$ или $f(x_k) - t \geqslant \rho(t)/2 > 0$ для всех $k \geqslant k_0$. Тогда $\varliminf\limits_{k\to\infty} f(x_k) \geqslant t$ при любом $t < t_*$. Устремляя $t \to t_* - 0$, отсюда получим неравенство $\varliminf\limits_{k\to\infty} f(x_k) \geqslant t_* = f_*$.

Рассмотрим случай (12). Пусть $\varliminf\limits_{k\to\infty} f(x_k) = \lim\limits_{r\to\infty} f(x_{k_r}) = a$. Имеются две возможности: либо $a \geqslant t_*$, либо $a < t_*$. Если $a \geqslant t_*$, то требуемое неравенство $\varliminf\limits_{k\to\infty} f(x_k) \geqslant t_* = f_*$ установлено. Остается рассмотреть возможность $a < t_*$. В этом случае величина a не может быть конечной.

Допустим противное: пусть $-\infty < a < t_*$. Тогда при $t = a$ получим

$$\varliminf_{k\to\infty} |f(x_k) - a| = |\varliminf_{r\to\infty} f(x_{k_r}) - a| = 0,$$

что противоречит условию (12). Таким образом, если $a < t_*$, то $a = -\infty$, т. е. $\varliminf\limits_{k\to\infty} f(x_k) = \varliminf\limits_{r\to\infty} f(x_{k_r}) = -\infty$. Тогда, взяв $t_r = f(x_{k_r})$, $r = 1, 2, \ldots$, получим $0 < \rho(t_r) \leqslant \Phi(x_{k_r}, f(x_{k_r})) = MP(x_{k_r}) \to 0$ при $r \to \infty$. Это значит, что $\varliminf\limits_{t\to-\infty} \rho(t) = \varliminf\limits_{r\to\infty} \rho(t_r) = 0$ и согласно определению 1 тогда $t_* = -\infty$. Но по условию $t_* = f_*$, поэтому $\varliminf\limits_{k\to\infty} f(x_k) = f_* = t_* = -\infty$, дело свелось к ранее рассмотренному случаю.

Тем самым установлено, что для любой последовательности $\{x_k\} \in X_0$, удовлетворяющей условиям (9), справедливо неравенство (10). Наконец, если такой последовательности $\{x_k\}$ не существует, т. е. $\inf\limits_{X_0} P(x) > 0$, то задача (1) имеет согласованную постановку по определению. Необходимость доказана.

Достаточность. Пусть задача (1) имеет согласованную постановку на X_0. Покажем, что тогда $t_* = f_*$. Сначала рассмотрим случай, когда $t_* = -\infty$. Это значит, что $\varliminf\limits_{t\to-\infty} \rho(t) = \varliminf\limits_{k\to\infty} \rho(t_k) = 0$, где $\{t_k\} \to -\infty$. По определению $\rho(t_k)$, согласно формуле (4), следует существование точки $x_k \in X_0$

такой, что $\rho(t_k) \leqslant \Phi(x_k, t_k) \leqslant \rho(t_k) + 1/k$, $k = 1, 2, \ldots$ Отсюда при $k \to \infty$ имеем $\lim\limits_{k\to\infty} \Phi(x_k, t_k) = 0$. Это означает, что $\lim\limits_{k\to\infty} P(x_k) = 0$ и $\lim\limits_{k\to\infty} \max\{f(x_k) - t_k; 0\} = 0$ в случае использования функции (2) и $\lim\limits_{k\to\infty} |f(x_k) - t_k| = 0$ — в случае функции (6). Но по построению $\{t_k\} \to -\infty$, поэтому последние два равенства возможны только при $\varliminf\limits_{k\to\infty} f(x_k) = -\infty = t_*$. С другой стороны, из $\{P(x_k)\} \to 0$ и формулы (3) следует выполнение условий (9). В силу (10) тогда $\varliminf\limits_{k\to\infty} f(x_k) \geqslant f_*$. Следовательно, $f_* = t_* = -\infty$.

Пусть теперь $t_* > -\infty$. В силу теоремы 1 тогда $f_* \geqslant t_* > -\infty$. Возьмем произвольное $t < f_*$. По определению $\rho(t)$ существует последовательность $\{x_k\} \in X_0$ такая, что $\lim\limits_{k\to\infty} \Phi(x_k, t) = \rho(t)$. Может случиться, что $\lim\limits_{k\to\infty} P(x_k) = d > 0$. Тогда из $\Phi(x_k, t) \geqslant MP(x_k)$ при $k \to \infty$ следует, что $\rho(t) \geqslant M \lim\limits_{k\to\infty} P(x_k) = Md > 0$. Если же $\lim\limits_{k\to\infty} P(x_k) = 0 = \lim\limits_{r\to\infty} P(x_{k_r})$, то $\lim\limits_{r\to\infty} g_i^+(x_{k_r}) = 0$, $i = 1, \ldots, s$. В силу (10) отсюда имеем $\varliminf\limits_{k\to\infty} f(x_{k_r}) \geqslant f_* > t$. А тогда $\rho(t) = \lim\limits_{r\to\infty} \Phi(x_{k_r}, t) \geqslant L(f_* - t)^{p_0} > 0$ как в случае использования функции (2), так и функции (6). Тем самым показано, что $\rho(t) > 0$ при всех $t < f_*$. Кроме того, в рассматриваемом случае $t_* > -\infty$ по определению 1 имеем $\varliminf\limits_{t\to-\infty} \rho(t) > 0$. Следовательно, $t_* \geqslant f_*$, что в силу теоремы 1 возможно только при $t_* = f_*$. Теорема доказана. □

В § 15 были приведены достаточные условия, гарантирующие согласованную постановку задачи (1) на X_0 (см. теорему 15.2, леммы 15.1, 15.5).

4. Подробнее остановимся на частном случае функции (2), когда

$$\Phi(x, t) = L \max\{f(x) - t; 0\} + MP(x), \quad x \in X_0, \tag{13}$$

где $L > 0$, $M > 0$, а функция $P(x)$ взята из (3) при некоторых $p_i \geqslant 1$, $i = 1, \ldots, s$. Оказывается, функция (13) и соответствующая ей функция $\rho(t)$ обладают рядом полезных свойств, облегчающих поиск минимального корня уравнения (5).

Теорема 3. *Функции $\Phi(x, t)$, $\rho(t)$, определяемые формулами* (13), (4), *монотонно убывают (вообще говоря, не строго) при возрастании t и удовлетворяют неравенствам*

$$|\Phi(x, t) - \Phi(x, \tau)| \leqslant L|t - \tau|, \tag{14}$$

$$|\rho(t) - \rho(\tau)| \leqslant L|t - \tau| \tag{15}$$

при всех $x \in X_0$ и любых t, τ. Если $f_{**} = \inf\limits_{X_0} f(x) > -\infty$, то

$$\Phi(x, t) = -Lt + Lf(x) + MP(x), \quad \rho(t) = -Lt + \inf_{X_0}(Lf(x) + MP(x)) \tag{16}$$

при всех $t \leqslant f_{**}$ — линейные функции по t.

Доказательство. Простым перебором возможных значений функции $\max\{a; b\}$ легко доказываются неравенства

$$\max\{f(x) - t; 0\} \geqslant \max\{f(x) - \tau; 0\}, \quad t \leqslant \tau, \quad x \in X_0,$$

$$|\max\{f(x) - t; 0\} - \max\{f(x) - \tau; 0\}| \leqslant |t - \tau|, \quad x \in X_0.$$

Отсюда следует невозрастание функции $\Phi(x, t)$ по переменной t и неравенство (14). Далее, для любых $t \leqslant \tau$ имеем $\Phi(x, t) \geqslant \Phi(x, \tau) \geqslant \rho(\tau)$ или $\Phi(x, t) \geqslant \rho(\tau)$ при каждом $x \in X_0$. Отсюда, переходя к нижней грани по $x \in X_0$, получим $\rho(t) \geqslant \rho(\tau)$ при всех $t \leqslant \tau$.

Докажем неравенство (15). Зафиксируем произвольные t, τ. По определению нижней грани при каждом $\varepsilon > 0$ существуют точки $x_t, x_\tau \in X_0$ такие, что

$$\rho(t) \leqslant \Phi(x_t, t) \leqslant \rho(t) + \varepsilon, \quad \rho(\tau) \leqslant \Phi(x_\tau, \tau) \leqslant \rho(\tau) + \varepsilon.$$

Тогда, учитывая уже доказанное неравенство (14), имеем $\rho(t) - \rho(\tau) \leqslant \Phi(x_\tau, t) - \Phi(x_\tau, \tau) + \varepsilon \leqslant L|t - \tau| + \varepsilon$, $\rho(t) - \rho(\tau) \geqslant \Phi(x_t, t) - \varepsilon - \Phi(x, \tau) \geqslant -L|t - \tau| - \varepsilon$, т. е. $|\rho(t) - \rho(\tau)| \leqslant L|t - \tau| + \varepsilon$

при любом $\varepsilon > 0$. Отсюда при $\varepsilon \to +0$ получим неравенство (15). Формулы (16) следуют из того, что $f(x) - t \geqslant f_{**} - t \geqslant 0$ при всех $x \in X_0$. Теорема 3 доказана. □

Если задача (1) имеет согласованную постановку на X_0 и $f_* > -\infty$, то, опираясь на теорему 3, можно предложить следующий итерационный метод определения f_*. Сначала выберем t_0 так, чтобы $\rho(t_0) > 0$ (например, если $f_{**} = \inf\limits_{X_0} f(x) > -\infty$, то можно взять любую точку $t_0 \leqslant f_{**}$).

Следующие приближения определим по формулам

$$t_{k+1} = t_k + \rho(t_k)/L, \quad k = 0, 1, \ldots \tag{17}$$

Теорема 4. *Пусть функция $\rho(t) \geqslant 0$ при всех t, $-\infty < t < +\infty$, удовлетворяет условию* (15), *пусть t_* — минимальный корень уравнения* (5) *в смысле определения* 1, $t_* > -\infty$. *Тогда при любом выборе начального приближения t_0, $-\infty < t_0 < t_*$, последовательность $\{t_k\}$, определяемая условиями* (17), *сходится к t_*.*

Доказательство. Так как $\rho(t) \geqslant 0$, то из (17) следует, что последовательность $\{t_k\}$ монотонно возрастает и поэтому существует $\lim\limits_{k\to\infty} t_k = a \leqslant \infty$. Покажем, что $a = t_*$. По условию $t_0 < t_*$. Допустим, что при некотором $k \geqslant 0$ оказалось $t_k < t_*$. Тогда $\rho(t) > 0$ при всех $t \leqslant t_k$. Возьмем произвольное t, $t_k \leqslant t < t_{k+1}$. С учетом условий (15), (17) имеем

$$\rho(t) = \rho(t_k) + [\rho(t) - \rho(t_k)] \geqslant \rho(t_k) - L(t - t_k) > \rho(t_k) - L(t_{k+1} - t_k) = 0, \quad t_k \leqslant t < t_{k+1}.$$

Это значит, что $\rho(t) > 0$ при всех $t < t_{k+1}$, т. е $t_{k+1} \leqslant t_*$. Может случиться, что $\rho(t_{k+1}) = 0$. Тогда $t_{k+1} = t_*$ — в этом случае итерации (17) заканчиваются. Если $\rho(t_{k+1}) > 0$, то $t_{k+1} < t_*$ и итерации продолжаются дальше.

Таким образом, имеются две возможности. Либо процесс (17) закончится тем, что $\rho(t_0) > 0, \ldots, \rho(t_{k-1}) > 0$, $\rho(t_k) = 0$ — тогда $t_k = t_* = a$, утверждение теоремы верно. Либо $\rho(t_k) > 0$, $t_k < t_*$, $\rho(t) > 0$ при $t < t_k$ для всех $k = 0, 1, \ldots$ — в этом случае $\lim\limits_{k\to\infty} t_k = a \leqslant t$ и $\rho(t) > 0$ при всех $t < a$. Покажем, что $a = t_*$. Если последовательность $\{t_k\}$ неограничена сверху, то $a = \infty = t_*$. Если же $t_k < a < \infty$, $k = 0, 1, \ldots$, то, учитывая непрерывность функции $\rho(t)$, из (17) при $k \to \infty$ получим $a = a + \rho(a)/L$ или $\rho(a) = 0$. Это значит, что $t_* = a$ при $a < \infty$. Теорема доказана. □

Заметим, что на каждом шаге метода (17) нужно вычислить одно значение функции $\rho(t)$, и для этого в свою очередь нужно решить задачу минимизации

$$\Phi(x, t) \to \inf; \quad x \in X_0. \tag{18}$$

Поскольку функция (13), вообще говоря, не является гладкой, то это обстоятельство может вызвать некоторые трудности при решении задачи (18). Однако имеющиеся методы решения негладких задач минимизации (см., например, [264; 265; 361; 386; 396; 426; 572; 586; 718; 769; 777]) позволяют надеяться на то, что вычисление приближенного значения $\rho(t)$ не окажется слишком трудным.

При изложении метода (17) предполагалось, что величины $\rho(t_k)$ известны точно. Однако задача (18) на практике, как правило, будет решаться приближенно, и точное значение $\rho(t_k)$ удастся вычислить лишь в редких случаях. Поэтому желательно обобщить итерационный процесс (17) на случай, когда значения функции известны неточно. Опишем одно из возможных таких обобщений [143].

Предположим, что вместо точных значений функции $\rho(t)$ известны лишь некоторые приближения $\rho_\nu(t)$, $\nu = 1, 2, \ldots$, удовлетворяющие условиям

$$\rho_\nu(t) \geqslant 0, \quad |\rho_\nu(t) - \rho(t)| \leqslant \gamma_\nu, \quad \nu = 1, 2, \ldots; \quad \lim_{\nu\to\infty} \gamma_\nu = 0. \tag{19}$$

Пусть t_0 — начальное приближение, $t_0 < t_*$. Пусть $(\nu-1)$-е приближение $t_{\nu-1}$ при некотором $\nu \geqslant 1$ уже известно. Для определения следующего приближения t_ν рассмотрим итерационный процесс

$$t_{\nu k+1} = t_{\nu k} + \rho_\nu(t_{\nu k})/L, \quad k = 0, 1, \ldots; \quad t_{\nu 0} = t_0, \tag{20}$$

аналогичный процессу (17). Поскольку функция $\rho_\nu(t)$ может не обращаться в нуль ни в одной точке даже в том случае, когда уравнение (5) имеет конечный минимальный корень (так будет, например, если $\rho_\nu(t) = \rho(t) + \gamma_\nu \geqslant \gamma_\nu > 0$), то процесс (20) следует прекращать не по критерию $\rho_\nu(t_{\nu k}) = 0$, как было выше в (17), а по условию вида $\rho_\nu(t_{\nu k}) < \theta_\nu$, где величина $\theta_\nu > 0$ стремится к нулю при $\nu \to \infty$ и как-то согласована с погрешностью γ_ν. Предположим, что такая последовательность $\{\theta_\nu\}$ уже задана (условия согласования $\{\theta_\nu\}$ и $\{\gamma_\nu\}$ будут обсуждаться ниже). Тогда имеются две возможности:

1) либо найдется номер $k = k_\nu \geqslant 0$ такой, что

$$\rho_\nu(t_{\nu k}) > \theta_\nu, \quad k = 0, \ldots, k_\nu - 1, \quad \rho_\nu(t_{\nu k_\nu}) \leqslant \theta_\nu; \tag{21}$$

в этом случае процесс (20) заканчивается и полагаем

$$t_\nu = t_{\nu k_\nu}; \tag{22}$$

2) либо

$$\rho_\nu(t_{\nu k}) > \theta_\nu \quad \text{при всех} \quad k = 0, 1, \ldots \tag{23}$$

Тогда, как будет показано ниже, при выполнении условий (15), (19) и согласованном изменении величин θ_ν и γ_ν, будет справедливо равенство

$$t_* = \infty. \tag{24}$$

Метод поиска минимального корня уравнения (5) при условиях (15), (19) описан.

Теорема 5. *Пусть функция $\rho(t)$ неотрицательная при всех t, не возрастает, удовлетворяет условию* (15), *а $t_* > -\infty$ — минимальный корень уравнения* (5) *в смысле определения* 1. *Кроме того, пусть функция $\{\rho_\nu(t)\}$ удовлетворяет условиям* (19) *и*

$$\theta_\nu \geqslant \gamma_\nu, \quad \nu = 1, 2, \ldots \tag{25}$$

Тогда последовательность $\{t_\nu\}$, определяемая методом (20)–(24), *сходится к t_* при любом выборе начального приближения t_0, $-\infty < t_0 < t_*$. При этом, если $t_* < \infty$, то итерации* (20) *при каждом $\nu \geqslant 1$ будут заканчиваться за конечное число шагов выполнением условий* (21); *случай* (23) *возможен лишь при $t_* = \infty$.*

Доказательство. Сначала рассмотрим случай $t_* < \infty$. Тогда при каждом фиксированном $\nu \geqslant 1$ имеются две возможности:

1) $t_{\nu k} \leqslant t_* < \infty$ при всех $k = 0, 1, \ldots$ В силу монотонности $\{t_{\nu k}\}$ тогда существует $\lim\limits_{k\to\infty} t_{\nu k} \leqslant t_* < \infty$. Переходя в (20) к пределу при $k \to \infty$, получим $\lim\limits_{k\to\infty} \rho_\nu(t_{\nu k}) = 0$. Это значит, что за конечное число итераций процесс (20) закончится выполнением условий (21).

2) Найдется номер $l \geqslant 0$ такой, что $t_{\nu l} \leqslant t_* < t_{\nu l+1}$. Тогда с учетом соотношений (15), (19), (20) получим

$$t_{\nu l+1} = t_{\nu l} + \rho_\nu(t_{\nu l})/L = t_{\nu l} + [\rho_\nu(t_{\nu l}) - \rho(t_{\nu l})]/L + \\ + [\rho(t_{\nu l}) - \rho(t_*)]/L \leqslant t_{\nu l} + \gamma_\nu/L + t_* - t_{\nu l} = t_* + \gamma_\nu/L. \tag{26}$$

Далее, в силу монотонности $\rho(t)$ имеем $\rho(t) \equiv 0$ при $t \geqslant t_*$, поэтому $\rho(t_{\nu l+1}) = 0$. Отсюда и из условий (19), (25) следует

$$\rho_\nu(t_{\nu l+1}) = \rho_\nu(t_{\nu l+1}) - \rho(t_{\nu l+1}) \leqslant \gamma_\nu \leqslant \theta_\nu. \tag{27}$$

Это значит, что условия (21) выполнятся при некотором $k_\nu \leqslant l + 1$.

Объединяя обе рассмотренные возможности, заключаем, что при $t_* < \infty$ процесс (20) при каждом $\nu \geqslant 1$ заканчивается за конечное число шагов k_ν выполнением условий (21), причем в силу (26)

$$t_\nu = t_{\nu k_\nu} \leqslant t_* + \gamma_\nu L^{-1}, \quad \nu = 1, 2, \ldots \tag{28}$$

Покажем, что $\lim\limits_{\nu\to\infty} t_\nu = t_*$. Из (28) имеем $\overline{\lim\limits_{\nu\to\infty}} t_\nu \leqslant t_*$. Пусть $\lim\limits_{\nu\to\infty} t_\nu = \lim\limits_{r\to\infty} t_{\nu_r} = a$. Тогда с учетом условий (15), (19), (21) получим

$$0 \leqslant \rho(a) \leqslant [\rho(a) - \rho(t_{\nu_r})] + [\rho(t_{\nu_r}) - \rho_{\nu_r}(t_{\nu_r})] + \rho_{\nu_r}(t_{\nu_r}) \leqslant L|a - t_{\nu_r}| + \gamma_{\nu_r} + \theta_{\nu_r} \to 0$$

при $r \to \infty$, т. е. $\rho(a) = 0$. Но t_* — минимальный корень уравнения (5), поэтому $a \geqslant t_*$. Следовательно, $\lim\limits_{\nu\to\infty} t_\nu = \lim\limits_{r\to\infty} t_{\nu_r} > t_* \geqslant \overline{\lim\limits_{\nu\to\infty}} t_\nu$. Это значит, что $\lim\limits_{\nu\to\infty} t_\nu = t_*$. Случай $t_* < \infty$ полностью рассмотрен.

Пусть теперь $t_* = \infty$ и пусть процесс (20) при каждом $\nu \geqslant 1$ заканчивается выполнением условий (21). Покажем, что тогда $\{t_\nu\} \to \infty$. Возьмем произвольное число $T > 0$. Согласно определению 1, если $t_* = \infty$, то $\lim\limits_{t\to-\infty} \rho(t) > 0$ и $\rho(t) > 0$ при всех t. Отсюда и из непрерывности функции $\rho(t)$ следует, что $\inf\limits_{t\leqslant T} \rho(t) = \rho_T > 0$. Так как $\{\theta_\nu\} \to 0$, $\{\gamma_\nu\} \to 0$, то найдется номер $\nu_0 = \nu_0(T)$ такой, что $\theta_\nu + \gamma_\nu < \rho_T$ при всех $\nu \geqslant \nu_0$. Тогда $\rho_\nu(t) \geqslant \rho(t) - \gamma_\nu \geqslant \rho_T - \gamma_\nu > \theta_\nu$ для всех $t \leqslant T$ и $\nu \geqslant \nu_0$. Это значит, что условия (21) не могут выполняться при $t_{\nu k} \leqslant T$, если $\nu \geqslant \nu_0$. Тогда согласно (21), (22) $t_\nu > T$ для всех $\nu \geqslant \nu_0 = \nu_0(T)$, что означает выполнение равенства $\lim\limits_{\nu\to\infty} t_\nu = \infty$.

Остается рассмотреть случай, когда при некотором $\nu \geqslant 1$ выполняются условия (23). Выше было установлено, что при $t_* < \infty$ процесс (20) при всех $\nu \geqslant 1$ закончится выполнением условий (21). Следовательно, если при каком-либо $\nu \geqslant 1$ реализуются условия (23), то $t_* = \infty$. Теорема 5 доказана. □

Заметим, что условие (25) в теореме 5 существенно: его нарушение может привести к тому, что метод (20)–(24) не будет сходиться.

Пример 9. Пусть $f(x) = x \to \inf$; $x \in X = X_0 = \{x \in E^1 : x \geqslant 0\}$ — это частный случай задачи (1), когда $g_i(x) \equiv 0$ (или $s = 0$). Тогда $\Phi(x, t) = \max\{x - t; 0\}$, $x \geqslant 0$ и $\rho(t) = \inf\limits_{x \geqslant 0} \Phi(x, t) = \max\{-t; 0\}$ (ср. с примером 1). Здесь $t_* = f_* = 0$, $x_* = 0$. Если $\rho_\nu(t) = \max\{-t; 0\} + \gamma_\nu$ и $\theta_\nu < \gamma_\nu$, то $\rho_\nu(t) \geqslant \gamma_\nu > \theta_\nu$ при всех t. Поэтому в методе (20)–(24) реализуется случай (23), и искомый минимальный корень $t_* = 0$ уравнения (5) в рассматриваемом случае не будет найден. Причина этого явления — нарушение условия (25).

Заметим также, что на практике вместо полуоси $t_0 \leqslant t < \infty$ часто приходится работать на каком-то отрезке $t_0 \leqslant t \leqslant T$, где величина T ограничена, например, разрядной сеткой ЭВМ. В этом случае метод (20)–(24) требует модификации. А именно, итерационный процесс (20) при каждом $\nu \geqslant 1$ здесь будет заканчиваться определением номера $k_\nu \geqslant 0$ такого, что будет выполнено одно из двух следующих условий:

$$\rho_\nu(t_{\nu k}) > \theta_\nu, \quad k = 0, \ldots, k_\nu - 1; \quad \rho_\nu(t_{\nu k_\nu}) \leqslant \theta_\nu, \quad t_{\nu k_\nu} \leqslant T, \tag{29}$$

или

$$\rho_\nu(t_{\nu k}) > \theta_\nu, \quad k = 0, \ldots, k_\nu - 1; \quad t_{\nu k_\nu - 1} \leqslant T \leqslant t_{\nu k_\nu}. \tag{30}$$

В качестве ν-го приближения t_ν будем брать

$$t_\nu = \min\{t_{\nu k_\nu}; T\}, \quad \nu = 1, 2, \ldots \tag{31}$$

Если выполнены все условия теоремы 5, то, немного видоизменив доказательство этой теоремы, нетрудно установить, что при $t_0 < t_* < T$ для достаточно больших номеров $\nu > \nu_0$ процесс (20) будет заканчиваться выполнением условий (29) и оценки (28), а последовательность $\{t_\nu\}$, определяемая методом (20), (29)–(31), сходится к числу $\min\{t_*; T\}$. Таким образом, метод (20), (29)–(31) позволяет определить, принадлежит ли t_* отрезку $[t_0, T]$, и в случае $t_* \in [t_0, T]$ позволяет найти t_* с нужной точностью.

Для определения t_* при условиях (19) может быть также использован метод деления отрезка пополам.

5. Кратко остановимся на случае, когда функция $\rho(t)$ из (4) определяется с помощью функции

$$\Phi(x, t) = L|f(x) - t| + MP(x), \quad x \in X_0, \tag{32}$$

где $L > 0$, $M > 0$, функция $P(x)$ взята из (3) при некоторых $p_i \geqslant 1$, $i = 1, \ldots, s$.

Поскольку $||f(x) - t| - |f(x) - \tau|| \leqslant |t - \tau|$, то, рассуждая так же, как при доказательстве теоремы 3, убеждаемся, что функции $\Phi(x, t)$, $\rho(t)$ из (4), (32) удовлетворяют условиям (14), (15). Это значит, что для поиска минимального корня уравнения (5) и в этом случае может быть применен описанный выше метод (20)–(24) или его модификация (20), (29)–(31). Только условие (25) здесь нужно заменить условием

$$\theta_\nu \geqslant 2\gamma_\nu, \quad \nu = 1, 2, \ldots \tag{33}$$

Такая замена связана с тем, что функция (32), в отличие от (13), а также соответствующая ей функция $\rho(t)$, вообще говоря, не будут монотонными (см. примеры 1–8). Справедлива

Теорема 6. *Пусть функция $\rho(t)$ неотрицательна при всех t, удовлетворяет условию* (15), *а $t_* > -\infty$ — минимальный корень уравнения* (5) *в смысле определения* 1. *Кроме того, пусть функции $\{\rho_\nu(t)\}$ удовлетворяют условиям* (19), *а последовательности $\{\theta_\nu\}$, $\{\gamma_\nu\}$ — условию* (33). *Тогда последовательность $\{t_\nu\}$, определяемая методом* (20)–(24), *сходится к t_* при любом выборе t_0, $-\infty < t_0 < t_*$. При этом, если $t_* < \infty$, то итерации* (20) *при каждом $\nu \geqslant 1$ будут заканчиваться за конечное число шагов выполнением условий* (21); *случай* (23) *возможен лишь при $t_* = \infty$.*

Доказательство проводится дословно так же, как доказательство теоремы 5. Нужно лишь неравенство (27), полученное в предположении монотонности $\rho(t)$ и условия (25), заменить следующим неравенством, вытекающим из условий (15), (19), (26), (33):

$$\rho_\nu(t_{\nu, l+1}) = [\rho_\nu(t_{\nu, l+1}) - \rho(t_{\nu, l+1})] + [\rho(t_{\nu, l+1}) - \rho(t_*)] \leqslant \gamma_\nu + L(t_{\nu, l+1} - t_*) \leqslant 2\gamma_\nu \leqslant \theta_\nu. \qquad \square$$

Нетрудно также показать, что при выполнении условий теоремы 6 последовательность $\{t_\nu\}$, полученная методом (20), (29)–(31), сходится к $\min\{t_*; T\}$.

Приведем пример, который показывает, что условие (33) в общем случае не может быть ослаблено.

Пример 10. Пусть $f(x) \equiv 0$, X — произвольное непустое множество вида (1). Тогда $\Phi(x, t) = |t| + MP(x)$, $x \in X_0$ и $\rho(t) = |t|$; $t_* = 0 = f_*$. Пусть $\rho_\nu(t) = |t| + \gamma_\nu$, $\nu = 1, 2, \ldots$ Предположим, что условие (33) нарушено, т. е. $\theta_\nu < 2\gamma_\nu$. Возьмем начальное приближение $t_0 = t_{\nu 0} < \min\{\gamma_\nu - \theta_\nu; 0\}$. Тогда $\rho_\nu(t_{\nu 0}) = |t_{\nu 0}| + \gamma_\nu = -t_{\nu 0} + \gamma_\nu > \theta_\nu$. Далее, $t_{\nu 1} = t_{\nu 0} + \rho_\nu(t_{\nu 0}) = \gamma_\nu > 0$, и снова $\rho_\nu(t_{\nu 1}) = \rho_\nu(\gamma_\nu) = 2\gamma_\nu > \theta_\nu$. Отсюда с учетом монотонности $\rho_\nu(t)$ при $t \geqslant 0$ имеем $\rho_\nu(t) \geqslant \rho_\nu(t_{\nu 1}) > \theta_\nu$ для всех $t \geqslant t_{\nu 1}$. Это значит, что $\rho_\nu(t_{\nu k}) > \theta_\nu$ для всех $k = 0, 1, \ldots$ — реализовался случай (23), и искомый минимальный корень $t_* = 0$ уравнения (5) здесь не будет найден.

Полезно заметить, что при описании методов (17), (20)–(24) и (20), (29)–(31), а также при формулировке и доказательстве теорем 4–6 никак не использовался тот факт, что функция $\rho(t)$ получена из (4) и как-то связана с задачей (1), с методом нагруженных функций. Это значит, что описанные методы могут быть использованы для поиска минимального корня уравнения (5) для любой неотрицательной функции $\rho(t)$, удовлетворяющей условию (15) и приближенно заданной посредством условий (19).

По поводу метода нагруженных функций см., например, [143; 154; 286; 608; 670].

Упражнения

1. Найти минимальное решение уравнения (5), где функции $\Phi(x, t)$, $\rho(t)$ определяются из (2), (3), (6), (13) или (32), и проверить условие $t_* = f_*$ для задачи: $f(x) \to \inf$; $x \in X = \{x \in E^1 : x \in X_0, g(x) \leqslant 0\}$, где:

а) $f(x) = \operatorname{arctg} x$, $g(x) = x^2 - 4$, $g(x = (x^2 - 4)(x^4 + 1)^{-1}$, $g(x) = x$, $g(x) = x(x^2 + 1)^{-1}$, $g(x) = x^2$, $X_0 = \{x \in E^1 : x \geqslant 1\}$, $X_0 = \{x \in E^1 : x \geqslant 0\}$, $X_0 = E^1$;

б) $f(x) = x \sin x$, $g(x) = x^2 - 1$; $g(x) = (x^2 - 1)(x^4 + 1)^{-1}$, $g(x) = (x^2 - 1)e^{-x^2}$; $X_0 = \{x \in E^1 : x \geqslant 0\}$;

в) $f(x)$ — произвольная функция, $X = X_0 = E^1$;

г) $f(x) = 1$ при $x \leqslant 1$, $f(x) = x^{-1}$ при $x > 1$; $g(x) = x - 1$, $g(x) = (x - 2)(x^2 + 1)^{-1}$, $g(x) = e^{-x^2}$, $X_0 = \{x \in E^1 : x \geqslant 0\}$ или $X_0 = E^1$;

д) $f(x) \equiv 1$, $g(x) = x$, $g(x) = x^2 + 1$, $g(x) = e^{-x^2}(x^2 + 1)$, $g(x) = x^2 e^{-x^2}$; $X_0 = E^1$, $X_0 = \{x \in E^1 : x \geqslant 0\}$.

2. Найти функцию $\rho(t)$ для задачи $f(x) \to \inf$; $x \in X = \{x \in E^1 = X_0 : g(x) = |x| - 1 \leqslant 0\}$, беря за основу функции $\Phi(x, t)$ из (13) и (32), сравнить результаты с примером 2.

3. Показать, что если функция $\rho(t)$ построена с помощью функций (2) или (6) при $p_0 > 1$, то условия (14), (15), вообще говоря, не будут иметь места (ни с какой константой L). Указание: рассмотреть задачу (1) при $f(x) \equiv 1$, $X = X_0$.

4. Указать такой способ задания множества X из примера 5, чтобы задача имела согласованную постановку на $X_0 = E^1$. Рассмотреть возможности $g(x) = |x| - 1$, $g(x) = x^2 - 1$, $g(x) = e^{-x^2}(x^2 - 1)$ (воспользуйтесь теоремой 2).

5. Выяснить геометрический смысл методов (17), (20)–(24) и (20), (29)–(31), а также геометрический смысл условий (25), (33).

6. Проверить, будут ли функции $\Phi(x, t)$ из (2), (6), (13), (32), а также соответствующая функция $\rho(t)$ из (4) выпуклы, если исходная задача (1) выпукла.

7. Пусть в задаче (1) $f_* > -\infty$, $X_* \neq \varnothing$, и эта задача имеет согласованную постановку на множестве X_0. Пусть последовательность $\{t_k\}$ построена методом (17), а последовательность $\{x_k\}$ определена условиями: $x_k \in X_0$, $\rho(t_k) = \Phi(x_k, t_k)$, $k = 0, 1, \ldots$ Можно ли ожидать, что $\{x_k\} \to X_*$? Приведите примеры.

8. Пусть функция Лагранжа задачи (1) имеет седловую точку $(x_*, \lambda^*) \in X_0 \times \Lambda_0$. Показать, что эта же точка (x_*, λ^*) является седловой точкой функции Лагранжа для задачи

$$\max\{f(x) - t; 0\} \to \inf; \quad x \in X$$

при всех $t \leqslant f_*$. Выяснить связь между множествами точек минимума последней задачи и задачи (1).

9. Для задачи (1) ввести функцию

$$\Phi_1(x, t) = L \max\{f(x); t\} + MP(x), \quad x \in X_0,$$

где $P(x)$ взята из (3), $L>0$, $M>0$, и положить $\rho_1(t)=\inf\limits_{x\in X_0}\Phi_1(x,t)$. Показать, что $f_*=\rho_1(f_*)$. Можно ли утверждать, что f_* будет минимальным корнем уравнения $\rho_1(t)-t=0$? Пользуясь равенством $\max\{f(x),t\}=\max\{f(x)-t;0\}+t$, установить связь между функциями $\Phi_1(x,t)$, $\rho_1(t)$ и функциями $\Phi(x,t)$, $\rho(t)$ из (2), (13).

10. Для задачи $f(x)\to\inf$; $x\in X=\{x\in X_0, g_1(x)\leqslant 0,\ldots,g_m(x)\leqslant 0\}$ ввести функцию $G(w,x)=\max\{f(w)-f(x);g_1(w),\ldots,g_m(w)\}$ или $G(w,x)=(f(w)-f(x))\ g_1(w)\ldots g_m(w)$ переменных x, $w\in X_0$ и рассмотреть итерационный процесс $G(x_{k+1},x_k)=\inf\limits_{w\in X_0}G(w,x_k)$; исследовать его сходимость (метод центров, [319; 345; 613]).

§ 19. О методе случайного поиска

Наряду с описанными выше методами минимизации функций переменных существует большая группа методов поиска минимума, объединенных под названием метода случайного поиска. Метод случайного поиска, в отличие от ранее рассмотренных методов, характеризуется намеренным введением элемента случайности в алгоритм поиска. Многие варианты метода случайного поиска сводятся к построению последовательности $\{x_k\}$ по правилу:

$$x_{k+1}=x_k+\alpha_k\xi,\quad k=0,1,\ldots, \tag{1}$$

где α_k — некоторая положительная величина, $\xi=(\xi^1,\ldots,\xi^n)$ — какая-либо реализация n-мерной случайной величины ξ с известным законом распределения. Например, координаты ξ^i случайного вектора ξ могут представлять независимые случайные величины, распределенные равномерно на отрезке $[-1,1]$. Как видим, метод случайного поиска минимума функции n переменных предполагает наличие датчика (или генератора) случайных чисел, обращаясь к которому, в любой нужный момент можно получить какую-либо реализацию n-мерного случайного вектора ξ с заданным законом распределения. Такие датчики, оформленные в виде стандартных программ, имеются в библиотеках подпрограмм на ЭВМ.

1. Приведем несколько вариантов метода случайного поиска минимума функции $f(x)$ на множестве $X\subseteq E^n$, предполагая, что k-е приближение $x_k\in X$; $k\geqslant 0$, уже известно.

а) *Алгоритм с возвратом при неудачном шаге.* Смысл этого алгоритма заключается в следующем. С помощью датчика случайного вектора получают некоторую его реализацию ξ и в пространстве E^n определяют точку $v_k=x_k+\alpha\xi$, $\alpha=\text{const}>0$. Если $v_k\in X$ и $f(v_k)<f(x_k)$, то сделанный шаг считается удачным, и в этом случае полагается $x_{k+1}=v_k$. Если $v_k\in X$, но $f(v_k)\geqslant f(x_k)$, или же $v_k\notin X$, то сделанный шаг считается неудачным и полагается $x_{k+1}=x_k$.

Если окажется, что $x_k=x_{k+1}=\ldots=x_{k+N}$ для достаточно больших N, то точка x_k принимается в качестве приближения искомой точки минимума.

б) *Алгоритм наилучшей пробы.* Берутся какие-либо s реализаций $\xi_1,\ldots,\xi_s$ случайного вектора ξ и вычисляются значения функции $f(x)$ в тех точках $x=x_k+\alpha\xi$, $i=1,\ldots,s$, которые принадлежат множеству X. Затем полагается $x_{k+1}=x_k+\alpha\xi_{i_0}$, где индекс i_0 определяется условием

$$f(x_k+\alpha\xi_{i_0})=\min_{\substack{x_k+\alpha\xi_i\in X\\ 1\leqslant i\leqslant s}}f(x_k+\alpha\xi_i).$$

Величины $s>1$ и $\alpha=\text{const}>0$ являются параметрами алгоритма.

в) *Алгоритм статистического градиента.* Берутся какие-либо s реализаций $\xi_1,\ldots,\xi_s$ случайного вектора ξ и вычисляются разности $\Delta f_{ki}=f(x_k+\gamma\xi_i)-f(x_k)$ для всех $x_k+\gamma\xi_i\in X$. Затем полагают $p_k=\frac{1}{\gamma}\sum\limits_i\xi_i\Delta f_{ki}$, где сумма берется по всем тем i, $1\leqslant i\leqslant s$, для которых $x_k+\gamma\xi_i\in X$. Если $x_k+\alpha p_k\in X$, то принимается $x_{k+1}=x_k+\alpha p_k$. Если же $x_k+\alpha p_k\notin X$, то повторяют описанный процесс с новым набором из s реализаций случайного вектора ξ. Величины $s>1$, $\alpha>0$, $\gamma>0$ являются параметрами алгоритма. Вектор p_k называют *статистическим градиентом*. Если $X\equiv E^n$, $s=n$, и векторы ξ_i являются неслучайными и совпадают с соответствующими единичными векторами $e_i=(0,\ldots,0,1,\ldots,0)$, $i=1,\ldots,n$, то описанный алгоритм, как нетрудно видеть, превращается в разностный аналог градиентного метода.

2. В описанных вариантах а)–в) метода случайного поиска предполагается, что закон распределения случайного вектора ξ не зависит от номера итерации. Такой поиск называют *случайным поиском без обучения*. Алгоритмы случайного поиска без обучения не обладают «способностью» анализировать результаты предыдущих итераций и выделять направления, более перспективные в смысле убывания минимизируемой функции, и сходятся, вообще говоря, медленно.

Между тем ясно, что от метода случайного поиска можно ожидать большей эффективности, если на каждой итерации учитывать накопленный опыт поиска минимума на предыдущих итерациях и перестраивать вероятностные свойства поиска так, чтобы направления ξ более перспективные в смысле убывания функции становились более вероятными. Иначе говоря, желательно иметь алгоритмы случайного поиска, которые обладают способностью к самообучению и самоусовершенствованию в процессе поиска минимума в зависимости от конкретных особенностей минимизируемой функции. Такой поиск называют *случайным поиском с обучением*. Обучение алгоритма осуществляют посредством целенаправленного изменения закона распределения случайного вектора ξ в зависимости от номера итерации и результатов предыдущих итераций таким образом, чтобы «хорошие» направления, по которым функция убывает, стали более вероятными, а другие направления — менее вероятными. Таким образом, на различных этапах метода случайного поиска с обучением приходится иметь дело с реализациями случайных векторов ξ с различными законами распределения. Имея в виду это обстоятельство, итерационный процесс (1) удобнее записать в виде

$$x_{k+1}=x_k+\alpha_k\xi_k,\quad k=0,1,\ldots, \tag{2}$$

подчеркнув зависимость случайного вектора ξ от k.

В начале поиска закон распределения случайного вектора $\xi=\xi_0$ выбирают с учетом имеющейся априорной информации о минимизируемой функции. Если такая информация отсутствует, то поиск обычно начинают со случайного вектора $\xi_0=(\xi_0^1,\ldots,\xi_0^n)$, компоненты ξ_0^i, $i=1,\ldots,n$, которого представляют собой независимые случайные величины, распределенные равномерно на отрезке $[-1,1]$.

Для обучения алгоритма в процессе поиска часто берут семейство случайных векторов $\xi=\xi(w)$, зависящих от параметров $w=(w^1,\ldots,w^n)$, и при переходе от k-й итерации к $(k+1)$-й итерации имеющиеся значения параметров w_k заменяют новыми значениями w_{k+1} с учетом результатов предыдущего поиска.

Приведем два варианта метода случайного поиска с обучением для минимизации функции $f(x)$ на всем пространстве.

а) *Алгоритм покоординатного обучения.* Пусть имеется семейство случайных векторов $\xi = \xi(w) = (\xi^1, \ldots, \xi^n)$, каждая координата ξ^i которых принимает два значения: $\xi^i = 1$ с вероятностью p^i и $\xi^i = -1$ с вероятностью $1 - p^i$, где вероятности p^i зависят от параметра w^i следующим образом:

$$p^i = \begin{cases} 0, & w^i < -1, \\ \frac{1}{2}(1 + w^i), & |w^i| \leqslant 1, \\ 1, & w^i > 1, \quad i = 1, \ldots, n. \end{cases} \tag{3}$$

Пусть начальное приближение x_0 уже выбрано. Тогда для определения следующего приближения x_1 в формуле (2) при $k = 0$ берется какая-либо реализация случайного вектора $\xi_0 = \xi(0)$, соответствующего значению параметров $w = w_0 = (0, 0, \ldots, 0)$. Приближение x_2 определяется по формуле (2) при $k = 1$ с помощью случайного вектора $\xi_1 = \xi(0)$. Пусть известны приближения $x_0, x_1, \ldots, x_k$ и значения параметров $w_{k-1} = (w^1_{k-1}, \ldots, w^n_{k-1})$ при некотором $k \geqslant 1$. Тогда полагаем

$$w^i_k = \beta w^i_{k-1} - \delta \operatorname{sign}[(f(x_{k-1}) - f(x_{k-2}))(x^i_{k-1} - x^i_{k-2})], \quad i = 1, \ldots, n, \quad k = 2, 3, \ldots, \tag{4}$$

где величина $\beta \geqslant 0$ называется *параметром забывания*, $\delta \geqslant 0$ — *параметром интенсивности обучения*, $\beta + \delta > 0$. При определении следующего приближения x_{k+1} в формуле (2) берем какую-либо реализацию случайного вектора $\xi_k = \xi(w_k)$, $w_k = (w^1_k, \ldots, w^n_k)$.

Из (3), (4) видно, что если переход от точки x_{k-2} к x_{k-1} привел к уменьшению значения функции, то вероятность выбора направления $x_{k-1} - x_{k-2}$ на следующем шаге увеличивается. И наоборот, если при переходе от x_{k-2} к x_{k-1} значение функции увеличилось, то вероятность выбора направления $x_{k-1} - x_{k-2}$ на последующем шаге уменьшается. Таким образом, формулы (4) осуществляют обучение алгоритма. Величина $\delta \geqslant 0$ в (4) регулирует скорость обучения: чем больше $\delta > 0$, тем быстрее обучается алгоритм; при $\delta = 0$, как видно, обучения нет. Величина $\beta \geqslant 0$ в формулах (4) регулирует влияние предыдущих значений параметров на обучение алгоритма; при $\beta = 0$ алгоритм «забывает» предыдущие значения w_{k-1}. Для устранения возможного чрезмерного детерминирования алгоритма и сохранения способности алгоритма к достаточно быстрому обучению, на параметры w^i_k накладываются ограничения $|w^i_k| \leqslant c_i$, и при нарушении этих ограничений w^i_k заменяются ближайшим из чисел c_i и $-c_i$, $i = 1, \ldots, n$. Величины β, δ, c_i являются параметрами алгоритма.

Вместо формул (4), посредством которых производится обучение алгоритма, часто пользуются другими формулами

$$w^i_k = \beta w^i_{k-1} - \delta(f(x_{k-1}) - f(x_{k-2}))(x^i_{k-1} - x^i_{k-2}), \quad i = 1, \ldots, n, \quad k = 2, 3, \ldots \tag{5}$$

Описанный алгоритм покоординатного обучения имеет тот недостаток, что поиск и обучение происходят лишь по одному из 2^n направлений $\xi = (\xi^1, \ldots, \xi^n)$, где либо $\xi^i = 1$, либо $\xi^i = -1$. Отсутствие «промежуточных» направлений делает покоординатное обучение немобильным в областях с медленно изменяющимися направлениями спуска. От этого недостатка свободен следующий алгоритм.

б) *Алгоритм непрерывного самообучения.* Пусть имеется семейство случайных векторов $\xi = \xi(w) = \frac{\eta + w}{|\eta + w|}$, где $w = (w^1, \ldots, w^n)$ — параметры обучения, $\eta = (\eta^1, \ldots, \eta^n)$ — случайный вектор, координаты η^i которого представляют собой независимые случайные величины, распределенные равномерно на отрезке $[-1, 1]$. Поиск начинается с рассмотрения случайных векторов $\xi_0 = \xi(0)$, $\xi_1 = \xi(0)$, реализации которых используются при определении приближений x_0, x_1 по формулам (2). Обучение алгоритма при $k \geqslant 2$ производится так же, как в алгоритме покоординатного обучения, с помощью формул (4) или (5). При больших значениях $|w_k|$ влияние случайной величины η уменьшается, и направление $\xi_k = \xi(w_k)$ становится более детерминированным и близким к направлению w_k. Во избежание излишней детерминированности метода на параметры $w_k = (w^1_k, \ldots, w^n_k)$ накладываются ограничения $|w_k| \leqslant c = \text{const}$, и при нарушении этих ограничений w_k заменяется на $\frac{w_k}{|w_k|} c$.

Приведенные алгоритмы случайного поиска с обучением показывают, что процесс обучения в ходе поиска сопровождается уменьшением фактора случайности и увеличением степени детерминированности алгоритма поиска минимума, направляя поиск преимущественно по направлению убывания функции. В то же время наличие случайного фактора в выборе направления дает возможность алгоритму «переучиваться», если свойства функции в районе поиска изменились или предыдущее обучение было неточным. Случайный поиск с обучением в некотором смысле занимает промежуточное положение между случайным поиском без обучения и детерминированными методами поиска минимума из предыдущих параграфов. Разумеется, и в методах предыдущих параграфов можно обнаружить в том или ином виде элементы самообучения алгоритма, однако наличие случайного фактора в алгоритме делает метод случайного поиска более гибким.

3. Весьма усложняет решение задачи минимизации функций многих переменных наличие помех, когда на значения функции $f(x)$ в каждой точке x накладываются случайные ошибки.

Задача минимизации функций при наличии случайных ошибок относится к задачам *стохастического программирования*. Более подробно о стохастическом программировании, о теоретических и вычислительных аспектах методов случайного поиска, стохастической аппроксимации см., например, в [34; 77; 128; 226; 251; 256; 262; 301; 302; 305; 308; 318; 374; 377; 401; 495; 518; 538; 542; 586; 610; 662; 709; 713; 720; 777].

§ 20. Общие замечания

Выше были рассмотрены лишь немногие из известных в настоящее время методов минимизации. Заметим, что по сей день интенсивно продолжается разработка все новых и новых методов решения экстремальных задач, о чем свидетельствует неуклонно растущее количество публикаций в научной печати по этой тематике.

1. Возникают естественные вопросы: чем руководствоваться при выборе метода для решения той или иной конкретной экстремальной задачи, какой же метод является наилучшим? Иногда считают, что тот метод лучше, у которого выше скорость сходимости на некотором фиксированном классе

задач. Однако при таком способе оценки методов не принимается во внимание такое важное качество, как трудоемкость каждой отдельно взятой итерации метода. Нередко бывает, что при решении конкретной задачи выгоднее применять метод, который сходится не очень быстро и для получения решения с нужной точностью требует довольно большого числа итераций, но тем не менее из-за того, что каждая итерация метода осуществляется просто, суммарный объем вычислений и, следовательно, общее машинное время для получения решения оказывается меньшим, чем при применении другого быстросходящегося метода, каждая итерация которого весьма трудоемка. Таким образом, при характеристике метода минимизации важным является не столько скорость его сходимости, сколько общий объем вычислений, общее машинное время, необходимое для получения решения с нужной точностью.

При практическом использовании методов значительная часть времени, отведенного на расчеты, часто затрачивается на вычисление значений минимизируемой функции или ее производных. Поэтому в тех случаях, когда вычисление значений функции намного проще вычисления ее производных, естественно, выгоднее пользоваться теми методами, которые для своей реализации требуют лишь вычисления значений функции. Конечно, возможны и такие ситуации, когда имеются простые аналитические выражения для производных минимизируемой функции, — в таких случаях, возможно, выгоднее применять методы, использующие градиент или производные более высокого порядка.

Важными характеристиками метода минимизации являются также область сходимости метода, устойчивость метода к погрешностям, объем памяти ЭВМ, необходимой для реализации метода, удобство программирования, широта класса задач, к которым применим метод, и т. п.

Большое количество разнообразных и отчасти противоречивых характеристик методов, недостаточная разработанность методики оценки упомянутых характеристик затрудняют сравнение методов друг с другом. Иногда для сравнения методов минимизации задают некоторый набор тестовых задач (набор таких задач см., например, в [613; 738]) и лучшим признают тот метод, с помощью которого удастся решить указанные тестовые задачи с нужной точностью за меньшее число итераций, меньшее число вычислений значений функции или ее производных, или за меньшее машинное время. Несомненно, такие «соревнования» методов полезны, хотя и на их основе нельзя делать окончательные выводы о преимуществах того или иного метода. Здесь следует также заметить, что один и тот же метод, примененный для минимизации одной и той же функции, может привести к различным результатам в зависимости от того, на каком алгоритмическом языке составлена программа, каково качество транслятора (квалификация программиста), на какой ЭВМ решается задача и т. д.

Конечно, хотелось бы иметь метод, наилучший во всех отношениях. Однако такого универсального метода пока нет, и вряд ли такой метод существует. Поэтому для эффективного решения конкретной задачи минимизации, по-видимому, нужно разумно сочетать различные методы с учетом всевозможной априорной информации о решаемой задаче (гладкость исходных данных, выпуклость, физические или какие-либо иные соображения об области возможного расположения точки минимума и т. д.), имеющихся вычислительных средств, ресурсов машинного времени и т. п. В тех случаях, когда нет никакой априорной информации о задаче, которую нужно решить,

по-видимому, сначала полезно попробовать применить не очень точные, но простые методы минимизации (например, метод перебора значений функций на сетке с небольшим числом узловых точек, метод покоординатного спуска, метод случайного поиска), а затем на основе накопленной информации при необходимости перейти к более точным методам.

2. Успешное решение различных классов прикладных экстремальных задач невозможно без пакета минимизации, состоящего из библиотеки подпрограмм, охватывающей достаточно много методов минимизации, а также управляющих и вспомогательных программ. Пакеты минимизации могут быть использованы в автоматизированном или диалоговом режиме.

При работе с пакетом в диалоговом режиме математик-вычислитель, получая сведения о текущих результатах, оперативно вмешивается в процесс минимизации, осуществляет переход от одного метода к другому, изменяет параметры методов, параметры программ. Диалоговый режим работы с пакетом минимизации позволяет лучшим образом использовать опыт и интуицию математика-вычислителя и предъявляет высокие требования к его профессиональным знаниям в области методов решения экстремальных задач.

В тех случаях, когда пользователь, т. е. специалист, проводящий расчеты, не является компетентным в области методов решения экстремальных задач, желательно иметь пакеты минимизации, работающие в автоматическом режиме. Для работы в этом режиме пакет должен содержать управляющую программу, обеспечивающую автоматический выбор наиболее подходящей последовательности используемых методов, их параметров в зависимости от конкретной решаемой задачи.

Принцип построения пакетов минимизации, примеры таких пакетов описаны в [286; 499]. Заметим, что создание эффективно действующих и достаточно универсальных пакетов минимизации, которые могут быть использованы в различных режимах, представляет собой важную и большую научно-техническую задачу.

3. Следует обратить внимание читателя на то, что первоначальная постановка прикладных задач минимизации зачастую бывает достаточно грубой, упрощенной и предполагает, что в процессе решения задача будет уточняться. Это значит, что первоначальный вариант задачи не всегда имеет смысл решать слишком точно. Иногда гораздо выгоднее с помощью простых методов, с небольшой затратой машинного времени получить грубые предварительные результаты и затем проанализировать их вместе с экспертами, с заказчиком. Уже при таком упрощенном анализе может выясниться, что некоторые параметры и ограничения, ранее казавшиеся несущественными и поэтому не учтенные в первоначальной постановке задачи, должны быть включены в нее, и наоборот, часть прежних параметров и ограничений могут оказаться несущественными и без ущерба для существа задачи могут быть опущены. Заметим, что процесс уточнения постановки задачи весьма удобно проводить с помощью пакета минимизации в диалоговом режиме.

Иногда стремятся учесть многие детали задачи и создать слишком подробную математическую модель исследуемого процесса, а затем пытаются найти наилучшие, оптимальные значения всех параметров процесса. Однако такой подход может привести к задаче минимизации с очень большим числом переменных, и численное решение такой задачи может встретить непреодолимые трудности. Но даже в тех случаях, когда удается найти оптимальные значения параметров, их практическое использование может

оказаться невозможным из-за того, что заказчик, будучи не в состоянии охватить полученную информацию, может не понять разумность выработанных на ее основе рекомендаций и может от них вообще отказаться. Поэтому на первых этапах исследования прикладных задач минимизации желательно пользоваться простыми моделями, учитывающими основные, определяющие параметры.

4. К сожалению, последнему совету удается следовать не всегда. Например, математические модели социально-экономических процессов, достаточно адекватно отражающих основные закономерности, как правило, чрезвычайно сложны, содержат большое число переменных и приводят к так называемым задачам минимизации большой размерности. Численное решение таких задач обычными методами становится невозможным даже при использовании самых мощных современных ЭВМ. Некоторые классы задач большой размерности допускают разбиение на ряд слабо связанных между собой подзадач, имеющих сравнительно небольшие размерности, решая которые, иногда удается получить приближенное решение исходной задачи. Следует заметить, что задачи минимизации большой размерности к настоящему времени изучены недостаточно. Некоторые методы решения таких задач см., например, в [222; 470; 564; 613; 711; 746; 747; 759].

5. В ряде методов минимизации, описанных выше, предполагалось, что начальная точка x_0, принадлежащая множеству X, известна. Для некоторых множеств, таких как, например, параллелепипед, шар, гиперплоскость, указать такую точку x_0 совсем нетрудно. Однако не следует думать, что определение точки x_0 из любого множества X всегда просто. Например, если

$$X = \{x \in E^n:\ g_i(x) = 0,\ i = 1, \ldots, s\}, \tag{1}$$

то для определения точки $x_0 \in X$ нужно решать систему уравнений (вообще говоря, нелинейных). Чтобы найти какую-либо точку множества

$$X = \{x \in E^n:\ x \in X_0,\ g_i(x) \leqslant 0,\ i = 1, \ldots, m,\ g_i(x) = 0,\ i = m+1, \ldots, s\}, \tag{2}$$

придется решать смешанную систему уравнений и неравенств. Определение решения систем линейных или нелинейных уравнений и неравенств представляет собой весьма серьезную задачу, которой посвящена обширная литература; см., например, [59; 74; 89; 131; 192; 214; 222; 286; 334; 444; 480; 481; 550; 613; 630–635; 671; 695; 704; 747; 752].

Полезно заметить, что задачу нахождения какой-либо точки x_0, принадлежащей множеству (1) или (2), можно переформулировать в виде задачи минимизации. А именно, в случае множества (1) введем функцию

$$P(x) = \sum_{i=1}^{s} g_i^2(x), \quad x \in E^n,$$

а в случае множества (2) — функцию

$$P(x) = \sum_{i=1}^{m} (\max\{g_i(x); 0\})^p + \sum_{i=m+1}^{s} |g_i(x)|^p, \quad x \in X_0, \quad p > 0.$$

и рассмотрим задачу минимизации

$$P(x) \to \inf; \quad x \in X_0.$$

Для решения этой задачи могут быть использованы любые подходящие методы минимизации. Если $X \neq \varnothing$, то условие $x_0 \in X$ равносильно условию $P(x_0) = 0 = \inf\limits_{E^n} P(x) = P_*$. Если $P_* > 0$, то $X = \varnothing$. Если же $P_* = 0$, но нижняя грань $P(x)$ на E^n не достигается, то также $X = \varnothing$. Здесь предполагается, что множество X_0 имеет столь простую структуру, что нахождение точки $x_0 \in X_0$ не вызывает трудностей.

6. Интересно проанализировать доказательства теорем сходимости градиентного метода, методов проекции градиента, возможных направлений, условного градиента и т. д. Такой анализ показывает, что проводимые рассуждения опираются на предположения одного и того же типа, содержат много общих моментов, техника получения оценок скорости сходимости имеет общие черты. Возникает вопрос, нельзя ли создать общую методику исследования сходимости если и не всех, то хотя бы некоторых достаточно широких семейств методов минимизации? Оказывается, это возможно. К настоящему времени сделаны весьма удачные и интересные попытки создания такой методики, позволяющей единообразно исследовать сходимость широких классов методов минимизации, получать оценку скорости сходимости. К сожалению, мы здесь не имеем возможности останавливаться на этих увлекательных вопросах и отсылаем читателя к работам [77; 286; 319; 374; 495; 542; 582; 586; 613].

Можно, конечно, задаться вопросом: зачем нужны теоремы сходимости методов? Ведь на практике мы все равно можем реализовать лишь конечное число шагов используемого метода, и факт сходимости метода как будто не так уж и важен. К тому же условия теорем сходимости зачастую труднопроверяемы, в них редко учитываются погрешности задания исходных данных, погрешности в реализации метода. В общем, основания для сомнений есть. В то же время нетрудно привести немало доводов в пользу теорем сходимости методов. Такие теоремы подтверждают «добротность» используемого метода, очерчивают границы его применимости, из них нередко можно узнать о качественном поведении метода, о его сильных и слабых сторонах, не прибегая к трудоемким численным экспериментам, в формулировках и доказательствах таких теорем могут содержаться полезные для практики конструктивные соображения. Роль теорем сходимости содержательно обсуждается в [586]. Завершим главу цитатой из [586]: «теоретические исследования методов оптимизации могут дать много информации вычислителю-практику. Нужно лишь при этом проявлять разумную осторожность и здравый смысл».

ГЛАВА 6

Принцип максимума Понтрягина

В этой главе рассматриваются задачи оптимального управления процессами, описываемыми системами обыкновенных дифференциальных уравнений. Этот класс экстремальных задач существенно отличается от рассмотренных: если в задачах минимизации функции конечного числа переменных искомая точка минимума являлась точкой n-мерного пространства, то в задачах оптимального управления искомая точка минимума представляет собой функцию, принадлежащую некоторому бесконечномерному функциональному пространству. Такие задачи имеют многочисленные приложения в механике космического полета, в вопросах управления электроприводами, химическими или ядерными реакторами, виброзащиты и т. д.

Эффективным средством исследования задач оптимального управления является принцип максимума Понтрягина [587], представляющий собой необходимое условие оптимальности в таких задачах. Принцип максимума, открытый коллективом российских математиков во главе с академиком Л. С. Понтрягиным, представляет собой одно из крупных достижений современной математики и является краеугольным камнем современной математической теории оптимального управления. Принцип максимума Понтрягина существенно обобщает и развивает основные результаты классического вариационного исчисления, созданного Эйлером, Лагранжем и другими выдающимися математиками прошлого. Появление принципа максимума стимулировало последующее бурное развитие теории экстремальных задач и методов их решения.

Из обширной литературы, посвященной различным аспектам современной теории оптимального управления и управляемых систем, их приложений, упомянем [6–10; 12; 14; 15; 21; 31; 34; 36–39; 40–44; 46; 47; 50; 55–58; 67; 72; 79; 81; 93; 96–101; 105; 106; 108; 110; 115; 121; 132; 134; 135; 137; 140; 141; 148; 151; 198–202; 204–207; 209; 210; 212; 217; 221; 225; 240; 241; 244; 245; 249; 253; 254; 267; 269; 274–280; 282; 283; 286; 288; 291; 293; 310–312; 321–325; 328; 329; 331; 332; 336; 358; 366; 376; 379; 380–386; 406; 409–414; 417; 418; 427; 429–431; 436; 440; 457; 476; 477; 486–488; 497; 498; 500; 502; 503; 505; 513; 528–533; 541; 545; 546; 566; 568–571; 578; 582; 583; 587; 589;602; 611; 616; 637; 643; 646; 653–656; 663; 677; 683; 687; 688; 695; 702; 703; 712; 715–717; 719; 720; 722–724; 726–729; 731–734; 739; 744; 753–756; 781; 787; 809; 818; 819].

§ 1. Постановка задачи оптимального управления

1. Приведем несколько конкретных задач оптимального управления.

Пример 1. Движение плоского маятника, подвешенного к точке опоры при помощи жесткого невесомого стержня (рис. 6.1), как известно, описывается уравнением

$$I\ddot{\theta} + b\dot{\theta} + mgl \sin\theta = M(\tau),$$

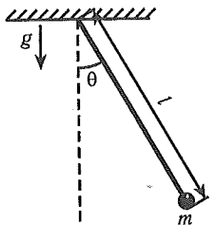

Рис. 6.1

где l — длина жесткого стержня маятника, m — масса, сосредоточенная в конце стержня, $I = ml^2$ — момент инерции, g — гравитационная постоянная (ускорение силы тяжести), $b \geqslant 0$ — коэффициент демпфирования, τ — время, $M(\tau)$ — внешний управляющий момент, $\theta = \theta(\tau)$ — угол отклонения стержня от точки устойчивого равновесия. Если сделать замену переменной $t = \tau\sqrt{mgl/I}$, то это уравнение можно привести к виду

$$\ddot{\varphi} + \beta\dot{\varphi} + \sin\varphi = u(t), \tag{1}$$

где

$$\varphi = \varphi(t) = \theta(t\sqrt{I/(mgl)}), \quad \beta = b/\sqrt{Imgl}, \quad u(t) = M(t\sqrt{I/(mgl)})/(mgl).$$

Обозначим $x^1(t) = \varphi(t)$ (угол отклонения маятника), $x^2(t) = \dot{\varphi}(t)$ (скорость маятника). Тогда уравнение (1) запишется в виде системы двух уравнений первого порядка:

$$\dot{x}^1(t) = x^2(t), \quad \dot{x}^2(t) = -\beta x^2(t) - \sin x^1(t) + u(t). \tag{2}$$

Пусть в начальный момент $t = 0$ маятник отклонился на угол $x^1(0) = x_0^1$ и имеет начальную скорость $x^2(0) = x_0^2$. Будем также считать, что функция $u(t)$ — управляющий момент (выбор которого может влиять на движение маятника) — удовлетворяет ограничению

$$|u(t)| \leqslant \gamma, \quad \gamma = \text{const} > 0, \quad t \geqslant 0. \tag{3}$$

Здесь возможны следующие постановки задач оптимального управления: выбрать управление $u(t)$, удовлетворяющее условиям (3) так, чтобы:

1) за минимальное время T остановить маятник в одной из точек устойчивого равновесия, т. е. добиться выполнения условий

$$x^1(T) = 2\pi k, \quad x^2(T) = 0 \tag{4}$$

при некотором k, $k = 0, \pm 1, \ldots$ (задача быстродействия);

2) за минимальное время T добиться выполнения условия

$$(x^1(T))^2 + (x^2(T))^2 \leqslant \varepsilon,$$

где $\varepsilon > 0$ — заданное число;

3) к заданному моменту времени T величина $(x^1(T))^2 + (x^2(T))^2$, или $\int_0^T (x^1(t))^2 dt$, или $\int_0^T ((x^1(t))^2 + (x^2(t))^2) dt$, или $\max\limits_{0 \leqslant t \leqslant T} |x^1(t)|$, или $\max\limits_{0 \leqslant t \leqslant T} \max\{|x^1(t)|, |x^2(t)|\}$ принимала минимально возможное значение, или

4) в заданный момент T выполнялось равенство $x^2(T)=0$, а величина $x^1(T)$ была максимально возможной (задача о накоплении возмущений), или

5) к заданному моменту T добиться выполнения условий (4) и минимизировать величину $\int_0^T u^2(t) dt$ (условие (3) здесь может быть опущено).

Если колебание маятника ограничено какими-либо упорами, то в перечисленных задачах нужно еще требовать выполнения условия вида

$$|x^1(t)| \leqslant \mu, \quad \mu = \text{const} > 0.$$

На управление $u(t)$ вместо условия (3) (или наряду с условием (3)) могут накладываться ограничения вида

$$\int_0^T u^2(t) dt \leqslant R,$$

где $R = \text{const} > 0$.

При изучении малых колебаний маятника часто полагают $\sin\varphi \approx \varphi$, и тогда уравнение (1) и эквивалентная ему система становятся линейными и будут иметь вид

$$\ddot{\varphi} + \beta\dot{\varphi} + \varphi = u(t)$$

и соответственно

$$\dot{x}^1(t) = x^2(t), \quad \dot{x}^2(t) = -\beta x^2(t) - x^1(t) + u(t).$$

Пример 2. Как известно [249, с. 129], движение центра масс космического аппарата и расход массы описывается системой дифференциальных уравнений

$$\dot{r} = v, \quad \dot{v} = gp/G + F, \quad \dot{G} = -gq, \quad 0 \leqslant t \leqslant T, \tag{5}$$

где t — время, $r = r(t) = (r_1(t), r_2(t), r_3(t))$ — радиус-вектор центра масс аппарата, $v = v(t) = (v_1(t), v_2(t), v_3(t))$ — скорость центра масс, $G = G(t)$ — текущий вес аппарата, g — коэффициент пропорциональности между массой и весом, $p = p(t) = (p_1(t), p_2(t), p_3(t))$ — вектор тяги двигателя, $q = q(t)$ — расход рабочего вещества, $F = F(r, t) = (F_1, F_2, F_3)$ — вектор ускорения от гравитационных сил.

В каждый момент времени t движение космического аппарата характеризуется величинами $r(t)$, $v(t)$, $G(t)$, называемыми фазовыми координатами. Пусть в начальный момент $t = 0$ фазовые координаты аппарата известны:

$$r(0) = r_0, \quad v(0) = v_0, \quad G(0) = G_0. \tag{6}$$

Величины $q = q(t)$, $p = p(t)$ являются управлением — задавая их поразному, можно получить различные фазовые траектории (решения) задачи (5), (6). Конструктивные возможности аппарата, ограниченность ресурсов рабочего вещества накладывают на управление $q(t)$, $p(t)$ ограничения, например, вида

$$p_{\min} \leqslant |p(t)| \leqslant p_{\max}, \quad q_{\min} \leqslant q(t) \leqslant q_{\max}, \quad o \leqslant t \leqslant T,$$

или $\int_0^T q^2(t)dt \leqslant R$, $R = \text{const} > 0$. Кроме того, на фазовые траектории задачи (5), (6) могут накладываться некоторые ограничения, вытекающие, например, из условий того, чтобы вес аппарата был не меньше определенной величины или траектория полета проходила вне определенных областей космического пространства (областей повышенной радиации) и др.

Здесь возникают задачи выбора управлений $q(t)$, $p(t)$ так, чтобы управления и соответствующие им траектории задачи (5), (6) удовлетворяли всем наложенным ограничениям, и кроме того, достигалась та или иная цель. Например, здесь возможны следующие задачи:

1) попасть в заданную точку или область космического пространства за минимальное время;

2) к заданному моменту времени попасть в заданную область пространства с заданной скоростью (совершить мягкую посадку, например) и с максимальным весом аппарата или с минимальной затратой энергии;

3) достичь определенной скорости за минимальное время и т. п.

Большое число прикладных задач оптимального управления, которые связаны с механикой полета летательных аппаратов в космосе и атмосфере, с работой электроприводов, химических и ядерных реакторов, с вопросами виброзащиты и амортизации, с математической экономикой и т. д., читатель найдет в [21; 53; 67; 108; 110; 240; 246–249; 253; 254; 293; 328; 417; 418; 431; 457; 466; 497; 498; 500; 568; 608; 611; 612; 616; 683; 712; 715; 716; 719; 726; 727; 750; 751; 753–756].

2. Приведенные в примерах 1, 2 задачи являются частным случаем более общей задачи оптимального управления, к формулировке которой мы переходим. Пусть движение некоторого управляемого объекта (течение управляемого процесса, изменение управляемой системы) описывается обыкновенными дифференциальными уравнениями

$$\dot{x}^i = f^i(x^1, x^2, \ldots, x^n, u^1, u^2, \ldots, u^r), \quad i = 1, \ldots, n,$$

которые в векторной форме можно записать в виде

$$\dot{x} = f(x, u, t), \tag{7}$$

где t — время, $x = (x^1, x^2, \ldots, x^n)$ — величины, характеризующие движение объекта в зависимости от времени и называемые *фазовыми координатами объекта*, $u = (u^1, u^2, \ldots, u^r)$ — *параметры управления* («положение рулей» объекта), выбором которых можно влиять на движение объекта, $f = (f^1, f^2, \ldots, f^n)$; функции $f^i(x, u, t)$, $i = 1, \ldots, n$, описывающие внутреннее устройство объекта и учитывающие различные внешние факторы, предполагаются известными.

Для того, чтобы фазовые координаты объекта (процесса, системы) (7) были определены в виде функций времени $x = x(t)$ на некотором отрезке $t_0 \leqslant t \leqslant T$, необходимо в начальный момент времени t_0 задать начальное условие $x(t_0) = x_0$ и параметры управления $u = (u^1, u^2, \ldots, u^r)$ как функции времени $u = u(t)$ при $t \in [t_0, T]$. Тогда фазовые координаты $x = x(t)$ будут определяться как решение следующей задачи Коши:

$$\dot{x}(t) = f(x(t), u(t), t), \quad t_0 \leqslant t \leqslant T, \tag{8}$$

$$x(t_0) = x_0. \tag{9}$$

Нетрудно видеть, что функции $u = u(t)$, называемые *управлениями*, должны удовлетворять определенным требованиям непрерывности, гладкости, так как, с одной стороны, при слишком «плохих» (слишком «разрывных») $u(t)$ задача (8), (9) может не иметь смысла, с другой стороны, слишком «плохая» функция $u(t)$ не будет иметь физического смысла управления. В большинстве прикладных задач в качестве управлений $u = u(t)$ могут быть взяты кусочно-непрерывные функции. Напоминаем, что функция $u(t)$ называется *кусочно-непрерывной* на отрезке $[t_0, T]$, если $u(t)$ непрерывна во всех точках $i \in [t_0, T]$, за исключением, быть может, лишь конечного числа точек $\tau_1, \ldots, \tau_p \in [t_0, T]$, в которых функция $u(t)$ может терпеть разрывы типа скачка, т. е. существуют конечные пределы

$$\lim_{t \to \tau_i - 0} u(t) = u(\tau_i - 0), \quad \lim_{t \to \tau_i + 0} u(t) = u(\tau_i + 0),$$

но, вообще говоря, $u(\tau_i - 0) \neq u(\tau_i + 0)$, $i = 1, \ldots, p$. В тех прикладных задачах, в которых разрывные управления технически нереализуемы, рассматриваются лишь непрерывные управления $u(t)$. Встречаются также задачи, в которых, кроме непрерывности, от $u(t)$ требуется существование кусочно-непрерывной производной $\dot{u}(t)$ — такие управления называют *кусочно-гладкими*.

В теоретических исследованиях упомянутые классы кусочно-непрерывных, кусочно-гладких управлений часто бывают слишком узкими, и вместо

них приходится рассматривать более широкие классы управлений, такие, как, например, пространство $L_p^r[t_0, T]$ при некотором p, $1 \leqslant p \leqslant \infty$ [393]. Через $L_p^r[t_0, T]$ при $1 \leqslant p \leqslant \infty$ будем обозначать пространство измеримых вектор-функций $u(t) = (u^1(t), \ldots, u^r(t))$, $t_0 \leqslant t \leqslant T$, для которых функция $|u(t)|^p$ суммируема на $[t_0, T]$ в смысле Лебега, и, следовательно, имеет смысл норма

$$\|u\|_{L_p} = \left(\int_{t_0}^{T} |u(t)|^p dt\right)^{1/p}, \quad 1 \leqslant p < \infty.$$

Если $p = \infty$, то под $L_\infty^r[t_0, T]$ понимается пространство ограниченных измеримых вектор-функций $u(t) = (u^1(t), \ldots, u^r(t))$, $t_0 \leqslant t \leqslant T$, с нормой

$$\|u\|_{L_\infty} = \operatorname{ess} \sup_{t_0 \leqslant t \leqslant T} |u(t)| = \inf_{v(t)} \sup_{t_0 \leqslant t \leqslant T} |v(t)|,$$

где $v(t)$ пробегает множество всех измеримых функций, совпадающих с $u(t)$ почти всюду на отрезке $[t_0, T]$. Если читатель недостаточно знаком с интегралом Лебега и пространствами $L_p^r[t_0, T]$ [393], то всюду в этой главе он может считать, что все рассматриваемые управления $u(t)$ суть кусочно-непрерывные функции.

Далее, как видно из примеров 1, 2, значения управлений не могут быть совершенно произвольными и подчиняются некоторым ограничениям. Такие ограничения можно описать условием

$$u(t) \in V(t), \quad t_0 \leqslant t \leqslant T, \tag{10}$$

где $V(t)$ — заданное множество из E^r при каждом $t \in [t_0, T]$. Например, в случае ограничений (3) $V(t) = \{u\colon u \in E^1, |u| \leqslant \gamma\}$ при всех t. Для кусочно-непрерывных управлений выполнение условий (10) требуется для всех $t \in [t_0, T]$, в которых управление непрерывно, а для измеримых управлений — почти всюду на $[t_0, T]$.

Кроме ограничений (10), возможны также и ограничения вида

$$\|u\|_{L_p}^p = \int_{t_0}^{T} |u(t)|^p dt \leqslant R^p, \quad R = \text{const} > 0, \tag{11}$$

при некотором p, $1 \leqslant p < \infty$.

Таким образом, постановка задачи оптимального управления прежде всего предполагает, что выбран некоторый класс функций — управлений (например, кусочно-непрерывные, кусочно-гладкие или функции из $L_p^r[t_0, T]$, $1 \leqslant p \leqslant \infty$) и указаны налагаемые на них ограничения (например, ограничения вида (10) или (11)).

Заметим, что в учебной литературе символом $z(t) = (z^1(t), \ldots, z^m(t))$ часто обозначают как значение функции в точке t, так и саму функцию, которая представляет собой отображение области определения функции в пространстве E^m, ставящее в соответствие каждой точке t из области определения некоторую точку из E^m. Отдавая дань традициям, мы будем продолжать пользоваться этим не вполне определенным символом в тех случаях, когда из контекста нетрудно понять, идет ли речь о функции в целом или о ее значении в конкретной точке. В тех случаях, когда обозначение $z(t)$ может привести к недоразумениям, за значением функции в точке t будем сохранять обозначение $z(t)$, а саму функцию будем обозначать, через $z(\cdot)$ или просто z. В свете этих обозначений подчеркнем, что ограничения (10)

являются ограничениями на значения функции, и поэтому было бы бессмысленно вместо (10) писать $u(\cdot) \in V(t)$. Ограничение (11), наоборот, накладывается на всю функцию $u(\cdot)$ в целом и не является ограничением на значения функции — функция $u(\cdot)$, удовлетворяющая этому ограничению, в отдельных точках или промежутках малой длины может принимать произвольные значения. Поэтому (11) можно записать в виде $\|u(\cdot)\|_{L_p} \leqslant R$, а обозначение $\|u(t)\|_{L_p} \leqslant R$, иногда встречающееся в литературе, не вполне удачное.

3. Итак, пусть заданы точка $x_0 \in E^n$ и некоторое кусочно-непрерывное управление $u = u(\cdot) = u(t)$, $t_0 \leqslant t \leqslant T$, или управление $u(\cdot) \in L_p^r[t_0, T]$ при некотором $p \geqslant 1$. Рассмотрим задачу Коши (8), (9). Сразу же возникает вопрос: что понимать под решением этой задачи? Если функции $u(t)$, $f(x, u, t)$ непрерывны, то, как обычно принято в учебниках по дифференциальным уравнениям [376; 588; 694], под решением задачи (8), (9) можно понимать функцию $x = x(\cdot) = x(t)$, $t_0 \leqslant t \leqslant T$, которая непрерывно дифференцируема на отрезке $[t_0, T]$ и удовлетворяет условиям (8), (9). Однако для случая кусочно-непрерывных или измеримых управлений, как видно из (8), требовать существование непрерывно дифференцируемого решения задачи (8), (9), вообще говоря, не имеет смысла. Поэтому мы будем пользоваться следующим более общим определением решения задачи (8), (9).

Определение 1. Непрерывную функцию $x = x(\cdot) = x(t)$, $t_0 \leqslant t \leqslant T$, удовлетворяющую равенству

$$x(t) = \int_{t_0}^{t} f(x(\tau), u(\tau), \tau) d\tau + x_0, \quad t_0 \leqslant t \leqslant T, \tag{12}$$

будем называть *решением* или *траекторией задачи* (8), (9), соответствующей начальному условию x_0 и управлению $u = u(\cdot)$, и будем обозначать через $x = x(\cdot, u, x_0) = x(\cdot, u(\cdot), x_0)$ или $x = x(t, u, x_0)$, или $x = x(t, u, x_0, t_0)$, $t_0 \leqslant t \leqslant T$. Начальную точку $x(t_0, u, x_0)$ будем называть *левым концом траектории* $x(\cdot, u, x_0)$, t_0 — *начальным моментом*, $x(T, u, x_0)$ — *правым концом траектории*, T — *конечным моментом*.

В тех случаях, когда ясно, какому именно управлению $u(\cdot)$ или начальному условию x_0 и начальному моменту t_0 соответствует траектория, в обозначении $x(\cdot, u, x_0, t_0)$ букву u или x_0, t_0 будем опускать и просто писать $x(\cdot, u)$, или $x(\cdot, x_0)$, или $x = x(\cdot) = x(t)$, $t_0 \leqslant t \leqslant T$.

Классические теоремы существования и единственности решения задачи Коши

$$\dot{x} = g(x, t), \quad x(t_0) = x_0,$$

в учебниках по дифференциальным уравнениям обычно доказываются при требовании непрерывности $g(x, t)$ и $g_x(x, t)$ по совокупности переменных в некоторой области, содержащей точку (x_0, t_0) (условие непрерывности $g_x(x, t)$ часто заменяется условием Липшица $g(x, t)$ по переменной x) [376; 588; 694]. Однако если $u(t) \in L_p^1[t_0, T]$, $p \geqslant 1$, то непрерывности $g(x, t) \equiv f(x, u(t), t)$ по переменной t ожидать не приходится, и классические теоремы существования и единственности решения здесь становятся недостаточными. Тем не менее, используя ту же технику доказательства упомянутых классических теорем, можно получить существование и единственность решения задачи (8), (9) и для кусочно-непрерывных управлений $u(t)$ или $u(t) \in L_p^r[t_0, T]$, $p \geqslant 1$. Мы здесь ограничимся доказательством следующей теоремы.

Теорема 1. *Пусть функция $f(x, u, t)$ определена и непрерывна по совокупности переменных при всех $(x, u, t) \in E^n \times E^r \times [t_0, T]$ и пусть*

$$|f(x, u, t) - f(y, u, t)| \leqslant L(t)|x - y| \tag{13}$$

при всех $(x, u, t), (y, u, t) \in E^n \times E^r \times [t_0, T]$, где $L(t)$ — неотрицательная функция, принадлежащая $L_1(t_0, T)$. Тогда для любого ограниченного измеримого управления $u(t)$ (т. е. $u(t) \in L_\infty^r[t_0, T]$) и начального условия x_0 задача (8), (9) имеет, и притом единственное, решение $x = x(t)$, определенное на всем отрезке $[t_0, T]$. Это решение имеет производную $\dot{x}(t)$ почти всюду на $[t_0, T]$, $\dot{x}(t) \in L_\infty^n[t_0, T]$ и удовлетворяет уравнению (8) при почти всех $t \in [t_0, T]$.

Доказательство. Пространство непрерывных вектор-функций $x(t) = (x^1(t), \ldots, x^n(t))$, $t_0 \leqslant t \leqslant T$, с нормой

$$\|x\|_C = \max_{t_0 \leqslant t \leqslant T} |x(t)|$$

обозначим через $C^n[t_0, T]$. Как известно [179], $C^n[t_0, T]$ — полное нормированное пространство. Зафиксируем какие-либо точки x_0 и ограниченное измеримое управление $u = u(t)$, $t_0 \leqslant t \leqslant T$. Можно показать [14], что тогда для любой функции $x(t) \in C^n[t_0, T]$ функция $f(x(t), u(t), t)$ будет ограниченной измеримой функцией переменной t на отрезке $[t_0, T]$. Определим отображение A:

$$z(t) = Ax = \int_{t_0}^{t} f(x(\tau), u(\tau), \tau)d\tau + x_0, \quad t_0 \leqslant t \leqslant T, \tag{14}$$

действующее из $C^n[t_0, T]$ в $C^n[t_0, T]$. Значение функции $z(\cdot) = Ax(\cdot)$ в точке t будем обозначать через $Ax(\cdot)(t)$. Покажем, что отображение A^m — m-я степень отображения A — при достаточно большом m будет сжимающим [393]. Для этого с помощью индукции докажем, что для любых $x(\cdot)$, $y(\cdot) \in C^n[t_0, T]$

$$|A^m x(\cdot)(t) - A^m y(\cdot)(t)| \leqslant \frac{1}{m!} \max_{t_0 \leqslant \tau \leqslant t} |x(\tau) - y(\tau)| \Big(\int_{t_0}^{t} L(\tau)d\tau \Big)^m \tag{15}$$

при всех t, $t_0 \leqslant t \leqslant T$, $m = 1, 2, \ldots$. Из (14) с учетом условия (13) имеем

$$|Ax(\cdot)(t) - Ay(\cdot)(t)| = \Big| \int_{t_0}^{t} [f(x(\tau), u(\tau), \tau) - f(y(\tau), u(\tau), \tau)]d\tau \Big| \leqslant$$
$$\leqslant \int_{t_0}^{t} L(\tau)|x(\tau) - y(\tau)|d\tau \leqslant \max_{t_0 \leqslant \tau \leqslant t} |x(\tau) - y(\tau)| \int_{t_0}^{t} L(\tau)d\tau. \tag{16}$$

Оценка (15) при $m = 1$ доказана. Пусть оценка (15) верна для некоторого $m \geqslant 1$. Тогда с помощью неравенства (16) получим

$$|A^{m+1}x(\cdot)(t) - A^{m+1}y(\cdot)(t)| = |A(A^m x(\cdot))(t) - A(A^m y(\cdot))(t)| \leqslant$$
$$\leqslant \int_{t_0}^{t} L(\tau)|A^m x(\cdot)(\tau) - A^m y(\cdot)(\tau)|d\tau \leqslant$$
$$\leqslant \int_{t_0}^{t} L(\tau) \frac{1}{m!} \max_{t_0 \leqslant \xi \leqslant \tau} |x(\xi) - y(\xi)| \Big(\int_{t_0}^{\tau} L(\xi)d\xi \Big)^m d\tau \leqslant$$
$$\leqslant \frac{1}{m!} \max_{t_0 \leqslant \xi \leqslant t} |x(\xi) - y(\xi)| \int_{t_0}^{t} L(\tau) \Big(\int_{t_0}^{\tau} L(\xi)d\xi \Big)^m d\tau =$$

$$= \frac{1}{(m+1)!} \max_{t_0 \leqslant \tau \leqslant t} |x(\tau) - y(\tau)| \int_{t_0}^{t} \frac{d}{d\tau} \Big(\int_{t_0}^{\tau} L(\xi)d\xi \Big)^{m+1} d\tau =$$
$$= \frac{1}{(m+1)!} \max_{t_0 \leqslant \tau \leqslant t} |x(\tau) - y(\tau)| \Big(\int_{t_0}^{t} L(\tau)d\tau \Big)^{m+1}$$

для любых $t \in [t_0, T]$. Оценка (15) доказана. Из этой оценки следует, что

$$\|A^m x(\cdot) - A^m y(\cdot)\|_C \leqslant \frac{1}{m!} \Big(\int_{t_0}^{T} L(\tau)d\tau \Big)^m \|x(\cdot) - y(\cdot)\|_C.$$

Так как $\lim\limits_{m \to \infty} \frac{1}{m!} \Big(\int_{t_0}^{T} L(\tau)d\tau \Big)^m = 0$, то при достаточно большом m будем иметь $\frac{1}{m!} \Big(\int_{t_0}^{T} L(\tau)d\tau \Big)^m < 1$. Таким образом, отображение A^m является сжимающим. Из принципа сжимающих отображений ([393, с. 82]) следует существование единственной функции $x(\cdot) \in C^n[t_0, T]$, для которой $x(\cdot) = Ax(\cdot)$, что равносильно выполнению равенства (12).

Из свойств интеграла Лебега с переменным верхним пределом (см. [393, с. 344]) и из (12) следует, что получившаяся функция $x(t)$, $t_0 \leqslant t \leqslant T$, абсолютно непрерывна, ее производная $\dot{x}(t) \in L_\infty^n[t_0, T]$, и уравнение (8) удовлетворяется почти всюду на $[t_0, T]$. Теорема 1 доказана. □

Замечание 1. Вместо условия (13) можно потребовать непрерывность $\frac{\partial f}{\partial x} = \left\{ \frac{\partial f^i(x, u, t)}{\partial x^j},\ i, j = 1, \ldots, n \right\}$ при $(x, u, t) \in E^n \times E^r \times [t_0, T]$, однако в этом случае существование решения задачи (8), (9) можно гарантировать, вообще говоря, лишь на отрезке $[t_0, t_0 + \alpha]$, где α — достаточно малое число.

Замечание 2. Если управление $u = u(t) \in L_p^r[t_0, T]$, $1 \leqslant p < \infty$, то теорема 1 и ее доказательство останутся в силе, если, например, дополнительно потребовать

$$|f(x, u, t)| \leqslant C_0(|x| + |u|^p) + C_1(t) \tag{17}$$

для всех $(x, u, t) \in E^n \times E^r \times [t_0, T]$, где $C_0 = \text{const} \geqslant 0$, $C_1(t) \geqslant 0$, $C_1(t) \in L_1[t_0, T)$. Условие (17) нужно для обеспечения включения $f(x, u(t), t) \in L_1[t_0, T]$ для любых $x(\cdot) \in C^n[t_0, T]$, $u(\cdot) \in L_p^r[t_0, T]$, чтобы отображение (14) имело смысл.

Более тонкие теоремы существования и единственности решения задачи (8), (9) для управлений $u(t) \in L_p^r[t_0, T]$, $1 \leqslant p \leqslant \infty$, можно найти в [14; 132; 212; 358; 457].

Остановимся еще на случае линейной системы, когда вместо (8) имеет место

$$\dot{x}(t) = A(t)x(t) + B(t)u(t) + f(t), \quad t_0 \leqslant t \leqslant T, \tag{18}$$

где $A(t) = \{a_{ij}(t)\}$, $B(t) = \{b_{ij}(t)\}$ — матрицы размеров $n \times n$ и $n \times r$ соответственно, $f(t) = (f^1(t), \ldots, f^n(t))$, $t_0 \leqslant t \leqslant T$.

Теорема 2. *Пусть элементы $\{a_{ij}(t)\}$, $\{b_{ij}(t)\}$ матриц $A(t)$, $B(t)$ принадлежат $L_\infty[t_0, T]$, а $f(t) \in L_1^n[t_0, T]$. Тогда для каждого управления $u = u(t) \in L_p^r[t_0, T]$, где p — какое-либо фиксированное число, $1 \leqslant p \leqslant \infty$, и любой точки $x_0 \in E^n$ задача Коши для системы (18) с начальным условием $x(t_0) = x_0$ имеет, и притом единственное, решение $x = x(t)$ в смысле определения 1, определенное на всем отрезке $[t_0, T]$. Это решение имеет производную $\dot{x}(t)$ почти всюду на $[t_0, T]$, $\dot{x}(t) \in L_1^n$*

$[t_0, T]$ *и удовлетворяет уравнению* (18) *почти всюду на* $[t_0, T]$. *Если кроме перечисленных условий, еще имеет место включение* $f(t) \in L_p^n[t_0, T]$, *то* $\dot{x}(t) \in L_p^n[t_0, T]$.

Доказательство. Нетрудно видеть, что правая часть уравнения (18) удовлетворяет условию (13) с $L(t) = \|A(t)\| \in L_\infty[t_0, T]$. Кроме того, $g(t) \equiv A(t)x(t) + B(t)u(t) + f(t) \in L_1^n[t_0, T]$ для любых $x(t) \in C^n[t_0, T]$, $u(t) \in L_p^r[t_0, T]$. Дальнейшее доказательство проводится так же, как доказательство теоремы 1. □

4. Вернемся к постановке задачи оптимального управления. Как видно из примеров 1, 2, не только на управляющие параметры объекта, но и на его фазовые координаты могут накладываться некоторые дополнительные ограничения, которые не вытекают из свойств системы (8) и ограничений на управления. Такие ограничения можно описать условием

$$x(t) = x(t, u(\cdot), x_0, t_0) \in G(t), \quad t_0 \leqslant t \leqslant T, \tag{19}$$

где $G(t)$ — некоторое заданное множество из E^n при каждом $t \in [t_0, T]$, например, $G(t) = \{x \in E^n\colon G^i(x, t) \leqslant 0,\ i = 1, \ldots, m_3;\ G^i(x, t) = 0,\ i = m_3 + 1, \ldots, s_3\}$, $G^i(x, t)$ — заданные функции, $t_0 \leqslant t \leqslant T$. Ограничения (19) часто называют *фазовыми ограничениями*.

Далее, начальный и конечный моменты времени t_0 и T, $t_0 \leqslant T$, характеризующие продолжительность движения объекта, могут зависеть от управления (например, в задачах быстродействия) и не всегда могут быть заданы заранее. В таких случаях обычно указывают ограничения

$$t_0 \in \Theta_0, \quad T \in \Theta_1, \tag{20}$$

где Θ_0, Θ_1 — заданные множества на числовой оси $\mathbb{R} = \{t\colon -\infty < t < +\infty\}$ (не исключается возможность, что $\Theta_0 = \mathbb{R}$ или $\Theta_1 = \mathbb{R}$).

Наконец остановимся на условиях, которым должны удовлетворять левый и правый концы траектории. Собственно говоря, из включений (19) при $t = t_0$ и $t = T$ уже следуют условия $x(t_0) \in G(t_0)$, $x(T) \in G(T)$, и в некоторых случаях нет необходимости как-то еще иначе выделять ограничения на концы траектории. Однако могут возникнуть ситуации (например, при $G(t) = E^n$, $t_0 < t < T$), когда такие ограничения удобнее выделить и рассматривать самостоятельно. В таких случаях будем считать, что в E^n при каждом $t_0 \in \Theta_0$ задано множество $S_0(t_0)$ и при каждом $T \in \Theta_1$ — множество $S_1(T)$, и условия на концах траекторий будем записывать в виде

$$x(t_0) \in S_0(t_0), \quad t_0 \in \Theta_0; \quad x(T) \in S_1(T), \quad T \in \Theta_1. \tag{21}$$

В задачах оптимального управления принята следующая классификация условий (20), (21). Если множество Θ_0 состоит из единственной точки t_0, то *начальный момент называют закрепленным*; если Θ_1 состоит из одной точки T, то *конечный момент называют закрепленным*. Если множество $S_0(t_0)$ [или $S_1(T)$] состоит из одной точки и не зависит от t_0, т. е. $S_0(t_0) = \{x_0\}$, $t_0 \in \Theta_0$ [или соответственно $S_1(T) = \{x_1\}$, $T \in \Theta_1$], то говорят, что *левый* [*правый*] *конец траектории закреплен*. Если $S_0(t_0) \equiv E^n$, $t_0 \in \Theta_0$ [или $S_1(T) \equiv E^n$, $T \in \Theta_1$], то *левый* [*правый*] *конец траектории называют свободным*. В остальных случаях *левый* [*соответственно правый*] *конец траектории называют подвижным*. Примером того, как могут задаваться множества $S_0(t_0)$, является

$$S_0(t_0) = \{x\colon x \in E^n,\ h_i(x, t_0) \leqslant 0,\ i = 1, \ldots, m_0,\ h_i(x, t_0) = 0,\ i = m_0 + 1, \ldots, s_0\}, \tag{22}$$

где функции $h_i(x, t)$, $i = 1, \ldots, s_0$, определены при $x \in E^n$, $t \in \Theta_0$. Аналогично, примером множества $S_1(T)$ служит

$$S_1(T) = \{y\colon y \in E^n,\ g_i(y, T) \leqslant 0,\ i = 1, \ldots, m_1,\ g_i(y, T) = 0,\ i = m_1 + 1, \ldots, s_1\}, \tag{23}$$

где функции $g_i(y, t)$, $i = 1, \ldots, s_1$, определены при $y \in E^n$, $t \in \Theta_1$.

В приложениях нередко возникают также задачи, в которых левый и правый концы траектории должны выбираться согласованно, в зависимости друг от друга. Это требование можно записать в виде

$$(x(t_0), x(T)) \in S(t_0, T), \quad t_0 \in \Theta_0, \quad T \in \Theta_1, \tag{24}$$

где $S(t_0, T)$ при каждом $(t_0, T) \in \Theta_0 \times \Theta_1$ представляет собой заданное множество из $E^n \times E^n$. Примером такого множества является

$$S_0(t_0, T) = \{(x, y) \in E^n \times E^n\colon g_i(x, y, t_0, T) \leqslant 0,\ i = 1, \ldots, m,\ g_i(x, y, t_0, T) = 0,\ i = m + 1, \ldots, s\}, \tag{25}$$

где $g_i(x, y, t, \mathrm{T})$ — заданные функции переменных $(x, y, t, \mathrm{T}) \in E^n \times E^n \times \Theta_0 \times \Theta_1$. Понятно, что множества (21) являются частным случаем множества (24), когда $S(t_0, T) = S_0(t_0) \times S_1(T)$; множества (22), (23) — частный случай (25).

5. Теперь перейдем к непосредственной формулировке задачи оптимального управления. Пусть заданы множества Θ_0, Θ_1 на числовой оси $\mathbb{R}$, $\inf \Theta_0 < \sup \Theta_1$; $V(t) \subseteq E^r$, $G(t) \subset E^n$ при всех t, $\inf \Theta_0 \leqslant t \leqslant \sup \Theta_1$; $S(t_0, T)$, $t_0 \in \Theta_0$, $T \in \Theta_1$. Пусть движение фазовой точки $x = (x^1, \ldots, x^n)$ описывается системой обыкновенных дифференциальных уравнений (7), где функция $f(x, u, t)$ определена при $x \in G(t)$, $u \in V(t)$, $t \in [t_0, T]$.

Набор $(x_0, u(\cdot), x(\cdot), t_0, T)$ назовем *допустимым* или *допустимым процессом*, если $t_0 \in \Theta_0$, $T \in \Theta_1$, $t_0 \leqslant T$, управление $u = u(\cdot) = (u^1(t), \ldots, u^r(t))$ определено и кусочно-непрерывно на отрезке $[t_0, T]$ и удовлетворяет ограничению (10) на этом отрезке, а $x = x(\cdot) = x(\cdot, u(\cdot), x_0)$ — траектория задачи (8), (9) (см. определение 1), которая определена на отрезке $[t_0, T]$ и удовлетворяет фазовому ограничению (19), $(x(t_0) = x_0, x(T)) \in S(t_0, T)$. Будем предполагать, что множество допустимых наборов (процессов) непусто.

Пусть на множестве допустимых наборов задана функция (или, как часто говорят, целевая функция или функционал)

$$J(x_0, u(\cdot), x(\cdot), t_0, T) = \int_{t_0}^{T} f^0(x(t), u(t), t)dt + g^0(x_0, x(T), t_0, T), \tag{26}$$

где $f^0(x, u, t)$, $g^0(x, y, t, T)$ — заданные функции при $x, y \in E^n$, $u \in V(t)$, $\inf \Theta_0 \leqslant t \leqslant \sup \Theta_1$, $T \in \Theta_1$.

Задача оптимального управления заключается в том, чтобы минимизировать или максимизировать функцию (26) на множестве допустимых наборов. Мы ограничимся рассмотрением лишь задач минимизации, так как задача минимизации J всегда может быть сведена к эквивалентной задаче минимизации $(-J)$.

Обозначим

$$J_* = \inf J(x_0, u(\cdot), x(\cdot), t_0, T),$$

где нижняя грань берется по всем допустимым наборам. Допустимый набор $(x_{0*}, u_*(\cdot), x_*(\cdot), t_{0*}, T_*)$ назовем *оптимальным процессом* или *решением задачи оптимального управления*, $u_*(\cdot)$ — *оптимальным управлением*, $x_*(\cdot)$ — *оптимальной траекторией*, если $J(x_{0*}, u_*(\cdot), x_*(\cdot), t_{0*}, T_*) = J_*$.

Сформулированную задачу оптимального управления можно записать в следующем кратком виде:

$$J(x_0, u(\cdot), x(\cdot), t_0, T) = \int_{t_0}^{T} f^0(x(t), u(t), t)dt + g^0(x_0, x(T), t_0, T) \to \inf, \quad (27)$$

$$\dot{x}(t) = f(x(t), u(t), t), \quad t_0 \leqslant t \leqslant T, \quad (28)$$

$$(x(t_0) = x_0,\ x(T)) \in S(t_0, T), \quad t_0 \in \Theta_0, \quad T \in \Theta_1, \quad (29)$$

$$x(t) \in G(t), \quad t_0 \leqslant t \leqslant T, \quad (30)$$

$$u(t) \in V(t), \quad (31)$$

подразумевая (если не оговорено другое), что управление $u = u(\cdot)$ — кусочно-непрерывно на отрезке $[t_0, T]$ и условие (31) выполняется во всех точках непрерывности управления $u(\cdot)$. Нетрудно видеть, что значение кусочно-непрерывного управления в точках разрыва не влияют ни на решение уравнения (28) (см. определение 1), ни на значение функции (27) и, следовательно, на задачу (27)–(31) в целом. Поэтому в задаче (27)–(31) не так важно, как определено управление $u(\cdot)$ в точках разрыва и выполняется ли в этих точках включение (31).

Заметим, что в задаче (27)–(31) управления можно брать также из класса ограниченных измеримых функций; в этом случае изменение управления $u(\cdot)$ на множестве меры нуль не влияет на задачу и поэтому естественно считать, что включение (31) выполняется для почти всех $t \in [t_0, T]$.

Если $f^0 \equiv 1$, $g^0 \equiv 0$, то $J(x_0, u(\cdot), x(\cdot), t_0, T) \equiv T - t_0$ — в этом случае задачу (27)–(31) называют *задачей быстродействия*. Если $f^i(x, u, t) \equiv f^i(x, u)$, $i = 0, \ldots, n$, $g^0(x, y, t, T) \equiv g^0(x, y)$, множества $S(t_0, T)$, $V(t)$, $G(t)$ не зависят от времени, то задачу (27)–(31) называют *автономной* или *стационарной*.

Если начальный момент закреплен, т. е. $\Theta_0 = \{t_0\}$, то в формулировке задачи (27)–(31) включение $t_0 \in \Theta_0$ опускают, вместо $J(x_0, u(\cdot), x(\cdot), t_0, T)$ пишут короче: $J(x_0, u(\cdot), x(\cdot), T)$ или $J(x_0, u(\cdot), T)$, вместо $S(t_0, T)$ пишут $S(T)$. Аналогично поступают, если закреплены конечный момент T или один из концов траектории. В том случае, когда $G(t) \equiv E^n$ при всех t или $S(t_0, T) = S_0(t_0) \times S_1(T)$, где $S_0(t) \equiv E^n$, или $S_1(t) \equiv E^n$, или $V(t) \equiv E^n$, то соответствующие из условий (29), (30), (31) в постановке задачи (27)–(31) также явно не указывают.

В приложениях встречаются задачи оптимального управления более общего вида, чем задача (27)–(31). Возможны ситуации, когда наряду с ограничениями на управления и фазовые координаты, записанными, так сказать, в разделенном виде (30), (31), имеются более сложные ограничения вида

$$(u(t), x(t)) \in W(t), \quad t_0 \leqslant t \leqslant T, \quad (32)$$

где $W(t)$ — заданное множество из $E^r \times E^n$, например, вида $W(t) = \{(u, x): R^i(u, x, t) \leqslant 0,\ i = 1, \ldots, m_4;\ R^i(u, x, t) = 0,\ i = m_4 + 1, \ldots, s_4\}$, $t_0 \leqslant t \leqslant T$, $R^i(u, x, t)$ — известные функции. Такие ограничения называются *смешанными*. Ограничения (30)–(32) накладываются на значения функций $u(t)$, $x(t)$ в каждой точке t, поэтому их можно назвать *точечными ограничениями*. Наряду с точечными ограничениями возможны также ограничения вида

$$J_i(x_0, u(\cdot), x(\cdot), t_0, T) \leqslant 0, \quad i = 1, \ldots, m_5,$$
$$J_i(x_0, u(\cdot), x(\cdot), t_0, T) = 0, \quad i = m + 1, \ldots, s_5, \quad (33)$$

где

$$J_i(x_0, u(\cdot), x(\cdot), t_0, T) = \int_{t_0}^{T} f_i^0(x(t), u(t), t)dt + g_i^0(x_0, x(T), t_0, T),$$

$f_i^0(x, u, t)$, $g_i^0(x, y, t, T)$, $i = 1, \ldots, s_2$ — заданные функции. Ограничения (33) накладываются на функции $u(\cdot)$, $x(\cdot) = x(\cdot, u(\cdot), x_0)$ в целом, поэтому их, в отличие от точечных, можно назвать *интегральными ограничениями*. Кстати, заметим, что ограничения вида (11) относятся к интегральным.

В теории оптимального управления рассматриваются также задачи, учитывающие запаздывание информации, задачи с параметрами, с дискретным временем, с более общим видом целевой функции, задачи для интегро-дифференциальных уравнений, для уравнений с частными производными, для стохастических уравнений и др. Важнейшим обобщением задач оптимального управления являются дифференциальные игры, описывающие конфликтно-управляемые системы, управляемые системы в условиях неопределенности. При исследовании управляемых систем наряду с задачами оптимизации описанного выше типа рассматриваются также и другие важные проблемы, такие, как управляемость, наблюдаемость, синтез, инвариантность, чувствительность, устойчивость, стабилизация, идентификация, фильтрация и т. д. У нас здесь нет возможности хотя бы бегло остановиться на перечисленных аспектах теории управляемых систем, и по этим вопросам мы отсылаем читателя к литературе, упомянутой во введении к настоящей главе.

§ 2. Формулировка принципа максимума. Примеры

1. Начнем с рассмотрения задачи оптимального управления с закрепленным временем. А именно, пусть требуется минимизировать функцию

$$J(x_0, u(\cdot), x(\cdot)) = \int_{t_0}^{T} f^0(x(t), u(t), t)dt + g^0(x_0, x(T)) \quad (1)$$

при условиях

$$\dot{x}(t) = f(x(t), u(t), t), \quad t_0 \leqslant t \leqslant T; \quad x(t_0) = x_0, \quad (2)$$

$$g^i(x_0, x(T)) \leqslant 0, \ i = 1, \ldots, m, \ g^i(x_0, x(T)) = 0, \ i = m + 1, \ldots, s, \quad (3)$$

$$u = u(t) \in V, \quad (4)$$

где моменты t_0, T предполагаются заданными, $u = (u^1, \ldots, u^r)$, $x = (x^1, \ldots, x^n)$, $f = (f^1, \ldots, f^n)$; управление $u = u(\cdot)$ является кусочно-непрерывной функцией на отрезке $[t_0, T]$ и удовлетворяет условию (4) во всех точках непрерывности; $f^i(x, u, t)$, $g^i(x, y)$ — заданные функции. Подчеркнем, что в задаче (1) — (4) множество $V \subset E^r$ не зависит от времени и фазовые ограничения при $t_0 < t < T$ отсутствуют. Очевидно, задача (1)–(4) является частным случаем задачи (1.27)–(1.31). В (3) не исключаются возможности, когда отсутствуют ограничения типа неравенств ($m = 0$), типа равенств ($s = m \geqslant 1$) или все ограничения (3) ($s = m = 0$). Предполагается,

что функции $f^j(x, u, t)$, $j=0, \ldots, n$, $g^j(x, y)$, $j=0, \ldots, s$, имеют частные производные $\partial f^j/\partial x^i = f^j_{x^i}$, $\partial g^j/\partial x^i = g^j_{x^i}$, $\partial g^j/\partial y^i = g^j_{y^i}$, $i=1, \ldots, n$. Обозначим $f^j_x = (f^j_{x^1}, \ldots, f^j_{x^n})$, $g^j_x = (g^j_{x^1}, \ldots, g^j_{x^n})$, $g^j_y = (g^j_{y^1}, \ldots, g^j_{y^n})$. Для формулировки принципа максимума введем функции:

$$H(x, u, t, \psi, a_0) = -a_0 f^0(x, u, t) + \psi_1 f^1(x, u, t) + \ldots + \psi_n f^n(x, u, t) = \\ = -a_0 f^0(x, u, t) + \langle \psi, f(x, u, t) \rangle, \quad (5)$$

$$l(x, y, a) = a_0 g^0(x, y) + a_1 g^1(x, y) + \ldots + a_s g^s(x, y);$$

здесь $\psi = (\psi_1, \ldots, \psi_n)$, $a = (a_0, a_1, \ldots, a_s)$ — вспомогательные переменные, определяемые ниже. Функцию $H(x, u, t, \psi, a_0)$ называют функцией *Гамильтона — Понтрягина*, а функцию $l(x, y, a)$ — *малым лагранжианом*.

Пусть $u = u(t)$ — кусочно-непрерывное управление на отрезке $[t_0, T]$, $x(t) = x(t, u, x_0)$ — решение задачи (2), соответствующее этому управлению $u = u(\cdot)$ и начальному условию x_0 и определенное на всем отрезке $[t_0, T]$. Паре $(u(t), x(t))$, $t_0 \leqslant t \leqslant T$, поставим в соответствие следующую систему линейных дифференциальных уравнений относительно переменных $\psi = \psi(t) = (\psi_1(t), \ldots, \psi_n(t))$:

$$\dot{\psi}_i(t) = -\frac{\partial H(x, u, t, \psi(t), a_0)}{\partial x^i}\Big|_{u=u(t), x=x(t)} = \\ = a_0 f^0_{x^i}(x(t), u(t), t) - \sum_{j=1}^{n} \psi_j(t) f^j_{x^i}(x(t), u(t), t), \quad t_0 \leqslant t \leqslant T, \quad (6)$$

называемую *сопряженной* системой. Систему (6) можно записать в векторной форме

$$\dot{\psi}(t) = -H_x(x(t), u(t), t, \psi(t), a_0), \quad t_0 \leqslant t \leqslant T, \quad (7)$$

где $H_x = (H_{x^1}, \ldots, H_{x^n})$. Подчеркнем, что в (6), (7) и всюду ниже запись вида $H_{x^i}(x(t), u(t), t, \psi(t), a_0)$, $l_{x^i}(x_0, x(T), a)$, $l_{y^i}(x_0, x(T), a)$, $g^j_{x^i}(x_0, x(T))$, $g^j_{y^i}(x_0, x(T))$, как это обычно принято, означает, что сначала вычисляется соответствующая частная производная функций $H(x, u, t, \psi, a_0)$, $g^j(x, y)$ и затем вместо аргументов подставляются их конкретные значения.

Если система (2) линейна относительно x, u, т. е.

$$\dot{x}(t) = A(t)x(t) + B(t)u(t) + f(t), \quad t_0 \leqslant t \leqslant T$$

(см. обозначения в (1.18)), то $H(x, u, t, \psi, a_0) = -a_0 f^0(x, u, t) + \langle \psi, A(t)x + B(t)u + f(t) \rangle$ и сопряженная система (7) может быть записана и виде

$$\dot{\psi}(t) = a_0 f^0_x(x(t), u(t), t) - (A(t))^\top \psi(t), \quad t_0 \leqslant t \leqslant T,$$

где $(A(t))^\top$ — матрица, полученная транспонированием матрицы $A(t)$.

Из теоремы 1.2 следует, что если зафиксировать постоянную a_0, момент времени t_1, $t_0 \leqslant t_1 \leqslant T$, и точку $\psi_0 \in E^n$, то линейная система (7) будет иметь и притом единственное решение $\psi(t) = \psi(t, u, x_0, \psi_0, t_1)$, удовлетворяющее условию $\psi(t_1) = \psi_0$ и определенное на всем отрезке $[t_0, T]$.

Теперь можем перейти к формулировке теоремы, выражающей необходимое условие оптимальности — принцип максимума для задачи (1)–(4).

Теорема 1. *Пусть функции $f^j(x, u, t)$, $j=0, \ldots, n$, $g^j(x, y)$, $j=0, \ldots, s$, имеют частные производные $f^j_{x^i}$, $g^j_{x^i}$, $g^j_{y^i}$, $i=1, \ldots, n$ и непрерывны вместе с этими производными по совокупности своих аргументов при $x \in E^n$, $y \in E^n$, $u \in \overline{V}$, $t \in [t_0, T]$, где $\overline{V}$ — замыкание множества V. Пусть $(x_0, u(t), x(t))$, $t_0 \leqslant t \leqslant T$, — решение задачи (1), (4). Тогда необходимо существуют числа $a_0, a_1, \ldots, a_s$ и вектор-функция $\psi(t) = (\psi_1(t), \ldots, \psi_n(t))$, $t_0 \leqslant t \leqslant T$, такие, что*

$$1) \quad a = (a_0, a_1, \ldots, a_s) \neq 0, \quad a_0 \geqslant 0, \ a_1 \geqslant 0, \ldots, a_m \geqslant 0; \quad (8)$$

2) $\psi(t)$ является решением сопряженной системы (6), соответствующей рассматриваемому решению $(x_0, u(\cdot), x(\cdot))$;

3) для всех $t \in [t_0, T]$, являющихся точками непрерывности оптимального управления $u(\cdot)$, функция $H(x(t), u, t, \psi(t), a_0)$ переменной $u = (u^1, \ldots, u^r)$ достигает своей верхней грани на множестве V при $u = u(t)$, т. е.

$$\max_{u \in V} H(x(t), u, t, \psi(t), a_0) = H(x(t), u(t), t, \psi(t), a_0); \quad (9)$$

4) выполнены условия

$$\psi_i(t_0) = l_{x^i}(x_0, x(T), a) = \sum_{j=0}^{s} a_j g^j_{x^i}(x_0, x(T)) = \frac{\partial l(x_0, x(T), a)}{\partial x^i}, \\ \psi_i(T) = l_{y^i}(x_0, x(T), a) = -\sum_{j=0}^{s} a_j g^j_{y^i}(x_0, x(T)) = -\frac{\partial l(x_0, x(T), a)}{\partial y^i}, \quad i=1, \ldots, n, \quad (10)$$

$$a_j g^j(x_0, x(T)) = 0, \quad j = 1, \ldots, m. \quad (11)$$

Условия (10) называют *условием трансверсальности*, условие (11) — *условием дополняющей нежесткости*. В том случае, когда управление $u(\cdot)$ является ограниченной измеримой функцией, т. е. $u(\cdot) \in L^r_\infty[t_0, T]$, то формулировка теоремы 1 полностью сохраняется, но равенство (9) (как и включение (4)) будет выполняться почти всюду на отрезке $[t_0, T]$.

Центральное место в теореме 1 занимает условие максимума (9): оказывается, если $u(\cdot)$ — оптимальное управление, а $x(\cdot)$ — оптимальная траектория, то непременно найдутся такие числа $a_0, a_1, \ldots, a_s$ и такое решение $\psi(t)$ системы (6), (10), что функция $H(x(t), u, t, \psi(t), a_0)$ переменной u будет достигать своего максимума на V именно при $u = u(t)$ во всех точках $t \in [t_0, T]$ непрерывности управления $u(\cdot)$. Поэтому теорему 1 и нижеследующую теорему 2, дающие необходимое условие оптимальности, принято называть *принципом максимума*.

2. Однако как практически пользоваться теоремой 1 для поиска решения задачи (1)–(4)? Здесь обычно поступают следующим образом. Рассматривают функцию $H(x, u, t, \psi, a_0)$ как функцию r переменных $u = (u^1, \ldots, u^r) \in V$, считая остальные переменные (x, t, ψ, a_0) параметрами, и при каждом фиксированном наборе (x, t, ψ, a_0) решают задачу максимизации:

$$H(x, u, t, \psi, a_0) \to \sup, \quad u \in V. \quad (12)$$

Отсюда находят функцию

$$u = u(x, t, \psi, a_0) \in V, \quad (13)$$

на которой достигается верхняя грань в задаче (12), т. е.

$$H(x, u(x, t, \psi, a_0), t, \psi, a_0) = \sup_{u \in V} H(x, u, t, \psi, a_0). \quad (14)$$

Если исходная задача (1)–(4) имеет решение, то, как следует из (9), функция (13) определена на непустом множестве.

В ряде случаев функция (13) может быть выписана в явном виде. Например, если

$$f^j(x, u, t) = f_0^j(x, t) + \sum_{i=1}^{r} f_{1i}^j(x, t)u^i, \quad j = 0, \ldots, n,$$

$$V = \{u = (u^1, \ldots, u^r) \in E^r\colon\ \alpha_i \leqslant u^i \leqslant \beta_i, \quad i = 1, \ldots, r\},$$

где α_i, β_i — заданные числа, то

$$H(x, u, t, \psi, a_0) = -a_0 f_0^0(x, t) + \sum_{j=1}^{n} \psi_j f_0^j(x, t) + \sum_{i=1}^{r} \varphi_i(x, t, \psi, a_0)u^i;$$

здесь для краткости обозначено

$$\varphi_i(x, t, \psi, a_0) = -a_0 f_{1i}^0(x, t) + \sum_{j=1}^{n} \psi_j f_{1i}^j(x, t), \quad i = 1, \ldots, r.$$

Ясно, что решением задачи (12) тогда будет вектор-функция $u(x, t, \psi, a_0)$ с координатами

$$u^i = u^i(x, t, \psi, a_0) = \begin{cases} \beta_i, \text{ при } \varphi_i(x, t, \psi, a_0) > 0, \\ \alpha_i, \text{ при } \varphi_i(x, t, \psi, a_0) < 0, \end{cases} \quad i = 1, \ldots, r.$$

В частности, если $\alpha_i = -1$, $\beta_i = +1$, то $u^i = \operatorname{sign} \varphi_i(x, t, \psi, a_0)$, $i = 1, \ldots, r$. Полученная формула дает довольно много информации о структуре оптимального управления: можно ожидать, что i-я координата оптимального управления является ступенчатой функцией со значениями α_i или β_i, причем точки переключения определяются условием $\varphi_i(x, t, \psi, a_0) = 0$. Обратим внимание читателя на возможный особый случай, когда $\varphi_i(x(t), t, \psi(t), a_0) = 0$ на каком-либо промежутке $[\tau_1, \tau_2] \subset [t_0, T]$. В этом случае функция H не будет зависеть от u^i и из условия (9) не удастся извлечь никакой полезной информации об i-й координате управления $u(\cdot)$ при $t \in [\tau_1, \tau_2]$; источником информации об $u^i(\cdot)$ на этом особом участке $[\tau_1, \tau_2]$ является само равенство $\varphi_i(x(t), t, \psi(t), a_0) = 0$.

Если множество V имеет вид

$$V = \left\{u \in E^r\colon\ |u| = \left(\sum_{i=1}^{r} (u^i)^2\right)^{1/2} \leqslant R\right\},$$

то, пользуясь известным неравенством Коши — Буняковского, также нетрудно выписать функцию (13) в явном виде:

$$u(x, t, \psi, a_0) = \frac{\varphi(x, t, \psi, a_0)}{|\varphi(x, t, \psi, a_0)|} R, \quad \varphi = (\varphi_1, \ldots, \varphi_r).$$

Ряд задач, в которых удастся получить явное выражение для функции (13), приводятся ниже в примерах.

Допустим, что функция (13) нам уже известна. Тогда можем рассмотреть следующую систему из $2n$ дифференциальных уравнений

$$\begin{aligned} \dot{x} &= f(x, u(x, t, \psi, a_0), t), \\ \dot{\psi} &= -H_x(x, u(x, t, \psi, a_0), t, \psi, a_0), \quad t_0 \leqslant t \leqslant T, \end{aligned} \tag{15}$$

относительно неизвестных $x(\cdot)$, $\psi(\cdot)$. Как известно [376; 588; 694] общее решение системы (15) зависит, вообще говоря, от $2n$ произвольных числовых параметров (например, такими параметрами могли бы служить начальные условия $x(t_0)$, $\psi(t_0)$) и для определения этих параметров нам нужно иметь $2n$ условий. Кроме того, параметры $a_0, a_1, \ldots, a_s$, встречающиеся в теореме 1, также неизвестны и для их определения нужно еще $s+1$ условие. Таким образом, для определения $2n+s+1$ неизвестных числовых параметров нам нужно $2n+s+1$ условие. Где их взять? Оказывается, эти условия также могут быть извлечены из теоремы 1. А именно, условия трансверсальности (10) и дополняющей нежесткости (11) нам дают $2n+m$ уравнений; еще $s-m$ уравнений

$$g^j(x_0, x(T)) = 0, \quad j = m+1, \ldots, s \tag{16}$$

вытекают из условий (3). Для получения еще одного уравнения заметим, что функция $H(x, u, t, \psi, a_0)$, определенная согласно (5), линейна и однородна относительно переменных $\psi_1, \ldots, \psi_n, a_0$, т. е. $H(x, u, t, \alpha\psi, \alpha a_0) = \alpha H(x, u, t, \psi, a_0)$ при любых α. Отсюда и из условия (14) тогда имеем

$$u(x, t, \alpha\psi, \alpha a_0) \equiv u(x, t, \psi, a_0) \quad \forall \alpha > 0. \tag{17}$$

Из (8), (10), (11), (17) следует, что если некоторый набор $a_0, \ldots, a_s$, $\psi_1, \ldots, \psi_n$ удовлетворяет условиям теоремы 1, то этим условиям удовлетворяет также набор $\alpha a_0, \ldots, \alpha a_s$, $\alpha\psi_1, \ldots, \alpha\psi_n$ при любых $\alpha > 0$. Это означает, что теорема 1 определяет величины $\alpha a_0, \ldots, \alpha a_s$, $\alpha\psi_1, \ldots, \alpha\psi_n$ лишь с точностью до положительного множителя, и этим множителем мы можем распорядиться по своему усмотрению. Например, опираясь на первое из условий (8), можно положить

$$|a|^2 = \sum_{i=0}^{s} a_i^2 = 1. \tag{18}$$

В тех задачах, в которых удается показать, что $a_0 > 0$, вместо условия нормировки (18) часто берут $a_0 = 1$.

Таким образом, для определения $2n+s+1$ параметров — $2n$ параметров общего решения системы (15) и параметров $a_0, a_1, \ldots, a_s$ — у нас имеется система $2n+s+1$ уравнений (10), (11), (16), (18). Разумеется, эти уравнения надо решать совместно с неравенствами

$$a_0 \geqslant 0, \ldots, a_m \geqslant 0, \quad g_i(x_0, x(T)) \leqslant 0, \quad i = 1, \ldots, m. \tag{19}$$

Если исходная задача (1)–(4) имеет решение, то согласно теореме 1 система (10), (11), (16), (18), (19) также имеет решение. Попутно заметим, что для тех i, $1 \leqslant i \leqslant m$, для которых $g_i(x_0, x(T)) < 0$ (неактивные ограничения), из (11) вытекает, что $a_i = 0$, и неопределенными остаются лишь a_i, с номерами i, для которых $g_i(x_0, x(T)) = 0$ (активные ограничения). Это означает, что из уравнений (11) существенное значение имеют лишь подсистемы $g_i(x_0, x(T)) = 0$, $i \in I$, состоящие из активных ограничений.

Итак, основываясь на теореме 1, от исходной задачи (1)–(4) мы пришли к специальной краевой задаче, состоящей из условия максимума (14), системы дифференциальных уравнений (15) и условий (10), (11), (16), (18), (19). Такую краевую задачу естественно назвать *краевой задачей принципа максимума* для задачи оптимального управления (1)–(4).

Можно ожидать, что имеются лишь отдельные, изолированные функции $(x(t), \psi(t))$, $t_0 \leqslant t \leqslant T$, и значения параметров $a_0, a_1, \ldots, a_s$, удовлетворяющие условиям (10), (11), (15), (16), (18), (19). Возьмем один из таких наборов $x(t)$, $\psi(t)$, $a_0, a_1, \ldots, a_s$, и подставим их в (13); получим функцию

$$u(t) = u(x(t), t, \psi(t), a_0), \quad t_0 \leqslant t \leqslant T. \tag{20}$$

Пусть эта функция оказалась кусочно-непрерывной на $[t_0, T]$. Из (13), (14), (20) тогда следует, что полученное таким образом управление $u(t)$, $t_0 \leqslant t \leqslant T$, удовлетворяет условию (9) и, следовательно, согласно теореме 1 может претендовать на роль оптимального управления задачи (1)–(4), а функция $x(t) = x(t, u(\cdot), x(t_0))$, $t_0 \leqslant t \leqslant T$, — на роль оптимальной траектории этой задачи. Будет ли найденная пара $(u(t), x(t))$, $t_0 \leqslant t \leqslant T$, в самом деле решением задачи (1)–(4), теорема 1 не гарантирует, так как эта теорема, вообще говоря, дает лишь необходимое условие оптимальности. Более того, ниже на примерах мы увидим, что бывают случаи, когда пара $(u(t), x(t))$ удовлетворяет условиям теоремы 1, но не является решением задачи (1)–(4). Однако, если из каких-либо соображений известно, что задача (1)–(4) имеет решение, а краевая задача принципа максимума однозначно определяет функции $(x(t), \psi(t))$ и параметры $a_0, a_1, \ldots, a_s$, то управление (20) будет оптимальным. Если информации о существовании решения задачи (1)–(4) нет или краевая задача принципа максимума имеет несколько решений, то для выяснения вопроса об оптимальности получаемых здесь управлений требуется дополнительное и порою весьма сложное исследование.

Как найти хотя бы одно решение краевой задачи принципа максимума? Напрашивается следующий подход: сначала попытаться определить решение системы (15) на отдельных отрезках времени, в пределах которых управление (13) является непрерывным, затем из полученных «кусков» составить («склеить») решение на всем отрезке протекания процесса, добиваясь удовлетворения краевых условий и условий на параметры. На этом пути, как увидим ниже на примерах, нередко удается найти решение краевой задачи принципа максимума и оптимальное управление исходной задачи (1)–(4). Полученное решение весьма информативно, наглядно, помогает лучше понять специфику задач оптимального управления. Примеры таких задач, в которых удается найти решение в аналитической форме, желательны как в теоретическом, так и в прикладном отношениях.

К сожалению, получить решение краевой задачи принципа максимума в аналитической форме можно лишь в частных случаях для задач со специальной структурой (линейность управляемой системы по фазовым переменным, малая размерность задачи, простые ограничения на управление и т. п.). В общем случае, для задач повышенной сложности речь можно вести только о численном решении. Однако и на этом пути имеются серьезные трудности, связанные прежде всего с разрывным характером функции (13) (выше мы уже видели, что даже в простейших случаях эта функция может иметь разрывы с неопределенностью на многообразиях разрыва). Это значит, что система (15) является, вообще говоря, разрывной по фазовым и сопряженным переменным, что делает проблематичным использование традиционных методов [59; 74; 89; 481; 635]. Поэтому для численного решения задач оптимального управления приходится разрабатывать специальные методы с учетом специфики таких задач (см., например [56; 57; 134; 135; 206; 282; 582; 653–656; 719; 753; 811; 818]; один из методов будет изложен в следующей главе).

Следует предупредить читателя, что краевая задача принципа максимума весьма своеобразна и таит в себе немало опасностей. Прежде всего, возможно вырождение принципа максимума, когда на отдельных участках времени функция $H(x, u, t, \psi, a_0)$ не зависит от переменной u, что приводит к появлению особых управлений, требующих отдельного исследования.

Кроме того, существуют задачи оптимального управления, которые с виду формулируются достаточно просто и полностью укладываются в схему задачи (1)–(4), но оптимальное управление в них имеет бесконечное число переключений на конечном отрезке времени (так называемые четтеринг-режимы). Это значит, что задача (1)–(4) не всегда имеет решение в классе кусочно-непрерывных управлений. Более того, задача (1)–(4) может не иметь решение даже в классе измеримых управлений, что приводит к необходимости использования более широких классов допустимых управлений (скользящие режимы, импульсные управления). Не имея возможности обсуждать возникающие здесь тонкие и интересные проблемы (укажем лишь на приводимые ниже примеры 6, 7.4.2, упражнения 7, 9, 16), отсылаем читателя к специальной литературе [57; 72; 108; 132; 199; 207; 212; 253; 254; 322; 323; 325; 417; 418; 656; 716; 717; 723; 787; 819].

3. Посмотрим, как выглядит краевая задача принципа максимума для задач оптимального управления (1)–(4) для некоторых конкретных классов функций $g^i(x, y)$, соответствующих различным режимам на левом и правом концах траектории. Начнем с рассмотрения случая, когда концы траектории закреплены:

$$x(t_0) = x_0, \quad x(T) = x_1. \tag{21}$$

Если положить $g^i(x, y) = x^i - x_0^i$, $i = 1, \ldots, n$, $g^i(x, y) = y^i - x_1^i$, $i = n+1, \ldots, 2n$, то условия (21) запишутся в виде (3): $g^i(x_0, x(T)) = 0$, $i = 1, \ldots, 2n = s$, $m = 0$. Поэтому условия (11) здесь отсутствуют, а условия трансверсальности (10) дадут

$$\begin{aligned} \psi(t_0) &= \sum_{j=0}^{2n} a_j g_x^j(x_0, x(T)) = a_0 g_x^0(x_0, x(T)) + (a_1, \ldots, a_n), \\ \psi(T) &= -\sum_{j=0}^{2n} a_j g_y^j(x_0, x(T)) = -a_0 g_y^0(x_0, x(T)) - (a_{n+1}, \ldots, a_{2n}). \end{aligned} \tag{22}$$

Оказывается, условие $a \neq 0$ из (8) здесь может быть заменено условием

$$|a_0| + |\psi(t)| \neq 0 \quad \forall t \in [t_0, T]. \tag{23}$$

В самом деле, если (23) не выполняется, то $a_0 = 0$, $\psi(t) \equiv 0$, $t_0 \leqslant t \leqslant T$. А тогда в силу (22) $\psi(t_0) = 0 = (a_1, \ldots, a_n)$, $\psi(T) = 0 = (a_{n+1}, \ldots, a_{2n})$ и, следовательно, $a = (a_0, a_1, \ldots, a_{2n}) = 0$, что противоречит (8). Это значит, что для задачи (1), (2), (4), (21) выполняется условие (23). Тогда условие нормировки (18) можно заменить условием

$$|a_0| + |\psi(t_1)| = 1, \tag{24}$$

где t_1 — какая-либо подходящая точка из отрезка $[t_0, T]$ (часто берут $t_1 = t_0$ или $t_1 = T$). Таким образом, краевая задача принципа максимума для задачи (1), (2), (4), (21) состоит из системы (15), граничных условий (21), (22), неравенства $a_0 \geqslant 0$ и условия нормировки (24). Так как неизвестные параметры $a_1, \ldots, a_{2n}$ входят лишь в условие трансверсальности (22) и не входят в (15), (21), (24), то эти параметры и условия (22) можно исключить из дальнейшего рассмотрения. В итоге, для определения функций $(x(t), \psi(t))$, $t_0 \leqslant t \leqslant T$, и параметра $a_0 \geqslant 0$ имеем краевую задачу принципа максимума, состоящую из системы $2n$ дифференциальных уравнений (15), $2n$ граничных условий (21) и условия нормировки (24). Конечно, при необходимости исключенные параметры $a_1, \ldots, a_{2n}$ могут быть определены из условия (22) после того, как уже будут найдены $x(t), \psi(t), a_0$ из (15), (21), (24).

Теперь рассмотрим задачу (1), (2), (4) при условиях, когда левый конец траектории закреплен: $x(t_0)=x_0$, а правый конец свободный. Этот случай граничных режимов соответствует задаче (1)–(4), в которой $g^i(x,y)=x^i-x_0^i$, $i=1,\ldots,n=s$, $m=0$. Поэтому условия (11) здесь отсутствуют, а условия трансверсальности (10) запишутся в виде:

$$\psi(t_0)=a_0 g_x^0(x_0,x(T))+(a_1,\ldots,a_n),\quad \psi(T)=-a_0 g_y^0(x_0,x(T)). \tag{25}$$

Покажем, что в рассматриваемом случае также выполняется условие (23) и, более того, можно гарантировать, что $a_0>0$. В самом деле, если $a_0=0$, то, как видно из (25), $\psi(T)=0$ и система однородных уравнений (6) будет иметь лишь тривиальное решение $\psi(t)\equiv 0$, а тогда $\psi(t_0)=0=(a_1,\ldots,a_n)$. Пришли к противоречию с первым условием (8): $a=(a_0,a_1,\ldots,a_n)\neq 0$. Следовательно, $a_0>0$, и можем принять условие нормировки $a_0=1$. Таким образом, краевая задача принципа максимума для задачи (1), (2), (4) с закрепленным левым концом и свободным правым концом состоит из системы (15), граничных условий (25), $x(t_0)=x_0$ и условия нормировки $a_0=1$. Так как параметры $a_1,\ldots,a_n$ входят лишь в первое из условий (25), то это условие и параметры $a_1,\ldots,a_n$ можно исключить из дальнейшего рассмотрения, и в результате краевая задача принципа максимума сведется к системе $2n$ дифференциальных уравнений (15), которая решается при условиях

$$x(t_0)=x_0,\quad \psi(T)=-g_y^0(x_0,x(T)),\quad a_0=1. \tag{26}$$

Аналогичные рассуждения показывают, что если в задаче (1), (2), (4) левый конец траектории свободный, правый конец закреплен, то соответствующая краевая задача принципа максимума представляет собой систему (15), которая должна решаться при условиях

$$x(T)=x_1,\quad \psi(t_0)=g_x^0(x_0,x(T)),\quad a_0=1. \tag{27}$$

Если в задаче (1), (2), (4) оба конца траектории свободны, то краевая задача принципа максимума состоит из системы (15) и условий

$$\psi(t_0)=g_x^0(x_0,x(T)),\quad \psi(T)=-g_y^0(x_0,x(T)),\quad a_0=1. \tag{28}$$

Далее, рассмотрим задачу (1), (2), (4) в случае, когда левый конец траектории закреплен: $x(t_0)=x_0$, а правый конец подвижный и удовлетворяет условиям

$$g^i(x(T))\leqslant 0,\quad i=1,\ldots,m_1;\quad g^i(x(T))=0,\quad i=m_1+1,\ldots,s_1. \tag{29}$$

Здесь мы имеем дело с задачей (1)–(4), в которой функции $g^i(x,y)$ определены так: $g^i(x,y)=g^i(y)$, $i=1,\ldots,s_1$; $g^i(x,y)=x^i-x_0^i$, $i=s_1+1,\ldots,s_1+n=s$, $m=m_1$. Условия (10), (11) на правом конце траектории с учетом (8) запишутся в виде

$$\psi(T)=-a_0 g_y^0(x_0,x(T))-\sum_{j=1}^{s_1}a_j g_y^j(x(T)), \tag{30}$$

$$a_j g^j(x(T))=0,\quad a_j\geqslant 0,\quad j=1,\ldots,m_1;\quad a_0\geqslant 0,$$

на левом конце — в виде:

$$\psi(t_0)=a_0 g_x^0(x_0,x(T))+(a_{s_1+1},\ldots,a_{s_1+n}). \tag{31}$$

Условие $a=(a_0,a_1,\ldots,a_{s_1+n})\neq 0$ здесь может быть заменено условием $(a_0,$

$a_1,\ldots,a_{s_1})\neq 0$. В самом деле, если бы $(a_0,a_1,\ldots,a_{s_1})=0$, то в силу (30) $\psi(T)=0$, и однородная система (6) будет иметь тривиальное решение $\psi(T)\equiv 0$; тогда $\psi(t_0)=0$ и из (31) будет следовать $(a_{s_1+1},\ldots,a_{s_1+n})=0$, что противоречит условию $a\neq 0$. Поэтому условие нормировки (18) можно заменить на

$$a_0^2+a_1^2+\ldots+a_{s_1}^2=1, \tag{32}$$

а условие (31) и параметры $a_{s_1+1},\ldots,a_{s_1+n}$ исключить из рассмотрения. В результате, краевая задача принципа максимума будет состоять из системы (15), начального условия $x(t_0)=x_0$, условий (29), (30), (32).

Аналогично показывается, что в задаче (1), (2), (4), когда левый конец траектории свободный, а на правом конце заданы условия (29), краевая задача принципа максимума будет состоять из системы (15), условий (29), (30) на правом конце, условия нормировки (32) и условия на левом конце

$$\psi(t_0)=a_0 g_x^0(x_0,x(T)). \tag{33}$$

Рассмотрим задачу (1), (2), (4), когда оба конца траектории подвижны, причем условия на правом конце задаются в виде (29), на левом конце пусть

$$h^i(x(t_0))\leqslant 0,\quad i=1,\ldots,m_0;\quad h^i(x(t_0))=0,\quad i=m_0+1,\ldots,s_0. \tag{34}$$

Здесь мы имеем дело с задачей (1)–(4), в которой функции $g^i(x,y)$ определены так: $g^i(x,y)=g^i(y)$, $i=1,\ldots,m_1$; $g^i(x,y)=h^{i-m_1}(x)$, $i=m_1+1,\ldots,m_1+m_0=m$, $g^i(x,y)=g^{i-m_0}(y)$, $i=m_1+m_0+1,\ldots,m_0+s_1$; $g^i(x,y)=h^{i-s_1}$, $i=m_0+s_1+1,\ldots,s_0+s_1=s$. Условия (10), (11) на правом конце траектории с учетом (8) запишутся в виде (30), а на левом конце получим

$$\psi(t_0)=a_0 g_x^0(x_0,x(T))+\sum_{j=1}^{s_0}b_j h_x^j(x_0),\quad b_j h^j(x_0)=0,\quad b_j\geqslant 0,\quad j=1,\ldots,m_0. \tag{35}$$

Таким образом, краевая задача принципа максимума, соответствующая задаче (1), (2), (4), (29), (34), состоит из системы (15), условий (29), (30), (34), (35), условия нормировки

$$a_0^2+a_1^2+\ldots+a_{s_1}^2+b_1^2+\ldots+b_{s_0}^2=1.$$

Если в (1), (2), (4) левый конец удовлетворяет условиям (34), а правый конец закрепленный, то систему (15) нужно решать при условиях (34), (35), $x(T)=x_1$, условии нормировки

$$a_0^2+b_1^2+\ldots+b_{s_0}^2=1. \tag{36}$$

Если в задаче (1), (2), (4) левый конец удовлетворяет условиям (34), а правый конец свободный, то система (15) решается при условиях (34)–(36), $\psi(T)=-a_0 g_y^0(x_0,x(T))$.

4. Сформулируем принцип максимума для задачи оптимального управления, когда начальный или конечный моменты времени не закреплены. А именно, пусть требуется минимизировать функцию

$$J(x_0,u(\cdot),x(\cdot),t_0,T)=\int_{t_0}^{T}f^0(x(t),u(t),t)dt+g^0(x_0,x(T),t_0,T) \tag{37}$$

при условиях

$$\dot{x}(t)=f(x(t),u(t),t),\quad t_0\leqslant t\leqslant T;\quad x(t_0)=x_0, \tag{38}$$

$$g^i(x_0,x(T),t_0,T)\leqslant 0,\ i=1,\ldots,m;\quad g^i(x_0,x(T),t_0,T)=0,\ i=m+1,\ldots,s, \tag{39}$$

$$u=u(t)\in V, \tag{40}$$

где один из моментов t_0 или T или оба эти момента заранее неизвестны и подлежат определению вместе с управлением $u(t)$ и траекторией $x(t)$ из условия минимума функции (37): $g^i(x, y, t, T)$, $i=0,\ldots,s$ — заданные функции переменных $x\in E^n$, $y\in E^n$, $t\in\mathbb{R}$, $T\in\mathbb{R}$, $t\leqslant T$; остальные обозначения те же, что и в задаче (1)–(4). В (39) не исключаются возможности, когда отсутствуют ограничения типа неравенств ($m=0$), типа равенств ($s=m\geqslant 1$) или все ограничения (39) ($s=m=0$). В этой задаче функция Гамильтона — Понтрягина $H(x, u, t, \psi, a_0)$ определяется также, как в (5), а функция $l(x, y, a)$ заменяется на $l(x, y, t, T, a)=\sum\limits_{i=0}^{s} a_i g^i(x, y, t, T)$.

Теорема 2. *Пусть функции $f^j(x, u, t)$, $j=0,\ldots,n$; $g^j(x, y, t, T)$, $j=0,\ldots,s$, имеют частные производные $f^j_{x^i}$, $g^j_{x^i}$, $g^j_{y^i}$, $i=1,\ldots,n$; g^j_t, g^j_T и непрерывны вместе с этими производными по совокупности своих аргументов при $x\in E^n$, $y\in E^n$, $u\in\overline{V}$, $t\in\mathbb{R}$, $T\in\mathbb{R}$, $t\leqslant T$. Пусть $(x_0, u(\cdot), x(\cdot), t_0, T)$ — решение задачи (37)—(40), управление $u(t)$, $t_0\leqslant t\leqslant T$, кусочно-непрерывно. Тогда необходимо существуют числа $a_0, a_1, \ldots, a_s$ и вектор-функция $\psi(t)=(\psi_1(t),\ldots,\psi_n(t))$, $t_0\leqslant t\leqslant T$, удовлетворяющие условиям 1)–3) теоремы 1, условиям трансверсальности*

$$\psi(t_0)=l_x(x_0, x(T), t_0, T, a)=\sum_{j=0}^{s} a_j g^j_x(x_0, x(T), t_0, T),$$

$$\psi(T)=l_y(x_0, x(T), t_0, T, a)=-\sum_{j=0}^{s} a_j g^j_y(x_0, x(T), t_0, T), \tag{41}$$

$$H(x(t_0), u(t_0+0), t_0, \psi(t_0), a_0)=$$

$$=-l_t(x_0, x(T), t_0, T, a)=-\sum_{j=0}^{s} a_j g^j_t(x_0, x(T), t_0, T) \tag{42}$$

(если t_0 закреплено, то условие (42) отсутствует);

$$H(x(T), u(T-0), T, \psi(T), a_0)=$$

$$=l_T(x_0, x(T), t_0, T, a)=\sum_{j=0}^{s} a_j g^j_T(x_0, x(T), t_0, T) \tag{43}$$

(если T закреплено, то условие (43) отсутствует) и условию дополняющей нежесткости

$$a_j g^j(x_0, x(T), t_0, T)=0,\qquad j=1,\ldots,m. \tag{44}$$

С помощью теоремы 2 задачу (37)–(40) можно свести к краевой задаче принципа максимума, действуя по той же схеме, что и в задаче (1)–(4). Это снова приведет нас к конечномерной задаче максимизации (12), откуда определяется функция (13), и к системе $2n$ дифференциальных уравнений (15) относительно функций $x(t)$, $\psi(t)$. Для определения $2n$ параметров, от которых зависит общее решение системы (15), и параметров $a_0, a_1, \ldots, a_s$ имеем $2n$ условий (41), m условий (44), $s-m$ условий типа равенств из (39), условие нормировки (18), а наличие неизвестных моментов t_0, T здесь компенсируется появлением дополнительных условий (42),

(43); разумеется, поиск упомянутых параметров нужно вести с учетом неравенств из (8), (39). Расшифровка условий трансверсальности (41), (42) для различных режимов на концах траекторий, когда каждый из концов может быть закрепленным, свободным или подвижным, проводится точно так же, как и выше; в частности, условия (41) здесь приведут к тем же условиям (21)–(36). Естественно, в задаче оптимального управления с незакрепленным временем функции g^i, h^i из условий (29), (34) наряду с x, y могут зависеть еще от переменных t_0, T, как например, в (1.22), (1.23).

5. Для иллюстрации теорем 1, 2 рассмотрим конкретные примеры задач оптимального управления.

Пример 1. Минимизировать функцию $J(u)=\int\limits_0^T (u^2(t)+x^2(t))dt$ при условиях $\dot{x}(t)=u(t)$, $0\leqslant t\leqslant T$, $x(0)=x(T)=0$.

Здесь момент $T>0$ задан; $V=E^1$. Задача, конечно, несложная: пара $(u(t)\equiv 0,\ x(t)\equiv 0)$, $0\leqslant t\leqslant T$, очевидно, является единственным ее решением. Продемонстрируем на этой простой задаче изложенную выше схему использования принципа максимума — теоремы 1. Выпишем функцию Гамильтона — Понтрягина $H(x, u, \psi, a_0)=-a_0(u^2+x^2)+\psi u$ и сопряженную систему $\dot{\psi}=-H_x=2a_0x$. Если $a_0=0$, то функция $H=\psi u$ может достигать своей верхней грани на множестве $V=E^1$ лишь при $\psi=0$. Однако соотношения $a_0=\psi=0$ противоречат условию (23). Следовательно, $a_0>0$. Тогда можем считать, что $a_0=1$. В этом случае функция $H=-u^2-x^2+\psi u$ достигает верхней грани на E^1 при $u=\psi/2$ — вот какой вид имеет функция (13) в рассматриваемой задаче. Тогда краевая задача принципа максимума запишется в виде

$$\dot{x}=\psi/2,\quad \dot{\psi}=2x,\quad 0\leqslant t\leqslant T;\quad x(0)=x(T)=0.$$

Отсюда однозначно определяем $x(t)\equiv\psi(t)\equiv 0$, $0\leqslant t\leqslant T$. Тогда $u(t)=\psi(t)/2\equiv 0$, $0\leqslant t\leqslant T$, — получили известное нам оптимальное управление.

Перейдем к рассмотрению более интересной задачи оптимального управления, которая в зависимости от величины конечного момента T имеет единственное решение или бесконечно много решений, или не имеет решения. Эта задача любопытна также и тем, что даже в том случае, когда она не имеет решения, краевая задача принципа максимума будет иметь одно или даже бесконечно много решений.

Пример 2. Пусть требуется минимизировать функцию $J(u)=\int\limits_0^T (u^2(t)-x^2(t))dt$ при условиях $\dot{x}(t)=u(t)$, $0\leqslant t\leqslant T$, $x(0)=x(T)=0$, $T>0$.

Функция Гамильтона — Понтрягина здесь имеет вид $H=-a_0(u^2-x^2)+\psi u$, сопряженная система такая: $\dot{\psi}=-H_x=-2a_0x$. Если $a_0=0$, то $H=\psi u$ достигает своей верхней грани на $V=E^1$ лишь при $\psi=0$, что противоречит условию (23). Следовательно, $a_0>0$. Можно считать, что $a_0=1$. Тогда $H=x^2-u^2+\psi u$ и $\sup\limits_{u\in E^1} H$ достигается при $u=\psi/2$. Краевая задача принципа максимума имеет вид $\dot{x}=\psi/2$, $\dot{\psi}=-2x$, $0\leqslant t\leqslant T$, $x(0)=x(T)=0$. Общее решение этой системы дается формулой $x(t)=C\sin t+D\cos t$, $\psi(t)=2C\cos t-2D\sin t$, где C, D — произвольные постоянные. С учетом условия $x(0)=0$ отсюда имеем $D=0$, и тогда $x(t)=C\sin t$, $\psi(t)=2C\cos t$. Условие $x(T)=0$ приводит к равенству $C\sin T=0$. Возможно, что $T\neq\pi k$, $k=1, 2, \ldots$; тогда $C=0$ и краевая задача принципа максимума будет иметь

единственное решение $x(t)\equiv 0$, $\psi(t)\equiv 0$, $0\leqslant t\leqslant T$, а управление, подозрительное на оптимальность, равно $u(t)=\psi(t)/2=0$, $0\leqslant t\leqslant T$. Если же $T=\pi k$, k — целое положительное число, то краевая задача принципа максимума имеет бесчисленное множество решений $x(t)=C\sin t$, $\psi(t)=2C\cos t$, зависящих от одного параметра C, и управлений, подозрительных на оптимальность, будет бесконечно много: $u(t)=C\cos t$, $0\leqslant t\leqslant T$.

Спрашивается, будут ли найденные управления оптимальными? Оказывается, ответ на этот вопрос зависит от величины T. Рассмотрим случаи $T>\pi$ и $0<T\leqslant\pi$.

1) $T>\pi$. Покажем, что тогда $\inf J(u)=-\infty$. Для этого возьмем последовательность управлений $u_m=u_m(t)=\frac{m\pi}{T}\cos\frac{\pi t}{T}$ и соответствующих им траекторий $x_m=x_m(t)=m\sin\frac{\pi t}{T}$, $0\leqslant t\leqslant T$, $m=1,2,\ldots$. Тогда $J(u_m)=\int\limits_0^T(u_m^2(t)-x_m^2(t))dt=\frac{1}{2}Tm^2\left(\frac{\pi^2}{T^2}-1\right)\to-\infty$ при $m\to\infty$. Следовательно, при $T>\pi$ рассматриваемая задача оптимального управления не имеет решения. В то же время краевая задача принципа максимума при всех $T>\pi$ разрешима, причем при $T=\pi k$, $k=2,3,\ldots$, она имеет бесконечно много решений, при остальных $T>\pi$ — единственное решение.

2) $0<T\leqslant\pi$. Тогда для любых кусочно-непрерывных $v(t)$, для которых существует решение $x(t)$ задачи $\dot x(t)=v(t)$, $0\leqslant t\leqslant T$, $x(0)=x(T)=0$, имеем

$$J(v)=\int\limits_0^T(v^2-x^2)dt=\int\limits_0^T(v^2+x^2\operatorname{ctg}^2 t-x^2\sin^{-2}t)dt=$$
$$=\int\limits_0^T(v^2+x^2\operatorname{ctg}^2 t-2x\dot x\operatorname{ctg}t)dt=\int\limits_0^T(v(t)-x(t)\operatorname{ctg}t)^2dt\geqslant 0.$$

Заметим, что проделанные преобразования законны, так как все подынтегральные функции, встретившиеся при этих преобразованиях, ограничены и $\lim\limits_{t\to+0}x^2(t)\operatorname{ctg}t=0$, а в случае $T=\pi$ еще и $\lim\limits_{t\to T-0}x^2(t)\operatorname{ctg}t=0$. Итак, $J(v)\geqslant 0$, а на управлениях $u(t)\equiv 0$ при $T<\pi$ и $u(t)=c\cos t$ при $T=\pi$ будем иметь $J(u)=0$. Таким образом, при $T<\pi$ рассматриваемая задача оптимального управления имеет единственное решение, при $T=\pi$ — бесчисленное множество решений, причем все решения найдены с помощью принципа максимума.

Пример 3. Минимизировать функцию $J(u)=\frac{1}{2}\int\limits_0^T(x^2(t)+u^2(t))dt$ при условиях $\dot x(t)=-ax(t)+u(t)$, $x(0)=x_0$.

Здесь x_0, $a>0$, $T>0$ — заданные постоянные, $V=E^1$, правый конец траектории свободный. Составим функцию Гамильтона — Понтрягина $H=-a_0(x^2+u^2)/2+\psi(-ax+u)$ и выпишем сопряженную систему

$$\dot\psi=-H_x=a_0x+a\psi,\quad 0\leqslant t\leqslant T.$$

Из условия (26) трансверсальности для свободного правого конца траектории имеем $\psi(T)=0$, $a_0=1$. Тогда функция $H=-(x^2+u^2)/2+\psi(-ax+u)$ достигает своей верхней грани по u на $V=E^1$ при $u=\psi$, и краевая задача принципа максимума запишется в виде

$$\dot x=-ax+\psi,\quad \dot\psi=a\psi+x,\quad x(0)=x_0,\quad \psi(T)=0.$$

Таким образом, подозрительным на оптимальность является управление

$$u(t)=\psi(t)=x_0\frac{e^{\lambda t}-e^{2\lambda T}e^{-\lambda T}}{(\lambda-a)+(\lambda+a)e^{2\lambda T}},\quad 0\leqslant t\leqslant T,\quad \lambda=\sqrt{a^2+1}$$

(см. пример 8.5.2).

Пример 4. Пусть точка движется по оси Ox по закону $\ddot x(t)=u(t)$, $t\geqslant 0$. Требуется найти кусочно-непрерывное управление $u(t)$, $|u(t)|\leqslant 1$, $0\leqslant t\leqslant T$, такое, чтобы точка, выйдя из начального положения $x(0)=1$ с нулевой скоростью, пришла в начало координат с нулевой скоростью за минимальное время T.

Положим, что $x^1=x$, $x^2=\dot x$ — фазовые координаты точки. Тогда задачу можно переформулировать так: быстрейшим образом перевести фазовую точку (x^1,x^2) из состояния $(1,0)$ в состояние $(0,0)$, считая, что движение подчиняется уравнениям $\dot x^1(t)=x^2(t)$, $\dot x^2(t)=u(t)$, $t\geqslant 0$. Здесь $V=\{u\in E^1\colon |u|\leqslant 1\}$, $f^0\equiv 1$, $g^0\equiv 0$.

Составим функцию $H=-a_0+\psi_1x^2+\psi_2u$ и выпишем сопряженную систему

$$\dot\psi_1=-H_{x^1}=0,\quad \dot\psi_2=-H_{x^2}=-\psi_1,\quad t\geqslant 0.$$

Отсюда имеем $\psi_1(t)=C$, $\psi_2(t)=-Ct+D$, где C, D — постоянные. Отметим, что $\psi_2(t)\not\equiv 0$, так как в противном случае $C=D=0$, а тогда $\psi_1(t)\equiv 0$ и равенство $H|_{t=T}=-a_0=0$, вытекающее из (43), приводит к противоречию с условием (23). Из условия $\max\limits_{|u|\leqslant 1}H$ следует $u(t)=\operatorname{sign}\psi_2(t)=\operatorname{sign}(-Ct+D)$, $t\geqslant 0$. Таким образом, оптимальное управление (если оно существует) является кусочно-постоянной функцией, принимающей значения $+1,-1$ и имеющей не более одной точки переключения t_1, при переходе через которую $u(t)$ меняет знак. Нетрудно убедиться, что траектория, выходящая из точки $(1,0)$ и соответствующая управлению $u(t)=+1$ при $t\geqslant 0$, или $u(t)\equiv -1$ при $t\geqslant 0$, или $u(t)=+1$, $0\leqslant t<t_1$, $u(t)=-1$, $t\geqslant t_1$, никогда не будет проходить через точку $(0,0)$. Остается рассмотреть управление $u(t)=-1$, $0\leqslant t<t_1$, $u(t)=+1$, $t\geqslant t_1$. Этому управлению соответствует траектория $(x^1(t),x^2(t))$:

$$x^1(t)=\begin{cases}1-t^2/2, & 0\leqslant t\leqslant t_1,\\ t^2/2-2t_1t+t_1^2+1, & t\geqslant t_1,\end{cases}\qquad x^2(t)=\begin{cases}-t, & 0\leqslant t\leqslant t_1,\\ t-2t_1, & t\geqslant t_1.\end{cases}$$

Из условия $x^1(T)=x^2(T)=0$ находим $t_1=1$, $T=2$. Тогда

$$x^1(t)=\begin{cases}1-t^2/2, & 0\leqslant t\leqslant 1,\\ (t-2)^2/2, & 1\leqslant t\leqslant 2,\end{cases}\qquad x^2(t)=\begin{cases}-t, & 0\leqslant t\leqslant 1,\\ t-2, & 1\leqslant t\leqslant 2.\end{cases}$$

В качестве величин a_0, ψ_1, ψ_2, участвующих в формулировке принципа максимума, могут служить $a_0=0$, $\psi_1(t)=-1$, $\psi_2(t)=t-1$, $0\leqslant t\leqslant 2$. Можно показать, что полученные управление и траектория в самом деле являются решением поставленной задачи быстродействия, об этом см. пример 7.4.4.

Пример 5. Требуется перевести точку $x=(x^1,x^2)$ из состояния $x_0=(2,-2)$ на множество $S_1=\{x\in E^2\colon g^1(x)\equiv x^1=0\}$ быстрейшим образом, предполагая, что движение точки подчиняется уравнениям $\dot x^1(t)=x^2(t)$, $\dot x^2(t)=u(t)$, причем $u(t)\in V=\{u\in E^1\colon |u|\leqslant 1\}$.

Как и в предыдущем примере, здесь $H=-a_0+\psi_1x^2+\psi_2u$, сопряженная система имеет вид: $\dot\psi_1=0$, $\dot\psi_2=-\psi_1$, откуда следует $\psi_1(t)\equiv C$, $\psi_2(t)=-Ct+D$,

$C, D = \text{const}$. Условия трансверсальности (30), (43) здесь дают

$$\psi_1(T) = -a_1, \quad \psi_2(T) = 0, \quad -a_0 + \psi_1(T)x^2(T) + \psi_2(T)u(T) = H|_{t=T} = 0.$$

Следовательно, $\psi_2(t) = C(T-t)$, $0 \leqslant t \leqslant T$. Заметим, что здесь $C \neq 0$, так как при $C = 0$ получим $\psi_1(t) \equiv \psi_2(t) \equiv 0$, $0 \leqslant t \leqslant T$, а тогда $a_1 = 0$, из условия $H|_{t=T} = 0$ вытекает $a_0 = 0$ — противоречие с условием (32). Итак, $C \neq 0$, $\psi(t) = C(T-t) \neq 0$ при $0 \leqslant t < T$. Из условия $\max\limits_{|u| \leqslant 1} H$ тогда имеем

$$u(t) = \operatorname{sign} \psi_2(t) = \operatorname{sign} C, \quad 0 \leqslant t \leqslant T.$$

Подозрительными на оптимальность здесь могут быть лишь управления $u(t) \equiv 1$ или $u(t) \equiv -1$, $0 \leqslant t \leqslant T$. Если $u(t) = 1$, то из краевой задачи

$$\dot{x}^1 = x^2, \quad \dot{x}^2 = 1, \quad 0 \leqslant t \leqslant T; \quad x^1(0) = 2, \quad x^2(0) = -2, \quad x^1(T) = 0$$

получим $T = 2$, $x^1(t) = (t-2)^2/2$, $x^2(t) = t-2$, $0 \leqslant t \leqslant 2$. Если $u(t) = -1$, то из

$$\dot{x}^1 = x^2, \quad \dot{x}^2 = -1, \quad 0 \leqslant t \leqslant T; \quad x^1(0) = 2, \quad x^2(0) = -2, \quad x^1(T) = 0$$

будем иметь $T = \sqrt{8} - 2$, $x^1(t) = 4 - (t+2)^2/2$, $x^2(t) = -t - 2$, $0 \leqslant t \leqslant \sqrt{8} - 2$.

Таким образом, краевая задача принципа максимума здесь дает два решения. Однако лишь управление $u(t) = -1$, $0 \leqslant t \leqslant T = \sqrt{8} - 2$, может претендовать на оптимальность, так как $T = 2 > \sqrt{8} - 2$, а управление $u(t) = 1$, $0 \leqslant t \leqslant T$ заведомо неоптимально.

Пример 6. Рассмотрим задачу минимизации функции

$$J(u) = \int_0^1 ((u^1(t))^2 + (u^2(t))^2)dt + x^1(1) + x^2(1) \tag{45}$$

при условиях

$$\begin{aligned} &\dot{x}^1(t) = u^1(t), \quad \dot{x}^2(t) = u^2(t), \quad 0 \leqslant t \leqslant 1, \\ &x^1(0) = 0, \quad x^2(0) = 0, \quad x^1(1) \leqslant 0, \quad (x^2(1))^2 - x^1(1) \leqslant 0, \end{aligned} \tag{46}$$

где $u = u(t) = (u^1(t), u^2(t))$. Эта задача является частным случаем задачи (1), (2), (4), (29) при $t_0 = 0$, $T = 1$, $n = r = 2$, $f^0 \equiv (u^1)^2 + (u^2)^2$, $f(x, u, t) = u$, $V = E^2$, $x = (x^1, x^2)$, $y = (y^1, y^2)$, $g^0(x, y) = y^1 + y^2$, $g^1(x, y) = y^1$, $g^2(x, y) = -y^1 + (y^2)^2$, $m_1 = s_1 = 2$; левый конец траектории закреплен. Из (46) видно, что правый конец любой допустимой траектории этой задачи удовлетворяет равенствам: $x^1(1) = 0$, $x^2(1) = 0$. Тогда $J(u) = \int\limits_0^1 |u(t)|^2 dt \geqslant 0$ для всех допустимых управлений. Поскольку $u = u(t) \equiv 0$ допустимое управление и $J(0) = 0$, то $J_* = 0$, $u(t) \equiv 0$ — единственное оптимальное управление задачи (45), (46). Функция Гамильтона — Понтрягина $H = -a_0((u^1)^2 + (u^2)^2) + \psi_1 u^1 + \psi_2 u^2$ не зависит от x, поэтому сопряженная система имеет вид $\dot{\psi}_1 = 0$, $\dot{\psi}_2 = 0$, $0 \leqslant t \leqslant 1$. Следовательно, $\psi_1(t) \equiv c_1$, $\psi_2(t) \equiv c_2$, c_1, c_2 — постоянные. Далее, здесь $l(x, y, a) = a_0(y^1 + y^2) + a_1 y^1 + a_2(-y^1 + (y^2)^2)$ и условия (30), (32) дают

$$\begin{aligned} -\psi_1(1) &= a_0 \cdot 1 + a_1 \cdot 1 + a_2(-1) = a_0 + a_1 - a_2, \\ -\psi_2(1) &= a_0 \cdot 1 + a_1 \cdot 0 + a_2 \cdot 2x^2(1) = a_0, \\ a_0^2 + a_1^2 + a_2^2 &= 1, \quad a_0 \geqslant 0, \quad a_1 \geqslant 0, \quad a_2 \geqslant 0. \end{aligned} \tag{47}$$

Покажем, что в этой задаче $a_0 = 0$. В самом деле, если $a_0 > 0$, то можем воспользоваться условием нормировки $a_0 = 1$. Тогда функция $H = -|u|^2 + \langle \psi, u \rangle$ достигает своего максимума на $V = E^2$ в точке $u = \psi/2 = c/2$, $c = (c_1, c_2)$. Соответствующая траектория $x(t) = tc/2$ условию $x(1) = 0$ может удовлетворять лишь при $c_1 = c_2 = 0$. Таким образом, $\psi(t) \equiv 0$, $0 \leqslant t \leqslant 1$. Из второго условия (47) тогда следует, что $a_0 = 0$, что противоречит равенству $a_0 = 1$. Следовательно, $a_0 = 0$. Но тогда линейная функция $H = \langle \psi, u \rangle$ на E^2 может иметь конечный максимум (который, кстати, должен достигаться на оптимальном управлении $u(t) \equiv 0$) лишь при $\psi = \psi(t) \equiv c = 0$. Из условий (47), учитывая, что $a_0 = 0$, получаем $a_1 = a_2 > 0$. Таким образом, краевая задача принципа максимума здесь дает $a = (a_0 = 0, a_1 = a, a_2 = a)$, $a > 0$, $\psi(t) \equiv 0$, $0 \leqslant t \leqslant 1$. Как видим, условие (23) в этой задаче не выполняется, функция $H \equiv 0$, и условие максимума (9) не позволяет определить оптимальное управление $u = u(t) \equiv 0$.

Говорят, что оптимальное управление $u(\cdot)$ является *особым* на отрезке $[\alpha, \beta] \subset [t_0, T]$, если $H(x(t), u, t, \psi(t), a_0)$ при $t \in [\alpha, \beta]$ не зависит от u. В этом случае для набора $(x = x(t), t, \psi = \psi(t), a_0)$ при $t \in [\alpha, \beta]$ условие (14) не дает никакой полезной информации об оптимальном управлении, функция (13) становится неопределенной и пользоваться формулой (20) невозможно. В частности, когда нарушается условие (23), т. е. $a_0 = 0$, $\psi(t) \equiv 0$, $t_0 \leqslant t \leqslant T$, то имеем дело с одним из типичных случаев появления особого управления. Так случилось в только что рассмотренном примере 6. Разумеется, условие (23) само по себе не исключает возможность появления особого управления, но тем не менее полезно подчеркивать случаи, когда оно выполняется.

К сожалению, условие $a = (a_0, \ldots, a_s) \neq 0$ из (8) само по себе не всегда приводит к условию (23). Так в задаче (1), (2), (4), (29), как показывает пример 6, в общем случае (23) не выполняется и приходится довольствоваться условием $a = (a_0, \ldots, a_{s_1}) \neq 0$ и вытекающим из него условием нормировки (32). Для того чтобы в задаче (1), (2), (4), (29) из $a \neq 0$ следовало условие (23), можно дополнительно потребовать, например, чтобы градиенты $g_y^1(x(T)), \ldots, g_y^{s_1}(x(T))$ были линейно независимы. Тогда либо $a_0 \neq 0$, либо $a_0 = 0$, но согласно (30) $\psi(T) \neq 0$ и однородная система (6) будет иметь нетривиальное решение $\psi(t)$, $t_0 \leqslant t \leqslant T$. Аналогичные требования, гарантирующие условие (23), можно сформулировать и для задачи (1), (2), (4), (29), (34) и других задач оптимального управления, рассмотренных выше. Об особых управлениях читатель может прочесть в [108, 199; 476].

В следующих примерах покажем, как выписывается краевая задача принципа максимума для некоторых задач оптимального управления движением математического маятника (см. пример 1.1).

Пример 7. Пусть требуется минимизировать функцию

$$J(u) = (x^1(T))^2 + (x^2(T))^2 \tag{48}$$

при условиях

$$\dot{x}^1(t) = x^2(t), \quad \dot{x}^2(t) = -\sin x^1(t) - \beta x^2(t) + u(t), \quad 0 \leqslant t \leqslant T, \quad x(0) = x_0, \tag{49}$$

$$u(t) \in V = \{u \in E^1 : |u| \leqslant 1\}, \tag{50}$$

где $x = (x^1, x^2)$ — фазовые координаты, $x_0 = (x_0^1, x_0^2)$ — заданная точка, $T > 0$ — заданный момент времени. В этой задаче правый конец траектории

свободен, $f^0 \equiv 0$, $g^0(y) = (y^1)^2 + (y^2)^2$. Выпишем функцию Гамильтона — Понтрягина

$$H(x, u, \psi) = \psi_1 x^2 + \psi_2(-\sin x^1 - \beta x^2 + u)$$

и сопряженную систему

$$\dot{\psi}_1 = -H_{x^1} = \psi_2 \cos x^1, \quad \dot{\psi}_2 = -H_{x^2} = -\psi_1 + \beta\psi_2, \quad 0 \leqslant t \leqslant T. \tag{51}$$

Из условий (26) имеем

$$\psi_1(T) = -g^0_{y^1}(x(T)) = -2x^1(T), \quad \psi_2(T) = -g^0_{y^2}(x(T)) = -2x^2(T). \tag{52}$$

Из условия $\max\limits_{|u|\leqslant 1} H$ следует $u = \operatorname{sign}\psi_2$. Тогда краевая задача принципа максимума запишется в виде

$$\dot{x}^1 = x^2, \quad \dot{x}^2 = -\sin x^1 - \beta x^2 + \operatorname{sign}\psi_2, \tag{53}$$

$$\dot{\psi}^1 = \psi_2 \cos x^1, \quad \dot{\psi}^2 = -\psi_1 + \beta\psi_2, \quad 0 \leqslant t \leqslant T,$$

$$x(0) = x_0, \quad \psi_1(T) = -2x^1(T), \quad \psi_2(T) = -2x^2(T). \tag{54}$$

Если краевая задача (53), (54) имеет решение $(x(t), \psi(t))$, $0 \leqslant t \leqslant T$, причем $\psi_2(t)$ обращается в нуль в конечном числе точек, то функция $u(t) = \operatorname{sign}\psi_2(t)$ будет управлением, подозрительным на оптимальность в задаче (48)–(50).

Заметим, что если для некоторого управления $v = v(\cdot)$ из (50) решение $x(\cdot, v)$ задачи Коши (49) таково, что $x(T, v) = 0$, то $J(v) = 0$. Это значит, что $v(\cdot)$ — оптимальное управление в задаче (48) — (50). Любопытно, что это управление является особым — его нельзя получить из принципа максимума. В самом деле, при $x(T, v) = 0$ из (51), (52) следует $\psi(t, v) \equiv 0$, $0 \leqslant t \leqslant T$, а тогда $H(x(t, v), u, \psi(t, v)) \equiv 0$, $0 \leqslant t \leqslant T$, при всех $u \in V$ и условие (9) не суживает исходное множество управлений, подозрительных на оптимальность.

Пример 8. Минимизировать функцию $J(u) = \int\limits_0^T u^2(t)dt$ при условиях (49) и закрепленном правом конце $x(T)=0$; $T>0$ — задано.

Здесь $V = E^1$, $H = -a_0 u^2 + \psi_1 x^2 + \psi_2(-\sin x^1 - \beta x^2 + u)$, сопряженная система имеет вид (51). В случае $a_0 = 0$ функция H может достигать своей верхней грани на E^1 лишь при $\psi_2 = 0$. Но если $\psi_2(t) \equiv o$, $0 \leqslant t \leqslant T$, то из второго уравнения (51) получим $\psi_1(t) \equiv 0$, что противоречит условию (23). Таким образом, можем считать $a_0 = 1$. Тогда из условия $\max\limits_{u \in E^1} H$ получим $u = \psi_2/2$. Краевая задача принципа максимума будет иметь вид

$$\dot{x}^1 = x^2, \quad \dot{x}^2 = -\sin x^1 - \beta x^2 + \psi_2/2,$$

$$\dot{\psi}^1 = \psi_2 \cos x^1, \quad \dot{\psi}^2 = -\psi_1 + \beta\psi_2, \quad 0 \leqslant t \leqslant T, \quad x(0) = x_0, \quad x(T) = 0.$$

Пример 9. Рассмотрим задачу быстрейшего перевода точки $x = (x^1, x^2)$ из состояния $x_0 \neq 0$ в начало координат $(0, 0)$, предполагая, что движение точки подчиняется условиям (49), (50). Эта задача является частным случаем задачи (37)–(40). Если $f^0 \equiv 1$, $g^0 \equiv 0$; функция Гамильтона — Понтрягина имеет вид

$$H = -a_0 + \psi_1 x^2 + \psi_2(-\sin x^1 - \beta x^2 + u). \tag{55}$$

Отсюда ясно, что сопряженная система будет иметь вид (51), а условие $\max\limits_{|u|\leqslant 1} H$ выделит функцию $u = \operatorname{sign}\psi_2$. Краевая задача принципа максимума

в этом случае будет состоять из системы (53), граничных условий $x(0) = x_0$, $x(T) = 0$, условия трансверсальности $H|_{t=T} = -a_0 + \psi_2(T)u(T) = 0$, вытекающего из (43), и условия (23). Отметим, что в этой задаче $\psi_2(t) \not\equiv 0$. В самом деле, если бы $\psi_2(t) \equiv 0$, то из (51) будем иметь $\psi_1(t) \equiv 0$, а из $H|_{t=T} = 0$ получим $a_0 = 0$, что противоречит (23).

Пример 10. Пусть требуется быстрейшим образом перевести точку $x = (x^1, x^2)$ из состояния x_0 в начальный момент $t_0 = 0$ в состояние, удаленное от точки $(0, 0)$ на расстояние, равное $\varepsilon > 0$, предполагая, что движение точки подчиняется условиям (49), (50). Это значит, что правый конец траектории является подвижным и удовлетворяет условиям $|x(T)|^2 = (x^1(T))^2 + (x^2(T))^2 = \varepsilon^2$. Здесь функция H имеет тот же вид (55), сопряженная система — вид (51), условие $\max\limits_{|u|\leqslant 1} H$ дает $u = \operatorname{sign}\psi_2$. Из условия трансверсальности (30), (43) имеем

$$\psi(T) = -2a_1 x(T), \quad H(x(T), u(T), T, \psi(T), a_0) = 0. \tag{56}$$

Оказывается, здесь $a_1 \neq 0$. В самом деле, если $a_1 = 0$, то из (56) будем иметь $\psi(T) = 0$, а из (51) будет следовать $\psi(t) \equiv 0$; тогда из (55), (56) получим $a_0 = 0$, что противоречит условию (32). Итак, $a_1 \neq 0$; тогда условие нормировки (32) можем заменить на равенство $|a_1| = 1$. Таким образом, краевая задача принципа максимума в рассматриваемой задаче состоит из системы (53), граничных условий $x(0) = x_0$, $\psi(T) = \pm 2x(T)$, $|x(T)|^2 = \varepsilon^2$, $H|_{t=T} = 0$, из которых нужно определить функции $x(t)$, $\psi(t)$, параметры a_0, T. Отметим, что здесь $\psi_2(t) \not\equiv 0$. В самом деле, если $\psi_2(t) \equiv 0$, то в силу (51) $\psi_1(t) \equiv 0$, а тогда $x(T) = 0$, что противоречит равенству $|x(T)|^2 = \varepsilon^2 > 0$.

Пример 11. В задаче из примера 10 условие $|x(T)| = \varepsilon^2$ на правом конце траектории заменим неравенством $|x(T)|^2 \leqslant \varepsilon^2$, считая, что $|x(t_0)| > \varepsilon^2$. Тогда функция H, сопряженная система (51), условия (56) останутся без изменений; здесь также выполняется условие дополняющей нежесткости $a_1(|x(T)|^2 - \varepsilon^2) = 0$. Оказывается, и в этой задаче можно показать, что $a_1 \neq 0$. В самом деле, если $a_1 = 0$, то в силу (56) $\psi(T) = 0$, из (51) будет следовать $\psi(t) \equiv 0$, а из $H|_{t=T} = 0$ получаем $a_0 = 0$, что противоречит условию (32). Итак, $a_1 \neq 0$. С учетом $a_1 \geqslant 0$ можем принять $a_1 = 1$. Таким образом, краевая задача принципа максимума будет состоять из системы (53), условий $x(0) = x_0$, $\psi(T) = -2x(T)$, $|x(T)|^2 = \varepsilon^2$, $H|_{t=T} = 0$. Как и в предыдущем примере можно показать, что $\psi_2(t) \not\equiv 0$.

Предлагаем читателю самостоятельно выписать краевую задачу принципа максимума для задач из примеров 10, 11 с заменой условий на правом конце одним из условий $x^i(T) = 0$, $|x^i(T)|^2 \leqslant \varepsilon^2$, $|x^i(T)|^2 = \varepsilon^2$, где $i = 1$ или 2.

6. В следующем параграфе будет приведено доказательство теорем 1, 2. Здесь мы приведем эвристические соображения, которые помогают понять, откуда появляется сопряженная система, условия максимума, условия трансверсальности, фигурирующие в теоремах 1, 2, и установить связь между принципом максимума Понтрягина и правилом множителей Лагранжа.

Для простоты рассмотрим задачу (1), (2), (4) при дополнительном предположении, что $V = E^r$, $g^0(x, y) = g^0(y)$, условия на правом конце траектории имеют вид

$$g^1(x(T)) = 0, \ldots, g^{s_1}(x(T)) = 0, \tag{63}$$

на левом конце

$$h^1(x(t_0)) = 0, \ldots, h^{s_0}(x(t_0)) = 0. \tag{64}$$

Воспользуемся процедурой исследования задач на условный экстремум (см. гл. 2) и введем множители Лагранжа: непрерывную вектор-функцию $\psi(t)=(\psi_1(t),\ldots,\psi_n(t))$ для учета ограничений (2), постоянные $(a_1,\ldots,a_{s_1})=a$ для учета ограничений (63), постоянные $(b_1,\ldots,b_{s_0})=b$ — для учета (64), постоянную a_0 — для функции (1), и составим функцию Лагранжа

$$\mathcal{L}(x(\cdot), u(\cdot), \psi(\cdot), a_0, a, b) = -a_0\left[\int_{t_0}^{T} f^0(x(t), u(t), t)dt + g^0(x(T))\right] +$$
$$+\int_{t_0}^{T}\langle\psi(t), f(x(t), u(t), t) - \dot{x}(t)\rangle dt - \sum_{j=1}^{s_1} a_j g^j(x(T)) - \sum_{j=1}^{s_0} b_j h^j(x(t_0)).$$

С помощью функции Гамильтона — Понтрягина (5) можно записать функцию Лагранжа в следующем виде:

$$\mathcal{L}(x(\cdot), u(\cdot), \psi(\cdot), a_0, a, b) = \int_{t_0}^{T}[H(x(t), u(t), t, \psi(t), a_0) -$$
$$-\langle\psi(t), \dot{x}(t)\rangle]dt - \sum_{j=0}^{s_1} a_j g^j(x(T)) - \sum_{j=1}^{s_0} b_j h^j(x(t_0)).$$

Все аргументы функции Лагранжа считаются независимыми переменными. Дадим приращения (вариации) переменным $x(t)$, $u(t)$, т. е. рассмотрим $x(t)+\delta x(t)$, $u(t)+\delta u(t)$, $t_0\leqslant t\leqslant T$, здесь $u(t)$, $\delta u(t)$ — кусочно-непрерывны, а $x(t)$, $\delta x(t)$ — непрерывно дифференцируемы на $[t_0, T]$. Тогда вариация функции Лагранжа, представляющая собой главную линейную часть приращения этой функции, имеет вид

$$\delta\mathcal{L} = \int_{t_0}^{T}[\langle H_x, \delta x\rangle + \langle H_u, \delta u\rangle - \langle\psi, \delta\dot{x}\rangle]dt -$$
$$-\sum_{j=0}^{s_1} a_j\langle g_x^j(x(T)), \delta x(T)\rangle - \sum_{j=1}^{s_0} b_j\langle h_x^j(x(t_0)), \delta x(t_0)\rangle;$$

для краткости аргументы у функций под интегралом опущены. Интегрируя по частям, находим

$$\int_{t_0}^{T}\langle\psi(t), \delta\dot{x}(t)\rangle dt = \langle\psi(t), \delta x(t)\rangle\Big|_{t=t_0}^{T} - \int_{t_0}^{T}\langle\dot{\psi}(t), \delta x(t)\rangle dt.$$

Тогда

$$\delta\mathcal{L} = \int_{t_0}^{T}[\langle H_x + \dot{\psi}(t), \delta x(t)\rangle + \langle H_u, \delta u\rangle]dt -$$
$$-\left\langle\sum_{j=0}^{s_1} a_j g_x^j(x(T)) + \psi(T), \delta x(T)\right\rangle + \left\langle -\sum_{j=1}^{s_0} b_j h_x^j(x(t_0)) + \psi(t_0), \delta x(t_0)\right\rangle.$$

Пользуясь независимостью вариаций $\delta x(\cdot)$, $\delta u(\cdot)$, из условия стационарности $\delta\mathcal{L}=0$ имеем

$$\dot{\psi} + H_x(x, u, t, \psi, a_0) = 0, \quad H_u(x, u, t, \psi, a_0) = 0,$$
$$\psi(T) = -\sum_{j=0}^{s_1} a_j g_x^j(x(T)), \quad \psi(t_0) = \sum_{j=1}^{s_0} b_j h_x^j(x(t_0)).$$

Таким образом, получены, почти все основные соотношения теоремы 1, кроме условий (8), (9). Вместо условия (9) мы получили равенство $H_u=0$, являющееся следствием (9) при $V=E^r$. Подчеркнем, что приведенные здесь

рассуждения, конечно, не могут считаться строгими, и являются лишь полезными наводящими соображениями при получении необходимых условий оптимальности. Систематическое и строгое изложение правила множителей Лагранжа применительно к различным классам задач на экстремум, в частности, к задачам оптимального управления дано в [14; 15; 209; 210].

Упражнения

1. С помощью принципа максимума решить задачу быстродействия при условиях $\dot{x}^1 = x^2$, $\dot{x}^2 = u(t)$, $x(0)\in S_0$, $x(T)\in S_1$; $u(t)\in V=\{u\in E^1: |u|\leqslant 1\}$, где $S_0=\{x\in E^2: h(x)\equiv|x|^2=1\}$ или $S_0=\{x\in E^2: h(x)\equiv|x|^2\leqslant 1\}$, или $S_0=\{(0,0)\}$, или $S_0=\{(1,0)\}$, а $S_1=\{x\in E^2, g(x)\equiv|x|^1=0\}$, или $S_1=\{x\in E^2: g(x)\equiv|x|^2=4\}$ или $S_1=\{x\in E^2: g(x)\equiv|x|^2\leqslant 4\}$.

2. Применить принцип максимума к задаче: $J(u)=\int_0^1|u(t)|^2dt\to\inf$; $\dot{x}^1=x^2$, $\dot{x}^2=u^1(t)$, $\dot{x}^3=x^4$, $\dot{x}^4=u^2(t)-g$, $0\leqslant t\leqslant 1$; $x(0)=(-1,0,0,0)$, $x(T)=(0,0,0,0)$. Здесь $x=(x^1,x^2,x^3,x^4)$, $u=(u^1,u^2)$; $g=\text{const}\geqslant 0$.

3. Применить принцип максимума к задаче о мягкой посадки космического корабля на Луну с минимальной затратой горючего ([724], с. 36, 44, 54): $J(u)=m(T)\to\sup$; $\dot{h}(t)=v(t)$, $\dot{v}(t)=-g+u(t)/m(t)$, $\dot{m}(t)=-ku(t)$, $u(t)\in V=\{u\in E^1: 0\leqslant u\leqslant\alpha\}$, $0\leqslant t\leqslant T$; $h(0)=h_0>0$, $v(0)=v_0$, $m(0)=m_0>0$; $h(T)=0$; момент T заранее не задан. Здесь $m(t)$ — масса корабля, $h(t)$ — высота, $v(t)$ — вертикальная скорость корабля над Луной, $u(t)$ — тяга двигателя, g — гравитационное ускорение Луны, $\alpha>0$, $k>0$ — постоянные (ср. пример 1.2).

4. Рассмотреть задачи из примеров 7–11 для малых колебаний маятника, считая $\sin x^1\approx x^1$, $\cos x^1\approx 1$. Найти решения краевых задач принципа максимума.

5. Рассмотреть задачи из примеров 1, 2 при дополнительном ограничении $u(t)\in V=\{u\in E^1: |u|\leqslant 1\}$.

6. Применить принцип максимума к задаче: $J(u)=x^2(T)\to\inf$; $\dot{x}=u(t)$, $u(t)\in V=\{u\in E^1: |u|\leqslant 1\}$, $0\leqslant t\leqslant T$; $x(0)=x_0$; момент $T>0$ задан. Сколько решений имеет эта задача при $x_0=0$? $x_0=T$?

7. Показать, что в задаче $J(u)=\int_0^1(x^2(t)-u^2(t))dt\to\inf$; $\dot{x}=u(t)$, $u(t)\in V=\{u\in E^1: |u|\leqslant 1\}$, $0\leqslant t\leqslant 1$; $x(0)=0$, оптимальное управление не существует, $\inf J(u)=-1$ (см. пример 7.4.2). Что дает здесь применение принципа максимума?

8. Найти минимум функции $J(u)=\int_0^1\sin u(t)dt$ при условии $\dot{x}=\cos u(t)$, $u(t)\in V=\{u\in E^1: |u|\leqslant\pi/2\}$, $0\leqslant t\leqslant 1$; $x(0)=0$, $x(1)=1$. Показать, что $u(t)\equiv 0$, $0\leqslant t\leqslant 1$ — оптимальное управление, и убедиться в том, что в принципе максимума здесь надо принять $a_0=0$.

9. Применить принцип максимума к задаче: $J(u)=\int_0^T|x(t^2)|^2dt\to\inf$; $\dot{x}=u(t)$, $u(t)\in V=\{u\in E^1: |u|\leqslant 1\}$, $0\leqslant t\leqslant T$; $x(0)=1$, $x(T)=0$; момент $T>0$ задан. Показать, что при $T>1$ оптимальное управление: $u(t)\equiv -1$, $0\leqslant t<1$; $u(t)\equiv 0$, $1\leqslant t\leqslant T$, на участке $1\leqslant t\leqslant T$ является особым, т. е. его нельзя определить из принципа максимума.

10. С помощью принципа максимума исследовать задачу: $J(u)=\int_0^1 u(t)\cos\frac{\pi x(t)}{2}dt\to\inf$; $\dot{x}=u(t)$, $u(t)\in V=\{u\in E^1: |u|\leqslant 1\}$, $0\leqslant t\leqslant 1$; $x(0)=0$.

11. Показать, что задача: $J(u)=\int_0^T u^2(t)\,(u(t)-1)^2dt\to\inf$; $\dot{x}=u(t)$, $u(t)\in V=\{u\in E^1: 0\leqslant u\leqslant 1\}$, $0\leqslant t\leqslant T$; $x(0)=0$, $x(T)=1$ при $T=1$ имеет единственное решение, а при $T>1$ — бесконечно много решений. Изменится ли этот вывод, если $V=E^1$? Если $x(T)=-1$?

12. С помощью принципа максимума исследовать задачу: $J(u)=|x^2(1)|^2\to\inf$; $\dot x^1=x^2$, $\dot x^2=u(t)$, $u(t)\in V=\{u\in E^1\colon |u|\leqslant 1\}$, $x(0)=(x_0^1, x_0^2)$.

13. Пусть задача оптимального управления автономна (см. § 1), $(u(\cdot), x(\cdot))$ — ее решение, а $\psi(\cdot)$, $a_0, a_1, \ldots, a_s$, определены из принципа максимума (теоремы 1, 2). Показать, что тогда $H(x(t), u(t), \psi(t), \psi_0)\equiv\text{const}$, $t_0\leqslant t\leqslant T$.

14. Сформулировать принцип максимума для задачи оптимального управления с параметрами $w=(w^1, \ldots, w^p)$ (они не зависят от времени):

$$J(u(\cdot), w)=\int_{t_0}^{T} f^0(x(t), u(t), t, w)dt+g_0(x_0, x(T), t_0, T, w)\to\inf,$$

$$\dot x(t)=f(x(t), u(t), t, w),\quad t_0\leqslant t\leqslant T,\quad x(t_0)=x_0,$$

$$g^i(x_0, x(T), t_0, T, w)\leqslant 0,\quad i=1, \ldots, m,\quad g^i(x_0, x(T), t_0, T, w)=0,\quad i=m+1, \ldots, s,$$

$$u(t)\in V,\quad t_0\leqslant t\leqslant T,\quad w\in W,$$

где W — заданное множество из E^p, остальные обозначения см. выше. Рассмотреть случаи $W=E^p$ и $W=\{w\colon q_i(w)=0, i=1, \ldots, k\}$, где $q_i(w)$ — заданные гладкие функции на E^p, $i=1, \ldots, k$. Указание: ввести новые переменные x^{n+i} посредством условий $\dot x^{n+i}(t)=0$, $t_0\leqslant t\leqslant T$, $x^{n+i}(t_0)=w^i$, $i=1, \ldots, k$, в пространстве $(x^1, \ldots, x^{n+k})$ воспользоваться теоремой 1 или 2, а затем исключить переменные x^{n+i}, ψ_{n+i}, $i=1, \ldots, k$.

15. Сформулировать принцип максимума для задачи получающейся из задачи (1)–(4) или (37)–(40) добавлением условий (1.33). Указание: ввести новые переменные x^{n+i} посредством условий $\dot x^{n+i}(t)=f_i^0(x(t), u(t), t)$, $t_0\leqslant t\leqslant T$, $x^{n+i}(t_0)=0$, $i=1, \ldots, s_2$; $x^{n+i}(T)+g_i^0(x_0, x(T), t_0, T)\leqslant 0$, $i=1, \ldots, m_2$; $x^{n+i}(T)+g_i^0(x_0, x(T), t_0, T)=0$, $i=m_2+1, \ldots, s_2$, в пространстве $(x^1, \ldots, x^{n+s_2})$, воспользоваться теоремой 1 или 2, затем исключить переменные x^{n+i}, ψ_{n+i}, $i=1, \ldots, s_2$.

16. Показать, что в задаче

$$J(u)=\int_0^T x^2(t)dt\to\inf,\quad \dot x=y,\quad \dot y=u(t),\quad 0\leqslant t\leqslant T,$$

$$x(0)=x_0,\quad y(0)=y_0,\quad (x_0, y_0)\neq 0,\quad |u(t)|\leqslant 1,$$

при достаточно большом T оптимальное управление имеет счетное число переключений с $+1$ на -1, причем точки переключения лежат на кривой $x=Cy^2\operatorname{sign} y$, C — некоторая постоянная (четтеринг-режим) [322; 323; 819]. Указание: рассмотреть задачу быстрейшего попадания из точки (x_0, y_0) в точку $(0, 0)$ и, пользуясь принципом максимума, убедиться, что оптимальное управление не может иметь конечное число переключений.

§ 3. Доказательство принципа максимума

1. Доказательство теорем 2.1, 2.2 проведем при дополнительном предположении, что вектор-функция $f(x, u, t)=(f^1(x, u, t), \ldots, f^n(x, u, t))$ удовлетворяет условию Липшица по переменным (x, u), т. е.

$$|f(x, u, t)-f(y, v, t)|\leqslant(|x-y|+|u-v|),\quad L>0, \tag{1}$$

при всех $(x, u, t), (y, v, t)\in E^n\times E^n\times[t_0, T]$. Покажем, что тогда решение задачи Коши

$$\dot x=f(x, u(t), t),\quad t_0\leqslant t\leqslant T,\quad x(t_0)=x_0 \tag{2}$$

непрерывно зависит от начальной точки x_0 и управления $u=u(t)$. Сначала докажем одно утверждение, которое часто приводится в учебных пособиях по дифференциальным уравнениям и в литературе известно как неравенство Гронуолла.

Лемма 1. *Пусть функции $\varphi(t)$, $b(t)$ неотрицательны и непрерывны на отрезке $t_0\leqslant t\leqslant T$, $a=\text{const}\geqslant 0$. Пусть*

$$\varphi(t)\leqslant a\int_{t_0}^{t}\varphi(\tau)d\tau+b(t),\quad t_0\leqslant t\leqslant T. \tag{3}$$

Тогда

$$0\leqslant\varphi(t)\leqslant a\int_{t_0}^{t}b(\tau)e^{a(t-\tau)}d\tau+b(t),\quad t_0\leqslant t\leqslant T. \tag{4}$$

В частности, если $b(t)\equiv b=\text{const}\geqslant 0$, то

$$0\leqslant\varphi(t)\leqslant be^{a(t-t_0)},\ t_0\leqslant t\leqslant T. \tag{5}$$

Если же

$$\varphi(t)\leqslant a\int_{t}^{T}\varphi(\tau)d\tau+b(t),\quad t_0\leqslant t\leqslant T, \tag{6}$$

то

$$0\leqslant\varphi(t)\leqslant a\int_{t}^{T}b(\tau)e^{a(\tau-t)}d\tau+b(t),\quad t_0\leqslant t\leqslant T, \tag{7}$$

а при $b(t)\equiv b=\text{const}\geqslant 0$ будем иметь

$$0\leqslant\varphi(t)\leqslant be^{a(T-t)},\quad t_0\leqslant t\leqslant T. \tag{8}$$

Доказательство. Положим $R(t)=a\int_{t_0}^{t}\varphi(\tau)d\tau$. Заметим, что $R(t_0)=0$, $R(t)\geqslant 0$, $\dot R(t)=a\varphi(t)$, $t_0\leqslant t\leqslant T$. С учетом (3) и $a\geqslant 0$ имеем

$$\dot R(t)\leqslant aR(t)+ab(t)\quad\text{или}\quad \dot R(t)-aR(t)\leqslant ab(t),\quad t_0\leqslant t\leqslant T.$$

Умножая обе части последнего неравенства на $e^{-a(t-t_0)}$ получим

$$\frac{d}{dt}(R(t)e^{-a(t-t_0)})\leqslant ab(t)e^{-a(t-t_0)},\quad t_0\leqslant t\leqslant T.$$

Интегрирование этого неравенства от t_0 до t с учетом $R(t_0)=0$ приводит к оценке

$$R(t)e^{-a(t-t_0)}\leqslant a\int_{t_0}^{t}b(\tau)e^{-a(\tau-t_0)}d\tau,$$

или

$$R(t)\leqslant a\int_{t_0}^{t}b(\tau)e^{a(t-\tau)}d\tau,\quad t_0\leqslant t\leqslant T.$$

Подставив эту оценку в правую часть (3), сразу получим требуемое неравенство (4). Если $b(t)=b=\text{const}$, то непосредственно вычисляя интеграл в правой части (4), придем к оценке (5).

Неравенства (7), (8), вытекающие из (6), доказываются аналогично с помощью вспомогательной функции $R(t)=a\int_{t}^{T}\varphi(\tau)d\tau$, $t_0\leqslant t\leqslant T$. Лемма 1 доказана. □

Теорема 1. *Пусть вектор-функция $f(x, u, t)$ непрерывна по совокупности переменных $(x, u, t)\in E^n\times V\times[t_0, T]$, удовлетворяет условию (1). Тогда*

$$\max_{t_0\leqslant t\leqslant T}|x(t, v, y_0)-x(t, u, x_0)|\leqslant c_1|y_0-x_0|+c_2\int_{t_0}^{T}|v(t)-u(t)|dt, \tag{9}$$

где $x(t, u, x_0)$ — решение задачи Коши (2), соответствующее управлению $u = u(t) \in L_\infty^r[t_0, T]$: $u(t) \in V$, $t_0 \leqslant t \leqslant T$, и начальному условию x_0; $c_1 = e^{L(T-t_0)}$, $c_2 = c_1 L$.

Доказательство. Существование траекторий $x(t, u, x_0)$, $x(t, v, y_0)$, $t_0 \leqslant t \leqslant T$, следует из теоремы 1.1. Обозначим для краткости $\Delta u(t) = v(t) - u(t)$, $\Delta x(t) = x(t, v, y_0) - x(t, u, x_0)$, $\Delta x_0 = y_0 - x_0$. Тогда из (1.12) следует

$$\Delta x(t) = \int_{t_0}^{t} [f(x(\tau, v, y_0), v(\tau), \tau) - f(x(\tau, u, x_0), u(\tau), \tau)]d\tau + \Delta x_0.$$

Отсюда с учетом условия (1) имеем

$$|\Delta x(t)| \leqslant L \int_{t_0}^{t} |\Delta x(\tau)|d\tau + L \int_{t_0}^{T} |\Delta u(\tau)|d\tau + |\Delta x_0|.$$

Это неравенство запишется в виде (3), если принять $\varphi(t) = |\Delta x(t)|$, $a = L$, $b(t) \equiv b = L \int_{t_0}^{T} |\Delta u(t)|dt + |\Delta x_0|$. Отсюда и из леммы 1 следует оценка (9).

Более тонкие теоремы о непрерывной зависимости решений задачи (2) от исходных данных, справедливые без дополнительного требования (1) и удобные для использования при доказательстве принципа максимума, читатель найдет, например, в [14, 132; 212; 588].

2. Приступим к доказательству теорем 2.1, 2.2. Приводимое ниже простое и изящное доказательство этих теорем принадлежит А. В. Арутюнову [44].

Доказательство теоремы 2.1 будет состоять из трех этапов. Вначале исходная задача (2.1)–(2.4) аппроксимируется семейством конечномерных задач. Затем к конечномерной задаче будет применен метод штрафных функций для учета ограничений на концах траекторий и выведено необходимое условие оптимальности для получившейся штрафной задачи. Наконец, будет совершен предельный переход в полученных необходимых условиях и установлена справедливость принципа максимума.

Сначала доказательство проведем в предположении, что решение $(x_{0*}, u_*(t), x_*(t))$ задачи (2.1)–(2.4) единственно. Через $t_1, t_2, \ldots$ обозначим каким-либо образом занумерованные рациональные точки интервала (t_0, T), являющиеся точками непрерывности оптимального управления $u_*(t)$. Выберем во множестве V всюду плотную последовательность точек $v_1, v_2, \ldots$ Зафиксируем произвольный номер $N \geqslant 1$ и для любого натурального числа $l \leqslant N$ выберем такие точки $t_{lp} = t_{lp}(N) \in (t_0, T)$, $p = 1, \ldots, N+1$, что

$$t_l = t_{l1} < \ldots < t_{lp} < \ldots < t_{lN+1}, \quad t_{lN+1} - t_l \leqslant 1/N, \tag{12}$$
$$[t_l, t_{lN+1}] \cap [t_k, t_{kN+1}] = \varnothing \quad \forall l, k, \ l \neq k, \quad 1 \leqslant l, \ k \leqslant N,$$

причем, оптимальное управление $u(t)$ непрерывно на отрезках $[t_l, t_{lN+1}]$, $l = 1, \ldots, N$. Обозначим через ξ квадратную матрицу $\xi = \{\xi_{lp}\}$ размерности N, элементы которой удовлетворяют ограничению

$$0 \leqslant \xi_{lp} \leqslant d_N = \min_{1 \leqslant l, p \leqslant N} (t_{lp+1} - t_{lp}). \tag{12.A}$$

Для каждой такой матрицы ξ определим управление $u(\cdot, \xi)$ следующим образом:

$$u(t, \xi) = \begin{cases} v_p, & t_{lp} < t \leqslant t_{lp} + \xi_{lp}, \quad 1 \leqslant l, \ p \leqslant N, \\ u_*(t) & \text{в остальных точках отрезка } [t_0, T]. \end{cases} \tag{13}$$

Нетрудно видеть, что управление (13) кусочно-непрерывно, $u(t, \xi) \in V$ при всех $t \in [t_0, T]$. Рассмотрим задачу: минимизировать функцию

$$J_N(x_0, \xi) \equiv J(x_0, u(\cdot, \xi)) = \int_{t_0}^{T} f^0(x(t), u(t, \xi), t)dt + g^0(x_0, x(T)) \tag{14}$$

при условиях:

$$\dot{x}(t) = f(x(t), u(t, \xi), t), \quad t_0 \leqslant t \leqslant T, \quad x(t_0) = x_0, \tag{15}$$
$$g^i(x_0, x(T)) \leqslant 0, \ i = 1, \ldots, m, \ g^i(x_0, x(T)) = 0, \ i = m+1, \ldots, s, \tag{16}$$
$$\xi = \{\xi_{lp}\}: 0 \leqslant \xi_{lp} \leqslant d_N, \quad l, p = 1, \ldots, N, \tag{17}$$

где управление $u = u(t, \xi)$ определяется согласно (13). Подчеркнем, что при каждом фиксированном $N \geqslant 1$ задача (14)–(17) является конечномерной задачей минимизации в пространстве переменных (x_0, ξ): $x_0 = (x_0^1, \ldots, x_0^n)$, $\xi = \{\xi_{lp}, \ l, p = 1, \ldots, N\}$.

Нетрудно видеть, что любой набор $(x_0, u(\cdot, \xi))$, допустимый в задаче (14)–(17), является допустимым и в задаче (2.1)–(2.4), причем значения целевых функций в обеих задачах на таком наборе совпадают. Отсюда следует, что нижняя грань J_{N_*} целевой функции задачи (14)–(17) не меньше нижней грани J_* в задаче (2.1)–(2.4): $J_{N_*} \geqslant J_*$. В то же время оптимальный набор $(x_{0*}, u_*(\cdot))$ задачи (2.1)–(2.4) является допустимым в задаче (14)–(17), так как согласно (13) $u(t, 0) = u_*(t)$, $t_0 \leqslant t \leqslant T$. Тогда $J_* = J(x_{0*}, u_*(\cdot)) = J(x_{0*}, u(\cdot, 0)) = J_N(x_{0*}, 0) \geqslant J_{N_*}$. Следовательно, $J_* = J_{N_*} = J_N(x_{0*}, 0)$. По предположению задача (2.1)–(2.4) имеет единственное решение. Тогда и задача (14)–(17) также будет иметь единственное решение $(x_{0*}, \xi_* = 0)$ при каждом фиксированном $N \geqslant 1$.

Применим к задаче (14)–(17) метод штрафных функций. Обозначим $g_{iN}(x_0, \xi) = g_i(x_0, x(T)) = g_i(x_0, x(T, u(\cdot, \xi), x_0))$, $i = 1, \ldots, s$,

$$g_i^+(x, y) = \begin{cases} \max\{0; g^i(x, y)\}, & i = 1, \ldots, m, \\ g^i, & i = m+1, \ldots, s. \end{cases} \tag{18}$$

Рассмотрим задачу: минимизировать функцию

$$\Phi_k(x_0, \xi) = J_N(x_0, \xi) + A_k \sum_{i=1}^{s} (g_{iN}^+(x_0, \xi))^2 =$$
$$= \int_{t_0}^{T} f^0(x(t), u(t, \xi), t)dt + g^0(x_0, x(T)) + A_k \sum_{i=1}^{s} (g_i^+(x_0, x(T)))^2 \tag{19}$$

при условиях

$$\dot{x}(t) = f(x(t), u(t, \xi), t), \quad t_0 \leqslant t \leqslant T, \quad x(t_0) = x_0, \tag{20}$$
$$\xi = \{\xi_{lp}\}, \quad 0 \leqslant \xi_{lp} \leqslant d_N, \quad l, p = 1, \ldots, N, \tag{21}$$
$$|x_0 - x_{0*}| \leqslant 1, \tag{22}$$

где $A_k > 0$, $k = 0, 1, \ldots$; $\{A_k\} \to \infty$ (например, $A_k = k$). Если ввести множество

$$U_0 = \{(x_0, \xi): |x_0 - x_{0*}| \leqslant 1, \quad 0 \leqslant \xi_{lp} \leqslant d_N, \quad i, p = 1, \ldots, N\},$$

то задача (19)–(22) может быть кратко записана в виде

$$\Phi_k(x_0, \xi) \to \inf, \quad (x_0, \xi) \in U_0. \tag{23}$$

Покажем, что задача (23) имеет хотя бы одно решение. Сначала убедимся, что функции $J_N(x_0, \xi)$, $g_{iN}(x_0, \xi)$ непрерывны по совокупности (x_0, ξ) на множестве U_0. Пусть (x_0, ξ), $(x_0+\Delta x_0, \xi+\Delta\xi) \in U_0$. Согласно оценке (9) с учетом (13) имеем

$$\max_{t_0 \leqslant t \leqslant T} |x(t, u(\cdot, \xi+\Delta\xi), x_0+\Delta x_0) - x(t, u(\cdot, \xi), x_0)| \leqslant$$

$$\leqslant c_1|\Delta x_0| + c_2 \int_{t_0}^{T} |u(t, \xi+\Delta\xi) - u(t, \xi)|dt =$$

$$= c_1|\Delta x_0| + c_2 \sum_{l,p=1}^{N} \int_{t_{lp}+\xi_{lp}}^{t_{lp}+\xi_{lp}+\Delta\xi_{lp}} |v_p - u_*(t)|dt \leqslant c_1|\Delta x_0| + c_{2N}|\Delta\xi|, \quad (24)$$

где

$$c_{2N} = c_2 N^2 \sup_{1 \leqslant p \leqslant N} \sup_{t_0 \leqslant t \leqslant T} |v_p - u_*(t)|, \quad |\Delta\xi| = \max_{1 \leqslant l,p \leqslant N} |\Delta\xi_{lp}|.$$

Из непрерывности функции $f^0(x, u, t)$, $g^i(x, y)$ по совокупности своих аргументов и оценки (24) следует непрерывность функций $J_N(x_0, \xi)$, $g_{iN}(x_0, \xi)$, $g_{iN}^+(x_0, \xi)$, $\Phi_k(x_0, \xi)$ на компактном множестве U_0. Согласно теореме 2.1.1 $\Phi_{k*} = \inf_{U_0} \Phi_k(x_0, \xi) > -\infty$ и существует хотя бы одна точка $(x_{0k}, \xi_k) \in U_0$, в которой нижняя грань достигается. Покажем, что последовательность $(x_{0k}, \xi_k = \{\xi_{lp}^k\})$, $k = 0, 1, \ldots$, решений задач (23) сходится к решению $(x_{0*}, \xi_* = 0)$ задачи (14)–(17). Воспользуемся теоремами 5.15.1, 5.15.2. Проверим выполнение условий этих теорем. Непрерывность функций $J_N(x_0, \xi)$, $g_{iN}(x_0, \xi)$ мы уже установили. Из непрерывности $J_N(x_0, \xi)$ и компактности U_0 следует, что $J_{**} = \inf_{U_0} J_N(x_0, \xi) > -\infty$. Множество $U_\delta = \{(x_0, \xi) \in U_0: g_{iN}^+(x_0, \xi) \leqslant \delta,\ i = 1, \ldots, s\}$ ограничено в силу ограниченности U_0 при любом $\delta > 0$. Все условия теорем 5.15.1, 5.15.2 выполнены. Поэтому решение задачи (23) или (19)–(22) сходится к решению задачи (14)–(17) как по функции, так и по аргументу. Поскольку задача (14)–(17) имеет единственное решение $(x_{0*}, \xi_* = 0)$, то

$$\lim_{k\to\infty} \xi_k = \xi_* = 0, \quad \lim_{k\to\infty} x_{0k} = x_{0*}, \quad \lim_{k\to\infty} \Phi_k(x_{0k}, \xi_k) = J_{N_*} = J_*. \quad (25)$$

Из оценки (24) при $\xi = \xi_* = 0$, $x_0 = x_{0*}$, $\Delta\xi = \xi_k$, $\Delta x_0 = x_{0k} - x_{0*}$ получаем

$$\lim_{k\to\infty} \max_{t_0 \leqslant t \leqslant T} |x_k(t) - x_*(t)| = 0, \quad (26)$$

где $x_*(t) = x(t, u_*(\cdot), x_{0*}) = x(t, u(\cdot, 0), x_{0*})$ — оптимальная траектория в задачах (14)–(17), (2.1)–(2.4), $x_k(t) = x(t, u(\cdot, \xi_k), x_{0k})$ — оптимальная траектория в задаче (19)–(22). Кроме того, из теорем 5.15.1, 5.15.2 следует

$$\lim_{k\to\infty} g_i^+(x_{0k}, x_k(T)) = g_i^+(x_{0*}, x_*(T)) = 0, \quad i = 1, \ldots, s. \quad (27)$$

Задача (19)–(22) представляет собой задачу оптимального управления с подвижным левым концом и свободным правым концом траектории и поэтому для приращения функции (19) нетрудно получить формулу, которая понадобится нам при получении необходимых условий оптимальности для задачи (19)–(22). Будем рассматривать задачи (19)–(22) со столь большими номерами k, чтобы

$$|x_{0k} - x_{0*}| < 1, \quad 0 \leqslant \xi_{lp}^k < d_N, \quad l, p = 1, \ldots, N; \quad (28)$$

это возможно в силу равенств (25), из которых вытекает, что неравенства (28) будут выполняться для всех $k \geqslant k_0$, где k_0 — достаточно большое число. Оптимальному решению (x_{0k}, ξ_k) задачи (19)–(22) в силу (28) можно дать такие малые ненулевые приращения $(\Delta x_0, \Delta\xi)$, что

$$|x_{0k}+\Delta x_0 - x_{0*}| < 1, \ 0 \leqslant \xi_{lp}^k + \Delta\xi_{lp} < d_N, \ \Delta\xi_{lp} \geqslant 0, \ l, p = 1, \ldots, N, \ k \geqslant k_0. \quad (29)$$

Тогда оптимальное управление $u_k(t) = u(t, \xi_k)$ и траектория $x_k(t) = x(t, u_k(\cdot), x_{0k})$ задачи (19)–(22) получат приращение $\Delta u(t) = u(t, \xi_k + \Delta\xi_k) - u_k(t)$, $\Delta x(t) = x(t, u_k+\Delta u, x_{0k}+\Delta x_0) - x_k(t)$, $t_0 \leqslant t \leqslant T$. Приращение $\Delta x(t)$ удовлетворяет условиям

$$\Delta\dot{x}(t) = \Delta f = f(x_k(t)+\Delta x(t), u_k(t)+\Delta u(t), t) - f(x_k(t), u_k(t), t), \quad t_0 \leqslant t \leqslant T, \quad \Delta x(t_0) = \Delta x_0, \quad (30)$$

вытекающим из (20). Из (24) для $\Delta x(t)$ следует оценка

$$\max_{t_0 \leqslant t \leqslant T} |\Delta x(t)| \leqslant c_1|\Delta x_0| + c_{2N}|\Delta\xi|. \quad (31)$$

С учетом оптимальности (x_{0k}, ξ_k) для приращения функции (19) получим

$$0 \leqslant \Delta\Phi_k = \Phi_k(x_{0k}+\Delta x_0, \xi_k+\Delta\xi) - \Phi_k(x_{0k}, \xi_k) =$$

$$= \int_{t_0}^{T} [f^0(x_k(t)+\Delta x(t), u_k(t)+\Delta u(t), t) - f^0(x_k(t), u_k(t), t)]dt +$$

$$+ g^0(x_{0k}+\Delta x_0, x_k(T)+\Delta x(T)) - g^0(x_{0k}, x_k(T)) +$$

$$+ A_k \sum_{i=1}^{s} [(g_i^+(x_{0k}+\Delta x_0, x_k(T)+\Delta x(T)))^2 - (g_i^+(x_0, x_k(T)))^2] =$$

$$= \int_{t_0}^{T} \Delta f^0 dt + \left\langle g_x^0(x_{0k}, x_k(T)) + \sum_{i=1}^{s} 2A_k g_i^+(x_{0k}, x_k(T)) g_x^i(x_{0k}, x_k(T)), \Delta x_0 \right\rangle +$$

$$+ \left\langle g_y^0(x_{0k}, x_k(T)) + \sum_{i=1}^{s} 2A_k g_i^+(x_{0k}, x_k(T)) g_y^i(x_{0k}, x_k(T)), \Delta x(T) \right\rangle + R_{1k}, \quad (32)$$

где

$$\Delta f^0 = f^0(x_k(t)+\Delta x(t), u_k(t)+\Delta u(t), t) - f^0(x_k(t), u_k(t), t),$$

$$R_{1k} = \left\langle \left(g_x^0\begin{pmatrix}\theta\\ \ldots\end{pmatrix} - g_x^0(\ldots)\right) + \sum_{i=1}^{s} 2A_k \left(g_i^+\begin{pmatrix}\theta\\ \ldots\end{pmatrix} g_x^i\begin{pmatrix}\theta\\ \ldots\end{pmatrix} - g_i^+(\ldots) g_x^i(\ldots)\right), \Delta x_0 \right\rangle +$$

$$+ \left\langle \left(g_y^0\begin{pmatrix}\theta\\ \ldots\end{pmatrix} - g_y^0(\ldots)\right) + \sum_{i=1}^{s} 2A_k \left(g_i^+\begin{pmatrix}\theta\\ \ldots\end{pmatrix} g_y^i\begin{pmatrix}\theta\\ \ldots\end{pmatrix} - g_i^+(\ldots) g_y^i(\ldots)\right), \Delta x(T) \right\rangle; \quad (33)$$

для краткости здесь обозначено $\begin{pmatrix}\theta\\ \ldots\end{pmatrix} = (x_{0k}+\theta\Delta x_0, x_k(T)+\theta\Delta x(T))$, $0 < \theta < 1$, $(\ldots) = (x_{0k}, x_k(T))$.

Положим

$$a_{0k} = \left[1 + \sum_{i=1}^{s} (2A_k g_i^+(x_{0k}, x_k(T)))^2\right]^{-1/2},$$

$$a_{ik} = 2A_k g_i^+(x_{0k}, x_k(T)) a_{0k}, \quad i = 1, \ldots, s, \quad a_k = (a_{0k}, \ldots, a_{sk}). \quad (34)$$

Отсюда и из (18) следует, что

$$0 < a_{0k} \leqslant 1, \quad a_{1k} \geqslant 0, \quad \ldots, a_{mk} \geqslant 0, \quad |a_k|^2 = \sum_{i=0}^{s} a_{ik}^2 = 1. \tag{35}$$

Для преобразования правой части равенства (32) к более удобному виду, введем функции

$$\begin{gathered} H(x, u, t, \psi, a_0) = -a_0 f^0(x, u, t) + \langle \psi, f(x, u, t)\rangle, \\ l(x, y, a) = a_0 g^0(x, y) + \ldots + a_s g^s(x, y) \end{gathered} \tag{36}$$

и сопряженную задачу

$$\dot{\psi}_k(t) = -H_x(x, u, t, \psi, a_0)|_{x = x_k(t),\, u = u_k(t),\, \psi = \psi_k,\, a_0 = a_{0k}}, \tag{37}$$

$$\psi_k(T) = -\sum_{i=0}^{s} a_{ik} g_y^i(x_{0k}, x_k(T)) = -\frac{\partial l(x_{0k}, x_k(T), a_k)}{\partial y}, \tag{38}$$

где a_{ik} взяты из (34). Напоминаем, что в соответствии с определением 1.1 решением задачи (37), (38) называется непрерывная вектор-функция $\psi_k(t)$, являющаяся решением интегрального уравнения

$$\psi_k(t) = \int_t^T H_x(x_k(\tau), u_k(\tau), \tau, \psi_k(\tau), a_{0k})\, d\tau + \psi_k(T). \tag{39}$$

Система (37) линейна относительно ψ_k; существование и единственность решения задачи (37), (38) следует из теоремы 1.2. При сделанных предположениях относительно функций $f^i(x, u, t)$, $i = 0, \ldots, n$, как видно из (39), функция $\psi_k(t)$ во всех точках непрерывности управления $u_k(t)$ непрерывно дифференцируема и удовлетворяет уравнению (37).

Умножим равенство (32) на $a_{0k} > 0$ и с учетом обозначений (34), (38) перепишем его в виде

$$0 \leqslant a_{0k}\Delta\Phi_k = \int_{t_0}^{T} a_{0k}\Delta f^0\, dt + \left\langle \sum_{i=0}^{s} a_{ik} g_x^i(x_{0k}, x_k(T)), \Delta x_0 \right\rangle - \\ - \langle \psi_k(T), \Delta x(T)\rangle + a_{0k} R_{1k}. \tag{40}$$

Преобразуем третье слагаемое из правой части (40). С учетом соотношений (30), (37), (38) имеем

$$\langle \psi_k(T), \Delta x(T)\rangle = \int_{t_0}^{T} \frac{d}{dt}\langle \psi_k(t), \Delta x(t)\rangle\, dt + \langle \psi_k(t_0), \Delta x_0\rangle = \\ = \int_{t_0}^{T} (\langle \psi_k(t), \Delta \dot{x}(t)\rangle + \langle \dot{\psi}_k(t), \Delta x(t)\rangle)\, dt + \langle \psi_k(t_0), \Delta x_0\rangle = \\ = \int_{t_0}^{T} \langle \psi_k(t), \Delta f\rangle\, dt - \int_{t_0}^{T} \langle H_x(x_k(t), u_k(t), t, \psi_k(t), a_{0k}), \Delta x(t)\rangle\, dt + \\ + \langle \psi_k(t_0), \Delta x_0\rangle. \tag{40.A}$$

Подставим (40.A) в (40); с учетом обозначений (36) будем иметь

$$0 \leqslant a_{0k}\Delta\Phi_k = -\int_{t_0}^{T} [H(x_k(t) + \Delta x(t), u_k(t) + \Delta u(t), t, \psi_k(t), a_{0k}) - \\ - H(x_k(t), u_k(t), t, \psi_k(t), a_{0k})]\, dt + \left\langle \frac{\partial l(x_{0k}, x_k(T), a_k)}{\partial x} - \psi_k(t_0), \Delta x_0 \right\rangle + \\ + \int_{t_0}^{T} \langle H_x(x_k(t), u_k(t), t, \psi_k(t), a_{0k}), \Delta x(t)\rangle\, dt + a_{0k} R_{1k}. \tag{41}$$

В силу формулы конечных приращений

$$H(x_k + \Delta x, u_k + \Delta u, t, \psi_k, a_{0k}) = H(x_k, u_k + \Delta u, t, \psi_k, a_{0k}) + \\ + \langle H_k(x_k + \theta_1 \Delta x, u_k + \Delta u, t, \psi_k, a_{0k}), \Delta x\rangle, \quad 0 < \theta_1 < 1. \tag{41.A}$$

Отсюда и из (41) получим

$$0 \leqslant a_{0k}\Delta\Phi_k = -\int_{t_0}^{T} [H(x_k(t), u_k(t) + \Delta u(t), t, \psi_k(t), a_{0k}) - \\ - H(x_k(t), u_k(t), t, \psi_k(t), a_{0k})]\, dt + \left\langle \frac{\partial l(x_{0k}, x_k(T), a_k)}{\partial x} - \psi_k(t_0), \Delta x_0 \right\rangle + R_k, \\ R_k = a_{0k} R_{1k} + R_{2k}, \tag{42}$$

где

$$R_{2k} = -\int_{t_0}^{T} \langle H_x(x_k(t) + \theta_1 \Delta x(t), u_k(t) + \Delta u(t), t, \psi_k(t), a_{0k}) - \\ - H_x(x_k(t), u_k(t), t, \psi_k(t), a_{0k}), \Delta x(t)\rangle\, dt. \tag{43}$$

Покажем, что

$$R_k = o_k(|\Delta x_0| + |\Delta \xi|), \quad \lim_{\alpha \to 0} o_k(\alpha)/\alpha = 0. \tag{44}$$

Из непрерывности функции $g^i(x, y)$, $g_i^+(x, y)$, $g_x^i(x, y)$, $g_y^i(x, y)$, оценки (31) и выражения (33) для R_{1k} следует, что $a_{0k} R_{1k} = o_k(|\Delta x_0| + |\Delta \xi|)$. Далее, перепишем выражение (43) для R_{2k} в виде $R_{2k} = R_{3k} + R_{4k}$, где

$$R_{3k} = -\int_{t_0}^{T} \langle H_x(x_k(t) + \theta_1 \Delta x(t), u(t, \xi_k + \Delta \xi), t, \psi_k(t), a_{0k}) - \\ - H_x(x_k(t), u(t, \xi_k + \Delta \xi), t, \psi_k(t), a_{0k}), \Delta x(t)\rangle\, dt, \tag{45}$$

$$R_{4k} = -\int_{t_0}^{T} \langle H_x(x_k(t), u(t, \xi_k + \Delta \xi), t, \psi_k(t), a_{0k}) - \\ - H_x(x_k(t), u(t, \xi_k), t, \psi_k(t), a_{0k}), \Delta x(t)\rangle\, dt.$$

Напомним, что управления $u(t, \xi_k)$, $u(t, \xi_k + \Delta \xi)$ определены согласно (13). Отсюда и из неравенств (21), (22), (28), (29), (31) следует, что аргументы x, u, t, ψ функции H_x в формулах (45) принадлежат ограниченному замкнутому множеству $Q_{kN} = \{(x, u, t, \psi)$: $|x| \leqslant \max\limits_{t_0 \leqslant t \leqslant T} |x_k(t)| + c_1(|x_{0*}| + 1) + c_{2N} N^2 d_N$, $|u| \leqslant \sup\limits_{t_0 \leqslant t \leqslant T} |u_*(t)| + \max\limits_{1 \leqslant p \leqslant N} |v_p|$, $t_0 \leqslant t \leqslant T$, $|\psi| \leqslant \max\limits_{t_0 \leqslant t \leqslant T} |\psi_k(t)|\}$.

Непрерывная функция $H_x(x, u, t, \psi, a_{0k})$ переменных (x, u, t, ψ) на компактном множестве Q_{kN} будет равномерно непрерывна на этом множестве. Отсюда и из оценки (31) следует, что $R_{3k} = o_k(|\Delta x_0| + |\Delta\xi|)$. Далее, с учетом определения (13) управлений $u(t, \xi_k)$, $u(t, \xi_k + \Delta\xi)$ имеем

$$|R_{4k}| = \left| \sum_{l,p=1}^{N} \int_{t_{lp}+\xi_{lp}^k}^{t_{lp}+\xi_{lp}^k+\Delta\xi_{lp}} \langle H_x(x_k(t), v_p, t, \psi_k(t), a_{0k}) - \right.$$
$$\left. - H_x(x_k(t), u_*(t), t, \psi_k(t), a_{0k}), \Delta x(t)\rangle dt \right| \leqslant$$
$$\leqslant 2|\Delta\xi| \sup_{Q_{kN}} |H_x(x, u, t, \psi, a_{0k})||\Delta x(t)| = o_k(|\Delta x_0| + |\Delta\xi|).$$

Суммируя полученные оценки для R_{1k}, R_{3k}, R_{4k}, придем к оценке (44).

Итак, нужная формула (42) для приращения функции (19) с оценкой остаточного члена (44) получена. Положим $\Delta\xi = 0$, $\Delta x_0 = -\varepsilon\left(\frac{\partial l(x_{0k}, x_k(T), a_k)}{\partial x} - \psi_k(t_0)\right)$, где число $\varepsilon > 0$ столь мало, что выполняется первое из неравенств (29); тогда $\Delta u(t) \equiv 0$ и из (42), (44) получим $0 \leqslant a_{0k}\Delta\Phi_k = -\varepsilon\left|\frac{\partial l(x_{0k}, x_k(T), a_k)}{\partial x} - \psi_k(t_0)\right|^2 + o_k(\varepsilon)$ или

$$\left|\frac{\partial l(x_{0k}, x_k(T), a_k)}{\partial x} - \psi_k(t_0)\right|^2 \leqslant \frac{o_k(\varepsilon)}{\varepsilon}.$$

Отсюда при $\varepsilon \to 0$ имеем

$$\psi_k(t_0) = \frac{\partial l(x_{0k}, x_k(T), a_k)}{\partial x}, \quad k \geqslant k_0. \tag{46}$$

Далее, положим в (42) $\Delta x_0 = 0$, матрицу $\Delta\xi = \{\Delta\xi_{lp}\}$ возьмем так, чтобы элемент, находящийся на пересечении произвольным образом фиксированных l-й строки и p-го столбца, $1 \leqslant l, p \leqslant N$, равнялся $\varepsilon > 0$, а остальные элементы равны нулю, причем ε возьмем столь малым, чтобы выполнялось второе неравенство (29). Тогда согласно (13)

$$\Delta u(t) = u(t, \xi_k + \Delta\xi) - u(t, \xi_k) = \begin{cases} v_p - u_k(t) = v_p - u_*(t) \\ \text{при } t_{lp} + \xi_{lp}^k < t \leqslant t_{lp} + \xi_{lp}^k + \varepsilon, \\ 0 \quad \text{в остальных точках из } [t_0, T], \end{cases}$$

и из (42), (44) получим

$$0 \leqslant a_{0k}\Delta\Phi_k = -\int_{t_{lp}+\xi_{lp}^k}^{t_{lp}+\xi_{lp}^k+\varepsilon} [H(x_k(t), v_p, t, \psi_k(t), a_{0k}) -$$
$$- H(x_k(t), u_*(t), t, \psi_k(t), a_{0k})]dt = o_k(\varepsilon). \tag{47}$$

Заметим, что подынтегральная функция $g_k(t) = H(x_k(t), v_p, t, \psi_k(t), a_{0k}) - H(x_k(t), u_*(t), t, \psi_k(t), a_{0k})$ непрерывна на отрезке $[t_{lp} + \xi_{lp}^k, t_{lp} + \xi_{lp}^k + \varepsilon]$. Применяя теорему о среднем, из (47) имеем $0 \leqslant -g_k(t_{lp} + \xi_{lp}^k + \theta_2\varepsilon)\varepsilon + o_k(\varepsilon)$, $0 < \theta_2 < 1$. Разделим это неравенство на $\varepsilon > 0$ и совершим предельный переход при $\varepsilon \to 0$. Получим $0 \leqslant -g_k(t_{lp} + \xi_{lp}^k)$ или

$$[H(x_k(t), v_p, t, \psi_k(t), a_{0k}) - H(x_k(t), u_*(t), t, \psi_k(t), a_{0k})]_{t = t_{lp}+\xi_{lp}^k} \leqslant 0 \tag{48}$$

при всех $k \geqslant k_0$, $l = 1, \ldots, N$, $p = 1, \ldots, N$.

Заметим, что соотношения (35), (37), (38), (46), (48), представляющие собой необходимое условие оптимальности в задаче (19)–(22), вполне аналогичны соотношениям (2.7)–(2.10) из теоремы 2.1. Соотношения (35), (37), (38), (46), (48) получены при фиксированном $N \geqslant 1$ и справедливы при каждом $k \geqslant k_0$. Для завершения доказательства теоремы 2.1 остается сначала перейти в этих соотношениях к пределу при $k \to \infty$, считая N фиксированным, затем совершить предельный переход при $N \to \infty$. Поскольку построенные выше последовательности $\{a_k\}$, $\{\psi_k(t)\}$ зависят от N, то их предельные точки, вообще говоря, также будут зависеть от N. В дальнейшем нам будет полезно явно подчеркнуть эту зависимость и поэтому упомянутые последовательности и их предельные точки ниже будем снабжать индексом N.

Согласно (35) последовательность $a_k = (a_{0k}, a_{1k}, \ldots, a_{sk}) = a_k(N)$, $k = 0, 1, \ldots$, ограничена и, пользуясь теоремой Больцано — Вейерштрасса, из нее можно выбрать подпоследовательность, сходящуюся к некоторой точке $a = a(N) = (a_0(N), a_1(N), \ldots, a_k(N))$. Без умаления общности дальнейших рассуждений можем считать, что сама последовательность $\{a_k(N)\}$ сходится к $a(N)$. Из (35) следует

$$0 \leqslant a_0(N) \leqslant 1, \quad a_1(N) \geqslant 0, \ldots, a_m(N) \geqslant 0, \quad |a(N)|^2 = \sum_{i=0}^{s} a_i^2(N) = 1. \tag{49}$$

Далее, пользуясь непрерывностью $g_y^i(x, y)$ и равенствами (25), (26), из (38) при $k \to \infty$ получим

$$\lim_{k\to\infty} \psi_k(T; N) = \psi(T; N) = -\frac{\partial l(x_{0*}, x_*(T), a(N))}{\partial y}. \tag{50}$$

Покажем, что последовательность $\{\psi_k(t)\} = \{\psi_k(t, N)\}$ равномерно на $[t_0, T]$ сходится к решению $\psi(t; N)$ системы уравнений

$$\dot\psi(t; N) = -H_x(x_*(t), u_*(t), t, \psi(t; N), a_0(N)), \quad t_0 \leqslant t \leqslant T, \tag{51}$$

с начальным условием (50). Как и в задаче (37), (38) решение задачи (51), (50) существует, единственно, является непрерывным решением интегрального уравнения

$$\psi(t; N) = \int_t^T H_x(x_*(\tau), u_*(\tau), \tau, \psi(\tau; N), a_0(N))d\tau + \psi(T; N), \tag{52}$$

во всех точках непрерывности оптимального управления $u_*(t)$ непрерывно дифференцируемо и удовлетворяет уравнению (51).

Обозначим $\Delta\psi_k(t) = \psi_k(t; N) - \psi(t; N)$, $t_0 \leqslant t \leqslant T$. Из (39), (52) с учетом определения (36) функции H и ее производной H_x имеем

$$\Delta\psi_k(t) = \int_t^T \sum_{i=1}^{n} \Delta\psi_{ik}(\tau) f_x^i(x_k(\tau), u_k(\tau), \tau)d\tau + b_k(t; N) + \Delta\psi_k(T), \tag{53}$$

где

$$b_k(t; N) = -\int_t^T (a_{0k}(N) - a_0(N)) f_x^0(x_k(\tau), u_k(\tau), \tau)d\tau -$$
$$- a_0(N)\int_t^T [f_x^0(x_k(\tau), u_k(\tau), \tau) - f_x^0(x_*(\tau), u_k(\tau), \tau)]d\tau -$$
$$- a_0(N)\int_t^T [f_x^0(x_*(\tau), u_k(\tau), \tau) - f_x^0(x_*(\tau), u_*(\tau), \tau)]d\tau +$$

$$+\int_t^T \sum_{i=1}^n \psi_i(\tau; N)[f_x^i(x_k(\tau), u_k(\tau), \tau) - f_x^i(x_*(\tau), u_k(\tau), \tau)]d\tau +$$
$$+\int_t^T \sum_{i=1}^n \psi_i(\tau; N)[f_x^i(x_*(\tau), u_k(\tau), \tau) - f_x^i(x_*(\tau), u_*(\tau), \tau)]d\tau. \quad (54)$$

Покажем, что $\{b_k(t; N)\} \to 0$ при $k \to \infty$ (N фиксировано!) равномерно на отрезке $[t_0, T]$. С этой целью заметим, что в силу (21), (22), (28), оценки (24) при $\xi = 0$, $\Delta\xi = \xi_k$ и определения (13) управления $u_k(t) = u(\cdot, \xi_k)$ аргументы (x, u, t) функций $f_x^i(x, u, t)$, входящих в (53), (54), принадлежат компактному множеству $Q_N = \{(x, u, t)$: $|x| \leqslant \max\limits_{t_0 \leqslant t \leqslant T} |x_*(t)| + c_1(|x_{0*}| + 1) + c_{2N}N^2 d_N$, $|u| \leqslant \sup\limits_{t_0 \leqslant t \leqslant T} |u_*(t)| + \sup\limits_{1 \leqslant p \leqslant N} |v_p|$, $t_0 \leqslant t \leqslant T\}$ при всех $k \geqslant k_0$. Непрерывные функции $f_x^i(x, u, t)$ на Q_N будут ограничены и равномерно непрерывны по совокупности аргументов $(x, u, t) \in Q_N$. Следовательно, $\max\limits_{0 \leqslant i \leqslant n} \max\limits_{(x, u, t) \in Q_N} |f_x^i(x, u, t)| = L_N < \infty$. Отсюда и из $\{a_{0k}(N)\} \to a_0(N)$ получаем, что 1-е слагаемое из правой части (54) стремится к нулю равномерно на $[t_0, T]$. Равномерная сходимость к нулю 2-го слагаемого из (54) следует из равномерной непрерывности $f^0(x, u, t)$ на Q_N и равномерной сходимости $\{x_k(t)\}$ к $x_*(t)$, вытекающей из оценки (24) при $\xi = 0$, $\Delta\xi = \xi_k$. Для 3-го слагаемого из (54) с учетом определения (13) управлений $u_k(t) = u(t, \xi_k)$, $u_*(t) = u(t; 0)$ имеем

$$\left|a_0(N) \int_t^T [f_x^0(x_*(\tau), u_k(\tau), \tau) - f_x^0(x_*(\tau), u_*(\tau), \tau)]d\tau\right| =$$
$$= a_0(N) \sum_{l,p=1}^N \int_{t_{lp}}^{t_{lp}+\xi_{lp}^k} |f_x^0(x_*(\tau), v_p, \tau) - f_x^0(x_*(\tau), u_*(\tau), \tau)|d\tau \leqslant$$
$$\leqslant a_0(N)|\xi_k| \cdot 2 \max_{(x, u, t) \in Q_N} |f_x^0(x, u, t)| \to 0$$

при $k \to \infty$ равномерно на $[t_0, T]$, так как $|\xi_k| \to 0$ в силу (25). Аналогично доказывается равномерная на $[t_0, T]$ сходимость при $k \to \infty$ 4-го и 5-го слагаемых из (54). Таким образом,

$$\lim_{k \to \infty} \sup_{t_0 \leqslant t \leqslant T} |b_k(t; N)| = 0. \quad (55)$$

Далее, из (53) имеем

$$|\Delta\psi_k(t)| \leqslant \int_t^T L_N \sum_{i=1}^n |\Delta\psi_{ik}(\tau)|d\tau + |b_k(t; N)| + |\Delta\psi_k(T)| \leqslant$$
$$\leqslant L_N\sqrt{n} \int_t^T L_N|\Delta\psi_k(\tau)|d\tau + |b_k(t; N)| + |\Delta\psi_k(T)|, \quad t_0 \leqslant t \leqslant T.$$

Как видно, функция $\varphi(t) = |\Delta\psi_k(t)|$ удовлетворяет неравенству (6) с $a = L_N\sqrt{n}$, $b(t) = |b_k(t; N)| + |\Delta\psi_k(T)|$. Тогда в силу (7) получаем

$$|\Delta\psi_k(t)| = |\psi_k(t; N) - \psi(t; N)| \leqslant$$
$$\leqslant L_N\sqrt{n} \int_t^T (|b_k(t; N)| + |\Delta\psi_k(T)|) \exp\{L_N\sqrt{n}(\tau - t)\}d\tau +$$
$$+ |b_k(t; N)| + |\Delta\psi_k(T)|, \quad t_0 \leqslant t \leqslant T.$$

Отсюда и из (50), (55) следует

$$\lim_{k \to \infty} \max_{t_0 \leqslant t \leqslant T} |\psi_k(t; N) - \psi(t; N)| = 0. \quad (56)$$

Из (46) с учетом (26), $\{a_k(N)\} \to a(N)$ при $k \to \infty$ тогда имеем

$$\lim_{k \to \infty} \psi_k(t_0; N) = \psi(t_0; N) = \frac{\partial l(x_{0*}, x_*(T), a(N))}{\partial x}. \quad (57)$$

Перейдем к пределу при $k \to \infty$ в неравенстве (48). При этом заметим, что

$$|x_k(t + \Delta t) - x_k(t)| \leqslant \left|\int_t^{t+\Delta t} f(x_k(\tau), u_k(\tau, \xi_k), \tau)d\tau\right| \leqslant L_N|\Delta t|,$$
$$\forall t, t + \Delta t \in [t_0, T] \quad \forall k \geqslant k_0,$$

где $L_N = \sup\limits_{Q_{1N}} |f(x, u, t)|$, $Q_{1N} = \{(x, u, t)$: $|x| \leqslant \max\limits_{t \in [t_0, T]} |x_*(t)| + 1$, $|u| \leqslant \sup\limits_{t \in [t_0, T]} |u_*(t)| + \sup\limits_{1 \leqslant p \leqslant N} |v_p|$, $t_0 \leqslant t \leqslant T\}$, k_0 — достаточно большой номер; здесь учтено, что $\max\limits_{t \in [t_0, T]} |x_k(t)| \leqslant \max\limits_{t \in [t_0, T]} |x_*(t)| + 1$ $\forall k \geqslant k_0$ в силу (26), $\sup\limits_{t \in [t_0, T]} |u(t, \xi_k)| \leqslant \sup\limits_{t \in [t_0, T]} |u_*(t)| + \max\limits_{1 \leqslant p \leqslant N} |v_p|$ в силу (13). Отсюда и из (25), (26) имеем $\lim\limits_{k \to \infty} x_k(t_{lp} + \xi_{lp}^k) = x_*(t_{lp})$. Аналогично, с использованием (25), (26), (37), (38), (56) доказывается, что $\lim\limits_{k \to \infty} \psi_k(t_{lp} + \xi_{lp}^k) = \psi(t_{lp}; N)$. Отсюда и из (48) при $k \to \infty$ с учетом равенства $\lim\limits_{k \to \infty} a_{0k}(N) = a_0(N)$ и непрерывности функции H по совокупности своих аргументов получим

$$[H(x_*(t), v_p, t, \psi(t; N), a_0(N)) - H(x_*(t), u_*(t), t, \psi(t; N), a_0(N))]_{t=t_{lp}} \leqslant 0 \quad (58)$$

для всех $l, p = 1, \ldots, N$.

Наконец, совершим предельный переход при $N \to \infty$ в соотношениях (49)–(51), (57), (58). Выбирая при необходимости подпоследовательность из $\{a(N)\}$, можем считать, что сама последовательность $\{a(N)\} \to a = (a_0, a_1, \ldots, a_s)$. Тогда из (49) сразу получим

$$0 \leqslant a_0 \leqslant 1, \quad a_1 \geqslant 0, \ \ldots, \ a_m \geqslant 0, \quad |a|^2 = \sum_{i=0}^s a_i^2 = 1. \quad (59)$$

Из (50) при $N \to \infty$ имеем

$$\lim_{N \to \infty} \psi(T; N) = \psi(T) = -\frac{\partial l(x_{0*}, x_*(T), a)}{\partial y}. \quad (60)$$

Покажем, что последовательность $\{\psi(t; N)\}$ равномерно на $[t_0, T]$ сходится к решению $\psi(t)$ системы уравнений

$$\dot{\psi}(t) = -H_x(x_*(t), u_*(t), t, \psi(t), a_0), \quad t_0 \leqslant t \leqslant T; \quad (61)$$

с начальным условием (60). Существование и единственность решения задачи (61), (60) следует из теоремы 1.1. Выпишем интегральное уравнение, которому удовлетворяет решение задачи (61), (60):

$$\psi(t) = \int_t^T H_x(x_*(\tau), u_*(\tau), \tau, \psi(\tau), a_0)d\tau + \psi(T).$$

Отсюда и из (52) для разности $\Delta\psi(t)=\psi(t;N)-\psi(t)$ имеем

$$\Delta\psi(t)=\int_t^T[(-a_0(N)+a_0)f_x^0(x_*(\tau),u_*(\tau),\tau)+\sum_{i=1}^n\Delta\psi_i(\tau)f_x^i(x_*(\tau),u_*(\tau),\tau)]d\tau+\Delta\psi(T),\quad t_0\leqslant t\leqslant T.$$

Тогда

$$|\Delta\psi(t)|\leqslant\int_t^T L\sqrt{n}|\Delta\psi(\tau)|d\tau+L(T-t_0)|a_0(N)-a_0|+|\Delta\psi(T)|,\quad t_0\leqslant t\leqslant T,$$

где $L=\max\limits_{0\leqslant i\leqslant n}\sup\limits_{t_0\leqslant\tau\leqslant T}|f_x^i(x_*(\tau),u_*(\tau),\tau)|$. Отсюда и из неравенств (6)–(8) следует

$$|\Delta\psi(t)|=|\psi(t;N)-\psi(t)|\leqslant$$
$$\leqslant\exp\{L\sqrt{n}(T-t_0)\}(L(T-t_0)|a_0(N)-a_0|+|\Delta\psi(T)|),\quad t_0\leqslant t\leqslant T.$$

Тогда в силу (60), $\{a_0(N)\}\to a_0$ будем иметь

$$\lim_{N\to\infty}\max_{t_0\leqslant t\leqslant T}|\psi(t;N)-\psi(t)|=0. \tag{62}$$

Из (57) с учетом (62), $\{a(N)\}\to a$ получаем

$$\lim_{N\to\infty}\psi(t_0;N)=\psi(t_0)=\frac{\partial l(x_{0*},x_*(T),a)}{\partial x}. \tag{63}$$

Далее заметим, что $|\psi(t+\Delta t;N)-\psi(t;N)|=\left|\int_t^{t+\Delta t}H_x(x_*(\tau),u_*(\tau),\tau,\psi(\tau;N),a_0(N))d\tau\right|\leqslant L_1|\Delta t|\ \ \forall t,\ t+\Delta t\in[t_0,T],\ N\geqslant N_0$, где $L_1=\sup\limits_Q|H_x(x_*(\tau),u_*(\tau),\tau,\psi,\mu)|$, $Q=\{(\psi,\mu):|\psi|\leqslant\max\limits_{t\in[t_0,T]}|\psi(t)|+1;\ |\mu|\leqslant a_0+1\}$; здесь учтены равенства (62) и $\lim\limits_{N\to\infty}a_0(N)=a_0$. Отсюда из (12), (62) имеем $\lim\limits_{N\to\infty}\psi(t_{l_p};N)=\psi(t_l)$. Переходя к пределу при $N\to\infty$, из (58) получаем

$$H(x_*(t_l),v_p,t_l,\psi(t_l),a_0)-H(x_*(t_l),u_*(t_l),t_l,\psi(t_l),a_0)\leqslant 0,\quad l=1,2,\ldots$$

В силу плотности последовательности точек $\{v_p\}$ во множестве V отсюда получаем

$$H(x_*(t_l),v,t_l,\psi(t_l),a_0)\leqslant H(x_*(t_l),u_*(t_l),t_l,\psi(t_l),a_0)\quad\forall v\in V,\ l=1,2,\ldots$$

Последовательность $\{t_l\}$ рациональных точек интервала (t_0,T), являющихся точками непрерывности оптимального управления $u_*(t)$, всюду плотна на отрезке $[t_0,T]$. Поэтому предыдущее неравенство справедливо для всех $t\in[t_0,T]$, в которых $u_*(t)$ непрерывно:

$$H(x_*(t),v,t,\psi(t),a_0)\leqslant H(x_*(t),u_*(t),t,\psi(t),a_0)\quad\forall v\in V.$$

Поскольку при $v=u_*(t)\in V$ здесь получаем равенство, то

$$\max_{v\in V}H(x_*(t),v,t,\psi(t),a_0)=H(x_*(t),u_*(t),t,\psi(t),a_0) \tag{64}$$

во всех точках непрерывности $u_*(t)$.

Далее, докажем, что

$$a_i g^i(x_{0*},x_*(T))=0,\quad i=1,\ldots,m. \tag{65}$$

Для тех номеров i, $1\leqslant i\leqslant m$, для которых $g^i(x_{0*},x_*(T))=0$, равенства (65), конечно, выполняются. Пусть $g^i(x_{0*},x_*(T))<0$. Из (25), (26) тогда следует: $g^i(x_{0k},x_k(T))<0$ при всех $k\geqslant k_0$, где k_0 достаточно большое число. Поэтому из формул (18), (34) получаем $a_{ik}=a_{ik}(N)=0$ для всех $k\geqslant k_0$. Тогда $\lim\limits_{k\to\infty}a_{ik}(N)=a_i(N)=0$ для каждого $N\geqslant 1$. Следовательно, $\lim\limits_{N\to\infty}a_i(N)=a_i=0$ для тех номеров i, $1\leqslant i\leqslant m$, для которых $g^i(x_{0*},x_*(T))<0$. Равенства (65) доказаны.

Соотношения (59)–(61), (63)–(65), составляющие основное содержание теоремы 2.1, установлены. Тем самым теорема 2.1 полностью доказана для случая, когда задача (2.1)–(2.4) имеет единственное решение $(x_{0*},u_*(t),x_*(t))$. Общий случай, когда задача (2.1)–(2.4) имеет более чем одно решение, легко сводится к рассмотренному случаю. А именно, пусть $(x_{0*},u_*(t),x_*(t))$ одно из решений задачи (2.1)–(2.4). Введем новую фазовую координату x^{n+1} и к системе (2.2) добавим еще одно уравнение

$$\dot x^{n+1}(t)=|u(t)-u_*(t)|=f^{n+1}(u(t),t),\quad t_0\leqslant t\leqslant T,\quad x^{n+1}(t_0)=x_0^{n+1}. \tag{66}$$

Введем функцию

$$J_1(x_0,x_0^{n+1},u(\cdot))=\int_{t_0}^T f^0(x(t),u(t),t)dt+g^0(x_0,x(T))+(x^{n+1}(t_0))^2+$$
$$+|x_0-x_{0*}|^2=\int_{t_0}^T f^0(x(t),u(t),t)dt+g_1^0(x_0,x(T),x^{n+1}(t_0)), \tag{67}$$

где $g_1^0(x,x^{n+1},y)=g^0(x,y)+(x^{n+1})^2+|x-x_{0*}|^2$. Рассмотрим задачу минимизации функции (67) при условиях (2.2)–(2.4), (66). Нетрудно видеть, что $J_1(x_0,x_0^{n+1},u(\cdot))>J(x_0,u(\cdot))\geqslant J_*$ для всех допустимых наборов $(x_0,x_0^{n+1},u(\cdot),x(\cdot),x^{n+1}(\cdot))$ задачи (67), (2.2)–(2.4), (66), для которых $x_0\neq x_{0*}$, $x_0^{n+1}\neq 0$, $u(\cdot)\neq u_*(\cdot)$. В то же время $J_1(x_{0*},0,u_*(\cdot))=J(x_{0*},u_*(\cdot))=J_*$. Это значит, что функция (67) при условиях (2.2)–(2.4), (66) достигает своей нижней грани на единственном допустимом наборе $(x_{0*},x_{0*}^{n+1}=0,u_*(\cdot),x_*(\cdot),x_*^{n+1}(\cdot)\equiv 0)$. Следовательно, для задачи (67), (2.2)–(2.4), (66) справедлив принцип максимума. Конечно, для полной строгости надо оговорить, что функция $f^{n+1}(u,t)=|u-u_*(t)|$, находящаяся в правой части уравнения (66), необязательно непрерывна по $t\in[t_0,T]$ и, строго говоря, не удовлетворяет условиям теоремы 2.1. Но, тем не менее, благодаря тому, что $f^{n+1}(u,t)$ кусочно-непрерывна по t, не зависит от x, удовлетворяет условию (1), нетрудно проследить, что все вышеприведенные рассуждения, приведшие к соотношениям (59)–(61), (63)–(65), сохраняют силу и для задачи (67), (2.2)–(2.4), (66). В этой задаче функция Гамильтона — Понтрягина имеет вид

$$H_1(x,x^{n+1},u,t,\psi,\psi_{n+1},a_0)=-a_0f^0(x,u,t)+\langle\psi,f(x,u,t)\rangle+$$
$$+\psi_{n+1}|u-u_*(t)|=H(x,u,t,\psi,a_0)+\psi_{n+1}|u-u_*(t)|.$$

Согласно уже доказанному случаю принципа максимума найдутся числа $a_0, a_1, \ldots, a_s$ и вектор-функция $(\psi(t), \psi_{n+1}(t))$, $t_0 \leqslant t \leqslant T$, такие, что

$$a = (a_0, a_1, \ldots, a_s) \neq 0, \quad a_0 \geqslant 0,\ a_1 \geqslant 0,\ \ldots,\ a_m \geqslant 0,$$

$$\dot{\psi}(t) = -H_{1x}(x_*(t), x_*^{n+1}(t), u_*(t), t, \psi(t), \psi_{n+1}(t), a_0) = -H_x(x_*(t), u_*(t), t, \psi(t), a_0), \quad t_0 \leqslant t \leqslant T,$$

$$\dot{\psi}_{n+1}(t) = -H_{1x^{n+1}} \equiv 0, \quad t_0 \leqslant t \leqslant T;$$

условие максимума

$$\max_{v \in V} H_1(x_*(t), x_*^{n+1}(t), v, t, \psi(t), \psi_{n+1}(t), a_0) = H_1(x_*(t), x_*^{n+1}(t), u_*(t), t, \psi(t), \psi_{n+1}(t), a_0)$$

выполняется во всех точках непрерывности $u_*(t)$; справедливы условия трансверсальности

$$\psi(t_0) = a_0 g^0_{1x}(x_{0*}, x_*^{n+1}(t_0) = 0, x_*(T)) + \sum_{j=1}^{s} a_j g^j_x(x_{0*}, x_*(T)) = \frac{\partial l(x_{0*}, x_*(T), a)}{\partial x},$$

$$\psi_{n+1}(t_0) = a_0 g^0_{1x^{n+1}} + \sum_{j=1}^{s} a_j g^j_{x^{n+1}} = 0,$$

$$\psi(T) = -a_0 g^0_{1y}(x_{0*}, x_*^{n+1}(t_0) = 0, x_*(T)) - \sum_{j=1}^{s} a_j g^j_y(x_{0*}, x_*(T)) = -\frac{\partial l(x_{0*}, x_*(T), a)}{\partial y},$$

$$\psi_{n+1}(T) = -a_0 g^0_{1y^{n+1}}(x_{0*}, x_*^{n+1}(t_0) = 0, x_*(T)) - \sum_{j=1}^{s} a_j g^j_{y^{n+1}}(x_{0*}, x_*(T)) = 0$$

и условия дополняющей нежесткости

$$a_i g_i(x_{0*}, x_*(T)) = 0, \quad i = 1, \ldots, m.$$

Из этих условий следует, что $\psi_{n+1}(t) \equiv 0$. Учитывая это равенство в полученных условиях, снова придем к соотношениям (59)–(61), (63)–(65) и в том случае, когда в задаче (2.1)–(2.4) оптимальное решение не единственно. Теорема 2.1 доказана. □

2. Теорема 2.2 доказывается аналогично теореме 2.1, поэтому мы здесь наметим лишь схему доказательства. Будем предполагать, что условие (1) выполнено при всех $t \in \mathbb{R}$. Сначала рассмотрим случай, когда задача (2.37)–(2.40) имеет единственное решение $(x_{0*}, u_*(t), x_*(t), t_{0*}, T_*)$, $t_{0*} < T_*$. Доопределим $u_*(t) \equiv u_*(t_{0*} + 0)$ при $t \leqslant t_{0*}$, $u_*(t) \equiv u_*(T_* - 0)$ при $t \geqslant T_*$. Как и выше, все рациональные точки, принадлежащие интервалу (t_{0*}, T_*), в которых управление $u_*(t)$ непрерывно, занумеруем каким-либо образом в виде последовательности $t_1, t_2, \ldots, t_l, \ldots$ В множестве V выберем некоторое счетное всюду плотное множество точек $v_1, v_2, \ldots, v_p, \ldots$ Зафиксируем произвольный номер $N \geqslant 1$ и для каждого $l = 1, 2, \ldots, N$ выберем точки $t_{lp} \in (t_{0*}, T_*)$, $p = 1, \ldots, N+1$, и матрицу $\xi = \{\xi_{lp}\}$, удовлетворяющие условиям (12), (12.A). По аналогии с (14)–(17) рассмотрим следующую конечномерную аппроксимацию задачи (2.37)–(2.40):

$$J_N(x_0, \xi, t_0, T) = J(x_0, u(\cdot, \xi), x(\cdot), t_0, T) = \int_{t_0}^{T} f^0(x(t), u(t, \xi), t)dt + g^0(x_0, x(T), t_0, T) \to \inf, \quad (68)$$

$$\dot{x}(t) = f(x(t), u(t, \xi), t), \quad t_0 \leqslant t \leqslant T, \quad x(t_0) = x_0, \quad (69)$$

$$g^i(x_0, x(T), t_0, T) \leqslant 0, \quad i = 1, \ldots, m,$$
$$g^i(x_0, x(T), t_0, T) = 0, \quad i = m+1, \ldots, s, \quad (70)$$
$$\xi = \{\xi_{lp}\}: 0 \leqslant \xi_{lp} \leqslant d_N,$$

$$u(t, \xi) = \begin{cases} v_p, & t_{lp} < t \leqslant t_{lp} + \xi_{lp}, \quad 1 \leqslant l, p \leqslant N, \\ u_*(t) & \text{для других} \quad t \in \mathbb{R}. \end{cases} \quad (71)$$

Так как уравнение $u_*(t)$ кусочно-непрерывно на отрезке $[t_{0*}, T_*]$, то найдутся столь малые положительные числа Δt_{0N}, ΔT_N, что $t_{0*} + \Delta t_{0N} < T_* - \Delta T_N$ и на отрезках $[t_{0*}, t_{0*} + \Delta t_{0N}]$, $[T_* - \Delta T_N, T_*]$ управление $u_*(t)$ непрерывно и эти отрезки не содержат точек $\{t_{lp}\}$, $l = 1, \ldots, N$, $p = 1, \ldots, N+1$. Задачу (68)–(71) будем рассматривать при дополнительных ограничениях

$$(x_0, \xi, t_0, T) \in U_0 = \{|x_0 - x_{0*}| \leqslant 1, \quad 0 \leqslant \xi_{lp} \leqslant d_N, \quad l, p = 1, \ldots, N, \quad |t_0 - t_{0*}| \leqslant \Delta t_{0N}, \quad |T - T_*| \leqslant \Delta T_N\}. \quad (72)$$

Любой набор (x_0, ξ, t_0, T), допустимый в задаче (68)–(72), порождает допустимый набор $(x_0, u(t) = u(t, \xi), x(t), t_0, T)$ задачи (2.37)–(2.40), причем значения целевых функций в обеих задачах на таких наборах совпадают. В то же время оптимальное решение $(x_{0*}, u_*(t), x_*(t), t_{0*}, T_*)$ задачи (2.37)–(2.40) соответствует допустимому набору $(x_{0*}, \xi_* = 0, t_{0*}, T_*)$ задачи (68)–(72). Рассуждая также, как в задаче (14)–(17), отсюда заключаем, что набор $(x_{0*}, \xi_* = 0, t_{0*}, T_*)$ является единственным решением задачи (68)–(72), причем $J_N(x_{0*}, \xi_*, t_{0*}, T_*) = J_*$.

Применяя к задаче (68)–(72) метод штрафных функций, придем к задаче

$$\Phi_k(x_0, \xi, t_0, T) = J_N(x_0, \xi, t_0, T) + k \sum_{i=1}^{s} (g^+_{iN}(x_0, \xi, t_0, T))^2 = \int_{t_0}^{T} f^0(x(t), u(t, \xi), t)dt + g^0(x_0, x(T), t_0, T) + k \sum_{i=1}^{s} (g_i^+(x_0, x(T), t_0, T))^2 \to \inf \quad (73)$$

$$\dot{x}(t) = f(x(t), u(t, \xi), t), \quad t_0 \leqslant t \leqslant T, \quad x(t_0) = x_0 \quad (74)$$

$$(x_0, \xi, t_0, T) \in U_0, \quad (75)$$

где по аналогии с (18) обозначено $g_{iN}(x_0, \xi, t_0, T) = g^i(x_0, x(T, u(\cdot, \xi), x_0, t_0), t_0, T)$,

$$g_i^+(x, y, t_0, T) = \begin{cases} \max\{0; g^i(x, y, t_0, T)\}, & i = 1, \ldots, m, \\ g^i(x, y, t_0, T), & i = m+1, \ldots, s. \end{cases}$$

При сделанных предположениях относительно функции $f(x,u,t)$ в силу теоремы 1.1 решение задач Коши (69), (74) существует на любом конечном отрезке $[a, b]$ и единственно для всех (x_0, ξ, t_0), $|x_0 - x_{0*}| \leqslant 1$, $0 \leqslant \xi_{lp} \leqslant d_N$, $|t_0 - t_{0*}| \leqslant \Delta t_{0N}$. Зафиксируем какие-либо a, b так, чтобы $a < t_{0*} < T_* < b$, и все управления $u(t, \xi)$ и траектории $x(t) = x(t, u(\cdot, \xi), x_0, t_0)$ в задачах (68)–(72), (73)–(75) будем рассматривать лишь при $t \in [a, b]$. Величины Δt_{0N},

ΔT_N в (72) можем считать столь малыми, что $\Delta t_{0N} < t_{0*} - a$, $\Delta T_N < b - T_*$. Справедливы неравенства:

$$\sup_{a\leqslant t\leqslant b}|u(t,\xi)| \leqslant \sup_{a\leqslant t\leqslant b}|u_*(t)| + \max_{1\leqslant p\leqslant N}|v_p| = c_{1N}, \tag{76}$$

$$\max_{a\leqslant t\leqslant b}|x(t,u(\cdot,\xi),x_0,t_0)| \leqslant \max_{a\leqslant t\leqslant b}|x_*(t)| + c_{2N}(1+\Delta t_{0N}+d_N) = c_{3N}, \tag{77}$$

$$\max_{a\leqslant t\leqslant b}|x(t,u(\cdot,\xi),x_0,t_0) - x_*(t)| \leqslant c_{2N}(|x_0-x_{0*}| + |t_0-t_{0*}| + |\xi|), \quad |\xi| = \max_{1\leqslant l,p\leqslant N}|\xi_{lp}|, \tag{78}$$

$$\max_{a\leqslant t\leqslant b}|x(t,u(\cdot,\xi+\Delta\xi),x_0+\Delta x_0,t_0+\Delta t_0) - x(t,u(\cdot,\xi),x_0,t_0)| \leqslant c_{4N}(|\Delta x_0| + |\Delta t_0| + |\Delta\xi|), \tag{79}$$

$$|x(t_0+\Delta t_0,u(\cdot,\xi+\Delta\xi),x_0+\Delta x_0,t_0+\Delta t_0) - x(t_0,u(\cdot,\xi),x_0,t_0)| \leqslant c_{5N}(|\Delta x_0| + |\Delta t_0| + |\Delta\xi|), \tag{80}$$

$$|x(T+\Delta T,u(\cdot,\xi+\Delta\xi),x_0+\Delta x_0,t_0+\Delta t_0) - x(T,u(\cdot,\xi),x_0,t_0)| \leqslant c_{6N}(|\Delta x_0| + |\Delta t_0| + |\Delta T| + |\Delta\xi|), \tag{81}$$

$$|x(t,u(\cdot,\xi),x_0,t_0) - x(\tau,u(\cdot,\xi),x_0,t_0)| \leqslant c_{7N}|t-\tau|, \quad t,\tau\in[a,b], \tag{82}$$

где через c_{iN} обозначаются положительные постоянные, не зависящие от Δx_0, Δt_0, ΔT, $\Delta\xi$, но, вообще говоря, зависящие от N; в (76)–(82) предполагается, что (x_0,ξ,t_0,T), $(x_0+\Delta x_0,\xi+\Delta\xi,t_0+\Delta t_0,T+\Delta T)\in U_0$. Оценка (76) вытекает из определения $u(t,\xi)$ (см. (71)), (77) является следствием (78), оценки (78), (79) получаются, как и аналогичные (11), (24), с помощью леммы 1. При оценке интегралов вида $\int_{t_0}^{t_0+\Delta t_0} f(x(t,u(\cdot,\xi),x_0,t_0),u(t,\xi),t)dt$, $\int_{T}^{T+\Delta T} f(x(t,u(\cdot,\xi),x_0,t_0),u(t,\xi),t)dt$, $\int_{\tau}^{t} f(x(\theta,u(\cdot,\xi),x_0,t_0),u(\theta,\xi),\theta)d\theta$, возникающих при доказательстве неравенств (80)–(82), нужно иметь в виду, что здесь аргументы непрерывной функции $f(x,u,t)$ в силу оценок (76), (77) принадлежат компактному множеству $Q_N = \{(x,u,t)\colon |x|\leqslant c_{3N}, |u|\leqslant c_{1N}, t\in[a,b]\}$ и поэтому $\sup_{Q_N}|f(x,u,t)| < \infty$.

Вернемся к задаче (73)–(75). Из оценок (76)–(82) следует, что функция $\Phi_k(x_0,\xi,t_0,T)$ непрерывна на компактном множестве U_0 и в силу теоремы 2.1.1 задача (73)–(75) при каждом фиксированном k, N имеет хотя бы одно решение $(x_{0k},\xi_k,t_{0k},T_k)\in U_0$. Нетрудно проверить, что здесь выполнены все условия теорем 5.15.1, 5.15.2, из которых следует

$$\lim_{k\to\infty}x_{0k} = x_{0*}, \quad \lim_{k\to\infty}\xi_k = \xi_* = 0, \quad \lim_{k\to\infty}t_{0k} = t_{0*}, \quad \lim_{k\to\infty}T_k = T_*, \tag{83}$$

так как задача (68)–(72) имеет единственное решение $(x_{0*},\xi_*=0,t_{0*},T_*)$. Оптимальную траекторию $x(t,u(\cdot,\xi_k),x_{0k},t_{0k})$, $t_{0k}\leqslant tT_k$, задачи (73)–(75) для краткости будем обозначать через $x_k(t)$. Из (83), оценки (78) при $\xi=\xi_k$, $x_0=x_{0k}$, $t_0=t_{0k}$ и оценок (80)–(82) при $\xi=0$, $x_0=x_{0*}$, $t_0=t_{0*}$, $T=T_*$, $\Delta\xi=\xi_k$, $\Delta x_0 = x_{0k}-x_{0*}$, $\Delta t_0 = t_{0k}-t_{0*}$, $\Delta T = T_k - T_*$ имеем

$$\lim_{k\to\infty}\max_{a\leqslant t\leqslant b}|x_k(t)-x_*(t)|=0, \quad \lim_{k\to\infty}x_k(t_{0k})=x_*(t_{0*}), \quad \lim_{k\to\infty}x_k(T_k)=x_*(T_*). \tag{84}$$

Из (83) следует, что $|x_{0k}-x_{0*}|<1$, $|\xi_k|<d_N$, $|t_{0k}-t_{0*}|<\Delta t_{0N}$, $|T_k-T_*|<\Delta T_N$ $\forall k\geqslant k_0$, т. е. $(x_{0k},\xi_k,t_{0k},T_k)\in\operatorname{int}U_0$, $k\geqslant k_0$.

Решению $(x_{0k},\xi_k,t_{0k},T_k)$ задачи (73)–(75) дадим приращение $(\Delta x_0,\Delta\xi,\Delta t_0,\Delta T)$ такое, что $(x_{0k}+\Delta x_0,\xi_k+\Delta\xi,t_{0k}+\Delta t_0,T_k+\Delta T)\in U_0$. Тогда оптимальное управление $u_k(t)=u(t,\xi_k)$ и траектория $x_k(t)$ получает приращение $\Delta u(t)=u(t,\xi_k+\Delta\xi)-u_k(t)$, $\Delta x(t)=x(t,u(\cdot,\xi_k+\Delta\xi),x_{0k}+\Delta x_0,t_{0k}+\Delta t_0)-x_k(t)$, причем

$$\begin{gathered}\Delta\dot x(t)=\Delta f=f(x_k(t)+\Delta x(t),u_k(t)+\Delta u(t),t)-f(x_k(t),u_k(t),t),\\ a\leqslant t\leqslant b,\\ \Delta x(t_{0k})=x(t_{0k},u(\cdot,\xi_k+\Delta\xi),x_{0k}+\Delta x_0,t_{0k}+\Delta t_0)-x(t_{0k}).\end{gathered} \tag{85}$$

Из оценок (76)–(82) следует, что

$$\begin{gathered}\max\{\sup_{a\leqslant t\leqslant b}|u(t,\xi_k+\Delta\xi)|,\ \sup_{a\leqslant t\leqslant b}|u(t,\xi_k)|\}\leqslant c_{1N},\\ \max\{\max_{a\leqslant t\leqslant b}|x_k(t)+\Delta x(t)|,\ \max_{a\leqslant t\leqslant b}|x_k(t)|\}\leqslant c_{3N},\\ \max_{a\leqslant t\leqslant b}|\Delta x(t)|\leqslant c_{4N}(|\Delta x_0|+|\Delta t_0|+|\Delta\xi|),\end{gathered} \tag{86}$$

$$|x_k(t_{0k}+\Delta t_0)+\Delta x(t_{0k}+\Delta t_0)-x_k(t_{0k})|\leqslant c_{5N}(|\Delta x_0|+|\Delta t_0|+|\Delta\xi|),$$
$$|x_k(T_k+\Delta T)+\Delta x(T_k+\Delta T)-x_k(T_k)|\leqslant c_{6N}(|\Delta x_0|+|\Delta t_0|+|\Delta T|+|\Delta\xi|).$$

С учетом оптимальности набора $(x_{0k},\xi_k,t_{0k},T_k)$ для приращения функции (73) получим:

$$\begin{aligned}0\leqslant\Delta\Phi_k &= \Phi_k(x_{0k}+\Delta x_0,\xi_k+\Delta\xi,t_{0k}+\Delta t_0,T_k+\Delta T)-\\ &-\Phi_k(x_{0k},\xi_k,t_{0k},T_k)=\int_{t_{0k}}^{T_k}\Delta f^0dt+\langle\Delta x_0,g_x^0(x_{0k},x_k(T_k),t_{0k},T_k)+\\ &+\sum_{i=1}^{s}2kg_i^+(x_{0k},x_k(T_k),t_{0k},T_k)\cdot g_x^i(x_{0k},x_k(T_k),t_{0k},T_k)\rangle+\\ &+\langle\Delta x(T_k),g_y^0(x_{0k},x_k(T_k),t_{0k},T_k)+\sum_{i=1}^{s}2kg_i^+(x_{0k},x_k(T_k),t_{0k},T_k)\times\\ &\times g_y^i(x_{0k},x_k(T_k),t_{0k},T_k)\rangle+\Delta t_0[-f^0(x_k(t_{0k}),u_*(t_{0k}),t_{0k})+\\ &+g_t^0(x_{0k},x_k(T_k),t_{0k},T_k)+\sum_{i=1}^{s}2kg_i^+(x_{0k},x_k(T_k),t_{0k},T_k)\times\\ &\times g_t^i(x_{0k},x_k(T_k),t_{0k},T_k)]+\Delta T[f^0(x_k(T_k),u_*(T_k),T_k)+\\ &+\langle g_y^0(x_{0k},x_k(T_k),t_{0k},T_k)+\sum_{i=1}^{s}2kg_i^+(x_{0k},x_k(T_k),t_{0k},T_k)\times\\ &\times g_y^i(x_{0k},x_k(T_k),t_{0k},T_k),f(x_k(T_k),u_*(T_k),T_k)\rangle+\\ &+g_T^0(x_{0k},x_k(T_k),t_{0k},T_k)+\sum_{i=1}^{s}2kg_i^+(x_{0k},x_k(T_k),t_{0k},T_k)\times\\ &\times g_T^i(x_{0k},x_k(T_k),t_{0k},T_k)]+R_{1k},\end{aligned} \tag{87}$$

где $\Delta f^0 = f^0(x_k(t)+\Delta x(t), u_k(t)+\Delta u(t), t) - f^0(x_k(t), u_k(t), t)$, $R_{1k} = o_k(|\Delta x_0|+|\Delta\xi|+|\Delta t_0|+|\Delta T|)$, $\lim\limits_{\alpha\to 0} o_k(\alpha)/\alpha = 0$. Формула (87) обобщает формулу (32) на случай задач с незакрепленным временем и доказывается аналогично на основе оценок (86) с учетом непрерывности функций $g_x^i, g_y^i, g_t^i, g_T^i, f^j$. В частности, при выводе (87) учтено, что $u_k(t_{0k}) = u_*(t_{0k})$, $u_k(T_k) = u_*(T_k)$,

$$\int\limits_{t_{0k}}^{t_{0k}+\Delta t_0} f^0(x_k(t)+\Delta x(t), u_k(t)+\Delta u(t), t)dt = f^0(x_k(t_{0k}), u_*(t_{0k}), t_{0k})\Delta t_{0k} + o_k(|\Delta x_0|+|\Delta t_0|+|\Delta\xi|),$$

$$\int\limits_{T_k}^{T_k+\Delta T} f^0(x_k(t)+\Delta x(t), u_k(t)+\Delta u(t), t)dt = f^0(x_k(T_k), u_*(T_k), T_k)\Delta T + o_k(|\Delta x_0|+|\Delta t_0|+|\Delta T|+|\Delta\xi|),$$

$$x_k(T_k+\Delta T)+\Delta x(T_k+\Delta T) - x_k(T_k) = (x(T_k+\Delta T, u(\cdot, \xi_k+\Delta\xi), x_{0k}+\Delta x_0, t_{0k}+\Delta t_0) - x(T_k, u(\cdot, \xi_k+\Delta\xi), x_{0k}+\Delta x_0, t_{0k}+\Delta t_0)) + \Delta x(T_k) = \int\limits_{T_k}^{T_k+\Delta T} f(x(t, u(\cdot, \xi_k+\Delta\xi), x_{0k}+\Delta x_0, t_{0k}+\Delta t_0),$$

$$u(t, \xi_k+\Delta\xi), t)dt + \Delta x(T_k) = f(x_k(T_k), u_*(T_k), T_k)\Delta T + \Delta x(T_k) + o_k(|\Delta x_0|+|\Delta t_0|+|\Delta T|+|\Delta\xi|).$$

Обозначим

$$a_{0k} = \left(1 + \sum_{i=1}^{s} (2kg_i^+(x_{0k}, x_k(T_k), t_{0k}, T_k))^2\right)^{-1/2}, \qquad (88)$$

$$a_{ik} = a_{0k} 2kg_i^+(x_{0k}, x_k(T_k), t_{0k}, T_k), \quad i=1, 2, \ldots, s; \quad a_k = (a_{0k}, a_{1k}, \ldots, a_{sk}).$$

Нетрудно видеть, что

$$0 < a_{0k} \leqslant 1, \quad a_{1k} \geqslant 0, \ldots, a_{mk} \geqslant 0, \quad |a_k|^2 = \sum_{i=0}^{s} a_{ik}^2 = 1. \qquad (89)$$

Для дальнейшего преобразования формулы (87) по аналогии с (36)–(38) воспользуемся функциями

$$H(x, u, t, \psi, a_0) = -a_0 f^0(x, u, t) + \langle\psi, f(x, u, t)\rangle,$$
$$l(x, y, t, T, a) = \sum_{i=0}^{s} a_i g^i(x, y, t, T) \qquad (90)$$

и сопряженной задачей

$$\dot\psi_k(t) = -H_x(x_k(t), u_k(t), t, \psi_k(t), a_{0k}), \quad t \in [a, b],$$
$$\psi_k(T_k) = -\sum_{i=0}^{s} a_{ik} g_y^i(x_{0k}, x_k(T_k), t_{0k}, T_k) = -\frac{\partial l((x_{0k}, x_k(T_k), t_{0k}, T_k, a_k)}{\partial y}. \qquad (91)$$

Существование решения задачи (91) на всем отрезке $[a, b]$ и его единственность доказывается также, как в задаче (37), (38). Справедливо равенство

$$\langle\psi_k(T_k), \Delta x(T_k)\rangle = \int\limits_{t_{0k}}^{T_k} \langle\psi_k(t), f(x_k(t)+\Delta x(t), u_k(t)+\Delta u(t), t) - f(x_k(t), u_k(t), t)\rangle dt + \langle\psi_k(t_{0k}), \Delta x_0\rangle - \langle\psi_k(t_{0k}), f(x_k(t_{0k}), u_*(t_{0k}), t_{0k})\rangle\Delta t_0 -$$

$$- \int\limits_{t_{0k}}^{T_k} \langle H_x(x_k(t), u_k(t), t, \psi_k(t), a_{0k}), \Delta x(t)\rangle dt + o_k(|\Delta x_0|+|\Delta\xi|+|\Delta t_0|+|\Delta T|), \qquad (92)$$

которое устанавливается аналогично (40.A) с учетом (85), (91) и равенства

$$\Delta x(t_{0k}) = \int\limits_{t_{0k}+\Delta t_0}^{t_{0k}} f(x(t, u(\cdot, \xi_k+\Delta\xi), x_{0k}+\Delta x_0, t_{0k}+\Delta t_0), u(t, \xi_k+\Delta\xi), t)dt + \Delta x_0 = f(x_k(t_{0k}), u_*(t_{0k}), t_{0k})(-\Delta t_0) + \Delta x_0 + o_k(|\Delta x_0|+|\Delta\xi|+|\Delta t_0|).$$

Умножим неравенство (87) на $a_{0k} > 0$ и преобразуем правую часть полученного выражения с учетом (41.A), (88), (90), (92). Получим

$$0 \leqslant a_{0k}\Delta\Phi_k = -\int\limits_{t_{0k}}^{T_k} [H(x_k(t), u_k(t)+\Delta u(t), t, \psi_k(t), a_{0k}) - H(x_k(t), u_k(t), t, \psi_k(t), a_{0k})]dt + \langle\Delta x_0, \frac{\partial l(x_{0k}, x_k(T_k), t_{0k}, T_k, a_k)}{\partial x} - \psi_k(t_{0k})\rangle + \Delta t_0\left[H(x_k(t_{0k}), u_*(t_{0k}), t_{0k}, \psi_k(t_{0k}), a_{0k}) + \frac{\partial l(x_{0k}, x_k(T_k)t_{0k}, T_k, a_k)}{\partial t}\right] + \Delta T\left[-H(x_k(T_k), u_*(T_k), T_k, \psi_k(T_k), a_{0k}) + \frac{\partial l(x_{0k}, x_k(T_k), t_{0k}, T_k, a_k)}{\partial T}\right] + o_k(|\Delta x_0|+|\Delta\xi|+|\Delta t_0|+|\Delta T|), \quad k = 1, 2, \ldots \qquad (93)$$

Так как точка $z_k = (x_{0k}, \xi_k, t_{0k}, T_k)$ минимума функции $\Phi_k(x_0, \xi, t_0, T)$ на множестве U_0 является внутренней точкой этого множества при всех $k \geqslant k_0$, то $\frac{\partial\Phi_k}{\partial x_0}\Big|_{z_k} = 0$, $\frac{\partial\Phi_k}{\partial t_0}\Big|_{z_k} = 0$, $\frac{\partial\Phi_k}{\partial T}\Big|_{z_k} = 0$ $\forall k \geqslant k_0$. Отсюда и из (93) имеем:

$$\psi_k(t_{0k}) = \frac{\partial l(x_{0k}, x_k(T_k), t_{0k}, T_k, a_k)}{\partial x},$$
$$H(x_k(t_{0k}), u_*(t_{0k}), t_{0k}, \psi_k(t_{0k}), a_{0k}) = -\frac{\partial l(x_{0k}, x_k(T_k)t_{0k}, T_k, a_k)}{\partial t}, \qquad (94)$$
$$H(x_k(T_k), u_*(T_k), T_k, \psi_k(T_k), a_{0k}) = \frac{\partial l(x_{0k}, x_k(T_k)t_{0k}, T_k, a_k)}{\partial T}, \quad \forall k \geqslant k_0.$$

Кроме того, рассуждая также, как при выводе неравенства (48), из (93) получим

$$[H(x_k(t), v_p, t, \psi_k(t), a_{0k}) - H(x_k(t), u_*(t), t, \psi_k(t), a_{0k})]\Big|_{t=t_{lp}+\xi_{lp}} \leqslant 0 \qquad (95)$$
$$\forall k \geqslant k_0, \quad l, p = 1, \ldots, N.$$

Далее, пользуясь соотношениями (83), (84), (88), (89), рассуждая также, как при доказательстве теоремы 2.1 (см. текст после формулы (48)), совершим предельные переходы в (91), (94), (95) сначала при $k \to \infty$, затем при $N \to \infty$ и получим все утверждения теоремы 2.2. Напоминаем, что теоре-

ма 2.2 пока доказана в предположении, что задача (2.37)–(2.40) имеет единственное решение. Если эта задача имеет неединственное решение, то перейдем к задаче минимизации функции

$$J_1(x_0, x_0^{n+1}, u(\cdot), t_0, T) = J(x_0, u(\cdot), t_0, T) + \\ + (x_0^{n+1})^2 + |x_0 - x_{0*}|^2 + (t_0 - t_{0*})^2 + (T - T_*)^2$$

при условиях (2.38)–(2.40), (66), которая имеет единственное решение; применим к ней уже доказанное, а затем, учитывая, что оптимальные $x_*^{n+1}(t) \equiv 0$, $\psi_{n+1}(t) \equiv 0$, убедимся в справедливости теоремы 2.2 в общем случае.

3. Условие максимума (2.9)

$$\max_{v \in V} H(x(t), v, t, \psi(t), a_0) = H(x(t), u(t), t, \psi(t), a_0) \tag{96}$$

выше было установлено лишь в точках непрерывности оптимального управления. В следующей теореме показывается, как можно продолжить это равенство на весь отрезок $[t_0, T]$, устанавливается непрерывность функции

$$H(t) = \sup_{v \in V} H(x(t), v, t, \psi(t), a_0), \quad t \in [t_0, T], \tag{97}$$

обсуждаются ее дифференциальные свойства.

Теорема 1. *Пусть вычислены все условия теоремы* 2.2, *пусть* $(x_0, u(\cdot), x(\cdot), t_0, T)$ *— решение задачи* (2.37)–(2.40). *Тогда функция* (97) *непрерывна на отрезке* $[t_0, T]$, *причем верхняя грань в* (97) *реализуется как при* $v = u(t+0)$, *так и при* $v = u(t-0)$ *для всех* $t \in [t_0, T]$ (*здесь по определению принимается, что* $u(t_0 - 0) = u(t_0 + 0)$, $u(T+0) = u(T-0)$). *Если в дополнение к условиям теоремы 2.2 функции* $f^i(x, u, t)$, $i = 0, 1, \ldots, n$, *имеют частные производные* $f_t^i(x, u, t)$, *непрерывные по совокупности аргументов* $(x, u, t) \in E^n \times E^r \times \mathbb{R}$, *то функция* (97) *имеет левую и правую производные во всех точках* $t \in [t_0, T]$, *причем*

$$\frac{dH(t \pm 0)}{dt} = H_t(x(t), u(t \pm 0), t, \psi(t), a_0) \quad t \in [t_0, T], \tag{98}$$

где $H_t(x,u,t,\psi,a_0) = -a_0 f_t^0(x,u,t) + \langle \psi, f_t(x,u,t) \rangle$ *— частная производная функции* H *по переменной* t; *производная* $\frac{dH(t)}{dt}$ *существует и непрерывна во всех точках непрерывности оптимального управления* $u(\cdot)$.

Доказательство. Из (96), (97) следует

$$H(x(\tau), v, \tau, \psi(\tau), a_0) \leqslant H(\tau) = H(x(\tau), u(\tau), \tau, \psi(\tau), a_0) \tag{99}$$

во всех точках $\tau \in (t_0, T)$, в которых управление $u(\cdot)$ непрерывно. Это значит, что в точках непрерывности $u(\cdot)$ функция $H(\cdot)$ непрерывна. Пусть $t \in (t_0, T]$ — точка разрыва управления $u(\cdot)$. В силу кусочной непрерывности $u(\cdot)$ найдется $\delta > 0$, что $u(\cdot)$ непрерывна во всех точках $\tau \in [t_0, T]$, $t - \delta < \tau < t$. Поэтому из (99) при $\tau \to t - 0$ имеем

$$H(x(t), v, t, \psi(t), a_0) \leqslant H(t-0) = H(x(t), u(t-0), t, \psi(t), a_0) \quad \forall v \in V.$$

Переходя к верхней грани по $v \in V$ отсюда получим

$$H(t) \leqslant H(t-0) = H(x(t), u(t-0), t, \psi(t), a_0). \tag{100}$$

С другой стороны, учитывая включение $u(\tau) \in V$, имеем $H(t) \geqslant H(x(t), u(\tau), t, \psi(t), a_0)$, $t - \delta < \tau < t$. При $\tau \to t - 0$ отсюда следует

$$H(t) \geqslant H(x(t), u(t-0), t, \psi(t), a_0) = H(t-0).$$

Сравнивая это неравенство с (100) получаем

$$H(t) = H(t-0) = H(x(t), u(t-0), t, \psi(t), a_0), \tag{101}$$

т. е. функция $H(t)$ непрерывна слева в точке $t \in (t_0, T]$ разрыва $u(\cdot)$. С учетом (99) заключаем, что равенство (101) справедливо при всех $t \in (t_0, T]$. Аналогично доказывается, что функция $H(t)$ непрерывна справа на промежутке $[t_0, T)$ и имеет место равенство

$$H(t) = H(t+0) = H(x(t), u(t+0), t, \psi(t), a_0) \quad \forall t \in [t_0, T). \tag{102}$$

Таким образом, непрерывность функции (97) на отрезке $[t_0, T]$ установлена.

Для изучения дифференциальных свойств функции (97), следуя [44], введем в рассмотрение так называемую v-задачу:

$$J_1(\widetilde{x}_0, \widetilde{u}(\cdot), v(\cdot), \widetilde{x}(\cdot), \chi(\theta), \theta_1, \theta_2) = \int_{\theta_1}^{\theta_2} f^0(\widetilde{x}(\theta), \widetilde{u}(\theta), \chi(\theta) \cdot (1 + v(\theta)) d\theta + \\ + g^0(\widetilde{x}(\theta_1), \widetilde{x}(\theta_2), \chi(\theta_1), \chi(\theta_2)) \to \inf, \tag{103}$$

$$\frac{d\widetilde{x}(\theta)}{d\theta} = f(\widetilde{x}(\theta), \widetilde{u}(\theta), \chi(\theta))(1 + v(\theta)), \quad \theta_1 \leqslant \theta \leqslant \theta_2; \quad \widetilde{x}(\theta_1) = \widetilde{x}_0 \tag{104}$$

$$\frac{d\chi(\theta)}{d\theta} = 1 + v(\theta), \quad \theta_1 \leqslant \theta \leqslant \theta_2, \quad \chi(\theta_1) = \chi_0. \tag{105}$$

$$g^i(\widetilde{x}(\theta_1), \widetilde{x}(\theta_2), \chi(\theta_1), \chi(\theta_2)) \leqslant 0, \quad i = 1, \ldots, m; \\ g^i(\widetilde{x}(\theta_1), \widetilde{x}(\theta_2), \chi(\theta_1), \chi(\theta_2)) = 0, \quad i = m+1, \ldots, s, \tag{106}$$

управление $(\widetilde{u}(\theta), v(\theta))$, $\theta_1 \leqslant \theta \leqslant \theta_2$, кусочно-непрерывно и

$$\widetilde{u}(\theta) \in V, \quad |v(\theta)| \leqslant 1/2, \quad \theta_1 \leqslant \theta \leqslant \theta_2, \tag{107}$$

где функции $f^i(x, u, t)$, $i = 0, 1, \ldots, n$, $g^j(x, y, t, T)$, $j = 0, 1, \ldots, s$, взяты из исходной задачи (2.37)–(2.40). Так как перечисленные функции от времени θ не зависят, то задача (103)–(107) автономна, что существенно будет использовано ниже.

Установим связь между задачами (2.37)–(2.40) и (103)–(107). Положим $t = \chi(\theta)$, $t_0 = \chi(\theta_1)$, $T = \chi(\theta_2)$. Так как $|v(\theta)| \leqslant 1/2$, то $1/2 \leqslant \frac{d\chi(\theta)}{d\theta} = 1 + v(\theta) \leqslant 3/2$. Поэтому

$$\frac{1}{2}(s_2 - s_1) \leqslant \chi(s_2) - \chi(s_1) = \int_{s_1}^{s_2} \frac{d\chi(\theta)}{d\theta} d\theta \leqslant \frac{3}{2}(s_2 - s_1) \quad \forall s_1, s_2 \in [\theta_1, \theta_2], \quad s_1 < s_2. \tag{108}$$

Как видим, функция $\chi(\theta)$ строго монотонно возрастает на отрезке $[\theta_1, \theta_2]$ от $\chi(\theta_1) = t_0$ до $\chi(\theta_2) = T$. Поэтому существует обратная функция $\theta = \chi^{-1}(t)$, определенная на отрезке $[t_0, T]$, строго монотонно возрастающая от значения θ_1 до θ_2. Нетрудно убедиться, что функция $\chi^{-1}(\theta)$ также

удовлетворяет условию вида (108). В самом деле, пусть $\theta_1 \leqslant s_1 < s_2 \leqslant \theta_2$. Тогда $\chi(s_1) = t_1 < \chi(s_2) = t_2$, $s_1 = \chi^{-1}(t_1) < s_2 = \chi^{-1}(t_2)$ и в силу (108) $\dfrac{\chi^{-1}(t_2) - \chi^{-1}(t_1)}{t_2 - t_1} = \dfrac{s_2 - s_1}{\chi^{-1}(s_2) - \chi^{-1}(s_1)} \in [2/3, 2]$. Следовательно,

$$\frac{2}{3}(t_2 - t_1) \leqslant \chi^{-1}(t_2) - \chi^{-1}(t_1) \leqslant 2(t_2 - t_1) \quad \forall t_1, t_2 \in [t_0, T], \quad t_1 \leqslant t_2.$$

Отсюда следует, что функция $\theta = \chi^{-1}(t)$ удовлетворяет условию Липшица. Тогда она абсолютно непрерывна и почти всюду дифференцируема на $[t_0, T]$ [393], причем

$$\frac{2}{3} \leqslant \frac{d\chi^{-1}(t)}{dt} = \frac{1}{\left.\frac{d\chi(\theta)}{d\theta}\right|_{\theta = \chi^{-1}(t)}} = \frac{1}{1 + v(\chi^{-1}(t))} \leqslant 2.$$

Пусть $(\tilde{x}_0, \tilde{u}(\cdot), v(\cdot), \tilde{x}(\cdot), \chi(\cdot), \theta_1, \theta_2)$ — допустимый процесс задачи (103)–(107). Положим

$$x(t) = \tilde{x}(\chi^{-1}(t)), \quad u(t) = \tilde{u}(\chi^{-1}(t)), \quad t_0 \leqslant t \leqslant T; \quad x_0 = \tilde{x}_0. \tag{109}$$

Пользуясь заменой переменной $\theta = \chi^{-1}(t)$ в задаче (103)–(107), нетрудно убедиться, что $(x_0, u(\cdot), x(\cdot), t_0, T)$ — допустимый процесс задачи (2.37)–(2.40) со значением функции (2.37), совпадающим со значением $J_1(\tilde{x}_0, \tilde{u}(\cdot), v(\cdot), \tilde{x}(\cdot), \chi(\cdot), \theta_1, \theta_2)$. Обратно: пусть $(x_0, u(\cdot), x(\cdot), t_0, T)$ какой-либо допустимый процесс задачи (2.37)–(2.40). Тогда ему соответствует процесс $(\tilde{x}_0, \tilde{u}(\cdot), v(\cdot), \tilde{x}(\cdot), \chi(\cdot), \theta_1, \theta_2)$, где

$$\begin{gathered} \tilde{x}_0 = x_0, \quad \chi(\theta) = \theta, \quad t_0 = \theta_1 \leqslant \theta \leqslant \theta_2 = T, \\ \tilde{u}(\theta) = u(\chi(\theta)), \quad \tilde{x}(\theta) = x(\chi(\theta)), \quad v(\theta) \equiv 0, \end{gathered} \tag{110}$$

являющийся допустимым для задачи (103)–(107), причем функции (2.37) и (103) на соответствующих процессах принимают одинаковые значения. Из построенных соответствий (109), (110) между допустимыми процессами рассматриваемых задач (2.37)–(2.40) и (103)–(107) следует, что нижние грани функций (2.37) и (103) совпадают и, зная оптимальный процесс задачи (103)–(107), по формулам (109) можно построить оптимальный процесс задачи (2.37)–(2.40) и обратно, имея решение задачи (2.37)–(2.40), по формулам (110) нетрудно восстановить решение задачи (103)–(107). Пусть $(x_0, u(\cdot), x(\cdot), t_0, T)$ — решение задачи (2.37)–(2.40). Тогда согласно (110)

$$\begin{gathered} \tilde{x}_0 = x_0, \quad \tilde{u}(\theta) = u(\theta), \quad v(\theta) = 0, \quad \tilde{x}(\theta) = x(\theta), \\ \chi(\theta) = \theta, \quad \theta_1 = t_0, \quad \theta_2 = T \end{gathered} \tag{111}$$

решение задачи (103)–(107). Применим к этой задаче теорему 2.2. Функции Понтрягина H и $\tilde{H}$ задач (2.37)–(2.40) и (103)–(107) связаны так

$$\begin{gathered} \tilde{H}(\tilde{x}, \tilde{u}, v, \chi, \psi, \psi_0, a_0) = -a_0 f^0(\tilde{x}, \tilde{u}, \chi)(1 + v) + \\ + < \psi, f(\tilde{x}, \tilde{u}, \chi)(1 + v) > + \psi_0(1 + v) = (H(\tilde{x}, \tilde{u}, \chi, \psi, a_0) + \psi_0)(1 + v) \\ \tilde{H}|_{v=0} = H + \psi_0, \quad \tilde{H}_{\tilde{x}}|_{v=0} = H_x, \quad \tilde{H}_{\chi}|_{v=0} = H_t. \end{gathered} \tag{112}$$

Для решения (111) задачи (103)–(107) найдутся $a = (a_0, \ldots, a_s)$, $\psi(\theta)$, $\psi_0(\theta)$, $\theta_1 \leqslant \theta \leqslant \theta_2$, такие, что

$$\begin{gathered} \frac{d\psi(\theta)}{d\theta} = -H_x(\tilde{x}(\theta), \tilde{u}(\theta), \chi(\theta) = \theta, \psi(\theta), a_0), \quad \theta_1 \leqslant \theta \leqslant \theta_2, \\ \frac{d\psi_0(\theta)}{d\theta} = -H_t(\tilde{x}(\theta), \tilde{u}(\theta), \theta, \psi(\theta), a_0), \quad \theta_1 \leqslant \theta \leqslant \theta_2, \end{gathered} \tag{113}$$

$$\begin{gathered} \sup_{(w, v) \in V \times \{|v| \leqslant 1/2\}} \tilde{H}(\tilde{x}(\theta), w, v, \theta, \psi(\theta), \psi_0(\theta), a_0) = \\ = \tilde{H}(\tilde{x}(\theta), \tilde{u}(\theta), v(\theta) = 0, \theta, \psi(\theta), a_0) = \\ = H(\tilde{x}(\theta), \tilde{u}(\theta), \theta, \psi(\theta), a_0) + \psi_0(\theta) \end{gathered} \tag{114}$$

во всех точках $\theta \in (\theta_1, \theta_2)$, в которых оптимальное управление $(\tilde{u}(\theta), v(\theta) = 0)$ непрерывно; остальные утверждения теоремы 2.2 пока нам не понадобятся.

Как видно из (112), функция $\tilde{H}$ зависит от v линейно, поэтому условие (114) при $v = v(\theta) \equiv 0$ может реализоваться только тогда, когда коэффициент при v у функции $\tilde{H}$ равен нулю. Отсюда и из (114), учитывая, что

$$\sup_{(w, v) \in V \times \{|v| \leqslant 1/2\}} \tilde{H} = \sup_{w \in V} \sup_{|v| \leqslant 1/2} \tilde{H} = [\sup_{w \in V} H(\tilde{x}(\theta), w, \theta, \psi(\theta), a_0) + \psi_0(\theta)](1 + v)\Big|_{v=0},$$

получаем

$$\sup_{v \in V} H(\tilde{x}(\theta), v, \theta, \psi(\theta), a_0) + \psi_0(\theta) = H(\tilde{x}(\theta), \tilde{u}(\theta), \theta, \psi(\theta), a_0) + \psi_0(\theta) \equiv 0 \tag{115}$$

во всех точках $\theta \in (\theta_1, \theta_2)$, в которых $\tilde{u}(\cdot)$ непрерывна. Пользуясь уже доказанной частью настоящей теоремы, можем сказать, что функция

$$\tilde{H}(\theta) = \sup_{w \in V} \sup_{|v| \leqslant 1/2} \tilde{H}(\tilde{x}(\theta), w, v, \theta, \psi(\theta), \psi_0(\theta), a_0)$$

непрерывна во всех точках $\theta \in [\theta_1, \theta_2]$ и в силу (101), (102), (112)

$$\tilde{H}(\theta) = H(\tilde{x}(\theta), \tilde{u}(\theta \pm 0), \theta, \psi(\theta), a_0) + \psi_0(\theta) \quad \forall \theta \in [\theta_1, \theta_2].$$

Отсюда и из (115) имеем

$$\begin{gathered} \psi_0(\theta) = -\sup_{v \in V} H(\tilde{x}(\theta), v, \theta, \psi(\theta), a_0) = -H(\theta) = \\ = -H(\tilde{x}(\theta), \tilde{u}(\theta \pm 0), \theta, \psi(\theta), a_0) \quad \forall \theta \in [\theta_1, \theta_2]. \end{gathered} \tag{116}$$

Как следует из второго уравнения (113), левая часть равенства (116) имеет непрерывную производную во всех точках непрерывности $\tilde{u}(\cdot)$, левую и правую производные во во всех точках $\theta \in [\theta_1, \theta_2]$, причем

$$\frac{dH(\theta \pm 0)}{d\theta} = H_t(\tilde{x}(\theta), \tilde{u}(\theta \pm 0), \theta, \psi(\theta), a_0) \quad \forall \theta \in [\theta_1, \theta_2].$$

Отсюда, возвращаясь к исходным обозначениям в силу (111), с заменой θ на t, приходим к равенству (98). Теорема 1 доказана. □

Замечание 1. Более тонкие рассуждения показывают [212], что для всех допустимых процессов, удовлетворяющих необходимым условиям оп-

тимальности, сформулированным в теореме 2.2 (но, возможно, не являющихся оптимальными), также справедливы утверждения теоремы 1.

Замечание 2. Условия трансверсальности (2.42), (2.43) с учетом (97), (101), (102) могут быть записаны в виде

$$\begin{aligned}&\sup_{v\in V} H(x(t_0), v, t_0, \psi(t_0), a_0) = -l_t(x_0, x(T), t_0, T, a),\\ &\sup_{v\in V} H(x(T), v, T, \psi(T), a_0) = l_T(x_0, x(T), t_0, T, a).\end{aligned} \tag{117}$$

Поскольку замкнутость множества V не предполагается, мы не можем утверждать, что в точках t разрыва управления $u(\cdot)$ предельные значения $u(t-0)$, $u(t+0)$ принадлежат множеству V. Поэтому в (97), (117) в общем случае знак sup нельзя заменить на max. Равенство (98) с учетом непрерывности обеих частей равенства (101) и второго условия (117) можно записать в интегральной форме

$$\begin{aligned}&H(t) = H(x(t), u(t-0), t, \psi(t), a_0) =\\ &= -\int_t^T H_t(x(\tau), u(\tau), \tau, \psi(\tau), a_0)d\tau + l_T(x_0, x(T), t_0, T) \quad \forall t\in[t_0, T].\end{aligned} \tag{118}$$

Аналогичное равенство получается при использовании равенств (98), (102) и второго условия (117).

Замечание 3. Нетрудно убедиться, что теорема 1 и замечания 1, 2 сохраняют силу для задачи (2.1)–(2.4) с закрепленным временем и задачи (2.37)–(2.40), когда один из моментов t_0 или T закреплены. Для задачи (2.1)–(2.4) равенство (118) будет иметь вид

$$\begin{aligned}H(x(t), u(t-0), t, \psi(t), a_0) = -\int_t^T H_t(x(\tau), u(\tau), \tau, \psi(\tau), a_0)d\tau +\\ + H(x(T), u(T), T, \psi(T), a_0) \quad \forall t\in[t_0, T].\end{aligned}$$

§ 4. Принцип максимума для задач оптимального управления с фазовыми ограничениями

1. В этом параграфе мы будем предполагать, что читатель знаком с элементами теории меры Лебега — Стилтьеса, интеграла Римана — Стилтьеса в объеме книг [284; 393]. Кратко напомним некоторые сведения из [393], а также сформулируем и докажем несколько утверждений, необходимых в дальнейшем. Как известно [393], каждая неубывающая функция $\eta(t)$, $a\leqslant t\leqslant b$, порождает меру Лебега — Стилтьеса, причем мера промежутков $[\alpha, \beta]$, $[\alpha, \beta)$ $(\alpha, \beta]$, (α, β), принадлежащих отрезку $[a, b]$, определяются следующим образом:

$$\begin{aligned}&m[\alpha, \beta] = \eta(\beta+0) - \eta(\alpha-0), \quad m[\alpha, \beta) = \eta(\beta-0) - \eta(\alpha-0),\\ &m(\alpha, \beta] = \eta(\beta+0) - \eta(\alpha+0), \quad m(\alpha, \beta) = \eta(\beta-0) - \eta(\alpha+0),\end{aligned} \tag{1}$$

где $\eta(t-0) = \lim\limits_{\tau\to t-0}\eta(\tau)$, $\eta(t+0) = \lim\limits_{\tau\to t+0}\eta(\tau)$, по определению принимается

$$\eta(a-0) = \eta(a), \quad \eta(b+0) = \eta(b), \tag{2}$$

$m(A)$ — мера множества $A\in[a, b]$. В частности, мера точки t равна $m\{t\} = m[t, t] = \eta(t+0) - \eta(t-0)$. Функцию $\eta(t)$, которая порождает указанную меру m, называют *производящей* функцией этой меры. Напомним определение интеграла Римана — Стилтьеса. Поскольку точки $t=a$, $t=b$ могут иметь положительную меру, то различают интегралы по промежуткам $[a, b]$, $[a, b)$, $(a, b]$, (a, b). Сначала определим интеграл Римана — Стилтьеса по отрезку $[a, b]$ функции $f(t)$, определенной и принимающей конечные значения при всех $t\in[a, b]$. Возьмем произвольное разбиение

$$a = t_0 < t_1 < \ldots < t_{n-1} < t_n = b \tag{3}$$

отрезка $[a, b]$ на элементы $[t_{i-1}, t_i)$, $i=1, \ldots, n-1$, $[t_{n-1}, b]$, в каждом таком элементе выберем какую-либо точку ξ_i, составим интегральную сумму

$$S_n = f(\xi_1)m[t_0, t_1) + \sum_{i=2}^{n-1} f(\xi_i)m[t_{i-1}, t_i) + f(\xi_n)m[t_{n-1}, t_n]. \tag{4}$$

Если при $d_n = \max\limits_{1\leqslant i\leqslant n}(t_i - t_{i-1})\to 0$ суммы (4) стремятся к некоторому пределу, не зависящему от способа разбиения (3) отрезка $[a, b]$ и от выбора точек ξ_i, то этот предел называется *интегралом Римана — Стилтьеса* функции f по мере, порожденной неубывающей функцией $\eta(t)$, $t\in[a, b]$, и обозначается символом $\int_a^{b-0} f(t)d\eta$. Также определяются интегралы $\int_a^b f(t)d\eta$, $\int_{a+0}^b f(t)d\eta$, $\int_{a+0}^{b-0} f(t)d\eta$ от функции $f(t)$ по промежуткам $[a, b)$, $(a, b]$, (a, b) соответственно, причем в сумме (4) в первом и последнем слагаемом меры $m[t_0, t_1)$, $m[t_{n-1}, t_n]$ заменяются на $m(t_0, t_1)$, $m[t_{n-1}, t_n)$ в зависимости от того, принадлежат $t_0=a$, $t_n=b$ промежутку интегрирования или нет. Если $f(t)$ непрерывна на $[a, b]$, то все перечисленные интегралы существуют [393]. Интегралы по разным промежуткам связаны между собой так:

$$\begin{aligned}\int_a^b f(t)d\eta &= \int_a^{b-0} f(t)d\eta + f(b)(\eta(b) - \eta(b-0)) =\\ &= \int_{a+0}^b f(t)d\eta + f(a)(\eta(a+0) - \eta(a)) = \int_{a+0}^{b-0} f(t)d\eta +\\ &+ f(b)(\eta(b) - \eta(b-0)) + f(a)(\eta(a+0) - \eta(a)).\end{aligned} \tag{5}$$

Замечание 1. Из (1) и определения интегралов видно, что мера Лебега — Стилтьеса на $[a, b]$, интеграл Римана — Стилтьеса на зависят от того, как определена неубывающая функция $\eta(t)$ в точках разрыва, принадлежащих интервалу (a, b). Имея это в виду, функцию $\eta(t)$ часто считают непрерывной слева на (a, b), переопределяя ее в точках разрыва $t\in(a, b)$ значением $\eta(t) = \eta(t-0)$.

Пусть $a\leqslant c < d\leqslant b$. Тогда

$$\begin{aligned}\int_a^b f(t)d\eta &= \int_a^{c-0} f(t)d\eta + \int_c^d f(t)d\eta + \int_{d+0}^b f(t)d\eta =\\ &= \int_a^c f(t)d\eta + \int_{c+0}^{d-0} f(t)d\eta + \int_d^b f(t)d\eta;\end{aligned} \tag{6}$$

аналогично выписываются другие формулы с использованием интегралов по промежуткам $[c, d)$, $(c, d]$. Если $c=a$, $d=b$, то в (6) по определению

считается $\int_a^{a+0} f(t)d\eta = f(a)(\eta(a+0)-\eta(a))$, $\int_{b-0}^{b} f(t)d\eta = f(b)(\eta(b)-\eta(b-0))$, и формула (6) превращается в (5), а интегралы по пустым промежуткам $[a, a-0)$, $(b+0, b]$ в (6) опускаются. В (6) подразумевается, что на отрезках $[a, c], [c, d], [d, b]$ мы имеем дело с мерой, которая индуцируется мерой с исходного отрезка $[a, b]$, заданной с помощью производящей функции $\eta(t)$, $a \leqslant t \leqslant b$.

Определение 1. Мера m_1 на $[c, d]$ называется *индуцированной* с отрезка $[a, b]$, если для любого измеримого множества $A \in [c, d]$ с мерой $m_1(A)$ множество A, как подмножество отрезка $[a, b]$, также измеримо и его мера $m(A)$ на отрезке $[a, b]$ равна $m_1(A)$.

Зададимся вопросом: какова производящая функция такой меры на перечисленных отрезках, например, на $[c, d]$? Может показаться, что мера на $[c, d]$ индуцированная с отрезка $[a, b]$, задается производящей функцией $\eta(t)$, $t \in [c, d]$. Однако тогда мера точки c на отрезке $[c, d]$ будет равна $m_1\{c\} = \eta(c+0) - \eta(c)$, в то время как на отрезке $[a, b]$ эта точка имела меру $m\{c\} = \eta(c+0) - \eta(c-0) \neq m_1(c)$ при $\eta(c-0) \neq \eta(c)$. Аналогично, если $\eta(d+0) \neq \eta(d)$, то меры точки $t = d$ на $[c, d]$ и на $[a, b]$ будут различными. Нетрудно убедиться, что производящей функцией индуцированной на $[c, d]$ меры будет функция

$$\eta_1 = \eta_1(t) = \begin{cases} \eta(c-0), & t = c, \\ \eta(t), & t \in (c, d), \\ \eta(d+0), & t = d. \end{cases} \tag{7}$$

Тогда $\int_c^d f(t)d\eta$, взятый из (6), равен $\int_c^d f(t)d\eta_1$, причем имея функцию η_1, последний интеграл можно уже рассматривать самостоятельно, независимо от других слагаемых из (6) и независимо от «предыстории» меры на $[c, d]$. Предлагаем читателю выписать производящие функции для индуцированной меры на отрезках $[a, c], [d, b]$, учитывая равенства (2). Кстати, если $c = d$, $d = b$, то из (7) с учетом (2) имеем $\eta_1(t) = \eta(t)$, $t \in [a, b]$.

Справедливы равенства

$$\begin{aligned} &\int_a^b (f(t)+g(t))d\eta = \int_a^b f(t)d\eta + \int_a^b g(t)d\eta, \\ &\int_a^b f(t)d(\eta_1+\eta_2) = \int_a^b f(t)d\eta_1 + \int_a^b f(t)d\eta_2, \end{aligned} \tag{8}$$

вытекающие из того, что сумма $\eta_1+\eta_2$ неубывающих функций $\eta_1\eta_2$ также является неубывающей, и из справедливости соответствующих равенств для интегральных сумм вида (4) для каждого разбиения (3), которое сохраняется в пределе для интегралов.

Ниже нам понадобятся оценки:

$$\begin{aligned} &\left|\int_a^b f(t)d\eta\right| \leqslant \|f\|(\eta(b)-\eta(a)), && \left|\int_a^{b-0} f(t)d\eta\right| \leqslant \|f\|(\eta(b-0)-\eta(a)), \\ &\left|\int_{a+0}^b f(t)d\eta\right| \leqslant \|f\|(\eta(b)-\eta(a+0)), && \left|\int_{a+0}^{b-0} f(t)d\eta\right| \leqslant \|f\|(\eta(b-0)-\eta(a+0)), \end{aligned} \tag{9}$$

где $\|f\| = \sup_{t \in [a, b]} |f(t)|$. Докажем первое из неравенств (9). Для интегральной суммы (4) имеем $|S_n| \leqslant \|f\|(\eta(t_1-0)-\eta(a)+\sum_{i=2}^{n-1}(\eta(t_i-0)-\eta(t_{i-1}-0))+\eta(b)-\eta(t_{n-1}-0)) = \|f\|(\eta(b)-\eta(a))$. Отсюда при $d_n \to 0$ получим требуемое. Остальные неравенства (9) доказываются аналогично.

Теорема 1 (первая теорема Хелли). *Пусть $\eta_1(t), \ldots, \eta_k(t), \ldots$ — последовательность неубывающих функций на отрезке $[a, b]$, $\sup_{k \geqslant 1} \max\{|\eta_k(a)|; |\eta_k(b)|\} \leqslant c_0 < \infty$, пусть $\lim_{k \to \infty} \eta_k(t) = \eta(t)$ в каждой точке $t \in [a, b]$. Тогда $\eta(t)$ также не убывает на $[a, b]$, и для каждой непрерывной на $[a, b]$ функции $f(t)$ справедливо равенство*

$$\lim_{k \to \infty} \int_a^b f(t)d\eta_k = \int_a^b f(t)d\eta.$$

Теорема 2 (вторая теорема Хелли). *Пусть $\Phi_1(t), \ldots, \Phi_k(t), \ldots$ — последовательность функций с ограниченным изменением на отрезке $[a, b]$ такая, что*

$$\sup_{k \geqslant 1} |\Phi_k(t_0)| \leqslant c_0 < \infty, \quad \sup_{k \geqslant 1} V_a^b(\Phi_k) \leqslant c_1 < \infty,$$

где t_0 какая-либо точка из $[a, b]$, $V_a^b(\Phi_k) = \sup_{\{t_i\}} \sum_{i=1}^{n} |\Phi_k(t_i) - \Phi_k(t_{i-1})|$ — полное изменение функции Φ_k на $[a, b]$, c_0, c_1 — некоторые положительные постоянные. Тогда существует подпоследовательность $\{\Phi_{k_r}(t)\}$, сходящаяся к некоторой функции $\Phi(t)$ в каждой точке $t \in [a, b]$, причем $\Phi(t)$ имеет ограниченное изменение на отрезке $[a, b]$.

Заметим, что при сделанных предположениях $\sup_{k \geqslant 1} \sup_{t \in [a, b]} |\Phi_k(t)| \leqslant c_0 + c_1$. Теорема 2 верна и для функций $\{\eta_k(t)\}$ из теоремы 1, так как всякая неубывающая на $[a, b]$ функция $\eta(t)$ имеет ограниченное изменение, причем $V_a^b(\eta) \leqslant \eta(b)-\eta(a)$. Более общее определение интеграла Римана — Стилтьеса, более общую формулировку теорем 1, 2, их доказательства читатель найдет в [284; 393].

Лемма 1. *Пусть функция $\eta(t)$ определена и не убывает на отрезке $[a, b]$, пусть последовательность $\{t_k\} \in [a, b]$ и $\lim_{k \to \infty} t_k = t$. Если $t \in (a, b]$, $a \leqslant t_k < t$, $k = 1, 2, \ldots$, то*

$$\lim_{k \to \infty} \eta(t_k) = \lim_{k \to \infty} \eta(t_k - 0) = \lim_{k \to \infty} \eta(t_k + 0) = \eta(t - 0). \tag{10}$$

Если $t \in [a, b)$, $t < t_k \leqslant b$, $k = 1, 2, \ldots$, то

$$\lim_{k \to \infty} \eta(t_k) = \lim_{k \to \infty} \eta(t_k - 0) = \lim_{k \to \infty} \eta(t_k + 0) = \eta(t + 0). \tag{11}$$

Доказательство. Рассмотрим случай $t \in (a, b]$, $a \leqslant t_k < t$, $k = 1, 2, \ldots$. Возьмем произвольную точку $\tau \in [a, b]$, $\tau < t$. Тогда $\tau < t_k < t$ $\forall k \geqslant k_0$, где k_0 — достаточно большое число. Так как функция $\eta(t)$ не убывает, то $\eta(\tau) \leqslant \eta(t_k - 0) \leqslant \eta(t_k) \leqslant \eta(t_k + 0) \leqslant \eta(t - 0)$, $k = 1, 2, \ldots$. Учитывая, что $\lim_{k \to \infty} \eta(t_k) = \eta(t - 0)$, отсюда имеем $\eta(\tau) \leqslant \underline{\lim}_{k \to \infty} \eta(t_k - 0) \leqslant \overline{\lim}_{k \to \infty} \eta(t_k - 0) \leqslant \eta(t - 0) = \lim_{k \to \infty} \eta(t_k + 0)$ $\forall \tau < t$. Переходя здесь к пределу при $\tau \to t - 0$, получаем равенства (10). Равенства (11) доказываются аналогично. □

Лемма 2. *Пусть* $\{\eta_k(t)\}$ *— последовательность неубывающих функций на отрезке* $[a, b]$, $\sup\limits_{k\geqslant 1}\max\{|\eta_k(a)|; |\eta_k(b)|\} \leqslant c_0 < \infty$, *пусть* $\lim\limits_{k\to\infty}\eta_k(t) = \eta(t)$ *в каждой точке* $t \in [a, b]$. *Тогда функция* $\eta(t)$ *не убывает на* $[a, b]$, *причем*

$$\eta(t-0) \leqslant \varliminf_{k\to\infty}\eta_k(t-0) \leqslant \varlimsup_{k\to\infty}\eta_k(t-0) \leqslant \eta(t) \leqslant$$
$$\leqslant \varliminf_{k\to\infty}\eta_k(t+0) \leqslant \varlimsup_{k\to\infty}\eta_k(t+0) \leqslant \eta(t+0) \quad \forall t \in [a, b] \quad (12)$$

(*в* (12) *подразумевается, что в точках* $t = a$, $t = b$ *для функций* $\eta_k(t)$, $\eta(t)$ *по определению выполняются равенства* (2)).

Доказательство. Так как $\eta_k(t) \leqslant \eta_k(\tau)$ $t \leqslant \tau$, то при $k \to \infty$ отсюда имеем $\eta(t) \leqslant \eta(\tau)$, $\forall t \leqslant \tau$, т. е. функция $\eta(t)$ не убывает на $[a, b]$. Рассмотрим случай, когда $a < t < b$. Возьмем произвольные точки $\tau_1, \tau_2 \in [a, b]$, $\tau_1 < t < \tau_2$. Тогда

$$\eta_k(\tau_1) \leqslant \eta_k(t-0) \leqslant \eta_k(t) \leqslant \eta_k(t+0) \leqslant \eta_k(\tau_2).$$

Отсюда переходя к пределу сначала при $k \to \infty$, затем при $\tau_1 \to t-0$, $\tau_2 \to t+0$, приходим к неравенствам (12). Случаи $t = a$ или $t = b$ рассматриваются аналогично с учетом (2). □

Приведем пример, показывающий, что все указанные в лемме 2 ситуации возможны.

Пример 1. Пусть числа c_{1k}, c_{2k}, c_k таковы, что $0 \leqslant c_{1k} \leqslant c_k \leqslant c_{2k} \leqslant 1$, $k = 1, 2, \ldots$, $\lim\limits_{k\to\infty} c_k = c$, последовательности $\{c_{1k}\}, \{c_{2k}\}$ необязательно сходятся. Положим

$$\eta_{1k}(t) = \begin{cases} 0, & t \in [-1, -\frac{1}{k}), \\ c_{1k}, & t \in [-\frac{1}{k}, 0), \\ c_k, & t = 0, \\ c_{2k}, & t \in (0, \frac{1}{k}), \\ 1, & t \in [\frac{1}{k}, 1], \end{cases} \qquad \eta_{2k}(t) = \begin{cases} 0, & t \in [-1, -\frac{c_k}{k}], \\ kt + c_k, & t \in (-\frac{c_k}{k}, \frac{1-c_k}{k}), \\ 1, & t \in [\frac{1-c_k}{k}, 1], \end{cases} \quad k = 1, 2, \ldots$$

Нетрудно убедиться, что функции $\eta_{1k}(t), \eta_{2k}(t)$ не убывают и $\lim\limits_{k\to\infty}\eta_{1k}(t) = \lim\limits_{k\to\infty}\eta_{2k}(t) = \eta(t)$ всюду на $[-1, 1]$, где $\eta(t) = 0$ при $-1 \leqslant t < 0$, $\eta(0) = c$, $\eta(t) = 1$ при $0 < t \leqslant 1$. Выбирая по разному c, $0 \leqslant c \leqslant 1$, и параметры c_{1k}, c_{2k}, можно реализовать все возможные ситуации, допускаемые неравенствами (12). Отметим, что функции $\eta_{2k}(t)$ непрерывны на $[-1, 1]$.

Лемма 3. *Пусть функции* $\eta_k(t), \eta(t)$ *удовлетворяют условиям леммы* 2, *пусть последовательность* $\{t_k\} \in [a, b]$, $\lim\limits_{k\to\infty} t_k = t$. *Если* $t \in (a, b]$, $a \leqslant t_k < t$, $k = 1, 2, \ldots$, *то*

$$\eta(t-0) \leqslant \varliminf_{k\to\infty}\eta_k(t_k-0) \leqslant \varlimsup_{k\to\infty}\eta_k(t_k-0) \leqslant \eta(t),$$
$$\eta(t-0) \leqslant \varliminf_{k\to\infty}\eta_k(t_k) \leqslant \varlimsup_{k\to\infty}\eta_k(t_k) \leqslant \eta(t), \quad (13)$$
$$\eta(t-0) \leqslant \varliminf_{k\to\infty}\eta_k(t_k+0) \leqslant \varlimsup_{k\to\infty}\eta(t_k+0) \leqslant \eta(t).$$

Если $t \in [a, b)$, $t < t_k \leqslant b$, $k = 1, 2, \ldots$, *то*

$$\eta(t) \leqslant \varliminf_{k\to\infty}\eta_k(t_k-0) \leqslant \varlimsup_{k\to\infty}\eta_k(t_k-0) \leqslant \eta(t+0),$$
$$\eta(t) \leqslant \varliminf_{k\to\infty}\eta_k(t_k) \leqslant \varlimsup_{k\to\infty}\eta_k(t_k) \leqslant \eta(t+0), \quad (14)$$
$$\eta(t) \leqslant \varliminf_{k\to\infty}\eta_k(t_k+0) \leqslant \varlimsup_{k\to\infty}\eta(t_k+0) \leqslant \eta(t+0).$$

Доказательство. Рассмотрим случай $t \in (a, b]$, $a \leqslant t_k < t$, $k = 1, 2, \ldots$ Возьмем произвольную точку $\tau_1 \in [a, b]$, $\tau_1 < t$. Тогда $\tau_1 < t_k < t$ $\forall k \geqslant k_0$, и в силу монотонности $\eta_k(t)$ тогда $\eta_k(\tau_1) \leqslant \eta_k(t_k-0) \leqslant \eta_k(t_k) \leqslant \eta_k(t_k+0) \leqslant \eta_k(t)$, $k \geqslant k_0$. Отсюда, переходя к пределу сначала при $k \to \infty$, затем при $\tau_1 \to t-0$, получим неравенства (13). Неравенства (14) доказываются аналогично. □

Лемма 4. *Пусть функции* $\eta_k(t)$, $\eta(t)$ *удовлетворяют условиям леммы* 2, *пусть последовательность* $\{t_k\} \in [a, b]$, $\lim\limits_{k\to\infty} t_k = t$. *Тогда*

$$0 \leqslant \varlimsup_{k\to\infty}(\eta_k(t_k+0) - \eta_k(t_k-0)) \leqslant \eta(t+0) - \eta(t-0) \quad (15)$$

(*в* (15) *подразумевается, что в точках* $t = a$, $t = b$ *для функций* $\eta_k(t)$, $\eta(t)$ *по определению выполняются равенства* (2)).

Доказательство. Пусть верхний предел последовательности $a_k = \eta_k(t_k+0) - \eta_k(t_k-0)$, $k = 1, 2, \ldots$ реализуется на подпоследовательности $\{a_{k_p}\}$, т. е. $\varlimsup\limits_{k\to\infty} a_k = \varlimsup\limits_{p\to\infty} a_{k_p}$. Перенумеровав подпоследовательность $\{a_{k_p}\}$, можем считать, что $\varlimsup\limits_{k\to\infty} a_k = \lim\limits_{k\to\infty} a_k$ для самой последовательности $\{a_k\}$ и, значит, для любой ее подпоследовательности. Подробнее рассмотрим случай $a < t < b$. Может оказаться, что у последовательности $\{t_k\}$ существует подпоследовательность $\{t_{k_l}\}$, $a \leqslant t_{k_l} < t$, $l = 1, 2, \ldots$ Тогда с учетом (13) имеем

$$\varlimsup_{k\to\infty} a_k = \lim_{l\to\infty} a_{k_l} \leqslant \varlimsup_{l\to\infty}\eta_{k_l}(t_{k_l}+0) - \varliminf_{l\to\infty}\eta_{k_l}(t_{k_l}-0) \leqslant$$
$$\leqslant \eta(t) - \eta(t-0) \leqslant \eta(t+0) - \eta(t-0). \quad (16)$$

Если последовательность $\{t_k\}$ содержит подпоследовательность $\{t_{k_l}\}$, для которой $t < t_{k_l} \leqslant b$, $l = 1, 2, \ldots$, то с помощью неравенств (14) аналогично получаем

$$\varlimsup_{k\to\infty} a_k = \lim_{l\to\infty} a_{k_l} \leqslant \eta(t+0) - \eta(t) \leqslant \eta(t+0) - \eta(t-0). \quad (17)$$

Наконец, если последовательность $\{t_k\}$ не содержит указанные выше подпоследовательности $\{t_{k_l}\}$, то $t_k = t$ $\forall k \geqslant k_0$. Отсюда и из (12) тогда имеем

$$\varlimsup_{k\to\infty} a_k = \lim_{k\to\infty} a_k \leqslant \varlimsup_{k\to\infty}\eta_k(t+0) - \varliminf_{k\to\infty}\eta_k(t-0) \leqslant \eta(t+0) - \eta(t-0). \quad (18)$$

Объединяя все возможные случаи (16)–(18), приходим к неравенству (15). Случаи $t = a$ или $t = b$ рассматриваются аналогично с учетом равенств (2). □

Далее, зафиксируем произвольную непрерывную функцию $f(t)$, $t\in[a,b]$, и определим функции

$$F_k(t)=\int_t^b f(\tau)d\eta_k,\quad k=1,2,\ldots;\quad F(t)=\int_t^b f(\tau)d\eta,\quad t\in[a,b],\qquad(19)$$

где функции $\eta_k(t),\eta(t)$ взяты из леммы 2, интегралы в (19) берутся по индуцированным с отрезка $[a,b]$ на $[t,b]$ мерам, порожденным функциями $\eta_k(t),\eta(t)$, $t\in[a,b]$. Из теоремы 1 вытекает, что $\lim\limits_{k\to\infty}F_k(a)=F(a)$. Может показаться, что

$$\lim_{k\to\infty}F_k(t)=F(t)\qquad(20)$$

и при других $t\in(a,b]$. Приведем пример, показывающий, что если t — точка разрыва функции $\eta(\tau)$, то равенство (20), вообще говоря, не имеет места и, более того, предел в левой части (20) может не существовать.

Пример 2. Возьмем функции $\eta_k(t)=\eta_{1k}(t)$, $t\in[-1,1]$ из примера 1, считая $c_k=c=1/2$, $c_{1k}=1/4$ при четных k, $c_{1k}=1/2$ при нечетных k, $c_{2k}=1/2, k=1,2,\ldots$ Пусть $f(t)\equiv1$, $t\in[-1,1]$. Тогда $F_k(t)=\int\limits_t^1 d\eta_k=\eta_k(1)-\eta_k(t-0)$, $F(t)=\int\limits_t^1 d\eta=\eta(1)-\eta(t-0)$. При $t=0$ имеем $F_k(0)=1-c_{1k}$, $k=1,2,\ldots$; $F(0)=1$. Ясно, что $\underline{\lim\limits_{k\to\infty}}F_k(0)=1/2$, $\overline{\lim\limits_{k\to\infty}}F_k(0)=3/4$, $\lim\limits_{k\to\infty}F_k(0)$ не существует.

Лемма 5. *Пусть функции $\eta_k(\tau),\eta(\tau)$ удовлетворяют условиям леммы 2, пусть $f(t)$ — произвольная непрерывная функция на $[a,b]$. Тогда*

$$\overline{\lim_{k\to\infty}}|F_k(t)-F(t)|=\overline{\lim_{k\to\infty}}\left|\int_t^b f(\tau)d\eta_k-\int_t^b f(\tau)d\eta\right|\leqslant$$
$$\leqslant2\|f\|_C(\eta(t+0)-\eta(t-0))\quad\forall t\in[a,b],\qquad(21)$$

где $\|f\|_C=\max\limits_{t\in[a,b]}|f(t)|$ (при $t=a$ или $t=b$ в (21) подразумеваются равенства (2)). Если $t=a$ или $t\in(a,b]$ точка непрерывности функции $\eta(\tau)$, то справедливо равенство (20) (при $t=b$ здесь имеется в виду непрерывность слева).

Доказательство. Пусть $t\in(a,b]$. Положим $\widetilde{\eta}_k(\tau)=\eta_k(\tau)$, $\widetilde{\eta}(\tau)=\eta(\tau)$ для всех $\tau\in[t,b]$, и определим функции $\widetilde{F}_k(t)=\int\limits_t^b f(\tau)d\widetilde{\eta}_k$, $\widetilde{F}(t)=\int\limits_t^b f(\tau)d\widetilde{\eta}$. Так как $\lim\limits_{k\to\infty}\widetilde{\eta}_k(\tau)=\widetilde{\eta}(\tau)$ во всех точках $\tau\in[t,b]$, то, применяя теорему 1 к отрезку $[t,b]$, имеем

$$\lim_{k\to\infty}|\widetilde{F}_k(t)-\widetilde{F}(t)|=0.\qquad(22)$$

Далее, используя формулы (5) применительно к отрезку $[t,b]$, получаем

$$F_k(t)=\int_{t+0}^b f(\tau)d\eta_k+f(t)(\eta_k(t+0)-\eta_k(t-0)),$$
$$F(t)=\int_{t+0}^b f(\tau)d\eta+f(t)(\eta(t+0)-\eta(t-0)),$$

$$\widetilde{F}_k(t)=\int_{t+0}^b f(\tau)d\eta_k+f(t)(\eta_k(t+0)-\eta_k(t)),$$
$$\widetilde{F}(t)=\int_{t+0}^b f(\tau)d\eta+f(t)(\eta(t+0)-\eta(t)).$$

Отсюда следуют оценки:

$$|\widetilde{F}(t)-F(t)|=|f(t)|(\eta(t)-\eta(t-0))\leqslant\|f\|_C(\eta(t+0)-\eta(t-0)),$$
$$|\widetilde{F}_k(t)-F_k(t)|=|f(t)|(\eta_k(t)-\eta_k(t-0))\leqslant\|f\|_C(\eta_k(t+0)-\eta_k(t-0)),\qquad(23)$$
$$k=1,2,\ldots$$

Из последнего неравенства и леммы 2 при $k\to\infty$ имеем

$$\overline{\lim_{k\to\infty}}|\widetilde{F}_k(t)-F_k(t)|\leqslant\|f\|_C(\eta(t+0)-\eta(t-0)).\qquad(24)$$

Переходя к пределу при $k\to\infty$ в неравенстве

$$|F_k(t)-F(t)|\leqslant|F_k(t)-\widetilde{F}_k(t)|+|\widetilde{F}_k(t)-\widetilde{F}(t)|+|\widetilde{F}(t)-F(t)|,$$

с учетом соотношений (22)–(24) получим требуемую оценку (21) при всех $t\in(a,b]$. Справедливость равенства (20) и тем более оценки (21) при $t=a$, как уже упоминалось выше, следует из теоремы 1. В точках $t\in(a,b]$, в которых функция $\eta(\tau)$ непрерывна, равенство (20) вытекает непосредственно из оценки (21). □

Лемма 6. *Пусть функции $\eta_k(\tau),\eta(\tau)$ удовлетворяют условиям леммы 2, пусть последовательность непрерывных функций $\{f_k(t)\}$ сходится к $f(t)$ равномерно на $[a,b]$, т. е. $\lim\limits_{k\to\infty}\|f_k-f\|_C=0$. Тогда*

$$\overline{\lim_{k\to\infty}}\left|\int_t^b f_k(\tau)d\eta_k-\int_t^b f(\tau)d\eta\right|\leqslant2\|f\|_C(\eta(t+0)-\eta(t-0))\quad\forall t\in[a,b]\qquad(25)$$

(при $t=a$ или $t=b$ в (25) подразумеваются равенства (2)). Если $t=a$ или $t\in(a,b]$ — точка непрерывности функции $\eta(\tau)$, то

$$\lim_{k\to\infty}\int_t^b f_k(\tau)d\eta_k=\int_t^b f(\tau)d\eta.\qquad(26)$$

Доказательство. С учетом (8), (9), (19) имеем

$$\left|\int_t^b f_k(\tau)d\eta_k-\int_t^b f(\tau)d\eta\right|\leqslant\left|\int_t^b f_k(\tau)d\eta_k-\int_t^b f(\tau)d\eta_k\right|+$$
$$+|F_k(t)-F(t)|\leqslant\|f_k-f\|_C(\eta(b)-\eta(a))+|F_k(t)-F(t)|,\quad k=1,2,\ldots$$

Отсюда и из леммы 5 следуют соотношения (25), (26). □

Лемма 7. *Пусть функции $\eta_k(\tau),\eta(\tau)$ удовлетворяют условиям леммы 2, $f(t)$ — непрерывная функция на $[a,b]$, пусть последовательность $\{t_k\}\in[a,b]$, $\lim\limits_{k\to\infty}t_k=t$. Тогда*

$$\overline{\lim_{k\to\infty}}|F_k(t_k)-F(t)|\leqslant3\|f\|_C(\eta(t+0)-\eta(t-0))\quad\forall t\in[a,b]\qquad(27)$$

(при $t=a$ или $t=b$ здесь подразумеваются равенства (2)).

Доказательство. Можем считать, что верхний предел в (27) реализуется на самой последовательности $\{|F_k(t_k)-F(t)|\}$ и, стало быть, на любой ее подпоследовательности $\{|F_{k_p}(t_{k_p})-F(t)|\}$. Пусть $a<t<b$. Может случиться, что у последовательности $\{t_k\}$ существует подпоследовательность $\{t_{k_p}\}$, $a\leqslant t_{k_p}<t$, $p=1,2,\ldots$ Тогда с учетом (6), (9) имеем

$$|F_{k_p}(t_{k_p})-F(t)|\leqslant|F_{k_p}(t)-F(t)|+\left|\int\limits_{t_{k_p}}^{t-0}f(\tau)d\eta_{k_p}\right|\leqslant$$

$$\leqslant|F_{k_p}(t)-F(t)|+\|f\|_C(\eta_{k_p}(t-0)-\eta_{k_p}(t_{k_p}-0)). \quad (28)$$

Из лемм 2, 3 следует, что $\overline{\lim\limits_{p\to\infty}}(\eta_{k_p}(t-0)-\eta_{k_p}(t_{k_p}-0))\leqslant\overline{\lim\limits_{p\to\infty}}\eta_{k_p}(t-0)-\underline{\lim\limits_{p\to\infty}}\eta_{k_p}(t_{k_p}-0)\leqslant\eta(t)-\eta(t-0)$. Отсюда и из (28) с учетом (21) получим оценку (27). Если последовательность $\{t_k\}$ имеет подпоследовательность $\{t_{k_p}\}$, $t<t_{k_p}\leqslant b$, то

$$|F_{k_p}(t_{k_p})-F(t)|\leqslant|F_{k_p}(t)-F(t)|+\left|\int\limits_{t}^{t_{k_p}-0}f(\tau)d\eta_{k_p}\right|\leqslant$$

$$\leqslant|F_k(t)-F(t)|+\|f\|_C(\eta_{k_p}(t_{k_p}-0)-\eta_{k_p}(t-0)). \quad (29)$$

Из лемм 2, 3 следует, что $\overline{\lim\limits_{p\to\infty}}(\eta_{k_p}(t_{k_p}-0)-\eta_{k_p}(t-0))\leqslant\eta(t+0)-\eta(t-0)$. Отсюда и из (21), (29) получим оценку (27). Наконец, если последовательность $\{t_k\}$ не содержит подпоследовательностей $\{t_{k_p}\}$ рассмотренных выше типов, то $t_k=t$ $\forall k\geqslant k_0$. Тогда $|F_k(t_k)-F(t)|=|F_k(t)-F(t)|$ $\forall k\geqslant k_0$. Отсюда и из (21) сразу получаем (27). Таким образом, оценка (27) доказана для всех $t\in(a,b)$. В случае $t=a$ или $t=b$ оценка (27) доказывается аналогично. □

Лемма 8. *Пусть функция $\eta(\tau)$ не убывает на отрезке $[a,b]$, $f(t)$ — произвольная непрерывная функция на $[a,b]$. Тогда функция $F=F(t)=\int\limits_t^b f(\tau)d\eta$ имеет ограниченное изменение на $[a,b]$, непрерывна слева во всех точках $t\in(a,b]$. Если $t\in(a,b)$, функция $\eta(\tau)$ непрерывна в точке $\tau=t$, то функция F также непрерывна в точке $\tau=t$. Если $\eta(\tau)$ непрерывна справа в точке $t=a$, то F непрерывна справа в этой точке.*

Доказательство. Пусть $\{t_i\}$ — произвольное разбиение (3) отрезка $[a,b]$. Тогда

$$\sum_{i=1}^n|F(t_i)-F(t_{i-1})|=\sum_{i=1}^n\left|\int\limits_{t_{i-1}}^{t_i-0}f(\tau)d\eta\right|\leqslant$$

$$\leqslant\|f\|_C\left(\sum_{i=1}^n(\eta(t_i-0)-\eta(t_{i-1}-0)\right)=\|f\|_C(\eta(b-0)-\eta(a)).$$

Отсюда следует, что функция $F(t)$ имеет ограниченное изменение на $[a,b]$ и ее полная вариация

$$V_a^b(F)=\sup_{\{t_i\}}\sum_{i=1}^n|F(t_i)-F(t_{i-1})|\leqslant\|f\|_C(\eta(b-0)-\eta(a)).$$

Далее, пусть $\{t_k\}$ — произвольная последовательность такая, что $\{t_k\}\in[a,b]$,

$t_k\neq t$, $k=1,2,\ldots$, $\lim\limits_{k\to\infty}t_k=t$. Если $t\in(a,b]$, $a\leqslant t_k<t$, $k=1,2,\ldots$, то $|F(t_k)-F(t)|=\left|\int\limits_{t_k}^{t-0}f(\tau)d\eta\right|\leqslant\|f\|_C(\eta(t-0)-\eta(t_k-0))$. Отсюда и из (10) следует, что $\lim\limits_{k\to\infty}F(t_k)=F(t)$. В силу произвольности выбора последовательности $\{t_k\}\to t-0$ отсюда заключаем, что функция F непрерывна слева во всех точках $t\in(a,b]$.

Теперь пусть $t\in[a,b)$, $t<t_k\leqslant b$. Тогда $|F(t_k)-F(t)|=\left|\int\limits_t^{t_k-0}f(\tau)d\eta\right|\leqslant$ $\leqslant\|f\|_C(\eta(t_k-0)-\eta(t-0))$, $k=1,2,\ldots$. Отсюда и из (11) при $k\to\infty$ имеем $0\leqslant\underline{\lim\limits_{k\to\infty}}|F(t_k)-F(t)|\leqslant\overline{\lim\limits_{k\to\infty}}|F(t_k)-F(t)|\leqslant\|f\|_C(\eta(t+0)-\eta(t-0))$. Если функция $\eta(\tau)$ непрерывна в точке $\tau=t$, то $\eta(t+0)=\eta(t-0)=\eta(t)$ и из последних неравенств получим $\lim\limits_{k\to\infty}F(t_k)=F(t)$. В силу произвольности выбора последовательности $\{t_k\}\to t+0$ заключаем, что функция F непрерывна справа во всех точках $t\in[a,b)$ непрерывности функции $\eta(\tau)$ (непрерывности справа в точке $t=a$). Отсюда и из непрерывности слева функции $F(t)$ во всех точках $t\in(a,b]$ следует, что эта функция непрерывна во всех точках $t\in(a,b)$, в которых $\eta(\tau)$ непрерывна. Лемма 8 доказана. □

2. Рассмотрим задачу оптимального управления с ограничениями на фазовые координаты в следующей постановке: минимизировать функцию

$$J(x_0,u(\cdot),x(\cdot),t_0,T)=\int\limits_{t_0}^T f^0(x(t),u(t),t)dt+g^0(x_0,x(T),t_0,T) \quad (30)$$

при условиях

$$\dot x(t)=f(x(t),u(t),t),\quad t_0\leqslant t\leqslant T,\quad x(t_0)=x_0, \quad (31)$$

$$g^i(x_0,x(T),t_0,T)\leqslant 0,\quad i=1,\ldots,m,$$
$$g^i(x_0,x(T),t_0,T)=0,\quad i=m+1,\ldots,s, \quad (32)$$

$$G^1(x(t),t)\leqslant 0,\ldots,G^q(x(t),t)\leqslant 0,\quad t_0\leqslant t\leqslant T, \quad (33)$$

$$u=u(t)\in V\subseteq E^r, \quad (34)$$

где $G^j(x,t)$, $j=1,\ldots,q$, — заданные функции переменных $x\in E^n$, $t\in\mathbb{R}$; остальные обозначения те же, что и в задаче (2.37)–(2.40). Как в § 2, 3 будем считать, что допустимое управление в этой задаче также кусочно-непрерывно, а включение (34) справедливо во всех точках $t\in[t_0,T]$, в которых управление $u(\cdot)$ непрерывно. Как и выше, через J_* будем обозначать нижнюю грань функции (30) при условиях (31)–(34), предполагая, что множество допустимых процессов этой задачи непусто. Предполагается также, что $J_*>-\infty$, и задача (30)–(34) имеет хотя бы одно решение $(x_0,u(\cdot),x(\cdot),t_0,T)$. Следуя [44], задачу (30)–(34) будем рассматривать при дополнительном условии, что фазовые ограничения (33) согласованы с концевыми ограничениями (32) в следующем смысле.

Определение 1. Пусть точка $z_*=(x_*,y_*,t_*,T_*)$ удовлетворяет концевым ограничениям (32), т. е. $g^i(z_*)\leqslant 0$, $i=1,\ldots,m$, $g^i(z_*)=0$, $i=m+1,\ldots,s$. Говорят, что в точке z_* *фазовые ограничения* (33) *согласованы с концевыми*, если существует число $\varepsilon>0$ такое, что множество $\{z=(x,y,t,T)\colon|z-z_*|\leqslant\varepsilon,\ g^i(z)\leqslant 0,\ i=1,\ldots,m;\ g^i(z)=0,\ i=m+1,\ldots,s\}\subseteq\{z=(x,y,t,T)\colon G^i(x,t)\leqslant 0,\ G^i(y,T)\leqslant 0,\ i=1,\ldots,q\}$.

Нетрудно видеть, что если в задаче (30)–(34) фазовые ограничения отсутствуют ($q = 0$) или функции $G^i(x, t)$ непрерывны и $G^i(x_*, t_*) < 0$, $G^i(y_*, T_*) < 0$, $i = 1, \ldots, q$, то условие согласования указанных ограничений в точке z_* автоматически выполнено. Заметим, что если условия согласования фазовых и концевых ограничений не выполняются для концов какого-либо допустимого процесса задачи (30)–(34), то, добавив к s ограничениям (32) еще $2q$ концевых ограничений $G^i(x_0, t_0) \leqslant 0$, $G^i(x(T), T) \leqslant 0$, $i = 1, \ldots, q$, придем к задаче оптимального управления того же вида, но в которой, как нетрудно проверить, фазовые ограничения уже будут согласованы с концевыми во всех точках $z_* = (x_0, x(T), t_0, T)$ для любого допустимого процесса $(x_0, u(\cdot), x(\cdot), t_0, T)$. Поэтому требуемое ниже условие согласования фазовых и концевых ограничений в концах оптимального допустимого процесса не является слишком жестким.

Для формулировки принципа максимума для задачи (30)–(34) нам снова понадобятся функции

$$H(x, u, t, \psi, a_0) = -a_0 f^0(x, u, t) + \langle \psi, f(x, u, t) \rangle,$$

$$l(x, y, t, T, a) = \sum_{i=0}^{s} a_i g^i(x, y, t, T).$$

Теорема 3 [44]. *Пусть функции $f^j(x, u, t)$, $j = 0, \ldots, n$, $g^j(x, y, t, T)$, $j = 0, \ldots, s$, $G^j(x, t)$, $j = 1, \ldots, q$, имеют частные производные $f^j_{x^i}, g^j_{x^i}, g^j_{y^i}, G^j_{x^i}$, $i = 1, \ldots, n$, g^j_t, g^j_T и непрерывны вместе с этими производными по совокупности своих аргументов при $x \in E^n$, $y \in E^n$, $u \in V$, $t \in \mathbb{R}$, $T \in \mathbb{R}$. Пусть $(x_0, u(\cdot), x(\cdot), t_0, T)$ — решение задачи (30) — (34), пусть фазовые ограничения (33) согласованы с концевыми в точке $z_* = (x_0, x(T), t_0, T)$. Тогда необходимо существуют числа $a = (a_0, \ldots, a_s)$, вектор-функции $\psi(t) = (\psi_1(t), \ldots, \psi_n(t))$, $\eta(t) = (\eta^1(t), \ldots, \eta^q(t))$, $t_0 \leqslant t \leqslant T$, такие, что*

1) функции $\eta^j(t)$ определены всюду на отрезке $[t_0, T]$, неотрицательны, не убывают, непрерывны слева при всех $t \in (t_0, T)$, постоянны на каждом интервале, принадлежащем множеству $\{t \in (t_0, T): G^j(x(t), t) < 0\}$; $j = 1, \ldots, q$; $\eta(t_0) = 0$;

2) $(a, \eta(T)) \neq 0$; $a_0 \geqslant 0, a_1 \geqslant 0, \ldots, a_m \geqslant 0$;

3) функция $\psi(t)$ определена всюду на $[t_0, T]$, имеет ограниченное изменение на $[t_0, T]$, непрерывна слева при всех $t \in (t_0, T]$, непрерывна во всех точках $t \in (t_0, T)$, в которых непрерывна функция $\eta(\cdot)$, непрерывна в точке $t = t_0$ справа, если $\eta(\cdot)$ непрерывна в этой точке справа, является решением интегрального уравнения

$$\psi(t) = \int_t^T H_x(x(\tau), u(\tau), \tau, \psi(\tau), a_0) d\tau - \sum_{i=1}^{q} \int_t^T G^i_x(x(\tau, \tau) d\eta^i - l_y(x_0, x(T), t_0, T, a) \quad (35)$$

при всех $t \in [t_0, T]$, где $\int_t^T G^i_x(x(\tau), \tau) d\eta^i$ — интеграл Римана — Стилтьеса по мере, порожденной функцией $\eta^i(t)$, $t_0 \leqslant t \leqslant T$;

4) справедливо равенство

$$\max_{v \in V} H(x(t), v, t, \psi(t), a_0) = H(x(t), u(t), t, \psi(t), a_0) \quad (36)$$

во всех точках $t \in (t_0, T]$, в которых оптимальное управление $u(\cdot)$ непрерывно;

5) выполнены условия трансверсальности

$$\psi(t_0) = l_x(x_0, x(T), t_0, T, a); \quad (37)$$

$$H(x(t_0), u(t_0 + 0), t_0, l_x(x_0, x(T), t_0, T, a), a_0) = -l_t(x_0, x(T), t_0, T, a) \quad (38)$$

(если в задаче (30)–(34) время t_0 закреплено, то условие (38) отсутствует);

$$H(x(T), u(T - 0), T, -l_y(x_0, x(T), t_0, T, a), a_0) = l_T(x_0, x(T), t_0, T, a) \quad (39)$$

(если в задаче (30)–(34) время T закреплено, то условие (39) отсутствует);

6) выполнено условие дополняющей нежесткости

$$a_i g^i((x_0, x(T), t_0, T) = 0, \quad i = 1, \ldots, m. \quad (40)$$

Из (35) при $t = T$ имеем $\psi(T) = -\sum_{i=1}^{q} G^i_x(x(T), T)(\eta^i(T) - \eta^i(T - 0)) - l_y(x_0, x(T), t_0, T, a)$. Как видим, это равенство превращается в условие трансверсальности (2.41) при $t = T$ лишь в том случае, когда сумма в правой части равна нулю. Предлагаем читателю провести сравнительный анализ остальных утверждений этой теоремы с соответствующими утверждениями теорем 2.1, 2.2.

Доказательство теоремы 3 проведем при дополнительном предположении, что вектор-функция $f(x, u, t)$ удовлетворяет условию Липшица (3.1) при всех $t \in \mathbb{R}$. Оно проводится по той же схеме, как и теоремы 2.1, 2.2, поэтому, опуская повторяющиеся детали из предыдущих доказательств, мы ниже подробнее остановимся лишь на тех тонкостях, которые связаны с наличием фазовых ограничений. Пусть $(x_{0*}, u_*(t), x_*(t), t_{0*}, T_*)$, $t_{0*} < T_*$, единственное решение задачи (30)–(34). Для построения конечномерной аппроксимации задачи (30)–(34) воспользуемся задачей (3.68)–(3.72) с добавлением к ней фазовых ограничений (33). Как и выше доказываем, что задача (3.68)–(3.72), (4) имеет единственное решение $(x_{0*}, \xi_* = 0, t_{0*}, T_*)$. Применяя метод штрафных функций для учета ограничений (3.70), (4), придем к задаче:

$$\Phi_k(x_0, \xi, t_0, T) = \int_{t_0}^{T} f^0(x(t), u(t, \xi), t) dt + g^0(x_0, x(T), t_0, T) + k_1 \sum_{i=1}^{s} (g_i^+(x_0, x(T), t_0, T))^2 + k_2 \sum_{i=1}^{q} \int_{t_0}^{T} (G_i^+(x(t), t))^2 dt \to \inf, \quad (41)$$

$$\dot{x}(t) = f(x(t), u(t, \xi), t), \quad t_0 \leqslant t \leqslant T; \quad x(t_0) = x_0, \quad (42)$$

$$(x_0, \xi, t_0, T) \in U_0 = \{|x_0 - x_{0*}| \leqslant 1, \ 0 \leqslant \xi_{lp} \leqslant d_N, \ l, p = 1, \ldots, N, \ |t_0 - t_{0*}| \leqslant \Delta t_{0N}, \ |T - T_*| \leqslant \Delta T_N\}; \quad (43)$$

здесь сохранены все обозначения аналогичной задачи (3.73)–(3.75); $G_i^+ = \max\{0; G^i\}$, $i = 1, \ldots, q$. В задаче (41)–(43) штрафной коэффициент $k = (k_1, k_2)$ представляет собой набор из двух независимых параметров k_1, $k_2 =$

$=1, 2, \ldots$; в дальнейшем нам придется совершить предельные переходы сначала при $k_1 \to \infty$, а затем при $k_2 \to \infty$.

Оценки (3.76)–(3.82) остаются верными и для задачи (41)–(43). Поэтому функция $\Phi_k(x_0, \xi, t_0, T)$ непрерывна на компактном множестве U_0 и задача (41)–(43) имеет хотя бы одно решение $(x_{0k}, \xi_k, t_{0k}, T_k) \in U_0$. Из теорем 5.15.1, 5.15.2 следует, что

$$\lim_{k\to\infty} x_{0k} = x_{0*}, \quad \lim_{k\to\infty} \xi_k = \xi_* = 0, \quad \lim_{k\to\infty} t_{0k} = t_{0*}, \quad \lim_{k\to\infty} T_k = T_*. \tag{44}$$

В (44) подразумевается двойной предельный переход при $k = (k_1, k_2) \to \infty$, поэтому найдется достаточно большой номер k_0 такой, что $|x_{0k} - x_{0*}| < 1/2$, $|\xi_k| < d_N/2$, $|t_{0k} - t_{0*}| < \Delta t_{0N}/2$, $|T_k - T_*| < \Delta T_N/2$, т. е. $(x_{0k}, \xi_k, t_{0k}, T_k) \in \in \operatorname{int} U_0$ $\forall k = (k_1, k_2)$, $k_1 \geqslant k_0$, $k_2 \geqslant k_0$. По аналогии с (3.86) для приращения функции (41) имеем формулу

$$\begin{aligned}
0 \leqslant \Delta\Phi_k &= \Phi_k(x_{0k}+\Delta x_0, \xi_k+\Delta\xi, t_{0k}+\Delta t_0, T_k+\Delta T) - \Phi(x_{0k}, \xi_k, t_{0k}, T_k) = \\
&= \int_{t_{0k}}^{T_k} [\Delta f^0 + \sum_{i=1}^{q} 2k_2 G_i^+(x_k(t), t)\langle G_x^i(x_k(t), t), \Delta x(t)\rangle] dt + \langle \Delta x_0, g_x^0(z_k) + \\
&\quad + \sum_{i=1}^{s} 2k_1 g_i^+(z_k) g_x^i(z_k)\rangle + \langle \Delta x(T_k), g_y^0(z_k) + \sum_{i=1}^{s} 2k_1 g_i^+(z_k) g_y^i(z_k)\rangle + \\
&\quad + \Delta t_0 [-f^0(x_k(t_{0k}), u_*(t_{0k}), t_{0k}) + g_t^0(z_k) + \sum_{i=1}^{s} 2k_1 g_i^+(z_k) g_t^i(z_k) - \\
&\quad - \sum_{i=1}^{q} k_2 (G_i^+(x_k(t_{0k}), t_{0k}))^2] + \Delta T [f^0(x_k(T_k), u_*(T_k), T_k) + g_T^0(z_k) + \\
&\quad + \sum_{i=1}^{s} 2k_1 g_i^+(z_k) g_T^i(z_k) + \langle g_y^0(z_k) + \sum_{i=1}^{s} 2k_1 g_i^+(z_k) g_y^i(z_k), \\
&\quad f(x_k(T_k), u_*(T_k), T_k)\rangle + \sum_{i=1}^{q} k_2 (G_i^+(x_k(T_k), T_k))^2] + \\
&\quad + o_k(|\Delta x_0| + |\Delta\xi| + |\Delta t_0| + |\Delta T|),
\end{aligned} \tag{45}$$

где $(x_{0k}+\Delta x_0, \xi_k+\Delta\xi, t_{0k}+\Delta t_0, T_k+\Delta T) \in U_0$, $z_k = (x_{0k}, x_k(T_k), t_{0k}, T_k)$, $x_k(t) = x(t, u(\cdot, \xi_k), x_{0k}, t_{0k})$, $t_{0k} \leqslant t \leqslant T_k$, приращение $\Delta x(t)$ траектории системы (42) удовлетворяет тем же условиям (3.84), $\Delta f^0 = f^0(x_k(t) + \Delta x(t), u(t, \xi_k + \Delta\xi), t) - f^0(x_k(t), u(t, \xi_k), t)$, $\lim_{\alpha\to 0} o_k(\alpha)/\alpha = 0$. Обозначим

$$\begin{aligned}
a_{0k} &= \Big[1 + \sum_{i=1}^{s} (2k_1 g_i^+(z_k))^2 + \sum_{i=1}^{q} \Big(\int_{t_{0k}}^{T_k} 2k_2 G_i^+(x_k(t), t) dt\Big)^2\Big]^{-1/2}, \\
a_{ik} &= a_{0k} \cdot 2k_1 g_i^+(z_k), \quad i = 1, \ldots, s; \\
\tilde{a}_{ik} &= a_{0k} \cdot 2k_2 \int_{t_{0k}}^{T_k} G_i^+(x_k(t), t) dt, \quad i = 1, \ldots, q; \\
A_k &= (a_{0k}, a_{1k}, \ldots, a_{sk}, \tilde{a}_{1k}, \ldots, \tilde{a}_{qk}).
\end{aligned} \tag{46}$$

Ясно, что

$$\begin{aligned}
&0 < a_{0k} \leqslant 1, \quad a_{1k} \geqslant 0, \ldots, a_{mk} \geqslant 0, \tilde{a}_{1k} \geqslant 0, \ldots, \tilde{a}_{qk} \geqslant 0, \\
&|A_k|^2 = \sum_{i=0}^{s} a_{ik}^2 + \sum_{i=1}^{q} \tilde{a}_{ik}^2 = 1.
\end{aligned} \tag{47}$$

Как и в § 3, для дальнейшего преобразования формулы (45) воспользуемся функциями $H(x, u, t, \psi, a_0)$, $l(x, y, t, T, a)$ и сопряженной задачей

$$\begin{aligned}
\dot{\psi}_k(t) &= -H_x(x_k(t), u(t, \xi_k), t, \psi_k(t), a_{0k}) + \\
&\quad + \sum_{i=1}^{q} a_{0k} \cdot 2k_2 G_i^+(x_k(t), t) G_x^i(x_k(t), t), \quad t_{0k} \leqslant t \leqslant T_k; \\
\psi_k(T_k) &= -\sum_{i=0}^{s} a_{ik} g_y^i(z_k) = -l_y(z_k, a_k),
\end{aligned} \tag{48}$$

где $a_k = (a_{0k}, \ldots, a_{sk})$. Справедливо равенство

$$\begin{aligned}
\langle \psi_k(T_k), \Delta x(T_k)\rangle &= \int_{t_{0k}}^{T_k} \langle \psi_k(t), f(x_k(t) + \Delta x(t), u(t, \xi_k + \Delta\xi), t) - \\
&\quad - f(x_k(t), u(t, \xi_k), t)\rangle dt + \langle \psi_k(t_{0k}), \Delta x_0\rangle - \\
&\quad - \langle \psi_k(t_{0k}), f(x_k(t_{0k}), u_*(t_{0k}), t_{0k})\rangle \Delta t_0 - \\
&\quad - \int_{t_{0k}}^{T_k} < H_x(x_k(t), u(t, \xi_k), t, \psi_k(t), a_{0k}), \Delta x(t) > dt + \\
&\quad + \sum_{i=1}^{q} \int_{t_{0k}}^{T_k} a_{0k} \cdot 2k_2 G_i^+(x_k(t), t)\langle G_x^i(x_k(t), t), \Delta x(t)\rangle dt + \\
&\quad + o_k(|\Delta x_0| + |\Delta\xi| + |\Delta t_0| + |\Delta T|),
\end{aligned} \tag{49}$$

аналогичное равенствам (3.40.A), (3.92). Умножим (45) на $a_{0k} > 0$ и преобразуем правую часть полученного выражения с учетом (3.41.A), (3.90), (46), (49). Получим

$$\begin{aligned}
0 \leqslant a_{0k}\Delta\Phi_k &= -\int_{t_{0k}}^{T_k} [H(x_k(t), u(t, \xi_k + \Delta\xi), t, \psi_k(t), a_{0k}) - \\
&\quad - H(x_k(t), u(t, \xi_k), t, \psi_k(t), a_{0k})] dt + \langle \Delta x_0, l_x(z_k, a_k) - \psi_k(t_{0k})\rangle + \\
&\quad + \Delta t_0 [H(x_k(t_{0k}), u_*(t_{0k}), t_{0k}, \psi(t_{0k}), a_{0k}) + l_t(z_k, a_k) - \\
&\quad - \sum_{i=1}^{q} k_2 a_{0k} (G_i^+(x_k(t_{0k}), t_{0k}))^2] + \Delta T [-H(x_k(T_k), u_*(T_k), T_k, \psi_k(T_k), a_{0k}) + \\
&\quad + l_T(z_k, a_k) + \sum_{i=1}^{q} k_2 a_{0k} (G_i^+(x_k(T_k), T_k))^2] + \\
&\quad + o_k(|\Delta x_0| + |\Delta\xi| + |\Delta t_0| + |\Delta T|).
\end{aligned} \tag{50}$$

Точка $(x_{0k}, \xi_k, t_{0k}, T_k)$ минимума функции $\Phi_k(x_0, \xi_k, t_0, T)$ на множестве U_0 является внутренней точкой этого множества при $k = (k_1, k_2)$, $k_1 \geqslant k_0$, $k_2 \geqslant k_0$. Отсюда и из (50) по аналогии с (3.94), (3.95) получаем

$$\begin{aligned}
&\psi_k(t_{0k}) = l_x(z_k, a_k), \\
&H(x_k(t_{0k}), u_*(t_{0k}), t_{0k}, \psi_k(t_{0k}), a_{0k}) = -l_t(z_k, a_k) + \sum_{i=1}^{q} k_2 a_{0k} (G_i^+(x_k(t_{0k}), t_{0k}))^2, \\
&H(x_k(T_k), u_*(T_k), T_k, \psi_k(T_k), a_{0k}) = l_T(z_k, a_k) - \sum_{i=1}^{q} k_2 a_{0k} (G_i^+(x_k(T_k), T_k))^2, \\
&[H(x_k(t), v_p, t, \psi_k(t), a_{0k}) - H(x_k(t), u_*(t), t, \psi_k(t), a_{0k})]\Big|_{t = t_{ip} + \xi_{ip}^k} \leqslant 0,
\end{aligned} \tag{51}$$

при всех $k = (k_1, k_2)$, $k_1 \geqslant k_0$, $k_2 \geqslant k_0$.

Зафиксируем в (46)–(48), (51) номер k_2 и перейдем к пределу при $k_1 \to \infty$. Последовательности $\{A_k = A_{k_1 k_2}\}$, $\{x_{0k} = x_{0k_1 k_2}\}$, $\{\xi_k = \xi_{k_1 k_2}\}$, $\{t_{0k} = t_{0k_1 k_2}\}$, $\{T_k = T_{k_1 k_2}\}$ ограничены, поэтому выбирая при необходимости подпоследовательности, можем считать

$$\lim_{k_1 \to \infty} A_{k_1 k_2} = A_{k_2} = (a_{k_2}, \widetilde{a}_{k_2}), \quad a_{k_2} = (a_{0k_2}, \ldots, a_{sk_2}),$$
$$\widetilde{a}_{k_2} = (\widetilde{a}_{1k_2}, \ldots, \widetilde{a}_{qk_2}), \quad \lim_{k_1 \to \infty} x_{0k_1 k_2} = x_{0k_2}, \quad \lim_{k_1 \to \infty} \xi_{k_1 k_2} = \xi_{k_2} = \{\xi_{lp}^{k_2}\}, \tag{52}$$
$$\lim_{k_1 \to \infty} t_{0k_1 k_2} = t_{0k_2}, \quad \lim_{k_1 \to \infty} T_{k_1 k_2} = T_{k_2}.$$

Из (47) и замкнутости U_0 следует, что

$$a_{0k_2} \geqslant 0, a_{1k_2} \geqslant 0, \ldots, a_{mk_2} \geqslant 0, \quad \widetilde{a}_{k_2} \geqslant 0, \quad |A_{k_2}| = 1, \quad (x_{0k_2}, \xi_{k_2}, t_{0k_2}, T_{k_2}) \in U_0. \tag{53}$$

Нетрудно видеть, что точка $(x_{0k_2}, \xi_{k_2}, t_{0k_2}, T_{k_2})$ является решением задачи

$$\Phi_{k_2}(x_0, \xi, t_0, T) = \int_{t_0}^{T} f^0(x(t), u(t, \xi), t)dt + g^0(x_0, x(T), t_0, T) + k_2 \sum_{i=1}^{q} \int_{t_0}^{T} (G_i^+(x(t), t))^2 dt \to \inf,$$
$$\dot{x}(t) = f(x(t), u(t, \xi), t), \quad t_0 \leqslant t \leqslant T; \quad x(t_0) = x_0, \tag{54}$$
$$g^i(x_0, x(T), t_0, T) \leqslant 0, \quad i = 1, \ldots, m,$$
$$g^i(x_0, x(T), t_0, T) = 0, \quad i = m+1, \ldots, s, \quad (x_0, \xi, t_0, T) \in U_0.$$

В самом деле, если к конечномерной задаче (54) при каждом фиксированном k_2 применяется метод штрафных функций, учитывая концевые ограничения с помощью штрафов $k_1 \sum_{i=1}^{s} (g_i^+)^2$, то придем как раз к задаче (41)–(43). Из теорем 5.15.1, 5.15.2 следует, что каждая предельная при $k_1 \to \infty$ точка $(x_{0k_1 k_2}, \xi_{k_1 k_2}, t_{0k_1 k_2}, T_{k_1 k_2})$ последовательности $(x_{0k_1 k_2}, \xi_{k_1 k_2}, t_{0k_1 k_2}, T_{k_1 k_2})$ решений задачи (41)–(43) будет решением задачи (54). Отсюда же имеем, что

$$g^i(x_{0k_2}, x_{k_2}(T_{k_2}), t_{0k_2}, T_{k_2}) \leqslant 0, \quad i = 1, \ldots, m;$$
$$g^i(x_{0k_2}, x_{k_2}(T_{k_2}), t_{0k_2}, T_{k_2}) = 0, \quad i = m+1, \ldots, s, \tag{55}$$

где $x_{k_2}(t) = x(t, u(\cdot, \xi_{k_2}), x_{0k_2}, t_{0k_2})$ — оптимальная траектория задачи (54). Из оценок (3.79)–(3.82), из (46), (52) получаем

$$\lim_{k_1 \to \infty} \max_{a \leqslant t \leqslant b} |x_{k_1 k_2}(t) - x_{k_2}(t)| = 0, \quad \lim_{k_1 \to \infty} x_{k_1 k_2}(t_{0k_1 k_2}) = x_{0k_2},$$
$$\lim_{k_1 \to \infty} x_{k_1 k_2}(T_{k_1 k_2}) = x_{k_2}(T_{k_2}), \quad \lim_{k_1 \to \infty} x_{k_1 k_2}(t_{lp} + \xi_{lp}^{k_1 k_2}) = x_{k_2}(t_{lp} + \xi_{lp}^{k_2}), \tag{56}$$
$$\lim_{k_1 \to \infty} \widetilde{a}_{ik_1 k_2} = \widetilde{a}_{ik_2} = k_2 a_{0k_2} \int_{t_{0k_2}}^{T_{k_2}} G_i^+(x_{k_2}(t), t))dt, \quad i = 1, 2, \ldots, q.$$

Здесь, как и в § 3, $[a, b]$ — это какой-либо фиксированный отрезок, $a < t_{0*} < T_* < b$. В силу (44), выбирая номера $k = (k_1, k_2)$ достаточно большими, можем считать, что $a < t_{0k} < T_k < b$, $a < t_{0k} + \Delta t_0 < T_k + \Delta T < b$ при всех малых $\Delta t_0, \Delta T$.

Рассуждая также, как в § 3, для решения $\psi_k(t) = \psi_{k_1 k_2}(t)$ задачи (48) и решения $\psi_{k_2}(t)$ задачи

$$\dot{\psi}_{k_2}(t) = -H_x(x_{k_2}(t), u(t, \xi_{k_2}), t, \psi_{k_2}(t), a_{0k_2}) + 2k_2 a_{0k_2} \sum_{i=1}^{q} G_i^+(x_{k_2}(t), t) G_x^i(x_{k_2}(t), t),$$
$$t_{0k_2} \leqslant t \leqslant T_{k_2}, \quad \psi_{k_2}(T_{k_2}) = -l_y(z_{k_2}, a_{k_2}), \tag{57}$$

где $z_{k_2} = (x_{0k_2}, x_{k_2}(T_{k_2}), t_{0k_2}, T_{k_2})$, доказываем оценки для $\max_{a \leqslant t \leqslant b} |\psi_{k_1 k_2}(t)|$, $\max_{a \leqslant t \leqslant b} |\psi_{k_1 k_2}(t) - \psi_{k_2}(t)|$, $|\psi_{k_1 k_2}(t) - |\psi_{k_1 k_2}(\tau)| \leqslant L_{k_2, N} |t - \tau| \ \forall t, \tau \in [a, b]$, аналогичные оценкам (3.77)–(3.82), из которых получаем равенства:

$$\lim_{k_1 \to \infty} \max_{a \leqslant t \leqslant b} |\psi_{k_1 k_2}(t) - \psi_{k_2}(t)| = 0, \quad \lim_{k_1 \to \infty} \psi_{k_1 k_2}(t_{0k_1 k_2}) = \psi_{k_2}(t_{0k_2}),$$
$$\lim_{k_1 \to \infty} \psi_{k_1 k_2}(T_{k_1 k_2}) = \psi_{k_2}(T_{k_2}), \quad \lim_{k_1 \to \infty} \psi_{k_1 k_2}(t_{lp} + \xi_{lp}^{k_1 k_2}) = \psi_{k_2}(t_{lp} + \xi_{lp}^{k_2}). \tag{58}$$

Пользуясь (52), (56), (58), непрерывностью функций H, l, кусочной непрерывностью $u_*(t)$, определением точек t_{lp}, ξ_{lp}^k, можем перейти к пределу при $k_1 \to \infty$ в соотношениях (51). Получим

$$\psi_{k_2}(t_{0k_2}) = l_x(z_{k_2}, a_{k_2}), \tag{59}$$

$$H(x_{k_2}(t_{0k_2}), u_*(t_{0k_2}), t_{0k_2}, \psi_{k_2}(t_{0k_2}), a_{0k_2}) = -l_t(z_{k_2}, a_{k_2}) + \sum_{i=1}^{q} k_2 a_{0k_2} (G_i^+(x_{k_2}(t_{0k_2}), t_{0k_2}))^2, \tag{60}$$

$$H(x_{k_2}(T_{k_2}), u_*(T_{k_2}), T_{k_2}, \psi_{k_2}(T_{k_2}), a_{0k_2}) = l_T(z_{k_2}, a_{k_2}) - \sum_{i=1}^{q} k_2 a_{0k_2} (G_i^+(x_{k_2}(T_{k_2}), T_{k_2}))^2,$$
$$[H(x_{k_2}(t), v_p, t, \psi_{k_2}(t), a_{0k_2}) - H(x_{k_2}(t), u_*(t), t, \psi_{k_2}(t), a_{0k_2})]\Big|_{t = t_{lp} + \xi_{lp}^{k_2}} \leqslant 0. \tag{61}$$

Далее нам нужно перейти к пределу при $k_2 \to \infty$ в полученных соотношениях. Убедимся, что

$$\lim_{k_2 \to \infty} x_{0k_2} = x_{0*}, \quad \lim_{k_2 \to \infty} \xi_{k_2} = \xi_* = 0, \quad \lim_{k_2 \to \infty} t_{0k_2} = t_{0*}, \quad \lim_{k_2 \to \infty} T_{k_2} = T_*. \tag{62}$$

Согласно (44) для любого $\varepsilon > 0$ найдется номер k_0 такой, что

$$|x_{0k} - x_{0*}| < \varepsilon, \quad |\xi_k| < \varepsilon, \quad |t_{0k} - t_{0*}| < \varepsilon, \quad |T_k - T_*| < \varepsilon$$

для всех $k = (k_1, k_2)$, $k_1 \geqslant k_0$, $k_2 \geqslant k_0$. Отсюда следует, что все предельные при $k_1 \to \infty$ точки $(x_{0k_2}, \xi_{k_2}, t_{0k_2}, T_{k_2})$ последовательности $\{(x_{0k}, \xi_k, t_{0k}, T_k)\}$ удовлетворяют неравенствам: $|x_{0k_2} - x_{0*}| < \varepsilon$, $|\xi_{k_2}| < \varepsilon$, $|t_{0k_2} - t_{0*}| < \varepsilon$, $|T_{k_2} - T_*| < \varepsilon \ \forall k_2 \geqslant k_0$. В силу произвольности $\varepsilon > 0$ отсюда получаем равенства (62). Далее, пользуясь неравенствами (3.78), (3.82) при $\xi = \xi_{k_2}$, $x_0 = x_{0k_2}$, $t_0 = t_{0k_2}$, (3.80), (3.81) при $x_0 = x_{0*}$, $\xi = \xi_* = 0$, $t_0 = t_{0*}$, $T = T_*$, $\Delta x_0 = x_{0k_2} - x_{0*}$, $\Delta \xi = \xi_{k_2} - \xi_*$, $\Delta t_0 = t_{0k_2} - t_{0*}$, $\Delta T = T_{k_2} - T_*$ с учетом (62) получим

$$\lim_{k_2 \to \infty} \max_{a \leqslant t \leqslant b} |x_{k_2}(t) - x_*(t)| = 0, \quad \lim_{k_2 \to \infty} x_{k_2}(t_{0k_2}) = x_*(t_{0*}),$$
$$\lim_{k_2 \to \infty} x_{k_2}(T_{k_2}) = x_*(T_*), \quad \lim_{k_2 \to \infty} x_{k_2}(t_{lp} + \xi_{lp}^{k_2}) = x_*(t_{lp}). \tag{63}$$

Так как фазовые ограничения согласованы с концевыми в точке $z_* = (x_{0*}, x_*(T_*), t_{0*}, T_*)$, а точка $z_{k_2} = (x_{0k_2}, x_{k_2}(T_{k_2}), t_{0k_2}, T_{k_2})$ удовлетворяет условиям (55) и $\{z_{k_2}\} \to z_*$ в силу (62), (63), то $G^i(x_{k_2}(t_{0k_2}), t_{0k_2}) \leqslant 0$, $G^i(x_{k_2}(T_{k_2}), T_{k_2}) \leqslant 0$, $i = 1, \ldots, q$, при всех $k_2 \geqslant k_0$. Тогда $G_i^+(x_{k_2}(t_{0k_2}), t_{0k_2}) = 0$, $G_i^+(x_{k_2}(T_{k_2}), T_{k_2}) = 0$, $i = 1, \ldots, q$, и равенства (60) перепишутся в виде

$$H(x_{k_2}(t_{0k_2}), u_*(t_{0k_2}), t_{0k_2}, l_x(z_{k_2}, a_{k_2}), a_{0k_2}) = -l_t(z_{k_2}, a_{k_2}),$$
$$H(x_{k_2}(T_{k_2}), u_*(T_{k_2}), T_{k_2}, -l_y(z_{k_2}, a_{k_2}), a_{0k_2}) = l_T(z_{k_2}, a_{k_2}) \quad \forall k_2 \geqslant k_0. \tag{64}$$

Так как последовательность $\{A_{k_2}\}$ из (52), (53) ограничена, то, выбирая при необходимости подпоследовательность, можем считать, что

$$\lim_{k_2 \to \infty} A_{k_2} = A_N = (a_N, \widetilde{a}_N), \quad a_N = (a_{0N}, \ldots, a_{sN}), \quad \widetilde{a}_N = (\widetilde{a}_{1N}, \ldots, \widetilde{a}_{qN}),$$
$$a_{0N} \geqslant 0, \ \ldots, \ a_{mN} \geqslant 0, \quad |A_N|^2 = \sum_{i=0}^{s} a_{iN}^2 + \sum_{i=1}^{q} \widetilde{a}_{iN} = 1. \tag{65}$$

Объясним появление индекса N в обозначении A_N предельной точки последовательности $\{A_{k_2}\}$. Напоминаем, что все рассуждения до сих пор мы проводили при фиксированном N, но для краткости записей зависимость используемых величин от N явно не подчеркивали. Однако в дальнейшем нам еще предстоит совершить предельный переход при $N \to \infty$, поэтому предельные точки всех последовательностей, зависящих от N, мы будем далее снабжать индексом N.

Для обоснования предельного перехода при $k_2 \to \infty$ задачу (57) удобнее записать в интегральной форме

$$\psi_{k_2}(t) = \int_t^{T_{k_2}} H_x(x_{k_2}(\tau), u(\tau, \xi_{k_2}), \tau, \psi_{k_2}(\tau), a_{0k_2}) d\tau -$$
$$- \sum_{i=1}^{q} \int_t^{T_{k_2}} 2k_2 a_{0k_2} \max\{G^i(x_{k_2}(\tau); 0\} G_x^i(x_{k_2}(\tau), \tau) d\tau - l_y(z_{k_2}, a_{k_2}), \quad t_{0k_2} \leqslant t \leqslant T_{k_2}. \tag{66}$$

Кроме того, интегралы из второго слагаемого правой части (66) ниже запишем в виде интеграла Римана — Стилтьеса. С этой целью введем функцию

$$\eta_{k_2}^i(t) = \begin{cases} 0, & a \leqslant \tau < t_{0k_2}, \\ \int\limits_{t_{0k_2}}^{t} 2k_2 a_{0k_2} \max\{G^i(x_{k_2}(\tau), \tau); 0\} d\tau, & t_{0k_2} \leqslant t \leqslant T_{k_2}, \\ \int\limits_{t_{0k_2}}^{T_{k_2}} 2k_2 a_{0k_2} \max\{G^i(x_{k_2}(\tau), \tau); 0\} d\tau = \widetilde{a}_{ik_2}, & T_{k_2} < \tau \leqslant b. \end{cases} \tag{67}$$

Очевидно, функция $\eta_{k_2}^i(t)$ непрерывна на отрезке $[a, b]$, не убывает, дифференцируема во всех точках $t \in [a, b]$, кроме, быть может, точек $t = t_{0k_2}$, $t = T_{k_2}$, причем ее производная

$$\frac{d\eta_{k_2}^i(t)}{dt} = \begin{cases} 0, & t \in [a, t_{0k_2}) \cup (T_{k_2}, b], \\ 2k_2 a_{0k_2} \max\{G^i(x_{k_2}(t), t); 0\} & t \in (t_{0k_2}, T_{k_2}). \end{cases} \tag{68}$$

Кроме того, с учетом условий (53), (56) мы имеем

$$\eta_{k_2}^i(a) = 0 \leqslant \eta_{k_2}^i(t) \leqslant \eta_{k_2}^i(b) = \widetilde{a}_{ik_2} \leqslant 1, \quad k_2 = 1, 2, \ldots \tag{69}$$

По теореме 2 из последовательности $\{\eta_{k_2}^i(t)\}$ можно выбрать подпоследовательность, сходящуюся к некоторой неубывающей функции $\widetilde{\eta}_N^i(t)$ в каждой точке $t \in [a, b]$. Не умаляя общности, можем считать, что

$$\lim_{k_2 \to \infty} \eta_{k_2}^i(t) = \widetilde{\eta}_N^i(t) \quad \forall t \in [a, b]. \tag{70}$$

Так как $\{t_{0k_2}\} \to t_{0*}$, $\{T_{k_2}\} \to T_*$, то из (67), (70) следует, что

$$\widetilde{\eta}_N^i(t) \equiv 0 \quad \forall t \in [a, t_{0*}), \quad \widetilde{\eta}_N^i(t) \equiv \widetilde{a}_{iN} \quad \forall t \in (T_*, b], \tag{71}$$

и поэтому последнее из равенств (65) можно записать в виде

$$|A_N|^2 = \sum_{i=0}^{s} a_{iN}^2 + \sum_{i=1}^{q} (\widetilde{\eta}_N^i(t))^2 = 1, \quad \forall t \in (T_*, b]. \tag{72}$$

Покажем, что функция $\widetilde{\eta}_N^i(t)$, $t \in [a, b]$, постоянна на каждом интервале, принадлежащем множеству $\{t \in (t_{0*}, T_*)$: $G^i(x_*(t), t) < 0\}$. Это множество открыто в силу непрерывности функции $G^i(x_*(t), t)$ на $[t_{0*}, T_*]$, поэтому является [393] объединением не более чем счетного множества интервалов (α_j, β_j). Покажем, что $\widetilde{\eta}_N(t) \equiv c_j^i$ $\forall t \in (\alpha_j, \beta_j)$.Возьмем произвольный отрезок $[\alpha, \beta] \subset (\alpha_j, \beta_j)$. Так как $G^i(x_*(t), t) < 0$ $\forall t \in (\alpha_j, \beta_j)$, то $\max\limits_{t \in [\alpha, \beta]} G^i(x_*(t), t) = -\gamma < 0$, а из

$$\lim_{k_2 \to \infty} \max_{t \in [\alpha, \beta]} |G^i(x_{k_2}(t), t) - G^i(x_*(t), t)| = 0 \tag{73}$$

следует, что $|G^i(x_{k_2}(t), t) - G^i(x_*(t), t)| \leqslant \frac{\gamma}{2}$ $\forall t \in [a, b]$ $\forall k_2 \geqslant k_0$. Поэтому

$$G^i(x_{k_2}(t), t) \leqslant |G^i(x_{k_2}(t), t) - G^i(x_*(t), t)| + G^i(x_*(t), t)) \leqslant -\frac{\gamma}{2}$$
$$\forall t \in [\alpha, \beta], \quad \forall k_2 \geqslant k_0,$$

и в силу определения (67) функции $\eta_{k_2}^i(t)$ имеем: $\eta_{k_2}^i(t) \equiv \eta_{k_2}^i(\alpha) = \text{const}$ $\forall t \in [\alpha, \beta]$, $\forall k_2 \geqslant k_0$. Отсюда и из (70) следует, что $\widetilde{\eta}_N^i(t) \equiv \widetilde{\eta}_N^i(\alpha)$ $\forall t \in [\alpha, \beta]$. В силу произвольности выбора отрезка $[\alpha, \beta]$ из интервала (α_j, β_j) заключаем, что

$$\widetilde{\eta}_N^i(t) \equiv \widetilde{\eta}_N^i(\alpha) = c_j^i \quad \forall t \in (\alpha_j, \beta_j). \tag{74}$$

Введем функцию

$$\widetilde{\psi}_{k_2}(t) = \begin{cases} \psi_{k_2}(t_{0k_2}), & a \leqslant t < t_{0k_2}, \\ \psi_{k_2}(t), & uadt_{0k_2} \leqslant t \leqslant T_{k_2}, \\ \psi_{k_2}(T_{k_2}) = -l_y(z_{k_2}, a_{k_2}), & T_{k_2} < t \leqslant b. \end{cases} \tag{75}$$

Из (66)–(68) следует, что функция $\widetilde{\psi}_{k_2}(t)$ является решением интегрального уравнения

$$\widetilde{\psi}_{k_2}(t) = \int_t^b \widetilde{H}_{xk_2}(\tau, \widetilde{\psi}_{k_2}(\tau)) d\tau - \sum_{i=1}^{q} \int_t^b G_x^i(x_{k_2}(\tau), \tau) d\eta_{k_2}^i - l_y(z_{k_2}, a_{k_2}), \quad a \leqslant t \leqslant b, \tag{76}$$

где

$$\widetilde{H}_{xk_2}(\tau,\psi)=\begin{cases} H_x(x_{k_2}(\tau),u(\tau,\xi_{k_2}),\tau,\psi,a_{0k_2}), & \tau\in[t_{0k_2},T_{k_2}],\\ 0, & \tau\in[a,t_{0k_2})\cup(T_{k_2},b],\end{cases} \tag{77}$$

а $\int_t^b G_x^i(x_{k_2}(\tau),\tau)d\eta_{k_2}^i=\int_t^b G_x^i(x_{k_2}(\tau),\tau)\frac{d\eta_{k_2}^i(\tau)}{d\tau}d\tau$ представляет собой интеграл Римана — Стилтьеса от непрерывной вектор-функции $G_x^i(x_{k_2}(\tau),\tau)$ по мере, порожденной функцией $\eta_{k_2}^i(\tau)$ из (67). Из (76) с помощью леммы 3.1, учитывая (62), (63), (69), (77), (3.76), непрерывность функций H_x, G_x^i, l_y по совокупности своих аргументов, получаем оценку

$$\max_{a\leqslant t\leqslant b}|\widetilde{\psi}_{k_2}(t)|\leqslant c_{0N}\quad\forall k\geqslant k_0 \tag{78}$$

и убеждаемся, что непрерывная функция $\widetilde{\psi}_{k_2}$ имеет ограниченное изменение на $[a,b]$, причем полное ее изменение оценивается так:

$$V_a^b(\widetilde{\psi}_{k_2})\leqslant(b-a)[\sup_{k_2\geqslant k_0}\sup_{\tau\in[a,b]}\sup_{|\psi|\leqslant c_{0N}}|\widetilde{H}_{xk_2}(\tau,\psi)|+$$
$$+\sum_{i=1}^q\sup_{k_2\geqslant k_0}\sup_{\tau\in[a,b]}|G_x^i(x_{k_2}(\tau),\tau)|]=c_{1N} \tag{79}$$

для $\forall k_2\geqslant k_0$; здесь и ниже через c_{iN} обозначаются константы, не зависящие от k_2, но, возможно, зависящие от N. Из (78), (79) и теоремы 2 следует, что последовательность $\{\widetilde{\psi}_{k_2}(t)\}$ содержит подпоследовательность, которая всюду на $[a,b]$ сходится к некоторой функции $\widetilde{\psi}_N(t)$ с ограниченным на $[a,b]$ изменением. Можем считать, что

$$\lim_{k_2\to\infty}\widetilde{\psi}_{k_2}(t)=\widetilde{\psi}_N(t),\quad\forall t\in[a,b]. \tag{80}$$

Из определения (77) функции $\widetilde{H}_{xk_2}(\tau,\psi)$, из соотношений (3.71), (62), (63), (80) вытекает, что

$$\lim_{k_2\to\infty}\widetilde{H}_{xk_2}(\tau,\widetilde{\psi}_{k_2}(\tau))=\widetilde{H}_{xN}(\tau,\widetilde{\psi}_N(\tau))=$$
$$=\begin{cases} H_x(x_*(\tau),u_*(\tau),\tau,\widetilde{\psi}_N(\tau),a_{0N}), & \tau\in[t_{0*},T_*],\\ 0, & \tau\in[a,t_{0*})\cup(T_*,b]\end{cases} \tag{81}$$
$$\sup_{\tau\in[a,b]}|\widetilde{H}_{xN}(\tau,\widetilde{\psi}_N(\tau))|\leqslant c_{2N}.$$

Пользуясь теоремой Лебега о предельном переходе под знаком интеграла [393], имеем

$$\lim_{k_2\to\infty}\int_t^b\widetilde{H}_{xk_2}(\tau,\widetilde{\psi}_{k_2}(\tau))d\tau=\int_t^b\widetilde{H}_{xN}(\tau,\widetilde{\psi}_N(\tau))d\tau\quad\forall t\in[a,b]. \tag{81.A}$$

Из (70), (73) и леммы 6 следует

$$\lim_{k_2\to\infty}\int_t^b G_x^i(x_{k_2}(\tau),\tau)d\eta_{k_2}^i=\int_t^b G_x^i(x_*(\tau),\tau)d\widetilde{\eta}_N^i$$

в точке $t=a$ и во всех точках $t\in(a,b]$, в которых $\widetilde{\eta}_N^i$ непрерывна. Кроме

того, $\lim_{k_2\to\infty}l_y(z_{k_2},a_{k_2})=l_y(z_{k_2},a_N)$ в силу (62), (63), (65). Таким образом, из равенства (76) при $k_2\to\infty$ с учетом (80) имеем

$$\widetilde{\psi}_N(t)=\int_t^b\widetilde{H}_{xN}(\tau,\widetilde{\psi}_N(\tau))d\tau-\sum_{i=1}^q\int_t^b G_x^i(x_*(\tau),\tau)d\widetilde{\eta}_N^i-l_y(z_*,a_N) \tag{82}$$

в точке $t=a$ и во всех точках $t\in(a,b]$, в которых функция $\widetilde{\eta}_N(t)=(\widetilde{\eta}_N^1(t),\ldots,\widetilde{\eta}_N^q(t))$ непрерывна. Введем функцию

$$\psi_N(t)=\int_t^b\widetilde{H}_{xN}(\tau,\widetilde{\psi}_N(\tau))d\tau-\sum_{i=1}^q\int_t^b G_x^i(x_*(\tau),\tau)d\widetilde{\eta}_N^i-l_y(z_*,a_N), \tag{83}$$

совпадающую с правой частью равенства (82). Так как все интегралы в (83) имеют смысл при всех $t\in[a,b]$, то функция $\psi_N(t)$ определена при всех $t\in[a,b]$. Функция $\widetilde{\psi}_N(t)$ из (80) также определена при всех $t\in[a,b]$, причем, как следует из (82), (83), равенство

$$\widetilde{\psi}_N(t)=\psi_N(t) \tag{84}$$

справедливо при $t=a$ и во всех точках $t\in[a,b]$, в которых $\widetilde{\eta}_N$ непрерывна. Так как мощность множества точек разрыва неубывающей функции не более, чем счетна [14], то можем сказать, что равенство (84) выполняется почти всюду на $[a,b]$. Тогда $\widetilde{H}_{xN}(\tau,\widetilde{\psi}_N(\tau))=\widetilde{H}_{xN}(\tau,\psi_N(\tau))$ почти всюду на $[a,b]$ и, следовательно,

$$\int_t^b\widetilde{H}_{xN}(\tau,\widetilde{\psi}_N(\tau))d\tau=\int_t^b\widetilde{H}_{xN}(\tau,\psi_N(\tau))d\tau\quad\forall t\in[a,b].$$

Отсюда и из (83) имеем

$$\psi_N(t)=\int_t^b\widetilde{H}_{xN}(\tau,\psi_N(\tau))d\tau-\sum_{i=1}^q\int_t^b G_x^i(x_*(\tau),\tau)d\widetilde{\eta}_N^i-l_y(z_*,a_N)\quad\forall t\in[a,b]. \tag{85}$$

Сравнение (85) с (82) показывает, что функция ψ_N в отличие от $\widetilde{\psi}_N$ является решением интегрального уравнения (85) всюду на $[a,b]$. В дополнение к равенству (84) оценим отклонение значений ψ_N и $\widetilde{\psi}_N$ в точках разрыва функции $\widetilde{\eta}_N$. Нетрудно убедиться, что

$$|\psi_N(t)-\widetilde{\psi}_N(t)|\leqslant 2c_1|\widetilde{\eta}_N(t+0)-\widetilde{\eta}_N(t-0)|,$$
$$c_1=\Big(\sum_{i=1}^q\|G_x^i(x_*(t),t)\|^2_{C[a,b]}\Big)^{1/2}\quad\forall t\in[a,b]. \tag{86}$$

В самом деле, из (70), 76), (80), (85) и леммы 6 имеем $|\widetilde{\psi}_N(t)-\psi_N(t)|=$ $=\lim_{k_2\to\infty}|\widetilde{\psi}_{k_2}(t)-\psi_N(t)|\leqslant\overline{\lim_{k_2\to\infty}}\sum_{i=1}^q\Big|\int_t^b G_x^i(x_{k_2}(\tau),\tau)d\widetilde{\eta}_{k_2}^i-\int_t^b G_x^i(x_*(\tau),\tau)d\eta_N^i\Big|\leqslant$ $\leqslant 2c_1|\widetilde{\eta}_N(t+0)-\widetilde{\eta}_N(t-0)|\ \forall t\in[a,b]$. Заметим, что в силу (81), (84)

$$\int_t^b\widetilde{H}_{xN}(\tau,\psi_N(\tau))d\tau=\begin{cases} 0, & t\in(T_*,b],\\ \int_t^{T_*}H_x(x_*(\tau),u_*(\tau),\tau,\psi_N(\tau),a_{0N})d\tau, & t\in[t_{0*},T_*],\\ \int_{t_{0*}}^{T_*}H_x(x_*(\tau),u_*(\tau),\tau,\psi_N(\tau),a_{0N})d\tau, & t\in[a,t_{0*}).\end{cases} \tag{87}$$

Кроме того, с учетом формул (6), (71) мы имеем

$$\int_t^b G_x^i(x_*(\tau),\tau)\,d\tilde{\eta}_N^i =$$

$$= \begin{cases} \int_t^b g(\tau)\,d\tilde{\eta}_N^i = 0, & t\in(T_*,b], \\ \int_t^{T_*} g(\tau)\,d\tilde{\eta}_N^i + \int_{T_*+0}^{b} g(\tau)\,d\tilde{\eta}_N^i = \int_t^{T_*} g(\tau)\,d\tilde{\eta}_N^i, & t\in[t_{0*},T_*], \\ \int_t^{t_{0*}-0} g(\tau)\,d\tilde{\eta}_N^i + \int_{t_{0*}}^{T_*} g(\tau)\,d\tilde{\eta}_N^i + \int_{T_*+0}^{b} g(\tau)\,d\tilde{\eta}_N^i = \int_{t_{0*}}^{T_*} g(\tau)\,d\tilde{\eta}_N^i, & t\in[a,t_{0*}), \end{cases} \tag{88}$$

где для краткости мы обозначили $g(\tau)=G_x^i(x_*(\tau),\tau)$, $\tau\in[a,b]$. Далее при каждом i, $1\leqslant i\leqslant q$, в силу (7) для меры на отрезке $[t_{0*},T_*]$, индуцированной с отрезка $[a,b]$, производящей функцией является

$$\eta_N^i(t)=\begin{cases} \tilde{\eta}_N^i(t_{0*}-0)=0, & t=t_{0*}, \\ \tilde{\eta}_N^i(t), & t\in(t_{0*},T_*), \\ \tilde{\eta}_N^i(T_*+0)=\tilde{a}_{iN}, & t=T_*; \quad N=1,2,\ldots \end{cases} \tag{89}$$

Переходя с отрезка $[a,b]$ к отрезку $[t_{0*},T_*]$, равенство (85) с учетом (87)–(89) можем записать в виде

$$\psi_N(t)=\int_t^{T_*} H_x(x_*(\tau),u_*(\tau),\tau,\psi_N(\tau),a_{0N})\,d\tau -$$

$$-\sum_{i=1}^{q}\int_t^{T_*} G_x^i(x_*(\tau),\tau)\,d\eta_N^i - l_y(z_*,a_N) \quad \forall t\in[t_{0*},T_*], \quad N=1,2,\ldots \tag{90}$$

Сразу заметим, что из (72), (89) следует равенство

$$|A_N|^2=\sum_{i=0}^{s}a_{iN}^2+\sum_{i=1}^{q}(\eta_N^i(T_*))^2=1, \quad N=1,2,\ldots \tag{91}$$

Докажем, что

$$\lim_{k_2\to\infty}\psi_{k_2}(t_{0k_2})=\psi_N(t_{0*}). \tag{92}$$

Из (76) при $t=a$ с учетом (75) имеем

$$\psi_{k_2}(t_{0k_2})=\tilde{\psi}_{k_2}(a)=\int_a^b \tilde{H}_{xk_2}(\tau,\tilde{\psi}_{k_2}(\tau))\,d\tau-\sum_{i=1}^{s}\int_a^b G_x^i(x_*(\tau),\tau)\,d\eta_{k_2}^i - l_y(z_{k_2},a_{k_2}),$$

$k_2=1,2,\ldots$. При $k_2\to\infty$ отсюда и из (26) при $t=a$, из (62), (63), (65), (70), (81.А) следует

$$\lim_{k_2\to\infty}\psi_{k_2}(t_{0k_2})=\tilde{\psi}_N(a)=\int_a^b \tilde{H}_{xN}(\tau,\tilde{\psi}_N(\tau))\,d\tau-\sum_{i=1}^{q}\int_a^b G_x^i(x_*(\tau),\tau)\,d\tilde{\eta}_N^i - l_y(z_*,a_N),$$

что с учетом (83)–(85), (87)–(90) равносильно (92).

Теперь можем перейти к пределу при $k_2\to\infty$ в равенствах (59), (64). С учетом (62), (63), (65), (92) и равенств $u_*(t)\equiv u_*(t_{0*}+0)$ $\forall t\leqslant t_{0*}$, $u_*(t)\equiv u_*(T_*-0)$ $\forall t\geqslant T_*$ имеем:

$$\psi_N(t_{0*})=l_x(z_*,a_N),$$

$$H(x_*(t_{0*}),u_*(t_{0*}+0),t_{0*},l_x(z_*,a_N),a_{0N})=-l_t(z_*,a_N), \tag{93}$$

$$H(x_*(T_*),u_*(T_*-0),T_*,-l_y(z_*,a_N),a_{0N})=l_T(z_*,a_N), \quad N=1,2,\ldots$$

Переход к пределу при $k_2\to\infty$ в неравенстве (61) будет более сложным, так как у нас нет оснований ожидать, что $\lim\limits_{k_2\to\infty}\tilde{\psi}_{k_2}(t_{lp}+\xi_{lp}^{k_2})=\tilde{\psi}_N(t_{lp})$. Покажем, что

$$\overline{\lim_{k_2\to\infty}}|\tilde{\psi}_{k_2}(t_{lp}+\xi_{lp}^{k_2})-\tilde{\psi}_N(t_{lp})|\leqslant 3c_1|\tilde{\eta}_N(t_{lp}+0)-\tilde{\eta}_N(t_{lp}-0)|, \quad l,p=1,\ldots,N, \tag{94}$$

где постоянная c_1 взята из (86). С этой целью воспользуемся равенствами (76) при $t=t_{lp}+\xi_{lp}^{k_2}$ и (82) при $t=t_{lp}$. Будем иметь:

$$\overline{\lim_{k_2\to\infty}}|\tilde{\psi}_{k_2}(t_{lp}+\xi_{lp}^{k_2})-\tilde{\psi}_N(t_{lp})|\leqslant \lim_{k_2\to\infty}\left|\int_{t_{lp}}^{b}[\tilde{H}_{xk_2}(\tau,\tilde{\psi}_{k_2}(\tau))-\tilde{H}_{xN}(\tau,\tilde{\psi}_N(\tau))]\,d\tau\right| +$$

$$+\lim_{k_2\to\infty}\left|\int_{t_{lp}}^{t_{lp}+\xi_{lp}^{k_2}}\tilde{H}_{xk_2}(\tau,\tilde{\psi}_{k_2}(\tau))\,d\tau\right| +$$

$$+\sum_{i=1}^{q}\lim_{k_2\to\infty}\left|\int_{t_{lp}+\xi_{lp}^{k_2}}^{b}(G_x^i(x_{k_2}(\tau),\tau)-G_x^i(x_*(\tau),\tau))\,d\eta_{k_2}^i\right| +$$

$$+\sum_{i=1}^{q}\overline{\lim_{k_2\to\infty}}\left|\int_{t_{lp}+\xi_{lp}^{k_2}}^{b}G_x^i(x_*(\tau),\tau)\,d\eta_{k_2}^i-\int_{t_{lp}}^{b}G_x^i(x_*(\tau),\tau)\,d\tilde{\eta}_N^i\right| +$$

$$+\lim_{k_2\to\infty}|l_y(z_{k_2},a_{k_2})-l_y(z_*,a_N)|.$$

Четвертое слагаемое из правой части этого неравенства не превышает правой части (94) в силу леммы 7, а остальные слагаемые равны нулю в силу (9), (62), (63), (65), (69), (78), (81). Оценка (94) доказана. Из (86), (89), (94), включений $t_{lp}\in(t_{0*},T_*)$ получим

$$\overline{\lim_{k_2\to\infty}}|\tilde{\psi}_{k_2}(t_{lp}+\xi_{lp}^{k_2})-\psi_N(t_{lp})|\leqslant 5c_1|\eta_N(t_{lp}+0)-\eta_N(t_{lp}-0)|, \quad l,p=1,\ldots,N. \tag{95}$$

Далее, пользуясь линейностью функции $H(x,u,t,\psi,a_0)$ по переменной ψ, получим

$$B_{k_2}^1\equiv|H(x_{k_2}(t),v_p,t,\tilde{\psi}_{k_2}(t),a_{0k_2})|_{t=t_{lp}+\xi_{lp}^{k_2}}-H(x_*(t),v_p,t,\psi_N(t),a_{0N})|_{t=t_{lp}}|\leqslant$$

$$\leqslant|a_{0k_2}f^0(x_{k_2}(t_{lp}+\xi_{lp}^{k_2}),v_p,t_{lp}+\xi_{lp}^{k_2})-a_{0N}f^0(x_*(t_{lp}),v_p,t_{lp})| +$$

$$+|\langle\tilde{\psi}_{k_2}(t_{lp}+\xi_{lp}^{k_2}),f(x_{k_2}(t_{lp}+\xi_{lp}^{k_2}),v_p,t_{lp}+\xi_{lp}^{k_2})-f(x_*(t_{lp}),v_p,t_{lp})\rangle| +$$

$$+|\langle\tilde{\psi}_{k_2}(t_{lp}+\xi_{lp}^{k_2})-\psi_N(t_{lp}),f(x_*(t_{lp}),v_p,t_{lp})\rangle|.$$

Первое и второе слагаемые из правой части этого неравенства стремятся к нулю при $k_2\to\infty$ в силу (62), (63), (65), (78), а верхний предел последнего

слагаемого оценивается сверху величиной из правой части (95), умноженной на $c_2 = \max\{\sup_{t\in[a,b]} |f(x_*(t), v_p, t)|; \sup_{t\in[a,b]} |f(x_*(t), u_*(t), t)|\}$. Поэтому

$$\overline{\lim_{k_2\to\infty}} B^1_{k_2} \leqslant 5c_1c_2|\eta_N(t_{lp}+0) - \eta_N(t_{lp}-0)|, \quad l, p = 1, \ldots, N \tag{96}$$

Аналогично для величины $B^2_{k_2} = |H(x_{k_2}(t), u_*(t), t, \widetilde{\psi}_{k_2}(t), a_{0k_2})|_{t=t_{lp}+\xi^{k_2}_{lp}} - H(x_*(t), u_*(t), t, \psi_N(t), a_{0N})|_{t=t_{lp}}|$ получаем

$$\overline{\lim_{k_2\to\infty}} B^2_{k_2} \leqslant 5c_1c_2|\eta_N(t_{lp}+0) - \eta_N(t_{lp}-0)|, \quad l, p = 1, \ldots, N. \tag{97}$$

Кроме того из (61), (75) следует

$$[H(x_*(t), v_p, t, \psi_N(t), a_{0N}) - H(x_*(t), u_*(t), t, \psi_N(t), a_{0N})]\Big|_{t=t_{lp}} \leqslant$$
$$\leqslant B^1_{k_2} + B^2_{k_2} + [H(x_{k_2}(t), v_p, t, \widetilde{\psi}_{k_2}(t), a_{0k_2}) -$$
$$- H(x_{k_2}(t), u_*(t), t, \widetilde{\psi}_{k_2}(t), a_{0k_2})]\Big|_{t=t_{lp}+\xi^{k_2}_{lp}} \leqslant B^1_{k_2} + B^2_{k_2}, \quad k_2 = 1, 2, \ldots$$

Отсюда при $k_2 \to \infty$ с учетом оценок (96), (97) имеем

$$[H(x_*(t), v_p, t, \psi_N(t), a_{0N}) - H(x_*(t), u_*(t), t, \psi_N(t), a_{0N})]\Big|_{t=t_{lp}} \leqslant$$
$$\leqslant 10c_1c_2|\eta_N(t_{lp}+0) - \eta_N(t_{lp}-0)|, \quad l, p = 1, \ldots, N. \tag{98}$$

Подчеркнем, что постоянные c_1, c_2 в (98) не зависят от N.

Покажем, что

$$a_{iN}g^i(z_*) = 0, \quad i = 1, \ldots, m. \tag{99}$$

Если $g^i(z_*) = 0$, то (99) очевидно. Пусть $g^i(z_*) < 0$. Так как $\lim_{k=(k_1,k_2)\to\infty} z_k = z_*$, то $\lim_{k\to\infty} g(z_k) = g(z_*) < 0$. Поэтому $g^i(z_{k_1k_2}) < 0$ $\forall k_1 \geqslant k_0$, $k_2 \geqslant k_0$ при достаточно большом номере k_0. Отсюда и из (46) следует, что $a_{ik_1k_2} = 0$ $\forall k_1 \geqslant k_0$, $k_2 \geqslant k_0$, так что $\lim_{k\to\infty} a_{ik} = a_{iN} = 0$ и, следовательно, равенство (99) верно и при $g^i(z_*) < 0$.

Наконец, остается перейти к пределу при $N \to \infty$ в соотношениях (65), (91), (93), (98), (99). Так как $|A_N| = 1$, то, переходя при необходимости к подпоследовательности, можем считать, что $\lim_{N\to\infty} A_N = A = (a, \widetilde{a})$, $a = (a_0, \ldots, a_s)$, $\widetilde{a} = (\widetilde{a}_1, \ldots, \widetilde{a}_q)$, причем в силу (65)

$$a_0 \geqslant 0, a_1 \geqslant 0, \ldots, a_m \geqslant 0, \quad |A|^2 = |a|^2 + |\widetilde{a}|^2 = 1. \tag{100}$$

Функции $\eta^i_N(t)$ из (89) не убывают на отрезке $[t_{0*}, T_*]$, $\eta^i_N(t_{0*}) = 0 \leqslant \eta^i_N(t) \leqslant \eta^i_N(T_*) = \widetilde{a}_{iN} \leqslant 1$ $\forall t \in [t_{0*}, T_*]$, $N = 1, 2, \ldots$, и согласно теореме 2, выбирая при необходимости подпоследовательность, можем считать, что

$$\lim_{N\to\infty} \eta^i_N(t) = \eta^i(t) \quad \forall t \in [t_{0*}, T_*], \tag{101}$$

где $\eta^i(t)$ — неубывающая функция, $\eta^i(t_{0*}) = 0 \leqslant \eta^i(t) \leqslant \eta^i(T_*) = \widetilde{a}_i \leqslant 1$. Положим $\eta(t) = (\eta^1(t), \ldots, \eta^q(t))$. Из (91) следует, что $\sum_{i=0}^{s} a_i^2 + |\eta(T_*)|^2 = 1$, так что набор $(a, \eta(T_*))$ нетривиален.

Кроме того, в силу (74), (101) на любом интервале (α_j, β_j), принадлежащем множеству $\{t \in [t_{0*}, T_*]: G^i(x_*(t), t) < 0\}$ функция $\eta^i(t)$ принимает постоянные значения. Функции $\eta^i(t)$ можем считать непрерывными слева во всех точках $t \in (t_{0*}, T_*)$ в силу замечания 1.

Функция $\widetilde{\psi}_N(t)$ из (80) имеет ограниченное изменение, поэтому $\widetilde{\psi}_N(t) \in L^n_1[a, b]$. Отсюда и из (84) следует, что $\psi_N(t) \in L^n_1[a, b]$. Пользуясь леммой 3.1, из (90) получаем оценку

$$|\psi_N(t)| \leqslant [f^0_{x\max}(T_* - t_{0*})(a_0+1) + c_1 + \max_{|d|\leqslant 1} |l_y(z_*, d)|] \cdot \exp(f_{x\max}(T_* - t_{0*})) = c_3,$$
$$\forall t \in [t_{0*}, T_*], \quad \forall N \geqslant N_0, \tag{102}$$

где N_0 достаточно большой номер, $f^0_{x\max} = \sup_{t\in[t_{0*}, T_*]} |f^0_x(x_*(t), u_*(t), t)|$, $f_{x\max} = \sup_{t\in[t_{0*}, T_*]} \|f_x(x_*(t), u_*(t), t)\|$, постоянная c_1 взята из (86). Кроме того, из (90) с помощью леммы 8 и свойств интеграла Лебега заключаем, что функция $\psi_N(t)$ имеет ограниченное изменение и ее полное изменение оценивается сверху следующим образом

$$V^{T_*}_{t_{0*}}(\psi_N) \leqslant \int_{t_{0*}}^{T_*} \sup_{|\psi|\leqslant c_3} \sup_{|\mu|\leqslant 1} |H_x(x_*(t), u_*(t), t, \psi, \mu)|d\tau + c_1 = c_4 \quad \forall N \geqslant N_0. \tag{103}$$

Из (102), (103) и теоремы 2 следует, что из $\{\psi_N(t)\}$ можно выбрать подпоследовательность, сходящуюся всюду на $[t_{0*}, T_*]$ к некоторой функции $\widetilde{\psi}(t)$ с ограниченным изменением. Можем считать, что

$$\lim_{N\to\infty} \psi_N(t) = \widetilde{\psi}(t), \quad \forall t \in [t_{0*}, T_*]. \tag{104}$$

Перейдем в равенстве (90) к пределу при $N \to \infty$. Пользуясь теоремой Лебега о предельном переходе под знаком интеграла в первом слагаемом правой части (90), леммой 5 — во втором слагаемом и равенствами (104), $\lim_{N\to\infty} a_N = a$, получим

$$\widetilde{\psi}(t) = \int_t^{T_*} H_x(x_*(\tau), u_*(\tau), \tau, \widetilde{\psi}(\tau), a_0)d\tau - \sum_{i=1}^{q} \int_t^{T_*} G^i_x(x_*(\tau), \tau)d\eta^i - l_y(z_*, a)$$

при $t = t_{0*}$ и во всех точках $t \in (t_{0*}, T_*]$ непрерывности функции $\eta(\tau) = (\eta^1(\tau), \ldots, \eta^q(\tau))$. Правую часть этого равенства обозначим через $\psi(t)$ и, как и в аналогичных равенствах (82)–(85), (90), показываем, что функция $\psi(t)$ определена на $[t_{0*}, T_*]$ всюду, удовлетворяет уравнению

$$\psi(t) = \int_t^{T_*} H_x(x_*(\tau), u_*(\tau), \tau, \psi(\tau), a_0)d\tau -$$
$$- \sum_{i=1}^{q} \int_t^{T_*} G^i_x(x_*(\tau), \tau)d\eta^i - l_y(z_*, a) \quad \forall t \in [t_{0*}, T_*], \tag{105}$$

и неравенству

$$|\widetilde{\psi}(t) - \psi(t)| \leqslant 2c_1|\eta(t+0) - \eta(t-0)| \quad \forall t \in [t_{0*}, T_*], \tag{106}$$

где постоянная c_1 взята из аналогичного неравенства (86), причем $\widetilde{\psi}(t) = \psi(t)$ при $t = t_{0*}$ и во всех точках непрерывности $\eta(t)$. Кроме того, функция $\psi(t)$ непрерывна слева в каждой точке $t \in (t_{0*}, T_*]$ — это следует из леммы 8, примененной ко второму слагаемому из правой части (105) и непрерывности остальных слагаемых. Из (93) при $N \to \infty$ имеем:

$$\begin{gathered}\psi(t_{0*}) = l_x(z_*, a),\\ H(x_*(t_{0*}), u_*(t_{0*}+0), t_{0*}, l_x(z_*, a), a_0) = -l_t(z_*, a),\\ H(x_*(T_*), u_*(T_*-0), T_*, -l_y(z_*, a), a_0) = l_T(z_*, a).\end{gathered} \tag{107}$$

Из (99) при $N \to \infty$ получаем

$$a_i g^i(z_*) = 0, \quad i = 1, \ldots, m. \tag{108}$$

Остается совершить предельный переход при $N \to \infty$ в неравенстве (98). Вспомним, что $\lim\limits_{N\to\infty} t_{l_p} = t_l$ в силу (3.12). Рассуждая также, как при доказательстве неравенства (94), устанавливаем, что

$$\overline{\lim_{N\to\infty}} |\psi_N(t_{l_p}) - \widetilde{\psi}(t_l)| \leqslant 3c_1 |\eta(t_l+0) - \eta(t_l-0)|.$$

Отсюда и из (106) следует

$$\overline{\lim_{N\to\infty}} |\psi_N(t_{l_p}) - \psi(t_l)| \leqslant 5c_1 |\eta(t_l+0) - \eta(t_l-0)|.$$

Тогда по аналогии с (96), (97) для величин

$$B_N^1 = |H(x_*(t), v_p, t, \psi_N(t), a_{0N})|_{t=t_{l_p}} - |H(x_*(t), v_p, t, \psi(t), a_0)|_{t=t_l},$$
$$B_N^2 = |H(x_*(t), u_*(t), t, \psi_N(t), a_{0N})|_{t=t_{l_p}} - |H(x_*(t), u_*(t), t, \psi(t), a_0)|_{t=t_l},$$
$$N = 1, 2, \ldots$$

имеем

$$\max\{\overline{\lim_{N\to\infty}} B_N^1, \overline{\lim_{N\to\infty}} B_N^2\} \leqslant 5c_1 c_2 |\eta(t_l+0) - \eta(t_l-0)| \tag{109}$$

с теми же постоянными c_1, c_2. Кроме того, из (98) следует

$$\begin{aligned}&[H(x_*(t), v_p, t, \psi(t), a_0) - H(x_*(t), u_*(t), t, \psi(t), a_0)]\Big|_{t=t_l} \leqslant\\ &\leqslant B_N^1 + B_N^2 + [H(x_*(t), v_p, t, \psi_N(t), a_{0N}) -\\ &- H(x_*(t), u_*(t), t, \psi_N(t), a_{0N})]\Big|_{t=t_{l_p}} \leqslant B_N^1 + B_N^2 +\\ &+ 10c_1 c_2 |\eta_N(t_{l_p}+0) - \eta_N(t_{l_p}-0)|, \quad l, p = 1, \ldots, N; \quad N = 1, 2, \ldots\end{aligned}$$

Отсюда при $N \to \infty$ с помощью леммы 4 и неравенства (109) получим

$$\begin{aligned}&H(x_*(t_l), v_p, t_l, \psi(t_l), a_0) - H(x_*(t_l), u_*(t_l), t_l, \psi(t_l), a_0) \leqslant\\ &\leqslant 20c_1 c_2 |\eta(t_l+0) - \eta(t_l-0)|.\end{aligned} \tag{110}$$

Так как функция $\eta^i(t)$ не убывает и $0 \leqslant \eta^i(t) \leqslant 1$, $t \in [t_{0*}, T_*]$, то отрезки $[\eta^i(t_l-0), \eta^i(t_l+0)] \in [0, 1]$ не пересекаются, и суммарная длина этих отрезков не превосходит 1. Это означает, что ряд $\sum\limits_{l=1}^{\infty} [\eta^i(t_l+0) - \eta^i(t_l-0)]$

с неотрицательными членами сходится и, следовательно, $\lim\limits_{l\to\infty} [\eta^i(t_l+0) - \eta^i(t_l-0)] = 0$, $i = 1, \ldots, q$. Зафиксируем произвольное число $\varepsilon > 0$. Тогда найдется номер l_0 такой, что

$$|\eta(t_l+0) - \eta(t_l-0)| \leqslant \varepsilon \quad \forall l \geqslant l_0.$$

Отсюда и из (110) имеем

$$H(x_*(t_l), v_p, t_l, \psi(t_l), a_0) - H(x_*(t_l), u_*(t_l), t_l, \psi(t_l), a_0) \leqslant 20c_1 c_2 \varepsilon \quad \forall l \geqslant l_0. \tag{111}$$

Множество точек $\{t_l, l \geqslant l_0\}$ образует плотное множество на $[t_{0*}, T_*]$. Поэтому для каждой точки $t \in [t_{0*}, T_*]$ непрерывности функций $u_*(t), \psi(t)$ найдется подпоследовательность $\{t_{l_\nu}\} \to t$ при $\nu \to \infty$. Из (111) при $l = l_\nu \to \infty$ получим

$$H(x_*(t), v_p, t, \psi(t), a_0) - H(x_*(t), u_*(t), t, \psi(t), a_0) \leqslant 20c_1 c_2 \varepsilon.$$

В силу произвольности $\varepsilon > 0$ отсюда имеем

$$H(x_*(t), v_p, t, \psi(t), a_0) - H(x_*(t), u_*(t), t, \psi(t), a_0) \leqslant 0$$

во всех точках $t \in [t_{0*}, T_*]$, в которых $u_*(t), \psi(t)$ непрерывны, и любых v_p из всюду плотного в V множества точек $\{v_p\}$. А тогда

$$H(x_*(\tau), v, \tau, \psi(\tau), a_0) - H(x_*(\tau), u_*(\tau), \tau, \psi(\tau), a_0) \leqslant 0$$

для всех $v \in V$ во все моменты τ непрерывности функций $u_*(t), \psi(t)$. Однако нам уже известно, что функция $\psi(t)$ непрерывна слева во всех точках $t \in (t_{0*}, T_*]$. Поэтому с помощью предельного перехода при $\tau \to t-0$ убеждаемся в справедливости последнего неравенства во всех точках промежутка $(t_{0*}, T_*]$, в которых управление $u_*(t)$ непрерывно, не взирая на возможные разрывы $\psi(t)$. Отсюда следует справедливость равенства (36).

Таким образом, доказаны все утверждения теоремы 3 в предположении, что задача (30)–(31) имеет единственное решение $(x_{0*}, u_*(t), x_*(t), t_{0*}, T_*)$. Если эта задача имеет много решений, то переходим к задаче

$$\begin{aligned}&J_1(x_0, x_0^{n+1}, u(\cdot), x(\cdot), x^{n+1}(\cdot), t_0, T) =\\ &= \int_{t_0}^{T} f^0(x(t), u(t), t)dt + g^0(x_0, x(T), t_0, T) +\\ &+ |x_0 - x_{0*}|^2 + |x_0^{n+1}|^2 + |t_0 - t_{0*}|^2 + (T - T_*)^2 \to \inf\end{aligned} \tag{112}$$

при условиях (31)–(34) и

$$\dot{x}^{n+1}(t) = |u(t) - u_*(t)|^2, \quad t_0 \leqslant t \leqslant T, \quad x^{n+1}(t_0) = x_0^{n+1}. \tag{113}$$

Задача (112), (31)–(34), (113) имеет единственное решение и для нее справедлив уже доказанный принцип максимума. Далее, рассуждая также, как в § 3 в аналогичной ситуации, убеждаемся, что в этой задаче оптимальные $x_*^{n+1}(t) \equiv 0$, $\psi_{n+1}(t) \equiv 0$ и, исключая из рассмотрений переменные x^{n+1}, ψ_{n+1}, убеждаемся в справедливости теоремы 3 в общем случае. □

Основываясь на теореме 3, по схеме, приведенной в § 2, нетрудно выписать краевую задачу принципа максимума для задачи (30)–(34). Следует сказать, что получающаяся при этом краевая задача сложна для исследования, и методы ее решения пока что разработаны мало [274; 332; 505].

3. Для иллюстрации теоремы 3 рассмотрим

Пример 3 [44, стр. 111]. Пусть требуется минимизировать функцию $J(u)=\int\limits_0^1 u(t)dt$ при условиях: $\dot{x}(t)=tu(t)$, $0\leqslant t\leqslant 1$; $x(0)=0$, $x(1)=0$; $x(t)\geqslant 0$, $t\in[0,1]$. Эта задача, очевидно, является частным случаем задачи (30)–(34) при $V=E^1$, $t_0=0$, $T=1$, $f^0=u$, $f=tu$, $G(x)=-x$, концы траекторий закреплены. Так как $|x(t)|=\left|\int\limits_0^1 \tau u(\tau)d\tau\right|\leqslant \frac{1}{2}t^2\sup\limits_{t\in[0,1]}|u(t)|$, то

$$\int\limits_t^1 u(\tau)d\tau=\int\limits_t^1\frac{\dot{x}(\tau)}{\tau}d\tau=-\frac{x(t)}{t}+\int\limits_t^1\frac{x(\tau)}{\tau^2}d\tau\geqslant-\frac{x(t)}{t}\geqslant-\frac{1}{2}t\sup_{t\in[0,1]}|u(t)|\quad\forall t>0.$$

Отсюда при $t\to+0$ получаем: $J(u)\geqslant 0$ для всех допустимых управлений $u(t)$. Но $J(0)=0$, так что $J_*=0$ и $u_*(t)\equiv 0$ — оптимальное управление, $x_*(t)\equiv 0$ — оптимальная траектория. Функция $H=-a_0u+\psi ut=u(-a_0+t\psi)$ достигает максимума на множестве $V=E^1$ только при $-a_0+t\psi=0$. Следовательно, $\psi(t)=\frac{a_0}{t}$, $0<t\leqslant 1$. Согласно теореме 3 функция $\psi(t)$ должна быть определена всюду на $[0,1]$ и иметь ограниченное изменение, что возможно лишь только при $a_0=0$. Таким образом, $\psi(t)\equiv 0$, $0<t\leqslant 1$. Далее, $l(x,y,a)=a_1x+a_2y$ и из условия трансверсальности при $t_0=0$ имеем: $\psi(0)=l_x(0,0,a_1)=a_1$. Так как $H_x\equiv 0$, $l_y=a_2$, то интегральное уравнение (35) запишется в виде $\psi(t)=\int\limits_t^1 d\eta-a_2=\eta(1)-\eta(t-0)-a_2$, $t\in[0,1]$, и, следовательно,

$$\psi(t)=\begin{cases}a_1=\eta(1)-\eta(0)-a_2, & t=0,\\ \eta(1)-\eta(t-0)-a_2=0, & 0<t\leqslant 1.\end{cases}$$

Тогда можно взять

$$\eta(t)=\begin{cases}0, & t=0,\\ A, & 0<t<1,\\ B, & t=1,\end{cases}$$

$a_1=A$, $a_2=B-A$, где A, B — произвольные числа, лишь бы $0\leqslant A\leqslant B\leqslant 1$, причем либо $A\neq 0$, либо $B-A\neq 0$. Тогда набор $(a_1,a_2,\eta(1))\neq 0$, причем функция $\eta(t)$ имеет скачок хотя бы в одном из концов отрезка $[0,1]$.

Отметим, что в этом примере функция $H=H(u,t,\psi,a_0)\equiv 0$ и условие максимума (36) не несет никакой полезной информации, так как оно не сужает класс допустимых управлений, подозрительных на экстремум. Оказывается [44], такое вырождение принципа максимума довольно характерно для задач оптимального управления с фазовыми ограничениями. Достаточные условия, когда в таких задачах условие максимума (36) не вырождается, получены в [44].

4. Сформулируем аналог теоремы 3.1 для задачи (30)–(34).

Теорема 4. *Пусть в дополнение к условиям теоремы* 1 *функции* $f^i(x,u,t)$, $i=0,1,\ldots,n$, $G^i(x,t)$, $i=1,\ldots,q$, *имеют производные* $f^i_t(x,u,t)$, $G^i_t(x,t)$, *непрерывные по совокупности аргументов* $(x,u,t)\in E^n\times V\times\mathbb{R}$, *пусть* $(x_0,u(\cdot),x(\cdot),t_0,T)$ *решение задачи* (30)–(34). *Тогда функция*

$$H(t)=\sup_{v\in V}H(x(t),v,t,\psi(t),a_0)\tag{114}$$

определена и принимает конечные значения на $[t_0,T]$, *непрерывна слева при всех* $t\in(t_0,T]$, *причем*

$$H(t)=H(t-0)=H(x(t),u(t-0),t,\psi(t),a_0)\quad\forall t\in(t_0,T];\tag{115}$$

непрерывна справа во всех точках непрерывности $\psi(\cdot)$, *причем*

$$H(t)=H(t+0)=H(x(t),u(t+0),t,\psi(t),a_0)\quad\forall t\in[t_0,T);\tag{116}$$

справедливо равенство

$$H(x(t),u(t-0),t,\psi(t),a_0)=-\int\limits_t^T H_t(x(\tau),u(\tau),\tau,\psi(\tau),a_0)d\tau+$$
$$+\sum_{i=1}^q\int\limits_t^T G^T_t(x(\tau),\tau)d\eta^i+l_T(x_0,x(T),t_0,T,a),\forall t\in[t_0,T]\tag{117}$$

(*при* $t=t_0$ *в* (117) *по определению считается* $u(t_0-0)=u(t_0+0)$).

Равенство (117) обобщает (3.118) на случай задачи (30)–(34).

Доказательство. Конечность значений функции (114) и ее непрерывность слева на промежутке $(t_0,T]$, равенства (115), (116) устанавливаются также как аналогичные утверждения в теореме 3.1. Для доказательства равенства (117) обратимся к v-задаче, которая получается из задачи (3.103)–(3.107) добавлением фазовых ограничений

$$G^i(\tilde{x}(\theta),\chi(\theta))\leqslant 0,\quad\theta_1\leqslant\theta\leqslant\theta_2,\quad i=1,\ldots,q.\tag{118}$$

Пусть $(x_0,u(\cdot),x(\cdot),t_0,T)$ — оптимальный процесс задачи (30)–(34). Тогда согласно формулам (3.111) процесс

$$(\tilde{x}_0=x_0,\ \tilde{u}(\theta)=u(\theta),\ v(\theta)=0,\ \tilde{x}(\theta)=x(\theta),\ \chi(\theta)=\theta,\ \theta_1=t_0,\ \theta_2=T)\tag{119}$$

будет оптимальным для задачи (3.103)–(3.107), (118). Нетрудно проверить, что если фазовые ограничения (33) согласованы с концевыми в точке $z_*=(x_0,x(T),t_0,T)$, то для процесса (119) фазовые ограничения (118) также будут согласованы с концевыми (3.106) в точке $\tilde{z}_*=(\tilde{x}(\theta_1),\tilde{x}(\theta_2),\chi(\theta_1),\chi(\theta_2))$. Функции Понтрягина H и $\tilde{H}$ задач (30)–(34) и (3.103)–(3.107), (118) связаны посредством тех же формул (3.112). В силу теоремы 3 для процесса (119) существуют $a=(a_0,a_1,\ldots,a_s)$, $\eta(\theta)=(\eta^1(\theta),\ldots,\eta^q(\theta))$, $\psi(\theta)=(\psi_1(\theta),\ldots,\psi_n(\theta),\psi_0(\theta))$, $\theta_1\leqslant\theta\leqslant\theta_2$, такие, что

$$\psi(\theta)=\int\limits_\theta^{\theta_2}H_x(\tilde{x}(\tau),\tilde{u}(\tau),\tau,\psi(\tau),a_0)d\tau-$$
$$-\sum_{i=1}^q\int\limits_\theta^{\theta_2}G^i_x(\tilde{x}(\tau),\tau)d\eta^i-l_y(\tilde{z}_*,a)\quad\forall\theta\in[\theta_1,\theta_2],\tag{120}$$

$$\psi_0(\theta)=\int_{\theta}^{\theta_2} H_t(\tilde{x}(\tau),\tilde{u}(\tau),\tau,\psi(\tau),a_0)d\tau-$$
$$-\sum_{i=1}^{q}\int_{\theta}^{\theta_2} G_t^i(\tilde{x}(\tau),\tau)d\eta^i-l_T(\tilde{z}_*,a)\quad\forall\theta\in[\theta_1,\theta_2],\qquad(121)$$

$$\tilde{H}(\theta)=\sup_{(w,v)\in V\times\{|v|\leqslant 1/2\}}\tilde{H}(\tilde{x}(\theta),w,v,\theta,\psi(\theta),\psi_0(\theta),a_0)=$$
$$=\tilde{H}(\tilde{x}(\theta),\tilde{u}(\theta),v(\theta)=0,\theta,\psi(\theta),\psi_0(\theta),a_0)=$$
$$=H(\tilde{x}(\theta),\tilde{u}(\theta),\theta,\psi(\theta),a_0)+\psi_0(\theta)\qquad(122)$$

во всех точках $\theta\in(\theta_1,\theta_2]$, в которых управление $(\tilde{u}(\tau),v(\theta)\equiv 0)$ непрерывно; функция $\tilde{H}(\theta)$ непрерывна слева на промежутке $(\theta_1,\theta_2]$;

$$\psi(\theta_1)=l_x(\tilde{z}_*,a),\quad\psi_0(\theta_1)=l_t(\tilde{z}_*,a),\qquad(123)$$

$$\tilde{H}(\tilde{x}(\theta_1),\tilde{u}(\theta_1+0),v(\theta_1)=0,\theta_1,l_x(\tilde{z}_*,a),l_t(\tilde{z}_*,a),a_0)=$$
$$=H(\tilde{x}(\theta_1),\tilde{u}(\theta_1+0),\theta_1,l_x(\tilde{z}_*,a),a_0)+l_t(\tilde{z}_*,a)=-l_{\theta_1}(\tilde{z}_*,a)=0,\qquad(124)$$

$$\tilde{H}(\tilde{x}(\theta_2),\tilde{u}(\theta_2-0),v(\theta_2)=0,\theta_2,-l_y(\tilde{z}_*,a),-l_T(\tilde{z}_*,a),a_0)=$$
$$=H(\tilde{x}(\theta_2),\tilde{u}(\theta_2-0),\theta_2,-l_y(\tilde{z}_*,a),a_0)-l_T(\tilde{z}_*,a)=l_{\theta_2}(\tilde{z}_*,a)=0;\qquad(125)$$

остальные утверждения теоремы 1 нам явно не понадобятся. Так как $v(\theta)\equiv\equiv 0$ является внутренней точкой множества $|v|\leqslant 1/2$, функция $\tilde{H}$ линейна по v, то, рассуждая также, как при выводе равенства (3.116), из (122) получаем

$$\psi_0(\theta)=-H(\tilde{x}(\theta),\tilde{u}(\theta-0),\theta,\psi(\theta),a_0)\quad\forall\theta\in(\theta_1,\theta_2].\qquad(126)$$

Кроме того, из (123), (124) следует

$$\psi_0(\theta_1)=l_t(\tilde{z}_*,a)=-H(\tilde{x}(\theta_1),\tilde{u}(\theta_1+0),\theta_1,\psi(\theta_1),a_0),$$

так что равенство (126) имеет место и при $\theta=\theta_1$ (напоминаем, что $\tilde{u}(\theta_1+$ $+0)=\tilde{u}(\theta_1-0)$ по определению). Отсюда и из (121) получаем

$$H(\tilde{x}(\theta),\tilde{u}(\theta-0),\theta,\psi(\theta),a_0)=-\int_{\theta}^{\theta_2} H_t(\tilde{x}(\tau),\tilde{u}(\tau),\tau,\psi(\tau),a_0)d\tau+$$
$$+\sum_{i=1}^{q}\int_{\theta}^{\theta_2} G_t(\tilde{x}(\tau),\tau)d\eta^i+l_T(\tilde{x}_0,\tilde{x}(\theta_2),\theta_1,\theta_2),a)\quad\forall\theta\in[\theta_1,\theta_2].$$

Возвращаясь к исходным обозначениям в силу (119) с заменой θ на t, приходим к равенству (117). Теорема 4 доказана. □

Предлагаем читателю сформулировать и доказать варианты теорем 3, 4 для задачи (30)–(34), когда один из моментов t_0 или T или оба эти момента закреплены.

Другие варианты принципа максимума для задач с фазовыми ограничениями при различных предположениях о динамике системы, о класса допустимых управлений читатель найдет, например, в [36–39; 41; 42; 44; 46; 50; 212; 225; 274; 276; 278–280; 332; 358; 380; 429; 430; 505; 578; 587].

§ 5. Связь между принципом максимума и классическим вариационным исчислением

Основной задачей классического вариационного исчисления, как известно [101; 121; 217; 376; 406; 409; 781], является следующая задача: среди всех непрерывных кривых $x=x(t)$, $t_0\leqslant t\leqslant T$, имеющих кусочно-непрерывные производные $\dot{x}(t)$ и удовлетворяющих условиям $x(t_0)\in S_0$, $x(T)\in S_1$, найти такую, которая доставляет функции (функционалу)

$$J=\int_{t_0}^{T} f^0(x(t),\dot{x}(t))dt$$

минимальное значение. Здесь $x(t)=(x^1(t),\ldots,x^n(t))$, S_0 и S_1 — заданные множества в E^n. Будем предполагать, что функция $f^0(x,u,t)$ непрерывна и имеет непрерывные производные f^0_x, f^0_u, f^0_t, f^0_{ut}, f^0_{ux}, f^0_{uu} при $(x,u,t)\in E^n\times E^n\times[t_0,\infty)$. Далее, в этом параграфе для простоты мы ограничимся рассмотрением случая закрепленного левого конца: $x(t_0)=x_0$, t_0 — задано, а правый конец $x(T)$ либо закреплен: $x(T)=x_1$, T — задано, либо свободный: $S_1\equiv E^n$, T — задано, либо является подвижным и лежит на заданной гладкой кривой

$$S_1=S_1(T)=\{y\in E^n:\ g(y,T)=y-\varphi(T)=0\},\quad T\in\mathbb{R}=\{-\infty<t<\infty\}.$$

Обозначим $\dot{x}(t)=u(t)$ и запишем рассматриваемую задачу в эквивалентном виде как задачу оптимального управления:

$$J(u)=\int_{t_0}^{T} f^0(x(t),u(t),t)dt\to\inf,$$
$$\dot{x}(t)=u(t),\quad t_0\leqslant t\leqslant T,\quad x(t_0)=x_0,\quad x(T)\in S_1(T).$$

Для исследования этой задачи воспользуемся принципом максимума Понтрягина. Выпишем функцию Гамильтона — Понтрягина

$$H(x,u,t,\psi,a_0)=-a_0f^0(x,u,t)+\langle\psi,u\rangle\qquad(1)$$

и сопряженную систему

$$\dot{\psi}=-H_x=a_0f^0_x(x,u,t),\quad a_0\geqslant 0.\qquad(2)$$

Для решения $(u(t),x(t))$, $t_0\leqslant t\leqslant T$, рассматриваемой задачи должно выполняться необходимое условие

$$H(x(t),u(t),t,\psi(t),a_0)=\sup_{u\in E^n}H(x(t),u,t,\psi(t),a_0),\quad t_0\leqslant t\leqslant T,\qquad(3)$$

где $\psi(t)$ — решение системы (2) при $u=u(t)$, $x=x(t,u)$, $t_0\leqslant t\leqslant T$. Так как в данном случае множество V совпадает со всем пространством E^n, то условие (3) может соблюдаться лишь в стационарной точке, т. е.

$$H_u\equiv-a_0f^0_u(x(t),u(t),t)+\psi(t)=0,\quad t_0\leqslant t\leqslant T.\qquad(4)$$

Отсюда ясно, что $a_0\neq 0$, так как при $a_0=0$ из (4) получаем $\psi(t)\equiv 0$, что

противоречит теоремам 2.1, 2.2. Следовательно, можно считать, что $a_0 = 1$. Тогда соотношения (1)–(4) перепишутся соответственно в виде

$$H(x, u, t, \psi) = -f^0(x, u, t) + \langle \psi, u \rangle, \tag{5}$$

$$\dot{\psi}(t) = f_x^0(x(t), u(t), t), \quad t_0 \leqslant t \leqslant T, \tag{6}$$

$$H(x(t), u(t), t, \psi(t)) = \sup_{u \in E^n} H(x(t), u, t, \psi(t)), \quad t_0 \leqslant t \leqslant T, \tag{7}$$

$$\psi(t) = f_u^0(x(t), u(t), t), \quad t_0 \leqslant t \leqslant T. \tag{8}$$

Из уравнения (6) имеем: $\psi(t) = \int\limits_{t_0}^{t} f_x^0(x(\tau), u(\tau), \tau) d\tau + \psi(t_0)$. С учетом (8) отсюда получаем

$$f_u^0(x(t), u(t), t) = \int\limits_{t_0}^{t} f_x^0(x(\tau), u(\tau), \tau) d\tau + \psi(t_0), \quad t_0 \leqslant t \leqslant T. \tag{9}$$

Уравнение (9) называется *уравнением Эйлера в интегральной форме*; здесь $u(t) = \dot{x}(t)$, $t_0 \leqslant t \leqslant T$. Если (9) продифференцировать по t, то получим *уравнение Эйлера* классического вариационного исчисления *в дифференциальной форме*

$$\frac{d}{dt}(f_u^0(x(t), u(t), t)) - f_x^0(x(t), u(t), t) = 0, \quad u(t) = \dot{x}(t), \quad t_0 \leqslant t \leqslant T.$$

Далее, необходимым условием для достижения функцией $H(x(t), u(t), t, \psi(t))$ максимума при $u = u(t)$ является неположительность следующей квадратичной формы (теорема 2.2.1):

$$\sum_{i,j=1}^{n} H_{u^i u^j}(x(t), u(t), t, \psi(t)) \xi_i \xi_j \leqslant 0$$

при любых

$$\xi = (\xi_1, \xi_2, \ldots, \xi_n), \quad t_0 \leqslant t \leqslant T.$$

Отсюда, учитывая выражение (5) для H, имеем

$$\sum_{i,j=1}^{n} f_{u^i u^j}(x(t), u(t), t) \xi_i \xi_j \geqslant 0, \quad \xi \in E^n, \quad t_0 \leqslant t \leqslant T. \tag{10}$$

Условие (10) называется необходимым *условием Лежандра*. В частности, при $n = 1$ отсюда имеем

$$f_{uu}^0(x(t), u(t), t) \geqslant 0, \quad t_0 \leqslant t \leqslant T.$$

Теперь выведем необходимое условие Вейерштрасса. Для этого перепишем условие (7) с учетом (5), (8) в следующем виде:

$$0 \leqslant H(x(t), u(t), t, \psi(t)) - H(x(t), v, t, \psi(t)) =$$

$$= f^0(x(t), v, t) - f^0(x(t), u(t), t) - \langle v - u(t), f_u^0(x(t), u(t), t) \rangle. \tag{11}$$

Это неравенство справедливо при любых $v \in E^n$, $t \in [t_0, T]$, если $(u(t), x(t))$, $t_0 \leqslant t \leqslant T$ — решение исходной задачи. Введем в рассмотрение функцию

$$E(t, x, u, v) \equiv f^0(x, v, t) - f^0(x, u, t) - \langle v - u, f_u^0(x, u, t) \rangle, \tag{12}$$

называемую *функцией Вейерштрасса*. Известно, что в классическом вариационном исчислении *необходимое условие Вейерштрасса*

$$E(t, x(t), u(t), v) \geqslant 0, \quad t_0 \leqslant t \leqslant T, \quad v \in E^n,$$

является следствием неравенства (11). Далее, из теорем 2.1, 2.3, 3.1 следует непрерывность функций $\psi(t)$ и $H(t) = \sup\limits_{u \in E^n} H(x(t), u, t, \psi(t))$ на отрезке $[t_0, T]$. Поэтому с учетом соотношений (5), (7), (8) имеем

$$\begin{gathered} [f_u^0(x(t), u(t), t)]_t = 0, \\ [\langle u(t), f_u^0(x(t), u(t), t) \rangle - f^0(x(t), u(t), t)]_t = 0, \quad t_0 \leqslant t \leqslant T; \end{gathered} \tag{13}$$

здесь принято обозначение: $[z(t)]_t = z(t+0) - z(t-0)$. Поскольку равенства (13) выполнены при всех t, $t_0 \leqslant t \leqslant T$, то они сохраняют силу, в частности, и в те моменты t, когда функция $x(t)$ может иметь излом, т. е. производная $\dot{x}(t)$ терпит разрыв. Таким образом, если учесть связь $u(t) \equiv \dot{x}(t)$, условия (13) превращаются в известные из классического вариационного исчисления *условия Эрдмана — Вейерштрасса* в точках излома кривой $x(t)$, $t_0 \leqslant t \leqslant T$.

Перейдем к рассмотрению условий на правом конце оптимальной кривой $x(t)$, $t_0 \leqslant t \leqslant T$. Если конец $x(T)$ свободен, то в силу условия (2.26) тогда $\psi(T) = 0$. Отсюда с учетом выражения (8) имеем

$$f_u^0(x(T), u(T), T) = 0. \tag{14}$$

Если правый конец $x(T)$ подвижен, точнее, $x(T) \in S_1(T) = \{y \in E^n: g^j(y, T) \equiv y^j - \varphi_j(T) \equiv 0, \ j = 1, \ldots, n\}$, то согласно условиям (2.30), (2.41) существуют постоянные $a_1, \ldots, a_n$ такие, что

$$\psi_i(T) = -\sum_{j=1}^{n} a_j g_{y^i}^j(x(T), T) = -a_i,$$

$$H(x(T), u(T), T, \psi(T)) = \sum_{j=1}^{n} a_j g_t^j(x(T), T) =$$

$$= -\sum_{j=1}^{n} a_j \dot{\varphi}_j(T) = \sum_{j=1}^{n} \psi_j(T) \dot{\varphi}_j(T) = \langle \psi(T), \dot{\varphi}(T) \rangle.$$

Так как $H(x, u, t, \psi) \equiv \langle \psi, u \rangle - f^0(x, u, t)$ и $\psi(t)$ выражается формулой (8), то последнее равенство можно переписать так:

$$f^0(x(T), u(T), T) + \langle f_u^0(x(T), u(T), T), \dot{\varphi}(T) - u(T) \rangle = 0. \tag{15}$$

Условия (14), (15) при учете связи $\dot{x}(t) \equiv u(t)$ выражают собой известные в классическом вариационном исчислении условия *трансверсальности* для свободного и соответственно подвижного правого конца.

Таким образом, в случае $V \equiv E^n$ из принципа максимума следуют все основные необходимые условия, известные в классическом вариационном исчислении [101; 121; 217; 376; 406; 409; 781]. Однако, если V — замкнутое множество и $V \neq E^n$, то соотношение (4), вообще говоря, не выполняется. Более того, имеются примеры, когда и условие Вейерштрасса в этом случае не имеет места ([587, с. 284]). Условие максимума (2.9), являясь естественным обобщением условия Вейерштрасса из классического вариационного исчисления, имеет то существенное преимущество перед условием Вейерштрасса, что оно применимо для любого (в частности, и замкнутого) множества $V \in E^r$ и для более общих задач. Заметим, что случай замкнутого множества наиболее интересен в прикладных вопросах, поскольку значения оптимальных управлений чаще всего лежат на границе V.

ГЛАВА 7

Динамическое программирование

В этой главе мы остановимся на методе динамического программирования, часто используемом для численного решения задач оптимального управления при наличии фазовых ограничений, конечномерных задач минимизации специального вида. С помощью динамического программирования можно также наметить пути решения проблемы синтеза, сформулировать достаточные условии оптимальности для задач оптимального управления и т. д. Изложение метода динамического программирования начнем с простейшей схемы Беллмана для задачи оптимального управления, затем опишем более совершенную и удобную для практики схему Моисеева, а также обсудим некоторые другие аспекты этого метода [13; 16; 78–81; 92–94; 105–108; 140; 202; 253; 254; 332; 369; 417; 418; 442; 471; 493; 497–499; 517; 541; 614; 616; 620; 643; 674; 719; 724; 737; 744; 753; 754].

§ 1. Схема Беллмана. Проблема синтеза для дискретных систем

1. Рассмотрим следующую задачу оптимального управления:

$$J(t_0, x_0, u(\cdot)) = \int_{t_0}^{T} f^0(x(\tau), u(\tau), \tau)d\tau + \Phi(x(T)) \to \inf, \tag{1}$$

$$\dot{x}(\tau) = f(x(\tau), u(\tau), \tau), \quad t_0 \leqslant \tau \leqslant T, \quad x(t_0) = x_0, \tag{2}$$

$$x(\tau) \in G(\tau), \quad t_0 \leqslant \tau \leqslant T, \tag{3}$$

$$u = u(\tau) \in V(\tau), \quad t_0 \leqslant \tau \leqslant T; \quad u(\tau) \text{ — кусочно-непрерывна}; \tag{4}$$

моменты времени t_0, T будем считать заданными (описание обозначений см. в § 6.1).

Для приближенного решения этой задачи разобьем отрезок $t_0 \leqslant t \leqslant T$ на N частей точками $t_0 < t_1 < \ldots < t_{N-1} < t_N = T$, и, приняв эти точки в качестве узловых, интеграл в (1) заменим квадратурной формулой прямоугольников, уравнения (2) — разностными уравнениями с помощью явной схемы Эйлера [59; 74; 89; 481; 635]. В результате придем к следующей дискретной задаче оптимального управления

$$I_0(x; [u_i]_0) = \sum_{i=0}^{N-1} F_i^0(x_i, u_i) + \Phi(x_N) \to \inf, \tag{5}$$

$$x_{i+1} = F_i(x_i, u_i), \quad i = 0, \ldots, N-1, \quad x_0 = x, \tag{6}$$

$$x_i \in G_i, \quad i = 0, \ldots, N, \tag{7}$$

$$[u_i]_0 = (u_0, u_1, \ldots, u_{N-1}): \; u_i \in V_i, \quad i = 0, \ldots, N-1, \tag{8}$$

где $F_i^0(x, u) = f^0(x, u, t_i)(t_{i+1} - t_i)$, $F_i(x, u) = x + f(x, u, t)(t_{i+1} - t_i)$, $G_i = G(t_i)$, $V_i = V(t_i)$. Заметим, что задача (5)–(8) имеет также и самостоятельный интерес и возникает при описании управляемых дискретных (импульсных) систем, в которые сигналы управления поступают в дискретные моменты времени, фазовые координаты также меняются дискретно [106; 202; 602; 750; 751].

Если задать какое-либо дискретное управление $[u_i]_0 = (u_0, u_1, \ldots, u_{N-1})$ и начальное условие $x_0 = x \in G_0$, то система (6) однозначно определяет соответствующую дискретную траекторию $[x_i]_0 = [x_i(x; [u_i]_0)]_0 = (x_0, x_1, \ldots, x_N)$. Зафиксируем некоторое $x \in G_0$ и через $\Delta_0(x)$ обозначим множество управлений $[u_i]_0$ таких, что: 1) выполнены условия (8); 2) дискретная траектория $[x_i]_0$, соответствующая управлению $[u_i]_0$ и выбранному начальному условию $x_0 = x$, удовлетворяет ограничениям (7). Пару $([u_i]_0, [x_i]_0)$, состоящую из управления и траектории, будем называть допустимой для задачи (5)–(8) или, короче, *допустимой парой*, если эта пара удовлетворяет всем условиям (6)–(8) или, иначе говоря, $[u_i]_0 \in \Delta_0(x_0)$.

Множество $\Delta_0(x_0)$ может быть пустым или непустым. Если $\Delta_0(x_0) = \varnothing$ при всех $x \in G_0$, то условия (6)–(8) несовместны и функция (5) будет определена на пустом множестве. Поэтому, чтобы задача (5)–(8) имела смысл, естественно требовать существование хотя бы одной точки $x \in G_0$, для которой $\Delta_0(x) \neq \varnothing$. Обозначим $X_0 = \{x: x \in G_0, \Delta_0(x) \neq \varnothing\}$. Тогда задача (5)–(8) может быть сформулирована совсем кратко: минимизировать функцию $I_0(x, [u_i]_0)$ при $[u_i]_0 \in \Delta_0(x)$, $x \in X_0$. Положим

$$I_0^* = \inf_{x \in X_0} \inf_{[u_i]_0 \in \Delta_0(x)} I_0(x, [u_i]_0).$$

Допустимую пару $([u_i^*]_0, [x_i^*]_0)$ назовем решением задачи (5)–(8), если $I_0(x_0^*, [u_i^*]_0) = I_0^*$, то $[u_i^*]_0$ назовем оптимальным управлением, $[x_i^*]_0$ — оптимальной траекторией задачи (5)–(8).

Как видим, задача (5)–(8) является уже известной нам задачей минимизации функции $n + Nr$ переменных $x, u_0, u_1, \ldots, u_{N-1}$ и для ее решения в принципе могут быть использованы методы, описанные в главах 1–3, 5. Однако в практических задачах число $n + Nr$ обычно бывает столь большим, что непосредственное использование методов глав 1–3, 5, вообще говоря, сильно осложняется: вызывает трудности также и то обстоятельство, что множества $\Delta_0(x)$, X_0, на которых минимизируется функция $I_0(x, [u_i]_0)$, заданы неявно. Для преодоления этих трудностей здесь часто пользуются методом динамического программирования, с помощью которого задачу (5)–(8) большого числа переменных удается свести к последовательности конечного числа задач минимизации функций меньшего числа переменных.

2. Для изложения метода динамического программирования нам понадобятся следующие вспомогательные задачи:

$$I_k(x, [u_i]_k) = \sum_{i=k}^{N-1} F_i^0(x_i, u_i) + \Phi(x_N) \to \inf, \tag{9}$$

$$x_{i+1} = F_i(x_i, u_i), \quad i = k, \ldots, N-1, \quad x_k = x, \tag{10}$$

$$x_i \in G_i, \quad i = k, \ldots, N, \tag{11}$$

$$[u_i]_k = (u_k, u_{k+1}, \ldots, u_{N-1}): \; u_i \in V_i, \quad i = k, \ldots, N-1, \tag{12}$$

где точка x и целое число k фиксированы, $x \in G_k$, $0 \leqslant k \leqslant N-1$. При $k = 0$ отсюда получим исходную задачу (5)–(8). Через $\Delta_k(x)$ обозначим множество всех управлений $[u_i]_k$, удовлетворяющих условиям (12) и таких,

что соответствующая ему траектория $[x_i]_k = (x_k = x, x_{k+1}, \ldots, x_N)$ из (10) удовлетворяет фазовым ограничениям (11). Пару $([u_i]_k, [x_i]_k)$ назовем допустимой парой для задачи (9)–(12), если $[u_i]_k \in \Delta_k(x)$. Допустимую пару $([u_i^*]_k, [x_i^*]_k)$ будем называть решением задачи (9)–(12), если $I_k(x, [u_i^*]_k) = I_k^*(x) = \inf\limits_{\Delta_k(x)} I_k(x, [u_i]_k)$; $[u_i^*]_k$ назовем оптимальным управлением, $[x_i^*]_k$ — оптимальной траекторией задачи (9)–(12).

Нетрудно видеть, что если $X_0 \neq \varnothing$, то $\Delta_k(x) \neq \varnothing$ хотя бы для одного $x \in G_k$. Введем функцию

$$B_k(x) = \inf_{\Delta_k(x)} I_k(x, [u_i]_k), \quad k = 0, \ldots, N-1,$$

называемую *функцией Беллмана* задачи (5)–(8). Область определения функции $B_k(x)$ представляет собой множество $X_k = \{x \in G_k\colon\ \Delta_k(x) \neq \varnothing\}$. Покажем, что функция Беллмана задачи (5)–(8) удовлетворяет некоторым рекуррентным соотношениям, называемым *уравнением Беллмана*.

Теорема 1. *Функция Беллмана задачи* (5)–(8) *необходимо является решением уравнения*

$$B_k(x) = \inf_{u \in D_k(x)} [F_k^0(x, u) + B_{k+1}(F_k(x, u))], \quad x \in X_k, \quad k = 0, \ldots, N-1, \tag{13}$$

где

$$B_N(x) = \Phi(x), \quad x \in G_N, \tag{14}$$

$D_k(x)$ *— множество всех тех* $u \in V_k$, *для которых существует хотя бы одно управление* $[u_i]_k \in \Delta_k(x)$ *с компонентой* $u_k = u$. *Верно и обратное: функция* $B_k(x)$, $x \in X_k$, $k = 0, \ldots, N-1$, *определяемая условиями* (13), (14), *является функцией Беллмана задачи* (5)–(8).

Доказательство проведем в предположении, что все упоминаемые в теореме 1 нижние грани конечны.

Необходимость. Пусть $B_k(x) = \inf\limits_{\Delta_k(x)} I_k(x, [u_i]_k)$, $x \in X_k$, $k = 0, \ldots, N-1$; $B_N(x) = \Phi(x)$, $x \in G_N$. Покажем, что эти функции удовлетворяют уравнению (13). Из определения множеств $\Delta_k(x)$, $D_k(x)$ видно, что множества $D_k(x)$ и $\Delta_k(x)$ оба пусты или непусты одновременно, и поскольку $x_{k+1} = F_k(x, u)$, то для непустоты этих множеств необходимо и достаточно, чтобы $\Delta_{k+1}(F_k(x, u)) \neq \varnothing$. Справедливость соотношения (13) при $k = N-1$ очевидным образом вытекает из условия $B_N(x) \equiv \Phi(x)$ и представления $I_{N-1}(x, [u_i]_{N-1}) \equiv F_{N-1}^0(x, u) + \Phi(F_{N-1}(x, u))$, верного для любого $u \in D_{N-1}(x) \equiv \Delta_{N-1}(x) \equiv \{u\colon u \in V_{N-1},\ x_N = F_{N-1}(x, u) \in G_N,\ x \in G_{N-1}\}$. Докажем (13) при k, $0 \leqslant k < N-1$. Для этого сначала убедимся в том, что

$$B_k(x) \leqslant \inf_{u \in D_k(x)} [F_k^0(x, u) + B_{k+1}(F_k(x, u))], \quad x \in X_k. \tag{15}$$

Возьмем произвольное $u \in D_k(x)$ и с этим управлением «выйдем» из точки x в момент k. В момент $k+1$ придем в точку $x_{k+1} = F_k(x, u)$, для которой $\Delta_{k+1}(x_{k+1}) \neq \varnothing$. По определению $B_{k+1}(x_{k+1}) = \inf\limits_{\Delta_{k+1}(x_{k+1})} I_{k+1}(x_{k+1}, [u_i]_{k+1})$ для любого $\varepsilon > 0$ найдется управление $[u_i^\varepsilon]_{k+1} \in \Delta_{k+1}(x_{k+1})$ такое, что $B_{k+1}(x_{k+1}) \leqslant I_{k+1}(x_{k+1}, [u_i^\varepsilon]_{k+1}) \leqslant B_{k+1}(x_{k+1}) + \varepsilon$. Поскольку $[\overline{u}_i]_k = (u, u_{k+1}^\varepsilon, \ldots, u_{N-1}^\varepsilon) \in \Delta_k(x)$, то $B_k(x) \leqslant I_k(x, [\overline{u}_i]_k) = F_k^0(x, u) + I_{k+1}(x_{k+1}, [u_i^\varepsilon]_{k+1}) \leqslant F_k^0(x, u) + B_{k+1}(x_{k+1}) + \varepsilon \equiv F_k^0(x, u) + B_{k+1}(F_k(x, u)) + \varepsilon$. В силу произвольности $u \in D_k(x)$ и величины $\varepsilon > 0$ отсюда следует неравенство (15).

Теперь покажем, что в (15) на самом деле знак неравенства можно заменить знаком равенства. По определению $\inf\limits_{\Delta_k(x)} I_k(x, [u_i]_k) = B_k(x)$ для каждого $\varepsilon > 0$ найдется такое управление $[v_i^\varepsilon]_k \in \Delta_k(x)$, что $B_k(x) \leqslant I_k(x, [v_i^\varepsilon]_k) \leqslant B_k(x) + \varepsilon$. Но $\overline{[v_i]}_{k+1} \equiv (v_{k+1}^\varepsilon, \ldots, v_{N-1}^\varepsilon) \in \Delta_{k+1}(F_k(x, v_k^\varepsilon))$, поэтому

$$F_k^0(x, v_k^\varepsilon) + B_{k+1}(F_k(x, v_k^\varepsilon)) \leqslant F_k^0(x, v_k^\varepsilon) + I_{k+1}(F_k(x, v_k^\varepsilon), \overline{[v_i]}_{k+1}) = I_k(x, [v_i^\varepsilon]_k) \leqslant B_k(x) + \varepsilon.$$

Так как $v_k^\varepsilon \in D_k(x)$, то отсюда имеем: $\inf\limits_{u \in D_k(x)} [F_k^0(x, u) + B_{k+1}(F_k(x, u))] \leqslant B_k(x) + \varepsilon$, или в силу произвольности $\varepsilon > 0$:

$$\inf_{u \in D_k(x)} [F_k^0(x, u) + B_{k+1}(F_k(x, u))] \leqslant B_k(x), \quad x \in X_k.$$

Отсюда и из (15) немедленно следует равенство (13).

Достаточность. Пусть функции $B_k(x)$, $x \in X_k$, $k = 0, \ldots, N-1$, определены из условий (13), (14). Спрашивается: какое отношение имеют эти функции к задаче (5)–(8)? Покажем, что при каждом $x \in X_k$ величина $B_k(x)$ равна нижней грани функции (9) при условиях (10)–(12). Отметим, что условия (13), (14) однозначно определяют функции $B_k(x)$, $x \in X_k$, $k = 0, \ldots, N$. Это легко доказывается с помощью индукции последовательным перебором в (13), (14) номеров $k = N, N-1, \ldots, 0$ с учетом того, что функции $\Phi(x)$, $F_k^0(x, u)$, $F_k(x, u)$ однозначны и нижняя грань функций определяется также однозначно. С другой стороны, как было установлено выше, функции $\inf\limits_{\Delta_k(x)} I_k(x, [u_i]_k)$, $x \in X_k$, также являются решением уравнения (13) при условии (14). Из единственности решения системы (13), (14) тогда следует, что $B_k(x) = \inf\limits_{\Delta_k(x)} I_k(x, [u_i]_k)$, $x \in X_k$, $k = 0, \ldots, N-1$. Теорема 1 доказана. □

3. Пользуясь условиями (13), (14), можно последовательно определить функции $B_k(x)$ и их области определения X_k, $k = N, N-1, \ldots, 1, 0$. А именно $B_N(x) \equiv \Phi(x)$, $x \in G_N \equiv X_N$ — известны. Если известны $B_{k+1}(x)$ и X_{k+1}, $k \leqslant N-1$, то для определения $B_k(x)$ нужно решить задачу минимизации функции $\varphi(x, u) \equiv F_k^0(x, u) + B_{k+1}(F_k(x, u))$ переменных $u = (u^1, \ldots, u^r)$ на известном множестве $D_k(x) = \{u\colon u \in V_k,\ F_k(x, u) \in X_{k+1}\}$. Здесь могут быть использованы методы глав 1–3, 5. Очевидно, функция $B_k(x)$ определена в точке x тогда и только тогда, когда $D_k(x) \neq \varnothing$. Таким образом, при определении значений функции $B_k(x)$ одновременно находится и область ее определения $X_k = \{x\colon x \in G_k,\ D_k(x) \neq \varnothing\} \equiv \{x\colon x \in G_k,\ \Delta_k(x) \neq \varnothing\}$. Так как $\Delta_k(x) \neq \varnothing$ хотя бы при одном $x \in G_k$, то $X_k \neq \varnothing$, $k = N, N-1, \ldots, 1, 0$.

Предположим, что нам удалось найти функции $B_k(x)$ из условий (13), (14) и, кроме того, пусть также известны функции $u_k(x) \in D_k(x)$, $x \in X_k$, $k = 0, \ldots, N-1$, на которых достигается нижняя грань в правой части (13). Тогда, оказывается, решения задач (5)–(8) и (9)–(12) выписываются совсем просто. А именно, оптимальное управление $[u_i^*]_0$ и соответствующая траектория $[x_i^*]_0$ для задачи (5)–(8) определяются следующим образом: сначала из условия

$$\inf_{x \in X_0} B_0(x) = B_0(x_0^*) \tag{16}$$

находят $x_0^* \in X_0$, затем последовательно полагают

$$\begin{aligned} &u_0^* = u_0(x_0^*), \quad x_1^* = F_0(x_0^*, u_0^*), \quad u_1^* = u_1(x_1^*), \\ &x_2^* = F_1(x_1^*, u_1^*), \ \ldots, \ x_N^* = F_{N-1}(x_{N-1}^*, u_{N-1}^*). \end{aligned} \tag{17}$$

Оптимальное управление $[u_i^*]_k$ и траектория $[x_i^*]_k$ для задачи (9)–(12) определяются аналогично:

$$x_k^* = x, \quad u_k^* = u_k(x_k^*), \\ x_{k+1}^* = F_k(x_k^*, u_k^*), \ \ldots, \ x_N^* = F_{N-1}(x_{N-1}^*, u_{N-1}^*). \tag{18}$$

Для доказательства этих утверждений введем вспомогательные функции:

$$R_i(x, u) \equiv B_{i+1}(F_i(x, u)) - B_i(x) + F_i^0(x, u), \quad i = 0, \ldots, N-1. \tag{19}$$

Очевидно, уравнения Беллмана (13) тогда можно переписать в эквивалентном виде:

$$\inf_{u \in D_k(x)} R_k(x, u) \equiv R_k(x, u_k(x)) = 0, \quad i = 0, \ldots, N-1. \tag{20}$$

Кроме того, с помощью функций $R_i(x, u)$ значение функции (9) на любом управлении $[u_i]_k \in \Delta_k(x)$ и $x \in X_k$ можно выразить формулой

$$I_k(x, [u_i]_k) = \sum_{i=k}^{N-1} R_i(x_i, u_i) + B_k(x) \tag{21}$$

при всех $k = 0, \ldots, N-1$. В самом деле, учитывая равенство $B_N(x) \equiv \Phi(x)$, из (10), (19) имеем $\sum_{i=k}^{N-1} R_i(x_i, u_i) = \sum_{i=k}^{N-1} [B_{i+1}(x_{i+1}) - B_i(x_i) + F_i^0(x_i, u_i)] = B_N(x_N) - B_k(x) + \sum_{i=k}^{N-1} F_i^0(x_i, u_i) = I_k(x, [u_i]_k) - B_k(x)$, что равносильно (21).

Теорема 2. *Пусть найдены функции $B_k(x)$ из (13), (14) и их области определения X_k, а также функции $u = u_k(x)$ $x \in X_k$, $k = 0, \ldots, N-1$, на которых достигается нижняя грань в уравнении (13) или (20), и пусть x_0^* определена условием (16). Тогда оптимальное управление $[u_i^*]_0$ и траектория $[x_i^*]_0$ для задачи (5)–(8) определяются соотношениями (16), (17).*

Доказательство. Из определения $u(x)$, $[u_i^*]_0$, $[x_i^*]_0$ и эквивалентности записей уравнения Беллмана (1.3) и (20) имеем

$$R_i(x_i^*, u_i^*) \equiv R_i(x_i^*, u_i(x_i^*)) = \inf_{u \in D_i(x_i^*)} R_i(x_i^*, u) = 0, \quad i = 0, \ldots, N-1. \tag{22}$$

Возьмем произвольные $x \in X_0$, управление $[u_i]_0 \in \Delta_0(x)$ с соответствующей траекторией $[x_i]_0$ из (6). Так как $u_i \in D_i(x_i)$, то из уравнения (20) и определения $u_i(x)$ следует

$$R_i(x_i, u_i) \geqslant \inf_{u \in D_i(x_i)} R_i(x_i, u) = R_i(x_i, u_i(x_i)) = 0, \quad i = 0, \ldots, N-1. \tag{23}$$

С помощью формулы (21) при $k = 0$ с учетом соотношений (16), (22), (23) получаем $I_0(x, [u_i]_0) - I_0(x_0^*, [u_i^*]_0) = \sum_{i=0}^{N-1} [R_i(x_i, u_i) - R_i(x_i^*, u_i^*)] + B_0(x) - B_0(x_0^*) \geqslant 0$ для любых $x \in X_0$ и $[u_i]_0 \in \Delta_0(x)$, что и требовалось. □

Теорема 3. *Пусть известны $B_k(x)$, $x \in X_k$ из (13), (14), а также функции $u_k(x)$, на которых достигается нижняя грань в уравнении (13) (или (20)). Тогда оптимальное управление $[u_i^*]_k$ и траектория $[x_i^*]_k$ для задачи (9)–(12) определяются формулами (18).*

Доказательство. Возьмем произвольное управление $[u_i]_k \in \Delta_k(x)$ и соответствующую траекторию $[x_i]_k$ из (10). Очевидно, соотношения (22), (23) остаются справедливыми и здесь при всех $i = k, \ldots, N-1$. Отсюда с помощью (21) получим

$$I_k(x, [u_i]_k) - I_k(x, [u_i^*]_k) = \sum_{i=k}^{N-1} [R_i(x_i, u_i) - R_i(x_i^*, u_i^*)] \geqslant 0,$$

что и требовалось. □

4. В теории оптимального управления и ее приложениях важное место занимает так называемая *проблема синтеза*, заключающаяся в построении функции $u = u_k(x)$, выражающей собой оптимальное управление при условии, что в момент k объект находится в точке x фазового пространства. Такая функция $u_k(x)$ называется *синтезирующей*.

Теорема 3 показывает, что решение уравнения Беллмана (13) равносильно решению проблемы синтеза для задачи (5)–(8). А именно, функция $u_k(x)$, на которой достигается нижняя грань в (13), является синтезирующей: если в момент k объект находится в точке $x \in X_k$, то дальнейшее оптимальное движение объекта определяется условиями: $x_{i+1} = F_i(x_i, u_i(x_i))$, $i = k, \ldots, N-1$, $x_k = x$ (если $x \notin X_k$, то $\Delta_k(x) = \varnothing$ — движение с соблюдением условий (10)–(12) невозможно). Достаточные условия существования функции Беллмана и синтезирующей функции для задачи (5)–(8) даются в следующей теореме.

Теорема 4. *Пусть множества G_k, $k = 0, \ldots, N$, замкнуты, множества V_k, $k = 0, \ldots, N-1$, замкнуты и ограничены, функция $F_k^0(x, u)$ полунепрерывна снизу, а функция $F_k(x, u)$ непрерывна по совокупности аргументов (x, u) при $x \in G_k$, $u \in V_k$, $k = 0, \ldots, N-1$, $\Phi(x)$ полунепрерывна снизу на множестве G_N. Тогда:* 1) *множества X_k, $k = 0, \ldots, N$, замкнуты, множества $D_k(x)$, $k = 0, \ldots, N-1$, замкнуты и ограничены равномерно по $x \in X_k$;* 2) *нижняя грань в правой части (13) достигается хотя бы при одном $u = u_k(x) \in D_k(x)$;* 3) *функция $B_k(x)$ полунепрерывна снизу на X_k, $k = 0, \ldots, N$.*

Доказательство. По условию $G_N \equiv X_N$ замкнуто, $\Phi(x) \equiv B_N(x)$ полунепрерывна снизу на X_N. Сделаем индуктивное предположение: пусть X_{k+1} замкнуто, $B_{k+1}(x)$ полунепрерывна снизу на X_{k+1} при некотором k, $0 \leqslant k \leqslant N-1$. Докажем, что тогда X_k замкнуто и на X_k справедливы все утверждения теоремы. Так как $D_k(x) = \{u: \ u \in V_k, \ F_k(x, u) \in X_{k+1}\} \subseteq V_k$ и V_k ограничено, то $D_k(x)$ ограничено равномерно по $x \in X_k$. Докажем замкнутость $D_k(x)$ при любом фиксированном $x \in X_k$. Пусть $v_m \in D_k(x)$, $m = 1, 2\ldots$, $v_m \to v$, при $m \to \infty$. Это значит, что $v_m \in V_k$, $F_k(x, v_m) \in X_{k+1}$, $m = 1, 2, \ldots$ Из замкнутости V_k, X_{k+1} и непрерывности $F_k(x, v)$ сразу имеем: $v \in V_k$, $\lim_{m \to \infty} F_k(x, v_m) = F_k(x, v) \in X_{k+1}$, т. е. $v \in D_k(x)$. Замкнутость $D_k(x)$ доказана.

Покажем замкнутость $X_k = \{x: x \in G_k, \ D_k(x) \neq \varnothing\}$. Пусть $y_m \in X_k$, $m = 1, 2, \ldots$, $y_m \to y$ при $m \to \infty$. Из замкнутости G_k следует $y \in G_k$. Если мы еще покажем, что $D_k(y) \neq \varnothing$, то это будет означать, что $y \in X_k$, и замкнутость X_k будет доказана. Так как $D_k(y_m) \neq \varnothing$, то существует такое $v_m \in V_k$, что $F_k(y_m, v_m) \in X_{k+1}$, $m = 1, 2, \ldots$ В силу компактности V_m из последовательности $\{v_m\}$ можно выбрать подпоследовательность $\{v_{m_p}\} \to v \in V_k$ при $p \to \infty$. Поскольку X_{k+1} замкнуто, $F_k(x, u)$ непрерывна, то $\lim_{p \to \infty} F_k(y_{m_p}, v_{m_p}) = F_k(y, v) \in X_{k+1}$, т. е. $v \in D_k(y)$. Таким образом, $D_k(y) \neq \varnothing$.

Далее, функция $\varphi(x, u) \equiv F_k^0(x, u) + B_{k+1}(F_k(x, u))$ полунепрерывна снизу по (x, u) при $x \in X_k$, $u \in D_k(x)$ — это следует из непрерывности $F_k(x, u)$ и полунепрерывности снизу $F_k^0(x, u)$, $B_{k+1}(x)$. Поскольку $D_k(x)$ — замкнутое ограниченное множество, то в силу теоремы 2.1.1 $\varphi(x, u)$ при каждом фиксированном $x \in X_k$ достигает своей нижней грани на $D_k(x)$ хотя бы в одной точке $u = u_k(x) \in D_k(x)$. Таким образом, $B_k(x) = \inf\limits_{u \in D_k(x)} \varphi(x, u) = \varphi(x, u_k(x))$, в силу уравнения Беллмана (13).

Остается еще доказать полунепрерывность снизу $B_k(x)$ на X_k. Пусть $x, y_m \in X_k$, $y_m \to x$ при $m \to \infty$, $B_k(y_m) = \varphi(y_m, u_k(y_m))$. Так как $u_k(y_m) \in D_k(y_m) \in V_k$, то в силу компактности V_k последовательность $\{u_k(y_m),\ m = 1, 2, \ldots\}$ имеет хотя бы одну предельную точку $v \in V_k$. Можем считать, что сама последовательность $\{u_k(y_m)\} \to v$ при $m \to \infty$. Поскольку $F_k(x, u)$ непрерывна, X_{k+1} замкнуто, кроме того, $F_k(y_m, v_k(y_m)) \in X_{k+1}$, то $\lim\limits_{m \to \infty} F_k(y_m, v_k(y_m)) = F_k(y, v) \in X_{k+1}$. Это значит, что $v \in D_k(x)$. Тогда

$$\lim_{m \to \infty} B_k(y_m) = \lim_{m \to \infty} \varphi(y_m, u_k(y_m)) \geqslant \varphi(x, v) \geqslant \inf_{u \in D_k(x)} \varphi(x, u) = B_k(x).$$

Полунепрерывность $B_k(x)$ на X_k доказана, что и требовалось. □

5. Нетрудно привести примеры задач типа (5)–(8), когда нижняя грань в (13) или (16) не достигается (см. ниже упражнение 2). В таких задачах, конечно, приходится пользоваться величинами, лишь приближенно реализующими нижнюю грань в (13), (16). Но даже в том случае, когда нижняя грань в (13), (16) достигается, получить точные выражения для функций $B_k(x)$, $u_k(x)$ и точки x_0^* из (13), (16) часто бывает затруднительно. Поэтому на практике часто пользуются соотношениями (16), (17) для приближенных $B_k(x)$, $u_k(x)$ и вместо точных управлений $[u_i^*]_0$ и траектории $[x_i^*]_0$ получают какие-то их приближения. Возникает вопрос, насколько отличается полученное таким образом приближенное решение задачи (5)–(8) от ее точного решения? Приводимая ниже оценка погрешности дает некоторый ответ на этот вопрос.

Пусть $K_i(x)$ — приближенное значение функции Беллмана $B_i(x)$, $i = 0, \ldots, N$. По аналогии с (19) введем функцию

$$S_i(x, u) \equiv K_{i+1}(F_i(x, u)) - K_i(x) + F_i^0(x, u), \quad i = 0, \ldots, N-1, \tag{24}$$

и, кроме того, положим

$$s_N(x) = \Phi(x) - K_N(x), \quad x \in G_N. \tag{25}$$

Возьмем произвольную допустимую пару $[u_i]_0 = (u_0, u_1, \ldots, u_{N-1})$, $[x_i]_0 = (x_0, x_1, \ldots, x_{N-1})$ задачи (5)–(8). Тогда $x_0 \in X_0$, $[u_i]_0 \in \Delta_0(x_0)$, $u_i \in D_i(x_i)$, $i = 0, \ldots, N-1$. Учитывая условие (6), из (24) имеем

$$S_i(x_i, u_i) = K_{i+1}(x_{i+1}) - K_i(x_i) + F_i^0(x_i, u_i), \quad i = 0, \ldots, N-1.$$

Суммируя эти равенства по i от 0 до $N-1$, с помощью (25) получим формулу

$$I_0(x_0, [u_i]_0) = \sum_{i=0}^{N-1} S_i(x_i, u_i) + s_N(x_N) + K_0(x_0). \tag{26}$$

Если $K_i(x) = B_i(x)$, то $s_N(x) \equiv 0$, и формула (26) превратится в знакомую нам формулу (21) при $k = 0$.

Предположим, что каким-либо образом нам удалось найти некоторое управление $[\bar{u}_i]_0$ и соответствующую ему траекторию $[\bar{x}_i]_0$, удовлетворяющую условиям (6)–(8), т. е. $\bar{x}_0 \in X_0$, $[\bar{u}_i]_0 \in \Delta_0(\bar{x}_0)$, $\bar{u}_i \in D_i(\bar{x}_i)$, $i = 0, \ldots, N-1$. Согласно (26) тогда

$$I_0(\bar{x}_0, [\bar{u}_i]_0) = \sum_{i=0}^{N-1} S_i(\bar{x}_i, \bar{u}_i) + s_N(\bar{x}_N) + K_0(\bar{x}_0).$$

Отсюда и из (26) имеем

$$\begin{aligned} -I_0(\bar{x}_0, [\bar{u}_i]_0) + I_0(x_0, [u_i]_0) = \\ = \sum_{i=0}^{N-1} [-S_i(\bar{x}_i, \bar{u}_i) + S_i(x_i, u_i)] - s_N(\bar{x}_N) + s_N(x_N) - K_0(\bar{x}_0) + K_0(x_0) \end{aligned} \tag{27}$$

для любых $x_0 \in X_0$, $[u_i]_0 \in \Delta_0(x_0)$. Учитывая, что $\bar{x}_0, x_0 \in X_0$, $[\bar{u}_i]_0 \in \Delta_0(\bar{x}_0)$, $[u_i]_0 \in \Delta_0(x_0)$, $(x_i, u_i) \in X_i \times D_i(x_i)$, перейдем к нижней грани по $(x_0, [u_i]_0)$ сначала в правой части (27), а затем в левой части (27). Тогда получим неравенства

$$\begin{aligned} 0 \leqslant I_0(\bar{x}_0, [\bar{u}_i]_0) - I_0^* \leqslant \sum_{i=0}^{N-1} \Big[S_i(\bar{x}_i, \bar{u}_i) - \inf_{x \in X_i} \inf_{u \in D_i(x)} S_i(x, u) \Big] + \\ + s_N(\bar{x}_N) - \inf_{X_N} s_N(x) + K_0(\bar{x}_0) - \inf_{X_0} K_0(x), \end{aligned} \tag{28}$$

представляющие собой оценку погрешности, которая будет допущена, если $[\bar{u}_i]_0, [\bar{x}_i]$ будут взяты в качестве приближенного решения задачи (5)–(8).

Если $K_i(x) = B_i(x)$, $i = 0, \ldots, N$, то $S_i(x, u) = R_i(x, u)$, $s_N(x) = 0$. Кроме того, из (20) следует, что $\inf\limits_{u \in D_i(x)} R_i(x, u) = 0$ для всех $x \in X_i$, так что $\inf\limits_{x \in X_i} \inf\limits_{u \in D_i(x)} R_i(x, u) = 0$. Поэтому при $K_i(x) = B_i(x)$ из (28) получим

$$0 \leqslant I_0(\bar{x}_0, [\bar{u}_i]_0) - I_0^* \leqslant \sum_{i=0}^{N-1} R_i(\bar{x}_i, \bar{u}_i) + B_0(\bar{x}_0) - \inf_{X_0} B_0(x). \tag{29}$$

Из оценки (29) следует, что если определить $\bar{u}_i(x) \in \Delta_i(x)$, $\bar{x}_0 \in X_0$ так, чтобы $R_i(x, \bar{u}_i(x))$ было поближе к $\inf\limits_{u \in D_i(x)} R_i(x, u) = 0$, $x \in X_i$, а $B_0(\bar{x}_0)$ — поближе к $\inf\limits_{X_0} B_0(x)$, и затем строить управление $[\bar{u}_i]$ и траекторию $[\bar{x}_i]$ следующим образом:

$$\bar{u}_0 = \bar{u}_0(\bar{x}_0), \quad \bar{x}_1 = F_0(\bar{x}_0, \bar{u}_0), \quad \bar{u}_1 = \bar{u}_1(x_1), \ldots, \quad x_N = F_{N-1}(\bar{x}_{N-1}, \bar{u}_{N-1}),$$

то и величина $I_0(\bar{x}_0, [\bar{u}_i]_0) - I_0^*$ будет небольшой.

Заметим, что поскольку на практике конструктивное описание множеств X_i, $D_i(x)$ часто отсутствует, то в правой части оценки (28) вместо X_i, $D_i(x)$ часто берут G_i и V_i соответственно — такая замена, очевидно, может привести лишь к увеличению правой части (28). В результате получим следующую апостериорную оценку погрешности

$$\begin{aligned} 0 \leqslant I_0(\bar{x}_0, [\bar{u}_i]_0) - I_0^* \leqslant \sum_{i=0}^{N-1} \Big[S_i(\bar{x}_i, \bar{u}_i) - \inf_{x \in G_i} \inf_{u \in V_i(x)} S(x, u) \Big] + \\ + s_N(\bar{x}_N) - \inf_{G_N} s_N(x) + K_0(\bar{x}_0) - \inf_{G_0} K_0(x). \end{aligned} \tag{30}$$

Конечно, при пользовании оценкой (30) надо помнить, что если правые части оценок (28), (30) отличаются намного, то оценка (30) может оказаться слишком грубой.

6. Оценка (28) полезна также и тем, что она указывает пути получения достаточных условий оптимальности для задачи (5)–(8).

Теорема 5. *Для того чтобы управление* $[\bar{u}_i]_0$ *и траектория* $[\bar{x}_i]_0$, *удовлетворяющие условиям* (6)–(8), *были решением задачи* (5)–(8), *достаточно, чтобы существовала функция* $K_i(x)$, $i=0,\ldots,N$ *такая, что*

$$S_i(\bar{x}_i,\bar{u}_i)=\inf_{x\in X_i}\inf_{u\in D_i(x)}S_i(x,u)=S_{i\min},\quad i=0,\ldots,N-1, \tag{31}$$

$$s_N(\bar{x}_N)=\inf_{X_N}s_N(x)=s_{N\min},\quad K_0(\bar{x}_0)=\inf_{X_0}K_0(x)=K_{0\min}, \tag{32}$$

где функции $S_i(x,u)$, $s_N(x)$ *определяются формулами* (24), (25).

Доказательство следует из того, что при выполнении условий (31), (32) правая часть оценки (28) обращается в нуль, и $I_0(\bar{x}_0,[\bar{u}_i]_0)=I_0^*$. □

С помощью оценки (28) нетрудно также получить условия, достаточные для того, чтобы та или иная последовательность допустимых пар управлений и траекторий была минимизирующей для задачи (5)–(8).

Теорема 6. *Пусть последовательность управлений* $[u_{im}]_0$ *и траекторий* $[x_{im}]_0$, $m=1,2,\ldots$, *удовлетворяет условиям* (6)–(8). *Для того чтобы*

$$\lim_{m\to\infty}I_0(x_{0m},[u_{im}]_0)=I_0^*, \tag{33}$$

достаточно существования функции $K_i(x)$, $i=0,\ldots,N$, *такой, что*

$$\lim_{m\to\infty}S_i(\bar{x}_{im},\bar{u}_{im})=S_{i\min},\quad i=0,1,\ldots,N-1, \tag{34}$$

$$\lim_{m\to\infty}s_N(x_{Nm})=s_{N\min},\quad \lim_{m\to\infty}K_0(x_{0m})=K_{0\min}, \tag{35}$$

где $S_{i\min}$, $s_{N\min}$, $K_{0\min}$ *определены в* (31), (32).

Доказательство. В оценку (28) вместо $[\bar{u}_i]_0$, $[\bar{x}_i]_0$ поставим соответственно $[u_{im}]_0$, $[x_{im}]_0$ и перейдем к пределу при $m\to\infty$. С учетом условий (34), (35) получим равенство (33). □

Утверждения, аналогичные теоремам 5, 6, доказаны в [253; 417; 418] для более общих задач, чем задача (5)–(8).

Всякую функцию $K_i(x)$, $i=0,\ldots,N$, удовлетворяющую условиям теоремы 5 (теоремы 6), назовем *функцией Кротова* задачи (5)–(8), соответствующей допустимой паре $[u_i]_0$, $[x_i]_0$ [или последовательности допустимых пар $[u_{im}]_0$, $[x_{im}]_0$, $m=1,2,\ldots$].

Заметим, что если существует хотя бы одна функция Кротова $K_i(x)$, $i=0,\ldots,N$, то функция $K_i(x)+\alpha_i$, $i=0,\ldots,N$ при любых α_i также является функцией Кротова. Поэтому без ограничения общности в теоремах 5, 6 можно принять $S_{i\min}=0$, $i=0,\ldots,N-1$, $s_{N\min}=0$, ибо в противном случае функцию $K_i(x)$ заменим новой функцией $K_i(x)+\alpha_i$, где

$$\alpha_i=S_{i\min}+S_{i+1,\min}+\ldots+S_{N-1,\min},\quad i=1,\ldots,N-1,\quad \alpha_N=s_{N\min}.$$

Таким образом, функция Кротова для допустимой пары $([\bar{u}_i]_0,[\bar{x}_i]_0)$ или последовательности $([u_{im}]_0,[x_{im}]_0)$, $m=1,2,\ldots$, согласно теоремам 5, 6 удовлетворяет условиям

$$S_i(x,u)\equiv K_{i+1}(F_i(x,u))-K_i(x)+F_i^0(x,u)\geqslant 0,$$
$$u\in D_i(x),\quad x\in X_i,\quad i=0,\ldots,N-1, \tag{36}$$

$$s_N(x)=-K_N(x)+\Phi(x)\geqslant 0,\quad x\in X_N=G_N,\quad K_0(x)\geqslant K_{0\min},\quad x\in X_0, \tag{37}$$

причем неравенства (36), (37) должны обратиться в равенства при $u=\bar{u}_i$, $x=\bar{x}_i$ или при $u=u_{im}$, $x=x_{im}$ в пределе при $m\to\infty$, $i=0,\ldots,N$.

Сравнение соотношений (36), (37) с (13), (14), (16), (20) показывает, что функция Беллмана всегда является функцией Кротова, а обратное, вообще говоря, неверно. Заметим также, что с помощью функции Кротова удается установить оптимальность допустимых пар, не решая проблемы синтеза, так как согласно условиям (36), (37) функция $K_i(x)$ выбираются с учетом индивидуальных свойств конкретной допустимой пары управления и траектории (или последовательности пар), подозрительных на оптимальность.

7. Остановимся на одном специальном классе задач минимизации функций большого числа переменных, которые с помощью метода динамического программирования могут быть сведены к последовательности задач минимизации функций меньшего числа переменных. А именно, пусть требуется минимизировать функцию

$$I_0([u_i]_0)=I_0(u_0,u_1,\ldots,u_{N-1})=\sum_{i=0}^{N-1}f_i^0(u_i) \tag{38}$$

при условиях

$$u_i\in V_i,\quad i=0,\ldots,N-1, \tag{39}$$

$$\sum_{i=0}^{N-1}g_i^j(u_i)\leqslant b^j,\quad j=1,\ldots,p,\quad \sum_{i=0}^{N-1}g_i^j(u_i)=b^j,\quad j=p+1,\ldots,n, \tag{40}$$

где V_i — заданные множества из E^r; $f_i^0(u)$, $g_i^j(u)$, $u\in V_i$ — заданные функции, b^j — заданные числа.

Задачу (38)–(40), оказывается, нетрудно записать в виде задачи (5)–(8). В самом деле, введем переменные x_i, $i=0,\ldots,N$, как решение системы

$$x_{k+1}=x_k+g_k(u_k),\quad k=0,\ldots,N-1,\quad x_0=0, \tag{41}$$

где $g_k(u)=(g_k^1(u),\ldots,g_k^n(u))$, $x_k=(x_k^1,\ldots,x_k^n)$, $k=0,\ldots,N$. Так как из (41) следует, что $x_N=\sum_{k=0}^{N-1}g_k(u_k)$, то ясно, что ограничения (40) равносильны условию

$$x_N\in G_N=\{x\colon x\in E^n,\ x^j\leqslant b^j,\ j=1,\ldots,p;\ x^j=b^j,\ j=p+1,\ldots,n\}. \tag{42}$$

Таким образом, задача (38)–(40) эквивалентна задаче минимизации функции (38) при условиях (39), (41), (42) и является частным случаем задачи (5)–(8) при $F_i^0(x,u)=f_i^0(u)$, $F_i(x,u)=x+g_i(u)$, $\Phi(x)\equiv 0$, $G_i=E^n$, $i=1,\ldots,N-1$; $G_0=\{0\}$, G_N определено соотношением (42). Это значит, что для исследования задачи (38)–(40) может быть применен метод динамического программирования, изложенный выше. Пользуясь введенными ранее обозначениями, можем переписать уравнение Беллмана (13), (14) применительно к задаче (38)–(40):

$$B_k(x)=\inf_{u\in D_k(x)}[f_k^0(u)+B_{k+1}(x+g_i(u))],$$
$$x\in X_k,\quad k=0,\ldots,N-1,\quad B_N(x)=0. \tag{43}$$

В том случае, когда ограничения (40) и, следовательно, (42) отсутствуют, то $G_N=E^n$ и в (43) можно положить $D_k(x)=V_k$, $k=0,\ldots,N-1$. Уравнениями (43) можно пользоваться для решения задачи (38)–(40), как это показано выше в п. 3. Предлагаем читателю в качестве упражнения переформулировать теоремы 1–6 применительно к задаче (38)–(40).

Подчеркнем, что в задаче (38)–(40) функция и ограничения имеют весьма специальный вид — это обстоятельство было весьма существенно для применения метода динамического программирования. Этот метод применим и к несколько более общим, чем (38)–(40), задачам — об этом см. подробнее, например, в [202].

8. Изложенный выше метод динамического программирования является достаточно эффективным средством решения задач вида (5)–(8) или (38)–(40) — с его помощью исходная задача сводится к последовательности вспомогательных и, вообще говоря, более простых задач минимизации функций меньшего числа переменных для определения $B_k(x)$, $u_k(x)$ (см. условия (13), (14) или (43)). Если эти вспомогательные задачи решены с достаточно хорошей точностью, то тем самым и в исходной задаче глобальный минимум функции будет найден с высокой точностью. Далее, метод динамического программирования позволяет решить важную в приложениях проблему синтеза. Как показано в п. 6, с помощью этого метода и его обобщений могут быть получены достаточные условия оптимальности для дискретных управляемых систем. Кроме того, этот метод дает значительный выигрыш в объеме вычислений по сравнению с простым перебором всевозможных допустимых управлений и траекторий, поскольку при определении $B_k(x)$, $u_k(x)$ рассматриваются лишь такие управления, которые переводят точку $x \in G_k$ в точку $x_{k+1} = F(x, u) \in G_{k+1}$, а дальнейшее движение из точки x_{k+1} осуществляется по оптимальной траектории, при этом неоптимальные траектории вовсе не рассматриваются. Указанные достоинства метода, динамического программирования, простота схемы, применимость к задачам оптимального управления с фазовыми ограничениями делают этот метод весьма привлекательным, и его широко используют при решении задач типа (5)–(8) или (38)–(40). Что касается задачи (1)–(4), с которой мы начали изложение, то можно показать, что при некоторых ограничениях решение дискретной задачи (5)–(8) при $\lim\limits_{N\to\infty} \max\limits_{0\leqslant i\leqslant N-1} (t_{i+1} - t_i) = 0$ будет приближаться в некотором смысле к решению задачи (1)–(4).

Заметим, что поскольку аналитическое выражение для $B_k(x)$, $u_k(x)$ при всех $x \in X_k$ в общем случае найти не удается, то на практике приходится ограничиваться приближенным вычислением $B_k(x)$, $u_k(x)$ в некоторых заранее выбранных узловых точках множества X_k. Однако согласно (13) при вычислении $B_k(x)$ нужно знать значение $B_{k+1}(F_k(x, u))$ при некоторых u, и здесь вполне возможны случаи, когда точка $x_{k+1} = F_k(x, u)$ не будет принадлежать заранее выбранному множеству узловых точек из X_{k+1} и нужное значение $B_{k+1}(x_{k+1})$ еще не будет вычислено. Если же мы захотим вычислить недостающее значение $B_{k+1}(x_{k+1})$, то здесь могут понадобиться значения ранее вычисленных функций $B_{k+2}(x), \ldots, B_N(x)$ в новых дополнительных точках, а для этого в свою очередь придется еще более расширить множества узловых точек в X_{k+2}, $X_{k+2}, \ldots$ и т. д. На практике в таких случаях недостающее значение $B_{k+1}(x)$ получают с помощью интерполяции по значениям $B_{k+1}(x)$ в близлежащих узловых точках, что, вообще говоря, снижает точность. Заметим также, что принятый выше способ аппроксимации задачи (1)–(4) с помощью разностной задачи (5)–(8) довольно груб, поскольку опирается на простейший метод ломаных Эйлера для интегрирования дифференциальных уравнений и квадратурную формулу прямоугольников. В следующем параграфе будет описана схема Моисеева, которая не требует интерполяции и оставляет достаточную свободу при выборе способа аппроксимации задачи (1)–(4).

В заключение мы отметим, что метод динамического программирования относится к классу *методов декомпозиции* — так называются методы минимизации, которые позволяют задачи большой размерности свести к задачам меньшей размерности; о методах декомпозиции см., например, в [222; 470; 746; 747; 759].

Упражнения

1. Найти функцию Беллмана для задачи

$$I_0([u_i]_0) = \sum_{i=0}^{N-1} [\langle a_i, x_i\rangle + b_i(u_i)] + \langle c, x_N\rangle \to \inf,$$

$$x_{i+1} = A_i x_i + C_i(u_i), \quad u_i \in V_i, \quad i = 0, \ldots, N-1, \quad x_0 = a,$$

где A_i — матрица порядка $n \times n$; $C_i(u) = (C_i^1(u), \ldots, C_i^n(u))$, C_i^j, $b_j(u)$ — функции переменной $u \in V_i \subseteq E^r$, a_i, c, a — n-мерные векторы, $i = 0, \ldots, N-1$. Указание: функцию Беллмана искать в виде $B_k(x) = \langle \psi_k, x\rangle$, $k = 0, \ldots, N$.

2. Найти функцию Беллмана для задачи: $I_0(u) = \Phi(x_1) \to \inf$: $x_1 = x_0 + u$, $x_i \in G_i$, $i = 0, 1$; $u \in V_0$, где $\Phi(x) = (1 + e^{-1/x})^{-1}$ при $x \neq 0$, $\Phi(0) = 1/2$; $G_0 = \{x \in E^1\colon |x| \leqslant 1/2\}$, $G_1 = E^1$; $V_0 = \{u \in E^1\colon |u| \leqslant 1\}$. Показать, что $X_0 = G_0$, $X_1 = E^1$, и убедиться, что для этой задачи нижняя грань в (13), (16) не достигается.

3. Пусть функция $B_k(x)$, $x \in X_k$, $k = 0, \ldots, N$, удовлетворяет условиям (13), (14), а функция $u_{km}(x) \in D_k(x)$ и точки $x_{om} \in X_0$, $m = 1, 2, \ldots$, таковы, что $\lim\limits_{m\to\infty} R_i(x, u_{im}(x)) = 0$, $\lim\limits_{m\to\infty} B_0(x_{0m}) = \inf\limits_{X_0} B_0(x)$. Пусть управление $[u_{im}]_0$ и траектория $[x_{im}]_0$ построены по правилу: $u_{0m} = u_{0m}(x_{0m})$, $x_{1m} = F_0(x_{0m}, u_{0m})$, $u_{1m} = u_{1m}(x_{1m}), \ldots, x_{Nm} = F_{N-1}(x_{N-1,m}, u_{N-1,m})$. Тогда последовательность пар $(x_{0m}, [u_{im}]_0)$ — минимизирующая для задачи (5)–(8), т. е. $\lim\limits_{m\to\infty} I_0(x_{0m}, [u_{im}]_0) = I_0^*$. Доказать.

4. Доказать, что последовательность функций $u_{im}(x)$, $i = 0, \ldots, N-1$, $m = 1, 2, \ldots$, из упражнения 3 дает приближенное решение проблемы синтеза для задачи (5)–(8), т. е. если в момент k система находится в точке $x_k = x \in X_k$, то движение по закону $x_{i+1,m} = F_i(x_{im}, u_{im}(x_{im}))$, $i = k, \ldots, N-1$; $x_{km} = x$, $m = 1, 2, \ldots$, при больших m доставляет функции $I_k(x, [u_i]_k)$ значения, близкие к I_k^*.

5. Для задачи (9)–(12) получить оценку погрешности, аналогичную оценке (28).

6. Вывести уравнения Беллмана и доказать теоремы, аналогичные теоремам 1–4 для задачи:

$$I_0(x, [u_i]_0) = \sum_{i=0}^{N} F_i^0(x_i, u_i) + \Phi_0(x_0) + \Phi_1(x_N) \to \inf,$$

$$x_{i+1} = F_i(x_i, u_i), \quad i = 0, \ldots, N-1, \quad x_i \in G_i, \quad u_i \in V_i, \quad i = 0, \ldots, N,$$

где $x_i = (x_i^1, \ldots, x_i^n)$, $F_i = (F_i^1, \ldots, F_i^n)$, $u_i = (u_i^1, \ldots, u_i^r)$; множества $G_i \subseteq E^n$, $V_i \subseteq E^r$, $i = 0, \ldots, N$, заданы; функции $F_i^j(x, u)$ определены при $x \in G_i$, $u \in V_i$, $i = 0, \ldots, N$, $j = 0, \ldots, n$; функции $\Phi_0(x), \Phi_1(x)$ определены при $x \in G_0$, $x \in G_N$ соответственно; i — дискретное время; момент N считается заданным.

7. Обобщить оценку (28) и теоремы 5,6 для задачи из упражнения 6.

8. Пусть в задаче (5)–(8) $F_i^0(x, u) \equiv 0$, $i = 0, \ldots, N-1$. Покажем, что тогда $B_k(x) = \Phi(x)$ для всех $k = 0, \ldots, N$, причем функции $B_k(x)$ при различных k отличаются друг от друга областями своего определения X_k.

9. Применить метод динамического программирования к задаче минимизации функции

$$I_0(u_0, u_1, \ldots, u_N) = \sum_{i=0}^{N-1} f_i(u_i, u_{i+1}) + f_N(u_N)$$

при условиях $u_i \in V_i$, $i = 0, \ldots, N$. Указание: ввести функцию

$$B_k(u) = \inf\Big(\sum_{i=k}^{N-1} f_i(u_i, u_{i+1}) + f_N(u_N)\Big),$$

где нижняя грань берется по всем наборам $(u_k = u, u_{k+1}, \ldots, u_N)$, $u_i \in V_i$, $i = k, \ldots, N$, и показать, что $B_k(u) = \inf\limits_{v \in V_{k+1}} [f_k(u, v) + B_{k+1}(v)]$, $k = 0, \ldots, N-1$; $B_N(u) = f_N(u)$.

§ 2. Схема Моисеева

1. Будем рассматривать задачу (1.1)–(1.4). Для приближенного решения этой задачи, как и раньше, разобьем отрезок $t_0 \leqslant t \leqslant T$ на N частей точками $t_0 < t_1 < \ldots < t_{N-1} < t_N + T$. На множестве $G_i \equiv G(t_i)$ возьмем некоторую дискретную сетку точек $x_{ij} \in G_i$; следуя [497; 498] множество всех точек выбранной сетки, будем называть *шкалой состояний* и обозначать через H_i, $i = 0, \ldots, N$. Шкалы состояний H_i и H_{i+1} будем называть *соседними*. На двух соседних шкалах H_i и H_{i+1} возьмем точки $x \in H_i$ и $y \in H_{i+1}$ и рассмотрим следующую вспомогательную задачу:

$$J_i(x, y, u(\cdot)) = \int_{t_i}^{t_{i+1}} f^0(x(t), u(t), t)dt \to \inf, \tag{1}$$

$$\dot{x}(t) = f(x(t), u(t), t), \quad t_i \leqslant t \leqslant t_{i+1}; \quad x(t_i) = x, \quad x(t_{i+1}) = y, \tag{2}$$

$$x(t) \in G(t), \quad t_i \leqslant t \leqslant t_{i+1}, \tag{3}$$

$$u = u(\cdot) \text{кусочно-непрерывна и } u(t) \in V(t) \text{при } t_i \leqslant t \leqslant t_{i+1}. \tag{4}$$

Следуя [497; 498], задачу (1)–(4) будем называть *элементарной операцией*, соединяющей точки x и y. Через $\Delta_i(x, y)$ обозначим множество всех управлений $u = u(\cdot)$ из (4), для которых соответствующая траектория $x(\cdot)$ удовлетворяет всем условиям (2), (3). Положим $M_i(x, y) = \inf\limits_{u \in \Delta_i(x, y)} J_i(x, y, u)$; если $\Delta_i(x, y) = \varnothing$, то $M_i(x, y) = +\infty$ по определению.

Пусть все точки всех соседних шкал попарно соединены элементарными операциями. Если $\inf\limits_{\Delta_i(x, y)} J_i(x, y, u)$ достигается на некотором управлении $u_i(\cdot) \in \Delta_i(x, y)$ и соответствующей траектории $x_i(\cdot)$, то величина

$$\sum_{i=0}^{N-1} M_i(x_{ij_i}, x_{i+1, j_{i+1}}) + \Phi(x_{Nj_N}) \tag{5}$$

выражает собой значение исходной функции (1.1) на управлении $u = u(t) = u_i(t)$, $t_i \leqslant t \leqslant t_{i+1}$, $i = 0, \ldots, N-1$, и соответствующей траектории $x(t, u) = x_i(t)$, $t_i \leqslant t \leqslant t_{i+1}$, $i = 0, \ldots, N-1$; $x(t_0) = x_{0j_0}$ при соблюдении всех ограничений (1.2)–(1.4). Поэтому исходную задачу (1.1)–(1.4) естественно аппроксимировать задачей отыскания минимума суммы (5) по всевозможным наборам точек $(x_{0j_0}, x_{1j_1}, \ldots, x_{Nj_N})$, $x_{ij_i} \in H_i$, $i = 0, \ldots, N$.

Такая аппроксимация задачи (1.1)–(1.4) имеет смысл, конечно, и в том случае, когда $\inf\limits_{\Delta_i(x, y)} J_i(x, y, u)$ не достигается при каких-либо x, y, i. Очевидно, приведенная аппроксимация задачи (1.1)–(1.4) более гибкая и лучше приспособлена для приближенного решения этой задачи, чем схема из § 1, ибо здесь представляется свобода в выборе способа разрешения элементарной операции (1)–(4) и, кроме того, рассмотрение лишь таких траекторий, концы которых лежат в известных точках соседних шкал, избавляет нас от необходимости интерполирования. На способах разрешения элементарной операции мы остановимся ниже.

Описанная аппроксимация задачи (1.1)–(1.4) имеет простой геометрический смысл. А именно, траекторию $x_i(t)$ из (2), на которой достигается $\inf\limits_{\Delta_i(x, y)} J_i(x, y, u)$ назовем *дугой*, соединяющей точки x, y, а числа $M_i(x, y)$ — *длиной* этой дуги, $i = 0, \ldots, N-1$. Дуги, последовательно соединяющие пары точек x_{ij_i}, $x_{i+1, j_{i+1}}$ соседних шкал H_i, H_{i+1}, $i = 0, \ldots, N-1$, назовем путем, соединяющим шкалы H_0 и H_N и проходящим через точки $x_{0i_0}, x_{1j_1}, \ldots, x_{Nj_N}$, и в качестве его длины примем величину (5). Тогда наша задача сведется к отысканию кратчайшего пути, соединяющего шкалы H_0 и H_N.

2. Обозначим

$$C_k(x) = \inf\Big\{\sum_{i=k}^{N-1} M_i(x_{ij_i}, x_{i+1}, j_{i+1}) + \Phi(x_{Nj_N})\Big\},$$

где нижняя грань берется по всем наборам точек $(x_{kj_k} = x, x_{k+1}, j_{k+1}, \ldots, x_{Nj_N})$, $x_{ij_i} \in H_i$, $i = k, \ldots, N$. Иначе говоря, $C_k(x)$ выражает собой кратчайшее расстояние между фиксированной точкой $x \in H_k$ и шкалой H_N. Покажем, что функции $C_k(x)$ удовлетворяют следующим рекуррентным соотношениям:

$$C_k(x) = \inf_{y \in H_{k+1}} \{M_k(x, y) + C_{k+1}(y)\}, \quad k = 0, \ldots, N-1; \quad C_N(x) \equiv \Phi(x), \tag{6}$$

аналогичным условиям (1.13), (1.14). Справедливость (6) при $k = N-1$ следует из определения $C_{N-1}(x)$, $C_N(x)$. Докажем (6) при других k, $0 \leqslant k \leqslant N-1$. Для этого сначала убедимся в том, что

$$C_k(x) \leqslant \inf_{y \in H_{k+1}} \{M_k(x, y) + C_{k+1}(y)\}, \quad x \in H_k. \tag{7}$$

Возьмем произвольное $y \in H_{k+1}$. По определению $C_{k+1}(y)$ для любого $\varepsilon > 0$ найдется путь, соединяющий точку y со шкалой H_N, длина которого не превышает $C_{k+1}(y) + \varepsilon$. Если этот путь «удлинить», добавив к нему дугу, соединяющую точки x и y, то получим путь, соединяющий точку $x \in H_k$ со шкалой H_N, длина которого не превышает $M_k(x, y) + C_{k+1}(y) + \varepsilon$. Поэтому заведомо $C_k(x) \leqslant M_k(x, y) + C_{k+1}(y) + \varepsilon$. В силу произвольности точки $y \in H_{k+1}$ и величины ε отсюда имеем неравенство (7).

Теперь покажем, что в (7) на самом деле знак неравенства можно заменить знаком равенства. По определению $C_k(x)$ для любого $\varepsilon > 0$ найдется путь, соединяющий точку $x \in H_k$ со шкалой H_N, длина которого не превосходит $C_k(x) + \varepsilon$. Пусть этот путь проходит через точку $y_\varepsilon \in H_{k+1}$. Ясно, что отрезок этого пути от y_ε до H_N не меньше $C_{k+1}(y_\varepsilon)$, и поэтому весь путь от x до H_N не меньше $M_k(x, y_\varepsilon) + C_{k+1}(y_\varepsilon)$. Следовательно, $M_k(x, y_\varepsilon) + C_{k+1}(y_\varepsilon) \leqslant C_k(x) + \varepsilon$. Так как $y_\varepsilon \in H_{k+1}$, то отсюда имеем $\inf\limits_{y \in H_{k+1}} \{M_k(x, y) + C_{k+1}(y)\} \leqslant C_k(x) + \varepsilon$, или в силу произвольности ε: $\inf\limits_{y \in H_{k+1}} \{M_k(x, y) + C_{k+1}(y)\} \leqslant C_k(x)$. Сравнивая это неравенство с (7), немедленно получаем требуемые соотношения (6).

3. Соотношения (6) могут быть использованы так же, как и уравнение Беллмана (1.13), (1.14). Опишем порядок работы с этими соотношениями в предположении, что каждая из шкал состояний H_i состоит из конечного

числа точек p_i, $i=0,\ldots,N$. Заметим, что в этом случае в (6) вместо inf можно писать min. Функция $C_N(x)\equiv\Phi(x)$, $x\in H_N$ нам известна. Для вычисления $C_{N-1}(x)$ с помощью элементарных операций соединим попарно все точки шкал H_{N-1} и H_N. Сравнивая p_{N-1}, p_N величин $M_{N-1}(x,y)+\Phi(y)$ при всевозможных $x\in H_{N-1}$, $y\in H_N$, найдем $\min\limits_{y\in H_N}[M_{N-1}(x,y)+\Phi(y)]=C_{N-1}(x)$, а также точку $y=y_{N-1}(x)\in H_N$, $x\in H_{N-1}$, на которой этот минимум достигается. Если $C_{k+1}(x)$ и $y=y_{k+1}(x)\in H_{k+2}$, $x\in H_{k+1}$ уже известны, то для вычисления $C_k(x)$, $x\in H_k$, соединим элементарными операциями всевозможные пары точек шкал H_k и H_{k+1}. Перебором p_kp_{k+1} величин $M_k(x,y)+C_{k+1}(y)$ при всех $x\in H_k$, $y\in H_{k+1}$ найдем $\min\limits_{y\in H_{k+1}}[M_k(x,y)+C_{k+1}(y)]$, а также точку $y=y_k(x)\in H_{k+1}$, $x\in H_k$, на которой этот минимум достигается, и т. д. для всех $k=N, N-1,\ldots,1,0$. На вычисление всех $C_k(x)$ в $y=y_k(x)$, $x\in H_k$, $k=0,\ldots,N-1$, понадобится перебор $\sum\limits_{i=0}^{N-1}p_ip_{i+1}$ величин. Наконец, находим $x_0^*\in H_0$ из условия $C_0(x_0^*)=\inf\limits_{x\in H_0}C_0(x)$ — для этого нужно перебрать еще p_0 величин, и определяем последовательно точки $x_1^*=y_0(x_0^*)$, $x_2^*=y_1(x_1^*),\ldots,x_N^*=y_{N-1}(x_{N-1}^*)$. Путь, проходящий через найденные точки $x_0^*, x_1^*,\ldots,x_N^*$, будет кратчайшим среди всех путей, соединяющих крайние шкалы H_0, H_N. В самом деле, по определению $y_k(x)$, $k=0,\ldots,N-1$, и $x_k^*\in H_k$, $k=0,\ldots,N$, имеем

$$C_k(x_k^*)=M_k(x_k^*,x_{k+1}^*)+C_{k+1}(x_{k+1}^*),\quad k=0,\ldots,N-1. \tag{8}$$

Возьмем произвольный путь, соединяющий шкалы H_0 и H_N и проходящий через точки $x_0, x_1,\ldots,x_N$, $x_i\in H_i$, $i=0,\ldots,N$. Так как $x_{k+1}\in H_{k+1}$, то согласно (6) будем иметь $C_k(x_k)\leqslant M_k(x_k,x_{k+1})+C_{k+1}(x_{k+1})$. Отсюда и из (8) тогда следует $M_k(x_k^*,x_{k+1}^*)+C_{k+1}(x_{k+1}^*)-C_k(x_k^*)=0\leqslant M_k(x_k,x_{k+1})+C_{k+1}(x_{k+1})-C_k(x_k)$, $k=0,\ldots,N-1$. Просуммируем это неравенство по k от нуля до $N-1$. Получим

$$\sum_{i=0}^{N-1}M_k(x_k^*,x_{k+1}^*)+C_N(x_N^*)-C_0(x_0^*)\leqslant\sum_{i=0}^{N-1}M_k(x_k,x_{k+1})+C_N(x_N)-C_0(x_0).$$

Но $C_0(x_0^*)=\inf\limits_{x\in H_0}C_0(x)\leqslant C_0(x_0)$, поэтому $\sum\limits_{i=0}^{N-1}M_k(x_k^*,x_{k+1}^*)+\Phi(x_N^*)\leqslant\sum\limits_{i=0}^{N-1}M_k(x_k,x_{k+1})+\Phi(x_N)$ для любых путей, соединяющих шкалы H_0 и H_N. Таким образом, путь, проходящий через точки $x_0^*, x_1^*,\ldots,x_N^*$, в самом деле кратчайший. Если $u_i^*(t)$ и $x_i^*(t)$, $t_i\leqslant t\leqslant t_{i+1}$, представляют собой то управление и соответствующую траекторию, на которых приближенно реализуется элементарная операция (1)–(4) при $x=x_i^*$, $y=x_{i+1}^*$, то в качестве приближенного решения исходной задачи (1.1)–(1.4) можно взять управление $u^*(t)=u_i^*(t)$ и траекторию $x^*(t)=x_i^*(t)$, $t_i\leqslant t\leqslant t_{i+1}$, $i=0,\ldots,N-1$.

Аналогично доказывается, что путь, проходящий через точки $x_k^*=x\in H_k$, $x_{k+1}^*=y_k(x_k^*)\in H_{k+1},\ldots,x_N^*=y_{N-1}(x_{N-1}^*)\in H_N$, является кратчайшим между точкой $x\in H_k$ и шкалой H_N. Это означает, что функция $y_k(x)$ дает нам приближенное решение проблемы синтеза для задачи (1.1)–(1.4).

Заметим, что определение кратчайшего пути между шкалами H_0 и H_N по описанной схеме потребовало перебора $\sum\limits_{i=0}^{N-1}p_ip_{i+1}+p_0$ величин, в то время как полный перебор всех путей, как нетрудно видеть, потребовал бы сравнения $p_0p_1\ldots p_N$ величин. Таким образом, уже при не слишком больших p_i, перебор с помощью соотношений (6) по сравнению с полным перебором дает существенную экономию памяти ЭВМ и машинного времени. Такая экономия достигается за счет того, что при вычислении $C_k(x)$ из (6) рассматриваются лишь пути, проходящие через всевозможные точки $x\in H_k$ и $y\in H_{k+1}$ и соединяющие точки $y\in H_{k+1}$ со шкалой H_N кратчайшим образом, и тем самым все некратчайшие пути, соединяющие $y\in H_{k+1}$ со шкалой H_N, из рассмотрения полностью исключаются.

4. Для получения более точного решения задачи (1.1)–(1.4) необходимо взять более густую сетку точек на шкалах и увеличить число шкал. Однако при этом число перебираемых величин даже при использовании описанной выше схемы перебора катастрофически быстро растет, и уже при небольших размерностях векторов x, u становится невозможным решить задачу о кратчайшем пути за разумное время с помощью самых лучших современных ЭВМ. В этом случае часто используют прием, известный под названием *метода блуждающих трубок* [753]. Суть этого приема заключается в следующем.

Сначала берут небольшое число шкал с небольшим количеством точек на них, и по описанной выше схеме поиска находят кратчайший путь l_1, соединяющий крайние шкалы H_0 и H_N. Затем уменьшают шаг сетки на каждой шкале, путь l_1 окружают некоторой «трубкой» из путей, проходящих вблизи l_1 по точкам новой сетки. Для построения трубки вокруг l_1 на каждой шкале обычно берут небольшие окрестности точки, через которую проходит l_1, и рассматривают пути, проходящие через выбранные точки. С помощью описанной выше схемы перебора находят кратчайший путь в полученной трубке; все пути вне этой трубки в переборе пока не участвуют. Таким образом, получают новый улучшенный путь l_2, длина которого не превышает длину l_1. Далее, сохраняя прежние шкалы и точки на них, окружают путь l_2 новой трубкой и находят следующее приближение l_3 и т. д., продолжая процесс до тех пор, пока трубка не перестает «блуждать» и впервые не получится равенство $l_s=l_{s+1}$. После этого измельчают сетку на каждой шкале, окружают путь l_s новой трубкой и продолжают поиск кратчайшего пути описанным приемом блуждающих трубок. Процесс измельчения сетки на шкалах и поиска кратчайшего пути указанным способом повторяют, пока два кратчайших пути, полученные после двух последовательных измельчений сетки, не совпадают с удовлетворительной точностью. Затем увеличивают число шкал, т. е. сгущают сетку по времени, и повторяют процесс поиска методом блуждающих трубок с постепенным измельчением сетки на шкалах состояний до удовлетворительного совпадения двух последовательных приближений. Попеременно измельчая сетку на шкалах состояния и сетку по времени, поиск с помощью блуждающих трубок продолжают до получения приближенного решения исходной задачи с достаточной точностью. Изменение шагов сеток на шкалах состояния и по времени должно быть согласованным; например, в случае равномерных сеток эти шаги должны удовлетворять соотношению $|\Delta x|=o(\Delta t)$ [497].

Оказывается, метод блуждающих трубок во многих случаях существенно сокращает перебор. В то же время следует заметить, что метод блуждающих трубок позволяет определить, вообще говоря, лишь локально кратчайший путь между крайними шкалами при фиксированной сетке, поскольку на каждом шаге в переборе участвуют лишь пути, попавшие в трубку.

5. Другой подход к поиску кратчайшего пути между шкалами дает *метод локальных вариаций* [497; 498; 753]. Этот метод предполагает, что какой-то путь l_1, соединяющий шкалы H_0 и H_N уже известен. Для определения следующего более короткого пути последовательно просматриваются шкалы $H_0, H_1, \ldots, H_N$. Допустим, что шкалы $H_0, \ldots, H_{i-1}$ уже просмотрены и получен путь $l_{1,i-1}$, соединяющий шкалу H_0 и H_N. Пусть $x_i(l_1)$ — точка пути l_1, лежащая на шкале H_i. Выберем на шкале H_i некоторое количество точек, расположенных близко к точке $x_i(l_1)$, и переберем пути, проходящие через эти выбранные точки шкалы H_i, а в остальном совпадающие с путем $l_{1,i-1}$. Если среди перебираемых путей найдется путь, имеющий меньшую длину, чем путь $l_{1,i-1}$, то его обозначаем через l_{1i} и просмотр шкалы H_i на этом заканчиваем. Если же длины всех перебираемых путей оказались не меньше длины $l_{1,i-1}$, то полагаем $l_{1i} = l_{1,i-1}$. После определения пути l_{1i} переходим к просмотру следующей шкалы H_{i+1}, и т. д. Такой перебор всех шкал $H_0, H_1, \ldots, H_N$ закончится получением пути $l_{1N} = l_2$. Если длина пути l_2 меньше длины l_1, то для пути l_2 повторяют описанный выше просмотр шкал $H_0, H_1, \ldots, H_N$ и находят следующий путь l_3 и т. д. Если же окажется, что $l_2 = l_1$, т. е. путь l_1 улучшить не удалось, то на шкалах берут более густую сетку точек, а также при необходимости измельчают сетку по времени и снова просматривают шкалы и т. д.

Метод локальных вариаций описан. Нетрудно видеть, что этот метод является аналогом метода покоординатного спуска. Поскольку здесь в переборе участвуют гораздо меньше путей, чем в методе, основанном на соотношениях (6), или методе блуждающих трубок, то ясно, что метод локальных вариаций гораздо экономичнее и позволяет увеличить размерность решаемых задач. С другой стороны, нетрудно привести примеры задач, когда этим методом не удается найти даже локально кратчайший путь между шкалами. На рис. 7.1 приведен такой пример — здесь $N = 3$, шкалы H_0, H_1, H_2, H_3 состоят соответственно из точек $\{A\}, \{B, C\}, \{D, E\}, \{F\}$; на дугах, соединяющих точки соседних шкал, указаны их длины. Кратчайшим путем, соединяющим шкалы H_0 и H_3, является путь $ACDF$. Если в качестве начального приближения l_1 взять путь $ABEF$, то улучшить его методом локальных вариаций не удается. Любопытно также заметить, что если за l_1 взять путь $ABDF$, то в зависимости от того, начнем ли просмотр со шкалы H_1 или H_2, придем к кратчайшему пути $ACDF$ или уже рассмотренному пути $ABEF$.

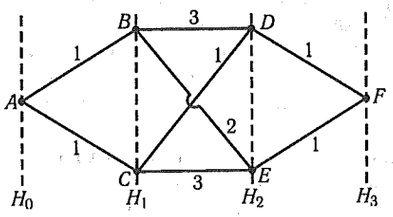

Рис. 7.1

Различные модификации метода локальных вариаций, примеры задач, к которым применялся этот метод, см. в [497; 498; 753].

6. Успех и применение описанных в этом параграфе методов приближенного решения задачи (1.1)–(1.4) во многом зависит от умения строить элементарные операции (1)–(4). По существу элементарная операция представляет собой экстремальную задачу той же трудности, что и исходная задача. Однако малость отрезка $t_i \leqslant t \leqslant t_{i+1}$ позволяет здесь сделать ряд упрощающих предположений. Прежде всего, если шаг сетки по времени достаточно мал, то элементарную операцию строят, исходя из условий (1), (2), (4), полагая, что дуги траекторий или не нарушают ограничений (3), или

этим нарушением можно пренебречь. Далее, вместо минимизации функции (1) при условиях (2), (4) часто ограничиваются построением какого-либо допустимого управления и соответствующей траектории, удовлетворяющих условиям (2), (4), и в качестве длины дуги $M_i(x, y)$ тогда берут значение функции (1) на полученных управлении и траектории.

Во многих задачах полезно задаться каким-либо семейством управлений $u(t) \equiv (t, c_1, \ldots, c_m)$, зависящих от параметров $c_1, \ldots, c_m$ $(m \geqslant n)$. Например, это могут быть алгебраические или тригонометрические многочлены с коэффициентами $c_1, \ldots, c_m$ или кусочно-постоянные функции со значениями $c_1, \ldots, c_m$ и т. д. Значения этих параметров затем можно определить из следующей системы n уравнений с m неизвестными $(m \geqslant n)$:

$$\int_{t_i}^{t_{i+1}} \dot{x}(t)dt = \int_{t_i}^{t_{i+1}} f(x(t), u(t, c_1, \ldots, c_m))dt = y - x, \tag{9}$$

$$u(t, c_1, \ldots, c_m) \in V(t_i), \quad t_i \leqslant t \leqslant t_{i+1},$$

используя для этого различные методы [59; 74; 89; 635]. Параметры $c_1, \ldots, c_m$ отсюда будут определяться, вообще говоря, неоднозначно, и свободные параметры можно использовать для минимизации функции (1). Для упрощения системы (9) дифференциальное уравнение (2) часто заменяют более простыми уравнениями:

$$\dot{x}(t) = f(x, u(t), t) \quad \text{или} \quad \dot{x}(t) = f\left(\frac{x+y}{2}, u(t), t\right)$$

или другими более точными разностными уравнениями; здесь возможно использование линеаризованной системы:

$$\dot{x}(t) = f(x, u(t), t) + \langle f_x(x, u(t), t), x(t) - x\rangle.$$

При решении задачи (1), (2), (4) часто применяется также принцип максимума в сочетании с различными упрощающими приемами, описанными выше. Другие способы построения элементарных операций, примеры решенных конкретных задач см. в [497; 498].

§ 3. Проблема синтеза для систем с непрерывным временем

Продолжим исследование задачи (1.1)–(1.4) с непрерывно меняющимся временем. *Проблема синтеза* для задачи (1.1)–(1.4) заключается в построении функции $u = u(x, t)$, называемой *синтезирующей* функцией этой задачи и представляющей собой значение оптимального управления при условии, что в момент t система (1.2) находится в точке x, т. е. $x(t) = x$. Умение решать проблему синтеза крайне важно в различных прикладных задачах оптимального управления. В самом деле, если известна синтезирующая функция $u(x, t)$, то техническое осуществление оптимального хода процесса может быть произведено по следующей схеме, называемой схемой с обратной связью; с измерительного прибора, замеряющего в каждый момент t фазовое состояние $x(t)$, на ЭВМ или какое-либо другое вычислительное средство подается величина $x(t)$, вычисляется значение управления $u(t) = u(x(t), t)$, после чего найденное значение $u(t)$ оптимального управления передается на исполнительный механизм, непосредственно реализующий требуемое течение управляемого процесса.

1. Проблема синтеза для задачи (1.1)–(1.4) сводится к решению следующей задачи: определить управление $u_*(\tau) = u_*(\tau, x, t)$, доставляющее функции

$$J(t, x, u(\cdot)) = \int_t^T f^0(x(\tau), u(\tau), \tau)d\tau + \Phi(x(T)) \tag{1}$$

минимальное значение при условиях

$$\dot{x}(\tau) = f(x(\tau), u(\tau), \tau), \quad t \leqslant \tau \leqslant T; \quad x(t) = x, \tag{2}$$

$$x(\tau) \in G(\tau), \quad t \leqslant \tau \leqslant T, \tag{3}$$

$$u = u(\tau) \in V(\tau), \quad t \leqslant \tau \leqslant T; \quad u(\cdot) \text{ — кусочно-непрерывно}, \tag{4}$$

где x — произвольная точка множества $G(t)$, а t — произвольный фиксированный момент времени, $t_0 \leqslant t \leqslant T$. Заметим, что при $t = t_0$ задача (1)–(4) превращается в исходную задачу (1.1)–(1.4).

Обозначим через $\Delta(x, t)$ множество всех управлений $u(\tau)$, $t \leqslant \tau \leqslant T$, удовлетворяющих условиям (4) и таких, что соответствующая траектория $x(\tau) = x(\tau, u)$, $t \leqslant \tau \leqslant T$, системы (2) определена на всем отрезке $[t, T]$ и удовлетворяет фазовому ограничению (3). Положим $X(t) = \{x: x \in G(t), \Delta(x, t) \neq \varnothing\}$, $t \leqslant \tau < T$; $X(T) = G(T)$.

Пару $(u(\tau), x(\tau))$, $t \leqslant \tau \leqslant T$, назовем *допустимой парой* задачи (1)–(4), если $x(t) = x \in X(t)$, $u(\cdot) \in \Delta(x, t)$ и $x(\cdot)$ является решением системы (2), соответствующим рассматриваемому управлению $u(\cdot)$. Аналогично, пару $(u(\tau), x(\tau))$, $t_0 \leqslant \tau \leqslant T$, будем называть допустимой парой исходной задачи (1.1)–(1.4), если $x(t_0) = x_0 \in X(t_0)$, $u(\cdot) \in \Delta(x_0, t_0)$ и выполнены все соотношения (1.2)–(1.4). Допустимую пару $(u_*(\tau), x_*(\tau))$, $t \leqslant \tau \leqslant T$, задачи (1)–(4) будем называть решением этой задачи, если $J(t, x, u_*(\cdot)) = \inf\limits_{\Delta(x, t)} J(t, x, u(\cdot))$, при этом $u_*(\cdot)$ назовем оптимальным управлением, а $x_*(\cdot)$ — оптимальной траекторией задачи (1)–(4).

Зная оптимальное управление $u_*(\tau) = u_*(\tau, x, t)$ задачи (1)–(4) при всех тех (x, t), $x \in G(t)$, $t_0 \leqslant t < T$, при которых эта задача имеет решение, нетрудно получить синтезирующую функцию задачи (1.1)–(1.4): достаточно положить $u(x, t) \equiv u_*(t, x, t)$. Однако получить явное аналитическое выражение для оптимального управления $u_*(\tau, x, t)$ задачи (1)–(4) удается лишь в редких случаях. Поэтому желательно иметь другие подходы к решению проблемы синтеза.

Вспомним, что при решении проблемы синтеза для дискретных систем важную роль играло уравнение Беллмана (1.13), (1.14). Оказывается, аналогичное уравнение может быть получено и для задачи (1)–(4). Введем функцию

$$B(x, t) = \inf_{\Delta(x, t)} J(t, x, u(\cdot)),$$

называемую *функцией Беллмана* для задачи (1.1)–(1.4). Если задача (1.1)–(1.4) удовлетворяет некоторым ограничениям и функция $B(x, t)$ непрерывно дифференцируема, то можно показать, что функция Беллмана удовлетворяет следующим условиям, называемым *уравнением* Беллмана задачи (1.1)–(1.4):

$$\inf_{u \in D(x, t)} [\langle B_x(x, t), f(x, u, t)\rangle + B_t(x, t) + f^0(x, u, t)] = 0, \quad x \in X(t), \quad t_0 \leqslant t < T, \tag{5}$$

$$B(x, T) = \Phi(x), \quad x \in X(T) = G(T), \tag{6}$$

где $B_x = (B_{x^1}, \ldots, B_{x^n})$, B_{x^i}, B_t — частные производные функции $B(x, t)$, а $D(x, t)$ — множество всех тех $u \in V(t)$, для которых существует хотя бы одно управление $u(\cdot) \in \Delta(x, t)$ со значением $u(t) = u(t - 0) = u$.

Приведем эвристические соображения, из которых следуют соотношения (5), (6). С этой целью воспроизведем уравнения (1.13), (1.14):

$$\inf_{u \in D(x, t)} [B_{k+1}(F_k(x, u)) - B_k(x) + F_k^0(x, u)] = 0, \quad x \in X_k, \quad k = 0, \ldots, N-1, \tag{7}$$

$$B_N(x) = \Phi(x), \quad x \in X_N = G_N. \tag{8}$$

Вспомним также обозначения, связывающие задачи (1.1)–(1.4) и (1.5)–(1.8)

$$F_k^0(x, u) = (t_{k+1} - t_k) f^0(x, u, t_k), \quad F_k(x, u) = x + (t_{k+1} - t_k) f(x, u, t_k)$$

и подставим их в (7), (8). Исключим из этих обозначений и соотношений (7), (8) индекс k, приняв $t_k = t$, $\Delta t = t_{k+1} - t_k$, $t_N = T$, $B_k(x) = B(x, t)$, $B_{k+1}(y) = B(y, t + \Delta t)$, $D_k(x) = D(x, t)$, $X_k = X(t)$. Тогда соотношения (7), (8) могут быть переписаны в следующей безындексной форме:

$$\inf_{u \in D(x, t)} [B(x + \Delta t f(x, u, t), t + \Delta t) - B(x, t) + \Delta t f^0(x, u, t)] = 0, \quad x \in X(t), \quad t_0 \leqslant t \leqslant T,$$

$$B(x, T) = \Phi(x), \quad x \in X(T) = G(T).$$

Если теперь поделим первое из этих равенств на Δt и совершим формальный предельный переход при $\Delta t \to 0$, то придем к соотношениям (5), (6). Подчеркнем, что приведенные рассуждения никоим образом не претендуют на какую-либо строгость м могут служить лишь наводящими соображениями при получении соотношений (5), (6). Аналогичным образом можно было бы «вывести» эти соотношения, опираясь на уравнения (2.6).

Заметим, что уравнение (5) является дифференциальным уравнением в частных производных первого порядка, левая часть которого осложнена взятием нижней грани, и вопросы существования и единственности решения задачи (5), (6), свойства ее решения в настоящее время исследованы слабо [419; 754]. Задача (5), (6) здесь нас будет интересовать с точки зрения решения проблемы синтеза.

Под решением задачи (5), (6) мы будем понимать функцию $B(x, t)$, которая определена и непрерывна при всех (x, t), $x \in X(t)$, $t_0 \leqslant t \leqslant T$, обладает кусочно-непрерывными частными производными B_x, B_t и удовлетворяет уравнению (5) всюду, где существуют эти производные, удовлетворяет условию (6) и, кроме того, для любой допустимой пары $(u(\cdot), x(\cdot))$ задачи (1)–(4) при всех $x \in X(t)$, $t_0 \leqslant t < T$, функция $B(x(\tau), \tau)$ переменной τ имеет кусочно-непрерывную производную (или $B(x(\tau), \tau)$ абсолютно непрерывна) на отрезке $[t, T]$.

Теорема 1. *Пусть $B(x, t)$ — решение задачи* (5), (6) *и, кроме того, пусть нижняя грань в левой части* (5) *достигается на кусочно непрерывной функции $u(x, t) \in D(x, t)$, $x \in X(t)$, $t_0 \leqslant t \leqslant T$. Тогда $u(x, t)$ — синтезирующая функция задачи* (1.1)–(1.4).

Доказательство. Возьмем произвольные t, $t_0 \leqslant t < T$, и $x \in X(t)$. Пусть $x_*(\tau)$, $t_0 \leqslant \tau \leqslant T$, является решением задачи Коши

$$\dot{x}(\tau) = f(x(\tau), u(x(\tau), \tau), \tau), \quad t_0 \leqslant \tau \leqslant T; \quad x(t) = x,$$

и пусть $x_*(\tau) \in X(\tau)$ при всех $\tau \in [t, T]$. Положим $u_*(\tau) = u(x_*(\tau), \tau)$, $t_0 \leqslant \tau \leqslant T$. Ясно, что $u_*(\cdot) \in \Delta(x, t)$ и $(u_*(\cdot), x_*(\cdot))$ — допустимая пара задачи (1)–(4). Для доказательства теоремы достаточно показать, что пара $(u_*(\cdot), x_*(\cdot))$ является решением задачи (1)–(4).

Сначала покажем, что для любой допустимой пары $(u(\cdot), x(\cdot))$ задачи (1)–(4) справедлива формула

$$J(t, x, u(\cdot)) = \int_t^T R(x(\tau), u(\tau), \tau) d\tau + B(x, t), \tag{9}$$

где

$$R(x, u, t) = \langle B_x(x, t), f(x, u, t)\rangle + B_t(x, t) + f^0(x, u, t). \tag{10}$$

В самом деле, по условию функция $B(x(\tau), \tau)$ переменной τ непрерывна и имеет кусочно-непрерывную производную. Тогда в силу уравнения (2) имеем

$$\frac{dB(x(\tau), \tau)}{d\tau} = R(x(\tau), u(\tau), \tau) - f^0(x(\tau), u(\tau), \tau)$$

всюду на $[t, T]$, за исключением, быть может, конечного числа точек. Интегрируя это тождество по τ на $[t, T]$, с учетом условия (6) получим

$$\Phi(x(T)) - B(x, t) = \int_t^T R(x(\tau), u(\tau), \tau) d\tau - \int_t^T f^0(x(\tau), u(\tau), \tau) d\tau,$$

что равносильно (9). Заметим, что формула (9) является аналогом формулы (1.21).

Уравнение (5) с помощью функции (10) можно переписать в виде $\inf\limits_{u \in D(x, \tau)} R(x, u, \tau) = 0$. Отсюда и из определения функции $u(x, \tau)$ имеем

$$R(x, u(x, \tau), \tau) = 0 = \inf_{u \in D(x, \tau)} R(x, u, \tau) \leqslant R(x, u, \tau) \tag{11}$$

для всех $u \in D(x, \tau)$, $x \in X(\tau)$, $t \leqslant \tau \leqslant T$. Если $(u(\cdot), x(\cdot))$ — допустимая пара задачи (1)–(4), то $x(\tau) \in X(\tau)$, $u(\tau) = u(\tau + 0) = u \in D(x(\tau), \tau)$, $t \leqslant \tau \leqslant T$. Поэтому из (11) получаем

$$R(x_*(\tau), u(x_*(\tau)), \tau) = R(x_*(\tau), u_*(\tau), \tau) = 0 \leqslant R(x(\tau), u(\tau), \tau), \tag{12}$$

$t \leqslant \tau \leqslant T$, для любой допустимой пары задачи (1)–(4). Отсюда и из формулы (9) с учетом условия $x_*(t) = x(t) = x$ имеем

$$J(t, x, u(\cdot)) - J(t, x, u_*(\cdot)) = \int_t^T R(x(\tau), u(\tau), \tau) d\tau \geqslant 0 \tag{13}$$

для всех допустимых пар $(u(\cdot), x(\cdot))$ задачи (1)–(4).

Из (9), (12), (13) следует, что

$$J(t, x, u(\cdot)) = \inf_{\Delta(x, t)} J(t, x, u(\cdot)) = B(x, t).$$

Тем самым доказано, что функция $B(x, t)$, определяемая соотношениями (5), (6), в самом деле является функцией Беллмана задачи (1.1)–(1.4), и функция $u(x, t)$, на которой достигается нижняя грань в левой части (5), является синтезирующей для этой задачи. □

С помощью функций $B(x, t)$, $u(x, t)$ нетрудно получить решение и для исходной задачи (1.1)–(1.4). А именно, верна

Теорема 2. *Пусть $B(x, t)$ — решение задачи* (5), (6) *и пусть нижняя грань в левой части* (5) *достигается на кусочно-непрерывной функции $u(x, t)$. Кроме того, пусть точка $x_0^* \in X(t_0)$ определена из условия*

$$B(x_0^*, t_0) = \inf_{x \in X(t_0)} B(x, t_0), \tag{14}$$

а пара $(u_*(\cdot), x_*(\cdot))$, где $x_*(\cdot)$ — решение задачи Коши

$$\dot{x}(\tau) = f(x(\tau), u(x(\tau), \tau), \tau), \quad t_0 \leqslant \tau \leqslant T; \quad x(t_0) = x_0^*,$$

и $u_*(\tau) = u(x_*(\tau), \tau)$, является допустимой парой задачи (1.1)–(1.4). *Тогда пара $(u_*(\cdot), x_*(\cdot))$ является решением задачи* (1.1)–(1.4), *т. е.*

$$J(t_0, x_0^*, u_*(\cdot)) = \inf_{x \in X(t_0)} \inf_{u(\cdot) \in \Delta(x, t_0)} J(t_0, x, u(\cdot)) = J_*.$$

Доказательство. Возьмем произвольную допустимую пару $(u(\tau), x(\tau))$, $t_0 \leqslant \tau \leqslant T$, $x(t_0) = x_0 \in X(t_0)$ задачи (1.1)–(1.4). Из формулы (9) и неравенств (12) при $t = t_0$ с учетом условия (14) имеем

$$J(t_0, x_0, u\cdot)) - J(t_0, x_0^*, u_*(\cdot)) = \int_{t_0}^{T} R(x(\tau), u(\tau), \tau) d\tau + B(x_0, t_0) - \inf_{x \in X(t_0)} B(x, t_0) \geqslant 0,$$

что и требовалось доказать. □

2. Таким образом согласно теореме 1 для решения рассматриваемой проблемы синтеза достаточно найти решение задачи (5), (6). Возникает вопрос, как же решить задачу (5), (6)? Прежде всего заметим, что удобное для работы конструктивное описание множеств $X(t)$, $D(x, t)$, входящих в формулировку задачи (5), (6), часто отсутствует, и поэтому на практике вместо задачи (5), (6) обычно пользуются следующей более конструктивной задачей:

$$\begin{gathered} \inf_{u \in V(t)} [\langle B_x(x, t), f(x, u, t)\rangle + B_t(x, t) + f^0(x, u, t)] = 0, \\ x \in G(t), \quad t_0 \leqslant t < T; \\ B(x, T) = \Phi(x), \quad x \in G(T), \end{gathered} \tag{15}$$

получающейся из (5), (6) заменой $D(x, t)$, $X(t)$ на $V(t)$, $G(t)$ соответственно. Конечно, здесь надо помнить, что задача (15) может и не иметь решения, в то время как задача (5), (6) может оказаться разрешимой.

Наиболее удобными и эффективными при решении задачи (5), (6) или задачи (15), по-видимому, являются методы, изложенные выше в § 1, 2, — рекуррентные соотношения (1.13), (1.14) и (2.6) по существу представляют собой некоторую дискретную аппроксимацию задач (5), (6) и (15), а функции $B_k(x)$, $C_k(x)$ являются приближенным значением для $B(x, t_k)$. Существуют и другие методы решения таких задач. В прикладных задачах часто удается получить явное выражение $u = u(x, t, B_x)$ для точки u, в которой достигается нижняя грань

$$\inf_{u \in V(t)} [\langle B_x, f(x, u, t)\rangle + f^0(x, u, t)]$$

при фиксированных значениях параметров (x, t, B_x). Подставив такое $u = u(x, t, B_x)$ в (15), приходим к следующей задаче Коши:

$$\begin{gathered} B_t + [\langle B_x, f(x, u, t)\rangle + f^0(x, u, t)]|_{u = u(x, t, B_x)} = 0, \\ x \in G(t), \quad t_0 \leqslant t < T; \quad B(x, T) = \Phi(x), \quad x \in G(T) \end{gathered}$$

для нелинейного уравнения с частными производными первого порядка. Для численного решения задачи Коши пользуются известным арсеналом методов — разностными методами, методом характеристик, методом прямых, методом коллокации и т. п. [59; 74; 89; 481; 635].

Иногда удается найти решение $B(x, t)$ задачи (15) в виде многочлена по переменным $x^1, \ldots, x^n$ с неопределенными коэффициентами, зависящими от времени:

$$B(x, t) = \sum_{i_1=0}^{m_1} \sum_{i_2=0}^{m_2} \cdots \sum_{i_n=0}^{m_n} \psi_{i_1 \ldots i_n}(t)(x^1)^{i_1} \ldots (x^n)^{i_n}.$$

Если подставим это выражение для $B(x, t)$ в (15), то для определения коэффициентов $\psi_{i_1 \ldots i_n}(t)$ получим дифференциальное уравнение следующего вида

$$\begin{gathered} B_t(x, t) = \sum_{i_1=0}^{m_1} \cdots \sum_{i_n=0}^{m_n} \dot{\psi}_{i_1 \ldots i_n}(t)(x^1)^{i_1} \ldots (x^n)^{i_n} = \\ = \inf_{u \in V(t)} F(\psi_{0 \ldots 0}(t), \ldots, \psi_{m_1 \ldots m_n}(t); x^1 \ldots x^n, u, t), \quad x \in G(t), \quad t_0 \leqslant t \leqslant T, \end{gathered} \tag{16}$$

с начальным условием

$$\sum_{i_1=0}^{m_1} \cdots \sum_{i_n=0}^{m_n} \psi_{i_1 \ldots i_n}(T)(x^1)^{i_1} \ldots (x^n)^{i_n} = \Phi(x), \quad x \in G(T). \tag{17}$$

Если $\Phi(x)$, $\inf\limits_{V(t)} F$, в свою очередь, являются многочленами относительно $x^1 \ldots x^n$, то, приравняв коэффициенты при одинаковых степенях в (16), (17), получим задачу Коши для системы обыкновенных дифференциальных уравнений относительно $\psi_{i_1 \ldots i_n}(t)$, записанной в нормальной форме Коши. Далее, здесь можно использовать различные численные методы решения задачи Коши, такие, как методы Эйлера, Адамса, Рунге — Кутта и т. д. [59; 74; 89; 481; 635].

Если $\Phi(x)$ или $\inf\limits_{V(t)} F$ не являются многочленами относительно $x^1, \ldots, x^n$, то условия (16), (17) не могут быть, вообще говоря, удовлетворены во всей области $G(t)$, $t_0 \leqslant t \leqslant T$, ни при каком выборе $N = (m_1 + 1) \ldots (m_n + 1)$ коэффициентов $\psi_{i_1 \ldots i_n}(t)$. В этом случае в [417] предлагается задать в области $G(t)$ N кривых $\xi_1(t), \ldots, \xi_N(t)$ и рекомендуется определять $\psi_{i_1 \ldots i_n}(t)$ из условия удовлетворения равенств (16), (17) не всюду в $G(t)$, а лишь на этих кривых. Этот подход перекликается с известными методами коллокаций и интегральных соотношений и приводит к задаче Коши для системы обыкновенных дифференциальных уравнений, не разрешенных относительно производной $\dot{\psi}_{i_1 \ldots i_n}(t)$ (отметим, что эти производные в уравнение будут входить линейно). Кривые $\xi_1(t), \ldots, \xi_N(t)$ обычно выбирают так, чтобы они имели достаточно простое аналитическое выражение (например, семейство прямых, параллельных осям координат, семейство парабол и т. п.) и задавали достаточно густую сетку в области $G(t)$, $t_0 \leqslant t \leqslant T$.

Для иллюстрации вышесказанного приведем пример.

Пример 1. Пусть требуется минимизировать функцию

$$J(u) = \int_0^T u^2(t) dt + \lambda x^2(T), \quad \lambda = \text{const} > 0,$$

при условиях $\dot{x} = u(t)$, $x(0) = x_0$, $u = u(t)$ — кусочно-непрерывная функция; числа T, x_0 заданы. Здесь $G(t) \equiv E^1$, $V(t) \equiv E^1$, $0 \leqslant t \leqslant T$.

Задача (15) в рассматриваемом случае имеет вид

$$\inf_{u \in E^1} [B_x(x, t)u + B_t(x, t) + u^2] = 0, \quad x \in E^1, \quad 0 \leqslant t \leqslant T, \tag{18}$$

$$B(x, T) = \lambda x^2, \quad x \in E^1. \tag{19}$$

Нижняя грань в (18) достигается при $u = -B_x/2$, поэтому уравнение (18) перепишется так:

$$B_t(x, t) - B_x^2(x, t)/4 = 0, \quad x \in E^1, \quad 0 \leqslant t \leqslant T. \tag{20}$$

Функцию $B(x, t)$ будем искать в виде многочлена

$$B(x, t) \equiv \psi_0(t) + \psi_1(t)x + \psi_2(t)x^2$$

переменной x. Подставим это выражение в (19), (20); получим

$$\begin{gathered} \dot{\psi}_0 + \dot{\psi}_1 x + \dot{\psi}_2 x^2 - (\psi_1 + 2\psi_2 x)^2/4 = 0, \quad x \in E^1, \quad 0 \leqslant t \leqslant T, \\ \psi_0(T) + \psi_1(T)x + \psi_2(T)x^2 = \lambda x^2, \quad x \in E^1. \end{gathered}$$

Приравнивая коэффициенты при одинаковых степенях x, придем к следующей задаче Коши:

$$\begin{gathered} \dot{\psi}_0 - \psi_1^2/4 = 0, \quad \dot{\psi}_1 - \psi_1\psi_2 = 0, \quad \dot{\psi}_2 - \psi_2^2 = 0, \quad 0 \leqslant t \leqslant T, \\ \psi_0(T) = 0, \quad \psi_1(T) = 0, \quad \psi_2(T) = \lambda. \end{gathered}$$

Отсюда находим

$$\psi_0(T) \equiv \psi_1(T) \equiv 0, \quad \psi_2(t) = \frac{\lambda}{1 - \lambda(t - T)}.$$

Таким образом, функция Беллмана здесь имеет вид

$$B(x, t)=\frac{\lambda x^2}{1-\lambda(t-T)};$$

синтезирующей является функция

$$u(x, t)=-\frac{B_x}{2}=\frac{\lambda x}{1-\lambda(t-T)}, \quad x\in E^1, \quad 0\leqslant t\leqslant T.$$

3. Предположим, что с помощью того или иного метода нам удалось получить некоторое приближенное решение $B(x, t)$ задачи (5), (6) или (15). Если это решение получено разностным методом (например, методами § 1, 2) на какой-то дискретной сетке точек, то доопределим ее (например, интерполяцией или с помощью сплайнов) во всех точках области $G(t)$, $0\leqslant t\leqslant T$, до некоторой непрерывной кусочно-гладкой функции $B(x, t)$. Тогда функцию $u=u(x, t)$, на которой реализуется точная или приближенная нижняя грань функции $R(x, u, t)$ из (10) на множестве $D(x, t)$ или $V(t)$, можем принять в качестве приближенного решения проблемы синтеза для задачи (1.1)–(1.4). Это значит, что приближенное решение $(\bar u(\cdot), \bar x(\cdot))$ задачи (1)–(4) будем определять из условий

$$\dot{\bar x}(\tau)=f(\bar x(\tau), u(\bar x(\tau), \tau), \tau), \quad t\leqslant\tau\leqslant T, \quad \bar x(T)=x,$$
$$\bar x(\tau)\in G(\tau), \quad \bar u(\tau)=u(\bar x(\tau), \tau), \quad t\leqslant\tau\leqslant T. \tag{21}$$

Приближенное решение исходной задачи (1.1)–(1.4) находится аналогично: сначала определяем точку $\bar x_0$, на которой точно или приближенно реализуется нижняя грань функции $B(x, t_0)$ на множестве $X(t_0)$ или $G(t_0)$, а затем решая задачу (21) при $t=t_0$, $x=\bar x_0$, находим траекторию $\bar x(\tau)$ и управление $\bar u(\tau)=u(\bar x(\tau))$, $t_0\leqslant\tau\leqslant T$. Найденную пару $(\bar u(\cdot), \bar x(\cdot))$ примем за приближенное решение задачи (1.1)–(1.4). Спрашивается, какая при этом будет допущена погрешность? Приводимая ниже оценка погрешности дает некоторый ответ на этот вопрос.

Пусть $K(x, t)$ — какая-либо функция, которая определена и непрерывна при всех $x\in X(t)$, $t_0\leqslant t\leqslant T$, обладает кусочно-непрерывными производными K_x, K_t и такова, что для любой допустимой пары $(u(\cdot), x(\cdot))$, задачи (1)–(4) при всех $x\in X(t)$, $t_0\leqslant t<T$, функция $K(x(\tau), \tau)$ переменной τ — кусочно-гладкая (или абсолютно непрерывная) на $[t, T]$. На практике в качестве функции $K(x, t)$ обычно берут какое-либо приближенное решение $B(x, t)$ задачи (5), (6) или (15).

По аналогии с (10) введем функцию

$$S(x, u, t)=\langle K_x(x, t), f(x, u, t)\rangle+K_t(x, t)+f^0(x, u, t),$$
$$x\in X(t), \quad t_0\leqslant t\leqslant T, \tag{22}$$

и, кроме того, положим

$$s(x, T)=\Phi(x)-K(x, T), \quad x\in G(T). \tag{23}$$

Возьмем произвольную допустимую пару $(u(\cdot), x(\cdot))$ задачи (1)–(4). В силу уравнения (2) тогда имеем

$$\frac{dK(x(\tau), \tau)}{d\tau}=S(x(\tau), u(\tau), \tau)-f^0(x(\tau), u(\tau), \tau), \quad t_0\leqslant\tau\leqslant T.$$

Учитывая непрерывность и кусочно-гладкую функцию $K(x(\tau), \tau)$, проинтегрируем это тождество по τ от t до T. Получим формулу

$$J(t, x, u(\cdot))=\int_t^T S(x(\tau), u(\tau), \tau)d\tau+s(x(T), T)+K(x, t). \tag{24}$$

Если $K(x, t)=B(x, t)$, то $S(x, u, \tau)=R(x, u, \tau)$, $s(x, T)\equiv 0$, и эта формула превратится в выведенную выше формулу (9).

Предположим, что каким-то образом мы получили пару $(\bar u(\tau), \bar x(\tau))$, $t\leqslant\tau\leqslant T$, удовлетворяющую условиям (3), (4) и уравнению (2) с начальным условием $\bar x(t)=\bar x\in X(t)$. Согласно (24) тогда

$$J(t, \bar x, \bar u(\cdot))-J(t, x, u(\cdot))=\int_t^T[S(\bar x(\tau), \bar u(\tau), \tau)-S(x(\tau), u(\tau), \tau)]d\tau+$$
$$+[s(\bar x(T), T)-s(x(T), T)]+[K(\bar x, t)-K(x, t)] \tag{25}$$

для любой допустимой пары $(u(\cdot), x(\cdot))$ задачи (1)–(4). Из (25) уже нетрудно получить требуемые оценки погрешности для задач (1)–(4) и (1.1)–(1.4). Обозначим

$$S_{\min}(\tau)=\inf_{x\in X(\tau)}\inf_{u\in D(x, \tau)} S(x, u, \tau), \quad s_{\min}=\inf_{x\in X(t)} s(x, T), \quad K_{\min}(t)=\inf_{x\in X(t)} K(x, t). \tag{26}$$

Пусть $(\bar u(\cdot), \bar x(\cdot))$ — некоторая допустимая пара задачи (1) — (4), которую мы хотим взять в качестве приближенного решения этой задачи. Учитывая, что для любой допустимой пары $(u(\cdot), x(\cdot))$ задачи (1)–(4) имеют место включения $x(t)=x\in X(t)$, $u(\cdot)\in\Delta(x, t)$, $x(\tau)\in X(\tau)$, $u(\tau)\in D(x(\tau), \tau)$, $t\leqslant\tau\leqslant T$, из (25), (26) получим требуемую оценку погрешности:

$$0\leqslant J(t, \bar x, \bar u(\cdot))-\inf_{\Delta(x, t)} J(t, x, u(\cdot))\leqslant$$
$$\leqslant\int_t^T[S(\bar x(\tau), \bar u(\tau), \tau)-S_{\min}(\tau)]d\tau+[s(\bar x(T), T)-s_{\min}]+[K(\bar x, t)-K_{\min}(t)]. \tag{27}$$

Если же $(\bar u(\tau), \bar x(\tau))$, $t_0\leqslant\tau\leqslant T$, — допустимая пара задачи (1.1)–(1.4), $\bar x(t_0)=\bar x_0$, которая берется за приближенное решение этой задачи, то из (25), (26) имеем такую оценку погрешности:

$$0\leqslant J(t_0, \bar x_0, \bar u(\cdot))-\inf_{x\in X(t_0)}\inf_{u(\cdot)\in\Delta(x, t_0)} J(t_0, x, u(\cdot))\leqslant$$
$$\leqslant\int_{t_0}^T[S(\bar x(\tau), \bar u(\tau), \tau)-S_{\min}(\tau)]d\tau+[s(\bar x(T), T)-s_{\min}]+K(\bar x_0, t_0)-K_{\min}(t_0). \tag{28}$$

Если $K(x, t)=B(x, t)$, то $S(x, u, t)=R(x, u, t)$, $s(x, T)=0$ и, кроме того, из (11) следует, что $\inf_{u\in D(x, t)} R(x, u, t)=0$ при всех $x\in X(t)$, так что $\inf_{x\in X(t)}\inf_{u\in D(x, t)} R(x, u, t)=0$. Поэтому при $K(x, t)=B(x, t)$ из (27), (28) соответственно получим

$$0\leqslant J(t, \bar x, \bar u(\cdot))-\inf_{u(\cdot)\in\Delta(x, t)} J(t, x, u(\cdot))\leqslant\int_t^T R(\bar x(\tau), \bar u(\tau), \tau)d\tau+B(\bar x, t)-\inf_{x\in X(t)} B(x, t), \tag{29}$$

$$0\leqslant J(t_0, \bar x_0, \bar u(\cdot))-\inf_{x\in X(t_0)}\inf_{u(\cdot)\in\Delta(x, t_0)} J(t_0, x, u(\cdot))\leqslant$$
$$\leqslant\int_{t_0}^T R(\bar x(\tau), u(\tau), \tau)d\tau+B(\bar x_0, t_0)-\inf_{x\in X(t_0)} B(x, t_0). \tag{30}$$

Из оценок (29), (30) следует, что если определить $\bar u(x, t)\in\Delta(x, t)$, $\bar x\in X(t)$ так, чтобы $R(x, \bar u(x, t), t)$ было поближе к $\inf_{\Delta(x, t)} R(x, u, t)$, $B(\bar x, t)$ — поближе к $\inf_{x\in X(t)} B(x, t)$, и затем найти пару $(\bar u(\tau), \bar x(\tau))$ из следующих условий

$$\dot{\bar x}(\tau)=f(\bar x(\tau), \bar u(\bar x(\tau), \tau)\tau), \quad \bar x(\tau)\in G(\tau), \quad t\leqslant\tau\leqslant T, \quad \bar x(t)=\bar x, \quad \bar u(\tau)=\bar u(\bar x(\tau), \tau),$$

то величина $J(t, \bar x, \bar u(\cdot))$ (в случае задачи (1.1)–(1.4) — величина $J(t_0, \bar x_0, \bar u(\cdot))$) будет мало отличаться от искомого оптимального значения, а функция $\bar u(x, t)$ будет хорошим приближением для синтезирующей функции. Заметим, что оценки (28), (30) являются аналогами оценок (1.28) и (1.29).

Заметим также, что в приложениях могут оказаться удобнее более грубые оценки, получающиеся из (26)–(30) при замене неконструктивно определенных множеств $\Delta(x, t)$, $X(t)$ на множества $V(t)$, $G(t)$ соответственно.

Упражнения

1. Решить проблему синтеза для задачи минимизации функции $J(x_0, u(\cdot))=\int_0^T(x^2(t)+u^2(t))dt$ при условиях $\dot x(t)=-x(t)+u(t)$, $x(0)=x_0$; здесь $G(t)\equiv E^1$, $V(t)\equiv E^1$ при всех $t\in[0, T]$.

2. Решить проблему синтеза для задачи минимизации функции $J(x_0, u(\cdot))=x^2(1)$ при условиях $\dot x(t)=u(t)$, $0\leqslant t\leqslant 1$, $x(0)=x_0$, $u(t)\in V=\{u\in E^1:\ 0\leqslant u\leqslant 1\}$, $x(t)\in G(t)\equiv E^1$ при $0\leqslant t\leqslant 1$. Показать, что в этой задаче синтезирующих функций бесконечно много. Убедиться, например, что синтезирующими являются функции

$$u(x, t)=\begin{cases}0, & x\geqslant 0,\\ 1, & x<0,\end{cases} \quad \text{или} \quad u(x, t)=\begin{cases}0, & x>t-1,\\ 1, & x\leqslant t-1.\end{cases}$$

3. Решить проблему синтеза для задачи минимизации функций

$$J(x_0, u(\cdot))=x^2(T) \quad \text{или} \quad J(x_0, u(\cdot))=\int_0^T x^2(t)dt$$

при условиях $\dot{x}(t)=u(t)$, $0\leqslant t\leqslant T$; $x(0)=x_0$, $u(t)\in V(t)=\{u\in E^1:\ -1\leqslant u\leqslant 1\}$, $x(t)\in G(t)\equiv E^1$, $0\leqslant t\leqslant T$. Будет ли в этих задачах синтезирующая функция единственной?

4. Написать уравнения Беллмана (5), (6) для задачи быстродействия:

$$J(x_0, t_0, u(\cdot))=T-t_0\to\inf,\quad \dot{x}(t)=f(x(t),u(t),t),\quad t_0\leqslant t\leqslant T,$$

$x(t_0)=x_0$, $x(T)=x_1$, $u(\cdot)$ — кусочно-непрерывна и принадлежит множеству $V\subseteq E^r$, $t\geqslant t_0$.

5. Показать, что функция Беллмана для задачи из примера 6.2.4 не является непрерывно дифференцируемой.

6. найти функцию Беллмана для задачи

$$J(t_0, x_0, u(\cdot))=\int_{t_0}^{T}[\langle a(t), x(t)\rangle+b(u(t),t)]dt+\langle c, x(T)\rangle\to\inf,$$

$$\dot{x}(t)=A(t)x(t)+C(t)u(t)+f(t),\quad t_0\leqslant t\leqslant T,\quad x(t_0)=x_0, \tag{31}$$

$$u=u(t)\in V(t),\quad u(\cdot)\text{ — кусочно-непрерывна}, \tag{32}$$

считая известными моменты времени t_0, T, матрицы $A(t)$, $C(t)$ порядка $n\times n$ и $n\times r$ соответственно, n-мерные вектор-функции $a(t)$, $f(t)$, скалярную функцию $b(u,t)$, n-мерные векторы c, x_0 и множества $V(t)\in E^r$, $t_0\leqslant t\leqslant T$. Указание: пользуясь уравнениями (5), (6) искать функцию $B(x,t)$ в виде $B(x,t)=\langle\psi(t),x\rangle$ — многочлена первой степени относительно переменных $x=(x^1,\ldots,x^n)$.

7. Исследовать уравнения (5), (6) для задачи минимизации функции

$$J(t_0, x_0, u(\cdot))=\alpha_1\int_{t_0}^{T}x^2(t)dt+\alpha_2x^2(T),\quad \alpha_1,\alpha_2=\text{const}\geqslant 0,$$

при условиях (31), (32). Рассмотреть случаи $V(t)=E^r$, $V(t)=\{u\in E^r:\ |u|\leqslant 1\}$, $V(t)=\{u=(u^1,\ldots,u^r):\ -1\leqslant u^i\leqslant 1,\ i=1,\ldots,r\}$.

§ 4. Достаточные условия оптимальности

При решении задач оптимального управлении часто возникает следующий вопрос: будут ли на самом деле оптимальными те управления и соответствующие им траектории, которые найдены с помощью каких-либо точных или приближенных методов? Такой вопрос, например, естественно возникает, когда управление и траектория найдены из краевой задачи принципа максимума, поскольку принцип максимума выражает собой необходимое условие оптимальности, не являясь, в общем случае, достаточным для оптимальности. Один из подходов, с помощью которого можно получить достаточные условия оптимальности, связан с методом динамического программирования [105; 140; 202; 253; 254; 417; 418; 616]. Этот подход уже был использован в § 1 для получения достаточных условий в дискретных системах. Покажем возможности этого подхода для задач оптимального управления с непрерывно меняющимся временем.

1. Начнем с рассмотрения задачи (1.1)–(1.4) с закрепленным временем. Будем пользоваться обозначениями и некоторыми формулами из § 3. Согласно (3.24) и (3.28) для каждой допустимой пары $(\bar{u}(t),\bar{x}(t))$, $t_0\leqslant t\leqslant T$, $\bar{x}(t_0)=\bar{x}_0$, задачи (1.1)–(1.4) справедливы формула

$$J(t_0,\bar{x}_0,\bar{u}(\cdot))=\int_{t_0}^{T}S(\bar{x}(t),\bar{u}(t),t)dt+s(x(T),T)+K(\bar{x}_0,t_0) \tag{1}$$

и оценка

$$0\leqslant J(t_0,\bar{x}_0,\bar{u}(\cdot))-J_*\leqslant\int_{t_0}^{T}[S(\bar{x}(t),\bar{u}(t),t)-S_{\min}(t)]dt+$$
$$+[s(\bar{x}(T),T)-s_{\min}]+[K(\bar{x}_0,t_0)-K_{\min}(t_0)], \tag{2}$$

где

$$S(x,u,t)=\langle K_x(x,t),f(x,u,t)\rangle+K_t(x,t)+f^0(x,u,t), \tag{3}$$

$$S(x,T)=\Phi(x)-K(x,T), \tag{4}$$

$$S_{\min}(t)=\inf_{x\in X(t)}\inf_{u\in D(x,t)}S(x,u,t),\quad s_{\min}=\inf_{x\in X(t)}s(x,T), \tag{5}$$

$$K_{\min}(t)=\inf_{x\in X(t)}K(x,t),\quad J_*=\inf_{x\in X(t_0)}\inf_{u(\cdot)\in\Delta(x,t_0)}J(t_0,x,u(\cdot));$$

остальные обозначения см. в § 3. Напомним, что для справедливости соотношений (1), (2) достаточно было, чтобы функция $K(x,t)$ была определена и непрерывна при всех $x\in G(t)$, $t\in[t_0,T]$, обладала кусочно-непрерывными производными K_x, K_t, а функция $K(x(\tau),\tau)$ переменной τ для любой допустимой пары $(u(\cdot),x(\cdot))$ задачи (1.1)–(1.4) была непрерывной и кусочно-гладкой на отрезке $[t_0,T]$.

Теорема 1. *Для того чтобы допустимая пара $(\bar{u}(t),\bar{x}(t))$, $t_0\leqslant t\leqslant T$, $\bar{x}(t_0)=\bar{x}_0$, задачи (1.1)–(1.4) была решением этой задачи, достаточно существования функции $K(x,t)$, для которой формула (1) верна для любой допустимой пары задачи (1.1)–(1.4) и*

$$S(\bar{x}(t),\bar{u}(t),t)=S_{\min}(t),\quad t_0\leqslant t\leqslant T, \tag{6}$$

$$s(\bar{x}(T),T)=s_{\min},\quad K(\bar{x}_0,t_0)=K_{\min}(t_0), \tag{7}$$

где $S(x,u,t)$, $s(x,t)$, $S_{\min}(t)$, $s_{\min}$, $K_{\min}(t)$ определяются (3)–(5).

Доказательство этой теоремы следует из того, что при выполнении условий (6), (7) правая часть оценки (2) обращается в нуль и $J(t_0,\bar{x}_0,\bar{u}(\cdot))=J_*$. □

С помощью оценки (2) нетрудно также получить условия, достаточные для того, чтобы та или иная последовательность допустимых управлений и траекторий была минимизирующей для задачи (1.1)–(1.4).

Теорема 2. *Для того чтобы некоторая последовательность $(u_m(t),x_m(t))$, $t_0\leqslant t\leqslant T$, $x_m(t_0)=x_{0m}$, $m=1,2,\ldots$, допустимых пар задачи (1.1)–(1.4) была минимизирующей, достаточно существования функции $K(x,t)$, для которой формула (1) верна для любой допустимой пары задачи (1.1)–(1.4) и*

$$\lim_{m\to\infty}\int_{t_0}^{T}S(x_m(t),u_m(t),t)dt=\int_{t_0}^{T}S_{\min}(t)dt, \tag{8}$$

$$\lim_{m\to\infty}s(x_m(T),T)=s_{\min},\quad \lim_{m\to\infty}K(x_{0m},t_0)=K_{\min}(t_0) \tag{9}$$

Доказательство. В оценке (2) вместо $\bar{u}(t)$, $\bar{x}(t)$ подставим $u_m(t)$, $x_m(t)$ и перейдем к пределу при $m\to\infty$. С учетом условий (8), (9) получим $\lim_{m\to\infty}J(t_0,x_{0m},u_m(\cdot))=J_*$, что и требовалось. □

Чтобы пользоваться приведенными в теоремах 1, 2 достаточными условиями оптимальности, нужно знать функцию $K(x,t)$, с помощью которой строятся функции (3)–(5). Как найти такую функцию $K(x,t)$, каким условиям она удовлетворяет?

Всякую функцию $K(x,t)$, удовлетворяющую условиям теоремы 1 [теоремы 2], назовем *функцией Кротова* задачи (1.1)–(1.4), соответствующей допустимой паре $(\bar{u}(t),\bar{x}(t))$, $t_0\leqslant t\leqslant T$ [или последовательности $(u_m(t),x_m(t))$ допустимых пар].

Заметим, что если существует какая-либо функция Кротова $K(x,t)$, то функцией Кротова является также функция $\overline{K}(x,t)=K(x,t)+\alpha(t)$, где $\alpha(t)$ — произвольная непрерывная, кусочно-гладкая (или абсолютно непрерывная) на $[t_0,T]$ функция. В частности, если $\alpha(t)=-\int_{T}^{t}S_{\min}(\tau)d\tau+s_{\min}$, то функции $\overline{S}(x,u,t)$, $\overline{s}(x,T)$, построенные по формулам (3), (4) с заменой K на $\overline{K}=K+\alpha$, таковы, что $\overline{S}(x,u,t)=S(x,u,t)-S_{\min}(t)$, $\overline{s}(x,T)=s(x,T)-s_{\min}$ и, следовательно, $\inf_{x\in X(t)}\inf_{u\in D(x,t)}\overline{S}(x,u,t)\equiv 0$, $t_0\leqslant t\leqslant T$, $\inf_{x\in X(T)}\overline{s}(x,T)=0$, а $\inf_{x\in X(t_0)}\overline{K}(x,t_0)=K_{\min}(t_0)+\alpha(t_0)$ отличается от $K_{\min}(t_0)$ на постоянную $\alpha(t_0)$. Поэтому в теоремах 1, 2 без ограничения общности можем принять $S_{\min}(t)\equiv 0$, $s_{\min}=0$.

С учетом этого замечания заключаем, что функция Кротова, соответствующая допустимой паре $(\bar{u}(t),\bar{x}(t))$ или последовательности $(u_m(t),x_m(t))$ допустимых пар задачи (1.1)–(1.4), согласно теоремам 1, 2 удовлетворяет условиям

$$S(x,u,t)=\langle K_x(x,t),f(x,u,t)\rangle+K_t(x,t)+f^0(x,u,t)\geqslant 0, \tag{10}$$

$$u\in D(x,t),\quad x\in X(t),\quad t_0\leqslant t\leqslant T;$$

$$s(x,T)=\Phi(x)-K(x,T)\geqslant 0,\quad x\in X(T),\quad K(x,t_0)\geqslant K_{\min}(t_0),\quad x\in X(t_0), \tag{11}$$

причем неравенства (10), (11) должны обратиться в равенства при $u=\bar{u}(t)$, $x=\bar{x}(t)$ или при $u=u_m(t)$, $x=x_m(t)$ в пределе при $m\to\infty$.

Задача (10), (11) для определения функции Кротова несколько необычна тем, что, во-первых, здесь мы имеем дело не с дифференциальным уравнением, а с дифференциальным неравенством (10) в частных производных, во-вторых, начальное условие при $t=T$ также за-

дано в виде неравенства и, наконец, задача (10), (11) тесно связана с конкретной допустимой парой $(u(\cdot), x(\cdot))$ или последовательностью $(u_m(\cdot), x_m(\cdot))$ допустимых пар, подозреваемых на оптимальность.

Сравнение соотношений (10), (11) с (3.5), (3.6), (3.14) показывает, что функция Беллмана всегда является функцией Кротова. С другой стороны, функция Кротова определяется из более широких условий (10), (11), и она может существовать даже тогда, когда функция Беллмана не существует.

В тех случаях, когда отсутствует удобное для работы конструктивное описание множеств $D(x, t)$, $X(t)$, функцию Кротова можно попытаться определить из условий, получающихся из (10), (11) при замене $D(x, t)$, $X(t)$ на $V(t)$, $G(t)$ соответственно.

Проиллюстрируем вышесказанное на примерах.

Пример 1. Пусть требуется минимизировать функцию $J(u(\cdot)) = \int_0^1 (x^2(t) - u(t))dt$ при условиях $\dot{x}(t) = u(t)$, $x(0) = x(1) = 0$, $u(t) \in V(t) = \{u \in E^1: |u| \leqslant 1\}$, $0 \leqslant t \leqslant 1$.

Здесь фазовые ограничения $G(t)$ такие: $G(0) = G(1) = \{0\}$, $G(t) = E^1$, $0 < t < 1$. Очевидно пара $(\bar{u}(t) \equiv 0, \bar{x}(t) \equiv 0)$ является допустимой для этой задачи. Покажем, что она является решением задачи. Для этого возьмем функцию $K(x, t) \equiv x$. Тогда

$$S(x, u, t) = x^2 \geqslant 0 = S(\bar{x}(t), \bar{u}(t), t) = S_{\min}(t),$$

$$s(x, 1) = -x = 0 = s(\bar{x}(1), 1) = \inf_{x \in G(1)} s(x, 1), \quad K(x, 0) = x = 0 = K(\bar{x}(0), 0) = \inf_{x \in G(0)} K(x, 0).$$

Кроме того, ясно, что формула (1) здесь будет верна для любой допустимой пары. Согласно теореме 1 тогда пара $(\bar{u}(t) \equiv 0, \bar{x}(t) \equiv 0)$ оптимальна. Предлагаем читателю найти функцию Беллмана и синтезирующую функцию этой задачи.

Пример 2. Пусть требуется минимизировать функцию $J(u(\cdot)) = \int_0^1 (x^2(t) - u^2(t))dt$ при условиях $\dot{x}(t) = u(t)$, $x(0) = 0$, $u(t) \in V(t) = \{u \in E^1: |u| \leqslant 1\}$, $0 \leqslant t \leqslant 1$.

Здесь фазовые ограничения $G(t)$ такие: $G(0) = \{0\}$, $G(t) \equiv E^1$, $0 < t \leqslant 1$. Возьмем последовательность пар функций $(u_m(\cdot), x_m(\cdot))$, где

$$u_m(t) = \begin{cases} 1, & \frac{p}{m} < t \leqslant \frac{p}{m} + \frac{1}{2m}, \\ -1, & \frac{p}{m} + \frac{1}{2m} < t \leqslant \frac{p}{m} + \frac{1}{m}, \end{cases} \qquad x_m(t) = \begin{cases} t - \frac{p}{m}, & \frac{p}{m} < t \leqslant \frac{p}{m} + \frac{1}{2m}, \\ -t + \frac{p+1}{m}, & \frac{p}{m} + \frac{1}{2m} < t \leqslant \frac{p}{m} + \frac{1}{m}, \end{cases}$$

$$p = 0, 1, \ldots, m-1, \quad m = 1, 2, \ldots$$

Нетрудно проверить, что пара $(u_m(\cdot), x_m(\cdot))$ допустима для рассматриваемой задачи при всех $m = 1, 2, \ldots$. Покажем, что последовательность этих пар является минимизирующей. Для этого возьмем функцию $K(x, t) = t - 1$. Тогда $S(x, u, t) = x^2 + 1 - u^2 \geqslant 0 = \min_{x \in E^1} \min_{|u| \leqslant 1} S(x, u, t) = \lim_{m \to \infty} S(x_m(t), u_m(t), t)$, $s(x, 1) = -K(x, 1) \equiv 0 = \inf_{x \in E^1} K(x, 1) = \lim_{m \to \infty} s(x_m(1), 1)$, $K(x, 0) \equiv -1 = \inf_{x \in G(0)} K(x, 0) = \lim_{m \to \infty} K(x_m(0), 0)$. Согласно теореме 2 последовательность $(u_m(\cdot), x_m(\cdot))$ будет минимизирующей: $\lim_{m \to \infty} (u_m(\cdot)) = -1 = J_*$.

Заметим, что в этом примере $\inf J(u) = -1$ не достигается ни на какой допустимой паре. В самом деле, если $x^2(t) \equiv 0$, то в силу уравнения $\dot{x} = u(t) \equiv 0$ и $J(0) = 0 > -1$. Если же $x^2(t) \not\equiv 0$, то $J(u) > -\int_0^1 u^2(t)dt \geqslant -1$. Таким образом, $J(u(\cdot) > -1$ для всех допустимых управлений и траекторий. В этой задаче мы имеем дело с так называемым *скользящим режимом* [57; 132; 212; 253; 254; 417; 418; 716]. Предлагаем читателю найти функцию Беллмана и приближенную синтезирующую функцию этой задачи.

2. Перейдем к рассмотрению следующей задачи оптимального управления с незакрепленным временем: минимизировать функцию

$$J(t_0, T, x_0, u(\cdot)) = \int_{t_0}^{T} f^0(x(t), u(t), t)dt + \Phi(x(T), T) \tag{12}$$

при условиях

$$\dot{x}(t) = f(x(t), u(t), t), \quad t_0 \leqslant t \leqslant T; \quad x(t_0) = x_0, \tag{13}$$

$$x(t) \in G(t), \quad t_0 \leqslant t \leqslant T; \quad x(T) \in S_1(T) \subseteq G(T), \tag{14}$$

$$u = u(t) \in V(t), \quad u(t) \text{ — кусочно-непрерывна}, \quad t_0 \leqslant t \leqslant T, \tag{15}$$

где моменты t_0, T, в отличие от задачи (1.1)–(1.4), неизвестные и таковы, что

$$t_0 \in \theta_0, \quad T \in \theta_1; \tag{16}$$

θ_0, θ_1 — некоторые множества на числовой оси $-\infty < t < \infty$; остальные обозначения см. в § 6.1. В частности, если $f^0 \equiv 1$, $\Phi \equiv 0$, то задача (12)–(16) превратится в задачу быстродействия. Всюду ниже будем предполагать, что $t_0 \leqslant T$.

Через $\Delta(x, t, T)$ обозначим множество всех управлений $u(\cdot)$, определенных на отрезке $[t_0, T]$, удовлетворяющих условиям (15) и таких, что траектория системы $\dot{x}(\tau) = f(x(\tau), u(\tau), \tau)$, $x(t) = x \in G(t)$ также определена на отрезке $[t, T]$, причем $x(\tau) \in G(\tau)$, $t \leqslant \tau \leqslant T$, $x(T) \in S_1(T)$. Положим $X(t, T) = \{x: x \in G(t), \Delta(x, t, T) \neq \varnothing\}$ при $t < T$; $X(T, T) \equiv S_1(T)$. Введем также множество $D(x, t, T)$ всех тех $u \in V(t)$, для которых существует хотя бы одно управление $u(\tau) \in \Delta(x, t, T)$ со значением $u(t) = u(t + 0) = u$. Пару $(u(t), x(t))$ назовем допустимой парой задачи (12)–(16), если функции $u(t)$, $x(t)$ определены на каком-либо отрезке $[t_0, T]$, где $t_0 \in \theta_0$, $T \in \theta_1$, и такие, что $x(t_0) = x_0 \in G(t_0)$, $u(\cdot) \in \Delta(x_0, t_0, T)$, $x(\cdot)$ — траектория системы (13). Если $(u(\tau), x(\tau))$, $t_0 \leqslant \tau \leqslant T$, является допустимой парой задачи (12)–(16), то $u(\cdot) \in \Delta(x(t), t, T)$, $x(t) \in X(t, T)$, $u(t) \in D(x(t), t, T)$ для всех t, $t_0 \leqslant t < T$.

Моменты времени $t_0^* \in \theta_0$, $T^* \in \theta_1$ и допустимую пару $(u_*(t), x_*(t))$, определенную на отрезке $[t_0^*, T^*]$, назовем решением задачи (12)–(16), если

$$J(t_0^*, T^*, x_*(t_0), u_*(\cdot)) = \inf_{t_0 \in \theta_0} \inf_{T \in \theta_1} \inf_{x_0 \in X(t_0, T)} \inf_{u(\cdot) \in \Delta(x_0, t_0, T)} J(t_0, T, x_0, u(\cdot)) = J_*.$$

Скажем, что последовательности моментов $\{t_{0m}\} \in \theta_0$, $\{T_m\} \in \theta_1$ и допустимых пар $(u_m(t), x_m(t))$, определенных на отрезке $[t_{0m}, T_m]$, являются минимизирующими для задачи (12)–(16), если $\lim_{m \to \infty} J(t_{0m}, T_m, x_m(t_{0m}), u_m(\cdot)) = J_*$.

Для формулировки достаточных условий оптимальности снова воспользуемся функциями $S(x, u, t)$, $s(x, T)$, определяемыми равенствами (3), (4), причем для случая рассматриваемой задачи (12)–(16) в (4) вместо $\Phi(x)$ будем брать $\Phi(x, T)$. Будем считать, что функции $K(x, t)$ такова, что формула (1) остается верной для любых допустимых пар задачи (12)–(16) — для этого достаточно, чтобы функция $K(x, y)$ была определена и непрерывна при всех $x \in G(t)$, $t \in [\inf \theta_0, \sup \theta_1]$, обладала кусочно-непрерывными K_x, K_t, а функция $K(x(t), t)$ переменной t для любой допустимой пары $(u(t), x(t))$, $t_0 \leqslant t \leqslant T$, задачи (12)–(16) была кусочно-гладкой на отрезке $[t_0, T]$.

Пусть $(u_*(t), x_*(t))$ и $(u(t), x(t))$ — какие-либо допустимые пары задачи (12)–(16), определенные на отрезках $[t_0^*, T^*]$ и $[t_0, T]$ соответственно. Тогда из формулы (1) получим

$$J(t_0^*, T^*, x_*(t_0^*), u_*(\cdot)) - J(t_0, T, x(t_0), u(\cdot)) = \int_{t_0^*}^{T^*} S(x_*(t), u_*(t), t)dt - \int_{t_0}^{T} S(x(t), u(t), t)dt +$$
$$+ [s(x_*(T^*), T^*) - s(x(T), T)] + [K(x_*(t_0^*), t_0^*) - K(x(t_0), t_0)]. \tag{17}$$

Обозначим

$$S_{\min} = \inf_{t_0 \in \theta_0} \inf_{T \in \theta_1} \int_{t_0}^{T} \inf_{x \in X(t, T)} \inf_{u \in D(x, t, T)} S(x, u, t)dt,$$
$$s_{\min} = \inf_{T \in \theta_1} \inf_{x \in S_1(T)} s(x, T), \quad K_{0\min} = \inf_{t_0 \in \theta_0} \inf_{T \in \theta_1} \inf_{x \in X(t_0, T)} K(x, t_0). \tag{18}$$

Учитывая, что для любой допустимой пары $(u(t), x(t))$, $t_0 \leqslant t \leqslant T$, задачи (12)–(16) имеют место включения $u(t) \in D(x(t), t, T)$, $x(t) \in X(t, T)$ при всех t, $t_0 \leqslant t < T$, и $u(\cdot) \in \Delta(x(t_0), t_0, T)$, $x(t_0) \in X(t_0, T)$, $x(T) \in X(T, T) = S_1(T)$, $t_0 \in \theta_0$, $T \in \theta_1$, из (17), (18) получим следующие неравенства:

$$0 \leqslant J(t_0^*, T^*, x_*(t_0^*), u_*(\cdot)) - J_* \leqslant \int_{t_0^*}^{T^*} S(x_*(t), u_*(t), t)dt - S_{\min} +$$
$$+ [s(x_*(T^*), T^*) - s_{\min}] + [K(x_*(t_0^*), t_0^*) - K_{0\min}]. \tag{19}$$

Неравенства (19) обобщают формулу (2) на случай задач оптимального управления с незакрепленным временем и представляют собой оценку погрешности, которая будет допущена,

если допустимую пару $(u_*(t), x_*(t))$, $t_0^* \leqslant t \leqslant T^*$, задачи (12)–(16) возьмем за приближенное решение этой задачи. Оценка (19) станет более конструктивной, если в формулах (18), определяющих величины $S_{\min}$, $s_{\min}$, $K_{0\min}$, множества $D(x, t, T)$, $X(t, T)$ заменим на $V(t)$, $G(t)$ соответственно.

Опираясь на оценку (19), сформулируем теперь достаточные условия оптимальности для задачи (12)–(16).

Теорема 3. *Для того чтобы допустимая пара $(u_*(t), x_*(t))$, $t_0^* \leqslant t \leqslant T^*$, задачи (12)–(16) была решением этой задачи, достаточно существования функции $K(x, t)$, для которой формула (1) верна для любой допустимой пары задачи (12)–(16) и*

$$\int_{t_0^*}^{T^*} S(x_*(t), u_*(t), t)dt = S_{\min}, \tag{20}$$

$$s(x_*(T^*), T^*) = s_{\min}, \quad K(x_*(t_0^*), t_0^*) = K_{0\min}. \tag{21}$$

Доказательство. При выполнении условий (20), (21) правая часть оценки (19) обращается в нуль, откуда и следует утверждение теоремы. □

Теорема 4. *Для того чтобы некоторая последовательность $(u_m(t), x_m(t))$, $t_{0m} \leqslant t \leqslant T_m$, $m = 1, 2\ldots$, допустимых пар задачи (12)–(16) была минимизирующей для этой задачи, достаточно существования функции $K(x, t)$, для которой формула (1) верна для любой допустимой пары задачи (12)–(16) и*

$$\lim_{m\to\infty} \int_{t_{0m}}^{T_m} S(x_m(t), u_m(t), t)dt = S_{\min}, \tag{22}$$

$$\lim_{m\to\infty} s(x_m(T_m), T_m) = s_{\min}, \quad \lim_{m\to\infty} K(x_m(t_{0m}), t_{0m}) = K_{0\min}. \tag{23}$$

Доказательство. В оценке (19) вместо $u_*(t)$, $x_*(t)$, t_0^*, T^* подставим соответственно $u_m(t)$, $x_m(t)$, t_{0m}, T_m и перейдем к пределу при $m \to \infty$. С учетом условий (22), (23) получим утверждение теоремы. □

Предлагаем читателю самостоятельно выписать условия, аналогичные условиям (10), (11), для определения функции Кротова $K(x, t)$ для задачи (12)–(16).

Пример 3. Требуется наибыстрейшим образом перевести точку $(x, y) \in E^2$ из начала координат $(0, 0)$ в точку $(1, 0)$, предполагая, что движение точки подчиняется условиям $\dot{x}(t) = -y^2(t) + u^2(t)$, $\dot{y}(t) = u(t)$, $u(t) \in V(t) = \{u \in E^1: |u| \leqslant 1\}$, $0 \leqslant t \leqslant T$.

Здесь $G(0) = \{(0, 0)\}$, $G(t) \equiv E^2$ при $0 < t \leqslant T$, $S_1(T) = \{(1, 0)\}$, $\theta_0 = \{t_0 = 0\}$, $\theta_1 = \{T: \geqslant 0\}$, $f^0 \equiv 1$, $\Phi \equiv 0$, $J(T, u(t)) = T$.

Пусть T_m — корень уравнения $(T-1)^3 = 12(T-1) - m^{-2}$, расположенный в пределах $1 < T < 1 + m^{-2/3}$, $m = 1, 2, \ldots$ Положим

$$u_m(t) = \begin{cases} 1 & \text{при } \frac{p}{m} < t \leqslant \frac{p}{m} + \frac{1}{2m} \text{ и } 1 \leqslant t \leqslant \frac{T_m + 1}{2}, \\ -1 & \text{при } \frac{p}{m} + \frac{1}{2m} < t \leqslant \frac{p}{m} + \frac{1}{m} \text{ и } \frac{T_m + 1}{2} < t \leqslant T_m, \end{cases}$$

$p = 0, 1, \ldots, m-1$, $m = 1, 2, \ldots$ Через $(x_m(t), y_m(t))$, $0 \leqslant t \leqslant T_m$, обозначим траекторию точки, соответствующую управлению $u_m(\cdot)$ и начальным условиям $x_m(0) = y_m(0) = 0$. Нетрудно видеть, что $x_m(T_m) = 1$, $y_m(T_m) = 0$, $(1 - m^{-4/3})t \leqslant x_m(t) \leqslant t$, $0 \leqslant y_m(t) \leqslant m^{-2/3}$ при $0 \leqslant t \leqslant T_m$, $m = 1, 2, \ldots$

Покажем, что $\lim\limits_{m\to\infty} T_m = 1 = T^*$ — оптимальное время. Для этого возьмем функцию $K(x, y, t) = -x$. Тогда

$$S(x, y, u, t) = y^2 - u^2 + 1, \quad \inf_{(x,y)\in E^2} \inf_{|u|\leqslant 1} S(x, y, u, t) \equiv 0$$

при всех $t \geqslant 0$, $S_{\min} = 0$, $\int_0^{T_m} S(x_m(t), y_m(t), u_m(t), t)dt = \int_0^{T_m} y_m^2(t)dt - \int_0^{T_m} (u_m^2(t) - 1)dt = \int_0^{T_m} y_m^2(t)dt \to$ $\to 0 = S_{\min}$ при $m \to \infty$; $s(x, y, T) = -K(x, y, T) = -x = -1$ при $(x, y) \in S_1(T)$, $s(x_m(T_m), y_m(T_m), T_m) = -1 = s_{\min}$, $K(x, y, 0) = 0$ при $(x, y) \in G(0)$, $K(x_m(0), y_m(0), 0) = 0 = K_{0\min}$, $m = 1, 2, \ldots$

Кроме того, очевидно, формула (1) для данной задачи будет справедлива при всех допустимых $(x(\cdot), y(\cdot), u(\cdot))$. Таким образом, для последовательностей $\{T_m\}$, $\{(x_m(t), y_m(t), u_m(t))\}$, $0 \leqslant t \leqslant T_m$, все условия теоремы 4 выполнены. Следовательно, $\lim\limits_{m\to\infty} J(T_m, u_m(\cdot)) = \lim\limits_{m\to\infty} T_m = 1$ — оптимальное время. Остается заметить, что в рассмотренной задаче $\inf J(T, u) = 1$ не достигается — здесь, как и в примере 2, мы имеем дело со скользящим режимом.

3. Ниже приведем еще одно достаточное условие оптимальности, касающееся задачи быстродействия.

Теорема 5. *Пусть в задаче (12)–(16) $f^0 \equiv 1$, $\Phi \equiv 0$, $\theta_0 = \{t_0\}$, т. е. начальный момент t_0 закреплен; $\theta_1 \equiv \{T: T \geqslant t_0\}$. Пусть имеется некоторая последовательность $(u_m(t), x_m(t))$, $t_0 \leqslant t \leqslant T_m$, $m = 1, 2, \ldots$, допустимых пар рассматриваемой задачи быстродействия, причем $\lim\limits_{m\to\infty} T_m = T^*$. Тогда для того чтобы T^* было оптимальным временем, достаточно существования функции $K(x, t)$, для которой формула (1) верна для любой допустимой пары u, кроме того,*

$$\lim_{m\to\infty} \int_{t_0}^{T_m} S(x_m(t), u_m(t), t)dt < \int_{t_0}^{T} S_{\min}(t, T)dt + T^* - T \tag{24}$$

для любого T, $t_0 \leqslant T < T^*$,

$$\lim_{m\to\infty} s(x_m(T_m), T_m) = s_{\min}, \quad \lim_{m\to\infty} K(x_m(t_0), t_0) = K_{0\min}, \tag{25}$$

где

$$S_{\min}(t, T) = \inf_{x\in X(t,T)} \inf_{u\in D(x,t,T)} S(x, u, t),$$

$$s_{\min} = \inf_{0\leqslant T\leqslant T^*} \inf_{x\in S_1(T)} s(x, T), \quad K_{0\min} = \inf_{0\leqslant T\leqslant T^*} \inf_{x\in X(t_0,T)} K(x, t_0).$$

Для получения формулировки достаточного условия оптимальности для фиксированной допустимой пары $(u_*(t), x_*(t))$, $t_0 \leqslant t \leqslant T^*$, в этой теореме надо принять $T_m = T^*$, $u_m(t) = u_*(t)$, $x_m(t) = x_*(t)$, $m = 1, 2, \ldots$, и в (24), (25) всюду опустить знак lim.

Доказательство. Пусть вопреки утверждению теоремы T^* не является оптимальным временем. Тогда существуют момент T, $t_0 \leqslant T < T^*$, и допустимая пара $(u(t), x(t))$, $t_0 \leqslant t \leqslant T$. В формуле (17) вместо t_0^*, T^*, $(u_*(t), x_*(t))$, $t_0^* \leqslant t \leqslant T^*$, примем соответственно t_0, T_m, $(u_m(t), x_m(t))$, $t_0 \leqslant t \leqslant T_m$, и перейдем к пределу при $m \to \infty$. С учетом условий (24), (25) будем иметь

$$\lim_{m\to\infty} J(t_0, T_m, x_m(t_0), u_m(\cdot)) - J(t_0, T, x(t_0), u(\cdot)) = T^* - T \leqslant$$
$$\leqslant \lim_{m\to\infty} \int_{t_0}^{T_m} S(x_m(t), u_m(t), t)dt - \int_{t_0}^{T} S(x(t), u(t), t)dt < T^* - T.$$

Полученное противоречивое неравенство доказывает теорему. □

Пример 4. Пусть требуется наибыстрейшим образом перевести точку $(x, y) \in E^2$ из положения $(1, 0)$ в начало координат $(0, 0)$, предполагая, что движение точки подчиняется условиям $\dot{x}(t) = y(t)$, $\dot{y}(t) = u(t)$, $u(t) \in V(t) = \{u \in E^1: |u| \leqslant 1\}$, $0 \leqslant t \leqslant T$.

Здесь $G(0) = \{(1, 0)\}$, $G(t) \equiv E^2$ при $0 \leqslant t < T$, $S_1(T) = \{(0, 0)\}$, $\theta_0 = \{t_0 = 0\}$, $\theta_1 = \{T \geqslant 0\}$.

В примере 6.2.4 с помощью принципа максимума была найдена допустимая пара $(x_*(t), y_*(t), u_*(t))$, $0 \leqslant t \leqslant T^* = 2$, где

$$u_*(t) = \begin{cases} -1, & 0 \leqslant t \leqslant 1, \\ 1, & 1 < t \leqslant 2, \end{cases} \quad x_*(t) = \begin{cases} 1 - t^2/2, & 0 \leqslant t \leqslant 1, \\ (t-2)^2/2, & 1 \leqslant t \leqslant 2, \end{cases} \quad y_*(t) = \begin{cases} -t, & 0 \leqslant t \leqslant 1, \\ t - 2, & 1 < t \leqslant 2. \end{cases}$$

Покажем, что эта пара является решением рассматриваемой задачи быстродействия. Возьмем функцию $K(x, y, t) = x - (t-1)y$. Тогда $S(x, y, u, t) = K_x y + K_y u + K_t + 1 = -(t - 1)u + 1$, $S_{\min}(t, T) = \inf\limits_{(x,y)\in E^2} \inf\limits_{|u|\leqslant 1} S(x, y, u, t) = -|t-1| + 1$ при всех T и $T \geqslant 0$, $S(x_*(t), y_*(t), u_*(t), t) = -|t-1| + 1$; $\int_0^{T^*} S(x_*(t), y_*(t), u_*(t), t)dt - \int_0^T S_{\min}(t, T)dt = \int_T^2 (1 - |t-1|)dt < 2 - T = T^* - T$ для всех T, $0 \leqslant T < T^* = 2$; $s(x, y, T) = -K(x, y, T) = 0$ при $(x, y) \in S_1(T)$, $s(x_*(T^*), y_*(T^*), T^*) = s_{\min}$; $K(x, y, 0) = 1$ при $(x, y) \in G(0) = X(0, T) = \{(1, 0)\}$, $K(x_*(0), y_*(0), 0) = 1 = K_{0\min}$.

Кроме того, очевидно, формула (1) для данной задачи будет справедлива при всех допустимых управлениях и траекториях. В силу теоремы 5 момент $T^* = 2$ и пара $(x_*(\cdot), y_*(\cdot), u_*(\cdot))$ являются оптимальными. Заметим, что функция Беллмана в этой задаче имеет разрывы первой производной именно на оптимальной траектории [105; 587], в то время как функция Кротова $K(x, t)$ является просто многочленом.

В заключение упомянем, что функции Беллмана, Кротова тесно связаны с функцией Ляпунова, широко используемой в теории устойчивости [328].

Упражнения

1. Рассмотреть задачу минимизации функции $J(u(\cdot)) = \int_0^T (u^2(t) - x^2(t))dt$ при условиях $x(0) = x(T) = 0$. Показать, что пара $(u_*(t) \equiv 0,\ x_*(t) \equiv 0)$ является оптимальной при $0 < T < \pi$. Указание: функцию Кротова искать в виде $K(x, t) = \psi(t)x^2$.

2. С помощью принципа максимума найти подозрительные на оптимальность управления и траектории, а затем доказать их оптимальность для следующей задачи быстродействия: наибыстрейшим образом перевести точку (x_0, y_0) из заданного состояния в начало координат $(0, 0)$, предполагая, что движение точки подчиняется одному из следующих условий:

а) $\dot{x}(t) = y(t),\ \dot{y}(t) = u(t),\ u(t) \in V(t) = \{u \in E^1 : |u| \leqslant 1\},\ 0 \leqslant t \leqslant T$;

б) $\dot{x}(t) = y(t),\ \dot{y}(t) = -x(t) + u(t),\ u(t) \in V(t) = \{u \in E^1 : |u| \leqslant 1\},\ 0 \leqslant t \leqslant T$;

в) $\dot{x}(t) = y(t) + u(t),\ \dot{y}(t) = -x(t) + v(t),\ (u(t), v(t)) \in V(t) = \{(u, v) \in E^2 : |u| \leqslant 1, |v| \leqslant 1\},\ 0 \leqslant t \leqslant T$. Указание: функцию Кротова искать в виде $K(x, t) = \psi_1(t)x + \psi_2(t)y$.

3. Перевести точку $(x, y, z) \in E^3$ из начала координат $(0, 0, 0)$ в точку $(a, 0, 0)$ быстрейшим образом, если $\dot{x}(t) = y(t),\ \dot{y}(t) = z(t),\ \dot{z}(t) = u(t),\ u(t) \in V(t) = \{u \in E^1 : |u| \leqslant 1\},\ 0 \leqslant t \leqslant T$; $a = \text{const}$. Показать, что оптимальное время $T^* = (32|a|)^{1/3}$. Указание: функцию Кротова искать в виде $K(x, t) = \psi_1(t)x + \psi_2(t)y + \psi_3(t)z$.

4. Рассмотреть задачу минимизации функции

$$J(t_0, T, x_0, u(\cdot)) = \int_{t_0}^{T} f^0(x(t), u(t), t)dt + \Phi_0(x(t_0), t_0) + \Phi(x(T), T)$$

при условиях (13)–(16). Для этой задачи сформулировать и доказать теоремы, аналогичные теоремам 1–4.

Часть II

МИНИМИЗАЦИЯ В ФУНКЦИОНАЛЬНЫХ ПРОСТРАНСТВАХ. РЕГУЛЯРИЗАЦИЯ. АППРОКСИМАЦИЯ

ГЛАВА 8

Методы минимизации в функциональных пространствах

В части I книги мы занимались задачами минимизации функций конечного числа переменных и задачами оптимального управления процессами, описываемыми системами обыкновенных дифференциальных уравнений. Наряду с этими задачами большой интерес для практики представляют задачи оптимального управления процессами, описываемыми уравнениями с частными производными, интегро-дифференциальными уравнениями, задачи наилучшего приближения функций и др. Оказывается, все перечисленные задачи можно трактовать как экстремальные задачи в подходящим образом выбранных функциональных пространствах, и для исследования этих задач использовать аппарат и методы функционального анализа. Такая трактовка позволяет выявить общие закономерности, присущие широким классам экстремальных задач, создавать и исследовать общие методы решения таких задач. Эти проблемы, а также вопросы аппроксимации и регуляризации экстремальных задач в функциональных пространствах составляют основное содержание части II книги.

В главе 8 мы кратко остановимся на элементах теории экстремальных задач в гильбертовых и банаховых пространствах, на методах их решения, рассмотрим некоторые классы задач оптимального управления процессами, описываемыми обыкновенными дифференциальными уравнениями и уравнениями с частными производными. Определение многих понятий, характеризующих задачи оптимизации (локальный и глобальный минимум, верхняя и нижняя грань функции, минимизирующая последовательность и т. п.), многих понятий выпуклого анализа (выпуклое множество, выпуклая функция, проекция точки, субградиент и т. п.) получаются из определений, приведенных в главах 1, 2, 4, нужно лишь в них под точкой теперь понимать элементы рассматриваемых банаховых и гильбертовых пространств, а вместо $|u|$, $\langle u, v \rangle$ понимать соответственно норму, скалярное произведение в этих пространствах. Поэтому мы здесь, как правило, не будем заново воспроизводить определения таких понятий и ограничимся ссылками на часть I книги. Многие теоремы, справедливые в конечномерных пространствах, без изменений остаются справедливыми и в бесконечномерных функциональных пространствах, и в таких случаях на соответствующее утверждение мы будем ссылаться в его прежней формулировке, указывая в контексте, о каком пространстве теперь идет речь. Следует предупредить неопытного читателя, что имеется немало утверждений, справедливых лишь в конечномерных пространствах, и их обобщение на бесконечномерные пространства требует определенной аккуратности и осторожности. В таких случаях мы будем приводить точные формулировки соответствующих утверждений, иллюстрировать их примерами и контрпримерами.

Напомним, что в главах 6, 7 мы уже рассматривали задачи оптимального управления, в которых управление принадлежит бесконечномерному функциональному пространству, но значения фазовой траектории в каждый фиксированный момент времени являются точкой конечномерного пространства. В задачах оптимального управления процессами, описываемыми уравнениями с частными производными, значения траекторий будут элементами бесконечномерных функциональных пространств. Для сохранения связи с главами 6, 7 мы далее в основном будем придерживаться обозначений из этих глав: целевую функцию (функционал) будем обозначать через $J(u)$, множество, на котором ищется экстремум этой функции — через U, элементы множества U — через u, фазовые траектории — через x, пространственную переменную — через s, а время, как обычно, будем обозначать через t.

Для понимания содержания излагаемого ниже материала достаточно знания начальных глав функционального анализа и элементов теории функций действительных переменных [258; 350; 357; 371; 393; 444; 705; 768]. Впрочем, заметим, что рассмотрение конкретных классов задач оптимального управления в главах 8–10 в основном ведется в терминах, связанных с этими задачами, и для понимания не требует знания элементов функционального анализа.

§ 1. Предварительные сведения. Обозначения

Здесь мы не будем приводить определения линейных, метрических, нормированных, банаховых и гильбертовых пространств — эти определения, а также основные свойства этих пространств читатель может найти в [393]. Ограничимся рассмотрением лишь вещественных банаховых и гильбертовых пространств, не оговаривая этого в дальнейшем. Элементы этих пространств часто будем называть точкой или вектором. Норму элемента в банаховом пространстве B будем обозначать через $\|u\|_B$, скалярное произведение двух элементов u, v из гильбертова пространства H — через $\langle u, v\rangle_H$. Напоминаем, что всякое гильбертово пространство H является банаховым пространством с нормой $\|u\|_H = (\langle u, v\rangle_H)^{1/2}$. Во всяком банаховом пространстве B можно ввести метрику, взяв в качестве расстояния $\rho(u, v)$ между точками $u, v \in B$ величину $\rho(u, v) = \|u - v\|_B$. В тех случаях, когда ясно, о каком банаховом или гильбертовом пространстве идет речь, знаки B и H в обозначениях $\|u\|_B$, $\langle c, u\rangle_H$ будем опускать и писать просто $\|u\|$, $\langle c, u\rangle$. Всюду ниже такие понятия, как ограниченность, сходимость, замкнутость, полунепрерывность сверху или снизу, компактность, будут пониматься в сильном смысле, т. е. в смысле нормы или метрики рассматриваемых банаховых пространств. Если эти понятия будут употребляться в слабом смысле, то будем говорить о слабой сходимости, слабой замкнутости, слабой полунепрерывности сверху или снизу, слабой компактности. Определение некоторых из этих понятий мы приведем и кратко поясним ниже по мере необходимости.

Кратко остановимся на понятии отображения. Пусть X и Y — два произвольных множества. Говорят, что на X определено *отображение*, если каждому элементу $x \in X$ поставлен в соответствие некоторый однозначно определяемый элемент $y \in Y$. Для обозначения отображения F из X в Y часто пользуются записью $y = F(x)$ или $y = Fx$ или $F\colon X \to Y$. В зависимости от того, какова природа множеств X и Y, вместо общего термина «отображение» в соответствии с установившимися традициями часто употребляются термины «функция», «функционал», «оператор» и т. д. В частности, если Y представляет собой множество на числовой оси E^1, то отображение $F\colon X \to E^1$ часто называют *функцией*. В классическом вариационном исчислении, когда в роли X выступают различные функциональные пространства, вместо термина «функция» часто употребляют термин «функционал». Мы ниже будем отождествлять термины «функция» и «функционал» — это позволит нам без изменения формулировок пользоваться многими определениями и теоремами из части I и в тех случаях, когда X представляет собой множество из метрического или банахова пространства.

Через B^* будем обозначать пространство, сопряженное к банахову пространству B. Напоминаем, что B^* состоит из линейных ограниченных функций (функционалов), определенных на B. Значение линейной функции $c \in B^*$ в точке $u \in B$ будем обозначать через $\langle c, u\rangle_B$ или $\langle c, u\rangle$. По определению, линейная ограниченная функция c такова, что

$$\langle c, \alpha u + \beta v\rangle = \alpha\langle c, u\rangle + \beta\langle c, v\rangle, \quad |\langle c, u\rangle| \leqslant M\|u\|$$

при всех $u, v \in B$ и всех вещественных числах α, β; M — неотрицательная постоянная, зависящая от функции c, но не зависящая от $u \in B$. Сопряженное пространство B^* само является банаховым с нормой $\|c\|_{B^*} = \sup\langle c, u\rangle_B$,

где верхняя грань берется по единичному шару $\|u\|_B \leqslant 1$. Отсюда следует, что $|\langle c, u\rangle_B| \leqslant \|c\|_{B^*}\cdot\|u\|_B$ при всех $u \in B$, $c \in B^*$.

Если H — гильбертово пространство, то для всякой линейной ограниченной функции на H найдется элемент $c \in H$ такой, что значение этой функции в любой точке $u \in H$ можно представить в виде скалярного произведения $\langle c, u\rangle_H$ [393]. Поэтому пространство H^*, сопряженное к гильбертову пространству H, можно отождествить с самим H, причем такое отождествление будет изометричным, т. е. $\|c\|_{H^*} = \sup\limits_{\|u\|_H \leqslant 1} \langle c, u\rangle_H = \|c\|_H$. Последнее равенство вытекает из *неравенства Коши — Буняковского*

$$|\langle u, v\rangle_H| \leqslant \|u\|_H\|v\|_H, \quad u, v \in H.$$

Гиперплоскостью в банаховом пространстве B называют множество

$$\Gamma = \{u\colon \langle c, u\rangle = \gamma\},$$

где $c \neq 0$ — фиксированный элемент из B^*, называемый *нормальным вектором гиперплоскости*, а γ — некоторое вещественное число.

Если X и Y — два банаховых пространства, то прямое произведение $B = X \times Y$ также является банаховым пространством с нормой $\|u\|_B = \|x\|_X + \|y\|_Y$ элемента $u = (x, y) \in B$, и сопряженное к B пространство B^* представимо в виде $B^* = X^* \times Y^*$.

В банаховых пространствах наряду с понятием сходимости по норме или, как еще говорят, сильной сходимости, важную роль играет понятие слабой сходимости. Напомним

Определение 1. Говорят, что последовательность $\{u_k\}$ из банахова пространства B *сходится к точке* $u \in B$ *слабо* в B, если

$$\lim_{k\to\infty}\langle c, u_k\rangle = \langle c, u\rangle \quad \text{при всех} \quad c \in B^*.$$

Если последовательность $\{u_k\}$ сходится к точке u сильно в B, т. е. $\lim\limits_{k\to\infty}\|u_k - u\| = 0$, то $\{u_k\}$ сходится к той же точке также и слабо в B, так как

$$|\langle c, u_k\rangle - \langle c, u\rangle| = |\langle c, u_k - u\rangle| \leqslant \|c\|_B\|u_k - u\| \to 0$$

при $k \to \infty$. Обратное неверно: из слабой сходимости последовательности, вообще говоря, не следует ее сильная сходимость.

Пример 1. Пусть H — гильбертово пространство, пусть $\{e_k\}$ — некоторая бесконечная ортонормированная система в H, т. е. $\langle e_i, e_k\rangle = 0$ при $i \neq k$ и $\langle e_k, e_k\rangle = 1$, где $i, k = 1, 2, \ldots$. Возьмем произвольный элемент $c \in H^* = H$. Тогда числа $c_k = \langle c, e_k\rangle$, $k = 1, 2, \ldots$, представляют собой коэффициенты Фурье элемента c по системе $\{e_k\}$. Согласно неравенству Бесселя ([393], стр. 151) $\sum\limits_{k=1}^{\infty} c_k^2 \leqslant \|c\|^2$, т. е. ряд $\sum\limits_{k=1}^{\infty} c_k^2$ сходится. Тогда $\lim\limits_{k\to\infty} c_k = \lim\limits_{k\to\infty}\langle c, e_k\rangle = 0 = \langle c, 0\rangle$ при всех $c \in H$. Это значит, что последовательность $\{e_k\}$ слабо в H сходится к нулю. Однако $\|e_k - e_m\|^2 = 2$ при любых $k \neq m$, поэтому последовательность $\{e_k\}$ не является фундаментальной в H и не может сильно сходиться в H.

В частности, пусть $H = L_2[a, b]$ — пространство Лебега функций $u = u(t)$, $a \leqslant t \leqslant b$, с нормой $\|u\|_{L_2} = \left(\int\limits_a^b |u(t)|^2 dt\right)^{1/2}$ и со скалярным произ-

ведением $\langle u, v\rangle_{L_2} = \int_a^b u(t)v(t)dt$. Тогда ортонормированные системы $\left\{e_k = \sqrt{\frac{2}{b-a}}\sin\frac{\pi k(t-a)}{b-a}\right\}$, $\left\{e_k = \sqrt{\frac{2}{b-a}}\cos\frac{\pi k(t-a)}{b-a}\right\}$ слабо в $L_2[a, b]$ сходятся к нулю, т. е. $\int_a^b c(t)e_k(t)dt \to 0$ при $k \to \infty$ для любой функции $c(t) \in L_2[a, b]$.

Так как сопряженное пространство B^* само является банаховым, то в свою очередь можно рассматривать второе сопряженное пространство $(B^*)^* = B^{**}$, состоящее из линейных ограниченных функций на B^*. Каждому элементу $u \in B$ можно поставить в соответствие линейную ограниченную функцию $\langle c, u\rangle$ переменной $c \in B^*$, т. е. некоторый элемент из B^{**}. Оказывается, это соответствие таково, что норма $\|u\|_B$ совпадает с нормой порожденной им функции $\langle c, u\rangle$, $c \in B^*$. Поэтому, отождествляя элемент из B с порожденной им функцией из B^{**}, получаем изометричное вложение пространства B в пространство B^{**}. В общем случае указанное вложение $B \subset B^{**}$ является строгим, т. е. возможно, что $B \neq B^{**}$. В тех случаях, когда это вложение таково, что $B = B^{**}$, банахово пространство B называется *рефлексивным*. Всякое гильбертово пространство H рефлексивно, так как $H = H^* = H^{**}$ [393; 705].

Отображение $A\colon X \to Y$, где X, Y — банаховы пространства, называют *линейным* оператором, если $A(\alpha x+\beta y)=\alpha Ax+\beta Ay$ для всех $x, y \in X$ и всех вещественных чисел α, β. Линейный оператор $A\colon X \to Y$ называется *ограниченным*, если существует постоянная $M \geqslant 0$ такая, что $\|Ax\|_Y \leqslant M\|x\|_X$ для всех $x \in X$. Если для каждого линейного ограниченного оператора A определить норму $\|A\| = \sup_{\|x\|_X \leqslant 1} \|Ax\|_Y$, то линейное пространство таких операторов превращается в банахово пространство, которое принято обозначать через $\mathcal{L}(X \to Y)$. Для каждого оператора $A \in \mathcal{L}(X \to Y)$ равенство

$$\langle c, Ax\rangle = \langle A^*c, x\rangle, \quad x \in X, \quad c \in Y^*$$

однозначно определяет оператор $A^* \in (Y^* \to X^*)$, называемый *сопряженным* к оператору A. Можно показать, что $\|A^*\| = \|A\|$ [393; 705].

Если $X = Y = H$ — гильбертово пространство, то $H = H^* = H^{**}$ и при каждом $A \in \mathcal{L}(H \to H)$ сопряженный оператор A^*, определяемый равенством $\langle Au, v\rangle_H = \langle u, A^*v\rangle_H$, также действует из H в H. Поэтому здесь возможно равенство $A = A^*$ — такой оператор A называют *самосопряженным*.

Приведем определения и обозначения некоторых конкретных банаховых и гильбертовых пространств, которые нам понадобятся в дальнейшем.

В конечномерном линейном вещественном пространстве R^n точек $u = (u^1, \ldots, u^n)$ наряду с евклидовой нормой $|u| = \left(\sum_{i=1}^n |u^i|^2\right)^{1/2}$ могут быть введены различные другие нормы. Например, полагая $|u|_p = \left(\sum_{i=1}^n |u^i|^p\right)^{1/p}$ при $1 \leqslant p < \infty$ или $|u|_\infty = \max_{1 \leqslant i \leqslant n} |u^i|$, получим различные банаховы пространства R_p^n, $1 \leqslant p \leqslant \infty$. Пространства R_p^n и R_q^n, где $p^{-1} + q^{-1} = 1$ при $1 < p < \infty$, $q = 1$ при $p = \infty$ и $q = \infty$ при $p = 1$, являются взаимно сопряженными. В частности, $(R_2^n)^* = R_2^n = E^n$. Заметим, что все нормы в R^n эквивалентны, т. е. если $\|u\|_I$ и $\|u\|_{II}$ — какие-либо две нормы в R^n, то найдутся числа $c_1, c_2 > 0$ такие, что $c_1\|u\|_I \leqslant \|u\|_{II} \leqslant c_2\|u\|_I$ при всех $u \in R^n$. Заметим также, что в любом конечномерном банаховом пространстве понятия сильной и слабой сходимости равносильны.

Через l_p, $1 \leqslant p < \infty$, будем обозначать банахово пространство последовательностей $u = (u^1, \ldots, u^k, \ldots)$ с конечной нормой $\|u\|_{l_p} = \left(\sum_{i=1}^\infty |u^k|^p\right)^{1/p}$. В случае $p = \infty$ под l_∞ понимают банахово пространство последовательностей $u = (u^1, \ldots, u^k, \ldots)$ с конечной нормой $\|u\|_{l_\infty} = \sup_k |u^k|$. Можно показать, что $\lim_{p \to \infty} \|u\|_{l_p} = \|u\|_{l_\infty}$ для всех $u \in l_\infty$. Сопряженным для l_p, $1 \leqslant p < \infty$, пространством является пространство l_q, где p, q связаны равенством $p^{-1} + q^{-1} = 1$ при $1 < p < \infty$ и $q = +\infty$ при $p = 1$. Описание сопряженного к l_∞ пространства см. в [258; 371]. Пространство l_p при $1 < p < \infty$ рефлексивно. Пространство l_2 является гильбертовым со скалярным произведением $\langle u, v\rangle_{l_2} = \sum_{i=1}^\infty u^i v^i$ и с нормой $\|u\|_{l_2} = (\langle u, u\rangle)^{1/2}$.

Пусть G — некоторое фиксированное измеримое по Лебегу множество из евклидова пространства E^n. Через $L_p^r(G)$, где $1 \leqslant p < \infty$, r — целое положительное число, будем обозначать банахово пространство измеримых вектор-функций $u = u(t) = (u^1(t), \ldots, u^r(t))$, $t \in G$, с конечной нормой

$$\|u\|_p = \left(\int_G |u(t)|_{E^r}^p dt\right)^{1/p}.$$

Если $p = \infty$, то через $L_\infty^r(G)$ будем обозначать банахово пространство ограниченных измеримых вектор-функций $u = u(t) = (u^1(t), \ldots, u^r(t))$ с нормой

$$\|u\|_{L_\infty} = \operatorname*{ess\,sup}_{t \in G} |u(t)|_{E^r} = \inf_v \sup_{t \in G} |v(t)|_{E^r},$$

где $v = v(t)$ пробегает множество всех измеримых вектор-функций, совпадающих с $u(t)$ почти всюду на G. Можно показать, что $\lim_{p \to \infty} \|u\|_{L_p} = \|u\|_{L_\infty}$ для всех $u \in L_\infty(G)$. Если $r = 1$, то вместо $L_p^r(G)$ будем писать просто $L_p(G)$, $1 \leqslant p \leqslant +\infty$. Если $p = 2$, то пространство $L_2^r(G)$ является гильбертовым пространством со скалярным произведением

$$\langle u, v\rangle_{L_2} = \int_G \langle u(t), v(t)\rangle_{E^r} dt = \int_G \left(\sum_{i=1}^r u^i(t)v^i(t)\right) dt;$$

тогда $\|u\|_{L_2}^2 = \langle u, u\rangle_{L_2}$. Пространство $L_p^r(G)$ при $1 < p < \infty$ является рефлексивным, а при $p = 1$ и $p = \infty$ оно нерефлексивно. Сопряженным для $L_p^r(G)$, $1 < p < \infty$, является пространство $L_q^r(G)$, где $1 < q < \infty$, $p^{-1} + q^{-1} = 1$, для $L_1^r(G)$ сопряженным является пространство $L_\infty^r(G)$; описание сопряженного пространства для $L_\infty^r(G)$ см. в [258; 371].

Через $C(G)$ будем обозначать банахово пространство непрерывных на замкнутом множестве G функций с нормой $\|u\|_C = \max_{t \in G} |u(t)|$; это пространство нерефлексивно; описание сопряженного к нему пространства см. в [258; 371].

Пусть множество G из E^n имеет непустую внутренность. Через $C^\infty(G)$ будем обозначать множество функций, бесконечно дифференцируемых на множестве G. Говорят, что функция $f(s) = f(s_1, \ldots, s_n) \in L_1(G)$ имеет *обобщенную производную* $\partial f(s)/\partial s_i = f_{s_i}(s)$ по переменной s_i в G, если $f_{s_i}(s) \in$

$\in L_1(G)$ и $\int\limits_G \varphi(s) f_{s_i}(s)ds = -\int\limits_G \varphi_{s_i}(s) f(s)ds$ для любой функции $\varphi(s) \in C^\infty(G)$, обращающейся в нуль в некоторой приграничной полосе множества G; здесь $\varphi_{s_i}(s)$ — частная производная $\varphi(s)$ по переменной s_i [441; 492; 648].

Через $H^1(G)$ (или $W_2^1(G)$) принято обозначать гильбертово пространство функций $f(s) \in L_2(G)$, обладающих обобщенными производными $f_{s_i}(s) \in L_2(G)$ по всем переменным $s_1, \ldots, s_n$, причем скалярное произведение в этом пространстве определяется так:

$$\langle f, g\rangle_{H^1} = \int\limits_G \Big(f(s)g(s) + \sum_{i=1}^{n} f_{s_i}(s) g_{s_i}(s) \Big) ds,$$

а норма имеет вид $\|f\|_{H^1} = (\langle f, f\rangle_{H^1})^{1/2}$.

Через $H^m(G)$ (или $W_2^m(G)$) обозначают гильбертово пространство функций $f(s) \in L_2(G)$, обладающих всеми обобщенными частными производными до порядка m включительно, принадлежащими $L_2(G)$; скалярное произведение в $H^m(G)$ определяется равенством

$$\langle f, g\rangle_{H^m} = \int\limits_G \sum_{0 \leqslant m_1 + \ldots + m_n \leqslant m} \frac{\partial^{m_1+\ldots+m_n} f(s)}{\partial s_1^{m_1} \ldots \partial s_n^{m_n}} \frac{\partial^{m_1+\ldots+m_n} g(s)}{\partial s_1^{m_1} \ldots \partial s_n^{m_n}} ds,$$

а норма имеет вид $\|f\|_{H^m} = (\langle f, f\rangle_{H^m})^{1/2}$ [492; 648; 649].

Ниже нам понадобятся пространства $H_r^m(G)$, $m \geqslant 1$, представляющие собой обобщение пространств $H^m(G)$ на случай r-мерных вектор-функций. Приведем соответствующие определения для случая, когда $G = [a, b] = \{t \in E^1: a \leqslant t \leqslant b\}$, $a < b$. Через $H_r^m[a, b]$ обозначим гильбертово пространство вектор-функций $u = u(t) = (u^1(t), \ldots, u^r(t)) \in L_2^r[a, b]$, обладающих обобщенными производными $\frac{d^i u(t)}{dt^i} = \Big(\frac{d^i u^1(t)}{dt^i}, \ldots, \frac{d^i u^r(t)}{dt^i}\Big)$, $i = 1, \ldots, m$, принадлежащими $L_2^r[a, b]$; скалярное произведение в этом пространстве определяется равенством

$$\langle u, v\rangle_{H_r^m} = \int\limits_a^b \Big(\langle u(t), v(t)\rangle_{E^r} + \sum_{i=1}^{m} \langle \frac{d^i u(t)}{dt^i}, \frac{d^i v(t)}{dt^i}\rangle_{E^r} \Big) dt,$$

норма равна

$$\|u\|_{H_r^m} = (\langle u, u\rangle_{H_r^m})^{1/2} = \Big(\int\limits_a^b \Big(|u(t)|_{E^r}^2 + \sum_{i=1}^{m} \Big| \frac{d^i u(t)}{dt^i} \Big|_{E^r}^2 \Big) dt \Big)^{1/2}.$$

Удобно считать, что $H_r^0[a, b] = L_2^r[a, b]$. Можно показать [95; 535; 649], что если $u(t) \in H_r^m[a, b]$, $m \geqslant 1$, то $u(t)$, $\frac{du(t)}{dt}, \ldots, \frac{d^{m-1}u(t)}{dt^{m-1}}$ представляют собой абсолютно непрерывные вектор-функции на отрезке $[a, b]$.

При $r = 1$ пространство $H_r^m[a, b]$, как и выше, будем обозначать просто $H^m[a, b]$. Подпространство функций из $H^m[a, b]$ со свойством $\frac{d^i u(a)}{dt^i} = 0$, $\frac{d^i u(b)}{dt^i} = 0$, $i = 1, \ldots, m-1$, будем обозначать через $H_0^m[a, b]$. Подпространство $H_0^m[a, b]$ замкнуто в $H^m[a, b]$ и поэтому само является гильбертовым пространством. В $H_0^m[a, b]$ наряду со скалярным произведением и нормой, индуцированными из $H^m[a, b]$, можно ввести скалярное произведение $\langle u, v\rangle_{H_0^m} = \int\limits_a^b \frac{d^m u(t)}{dt^m} \frac{d^m v(t)}{dt^m} dt$ и норму $\|u\|_{H_0^m} = \Big(\int\limits_a^b \Big(\frac{d^m u(t)}{dt^m} \Big)^2 dt \Big)^{1/2}$, эквивалентную норме $\|u\|_{H^m}$, т. е. $c_1 \|u\|_{H^m} \leqslant \|u\|_{H_0^m} \leqslant \|u\|_{H^m}$, $c_1 = \text{const} > 0$.

Пополнение $L_2[a, b]$ в норме $\sup\limits_{\varphi \in H_0^1[a, b]} \frac{\langle f, \varphi\rangle_{L_2[a, b]}}{\|\varphi\|_{H_0^1}}$ обозначают через $H^{-1}[a, b]$. Можно показать [458; 557], что получающееся гильбертово пространство $H^{-1}[a, b]$ изометрично пространству $(H_0^1[a, b])^*$, поэтому можно считать, что $H^{-1}[a, b] = (H_0^1[a, b])^*$. Пространство $H^{-1}[a, b]$ состоит из обобщенных функций (распределений), являющихся производными функций из $L_2[a, b]$ в смысле обобщенных функций (например, известная δ-функция Дирака, являющаяся производной функции скачка Хевисайда, принадлежит $H^{-1}[a, b]$).

Пусть $Q = G \times \{0 \leqslant t \leqslant T\}$, $G \subseteq E^n$, T — заданное положительное число. Через $H^{m,q}(Q)$ будем обозначать пространство функций $f(s, t) \in L_2(Q)$, обладающих обобщенными частными производными $\frac{\partial^{i_1+\ldots+i_n} f(s, t)}{\partial s_1^{i_1} \ldots s_n^{i_n}} \in L_2(Q)$; $0 \leqslant i_1 + \ldots + i_n \leqslant m$, $\frac{\partial^i f(s, f)}{\partial t^i} \in L_2(Q)$, $i = 1, \ldots, q$; это пространство является гильбертовым со скалярным произведением

$$\langle f, g\rangle_{H^{m,q}} = \int\limits_0^T \langle f(\cdot, t), g(\cdot, t)\rangle_{H^m} dt + \iint\limits_Q \sum_{i=1}^{q} \frac{\partial^i f(s, f)}{\partial t^i} \frac{\partial^i g(s, f)}{\partial t^i} ds dt$$

и нормой $\|f\|_{H^{m,q}} = (\langle f, g\rangle_{H^{m,q}})^{1/2}$.

При постановках краевых задач для уравнений с частными производными и связанных с ними задач оптимального управления важное значение имеет понятие следа функции, обобщающее понятие значения функции для классов разрывных функций. Мы здесь ограничимся следующим определением (более общие определения см. в [95; 492]).

Определение 2. Пусть $Q = \{(s, t): 0 \leqslant s \leqslant l, 0 \leqslant t \leqslant T\}$ и пусть функция $z = z(s, t) \in L_1(Q)$. Функция $g(s) \in L_1[0, l]$ называется *следом* функции $z(s, t)$ при $t = \tau$, если для любого $\varepsilon > 0$ найдется число $\delta > 0$ такое, что для почти всех $t \in [0, T]$, для которых $|t - \tau| < \delta$, имеет место неравенство

$$\int\limits_0^l |z(s, t) - g(s)| ds < \varepsilon.$$

Если след функции $z(s, t)$ при $t = \tau$ существует, то его будем обозначать через $z(s, \tau)$, $0 \leqslant s \leqslant l$, или $z(\cdot, \tau)$. Аналогично определяется след $z(s, \cdot) = z(s, t)$, $0 \leqslant t \leqslant T$, при каждом фиксированном $s \in [0, l]$. Можно показать, что если след функции существует, то он определяется единственным образом.

Если функция $z(s, t)$ непрерывна на Q, то след $z(\cdot, t)$ этой функции при каждом $t \in [0, T]$ совпадает со значением этой функции, представляющим собой функцию $z(s, t)$ переменной $s \in [0, l]$ при фиксированном t.

Пусть $z = z(s, t) \in L_1(Q)$. Напоминаем, что под элементом из $L_1(Q)$ понимается не одна функция, а класс эквивалентных функций, т. е. функций, отличающихся друг от друга на множестве нулевой меры. Поскольку двумерная мера множества $U_\tau = \{(s, t): 0 \leqslant s \leqslant l, t = \tau\}$ равна нулю, то эквивалентные функции на этом множестве могут принимать произвольные значения или даже могут быть не определены. Поэтому говорить о значениях функции $z(s, t) \in L_1(Q)$ при фиксированном t или s не имеет смысла, а введенное выше понятие следа функции естественным образом обобщает понятие значения функции для функций из $L_1(Q)$.

Однако в общем случае нельзя ожидать, что функция из $L_1(Q)$ будет иметь след при всех значениях $t \in [0, T]$ или $s \in [0, l]$.

Пример 2. Пусть $z(s,t)=0$ при $0\leqslant s\leqslant l$, $T/(2k)<t\leqslant T/(2k-1)$, $k=1,2,\ldots$, $z(s,t)=1$ при $0\leqslant s\leqslant l$; $T/(2k+1)<t\leqslant T/(2k)$, $k=1,2,\ldots$ Эта функция принадлежит $L_1(Q)$, но при $t=0$ не имеет следа.

Для того чтобы функция $z=z(s,t)\in L_1(Q)$ имела след при всех $t\in[0,T]$, на нее нужно наложить дополнительные ограничения. Например, функция $z(s,t)\in L_1(Q)$, обладающая обобщенной производной $z_t(s,t)\in L_1(Q)$, имеет след при каждом $t\in[0,T]$, и ее можно изменить на множестве двумерной меры нуль так, что она при всех $t\in[0,T]$ будет иметь значения, совпадающие со следом почти всюду на отрезке $0\leqslant s\leqslant l$. Замечательно то, что в этом случае справедлива формула, обобщающая формулу Ньютона — Лейбница:

$$\int_a^b z_t(s,t)dt=z(s,b)-z(s,a),$$

где $z(s,b)$, $z(s,a)$, $0\leqslant s\leqslant l$, — следы функции $z(s,t)$ при $t=b$ и $t=a$ соответственно; a, b — любые числа из отрезка $0\leqslant t\leqslant T$, причем в формуле равенство имеет место для почти всех $s\in[0,l]$. Если дополнительно известно, что $z(s,t)$, $z_t(s,t)\in L_p(Q)$, $1\leqslant p<\infty$, то следы такой функции принадлежат $L_p[0,l]$ и непрерывны по t в метрике $L_p[0,l]$, т. е.

$$\lim_{t\to\tau}\int_0^l|z(s,t)-z(s,\tau)|^p ds=0$$

при всех $\tau\in[0,T]$. В частности, если $z(s,t)\in H^1(Q)$, то такая функция имеет следы $z(\cdot,t)\in L_2[0,l]$ при всех $t\in[0,T]$ и $z(s,\cdot)\in L_2[0,T]$ при всех $s\in[0,l]$, причем указанные следы непрерывно зависят в метрике $L_2[0,l]$ и $L_2[0,T]$ соответственно [492; 648; 649].

Если для функции $z(s,t)\in L_2(Q)$ существует последовательность $\{z_k(s,t)\}\in C^\infty(Q)$ такая, что

$$\lim_{k\to\infty}\operatorname*{ess\,sup}_{t\in[0,T]}\int_0^l|z_k(s,t)-z(s,t)|^2ds=0,$$

то $z(s,t)$ также имеет след $z(\cdot,t)\in L_2[0,l]$ при каждом $t\in[0,T]$, причем существует эквивалентная функция, значения которой совпадают со следом $z(\cdot,t)$ при всех $t\in[0,T]$ [95; 492].

Остальные обозначения, определения и факты из функционального анализа будем приводить ниже по мере надобности.

§ 2. Теорема Вейерштрасса в функциональных пространствах

1. Пусть U — некоторое множество, а $J(u)$ — функция, определенная на этом множестве и принимающая на нем конечные вещественные значения. Для обозначения задачи минимизации [максимизации] функции $J(u)$ на множестве U, как и выше, будем пользоваться следующей краткой символической записью:

$$J(u)\to\inf,\quad u\in U\quad[J(u)\to\sup,\quad u\in U]. \tag{1}$$

Воспроизведем определения некоторых понятий, которые в главах 1, 2 были введены для задачи (1), когда множество U принадлежит конечномерному пространству E^n. Функция $J(u)$ называется *ограниченной снизу* [*сверху*] на множестве U, если существует число A такое, что $J(u)\geqslant A$ [$J(u)\leqslant A$] для всех $u\in U$. Функция $J(u)$ *не ограничена снизу* [*сверху*] на U, если существует последовательность $\{u_k\}\in U$, для которой $\lim\limits_{k\to\infty}J(u_k)=-\infty$ [$\lim\limits_{k\to\infty}J(u_k)=+\infty$].

Пусть функция $J(u)$ ограничена снизу [сверху] на U. Тогда существует число a называемое *нижней* [*верхней*] *гранью* функции $J(u)$ на множестве U и обладающее свойствами: 1) $J(u)\geqslant a$ [$J(u)\leqslant a$] при всех $u\in U$; 2) для любого $\varepsilon>0$ найдется точка $u_\varepsilon\in U$, для которой $J(u_\varepsilon)<a+\varepsilon$ [$J(u_\varepsilon)>a-\varepsilon$]. Если $J(u)$ не ограничена снизу [сверху], то в качестве нижней [верхней] грани $J(u)$ на U, по определению, принимают $a=-\infty$ [$a=+\infty$]. Нижнюю [верхнюю] грань $J(u)$ на U будем обозначать

$$\inf_U J(u)=J_*\quad[\sup_U J(u)=J^*].$$

Если $J_*>-\infty$ [$J^*<\infty$], то можно ввести множества

$$U_*=\{u\in U\colon J(u)=J_*\}\quad[U^*=\{u\in U\colon J(u)=J^*\}].$$

Если $U_*\neq\varnothing$ [$U^*\neq\varnothing$], то говорят, что нижняя [верхняя] грань в задаче (1) достигается, а точки $u_*\in U_*$ [$u^*\in U^*$] называются *точками минимума* [*максимума*] функции $J(u)$ на U.

Последовательность $\{u_k\}\in U$ называют *минимизирующей* [*максимизирующей*] для функции $J(u)$ на множестве U, если

$$\lim_{k\to\infty}J(u_k)=J_*\quad[\lim_{k\to\infty}J(u_k)=J^*].$$

Поскольку задача максимизации $J(u)\to\sup$, $u\in U$ равносильна задаче минимизации $(-J(u))\to\inf$, $u\in U$, поэтому в дальнейшем мы будем рассматривать, в основном, задачи минимизации.

Как и в главе 2, теоремами Вейерштрасса будем называть теоремы, содержащие утверждение о достижении нижней грани некоторой функции на каком-либо множестве.

2. Сначала приведем теорему Вейерштрасса, обобщающую теорему 2.1.1 на случай метрических пространств. Для ее формулировки нам понадобятся понятия компактного множества и полунепрерывности снизу функции в метрическом пространстве. Напоминаем [371; 393; 705], что множество M называется *метрическим пространством*, если каждой паре элементов $u,v\in M$ соответствует вещественное число $\rho(u,v)$, называемое расстоянием между элементами u и v, которое удовлетворяет условиям (аксиомам): 1) $\rho(u,v)\geqslant 0$ $\forall u,v\in M$, причем $\rho(u,v)=0$ тогда и только тогда, когда $u=v$, 2) $\rho(u,v)=\rho(v,u)$ $\forall u,v\in M$; 3) $\rho(u,v)\leqslant\rho(u,w)+\rho(w,v)$ $\forall u,v,w\in M$. Такая функция $\rho\colon M\times M\to\mathbb{R}$ называется *метрикой*.

Каждое банахово (и гильбертово) пространство B является линейным метрическим пространством с метрикой $\rho(u,v)=\|u-v\|_B$. Однако не каждое метрическое пространство M линейно. Множество $O(u,\varepsilon)=\{v\in M\colon \rho(v,u)<\varepsilon\}$ называется *ε-окрестностью точки u*. Точка $u\in M$ называется *предельной точкой* множества $U\subseteq M$, если любая окрестность точки u содержит хотя бы одну точку из U, отличную от u. Говорят, что последовательность точек $\{u_k\}$, $u_k\in M$, $k=1,2,\ldots$, *сходится* к точке u, если $\lim\limits_{k\to\infty}\rho(u_k,u)=0$. Точка u является предельной точкой множества U тогда и только тогда, когда существует последовательность $\{u_k\}\in U$, $u_k\neq u$,

$k = 1, 2, \ldots$, сходящаяся к u. Множество U называется *замкнутым*, если оно содержит все свои предельные точки. Множество U замкнуто тогда и только тогда, если любая точка u, к которой сходится хотя бы одна последовательность $\{u_k\} \in U$, принадлежит самому множеству U. Множество U называется *ограниченным*, если существует шар $S(u_0, R) = \{u \in M: \rho(u, u_0) \leqslant R\}$ радиуса R с центром в точке $u_0 \in M$, такой, что $U \subseteq S(u_0, R)$.

Определение 1. Множество U из метрического пространства M называется *относительно компактным* в метрике этого пространства, если из любой последовательности $\{u_k\} \in U$ можно выбрать хотя бы одну подпоследовательность $\{u_{k_m}\}$, которая сходится к некоторой точке $v \in M$. Если при этом любая такая точка v принадлежит самому множеству U, то такое множество U называется *компактным* в M.

Компактное множество замкнуто, а относительно компактное множество необязательно замкнуто. Если множество U относительно компактно, то его замыкание $\overline{U}$, которое получается присоединением к U всех его предельных точек, компактно.

Определение 2. Функцию $J(u)$, определенную на множестве U из метрического пространства M, называют *полунепрерывной снизу* [*сверху*] в точке $u \in U$, если для любой последовательности $\{u_k\} \in U$, сходящейся к точке u, имеет место неравенство

$$\varliminf_{k \to \infty} J(u_k) \geqslant J(u) \quad \left[\varlimsup_{k \to \infty} J(u_k) \leqslant J(u)\right]. \tag{2}$$

Функция $J(u)$ называется *полунепрерывной снизу* [*сверху*] на множестве U, если она полунепрерывна снизу [сверху] в каждой точке $u \in U$. Функция $J(u)$ называется *непрерывной в точке* $u \in U$ [*на множестве* U], если она полунепрерывна снизу и сверху в точке u [на множестве U].

Определение 3. Говорят, что последовательность $\{u_k\} \in M$ сходится ко множеству $U \subseteq M$, если $\lim\limits_{k \to \infty} \rho(u_k, U) = 0$, где $\rho(u, U) = \inf\limits_{v \in U} \rho(u, v)$ — расстояние от точки u до множества U.

Заметим, что определение 1, 2, 3 обобщают соответствующие определения 2.1.1, 2.1.3, 1.1.5 на случай метрических пространств. Нетрудно убедиться, что леммы 2.1.1, 2.1.2 сохраняют силу и в метрических пространствах. Справедлива

Теорема 1. *Пусть U — компактное множество из метрического пространства M, функция $J(u)$ определена и полунепрерывна снизу на U. Тогда $J_* = \inf\limits_U J(u) > -\infty$, множество $U_* = \{u \in U: J(u) = J_*\}$ непусто, компактно и любая минимизирующая последовательность $\{u_k\}$ сходится ко множеству U_*.*

Эта теорема доказывается так же, как и аналогичная теорема 2.1.1.

Для применения теоремы 1 к конкретным задачам минимизации полезно иметь критерии компактности и относительной компактности в наиболее часто встречающихся в приложениях функциональных пространствах. Например, в пространствах $C(G)$, $L_p(G)$, $1 \leqslant p < \infty$, где G — ограниченное замкнутое множество из E^n, критерий компактности может быть сформулирован следующим образом.

Теорема 2. *Замкнутое множество U в пространствах $C(G)$, $L_p(G)$, $1 \leqslant p < \infty$, компактно тогда и только тогда, когда*

1) *множество U равномерно ограничено, т. е.* $\sup\limits_{u \in U} \|u\| < \infty$;

2) *множество U равностепенно непрерывно, т. е. для любого $\varepsilon > 0$ найдется число $\delta > 0$ такое, что* $\sup\limits_{u \in U} \|u(t + \Delta t) - u(t)\| < \varepsilon$ *для всех t, $t + \Delta t \in G$, $|\Delta t| < \delta$; здесь $\|u\|$ означает норму пространства $C(G)$ или $L_p(G)$, $1 \leqslant p < \infty$.*

Если в сформулированной теореме откажемся от требования замкнутости множества U, то получим критерий относительной компактности в указанных пространствах; доказательство этого утверждения в $C(G)$ см., например, [371; 393; 705] (теорема Арцела), в $L_p(G)$, $1 \leqslant p < \infty$, см. в [371; 492; 648]. Критерии относительной компактности и компактности в различных других функциональных пространствах читатель может найти в [95; 258; 393; 535; 648; 649; 772]; некоторые такие критерии будут обсуждаться ниже в § 2.2.

В евклидовом пространстве E^n множество компактно тогда и только тогда, когда оно замкнуто и ограничено. Доказательство этого факта существенно опирается на известную теорему Больцано — Вейерштрасса, согласно которой из любой ограниченной последовательности $\{u_k\} \in E^n$ можно выбрать хотя бы одну сходящуюся подпоследовательность. Однако такая теорема в метрических пространствах, вообще говоря, неверна. По этой причине, оказывается, в метрических пространствах замкнутости и ограниченности множества, вообще говоря, недостаточно для его компактности.

Пример 1. Пусть H — гильбертово пространство, $S_1 = \{u \in H: \|u\| \leqslant 1\}$ — единичный шар в H, пусть $\{e_k\}$ — некоторая бесконечная ортонормированная система в H. Из последовательности $\{e_k\} \in S_1$ невозможно выбрать подпоследовательность, сходящуюся к какой-нибудь точке в метрике H. В самом деле, если бы такая последовательность $\{e_{k_m}\}$ существовала, то она была бы фундаментальной в H [393; 705]. Однако $\|e_k - e_m\|^2 = 2$ $\forall m \neq k$, и последовательность $\{e_k\}$ не может иметь ни одной фундаментальной подпоследовательности. Это означает, что шар в любом бесконечномерном гильбертовом пространстве H не может быть компактным в метрике H.

Пример 2. Пусть $U = \{u = u(t) \in L_2[0, 1]: |u(t)| \leqslant 1$ почти всюду на $[0, 1]\}$. Это множество не является компактным в метрике $L_2[0, 1]$. В самом деле, возьмем последовательность $u_k = \sin \pi k t$, $0 \leqslant t \leqslant 1$, $k = 1, 2, \ldots$ ... В примере 1.1 было замечено, что $\{u_k\} \to 0$ слабо в $L_2[0, 1]$. Однако $\|u_k - u_m\|^2_{L_2} = \int\limits_0^1 |\sin \pi k t - \sin \pi m t|^2 dt = 1$, $\forall k \neq m$, поэтому из $\{u_k\}$ нельзя выбрать подпоследовательность, которая по норме $L_2[0, 1]$ сходилась бы к нулю. Следовательно, U некомпактно в метрике $L_2[0, 1]$.

Пример 3. Множество $U = \{u = u(t) \in C[0, 1]: |u(t)| \leqslant 1, 0 \leqslant t \leqslant 1\}$ — шар в $C[0, 1]$ — некомпактно в метрике $C[0, 1]$. Это может быть доказано с помощью тех же рассуждений, которые проводились в предыдущем примере, с учетом того, что из сходимости в метрике $C[0, 1]$ следует сходимость в метрике $L_2[0, 1]$.

3. Заметим, что множества, подобные рассмотренным в примерах 1–3, часто встречаются в прикладных задачах оптимального управления. Отсутствие свойства компактности этих множеств не позволяет применять теорему 1 для доказательства существования оптимального управления в таких задачах. Поэтому желательно иметь такие теоремы Вейерштрасса в банаховых и гильбертовых пространствах, которые не требуют компактности

множества в метрике этих пространств. Для формулировки таких теорем введем несколько понятий, связанных с понятием слабой сходимости в банаховом пространстве.

Определение 4. Множество U из банахова пространства B называется *относительно слабо компактным*, если из любой последовательности $\{u_k\} \in U$ можно выбрать хотя бы одну подпоследовательность $\{u_{k_m}\}$, которая слабо в B сходится к некоторой точке $v \in U$. Если при этом такая точка v принадлежит самому множеству U, то такое множество называется *слабо компактным*.

Определение 5. Функцию $J(u)$, определенную на некотором множестве U из банахова пространства B, называют *слабо полунепрерывной снизу [сверху]* в точке $u \in U$, если для любой последовательности $\{u_k\} \in U$, которая слабо в B сходится к точке u, имеет место неравенство (2). Функция $J(u)$ называется *слабо полунепрерывной снизу [сверху] на множестве U*, если она слабо полунепрерывной снизу [сверху] в каждой точке $u \in U$. Функция $J(u)$ называется *слабо непрерывной в точке $u \in U$* [на множестве U], если она слабо полунепрерывна снизу и сверху в точке u [на множестве U].

Пример 4. Функция $J(u) = \langle c, u\rangle$, $u \in B$, где c — фиксированный элемент из сопряженного пространства B^*, слабо непрерывна на B, что вытекает непосредственно из определения слабой сходимости.

Пример 5. Функция $J(u) = \|u\|_H^2$ слабо полунепрерывна снизу во всех точках гильбертова пространства H. В самом деле, пусть $\{u_k\}$ произвольная последовательность из H, слабо сходящаяся к точке u. Тогда

$$J(u_k) = \|u_k\|^2 = \|(u_k - u) + u\|^2 = \|u_k - u\|^2 + 2\langle u, u_k - u\rangle + \|u\|^2 \geqslant$$
$$\geqslant 2\langle u, u_k - u\rangle + \|u\|^2, \quad k = 1, 2, \ldots$$

Отсюда, учитывая, что $\lim\limits_{k\to\infty} \langle u, u_k - u\rangle = 0$ по определению слабой сходимости, получаем: $\varliminf\limits_{k\to\infty} \|u_k\|^2 \geqslant \|u\|^2$. Это означает, что функция $J(u) = \|u\|_H^2$ слабо полунепрерывна снизу на H. Отметим, что последнее неравенство может быть строгим. А именно, возьмем $v_k = u + e_k$, $k = 1, 2, \ldots$, где $\{e_k\}$ — какая-либо бесконечная ортонормированная система из H. Так как $\{e_k\} \to 0$ слабо в H (пример 1.1), то $\{v_k\} \to u$ слабо в H. Кроме того,

$$\|v_k\|^2 = \|u\|^2 + 2\langle u, e_k\rangle + \|e_k\|^2 = \|u\|^2 + 2\langle u, e_k\rangle + 1, \quad k = 1, 2, \ldots,$$

так что $\lim\limits_{k\to\infty} \|v_k\|^2 = \|u\|^2 + 1 > \|u\|^2 \quad \forall u \in H$. Это означает, что функция $J(u) = \|u\|_H^2$ не является слабо полунепрерывной сверху и, тем более, слабо непрерывной ни в одной точке пространства H.

Пример 6. Функция $J(u) = -\|u\|_H^2$ слабо полунепрерывна сверху на H, то она не будет слабо полунепрерывной снизу и, тем более, слабо непрерывной ни в одной точке H.

Из примеров 5, 6 следует, что функция может быть сильно (в метрике H) непрерывной, но она необязательно слабо полунепрерывна снизу или сверху и, тем более, слабо непрерывна. Более того, нетрудно понять, что всякая слабо полунепрерывная снизу [сверху] функция сильно полунепрерывна снизу [сверху], слабо непрерывная функция — сильно непрерывна. В самом деле, поскольку из сильной сходимости последовательности вытекает ее слабая сходимость, то множество $O_1(u)$ слабо сходящихся к точке

u последовательностей гораздо богаче множества $O_2(u)$ последовательностей, сильно сходящихся к этой же точке. Поэтому неудивительно, что то из неравенств (2), которое выполняется на более богатом множестве $O_1(u)$, тем более будет выполняться на его подмножестве $O_2(u)$, что как раз и означает, что из слабой полунепрерывности снизу [сверху] функции следует ее сильная полунепрерывность снизу [сверху], а из слабой непрерывности — ее сильная непрерывность.

Далее, сформулируем один важный критерий относительно слабой компактности множеств в рефлексивных банаховых пространствах.

Теорема 3. *Множество U из рефлексивного банахова пространства B относительно слабо компактно тогда и только тогда, когда оно ограничено (в метрике B).*

Эта теорема представляет собой обобщение конечномерной теоремы Больцано — Вейерштрасса. Ее доказательство см. в [705, стр. 210].

Пример 7. В бесконечномерном гильбертовом пространстве H рассмотрим единичный шар $S_1 = \{u \in U: \|u\| \leqslant 1\}$. Выше мы выяснили (пример 1), что шар не является компактным в метрике H. Покажем, что шар S_1 слабо компактен. Возьмем произвольную последовательность $\{u_k\} \in S_1$. Так как $\|u_k\| \leqslant 1$, $k = 1, 2, \ldots$, то согласно теореме 3 из $\{u_k\}$ можно выбрать подпоследовательность $\{u_{k_m}\}$, которая слабо сходится к некоторой точке $v \in H$. Отсюда с учетом слабой полунепрерывности снизу функции $g(u) = \|u\|^2$ (пример 5) имеем: $g(v) = \|v\|^2 \leqslant \varliminf\limits_{m\to\infty} \|u_{k_m}\|^2 \leqslant \varlimsup\limits_{m\to\infty} \|u_{k_m}\|^2 \leqslant 1$. Это означает, что $v \in S_1$. Слабая компактность единичного шара в H установлена.

Из теоремы 3 следует, что открытый шар $\{u \in H: \|u\| < 1\}$ и сфера $\{u \in H: \|u\| = 1\}$ относительно слабо компактны, но не являются слабо компактными.

Другие примеры слабо компактных множеств и слабо полунепрерывных снизу функций будут приведены ниже. Сейчас сформулируем и докажем другой, так называемый слабый вариант теоремы Вейерштрасса.

Теорема 4. *Пусть U — слабо компактное множество из банахова пространства B, функция $J(u)$ определена и слабо полунепрерывна снизу на U. Тогда $J_* = \inf\limits_U J(u) > -\infty$, множество $U_* = \{u \in U: J(u) = J_*\}$ непусто, слабо компактно.*

Доказательство. Возьмем произвольную минимизирующую последовательность $\{u_k\}$: $u_k \in U$, $k = 1, 2, \ldots$, $\lim\limits_{k\to\infty} J(u_k) = J_*$. Так как U — слабо компактное множество, то $\{u_k\}$ имеет хотя бы одну подпоследовательность, слабо сходящуюся к некоторой точке из U. Пусть $u_* \in U$ — одна из таких точек и пусть подпоследовательность $\{u_{k_m}\}$ слабо сходится к u_*. Пользуясь определением нижней грани и слабой полунепрерывностью снизу функции $J(u)$, имеем

$$J_* \leqslant J(u_*) \leqslant \varliminf_{m\to\infty} J(u_{k_m}) = \lim_{k\to\infty} J(u_k) = J_*,$$

т. е. $J(u_*) = J_*$. Отсюда следует, что $J_* > -\infty$, $u_* \in U_*$, т. е. $U_* \neq \varnothing$. Покажем, что множество U_* слабо компактно. Возьмем произвольную последовательность $\{v_k\} \in U_*$. Так как $\{v_k\} \in U$ — слабо компактное множество, то существует подпоследовательность $\{v_{k_m}\}$, слабо сходящаяся к некоторой точке $v_* \in U$. Но $J(v_k) = J_*$, $k = 1, 2, \ldots$, поэтому $\lim\limits_{k\to\infty} J(v_k) = J_*$, т. е. $\{v_k\}$ — минимизирующая последовательность. Из вышедоказанного следует, что $v_* \in U_*$. Это значит, что множество U_* слабо компактно. Теорема 4 доказана.

4. Для удобства пользования теоремой 4 желательно иметь набор сравнительно легко проверяемых достаточных условий слабой компактности множеств и слабой полунепрерывности снизу функций в банаховых пространствах. Приведем несколько таких условий. Для их формулировки нам понадобятся понятия выпуклого множества, выпуклой функции, которые в банаховых пространствах определяются дословно также, как в соответствующих определениях 4.1.1, 4.2.1. Кроме того, ниже нам понадобится следующая важная

Теорема 5 (Мазур). *Если последовательность $\{u_k\}$ из банахова пространства B сходится к точке u слабо в B, то существует последовательность $\{v_k\}$ выпуклых комбинаций точек $\{u_k\}$, т. е.*

$$v_k = \sum_{i=1}^{k} \alpha_{ki} u_i, \quad \sum_{i=1}^{k} \alpha_{ki} = 1, \quad \alpha_{ki} \geqslant 0, \quad i = 1, \ldots, k,$$

такая, что $\{v_k\}$ сходится сильно в B к той же точке u.

Доказательство этой теоремы см., например, в [357, стр. 173].

Теорема 6. *Всякое выпуклое замкнутое ограниченное множество U из рефлексивного банахова пространства B слабо компактно.*

Доказательство. Пусть $\{u_k\}$ — произвольная последовательность из U. Так как по условию множество U ограничено (в метрике B), то последовательность $\{u_k\}$ ограничена и по теореме 3 найдется подпоследовательность $\{u_{k_m}\}$, которая слабо сходится к некоторой точке $v \in B$. В силу теоремы 5 тогда найдется последовательность $\{v_{k_m}\}$ выпуклых комбинаций точек u_{k_m}, $m = 1, 2, \ldots$, которая сильно сходится к той же точке v. Однако $v_{k_m} \in U$, $m = 1, 2, \ldots$, в силу выпуклости U. Из сильной сходимости $\{v_{k_m}\}$ к v и сильной (в метрике B) замкнутости U следует, что $v \in U$. Это означает, что множество U слабо компактно. □

Приведем пример, показывающий, что в теоремах 3, 6 требование рефлексивности пространства существенно.

Пример 8. Рассмотрим множество $S_1 = \{u = u(t) \in L_1[-1, 1]: \|u\|_{L_1} = \int_{-1}^{1} |u(t)|dt \leqslant 1\}$ — единичный шар в банаховом пространстве $L_1[-1, 1]$. Сопряженное к $L_1[-1, 1]$ пространство совпадает (изометрично) с пространством $L_\infty[-1, 1]$ и значение произвольного линейного непрерывного функционала $c = c(t) \in L_\infty[-1, 1]$ на элементе $u = u(t) \in L_1[-1, 1]$ равно $\langle c, u\rangle = \int_{-1}^{1} c(t)u(t)dt$. Возьмем последовательность

$$u_k = u_k(t) = \begin{cases} k & \text{при } |t| \leqslant \frac{1}{2k}, \\ 0 & \text{при } \frac{1}{2k} < |t| \leqslant 1, \end{cases} \quad k = 1, 2, \ldots$$

Очевидно, $\|u_k\|_{L_1} = 1$, $k = 1, 2, \ldots$, так что $\{u_k\} \in S_1$. Допустим, что из последовательности $\{u_k(t)\}$ можно выделить подпоследовательность $\{u_{k_m}(t)\}$, сходящуюся к некоторой функции $v = v(t)$ слабо в $L_1[-1, 1]$, т. е. $\lim\limits_{m\to\infty} \int_{-1}^{1} c(t)u_{k_m}(t)dt = \int_{-1}^{1} c(t)v(t)dt \ \ \forall c = c(t) \in L_\infty[-1, 1]$. В частности, для $c = c(t) \equiv 1$ отсюда имеем: $\lim\limits_{m\to\infty} \int_{-1}^{1} 1 \cdot u_{k_m}(t)dt = 1 = \int_{-1}^{1} v(t)dt$. Это значит,

что $|v(t)| > 0$ на множестве $E \subseteq [-1, 1]$, имеющем меру $\mathrm{mes}E > 0$. Далее, воспользуемся неравенством Чебышева:

$$\mathrm{mes}\{t: |t| \leqslant 1, |v(t)| \geqslant A\} \leqslant \frac{1}{A} \int_{-1}^{1} |v(t)|dt,$$

справедливым при всех $A > 0$. Это неравенство вытекает из оценки: $\int_{-1}^{1} |v(t)|dt = \int\limits_{\{|v(t)|\geqslant A\}} |v(t)|dt + \int\limits_{\{|v(t)|<A\}} |v(t)|dt \geqslant A\,\mathrm{mes}\{t: |t| \leqslant 1, |v(t)| \geqslant A\}$.

Возьмем число A столь большим, чтобы $\mathrm{mes}\{t: |t| \leqslant 1, |v(t)| \geqslant A\} < \frac{1}{4}\mathrm{mes}E$. Кроме того, выберем число $\delta > 0$ столь малым, чтобы $2\delta < \frac{1}{4}\mathrm{mes}E$. Введем множество $E_1 = E \setminus (\{t: |t| \leqslant \delta\} \cup \{t: |t| \leqslant 1, |v(t)| \geqslant A\})$. Нетрудно видеть, что $\mathrm{mes}E_1 > \frac{1}{2}\mathrm{mes}E > 0$ и $|v(t)| > 0 \ \ \forall t \in E_1$. Определим функцию $c_0(t)$ следующим образом: $c_0(t) = v(t) \ \ \forall t \in E_1$, $c_0(t) = 0$ $\forall t \notin E_1$. Так как $|c_0(t)| \leqslant A \ \ \forall t \in [-1, 1]$ и измерима, то $c_0(t) \in L_\infty[-1, 1]$. Кроме того, $\int_{-1}^{1} c_0(t)v(t)dt = \int_{E_1} v^2(t)dt > 0$, поэтому $\lim\limits_{m\to\infty} \int_{-1}^{1} c_0(t)u_{k_m}(t)dt = \int_{-1}^{1} c_0(t)v(t)dt > 0$. С другой стороны, $c_0(t)u_{k_m}(t) \equiv 0 \ \ \forall t \in [-1, 1]$ для всех номеров k_m, для которых $\frac{1}{2k_m} < \delta$, поэтому $\lim\limits_{m\to\infty} \int_{-1}^{1} c_0(t)u_{k_m}(t)dt = 0$. Противоречие. Это означает, что из последовательности $\{u_k\} \in S_1$ невозможно выбрать подпоследовательность $\{u_{k_m}\}$, слабо сходящуюся к какой-либо точке из $L_1[-1, 1]$. Следовательно, единичный шар в $L_1[-1, 1]$ не является относительно слабо компактным и, тем более, слабо компактным. Остается заметить, что S_1 — выпуклое замкнутое ограниченное множество из $B = L_1[-1, 1]$, однако условие рефлексивности пространства B, требуемое в теоремах 3, 6, здесь не выполнено.

Приведем два примера слабо компактных множеств в пространстве $L_2^r(G)$, где G — ограниченное замкнутое множество из E^n.

Пример 9. Пусть V — выпуклое замкнутое множество из E^r, пусть $U = \{u = u(t) \in L_2^r(G): u(t) \in V$ почти всюду на $G\}$. Нетрудно видеть, что U — выпуклое множество. Убедимся, что U замкнуто в метрике $L_2^r(G)$. Возьмем произвольную последовательность $\{u_k(t)\} \in U$, сходящуюся в метрике $L_2^r(G)$ к некоторой функции $u(t) \in L_2^r(G)$. Тогда найдется [393, стр. 388] подпоследовательность $\{u_{k_m}(t)\}$, сходящаяся к той же функции $u(t)$ почти всюду на G. Поскольку V — замкнутое множество и $u_{k_m}(t) \in V$ почти всюду на G, то при $m \to \infty$ отсюда получим $u(t) \in V$ почти всюду на G. Замкнутость множества U доказана. Если множество V еще и ограничено, то U — ограничено в метрике $L_2^r(G)$. По теореме 6 тогда множество U слабо компактно в $L_2^r(G)$.

Пример 10. Пусть $U = \{u = u(t) = (u^1(t), \ldots, u^r(t)) \in L_2^r(G)$: $\alpha_i(t) \leqslant u^i(t) \leqslant \beta_i(t)$ почти всюду на G, $i = 1, \ldots, r\}$, где $\alpha_i(t) \leqslant \beta_i(t)$, $i = 1, \ldots, r$, — заданные функции из $L_2(G)$. Покажем, что U замкнуто в метрике $L_2^r(G)$. Пусть последовательность $\{u_k(t)\} \in U$ сходится к $u = u(t)$ по норме $L_2^r(G)$. Тогда найдется [393, стр. 388] подпоследовательность $\{u_{k_m}(t)\}$, сходящаяся к той же функции $u(t)$ почти всюду на G. Но $\alpha_i(t) \leqslant u_{k_m}^i(t) \leqslant u_{k_m}^i(t) \leqslant \beta_i(t)$ почти всюду на G, $i = 1, \ldots, r$, $m = 1, 2, \ldots$

Отсюда при $m \to \infty$ получим $\alpha_i(t) \leqslant u^i(t) \leqslant \beta_i(t)$ почти всюду на G для всех $i = 1, \ldots, r$. Следовательно, $u(t) \in U$, т. е. U замкнуто в метрике $L_2^r(G)$. Нетрудно видеть, что U выпукло и ограничено в метрике $L_2^r(G)$. Согласно теореме 6 множество U слабо компактно. В метрике $L_2^r(G)$ это множество, как видно из примера 2, при $\alpha_i(t) \neq \beta_i(t)$, $i = 1, \ldots, r$, не является компактным.

Приведем один критерий слабой полунепрерывности снизу функции.

Теорема 7. *Пусть U — выпуклое множество из банахова пространства B. Выпуклая на U функция $J(u)$ слабо полунепрерывна снизу на U тогда и только тогда, когда $J(u)$ сильно (в метрике B) полунепрерывна снизу на U.*

Доказательство. Необходимость. Пусть $J(u)$ слабо полунепрерывна снизу на U. Возьмем произвольные точку $u \in U$ и последовательность $\{u_k\} \in U$, сходящуюся к точке u в метрике B. Тогда $\{u_k\}$ сходится к u слабо в B, и $\varliminf\limits_{k \to \infty} J(u_k) \geqslant J(u)$. Полунепрерывность снизу на U функции $J(u)$ в метрике B доказана. Заметим, что выпуклость U и $J(u)$ здесь не использовалась.

Достаточность. Пусть $J(u)$ полунепрерывна снизу на U. Возьмем произвольную последовательность $\{u_k\} \in U$, слабо в B сходящуюся к точке $u \in U$. Выбирая при необходимости подпоследовательность, можем считать, что $\varliminf\limits_{k \to \infty} J(u_k) = \lim\limits_{k \to \infty} J(u_k)$. Из теоремы 5 следует, что точка u принадлежит замыканию выпуклой оболочки точек $\{u_k, u_{k+1}, \ldots\}$, где k — любое фиксированное натуральное число. Это значит, что для каждого номера $k = 1, 2, \ldots$ найдутся целое число $m \geqslant k$ и вещественные числа $\alpha_{kmi} \geqslant 0$, $i = k, k+1, \ldots, m$, $\sum\limits_{i=k}^{m} \alpha_{kmi} = 1$ такие, что последовательность $v_k = \sum\limits_{i=k}^{m} \alpha_{kmi} u_i$ будет сходиться к точке u в метрике B, т. е. $\lim\limits_{k \to \infty} \|v_k - u\| = 0$. Тогда в силу полунепрерывности снизу $J(u)$ в точке u имеем $\varliminf\limits_{k \to \infty} J(v_k) \geqslant J(u)$. Из выпуклости $J(u)$ следует (неравенство Йенсена (4.2.2)):

$$J(v_k) = J\left(\sum_{i=k}^{m} \alpha_{kmi} u_i\right) \leqslant \sum_{i=k}^{m} \alpha_{kmi} J(u_i) \leqslant \sup_{i \geqslant k} J(u_i), \quad k = 1, 2, \ldots,$$

однако $\lim\limits_{k \to \infty} \sup\limits_{i \geqslant k} J(u_i) = \lim\limits_{k \to \infty} J(u_k)$, поэтому, переходя к пределу в предыдущем неравенстве, получим

$$J(u) \leqslant \varliminf_{k \to \infty} J(v_k) \leqslant \lim_{k \to \infty} \sup_{i \geqslant k} J(u_i) = \lim_{k \to \infty} J(u_k) = \varliminf_{k \to \infty} J(u_k).$$

Слабая полунепрерывность снизу на U функции $J(u)$ доказана. □

Пример 11. Пусть $J(u) = \|u\|$ — норма в банаховом пространстве B. Так как $\|\alpha u + (1-\alpha)v\| \leqslant \alpha\|u\| + (1-\alpha)\|v\|$ при всех $u, v \in B$, $0 \leqslant \alpha \leqslant 1$, то $J(u)$ выпукла на B. Далее, из неравенства $|\|u_k\| - \|u\|| \leqslant \|u - u_k\|$ следует, что $J(u) = \|u\|$ непрерывна в метрике B во всех точках $u \in B$. Согласно теореме 7 тогда $J(u) = \|u\|$ слабо полунепрерывна снизу на B. Из примера 5 следует, что норма в B не будет слабо непрерывной функцией.

Из теорем 4, 6, 7 следует

Теорема 8. *Пусть U — выпуклое замкнутое ограниченное множество из рефлексивного банахова пространства B, функция $J(u)$ выпукла и полунепрерывна снизу на U. Тогда $J_* > -\infty$, множество U_* непусто, выпукло, замкнуто, ограничено.*

Приведем несколько примеров задач минимизации, показывающих, что условия теорем 4, 8 не могут быть существенно ослаблены.

Пример 12. Рассмотрим задачу минимизации функции

$$J(u) = \int_0^1 (x^2(t) - u^2(t))dt$$

при условиях $\dot{x}(t) = u(t)$, $0 \leqslant t \leqslant 1$; $x(0) = 0$, $u = u(t) \in U = \{u(t) \in L_2[0, 1]: |u(t)| \leqslant 1$ почти всюду на $[0, 1]\}$. Как было показано в примере 7.4.2, в этой задаче функция $J(u)$ не достигает на U своей нижней грани $J_* = -1$. Заметим, что здесь множество слабо компактно (см. примеры 2, 10). Функция $J(u)$ непрерывна на U в метрике $L_2[0, 1]$, но она не является слабо полунепрерывной снизу. В самом деле, возьмем последовательность $u_k = \sin \pi k t$, $0 \leqslant t \leqslant 1$, $k = 1, 2, \ldots$, слабо сходящуюся к нулю. Нетрудно проверить, что $\lim\limits_{k \to \infty} J(u_k) = -1/2 < J(0) = 0$, т. е. слабой полунепрерывности снизу нет.

Пример 13. Рассмотрим задачу минимизации функции

$$J(u) = \int_0^1 \operatorname{sign}(\tfrac{1}{2} - t) u(t) dt$$

при $u = u(t) \in U = \{u(t) \in C[0, 1]: |u(t)| \leqslant 1, 0 \leqslant t \leqslant 1\}$. Здесь U — единичный шар в $C[0, 1]$. Нетрудно видеть, что $J(u) > -1$ при всех $u \in U$, но $\lim\limits_{k \to \infty} J(u_k) = -1$, где $u_k = u_k(t) = k(t - \frac{1}{2})$ при $|t - \frac{1}{2}| \leqslant \frac{1}{k}$, $u_k(t) = \operatorname{sign}(t - \frac{1}{2})$ при $|t - \frac{1}{2}| > \frac{1}{k}$, $k = 1, 2, \ldots$ Следовательно, $J_* = -1$, но нижняя грань $J(u)$ на U не достигается. Заметим, что здесь $J(u)$ выпукла и непрерывна в метрике $C[0, 1]$ и согласно теореме 7 она слабо полунепрерывна снизу на $C[0, 1]$ (точнее говоря, она даже линейна и слабо непрерывна на $C[0, 1]$). Кроме того, множество U выпукло, замкнуто и ограничено в $C[0, 1]$. Однако пространство $C[0, 1]$ не является рефлексивным и множество U не будет слабо компактным. Любопытно, что на более широком множестве кусочно непрерывных функций, удовлетворяющих условию $|u(t)| \leqslant 1$, или на единичном шаре из $L_\infty[0, 1]$, рассматриваемая функция достигает своей нижней грани при $u_* = \operatorname{sign}(t - \frac{1}{2})$, $0 \leqslant t \leqslant 1$.

5. Теоремы 4 и 8 отличаются от теоремы 1 тем, что требования к множеству U в теоремах 4, 8 ослаблены по сравнению с теоремой 1, но зато на минимизируемую функцию накладываются более жесткие ограничения. Действуя в этом же направлении, в частности, отказываясь от требования ограниченности множества U, можно получить другие теоремы Вейерштрасса. Приведем несколько таких теорем.

Теорема 9. *Пусть U — выпуклое замкнутое множество из рефлексивного банахова пространства B, функция $J(u)$ выпукла, полунепрерывна снизу на U и для некоторой фиксированной точки $v \in U$ множество Лебега $M(v) = \{u \in U: J(u) \leqslant J(v)\}$ ограничено. Тогда $J_* > -\infty$, множество $U_* = \{u \in U: J(u) = J_*\}$ выпукло, замкнуто, ограничено.*

Доказательство этой теоремы проводится так же, как и доказательство аналогичной теоремы 2.1.2. А именно, сначала устанавливаем, что нижняя грань $J(u)$ на U может достигаться лишь в точках множества $M(v)$. Затем доказываем, что множество $M(v)$ выпукло и замкнуто в метрике B. Кроме того, $M(v)$ ограничено по условию. Применяя теорему 8 к задаче: $J(u) \to \inf$, $u \in M(v)$ получаем все утверждения теоремы 9. □

Достаточным для ограниченности множества $M(v)$ является условие $\lim\limits_{k\to\infty} J(u_k)=+\infty$, которое должно выполняться для любой последовательности $\{u_k\}\in U$, $\lim\limits_{k\to\infty}\|u_k\|=+\infty$ (ср. с теоремой 2.1.3). К функциям, удовлетворяющим последнему условию, относятся сильно выпуклые и равномерно выпуклые функции. Определения 4.3.1, 4.7.1 этих функций сохраняются и в банаховом пространстве B, надо лишь в них заменить $|u-v|$ на норму $\|u-v\|_B$. Примером сильно выпуклой функции в гильбертовом пространстве H является функция $J(u)=\|u\|_H^2$, что следует из тождества (4.3.2), которое справедливо в любом гильбертовом пространстве. Функция $J(u)=\|u\|_H^\gamma$ строго равномерно выпукла на H при всех $\gamma\geqslant 2$ с модулем выпуклости $\delta(t)=\left(\frac{1}{2}\right)^{\gamma-2}t^\gamma$. В пространствах $L_p^r(G)$ и l_p функция $J(u)=\|u\|^\gamma$ строго равномерно выпукла на всем пространстве при всех $\gamma\geqslant p\geqslant 2$ с модулем выпуклости $\delta(t)=\left(\frac{1}{2}\right)^{\gamma-1}t^\gamma$, а при $1<p<2$ эта функция строго равномерно выпукла при всех $\gamma>1$ на любом выпуклом ограниченном множестве из $L_p^r(G)$ или l_p [191].

Следует сказать, что не во всяком банаховом пространстве существуют сильно выпуклые функции. Более того, известно, что (с некоторыми оговорками) в банаховом пространстве B, в котором существует сильно выпуклая функция на B, можно ввести скалярное произведение и превратить его в гильбертово пространство. Не обсуждая возникающие здесь тонкие вопросы, мы ограничимся рассмотрением сильно выпуклых функций лишь на выпуклых множествах из гильбертовых пространств. Отметим, что сумма любой выпуклой функции и сильно [равномерно] выпуклой функции является сильно [равномерно] выпуклой, так что классы сильно [равномерно] выпуклых функций в упомянутых пространствах достаточно богаты.

Приведем формулировки теорем Вейерштрасса для сильно выпуклых и равномерно выпуклых функций, аналогичных теоремам 4.3.1, 4.7.1.

Теорема 10. *Пусть U — выпуклое замкнутое множество из гильбертова пространства H, а функция $J(u)$ сильно выпукла и полунепрерывна снизу на U. Тогда:*

1) *множество Лебега $M(v)=\{u\in U\colon J(u)\leqslant J(v)\}$ выпукло, замкнуто и ограничено при всех $v\in U$;*

2) *$J_*>-\infty$, $U_*\neq\varnothing$, причем U_* состоит из единственной точки u_*;*

3) *любая минимизирующая последовательность $\{u_k\}$ сходится к точке u_* по норме H, причем*

$$\tfrac{1}{2}\varkappa\|u_k-u_*\|^2\leqslant J(u_k)-J(u_*),\quad k=1,2,\ldots$$

Теорема 11. *Пусть U — выпуклое замкнутое множество из рефлексивного банахова пространства B, функция $J(u)$ равномерно выпукла и полунепрерывна снизу на U. Тогда:*

1) *множество Лебега $M(v)=\{u\in U\colon J(u)\leqslant J(v)\}$ выпукло, замкнуто и ограничено при всех $v\in U$;*

2) *$J_*>-\infty$, $U_*\neq\varnothing$, причем U_* выпукло, замкнуто и ограничено;*

3) *если, кроме того, функция $J(u)$ строго равномерно выпукла на U, то U_* состоит из единственной точки u_* и всякая минимизирующая последовательность $\{u_k\}$ сходится к точке u_* по норме B, причем*

$$\delta(\|u_k-u_*\|)\leqslant J(u_k)-J(u_*),\quad k=1,2,\ldots$$

Доказательство теорем 10, 11 полностью аналогично доказательству теорем 4.3.1, 4.7.1 соответственно. При рассмотрении возникающей по ходу доказательства задачи: $J(u)\to\inf$, $u\in M(v)$ вместо теоремы 2.1.2 здесь нужно ссылаться на теорему 9.

Теоремы 1, 4, 8–11 широко используются в различных приложениях: при доказательстве существования решения в задачах оптимального управления, при доказательстве существования элемента наилучшего приближения в теории приближения функций и т. д. Некоторые из таких приложений будут рассмотрены ниже. Более тонкие теоремы существования, учитывающие специфику конкретных классов задач оптимизации, можно найти в [132; 211; 393; 399; 465; 645; 687; 722] и др.

6. Отдельно остановимся на одном важном классе задач минимизации:

$$J(u)=\|Au-b\|_F^2\to\inf,\quad u\in U, \tag{3}$$

где A — линейный непрерывный оператор, действующий из гильбертова пространства H в гильбертово пространство F, т. е. $A\in\mathcal{L}(H\to F)$, b — заданный элемент из F, U — заданное множество из H. Функцию $J(u)=\|Au-b\|_F^2$ называют *квадратичной*. Задача (3) возникает во многих приложениях, ей посвящена обширная литература (см., например, [459; 496], библиографию к ним). В частности, если A — матрица размера $m\times n$, $U=H=E^n$, к задаче (3) приводит *метод наименьших квадратов* решения системы линейных алгебраических уравнений $Au=b$. Ниже мы увидим, что многие задачи оптимального управления линейными системами могут быть записаны в виде задачи (3).

Покажем, что функция $J(u)=\|Au-b\|_F^2$ слабо полунепрерывна снизу на H. Этот факт нетрудно установить с помощью теоремы 7, однако мы здесь приведем для него прямое доказательство. Возьмем произвольную точку $u\in H$ и произвольную последовательность $\{u_k\}$, слабо в H сходящуюся к этой точке. Покажем, что тогда последовательность $\{Au_k\}$ слабо в F сходится к точке Au. С этой целью воспользуемся равенством $\langle Au_k, c\rangle_F=\langle u_k, A^*c\rangle$, справедливым для всех $c\in F$, $k=1,2,\ldots$, где $A^*\in\mathcal{L}(F\to H)$ — сопряженный к A оператор. Перейдем в этом равенстве к пределу при $k\to\infty$. Учитывая, что $\{u_k\}\to u$ слабо в H, получим

$$\lim_{k\to\infty}\langle Au_k, c\rangle_F=\lim_{k\to\infty}\langle A^*c, u_k\rangle_H=\langle A^*c, u\rangle_H=\langle c, Au\rangle_F\quad\forall c\in F.$$

Это означает, что $\{Au_k\}\to Au$ слабо в F. Тогда $\{Au_k-b\}\to Au-b$ слабо в F. Пользуясь тем, что функция $\|f\|_F^2$ слабо полунепрерывна снизу (см. пример 5), имеем: $\lim\limits_{k\to\infty}J(u_k)=\lim\limits_{k\to\infty}\|Au_k-b\|_F^2\geqslant\|Au-b\|_F^2=J(u)$. Слабая полунепрерывность снизу функции $J(u)$ на H установлена. Отсюда и из теоремы 4 вытекает

Теорема 12. *Пусть U — слабо компактное множество из H. Тогда задача* (3) *имеет хотя бы одно решение, т. е. $J_*\geqslant 0$, $U_*=\{u\in U\colon J(u)=J_*\}\neq\varnothing$.*

Задачу (3) и теорему 12 проиллюстрируем на примерах задач оптимального управления.

Пример 14. Рассмотрим задачу:

$$J(u)=|x(T;u)-b|_{E^n}^2\to\inf,\quad u\in U, \tag{4}$$

где $x=x(t)=x(t;u)=(x^1(t),\ldots,x^n(t))$, $t_0\leqslant t\leqslant T$, — решение системы

$$\dot{x}(t)=D(t)x(t)+B(t)u(t),\quad t_0\leqslant t\leqslant T;\quad x(t_0)=0, \tag{5}$$

t — время, $u = u(t) = (u^1(t), \ldots, u^r(t))$, $t_0 \leqslant t \leqslant T$, моменты t_0, T заданы, U — заданное множество из $L_2^r[t_0, T]$, $D(t) = \{d_{ij}(t)\}$ — матрица размера $n \times n$, $B(t) = \{b_{ij}(t)\}$ — матрица размера $n \times r$, $d_{ij}(t) \in L_\infty[t_0, T]$, $b_{ij}(t) \in L_\infty[t_0, T]$ (например, $d_{ij}(t)$, $b_{ij}(t)$ — кусочно-непрерывные функции на $[t_0, T]$), b — заданный вектор из E^n. Согласно теореме 6.1.2 задача Коши (5) при каждом фиксированном $u = u(t) \in L_2^r[t_0, T]$ имеет, притом единственное решение $x(t; u)$, $t_0 \leqslant t \leqslant T$ (см. определение 6.1.1). Поэтому функция (4) определена при всех $u \in L_2^r[t_0, T]$. Задача (4), (5) имеет простой геометрический смысл: среди всех управлений $u \in U$ ищется такое, для которого правый конец $x(T; u)$ траектории был бы как можно ближе к заданной точке b. Функцию $J(u)$ из (4) часто называют *терминальной*.

Покажем, что задачу (4), (5) можно представить в виде задачи (3). С этой целью введем оператор

$$Au = x(T; u), \tag{6}$$

который каждому управлению $u = u(t) \in L_2^r[t_0, T]$ ставит в соответствие правый конец $x(T; u) \in E^n$ траектории $x(t; u)$. Из единственности решения задачи Коши (5) следует, что

$$\begin{gathered} x(t; \alpha u + \beta v) = \alpha x(t; u) + \beta x(t; v) \\ \forall t \in [t_0, T], \quad \forall u, v \in L_2^r[t_0, T], \quad \forall \alpha, \beta \in \mathbb{R}. \end{gathered} \tag{7}$$

Это означает, что оператор A, определенный согласно (6), линейный. Убедимся, что этот оператор, действующий из гильбертова пространства $H = L_2^r[t_0, T]$ в гильбертово пространство $F = E^n$, является ограниченным и, следовательно, непрерывным [393; 705]. В самом деле, из (5) имеем

$$|x(t; u)| = \left| \int_{t_0}^{t} D(\tau)x(\tau; u) + B(\tau)u(\tau)d\tau \right| \leqslant$$
$$\leqslant D_{\max} \int_{t_0}^{t} |x(\tau; u)|d\tau + B_{\max} \int_{t_0}^{T} |u(\tau)|d\tau, \quad t_0 \leqslant t \leqslant T,$$

где $D_{\max} = \|D(t)\|_{L_\infty}$, $B_{\max} = \|B(t)\|_{L_\infty}$. Отсюда, пользуясь леммой 6.3.1 при $\varphi(t) = |x(t; u)|$, $a = D_{\max}$, $b = B_{\max} \int_{t_0}^{T} |u(\tau)|d\tau$, и неравенством Коши — Буняковского получаем оценку

$$|x(t;u)|_{E^n} \leqslant e^{D_{\max}(T-t_0)} B_{\max} \int_{t_0}^{T} |u(t)|dt \leqslant C_0 \left(\int_{t_0}^{T} |u(t)|^2 dt \right)^{1/2} \quad \forall t \in [t_0, T] \tag{8}$$

при всех $u \in L_2^r[t_0, T]$; здесь $C_0 = e^{D_{\max}(T-t_0)} B_{\max} \sqrt{T - t_0}$. Отсюда при $t = T$ с учетом (6) имеем

$$|x(T; u)| = \|Au\|_{E^n} \leqslant C_0 \|u\|_{L_2^r[t_0, T]}. \tag{9}$$

Оценка (9) означает, что A — непрерывный оператор, действующий из $H = L_2^r[t_0, T]$ в $F = E^n$. Таким образом, функция (4) при условиях (5) представлена в виде $J(u) = \|Au - f\|_F^2$, и задача (4), (5) сведена к задаче (3). Отсюда и из теоремы 12 следует, что если множество U слабо компактно в $L_2^r[t_0, T]$ (см. примеры 9, 10 при $G = [t_0, T]$), то задача (4), (5) имеет хотя бы одно решение.

Пример 15. Рассмотрим задачу

$$J_1(u) = |x_1(T; u) - b_1|_{E^n}^2 \to \inf, \quad u \in U, \tag{10}$$

где $x_1 = x_1(t; u)$ — решение задачи Коши

$$\dot{x}_1(t) = D(t)x_1(t) + B(t)u(t) + f(t), \quad t_0 \leqslant t \leqslant T; \quad x_1(t_0) = x_0, \tag{11}$$

$f(t) = (f^1(t), \ldots, f^n(t)) \in L_2^n[t_0, T]$, $x_0, b \in E^n$; остальные обозначения и предположения те же, что и в примере 14. Очевидно, при $f(t) \equiv 0$, $x_0 = 0$ задача (10), (11) превращается в задачу (4), (5). Нетрудно видеть, что решение задачи Коши (11) можно записать в виде

$$\dot{x}_1(t; u) = x(t; u) + x_0(t), \quad t_0 \leqslant t \leqslant T, \tag{12}$$

где $x(t; u)$ — решение задачи (5), а $x_0(t)$ — решение задачи

$$\dot{x}_0(t) = D(t)x_0(t) + f(t), \quad t_0 \leqslant t \leqslant T; \quad x(t_0) = x_0, \tag{13}$$

получающейся из (11) при $u(t) \equiv 0$. Из (6), (12) следует, что $x_1(T, u) = Au + x_0(T)$, и функцию (10) можем представить в виде

$$J_1(u) = |x(T; u) - b|_{E^n}^2 = \|Au - b\|_{E^n}^2,$$

где $b = b_1 - x_0(T)$, $x(t;u)$ — решение задачи (5). Таким образом, задача (10), (11) сведена к задаче (4), (5). Если в (10) множество U слабо компактно в $L_2^r[t_0, T]$, то из теоремы 12 следует существование решения задачи (10), (11).

Пример 16. Задача:

$$J(u) = \int_{t_0}^{T} |x(t; u) - b(t)|_{E^n}^2 dt \to \inf, \quad u \in U, \tag{14}$$

где $x(t; u)$ — решение задачи Коши (5), $b = b(t) \in L_2^n[t_0, T]$, U — заданное множество из $L_2^r[t_0, T]$; остальные обозначения взяты из задачи (4), (5). Функцию (14), в отличие от (4), часто называют *интегральной*. Покажем, что задачу (14), (5) также можно представить в виде задачи (3). Введем оператор

$$Au = x(t; u), \quad t_0 \leqslant t \leqslant T, \tag{15}$$

который каждому управлению $u = u(t) \in L_2^r[t_0, T]$ ставит в соответствие траекторию $x(t; u)$ задачи (5). Поскольку $x(t; u)$ непрерывна на $[t_0, T]$, то $x(t; u) \in L_2^n[t_0, T]$, и мы можем считать, что оператор (15) действует из гильбертова пространства $H = L_2^r[t_0, T]$ в гильбертово пространство $F = L_2^n[t_0, T]$. Из равенства (7) следует, что этот оператор линейный. Кроме того, из неравенства (8) имеем

$$\|Au\|_F = \left(\int_{t_0}^{T} |x(t; u)|^2 dt \right)^{1/2} \leqslant C_1 \|u\|_{L_2^r[t_0, T]}, \quad C_1 = (T - t_0)C_0. \tag{16}$$

Оценка (16) означает, что $A \in \mathcal{L}(H \to F)$, где $H = L_2^r[t_0, T]$, $F = L_2^n[t_0, T]$. Таким образом, задача (14), (5) сведена к задаче (3). Если множество U слабо компактно в $L_2^r[t_0, T]$, то из теоремы 12 следует разрешимость задачи (14), (5).

Пример 17. Задача:

$$J_1(u)=\int_{t_0}^{T}|x_1(t;u)-b_1(t)|^2_{E^n}dt\to\inf,\quad u\in U, \tag{17}$$

где $x_1=x_1(t;u)$ — решение задачи Коши (11), $b(t)\in L_2^n[t_0,T]$; остальные обозначения и предположения те же, что и в примерах 14, 15. Пользуясь представлением решения $x_1(t;u)$ в виде (12) и оператором A, определенным согласно (15), функцию (17) можно представить в виде

$$J_1(u)=\int_{t_0}^{T}|x(t;u)-b(t)|^2_{E^n}dt=\|Au-b\|^2_F,$$

где $b=b(t)=b_1(t)-x_0(t)$, $x(t;u)$ — решение задачи (5), $x_0(t)$ — решение задачи (13). Таким образом, задача (17), (11) также свелась к задаче (3). Если множество U слабо компактно в $L_2^r[t_0,T]$, то разрешимость задачи (17), (11) вытекает из теоремы 12.

Если в задаче (3) оператор A обладает некоторыми дополнительными свойствами, то в теореме 12 требование на множество U может быть ослаблено.

Теорема 13. *Пусть U — выпуклое замкнутое множество из гильбертова пространства H, множество AU значений оператора A на множестве U замкнуто в H. Тогда задача* (3) *имеет хотя бы одно решение, т. е. $J_*\geqslant 0$, $U_*\neq\varnothing$.*

В отличие от теоремы 12 здесь ограниченность множества U не предполагается, поэтому U необязательно слабо компактно (см. теорему 3).

Доказательство. Так как множество U выпукло, то AU также выпукло. В самом деле, пусть $f_1, f_2\in AU$. Нам надо показать, что тогда отрезок $[f_1,f_2]=\{f_\alpha\in F\colon f_\alpha=\alpha f_1+(1-\alpha)f_2,\ 0\leqslant\alpha\leqslant 1\}\in AU$. Так как $f_1, f_2\in AU$, то существуют $u_1, u_2\in U$ такие, что $Au_1=f_1$, $Au_2=f_2$. Поскольку U выпукло, то $u_\alpha=\alpha u_1+(1-\alpha)u_2\in U\ \ \forall\alpha\in[0,1]$. В силу линейности оператора A тогда $f_\alpha=\alpha Au_1+(1-\alpha)Au_2=A(\alpha u_1+(1-\alpha)u_2)=Au_\alpha$, т. е. $f_\alpha\in AU$ $\forall\alpha\in[0,1]$. Таким образом, AU — выпуклое замкнутое множество из гильбертова пространства F. Тогда существует $b_A=P_{AU}(b)$ — проекция точки b на множество AU (см. определение 4.4.1). Этот факт вытекает из сильной выпуклости функции $g(f)=\|f-b\|^2_F$, $f\in AU$ и теоремы 10. Так как $b_A\in AU$, то $b_A=Av$ при некотором $v\in U$. Кроме того, функция $g(f)$ дифференцируема и ее производная $g'(f)=2(f-b)$ (подробнее см. § 3). Отсюда и из теорем 4.2.3, 4.4.1, которые остаются справедливыми в гильбертовых пространствах, вытекает, что $\langle b_A-b, Au-b_A\rangle\geqslant 0\ \ \forall u\in U$. Тогда

$$\begin{aligned}J(u)=\|Au-b\|^2&=\|(Au-b_A)+(b_A-b)\|^2=\\&=\|Au-b_A\|^2+2\langle b_A-b,Au-b_A\rangle+\|b_A-b\|^2\geqslant\\&\geqslant\|b_A-b\|^2=\|Av-b\|^2=J(v)\quad\forall u\in AU.\end{aligned}\tag{18}$$

Это значит, что $J_*=\inf\limits_U J(u)=J(v)$ и, следовательно, $v\in U_*=\{u\in U\colon J(u)=J_*\}$. Более того, из (18) следует, что $U_*=\{u\in U\colon Au=Av=b_A\}$. Теорема 13 доказана.

Для иллюстрации теоремы 13 рассмотрим задачу (4), (5) при $U=L_2^r[t_0,T]$. Тогда множество AU в силу линейности оператора A является

подпространством в $F=E^n$. Как известно [192], в E^n всякое подпространство замкнуто. Таким образом, в задаче (4), (5) при $U=L_2^r[t_0,T]$ множество AU замкнуто. Отсюда и из теоремы 13 следует, что функция (4) при условиях (5), $U=L_2^r[t_0,T]$ достигает своей нижней грани хотя бы на одном управлении. Аналогично доказывается разрешимость задачи (10), (11) при $U=L_2^r[t_0,T]$.

Отметим, что задачи (14), (5) и (17), (11) при $U=L_2^r[t_0,T]$, вообще говоря, не имеют решения. Покажем это на примере.

Пример 18.

$$J(u)=\int_0^1|x(t;u)-1|^2dt\to\inf,\quad u=u(t)\in U=L_2[0,1],$$
$$\dot x(t)=u(t),\quad 0\leqslant t\leqslant 1;\quad x(0)=0.$$

Очевидно, если $u=u(t)\equiv 0$, то $J(0)>0$, если $u=u(t)\neq 0$, то также $J(u)>0$. Рассмотрим последовательность

$$u_k=u_k(t)=\begin{cases}k, & t\in[0,\frac1k],\\ 0, & t\in(\frac1k,1].\end{cases}$$

Ей соответствуют траектории

$$x_k(t)=x(t;u_k)=\begin{cases}kt, & t\in[0,\frac1k],\\ 1, & t\in(\frac1k,1].\end{cases}$$

Следовательно, $J(u_k)=\int\limits_0^{1/k}(kt-1)^2dt=\frac1k\to 0$ при $k\to\infty$. Таким образом, последовательность $\{u_k\}$ в этой задаче является минимизирующей, $J_*=0$, однако нижняя грань не достигается. Остается заметить, что здесь множество AU не замкнуто в $F=L_2[0,1]$: траектории $x_k(t)\to 1$ в норме $L_2[0,1]$, но $x(t)\equiv 1$ не является решением задачи $\dot x(t)=u(t)$, $x(0)=0$ ни при каком управлении $u(t)\in L_2[0,1]$.

7. В различных руководствах по функциональному анализу [258; 348; 371; 393; 772] читатель обнаружит и другие, отличные от приведенных выше, варианты теоремы Вейерштрасса, основанные на других определениях понятий компактности и полунепрерывности функций. Для того, чтобы облегчить читателю ориентацию в этих вопросах, совершим небольшой экскурс в топологические пространства. Сначала напомним некоторые определения [371; 393; 772].

Определение 6. Пусть X — некоторое множество. Говорят, что на X задана *топология*, если в X выделена система τ его подмножеств, удовлетворяющая следующим трем условиям (аксиомам): 1) $\varnothing, X\in\tau$; 2) объединение любого числа множеств из τ является множеством из τ; 3) пересечение конечного числа множеств из τ является множеством из τ. Все множества $G\in\tau$ называются *открытыми*, а их дополнения $F=X\setminus G$ — *замкнутыми*. Множество X с заданной на нем топологией τ называют *топологическим пространством* и обозначают через (X,τ).

Заметим, что всякое метрическое пространство M превращается в топологическое пространство, если открытым считать всякое множество $G\subseteq M$, которое вместе с каждой своей точкой $u\in G$ содержит и некоторую ε-окрестность этой точки.

Определение 7. Пусть (X,τ) — топологическое пространство. *Окрестностью точки* $u\in X$ называется всякое множество $V\subset X$, содержащее некоторое множество $G\subset\tau$, которое в свою очередь содержит точку u. *Окрестностью множества* $U\subset X$ называется всякое множество $V\subset X$, содержащее множество $G\subset\tau$, которое в свою очередь содержит U. Точка $u\in X$ называется *точкой прикосновения* множества $U\subset X$, если каждая окрестность точки u содержит хотя бы одну точку из U. Совокупность всех точек прикосновения множества U называется *замыканием множества* U и обозначается через $\overline{U}_X$ или просто $\overline{U}$. Пусть $U\subset W\subset X$. Совокупность всех точек прикосновения множества U, взятых из множества W, называется *замыканием U во множестве W* и обозначается через $\overline{U}_W$.

Можно показать, что $U=\overline{U}_X$ тогда и только тогда, когда U — замкнутое множество [393].

Определение 8. Пусть (X, τ) — топологическое пространство, U — множество из X. Точка $u \in X$ называется *предельной точкой* множества U, если каждая окрестность точки u содержит хотя бы одну точку из U, отличную от точки u.

Определение 9. Пусть (X, τ) — топологическое пространство. Говорят, что последовательность $\{u_k\} \in X$ *сходится к точке* $u \in X$ *в топологии* τ или, короче, $\{u_k\}$ *τ-сходится к точке* u, если для любой окрестности G точки u найдется номер $N = N(G)$ такой, что $u_k \in G$ при всех $k \geqslant N$; точку u называют τ-пределом последовательности $\{u_k\}$. Последовательность $\{u_k\}$ τ-сходится к множеству U, если для любой окрестности G множества U найдется номер N такой, что $u_k \in G$ для всех $k \geqslant N$.

Топологическое пространство (X, τ) называется *отделимым* (или *хаусдорфовым*), если для любых двух различных точек u_1 и $u_2 \in X$ найдутся окрестность G_1 точки u_1 и окрестность G_2 точки u_2 такие, что $G_1 \cap G_2 = \varnothing$. В отделимых топологических пространствах всякая сходящаяся последовательность имеет единственный предел.

Банахово пространство B превращается в топологическое пространство, если в нем открытые множества ввести как объединение любого числа открытых шаров $O(u, \varepsilon) = \{v \in B: \|v - u\| < \varepsilon\}$, где u — произвольная точка из B, ε — произвольное положительное число. Введенная таким образом топология называется *сильной топологией* банахова пространства B. Сходимость последовательности $\{u_k\}$ к точке u в сильной топологии B эквивалентна сходимости этой последовательности к той же точке по норме (в метрике) B.

В банаховых пространствах могут быть введены и другие топологии. Для нас наибольший интерес представляет так называемая *слабая топология*. Открытыми множествами в слабой топологии банахова пространства B называются множества, представимые в виде объединения любого числа множеств вида

$$G(u, c_1, \ldots, c_m, \varepsilon_1, \ldots, \varepsilon_m) = \{v \in B: |\langle c_i, v\rangle - \langle c_i, u\rangle| \leqslant \varepsilon_i,\ i = 1, \ldots, m\}, \qquad (19)$$

где u — произвольная точка из B, m — произвольное натуральное число, $c_1, \ldots, c_m$ — произвольные элементы из сопряженного пространства B^*, $\varepsilon_1, \ldots, \varepsilon_m$ — произвольные положительные числа. Заметим, что сходимость последовательности $\{u_k\}$ в слабой топологии B равносильна слабой сходимости $\{u_k\}$ в смысле определения 1.1.

Следует заметить, что в общем случае топологические пространства могут иметь весьма замысловатую структуру, в них многие «привычные» представления могут нарушаться. Так, например, мы «привыкли» к тому, что в метрическом пространстве M точка u является предельной точкой множества $U \subset M$ тогда и только тогда, когда существует последовательность $\{u_k\} \in U$, $u_k \neq u$, $k = 1, 2, \ldots$, которая сходится к точке u в метрике M. Однако в топологических пространствах это, вообще говоря, не так: предельная точка множества может не быть τ-пределом какой-либо последовательности, принадлежащей этому множеству. Рассмотрим

Пример 19. В пространстве l_2 в качестве топологии возьмем слабую топологию. В этом пространстве возьмем множество U, состоящее из точек $u_k = \sqrt{k}\, e_k$, $k = 1, 2, \ldots$, где $e_k = (0, \ldots, 1, 0, \ldots)$ с единицей на k-м месте. Покажем, что точка $u = 0 \in l_2$ является предельной точкой множества U в слабой топологии l_2 (см. определение 8). Будем рассуждать от противного: пусть $u = 0$ не является предельной точкой. Это означает, что существует окрестность G_0 точки $u = 0$, которая не содержит ни одной точки множества U. Согласно (19) тогда найдутся натуральное число m, числа $\varepsilon_1 > 0, \ldots, \varepsilon_m > 0$, элементы $c_1, \ldots, c_m \in l_2$, такие, что $u_k \notin G_0 = G(u = 0, c_1, \ldots, c_m, \varepsilon_1, \ldots, \varepsilon_m)$ при всех $k = 1, 2, \ldots$, т. е. $|\langle c, u_k\rangle| = \left|\sum\limits_{j=1}^{\infty} c_k^j u_k^j\right| = c_i^k \sqrt{k} \geqslant \varepsilon_0 = \min\{\varepsilon_1, \ldots, \varepsilon_m\} > 0$. Тогда $|c_0^k| \geqslant \dfrac{\varepsilon_0}{\sqrt{k}}$, $k = 1, 2, \ldots$, $i = 1, \ldots, m$, т. е. $\|c_i\|^2 = \sum\limits_{k=1}^{\infty} c_i^k \geqslant \sum\limits_{k=1}^{\infty} \left(\dfrac{\varepsilon_0}{\sqrt{k}}\right)^2 = +\infty$. Однако это противоречит условию $c_i \in l_2$, $i = 1, \ldots, m$. Следовательно, $u = 0$ — предельная точка множества U. Остается заметить, что никакая подпоследовательность $\{u_{k_m}\} \in U$ не может сходиться слабо в l_2 к точке $u = 0$, так как $\{u_{k_m}\}$ не ограничена в метрике l_2.

Отметим, что для множества $V = U \cup \{0\}$ точка $u = 0$ также является предельной и единственной последовательностью $\{v_k\} \in V$, которая слабо в l_2 сходится к точке $u = 0$ и является стационарной последовательностью $v_k = 0$, $k = 1, 2, \ldots$

Приведем несколько вариантов теоремы Вейерштрасса в топологических пространствах. Начнем с формулировки различных определений понятия компактности, общепринятых в топологии [258; 371; 393; 772].

Определение 10. Множество U из топологического пространства (X, τ) называется *относительно секвенциально компактным*, если из любой последовательности $\{u_k\} \in U$ можно выбрать хотя бы одну подпоследовательность, τ-сходящуюся к некоторой точке $v \in X$.

Если при этом любая такая точка v принадлежит самому множеству U, то такое множество называется *секвенциально компактным*.

Определение 11. Множество U из топологического пространства (X, τ) называется *относительно счетно-компактным*, если каждое его бесконечное подмножество имеет хотя бы одну предельную точку $v \in X$. Если любая такая точка v принадлежит самому множеству U, то такое множество называется *счетно-компактным*.

Определение 12. Система множеств $\{G_\alpha\}$ из топологического пространства (X, τ) называется *покрытием* множества U из X, если $U \in \bigcup\limits_{\alpha} G_\alpha$. Покрытие $\{G_\alpha\}$ называется *открытым*, если все G_α — открытые множества. Если некоторое подсемейство $\{G_{\alpha\beta}\}$ покрытия $\{G_\alpha\}$ само образует покрытие U, то $\{G_{\alpha\beta}\}$ называется *подпокрытием* покрытия $\{G_\alpha\}$. Множество U из X называется *компактным*, если из любого открытого покрытия $\{G_\alpha\}$ множества U можно выбрать конечное подпокрытие $\{G_{\alpha_1}, G_{\alpha_2}, \ldots, G_{\alpha_n}\}$. Множество U называется *относительно компактным*, если его замыкание $\overline{U}_X$ компактно.

В метрических пространствах (в частности, в E^n и в банаховых пространствах с их сильной топологией) все три понятия компактности равносильны [371, стр. 43]. Замечательно и то, что в слабой топологии банаховых пространств эти три понятия компактности также равносильны (теория Эберлейна — Шмульяна [371], стр. 288–292). В других топологических пространствах эти понятия, вообще говоря, не будут равносильными, и можно лишь сказать, что в общем случае из компактности (определение 12) и секвенциальной компактности (определение 10) следует счетная компактность [371; 393].

Определение 13. Функция $J(u)$ называется *секвенциально полунепрерывной снизу в точке* $u \in U$, если для любой последовательности $\{u_k\} \in U$, τ-сходящейся к точке u, справедливо неравенство $\varliminf\limits_{k \to \infty} J(u_k) \geqslant J(u)$. Функция $J(u)$ называется *секвенциально полунепрерывной снизу на множестве* U, если она секвенциально полунепрерывна снизу в каждой точке $u \in U$.

Теорема 14. *Пусть множество U из топологического пространства (X, τ) секвенциально компактно, а функция $J(u)$ определена на множестве U, принимает конечные значения и секвенциально полунепрерывна снизу на U. Тогда $J_* = \inf\limits_U J(u) > -\infty$, множество $U_* = \{u \in U: J(u) = J_*\}$ непусто, секвенциально компактно, и любая минимизирующая последовательность τ-сходится к U_*.*

Доказательство этой теоремы аналогично доказательству теоремы 4.

Определение 14. Функция $J(u)$ называется *полунепрерывной снизу в точке* $u \in U$, если для любого числа $\varepsilon > 0$ найдется окрестность $G_\varepsilon \subset \tau$ точки u такая, что $J(v) > J(u) - \varepsilon$ для всех $v \in G_\varepsilon \cup U$. Функция $J(u)$ называется *полунепрерывной снизу на множестве* U, если она полунепрерывна снизу в каждой точке $u \in U$.

Можно доказать, что если U — замкнутое множество, то функция $J(u)$ полунепрерывна снизу на U тогда и только тогда, когда множество $M(c) = \{u \in U: J(u) \leqslant c\}$ замкнуто при каждом c (ср. с леммой 2.1.1).

Теорема 15. *Пусть множество U из топологического пространства (X, τ) компактно (определение 12), а функция $J(u)$ определена на множестве U, принимает конечные значения и полунепрерывна снизу на U (определение 14). Тогда $J_* = \inf\limits_U J(u) > -\infty$, множество $U_* = \{u \in U: J(u) = J_*\}$ непусто, компактно.*

Доказательство. На множестве U введем топологию, индуцированную топологией τ пространства (X, τ). А именно, открытыми множествами в U назовем все множества вида $G \cap U$, где $G \subset \tau$. Нетрудно проверить, что все аксиомы топологического пространства здесь выполняются. Получившееся топологическое пространство обозначим через (U, τ). Множества $M(c) = \{u \in U: J(u) \leqslant c\}$, замкнутые в (X, τ), будут замкнутыми и в (U, τ), так что функция $J(u)$ полунепрерывна снизу на U и в топологии пространства (U, τ). Далее заметим, что произвольное открытое покрытие $\{G_\xi\}$ множества U в пространстве (X, τ), порождает открытое покрытие $\{G_\xi \cap U\}$ множества U в пространстве (U, τ). Поскольку множество U компактно в (X, τ), то из $\{G_\xi\}$ можно выбрать конечное подпокрытие $G_{\xi_1}, \ldots, G_{\xi_m}$. Тогда множества $G_{\xi_1} \cap U, \ldots, G_{\xi_m} \cap U$ образуют конечное покрытие множества U в пространстве (U, τ). Это означает, что множество U компактно и в (U, τ).

Покажем, что $J_* > -\infty$. Зафиксируем какую-либо числовую последовательность $\{a_k\} \to -\infty$. Так как множества $A_k = \{u \in U: J(u) \leqslant a_k\}$ замкнуты в силу полунепрерывности снизу функции $J(u)$, то множества $B_k = U \setminus A_k = \{u \in U: J(u) > a_k\}$, $k = 1, 2, \ldots$ открыты в (U, τ). Кроме того, произвольная точка $v \in U$ будет принадлежать множествам B_k для всех

номеров k, для которых $J(v) > a_k$. Следовательно, система $\{B_k\}$ образует открытое покрытие множества U. В силу компактности U тогда существует конечное подпокрытие $B_{k_1}, \ldots, B_{k_m}$ множества U. Тогда $J(u) \geqslant \min\limits_{1 \leqslant i \leqslant m} a_{k_i} > -\infty$ $\forall u \in U$. Следовательно, $J_* \geqslant \min\limits_{1 \leqslant i \leqslant m} a_{k_i} > -\infty$.

Допустим, что функция $J(u)$ не достигает своей нижней грани J_* на множестве U, т. е. $J(u) > J_*$ $\forall u \in U$. Зафиксируем какую-либо положительную последовательность $\{\varepsilon_k\} \to 0$. Как и выше, устанавливаем, что множества $C_k = \{u \in U\colon J(u) > J_* + \varepsilon_k\}$, $k = 1, 2, \ldots$, образуют открытое покрытие множества U. В силу компактности U тогда найдется конечное подпокрытие $C_{k_1}, \ldots, C_{k_p}$ множества U. Поэтому $J(u) \geqslant J_* + \min\limits_{1 \leqslant i \leqslant p} \varepsilon_{k_i} = J_* + \varepsilon$, $\varepsilon > 0$, $\forall u \in U$, что противоречит определению нижней грани J_*. Полученное противоречие показывает, что $U_* \neq \varnothing$. Теорема 15 доказана. □

Читатель, конечно, заметил, что выше при формулировке и доказательстве теорем Вейерштрасса 1, 4, 8–13 этого параграфа мы пользовались понятиями секвенциальной компактности множества и секвенциальной полунепрерывности снизу функции в соответствующих топологических пространствах. Это обстоятельство может быть оправдано двумя причинами: во-первых, проверку секвенциальной компактности множества и секвенциальной полунепрерывности снизу функции во многих прикладных задачах можно осуществить гораздо проще, чем других понятий компактности и полунепрерывности снизу, и, во-вторых, для задач оптимизации, рассматриваемых в метрических и банаховых пространствах с их сильной и слабой топологией, в которых, как было замечено, все три понятия компактности равносильны; принятый выше в пп. 2–6 подход не приводит к потере общности.

Упражнения

1. Показать, что функция $J(u) = \int\limits_0^1 u^2(t)\,dt$ непрерывна на $C[0, 1]$ в метрике $C[0, 1]$. Установить, что на множестве $U = \{u = u(t) \in C[0, 1]\colon u(0) = 0, u(1) = 1\}$ эта функция не достигает своей нижней грани $J_* = 0$. Ограничена ли эта функция на U сверху?

2. Пусть $J(u) = \int\limits_0^1 u(t)\,dt - u\left(\frac{1}{2}\right)$, $u = u(t) \in C[0, 1]$. Доказать, что: 1) эта функция линейна и непрерывна на $C[0, 1]$ в метрике этого пространства; 2) на единичном шаре $U = \{u = u(t) \in C[0, 1]\colon |u(t)| \leqslant 1, 0 \leqslant t \leqslant 1\}$ эта функция ограничена сверху и снизу, но не достигает на U ни нижней грани $J_* = -2$, ни верхней грани $J^* = 2$.

3. Доказать, что в задачах оптимального управления из примеров 14–17 целевые функции слабо непрерывны на $L_2^r[t_0, T]$. Указание: установить, что если последовательность $\{u_k\} \in L_2^r[t_0, T]$ сходится слабо в $L_2^r[t_0, T]$ к функции $u = u(t)$, то последовательности $\{x(t; u_k)\}$, $\{x_1(t; u_k)\}$ траекторий задач (5), (11) соответственно сходятся к $x(t; u)$, $x_1(t; u)$ равномерно на $[t_0, T]$.

4. Доказать, что в задачах оптимального управления из примеров 14–17 целевые функции достигают своей верхней грани на каждом выпуклом замкнутом ограниченном множестве из $L_2^r[t_0, T]$.

5. Выяснить, каким дополнительным условиям должны удовлетворять функции $b(t), b_1(t) \in L_2^n[t_0, T]$, чтобы задачи (14), (5) и (17), (11) были разрешимы на множестве $U = L_2^r[t_0, T]$.

6. Пусть $J(u) = \int\limits_0^1 f(u(t))\,dt$. 1) Доказать, что если $f(u)$ непрерывна на E^1, то функция $J(u)$ непрерывна на $C[0, 1]$ в метрике этого пространства.

2) Доказать, что если $|f(u+s) - f(u)| \leqslant C(|u||s| + |s|^2)$ при всех $u, s \in E^1$, $C = \text{const} \geqslant 0$, то функция $J(u)$ непрерывна на $L_2[0, 1]$ в метрике $L_2[0, 1]$.

3) Если $f(u)$ полунепрерывна снизу на E^1, то будет ли $J(u)$ полунепрерывной снизу в метрике $C[0, 1]$ или $L_2[0, 1]$?

4) Доказать, что если $f(u)$ выпукла [сильно выпукла] на E^1, то $J(u)$ выпукла [сильно выпукла] на $L_2[0, 1]$.

5) Доказать, что если $f(u)$ выпукла и удовлетворяет условию п. 2), то $J(u)$ достигает своей нижней грани на любом выпуклом замкнутом ограниченном множестве U из $L_2[0, 1]$.

7. Доказать, что в рефлексивном банаховом пространстве B шар $\{u \in B\colon \|u - u_0\| \leqslant R\}$ слабо компактен.

8. Доказать, что в рефлексивном банаховом пространстве B справедливо равенство $\|c\|_{B^*} = \max\limits_{\|u\| \leqslant 1} \langle c, u\rangle$ при всех $c \in B^*$.

9. Доказать, что функция $J(u) = \|u - \overline{u}\|$ на любом выпуклом замкнутом множестве U рефлексивного банахова пространства достигает своей нижней грани (иначе говоря, существует элемент наилучшего приближения из U для заданного элемента $\overline{u} \in B$).

10. Пусть оператор $A \in \mathcal{L}(B \to F)$, где B, F — банаховы пространства, B — рефлексивно, элемент $b \in F$ задан, U — выпуклое замкнутое ограниченное множество из B. Доказать, что функция $J(u) = \|Au - b\|_F^2$ на множестве U достигает своей нижней грани (ср. с теоремой 12).

11. В пространстве $L_p^r(G)$, $1 < p < \infty$, рассмотрим два множества: $U_1 = \{u = u(t) = (u^1(t), \ldots, u^r(t)) \in L_p^r(G)\colon \alpha_i(t) \leqslant u^i(t) \leqslant \beta_i(t)$ для почти всех $t \in G$, $i = 1, \ldots, r\}$, $U_2 = \{u = u(t) \in L_p^r(G)\colon \int\limits_G |u(t) - \overline{u}(t)|^p\,dt \leqslant R^p\}$, где функции $\overline{u} = \overline{u}(t) \in L_p^r(G)$, $\alpha_i(t), \beta_i(t) \in L_p(G)$, $i = 1, \ldots, r$, и число $R > 0$ заданы. Доказать, что эти множества слабо компактны в $L_p^r(G)$, $1 < p < \infty$, но не компактны в метрике этого пространства. Будут ли эти множества слабо компактными в $L_1^r(G)$ или $L_\infty^r(G)$? (см. примеры 7, 9, 10).

12. Доказать, что функция $J(u) = \int\limits_0^1 u^4(t)\,dt$ не является непрерывной в метрике $L_2[0, 1]$. Будет ли она полунепрерывной снизу в метрике $L_2[0, 1]$?

13. Доказать, что функция $J(u) = \int\limits_0^1 |\dot{u}(t)|^2\,dt$, определенная на $H^1[0, 1]$, разрывна в метрике $C[0, 1]$. Будет ли она полунепрерывной снизу в этой метрике?

14. Пусть P — линейное нормированное пространство всех алгебраических многочленов на отрезке $[0, 1]$ с нормой $\|u(t)\| = \max\limits_{0 \leqslant t \leqslant 1} |u(t)|$. Положим $J(u) = \sum\limits_{i=0}^n |\alpha_i|$ для $u = u(t) = \sum\limits_{i=0}^n \alpha_i t^i \in P$. Доказать, что $J(u)$ выпукла на P, но не является непрерывной на P. Указание: рассмотреть последовательность $u_k = u_k(t) = t^k - t^{k+1}$, $0 \leqslant t \leqslant 1$. Будет ли $J(u)$ полунепрерывной снизу на P?

15. Пусть $J(u)$ — выпуклая функция на открытом выпуклом множестве U банахова пространства B. Доказать, что следующие четыре утверждения эквивалентны: 1) $J(u)$ полунепрерывна сверху в точке $v \in U$ в метрике B; 2) $J(u)$ непрерывна в точке v в метрике B; 3) $J(u)$ ограничена в некоторой окрестности точки v; 4) $J(u)$ ограничена сверху в некоторой окрестности точки v [233].

§ 3. Дифференцирование. Условия оптимальности

При исследовании экстремальных задач в банаховых пространствах, как и в случае n-мерного пространства E^n, важную роль играет понятие производной (градиента) функции.

1. Сначала приведем определение более общего понятия производной отображения.

Определение 1. Пусть X и Y — два нормированных пространства и F — отображение, действующее из X в Y и определенное в окрестности $O(u, \gamma) = \{v \in X\colon \|v - u\|_X < \gamma\}$ точки u. Говорят, что отображение F *дифференцируемо* в точке u, если существует линейный непрерывный оператор $F'(u) \in \mathcal{L}(X \to Y)$ такой, что

$$F(u + h) - F(u) = F'(u)h + \alpha(u, h) \quad \forall h, \quad \|h\|_X < \gamma, \tag{1}$$

где $\lim\limits_{\|h\|_X \to 0} \frac{\|\alpha(u,h)\|_Y}{\|h\|_X} = 0$. Оператор $F'(u)$ называется *производной (производной Фреше)* отображения F в точке u.

Если производная существует, то она определяется однозначно. В самом деле, если $F_1'(u)$ и $F_2'(u)$ — производные отображения F в точке u, то из (1) имеем

$$F_1'(u)h - F_2'(u)h = (F_1'(u) - F_2'(u))h = \alpha_1(u,h) - \alpha_2(u,h)$$

при всех h, $\|h\| < \gamma$. Возьмем произвольный элемент $e \in X$, $e \neq 0$, и в предыдущем равенстве положим $h = te$, $0 < t < t_0 = \gamma/\|e\|$. Получим $t(F_1'(u) - F_2'(u))e = o(t)$ или $(F_1'(u) - F_2'(u))e = \frac{o(t)}{t}$. Отсюда, учитывая, что $\lim\limits_{t \to 0} \frac{o(t)}{t} = 0$, имеем $F_1'(u)e = F_2'(u)e$ $\forall e \in X$. Это означает, что $F_1(u) = F_2(u)$.

Из (1) следует, что дифференцируемое в точке u отображение непрерывно в этой точке. Перечислим некоторые простейшие свойства производной:

1) если $F(u) = y_0$ $\forall u \in X$, где y_0 — фиксированная точка из Y, то $F'(u) \equiv 0$ $\forall u \in X$;

2) если $F(u) = Au$, где $A \in \mathcal{L}(X \to Y)$, то $F'(u) = A$ $\forall u \in X$;

3) если отображения F и G, действующие из X в Y, дифференцируемы в точке u, то отображение $\alpha F + \beta G$, где $\alpha, \beta \in \mathbb{R}$, также дифференцируемо в точке u, причем

$$(\alpha F + \beta G)'(u) = \alpha F'(u) + \beta G'(u).$$

Приведем обобщение известной из классического анализа формулы производной сложной функции [393]. Пусть X, Y, Z — три нормированных пространства, $O(u, \gamma_1)$ — окрестность точки $u \in X$, отображение F отображает окрестность $O(u, \gamma_1)$ в Y, $y = F(u)$, $O(y, \gamma_2)$ — окрестность точки $y \in Y$, и отображение G отображает окрестность $O(y, \gamma_2)$ в Z. Тогда, если отображение F, дифференцируемое в точке u, а G — дифференцируемо в точке y, то отображение GF, которое определено в окрестности $O(u, \gamma_1)$, дифференцируемо в точке u, причем

$$(GF)'(u) = G'(y)F'(u). \tag{2}$$

Производная $F'(u)$ отображения F согласно определению 1 является отображением, действующим из нормированного пространства X в нормированное пространство $Y_1 = \mathcal{L}(X \to Y)$, и, в свою очередь, можно поставить вопрос о его дифференцируемости.

Определение 2. Пусть отображение F, действующее из X в Y, дифференцируемо в каждой точке некоторой окрестности $O(u, \gamma) = \{v \in X: \|v - u\|_X < \gamma\}$ точки u. Говорят, что отображение *дважды дифференцируемо в точке* u, если $F'(u)$ дифференцируемо в этой точке (определение 1), т. е. существует линейный непрерывный оператор $F''(u) \in \mathcal{L}(X \to Y_1) = \mathcal{L}(X \to \mathcal{L}(X \to Y))$ такой, что

$$F'(u+g) - F'(u) = F''(u)g + \alpha_1(u,g) \quad \forall g \in X,$$
$$\|g\|_X < \gamma, \quad \lim_{\|g\|_X \to 0} \frac{\|\alpha_1(u,g)\|_{Y_1}}{\|g\|_X} = 0. \tag{3}$$

Оператор $F''(u)$ называют *второй производной* отображения F в точке u.

Можно показать, что $F''(u)$ — симметричный оператор, т. е. $(F''(u)g)h = (F''(u)h)g$ [768].

Аналогично определяются производные третьего и более высокого порядков [393; 768; 705]. В настоящей книге мы ограничимся рассмотрением лишь первой и второй производных.

Пусть отображение F определено и дважды дифференцируемо в шаре $O(u, \gamma) = \{v \in X: \|v - u\| < \gamma\}$, причем $F''(u)$ непрерывна в точке u. Тогда справедлива формула Тейлора [768, стр. 156]:

$$F(u+h) - F(u) = F'(u)h + \tfrac{1}{2}(F''(u)h)h + o(\|h\|^2) \quad \forall h \in X, \quad \|h\|_X < \gamma. \tag{4}$$

2. Переформулируем определения 1, 2 на случай, когда $X = B$ — банахово пространство, $Y = E^1$ — числовая ось, т. е. отображение F представляет из себя функцию (функционал), принимающую вещественные значения. Тогда $\mathcal{L}(X \to Y) = \mathcal{L}(B \to E^1) = B^*$ — пространство линейных непрерывных функционалов, определенных на B. Напоминаем, что через $\langle c, u \rangle$ мы обозначаем значения линейного функционала $c \in B^*$ на элементе $u \in B$. В рассматриваемом случае определения 1, 2 перепишутся в следующей форме.

Определение 3. Пусть B — банахово пространство, пусть функция $J(u)$ определена в некоторой γ-окрестности $O(u, \gamma) = \{v \in B: \|v - u\| < \gamma\}$ точки u. Говорят, что функция $J(u)$ *дифференцируема* в точке u, если существует элемент $J'(u) \in B^*$ такой, что приращение функции $J(u)$ можно представить в виде

$$\Delta J(u) = J(u+h) - J(u) = \langle J'(u), h \rangle + \alpha(u,h), \quad \forall h \in B,$$
$$\|h\| < \gamma, \quad \lim_{\|h\|_X \to 0} \frac{\alpha(u,h)}{\|h\|} = 0. \tag{5}$$

Элемент $J'(u)$ называется *первой производной* (*производной Фреше*) или *градиентом* функции $J(u)$ в точке u.

В том случае, когда $B = H$ — гильбертово пространство, то $H^* = H$ и первая производная $J'(u) \in H$, и величина $\langle J'(u), h \rangle$ превращается в скалярное произведение в H.

Определение 4. Пусть B — банахово пространство, функция $J(u)$ дифференцируема в каждой точке некоторой γ-окрестности $O(u, \gamma) = \{v \in B: \|v - u\| < \gamma\}$ точки u. Говорят, что функция $J(u)$ *дважды дифференцируема* в точке u, если существует линейный непрерывный оператор $J'' \in \mathcal{L}(B \to B^*)$ такой, что

$$F'(u+g) - F'(u) = F''(u)g + \alpha_1(u,g) \quad \forall g \in B,$$
$$\|g\| < \gamma, \quad \lim_{\|g\| \to 0} \frac{\alpha_1(u,g)}{\|g\|} = 0. \tag{6}$$

Оператор J'' называют *второй производной* функции $J(u)$ в точке u.

Если $B = H$ — гильбертово пространство, то $J'' \in \mathcal{L}(H \to H)$.

Определение 5. Функция $J(u)$ называется *непрерывно дифференцируемой* [*дважды непрерывно дифференцируемой*] *на множестве* U из банахова пространства B, если она дифференцируема [дважды дифференцируема] во всех точках $u \in U$ и $\|J'(u+h) - J'(u)\|_{B^*} \to 0$ [$\|J''(u+h) - J''(u)\|_{\mathcal{L}(B \to B^*)} \to 0$] при $\|h\| \to 0$. Множество всех функций, непрерывно дифференцируемых на U, будем обозначать через $C^1(U)$, дважды непрерывно дифференцируемых на U — через $C^2(U)$.

Если $J(u)\in C^2(U)$, то справедлива формула (4), которая здесь запишется в виде

$$J(u+h)-J(u)=\langle J'(u),h\rangle+\tfrac{1}{2}\langle J''(u)h,h\rangle+o(\|h\|^2),\quad \|h\|<\gamma \qquad (7)$$

(ср. с формулой (2.2.5)).

Замечание 1. В определении 3 предполагается, что если функция $J(u)$ дифференцируема в точке u, то она определена в некоторой окрестности этой точки. Поэтому, говоря о принадлежности функции $J(u)$ множеству $C^1(U)$, обычно подразумевают существование некоторого открытого множества W из B, которое содержит U и на котором определена эта функция. Аналогично, если $J(u)\in C^2(U)$, то считается, что $J'(u)$ определена на открытом множестве W, содержащем U.

Замечание 2. Формулы (4), (7) нередко используются для определения второй производной. Однако при этом надо сознавать, что определение второй производной из формул (4), (7), вообще говоря, отличается от определений 2, 4. Рассмотрим

Пример 1. Пусть $J(u)=u^3\sin\frac{1}{u}$ при $u\neq 0$, $J(0)=0$. При любом $h\neq 0$ имеем

$$J(0+h)-J(0)=J(h)=h^3\sin\frac{1}{u}=0\cdot h+\frac{1}{2}0\cdot h^2+o(|h|^2),$$

так что здесь формула (7) справедлива. Отсюда может показаться, что функция $J(u)$ имеет производную $J''(0)$, причем $J''(0)=0$. Однако нетрудно убедиться, что у этой функции вторая производная в точке $u=0$ не существует.

Следуя [768, стр. 161], приведем достаточные условия, когда оба подхода к определению производной $F''(u)$ совпадают. Пусть при всех $u, u+h\in O(u_0,\gamma)=\{v\in X\colon |v-u_0|<\gamma\}$ выполняется соотношение

$$F(u+h)-F(u)=a_1(u)h+(a_2(u)h)h+R_2(u,h),$$

где операторы $a_1(u)\in\mathcal{L}(X\to Y)$, $a_2(u)\in\mathcal{L}(X\to\mathcal{L}(X\to Y))$ ограничены и непрерывны в $O(u_0,\gamma)$, $a_2(u)$ — симметричный оператор, остаточный член $R_2(u,h)$ обладает следующим свойством: для $\forall\varepsilon>0\ \exists\delta>0$, что $\|R_2(u,h)\|\leqslant\varepsilon\|h\|^2\ \forall u, u+h\in O(u_0,\gamma)$. Тогда отображение F дважды дифференцируемо в $O(u_0,\gamma)$ (в смысле определения 2) и $F'(u)=a_1(u)$, $F''(u)=2a_2(u)$.

3. Если функция $J(u)\in C^1(U)$ или $C^2(U)$, точка $u+th\in U$ при всех t, $0\leqslant t\leqslant 1$, то функция $f(t)=J(u+th)$ переменной t принадлежит $C^1[0,1]$ или $C^2[0,1]$ соответственно, причем, как следует из определений 1–4 и формулы (2):

$$f'(t)=\langle J'(u+th),h\rangle,\quad f''(t)=\langle J''(u+th)h,h\rangle,\quad 0\leqslant t\leqslant 1. \qquad (8)$$

Справедливы формулы конечных приращений

$$J(u+h)-J(u)=\int_0^1\langle J'(u+th),h\rangle dt=\langle J'(u+\theta_1 h),h\rangle=$$
$$=\langle J'(u),h\rangle+\tfrac{1}{2}\langle J''(u+\theta_2 h)h,h\rangle,\quad 0<\theta_1,\theta_2<1, \qquad (9)$$

$$\langle J'(u+h)-J'(u),h\rangle=\langle J''(u+\theta_3 h)h,h\rangle,\quad 0<\theta_3<1, \qquad (10)$$

$$J'(u+h)-J'(u)=\Big(\int_0^1 J''(u+th)dt\Big)h, \qquad (11)$$

которые следуют из (8) также, как и аналогичные формулы (2.6.2)–(2.6.5) (см. ниже упражнение 25). Заметим, что интеграл в правой части (11) понимается как предел в норме пространства $\mathcal{L}(B\to B^*)$ интегральных сумм Римана функции $J''(u+th)$ переменной $t\in[0,1]$ (подробности см. в [393, гл. 10]. В частности, если U — выпуклое множество из банахова пространства, то формулы (9)–(11) справедливы для всех $u, u+h\in U$. Опираясь на формулу (9), нетрудно убедиться, что лемма 2.6.1 остается верной и в банаховых пространствах, нужно лишь в определении 2.6.1 класса функций $C^{1,1}(U)$ условие Липшица (2.6.6) понимать в смысле неравенства: $\|J'(u)-J'(v)\|_{B^*}\leqslant L\|u-v\|_B\ \forall u,v\in U$. Свойства выпуклых и сильно выпуклых функций, описанные в теоремах 4.2.1, 4.2.2, 4.2.4, 4.3.3 и их доказательства остаются справедливыми и в банаховых пространствах. Критерии выпуклости и сильной выпуклости функции, составляющие содержание теорем 4.2.5, 4.4.4, в банаховых пространствах следует переформулировать несколько иначе.

Теорема 1. *Пусть U — выпуклое множество из банахова пространства B [гильбертова пространства H], функция $J(u)\in C^2(U)$ и, кроме того,*

$$\langle J''(u)h,h\rangle\geqslant 0\quad\forall h\in B,\quad\forall u\in U \qquad (12)$$

$$[\langle J''(u)h,h\rangle\geqslant\mu\|h\|_H^2\quad\forall h\in B,\quad\forall u\in U,\quad \mu=\text{const}>0]. \qquad (13)$$

Тогда функция $J(u)$ выпукла [сильно выпукла] на U.

Если $\operatorname{int}U\neq\varnothing$, *то верно и обратное: из выпуклости [сильной выпуклости] функции $J(u)\in C^2(U)$ на выпуклом множестве U необходимо следует условие* (12) [*условие* (13)].

4. Как и в конечномерных экстремальных задачах, с помощью первых и вторых производных могут быть сформулированы необходимые и достаточные условия экстремума функций на множествах из банаховых пространств.

Теорема 2. *Пусть функция $J(u)$ задана на банаховом пространстве B и пусть $J(u_*)=J_*=\inf\limits_{u\in B}J(u)$. Если $J(u)$ дифференцируема в точке u_*, то необходимо выполняется равенство*

$$J'(u_*)=0, \qquad (14)$$

а если $J(u)$ дважды непрерывно дифференцируема на B, то необходимо

$$J'(u_*)=0,\quad\langle J''(u_*)e,e\rangle\geqslant 0,\quad\forall e\in B. \qquad (15)$$

Доказательство. Возьмем произвольный элемент $e\in B$ и в (5) положим $u=u_*$, $h=te$, $-\infty<t<+\infty$. Так как в точке минимума $\Delta J(u_*)=J(u_*+te)-J(u_*)\geqslant 0$, то из (5) следует $0\leqslant\langle J'(u_*),e\rangle t+o(t)$ при всех t, где $\lim\limits_{t\to 0}o(t)/t=0$. Поделив это неравенство сначала на $t>0$, затем на $t<0$, и устремив t к нулю, получим $\langle J'(u_*),e\rangle=0$ для всех $e\in B$, что равносильно условию (14). Если $J(u)\in C^2(B)$, то из (7) при $u=u_*$, $h=te$ с учетом доказанного равенства (14) будем иметь $0\leqslant\Delta J(u_*)=\langle J''(u_*)e,e\rangle\frac{t^2}{2}+o(t^2)$. Отсюда, деля на $t\neq 0$ и устремляя $t\to 0$, придем ко второму из условий (15). □

Как видим, условия (14), (15) представляют собой обобщение известных необходимых условий безусловного минимума функций конечного числа переменных (теорема 2.2.1). Нетрудно видеть, что эти условия необходимы не

только для глобального, но и локального минимума функций на банаховом пространстве. Отметим, что, в отличие от конечномерных задач (теорема 2.2.2), в банаховых пространствах условия $J'(u_*)=0$, $\langle J''(u_*)e, e\rangle > 0$, $e\neq 0$, не являются достаточными для локального минимума (см. ниже упражнения 6, 7).

Следующая теорема дает необходимые и достаточные условия минимума гладких выпуклых функций на выпуклом множестве.

Теорема 3. *Пусть U — выпуклое множество из банахова пространства B, U_* — множество точек минимума функции $J(u)$ на U. Если в точке $u_*\in U_*$ функция $J(u)$ дифференцируема, то необходимо выполняется неравенство*

$$\langle J'(u_*), u-u_*\rangle \geqslant 0 \quad \forall u\in U, \tag{16}$$

которое в случае $u_*\in \operatorname{int} U$ превращается в равенство $J'(u_*)=0$. Если, кроме того, функция $J(u)$ выпукла на U, то условие (16) является достаточным для того, чтобы $u_*\in U_*$.

Доказательство этой теоремы проводится так же, как доказательство аналогичной теоремы 4.2.3.

5. Проиллюстрируем вышесказанное на примерах.

Пример 1. Пусть H — гильбертово пространство. Тогда функция $J(u)=\|u\|_H^2=\langle u,u\rangle_H$ дважды дифференцируема во всех точках $u\in H$. В самом деле

$$J(u+h)-J(u)=\|u+h\|_H^2-\|u\|_H^2=\langle 2u, h\rangle_H+\langle h, h\rangle_H \quad \forall u, h\in H.$$

Отсюда следует, что $J'(u)=2u$. Из формулы (6) тогда имеем $J''(u)=2I$, где I — единичный (тождественный) оператор на H. Таким образом, $J(u)=\|u\|_H^2\in C^2(H)$.

Пример 2. Функция $J(u)=\|u\|_H=\sqrt{\langle u, u\rangle_H}$ дважды дифференцируема во всех точках u гильбертова пространства, кроме точки $u=0$. В самом деле

$$J(u+h)-J(u)=\sqrt{\langle u+h, u+h\rangle}-\sqrt{\langle u, u\rangle}=$$

$$=\frac{\langle u+h, u+h\rangle-\langle u, u\rangle}{\sqrt{\langle u+h, u+h\rangle}+\sqrt{\langle u, u\rangle}}=\langle \frac{u}{\|u\|}, h\rangle+\alpha(u, h), \quad \forall u\neq 0,$$

где

$$\alpha(u, h)=\frac{\|h\|^2}{\|u+h\|+\|u\|}+\langle u, h\rangle\left(\frac{\|u\|-\|u+h\|}{\|u\|(\|u+h\|+\|u\|)}\right).$$

Отсюда видно, что $\lim\limits_{\|h\|\to 0}\frac{\alpha(u, h)}{\|h\|}=0$, так что первая производная $J'(u)=\frac{u}{\|u\|}$, $u\neq 0$. Эту же формулу нетрудно получить по правилу (2) дифференцирования сложной функции. Аналогично имеем

$$J'(u+h)-J'(u)=\frac{h}{\|u\|}-\frac{u\langle u, h\rangle}{\|u\|^3}+\alpha_1(u, h), \quad \lim_{\|h\|\to 0}\frac{\alpha_1(u, h)}{\|h\|}=0, \quad \forall u\neq 0.$$

Это означает, что функция $J(u)=\|u\|$ дважды дифференцируема на H при $u\neq 0$, причем оператор второй производной $J''(u)\in\mathcal{L}(H\to H)$ действует на элемент h по правилу: $J''(u)h=\frac{h}{\|u\|}-\frac{u\langle u, h\rangle}{\|u\|^3}$.

Пример 3. Пусть $J(u)=\int\limits_a^b f(u(t))dt$, где $f(u)$ — функция одной переменной $u\in E^1$, $u=u(t)\in C[a, b]=B$. Покажем, что если $f(u)\in C^2(E^1)$, то $J(u)\in C^2(B)$. Как известно [371, 393], сопряженное к $B=C[a, b]$ пространство $B^*=(C[a, b])^*$ состоит из функций с ограниченным изменением, т. е. для каждого элемента $g\in B^*$ существует функция $g=g(t)$, $a\leqslant t\leqslant b$, имеющая на $[a, b]$ ограниченное изменение, такая, что $\langle g, u\rangle_{B^*}=\int\limits_a^b u(t)dg$, где интеграл понимается в смысле Римана — Стилтьеса (некоторые определения см. в § 6.4). В частности, если функция $g(t)$ непрерывна и имеет кусочно-непрерывную производную $g'(t)$, то $\langle g, u\rangle_{B^*}=\int\limits_a^b u(t)g'(t)dt$.

Поскольку $f(u)\in C^2(E^1)$, то

$$J(u+h)-J(u)=\int_a^b[f'(u(t))h(t)+\tfrac{1}{2}f''(u(t))h^2(t)]dt+\alpha_2(u, h), \tag{17}$$

где $\alpha_2(u, h)=\int\limits_a^b[f''(u(t)+\theta(t)h(t))-f''(u(t))]h^2(t)dt$, $0<\theta(t)<1$. Так как $f''(u)$ непрерывна, то при $\|h\|_C\to 0$ имеем $f''(u(t)+\theta(t)h(t))-f''(u(t))\to 0$ равномерно по $t\in[a, b]$, так что $\lim\limits_{\|h\|_C\to 0}\frac{\alpha_2(u, h)}{\|h\|_C^2}=0$. Отсюда и из (17) следует, что функция $J(u)$ непрерывно дифференцируема во всех точках $u\in B=C[a, b]$, причем $\langle J'(u), h\rangle_{B^*}=\int\limits_a^b h(t)dg$, где $g=g(t)=\int\limits_a^t f'(u(\tau))d\tau=$ $=J'(u)\in C^1[a, b]$. Кроме того, из (17) вытекает (см. замечание 2) существование второй производной $J''(u)$, причем

$$\langle J''(u)h, h\rangle_{B^*}=\int_a^b(f''(u(t))h(t))h(t)dt=\int_a^b h(t)\left(\int_a^t f''(u(\tau))h(\tau)d\tau\right).$$

Это означает, что оператор второй производной $J''(u)$ действует на элемент $h=h(t)$ по правилу: $J''(u)h=\int\limits_a^t f''(u(\tau))h(\tau)d\tau\in C^1[a, b]$. Таким образом, если $f(u)\in C^1(E^1)$, то $J(u)\in C^1(B)$, а если $f(u)\in C^2(E^1)$, то $J(u)\in C^2(B)$.

Пример 4. Рассмотрим функцию

$$J(u)=\tfrac{1}{2}\langle Au, u\rangle-\langle b, u\rangle, \quad u\in H, \tag{18}$$

где H — гильбертово пространство, $A\in\mathcal{L}(H\to H)$, $b\in H$. Приращение этой функции представимо в виде

$$J(u+h)-J(u)=\tfrac{1}{2}(\langle Au, h\rangle+\langle u, Ah\rangle)-\langle b, h\rangle+\tfrac{1}{2}\langle Ah, h\rangle=$$

$$=\langle\tfrac{1}{2}(A+A^*)u-b, h\rangle+\tfrac{1}{2}\langle Ah, h\rangle,$$

где A^* — оператор, сопряженный к оператору A. Отсюда следует, что $J(u)\in C^2(H)$ и

$$J'(u)=\tfrac{1}{2}(A+A^*)u-b, \quad J''(u)=\tfrac{1}{2}(A+A^*).$$

В частности, если A — самосопряженный оператор, т. е. $A^*=A$, то $J'(u)=Au-b$, $J''(u)=A$. Если, кроме того, A — неотрицательно определенный, т. е. $\langle Ah, h\rangle \geqslant 0$ $\forall h \in H$, то $J(u)$ выпукла на H (теорема 1). Если при этом множество U_* точек минимума функции (18) на выпуклом множестве U непусто (теорема 2.8), то точка $u_* \in U_*$ тогда и только тогда, когда $\langle Au_*-b, u-u_*\rangle \geqslant 0$ $\forall u \in U$ (условие (16)). Если самосопряженный оператор A является положительно определенным, т. е. $\langle Ah, h\rangle \geqslant \mu\|h\|_H^2$ $\forall h \in H$, $\mu=\text{const}>0$, то функция (18) сильно выпукла на H (теорема 1) и на любом выпуклом замкнутом множестве U из H достигает своей нижней грани в единственной точке u_* (теорема 2.10), причем если $U=H$, то $Au_*=b$. Последнее равенство лежит в основе метода Ритца [496], сводящего задачу определения решения операторного уравнения $Au=b$ к поиску точки минимума функции (18) на H (см. пример 4.2.4).

Пример 5. Рассмотрим квадратичную функцию

$$J(u)=\|Au-b\|_F^2, \quad u \in H, \tag{19}$$

где $A \in \mathcal{L}(H \to F)$, H и F — гильбертовы пространства, $b \in F$ (см. задачу (2.3)). Приращение функции (19) представимо в виде

$$J(u+h)-J(u)=2\langle Au-b, Ah\rangle_F+\|Ah\|_F^2= \\ =\langle 2A^*(Au-b), h\rangle_H+\langle A^*Ah, h\rangle_H, \tag{20}$$

где $A^* \in \mathcal{L}(F \to H)$ — оператор, сопряженный к оператору A. Из формулы (20) следует, что $J(u) \in C^2(H)$, причем

$$J'(u)=2A^*(Au-b), \quad J''(u)=2A^*A \tag{21}$$

(ср. с формулами из примера 4.2.5). Так как $\langle J''(u)h, h\rangle=\langle 2A^*Ah, h\rangle_H=2\|Ah\|_F^2 \geqslant 0$ $\forall h \in H$, то функция (19) выпукла на H (теорема 1). Если $\|Ah\|_F^2 \geqslant \mu\|h\|_H^2$ $\forall h \in H$, $\mu=\text{const}>0$, то функция (19) сильно выпукла на H (теорема 1). Если множество U_* точек минимума функции (19) на выпуклом множестве U непусто (теоремы 2.12, 2.13), то точка $u_* \in U_*$ тогда и только тогда, когда

$$\langle A^*(Au_*-b), u-u_*\rangle_H \geqslant 0 \quad \forall u \in U. \tag{22}$$

Если $U=H$, то условие (22) эквивалентно равенству $A^*(Au_*-b)=0$ (теорема 3).

Пример 6. Рассмотрим функцию

$$J(u)=|x(T;u)-b|_{E^n}^2, \tag{23}$$

где $x=x(t)=x(t;u)$ — решение системы

$$\dot{x}(t)=D(t)x(t)+B(t)u(t), \quad t_0 \leqslant t \leqslant T; \quad x(t_0)=0, \tag{24}$$

управление $u=u(t) \in L_2^r[t_0, T]$; остальные обозначения см. в примере 2.14. В этом примере было показано, что оператор A, определенный формулой (2.6)

$$Au=x(T;u), \tag{25}$$

является линейным непрерывным оператором, действующим из гильбертова пространства $H=L_2^r[t_0, T]$ в гильбертово пространство $F=E^n$, и функция (23) с помощью этого оператора может быть представлена в виде (19).

Поэтому, пользуясь результатами примера 5, можем сказать, что функция (23) выпукла и дважды непрерывно дифференцируема на $H=L_2^r[t_0, T]$, причем ее производные имеют вид (21). Покажем, что оператор A^*, входящий в (21) и сопряженный к оператору (25), каждому элементу $c \in F=E^n$ ставит в соответствие элемент $A^*c \in H=L_2^r[t_0, T]$ по следующему правилу:

$$A^*c=B^\top(t)\psi(t;c), \quad t_0 \leqslant t \leqslant T, \tag{26}$$

где $\psi=\psi(t)=\psi(t;c)=(\psi^1(t), \ldots, \psi^n(t))$ — решение задачи Коши

$$\dot{\psi}(t)=-D^\top(t)\psi(t), \quad t_0 \leqslant t \leqslant T; \quad \psi(T)=c. \tag{27}$$

В силу теоремы 6.1.2 задача (27) имеет, притом единственное, решение при каждом $c \in E^n$, так что оператор (26) определен всюду на E^n. Очевидно, этот оператор линейный, действует из E^n в $L_2^r[t_0, T]$. Кроме того, с учетом (24)–(27) имеем:

$$\langle Au, c\rangle_{E^n}=\langle x(T;u), \psi(T;c)\rangle=\int_{t_0}^{T}\frac{d}{dt}\langle x(t;u), \psi(t;c)\rangle dt+\langle x(t_0;u), \psi(t_0;c)\rangle= \\ =\int_{t_0}^{T}(\langle \dot{x}(t;u), \psi(t;c)\rangle+\langle x(t;u), \dot{\psi}(t;c)\rangle)dt= \\ =\int_{t_0}^{T}(\langle D(t)x(t;u)+B(t)u(t), \psi(t;c)\rangle+\langle x(t;u), -D^\top(t)\psi(t;c)\rangle)dt= \\ =\int_{t_0}^{T}\langle u(t), B^\top(t)\psi(t;c)\rangle dt=\langle u, A^*c\rangle_{L_2^r} \quad \forall u \in H, \quad c \in F.$$

Это означает, что оператор A^*, определенный формулой (26), действительно является сопряженным к оператору (25).

Из (21), (25), (26) получаем следующее выражение для формулы градиента функции (23)

$$J'(u)=2B^\top(t)\psi(t;Au-b), \quad t_0 \leqslant t \leqslant T, \tag{28}$$

где $\psi(t;Au-b)$ —решение задачи Коши (27) при $c=Au-b=x(T;u)-b$, т. е.

$$\dot{\psi}(t)=-D^\top(t)\psi(t), \quad t_0 \leqslant t \leqslant T; \quad \psi(T)=x(T;u)-b. \tag{29}$$

Таким образом, для вычисления градиента функции (23) в некоторой точке $u=u(t) \in L_2^r[t_0, T]$ сначала нужно решить задачу Коши (24) и определить $x(t;u)$, затем подставить полученное значение $x(T;u)$ в (29) и, решая задачу Коши (29), определить $\psi(t;Au-b)$ и, наконец, по формуле (28) найти градиент $J'(u)$.

Пользуясь теоремой 3, теперь мы можем сформулировать следующий критерий оптимальности для задачи минимизации функции (23) на выпуклом множестве $U \subseteq L_2^r[t_0, T]$: для того, чтобы в точке $u=u_* \in U$ функция (23) достигала своей нижней грани на U, необходимо и достаточно, чтобы

$$\langle J'(u_*), u-u_*\rangle=\int_{t_0}^{T}\langle B^\top(t)\psi(t;x(T;u_*)-b), u(t)-u_*(t)\rangle_{E^r}dt \geqslant 0 \quad \forall u \in U. \tag{30}$$

Условие (30) тогда перепишем в виде

$$\min_{u(t)\in U}\int_{t_0}^{T}\langle B^{\top}(t)\psi(t;x(T;u_*)-b),u(t)\rangle dt=$$
$$=\int_{t_0}^{T}\langle B^{\top}(t)\psi(t;x(T;u_*)-b),u_*(t)\rangle dt. \quad (31)$$

Если ввести функцию Гамильтона — Понтрягина

$$H(x,u,t,\psi)=\langle\psi,D(t)x+B(t)u\rangle,$$

то условие (31) можно записать в так называемой *форме интегрального принципа минимума*:

$$\min_{u(t)\in U}\int_{t_0}^{T}H(x(t;u_*),u(t),t,\psi(t;x(T;u_*)-b))dt=$$
$$=\int_{t_0}^{T}H(x(t;u_*),u_*(t),t,\psi(t;x(T;u_*)-b))dt.$$

Если множество U имеет вид

$$U=\{u=u(t)\in L_2^r[t_0,T]\colon u(t)\in V \text{ почти всюду на } [t_0,T]\}, \quad (32)$$

где V — выпуклое множество из E^r, то из (31) можно получить

$$\min_{v\in V}\langle B^{\top}(t)\psi(t;x(T;u_*)-b),v\rangle=\langle B^{\top}(t)\psi(t;x(T;u_*)-b),u_*(t)\rangle$$

для почти всех $t\in[t_0,T]$. Взяв здесь вместо $\psi(t;x(T;u_*)-b)$ функцию $(-\psi(t;x(T;u_*)-b))$, придем к принципу максимума Понтрягина для задачи минимизации функции (23) на множестве (32). Заметим, что в главе 6 принцип максимума был доказан для более общей задачи оптимального управления без предположения выпуклости множества V. С другой стороны, условие (31) справедливо для всех выпуклых множеств U из $L_2^r[t_0,T]$, необязательно имеющих вид (32), и дает не только необходимое, но и достаточное условие оптимальности.

Вторая производная функции (23) согласно формуле (21) равна $J''(u)=2AA^*$, где операторы A, A^* имеют вид (25), (26). Формулу для $J''(u)$ можно записать в несколько иной форме, если воспользоваться известной формулой Коши

$$x(t;u)=\int_{t_0}^{t}\Phi(t,\tau)B(\tau)u(\tau)d\tau,\quad t_0\leqslant t\leqslant T, \quad (33)$$

для решения задачи (24), где $\Phi(t,\tau)$ — матрица порядка $n\times n$, определяемая условиями

$$\frac{d\Phi(t,\tau)}{dt}=D(t)\Phi(t,\tau),\quad t_0\leqslant t,\tau\leqslant T;\quad \Phi(\tau,\tau)=I, \quad (34)$$

I — единичная матрица порядка $n\times n$. Из (25), (33) следует, что значение оператора A на элементе $h=h(t)\in L_2^r[t_0,T]$ равно

$$Ah=x(T;h)=\int_{t_0}^{T}\Phi(t,\tau)B(\tau)h(\tau)d\tau. \quad (35)$$

Далее, с помощью той же матрицы $\Phi(t,\tau)$ можно получить следующее представление для решения задачи (27)

$$\psi(t;c)=\Phi^{\top}(T,t)c,\quad t_0\leqslant t\leqslant T, \quad (36)$$

где $\Phi^{\top}(t,\tau)$ — транспонированная матрица $\Phi(t,\tau)$ (см. упражнение 3). Из (26), (32) следует

$$A^*c=B^{\top}(t)\Phi^{\top}(T,t)c,\quad t_0\leqslant t\leqslant T. \quad (37)$$

С помощью представлений (35), (37) значение оператора $J''(u)=2A^*A$ на элементе $h=h(t)$ может быть выражено формулой

$$J''(u)h=2A^*Ah=2B^{\top}(t)\Phi^{\top}(T,t)\int_{t_0}^{T}\Phi(T,\tau)B(\tau)h(\tau)d\tau=$$
$$=\int_{t_0}^{T}2B^{\top}(t)\Phi^{\top}(T,t)\Phi(T,\tau)B(\tau)h(\tau)d\tau\quad t_0\leqslant t\leqslant T. \quad (38)$$

Как видим, оператор $J''(u)$ является оператором Фредгольма с симметричным ядром $K(t,\tau)=2B^{\top}(t)\Phi^{\top}(T,t)\Phi(T,\tau)B(\tau)$, $t_0\leqslant t,\tau\leqslant T$. Мы здесь не будем подробно останавливаться на свойствах матрицы $\Phi(t,\tau)$ (см. упражнение 3) и выводе формул (33), (36). Заметим лишь, что эти формулы весьма полезны при теоретическом исследовании задач оптимального управления, связанных с системой (24), но при численном решении таких задач они применяются крайне редко из-за трудностей в определении явного выражения матрицы $\Phi(t,\tau)$ из (34). Для получения приближенного решения задач Коши (24), (27), (34) могут быть использованы разностные методы [74; 89; 634].

Пример 7. Рассмотрим функцию

$$J_1(u)=|x_1(T;u)-b_1|^2_{E^n}, \quad (39)$$

где $x_1=x_1(t;u)$ — решение задачи Коши

$$\dot{x}_1(t)=D(t)x_1(t)+B(t)u(t)+f(t),\quad t_0\leqslant t\leqslant T,\quad x(t_0)=x_0, \quad (40)$$

$u=u(t)\in L_2^r[t_0,T]$, $f(t)\in L_2^n[t_0,T]$. Как и в примере 2.15, воспользуемся представлением решения $x_1(t;u)$ задачи (40) в виде суммы

$$x_1(t;u)=x(t;u)+x_0(t),\quad t_0\leqslant t\leqslant T, \quad (41)$$

где $x(t;u)$ — решение задачи (24), $x_0(t)$ — решение задачи (2.13): $\dot{x}_0(t)=D(t)x_0(t)+f(t)$, $x(t_0)=x_0$, и функцию (39) запишем в виде: $J_1(u)=|x(T;u)-b|^2=|Au-b|^2$, где $b=b_1-x_0(T)$, оператор A определен согласно (25). Отсюда и из формулы (28) имеем

$$J_1(u)=2B^{\top}(t)\psi(t;Au-b)=2B^{\top}(t)\psi(t;Au+x_0(T)-b_1)=$$
$$=2B^{\top}(t)\psi(t;x_1(T;u)-b_1),$$

где $\psi(t;x_1(T;u)-b_1)$ — решение задачи Коши (27) при $c=x_1(T;u)-b_1$, и, кроме того, справедливо равенство $J_1''=J''(u)=2A^*A$. Нетрудно видеть, что формулы (35)–(38) сохраняют силу и для функции (39).

Пример 8. Рассмотрим функцию

$$J(u)=\int_{t_0}^{T}|x(t;u)-b(t)|^2_{E^n}dt, \quad (42)$$

где $x = x(t; u)$ — решение задачи Коши (24), $b(t) \in L_2^n[t_0, T]$, $u = u(t) \in L_2^r[t_0, T]$. В примере 2.16 было показано, что оператор A, определенный формулой (2.15):

$$Au = x(t; u), \quad t_0 \leqslant t \leqslant T, \tag{43}$$

является линейным непрерывным оператором, действующим из гильбертова пространства $H = L_2^r[t_0, T]$ в гильбертово пространство $F = L_2^n[t_0, T]$, и функция (42) с помощью этого оператора представляется в виде (19). Отсюда и из примера 5 следует, что $J(u) \in C^2(H)$, причем ее производные имеют вид (21). Покажем, что оператор A^*, входящий в (21) и сопряженный к оператору (43), каждому элементу $c = c(t) \in F = L_2^n[t_0, T]$ ставит в соответствие элемент $A^*c \in H = L_2^r[t_0, T]$ по следующему правилу:

$$A^*c = B^\top(t)\psi(t; c), \quad t_0 \leqslant t \leqslant T, \tag{44}$$

где $\psi = \psi(t) = \psi(t; c)$ — решение задачи Коши

$$\dot\psi(t) = -D^\top(t)\psi(t) - c(t), \quad t_0 \leqslant t \leqslant T; \quad \psi(T) = 0. \tag{45}$$

В силу теоремы 6.1.2 задача (45) имеет, притом единственное решение при каждом $c \in F$, так что оператор (44) определен всюду на F. Очевидно, что этот оператор линейный, действует из $F = L_2^n[t_0, T]$ в $H = L_2^r[t_0, T]$. Кроме того, с учетом (24), (43)–(45) имеем:

$$\langle Au, c\rangle_H = \int_{t_0}^{T} \langle x(t; u), c(t)\rangle dt = \int_{t_0}^{T} \langle x(t; u), -\dot\psi(t; c) - D^\top(t)\psi(t; c)\rangle dt =$$

$$= -\langle x(t;u), \psi(t;c)\rangle\Big|_{t=0}^{T} + \int_{t_0}^{T} \langle \dot x(t;u), \psi(t;c)\rangle dt - \int_{t_0}^{T} \langle D(t)x(t;u), \psi(t;c)\rangle dt =$$

$$= \int_{t_0}^{T} \langle \dot x(t; u) - D(t)x(t; u), \psi(t; c)\rangle dt = \int_{t_0}^{T} \langle B(t)u(t), \psi(t; c)\rangle dt =$$

$$= \int_{t_0}^{T} \langle u(t), B^\top(t)\psi(t; c)\rangle dt = \langle u, A^*c\rangle_F \quad \forall u \in H, \quad c \in F.$$

Это означает, что оператор A^*, определенный формулой (44), действительно является сопряженным к оператору (43).

Из (21), (43), (44) получаем следующее выражение для формулы градиента функции (42)

$$J'(u) = 2B^\top(t)\psi(t; Au - b), \quad t_0 \leqslant t \leqslant T, \tag{46}$$

внешне схожее с (28), $\psi(t; Au - b)$ — решение задачи Коши (45) при $c = Au - b = x(t; u) - b(t)$, т. е.

$$\dot\psi(t) = -D^\top(t)\psi(t) - x(t; u) + b(t), \quad t_0 \leqslant t \leqslant T; \quad \psi(T) = 0. \tag{47}$$

Таким образом, для вычисления градиента функции (42) сначала нужно последовательно решить задачи Коши (24) и (47), а затем воспользоваться формулой (46).

Критерий оптимальности (теорема 3) для задачи минимизации функции (42) на выпуклом множестве $U \subseteq H$ запишется в том же виде (30)

или (31), но в качестве $\psi(t)$ в этих формулах здесь нужно взять решение задачи Коши (47) при $u = u_*$. В случае, когда множество U задается условиями (32), как и в примере 6, нетрудно установить связь между полученным условием оптимальности (31) и принципом максимума Понтрягина.

Вторая производная функции (42) согласно (21) имеет вид $J''(u) = 2A^*A$, где операторы A^*, A определены равенствами (43), (44). Приведем формулу для значения $J''(u)h$ оператора $J''(u)$ на элементе $h \in H = L_2^r[t_0, T]$, аналогичную (38). Из (33), (43) имеем:

$$Ah = \int_{t_0}^{t} \Phi(t, \xi)B(\xi)h(\xi)d\xi, \quad t_0 \leqslant t \leqslant T. \tag{48}$$

Решение задачи Коши (45) может быть записано в виде

$$\psi(t; c) = \int_{t}^{T} \Phi^\top(\tau, t)c(\tau)d\tau, \quad t_0 \leqslant t \leqslant T. \tag{49}$$

(см. упражнение 3). Из (44), (49) следует

$$A^*c = \int_{t}^{T} B^\top(t)\Phi^\top(\tau, t)c(\tau)d\tau, \quad t_0 \leqslant t \leqslant T. \tag{50}$$

Тогда значение $J''(u)h$ оператора $J''(u)$ на элементе $h \in H = L_2^r[t_0, T]$ выражается формулой

$$J''(u)h = 2A^*Ah = 2\int_{t}^{T} B^\top(t)\Phi^\top(\tau, t)\Big(\int_{t_0}^{\tau} \Phi(\tau, \xi)B(\xi)h(\xi)d\xi\Big)d\tau, \quad t_0 \leqslant t \leqslant T. \tag{51}$$

Квадратичная форма $\langle J''(u)h, h\rangle = 2\langle Ah, Ah\rangle$ равна

$$\langle J''(u)h, h\rangle = \int_{t_0}^{T} \langle 2\int_{t}^{T} B^\top(t)\Phi^\top(\tau, t)\Big(\int_{t_0}^{\tau} \Phi(\tau, \xi)B(\xi)h(\xi)d\xi\Big)d\tau, h(t)\rangle_{E^r} dt =$$

$$= 2\int_{t_0}^{T}\int_{t}^{T} \langle B^\top(t)\Phi^\top(\tau, t)\Big(\int_{t_0}^{\tau} \Phi(\tau, \xi)B(\xi)h(\xi)d\xi\Big), h(t)\rangle d\tau dt =$$

$$= 2\int_{t_0}^{T}\int_{t_0}^{\tau} \langle B^\top(t)\Phi^\top(\tau, t)\int_{t_0}^{\tau} \Phi(\tau, \xi)B(\xi)h(\xi)d\xi, h(t)dtd\tau =$$

$$= 2\int_{t_0}^{T}\Big(\int_{t_0}^{\tau}\int_{t_0}^{\tau} \langle B^\top(t)\Phi^\top(\tau, t)\Phi(\tau, \xi)B(\xi)h(\xi), h(t)\rangle d\xi dt\Big)d\tau. \tag{52}$$

Пример 9. Рассмотрим функцию

$$J_1(u) = \int_{t_0}^{t} |x_1(t; u) - b_1(t)|_{E^n}^2 dt, \tag{53}$$

где $x_1(t; u)$ — решение задачи Коши (40), $u = u(t) \in L_2^r[t_0, T]$. Пользуясь формулой (41), функцию (53) можем записать в виде (19): $J_1(u) = \int_{t_0}^{T} |x(t; u) - b(t)|_{E^n}^2 dt = \|Au - b\|_{L_2^n}^2$, где $b = b(t) = b_1(t) - x_0(t)$, оператор A определен согласно (43). Отсюда, пользуясь результатами примера 8, из формулы (46) имеем $J'(u) = 2B^\top(t)\psi(t; Au - b)$, где $\psi(t; Au - b)$ — реше-

ние задачи Коши (45) при $c=c(t)=Au-b(t)=Au+x_0(t)-b_1(t)=x_1(t;u)-b_1(t)$, $t_0\leqslant t\leqslant T$. Вторая производная функции (53) совпадает со второй производной функции (42); формулы (48)–(52) здесь также сохраняют силу.

Пример 10. Пусть

$$J(u)=\int_c^d\Big(\int_a^b K(t,\tau)u(\tau)d\tau-b(t)\Big)^2dt, \tag{54}$$

где $K(t,\tau)\in L_2(Q)$, $Q=\{(t,\tau)\colon a\leqslant\tau\leqslant b,\ c\leqslant t\leqslant d\}$, $b=b(t)\in L_2[c,d]$. Введем оператор Фредгольма [393; 705]:

$$Au=\int_a^b K(t,\tau)u(\tau)d\tau,\quad c\leqslant t\leqslant d, \tag{55}$$

действующий из $H=L_2[a,b]$ в $F=L_2[c,d]$, линейный, непрерывный. Тогда функцию (54) можем записать в виде (19) и воспользоваться результатами примера 5. Сопряженный к A оператор A^* имеет вид

$$A^*c=\int_c^d K(t,\tau)c(t)dt,\quad a\leqslant\tau\leqslant b. \tag{56}$$

Этот оператор действует из F в H, он линейный, непрерывный. Далее, пользуясь теоремой Фубини, имеем

$$\langle Au,c\rangle_F=\int_c^d\Big(\int_a^b K(t,\tau)u(\tau)d\tau\Big)c(t)dt=\iint_Q K(t,\tau)u(\tau)c(t)d\tau dt=$$
$$=\int_a^b\Big(\int_c^d K(t,\tau)c(t)dt\Big)u(\tau)d\tau=\langle A^*c,u\rangle_H\quad\forall u\in H,\quad\forall c\in F.$$

Это означает, что оператор (56) является сопряженным к оператору (55). Из (21), (55), (56) получим следующую формулу для градиента функции (54):

$$J'(u)=2A^*(Au-b)=2\int_c^d K(t,\tau)\Big(\int_a^b K(t,\xi)u(\xi)d\xi-b(t)\Big)dt,\quad a\leqslant\tau\leqslant b.$$

Вторая производная $J''(u)=2A^*A$ на каждый элемент $h\in H$ действует по правилу: $J''(u)h=2\int_c^d K(t,\tau)\Big(\int_a^b K(t,\xi)h(\xi)d\xi\Big)dt$, $a\leqslant\tau\leqslant b$. Из теоремы 3 следует, что функция (54) достигает своей нижней грани на множестве $U=L_2[a,b]$ в точке $u=u_*(t)$ тогда и только тогда, когда $u_*(t)$ является решением интегрального уравнения Фредгольма первого рода $\int_a^b\Big(\int_c^d K(t,\tau)K(t,\xi)dt\Big)u(\xi)d\xi=\int_c^d K(t,\tau)b(t)dt$, $a\leqslant\tau\leqslant b$.

6. Кратко остановимся на следующей задаче минимизации, когда множество U задается ограничениями типа равенств и неравенств

$$J(u)\to\inf,\quad u\in U=\{u\in U_0\colon g_i(u)\leqslant 0,\ i=1,\ldots,m;\ F(u)=0\}, \tag{57}$$

где U_0 — заданное множество из банахова пространства B, функции $J(u)$, $g_1(u)$, …, $g_m(u)$ определены и конечны на U_0, F — отображение, действующее из B в банахово пространство Y. Если $B=E^n$, $Y=E^{s-m}$, $F(u)=\begin{pmatrix}g_{m+1}(u)\\ \ldots\\ g_s(u)\end{pmatrix}$, то задача (57) превращается в задачу математического программирования, которая рассматривалась в главах 2–5.

В общем случае операторные ограничения $F(u)=0$ часто используются, например, в задачах оптимального управления для учета начально-краевых задач для дифференциальных уравнений. В частности, задача Коши (24) может быть записана в виде $F(u)=0$, где отображение $F\colon B=L_2^r[t_0,T]\to Y=H_n^1[t_0,T]$.

Приведем формулировки двух теорем, дающих необходимые условия оптимальности для задачи (57) [358]. С этой целью введем функцию Лагранжа задачи (57):

$$\mathcal{L}(u,\lambda_0,\lambda,c)=\lambda_0J(u)+\lambda_1g_1(u)+\ldots+\lambda_mg_m(u)+\langle c,F(u)\rangle,$$

где $u\in U_0$; $\lambda_0,\lambda_1,\ldots,\lambda_m$ — вещественные числа, $c\in Y^*$. Как и выше, будем пользоваться обозначениями: $J_*=\inf\limits_{u\in U}J(u)$, $U_*=\{u\in U\colon J(u)=J_*\}$.

Теорема 4. *Пусть в задаче* (57) $U_0=B$, *функции* $J(u),g_1(u),\ldots,g_m(u)$ *и отображения* $F(u)$ *непрерывны и дифференцируемы в некоторой окрестности точки* $u_*\in U_*$, *причем производная* $F'(u)$ *отображения* F *непрерывна в точке* u_* *и образ пространства* B *при отображении* $u\to F'(u_*)u$ *замкнут, пусть* $J_*>-\infty$, $U_*\neq\varnothing$. *Тогда существуют не равные одновременно нулю множители Лагранжа* $\lambda_0^*\geqslant 0$, $\lambda_1^*\geqslant 0,\ldots,\lambda_m^*\geqslant 0$, $c^*\in Y^*$ *такие, что*

$$\mathcal{L}_u(u_*,\lambda_0^*,\lambda^*,c^*)=\lambda_0^*J'(u_*)+\sum_{i=1}^m\lambda_i^*g_i'(u_*)+(F'(u_*))^*c^*=0,$$
$$\lambda_i^*g_i(u_*)=0,\quad i=1,\ldots,m.$$

Если, кроме того, множество значений оператора $F'(u_*)$ *совпадает с* Y *и существует такая точка* $\overline{u}\in B$, *что* $F'(u_*)\overline{u}=0$, $\langle g_i'(u_*),\overline{u}\rangle<0$ *для тех* i, $1\leqslant i\leqslant m$, *для которых* $g_i(u_*)=0$, *то* $\lambda_0^*>0$ (*можно принять* $\lambda_0^*=1$).

Теорема 5. *Пусть функции* $J(u),g_1(u),\ldots,g_m(u)$ *выпуклы на* B, U_0 — *выпуклое множество из* B, *а отображение* $F\colon B\to Y$ *является аффинным, т. е.* $F(u)=Au+y_0$, $A\in\mathcal{L}(B\to Y)$, $y_0\in Y$; *пусть* $J_*>-\infty$, $U_*\neq\varnothing$. *Тогда для любой точки* $u_*\in U_*$ *существуют не равные одновременно нулю множители Лагранжа* $\lambda_0^*\geqslant 0$, $\lambda_1^*\geqslant 0,\ldots,\lambda_m^*\geqslant 0$, $c^*\in Y^*$ *такие, что*

$$\mathcal{L}(u_*,\lambda_0^*,\lambda^*,c^*)=\min_{u\in U_0}\mathcal{L}(u,\lambda_0^*,\lambda^*,c^*),$$
$$\lambda_i^*g_i(u_*)=0,\quad i=1,\ldots,m. \tag{58}$$

Если, кроме того, образ множества U_0 *при отображении* $u\to F(u)$ *содержит окрестность нуля пространства* Y *и существует точка* $\overline{u}\in U_0$ *такая, что* $F(\overline{u})=0$, $g_i(\overline{u})<0$, $i=1,\ldots,m$, *то* $\lambda_0^*>0$ (*можно принять* $\lambda_0^*=1$). *В последнем случае условия* (58) *являются также и достаточными для того, чтобы* $u_*\in U_*$.

Теоремы 4, 5 представляют собой обобщения аналогичных конечномерных теорем из §§ 2.3, 4.8, 4.9. Другие утверждения, обобщающие правило множителей Лагранжа, теоремы Куна — Таккера на экстремальные задачи в банаховых пространствах можно найти в [6; 7; 14; 44; 211; 225; 233; 265; 278; 346–348; 366; 393; 434; 435; 465; 486; 604; 605; 638; 672; 673; 690; 699; 748; 780]. Условия оптимальности второго порядка, обобщающие теоремы из §§ 2.4, 2.5 на бесконечномерные пространства, а также условия более высокого порядка были исследованы в [44].

Напомним, что при доказательстве теорем из §§ 4.8, 4.9 были существенно использованы теоремы отделимости из § 4.5. Аналогичные теоремы отделимости играют важную роль при исследовании условий оптимальности в задаче (57) и во многих других вопросах выпуклого анализа в банаховых пространствах.

Приведем одну из таких теорем [393; 705].

Теорема 6. *Пусть* M,N — *выпуклые множества из банахова пространства* B, *причем* $\operatorname{int}M$ — *множество внутренних точек множества* M *непусто и* $\operatorname{int}M\cap N=\varnothing$. *Тогда существует гиперплоскость* $\langle c,u\rangle=\gamma$, *разделяющая эти два множества, а также их замыкания* $\overline{M}$ *и* $\overline{N}$, *т. е.* $\langle c,u\rangle\geqslant\gamma\geqslant\langle c,v\rangle$ $\forall u\in\overline{M}$, $\forall v\in\overline{N}$. *При этом, если* $\overline{M}$, $\overline{N}$ *имеют общую граничную точку* y, *то* $\gamma=\langle c,y\rangle$.

В теореме 6 в отличие от аналогичной конечномерной теоремы 4.5.2 требуется условие $\operatorname{int}M\neq\varnothing$. Приведем примеры, показывающие существенность этого условия для справедливости теоремы 6.

Пример 11. Пусть $U=\{u=(u^1,\ldots,u^n,\ldots)\in l_2\colon|u^n|\leqslant\frac{1}{n},\ n=1,2,\ldots\}$ — «гильбертов кирпич». Покажем, что это множество не имеет внутренних точек. Возьмем произвольную точку $u=(u^1,\ldots,u^n,\ldots)\in U$. Положим $e=(e^1,\ldots,e^n,\ldots)$, где $e^n=\dfrac{\operatorname{sign}u^n}{n^\alpha}$, $n=1,2,\ldots$; $\frac{1}{2}<\alpha<1$. Так как $\|e\|^2=\sum\limits_{n=1}^\infty|e^n|^2=\sum\limits_{n=1}^\infty n^{-2\alpha}<\infty$, то $e\in l_2$. Возьмем точку $u+\varepsilon e$, где $\varepsilon>0$.

Для каждого $\varepsilon>0$ найдется номер $N=N(\varepsilon)$ такой, что $|u^n+\varepsilon e^n|=|u^n|+\varepsilon n^{-\alpha}\geqslant \varepsilon n^{-\alpha}>n^{-1}$ при всех $n>N$. Это значит, что $u+\varepsilon e\notin U$ при всех $\varepsilon>0$. Таким образом, $\operatorname{int} U=\varnothing$, т. е. U состоит лишь из граничных точек.

Далее, множество U выпукло. В самом деле, если $|u^n|\leqslant n^{-1}$, $|v^n|\leqslant n^{-1}$, $n=1,2,\ldots$, то $|\alpha u^n+(1-\alpha)v^n|\leqslant n^{-1}$ при всех $n=1,2,\ldots$, $0\leqslant\alpha\leqslant 1$. Отсюда следует, что если $u, v\in U$, то $\alpha u+(1-\alpha)v\in U$ при всех α, $0\leqslant\alpha\leqslant 1$. Выпуклость U доказана.

Геометрические представления о выпуклых множествах «подсказывают» нам гипотезу о том, что через любую граничную точку выпуклого множества, по-видимому, можно провести опорную гиперплоскость. В евклидовом пространстве E^n эта гипотеза подтвердилась (теорема 4.5.1).

Посмотрим, справедлива ли эта гипотеза для «гильбертова кирпича». Возьмем любую точку $v=(v^1,\ldots,v^n,\ldots)\in U$ такую, что $|v^n|<n^{-1}$, $n=1,2,\ldots$ (например, $v=0$). Предположим, что через эту точку можно провести опорную гиперплоскость к множеству U, т. е. существуют вектор $c=(c^1,\ldots,c^n,\ldots)\neq 0$, $c\in l_2$, и число γ такие, что $\langle c,u\rangle\geqslant\gamma$ при всех $u\in U$ и $\langle c,v\rangle=\gamma$. Так как $c\neq 0$, то $c^n\neq 0$ для некоторого $n\geqslant 1$. Возьмем вектор $e=(0,\ldots,0,e_n=-\dfrac{\operatorname{sign} c^n}{n}-v^n,0,\ldots)\in l_2$. Так как $|v^n+e^n|=n^{-1}$, то $v+e\in U$, и поэтому должно выполняться неравенство $\langle c,v+e\rangle\geqslant\gamma$. Однако $\langle c,v+e\rangle=\gamma+\langle c,e\rangle=\gamma+c^ne^n=\gamma-|c^n|n^{-1}-c^nv^n\leqslant\gamma-|c^n|n^{-1}+|c^n|\cdot|v^n|=\gamma-|c^n|(n^{-1}-|v^n|)<\gamma$. Противоречие. Следовательно, множество U и ее граничная точка v неотделимы. Это значит, что не через всякую граничную точку рассматриваемого множества U можно провести опорную гиперплоскость.

Рассмотренный пример показывает, что условие $\operatorname{int} M=\varnothing$ в теореме 6 существенно. Любопытно заметить, что через любую точку $w=(w^1,\ldots,w^n,\ldots)$, имеющую хотя бы одну координату w^n, $|w^n|=n^{-1}$, можно провести опорную гиперплоскость к «гильбертову кирпичу». Достаточно взять $c=(0,\ldots,0,c_n=-\operatorname{sign} w^n,0,\ldots)$, $\gamma=-n^{-1}$, и получим $\langle c,w\rangle=\gamma$, $\langle c,u\rangle=-u^n\operatorname{sign} w^{-n}\geqslant-|u^n|\geqslant\gamma$ для всех $u\in U$.

Пример 12. Пусть $U=\{u=u(t)\in L_2[0,1]\colon |u(t)|\leqslant 1$ почти всюду на $[0,1]\}$. Покажем, что множество U не имеет внутренних точек в $L_2[0,1]$. Возьмем любую функцию $u=u(t)\in U$. Положим $e_k(t)=k^{1/4}$ при $0\leqslant t\leqslant\dfrac{1}{k}$, $e_k(t)=0$ при $\dfrac{1}{k}<t\leqslant 1$, $k=1,2,\ldots$ Ясно, что $u(t)+e_k(t)=u_k(t)\notin U$ при всех $k>16$, так как $|u_k(t)|\geqslant|e_k(t)|-|u(t)|>2-1=1$ при почти всех t, $0\leqslant t\leqslant\dfrac{1}{k}$, $k>16$. В то же время $\|u_k(t)-u(t)\|_{L_2}=\|e_k(t)\|_{L_2}=k^{-1/2}\to 0$ при $k\to\infty$. Это значит, что множество U не имеет внутренних точек. Очевидно, множество U выпукло. Покажем, что не через всякую точку из U можно провести опорную к U гиперплоскость. Возьмем, например, $v=v(t)\equiv 0$. Допустим, что существует $c=c(t)\in L_2[0,1]$, $c(t)\neq 0$, что $\langle c,u\rangle_{L_2}=\int\limits_0^1 c(t)u(t)dt\geqslant\langle c,v\rangle=0$ для всех $u\in U$. Возьмем $u_0=u_0(t)=-\operatorname{sign} c(t)$. Ясно, что $u_0\in U$, поэтому должно быть $\langle c,u_0\rangle\geqslant 0$. Однако, $\langle c,u_0\rangle=-\int\limits_0^1|c(t)|dt<0$. Противоречие. Следовательно, множество U и точка $v=0$ не могут быть отделены гиперплоскостью (см. упражнение 16).

Приведем формулировку еще одной теоремы об отделимости выпуклого множества и точки [705].

Теорема 7. *Пусть M — выпуклое замкнутое множество из банахова пространства B, точка y не принадлежит M. Тогда множество M и точка y сильно отделимы.*

7. Как видим, многие важные понятия теории экстремальных задач (градиент, выпуклое множество, выпуклая функция, отделяющая гиперплоскость и т. д.) представляют собой естественное обобщение соответствующих понятий, введенных для конечномерных евклидовых пространств E^n. Поэтому неудивительно, что многие утверждения, сформулированные и доказанные в главах 2, 4 для пространства E^n, остаются верными и в любых гильбертовых или банаховых пространствах. Однако, как показывают теорема 6 и примеры 11, 12, такая аналогия имеет место далеко не всегда: имеется немало утверждений, справедливых в E^n, но не имеющих аналога в общих банаховых и гильбертовых пространствах. Это значит, что теоремами, приведенными в главах 2, 4, можно пользоваться при исследовании экстремальных задач в конкретных банаховых или гильбертовых пространствах лишь после тщательной проверки того, что они верны и в рассматриваемом пространстве.

Еще раз возвращаясь к примерам 11, 12, заметим, что в банаховых и гильбертовых пространствах отделимость выпуклых множеств может быть гарантирована при более жестких ограничениях, чем в конечномерном пространстве E^n. Это обстоятельство приводит к тому, что ряд важных результатов теории экстремальных задач, опирающихся на конечномерные теоремы отделимости, не имеют аналога в банаховых и гильбертовых пространствах. В частности, как свидетельствует следующий пример, в задачах оптимального управления, в которых фазовое пространство является гильбертовым пространством, принцип максимума, сформулированный в главе 6 для задач с фазовым пространством E^n, в общем случае не имеет аналога.

Пример 13. Пусть управляемый процесс описывается системой уравнений [290]

$$\dot{x}^i(t)=u^i(t),\quad t>0;\quad x^i(0)=0,\quad i=1,2,\ldots, \tag{59}$$

где $u^i(t)$ — ограниченные измеримые на каждом конечном отрезке $0\leqslant t\leqslant T$, $i=1,2,\ldots$, функции, принимающие свои значения из множества

$$V=\{u=(u^1,\ldots,u^k,\ldots)\in l_2\colon |u^k|\leqslant\frac{1}{k}+\frac{1}{k^2},\quad k=1,2,\ldots\}.$$

Под решением системы (59), соответствующим допустимому управлению $u=u(t)=(u^1(t),\ldots,u^n(t),\ldots)$, $t>0$, будем понимать функцию $x(t,u)=(x^1(t),\ldots,x^n(t),\ldots)$, $t>0$, где $x^i(t)=x^i(t,u^i)=\int\limits_0^t u^i(\tau)d\tau$, $t>0$, $i=1,2,\ldots$, такую, что $x(t,u)\in l_2$ при всех $t>0$. Таким образом, фазовым пространством системы (59) является бесконечномерное гильбертово пространство l_2.

Рассмотрим задачу быстродействия: найти управление $u=u(t)\in V$, $t>0$, такое, чтобы соответствующее ему решение $x(t,u)$ системы (59) удовлетворяло условию

$$x(T;u)=x_1=(1,1/2,1/3,\ldots,1/n,\ldots)$$

при минимальном T.

Пользуясь принципом максимума из главы 6, нетрудно показать, что при каждом фиксированном $n\geqslant 1$ минимальное время перехода из точки $x^n(0)=0$ в точку $x^n(T)=\dfrac{1}{n}$ при движении по траектории дифференциального уравнения $\dot{x}^n(t)=u^n(t)$, $t>0$, $x^n(0)=0$, с ограничениями $|u^n(t)|\leqslant\dfrac{1}{n}+\dfrac{1}{n^2}$, $t\geqslant 0$, равно $t_{n*}=\dfrac{n}{n+1}$ и реализуется на управлении $u_{n*}(t)=\dfrac{1}{n}+\dfrac{1}{n^2}$, $t\geqslant 0$. Отсюда следует, что оптимальное время t_* в исходной задаче не может быть меньше t_{n*}, т. е. $t_*\geqslant t_{n*}$, $n=1,2,\ldots$ Отсюда при $n\to\infty$ получим $t_*\geqslant 1$. С другой стороны для управления $u_*=u_*(t)=(1,1/2,\ldots,1/n,\ldots)$ имеем $x(1,u_*)=x_1$. Это значит, что $t_*=1$ — оптимальное время, а $u=u_*$ — оптимальное управление, $x_*(t)=tu_*$ — оптимальная траектория в исходной задаче быстродействия.

Убедимся в том, что принцип максимума в этой задаче не выполняется. Для этого по аналогии с задачами оптимального управления из главы 6 напишем функцию Гамильтона — Понтрягина

$$H(x,u,\psi,a_0)=a_0+\langle\psi,u\rangle_{l_2},\quad \psi=(\psi_1,\ldots,\psi_n,\ldots)$$

и сопряженную систему

$$\dot{\psi}_i(t)=-H_{x^i}(x_*(t),u_*(t),\psi,a_0)\equiv 0,\quad t\geqslant 0,\quad i=1,2,\ldots$$

Отсюда имеем $\psi_i(t)\equiv c_i=\text{const}$, $\psi(t)=(c_1,c_2,\ldots,c_n,\ldots)\in l_2$. Если $c_n\neq 0$ для некоторого $n\geqslant 1$, то из условия $\max\limits_{u\in V}H(x_*(t),u,\psi(t),a_0)$ однозначно определится $u^n(t)=\left(\dfrac{1}{n}+\dfrac{1}{n^2}\right)\operatorname{sign} c_n$, $t\geqslant 0$, что не совпадает с $u_*^n(t)=\dfrac{1}{n}$. Это значит, что принцип максимума в рассматриваемой задаче может иметь место только в вырожденном случае $c^n=0$, $n=1,2,\ldots$, т. е. $\psi(t)\equiv 0$ и $H(x_*(t),u_*(t),\psi(t),a_0)=a_0=\text{const}$. Согласно условию трансверсальности (6.2.43) тогда $H(x_*(t_*),u_*(t_*),\psi(t_*),a_0)=0$, так что $a_0=0$. В результате получаем $(a_0,\psi(t))\equiv 0$, что противоречит принципу максимума (см. теорему 6.2.2).

Рассмотренный пример показывает, что для задач оптимального управления в банаховых пространствах принцип максимума в общем случае не имеет места. Тем не менее существуют классы задач оптимального управления, для которых принцип максимума остается верным и в том случае, когда фазовое пространство не является конечномерным [287–290; 641]. С другой стороны, можно указать и такие классы задач оптимального управления с конечномерным фазовым пространством, в которых принцип максимума не имеет места: такие задачи с дискретным временем см. ниже в § 6, с непрерывным временем — в [137].

8. В заключение отметим, что функцию, дифференцируемую в смысле определения 3, в литературе часто называют *сильно дифференцируемой* или *дифференцируемой по Фреше* [393]. Существуют и другие определения дифференцируемости функции, отличные от сильной дифференцируемости. Приведем одно из них.

Определение 6. Пусть функция $J(u)$ (определена в некоторой окрестности $O(u, \gamma)$ точки u из банахова пространства B. Говорят, что функция $J(u)$ *слабо дифференцируема* или *дифференцируема по Гато* в точке u, если существует предел

$$\lim_{t \to 0} \frac{J(u+th) - J(u)}{t} = J'(u, h) \tag{60}$$

при всех $h \in B$. Если существует $J'(u) \in B^*$, что $J'(u, h) = \langle J'(u), h\rangle \ \forall h \in B$, то $J'(u)$ называют *слабой производной* или *производной Гато* функции $J(u)$ в точке u.

Заметим, что из дифференцируемости по Гато не следует существование производной Гато, о чем свидетельствует

Пример 14. Пусть $J(u) = \dfrac{xy^2}{x^2+y^2}$ при $x^2+y^2 > 0$, $J(0) = 0$, $u = (x, y) \in E^2$. Очевидно, $J(tu) = tJ(u) \ \forall u \in E^2$, $\forall t \in \mathbb{R}$, так что $\lim\limits_{t \to 0} \dfrac{J(0+th) - J(0)}{t} = J(h) = J'(0, h) \ \forall h \in E^2$. Это означает, что рассматриваемая функция дифференцируема по Гато, но $J'(0, h)$ зависит от h нелинейно.

Нетрудно видеть, что из дифференцируемости по Фреше следует дифференцируемость по Гато, а также совпадение производных Фреше и Гато. Обратное утверждение неверно.

Пример 15. Пусть $u = (x, y) \in E^2$, $J(u) = 1$, если $y = x^2$, $J(u) \equiv 0$, если $y \neq x^2$. В точке $u = 0$ эта функция имеет производную Гато $J'(0) = 0$, но по Фреше она недифференцируема и, более того, разрывна в этой точке.

Такое различие между производными Фреше и Гато связано с тем, что в случае дифференцируемости по Фреше сходимость в (60) является равномерной по всем h, $\|h\| < \gamma$, а в случае дифференцируемости по Гато такая равномерность и представление $J'(u, h) = \langle J'(u), h\rangle$ необязательны.

Из (60) следует, что если функция $J(u)$ в точке u имеет производную Фреше или Гато, то она дифференцируема в этой точке по любому направлению e, $\|e\| = 1$, причем $\dfrac{dJ(u)}{de} = \langle J'(u), e\rangle$ (ср. с формулой (4.2.14)). Из существования производной по направлениям не следует дифференцируемость по Гато.

Пример 16. Функция $J(u) = \|u\|_B$ в точке $u = 0$ имеет производную по любому направлению, причем $\dfrac{dJ(0)}{de} = \|e\| = 1$. Однако отношение $\dfrac{(J(th) - J(0))}{t} = \dfrac{|t|}{t}\|h\|$ при $t \to 0$ не имеет предела, так что $J(u)$ недифференцируема по Гато в точке $u = 0$.

Понятие производной Гато широко используется при исследовании экстремальных задач в банаховых пространствах ([14; 393; 605; 638] и др.).

Упражнения

1. Результаты примера 5 вывести из примера 4. Указание: представить функцию (19) в виде $J(u) = \langle A^*Au, u\rangle - 2\langle A^*b, u\rangle + \|b\|^2$, $u \in H$.

2. Пусть B — банахово пространство, F — гильбертово пространство, оператор $A \in \mathcal{L}(H \to F)$, $b \in F$. Доказать, что функция $J(u) = \|Au - b\|_F^2$ выпукла и дважды непрерывно дифференцируема на B, производные $J'(u)$, $J''(u)$ представимы в виде (21), где $A^* \in \mathcal{L}(F \to B^*)$ — сопряженный к A оператор.

3. Доказать, что решение задачи Коши

$$\dot{x}(t) = D(t)x(t) + f(t), \quad t_0 \leqslant t \leqslant T; \quad x(t_0) = x_0$$

представимо в виде

$$x(t) = \Phi(t, t_0)x_0 + \int_{t_0}^{t} \Phi(t, \tau)f(\tau)d\tau, \quad t_0 \leqslant t \leqslant T;$$

решение задачи Коши

$$\dot{\psi}(t) = -D^\top(t)\psi(t) + g(t), \quad t_0 \leqslant t \leqslant T, \quad \psi(T) = \psi_1,$$

представимо в виде

$$\psi(t) = \Phi^\top(T, t)\psi_1 + \int_{T}^{t} \Phi^\top(\tau, t)g(\tau)d\tau, \quad t_0 \leqslant t \leqslant T,$$

где матрица $\Phi(t, \tau)$ — решение задачи (34), $\Phi^\top(t, \tau)$ — транспонированная матрица. Указание: предварительно установить следующие свойства матрицы $\Phi(t, \tau)$:

$$\Phi(t, \tau) = \Phi(t, \xi)\Phi(\xi, \tau) \quad \forall t, \tau, \xi \in [t_0, T]; \quad [\Phi(t, \tau)]^{-1} = \Phi(\tau, t) \quad \forall t, \tau \in [t_0, T];$$
$$\frac{d\Phi(t, \tau)}{d\tau} = -\Phi(t, \tau)D(\tau) \quad \forall t, \tau[t_0, T].$$

4. Доказать, что функция $J(x_0) = |x(T; x_0) - b|^2$, где $x(t; x_0)$ — решение задачи

$$\dot{x}(t) = D(t)x(t), \quad t_0 \leqslant t \leqslant T; \quad x(t_0) = x_0, \tag{61}$$

$D(t)$ — матрица размера $n \times n$ с элементами $d_{ij}(t) \in L_\infty[t_0, T]$, дважды непрерывно дифференцируема на E^n; вывести формулы для $J'(u)$, $J''(u)$. Указание: ввести оператор $Ax_0 = x(T; x_0) \in \mathcal{L}(E^n \to E^n)$ и воспользоваться результатами примера 5; показать, что сопряженный к A оператор A^* действует на элемент $c \in E^n$ по правилу: $A^*c = \psi(t_0; c)$, где $\psi(t; c)$ — решение задачи: $\dot{\psi}(t) = -D^\top(t)\psi(t)$, $t_0 \leqslant t \leqslant T$; $\psi(T) = c$; воспользоваться упражнением 3 и доказать формулы: $J'(u) = 2\Phi^\top(T, t_0)[\Phi(T, t_0)x_0 - b]$, $J''(u) = 2\Phi^\top(T, t_0)\Phi(T, t_0)$.

5. Доказать, что функция $J(x_0) = \int\limits_{t_0}^{T} |x(t; x_0) - b(t)|^2 dt$, где $x(t; x_0)$ — решение задачи (61), дважды непрерывно дифференцируема на $H = E^n$; вывести формулы для $J'(u)$, $J''(u)$. Указание: ввести оператор $Ax_0 = x(t; x_0) \in \mathcal{L}(H \to F)$, $F = L_2^n[t_0, T]$ и воспользоваться результатами примера 5; показать, что сопряженный к A оператор A^* действует на элемент $c = c(t) \in F$ по правилу: $A^*c = \psi(t_0; c)$, где $\psi(t; c)$ — решение задачи: $\dot{\psi}(t) = -D^\top(t)\psi(t) - c(t)$, $t_0 \leqslant t \leqslant T$; $\psi(T) = 0$; воспользоваться упражнением 3 и доказать формулы: $J'(u) = 2\int\limits_{t_0}^{T} \Phi^\top(\tau, t_0)[\Phi(\tau, t_0)x_0 - b(\tau)]d\tau$, $J''(u) = 2\int\limits_{t_0}^{T} \Phi^\top(\tau, t_0)\Phi(\tau, t_0)d\tau$.

6. Доказать, что функция $J(u, x_0) = |x(T; u, x_0) - b|^2$, где $x(t; u, x_0)$ — решение задачи

$$\dot{x}(t) = D(t)x(t) + B(t)u(t), \quad t_0 \leqslant t \leqslant T; \quad x(t_0) = x_0 \tag{62}$$

дважды непрерывно дифференцируема по совокупности переменных $z = (u, x_0) \in H = L_2^r[t_0, T] \times E^n$. Указание: ввести оператор $Az = x(T; u, x_0) \in \mathcal{L}(H \to E^n)$, воспользоваться результатами примера 6, упражнения 4.

7. Доказать, что функция $J(u, x_0) = \int\limits_{t_0}^{T} |x(t; u, x_0) - b(t)|^2 dt$, где $x(t; u, x_0)$ — решение задачи (62), дважды непрерывно дифференцируема по совокупности переменных $z = (u, x_0) \in H = L_2^r[t_0, T] \times E^n$. Указание: ввести оператор $Az = x(t; u, x_0) \in \mathcal{L}(H \to F)$, $F = L_2^n[t_0, T]$, воспользоваться результатами примера 8, упражнения 5.

8. Пользуясь теоремой 3, написать условия минимума выпуклой функции $J(u)$ из упражнений 4–7 на выпуклом множестве $U \subseteq H$.

9. Написать необходимое условие минимума функции из примера 3 для случая $U = C[a, b]$.

10. Пусть H — гильбертово пространство, функция $J(u) \in C^2(H)$. Показать, что если в некоторой точке $u_* \in H$ выполняются условия $J'(u_*) = 0$, $\langle J''(u_*)h, h\rangle > 0 \ \forall h \in H$, $h \neq 0$, то этого, вообще говоря, еще недостаточно для того, чтобы в точке u_* достигался локальный или глобальный минимум $J(u)$ на H. Указание: рассмотреть функцию $J(u) = \sum\limits_{n=1}^{\infty} \left(\dfrac{(u^n)^2}{n^3} - (u^n)^4 \right)$, $u \in l_2$, в точке $u_* = 0$; убедиться, что $J'(u_*) = 0$, $\langle J''(u_*)h, h\rangle = 2\sum\limits_{n=1}^{\infty} \dfrac{(h^n)^2}{n^3} > 0$ $\forall h \neq 0$, но в точке $u = (0, \ldots, 0, u^n = \frac{1}{n}, 0, \ldots)$ значение $J(u) < J(u_*) = 0$ [393].

11. Пусть H — гильбертово пространство, функция $J(u) \in C^2(H)$. Доказать, что если в некоторой точке $u_* \in H$ выполняются условия

$$J'(u_*) = 0, \quad \langle J''(u_*)h, h\rangle > \mu\|h\|^2, \quad \forall h \in H, \quad \mu = \text{const} > 0, \tag{63}$$

то u_* — точка строгого локального минимума функции $J(u)$ на H. Убедитесь, что при $H = E^n$ условия $J'(u_*) = 0$, $\langle J''(u_*)h, h\rangle > 0 \ \forall h \in E^n$, $h \neq 0$, равносильны (63).

12. Доказать, что функция $J(u)=\|u\|_B$ дифференцируема в $B=L_p[0,1]$, $1<p<\infty$, всюду, кроме $u=0$, и найти $J'(u)$. Будет ли $J(u)$ дифференцируема (по Фреше или по Гато) в $B=L_1[0,1]$? В $L_\infty[0,1]$? В $C[0,1]$?

13. Доказать дифференцируемость отображения $F\colon L_2[a,b]\to L_2[c,d]$, которое определяется равенством

$$F(u)=\int_a^b K(t,\tau)u(\tau)d\tau,\quad c\leqslant t\leqslant d,$$

где $K(t,\tau)\in L_2(Q)$, $Q=\{(t,\tau)\in E^2\colon c\leqslant t\leqslant d,\ a\leqslant\tau\leqslant b\}$.

14. Найти производную отображения (43), действующего из $L_2^r[t_0,T]$ в $H_n^1[t_0,T]$.

15. Пусть функции $J_1(u),\ldots,J_m(u)$ переменных $u=(u^1,\ldots,u^n)$ непрерывно дифференцируемы на E^n. Доказать, что тогда отображение $F(u)=\begin{pmatrix}f_1(u)\\ \cdots\\ f_m(u)\end{pmatrix}$, действующее из $X=E^n$ в $Y=E^m$, непрерывно дифференцируемо на E^n и его производная $F'(u)$ представляет собой линейный оператор с матрицей

$$F'(u)=\begin{pmatrix}\frac{\partial f_1(u)}{\partial u^1},&\ldots,&\frac{\partial f_1(u)}{\partial u^n}\\ \cdots\cdots\cdots&\cdots\cdots&\cdots\cdots\\ \frac{\partial f_m(u)}{\partial u^1},&\ldots,&\frac{\partial f_m(u)}{\partial u^n}\end{pmatrix},\quad u\in E^n$$

(матрица Якоби). Доказать, что если $f_i(u)\in C^2(E^n)$, то $F(u)\in C^2(E^n)$.

16. Показать, что множество U из примера 12 не имеет опорной гиперплоскости во всех точках $v=v(t)\in U$, для которых $|v(t)|<1$ почти всюду на $[0,1]$. Доказать, что если $v=v(t)\in U$ и $|v(t)|\equiv 1$ на множестве A положительной меры, то через такую точку v можно провести опорную к U гиперплоскость с нормальным вектором $c=c(t)\in L_2[0,1]$, где $c(t)=-\operatorname{sign}v(t)$ при $t\in A$ и $c(t)=0$ при $t\in[0,1]\setminus A$.

17. Пусть в пространстве $B=C[0,1]$ даны два множества $M=\{u=u(t)\in L_2[0,1]\colon |u(t)|\leqslant 1, 0\leqslant t\leqslant 1\}$ и $N=\left\{u=u(t)\in L_2[0,1]\colon \int_0^1 \operatorname{sign}\left(\frac{1}{2}-t\right)u(t)dt=1\right\}$. Доказать, что M и N выпуклы, замкнуты, не имеют общих точек, но не могут быть сильно отделимы (ср. с теоремой 4.5.3).

18. Пусть $U=\{u=(u^1,\ldots,u^n,\ldots)\in l_2\colon |u^n|\leqslant\frac{1}{n}+\frac{1}{n^2},\ n=1,2,\ldots\}$. Доказать, что в точке $v=(1,1/2,\ldots,1/n,\ldots)\in U$ нельзя провести опорную к U гиперплоскость (см. пример 13). Имеет ли U внутренние точки в l_2? Выяснить, к каким точкам из U можно провести опорную к U гиперплоскость.

19. Доказать, что «гильбертов кирпич» (см. пример 11) компактен в метрике l_2 (определение 2.1).

20. Пусть U — открытое выпуклое множество из банахова пространства B, функция $J(u)$ конечна, полунепрерывна снизу (в метрике B) и выпукла на U. Доказать, что $J(u)$ имеет субградиент во все точках $u\in U$ и слабо полунепрерывна снизу на U [233].

21. Доказать, что теорема 3 остается справедливой, если функция $J(u)$ имеет производную по Гато на множестве U.

22. Доказать, что если производная Гато функции $J(u)$ существует в некоторой окрестности точки u_0 и непрерывна в этой окрестности, то в точке u_0 существует производная Фреше, и она совпадает с производной Гато [393].

23. Пусть $J(x)=\int_a^b f(t,x(t),\dot{x}(t))dt$, где $f(t,x,y)$ — дважды непрерывно дифференцируемая функция по совокупности своих аргументов, $x(t)\in C^1[a,b]$. Доказать, что функция $J(x)$ дифференцируема в пространстве $C^1[a,b]$, и написать условие минимума этой функции [393, стр. 502].

24. Доказать, что в пространстве l_1 функция $y=|x|_1=\sum_{i=1}^\infty |x^i|$ недифференцируема ни в одной точке (ни по Фреше, ни по Гато).

25. Пусть отображение $F\colon X\to Y$, где X, Y — нормированные пространства, дифференцируемо на X. Показать, что $\|F(x+h)-F(x)\|\leqslant \sup\limits_{0\leqslant\theta\leqslant 1}\|F'(x+\theta h)\|\|h\|$ ([393], стр. 483). Убедиться, что это неравенство, вообще говоря, не может быть заменено равенством $F(x+h)-F(x)=F'(x+\theta h)h$ при каком-либо θ, $0\leqslant\theta\leqslant 1$ (ср. с формулой (9) при $Y=E^1$), рассмотрев пример: $F(t)=\begin{pmatrix}\cos 2\pi t\\ \sin 2\pi t\end{pmatrix}\colon E^1\to E^2$ на отрезке $0\leqslant t\leqslant 1$.

§ 4. Методы минимизации

Здесь мы будем предполагать, что читатель знаком с большинством из рассмотренных в части I методов минимизации. Заметим, что из этих методов лишь некоторые являются сугубо конечномерными, т. е. приспособленными для решения задач минимизации лишь в конечномерных пространствах — это симплекс-метод, метод покоординатного спуска и некоторые другие методы. Большинство же описанных в главе 5 методов минимизации вполне могут быть применены для минимизации функций на множествах из бесконечномерных банаховых и гильбертовых пространств — это градиентный метод, методы проекции градиента, условного градиента, возможных направлений, сопряженных градиентов, штрафных функций, Ньютона и др. Идеи и описание упомянутых методов минимизации в бесконечномерных пространствах по форме ничем не отличаются от их описания в конечномерном случае. Поэтому здесь мы ограничимся лишь кратким описанием некоторых из этих методов, отсылая читателя за подробностями к главе 5. Далее, формулировки и доказательства теорем сходимости для большинства упомянутых методов в бесконечномерном случае могут быть получены путем небольшой корректировки соответствующих конечномерных теорем и их доказательства из главы 5. Для того чтобы показать, как это делается, мы ниже приведем несколько таких теорем. Некоторые из излагаемых методов будем иллюстрировать на примере следующей задачи минимизации квадратичной функции

$$J(u)=\|Au-b\|_F^2\to\inf,\quad u\in U, \tag{1}$$

где $A\in\mathcal{L}(H\to F)$, H, F — гильбертовы пространства, $b\in F$, U — выпуклое замкнутое множество из H (задача (2.3)). Напоминаем формулы производных этой функции (пример 3.5):

$$J'(u)=2A^*(Au-b),\quad J''(u)=2A^*A. \tag{2}$$

Для иллюстрации методов также будем использовать задачу оптимального управления:

$$J(u)=|x(T,u)-b|^2\to\inf; \tag{3}$$

$$\dot{x}(t)=D(t)x(t)+B(t)u(t)+f(t),\quad t_0\leqslant t\leqslant T;\quad x(t_0)=x_0; \tag{4}$$

$$u=u(t)\in U\subseteq L_2^r[t_0,T] \tag{5}$$

(обозначения см. в примерах 2.14, 2.15). Градиент функции (3) имеет вид

$$J'(u)=2B^\top(t)\psi(t;u),\quad t_0\leqslant t\leqslant T, \tag{6}$$

где $\psi(t;u)$ — решение задачи Коши

$$\dot{\psi}(t)=-D^\top(t)\psi(t),\quad t_0\leqslant t\leqslant T;\quad \psi(T)=x(T;u)-b \tag{7}$$

(см. примеры 3.6, 3.7). Примеры других задач оптимального управления см. ниже в §§ 5–12.

1. Градиентный метод может применяться для поиска приближенного решения задачи

$$J(u) \to \inf; \quad u \in H,$$

где H — гильбертово пространство, $J(u) \in C^1(H)$. Этот метод заключается в построении последовательности $\{u_k\}$ по правилу

$$u_{k+1} = u_k - \alpha_k J'(u_k), \quad k = 0, 1, \ldots, \tag{8}$$

где u_0 — некоторая заданная начальная точка, α_k — положительная величина. Если $J'(u_k) \neq 0$, то α_k можно выбрать так, чтобы $J(u_{k+1}) < J(u_k)$. В самом деле, из равенства (3.5) имеем

$$J(u_{k+1}) - J(u_k) = \alpha_k(-\|J'(u_k)\|^2 + o(\alpha_k)/\alpha_k) < 0$$

при всех достаточно малых $\alpha_k > 0$. Если $J'(u_k) = 0$, то процесс (8) прекращается и при необходимости проводится дополнительное исследование поведения функции в окрестности точки u_k для выяснения того, будет ли u_k принадлежать U_* или нет. В частности, если $J(u)$ — выпуклая функция на H, то согласно теореме 3.3 $u_k \in U_*$. Различные способы выбора величины α_k в методе (8) описаны в § 5.1. Упомянем два из них: первый — α_k выбирается из условия

$$f_k(\alpha_k) = \inf_{\alpha \geqslant 0} f_k(\alpha), \quad f_k(\alpha) = J(u_k - \alpha J'(u_k)) \tag{9}$$

(этот вариант градиентного метода называют методом скорейшего спуска), второй — когда $J(u) \in C^{1,1}(H)$ (см. определение 2.6.1) и известна постоянная $L > 0$ из неравенства

$$\|J'(u) - J'(v)\| \leqslant L\|u - v\|, \quad u, v \in H, \tag{10}$$

величину α_k в (8) можно взять из условий

$$0 < \varepsilon_0 \leqslant \alpha_k \leqslant 2/(L + 2\varepsilon), \tag{11}$$

где ε, ε_0 — положительные числа, являющиеся параметрами метода.

На практике итерации (8) продолжают до тех пор, пока не выполнится какой-либо критерий окончания счета, описанный, например, в § 5.1.

Посмотрим, как выглядит метод (8), (9) на примере задачи (1) при $U = H$. С учетом формулы (2) для градиента процесс (8) можем записать в виде

$$u_{k+1} = u_k - \alpha_k(2A^*(Au_k - b)), \quad k = 1, 2, \ldots \tag{12}$$

Нетрудно убедиться, что функция $f_k(\alpha)$ здесь является квадратным трехчленом. В самом деле, для любых $u, h \in H$ имеем:

$$\begin{aligned} g(\alpha) = J(u + \alpha h) &= \|A(u + \alpha h) - b\|^2 = \|\alpha Ah + (Au - b)\|^2 = \\ &= \alpha^2\|Ah\|^2 + 2\alpha\langle Ah, Au - b\rangle + \|Au - b\|^2 = \\ &= \alpha^2\|Ah\|^2 + 2\alpha\langle h, J'(u)\rangle + J(u) \quad \forall \alpha \in \mathbb{R}. \end{aligned} \tag{13}$$

Из формулы (13) при $u = u_k$, $h = -J'(u_k) = -2A^*(Au_k - b)$ следует, что

$$f_k(\alpha) = \alpha^2\|2A(A^*(Au_k - b))\|^2 - \alpha\|2A^*(Au_k - b)\|^2 + J(u_k), \quad \alpha \in \mathbb{R}. \tag{14}$$

Отсюда видно, что при $AA^*(Au_k - b) \neq 0$ условие (9) однозначно определяет величину

$$\alpha_k = \frac{\|A^*(Au_k - b)\|^2}{\|AA^*(Au_k - b)\|^2} > 0. \tag{15}$$

Если $AA^*(Au_k - b) = 0$, то процесс (12) останавливается, так как тогда $\|A^*(Au_k - b)\|^2 = \langle AA^*(Au_k - b), Au_k - b\rangle = 0$, $f_k(\alpha) \equiv J(u_k)$, $J'(u_k) = 0$ и $u_k \in U_*$ (теорема 3.3). Таким образом, для задачи (1) метод скорейшего спуска (8), (9) порождает процесс (12), (15).

Явные формулы для метода (8), (9) можно выписать и в случае задачи (3)–(5) с $U = L_2^r[t_0, T]$. Их нетрудно получить, если эту задачу сначала свести к задаче (1) (см. пример 2.15) и далее воспользоваться формулами (12), (15). Получим процесс

$$u_{k+1}(t) = u_k(t) - \alpha_k \cdot 2B^\top(t)\psi(t, u_k), \quad t_0 \leqslant t \leqslant T, \quad k = 1, 2, \ldots, \tag{16}$$

где $\psi(t, u_k)$ — решение задачи (7) при $u = u_k$,

$$\begin{aligned} \alpha_k &= -\frac{\langle x(T, u_k) - b, x(T, u_k - J'(u_k)) - x(T, u_k)\rangle}{|x(T, u_k - J'(u_k)) - x(T, u_k)|^2} = \\ &= \frac{1}{2}\frac{\int_{t_0}^{T} |B^T(t)\psi(t, u_k)|^2 dt}{|x(T, u_k - J'(u_k)) - x(T, u_k)|^2} > 0 \end{aligned} \tag{17}$$

(подробнее вывод формул (16), (17) см. в [151]).

Приведем теорему сходимости метода скорейшего спуска, представляющую собой обобщение теорем 5.1.1–5.1.3 на случай гильбертовых пространств.

Теорема 1. *Пусть функция $J(u)$ определена на гильбертовом пространстве H, $J(u) \in C^{1,1}(H)$, $J_* = \inf_H J(u) > -\infty$. Пусть $\{u_k\}$ — последовательность, полученная методом* (8), (9) *при некотором начальном приближении $u_0 \in H$. Тогда последовательность $\{J(u_k)\}$ монотонно убывает и*

$$\lim_{k \to \infty} \|J'(u_k)\| = 0.$$

Если, кроме того, функция $J(u)$ выпукла на H, и множество $M(u_0) = \{u \in U: J(u) \leqslant J(u_0)\}$ ограничено, то последовательность $\{u_k\}$ минимизирует эту функцию на H, причем справедлива оценка скорости сходимости

$$0 \leqslant J(u_k) - J_* \leqslant c_1/k, \quad k = 1, 2, \ldots, \quad c_1 = \text{const} \geqslant 0.$$

Если функция $J(u)$ еще и сильно выпукла на H, то $\{u_k\}$ сходится к единственной точке минимума u_* по норме H, причем

$$\begin{aligned} 0 \leqslant J(u_k) - J_* &\leqslant (J(u_0) - J_*)q^k, \\ \|u_k - u_*\|^2 &\leqslant (2/\mu)(J(u_0) - J_*)q^k, \quad k = 0, 1, \ldots, \end{aligned}$$

где $q = 1 - \mu/L$, $0 \leqslant q < 1$, $\mu > 0$ — постоянная из теоремы 2.2.

Эта теорема доказывается также, как аналогичные теоремы 5.1.1–5.1.3.

Покажем, что градиенты функций (1), (3) удовлетворяют условию (10) и, следовательно, для их минимизации можно использовать метод (8) с выбором шага α_k по правилу (11). Для функции (1) условие (10) является простым следствием формулы (2):

$$\|J'(u) - J'(v)\| = \|2A^*A(u-v)\| \leqslant 2\|A^*A\|\|u-v\| \quad \forall u, v \in H, \tag{18}$$

так что здесь $L = 2\|A^*A\|$. Из (18) с учетом связи задач (1) и (3)–(5) (см. примеры 2.15, 3.7) можем сказать, что градиент функции (9) также удовлетворяет условию (10). Приведем оценку сверху постоянной L в этом случае. Приращение $\Delta x(t) = x(t;u) - x(t;v)$, $t_0 \leqslant t \leqslant T$, согласно (4) является решением задачи Коши

$$\Delta\dot{x}(t) = D(t)\Delta x(t) + B(t)(u(t) - v(t)), \quad t_0 \leqslant t \leqslant T; \quad \Delta x(t_0) = 0.$$

Отсюда, пользуясь оценкой (2.9), имеем

$$\begin{gathered} |x(t;u) - x(t;v)| \leqslant c_0\|u-v\|_{L_2^r} \quad \forall t \in [t_0, T]; \\ c_0 = \sqrt{T-t_0}\,B_{\max}e^{D_{\max}(T-t_0)}. \end{gathered} \tag{19}$$

В силу (7) функция $\Delta\psi(t) = \psi(t;u) - \psi(t;v)$, $t_0 \leqslant t \leqslant T$, является решением задачи Коши

$$\Delta\dot{\psi}(t) = -D^\top(t)\Delta\psi(t), \quad t_0 \leqslant t \leqslant T; \quad \Delta\psi(T) = x(T;u) - x(T;v),$$

откуда, рассуждая также, как при выводе оценки (2.9), получаем:

$$|\psi(t;u) - \psi(t;v)| \leqslant e^{A_{\max}(T-t_0)}|x(T;u) - x(T;v)| \quad \forall t \in [t_0, T]. \tag{20}$$

Подставим оценку (19) при $t = T$ в (20). Будем иметь

$$|\psi(t;u) - \psi(t;v)| \leqslant e^{A_{\max}(T-t_0)}c_0\|u-v\|_{L_2^r} \quad \forall t \in [t_0, T]. \tag{21}$$

Из формулы (6) и оценки (21) следует, что

$$\begin{gathered} \|J'(u) - J'(v)\|_{L_2^r} = \left(\int\limits_{t_0}^{T} |2B^\top(t)\Delta\psi(t)|^2 dt\right)^{1/2} \leqslant L\|u-v\|_{L_2^r}, \\ L = 2(T-t_0)B_{\max}^2 E^{2A_{\max}(T-t_0)}. \end{gathered} \tag{22}$$

Таким образом, установлено, что функции (1), (3) принадлежат классу $C^{1,1}(H)$ и в (12), (16) величину α_k можно взять из (11) (например, $\alpha_k = \frac{1}{L}$). Конечно, приведенная в (22) оценка для L может оказаться довольно грубой и шаг α_k из (11) тогда будет слишком маленьким, метод (16) будет сходиться медленно.

Теорема, аналогичная теореме 5.2.4, при $U = H$, $\delta_k = 0$, $k = 0, 1, \ldots$ остается справедливой и для метода (8), (11).

Отметим, что градиентный метод выше мы изложили применительно к гильбертовым пространствам. В банаховом пространстве B градиентный метод в форме (9) писать нельзя, так как слагаемые в правой части (9) принадлежат разным пространствам ($u_k \in B$, $-\alpha_k J'(u_k) \in B^*$) и их сумма не

имеет смысла. Не вдаваясь в детали, укажем, что в банаховом пространстве аналогом (9) является процесс $u_{k+1} = u_k - \alpha_k p_k$, где направление $p_k \in B$ определяется из условия $\min\limits_{\|h\| \leqslant 1} \langle J'(u_k), h\rangle = \langle J'(u_k), p_k\rangle$.

2. Метод проекции градиента может применяться для поиска приближенного решения задачи

$$J(u) \to \inf; \quad u \in U, \tag{23}$$

где U — выпуклое замкнутое множество из гильбертова пространства H, $J(u) \in C^1(U)$. Для описания этого метода нам понадобится понятие проекции точки на множество.

Определение 4.4.1 проекции $\mathcal{P}_U(u)$ точки $u \in H$ на множество U, а также свойства проекции и условия оптимальности для задачи (23), выраженные в теоремах 4.4.1, 4.4.2, 4.4.4, сохраняют силу и в гильбертовом пространстве.

Метод проекции градиента для решения задачи (23) заключается в построении последовательности $\{u_k\}$ по правилу

$$u_{k+1} = \mathcal{P}_U(u_k - \alpha_k J'(u_k)), \quad k = 0, 1, \ldots, \tag{24}$$

где α_k — положительная величина. Если при некотором k оказалось, что $u_{k+1} = u_k$, то процесс (24) прекращают. В этом случае точка u_k удовлетворяет необходимому условию оптимальности, а если $J(u)$ — выпуклая функция, то $u_k \in U_*$ (теорема 4.4.4).

Различные способы выбора величины α_k в методе (24) описаны в § 5.2.

Заметим, что метод (24) при $U = H$ переходит в градиентный метод.

Методом (24) удобно пользоваться лишь в тех случаях, когда имеется явная формула для проекции точки на множество. Укажем несколько примеров множеств, когда нетрудно получить такую формулу. Проекция точки $u \in H$ на шар

$$U = S(\overline{u}, R) = \{u \in H: \|u - \overline{u}\| \leqslant R\}$$

представима в виде

$$\mathcal{P}_U(u) = \begin{cases} \overline{u} + R(u-\overline{u})/\|u-\overline{u}\| & \text{при} \quad \|u-\overline{u}\| > R, \\ u & \text{при} \quad \|u-\overline{u}\| \leqslant R; \end{cases} \tag{25}$$

проекция на гиперплоскость

$$\Gamma = \{u \in H: \langle c, u\rangle = \gamma\}$$

выражается формулой

$$\mathcal{P}_U(u) = u + (\gamma - \langle c, u\rangle)c/\|c\|^2;$$

если

$$U = \{u \in H: Au = b\},$$

где $A \in \mathcal{L}(H \to H)$, $b \in H$, оператор AA^* имеет обратный, то проекция точки $u \in H$ может быть записана в виде

$$\mathcal{P}_U(u) = u - A^*(AA^*)^{-1}(Au - b).$$

Приведенные формулы для проекций доказываются так же, как это делалось в примерах 4.4.1–4.4.3.

Метод (24) для задачи (1) запишется в виде

$$u_{k+1} = \mathcal{P}_U(u_k - \alpha_k(2A^*(Au_k - b))), \quad k = 0, 1, \ldots$$

Посмотрим, как выглядит метод проекции градиента для задачи (3)–(5), когда множество U имеет вид

$$U=\{u=u(t)=(u^1(t),\ldots,u^r(t))\in L_2^r[t_0,T]:\ \alpha_i(t)\leqslant u^i(t)\leqslant\beta_i(t)\quad\text{почти всюду на}\quad[t_0,T],\ i=1,\ldots,r\};\qquad(26)$$

здесь $\alpha_i(t)$, $\beta_i(t)$ — заданные функции из $L_2[t_0,T]$. Проекция любой точки $u=u(t)=(u^1(t),\ldots,u^r(t))\in L_2^r[t_0,T]$ на это множество представляет собой вектор-функцию $\mathcal{P}_U(u)=(w^1(t),\ldots,w^r(t))$, $t_0\leqslant t\leqslant T$, где

$$w^i(t)=\begin{cases}\alpha_i(t) & \text{при } u^i(t)<\alpha_i(t),\\ u^i(t) & \text{при } \alpha_i(t)\leqslant u^i(t)\leqslant\beta_i(t),\\ \beta_i(t) & \text{при } \beta_i(t)<u^i(t),\end{cases}\quad i=1,\ldots,r$$

(ср. с примером 4.4.5). Поэтому $(k+1)$-е приближение $u_{k+1}(t)=(u^1_{k+1}(t),\ldots,u^r_{k+1}(t))$, $t_0\leqslant t\leqslant T$, метода проекции градиента для задачи (3)–(5), (26) будет получаться по правилу

$$u^i_{k+1}(t)=\begin{cases}\alpha_i(t) \text{ при } u^i_k(t)-\alpha_k(2B^T(t)\psi(t;u_k))_i<\alpha_i(t),\\ u^i_k(t)-\alpha_k(2B^T(t)\psi(t;u_k))_i \text{ при}\\ \alpha_i(t)\leqslant u^i_k(t)-\alpha_k(2B^T(t)\psi(t;u_k))_i\leqslant\beta_i(t),\\ \beta_i(t) \text{ при } u^i_k(t)-\alpha_k(2B^T(t)\psi(t;u_k))_i>\beta_i(t),\end{cases}\qquad(27)$$

$i=1,\ldots,r$; здесь $\psi(t;u)$ — решение задачи (7), $(B^T(t)\psi(t;u_k))_i$ — i-я координата вектора $B^T(t)\,\psi(t;u_k)$. Согласно (22) градиент функции (3) удовлетворяет условию Липшица с постоянной $L=2(T-t_0)B^2_{\max}e^{2A_{\max}(T-t_0)}$. Поэтому при выборе α_k в (27) можно воспользоваться условием (11).

Если в задаче (3)–(5) множество U является шаром из $L_2^r[t_0,T]$, т. е.

$$U=\left\{u=u(t)\in L_2^r[t_0,T]:\ \int_{t_0}^{T}|u(t)-\overline{u}(t)|^2dt\leqslant R^2\right\},\qquad(28)$$

где $u=\overline{u}(t)$ — заданная функция из $L_2^r[t_0,T]$, R — заданное положительное число, то в силу формулы (25) метод проекции градиента приведет к последовательности, которая строится по правилу

$$u_{k+1}(t)=\begin{cases}\overline{u}(t)+R\dfrac{u_k(t)-\alpha_k2B^T(t)\psi(t;u_k)-\overline{u}(t)}{\left(\int_{t_0}^{T}|u_k(t)-\alpha_kB^T(t)\psi(t;u_k)-\overline{u}(t)|^2dt\right)^{1/2}}\\ \text{при } \int_{t_0}^{T}|u_k(t)-\alpha_k2B^T(t)\psi(t;u_k)-\overline{u}(t)|^2dt>R^2,\\ u_k(t)-\alpha_k2B^T(t)\psi(t;u_k)\\ \text{при } \int_{t_0}^{T}|u_k(t)-\alpha_k2B^T(t)\psi(t;u_k)-\overline{u}(t)|^2dt\leqslant R^2.\end{cases}\qquad(29)$$

Приведем теорему сходимости метода (24), (11), обобщающую теоремы 5.2.1, 5.2.2 на случай гильбертовых пространств.

Теорема 2. *Пусть функция $J(u)$ определена на выпуклом замкнутом множестве U гильбертова пространства H, $J(u)\in C^{1,1}(U)$, $J_*=\inf\limits_U J(u)>-\infty$. Пусть $\{u_k\}$ — последовательность, полученная методом (24), (11) при произвольном начальном приближении $u_0\in U$. Тогда последовательность $\{J(u_k)\}$ монотонно убывает и $\lim\limits_{k\to\infty}\|u_k-u_{k+1}\|=0$. Если, кроме того, функция $J(u)$ выпукла на H и множество $M(u_0)=\{u\in U:\ J(u)\leqslant J(u_0)\}$ ограничено, то последовательность $\{u_k\}$ минимизирует эту функцию на U, причем справедлива оценка*

$$0\leqslant J(u_k)-J_*\leqslant c_1/k,\quad k=1,2,\ldots;\quad c_1=\text{const}\geqslant 0.$$

Если функция $J(u)$ еще и сильно выпукла на U, то $\{u_k\}$ сходится к единственной точке минимума u_* по норме H, причем

$$\|u_k-u_*\|^2\leqslant c_2/k,\quad k=1,2,\ldots;\quad c_1=\text{const}\geqslant 0.$$

Для сильно выпуклых функций можно предложить другой вариант метода проекции градиента, имеющий более высокую скорость сходимости.

Теорема 3. *Пусть U — выпуклое замкнутое множество из H, функция $J(u)$ принадлежит $C^{1,1}(U)$ и сильно выпукла на U, и пусть $0<\alpha<2\mu L^{-2}$, где постоянные μ, L, $\mu\leqslant L$, взяты из (10) и теоремы 4.3.3. Тогда последовательность $\{u_k\}$, получаемая из (24) при $\alpha_k=\alpha$, $k=0,1,\ldots$, сходится к точке минимума u_* по норме H, причем справедлива оценка*

$$\|u_k-u_*\|\leqslant\|u_0-u_*\|(q(\alpha))^k,\quad k=0,1,\ldots,$$

где $q(\alpha)=(1-2\mu\alpha+\alpha^2L^2)^{1/2}$, $0<q(\alpha)<1$.

Теоремы 2, 3 доказываются также, как аналогичные теоремы 5.2.1–5.2.3.

3. Метод условного градиента может применяться для поиска приближенного решения задачи

$$J(u)\to\inf;\qquad u\in U,$$

где U — выпуклое замкнутое ограниченное множество из гильбертова пространства H, $J(u)\in C^1(U)$. Этот метод заключается в построении последовательности по следующему правилу: по известному k-му приближению находят вспомогательное приближение $\overline{u}_k\in U$ из условия

$$\langle J'(u_k),\overline{u}_k-u_k\rangle=\inf_U\langle J'(u_k),u-u_k\rangle,$$

или, что равносильно, из условия

$$\overline{u}_k\in U,\quad\langle J'(u_k),\overline{u}_k\rangle=\inf_U\langle J'(u_k),u\rangle,\qquad(30)$$

и затем полагают

$$u_{k+1}=u_k+\alpha_k(\overline{u}_k-u_k),\qquad 0\leqslant\alpha_k\leqslant 1.\qquad(31)$$

Заметим, что линейная функция $\langle J'(u_k),u\rangle$ слабо непрерывна на U, а множество U согласно теореме 2.6 слабо компактно. Отсюда и из теоремы 2.4 следует, что нижняя грань в условии (30) достигается хотя бы на одном элементе $\overline{u}_k\in U$.

Если при некотором k оказалось, что $\overline{u}_k = u_k$, то процесс (30), (31) прекращают. В этом случае в силу условия (30) имеем $\inf\limits_U \langle J'(u_k), u - u_k\rangle = \langle J'(u_k), u_k - u_k\rangle = 0$, т. е. $\langle J'(u_k), u - u_k\rangle \geqslant 0$ при всех $u \in U$. Согласно теореме 3.3 это означает, что точка u_k удовлетворяет необходимому условию оптимальности, а если при этом $J(u)$ — выпуклая функция, то $u_k \in U_*$.

Различные способы выбора величины α_k в (31) описаны в § 5.4. Например, α_k может выбираться из условия

$$f_k(\alpha_k) = \inf_{0 \leqslant \alpha \leqslant 1} f_k(\alpha) = f_{k*}, \quad f_k(\alpha) = J(u_k + \alpha(\overline{u}_k - u_k)). \tag{32}$$

Посмотрим, как выглядит метод условного градиента для задачи (3)–(5), когда множество U имеет вид (26) или (28). Согласно (30) для определения $\overline{u}_k = \overline{u}_k(t)$ нужно на множестве U минимизировать линейную функцию

$$\int_{t_0}^{T} \langle B^T(t)\psi(t; u_k), u(t)\rangle_{E^r} dt = \int_{t_0}^{T} \sum_{i=1}^{r} (B^T(t)\psi(t; u_k))_i u^i(t) dt.$$

Отсюда видно, что в случае множества (26) будем иметь

$$\overline{u}_k(t) = (\overline{u}_k^1(t), \ldots, \overline{u}_k^r(t)), \qquad t_0 \leqslant t \leqslant T, \quad \text{где}$$

$$\overline{u}_k^i(t) = \begin{cases} \alpha_i(t) & \text{при } (B^T(t)\psi(t; u_k))_i \geqslant 0, \\ \beta_i(t) & \text{при } (B^T(t)\psi(t; u_k))_i < 0, \end{cases}$$

а если U — шар (28), то с помощью неравенства Коши — Буняковского получим

$$\overline{u}_k(t) = \overline{u}(t) - R \frac{B^T(t)\psi(t; u_k)}{\left(\int_{t_0}^{T} |B^T(t)\psi(t; u_k)|^2_{E^r} dt\right)^{1/2}}.$$

Параметр α_k, определяемый условиями (32) в рассматриваемой задаче, может быть выписан явно. Пользуясь тождеством (27) при $u = \overline{u}_k$, $v = u_k$, $\beta = 1 - \alpha$ и формулой (6), нетрудно показать, что [151]:

$$f_k(\alpha) = J(u_k) + 2\alpha\langle x(T, u_k) - b, x(T, \overline{u}_k) - x(T, u_k)\rangle + \alpha^2 |x(T, \overline{u}_k) - x(T, u_k)|^2 =$$

$$= J(u_k) + 2\alpha \int_{t_0}^{T} \langle B^T(t)\psi(t; u_k), \overline{u}_k(t) - u_k(t)\rangle_{E^r} dt +$$

$$+ \alpha^2 |x(T, \overline{u}_k) - x(T, u_k)|^2, \qquad -\infty < \alpha < +\infty. \tag{33}$$

Если $x(T, \overline{u}_k) - x(T, u_k) = 0$, то $f_k(\alpha) \equiv J(u_k) = \text{const}$ при всех α. Следовательно, $f_k'(\alpha) = \langle J'(u_k + \alpha(\overline{u}_k - u_k)), \overline{u}_k - u_k\rangle = 0$ при всех α. Отсюда с учетом условия (30) получаем $f_k'(0) = 0 = \langle J'(u_k), \overline{u}_k - u_k\rangle \leqslant \langle J'(u_k), u - u_k\rangle$ для любого $u \in U$. Согласно теореме 3.3 тогда $u_* = u_k = u_k(t)$ — оптимальное управление в задаче (3)–(5).

Рассмотрим случай $x(T, \overline{u}_k) \neq x(T, u_k)$. Тогда функция (33) представляет собой квадратный трехчлен и достигает своей нижней грани на числовой оси при

$$\alpha = \alpha_k^* = -\frac{\int_{t_0}^{T} \langle B^T(t)\psi(t; u_k), \overline{u}_k(t) - u_k(t)\rangle_{E^r} dt}{|x(T, \overline{u}_k) - x(T, u_k)|^2_{E^n}} =$$

$$= -\frac{\langle x(T, u_k) - b, x(T, \overline{u}_k) - x(T, u_k)\rangle_{E^n}}{|x(T, \overline{u}_k) - x(T, u_k)|^2_{E^n}}. \tag{34}$$

Так как в силу условия (30)

$$2\int_{t_0}^{T} \langle B^T(t)\psi(t, u_k), \overline{u}_k(t) - u_k(t)\rangle_{E^r} dt = \langle J'(u_k), \overline{u}_k - u_k\rangle \leqslant \langle J'(u_k), u_k - u_k\rangle = 0,$$

то ясно, что $\alpha_k^* \geqslant 0$. Возможен случай $\alpha_k^* = 0$. Согласно условию (30) и формуле (34) это значит, что

$$0 = \langle J'(u_k), \overline{u}_k(t) - u_k(t)\rangle \leqslant \langle J'(u_k), u - u_k\rangle$$

при всех $u \in U$. В силу теоремы 3.3 тогда $u_* = u_k = u_k(t)$ — оптимальное управление в задаче (3) — (5).

Если $\alpha_k^* < 0$, то квадратный трехчлен достигает своей нижней грани на отрезке $0 \leqslant \alpha \leqslant 1$ при

$$\alpha_k = \min\{1; \alpha_k^*\}. \tag{35}$$

Это и есть искомое явное выражение для α_k, удовлетворяющее условию (32). Для получения $(k+1)$-го приближения остается положить

$$u_{k+1}(t) = u_k(t) + \alpha_k(\overline{u}_k(t) - u_k(t)), \qquad t_0 \leqslant t \leqslant T.$$

Сходимость метода условного градиента для задачи (3)–(5) вытекает из следующей теоремы.

Теорема 4. *Пусть U — выпуклое замкнутое ограниченное множество из гильбертова пространства H, функция $J(u)$ принадлежит $C^{1,1}(U)$. Тогда для последовательности $\{u_k\}$, определяемой методом (30)–(32) при любом выборе начального приближения $u_0 \in U$, справедливо равенство*

$$\lim_{k \to \infty} \langle J'(u_k), \overline{u}_k - u_k\rangle = 0. \tag{36}$$

Если, кроме того, функция $J(u)$ выпукла на U, то последовательность $\{u_k\}$ минимизирует эту функцию на U, причем справедлива оценка

$$0 \leqslant J(u_k) - J_* \leqslant c_1/k, \quad k = 1, 2, \ldots; \quad c_1 = \text{const} \geqslant 0. \tag{37}$$

Если $J(u)$ еще и сильно выпукла на U, то $\{u_k\}$ сходится к единственной точке минимума u_* по норме H, причем

$$\|u_k - u_*\|^2 \leqslant c_2/k, \quad k = 1, 2, \ldots; \quad c_2 = \text{const} \geqslant 0.$$

Доказательство. Прежде всего заметим, что из ограниченности множества U, условия $J(u) \in C^{1,1}(U)$ и леммы 2.6.1 следует, что

$$|J(u)| \leqslant |J(u_0)| + \|J'(u_0)\|\|u - u_0\| + L\|u - u_0\|^2/2 \leqslant$$

$$\leqslant |J(u_0)| + \|J'(u_0)\|d + Ld^2/2 < \infty$$

при всех $u \in U$; здесь $d = \sup\limits_{u, v \in U} \|u - v\|$ — диаметр множества U. Это значит, что функция $J(u)$ ограничена на U. Следовательно, $J_* > -\infty$. Обозначим $J_k(u) = \langle J'(u_k), u - u_k\rangle$. Из (30) следует, что

$$J_k(\overline{u}_k) \leqslant J_k(u_k) = 0, \quad k = 0, 1, \ldots$$

Далее, справедливо неравенство

$$J(u_k) - J(u_{k+1}) \geqslant \alpha|J_k(\overline{u}_k)| - \alpha^2 L d^2/2 \tag{38}$$

при всех α, $0 \leqslant \alpha \leqslant 1$, $k = 0, 1, \ldots$, которое доказывается так же, как аналогичное неравенство (5.4.18). Отсюда

$$0 \leqslant |J_k(\overline{u}_k)| \leqslant \alpha L d^2/2 + (J(u_k) - J(u_{k+1}))/\alpha, \tag{39}$$

$k = 0, 1, \ldots$; $0 < \alpha \leqslant 1$. Так как $J(u_k)$ не возрастает и $J(u_k) \geqslant J_* > -\infty$, то $\{J(u_k)\}$ сходится и $J(u_k) - J(u_{k+1}) \to 0$ при $k \to \infty$. Поэтому, переходя в неравенстве (39) к пределу при $k \to \infty$, будем иметь

$$0 \leqslant \underline{\lim_{k\to\infty}} |J_k(\overline{u}_k)| \leqslant \overline{\lim_{k\to\infty}} |J_k(\overline{u}_k)| \leqslant \alpha L d^2/2$$

при всех α, $0 < \alpha \leqslant 1$. Отсюда при $\alpha \to +\infty$ получим равенство (26).

Пусть теперь $J(u)$ выпукла на U. Согласно теореме 2.8 тогда $U_* \neq \varnothing$. Возьмем произвольную точку $u_* \in U_*$. Из теоремы 4.2.2 и условия (30) имеем

$$0 \leqslant a_k = J(u_k) - J(u_*) \leqslant \langle J'(u_k), u_k - u_* \rangle = \\ = -J_k(u_*) \leqslant -J_k(\overline{u}_k) = |J_k(\overline{u}_k)|, \quad k = 0, 1, \ldots \tag{40}$$

Отсюда и из равенства (36) следует, что $\{u_k\}$ — минимизирующая последовательность.

Докажем оценку (37). Так как $J_k(\overline{u}_k) \to 0$ при $k \to \infty$, то найдется номер k_0 такой, что $0 \leqslant \gamma_k = |J_k(\overline{u}_k)|/(L d^2) \leqslant 1$ при всех $k \geqslant k_0$. Тогда максимальное значение функции $\alpha |J_k(\overline{u}_k)| - \alpha^2 d^2 L/2$ переменной α при $-\infty < \alpha < +\infty$, которое достигается при $\alpha = \gamma_k$, будет совпадать с максимальным значением этой функции на отрезке $0 \leqslant \alpha \leqslant 1$ при всех $k \geqslant k_0$. Поэтому, полагая в оценке (38) $\alpha = \gamma_k$, получим

$$a_k - a_{k+1} = J(u_k) - J(u_{k+1}) \geqslant |J_k(\overline{u}_k)|^2/(2L d^2), \quad k \geqslant k_0.$$

Отсюда и из неравенств (40) следует

$$a_k - a_{k+1} \geqslant a_k^2/(2L d^2), \quad k \geqslant k_0.$$

Остается применить лемму 2.6.4 и убедиться в справедливости оценки (37). Последнее утверждение теоремы для сильно выпуклых функций следует из оценки (37) и неравенства (4.3.3). □

Заметим, что описание метода условного градиента и теорема 4 сохраняют силу и в том случае, когда U — выпуклое замкнутое ограниченное множество из рефлексивного банахова пространства.

4. Метод возможных направлений может применяться для поиска приближенного решения задачи

$$J(u) \to \inf; \quad u \in U = \{u \in B: g_i(u) \leqslant 0, \ i = 1, \ldots, m\}, \tag{41}$$

где B — банахово пространство, $J(u), g_1(u), \ldots, g_m(u) \in C^1(B)$. Для описания этого метода нам понадобятся понятия возможного направления и возможного направления убывания функции, введенные определениями 4.2.3 и 5.5.1.

Метод возможных направлений заключается в следующем. Пусть $u_0 \in U$, $\varepsilon_0 > 0$ — некоторое начальное приближение. Допустим, что k-е приближение (u_k, ε_k), $u_k \in U$, $\varepsilon_k > 0$ при каком-то $k \geqslant 0$ уже известно. Определим множество номеров

$$I_k = \{i: 1 \leqslant i \leqslant m, \ -\varepsilon_k \leqslant g_i(u_k) \leqslant 0\}$$

и в пространстве переменных $z = (e, \sigma) \in B \times E^1$ рассмотрим вспомогательную задачу

$$\sigma \to \inf; \quad z = (e, \sigma) \in W_k = \{(e, \sigma): \langle J'(u_k), e \rangle \leqslant \sigma, \\ \langle g_i'(u_k), e \rangle \leqslant \sigma, \ i \in I_k; \ \|e\| \leqslant 1\}. \tag{42}$$

Множество W_k выпукло, замкнуто и ограничено, поэтому если B является рефлексивным банаховым пространством, то согласно теореме 2.8 задача (42) имеет решение. Пусть (e_k, σ_k) — решение задачи (42), т. е. $(e_k, \sigma_k) \in W_k$ и $\sigma_k = \inf\limits_{W_k} \sigma$. Так как $z = (0, 0) \in W_k$, то ясно, что $\sigma_k \leqslant 0$. Имеются две возможности:

1) $\sigma_k \leqslant -\varepsilon_k$. Тогда направление $e_k \neq 0$ является возможным направлением убывания функции $J(u)$ в точке u_k на множестве U. Полагаем

$$u_{k+1} = u_k + \alpha_k e_k, \quad 0 < \alpha_k \leqslant \beta_k; \quad \varepsilon_{k+1} = \varepsilon_k, \tag{43}$$

где

$$\beta_k = \sup\{\alpha: u_k + t e_k \in U, \ 0 \leqslant t \leqslant \alpha\} > 0.$$

2) $-\varepsilon_k < \sigma_k \leqslant 0$. Тогда полагаем

$$u_{k+1} = u_k, \quad \varepsilon_{k+1} = \theta \varepsilon_k, \quad 0 < \theta < 1,$$

где θ — параметр метода, и снова переходим к рассмотрению задачи (42) с заменой множества I_k на множество $I_{k+1} = \{i: 1 \leqslant i \leqslant m, -\varepsilon_{k+1} \leqslant g_i(u_k) \leqslant 0\}$.

Различные способы выбора величины α_k в (43) описаны в § 5.5.

Заметим, что задача (42) далеко не всегда просто решается. Поэтому методом возможных направлений пользуются для решения лишь таких задач (41), в которых решение вспомогательных задач (42) может быть легко найдено.

Предлагаем читателю самостоятельно исследовать сходимость этого метода и рассмотреть, в частности, возможность обобщения теоремы 5.5.2 на случай банаховых пространств.

5. Проксимальный метод применяют для решения задач минимизации

$$J(u) \to \inf, \quad u \in U, \tag{44}$$

где U — выпуклое замкнутое множество из гильбертова пространства H, функция $J(u)$ выпукла и полунепрерывна снизу на U. Этот метод заключается в построении последовательности $\{u_k\}$ по следующему правилу. Начальная точка $u_0 \in H$ предполагается заданной. Если уже известна точка u_k, $k \geqslant 0$, то следующее приближение u_{k+1} определяется как решение задачи:

$$\varphi(z, u_k) = \tfrac{1}{2}\|z - u_k\|^2 + \alpha J(z) \to \inf, \quad z \in U,$$

где $\alpha > 0$ — параметр метода. Так как функция $\varphi(z, u_k)$ переменной z полунепрерывна снизу и сильно выпукла на U, множество U — выпукло и замкнуто, то условия

$$u_{k+1} \in U, \quad \inf_{z \in U} \varphi(z, u_k) = \varphi(u_{k+1}, u_k) \tag{45}$$

однозначно определяют точку u_{k+1} (теорема 2.10).

Для дифференцируемых функций $J(u)$ описанный метод превращается в метод проекции градиента в неявной форме:

$$u_{k+1} = \mathcal{P}_U(u_k - \alpha J'(u_{k+1})), \quad k = 0, 1, \ldots$$

(см. метод (5.6.14)).

В задаче (1) на каждом шаге проксимального метода нужно решить задачу минимизации:

$$\varphi(z, u_k) = \frac{1}{2}\|z - u_k\|^2 + \alpha\|Az - b\|^2 \to \inf, \quad z \in U.$$

Если $U = H$, то эта задача равносильна задаче определения решения линейного операторного уравнения

$$\varphi_z(z, u_k) = z - u_k + 2\alpha A^*(Az - b) = 0$$

или

$$(2\alpha A^* A + I)z = 2\alpha A^* b + u_k.$$

Предлагаем читателю самостоятельно расписать проксимальный метод для задачи (3)–(5). Аналогично теореме 5.6.5 доказывается

Теорема 5. *Пусть U — выпуклое замкнутое множество из гильбертова пространства H, функция $J(u)$ полунепрерывна снизу и сильно выпукла на U с постоянной сильной выпуклости $\varkappa$. Пусть последовательность $\{u_k\}$ определена методом* (45) *с произвольным $u_0 \in H$. Тогда $\{u_k\}$ сильно в H сходится к точке u_* — решению задачи* (44), *причем*

$$\|u_k - u_*\|^2 \leqslant \|u_0 - u_*\|^2 q^k, \quad k = 0, 1, \ldots, \quad q = \frac{1}{1 + 2\alpha\varkappa}.$$

6. Метод сопряженных направлений применяют для поиска решения задачи

$$J(u) \to \inf; \qquad u \in H,$$

где H — гильбертово пространство, $J(u) \in C^1(H)$. Этот метод заключается в построении последовательности $\{u_k\}$ по правилу

$$u_{k+1} = u_k - \alpha_k p_k, \quad k = 0, 1, \ldots, \tag{46}$$

где

$$p_0 = J'(u_0), \quad p_k = J'(u_k) - \beta_k p_{k-1}, \quad k = 1, 2, \ldots, \tag{47}$$

величина β_k определяется по одной из формул

$$\beta_k = \langle J'(u_k), J'(u_{k-1}) - J'(u_k)\rangle \|J'(u_{k-1})\|^{-2},$$

или

$$\beta_k = -\|J'(u_k)\|^2 \|J'(u_{k-1})\|^{-2},$$

а α_k в (46) находят из условия

$$f_k(\alpha_k) = \inf_{\alpha \geqslant 0} f_k(\alpha), \quad f_k(\alpha) = J(u_k - \alpha p_k), \quad \alpha \geqslant 0. \tag{48}$$

Как показывает практика, погрешности, неизбежно появляющиеся при определении α_k из условия (48), могут привести к тому, что векторы $\{p_k\}$ из (47) перестают указывать направление убывания функции и сходимость метода может нарушиться. Чтобы бороться с этим явлением, метод сопряженных градиентов время от времени обновляют, полагая в (47) $\beta_k = 0$.

В отличие от конечномерных пространств (см. § 5.6), в гильбертовых пространствах нельзя ожидать, что точку минимума сильно выпуклой квадратичной функции

$$J(u) = \frac{1}{2}\langle Au, u\rangle - \langle b, u\rangle, \quad u \in H, \quad A \in \mathcal{L}(H \to H), \quad b \in H,$$

удастся найти за конечное число шагов метода сопряженных градиентов.

Предлагаем читателю самостоятельно сформулировать и доказать аналог теоремы 5.6.1 для гильбертовых пространств и рассмотреть возможность применения метода сопряженных градиентов к задачам (1), (3)–(5).

7. Метод Ньютона может быть использован для поиска решения задачи

$$J(u) \to \inf; \qquad u \in U,$$

где U — выпуклое замкнутое множество из банахова пространства B, $J(u) \in C^2(U)$. Этот метод заключается в построении последовательности $\{u_k\}$ по следующему правилу: по известному k-му приближению $u_k \in U$ находят вспомогательное приближение $\overline{u}_k \in U$ из условия

$$J_k(\overline{u}_k) = \inf_U J_k(u), \tag{49}$$

где

$$J_k(u) = \langle J'(u_k), u - u_k\rangle + \frac{1}{2}\langle J''(u_k)(u - u_k), u - u_k\rangle, \quad u \in U,$$

и затем полагают

$$u_{k+1} = u_k + \alpha_k(\overline{u}_k - u_k), \qquad 0 \leqslant \alpha_k \leqslant 1. \tag{50}$$

В частности, если $U = B$, то в точке минимума функции $J_k(u)$ ее производная $J_k'(u)$ обращается в нуль, т. е.

$$J_k'(\overline{u}_k) = J'(u_k) + J''(u_k)(\overline{u}_k - u_k) = 0.$$

Если оператор $J''(u_k)$ имеет обратный оператор $(J''(u_k))^{-1}$, то отсюда имеем

$$\overline{u}_k = u_k - (J''(u_k))^{-1} J'(u_k).$$

Подставляя это выражение для $\overline{u}_k$ в (50), получим

$$u_{k+1} = u_k - \alpha_k (J''(u_k))^{-1} J'(u_k), \quad k = 0, 1, \ldots$$

Таким образом, при $U = B$ метод (49), (50) представляет собой широко известный метод Ньютона для решения операторных уравнений (в данном случае уравнения $J'(u) = 0$).

Различные способы выбора величины α_k в (50) описаны в § 5.10.

Предлагаем читателю самостоятельно сформулировать и доказать аналоги теорем 5.10.1–5.10.3 о сходимости метода Ньютона для задач минимизации в банаховых пространствах.

8. Метод штрафных функций может быть использован для поиска решения задачи

$$J(u) \to \inf; \qquad u \in U, \tag{51}$$

$$U = \{u \in U_0: g_i(u) \leqslant 0, \ i = 1, \ldots, m; \ g_i(u) = 0, \ i = m+1, \ldots, s\}, \tag{52}$$

где U_0 — заданное множество из банахова пространства B, функции $J(u), g_1(u), \ldots, g_s(u)$ определены на U_0.

Определение 5.15.1 штрафной функции без изменений сохраняется и в банаховых пространствах. Примером штрафной функции множества (52) на множестве U_0 является функция

$$P_k(u) = A_k P(u), \quad u \in U_0, \tag{53}$$

где

$$P(u) = \sum_{i=1}^{m} (\max\{g_i(u); 0\})^p + \sum_{i=m+1}^{s} |g_i(u)|^p, \quad u \in U_0, \tag{54}$$

числа A_k, называемые *штрафными коэффициентами*, таковы, что

$$A_k > 0, \quad k = 1, 2, \ldots; \quad \lim_{k \to \infty} A_k = +\infty, \quad p \geqslant 1;$$

другие примеры штрафных функций см. в § 5.15.

Метод штрафных функций для задачи (51), (52) заключается в том, что вводят функции

$$\Phi_k(u) = J(u) + P_k(u), \quad u \in U_0, \quad k = 1, 2, \ldots, \tag{55}$$

и определяют последовательность $\{u_k\}$ условиями

$$u_k \in U_0, \quad \Phi_k(u_k) \leqslant \Phi_{k*} + \varepsilon_k, \tag{56}$$

где $P_k(u)$ — штрафная функция множества U (например, функция (53), (54)), $\Phi_{k*} = \inf_{U_0} \Phi_k(u)$, $\varepsilon_k > 0$, $k = 1, 2, \ldots$, $\lim_{k \to \infty} \varepsilon_k = 0$. Если существует точка $u_k \in U_0$, для которой $\Phi_k(u_k) = \Phi_{k*}$, то в (56) допускается возможность $\varepsilon_k = 0$. Заметим, что точка u_k, удовлетворяющая условию (56), вообще говоря, не принадлежит множеству U. При описании метода штрафной функции (55), (56) подразумевается, что $\Phi_{k*} > -\infty$ при всех $k = 1, 2, \ldots$

Приведем две теоремы о сходимости метода штрафных функций.

Теорема 6. *Пусть функции $J(u)$, $g_i(u)$, $i = 1, \ldots, s$, определены на множестве U_0, а последовательность $\{u_k\}$ определена из условий* (53)–(56). *Тогда*

$$\overline{\lim_{k \to \infty}} J(u_k) \leqslant \overline{\lim_{k \to \infty}} \Phi_k(u_k) = \overline{\lim_{k \to \infty}} \Phi_{k*} \leqslant J_*. \tag{57}$$

Если, кроме того, $J_{**} = \inf_{U_0} J(u) > -\infty$, то

$$P_k(u) = O(A_k^{-1}), \qquad k = 1, 2, \ldots, \tag{58}$$

$$\overline{\lim_{k \to \infty}} g_i(u_k) \leqslant 0, \quad i = 1, \ldots, m; \quad \lim_{k \to \infty} g_i(u_k) = 0, \quad i = m+1, \ldots, s. \tag{59}$$

Доказательство этой теоремы проводится дословно так же, как и теоремы 5.15.1. В § 5.15 приведен пример, показывающий, что в (57) неравенства могут быть строгими.

Теорема 7. *Пусть U_0 — выпуклое замкнутое множество из рефлексивного банахова пространства B, функции $J(u)$, $g_1(u), \ldots, g_m(u)$, $|g_{m+1}(u)|, \ldots, |g_s(u)|$ слабо полунепрерывны снизу на U_0 (например, эти функции выпуклы и полунепрерывны снизу в метрике B), $J_{**} = \inf_{U_0} J(u) > -\infty$; множество*

$$U(\delta) = \{u \in U_0 \colon g_i(u) \leqslant \delta, \ i = 1, \ldots, m; \quad |g_i(u)| \leqslant \delta, \ i = m+1, \ldots, s\}$$

выпукло и ограничено при некотором $\delta > 0$. Тогда последовательность $\{u_k\}$, определенная условиями (53)–(56), *такова, что любая точка v_*, являющаяся слабым пределом какой-либо ее подпоследовательности, принадлежит U_* и $\lim_{k \to \infty} J(u_k) = J_*$.*

Доказательство. При сделанных предположениях для $\{u_k\}$ соотношения (57)–(59) сохраняют силу. Из (59) следует, что $u_k \in U(\delta)$ при всех $k \geqslant k_0$. Однако согласно теореме 2.6 множество $U(\delta)$ слабо компактно. Тогда существует хотя бы одна точка $v_* \in U(\delta)$, к которой слабо сходится

некоторая подпоследовательность $\{u_{k_r}\}$. В силу слабой полунепрерывности снизу указанных в теореме функций и соотношений (59) имеем

$$g_i(v_*) \leqslant \lim_{r \to \infty} g_i(u_{k_r}) \leqslant \overline{\lim_{k \to \infty}} g_i(u_k) \leqslant 0, \quad i = 1, \ldots, m;$$

$$|g_i(v_*)| \leqslant \lim_{r \to \infty} |g_i(u_{k_r})| \leqslant \lim_{k \to \infty} |g_i(u_k)| = 0, \quad i = m+1, \ldots, s.$$

Это значит, что $v_* \in U$. Тогда с учетом (57) получаем

$$J_* \leqslant J(v_*) \leqslant \lim_{r \to \infty} J(u_{k_r}) \leqslant \overline{\lim_{k \to \infty}} J(u_k) \leqslant J_*,$$

так что

$$J(v_*) = \lim_{r \to \infty} J(u_{k_r}) = J_* \quad \text{или} \quad v_* \in U_*.$$

Таким образом, показано, что любая точка v_*, являющаяся слабым пределом какой-либо подпоследовательности $\{u_{k_r}\}$, принадлежит U_* и $\lim_{r \to \infty} J(u_{k_r}) = J_*$. Отсюда следуют утверждения теоремы 7. □

Доказанная теорема является аналогом теоремы 5.15.2. Читателю предлагаем самостоятельно исследовать возможность обобщения других теорем из § 5.15 на случай банаховых пространств.

Для иллюстрации метода штрафных функций рассмотрим задачу (3)–(5) при дополнительных фазовых ограничениях вида

$$a_i \leqslant x^i(t, u) \leqslant b_i, \quad t_0 \leqslant t \leqslant T, \quad i = 1, \ldots, m; \quad m \leqslant n, \tag{60}$$

или

$$x^i(T, u) = x_I^i, \quad i = 1, \ldots, s; \quad s < n, \tag{61}$$

где a_i, b_i, x_I^i — заданные числа. Для учета ограничений (60) можно взять штрафную функцию

$$P_k(u) = A_k \sum_{i=1}^{m} \int_{t_0}^{T} [(\max\{x^i(t, u) - b_i; 0\})^2 + (\max\{a_i - x^i(t, u); 0\})^2] dt,$$

где $A_k > 0$, $\lim_{k \to \infty} A_k = +\infty$; ограничения (61) можно учесть с помощью штрафной функции

$$P_k(u) = A_k \sum_{i=1}^{s} (x^i(T, u) - x_I^i)^2, \quad k = 1, 2, \ldots$$

Тогда задача (3)–(5), (60) или (61) сведется к решению последовательности задач минимизации функции

$$\Phi_k(u) = |x(T, u) - b|^2 + P_k(u), \quad k = 1, 2, \ldots, \tag{62}$$

при условиях (2), (3). Вопрос о том, как определить градиент функции (62) при условиях (4), (5) и как минимизировать функцию (62), будет обсужден в следующем параграфе.

Предлагаем читателю самостоятельно исследовать возможность обобщения других методов главы 5 на случай задач минимизации на множествах из гильбертовых или банаховых пространств, исследовать сходимость этих методов [65; 103; 112; 151; 638; 801].

Упражнения

1. Пусть $U=\{u(t)=(u^1(t),\ldots,u^r(t))\in L_2^r[t_0,T]\colon u^i(t)\geqslant 0$ почти всюду на отрезке $t_0\leqslant t\leqslant T,\ i=1,\ldots,r\}$. Найти проекцию любой точки из $L_2^r[t_0,T]$ на множество U. Описать метод проекции градиента для задачи (3)–(5) с этим множеством.

2. Описать метод сопряженных градиентов для задачи (3)–(5) при $U=L_2^r[t_0,T]$.

3. Описать градиентный метод, методы проекции градиента, условного градиента, сопряженных направлений для задач минимизации функций

$$J(u)=|x(T,u)-b|^2+\alpha\int_{t_0}^{T}|u(t)|^2dt,\quad \alpha=\text{const}>0,$$

$$J(u)=\int_{t_0}^{T}|x(t,u)-b(t)|^2dt,\quad b(t)\in L_2^n[t_0,T],$$

$$J(u)=\int_{t_0}^{T}|x(t,u)-b(t)|^2dt+\alpha\int_{t_0}^{T}|u(t)|^2dt,\quad \alpha=\text{const}>0$$

при условиях (4), (5). Исследовать сходимость методов.

4. Описать метод возможных направлений для задачи (41) в предположении, что $U=H$ — гильбертово пространство, взяв во вспомогательной задаче (42) поиска возможного направления убывания условие нормировки $\|e\|=1$.

5. Описать градиентный метод, методы проекции градиента, условного градиента, сопряженных направлений для задач минимизации функций из примера 3.4. Рассмотреть случаи, когда $U=H$, или U — шар или гиперплоскость в H.

6. Сформулировать и доказать теоремы сходимости методов барьерных, нагруженных функций из §§ 5.17–5.18 для множества (52) из гильбертова пространства.

§ 5. Градиент в задаче оптимального управления со свободным правым концом

Рассмотрим следующую задачу оптимального управления: минимизировать функцию

$$J(u)=\int_{t_0}^{T}f^0(x(t),u(t),t)+\Phi(x(T)) \tag{1}$$

при условиях

$$\dot{x}(t)=f(x(t),u(t),t),\quad t_0\leqslant t\leqslant T;\quad x(t_0)=x_0, \tag{2}$$

$$u=u(t)\in U\subseteq L_2^r[t_0,T], \tag{3}$$

где $x=(x^1,x^2,\ldots,x^n)$, $u=(u^1,\ldots,u^r)$, функции $f^0(x,u,t)$, $f(x,u,t)=(f^1(x,u,t),\ldots,f^n(x,u,t))$, $\Phi(x)$ переменных $(x,u,t)\in E^n\times E^r\times[t_0,T]$ считаются известными, U — заданное множество из $L_2^r[t_0,T]$, моменты времени t_0, T и начальная точка x_0 заданы.

Определение решения (или траектории) $x=x(t)=x(t,u)$, $t_0\leqslant t\leqslant T$, задачи Коши (2), соответствующего управлению $u=u(t)\in L_2^r[t_0,T]$, а также условия его существования обсуждались в гл. 6 (см. теорему 6.1.1); там же был доказан принцип максимума для задачи (1)–(3) (см. теорему 6.2.1, а также равенства (6.2.25), (6.2.26), § 6.3).

1. Ниже будут сформулированы достаточные условия дифференцируемости функции (1) на $L_2^r[t_0,T]$, и получена формула для ее градиента. Примем обозначения

$$f_x=\frac{\partial f}{\partial x}=\begin{pmatrix}f^1_{x^1},&\ldots,&f^1_{x^n}\\ f^n_{x^1},&\ldots,&f^n_{x^n}\end{pmatrix}=(f^1_x,\ldots,f^n_x)^T,$$

$$f_u=\frac{\partial f}{\partial u}=\begin{pmatrix}f^1_{u^1},&\ldots,&f^1_{u^r}\\ f^n_{u^1},&\ldots,&f^n_{u^r}\end{pmatrix}=(f^1_u,\ldots,f^n_u)^T,$$

$$f^i_x=\begin{pmatrix}f^i_{x^1}\\ \vdots\\ f^i_{x^n}\end{pmatrix},\quad f^i_u=\begin{pmatrix}f^i_{u^1}\\ \vdots\\ f^i_{u^r}\end{pmatrix},\quad i=0,\ldots,n;\quad \Phi_x=\begin{pmatrix}\Phi_{x^1}\\ \vdots\\ \Phi_{x^n}\end{pmatrix}.$$

Здесь $f^i_{x^j}=\dfrac{\partial f^i}{\partial x^j}$ — частная производная функции f^i по переменной x^j, T — знак транспонирования матрицы. Напомним, что *нормой матрицы* $A=\{a_{ij},\ i=1,\ldots,n;\ j=1,\ldots,m\}$ порядка $n\times m$ называется число $\|A\|=\sup|Az|_{E^n}$, где верхняя грань берется по единичному шару $|z|_{E^m}\leqslant 1$. Справедливо неравенство $|Az|_{E^n}\leqslant\|A\||z|_{E^m}$ при всех $z\in E^m$. Введем знакомую нам по гл. 6 функцию Гамильтона — Понтрягина

$$H(x,u,t,\psi)=-f^0(x,u,t)+\langle f(x,u,t),\psi\rangle,\quad (\psi)^T=(\psi_1,\psi_2,\ldots,\psi_n). \tag{4}$$

Обозначим

$$H_x=\begin{pmatrix}H_{x^1}\\ \vdots\\ H_{x^n}\end{pmatrix},\quad H_u=\begin{pmatrix}H_{u^1}\\ \vdots\\ H_{u^r}\end{pmatrix}.$$

Теорема 1. *Пусть функции f^0, f, Φ непрерывны по совокупности своих аргументов вместе со своими частными производными по переменным x, u при $(x,u,t)\in E^n\times E^r\times[t_0,T]$ и, кроме того, выполнены следующие условия:*

$$|f(x+\Delta x,u+h,t)-f(x,u,t)|\leqslant L(|\Delta x|+|h|), \tag{5}$$

$$\|f_x(x+\Delta x,u+h,t)-f_x(x,u,t)\|\leqslant L(|\Delta x|+|h|), \tag{6}$$

$$|f^0_x(x+\Delta x,u+h,t)-f^0_x(x,u,t)|\leqslant L(|\Delta x|+|h|), \tag{7}$$

$$\|f_u(x+\Delta x,u+h,t)-f_u(x,u,t)\|\leqslant L(|\Delta x|+|h|), \tag{8}$$

$$|f^0_u(x+\Delta x,u+h,t)-f^0_u(x,u,t)|\leqslant L(|\Delta x|+|h|), \tag{9}$$

$$|\Phi_x(x+\Delta x)-\Phi_x(x)|\leqslant L|\Delta x| \tag{10}$$

при всех $(x+\Delta x,u+h,t)$, $(x,u,t)\in E^n\times E^r\times[t_0,T]$, где $L=\text{const}\geqslant 0$.

Тогда функция (1) при условиях (2) непрерывна и дифференцируема по $u=u(t)$ в норме $L_2^r[t_0,T]$ всюду на $L_2^r[t_0,T]$, причем ее градиент $J'(u)=J'(u,t)\in L_2^r[t_0,T]$ в точке $u=u(t)$ представим в виде

$$J'(u)=-H_u(x,u,t,\psi)\big|_{x=x(t,u),\,u=u(t),\,\psi=\psi(t,u)}=$$

$$=f^0_u(x(t,u),u(t),t)-(f_u(x(t,u),u(t),t))^T\psi(t,u),\quad t_0\leqslant t\leqslant T, \tag{11}$$

где $x(t)=x(t,u)$, $t_0\leqslant t\leqslant T$, — решение задачи (2), соответствующее управлению $u=u(t)$, а $\psi(t)=\psi(t,u)$, $t_0\leqslant t\leqslant T$, является решением сопряженной системы

$$\dot{\psi}(t)=-H_x(x,u,t,\psi(t))\big|_{x=x(t,u),\,u=u(t)}=$$

$$=f^0_x(x(t,u),u(t),t)-(f_x(x(t,u),u(t),t))^T\psi(t),\quad t_0\leqslant t\leqslant T, \tag{12}$$

при начальных условиях

$$\psi(T)=-\Phi_x(x)\big|_{x=x(T,u)}. \tag{13}$$

Доказательство. Пусть $u=u(t)$, $u+h=u(t)+h(t)\in L_2^r[t_0, T]$, а $x(t, u)$, $x(t, u+h)$, $t_0\leqslant t\leqslant T$, — соответствующие этим управлениям решения задачи (2). Из условия теоремы имеем

$$|f(x, u, t)|\leqslant|f(x, u, t)-f(0, 0, t)|+|f(0, 0, t)|\leqslant$$
$$\leqslant L(|x|+|u|)+\sup_{t_0\leqslant t\leqslant T}|f(0, 0, t)|, \quad (x, u, t)\in E^n\times E^r\times[t_0, T]. \quad (14)$$

Из (14) и из теоремы 6.1.1 и замечания 2 к ней следует существование и единственность решения задачи (2) при всех $u=u(t)\in L_2^r[t_0, T]$. Приращение $\Delta x(t)=x(t, u+h)-x(t, u)$, $t_0\leqslant t\leqslant T$ удовлетворяет условиям:

$$\Delta\dot{x}(t)=f(x(t)+\Delta x(t), u(t)+h(t), t)-f(x(t), u(t), t),$$
$$t_0\leqslant t\leqslant T, \quad \Delta x(t_0)=0. \quad (15)$$

Из (15) и из теоремы 6.3.1 следует оценка

$$|\Delta x(t)|\leqslant C_1\int_{t_0}^{T}|h(t)|dt\leqslant C_1(T-t_0)^{1/2}\|h\|_{L_2}, \quad t_0\leqslant t\leqslant T; \quad (16)$$

здесь и ниже через $C_1, C_2, \ldots$ будем обозначать константы, не зависящие от выбора $u=u(t)\in L_2^r[t_0, T]$.

Покажем, что приращение функции (1) представимо в виде:

$$\Delta J(u)=J(u+h)-J(u)=-\int_{t_0}^{T}H_u(x(t), u(t), t, \psi(t))dt+R, \quad (17)$$

где $R=R_1+R_2+R_3$,

$$R_1=\langle\Phi_x(x(T)+\theta_1\Delta x(T))-\Phi_x(x(T)), \Delta x(T)\rangle, \quad 0<\theta_1<1,$$
$$R_2=-\int_{t_0}^{T}\langle H_x(x+\theta_2\Delta x, u+\theta_2 h, t, \psi)-H_x(x, u, t, \psi), \Delta x\rangle dt,$$
$$R_3=-\int_{t_0}^{T}\langle H_u(x+\theta_2\Delta x, u+\theta_2 h, t, \psi)-H_u(x, u, t, \psi), h(t)\rangle dt, \quad 0<\theta_2<1.$$

Так как $\Phi(x+\Delta x)-\Phi(x)=\langle\Phi_x(x+\theta_1\Delta x), 0\leqslant\theta_1<1$, то

$$\Delta J(u)=\int_{t_0}^{T}[f^0(x(t)+\Delta x(t), u(t)+h(t), t)-f^0(x(t), u(t), t)]dt+$$
$$+\langle\Phi_x(x(T)), \Delta x(T)\rangle+R_1, \quad (18)$$

Преобразуем второе слагаемое правой части (18). С учетом соотношений (12), (13), (15) имеем

$$\langle\Phi_x(x(T)), \Delta x(T)\rangle=-\langle\psi(T), \Delta x(T)\rangle=-\int_{t_0}^{T}\frac{d}{dt}\langle\psi(t), \Delta x(t)\rangle dt-$$
$$-\langle\psi(t_0), \Delta x(t_0)\rangle=-\int_{t_0}^{T}[\langle\psi(t), \Delta\dot{x}(t)\rangle+\langle\dot{\psi}(t), \Delta x(t)\rangle]dt=$$
$$=-\int_{t_0}^{T}\langle\psi(t), f(x(t)+\Delta x(t), u(t)+h(t), t)-$$
$$-f(x(t), u(t), t)\rangle dt+\int_{t_0}^{T}\langle H_x(x(t), u(t), t, \psi(t)), \Delta x(t)\rangle dt.$$

Подставим полученное выражение в (18), откуда с помощью функции (4) будем иметь

$$\Delta J(u)=-\int_{t_0}^{T}[H(x(t)+\Delta x(t), u(t)+h(t), t, \psi(t))-$$
$$-H(x(t), u(t), t, \psi(t))]dt+\int_{t_0}^{T}\langle H_x(x(t), u(t), t, \psi(t)), \Delta x(t)\rangle dt+R_1.$$

Отсюда, пользуясь формулой конечных приращений

$$H(x+\Delta x, u+h, t, \psi)-H(x, u, t, \psi)=\langle H_x(x+\theta_2\Delta x, u+\theta_2 h, t, \psi), \Delta x\rangle+$$
$$+\langle H_u(x+\theta_2\Delta x, u+\theta_2 h, t, \psi), h\rangle, \quad 0<\theta_2<1,$$

приходим к требуемому представлению (17) приращения функции (1). Так как $H_x(x, u, t, \psi)=-f_x^0(x, u, t)+(f_x(x, u, t))^{\top}\psi$, $H_u(x, u, t, \psi)=-f_u^0(x, u, t)+(f_u(x, u, t))^{\top}\psi$, то с учетом условий (6)–(10) и оценки (16) для величин R_1, R_2, R_3 из (17) имеем

$$|R_1|\leqslant L|\Delta x(T)|^2\leqslant C_2\|h\|_{L_2^r}^2,$$
$$|R_2|\leqslant(1+\sup_{t_0\leqslant t\leqslant T}|\psi(t, u)|)L\int_{t_0}^{T}(|\Delta x(t)|^2+|\Delta x(t)||h(t)|)dt\leqslant C_3\|h\|_{L_2^r}^2,$$
$$|R_3|\leqslant\Big(1+\sup_{t_0\leqslant t\leqslant T}|\psi(t, u)|\Big)L\int_{t_0}^{T}(|\Delta x(t)||h(t)|+|h(t)|^2)dt\leqslant C_4\|h\|_{L_2}^2.$$

Суммируя оценки для R_1, R_2, R_3, получим

$$|R|\leqslant C_5\int_{t_0}^{T}|h(t)|^2dt=C_5\|h\|_{L_2}^2. \quad (19)$$

Из формулы (17) и оценки (19) следует дифференцируемость функции (1) и формула (11) для градиента. Теорема 1 доказана. □

Таким образом, для вычисления градиента функции (1) при условиях (2) нужно последовательно решить две задачи Коши: сначала из задачи (2) нужно определить $x(t, u)$, затем из (12), (13) — $\psi(t, u)$ и, наконец, по формуле (11) найти искомый градиент. При решении упомянутых задач Коши (2) и (12), (13) можно использовать различные приближенные методы [74; 89; 482; 630; 634].

Заметим, что дифференцируемость функции (1) в теореме 1 доказана при довольно жестких ограничениях на исходные данные задачи (1)–(3). На самом деле, формулу (17) с остаточным членом R, $R/\|h\|_{L_2}\to 0$, при $\|h\|_{L_2}\to 0$, можно получить при несколько меньших требованиях. Например, вместо условия (10) можно требовать $\Phi(x)\in C^1(E^n)$, а вместо условий (6)–(9) для производных f_x^i, f_u^i, $i=0, \ldots, n$, можно ограничиться условиями типа

$$|f_x^i(x+\Delta x, u+h, t)-f_x^i(x, u, t)|\leqslant L|h|^{\gamma}+o(|\Delta x|),$$

где $0<\gamma\leqslant 1$, $o(\alpha)/\alpha\to 0$ при $\alpha\to 0$.

Заметим, что формулу градиента функции (1) иногда записывают в несколько ином виде: вместо функции (4) берут

$$H(x, u, t, \psi)=f^0(x, u, t)+\langle f(x, u, t), \psi\rangle, \quad (20)$$

сопряженную задачу (12), (13) заменяют на задачу

$$\dot{\psi}(t) = -H_x(x, u, t, \psi(t))|_{x=x(t,u),\, u=u(t)}, \quad t_0 \leqslant t \leqslant T, \tag{21}$$

$$\psi(T) = \Phi_x(x)|_{x=x(T,u)}, \tag{22}$$

и тогда вместо формулы (11) будем иметь

$$J'(u) = H_u(x, u, t, \psi)|_{x=x(t,u),\, u=u(t),\, \psi=\psi(t,u)}, \quad t_0 \leqslant t \leqslant T. \tag{23}$$

Нетрудно видеть, что функции H и $\psi(t, u)$ и, следовательно, производная H_u из (4), (11)–(13) и (20)–(23) отличаются друг от друга лишь знаком.

Пользуясь формулами (20)–(23), найдем градиент в задаче (4.3)–(4.5), являющейся частным случаем задачи (1)–(3) при $f^0=0$, $f(x, u, t)=A(t)x+B(t)u+f(t)$, $\varphi(x)=|x-b|^2$. Тогда $H(x, u, t, \psi)=\langle\psi, A(t)x+B(t)u+f(t)\rangle=\langle A^T(t)\psi, x\rangle+\langle B^T(t)\psi, u\rangle+\langle\psi, f(t)\rangle$; сопряженная задача имеет вид

$$\dot{\psi} = -H_x = -A^T(t)\psi(t), \quad t_0 \leqslant t \leqslant T; \quad \psi(T) = 2(x(T, u) - b),$$

градиент равен $J'(u)=B^T(t)\psi(t,u)$, $t_0 \leqslant t \leqslant T$ (ср. с формулами (4.6), (4.7)).

Для иллюстрации теоремы 1 приведем пример задачи оптимального управления для нелинейной системы.

Пример 1. Рассмотрим задачу об оптимальном успокоении математического маятника (см. примеры 6.1.1 и 6.2.6):

$$J(u) = (x^1(T))^2 + (x^2(T))^2 = |x(T)|^2 \to \inf; \tag{24}$$

$$\dot{x}^1(t) = x^2(t), \quad \dot{x}^2(t) = -\sin x^1(t) - \beta x^2(t) + u(t), \quad 0 \leqslant t \leqslant T; \tag{25}$$

$$x(0) = (x^1(0), x^2(0)) = x_0 = (x_0^1, x_0^2); \tag{26}$$

$$u = u(t) \in U \subseteq L_2^r[t_0, T]; \tag{27}$$

здесь момент $T>0$, постоянная β, начальная точка x_0 считаются известными. Задача (24)–(27) является частным случаем задачи (1)–(3), когда $f^0=0$, $\Phi(x)=\Phi(x^1, x^2)=(x^1)^2+(x^2)^2$, $f^1(x, u, t)=x^2$, $f^2(x, u, t)=-\sin x^1-\beta x^2+u$, $t_0=0$, $n=2$, $r=1$. Все условия теоремы 1 для задачи (24)–(27) выполнены. Для вычисления градиента функции (24) составим функцию Гамильтона — Понтрягина

$$H(x, u, \psi) = \psi_1 x^2 + \psi_2(-\sin x^1 - \beta x^2 + u).$$

Поскольку $H_{x^1}=-\psi_2\cos x^1$, $H_{x^2}=\psi_1-\beta\psi_2$, $\Phi_{x^1}=2x^1$, $\Phi_{x^2}=2x^2$, то сопряженная задача (21), (22) запишется в виде

$$\dot{\psi}_1 = \psi_2 \cos x^1(t, u), \quad \dot{\psi}_2 = -\psi_1 + \beta\psi_2, \quad 0 \leqslant t \leqslant T,$$

$$\psi_1(T) = 2x^1(T, u), \quad \psi_2(T) = 2x^2(T, u).$$

Так как $H_u=\psi_2$, то согласно формуле (23) градиент равен

$$J'(u) = \psi_2(t, u), \quad 0 \leqslant t \leqslant T.$$

2. Имея формулу градиента, нетрудно расписать градиентный метод, методы проекции градиента, условного градиента — это делается так же, как было сделано в § 4 для задачи (4.3)–(4.5). Заметим, что в задаче (1)–(3) в общем случае, конечно, нельзя ожидать, что функция $f(\alpha)=J(u+\alpha h)$ переменной α будет квадратным трехчленом, значит параметр α_k из условий типа (4.9), (4.32) будет определяться не так просто, как в задаче (4.3)–(4.5).

Во многих теоремах о сходимости методов минимизации требуется, чтобы минимизируемая функция принадлежала классу $C^{1,1}(U)$. Приведем достаточные условия принадлежности $C^{1,1}(U)$ для функции (1) при условиях (2).

Теорема 2. *Пусть выполнены все условия теоремы* 1 *и* $U=\{u=u(t)\in L_2^r[t_0, T]$: $u(t)\in V(t)$ *почти всюду на* $[t_0, T]\}$, *где* $V(t)$ *— заданные множества из* E^r, *причем*

$$\sup_{t_0 \leqslant t \leqslant T} \sup_{u \in V(t)} |u| \leqslant R < \infty.$$

Тогда

$$\|J'(u) - J'(v)\|_{L_2} \leqslant L_1\|u - v\|_{L_2}, \quad L_1 = \text{const} \geqslant 0, \tag{28}$$

при любых $u, v \in U$.

Доказательство. Возьмем произвольные $u=u(t)$, $v=v(t)\in U$. Из оценки (16) для $\Delta x(t)=x(t, u)-x(t, v)$, $t_0 \leqslant t \leqslant T$, следует

$$\|\Delta x(t)\|_C = \max_{t_0 \leqslant t \leqslant T} |\Delta x(t)| \leqslant C_6\|u - v\|_{L_2}. \tag{29}$$

Далее, с учетом неравенства (14) имеем

$$|x(t, u)| = \left|x_0 + \int_{t_0}^{t} f(x(\tau, u), u(\tau), \tau)d\tau\right| \leqslant L\int_{t_0}^{t}|x(\tau, u)|d\tau + L\int_{t_0}^{T}|u(\tau)|d\tau +$$

$$+ |x_0| + \sup_{t_0 \leqslant t \leqslant T}|f(0, 0, t)|(T - t_0) \leqslant L\int_{t_0}^{t}|x(\tau, u)|d\tau + L(T - t_0)R + C_7.$$

Отсюда с помощью леммы 6.3.1 получим

$$\sup_{u \in U}\|x(t, u)\|_C \leqslant e^{L(T-t_0)}(C_7 + L(T - t_0)R) = C_8. \tag{30}$$

Оценим $|\psi(t, u)|$. С этой целью заметим, что $(x(t, u), u(t), t) \in G = \{(x, u, t)\in E^n\times E^r\times[t_0, T]$: $|x|\leqslant C_8$, $|u|\leqslant R$, $t_0\leqslant t\leqslant T\}$ при всех $u\in U$. Так как функции Φ_x, f_x^0, f_x, f_u непрерывны по совокупности аргументов на замкнутом ограниченном множестве G, то

$$\max_G \max\{|\Phi_x|, |f_x^0|, \|f_x\|, \|f_u\|\} = C_9 < \infty. \tag{31}$$

Отсюда и из (12), (13) имеем

$$|\psi(t, u)| = \Bigg|\Phi_x(x(T, u)) + \int_t^T [f_x^0(x(\tau, u), u(\tau), \tau) -$$

$$- f_x(x(\tau, u), u(\tau), (\tau))^T\psi(\tau, u)]d\tau\Bigg| \leqslant$$

$$\leqslant C_9 + C_9(T - t_0) + C_9\int_t^T|\psi(\tau, u)|d\tau, \quad t_0 \leqslant t \leqslant T.$$

Тогда из леммы 6.3.1 следует

$$\sup_{u \in U}\|\psi(t, u)\|_C \leqslant C_9(1 + T - t_0)e^{C_9(T-t_0)} = C_{10}. \tag{32}$$

Далее, оценим $\Delta\psi(t)=\psi(t, u)-\psi(t, v)$, $t_0\leqslant t\leqslant T$. Из (12), (13), оценок (29)–(32) и условий (6), (7), (10) имеем

$$|\Delta\psi(t)| \leqslant |\Phi_x(x(T, u)) - \Phi_x(x(T, v))| +$$

$$+\int_t^T |f_x^0(x(\tau,u),u(\tau),\tau)-f_x^0(x(\tau,v),v(\tau),\tau)|d\tau+$$
$$+\int_t^T(|\psi(\tau,u)-\psi(\tau,v)|\|f_x(x(\tau,u),u(\tau),\tau)\|+$$
$$+|\psi(\tau,v)|\|f_x(x(\tau,u),u(\tau),\tau)-f_x(x(\tau,v),v(\tau),\tau)\|)d\tau\leqslant$$
$$\leqslant L|\Delta x(T)|+L\int_{t_0}^T(|\Delta x(t)|+|u(t)-v(t)|)dt+$$
$$+C_9\int_t^T|\psi(\tau,u)-\psi(\tau,v)|d\tau+C_{10}L\int_{t_0}^T(|\Delta x(t)|+|u(t)-v(t)|)dt\leqslant$$
$$\leqslant C_9\int_t^T|\psi(\tau,u)-\psi(\tau,v)|d\tau+C_{11}\|u-v\|_{L_2},\quad t_0\leqslant t\leqslant T.$$

Отсюда и из леммы 6.3.1 следует

$$\|\psi(t,u)-\psi(t,v)\|_C\leqslant C_{11}E^{C_9(T-t_0)}\|u-v\|_{L_2}=C_{12}\|u-v\|_{L_2}. \qquad (33)$$

Наконец, из формулы (11), оценок (29)–(33) и условий (8), (9) получим требуемое неравенство (28):

$$\|J'(u)-J'(v)\|_{L_2}=\Big(\int_{t_0}^T|H_u(x(t,u),u(t),t,\psi(t,u))-$$
$$-H_u(x(t,v),v(t),t,\psi(t,v))|^2dt\Big)^{1/2}\leqslant$$
$$\leqslant L(1+\|\psi(t,u)\|_C)\Big(\int_{t_0}^T(|\Delta x(t)|+|u(t)-v(t)|)^2dt\Big)^{1/2}+$$
$$+C_9\Big(\int_{t_0}^T|\Delta\psi(t)|^2dt\Big)^{1/2}\leqslant L_1\|u-v\|_{L_2},\quad u,v\in U.$$

Теорема 2 доказана. □

3. Отдельно остановимся на одном частном случае задачи (1)–(3), когда система (2) линейна по x, u, т. е.

$$\dot x(t)=A(t)x(t)+B(t)u(t)+f(t),\quad t_0\leqslant t\leqslant T,\quad x(t_0)=x_0, \qquad (34)$$

где $A(t)$, $B(t)$, $f(t)$ — заданные матрицы порядка $n\times n$, $n\times r$, $n\times 1$ соответственно. Для задачи (1)–(3), (34) принадлежность классу $C^{1,1}(U)$ может быть установлена при меньших требованиях, чем в теореме 2, и, кроме того, удается сформулировать условия, гарантирующие выпуклость и сильную выпуклость функции (1).

Теорема 3. *Пусть функции $f^0(x,u,t)$, $\Phi(x)$ удовлетворяют условиям теоремы 1 и матрицы $A(t)$, $B(t)$, $f(t)$ кусочно непрерывны на отрезке $[t_0,T]$. Тогда функция (1) при условиях (34) принадлежит классу $C^{1,1}$ на всем пространстве $L_2^r[t_0,T]$, причем ее градиент $J'(u)$ в точке $u=u(t)\in L_2^r[t_0,T]$ вычисляется по формуле*

$$J'(u)=f_u^0(x(t,u),u(t),t)-B^T(t)\psi(t,u),\quad t_0\leqslant t\leqslant T, \qquad (35)$$

где $x(t,u)$ — решение задачи (34), $\psi(t,u)$ — решение задачи

$$\dot\psi(t)=f_x^0(x(t,u),u(t),t)-A^T(t)\psi(t),\quad t_0\leqslant t\leqslant T;$$
$$\psi(T)=-\Phi_x(x(T,u)). \qquad (36)$$

Доказательство. Заметим, что здесь все условия (5)–(10) заведомо выполнены. Кроме того, для задачи (34) справедливы теорема 6.1.2, оценка (6.3.9) (см. также оценку (4.19)). Поэтому, рассуждая также, как при доказательстве теоремы 1, приходим к соотношениям (35), (36). Для $\Delta\psi(t)=\psi(t,u)-\psi(t,v)$ с учетом условий (7), (10), (36) будем иметь

$$|\Delta\psi(t)|\leqslant A_{\max}\int_t^T|\Delta\psi(\tau)|d\tau+L|\Delta x(T)|+L\int_{t_0}^T(|\Delta x(t)|+|u(t)-v(t)|)dt.$$

Отсюда и из леммы 6.3.1 следует оценка (33). Наконец, из формулы (35) и условия (9) с учетом оценок (16), (33) получим

$$\|J'(u)-J'(v)\|_{L_2}\leqslant\Big(\int_{t_0}^T|f_u^0(x(t,u),u(t),t)-f_u^0(x(t,v),v(t),t)|^2dt\Big)^{1/2}+$$
$$+B_{\max}\Big(\int_{t_0}^T|\psi(t,u)-\psi(t,v)|^2dt\Big)^{1/2}\leqslant L_1\|u-v\|_{L_2}.\quad\square$$

Укажем достаточные условия выпуклости и сильной выпуклости функции (1) при условиях (34), кратко обсудим условия оптимальности в задаче (1)–(3), (34).

Теорема 4. *Пусть матрицы $A(t)$, $B(t)$, $f(t)$ кусочно непрерывны на отрезке $[t_0,T]$, функция $\Phi(x)$ выпукла на E^n, а $f^0(x,u,t)$ выпукла по совокупности переменных (x,u), т. е.*

$$f^0(\alpha x+(1-\alpha)y,\alpha u+(1-\alpha)v,t)\leqslant\alpha f^0(x,u,t)+(1-\alpha)f^0(y,v,t) \qquad (37)$$

при всех $(x,u,t),(y,v,t)\in E^n\times E^r\times[t_0,T]$, $0\leqslant\alpha\leqslant 1$, и, кроме того, пусть $f^0(x(t),u(t),t)\in L_1[t_0,T]$ при каждой непрерывной функции $x(t)$, $t_0\leqslant t\leqslant T$, $u(t)\in L_2^r[t_0,T]$. Тогда функция (1) при условиях (34) будет определена и выпукла на $L_2^r[t_0,T]$.

Если при этом функции $f^0(x,u,t)$, $\Phi(x)$ удовлетворяют условиям теоремы 1, то функция (1) при условиях (34) достигает своей нижней грани на всяком выпуклом замкнутом ограниченном множестве $U\subset L_2^r[t_0,T]$, причем для оптимальности управления $u_*=u_*(t)\in U$ необходимо и достаточно выполнение неравенства

$$\int_{t_0}^T\langle f_u^0(x(t,u_*),u_*(t),t)-B^T(t)\psi(t,u_*),u(t)-u_*(t)\rangle dt\geqslant 0 \qquad (38)$$

при всех $u(t)\in U$. Если u_* — внутренняя точка множества U, то условие (38) равносильно условию

$$f_u^0(x(t,u_*),u_*(t),t)-B^T(t)\psi(t,u_*)=0,\quad t_0\leqslant t\leqslant T. \qquad (39)$$

Если же вместо (37) справедливо неравенство

$$f^0(\alpha x+(1-\alpha)y,\alpha u+(1-\alpha)v,t)\leqslant\alpha f^0(x,u,t)+(1-\alpha)f^0(y,v,t)-$$
$$-\alpha(1-\alpha)\frac{\varkappa}{2}|u-v|^2,\quad\varkappa=\mathrm{const}>0 \qquad (40)$$

при всех α, $0\leqslant\alpha\leqslant 1$, $(x,u,t),(y,v,t)\in E^n\times E^r\times[t_0,T]$, то функция (1) при условиях (34) является сильно выпуклой на $L_2^r[t_0,T]$ и будет достигать своей нижней грани на любом выпуклом замкнутом множестве $U\subseteq L_2^r[t_0,T]$, причем оптимальное управление единственно.

Доказательство. Нетрудно видеть, что решения задачи (34) обладают свойством

$$x(t, \alpha u + (1-\alpha)v) = \alpha x(t, u) + (1-\alpha)x(t, v), \quad \forall t \in [t_0, T],$$

при всех $u, v \in L_2^r[t_0, T]$ и $\alpha \in \mathbb{R}$ (ср. с (2.7)). Тогда выпуклость [сильная выпуклость] функции (1) при условиях (34) является простым следствием выпуклости $\Phi(x)$ и условия (37) [условия (40)]. Остальные утверждения теоремы вытекают из теорем 2.8, 2.10, 3.3. □

Пример 2. Пусть требуется минимизировать функцию

$$J(u) = \frac{1}{2}\int_0^T (x^2(t) + u^2(t))dt$$

при условиях $\dot{x}(t) = -ax(t) + u(t)$, $0 \leqslant t \leqslant T$; $x(0) = x_0$; $u = u(t) \in L_2[0, T] = U$, где $a, x_0, T > 0$ — заданные числа.

Эта задача является частным случаем задачи (1)–(3), (34) при $f^0 = (x^2 + u^2)/2$, $\Phi(x) \equiv 0$, $f = -ax + u$, $n = r = 1$ и к ней применимы теоремы 1–4. Поскольку здесь функция $f^0(x, u)$ удовлетворяет условию (40) при $\varkappa = 1/2$, то функция $J(u)$ сильно выпукла на $L_2[0, T]$ и достигает на $L_2[0, T]$ своей нижней грани в единственной точке $u_* = u_*(t) \in L_2[0, T]$. Поскольку

$$H(x, u, \psi) = -(x^2 + u^2)/2 + \psi(-ax + u),$$

то сопряженная задача (36) (или (12), (13)) здесь имеет вид

$$\dot{\psi}(t) = a\psi(t) + x(t, u), \quad 0 \leqslant t \leqslant T; \quad \psi(T) = 0,$$

а градиент согласно формуле (35) (или (11)) равен

$$J'(u) = u(t) - \psi(t, u), \quad 0 \leqslant t \leqslant T.$$

Условие (39) для оптимального управления тогда приведет к равенству

$$u_*(t) = \psi(t, u_*), \quad 0 \leqslant t \leqslant T.$$

Этот же результат был получен в примере 6.2.3 с помощью принципа максимума. В силу теоремы 4 последнее равенство является не только необходимым, но и достаточным для оптимальности управления $u_* = u_*(t)$.

Заметим, что, пользуясь условием (37) выпуклости функции $f^0(x, u, t)$, формулой (11) и теоремой 4.2.2, неравенство (38) можно переписать в эквивалентном виде

$$\int_{t_0}^T [H(x(t, u_*), u_*(t), t, \psi(t, u_*)) - H(x(t, u_*), u(t), t, \psi(t, u_*))]dt \geqslant 0, \quad u(t) \in U, \tag{41}$$

где

$$H(x, u, t, \psi) = -f^0(x, u, t) + \langle \psi, A(t) + B(t)u + f(t) \rangle.$$

Предлагаем читателю установить связь между принципом максимума и условием оптимальности (41) — это может быть сделано так же, как в примере 3.6 (формулы (3.30)–(3.32)).

4. Рассмотренная выше задача (1)–(3) является частным случаем задачи оптимального управления, когда правый конец траектории свободен. Более общие задачи оптимального управления, когда, например, правый конец

траектории закреплен или подвижен, или имеются какие-либо другие ограничения на фазовые координаты и управление, могут быть сведены к задаче вида (1)–(3) с помощью штрафных функций (см. § 4, п. 8).

Например, если задача (1)–(3) рассматривается при дополнительном условии $x(T) = x_I$ (правый конец закреплен), то в качестве штрафной функции для этого условия можно взять

$$P_k(u) = A_k |x(T, u) - x_I|^2, \quad k = 1, 2, \ldots,$$

и рассмотреть задачу минимизации функции

$$\Phi_k(u) = \int_{t_0}^T f^0(x(t, u), u(t), t)dt + \Phi(x(T, u)) + A_k |x(T, u) - x_I|^2$$

при условиях (1)–(3); здесь и ниже $\{A_k\}$ — некоторая заданная положительная последовательность, стремящаяся к бесконечности. Нетрудно видеть, что если функции f^0, f, Φ удовлетворяют условиям теоремы 1, функция $\Phi_k(u)$ дифференцируема и ее градиент определяется той же формулой (11), нужно лишь условие (13) для $\psi(T, u)$ заменить на

$$\psi(T, u) = -\Phi_x(x(T, u)) - 2A_k(x(T, u) - x_I).$$

Если задача (1)–(3) рассматривается при дополнительных фазовых ограничениях вида

$$a_i \leqslant x^i(t, u) \leqslant b_i, \quad t_0 \leqslant t \leqslant T, \quad i = 1, \ldots, m, \quad m \leqslant n, \tag{42}$$

где a_i, b_i — заданные постоянные, то штрафом может служить функция

$$P_k(u) = A_k \sum_{i=1}^m \int_{t_0}^T [(\max\{x^i(t, u) - b_i; 0\})^2 + (\max\{a_i - x^i(t, u); 0\})^2]dt.$$

Тогда задача (1)–(3), (42) сведется к решению последовательности задач минимизации функции

$$\Phi_k(u) = J(u) + P_k(u) = \int_{t_0}^T F_k^0(x(t,u), u(t), t)dt + \Phi(x(T,u)), \quad k = 1, 2, \ldots \tag{43}$$

при условиях (2), (3), где

$$F_k^0(x, u, t) = f^0(x, u, t) + A_k \sum_{i=1}^m [(\max\{x^i - b_i; 0\})^2 + (\max\{a_i - x^i; 0\})^2].$$

При каждом $k = 1, 2, \ldots$ задача (43) (2), (3) имеет тот же вид, что и задача (1)–(3). Заметим, что

$$F_{ku}^0 = f_u^0, \quad F_{kx^i}^0(x, u, t) = f_{x^i}^0(x, u, t) + 2A_k \max\{x^i - b_i; 0\} - 2A_k \max\{a_i - x^i; 0\},$$
$$i = 1, \ldots, m; \quad F_{kx^i}^0 = f_{x^i}^0, \quad i = m+1, \ldots, n.$$

Отсюда ясно, что если функции f^0, f, Φ удовлетворяют условиям теоремы 1, то и для задачи (43), (2), (3) также будут выполнены условия теоремы 1, и формула градиента для функции (43) будет определяться теми же формулами (4), (11)–(13), нужно лишь в них f^0 заменить на F_k^0.

Если задача (1)–(3) рассматривается при дополнительном условии

$$g(u)=\int_{t_0}^{T} G(x(t,u),u(t),t)dt+\Phi_1(x(T,u))\leqslant 0,$$

то штрафной функцией для этого неравенства можно взять

$$P_k(u)=A_k(\max\{g(u);0\})^2.$$

Возможно использование и других штрафных функций, аналогичных приведенным в § 5.15. Комментарии к методу штрафных функций, сделанные в § 5.15, сохраняют силу и для задач оптимального управления.

Отметим, что метод штрафных функций в главе 6 был использован для доказательства принципа максимума.

Упражнения

1. Рассмотреть задачу минимизации функции

$$J(u)=\int_0^T u^2(t)dt$$

при условиях (25)–(27) и с закрепленным правым концом траектории: $x(T)=x_I$; моменты T и точка x_I заданы. Применить метод штрафных функций для учета условия на правом конце; найти градиент штрафной функции.

2. Доказать, что при выполнении условий теоремы 1 функция (1) при условиях (2) дифференцируема по переменной $x_0\in E^n$, и по совокупности переменных $(x_0,u)\in E^n\times L_2^r[t_0,T]$; найти градиент.

3. Рассмотреть функцию

$$J(w)=\int_{t_0}^{T} f^0(x(t,w),w,t)dt+\Phi(x(T,w))$$

при условиях $\dot x(t)=f(x(t),w,t)$, $t_0\leqslant t\leqslant T$; $x(t_0)=x_0$, $w=(w^1,\ldots,w^r)$ — управляющие параметры, не зависящие от времени. Показать, что если функции f^0, f, Φ непрерывны и имеют непрерывные частные производные по x, w и $|f(x+\Delta x,w,t)-f(x,w,t)|\leqslant L|\Delta x|$ при всех $(x+\Delta x,w,t),(x,w,t)\in E^n\times E^r\times[t_0,T]$, то функция $J(w)$ дифференцируема и ее градиент будет равен

$$J'(w)=\int_{t_0}^{T} H_w(x(t,w),w,t,\psi(t,w))dt,$$

где $H(x,w,t,\psi)=f^0(x,w,t)+\langle f(x,w,t),\psi\rangle$, $\psi(t,w)$ — решение задачи

$$\dot\psi(t)=-H_x(x(t,w),w,t,\psi(t)),\quad t_0\leqslant t\leqslant T;\quad \psi(T)=\Phi_x(x(T,w)).$$

Указание: воспользоваться техникой доказательства теоремы 1.

4. Пусть выполнены все условия теорем 3, 4 (кроме, быть может, условия (40)) и пусть $U=\{u=u(t)\in L_2^r[t_0,T]: u(t)\in V$ почти всюду на $[t_0,T]\}$, где V — выпуклое множество из E^r. Доказать, что тогда принцип максимума является необходимым и достаточным условием оптимальности в задаче (1)–(3), (34). Указание: доказать выпуклость функции $H(x,u,t,\psi)$ по u и воспользоваться неравенством (38).

5. Пусть $J(u)=\int_0^1(u^2(t)-ax^2(t))dt$, где $\dot x(t)=u(t)\in L_2[0,1]$, $x(0)=0$, a — постоянная. При каких значениях параметра a функция $J(u)$ будет выпуклой или сильно выпуклой на $L_2[0,T]$? Показать, что $J(u)\in C^{1,1}(L_2)$, и найти градиент.

6. Пусть функции $f^0(x,u,t), f(x,u,t), \Phi(x)$ непрерывны по $(x,u,t)\in E^n\times E^r\times[t_0,T]$, выполняется условие (5) и $|f^0(x,u+h,t)-f^0(x,u,t)|\leqslant L(|h|^2+|u||h|)$, $L=\text{const}\geqslant 0$, при всех $(x,u+h,t),(x,u,t)\in E^n\times E^r\times[t_0,T]$. Доказать, что тогда функция (1) при условиях (2) непрерывна на $L_2^r[t_0,T]$ в метрике этого пространства.

7. Пусть функции $f^0(x,u,t), \Phi(x)$ непрерывны по $(x,u,t)\in E^n\times E^r\times[t_0,T]$, $f^0(x,u,t)$ выпукла по переменной $u\in E^r$ при каждом фиксированном $(x,t)\in E^n\times[t_0,T]$. Доказать, что тогда функция (1) при условиях (34) достигает своей нижней грани на любом выпуклом замкнутом ограниченном множестве $U\in L_2^r[t_0,T]$. Указание: установить, что $J(u)$ слабо полунепрерывна снизу на $L_2^r[t_0,T]$.

8. Пусть выполнены все условия теоремы 4 (кроме, быть может, условия (40)), U — выпуклое замкнутое ограниченное множество из $L_2^r[t_0,T]$, функция $g(x,t)$ непрерывна по $(x,t)\in E^n\times[t_0,T]$ и выпукла по $x\in E^n$ при каждом фиксированном $t\in[t_0,T]$. Пусть существует хотя бы одна траектория $x(t,u_0)$ задачи (34), $u_0\in U$, такая, что $g(x(t,u_0),t)\leqslant 0$ при всех $t\in[t_0,T]$. Доказать, что тогда функция (1) при условиях (34) и ограничении $g(x(t,u),t)\leqslant 0$, $t_0\leqslant t\leqslant T$, достигает на U своей нижней грани.

9. Пусть $f^0(x,u,t)=a(x,t)+\langle b(x,t),u\rangle$, $f(x,u,t)=A(x,t)+B(x,t)u$ и пусть матрицы $A(x,t), B(x,t), b(x,t)$ — порядков $n\times 1$, $n\times r$, $n\times 1$ соответственно и функции $a(x,t), \Phi(x)$ непрерывны по $(x,t)\in E^n\times[t_0,T]$, $\|A(x,t)\|\leqslant C_0|x|+C_1$, $\|B(x,t)\|\leqslant C_2$, где C_0, C_1, C_2 — неотрицательные постоянные. Пусть U — выпуклое замкнутое ограниченное множество из $L_2^r[t_0,T]$ и существует управление $u_0\in U$ такое, что соответствующее решение $x(t,u_0)$ задачи (2) удовлетворяет условию $x(T,u_0)=x_I$. Показать, что тогда функция (1) при условиях (2) и дополнительном условии $x(T,u)=x_I$ достигает своей нижней грани на U. Указание: установить, что если $\{u_k\}$ — минимизирующая последовательность, слабо сходящаяся к точке u_*, то $J(u_k)\to J(u_*)$.

§ 6. Градиент в задаче оптимального управления с дискретным временем

1. Рассмотрим следующую задачу: минимизировать функцию

$$I([u_i])=\sum_{i=0}^{N-1} F_i^0(x_i,u_i)+\Phi(x_N) \tag{1}$$

при условиях

$$x_{i+1}=F_i(x_i,u_i),\quad i=0,\ldots,N-1;\quad x_0=a, \tag{2}$$

$$[u_i]=(u_0,\ldots,u_{N-1});\quad u_i\in V_i,\quad i=0,\ldots,N-1, \tag{3}$$

где $x_i=(x_i^1,\ldots,x_i^n)$, $u_i=(u_i^1,\ldots,u_i^r)$, функции $F_i=(F_i^1,\ldots,F_i^n)$, F_i^0, Φ предполагаются известными, V_i — заданное множество из E^r, натуральное число $N\geqslant 1$ и начальная точка a заданы.

Задача (1)–(3) уже изучалась нами выше: в § 7.1 с помощью динамического программирования исследовалась проблема синтеза для этой задачи. В настоящем параграфе сформулируем достаточные условия дифференцируемости, выпуклости функции (1) при условиях (2), (3), а также выведем необходимые условия оптимальности. Будем пользоваться следующими обозначениями:

$$F_{ix}=\begin{pmatrix} F_{ix^1}^1 & \ldots & F_{ix^n}^1\\ \ldots\ldots & & \ldots\ldots\\ F_{ix^1}^n & \ldots & F_{ix^n}^n\end{pmatrix}=\begin{pmatrix} F_{ix}^1\\ \vdots\\ F_{ix}^n\end{pmatrix},\quad F_{iu}=\begin{pmatrix} F_{iu^1}^1 & \ldots & F_{iu^r}^1\\ \ldots\ldots & & \ldots\ldots\\ F_{iu^1}^n & \ldots & F_{iu^r}^n\end{pmatrix}=\begin{pmatrix} F_{iu}^1\\ \vdots\\ F_{iu}^n\end{pmatrix},$$

$$(F_{ix}^0)^T=(F_{ix^1}^0,\ldots,F_{ix^n}^0),\quad (F_{iu}^0)^T=(F_{iu^1}^0,\ldots,F_{iu^r}^0),\quad (\Phi_x)^T=(\Phi_{x^1},\ldots,\Phi_{x^n}).$$

Через $L_2^r[0,N]$ обозначим гильбертово пространство вектор-функций дискретной переменной $[u_i]=(u_0,u_1,\ldots,u_{N-1})$ со скалярным произведением

$\langle [u_i], [v_i] \rangle_{L_2} = \sum_{i=0}^{N-1} \langle u_i, v_i \rangle_{E^r}$ и с нормой $\|[u_i]\| = \left(\sum_{i=0}^{N-1} |u_i|^2_{E^r} \right)^{1/2}$. Пусть U — множество всех дискретных управлений $[u_i] = (u_0, \ldots, u_{N-1})$, удовлетворяющих условию (3); очевидно, $U \subseteq L_2^r[0, N]$.

Заметим, что (1) представляет собой функцию Nr переменных $u_0, u_1, \ldots, u_{N-1}$. Если функции $F_i(x, u)$ непрерывны, а F_i^0, Φ полунепрерывны снизу по совокупности переменных $(x, u) \in E^n \times V_i$, множества V_i замкнуты и ограничены в E^r, $i = 0, \ldots, N-1$, то функция $I([u_i])$ полунепрерывна снизу и существование оптимального управления $[u_{i*}]$, на котором функция (1) достигает своей нижней грани при условиях (2), (3), следует из теоремы 2.1.1. Для приближенного решения задачи (1)–(3) могут быть использованы методы гл. 5. Из-за большого числа переменных задачу (1)–(3), по-видимому, удобнее рассматривать в пространстве $L_2^r[0, N]$, считая функцию (1) зависящей от N векторных переменных $u_0, u_1, \ldots, u_{N-1}$.

Выведем формулу градиента функции (1) при условиях (2), (3) в пространстве $L_2^r[0, N]$.

Теорема 1. *Пусть функции F_i^0, F_i, Φ непрерывны по совокупности своих аргументов вместе со своими частными производными по переменным x, и при $x \in E^n$, $u \in V_i$, $i = 0, \ldots, N-1$. Кроме того, пусть*

$$|F_i(x + \Delta x, u + h) - F_i(x, u)|_{E^n} \leqslant L(|\Delta x|_{E^n} + |h|_{E^r}) \tag{4}$$

при всех x, $x + \Delta x$ и всех u, $u + h \in V_i$, $i = 0, \ldots, N-1$. Тогда функция (1) при условиях (2), (3) непрерывна и дифференцируема в норме $L_2^r[0, N]$, причем ее градиент $I'([u_i])$ в точке $[u_i] \in U$ представим в виде

$$I'([u_i]) = \{H_{iu}(x_i, \psi_i, u_i),\ i = 0, \ldots, N-1\} \in L_2^r[0, N], \tag{5}$$

где

$$H_i(x, \psi, u) = F_i^0(x, u) + \langle \psi, F_i(x, u) \rangle, \quad H_{iu} = (H_{iu^1}, \ldots, H_{iu^r}), \tag{6}$$

$[x_i] = (x_0, \ldots, x_N)$ — дискретная траектория задачи (2), соответствующая выбранному управлению $[u_i] \in U$, а вектор-функция $[\psi_i] = (\psi_{-1}, \psi_0, \ldots, \psi_{N-1})$ определяется из условий

$$\psi_{i-1} = H_{ix}(x_i, \psi_i, u_i), \quad i = 0, \ldots, N-1, \quad \psi_{N-1} = \Phi_x(x_N). \tag{7}$$

Доказательство. Пусть $[u_i], [u_i] + [h_i] \in U$ и пусть $[x_i]$ и $[x_i] + [\Delta x_i]$ — соответствующие этим управлениям дискретные траектории задачи (2), а $I([u_i])$ и $I([u_i]+[h_i]) = I([u_i]) + \Delta I$ — соответствующие значения функции (1). Из (2) следует, что приращение $[\Delta x_i]$ удовлетворяет условиям

$$\Delta x_{i+1} = F_i(x_i + \Delta x_i, u_i + h_i) - F_i(x_i, u_i), \quad i = 0, \ldots, N-1, \quad \Delta x_0 = 0. \tag{8}$$

Так как $\Phi(x + \Delta x) - \Phi(x) = \langle \Phi_x(x + \theta \Delta x), \Delta x \rangle$, $0 \leqslant \theta \leqslant 1$, то из (1) получим

$$\Delta I = \sum_{i=0}^{N-1} [F_i^0(x_i + \Delta x_i, u_i + h_i) - F_i^0(x_i, u_i)] + \langle \Phi_x(x_N), \Delta x_N \rangle + R_1, \tag{9}$$

где

$$R_1 = \langle \Phi_x(x_N + \theta \Delta x_N) - \Phi_x(x_N), \Delta x_N \rangle. \tag{10}$$

С учетом соотношений (7), (8) имеем

$$\langle \Phi_x(x_N), \Delta x_N \rangle = \langle \psi_{N-1}, \Delta x_N \rangle = \sum_{i=0}^{N-1} [\langle \psi_i, \Delta x_{i+1} \rangle - \langle \psi_{i-1}, \Delta x_i \rangle] =$$
$$= \sum_{i=0}^{N-1} \langle \psi, F_i(x_i + \Delta x_i, u_i + h_i) - F_i(x_i, u_i) \rangle - \sum_{i=0}^{N-1} \langle H_{ix}(x_i, \psi_i, u_i), \Delta x_i \rangle.$$

Подставляя полученное выражение в (9) и используя функцию $H_i(x, \psi, u)$, получим следующее представление для приращения функции:

$$\Delta I = \sum_{i=0}^{N-1} [H_i(x_i + \Delta x_i, \psi_i, u_i + h_i) - H_i(x_i, \psi_i, u_i) - \langle H_{ix}(x_i, \psi_i, u_i), \Delta x_i \rangle] + R_1.$$

Из формулы конечных приращений следует

$$H_i(x_i + \Delta x_i, \psi_i, u_i + h_i) =$$
$$= H_i(x_i, \psi_i, u_i) + \langle H_{ix}(x_i + \theta_i \Delta x_i, \psi_i, u_i + \theta h_i), \Delta x_i \rangle +$$
$$+ \langle H_u(x_i + \theta \Delta x_i, \psi_i, u_i + \theta h_i), h_i \rangle, \quad 0 \leqslant \theta_i \leqslant 1, \quad i = 0, \ldots, N-1.$$

Подставим это равенство в предыдущее представление для ΔI; будем иметь

$$\Delta I = \sum_{i=0}^{N-1} \langle H_{iu}(x_i, \psi_i, u_i), h_i \rangle + R, \quad R = R_1 + R_2 + R_3, \tag{11}$$

$$R_2 = \sum_{i=0}^{N-1} \langle H_{ix}(x_i + \theta_i \Delta x_i, \psi_i, u_i + \theta_i h_i) - H_{ix}(x_i, \psi_i, u_i), \Delta x_i \rangle, \tag{12}$$

$$R_3 = \sum_{i=0}^{N-1} \langle H_{iu}(x_i + \theta_i \Delta x_i, \psi_i, u_i + \theta_i h_i) - H_{iu}(x_i, \psi_i, u_i), h_i \rangle. \tag{13}$$

Для оценки остаточного члена R формулы (11) нам понадобится одна лемма, представляющая собой дискретный аналог леммы 6.3.1.

Лемма 1. *Если некоторые величины φ_i, $i = 0, \ldots, N$, удовлетворяют неравенствам*

$$0 \leqslant \varphi_0 \leqslant a, \quad 0 \leqslant \varphi_{i+1} \leqslant a + b \sum_{m=0}^{i} \varphi_m, \quad i = 0, \ldots, N-1, \quad b \geqslant 0, \tag{14}$$

то справедлива оценка

$$0 \leqslant \varphi_i \leqslant a(1 + b)^i, \quad i = 0, \ldots, N, \tag{15}$$

если же

$$0 \leqslant \varphi_{i-1} \leqslant a + b \sum_{m=i}^{N-1} \varphi_m, \quad i = 0, \ldots, N-1; \quad 0 \leqslant \varphi_{N-1} \leqslant a, \tag{16}$$

то верна оценка

$$0 \leqslant \varphi_i \leqslant a(1 + b)^{N-i-1}, \quad i = 0, \ldots, N-1. \tag{17}$$

Доказательство можно провести по индукции. При $k = 0$ по условию $0 \leqslant \varphi_0 \leqslant a$. Если неравенство $0 \leqslant \varphi_m \leqslant a(1 + b)^m$ верно при всех $m = 0, \ldots, i$, то из (14) следует $0 \leqslant \varphi_{i+1} \leqslant a + b \sum_{m=0}^{i} a(1 + b)^m = a(1 + b)^{i+1}$.

Аналогично можно убедиться, что из (16) вытекает оценка (17). □

Продолжим доказательство теоремы 1. Из соотношений (8) для $[\Delta x_i]$ с учетом условия (4) имеем

$$|\Delta x_{i+1}| = \left| \sum_{m=0}^{i} (\Delta x_{m+1} - \Delta x_m) \right| \leqslant \left| \sum_{m=0}^{i} (\Delta F_m(x_m, u_m) - \Delta x_m) \right| \leqslant$$
$$\leqslant (1 + L) \sum_{m=0}^{i} |\Delta x_m| + L \sum_{m=0}^{N-1} |h_m|.$$

Полагая в (14), (15) $a = L \sum_{i=0}^{N-1} |h_i|$, $b = (1 + L)$, $\varphi_i = |\Delta x_i|$, получим оценку

$$|\Delta x_i| \leqslant C_1 \sum_{i=0}^{N-1} |h_i|, \quad C_1 = L(2 + L)^N, \quad i = 0, \ldots, N. \tag{18}$$

Из условий теоремы следует непрерывность функций Φ_x, H_{ix}, H_{iu} по совокупности своих аргументов. Тогда с учетом оценки (18) из выражений (10), (12), (13) заключаем, что остаточный член R в формуле (11) имеет порядок $o(\|[h_i]\|)$. Таким образом, в формуле (11) приращения функции первое слагаемое является линейной ограниченной функцией на $L_2^r[0, N]$ относительно $[h_i]$, а второе слагаемое имеет порядок $o(\|[h_i]\|)$. Это значит, что функция (1) при условиях (2), (3) дифференцируема и ее градиент имеет вид (5). □

2. Зная формулу градиента, нетрудно расписать методы минимизации применительно к задаче (1)–(3). Например, метод проекции градиента здесь приводит к построению последовательности $[u_i]_k = (u_{0k}, \ldots, u_{N-1,k})$ по формулам

$$u_{i,k+1} = P_{V_i}(u_{i,k} - \alpha_k H_{iu}(x_{ik}, \psi_{ik}, u_{ik})), \quad i = 0, \ldots, N-1, \quad k = 0, 1, \ldots,$$

где $\alpha_k > 0$ выбирается, как в § 5.2; $[x_i]_k = (x_{0k}, \ldots, x_{Nk})$, $[\psi_i]_k = (\psi_{-1,k}, \ldots, \psi_{N-1,k})$ — решения задач (2) и (7) соответственно при $[u_i] = [u_i]_k$.

Приведем достаточные условия для того, чтобы градиент функции (1) при ограничениях (2), (3) удовлетворял условию Липшица.

Теорема 2. *Пусть выполнены все условия теоремы* 1 *и пусть функции* Φ_x, F_{ix}, F_{ix}^0, F_{iu}, F_{iu}^0 *удовлетворяют условию Липшица по совокупности* $(x, u) \in E^n \times V_i$ *с константой* $L > 0$, $i = 0, \ldots, N-1$. *Пусть, кроме того,*

$$|F_i(x, u)| \leqslant A_1 + A_2|x|; \quad A_1, A_2 = \text{const} \geqslant 0, \tag{19}$$

при всех $x \in E^n$, $u \in V_i$, *и множества* V_i *из* (3) *замкнуты и ограничены в* E^r, $i = 0, \ldots, N-1$. *Тогда градиент функции* (1) *при ограничениях* (2), (3) *удовлетворяет условию Липшица.*

Доказательство этой теоремы полностью аналогично доказательству теоремы 5.2; предлагаем читателю провести его самостоятельно.

3. Используя полученную формулу градиента, выведем необходимые условия оптимальности для задачи (1)–(3).

Теорема 3. *Пусть выполнены все условия теоремы* 1. *Пусть* $[u_{i*}]$ — *оптимальное управление,* $[x_{i*}]$ — *соответствующая ему траектория системы* (2), *т. е.* $I([u_{i*}]) = \inf I([u_i])$, *где нижняя грань берется по всем* $[u_i]$ *из условий* (2), (3). *Пусть* $[\psi_{i*}]$ — *решение задачи* (7), *соответствующее управлению* $[u_{i*}]$. *Тогда необходимо выполняются неравенства*

$$\langle H_{iu}(x_{i*}, \psi_{i*}, u_{i*}), u_i - u_{i*} \rangle \geqslant 0, \quad i = 0, \ldots, N-1, \tag{20}$$

при всех $u_i \in V_i$, *для которых направление* $e = u_i - u_{i*}$ *является возможным для множества* V_i *в точке* u_{i*}, *причем если* u_{i*} — *внутренняя точка множества* V_i, *то*

$$H_{iu}(x_{i*}, \psi_{i*}, u_{i*}) = 0; \tag{21}$$

функции $H_i(x, \psi, u)$ *определяются равенствами* (6).

Доказательство. Положим в формуле (11) $x_i = x_{i*}$, $\psi_i = \psi_{i*}$, $u_i = u_{i*}$; получим

$$\Delta I = \sum_{m=0}^{N-1} \langle H_{mu}(x_{m*}, \psi_{m*}, u_{m*}), h_m \rangle + o(\|[h_i]\|). \tag{22}$$

Пусть u_i — произвольная точка из V_i, для которой $u_{i*} + \alpha(u_i - u_{i*}) \in V_i$ при всех α, $0 < \alpha \leqslant \alpha_0$. Возьмем в (22) $[h_m] = (0, \ldots, 0, h_i = \alpha(u_i - u_{i*}), 0, \ldots, 0)$. Очевидно, при таком выборе $[h_m]$ управление $[u_{i*}] + [h_i] \in U$, и в силу оптимальности $[u_{i*}]$ из (22) тогда получим

$$0 \leqslant \Delta I = \langle H_{iu}(x_{i*}, \psi_{i*}, u_{i*}), \alpha(u_i - u_{i*}) \rangle + o(\alpha\|[h_i]\|), \quad 0 < \alpha \leqslant \alpha_0.$$

Поделив обе части этого неравенства на $\alpha > 0$ и перейдя к пределу при $\alpha \to +0$, сразу придем к неравенству (20). Если u_{i*} — внутренняя точка V_i, то в (20) можно положить $u_i = u_{i*} - \varepsilon H_{iu}(x_{i*}, \psi_{i*}, u_{i*}) \in V_i$ при некотором $\varepsilon > 0$, что сразу приведет к равенству (21). Если V_i — выпуклое множество, то условие (20), очевидно, выполнено для любого $u_i \in V_i$. Если V_i выпуклы при всех $i = 0, 1, \ldots, N-1$, то неравенства (20) в силу формулы (5) равносильны одному неравенству $\langle I'([u_{i*}]), [u_i] - [u_{i*}] \rangle_{L_2^r} \geqslant 0$ при всех $[u_i] \in U$, что совпадает с условием оптимальности из теоремы 3.3.

Таким образом, согласно теореме 3 оптимальным может быть лишь управление $[u_{i*}] \in U$, удовлетворяющее условиям (20). Однако, как связано управление $[u_{i*}]$ с экстремальными точками функции $H_i(x_{i*}, \psi_{i*}, u)$ на множестве V_i, условия (20) на это не дают ответа. В частности, возникает естественный вопрос: нельзя ли по аналогии с системами с непрерывным временем утверждать, что оптимальное управление $[u_{i*}]$ удовлетворяет принципу минимума

$$H_i(x_{i*}, \psi_{i*}, u_{i*}) = \min_{u \in V_i} H_i(x_{i*}, \psi_{i*}, u), \quad i = 0, \ldots, N-1 \tag{23}$$

(чтобы здесь, как и в гл. 6, можно было говорить о принципе максимума, нужно изменить знаки функций H_i, ψ_i). Ведь необходимое условие оптимальности тем ценнее, чем меньше управлений, подозрительных на оптимальность, оно выделяет. В этом смысле условие (23) явно имело бы преимущество перед условиями (20), так как неравенства (20) могут выполняться не только в тех точках, где имеет место (23), но и в других точках, в которых, например, $H_{iu}(x_{i*}, \psi_{i*}, u_{i*}) = 0$. К сожалению, оказывается, в управляемых системах с дискретным временем принцип минимума, вообще говоря, не имеет места: на оптимальном управлении функция $H_i(x_{i*}, \psi_{i*}, u)$ может и не достигать своего абсолютного минимума по $u \in V_i$.

Пример 1. Пусть фазовое состояние системы описывается двумя координатами (x_i, y_i), $i = 0, 1, 2$, причем

$$x_{i+1} = x_i + 2u_i, \quad y_{i+1} = -x_i^2 + y_i + u_i^2, \quad i = 0, 1; \quad x_0 = 3; \quad y_0 = 0.$$

Пусть требуется минимизировать функцию $I(u_0, u_1) = -y_2 \equiv \Phi(x_2, y_2)$ при условии $[u_i] = (u_0, u_1) \in U = \{(u_0, u_1): |u_i| \leqslant 5, i = 0, 1\}$. Нетрудно вычислить явное выражение $I(u_0, u_1) = 3(u_0+2)^2 - u_1^2 + 6$. Отсюда следует, что оптимальное управление (u_{0*}, u_{1*}) имеет вид: $u_{0*} = -2$, $u_{1*} = 5$ (возможность $u_{1*} = -5$ предоставляем читателю рассмотреть самостоятельно). Оптимальная траектория тогда такая: $x_{0*} = 3$, $x_{1*} = -1$, $x_{2*} = 9$; $y_{0*} = 0$, $y_{1*} = -5$, $y_{2*} = -19$; минимальное значение функционала равно $I_* = -19$.

Составим функцию $H_i(x_i, y_i, \psi_{1i}, \psi_{2i}) = \psi_{1i}(x_i + 2u_i) + \psi_{2i}(-x_i^2 + y_i + u_i^2)$. Система (7) здесь имеет вид

$$\psi_{1,i-1} = \psi_{1i} - 2\psi_{2i}x_i, \quad \psi_{2,i-1} = \psi_{2i}, \quad i = 1, 0,$$

$$\psi_{11} = \Phi_{x_2} = 0, \quad \psi_{21} = \Phi_{y_2} = -1.$$

Подставив сюда оптимальные (x_{i*}, y_{i*}), получим $\psi_{11*}=0$, $\psi_{10*}=-2$, $\psi_{21*}=-1$, $\psi_{20*}=-1$. Тогда $H_0(x_{0*}, y_{0*}, \psi_{10*}, \psi_{20*}, u)=-(u+2)^2+7$. Как видим, на оптимальном управлении $u=u_{0*}=-2$ функция $H_0(x_{0*}, y_{0*}, \psi_{10*}, \psi_{20*}, u)$ достигает своего абсолютного максимума, в то время как ее минимальное значение при $|u|\leqslant 5$ достигается в точке $u=5$. Поэтому для управляемых систем с дискретным временем принцип минимума, вообще говоря, не имеет места.

4. Рассмотрим задачу оптимального управления линейными дискретными системами: минимизировать функцию (1) при условиях (2), (3), если

$$F_i(x_i, u_i) = A_i x_i + B_i u_i + f_i, \quad i=0, \ldots, N-1, \tag{24}$$

где A_i, B_i, f_i — заданные матрицы порядков $n\times n$, $n\times r$ и $n\times 1$ соответственно.

Теорема 4. *Пусть функции F_i^0, Φ удовлетворяют условиям теоремы* 1. *Тогда функция* (1) *при условиях* (2), (3), (24) *дифференцируема в $L_2^r[0, N]$ и ее градиент $I'([u_i])$ в точке $[u_i]$ вычисляется по формуле*

$$I'([u_i]) = \{F_{iu}^0(x_i, u_i) + B_i^T \psi_i, \quad i=0, \ldots, N-1\}, \tag{25}$$

где $[x_i]=(x_0, \ldots, x_N)$ — решение задачи (2), *соответствующее выбранному управлению $[u_i]$, а $[\psi_i]=(\psi_{-1}, \ldots, \psi_{N-1})$ определяется из условий*

$$\psi_{i-1} = A_i^T \psi_i - F_{ix}^0(x_i, u_i), \quad i=0, \ldots, N-1, \quad \psi_{N-1} = \Phi_x(x_N), \tag{26}$$

матрицы A_i^T, B_i^T получены транспонированием матриц A_i, B_i. Если, кроме того, Φ_x, F_{ix}^0, F_{iu}^0 удовлетворяют условию Липшица по совокупности $(x, u)\in E^n\times E^r$, то градиент $I'([u_i])$ удовлетворяет условию Липшица на всем пространстве $L_2^r[0, N]$.

Формулы (25), (26) вытекают из теоремы 1; условие Липшица для градиента доказывается аналогично тому, как это делалось в теореме 5.3.

Укажем достаточные условия выпуклости и сильной выпуклости функции (1) при условиях (2), (3), (24).

Теорема 5. *Пусть выполнены условия* (24), *функция $\Phi(x)$ выпукла по x на E^n, а $F_i^0(x, u)$ выпукла по совокупности переменных (x, u), т. е.*

$$F_i^0(\alpha x + (1-\alpha)y, \alpha u + (1-\alpha)v) \leqslant \alpha F_i^0(x, u) + (1-\alpha)F_i^0(y, v) \tag{27}$$

при любых $x, y\in E^n$; $u, v\in E^r$, $\alpha\in[0, 1]$ и $i=0, \ldots, N-1$. Тогда функция (1) *выпукла на $L_2^r[0, N]$. Если при этом функции Φ, F_i^0 удовлетворяют условиям теоремы* 1, *множества V_i, $i=0, \ldots, N-1$, выпуклы, то для оптимальности управления $[u_{i*}]$ необходимо и достаточно, чтобы*

$$\langle F_{iu}^0(x_{i*}, u_{i*}) + B_i^T \psi_{i*}, u_i - u_{i*}\rangle \geqslant 0, \quad u_i\in V_i, \quad i=0, \ldots, N-1. \tag{28}$$

Доказательство. Решение задачи (2), (24), очевидно, обладает свойством $x_i(\alpha[u_i]+(1-\alpha)[v_i]) = \alpha x_i([u_i]) + (1-\alpha)x_i([v_i])$, $i=0, \ldots, N$, при любых α и любых $[u_i], [v_i]\in L_2^r[0, N]$. Тогда выпуклость функции (1) на $L_2^r[0, N]$ является простым следствием выпуклости $\Phi(x)$ и условия (27). Условие оптимальности (28) вытекает из теоремы 3.3 и формулы (25). □

С помощью теоремы 2.10 аналогично доказывается

Теорема 6. *Пусть выполнены условия* (24), *функция $\Phi(x)$ выпукла по $x\in E^n$ и, кроме того,*

$$F_i^0(\alpha x + (1-\alpha)y, \alpha u + (1-\alpha)v) \leqslant \alpha F_i^0(x, u) + (1-\alpha)F_i^0(y, v) - \alpha(1-\alpha)\tfrac{\varkappa}{2}|u-v|^2, \quad \varkappa = \text{const} > 0,$$

при любых $x, y\in E^n$; $u, v\in E^r$, $\alpha\in[0, 1]$ и $i=0, \ldots, N-1$. Тогда функция (1) *сильно выпукла на $L_2^r[0, N]$ и задача* (1)–(3), (24) *имеет, и притом единственное, решение на любом замкнутом выпуклом множестве $U\subseteq L_2^r[0, N]$.*

Упражнения

1. Доказать, что $I([u_i]) \equiv \alpha\sum_{i=0}^{N-1}|u_i|^2 + \beta\sum_{i=0}^{N-1}|x_i|^2 + \gamma x_N^2$, $\alpha, \beta, \gamma = \text{const}$, при условиях (2), (24) дважды дифференцируема в $L_2^r[0, N]$. Доказать, что если при этом $\alpha, \beta, \gamma\geqslant 0$, то $I([u_i])$ выпукла, а если $\alpha>0, \beta, \gamma\geqslant 0$, то сильно выпукла на $L_2^r[0, N]$.

2. Пусть выполнены все условия теоремы 5. Доказать, что тогда условие (23) является необходимым и достаточным условием оптимальности в задаче (1)–(3), (24).

§ 7. Оптимальное управление процессом нагрева стержня

Рассмотренные выше (см. гл. 6, 7, § 8.2–8.5) задачи оптимального управления относились к управляемым системам, описываемым обыкновенными дифференциальными уравнениями. В приложениях также возникает большое количество задач оптимального управления процессами, описываемыми дифференциальными (или интегро-дифференциальными) уравнениями с частными производными. Таким задачам посвящена обширная литература, для многих классов таких задач исследованы вопросы существования и единственности оптимального управления, получены условия оптимальности, разработаны методы их решения (см., например, [11; 35; 51; 67-69; 86; 104; 107; 118; 119; 122–124; 134–136; 138–140; 144; 187; 195–197; 227; 228; 285; 287; 289; 290; 360; 366; 391; 395; 402; 407; 408; 419; 420; 424; 425; 456; 459; 460; 463; 467–469; 475; 478; 479; 483–487; 489; 504; 512; 514; 515; 524; 540; 544; 547; 551–560; 573–577; 609; 630; 640–642; 654; 664–669; 682; 701; 706; 707; 730; 761; 778–780; 784; 790; 801; 802; 805; 818] и др.). Ниже будут рассмотрены некоторые из таких задач.

В этом параграфе мы займемся задачами оптимального управления процессами, описываемыми параболическими уравнениями. Такие задачи возникают при изучении управляемых процессов теплопроводности, диффузии, фильтрации и т. п. [107; 122–124; 228; 287; 289; 467; 504; 779].

1. Рассмотрим задачу, которая в теплофизических терминах может быть сформулирована следующим образом. Имеется однородный стержень $0\leqslant s\leqslant l$, левый конец $s=0$ которого теплоизолирован, на правом конце $s=l$ происходит теплообмен с внешней средой и, кроме того, в стержне имеются источники (или стоки) тепла. Через $x=x(s, t)$ обозначим температуру стержня в точке s в момент t. Пусть $x(s, 0)=\varphi(s)$, $0\leqslant s\leqslant l$, — распределение температуры в стержне в начальный момент времени $t=0$. Требуется, управляя температурой внешней среды и плотностью источников тепла в стержне, к заданному моменту $T>0$ распределение температуры в стержне сделать как можно ближе к заданному распределению $b(s)$, $0\leqslant s\leqslant l$.

Математическая формулировка этой задачи: требуется минимизировать функцию

$$J(u) = \int_0^l |x(s, T; u) - b(s)|^2 ds \tag{1}$$

при условии, что $x = x(s, t) = x(s, t; u)$ является решением краевой задачи:

$$x_t = a^2 x_{ss} + q(s, t), \quad (s, t) \in Q = \{0 < s < l,\ 0 < t \leqslant T\} \tag{2}$$

$$x_s|_{s=0} = 0, \quad 0 < t \leqslant T; \tag{3}$$

$$x_s|_{s=l} = \nu[p(t) - x(l, t)], \quad 0 < t \leqslant T, \tag{4}$$

$$x|_{t=0} = \varphi(s), \quad 0 \leqslant s \leqslant l, \tag{5}$$

где a^2, l, ν, T — заданные положительные величины, $p(t)$ — температура внешней среды, $q(s, t)$ — плотность источников тепла; $\varphi(s)$, $b(s)$ — заданные функции из $L_2[0, l]$. Предполагается, что $u = (p(t), f(s, t))$ — управление — принадлежит множеству

$$U = \{u = (p(t), q(s, t)) \in L_2[0, T] \times L_2(Q):$$

$$p_{\min} \leqslant p(t) \leqslant p_{\max} \text{ почти всюду на } [0, T],\ \iint\limits_Q q^2(s, t)dsdt \leqslant R^2\}, \tag{6}$$

где $p_{\min} < p_{\max}$, $R > 0$ — заданные числа.

Гильбертово пространство пар $u = (p(t), q(s, t))$, $p(t) \in L_2[0, T]$, $q(s, t) \in L_2(Q)$, со скалярным произведением

$$\langle u_1, u_2 \rangle = \int\limits_0^T p_1(t)p_2(t)dt + \iint\limits_Q q_1(s, t)q_2(s, t)dsdt$$

и с нормой $\|u\|_H = (\|p\|^2_{L_2} + \|q\|^2_{L_2})^{1/2}$, будем обозначать $H = L_2[0, T] \times L_2(Q)$.

Так как управление $u = (p(t), q(s, t)) \in U \subset H$ может иметь разрывы, то краевая задача (2)–(5), вообще говоря, не имеет классического решения [698]. Поэтому понятие решения этой краевой задачи требует уточнения.

Определение 1. *Обобщенным решением* краевой задачи (2)–(5), соответствующим управлению $u = (p(t), q(s, t)) \in H$, будем называть функцию $x = x(s, t) = x(s, t; u) \in H^{1,0}(Q)$ (см. обозначение в § 1), имеющую следы $x(s, \cdot) \in L_2[0, T]$ при всех $s \in [0, l]$ непрерывные в метрике $L_2[0, l]$, следы $x(\cdot, t) \in L_2[0, l]$ при всех $t \in [0, T]$ непрерывные в метрике $L_2[0, T]$, и удовлетворяющую интегральному тождеству

$$\int\limits_0^l x(s, T)\psi(s, T)ds - \int\limits_0^l \varphi(s)\psi(s, 0)ds - \iint\limits_Q (x\psi_t - a^2 x_s \psi_s)dsdt -$$

$$- \iint\limits_Q q\psi dsdt - a^2\nu \int\limits_0^T (p(t) - x(l, t))\psi(l, t)dt = 0$$

при всех $\psi = \psi(s, t) \in H^1(Q)$ и, кроме того, след $x(\cdot, t)$ при $t = 0$ совпадает с функцией $\varphi(s)$ почти всюду на $[0, l]$.

Можно доказать, что при каждом $u \in H$ краевая задача (2)–(5) имеет, и притом единственное, решение — по этим вопросам отсылаем читателя к книгам [441; 492]. Таким образом, функция (1) при условии (2)–(5) определена всюду на H.

Покажем, что задачу (1)–(6) можно свести к задаче (2.3) минимизации квадратичной функции. С этой целью решение краевой задачи (2)–(5) представим в виде суммы

$$x(s, t; u) = y(s, t; u) + x_0(s, t), \quad (s, t) \in Q, \tag{7}$$

где $y = y(s, t) = y(s, t; u)$ — решение задачи

$$y_t = a^2 y_{ss} + q(s, t), \quad (s, t) \in Q, \tag{8}$$

$$y_s|_{s=0} = 0, \quad y_s|_{s=l} = \nu(p(t) - y(l, t)), \quad 0 < t \leqslant T,$$

$$y|_{t=0} = 0, \quad 0 \leqslant s \leqslant l, \tag{9}$$

получающейся из задачи (2)–(5) при $\varphi \equiv 0$; $x_0(s, t)$ — решение задачи, получающейся из (2)–(5) при $p(t) \equiv 0$, $q(s, t) \equiv 0$. Введем оператор

$$Au = y(s, T; u), \quad s \in [0, l], \tag{10}$$

который действует из пространства $H = L_2[0, T] \times L_2(Q)$ в пространство $F = L_2[0, l]$. Из единственности решения задачи (8), (9) следует, что

$$y(s, t; \alpha u + \beta v) = \alpha y(s, t; u) + \beta y(s, t; v), \quad (s, t) \in Q \tag{11}$$

при всех $u, v \in H$, $\alpha, \beta \in \mathbb{R}$. Это означает, что оператор A, определенный согласно (10), линейный. Убедимся, что этот оператор ограниченный. Для этого умножим уравнение (8) на $y = y(s, t; u)$ и проинтегрируем по прямоугольнику Q. Преобразуем полученный интеграл с учетом условий (9):

$$0 = \iint\limits_Q (y_t - a^2 y_{ss} - q)ydsdt = \int\limits_0^l \int\limits_0^T \left(\tfrac{1}{2}(y^2)_t - a^2 y y_{ss} - qy\right)dsdt =$$

$$= \int\limits_0^l \tfrac{1}{2} y^2(s, t)\Big|_{t=0}^{t=T} ds - \int\limits_0^T a^2 y y_s\Big|_{s=0}^{s=l} dt + \iint\limits_Q a^2 y_s^2 dsdt - \iint\limits_Q qydsdt =$$

$$= \tfrac{1}{2}\int\limits_0^l y^2(s, T)ds + a^2\nu \int\limits_0^T y^2(l, t)dt - a^2\nu \int\limits_0^T p(t)y(l, t)dt +$$

$$+ a^2 \iint\limits_Q y_s^2 dsdt - \iint\limits_Q qydsdt. \tag{12}$$

Далее воспользуемся элементарными неравенствами

$$(a + b)^2 \leqslant 2a^2 + 2b^2, \quad |ab| \leqslant \tfrac{1}{2}\varepsilon a^2 + \tfrac{1}{2\varepsilon} b^2,$$

справедливыми для любых $a, b, \varepsilon > 0$. Отсюда и из (12) имеем

$$\tfrac{1}{2}\int\limits_0^l y^2(s, T)ds + a^2\nu \int\limits_0^T y^2(l, t)dt + a^2 \iint\limits_Q y_s^2 dsdt =$$

$$= a^2\nu \int\limits_0^T p(t)y(l, t)dt + \iint\limits_Q qydsdt \leqslant a^2\nu\left(\frac{\varepsilon_1}{2}\int\limits_0^T p^2(t)dt + \frac{1}{2\varepsilon_1}\int\limits_0^T y^2(l, t)dt\right) +$$

$$+ \varepsilon_2 \iint\limits_Q y^2(s, t)dsdt + \frac{1}{2\varepsilon_2}\iint\limits_Q q^2(s, t)dsdt \tag{13}$$

для любых $\varepsilon_1 > 0$, $\varepsilon_2 > 0$. Кроме того, поскольку

$$y^2(s, t) = \left(\int\limits_s^l y_s(\xi, t)d\xi - y(l, t)\right)^2 \leqslant 2\left(\int\limits_s^l y_s(\xi, t)d\xi\right)^2 + 2y^2(l, t) \leqslant$$

$$\leqslant 2l\int\limits_0^l y_s^2(\xi, t)d\xi + 2y^2(l, t), \quad (s, t) \in Q,$$

то

$$\iint\limits_Q y^2(s, t)dsdt \leqslant 2l^2 \iint\limits_Q y_s^2(s, t)dsdt + 2l\int\limits_0^T y^2(l, t)dt. \tag{14}$$

В правую часть неравенства (13) подставим оценку (14). После приведения подобных членов получим

$$\frac{1}{2}\int_0^l y^2(s,T)ds+(a^2\nu-\frac{1}{2}a^2\nu\varepsilon_1-\varepsilon_2 l^2)\int_0^T y^2(l,t)dt+(a^2-\varepsilon_2 l^2)\iint_Q y_s^2(s,t)dsdt \leqslant$$
$$\leqslant \frac{a^2\nu}{2\varepsilon_1}\int_0^T p^2(t)dt+\frac{1}{2\varepsilon_1}\iint_Q q^2(s,t)dsdt, \quad \forall\varepsilon_1>0, \quad \forall\varepsilon_2>0. \quad (15)$$

Возьмем числа ε_1, ε_2 столь малыми, чтобы $a^2\nu-\frac{1}{2}a^2\nu\varepsilon_1-l^2\varepsilon_2>0$, $a^2-l^2\varepsilon_2>0$, например, $\varepsilon_2=a^2\varepsilon_1$, $\varepsilon_1=\min\left\{\frac{1}{2l^2};\frac{\nu}{\nu+2l^2}\right\}$. Тогда из (15) получим оценку

$$\|Au\|_F^2=\int_0^l y^2(s,T;u)ds\leqslant C_0^2(\|p\|^2_{L_2[0,T]}+\|q\|^2_{L_2(Q)})=C_0^2\|u\|^2, \quad (16)$$

где $C_0^2=\max\left\{\frac{a^2\nu}{\varepsilon_1};\frac{1}{a^2\varepsilon_1}\right\}$. Это означает, что оператор (10) ограниченный. Следовательно, $A\in\mathcal{L}(H\to F)$, причем $\|A\|\leqslant C_0$. С учетом (7), (10) имеем $x(s,T;u)=Au+x_0(s,T)$, $s\in[0,l]$. Поэтому функцию (1) можем представить в виде

$$J(u)=\int_0^l|y(s,T;u)+x_0(s,T)-b(s)|^2ds=\|Au-b_1\|_F^2,$$

где $b_1=b_1(s)=b(s)-x_0(s,T)\in L_2[0,l]$. Таким образом, задача (1)–(6) сведена к задаче (2.3). Отсюда и из теоремы 2.12 с учетом слабой компактности множества (6) (теорема 2.6) заключаем, что задача (1)–(6) имеет хотя бы одно решение. Из примера 3.5 также следует, что функция (1) дважды непрерывно дифференцируема на H и ее производные равны

$$J'(u)=2A^*(Au-b_1), \quad J''(u)=2A^*A. \quad (17)$$

Тогда

$$\|J'(u)-J'(v)\|\leqslant 2\|A\|^2\|u-v\|\leqslant 2C_0^2\|u-v\| \quad \forall u,v\in H, \quad (18)$$

где постоянная C_0 взята из (16) и учтено, что $\|A\|=\|A^*\|\leqslant C_1$. Это значит, что $J(u)\in C^{1,1}(H)$ (определение 2.6.1).

Посмотрим, как в рассматриваемой задаче действует оператор A^*, сопряженный к оператору (10). Покажем, что оператор A^* каждому элементу $c=c(s)\in F=L_2[0,l]$ ставит в соответствие элемент $A^*c\in H=L_2[0,T]\times L_2(Q)$ по следующему правилу

$$A^*c=(a^2\nu\psi(l,t;c);\psi(s,t;c)), \quad (19)$$

где $\psi=\psi(s,t)=\psi(s,t;c)$ — обобщенное решение краевой задачи

$$\psi_t=-a^2\psi_{ss}, \quad (s,t)\in Q,$$
$$\psi_s|_{s=0}=0, \quad \psi_s|_{s=l}=-\nu\psi(l,t), \quad 0<t<T, \quad \psi|_{t=T}=c(s), \quad 0\leqslant s\leqslant l. \quad (20)$$

В самом деле, из (2)–(5), (10), (20), (21) имеем

$$\langle Au,c\rangle_F=\int_0^l y(s,T;u)c(s)ds=\int_0^l y(s,T;u)\psi(s,T;c)ds=$$

$$=\int_0^l\left(\int_0^T\frac{\partial}{\partial t}(y(t,s)\psi(t,s))dt+y\psi|_{t=0}\right)ds=$$
$$=\iint_Q(y_t\psi+y\psi_t)dsdt=\iint_Q[a^2y_{ss}\psi+f\psi+y(-a^2\psi_{ss})]dsdt=$$
$$=a^2\int_0^T(y_s\psi-y\psi_s)\Big|_{s=0}^{s=l}dt-a^2\iint_Q(y_s\psi_s-y_s\psi_s)dsdt+\iint_Q f\psi dsdt=$$
$$=a^2\nu\int_0^T p(t)\psi|_{s=l}ds+\iint_Q f\psi dsdt=\langle u,A^*c\rangle_H. \quad (21)$$

Формула (19) установлена. Из (10) (17), (19) следует, что градиент функции (1) равен

$$J'(u)=2(a^2\nu\psi(l,t);\psi(s,t)), \quad (22)$$

где $\psi(s,t)$ — решение задачи (20), (21) при $c=Au-b_1=y(s,T;u)-b_1(s)=y(s,T;u)+x_0(s,T)-b(s)=x(s,T;u)-b(s)$. Теперь, учитывая выпуклость функции (1), можем сказать: для того чтобы управление $u_*=(p_*(t),q_*(s,t))$ принадлежало множеству $U_*=\{u\in U: J(u)=J_*=\inf\limits_U J(u)\geqslant 0\}$ решений задачи (1)–(6), необходимо и достаточно, чтобы

$$\langle J'(u_*),u-u_*\rangle_H=\int_0^T a^2\nu\psi(l,t)(p(t)-p_*(t))dt+$$
$$+\iint_Q\psi(s,t)(q(s,t)-q_*(s,t))dsdt\geqslant 0 \quad (23)$$

при всех $u=(p(t),q(s,t))\in U$. Вариационное неравенство (23) вытекает из формулы (22) и теоремы 3.3.

Замечание 1. Приведенные выше доказательства оценки (16), равенства (21) нельзя признать вполне строгими, так как существование некоторых встретившихся в выкладках интегралов, законность операций интегрирования по частям не всегда вытекают из определения решения рассматриваемых краевых задач и остались необоснованными. Для строгого доказательства нужно было бы сначала сгладить функции $u(t)$, $\varphi(s)$, $q(s,t)$, $b(s)$, провести указанные преобразования для классических решений соответствующих сглаженных краевых задач, а затем перейти к пределу по параметру сглаживания и прийти к требуемым соотношениям для обобщенных решений краевых задач. Полная реализация намеченной здесь схемы строгого доказательства соотношений (16), (21) довольно громоздка для изложения, поэтому мы здесь вынуждены ограничиться приведенными выше рассуждениями, а читателя отсылаем за подробностями к руководствам и монографиям по уравнениям с частными производными [441; 492].

2. Для численного решения задачи (1)–(6) могут быть использованы методы проекции градиента и условного градиента (см. § 4).

Метод проекции градиента в задаче (1)–(6) сведется к построению последовательности $\{u_k=(p_k(t),q_k(s,t))\}$ по правилу

$$p_{k+1}(t)=\begin{cases}p_k(t)-2\alpha_k a^2\nu\psi(l,t;u_k) & \text{при}\\ p_{\min}\leqslant p_k(t)-2\alpha_k a^2\nu\psi(l,t;u_k)\leqslant p_{\max},\\ p_{\min} \quad \text{при} \quad p_k(t)-2\alpha_k a^2\nu\psi(l,t;u_k)<p_{\min},\\ p_{\max} \quad \text{при} \quad p_k(t)-2\alpha_k a^2\nu\psi(l,t;u_k)>p_{\max},\end{cases} \quad (24)$$

$$q_{k+1}(s,t)=\begin{cases} q_k(s,t)-2\alpha_k\psi(s,t;u_k) & \text{при} \\ \iint\limits_Q |q_k(s,t)-2\alpha_k\psi(s,t;u_k)|^2dsdt \leqslant R^2, \\ \dfrac{R(q_k(s,t)-2\alpha_k\psi(s,t;u_k))}{\left(\iint\limits_Q |q_k(s,t)-2\alpha_k\psi(s,t;u_k)|^2dsdt\right)^{1/2}} & \text{при} \\ \iint\limits_Q |q_k(s,t)-2\alpha_k\psi(s,t;u_k)|^2dtds > R^2, \end{cases} \tag{25}$$

где через $\psi(s,t;u_k)$ мы переобозначили решение задачи (20) при $c=c_k(s)=x(s,T;u_k)-b(s)$ (ср. с формулами (4.27) и (4.29)); выбор параметра α_k можно проводить с помощью одного из описанных в § 4.2 приемов. В частности, оценка (18) указывает на возможность выбора α_k из условий (5.2.4):

$$0<\varepsilon_0\leqslant\alpha_k\leqslant 2/(L+2\varepsilon),\quad \varepsilon>0,\quad L=2C_0^2. \tag{26}$$

Метод условного градиента в задаче (1)–(6) сведется к построению последовательности $\{u_k=(p_k(t),q_k(s,t))\}$ по правилу

$$\begin{aligned} p_{k+1}(t)&=p_k(t)+\alpha_k(\bar p_k(t)-p_k(t)),\quad 0\leqslant t\leqslant T, \\ q_{k+1}(s,t)&=q_k(s,t)+\alpha_k(\bar q_k(s,t)-q_k(s,t)),\quad (s,t)\in Q, \end{aligned} \tag{27}$$

где

$$\bar p_k(t)=\begin{cases} p_{\min} & \text{при } \psi(l,t;u_k)\geqslant 0, \\ p_{\max} & \text{при } \psi(l,t;u_k)<0, \end{cases} \tag{28}$$

$$\bar q_k(s,t)=\frac{R\psi(s,t;u_k)}{\left(\iint\limits_Q |\psi(s,t;u_k)|^2dsdt\right)^{1/2}}, \tag{29}$$

а параметр α_k, $0\leqslant\alpha_k\leqslant 1$, может быть выбран одним из указанных в § 5.4 приемов. Вспомогательное приближение $\bar u_k=(\bar p_k(t),\bar q_k(s,t))\in U$ здесь определено из условия минимума линейной функции

$$\langle J'(u_k),u\rangle_H=\int\limits_0^T 2a^2\nu\psi(l,t;u_k)p(t)dt+\iint\limits_Q 2\psi(s,t;u_k)q(s,t)dsdt$$

при ограничениях (6).

Заметим, что из равенства (11) при $u=\bar u_k$, $v=u_k$, $\beta=1-\alpha$ и формулы (22) следует, что функция

$$\begin{aligned} f_k(\alpha)&=J(u_k+\alpha(\bar u_k-u_k))=J(u_k)+2\alpha\int\limits_0^l(x(s,T;u_k)-y(s))\times \\ &\times(x(s,T;\bar u_k)-x(s,T;u_k))ds+\alpha^2\int\limits_0^l|x(s,T;\bar u_k)-x(s,T;u_k)|^2ds= \\ &=J(u_k)+\alpha\langle J'(u_k),\bar u_k-u_k\rangle_H+\alpha^2\int\limits_0^l|x(s,T;\bar u_k)-x(s,T;u_k)|^2ds \end{aligned}$$

при $x(s,T,\bar u_k)\neq x(s,T;u_k)$ является квадратным трехчленом относительно переменной α. Поэтому, рассуждая так же, как при выводе формулы (4.35), из условий

$$f_k(\alpha_k)=\min_{0\leqslant\alpha\leqslant 1}f_k(\alpha),\quad 0\leqslant\alpha_k\leqslant 1, \tag{30}$$

мы получим, что

$$\alpha_k=\min\{1;\alpha_k^*\}, \tag{31}$$

где

$$\begin{aligned} \alpha_k&=-\frac{\int\limits_0^l(x(s,T;u_k)-b(s))(x(s,T;\bar u_k)-x(s,T;u_k))ds}{\int\limits_0^l|x(s,T;\bar u_k)-x(s,T;u_k)|^2ds}= \\ &=-\frac{\int\limits_0^T a^2\nu\psi(l,t;u_k)(\bar p_k(t)-p_k(t))dt}{\int\limits_0^l|x(s,T;\bar u_k)-x(s,T;u_k)|^2ds}+\frac{\iint\limits_Q\psi(s,t;u_k)(\bar q_k(s,t)-q_k(s,t))dsdt}{\int\limits_0^l|x(s,T;\bar u_k)-x(s,T;u_k)|^2ds}\geqslant 0. \end{aligned}$$

В случае, когда $x(s,T;\bar u_k)=x(s,T;u_k)$, $0\leqslant s\leqslant l$, или $\alpha_k^*=0$, то $u_k=u_*$ — оптимальное управление задачи (1)–(6).

Согласно теоремам 4.2 и 4.4 последовательность $\{u_k\}$, построенная методом (24)–(26) или методом (27)–(31), является минимизирующей для задачи (1)–(6).

На практике приходится пользоваться разностными аналогами этих методов: встречающиеся в (24)–(26) и (27)–(31) интегралы вычисляются с помощью формул численного интегрирования (например, формулы прямоугольников или трапеций), а при решении краевых задач (2)–(5) и (20) можно пользоваться, например, неявной разностной схемой в сочетании с прогонкой [74; 89; 362; 480; 481; 615; 630–635, 698; 362] (см. также ниже § 10.7).

3. Перейдем к рассмотрению более сложной задачи минимизации функции (1), когда наряду с условиями (2)–(6) требуется, чтобы температура стержня не превышала некоторой заданной величины $\bar x$, т. е.

$$x(s,t;u)\leqslant\bar x,\quad (s,t)\in Q. \tag{32}$$

Такие задачи возникают при исследовании таких тепловых процессов, когда перегрев материала выше определенной критической температуры $\bar x$ не допустим.

Для решения задачи (1)–(6), (32) воспользуемся методом штрафных функций. Для учета ограничения (32) возьмем штрафную функцию

$$P_k(u)=A_k\iint\limits_Q|\max\{x(s,t;u)-\bar x;0\}|^2dsdt,$$

где $\{A_k\}$ — заданная положительная последовательность, $\{A_k\}\to\infty$ и при каждом $k=1,2,\ldots$ будем рассматривать задачу минимизации функции

$$\Phi_k(u)=\int\limits_0^l|x(s,T;u_k)-b(s)|^2ds+P_k(u) \tag{33}$$

при условиях (2)–(6). Функция (33) дифференцируема на H и ее градиент имеет вид

$$\Phi_k'(u)=2(a^2\nu\psi_k(l,t;u);\psi_k(s,t;u))\in H, \tag{34}$$

где $\psi_k(s,t;u)$ — решение уравнения

$$\psi_t=-a^2\psi_{ss}-2A_k\max\{x(s,t;u)-\bar x;0\},\quad (s,t)\in Q, \tag{35}$$

при краевых и начальных условиях из (20), $c(s)=x(s,T;u)-b(s)$.

В самом деле, приращение $\Delta\Phi_k(u)=\Phi_k(u+h)-\Phi_k(u)$ здесь представимо в виде

$$\Delta\Phi_k(u)=\int_0^l 2(x(s,T;u)-b(s))\Delta x(s,T)ds+ \\ +\iint_Q 2A_k\max\{x(s,t;u)-\bar{x};0\}\Delta x(s,t)dsdt+R_k, \quad (36)$$

где $\Delta x(s,t)=x(s,t;u+h)-x(s,t;u)$ — решение краевой задачи (8), (9), в которой $u=(p,q)$ заменено на $h=(\Delta p,\Delta q)$. Рассуждая так же как при выводе оценки (16), из (13), (14) нетрудно получить оценку

$$\iint_Q|\Delta x(s,t)|^2dsdt+\int_0^l|\Delta x(s,T)|^2ds\leqslant\widetilde{C}_0\|h\|_H^2.$$

Поэтому остаточный член R_k в (36) оценивается так:

$$|R_k|\leqslant\int_0^l|\Delta x(s,T)|^2ds+2A_k\iint_Q|\Delta x(s,t)|^2dsdt\leqslant C_k\|h\|_H^2,\ \ C_k=\widetilde{C}_0(1+2A_k). \quad (37)$$

Справедливо равенство

$$\int_0^l 2(x(s,T;u)-b(s))\Delta x(s,T)ds= \\ =\int_0^T a^2\nu\psi_k(l,t;u)\Delta p(t)dt+\iint_Q\psi_k(s,t;u)\Delta q(s,t)dsdt- \\ -\iint_Q 2A_k\max\{x(s,t;u)-\bar{x};0\}\Delta x(s,t)dsdt,$$

которое вытекает из цепочки равенств, аналогичных (21). С учетом этого равенства из (36), (37) получим

$$\Delta\Phi_k(u)=\int_0^T a^2\nu\psi_k(l,t;u)\Delta p(t)dt+\iint_Q\psi_k(s,t;u)\Delta q(s,t)dsdt+R_k, \quad (38)$$

при этом $|R_k|\leqslant C_k\|h\|_H^2$, $C_k=\mathrm{const}\geqslant 0$. Из (38) вытекает дифференцируемость функции (33), и получается формула (34).

Как видим, формула (34) вполне аналогична формуле (21), и поэтому нет ничего удивительного в том, что методы проекции градиента и условного градиента для задачи (33), (2)–(6) реализуются по тем же формулам (24), (25) и (27)–(29) с заменой ψ на ψ_k.

4. В рассмотренных задачах при выводе формулы градиента важную роль играли вспомогательные краевые задачи вида (20), которые принято называть *сопряженными краевыми задачами*, соответствующими исходной задаче оптимального управления. Возникает вопрос, откуда берется сопряженная краевая задача, по каким правилам она составляется?

Здесь мы приведем некоторые эвристические соображения, помогающие в составлении сопряженной краевой задачи, установим связь между решением сопряженной краевой задачи и множителем Лагранжа задачи оптимального управления [112]. Все построения проведем на примере задачи (1)–(6), считая, что

$$u=(p,q)\in U=H=L_2[0,T]\times L_2(Q).$$

Следуя уже известной нам процедуре исследования задач на условный экстремум (см. §§ 2.3, 4.8, 4.9, 6.2 и § 2), составим функцию Лагранжа задачи (1)–(6):

$$L(x,p,q,\psi)=\int_0^l|x(s,T)-y(s)|^2ds+\iint_Q\psi_k(s,t)(-x_t(s,t)+a^2x_{ss}(s,t)+q(s,t))dsdt, \quad (39)$$

где $\psi(s,t)$ — множитель Лагранжа, соответствующий ограничению (2). Будем предполагать, что функции $x(s,t)$, $\psi(s,t)$ являются достаточно гладкими на $\overline{Q}=\{(s,t)\colon 0\leqslant s\leqslant l,\ 0\leqslant t\leqslant T\}$. Поскольку уравнение (2) уже учтено в (39), то от функции $x(s,t)$ будем требовать лишь удовлетворения граничным и начальным условиям (3)–(5); дополнительные условия на функцию $\psi(s,t)$ будут наложены ниже.

Дадим приращения (вариации) переменным x,p,q, т. е. рассмотрим функции $x(s,t)+\delta x(s,t)$, $p(t)+\delta p(t)$, $q(s,t)+\delta q(s,t)$, $(s,t)\in\overline{Q}$, такие, что

$$\delta x_s|_{s=0}=0,\quad \delta x_s|_{s=l}=\nu(\delta p(t)-\delta x(l,t)),\quad \delta x|_{t=0}=0. \quad (40)$$

Вариация функции Лагранжа (39), представляющая собой главную линейную часть приращения этой функции, имеет вид

$$\delta L=\int_0^l 2(x(s,T)-y(s))\delta x(s,T)ds+\iint_Q\psi(-\delta x_t+a^2\delta x_{ss}+\delta q)dsdt.$$

Учитывая условия (40), преобразуем двойной интеграл с помощью интегрирования по частям. Получим

$$\delta L=\int_0^l[2(x(s,T)-y(s))-\psi(s,T)]\delta x(s,T)ds+\iint_Q(\psi_t+a^2\psi_{ss})\delta x dsdt+\iint_Q\psi\delta q dsdt+ \\ +a^2\int_0^T(\nu\psi(l,t)+\psi_s(l,t))\delta x(l,t)dt+a^2\nu\int_0^T\psi(l,t)\delta p(t)dt+\int_0^T\psi_s(0,t)\delta x(0,t)dt.$$

Считая, что в оптимальной точке выполняется условие стационарности $\delta L=0$, и пользуясь достаточно большим произволом в выборе $\delta x(s,t)$, приравняем нулю коэффициенты при вариациях $\delta x(s,T)$, $\delta x(s,t)$, $\delta x(l,t)$, $\delta x(0,t)$ и придем к условиям для множителя Лагранжа $\psi(s,t)$, полностью совпадающим с сопряженной краевой задачей (20); приравнивая нулю коэффициенты при $\delta p(t)$ и $\delta q(s,t)$, получим условия $\psi(l,t)=0$, $\psi(s,t)=0$, которые согласно формуле (22) означают равенство $J'(u)=0$ — условие оптимальности в задаче (1)–(6) при $u\in U=H$.

Изложенный на примере задачи (1)–(6) подход к получению сопряженной краевой задачи применим для широкого класса задач оптимального управления процессами, описываемыми как обыкновенными дифференциальными уравнениями, так и уравнениями с частными производными. Этот подход кратко можно сформулировать в виде следующих правил:

1) сначала нужно записать задачу минимизации в виде

$$J(u)\to\inf,\quad L_i(x,u,\xi)=0,\quad \xi\in G_i,\quad i=1,\ldots,m;\quad l_i(x,u)=0,\quad i=1,\ldots,p,$$

где G_i — заданная область из евклидова пространства E^{n_i}, $x=x(\xi)=(x^1(\xi),\ldots,x^n(\xi))$ — фазовые переменные, $u=(u^1(\xi),\ldots,u^r(\xi))$ — управления, L_i — дифференциальный оператор, l_i — операторы граничных и начальных условий, и составить функцию Лагранжа

$$L(x,u,\psi)=J(u)+\sum_{i=1}^m\int_{G_i}\psi_i(\xi)L_i(x(\xi),u(\xi),\xi)d\xi;$$

2) затем нужно найти вариацию функции Лагранжа по фазовым переменным и управлениям с соблюдением граничных и начальных условий $l_i(x,u)=0$, $i=1,\ldots,p$, и с помощью интегрирования по частям (или с помощью формулы Гаусса — Остроградского для сложных многомерных областей) с учетом граничных и начальных условий преобразовать полученную вариацию так, чтобы выражения под знаками интегралов по областям G_i не содержали частных производных вариаций фазовых переменных;

3) наконец, пользуясь условием стационарности функции Лагранжа и произволом в выборе вариаций фазовых переменных, приравнять нулю коэффициенты при соответствующих вариациях; совокупность полученных при этом равенств представляет собой условия на множители Лагранжа и образует искомую сопряженную краевую задачу.

Предлагаем читателю, пользуясь этими правилами, самостоятельно вывести сопряженную краевую задачу для задачи (33), (2)–(6), а также для рассматриваемых ниже задач оптимального управления.

Подчеркнем, что приведенные в этом пункте рассуждения не могут считаться строгими и являются лишь полезными наводящими соображениями при получении сопряженной краевой задачи, выводе формулы градиента, необходимых условий оптимальности. Для полной

строгости нужно еще выполнить большую и трудную работу и определить, что понимается под решением исходной и сопряженной краевых задач, исследовать вопросы существования и единственности решения этих задач, дать строгий вывод формулы приращения с оценкой остаточного члена и т. п.

5. Выше были подробно рассмотрены задачи оптимального управления для простейшего уравнения теплопроводности с одной пространственной переменной. Из этих рассмотрений видно, что, хотя окончательные расчетные формулы, реализующие методы проекции градиента и условного градиента для ограничений вида (6) достаточно просты и удобны для использования на ЭВМ, вывод этих формул связан с довольно громоздкими оценками, и строгое исследование таких задач является весьма тонким и хлопотным делом. Еще более трудным и громоздким становится исследование задач оптимального управления системами, описываемыми более общими параболическими уравнениями при более сложных функционалах, граничных условиях, ограничениях на управления и на решения. Здесь мы ограничимся лишь приведением формул градиента для следующей задачи.

Пусть Ω — заданная область в евклидовом пространстве E^n переменных $s=(s_1,\ldots,s_n)$ с кусочно гладкой границей Γ; пусть Γ_1 и Γ_2 — кусочно гладкие части границы Γ, не имеющие общих точек, причем $\Gamma_1\cup\Gamma_2=\Gamma$ (в частности, одно из этих множеств Γ_1 или Γ_2 может быть пустым). Пусть t_0, T — заданные моменты времени. Обозначим $Q=\{(s,t)\colon s\in\Omega,\ t_0\leqslant t\leqslant T\}$. Пусть в Q имеется конечное число кусочно гладких поверхностей, разбивающих Q на конечное число подобластей Q_i, $i=1,2,\ldots,p$ (случай $p=1$, $Q_1=Q$ не исключается). Поверхность, которая служит границей для подобластей Q_l и Q_k, обозначим через Γ_{kl}. Будем считать, что физические характеристики рассматриваемой области Q (плотность, теплопроводность, удельная теплоемкость и т. п.) непрерывны внутри каждой подобласти Q_i и могут терпеть разрывы типа скачка лишь на поверхностях Γ_{kl}.

Рассмотрим задачу минимизации функции (функционала):

$$J(u)=\int_{t_0}^{T}\int_{\Omega}\Phi_0(s,t,x(s,t;u),u_0(s,t))dsdt+\int_{t_0}^{T}\int_{\Gamma_1}\Phi_1\left(s,t,\frac{\partial x(s,t;u)}{\partial N},u_1(s,t)\right)d\Gamma_1dt+$$
$$+\int_{t_0}^{T}\int_{\Gamma_2}\Phi_2(s,t,x(s,t;u),u_2(s,t))d\Gamma_2dt+\int_{\Omega}\Phi_3(s,x(s,T;u),u_3(s))ds \quad (41)$$

при условиях

$$x_t=\sum_{i,j=1}^{n}a_{ij}(s,t)x_{s_is_j}+\sum_{i=1}^{n}b_i(s,t)x_{s_i}+c(s,t)x-\varphi_0(s,t,u_0(s,t)),\quad (s,t)\in Q; \quad (42)$$

$$[x(s,t)]_{\Gamma_{kl}}=0,\quad \left[\frac{\partial x(s,t)}{\partial N}\right]_{\Gamma_{kl}}=0,\quad (s,t)\in\Gamma_{kl}; \quad (43)$$

$$x(s,t)|_{s\in\Gamma_1}=\varphi_1(s,t,u_1(s,t))|_{s\in\Gamma_1},\quad t_0\leqslant t\leqslant T; \quad (44)$$

$$\left(\frac{\partial x(s,t)}{\partial N}+\gamma(s,t)x(s,t)\right)\Big|_{s\in\Gamma_2}=\varphi_2(s,t,u_2(s,t))|_{s\in\Gamma_2},\quad t_0<t\leqslant T; \quad (45)$$

$$x(s,t)|_{t=t_0}=\varphi_3(s,u_3(s)),\quad s\in\Omega, \quad (46)$$

где $\frac{\partial x(s,t)}{\partial N}=\sum_{i,j=1}^{n}a_{ij}(s,t)\cos(\widehat{n,s_i})\frac{\partial x(s,t)}{\partial s_j}$ — производная функции x по конормали к границе подобласти Q_k в точке (s,t); $n=n(s,t)$ — внешняя для Q_k нормаль в той же точке (s,t) с направляющими косинусами $\cos(\widehat{n,s_i})$, $i=1,2,\ldots,n$; $[z(s,t)]_{\Gamma_{kl}}$ — разность предельных значений функции $z(\xi,\tau)$ при стремлении точки (ξ,τ) к $(s,t)\in\Gamma_{kl}$ изнутри подобласти Q_k и Q_l соответственно; $a_{ij}(s,t)$, $b_i(s,t)$, $\varphi_i(s,t,u)$, $c(s,t)$, $\gamma(s,t)$, $\Phi_i(s,t,z,u)$ — заданные функции своих аргументов, $\sum_{i,j=1}^{n}a_{ij}(s,t)\xi_i\xi_j\geqslant\alpha\sum_{i=1}^{n}\xi_i^2$, $a_{ij}\equiv a_{ji}$, при всех $(s,t)\in Q$, $\xi=(\xi_1,\ldots,\xi_n)$; $\alpha=\text{const}>0$; $a_{ij}(s,t)$, $b_i(s,t)$ непрерывны внутри подобластей Q_i и могут терпеть разрывы типа скачка лишь на поверхностях Γ_{kl}; функции $u=(u_0(s,t),u_1(s,t),u_2(s,t),u_3(s))$ являются управлениями, подлежащими определению из условия минимума функции (41). Будем считать, что $u_0(s,t)\in L_2(Q)$, $u_i(s,t)\in L_2(\Gamma_i)$, $i=1,2$, $u_3(s)\in L_2(\Omega)$ и удовлетворяют ограничениям типа (6).

Нетрудно видеть, что рассмотренные выше задачи оптимального управления для уравнения теплопроводности (2) являются простым частным случаем задачи (41)–(46). Заметим, что некоторые из управлений u_j, $j=0,1,2,3$, могут отсутствовать в задаче (41)–(46), — в этом случае функции Φ_i, φ_i не зависят от u_j.

Управление $u=(u_0(s,t),u_1(s,t),u_2(s,t),u_3(s))$ в задаче (41)–(46) удобно считать элементом гильбертова пространства $H=L_2(Q)\times L_2(\Gamma_1)\times L_2(\Gamma_2)\times L_2(\Omega)$, в котором скалярное произведение двух любых элементов $u_i=(u_{0i}(s,t),u_{1i}(s,t),u_{2i}(s,t),u_{3i}(s))$, $i=1,2$, определяются посредством формулы

$$\langle u_1,u_2\rangle_H=\iint_Q u_{01}(s,t)u_{02}(s,t)dsdt+\int_{t_0}^{T}\int_{\Gamma_1}u_{11}(s,t)u_{12}(s,t)d\Gamma_1dt+$$
$$+\int_{t_0}^{T}\int_{\Gamma_2}u_{21}(s,t)u_{22}(s,t)dsdt+\int_{\Omega}u_{31}(s)u_{32}(s)ds,$$

а норма — формулой $\|u\|_H=(\langle u_1,u_2\rangle_H)^{1/2}$. При некоторых требованиях к функциям a_{ij}, b_i, c, γ, φ_i, Φ_i и к области Q можно доказать, что при каждом $u\in H$ обобщенное решение $x(s,t,u)$ краевой задачи (42)–(46) существует и единственно [492; 698], а функция (41) при условиях (42)–(46) дифференцируема в H и ее градиент имеет вид

$$J'(u)=\left\{-\psi(s,t,u)\varphi_{0u}+\Phi_{0u};-\left[\sum_{i=1}^{n}\sum_{j=1}^{n}(a_{ij}\psi)_{sj}\cos(\widehat{n,s_i})+\sum_{i=1}^{n}b_i\cos(\widehat{n,s_i})\Phi_{1z}\right]\varphi_{1u}+\Phi_{1u}|_{s\in\Gamma_1};\right.$$
$$\left.(\psi\varphi_{2u}+\Phi_{2u})|_{s\in\Gamma_2};\psi(s,t_0,u)\varphi_{3u}+\Phi_{3u}\right\}\in H,$$

где частные производные φ_{iu} вычислены для аргументов $(s,t,u_i\ (s,t))$, $i=0,1,2,3$, а Φ_{iu} и Φ_{1z} — для тех же аргументов, с которыми функция Φ_i входит в (41), $i=0,1,2,3$; $\psi=\psi(s,t,u)$ — решение сопряженной краевой задачи

$$\psi_t=-\sum_{i,j=1}^{n}(a_{ij}\psi)_{s_is_j}+\sum_{i=1}^{n}(b_i\psi)_{s_i}-c\psi-\frac{\partial\Phi_0(s,t,x(s,t,u),u_0(s,t))}{\partial z},\quad (s,t)\in Q;\quad [\psi]_{\Gamma_{kl}}=0;$$
$$-\left[\sum_{i,j=1}^{n}(a_{ij}\psi)_{s_j}\cos(\widehat{n,s_i})-\sum_{i=1}^{n}b_i\cos(\widehat{n,s_i})\psi+\psi\cos(\widehat{n,t})\right]_{\Gamma_{kl}}=0,\quad (s,t)\in\Gamma_{kl};$$
$$\psi(s,t)\Big|_{s\in\Gamma_1}=-\frac{\partial\Phi_1\left(s,t,\frac{\partial x(s,t,u)}{\partial N},u_1(s,t)\right)}{\partial z}\Big|_{s\in\Gamma_1},\quad t_0<t<T;$$
$$\sum_{i,j=1}^{n}(a_{ij}\psi)_{s_j}\cos(\widehat{n,s_i})+\gamma(s,t)\psi-\sum_{i=1}^{n}b_i\cos(\widehat{n,s_i})\psi\Big|_{s\in\Gamma_2}=$$
$$=\frac{\partial\Phi_2(s,t,x(s,t,u),u_2(s,t))}{\partial z}\Big|_{s\in\Gamma_2},\quad t_0<t<T;$$
$$\psi(s,T)=\frac{\partial\Phi_3(s,t,x(s,T,u),u_3(s,t))}{\partial z},\quad s\in\Omega.$$

Имея формулу градиента и опираясь на общую схему методов проекции градиента и условного градиента из § 4, пп. 2, 3, нетрудно расписать формулы, реализующие эти методы применительно к задаче (41)–(46).

Упражнения

1. Рассмотреть задачу (1)–(6), заменив условие $p_{\min}\leqslant p(t)\leqslant p_{\max}$ на условие $p(t)\in L_2[0,T]$, $\int_0^T|p(t)-\bar p(t)|^2dt\leqslant R_0^2$, где $\bar p(t)\in L_2[0,T]$ и число $R_0>0$ заданы. Описать методы проекции градиента и условного градиента.

2. Найти градиенты функции (1) при условиях (2)–(5) по каждой из переменных $\varphi=\varphi(x)\in L_2[0,l]$, $p(t)\in L_2[0,l]$, $q(s,t)\in L_2(Q)$ и по совокупности переменных $u=(\varphi(x),p(t),q(s,t))\in H=L_2[0,l]\times L_2[0,T]\times L_2(Q)$. Описать методы проекции градиента и условного градиента, считая, что $p(t)$, $q(s,t)$ удовлетворяют условиям (6), и, кроме того, $\varphi_{\min}\leqslant\varphi(s)\leqslant\varphi_{\max}$ почти всюду на $[0,l]$.

3. Рассмотреть функцию $J(u)=\beta_0\int_0^l|x(s,T,u)-b(s)|^2ds+\beta_1\int_0^T|p(t)|^2\,dt+\beta_2\iint_Q|q(s,t)|^2dsdt$ при условиях (2)–(5), считая $\beta_i=\text{const}>0$, $i=0,1,2$; $b(s)\in L_2[0,l]$. Доказать, что эта функ-

ция сильно выпукла на H; найти ее градиент; описать градиентный метод при $U=H$ и методы проекции градиента и условного градиента при ограничениях (6).

4. Рассмотреть функцию $J(p)=\int_0^l |x(s,T,u)-b(s)|^2 ds+\beta\int_0^T |p(t)|^2\,dt$, $\beta=\text{const}>0$, при условиях (2)–(6), считая функцию $f(s,t)\in L_2(Q)$ заданной. Доказать, что тогда управление $p_*=p_*(t)$ минимизирует $J(p)$ тогда и только тогда, когда $H(p_*(t),\psi(l,t,p_*))=\min H(p,\psi(l,t,p_*))$, $0\leqslant t\leqslant T$, где минимум берется по отрезку $p_{\min}\leqslant p(t)\leqslant p_{\max}$, $H(p,\psi)=a^2\nu p\psi+\beta p^2$, $\psi(s,t,p)$ — решение задачи (20). Указание: заметить, что $H(p,\psi)$ выпукла по p, $J'(p)=H_p(p(t),\psi(l,t,p))$, и воспользоваться теоремой 3.3.

5. Рассмотреть задачу (1)–(6) при ограничениях $\underline{x}\leqslant x(s,t,u)\leqslant\overline{x}$, где $\underline{x},\overline{x}$ — заданные величины; учесть эти ограничения с помощью штрафной функции.

6. Требуется минимизировать функцию $J(p)=\int_0^T p^2(t)dt$ при условиях (2)–(6), считая функцию $f(s,t)\in L_2(Q)$ заданной, и дополнительном условии $x(s,T,p)=y(s)$, $0\leqslant s\leqslant l$, $y(s)\in L_2[0,l]$. Указать штрафную функцию $P_k(p)$ для дополнительного условия; найти градиент функции $\Phi_k(p)=J(p)+P_k(p)$; описать методы проекции градиента и условного градиента.

§ 8. Оптимальное управление колебательными процессами

Задачи оптимального управления колебательными процессами имеют многочисленные приложения: к ним, например, приводят задачи об успокоении качки судна или стрелы подъемного крана, о работе вибротранспортеров, об организации виброзащиты, амортизации и т. п. [122–124; 354; 355; 395; 557; 596; 599; 641; 698; 706; 802; 812]. Здесь мы рассмотрим несколько задач оптимального управления процессами, описываемыми уравнением колебания струны и уравнением поперечных колебаний стержня.

1. Пусть имеется однородная упругая гибкая струна, один конец которой свободен, на другой ее конец действует внешняя сила, и, кроме того, к каждой точке струны также приложена внешняя сила. Требуется, управляя указанными внешними силами, к заданному моменту времени привести струну в состояние, характеризуемое смещением и скоростью точек струны, как можно меньше отличающееся от некоторого заданного состояния (например, состояния покоя, когда смещение и скорость равны нулю).

Математическая формулировка этой задачи: минимизировать функцию

$$J(u)=\int_0^l (|x(s,T;u)-y_0(s)|^2+|x_t(s,T;u)-y_1(s)|^2)ds \tag{1}$$

при условиях

$$x_{tt}=a^2x_{ss}+q(s,t),\quad (s,t)\in Q=\{0<s<l,\ 0<t<T\}, \tag{2}$$

$$\begin{gathered} x_s|_{s=0}=p(t),\quad x_s|_{s=l}=0,\quad 0<t<T;\\ x|_{t=0}=\varphi_0(s),\quad x_t|_{t=0}=\varphi_1(s),\quad 0<s<l, \end{gathered} \tag{3}$$

$$u=(p(t),q(s,t))\in U\subseteq H=L_2[0,T]\times L_2(Q), \tag{4}$$

где $a^2>0$, $l>0$, $T>0$, $\varphi_i(s)$, $y_i(s)$, $i=0,1$, $0\leqslant s\leqslant l$ — заданные функции, причем $\varphi_0(s)\in H^1[0,l]$, $\varphi_1(s)$, $y_0(s)$, $y_1(s)\in L_2[0,l]$, U — заданное множество из H. В частности, если $y_0(s)=y_1(s)=0$, то здесь можно говорить о задаче наилучшего успокоения струны.

Определение 1. Под обобщенным решением краевой задачи (2), (3), соответствующим управлению $u=(p(t),q(s,t))\in H$, будем понимать функцию $x=x(s,t)=x(s,t;u)\in H^1(Q)$, след которой при $t=0$ совпадает с $\varphi_0(s)$ и которая удовлетворяет интегральному тождеству

$$\iint_Q (a^2x_s\psi_s-x_t\psi_t-\psi q)dsdt+a^2\int_0^T\psi(0,t)p(t)dt-\int_0^l\psi(s,0)\varphi_1(s)ds=0$$

для всех функций $\psi=\psi(s,t)\in H^1(Q)$, след которых при $t=T$ равен нулю (см. обозначения в § 8.1).

Можно показать, что краевая задача (2), (3) при каждом $u\in H$ имеет, и притом единственное, обобщенное решение [492; 472]. Это решение может быть представлено с помощью формулы Даламбера в виде [698]:

$$x(s,t;u)=\frac{[\Phi_0(s+at)+\Phi_0(s-at)]}{2}+\frac{1}{2a}\int_{s-at}^{s+at}\Phi_1(\xi)d\xi+$$

$$+\frac{1}{2a}\int_0^t\int_{s-a(t-\tau)}^{s+a(t-\tau)}F_0(\xi,\tau)d\xi d\tau+a\sum_{m=1}^{\infty}P_1\Big(t-\frac{2ml-s}{a}\Big)+a\sum_{m=0}^{\infty}P_1\Big(t-\frac{2ml+s}{a}\Big), \tag{5}$$

где $\Phi_0(s),\Phi_1(s),F_0(s,t)$ — четные относительно $s=0$, $2l$-периодические по s продолжения функций $\varphi_0(s),\varphi_1(s),q(s,t)$ соответственно, а $P_1(t)=-\int_0^t p(\xi)d\xi$ при $t>0$, $P_1(t)=0$ при $t\leqslant 0$. Формулы для обобщенных производных x_s, x_t могут быть получены из (5) формальным дифференцированием по s и t соответственно. Из формулы (5) следует, что

$$\max_{0\leqslant t\leqslant T}\int_0^l[x^2(s,t;u)+x_t^2(s,t;u)]ds\leqslant$$

$$\leqslant C_0\Big(\int_0^T p^2(t)dt+\iint_Q q^2(s,t)dsdt+\int_0^l(\varphi_0^2(s)+\varphi_1^2(s))ds\Big), \tag{6}$$

где C_0 — постоянная, не зависящая от $(p,q,\varphi_0,\varphi_1)$ [441; 492].

Считая, что в (2), (3) $\varphi_0(s)=0$, $\varphi_1(s)=0$, введем оператор

$$Au=(x(s,T;u),x_t(s,T;u)),\quad 0\leqslant s\leqslant l, \tag{7}$$

действующий из гильбертова пространства $H=L_2[0,T]\times L_2(Q)$ в гильбертово пространство $F=L_2[0,l]\times L_2[0,l]$. Легко убедиться, что A — линейный оператор. Из оценки (6) следует ограниченность этого оператора, так что $A\in\mathcal{L}(H\to F)$. Это значит, что задачу (1)–(4) при $\varphi_0=\varphi_1=0$ можно записать в виде задачи (2.3):

$$J(u)=\|Au-b\|_F^2\to\inf,\quad u\in U, \tag{8}$$

где $b=(y_0(s),y_1(s))\in F$. В общем случае, когда $(\varphi_0(s),\varphi_1(s))\neq 0$, решение задачи (2), (3) можно представить в виде

$$x(s,t;u)=x_1(s,t;u)+x_0(s,t),\quad (s,t)\in Q, \tag{9}$$

где $x_1(s,t;u)$ — решение задачи (2), (3) при $\varphi_0=\varphi_1=0$, а $x_0(s,t)$ — решение той же задачи при $p(t)\equiv 0$, $q(s,t)\equiv 0$. Следовательно, и в общем случае

задачу (1)–(4) можно записать в виде задачи (8), где с учетом (9) надо принять

$$Au = (x_1(s, T; u), x_{1t}(s, T; u)), \quad b = (y_0(s) - x_0(s, T), y_1(s) - x_{0t}(s, T)).$$

Если U — выпуклое замкнутое ограниченное множество из H, то, как следует из теоремы 2.12, задача (1)–(4) имеет хотя бы одно решение. Пользуясь представлением этой задачи в форме (8) и результатами примера 3.5, убедимся, что функция (1) дважды непрерывно дифференцируема и получим формулы для ее производных. С этой целью покажем, что оператор A^*, сопряженный к оператору (7), на каждый элемент $c = (c_0(s), c_1(s)) \in F$ действует по правилу

$$A^*c = (a^2\psi(0, t; c); \psi(s, t; u)) \in H, \tag{10}$$

где $\psi = \psi(s, t; c)$ — решение краевой задачи

$$\psi_{tt} = a^2\psi_{ss}, \quad (s, t) \in Q, \tag{11}$$

$$\psi_s|_{s=0} = 0, \quad \psi_s|_{s=l} = 0, \quad 0 < t < T;$$
$$\psi|_{t=T} = c_1(s), \quad \psi_t|_{t=T} = -c_0(s), \quad 0 \leqslant s \leqslant l. \tag{12}$$

Краевая задача (11), (12) заменой $\tau = T - t$ формально может быть сведена к задаче вида (2), (3), однако функции $c_0(s)$, $c_1(s)$ в (12) менее гладкие, чем $\varphi_0(s)$, $\varphi_1(s)$ в (2), (3), поэтому здесь можно ожидать существование лишь более слабого решения, чем в задаче (2), (3).

Определение 2. Под обобщенным решением краевой задачи (11), (12) будем понимать функцию $\psi = \psi(s, t) = \psi(s, t; c) \in L_2(Q)$, имеющую следы $\psi(s, \cdot) \in L_2[0, T]$, $\psi(\cdot, t) \in L_2[0, l]$ при всех $s \in [0, l]$, $t \in [0, T]$ и удовлетворяющую интегральному тождеству

$$\iint_Q (\Phi_{tt} - a^2\Phi_{ss})\psi ds dt = -\int_0^l (c_0(s)\Phi(s, T) - c_1(s)\Phi_t(s, T))ds,$$

справедливому при всех $\Phi = \Phi(s, t) \in H^2(Q)$ со следом $\Phi|_{t=0} = \Phi_t|_{t=0} = \Phi_s|_{s=0} = \Phi_s|_{s=l} = 0$.

Решение задачи (11), (12) также может быть представлено с помощью формулы Даламбера вида (5), откуда, кстати, и следует его существование.

Из (2), (3) при $\varphi = 0$ и (11), (12) имеем

$$\langle Au, c\rangle_F = \int_0^l (x(s, T; u)c_0(s) + x_t(s, T; u)c_1(s))ds =$$
$$= \int_0^l (-x\psi_t + x_t\psi)|_{t=T} = \int_0^l \left(\int_0^T \frac{\partial}{\partial t}(-\psi_t x + \psi x_t)dt\right) ds =$$
$$= \iint_Q (-\psi_{tt}x + \psi x_{tt})ds dt = \iint_Q [a^2(-\psi_{ss}x + \psi x_{ss}) + \psi q]ds dt =$$
$$= \iint_Q [a^2\frac{\partial}{\partial s}(-\psi_s x + \psi x_s) + \psi q]ds dt = \int_0^T a^2(-\psi_s x + \psi x_s)|_{s=0}^l dt +$$
$$+ \iint_Q \psi q ds dt = \int_0^T a^2\psi(0, t)p(t)dt + \iint_Q \psi q ds dt = \langle u, A^*c\rangle_H.$$

С учетом замечания 7.1 можем считать, что формула (10) доказана.

Из формул $J'(u) = 2A^*(Au - f)$, $J'' = 2A^*A$ и из (7)–(12) следует, что функция (1) дважды непрерывно дифференцируема и ее градиент равен

$$J'(u) = (a^2\psi(0, t; c), \psi(s, t; u)), \tag{13}$$

где $\psi = \psi(s, t; c)$ решение задачи (11), (12) при $c = (c_0(s), c_1(s))$, $c_0(s) = 2(x(s, T; u) - y_0(s))$, $c_1(s) = 2(x_t(s, T; u) - y_1(s))$, $0 \leqslant s \leqslant l$.

Так как функция (1) выпукла на H, то согласно теореме 3.3 эта функция на выпуклом множестве $U \subseteq H$ будет достигать своей нижней грани в точке $u_* = (p_*(t), q_*(s, t)) \in U$ тогда и только тогда, когда

$$\langle J'(u_*), u - u_*\rangle_H = \int_0^T a^2\psi(0, t, u_*)(p(t) - p_*(t))dt +$$
$$+ \iint_Q \psi(s, t, u_*)(q(s, t) - q_*(s, t))ds dt \geqslant 0$$

при всех $u = (p(t), q(s, t)) \in U$.

Для решения задачи (1)–(4) могут быть использованы описанные выше методы минимизации. Кратко остановимся на методах проекции градиента и условного градиента, предполагая, что множество U состоит из управлений $u = (p(t), q(s, t)) \in H$, удовлетворяющих условиям

$$\int_0^T p^2(t)dt \leqslant R_0^2, \quad \iint_Q q^2(s, t)ds dt \leqslant R_1^2, \tag{14}$$

где R_0, R_1 — заданные положительные числа.

Метод проекции градиента для задачи (1)–(4), (14) с учетом формулы (13) сведется к построению последовательности $\{u_k = (p_k(t), q_k(s, t))\}$ по правилам

$$p_{k+1}(t) = \begin{cases} p_k(t) - \alpha_k a^2\psi(0, t, u_k) & \text{при} \\ \int_{t_0}^T |p_k(t) - \alpha_k a^2\psi(0, t, u_k)|^2 dt \leqslant R_0^2, & \\ \dfrac{R_0(p_k(t) - \alpha_k a^2\psi(0, t, u_k))}{\left(\int_0^T |p_k(t) - \alpha_k a^2\psi(0, t, u_k)|^2 dt\right)^{1/2}} & \text{при} \\ \int_{t_0}^T |p_k(t) - \alpha_k a^2\psi(0, t, u_k)|^2 dt > R_0^2, & \end{cases} \tag{15}$$

$$q_{k+1}(s, t) = \begin{cases} q_k(s, t) - \alpha_k\psi(s, t, u_k) & \text{при} \\ \iint_Q |q_k(s, t) - \alpha_k\psi(s, t, u_k)|^2 ds dt \leqslant R_1^2, & \\ \dfrac{R_1(q_k(s, t) - \alpha_k\psi(s, t, u_k))}{\left(\iint_Q |q_k(s, t) - \alpha_k\psi(s, t, u_k)|^2 ds dt\right)^{1/2}} & \text{при} \\ \iint_Q |q_k(s, t) - \alpha_k\psi(s, t, u_k)|^2 ds dt > R_1^2, & \end{cases}$$

где параметр $\alpha_k > 0$ выбирается одним из способов, описанных в § 5.2.

Одна итерация метода условного градиента для задачи (1)–(4), (14) будет выглядеть так:

$$\begin{cases} p_{k+1}(t) = p_k(t) + \alpha_k(\bar{p}_k(t) - p_k(t)), \\ q_{k+1}(s,t) = q_k(s,t) + \alpha_k(\bar{q}_k(s,t) - q_k(s,t)), \end{cases} \tag{16}$$

где

$$\bar{p}_k(t) = -\frac{R_0\psi(0,t,u_k)}{\left(\int\limits_0^T |\psi(0,t,u_k)|^2 dt\right)^{1/2}}, \quad \bar{q}_k(s,t) = -\frac{R_1\psi(s,t,u_k)}{\left(\iint\limits_Q |\psi(s,t,u_k)|^2 ds dt\right)^{1/2}}, \tag{17}$$

а величина α_k, $0 \leqslant \alpha_k \leqslant 1$, может быть выбрана одним из указанных в § 5.4 способов. Функция $f_k(\alpha) = J(u_k + \alpha(\bar{u}_k - u_k))$ переменной α представляет собой квадратный трехчлен, поэтому, рассуждая так же, как при выводе формулы (4.35), из условия $f_k(\alpha_k) = \min\limits_{0 \leqslant \alpha \leqslant 1} f_k(\alpha)$ можно определить $\alpha_k = \min\{1, \alpha_k^*\}$, где

$$\alpha_k^* = \frac{1}{2}\left[\int\limits_0^T a^2\psi(0,t,u_k)(p_k(t) - \bar{p}_k(t))dt + \iint\limits_Q \psi(s,t,u_k)(q_k(s,t) - \bar{q}_k(s,t))ds dt\right] \times$$

$$\times \left[\int\limits_0^l (2|x(s,T,u_k) - x(s,T,\bar{u}_k)|^2 + 2|x_t(s,T,u_k) - x_t(s,T,\bar{u}_k)|^2)ds\right]^{-1}, \tag{18}$$

причем если выражение в первой или во второй квадратной скобке обращается в нуль, то $u_k = (p_k(t), q_k(s,t))$ — оптимальное управление в рассматриваемой задаче (1)–(4), (14).

2. Рассмотрим задачу

$$J(u) = \int\limits_0^l (|x(s,T;u) - y_0(s)|^2 + |x_t(s,T;u) - y_1(s)|^2)ds \to \inf, \tag{19}$$

где $x = x(s,t;u)$ — решение краевой задачи

$$x_{tt} = x_{ss} + r(s)u(t), \quad (s,t) \in Q = (0,l) \times (0,T),$$
$$x|_{s=0} = 0, \quad x|_{s=l} = 0, \quad 0 < t < T; \quad x|_{t=0} = 0, \quad x_t|_{t=0} = 0, \quad 0 < s < l, \tag{20}$$

управление

$$u = u(t) \in U \subseteq H = L_2[0,T]; \tag{21}$$

функции $y_0(s) \in H_0^1[0,l]$, $y_1(s) \in L_2[0,l]$, $r(s) \in L_2[0,l]$ — заданы.

Определение 3. Обобщенным решением краевой задачи (20), соответствующим управлению $u = u(t) \in L_2[0,T]$, будем называть функцию $x = x(s,t;u) \in H^1(Q)$, имеющую следы $x(\cdot,t) \in H_0^1[0,l]$, $x_t(\cdot,t) \in L_2[0,l]$ при всех $t \in [0,T]$, $x(s,\cdot), x_s(s,\cdot) \in L_2[0,T]$ при всех $s \in [0,l]$ и удовлетворяющую интегральному тождеству

$$\iint\limits_Q (x_s\psi_s - x_t\psi_t - r(s)u(t)\psi)ds dt + \int\limits_0^l \psi(s,T)x_t(s,T)ds = 0$$

при всех $\psi = \psi(s,t) \in H^1(Q)$, причем след $x|_{t=0} = 0$, $0 \leqslant s \leqslant l$.

Можно показать [441; 492], что задача (20) при каждом $u = u(t) \in L_2[0,T]$ имеет, притом единственное, решение $x = x(s,t;u)$, и справедлива оценка:

$$\max_{0 \leqslant t \leqslant T} \int\limits_0^l (x^2(s,t;u) + x_s^2(s,t;u) + x_t^2(s,t;u))ds \leqslant c_0 \int\limits_0^T u^2(t)dt, \quad c_0 = \text{const} > 0. \tag{22}$$

Решение задачи (20) можно записать в виде формулы Даламбера, аналогичной формуле (5) [698]. Приведем представление решения в виде ряда Фурье:

$$x(s,t;u) = \sum_{k=1}^{\infty} \frac{r_k e_k(s)}{\sqrt{\lambda_k}} \int\limits_0^t u(\xi)\sin\sqrt{\lambda_k}(t-\xi)d\xi, \tag{23}$$

где $e_k(s) = \sqrt{\frac{2}{l}}\sin\sqrt{\lambda_k}s$ — собственная функция оператора $-\frac{d^2\varphi}{ds^2}$, $\varphi(0) = 0$, $\varphi(l) = 0$, соответствующая собственному числу $\lambda_k = \left(\frac{\pi k}{l}\right)^2$, $r_k = \int\limits_0^l r(s)e_k(s)ds$, $k = 1, 2, \ldots$ Более тонкий анализ показывает [458; 557], что решение задачи (20) имеет производную y_{tt}, представляющую собой производную функции y_t в смысле обобщенной функции (распределения) и имеющую следы из $H^{-1}[0,l]$ почти для всех $t \in [0,l]$.

Задачу (19)–(21) можно записать в форме (8), где оператор A внешне определяется той же формулой (7), но теперь A действует из $H = L_2[0,T]$ в $F = L_2[0,l] \times L_2[0,l]$, $f = (y_0(s), y_1(s)) \in F$. Как и в случае операторов (7), (10), можно показать, что здесь сопряженный оператор A^* имеет вид

$$A^*c = \int\limits_0^l r(s)\psi(s,t;c)ds, \tag{24}$$

где $\psi = \psi(s,t;c)$ — решение краевой задачи

$$\psi_{tt} = \psi_{ss}, \quad (s,t) \in Q;$$
$$\psi|_{s=0} = 0, \quad \psi|_{s=l} = 0, \quad 0 < t < T; \tag{25}$$
$$\psi|_{t=T} = c_1(s), \quad \psi_t|_{t=T} = -c_0(s), \quad 0 \leqslant s \leqslant l, \quad c = (c_0(s), c_1(s)) \in F.$$

Обобщенное решение задачи (25) понимается в том же смысле, как в определении 2, и его также можно записать в виде формулы Даламбера [698]. Функция (19) принадлежит классу $C^2(H)$, и ее градиент равен

$$J'(u) = A^*c = \int\limits_0^l r(s)\psi(s,t;c)ds,$$

где $\psi(s,t;c)$ — решение задачи (25) при $c_0(s) = 2(x(s,T;u) - y_0(s))$, $c_1(s) = 2(x_t(s,T;u) - y_1(s))$. Если U — выпуклое замкнутое ограниченное множество, то задача (19)–(21) имеет хотя бы одно решение (теоремы 2.6 и 2.12). Критерий оптимальности на выпуклом множестве U в этой задаче (теорема 3.3) запишется в форме

$$\langle J'(u_*), u - u_*\rangle_H = \iint\limits_Q r(s)\psi(s,t;c_*)(u(t) - u_*(t))ds \geqslant 0 \quad \forall u \in U,$$

где $c_* = (2(x(s,T;u_*) - y_0(s)), 2(x_t(s,T;u_*) - y_1(s)))$. Предлагаем читателю,

действуя по аналогии с задачей (1)–(4), дать описание методов проекции градиента и условного градиента для задачи (19)–(21), когда

$$U = \{u \in L_2[0, T]: \int_0^T u^2(t)dt \leqslant R^2\}.$$

3. Пусть дан однородный упругий стержень, один конец которого жестко закреплен, другой конец свободен. Требуется, управляя внешней поперечной нагрузкой, привести стержень к заданному моменту времени как можно ближе к заданному состоянию. Эту задачу математически можно сформулировать в виде следующей задачи минимизации:

$$J(u) = \int_0^l (|x(s, T, u) - y_0(s)|^2 + |x_t(s, T, u) - y_1(s)|^2)ds \to \inf, \quad (26)$$

$$x_{tt} + a^2 x_{ssss} = u(s, t), \quad (s, t) \in Q = \{0 < s < l,\ 0 < t \leqslant T\}, \quad (27)$$

$$x|_{s=0} = x_s|_{s=0} = 0, \quad 0 < t \leqslant T, \quad (28)$$

$$x_{ss}|_{s=l} = x_{sss}|_{s=l} = 0, \quad 0 < t \leqslant T, \quad (29)$$

$$x|_{t=0} = \varphi_0(s), \quad x_t|_{t=0} = \varphi_1(s), \quad 0 \leqslant s \leqslant l, \quad (30)$$

$$u = u(s, t) \in U = \{u(s, t) \in L_2(Q): \iint_Q u^2(s, t)dsdt \leqslant R^2\}, \quad (31)$$

где a^2, l, T, R — заданные положительные числа; $\varphi_i(s), y_i(s)$, $i = 1, 2$, — заданные функции; $\varphi_0(s) \in H^2[0, l]$, $\varphi_0(0) = \varphi_0'(0) = 0$; $\varphi_1(s), y_0(s), y_1(s) \in L_2[0, l]$.

Определение 4. Обобщенным решением краевой задачи (20)–(23), соответствующим управлению $u = u(s, t) \in L_2(Q)$, будем называть функцию $x(s, t) = x(s, t, u) \in H^{2,1}(Q)$, имеющую следы $x(\cdot, t), x_t(\cdot, t) \in L_2(0, l]$ при всех $t \in [0, T]$, $x(s, \cdot), x_s(s, \cdot) \in L_2[0, l]$ при всех $s \in [0, l]$ и удовлетворяющую условиям (28), (30) в смысле равенства соответствующих следов и интегральному тождеству

$$\iint_Q (-x_t\psi_t + a^2 x_{ss}\psi_{ss} - u\psi)dsdt - \int_0^l \varphi_1(s)\psi(s, 0)ds - \int_0^l x_t(s, T)\psi(s, T)ds = 0$$

при всех $\psi = \psi(s, t) \in H^{2,1}(Q)$, $\psi|_{s=0} = \psi_s|_{s=0} = 0$, $0 \leqslant s \leqslant T$. Можно показать [458], что при каждом $u = u(s, t)$ решение задачи (27)–(30) существует и единственно.

Убедимся, что функция (26) дифференцируема на $L_2(Q)$. Проиллюстрируем, как можно получить формулу градиента, пользуясь лишь определением 8.3.3 градиента, явно не прибегая к задаче (8). Возьмем произвольные $u, u + h \in L_2(Q)$ и соответствующие им решения $x(s, t, u)$, $x(s, t, u + h)$ краевой задачи (27)–(30).

Обозначим $\Delta x(s, t) = x(s, t, u + h) - x(s, t, u)$. Из (27)–(30) следует, что $\Delta x(s, t)$ является решением краевой задачи

$$\Delta x_{tt} + a^2 \Delta x_{ssss} = h(s, t), \quad (s, t) \in Q, \quad (32)$$

$$\Delta x|_{s=0} = \Delta x_s|_{s=0} = 0, \quad \Delta x_{ss}|_{s=l} = \Delta x_{sss}|_{s=l} = 0, \quad 0 < t \leqslant T, \quad (33)$$

$$\Delta x|_{t=0} = \Delta x_t|_{t=0} = 0, \quad 0 \leqslant s \leqslant l. \quad (34)$$

Тогда приращение функции (26) запишется в виде

$$\Delta J(u) = J(u + h) - J(u) = \int_0^l [2(x(s, T, u) - y_0(s))\Delta x(s, T) + 2(x_t(s, T, u) - y_1(s))\Delta x_t(s, T)]ds + R, \quad (35)$$

где

$$R = \int_0^l (|\Delta x(s, T)|^2 + |\Delta x_t(s, T)|^2)ds.$$

Справедлива оценка

$$|R| \leqslant C_1 \|h\|^2_{L_2} = C_1 \iint_Q |h(s, t)|^2 dsdt, \quad C_1 = \text{const} \geqslant 0. \quad (36)$$

Наметим схему доказательства этой оценки. Умножим уравнение (32) на $\Delta x_t(s, t)$, проинтегрируем по прямоугольнику $Q_t = \{(s, \tau): 0 \leqslant s \leqslant l,\ 0 \leqslant \tau \leqslant t\}$ и получившееся равенство преобразуем с учетом условий (33), (34); будем иметь

$$\iint_{Q_t} h\Delta x_t dsd\tau = \iint_{Q_t} (\Delta x_{tt} + a^2 \Delta x_{ssss})\Delta x_t dsd\tau = \frac{1}{2}\int_0^l (\Delta x_t)^2|_{\tau=0}^t ds +$$

$$+ \int_0^t a^2 \Delta x_{sss}\Delta x_t|_{s=0}^l d\tau - \iint_{Q_t} a^2 \Delta x_{sss}\Delta x_{ts} dsdt = \frac{1}{2}\int_0^l |\Delta x_t(s, t)|^2 ds -$$

$$- \int_0^t a^2 \Delta x_{ss}\Delta x_{ts}|_{s=0}^l dt + \iint_{Q_t} a^2 \Delta x_{ss}\Delta x_{sst} dsdt = \frac{1}{2}\int_0^l |\Delta x_t(s, t)|^2 ds +$$

$$+ \frac{a^2}{2}\int_0^l (\Delta x_{ss})^2|_{\tau=0}^t ds = \frac{1}{2}\int_0^l |\Delta x_t(s, t)|^2 ds + \frac{a^2}{2}\int_0^l |\Delta x_{ss}(s, t)|^2 ds$$

при всех t, $0 < t \leqslant T$. Отсюда с помощью неравенства $|ab| \leqslant (a^2 + b^2)/2$ получим

$$\int_0^l |\Delta x_t(s, t)|^2 ds \leqslant \int_0^t \left(\int_0^l |\Delta x_t(s, \tau)|^2 ds\right) d\tau + \int_0^T \int_0^l h^2(s, \tau)dsd\tau, \quad 0 < t \leqslant T.$$

Тогда из леммы 6.3.1 при $\varphi(t) = \int_0^l |\Delta x_t(s, t)|^2 ds$, $b = \iint_Q h^2(s, t)dsdt$, $a = 1$ следует, что

$$\int_0^l |\Delta x_t(s, t)|^2 ds \leqslant e^T \iint_Q h^2(s, t)dsdt, \quad 0 < t \leqslant T. \quad (37)$$

В частности, при $t = T$ имеем

$$\int_0^l |\Delta x_t(s, T)|^2 ds \leqslant e^T \iint_Q h^2(s, t)dsdt. \quad (38)$$

Далее, из равенства

$$\Delta x(s, T) = \Delta x(s, 0) + \int_0^T \Delta x_t(s, t)dt$$

с учетом первого условия (34) и оценки (37) получим

$$\int_0^l |\Delta x(s,T)|^2 ds = \int_0^l \Big(\int_0^T \Delta x_t(s,t)dt\Big)^2 ds \leqslant$$
$$\leqslant T\int_0^T \Big(\int_0^l |\Delta x_t(s,t)|^2 ds\Big)^2 dt \leqslant T^2 e^T \iint_Q h^2(s,t)dsdt.$$

Сложив эту оценку с (38), придем к оценке (36).

Для преобразования правой части формулы приращения (35) введем сопряженную краевую задачу

$$\psi_{tt} + a^2\psi_{ssss} = 0, \quad (s,t)\in Q,$$
$$\psi|_{s=0} = \psi_s|_{s=0} = 0, \quad \psi_{ss}|_{s=l} = \psi_{sss}|_{s=l} = 0, \quad 0\leqslant t\leqslant T, \tag{39}$$
$$\psi|_{t=T} = c_1(s) = 2(x_t(s,T) - y_1(s)), \quad \psi_t|_{t=T} = -c_0(s) = -2(x(s,T) - y_0(s)),$$
$$0\leqslant s\leqslant l.$$

Под решением задачи (39) здесь понимается функция $\psi(s,t,u) = \psi(s,t)\in L_2(Q)$, удовлетворяющая интегральному тождеству

$$\iint_Q (\Phi_{tt} + a^2\Phi_{ssss})\psi dsdt = 2\int_0^l (x(s,T) - y_0(s))\Phi(s,T)ds +$$
$$+2\int_0^l (x_t(s,T) - y_1(s))\Phi_t(s,T)ds$$

для всех функций $\Phi = \Phi(s,t)\in H^{4,2}(Q)$, обладающих обобщенными производными Φ_{xt}, $\Phi_{xxt}\in L_2(Q)$ и таких, что $\Phi|_{s=0} = \Phi_s|_{s=0} = \Phi_{ss}|_{s=l} = \Phi_{sss}|_{s=l} = 0$ при $0\leqslant t\leqslant T$ и $\Phi|_{t=0} = \Phi_t|_{t=0} = 0$ при $0\leqslant s\leqslant l$.

С помощью решения $\psi(s,t,u)$ краевой задачи (39) приращение функции (26) можно представить в виде

$$\Delta J(u) = \iint_Q \psi(s,t,u)h(s,t)dsdt + R. \tag{40}$$

В самом деле, из (35) с учетом условий (32)–(34), (39) имеем

$$\Delta J(u) = \int_0^l (-\psi_t(s,T)\Delta x(s,T) + \psi(s,T)\Delta x_t(s,T))ds + R =$$
$$= \int_0^l\int_0^T \frac{\partial}{\partial t}(-\psi_t\Delta x + \psi\Delta x_t)dtds + R = \iint_Q (-\psi_{tt}\Delta x + \psi\Delta x_{tt})dtds + R =$$
$$= \iint_Q a^2(\psi_{ssss}\Delta x - \psi\Delta x_{ssss})dsdt + \iint_Q \psi h dsdt + R =$$
$$= a^2\int_0^T (\psi_{sss}\Delta x - \psi_{ss}\Delta x_s + \psi_s\Delta x_{ss} - \psi\Delta x_{sss})\Big|_{s=0}^{s=l} dt +$$
$$+ \iint_Q \psi h dsdt + R = \iint_Q \psi(s,t,u)h(s,t)dsdt + R.$$

Формула (40) получена. Из (40) и оценки (36) следует, что функция (26) дифференцируема на $L_2(Q)$, и ее градиент равен

$$J'(u) = \psi(s,t,u), \quad (s,t)\in Q. \tag{41}$$

Конечно, как и в задаче (1)–(4), приведенный выше вывод формулы (40) и оценки (36) нельзя признать строгим; относительно строгого доказательства формулы (41) см. замечание 7.1. Можно показать, что функция (26) выпукла на $L_2(Q)$ и принадлежит $C^{1,1}(L_2)$. Отсюда и из теоремы 2.8 следует, что задача (26)–(31) имеет хотя бы одно решение. Согласно теореме 3.3 для оптимальности управления $u_* = u_*(s,t)\in U$ необходимо и достаточно, чтобы $\iint_Q \psi(s,t,u_*)(u(s,t) - u_*(s,t))dsdt \geqslant 0$ при всех $u = u(s,t)\in U$.

Предлагаем читателю самостоятельно написать итерации методов проекции градиента и условного градиента для задачи (26)–(31) и для $(k+1)$-го приближения получить формулы, аналогичные формулам (15)–(18).

Упражнения

1. Показать, что функция $J_1(u) = J(u) + \beta\int_0^T p^2(t)dt + \beta\iint_Q q^2(s,t)dsdt$, $\beta = \text{const} > 0$, где $J(u)$ взята из (1), при условиях (2), (3) сильно выпукла на $H = L_2[0,l]\times L_2(Q)$. Описать метод скорейшего спуска для задачи минимизации $J_1(u)$ на всем пространстве H.

2. Показать, что функция $J_1(u) = J(u) + \beta\iint_Q u^2(s,t)dsdt$, $\beta > 0$, где $J(u)$ определяется формулой (26), при условиях (27)–(30) сильно выпукла на $L_2(Q)$. Описать метод скорейшего спуска для минимизации $J_1(u)$ на $L_2(Q)$.

3. Найти градиент функций (1) и (26) по начальным условиям (φ_0, φ_1).

4. Пусть в задачах (1)–(4), (19)–(21), (26)–(31), имеются дополнительные ограничения $|x(s,t,u)|\leqslant\gamma_0$, $|x_t(s,t,u)|\leqslant\gamma_2$, $(s,t)\in Q$. Учесть эти ограничения с помощью штрафных функций, вывести формулу градиента для штрафной функции; описать метод штрафных функций в сочетании с методом проекции градиента или условного градиента.

5. Пусть требуется минимизировать функцию $J(u) = \int_0^T p^2(t)dt + \iint_Q q^2(s,t)dsdt$ при условиях (2)–(4), (14) или функцию $J(u) = \iint_Q u^2(s,t)dsdt$ при условиях (27)–(31) и дополнительных ограничениях $x(s,T,u) = 0$, $x_t(s,T,u) = 0$, $0\leqslant s\leqslant l$, где $T > 0$ — заданное время; учесть дополнительные ограничения с помощью штрафных функций; найти градиент штрафной функции.

6. Получите формулу (41), пользуясь результатами примера 3.5. Указание: запишите задачу (26)–(31) в форме (8), где A — оператор (7), действующий из пространства $H = L_2(Q)$ в пространство $F = L_2[0,l]\times L_2[0,l]$ при $(\varphi_0,\varphi_1) = 0$, покажите, что сопряженный к A оператор A^* действует из F в H по правилу: $A^*c = \psi(s,t;c)$, где $c = (c_0,c_1)$, $\psi = \psi(s,t;c)$ — решение краевой задачи (39); в случае $(\varphi_0,\varphi_1)\neq 0$ воспользуйтесь аналогичным (9) представлением решения задачи (27)–(30).

§ 9. Оптимальное управление процессами, описываемыми уравнением Гурса — Дарбу

При исследовании процессов сорбции, сушки и др. возникает следующая задача оптимального управления [51; 104; 135; 136; 268; 540; 575; 664; 698]: минимизировать функцию

$$J(u) = \iint_Q f^0(x(s,t), x_s(s,t), x_t(s,t), u(s,t), s, t)dsdt + \Phi(x(l,T)) \tag{1}$$

при условиях

$$x_{st}(s,t) = f(x(s,t), x_s(s,t), x_t(s,t), u(s,t), s, t), \quad (s,t)\in Q, \tag{2}$$
$$x(0,t) = \alpha(t), \quad 0\leqslant t\leqslant T; \quad x(s,0) = \beta(s), \quad 0\leqslant s\leqslant l, \tag{3}$$
$$u = u(s,t)\in U\subseteq L_2^r(Q), \tag{4}$$

где $x=(x^1,\ldots,x^n)$, $f=(f^1,\ldots,f^n)$, $u=(u^1,\ldots,u^r)$, $\alpha=(\alpha^1,\ldots,\alpha^n)$, $\beta=(\beta^1,\ldots,\beta^n)$, $Q=\{(s,t)\colon 0\leqslant s\leqslant l,\ 0\leqslant t\leqslant T\}$; l, T — заданные положительные числа, $f^i(x,p,q,u,s,t)$, $i=0,\ldots,n$, $\Phi(x)$, $\alpha^i(t)$, $\beta^i(s)$, $i=1,\ldots,n$, — заданные функции, U — заданное множество.

Эту задачу будем рассматривать при выполнении следующих условий:

1) функции $f^i(x,p,q,u,s,t)$, $i=0,\ldots,n$, и их частные производные $f^i_x, f^i_p, f^i_q, f^i_u$ непрерывны по совокупности аргументов $(x,p,q,u,s,t)\in E^n\times E^n\times E^n\times E^r\times[0,l]\times[0,T]$ и удовлетворяют условию Липшица по (x,p,q,u);

2) функция $\Phi(x)$ обладает непрерывными частными производными $\Phi_x(x)$ при всех $x\in E^n$;

3) $\alpha(t)\in H^1_n[0,T]$, $\beta(s)\in H^1_n[0,l]$; $\alpha(0)=\beta(0)$.

Под решением задачи (2), (3), соответствующим управлению $u=u(s,t)\in L^r_2(Q)$, будем понимать вектор-функцию $x(s,t)=x(s,t,u)\in L^n_2(Q)$, имеющую обобщенные производные $x_s(s,t)$, $x_t(s,t)$, $x_{st}(s,t)\in L^n_2(Q)$ и удовлетворяющую уравнению (2) почти всюду в Q, а условиям (3) — в смысле равенства соответствующих следов $x(0,\cdot)$, $x(\cdot,0)$.

При сделанных выше предположениях задача (2), (3) при любом $u=u(s,t)\in L^r_2(Q)$ имеет, и притом единственное, решение. Важно заметить, что любая вектор-функция $x(s,t)\in L^n_2(Q)$, обладающая обобщенными производными $x_s(s,t)$, $x_t(s,t)$, $x_{st}(s,t)\in L^n_2(Q)$, непрерывна в замкнутом прямоугольнике Q (точнее, $x(s,t)$ эквивалентна непрерывной на Q функции). Это значит, что решение $x(s,t,u)$ задачи (2), (3) можем считать непрерывной функцией Q, и тогда имеет смысл говорить о значении $x(l,T,u)$, о величине $\Phi(x(l,T,u))$. Таким образом, при сделанных выше предположениях функция (1) определена при всех $u=u(s,t)\in L^r_2(Q)$.

Можно показать, что непрерывная на Q вектор-функция $x(s,t)$ является решением краевой задачи (2), (3) тогда и только тогда, когда она удовлетворяет интегральному уравнению

$$x(s,t)=\alpha(t)+\beta(s)-\alpha(0)+\int_0^s\int_0^t f(x(\xi,\tau),x_s(\xi,\tau),x_t(\xi,\tau),u(\xi,\tau),\xi,\tau)\,d\xi\,d\tau. \quad (5)$$

Опираясь на это интегральное уравнение, существование и единственность решения задачи (2), (3) могут быть доказаны с помощью рассуждений, аналогичных рассуждениям из доказательства теоремы 6.1.1.

Покажем, что при некоторых сделанных дополнительных предположениях функция (1) дифференцируема на $L^r_2(Q)$. Возьмем произвольные u, $u+h\in L^r_2(Q)$ и соответствующие им решения $x(s,t,u)$, $x(s,t,u+h)$ задачи (2), (3). Тогда обозначим

$$\Delta x(s,t)=x(s,t,u+h)-x(s,t,u),$$

$$\Delta f^i=f^i(x(s,t,u+h),x_s(s,t,u+h),x_t(s,t,u+h),u(s,t)+h(s,t),s,t)-$$
$$-f^i(x(s,t),x_s(s,t),x_t(s,t),u(s,t),s,t),\quad i=0,\ldots,n.$$

Из условий (2), (3) следует, что

$$\Delta x_{st}=\Delta f,\quad (s,t)\in Q, \quad (6)$$

$$\Delta x(0,t)=0,\quad 0\leqslant t\leqslant T,\quad \Delta x(s,0)=0,\quad 0\leqslant s\leqslant l. \quad (7)$$

Можно доказать, что верна следующая оценка:

$$\max_Q|\Delta x(s,t)|+\operatorname{ess}\sup_{0\leqslant t\leqslant T}\int_0^l|\Delta x_s(s,t)|^2ds+$$
$$+\operatorname{ess}\sup_{0\leqslant s\leqslant l}\int_0^T|\Delta x_t(s,t)|^2dt\leqslant C_0\iint_Q h^2(s,t)\,ds\,dt,\quad C_0=\text{const}>0. \quad (8)$$

Приращение функции (1) равно

$$\Delta J(u)=J(u+h)-J(u)=\iint_Q\Delta f^0\,ds\,dt+\Delta\Phi, \quad (9)$$

где

$$\Delta\Phi=\Phi(x(l,T,u+h))-\Phi(x(l,T,u)).$$

Умножим уравнение (6) на некоторую функцию $\psi=\psi(s,t)\in L^n_2(Q)$ и проинтегрируем полученное равенство по прямоугольнику Q. Будем иметь

$$0=\iint_Q(\langle\psi,-\Delta x_{st}\rangle+\langle\psi,\Delta f\rangle)\,ds\,dt.$$

Сложим это равенство почленно с (9) и получим

$$\Delta J(u)=\iint_Q(\Delta f^0+\langle\Delta f,\psi\rangle-\langle\psi,\Delta x_{st}\rangle)\,ds\,dt+\Delta\Phi.$$

Если ввести функцию Гамильтона — Понтрягина

$$H(x,p,q,u,s,t,\psi)=f^0(x,p,q,u,s,t)+\langle f(x,p,q,u,s,t),\psi\rangle, \quad (10)$$

то приращение $\Delta J(u)$ можно переписать в виде

$$\Delta J(u)=\iint_Q(\Delta H-\langle\psi,\Delta x_{st}\rangle)\,ds\,dt+\Delta\Phi, \quad (11)$$

где $\Delta H=\Delta H(s,t)=H(x+\Delta x,x_s+\Delta x_s,x_t+\Delta x_t,u+h,s,t,\psi)-H(x,x_s,x_t,u,s,t,\psi)$; аргументы (s,t) функций x, Δx и их производных, u, h, ψ для краткости здесь опущены.

Учитывая ограничения, наложенные выше на функции f^i, $i=0,\ldots,n$, $\Phi(x)$, $\alpha(t)$, $\beta(t)$, с помощью формулы конечных приращений Лагранжа и оценки (8) из (11) имеем

$$\Delta J(u)=\iint_Q(\langle H_x,\Delta x\rangle+\langle H_p,\Delta x_s\rangle+\langle H_q,\Delta x_t\rangle+\langle H_u,h\rangle-\langle\psi,\Delta x_{st}\rangle)\,ds\,dt+$$
$$+\langle\Phi_x(x(l,T)),\Delta x(l,T)\rangle+R, \quad (12)$$

где остаточный член R удовлетворяет условию

$$|R|\,\|h\|^{-1}_{L_2}\to 0\quad\text{при}\quad\|h\|_{L_2}\to 0; \quad (13)$$

частные производные H_x, H_p, H_q, H_u в (12) вычислены в точке $(x(s,t),x_s(s,t),x_t(s,t),u(s,t),s,t,\psi(s,t))$.

Преобразуем первые три слагаемых в (12) интегрированием по частям с учетом условий (7). С помощью теоремы Фубини [393; 492] получим

$$\iint_Q\langle H_x,\Delta x\rangle\,ds\,dt=-\int_0^T\Big(\int_0^l\langle\frac{d}{ds}\int_s^l H_x(\xi,t)\,d\xi,\Delta x(s,t)\rangle\,ds\Big)\,dt=$$
$$=\iint_Q\langle\int_s^l H_x(\xi,t)\,d\xi,\Delta x_s(s,t)\rangle\,ds\,dt=\iint_Q\langle\int_t^T\int_s^l H_x(\xi,t)\,d\xi\,d\tau,\Delta x_{st}(s,t)\rangle\,ds\,dt;$$

$$\iint\limits_Q \langle H_p, \Delta x_s\rangle dsdt = \iint\limits_Q \langle \int\limits_t^T H_p(s,\tau)d\tau, \Delta x_{st}(s,t)\rangle dsdt,$$

$$\iint\limits_Q \langle H_q, \Delta x_t\rangle dsdt = \iint\limits_Q \langle \int\limits_s^l H_q(\xi,t)d\xi, \Delta x_{st}(s,t)\rangle dsdt.$$

Кроме того, в силу условий (7) имеем

$$\Delta x(l,T) = \int\limits_0^l \Delta x_s(\xi,T)d\xi = \int\limits_0^l \int\limits_0^T \Delta x_{st}(\xi,\tau)d\xi d\tau = \iint\limits_Q \Delta x_{st}dsdt,$$

поэтому

$$\langle \Phi_x(x(l,T)), \Delta x(l,T)\rangle = \iint\limits_Q \langle \Phi_x(x(l,T)), \Delta x_{st}(s,t)\rangle dsdt.$$

Подставим полученные равенства в формулу (12). Будем иметь

$$\Delta J(u) = \iint\limits_Q \langle -\psi(s,t) + \Phi_x(x(l,T)) + \int\limits_t^T H_p(s,\tau)d\tau + \int\limits_s^l H_q(\xi,t)d\xi +$$

$$+ \int\limits_t^T \int\limits_s^l H_x(\xi,\tau)d\xi d\tau, \Delta x_{st}(s,t)\rangle dsdt + \iint\limits_Q \langle H_u(s,t), h(s,t)\rangle dsdt + R. \quad (14)$$

До сих пор $\psi = \psi(s,t)$ была произвольной функцией из $L_2^n(Q)$. Теперь выберем эту функцию так, чтобы

$$\psi(s,t) = \Phi_x(x(l,T)) + \int\limits_t^T H_p(s,\tau)d\tau + \int\limits_s^l H_q(\xi,t)d\xi + \int\limits_t^T \int\limits_s^l H_x(\xi,\tau)d\xi d\tau, \quad (15)$$

$$(s,t) \in Q.$$

Так как в (15) $H_p(s,t), H_q(s,t), H_x(s,t)$ представляют собой частные производные H_p, H_q, H_x функции (10), вычисленные в точке $(x(s,t), x_s(s,t), x_t(s,t), u(s,t), s, t, \psi(s,t))$, причем $H(x,p,q,u,s,t,\psi)$ линейно зависит от переменной ψ, то (15) является линейным интегральным уравнением относительно $\psi(s,t)$. Уравнение (15) аналогично уравнению (5), и существование и единственность его решения $\psi(s,t) = \psi(s,t,u)$ при сделанных выше предположениях доказывается аналогично тому, как доказывалась теорема 6.1.1.

С учетом условия (15) из (14) имеем

$$\Delta J(u) = \iint\limits_Q \langle H_u(s,t), h(s,t)\rangle dsdt + R. \quad (16)$$

Отсюда и из условия (13) следует, что функция (1) дифференцируема и ее градиент равен

$$J'(u) = H_u(s,t) = H_u(x(s,t,u), x_s(s,t,u), x_t(s,t,u), u(s,t), s, t, \psi(s,t,u)). \quad (17)$$

Подчеркнем, что при выводе формулы (16) исходные функции f^0, f предполагались такими, что $H_u(s,t) \in L_2^r(Q)$.

Предлагаем читателю самостоятельно выписать, пользуясь формулой (17), необходимые условия оптимальности в задаче (1)–(4) для выпуклого множества U, сформулировать условия существования оптимального решения, условия выпуклости и сильной выпуклости функции (1), дать описание градиентного метода, методов проекции градиента и условного градиента.

Упражнения

1. Получить формулу градиента в задаче (1)–(3), считая, что $u(s,t) \equiv u(s) \in L_2^r[0,l]$ или $u(s,t) \equiv u(t) \in L_2^r[0,T]$, или $u(s,t) \equiv \omega \in E^r$.

2. Применить метод штрафных функций к задаче (1)–(4) при ограничениях $|x(s,t)| \leqslant 1$ или $|x^i(s,t)| \leqslant 1$, $i = 1,\ldots,n$, или $\iint\limits_Q |x(s,t)|^2 dsdt \leqslant 1$. Найти градиент штрафной функции.

3. Сформулировать и доказать принцип максимума для задачи (1)–(4), считая, что $U = \{u(s,t) \in L_2^r(Q)$: $u(s,t) \in V$ почти всюду на $Q\}$, где V — заданное множество из E^r [51].

4. Рассмотреть задачу [104; 119; 134; 135; 478]:

$$J(u) = \iint\limits_Q F^0(z(s,t;u), u(s,t), s, t)dsdt \to \inf, \quad u \in U, \quad (18)$$

где $z = z(s,t;u) = (z^1,\ldots,z^n)$ — решение задачи

$$z_s^i = F^i(z,u,s,t),\ i=1,\ldots,m;\ z_t^i = F^i(z,u,s,t),\ i=m+1,\ldots,n;\ (s,t)\in Q=(0,l)\times(0,T) \quad (19)$$

$$z^i(0,t) = \mu^i(t),\quad i=1,\ldots,m;\quad z^i(s,0) = \nu^i(s),\quad i=m+1,\ldots,n, \quad (20)$$

управление $u = u(s,t) = (u^1,\ldots,u^r) \in L_2^r(Q)$, U — заданное множество из $L_2^r(Q)$; $F^i(z,u,s,t)$, $i=0,\ldots,n$, $\mu^i(t)$, $i=1,\ldots,m$, $\nu^i(s)$, $i=m+1,\ldots,n$ — достаточно гладкие функции своих аргументов. Доказать, что функция (18) дифференцируема в $L_2^r(Q)$, найти формулу ее градиента; дать описание методов проекции градиента, условного градиента. Указание: показать, что задача (18)–(20) в пространстве переменных $x = (x^1,\ldots,x^n)$, $x^i(s,t) = \int\limits_0^t z^i(s,\tau)d\tau$, $i=1,\ldots,m$; $x^i(s,t) = \int\limits_0^s z^i(\xi,t)d\xi$, $i=m+1,\ldots,n$, запишется в виде задачи (1)–(4) с $f^i = F^i(x_t^1,\ldots,x_t^m, x_s^{m+1},\ldots,x_s^n)$, $i=0,\ldots,n$; $\alpha^i(t) = \int\limits_0^t \mu^i(\tau)d\tau$, $i=1,\ldots,m$, $\alpha^i(t)=0$, $i=m+1,\ldots,n$, $\beta^i(s)=0$, $i=1,\ldots,m$, $\beta^i(s) = \int\limits_0^s \nu^i(\xi)d\xi$, $i=m+1,\ldots,n$; воспользоваться формулами (15), (17) и в них вернуться к исходным переменным z.

5. Сформулировать и доказать принцип максимума для задачи (18)–(20) в случае множества U из упражнения 3 [134].

§ 10. Взаимодвойственные задачи управления и наблюдения

1. Под *задачей управления* в этом параграфе будем понимать задачу определения такого допустимого управления, которое переводит динамическую систему из некоторого начального состояния в заданное состояние к заданному моменту времени. Многие задачи управления можно записать в виде операторного уравнения

$$Au = f, \quad (1)$$

где A — линейный непрерывный оператор, действующий из гильбертова пространства H в гильбертово пространство F, т. е. $A \in \mathcal{L}(H \to F)$, элемент $f \in F$. Описание оператора A содержит информацию о начально-краевой задаче, характеризующей динамическую систему, элемент f связан с начальным или конечным состоянием системы. Поясним сказанное на примерах.

Пример 1. Пусть динамическая система описывается обыкновенным дифференциальным уравнением

$$\dot{x} = D(t)x + B(t)u(t), \quad 0 \leqslant t \leqslant T; \quad x(0) = 0, \quad (2)$$

где $D(t)$, $B(t)$ — кусочно-непрерывные матрицы размера $n \times n$, $n \times r$ соответственно, $u = u(t) = (u^1(t), \ldots, u^r(t)) \in L_2^r[0, T]$ — управление, $x = x(t) = x(t; u) = (x^1(t), \ldots, x^n(t))$ траектория системы (2), соответствующая управлению $u = u(t)$. Момент $T > 0$ задан. *Задача управления*: требуется найти управление $u = u(t) \in L_2^r[0, T]$ такое, что

$$x(T; u) = f, \tag{3}$$

где f — заданная точка из E^n. Введем оператор A следующим образом

$$Au = x(T; u). \tag{4}$$

Выше было показано (см. пример 2.14), что такой оператор $A \in \mathcal{L}(H \to F)$, где $H = L_2^r[0, T]$, $F = E^n$. Из (3), (4) следует, что задача управления (2), (3) может быть записана в виде уравнения (1).

Для более общей, чем (2), системы

$$\dot{y} = D(t)y + B(t)u(t) + f(t), \quad t_0 \leqslant t \leqslant T, \quad f(t) \in L_2^n[t_0, T], \tag{5}$$

задача определения управления $u = u(t)$, переводящего эту систему из начального состояния $y(t_0) = y_0$ в заданное конечное состояние $y(T) = y_1$, легко может быть сведена к задаче (2), (3). Прежде всего, начальный момент t_0 можем считать равным нулю, так как в противном случае сделаем замену $\tau = t - t_0$. Далее, любое решение $y = y(t; u)$ уравнения (5) представимо в виде (пример 2.15)

$$y(t; u) = x(t; u) + y_0(t), \quad 0 \leqslant t \leqslant T,$$

где $x(t; u)$ — решение системы (2), $y_0 = y_0(t)$ — решение системы (5) при $u(t) \equiv 0$. Отсюда следует, что задача перевода системы (5) из точки y_0 в точку y_1 равносильна задаче (2), (3) при $f = y_1 - y_0(T)$ к уравнению (1) с оператором (4).

Пример 2. Рассмотрим краевую задачу

$$\begin{gathered} x_t = x_{ss}, \quad (s, t) \in Q = (0, l) \times (0, T), \\ x_s|_{s=0} = 0, \quad (x_s + x)|_{s=l} = u(t), \quad 0 < t < T; \quad x|_{t=0} = 0, \quad 0 \leqslant s \leqslant l, \end{gathered} \tag{6}$$

описывающую температуру стержня (см. задачу (7.2), (7.3) при $q \equiv 0$, $\varphi \equiv 0$). Найдем управление $u = u(t) \in L_2[0, T]$ такое, чтобы в заданной момент T температура стержня имела заданное распределение $f(s) \in L_2[0, l]$, т. е.

$$x(s, T; u) = f(s), \quad 0 \leqslant s \leqslant l. \tag{7}$$

Введем оператор

$$Au = x(s, T; u), \quad 0 \leqslant s \leqslant l. \tag{8}$$

Как было отмечено в § 7, такой оператор $A \in \mathcal{L}(H \to F)$, где $H = L_2[0, l]$, $F = L_2[0, l]$. Отсюда следует, что искомое управление является решением уравнения (1), где оператор A определен согласно (8), а элемент $f = f(s)$ взят из (7).

Пример 3. Рассмотрим краевую задачу

$$\begin{gathered} x_{tt} = x_{ss}, \quad (s, t) \in Q = (0, l) \times (0, T), \\ x_s|_{s=0} = u(t), \quad x_s|_{s=l} = 0, \quad 0 < t < T; \quad x|_{t=0} = 0, \quad x_t|_{t=0} = 0, \quad 0 \leqslant s \leqslant l, \end{gathered} \tag{9}$$

описывающую колебания струны (см. задачу (8.2), (8.3) при $q = 0$, $\varphi_0 = 0$, $\varphi_1 = 0$). Будем искать управление $u = u(t) \in L_2[0, T]$ такое, чтобы в момент $T > 0$ состояние (смещение и скорость) струны совпало с заданным

$$x(s, T; u) = y_0(s), \quad x_t(s, T; u) = y_1(s), \quad 0 \leqslant s \leqslant l, \tag{10}$$

где $y_0(s)$, $y_1(s) \in L_2[0, l]$ — фиксированные функции. Введем оператор

$$Au = (x(s, T; u), x_t(s, T; u)), \tag{11}$$

который действует из $H = L_2[0, T]$ в $F = L_2[0, l] \times L_2[0, l]$. В § 8 было отмечено, что такой оператор $A \in \mathcal{L}(H \to F)$. Отсюда видно, что искомое управление является решением операторного уравнения (1), где A взят из (11), а $f = (y_0(s), y_1(s)) \in F$ — из (10).

Пример 4. Основываясь на краевой задаче (8.20):

$$\begin{gathered} x_{tt} = x_{ss} + r(s)u(t), \quad (s, t) \in Q = (0, l) \times (0, T), \\ x|_{s=0} = x|_{s=l} = 0, \quad 0 < t < T, \quad x|_{t=0} = x_t|_{t=0} = 0, \quad 0 \leqslant s \leqslant l, \end{gathered} \tag{12}$$

сформулируем еще одну задачу управления колебанием струны. Будем искать управление $u = u(t) \in L_2[0, T]$ такое, чтобы состояние струны в заданный момент T удовлетворяло условиям (10). Нетрудно видеть, что если ввести оператор A по той же формуле (11), задачу управления (12), (10) также можно записать в виде уравнения (1).

2. Предположим, что некоторая задача управления линейной динамической системой уже записана в виде операторного уравнения (1). Отождествляя задачу управления с уравнением (1), далее будем говорить о задаче управления системой A. Сформулируем такие важные понятия теории управления, как поточечная управляемость, полная управляемость в терминах уравнения (1).

Определение 1. Систему A будем называть *поточечно управляемой* или, иначе, *f-управляемой*, если уравнение (1) с фиксированной правой частью $f \in F$ имеет хотя бы одно решение. Скажем, что система A *вполне управляема*, если уравнение (1) имеет решение при всех $f \in F$.

Таким образом, если мы пожелаем узнать, будет ли та или иная линейная динамическая система поточечно или вполне управляемой, а также найти соответствующее управление, то можем обратиться к хорошо развитой теории линейных операторных уравнений, методам их решения [130; 334; 393; 416 и др.]

Приведем без доказательства несколько теорем из этой теории, которые могут быть полезны при исследовании конкретных задач управления. При формулировке этих теорем будем пользоваться обозначениями: $R(A)$, $R(A^*)$ — область значений оператора $A \in \mathcal{L}(H \to F)$ и сопряженного оператора $A^* \in \mathcal{L}(F \to H)$, где H, F — гильбертовы пространства, $\overline{R(A)}$, $\overline{R(A^*)}$ — замыкание $R(A)$, $R(A^*)$ в норме F и H соответственно, $N(A)$, $N(A^*)$ — нули (ядра) операторов A, A^* соответственно, $U_*(f)$ — множество решений уравнения (1) при фиксированном $f \in F$, u_* — нормальное решение уравнения (1), определяемое условием: $\|u_*\|_H = \inf\limits_{u \in U_*(f)} \|u\|_H$.

Теорема 1. *Если $U_*(f) \neq \varnothing$, то $U_*(f) = v_* + N(A)$, где v_* — произвольный фиксированный элемент из $U_*(f)$. Если $U_*(f) \neq \varnothing$, то существует, притом единственное, нормальное решение уравнения* (1). *Для*

того чтобы элемент $u_* \in U_*(f)$ был нормальным решением уравнения (1), *необходимо и достаточно, чтобы $u_* \in \overline{R(A^*)}$.*

Теорема 2. *Если уравнение*

$$\Lambda c = f, \quad \Lambda = AA^* \tag{13}$$

разрешимо, то система A f-управляема, причем $u_* = A^*c$ — нормальное решение уравнения (1), *где* c — *произвольное решение уравнения* (13). *Если $R(A^*) = \overline{R(A^*)}$, то верно и обратное: из f-управляемости системы A вытекает разрешимость уравнения* (13).

Теорема 3. *Эквивалентны следующие утверждения:*

1) *система A вполне управляема, т. е. $U_*(f) \neq \varnothing$ $\forall f \in F$;*

2) *$R(A) = F$;*

3) *$N(A^*) = \{0\}$, $R(A^*) = \overline{R(A^*)}$;*

4) *имеет место оценка: $\|A^*c\|_H \geqslant \mu \|c\|_F$ $\forall c \in F$, $\mu = \text{const} > 0$, так что оператор A^* обратим и $(A^*)^{-1} \in \mathcal{L}(R(A^*) \to F)$;*

5) *справедлива оценка $\langle \Lambda c, c \rangle_F \geqslant \mu \|c\|_F^2$ $\forall c \in F$;*

6) *уравнение* (13) *имеет, притом единственное, решение при всех $f \in F$.*

Доказательства теорем 1–4 читатель может найти в [334; 416]. Мы здесь лишь отметим, что их доказательство существенно опирается на разложение гильбертова пространства на прямую сумму взаимно ортогональных подпространств: нулей линейного непрерывного оператора и замыкание области изменения сопряженного оператора [334; 393; 705], т. е.

$$H = N(A) \oplus \overline{R(A^*)}, \quad F = N(A^*) \oplus \overline{R(A)}.$$

3. Проиллюстрируем применение теорем 1–3 на примере задачи управления (2), (3), в частности, докажем известные теоремы Н. Н. Красовского, Р. Е. Калмана [411]. Прежде всего заметим, что в задаче (2), (3) множество значений оператора (4) является подпространством конечномерного пространства E^n и, следовательно, замкнуто, т. е. $R(A) = \overline{R(A)}$. Тогда, как известно, замкнута и область значений оператора A^*, т. е. $R(A^*) = \overline{R(A^*)}$. Отсюда и из теоремы 2 следует, что система (2) поточечно управляема тогда и только тогда, когда уравнение (13) имеет решение. В рассматриваемой задаче управления (2), (3) оператор $\Lambda = AA^* \in \mathcal{L}(E^n \to E^n)$, и система (13) представляет собой систему n линейных алгебраических уравнений с n неизвестными $c = (c^1, \ldots, c^n)$. Выпишем матрицу Λ этой системы в явном виде, пользуясь представлениями (пример 3.6):

$$Au = \int_0^T \Phi(T, \tau) B(\tau) u(\tau) d\tau, \quad A^*c = B^T(t) \Phi^T(T, t) c, \tag{14}$$

где $\Phi(t, \tau)$ — решение системы (3.34)

$$\frac{d\Phi(t, \tau)}{dt} = D(t)\Phi(t, \tau), \quad 0 \leqslant t \leqslant T; \quad \Phi(\tau, \tau) = I.$$

Отсюда следует, что $AA^*c = \int_0^T \Phi(T, \tau) B(\tau) B^T(\tau) \Phi^T(T, \tau) c d\tau$, так что

$$\Lambda = AA^* = \int_0^T \Phi(T, \tau) B(\tau) (\Phi(T, \tau) B(\tau))^T d\tau. \tag{15}$$

Таким образом, для того чтобы узнать, является ли система (2) f-управляемой, достаточно выяснить разрешимость системы линейных алгебраических уравнений (13) с квадратной матрицей (15). Если $c = c_*$ — какое-либо решение системы (13), то

$$u = u_*(t) = A^*c_* = B^T(t) \Phi^T(T, t) c_*, \quad 0 \leqslant t \leqslant T \tag{16}$$

— нормальное решение задачи управления (2), (3). Если система (13), (15) не имеет решения, то система (2) не является f-управляемой.

Далее, пользуясь теоремой 3, получим следующий критерий полной управляемости системы (2).

Теорема 4. *Для того, чтобы система* (2) *была вполне управляемой, необходимо и достаточно выполнения одного из следующих трех условий:*

1) *ранг матрицы* (15) *равен n;*

2) *равенство*

$$A^*c = B^T(t) \Phi^T(T, t) c = 0 \tag{17}$$

справедливо почти всюду на $[0, T]$ тогда и только тогда, когда $c = 0$;

3) *матрица* (15) *положительно определена, т. е. $\langle \Lambda c, c \rangle > 0$ $\forall c \in E^n$, $c \neq 0$.*

Доказательство. Система (1), (4) и, следовательно, система (13), (15) имеет решение при всех $f \in E^n$ тогда и только тогда, когда $\det \Lambda \neq 0$ [192; 353], т. е. $\text{rang} \Lambda = n$. Далее, в силу представления (14) для оператора A^* условие $N(A^*) = \{c \in F: A^*c = 0\} = \{c = 0\}$ равносильно условию (17). Кроме того, здесь $R(A^*) = \overline{R(A^*)}$. Отсюда и из утверждения 3) теоремы 3 следует, что для полной управляемости системы (2) необходимо и достаточно, чтобы тождество (17) выполнялось только при $c = 0$. Наконец, утверждение 3) теоремы 4 равносильно утверждению 5) теоремы 3. □

Опираясь на теорему 4, сформулируем условие полной управляемости системы (2) непосредственно в терминах матриц $D(t)$, $B(t)$.

Теорема 5 (Красовский Н. Н.). *Пусть матрицы $D(t)$, $B(t)$ дифференцируемы на отрезке $[0, T]$ $n-1$ раз. Последовательно определим матрицы*

$$K_0(t) = B(t), \quad K_{m+1}(t) = \dot{K}_m(t) - D(t) K_m(t), \quad m = 0, 1, \ldots, n-1, \tag{18}$$

и составим блочную матрицу

$$K(t) = (K_0(t), \ldots, K_{n-1}(t)), \quad 0 \leqslant t \leqslant T, \tag{19}$$

размера $n \times nr$. Для полной управляемости системы (2) *достаточно, чтобы существовал хотя бы один момент $\tau \in [0, T]$ такой, что*

$$\text{rang} K(\tau) = n. \tag{20}$$

Доказательство. Возьмем произвольный вектор $c \in E^n$ и составим функцию

$$z(t, c) = c^T \Phi(T, t) B(t) = c^T \Phi(T, t) K_0(t), \quad 0 \leqslant t \leqslant T. \tag{21}$$

Поскольку $\frac{d\Phi(T, t)}{dt} = -\Phi(T, t) D(t)$, $0 \leqslant t \leqslant T$ (см. упражнение 3.3), то, пользуясь индукцией и равенствами (18), нетрудно доказать, что

$$\frac{d^m z(t, c)}{dt^m} = c^T \Phi(T, t) K_m(t), \quad 0 \leqslant t \leqslant T, \quad m = 0, 1, \ldots, n-1. \tag{22}$$

Покажем, что $z(t, c) \equiv 0$, $0 \leqslant t \leqslant T$, только при $c = 0$. В самом деле, если $z(t, c) \equiv 0$, то и все производные этой функции равны нулю на $[0, T]$. С учетом (19), (22) отсюда имеем

$$(c^T\Phi(T, t)K_0(t), \ldots, c^T\Phi(T, t)K_m(t), \ldots, c^T\Phi(T, t)K_{n-1}(t)) =$$
$$= c^T\Phi(T, t)K(t) \equiv 0 \quad \forall t \in [0, T].$$

В частности это верно и для момента $t = \tau$, взятого из (20), так что $c^T\Phi(T, \tau)K(\tau) = 0$ или $K^T(\tau)\Phi^T(T, \tau)c = 0$. Однако ранг матрицы $K^T(\tau)$ равен n в силу (20), поэтому последнее равенство возможно только при $\Phi^T(T, \tau)c = 0$. Отсюда, учитывая, что квадратная матрица $\Phi(T, \tau)$ невырожденная, получаем, что $c = 0$. Тем самым установлено, что $z(t, c) = c^T\Phi(T, t)B(t) \equiv 0$, $0 \leqslant t \leqslant T$, только при $c = 0$. Это значит, что равенство (17) возможно только при $c = 0$. Согласно теореме 4 система (2) вполне управляема. □

Для стационарных систем, когда $D(t), B(t)$ постоянные матрицы, оказывается, условие (20) необходимо для полной управляемости системы (2).

Теорема 6 (Калман Р. Е.). *Система* (20)

$$\dot{x} = Dx + Bu, \quad 0 \leqslant t \leqslant T, \tag{23}$$

где D, B — постоянные матрицы, вполне управляема тогда и только тогда, когда

$$\operatorname{rang}(B, DB, \ldots, D^{n-1}B) = n. \tag{24}$$

Доказательство. Достаточность. Для постоянных матриц D, B из (18), (19) следует, что $K_m = (-1)^m D^m B$, $K = (B, -DB, D^2B, \ldots, (-1)^{n-1}D^{n-1}B)$. Нетрудно видеть, что ранги матрицы K и матрицы (24) равны, так как столбцы этих матриц отличаются лишь знаками. Отсюда и из (24) следует, что $\operatorname{rang} K = n$, и в силу теоремы 5 система (23) вполне управляема.

Заметим, что поскольку матрица (19) здесь не зависит от t, то условие (20) выполняется при всех $\tau \in [0, T]$. Это значит, что при выполнении условия (24) стационарная система (23) вполне управляема на любом отрезке $[0, T]$, $T > 0$.

Необходимость. Пусть система (23) вполне управляема. Покажем, что тогда выполняется условие (24). Снова воспользуемся функцией (21)

$$z(t, c) = c^T\Phi(T, t)B, \quad 0 \leqslant t \leqslant T.$$

Так как матрицы B, D не зависят от t, то функция $z(t, c)$ бесконечно дифференцируема. Как и в (22) с помощью индукции нетрудно доказать, что

$$z^{(m)} \equiv \frac{d^m z(t, c)}{dt^m} = (-1)^m c^T\Phi(T, t)D^m B \quad \forall t, \quad m = 0, 1, \ldots \tag{25}$$

Далее, воспользуемся теоремой Кэли — Гамильтона [192; 353]: любая квадратная матрица D является корнем своего характеристического многочлена $\det(D - \lambda I) = 0$. Это значит, что если $\det(D - \lambda I) = (\lambda^n + \beta_1\lambda^{n-1} + \ldots + \beta_{n-1}\lambda + \beta_n)(-1)^n$, то $D^n + \beta_1 D^{n-1} + \ldots + \beta_{n-1}D + \beta_n I = 0$. Умножим это матричное равенство слева на $c^T\Phi(T, t)$, справа на B. С учетом равенств (25) получим

$$z^{(n)} - \beta_1 z^{(n-1)} + \beta_2 z^{(n-2)} + \ldots + (-1)^{n-1}\beta_{n-1}z' + (-1)^n\beta_n z = 0 \quad \forall t \in \mathbb{R}. \tag{26}$$

Таким образом, функция $z = z(t, c)$ при любом фиксированном $c \in E^n$ является решением линейного дифференциального уравнения n-го порядка с постоянными коэффициентами. Так как $\Phi(T, T) = I$, то $z = z(t, c)$ при $t = T$ удовлетворяет условиям

$$z^{(m)}(T, c) = (-1)^m c^T D^m B, \quad m = 0, 1, \ldots, n-1. \tag{27}$$

Так как по условию система (23) вполне управляема, то равенство (17) возможно только при $c = 0$. Отсюда и из (26) следует, что набор

$$(z(t, c), z'(t, c), \ldots, z^{(n-1)}(t, c)) \neq 0 \quad \forall t \in \mathbb{R}, \quad \forall c \neq 0. \tag{28}$$

В самом деле, если $(z(t_0, c_0), \ldots, z^{(n-1)}(t_0, c_0)) = 0$ при каких-то t_0 и $c_0 \neq 0$, то по теореме единственности решения задачи Коши для уравнения (26) [588; 694] $z(t, c_0) \equiv 0$ $\forall t \in \mathbb{R}$, что невозможно в силу (17). Положим в (28) $t = T$. С учетом равенств (27) получим $(c^TB, -c^TDB, c^TD^2B, \ldots, (-1)^{n-1}c^TD^{n-1}B) = c^T(B, -DB, D^2B, \ldots, (-1)^{n-1}D^{n-1}B) \neq 0$ при всех $c \in E^n$, $c \neq 0$, что равносильно (24). Теорема 6 доказана.

Таким образом, проблему полной управляемости стационарной системы (23) удалось свести к алгебраической задаче (24) определения ранга некоторой матрицы. Условия полной управляемости для задач управления, аналогичных задачам из примеров 2–4, исследованы, например, в [11; 122; 123; 175; 287; 354; 355; 404; 730; 784; 797; 802; 804; 812]. Заметим, что задачу управления (1), конечно, можно записать в виде задачи минимизации квадратичной функции (2.3): $J(u) = \|Au - f\|_F^2 \to \inf$, $u \in U = H$, из которой при $J_* = 0$, $U_* \neq \varnothing$ получим решение задачи (1).

4. Покажем, что с каждой задачей управления (1) тесно связана некоторая, так называемая двойственная *задача наблюдения*, которую сформулируем следующим образом. Пусть за системой A ведется наблюдение, и с помощью какой-то измерительной аппаратуры определяется элемент

$$g = A^*c \in H, \tag{29}$$

называемый сигналом; здесь A^* — оператор, сопряженный к оператору A из (1). Предполагается, что элемент $c \in F$ мы не можем наблюдать напрямую, и мы желаем по сигналу (29) определить проекцию неизвестного нам элемента c на заданный элемент $f \in F$, точнее, скалярное произведение $\langle f, c\rangle_F$. Будем искать такой элемент $u \in H$, называемый *восстанавливающим* величину $\langle f, c\rangle_F$, чтобы скалярное произведение элемента u на наблюдаемый сигнал g в точности равнялось искомой величине $\langle f, c\rangle_F$, каким бы ни был неизвестный элемент c, порождающий сигнал (29), т. е. выполнялось равенство

$$\langle u, g\rangle_H = \langle f, c\rangle_F \quad \forall c \in F. \tag{30}$$

Сформулированная задача наблюдения (29), (30) называется *двойственной* к задаче управления (1). На первый взгляд может показаться, что эти задачи слабо связаны друг с другом. Однако верна

Теорема 7. *Задачи* (1) *и* (29), (30) *одновременно разрешимы или неразрешимы. Если они разрешимы, то решение u каждой из них удовлетворяет операторному уравнению* (1).

Доказательство. Пусть задача наблюдения (29), (30) разрешима, пусть u — восстанавливающий элемент этой задачи. Тогда из равенств (29), (30) имеем:

$$\langle f, c\rangle_F = \langle u, g\rangle_H = \langle u, A^*c\rangle = \langle Au, c\rangle \quad \forall c \in F. \tag{31}$$

Отсюда следует, что $\langle Au - f, c\rangle_F = 0 \ \ \forall c \in F$. Полагая в этом равенстве $c = Au - f$, получим $Au - f = 0$, т. е. восстанавливающий элемент u является решением уравнения (1). Таким образом, если задача наблюдения (29), (30) разрешима, то задача управления (1) также разрешима. Обратно: если задача управления (1) имеет решение, то для решения этой задачи справедлива та же цепочка равенств (31), из которой следует, что любое решение u уравнения (1) является восстанавливающим элементом задачи наблюдения (29), (30). □

Заметим, что задачи (1) и (29), (30) являются взаимодвойственными и для их формулировки достаточно знать оператор A или A^* и элемент f. Здесь, конечно, мы учитываем, что $A \in \mathcal{L}(H \to F)$, где H, F — гильбертовы пространства, и справедливо равенство $A^{**} = A$.

Определение 2. Система A называется *f-наблюдаемой* или *поточечно наблюдаемой* по сигналу (29), если уравнение (1) при заданном $f \in F$ имеет хотя бы одно решение. Система A называется *вполне наблюдаемой* по сигналу (29), если уравнение (1) имеет решение при всех $f \in F$.

Сравнивая определения 1, 2, заключаем, что понятию поточечной управляемости [полной управляемости] в задаче управления соответствует понятие поточечной наблюдаемости [полной наблюдаемости] в двойственной задаче наблюдения.

В силу теоремы 7, связывающей взаимодвойственные задачи управления и наблюдения, при исследовании задачи наблюдения (29), (30) могут быть использованы теже теоремы 1–3. В частности, для выяснения того, будет ли система A f-наблюдаемой, можно воспользоваться уравнением (13). Если система A вполне наблюдаема, то в силу утверждения 4) теоремы 3, оператор A^* обратим, $(A^*)^{-1} \in \mathcal{L}(R(A^*) \to F)$. Отсюда следует, что равенство (29), рассматриваемое как уравнение относительно неизвестного (наблюдаемого) элемента c, однозначно разрешимо и $c = (A^*)^{-1}g$. Оператор $U = (A^*)^{-1}$ естественно назвать *восстанавливающим оператором*. Если F — сепарабельное пространство и система A вполне наблюдаема, то в качестве элемента f в (30) можно последовательно брать $f = f_k$, $k = 1, 2, \ldots$, где $\{f_k\}$ — ортонормированный базис в F. Каждое уравнение $Au = f_k$ имеет хотя бы одно решение $u = u_k$, $k = 1, 2, \ldots$ Тогда по формулам (30), (31) мы можем определить коэффициент Фурье $c_k = \langle f_k, c\rangle = \langle u_k, g\rangle$, $k = 1, 2, \ldots$, неизвестного элемента c и восстановить его по формуле $c = \sum\limits_{k=1}^{\infty} c_k f_k = \sum\limits_{k=1}^{\infty} \langle u_k, g\rangle f_k$.

5. Сформулируем задачи наблюдения, двойственные к задачам управления из примеров 1–4.

Пример 1*. Оператор A^*, сопряженный к оператору (4), имеет вид

$$A^*c = B^T(t)\psi(t; c), \quad 0 \leqslant t \leqslant T,$$

где $\psi = \psi(t; c)$ — решение задачи Коши:

$$\dot\psi = -D^T(t)\psi, \quad 0 \leqslant t \leqslant T; \quad \psi(T) = c \in E^n = F \tag{32}$$

(см. пример 3.6). Задача наблюдения (29), (30) для системы (32), двойственная к задаче управления (2), (3), формулируется так: найти элемент $u = u(t) \in H = L_2^r[0, T]$, восстанавливающий величину $\langle f, c\rangle_{E^n}$ по формуле (30), где $g(t) = B^T(t)\psi(t; c)$, $0 \leqslant t \leqslant T$ — наблюдаемый сигнал. Согласно теореме 7 восстанавливающий элемент $u = u(t)$ является решением уравнения (1), где A — оператор (4). Для решения задачи f-наблюдаемости для системы (32) можно воспользоваться уравнением (13) с матрицей (15). Теоремы 5, 6 содержат условия полной наблюдаемости системы (32), совпадающие с условиями полной управляемости системы (2).

Пример 2*. Оператор A^*, сопряженный к оператору (8), действует по формуле

$$A^*c = \psi(l, t; c), \quad 0 \leqslant t \leqslant T,$$

где $\psi = \psi(s, t; c)$ — решение краевой задачи (7.13), (7.14) с $c = c(s) \in F = L_2[0, l]$. Задача наблюдения для системы (7.13), (7.14), двойственная к задаче управления (6), (7), формулируется так: найти элемент $u = u(t) \in L_2[0, T] = H$, восстанавливающий величину $\langle f, c\rangle_F = \int\limits_0^l f(s)c(s)ds$ по формуле (30), где $g(t) = \psi(l, t; c)$ — наблюдаемый сигнал. Согласно теореме 7 восстанавливающий элемент $u = u(t)$ является решением уравнения (1), где A — оператор (8).

Пример 3*. Оператор A^*, сопряженный к оператору (11), имеет вид

$$A^*c = \psi(0, t; c),$$

где $\psi = \psi(s, t; c)$ — решение краевой задачи (8.11), (8.12) при $c = c(s) = (c_0(s), c_1(s)) \in F = L_2[0, l] \times L_2[0, l]$. Задача наблюдения для системы (8.11), (8.12), двойственная к задаче управления (9), (10), формулируется так. Наблюдаемый сигнал имеет вид

$$g = g(t) = A^*c = \psi(0, t; c), \quad 0 \leqslant t \leqslant T.$$

Требуется найти функцию $u = u(t) \in L_2[0, T] = H$, для которой

$$\langle u, g\rangle_H = \int\limits_0^T u(t)(A^*c)(t)dt = \langle f, c\rangle_F = \int\limits_0^l (f_0(s)c_0(s) + f_1(s)c_1(s))ds, \tag{33}$$

где $f = (f_0(s), f_1(s))$ — заданный элемент из F. Восстанавливающий элемент $u = u(t)$ является решением уравнения (1), где A — оператор (11).

Пример 4*. Оператор A^*, сопряженный к оператору (11), порожденному системой (12), согласно формуле (8.24), имеет вид

$$A^*c = \int\limits_0^l r(s)\psi(s, t; c)ds, \quad 0 \leqslant t \leqslant T, \tag{34}$$

где $\psi = \psi(s, t; c)$ — решение краевой задачи (8.25). Задача наблюдения для системы (8.25), двойственная к задаче управления (12), (10), формулируется так. Наблюдаемый сигнал имеет вид (34). Требуется найти функцию $u = u(t) \in L_2[0, T] = H$, для которой выполняется равенство (33). Восстанавливающий элемент $u = u(t)$ является решением уравнения (1), где A — оператор (11), соответствующий системе (12).

6. Описанная схема составления взаимодвойственных задач управления и наблюдения может быть применена и к другим классам линейных динамических систем, описываемых различными дифференциальными, интегро-дифференциальными, разностными уравнениями.

Параллельное изучение пар взаимодвойственных задач управления и наблюдения позволяет глубже понять многие аспекты теории таких задач, открывает возможности использования методов, разработанных для решения, скажем, задач управления, и применения их к задачам наблюдения и наоборот. В частности, отметим, что сформулированный выше класс задач наблюдения (29), (30) является подклассом обратных задач. Поэтому математический аппарат, разработанный в рамках теории и методов решения обратных задач (см., например, [127; 230; 269; 557; 618; 812]) может быть использован как для исследования задач наблюдения, так и двойственных к ним задач управления. Использование идей двойственности полезно также и при исследовании вопросов аппроксимации задач управления и наблюдения [170; 363; 433; 598; 599; 607; 797; 804].

Упражнения

1. Проверить, что система $\dot{x} = u(t)$, $0 \leqslant t \leqslant T$, вполне управляема в классе постоянных управлений при всех $T > 0$.

2. Доказать, что система $\dot{x} = u(t)$, $\dot{y} = -u(t)$, $u = u(t) \in L_2[0, T]$, поточечно управляема тогда и только тогда, когда $x(T) + y(T) = x(0) + y(0)$.

3. Исследовать, будут ли следующие системы поточечно или вполне управляемы [наблюдаемы]:

$$1)\begin{cases} \dot{x} = y + u(t) \\ \dot{y} = -x \end{cases} \quad 2)\begin{cases} \dot{x} = y \\ \dot{y} = u(t) \end{cases} \quad 3)\begin{cases} \dot{x} = y \\ \dot{y} = x + u(t) \end{cases} \quad 4)\begin{cases} \dot{x} = x + u(t) \\ \dot{y} = x + y \end{cases}$$

$$5)\begin{cases} \dot{x} = x + u(t) \\ \dot{y} = x + y + v(t) \end{cases} \quad 6)\begin{cases} \dot{x} = 2x - 2y + u(t) \\ \dot{y} = y - x - \frac{1}{2}u(t) \end{cases} \quad 7)\begin{cases} \dot{x}^1 = x^2 \\ \dot{x}^2 = x^3 \\ \dot{x}^3 = u(t) \end{cases}$$

4. Доказать, что система (2) не является вполне управляемой тогда и только тогда, если существует вектор $c_* \in E^n$, $c_* \neq 0$, что правые концы всех траекторий системы принадлежат гиперплоскости $\Gamma = \{x \in E^n : \langle c_*, x \rangle = 0\}$, т. е. $\langle c_*, x(T, u) \rangle = 0 \ \forall u \in L_2^r[0, T]$. Указание: воспользоваться утверждением 2) теоремы 4 (см. также [411; 712]).

5. Можно ли утверждать, что если $N(A^*) = \{0\}$, то $N(A)$ также состоит из единственной нулевой точки? Указание: рассмотреть оператор с матрицей $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}$, действующий из E^2 в E^3.

6. Доказать, что если $A \in \mathcal{L}(H \to F)$, H, F — гильбертовы пространства, то $R(A) = \overline{R(A)}$ тогда и только тогда, когда или 1) $R(A^*) = \overline{R(A^*)}$ [416], или 2) $\inf \|Au\| > 0$, где нижняя грань берется по всем $u \in H$, $\|u\| = 1$, $\langle u, v \rangle = 0 \ \forall v \in N(A)$ [131].

7. Докажите теоремы 1–3.

8. Доказать, что оператор A^+, который каждому $f \in R(A)$ ставит в соответствие нормальное решение уравнения (1), является линейным. Если $R(A) = \overline{R(A)}$, то $A^+ \in \mathcal{L}(R(A) \to F)$, $N(A^+) = N(A^*)$, $R(A^+) = R(A^*)$, $A^+A = I - \mathcal{P}_{N(A)}$ ($\mathcal{P}_U$ — оператор проектирования на U), $AA^+ = \mathcal{P}_{R(A)}$ (в частности, если $R(A) = F$, то $AA^+ = I$, поэтому оператор A^+ называют *псевдообратным* к оператору A), $AA^+A = A$, $A^+AA^+ = A^+$, $(AA^+)^* = AA^+$, $(A^+A)^* = A^+A$, $A^{++} = A$, $(A^*)^+ = (A^+)^*$.

9. Для системы (7.2), (7.3) задачу управления $x(s, T; u) = f(s)$, $0 \leqslant s \leqslant l$, записать в виде уравнения (1); сформулировать двойственную к ней задачу наблюдения. Указание: воспользоваться операторами (7.9), (7.13).

10. Для системы (8.2), (8.3) задачу управления (10) записать в виде уравнения (1); сформулировать двойственную к ней задачу наблюдения. Указание: воспользоваться операторами (8.7), (8.10).

11. Для системы (8.27)–(8.30) задачу управления (10) записать в виде уравнения (1); сформулировать двойственную к ней задачу наблюдения. Указание: по формуле (11) ввести оператор A, действующий из $H = L_2(Q)$ в $F = L_2[0, l] \times L_2[0, l]$; опираясь на оценку (8.34), убедиться, что $A \in \mathcal{L}(H \to F)$; показать, что сопряженный оператор A^* действует по правилу $A^*c = \psi(s, t; c)$, где $\psi = \psi(s, t; c)$ — решение краевой задачи (8.37).

12. Пусть колебания струны описываются краевой задачей

$$x_{tt} = x_{ss}, \quad (s, t) \in Q = (0, l) \times (0, T),$$
$$x|_{s=0} = u(t), \quad x|_{s=l} = 0, \quad 0 < t < T; \quad x|_{t=0} = 0, \quad x_t|_{t=0} = 0, \quad 0 \leqslant s \leqslant l.$$

Задача управления: найти $u = u(t) \in L_2[0, T]$, чтобы $x|_{t=T} = y_0(s)$, $x_t|_{t=T} = y_1(s)$, $0 \leqslant s \leqslant l$. Доказать [804], что эта задача при $T > 2l$ имеет решение при всех $y_0(s) \in L_2[0, T]$, $y_1(s) \in H^{-1}[0, l]$. Сформулировать двойственную задачу наблюдения. Указание: показать, что оператор A из (11) в этой задаче действует из $H = L_2[0, T]$ в $F = L_2[0, l] \times H^{-1}[0, l]$, непрерывен, а сопряженный к нему оператор A^* действует по правилу: $A^*c = \psi_s(0, t; c)$, где $\psi = \psi(s, t; c)$ — решение краевой задачи

$$\psi_{tt} = \psi_{ss}, \quad (s, t) \in Q; \quad \psi|_{s=0} = \psi|_{s=l} = 0, \quad \psi|_{t=T} = c_1(s), \quad \psi_t|_{t=T} = -c_0(s),$$

где $c = (c_0(s), c_1(s)) \in F^* = L_2[0, l] \times H_0^1[0, l]$; доказать оценку: $\|A^*c\|_H \geqslant \frac{T - 2l}{l} \|c\|_F$ [804] и воспользоваться утверждением 4) теоремы 3 (в других функциональных пространствах такая задача управления исследовалась в [354; 355]).

13. За динамической системой

$$\psi_t = \psi_{ss}, \ (s, t) \in Q = (0, l) \times (0, T), \ \psi|_{s=0} = \psi|_{s=l} = 0, \ \psi|_{t=T} = c(s)$$

ведется наблюдение и измеряется сигнал $g(t) = \psi_s(0, t; c)$, $0 \leqslant t \leqslant T$. Требуется найти элемент $u = u(t) \in L_2[0, T] = H$, восстанавливающий величину $\langle f, c \rangle_{L_2[0, l]}$ по правилу (30), где $f = f(s) \in F = L_2[0, l]$ — заданный элемент. Для этой задачи наблюдения сформулировать двойственную задачу управления.

§ 11. Метод моментов

1. Сначала изложим элементы теории моментов. Начнем с так называемой *конечномерной проблемы моментов*.

Пусть B — банахово пространство, B^* — сопряженное к B пространство. Пусть заданы элементы $\varphi_1, \varphi_2, \ldots, \varphi_n \in B$ и вещественные числа $\mu = (\mu_1, \ldots, \mu_n)$. Проблема моментов формулируется так: найти линейный функционал $u \in B^*$ такой, что

$$\langle u, \varphi_i \rangle = \mu_i, \quad i = 1, \ldots, n. \tag{1}$$

Напоминаем, что через $\langle u, \varphi \rangle$ мы обозначаем значение функционала u на элементе $\varphi \in B$. Рассмотрим условия, обеспечивающие существование решения проблемы (1). Если $\mu = 0$, то, очевидно, проблема (1) всегда имеет решение, например, $u = 0$. Пусть $\mu \neq 0$. Тогда, оказывается, важную роль при исследовании вопросов существования решения проблемы (1) играет следующая конечномерная задача минимизации:

$$g(c) = \Big\| \sum_{i=1}^{n} c^i \varphi_i \Big\|_B \to \inf, \quad c \in \Gamma_\mu = \{c = (c^1, \ldots, c^n) \in E^n : \langle \mu, c \rangle = 1\}. \tag{2}$$

Убедимся, что задача (2) имеет решение. Заметим, что Γ_μ — гиперплоскость в E^n — это непустое выпуклое замкнутое множество. Функция $g(c)$ непрерывна и, более того, удовлетворяет условию Липшица на E^n:

$$|g(c+\delta c)-g(c)| \leqslant \|\sum_{i=1}^{n}\delta c^i\varphi_i\| \leqslant \max_{1\leqslant i\leqslant n}|\delta c^i|\sum_{i=1}^{n}\|\varphi_i\| = L\|\delta c\|_\infty$$
$$\forall c,\ c+\delta c\in E^n.$$

Сначала рассмотрим случай, когда система $\{\varphi_1,\ldots,\varphi_n\}$ линейно независима. При этом $g(c)=0$ тогда и только тогда, когда $c=0$. Непрерывная функция $g(c)>0$ на компактном множестве $S=\{c\colon |c|=1\}$ достигает своей нижней грани хотя бы в одной точке $c_0\in S$, причем $g(c_0)=\min\limits_S g(c)>0$. Поэтому $g(c)=|c|\left\|\sum\limits_{i=1}^{n}\frac{c^i}{|c|}\varphi_i\right\|\geqslant |c|g(c_0)\to+\infty$ при $|c|\to\infty$. Согласно теореме 2.1.3 найдется точка $c_*\in\Gamma_\mu$, для которой $\inf\limits_{c\in\Gamma_\mu}g(c)=g(c_*)=g_*$. Так как $0\notin\Gamma_\mu$, то $c_*\neq 0$ и $g(c_*)=g_*>0$.

Пусть теперь система $\{\varphi_1,\ldots,\varphi_n\}$ линейно зависима. Не умаляя общности, можем считать, что элементы $\varphi_1,\ldots,\varphi_k$, $k<n$, линейно независимы, а остальные $\varphi_{k+1},\ldots,\varphi_n$ — линейно выражаются через них:

$$\varphi_j=\sum_{i=1}^{k}\alpha_{ij}\varphi_i,\quad (\alpha_{1j},\ldots,\alpha_{kj})\neq 0,\quad j=k+1,\ldots,n. \tag{3}$$

Тогда

$$g(c)=\left\|\sum_{i=1}^{k}c^i\varphi_i+\sum_{j=k+1}^{n}c^j\left(\sum_{i=1}^{k}\alpha_{ij}\varphi_i\right)\right\|=\left\|\sum_{i=1}^{k}\left(c^i+\sum_{j=k+1}^{n}c^j\alpha_{ij}\right)\varphi_i\right\|. \tag{4}$$

Допустим, что

$$\mu_j=\sum_{i=1}^{k}\alpha_{ij}\mu_i,\quad j=k+1,\ldots,n. \tag{5}$$

Условие $\langle\mu,c\rangle=1$ в этом случае запишется в виде

$$\sum_{i=1}^{k}c^i\mu_i+\sum_{j=k+1}^{n}c^j\left(\sum_{i=1}^{k}\alpha_{ij}\mu_i\right)=\sum_{i=1}^{k}\left(c^i+\sum_{j=k+1}^{n}c^j\alpha_{ij}\right)\mu_i=1. \tag{6}$$

Введем новые переменные $\bar c^i=c^i+\sum\limits_{j=k+1}^{n}c^j\alpha_{ij}$, $i=1,\ldots,k$, $\overline{\mu}=(\mu_1,\ldots,\mu_k)$, $\bar g(\bar c)=\left\|\sum\limits_{i=1}^{k}\bar c_i\varphi_i\right|$. Тогда с учетом соотношений (4), (6) имеем: $\inf\limits_{\langle\mu,c\rangle=1}g(c)=\inf\limits_{\langle\overline{\mu},\bar c\rangle=1}\bar g(\bar c)$. Так как элементы $\{\varphi_1,\ldots,\varphi_k\}$ линейно независимы, то как было установлено выше, найдется точка $\bar c_*=(\bar c_*^1,\ldots,\bar c_*^k)$, $\langle\overline{\mu},\bar c_*\rangle=1$, такая, что $\bar g(\bar c_*)=\inf\limits_{\langle\overline{\mu},\bar c\rangle=1}\bar g(\bar c)>0$. Тогда $c_*=(\bar c_*^1,\ldots,\bar c_*^k,0,\ldots,0)\in E^n$ удовлетворяет условию $\langle\mu,c_*\rangle=1$ и $\bar g(\bar c_*)=g(c_*)=\inf\limits_{\langle\mu,c\rangle=1}g(c)=g_*>0$. Таким образом, задача (2) разрешима и в случае (3), (5), причем $g_*>0$. Наконец, рассмотрим случай, когда $\varphi_1,\ldots,\varphi_k$ линейно независимы, выполнены условия (3), но (5) нарушено при некотором $j=j_0$, $k+1\leqslant j_0\leqslant n$, т. е. $\mu_{j_0}-\sum\limits_{i=1}^{k}\alpha_{ij_0}\mu_i=\alpha\neq 0$. Определим $c_*=(c_*^1,\ldots,c_*^n)$ так: $c_*^i=-\frac{\alpha_{ij_0}}{\alpha}$, $i=1,\ldots,k$, $c_*^{j_0}=\frac{1}{\alpha}$, остальные $c_*^j=0$. Тогда $\langle\mu,c_*\rangle=\sum\limits_{i=1}^{k}\mu_i\left(-\frac{\alpha_{ij_0}}{\alpha}\right)+\mu_{j_0}\frac{1}{\alpha}=(\mu_{j_0}-\sum\limits_{i=1}^{k}\mu_i\alpha_{ij_0})\frac{1}{\alpha}=1$, т. е. $c_*\in\Gamma_\mu$.

Кроме того, с учетом равенства (4) имеем: $g(c_*)=\left\|\sum\limits_{i=1}^{k}\left(-\frac{\alpha_{ij_0}}{\alpha}+\frac{\alpha_{ij_0}}{\alpha}\right)\varphi_i\right\|=0=\inf\limits_{\langle\mu,c\rangle=1}g(c)=g_*$. Таким образом, доказана

Теорема 1. *Пусть $\mu\neq 0$. Тогда задача* (2) *имеет решение, т. е. $\exists c_*\in\Gamma_\mu$, что $\inf\limits_{\langle\mu,c\rangle=1}g(c)=g(c_*)=g_*\geqslant 0$, причем:*

1) *$g_*>0$, если $\varphi_1,\ldots,\varphi_n$ линейно независимы или $\varphi_1,\ldots,\varphi_k$, $k<n$, линейно независимы и выполнены равенства* (3), (5);

2) *$g_*=0$, если $\varphi_1,\ldots,\varphi_k$, $k<n$, линейно независимы, выполнены равенства* (3), *но хотя бы одно из равенств* (5) *не выполнено.*

Теперь мы можем сформулировать следующий критерий существования решения проблемы моментов (1). Обозначим через U_* множество решений этой проблемы. Заметим, что проблема (1) в общем случае может иметь много решений. Решение u_* проблемы (1) назовем *нормальным решением*, если $\|u_*\|=\min\limits_{u\in U_*}\|u\|$. Если $\mu=0$, то $u_*=0$, очевидно, является нормальным решением проблемы (1).

Теорема 2. *Пусть $\mu\neq 0$. Тогда для того чтобы $U_*\neq\varnothing$, необходимо и достаточно, чтобы в задаче* (2) $g_*=\inf\limits_{\langle\mu,c\rangle=1}\left\|\sum\limits_{i=1}^{n}c_i\varphi_i\right\|>0$. *Если $U_*\neq\varnothing$, то существует нормальное решение u_* проблемы* (1), *причем $\|u_*\|=\frac{1}{g_*}$. Множество решений проблемы* (1) *представимо в виде: $U_*=u_*+U_{0*}$, где U_{0*} — множество решений проблемы* (1) *при $\mu=0$.*

Доказательство. Необходимость. Пусть $U_*\neq\varnothing$. Возьмем произвольные $u\in U_*$ и $c\in\Gamma_\mu$. Умножим i-е равенство (1) на c^i и просуммируем получившиеся равенства: $\sum\limits_{i=1}^{n}c^i\langle u,\varphi_i\rangle=\langle u,\sum\limits_{i=1}^{n}c^i\varphi_i\rangle=\langle\mu,c\rangle=1$. Тогда $1\leqslant\|u\|\left\|\sum\limits_{i=1}^{n}c^i\varphi_i\right\|$ $\forall c\in\Gamma_\mu$. Переходя к нижней грани по $c\in\Gamma_\mu$, отсюда получаем $1\leqslant\|u\|g_*$. Следовательно, $g_*>0$ и, кроме того,

$$0<\frac{1}{g_*}\leqslant\|u\|\quad\forall u\in U_*. \tag{7}$$

Достаточность. Пусть $g_*>0$. Покажем, что тогда проблема моментов (1) имеет хотя бы одно решение, причем $\exists u_*\in U_*$, для которого $\|u_*\|=\frac{1}{g_*}$. Введем два множества M и N следующим образом:

$$M=\{\varphi\in B\colon\|\varphi\|\leqslant g_*\},$$
$$N=\{\varphi\in B\colon\varphi=\sum_{i=1}^{n}c^i\varphi_i\text{ для какого-либо }c\in\Gamma_\mu\}.$$

Нетрудно проверить, что оба множества выпуклы, $\operatorname{int}M\neq\varnothing$. Кроме того, $\|\varphi\|\geqslant g_*$ $\forall\varphi\in N$ и $\|\varphi\|<g_*$ $\forall\varphi\in\operatorname{int}M$, так что $\operatorname{int}M\cap N=\varnothing$. По теореме 3.6 множества M и N отделимы, т. е. существует $u_*\in B^*$, что

$$\langle u_*,\varphi\rangle\leqslant 1\quad\forall\varphi\in M,\quad\langle u_*,\varphi\rangle\geqslant 1\quad\forall\varphi\in N. \tag{8}$$

Возьмем какую-либо точку $c_*\in\Gamma_\mu$, для которой $g(c_*)=\inf\limits_{\Gamma_\mu}g(c)=g_*$ (теорема 1), и положим $\varphi_*=\sum\limits_{i=1}^{n}c_*^i\varphi_i$. Нетрудно видеть, что $\varphi_*\in M\cap N$. Тогда из (8) следует, что $\langle u_*,\varphi_*\rangle=1$. Покажем, что

$$\langle u_*,\varphi\rangle=1\quad\forall\varphi\in N. \tag{9}$$

Возьмем произвольный элемент $\varphi \in N$. По определению множества N найдется точка $c \in \Gamma_\mu$, что $\varphi = \sum_{i=1}^{n} c^i \varphi_i$. Для точки $c_\alpha = \alpha c + (1-\alpha)c_*$ имеем $\langle \mu, c_\alpha \rangle = \alpha \langle \mu, c \rangle + (1-\alpha)\langle \mu, c* \rangle = 1 \quad \forall \alpha \in \mathbb{R}$. Тогда $\varphi_\alpha = \alpha\varphi + (1-\alpha)\varphi_* = \alpha \sum_{i=1}^{n} c^i \varphi_i + (1-\alpha) \sum_{i=1}^{n} c_*^i \varphi_i = \sum_{i=1}^{n} c_\alpha^i \varphi_i \in N \quad \forall \alpha \in \mathbb{R}$. В силу (8) $1 \leqslant \langle u_*, \varphi_\alpha \rangle = \langle u_*, \alpha\varphi + (1-\alpha)\varphi_* \rangle = \langle u_*, \varphi_* \rangle + \alpha \langle u_*, \varphi - \varphi_* \rangle = 1 + \alpha \langle u_*, \varphi - \varphi_* \rangle$ или $0 \leqslant \alpha \langle u_*, \varphi - \varphi_* \rangle \ \forall \alpha \in \mathbb{R}$. Это возможно только при $\langle u_*, \varphi - \varphi_* \rangle = 0$, т. е. $\langle u_*, \varphi \rangle = \langle u_*, \varphi_* \rangle = 1 \ \forall \varphi \in N$. Равенство (9) установлено.

Покажем, что u_* — решение проблемы (1). Так как $\mu \neq 0$, то $\mu_i \neq 0$ при каком-либо i, $1 \leqslant i \leqslant n$. Для таких μ_i положим $c = c_0 = (0, \ldots, 0, c_0^i = \frac{1}{\mu_i}, 0, \ldots, 0)$. Ясно, что $\langle \mu, c_0 \rangle = 1$, поэтому $\varphi_0 = \frac{1}{\mu_i}\varphi_i \in N$. Согласно (9) тогда $1 = \langle u_*, \varphi_0 \rangle = \frac{1}{\mu_i}\langle u_*, \varphi_i \rangle$, т. е. $\langle u_*, \varphi_i \rangle = \mu_i$ для всех i, для которых $\mu_i \neq 0$. Если же $\mu_i = 0$, то возьмем $c_\beta = (c_*^1, \ldots, c_*^{i-1}, c^i = \beta, c_*^{i+1}, \ldots, c_*^n)$, $\beta \in \mathbb{R}$. Так как $\langle \mu, c_\beta \rangle = \sum_{j=1, j\neq i} \mu_j c_*^j + 0 \cdot \beta = \langle \mu, c_* \rangle = 1 \ \forall \beta \in \mathbb{R}$, то $\varphi_\beta = \sum_{i=1}^{n} c_\beta^i \varphi_i \in N \ \forall \beta \in \mathbb{R}$. В силу (9) тогда $1 = \langle u_*, \varphi_\beta \rangle = \sum_{j=1, j\neq i} c_*^j \varphi_j + \beta \langle u_\alpha, \varphi_i \rangle \ \forall \beta \in \mathbb{R}$. Разделим это равенство на $\beta > 0$ и устремим $\beta \to +\infty$. Получим: $\langle u_*, \varphi_i \rangle = 0 = \mu_i$. Таким образом, показано, что $\langle u_*, \varphi_i \rangle = \mu_i$ при всех $i = 1, \ldots, n$, т. е. $u_* \in U_*$.

Убедимся, что u_* — нормальное решение проблемы (1). Так как $\langle u_*, \varphi \rangle \leqslant 1$ $\forall \varphi \in M$ в силу (8), то $\|u_*\| = \sup_{\|\varphi\| \leqslant 1} \langle u_*, \varphi \rangle = \sup_{\|\varphi\| \leqslant g_*} \langle u_*, \frac{\varphi}{g_*} \rangle = \frac{1}{g_*} \sup_{\varphi \in M} \langle u_*, \varphi \rangle \leqslant \frac{1}{g_*}$.

С другой стороны, для точки $u_* \in U_*$ из (7) имеем: $\|u_*\| \geqslant \frac{1}{g_*}$. Следовательно, $\|u_*\| = \inf_U \|u\| = \frac{1}{g_*}$, т. е. u_* — нормальное решение проблемы (1). Попутно получили, что

$$g_* \|u_*\| = 1 = \langle u_*, \varphi_* \rangle. \tag{10}$$

Равенство

$$U_* = u_* + U_{0*}$$

предлагаем читателю доказать самостоятельно. □

2. Рассмотрим так называемую конечномерную *l-проблему моментов*: найти функционал $u \in B^*$ такой, что

$$\langle u, \varphi_i \rangle = \mu_i, \quad i = 1, \ldots, n, \quad \|u\| \leqslant l, \tag{11}$$

где $l > 0$ — заданное число. Как видим, l—проблема моментов отличается от проблемы (1) наличием дополнительного ограничения $\|u\| \leqslant l$. Опираясь на теорему 2 нетрудно получить критерий разрешимости l-проблемы (11). Заметим тогда, что при $\mu = 0$ проблема (11) всегда имеет решение, например, $u_* = 0$.

Теорема 3. [441]. *Пусть $\mu \neq 0$. Для того чтобы проблема* (11) *имела хотя бы одно решение, необходимо и достаточно, чтобы*

$$g_* = \inf_{\langle \mu, c \rangle = 1} \left\| \sum_{i=1}^{n} c^i \varphi_i \right\| \geqslant \frac{1}{l}. \tag{12}$$

Доказательство. Необходимость. Пусть u — решение l— проблемы (11). Тогда u является решением и проблемы (1), и по теореме 2 необходимо $g_* > 0$, $l \geqslant \|u\| \geqslant \frac{1}{g_*}$.

Достаточность. Пусть $g_* \geqslant \frac{1}{l}$. Согласно теореме 2 проблема (1) имеет нормальное решение u_* с нормой $\|u_*\| = \frac{1}{g_*} \leqslant l$, так что u_* — решение проблемы (11). Теорема 3 доказана. □

Замечание 1. Основываясь на теоремах 1–3, можно наметить следующий порядок действий для определения решения проблем (1), (11). Можем считать, что $\mu \neq 0$. Тогда задача (2) имеет смысл, и, решая эту задачу, можем определить

$$c_* \in \Gamma_\mu, \quad g(c_*) = g_* = \|\varphi_*\|, \quad \varphi_* = \sum_{i=1}^{n} c_*^i \varphi_i.$$

Так как проекция точки на гиперплоскость Γ_μ легко вычисляется (пример 4.4.2), то при решении задачи (2) можно использовать метод проекции градиента или субградиента (см. § § 5.2, 5.3). Если при этом выяснится, что $g_* = 0$, то согласно теореме 2 проблема (1) не имеет решения. Если $g_* > 0$, то проблема (1) разрешима. Нормальное решение u_* проблемы (1) удовлетворяет равенствам (10), которые можно присоединить к системе уравнений (1). В частности, если равенства (10) однозначно определяют элемент u_*, то u_* — нормальное решение проблемы (1). Например, если $B = H$ — гильбертово пространство, то из (10) имеем $u_* = \frac{\varphi_*}{g_*^2}$. Если найденное нормальное решение u_* проблемы (1) удовлетворяет неравенству $\|u_*\| \leqslant l$, то u_* — решение l—проблемы (11). Если же $\|u_*\| > l$, то проблема (11) не имеет решения.

Замечание 2. Если $B = H$ — гильбертово пространство, то можно предложить следующий более простой метод решения проблемы (1), явно не использующий задачу (2). А именно, решение можем искать в виде

$$u = \sum_{i=1}^{n} a^i \varphi_i. \tag{13}$$

Подставив (13) в (1) для определения неизвестных коэффициентов $a^1, \ldots, a^n$, получим систему линейных алгебраических уравнений

$$\sum_{i=1}^{n} \langle \varphi_i, \varphi_j \rangle a^i = \mu_j, \quad j = 1, \ldots, n. \tag{14}$$

Нетрудно видеть, что элемент u из (13) будет решением проблемы (1) тогда и только тогда, когда $a = (a^1, \ldots, a^n)$ — решение системы (14). Матрица системы (14) симметрична и представляет собой матрицу Грама [89; 192; 353] системы векторов $\varphi_1, \ldots, \varphi_n$. Если эти векторы линейно независимы, то матрица Грама невырождена, и система (14) имеет единственное решение.

3. Остановимся на бесконечномерной проблеме моментов, также имеющей приложение в задачах управления. Пусть заданы некоторая последовательность элементов $\varphi_1, \varphi_2, \ldots, \varphi_n, \ldots$ из банахова пространства B и числовая последовательность $\mu_1, \mu_2, \ldots, \mu_n, \ldots$ Требуется найти функционал $u \in B^*$, удовлетворяющий условиям

$$\langle u, \varphi_i \rangle = \mu_i, \quad i = 1, 2, \ldots \tag{15}$$

Покажем, что разрешимость сформулированной проблемы моментов (15) тесно связана с разрешимостью следующих конечномерных проблем моментов, полученных «усечением» проблемы (15): ищется функционал $u_n \in B^*$ такой, что

$$\langle u_n, \varphi_i \rangle = \mu_i, \quad i = 1, \ldots, n; \quad n = 1, 2, \ldots \tag{16}$$

Согласно теореме 2 проблема (16) при каждом фиксированном n разрешима тогда и только тогда, когда $\mu_n = 0$ или $g_{n*} > 0$, где $g_{n*} = \inf_{\langle \mu_n, c_n \rangle = 1} \left\| \sum_{i=1}^{n} c_n^i \varphi_i \right\|$, $\mu_n = (\mu_1, \ldots, \mu_n)$, $c_n = (c_n^1, \ldots, c_n^n)$.

Теорема 4. *Пусть B — сепарабельное банахово пространство. Для того чтобы проблема* (15) *имела решение, необходимо и достаточно, чтобы проблемы* (16) *были разрешимы при каждом $n \geqslant 1$ и $\sup_{n \geqslant 1} \|u_{n*}\| < \infty$, где u_{n*} — нормальное решение проблемы* (16).

Доказательство. Необходимость. Пусть u — решение проблемы (15). Тогда, очевидно, u является решением проблемы (16) при каждом $n \geqslant 1$. Из разрешимости проблемы (16) при каждом $n \geqslant 1$ и теоремы 2 следует существование нормального решения u_{n*}, $n = 1, 2, \ldots$. Тогда $\|u_{n*}\| \leqslant \|u\|$, $n = 1, 2, \ldots$, так что $\sup_{n \geqslant 1} \|u_{n*}\| \leqslant \|u\| < \infty$.

Достаточность. Пусть при каждом $n \geqslant 1$ проблема (16) разрешима и $\sup_{n \geqslant 1} \|u_{n*}\| < \infty$, где u_{n*} — нормальное решение проблемы (16). Так как пространство B сепарабельно, то из ограниченной последовательности $\{u_{n*}\}$ можно выбрать подпоследовательность $\{u_{n_k*}\}$, которая сходится к некоторому элементу $v \in B^*$ в следующем смысле: $\lim_{k \to \infty} \langle u_{n_k*}, \varphi \rangle = \langle v, \varphi \rangle \quad \forall \varphi \in B$ (см. [393], гл. IV, § 3). Отсюда и из равенств $\langle u_{n_k*}, \varphi_i \rangle = \mu_i$, $i = 1, \ldots, n_k$, при $k \to \infty$ получим $\langle v, \varphi_i \rangle = \mu_i$ для всех $i = 1, 2, \ldots$ Следовательно, v — решение проблемы (15). Теорема 4 доказана. □

Добавив к (15), (16) условие

$$\|u\| \leqslant l, \tag{17}$$

получим бесконечномерную l-проблему моментов и ее конечномерные аппроксимации соответственно. Аналогично теореме 4 доказывается

Теорема 5. *Пусть B — сепарабельное банахово пространство. Тогда l-проблема моментов* (15), (17) *разрешима тогда и только тогда, когда при каждом $n \geqslant 1$ разрешима l-проблема* (16), (17), *причем $\|u_{n*}\| \leqslant l$ $\forall n = 1, 2, \ldots$, где u_{n*} — нормальное решение проблемы* (16), (17).

При поиске решения каждой из конечномерных проблем (16) или (16), (17) можно действовать по схеме, изложенной в замечании 1. Может случиться, что при каком-то $n \geqslant 1$ проблема (16) или (16), (17) не имеет решения. Тогда согласно теоремам 4, 5 соответствующая бесконечномерная проблема (15) или (15), (17) также не имеет решения.

Теоремы существования решений бесконечномерной проблемы моментов для конкретных систем $\{\varphi_i, i = 1, 2, \ldots\}$ см., например, в [11; 122; 123; 166; 287; 400; 415; 784; 802].

4. Проиллюстрируем метод моментов на задаче управления динамической системой

$$\dot{x} = D(t)x + B(t)u(t), \quad 0 \leqslant t \leqslant T; \quad x(0) = 0, \tag{18}$$

где момент $T > 0$ задан, $A(t)$, $B(t)$ — кусочно-непрерывные на $[0, T]$ матрицы $n \times n$, $n \times r$ соответственно. Управление $u = (u^1(t), \ldots, u^r(t))$ будем называть допустимым, если $u^i(t)$ — ограниченная измеримая функция, почти всюду на отрезке $[0, T]$ удовлетворяющая неравенству

$$|u^i(t)| \leqslant l, \quad i = 1, \ldots, r, \quad l > 0. \tag{19}$$

Пусть задана точка $x_1 \in E^n$, $x_1 \neq 0$. Будем искать такое допустимое управление, что

$$x(T; u) = x_1, \tag{20}$$

где $x = x(t; u)$ — траектория системы (18), соответствующая управлению u.

Сведем задачу (18)–(20) к l-проблеме моментов (11) в подходящим образом выбранных банаховых пространствах B, B^*. С этой целью воспользуемся уже известным нам представлением (см. пример 3.6) решения системы (18):

$$x(t; u) = \int_0^t \Phi(t, \tau) B(\tau) u(\tau) d\tau, \quad 0 \leqslant t \leqslant T,$$

где матрица $\Phi(t, \tau)$ размера $n \times n$ определена условиями $\frac{d\Phi(t, \tau)}{dt} = D(t)\Phi(t, \tau)$, $0 \leqslant t, \tau \leqslant T$; $\Phi(\tau, \tau) = I$. Тогда условие (20) можем записать в виде уравнения

$$\int_0^T \Phi(T, \tau) B(\tau) u(\tau) d\tau = x_1. \tag{21}$$

Обозначим i-ю строку матрицы $\Phi(T, \tau)B(\tau)$ через

$$\varphi_i = \varphi_i(T, \tau) = (\varphi_i^1(T, \tau), \ldots, \varphi_i^r(T, \tau)), \quad 0 \leqslant \tau \leqslant T; \quad i = 1, \ldots, n.$$

Тогда уравнение (21) перепишется в следующей покоординатной форме

$$\langle u, \varphi_i \rangle = \int_0^T \left(\sum_{j=1}^{r} \varphi_i^j(T, \tau) u^j(\tau) \right) d\tau = x_1^i, \quad i = 1, \ldots, n. \tag{22}$$

Функции $\varphi_i = \varphi_i(T, \tau)$, $i = 1, \ldots, n$, кусочно-непрерывны на $[0, T]$, и поэтому можем считать, что $\varphi_i \in B = L_1^r[0, T]$ — банахово пространство с нормой $\|\varphi_i\|_B = \int_0^T \sum_{j=1}^{r} |\varphi_i^j(T, \tau)| d\tau$. Сопряженным к B является пространство $B^* = L_\infty^r[0, T]$ с нормой $\|u\|_{B^*} = \operatorname{ess\,sup}_{0 \leqslant t \leqslant T} \max_{1 \leqslant i \leqslant r} |u^i(t)|$ [258]. Условие (19) теперь можем записать в виде:

$$\|u\|_{B^*} \leqslant l, \tag{23}$$

результат $\langle u, \varphi_i \rangle$ применения функционала $u \in B^*$ к элементу $\varphi_i \in B$ определяется формулой (22). Тем самым задача (18)–(20) сведена к задаче (22), (23), представляющей собой l-проблему моментов (11) при $\mu = x_1 \neq 0$ на паре пространств $B = L_1^r[0, T]$, $B^* = L_\infty^r[0, T]$, и для ее исследования могут быть использованы теоремы 1, 2 по схеме, изложенной в замечании 1. Поскольку в задаче (2) целевая функция $g(c)$ удовлетворяет условию Липшица, то при решении этой задачи можно использовать подходящие модификации метода покрытий из § 5.13.

При вычислении значений функции $g(c) = \left\| \sum_{i=1}^{n} c^i \varphi_i(T, \tau) \right\|_B$ в фиксированной точке c необязательно иметь явное выражение для матрицы $\Phi(T, \tau)$. Дело в том, что здесь

$$\sum_{i=1}^{n} c^i \varphi_i(T, \tau) = (\Phi_i(T, \tau) B(\tau))^T c = B^T(\tau) \Phi^T(T, \tau) c = B^T(\tau) \psi(\tau; c),$$

где $\psi = \psi(\tau; c) = \Phi^T(T, \tau)c$ является решением задачи Коши: $\dot{\psi} = -D^T \psi$, $0 \leqslant \tau \leqslant T$, $\psi(T) = c$. Следовательно, $g(c) = \int_0^T \sum_{j=1}^{r} \left| \sum_{i=1}^{n} b_{ij}(\tau) \psi^i(\tau; c) \right| d\tau$.

Для более общих систем вида (10.5) задачу управления при условиях (19), (20) можно свести к l-проблеме моментов (22), (23) с помощью уже известного нам приема, изложенного в примере 10.1 (см. также примеры 2.15, 3.7). Если условия на управления заданы в виде

$$\alpha_i(t) \leqslant u^i(t) \leqslant \beta_i(t), \quad 0 \leqslant t \leqslant T, \quad i = 1, \ldots, r,$$

то можно перейти к новому управлению $v = (v^1, \ldots, v^r)$ по формулам

$$v^i = [u^i - \tfrac{1}{2}(\alpha_i(t) + \beta_i(t))]\frac{2}{\beta_i(t) - \alpha_i(t)}, \quad i = 1, \ldots, r,$$

с ограничениями вида (19): $|v^i(t)| \leqslant 1$, $i = 1, \ldots, r$.

Замечание 3. Предполагая, что матрицы $D(t)$, $B(t)$ из (18) определены при всех $t \geqslant 0$ и на каждом конечном отрезке $[0, T]$ кусочно непрерывны, рассмотрим следующую задачу быстродействия: найти минимальное время $T = T_*$ и управление $u = u(t)$, $0 \leqslant t \leqslant T_*$, удовлетворяющее условиям (19), при которых уравнение (20) имеет решение. Для поиска решения этой задачи можно воспользоваться методом моментов. По определению оптимального времени T_* задача (18)–(20) разрешима при $T = T_*$, а при всех T, $0 < T < T_*$ эта задача не имеет решения. Согласно теореме 3 тогда $g_* = g_*(T) = \inf\limits_{\langle x_1, c\rangle = 1} \left\| \sum\limits_{i=1}^{r} c^i \varphi_i(T, \tau) \right\|_B < \frac{1}{l}$ при всех T, $0 < T < T_*$, и $g_*(T_*) \geqslant \frac{1}{l}$. В случае непрерывности функции $g_*(T)$ это означает, что время T_* является минимальным корнем уравнения $g_*(T) = \frac{1}{l}$, и для поиска T_* могут быть использованы известные методы. К задачам быстродействия мы еще вернемся в § 10.6.

Замечание 4. Мы рассмотрели метод моментов для задачи (18)–(20) при специальном выборе «базисных» функций $\varphi_1(T, t), \ldots, \varphi_n(T, t)$, $0 \leqslant t \leqslant T$, порожденных самой системой (18) и уравнением (21). Разумеется, здесь можно использовать и другие базисные функции $\varphi_1(t), \ldots, \varphi_n(t), \ldots$ (например, тригонометрическую систему). Однако в этом случае мы уже не можем гарантировать, что получающаяся при этом l-проблема моментов будет конечномерной.

Замечание 5. При переходе от задачи (18)–(20) к l-проблеме моментов (22), (23) выбор пространств $B = L_1^r[0, T]$, $B^* = L_\infty^r[0, T]$ был продиктован видом ограничений (19). Приведем примеры других ограничений на управления, когда задачу управления (18), (20) также удается свести к l-проблеме моментов (22) с ограничением вида

$$\|u\|_{B^*} \leqslant l \tag{24}$$

на других парах пространств B, B^*.

Если в задаче (18), (20) ограничение на управление имеет вид

$$|u(t)|_p = \left(\sum_{i=1}^{r} |u^i(t)|^p \right)^{1/p} \leqslant l \quad \text{почти всюду на } [0, T], \quad 1 \leqslant p < \infty,$$

то такая задача сводится к l-проблеме моментов (22), (24) в банаховом пространстве B^* с нормой $\|u\|_{B^*} = \operatorname{ess\,sup}\limits_{0 \leqslant t \leqslant T} |u(t)|_p$, сопряженном к банахову пространству B с нормой $\|\varphi\|_B = \int\limits_0^T |\varphi(t)|_q dt$, где $\frac{1}{p} + \frac{1}{q} = 1$ [258].

Если в задаче (18), (20) ограничение на управление имеет вид

$$\left(\int\limits_0^T |u(t)|_\alpha^p dt \right)^{1/p} \leqslant l, \quad 1 \leqslant p < \infty, \quad 1 \leqslant \alpha \leqslant \infty,$$

то в (22), (24) в качестве B^* надо принять банахово пространство с нормой $\|u\|_B = \left(\int\limits_0^T |u(t)|_\alpha^p dt \right)^{1/p}$, сопряженное к банахову пространству B с нормой $\|\varphi\|_B = \left(\int\limits_0^T |\varphi(t)|_\beta^q dt \right)^{1/q}$, где $\frac{1}{p} + \frac{1}{q} = 1$, $\frac{1}{\alpha} + \frac{1}{\beta} = 1$ (случай $p = \alpha = 2$, $l = \infty$ рассмотрен в п. 3 § 10).

5. Применим метод моментов к следующей задаче управления колебаниями струны:

$$\begin{gathered} x_{tt} = x_{ss} + r(t)u(t), \quad (s, t) \in Q = (0, l) \times (0, T), \\ x|_{s=0} = x|_{s=l} = 0, \quad 0 \leqslant t \leqslant T; \quad x|_{t=0} = 0, \quad x_t|_{t=0} = 0, \quad 0 \leqslant s \leqslant l, \end{gathered} \tag{25}$$

где $r = r(t) \in L_2[0, l]$ — заданная функция. Требуется найти управление $u = u(t) \in L_2[0, T]$ такое, что

$$x|_{t=T} = y_0(s), \quad x_t|_{t=T} = y_1(s), \quad 0 \leqslant s \leqslant l, \tag{26}$$

где $y_0(s) \in H_0^1[0, l]$, $y_1(s) \in L_2[0, l]$ — заданные функции (см. пример 11.4).

Как было замечено в § 8, решение $x = x(s, t; u)$ краевой задачи (25) представимо в виде ряда (8.23):

$$x(s, t; u) = \sum_{k=1}^{\infty} \frac{r_k}{\sqrt{\lambda_k}} e_k(s) \int\limits_0^t u(\xi) \sin \sqrt{\lambda_k}(t - \xi) d\xi, \quad (s, t) \in Q, \tag{27}$$

где $e_k(s) = \sqrt{\frac{2}{l}} \sin \frac{\pi k s}{l}$ — собственная функция оператора $-\frac{d^2\varphi}{ds^2}$, $\varphi(0) = \varphi(l) = 0$, соответствующая собственному числу $\lambda_k = \left(\frac{\pi k}{l} \right)^2$. Как известно (см., например, [557]), система функций $\{e_k(s), 0 \leqslant s \leqslant l, k = 1, 2, \ldots\}$ образует ортонормированный базис в пространствах $L_2[0, l]$, $H_0^1[0, l]$, $H^{-1}[0, l]$, поэтому справедливы разложения

$$y_0(s) = \sum_{k=1}^{\infty} y_{0k} e_k(s), \quad y_1(s) = \sum_{k=1}^{\infty} y_{1k} e_k(s), \quad r(s) = \sum_{k=1}^{\infty} r_k e_k(s), \quad 0 \leqslant s \leqslant l, \tag{28}$$

где

$$y_{0k} = \int\limits_0^l y_0 e_k(s) ds, \quad y_{1k} = \int\limits_0^l y_1 e_k(s) ds, \quad r_k = \int\limits_0^l r(s) e_k(s) ds, \quad k = 1, 2, \ldots$$

Производная функции (27) по переменной t имеет вид

$$x_t(s, t; u) = \sum_{k=1}^{\infty} r_k e_k(s) \int\limits_0^t u(\xi) \cos \sqrt{\lambda_k}(t - \xi) d\xi, \quad (s, t) \in Q. \tag{29}$$

Отметим, что ряд (27) сходится в норме $\max\limits_{0 \leqslant t \leqslant T} \left(\int\limits_0^l y_s^2(s, t) ds \right)^{1/2}$, ряд (29) — в норме $\max\limits_{0 \leqslant t \leqslant T} \left(\int\limits_0^l y^2(s, t) ds \right)^{1/2}$ (подробнее см., например, в [557]).

Подставим выражения (27)–(29) в равенства (26), умножим получившиеся равенства на $e_j(s)$, $j=1,2,\ldots$, скалярно в $L_2[0,l]$. С учетом ортонормированности системы $\{e_k\}$ в $L_2[0,l]$ получим систему уравнений для искомого управления $u=u(t)$:

$$\frac{r_k}{\sqrt{\lambda_k}}\int_0^T u(\xi)\sin\sqrt{\lambda_k}(T-\xi)d\xi = y_{0k},$$

$$r_k\int_0^T u(\xi)\cos\sqrt{\lambda_k}(T-\xi)d\xi = y_{1k}, \quad k=1,2,\ldots$$

Отсюда, считая, что

$$r_k\neq 0 \quad \forall k=1,2,\ldots \tag{30}$$

имеем

$$\begin{aligned}&\int_0^T u(\xi)\cos\sqrt{\lambda_k}(T-\xi)d\xi = \frac{y_{1k}}{r_k}=a_k,\\ &\int_0^T u(\xi)\sin\sqrt{\lambda_k}(T-\xi)d\xi = \frac{y_{0k}\sqrt{\lambda_k}}{r_k}=b_k, \quad k=1,2,\ldots\end{aligned} \tag{31}$$

Система (31) представляет собой бесконечномерную проблему моментов (15) относительно $u=u(t)\in B^*=L_2[0,T]$, в которой роль элементов $\{\varphi_k\}$ играют функции $\{\cos\sqrt{\lambda_k}(T-\xi), \sin\sqrt{\lambda_k}(T-\xi), k=1,2,\ldots\}$, принадлежащие пространству $B=L_2[0,T]$.

Предполагая, что наряду с (30) выполняются условия

$$\sum_{k=1}^{\infty}(a_k^2+b_k^2)=\sum_{k=1}^{\infty}\left[\left(\frac{y_{1k}}{r_k}\right)^2+\left(\frac{y_{0k}}{r_k}\right)^2\lambda_k\right]<\infty, \tag{32}$$

покажем, что проблема моментов (31) и, следовательно, задача управления (25), (26) разрешима при всех $T\geqslant 2l$.

Сначала рассмотрим случай $T=2l$. В интегралах (31) сделаем замену переменной $T-\xi=\tau$. Учитывая, что $T=2l$, получим

$$\begin{aligned}&\int_0^{2l} u(2l-\tau)\cos\sqrt{\lambda_k}\tau d\tau = a_k,\\ &\int_0^{2l} u(2l-\tau)\sin\sqrt{\lambda_k}\tau d\tau = b_k, \quad k=1,2,\ldots\end{aligned} \tag{33}$$

Поскольку $\left\{\frac{1}{\sqrt{l}}\cos\sqrt{\lambda_k}\tau, \frac{1}{\sqrt{l}}\sin\sqrt{\lambda_k}\tau\right\}$, где $\lambda_k=\left(\frac{\pi k}{l}\right)^2$, является ортонормированной системой в $L_2[0,2l]$, то по теореме Рисса — Фишера [393] при условии (32) функция

$$u_*(2l-\tau)=\frac{1}{\sqrt{l}}\sum_{k=1}^{\infty}(a_k\cos\sqrt{\lambda_k}\tau+b_k\sin\sqrt{\lambda_k}\tau), \quad 0\leqslant\tau\leqslant 2l \tag{34}$$

является решением проблемы моментов (33). Ряд (34) сходится в норме $L_2[0,2l]$, и его сумма $u_*(2l-\tau)\in L_2[0,2l]$. Перейдем в (34) к переменной $t=2l-\tau$. Учитывая, что $\cos\sqrt{\lambda_k}(2l-t)=\cos(2\pi k-\frac{\pi k}{l}t)=\cos\sqrt{\lambda_k}t$, $\sin\sqrt{\lambda_k}(2l-t)=\sin(2\pi k-\frac{\pi k}{l}t)=-\sin\sqrt{\lambda_k}t$, получим

$$u_*(t)=\frac{1}{\sqrt{l}}\sum_{k=1}^{\infty}(a_k\cos\sqrt{\lambda_k}t-b_k\sin\sqrt{\lambda_k}t), \quad 0\leqslant t\leqslant 2l, \tag{35}$$

где коэффициенты a_k, b_k взяты из (31). Заметим, что

$$\int_0^{2l} u_*(t)dt=0, \tag{36}$$

так как функции $\{\cos\sqrt{\lambda_k}t, \sin\sqrt{\lambda_k}t, k=1,2,\ldots\}$ ортогональны в $L_2[0,2l]$ к элементу $e_0(s)\equiv 1$. Тем самым показано, что при $T=2l$ задача управления (25), (26) имеет решение (35), обладающее свойством (36).

Рассмотрим случай $T>2l$. Нетрудно проверить, что функция

$$v(t)=\begin{cases}0, & 0\leqslant t\leqslant T-2l,\\ u_*(T-t), & T-2l\leqslant t\leqslant T,\end{cases} \tag{37}$$

является решением проблемы (31) и задачи управления (25), (26) при $T>2l$. Покажем, что при $T>2l$ задача (25), (26) имеет другое, в отличие от (37), $2l$-периодическое решение. А именно, пусть $T=2lN+\alpha$, $0\leqslant\alpha<2l$, где $N\geqslant 1$ — целое число. Через $v_*(t)$ обозначим $2l$-периодическую функцию, которая на одном периоде $[0,2l]$ определяется следующим образом:

$$v_*(t)=\begin{cases}\frac{1}{N+1}u_*(t), & 0\leqslant t\leqslant\alpha,\\ \frac{1}{N}u_*(t), & \alpha<t\leqslant 2l,\end{cases} \tag{38}$$

где функция $u_*(t)$ взята из (35). Убедимся, что $v_*(t)$, $0\leqslant t\leqslant T$, — решение проблемы моментов (31). Поскольку функции $v_*(t)$, $\cos\sqrt{\lambda_k}(T-\xi)$ являются $2l$-периодическими, то их произведение $v_*(t)\cos\sqrt{\lambda_k}(T-\xi)$ также $2l$-периодично. Поэтому, учитывая, что $u_*(t)$ — решение проблемы (33), имеем

$$\begin{aligned}\int_0^T v_*(\xi)\cos\sqrt{\lambda_k}(T-\xi)d\xi &= \int_0^{2Nl} v_*(\xi)\cos\sqrt{\lambda_k}(T-\xi)d\xi +\\ &+\int_{2Nl}^{2Nl+\alpha} v_*(\xi)\cos\sqrt{\lambda_k}(T-\xi)d\xi =\\ &= N\int_0^{2l} v_*(\xi)\cos\sqrt{\lambda_k}(T-\xi)d\xi+\int_0^{\alpha} v_*(\xi)\cos\sqrt{\lambda_k}(T-\xi)d\xi =\\ &= (N+1)\int_0^{\alpha} v_*(\xi)\cos\sqrt{\lambda_k}(T-\xi)d\xi+N\int_{\alpha}^{2l} v_*(\xi)\cos\sqrt{\lambda_k}(T-\xi)d\xi =\\ &= \int_0^{2l} u_*(\xi)\cos\sqrt{\lambda_k}(T-\xi)d\xi = a_k, \quad k=1,2,\ldots\end{aligned}$$

Аналогично доказывается, что $\int_0^T v_*(\xi)\sin\sqrt{\lambda_k}(T-\xi)d\xi=b_k$, $k=1,2,\ldots$; $\int_0^T v_*(t)dt=\int_0^T u_*(t)dt=0$. Таким образом, функция (38) является решением задачи управления (25), (26). Можно показать [802], что (38) — нормальное решение этой задачи:

Применения метода моментов к различным задачам управления читатель найдет, например, в [11; 122; 123; 166; 287; 400; 411; 415; 784; 802].

Упражнения

1. Пользуясь методом моментов, решить задачи управления из упражнения 10.3, считая, что управление удовлетворяет ограничениям 1) $|u(t)| \leqslant l$; 2) $0 \leqslant u(t) \leqslant l$; 3) $\int_0^T u^2(t)dt \leqslant l$ и, кроме того, $x(0) = y(0) = 0$, $x(T) = f_1$, $y(T) = f_2$.

2. Примените метод моментов к задачам управления из примеров 10.2–10.4, а также упражнений 10.9–10.13.

ГЛАВА 9

Методы решения неустойчивых задач оптимизации

При численном решении прикладных задач важное значение имеет тот факт, будет ли решение рассматриваемой задачи непрерывно зависеть от исходных данных или, иначе говоря, будет ли искомое решение устойчивым по отношению к возмущениям входных данных. Если решение устойчиво по входным данным, то можно быть уверенным в том, что достаточно малые погрешности в задании входных данных приведут к малым погрешностям в определении решения. Иное дело решать неустойчивую или, как еще говорят, некорректную задачу, решение которой не является непрерывно зависящим от входных данных: в этом случае приближенное решение задачи, соответствующее неточным входным данным, может как угодно сильно отличаться от искомого точного решения при сколь угодно малых погрешностях входных данных. Между тем неустойчивые задачи не такая уж большая редкость — они часто встречаются в самых различных областях физики, техники, экономики и т. д. [695]. Возникает важная проблема: как решать такие задачи?

Основы теории и методов решения неустойчивых задач заложены в работах А. Н. Тихонова, В. К. Иванова, М. М. Лаврентьева и др. К настоящему времени создана достаточно полная общая теория неустойчивых (некорректных) задач, разработаны приближенные методы решения таких задач. Из обширной литературы, в которой отражены современная теория неустойчивых задач, методы их решения, приложения, исторические аспекты, библиография, мы здесь в состоянии упомянуть лишь ее малую часть [17; 22; 23; 28; 60; 62–64; 91; 113; 115; 127; 130; 131; 142; 145–147; 149–153; 155–165; 167–174; 176–182; 184; 185; 188–190; 192; 215; 223; 224; 229; 230; 235–238; 242; 268–271; 275; 334; 335; 360; 363; 365; 367; 389; 391; 394; 403; 405; 421; 432; 438; 439; 443; 445; 450–455; 461; 462; 467; 474; 490; 501; 506–510; 519; 522; 524; 526; 537; 556; 557; 579; 592; 593; 596; 597; 599; 600; 618; 619; 621; 622; 625–628; 651; 658; 659; 679; 680; 691–693; 695–697; 708; 714; 740; 741; 757; 758; 764; 771; 782; 783; 788; 799; 800; 805; 808; 812; 817].

В этой главе мы рассмотрим лишь некоторые из методов решения неустойчивых задач минимизации.

§ 1. Постановка задачи. Устойчивые и неустойчивые задачи минимизации

Будем рассматривать задачу

$$J(u) \to \inf, \quad u \in U, \tag{1}$$

$$U = \{u \in U_0:\ g_i(u) \leqslant 0,\ i = 1, \ldots, m;\ g_i(u) = 0,\ i = m+1, \ldots, s\}, \tag{2}$$

где U_0 — заданное множество из некоторого метрического пространства $\mathcal{M}$ с метрикой $\rho(u, v)$, функции $J(u)$, $g_i(u)$, $1, \ldots, s$, определены на U_0 и принимают на этом множестве конечные значения. В (2) не исключаются возможности, когда отсутствуют либо ограничения типа неравенств ($m = 0$), либо типа равенств ($s = m$), либо оба вида таких ограничений ($m = s = 0$, $U = U_0$). Пусть $U \neq \varnothing$. Напоминаем (см. § § 1.1, 2.1), что задачу (1), (2) можно истолковать либо как *задачу первого типа*, когда ищется величина $J_* = \inf_U J(u)$ (или приближение к J_*) и при этом не предполагается, что указанная нижняя грань непременно достигается в какой-либо точке, либо как *задачу второго типа*, когда ищется не только J_*, но какая-либо точка $u_* \in U_0$, близкая ко множеству $U_* = \{u \in U:\ J(u) = J_*\}$, подразумевая при этом, что $J_* > -\infty$, $U_* \neq \varnothing$.

Будем считать, что множество U_0 известно точно (например, $U_0 = \mathcal{M}$), а вместо функций $J(u)$, $g_i(u)$ известны их приближения $J_\delta(u)$, $g_{i\delta}(u)$ такие, что

$$|J_\delta(u) - J(u)| \leqslant \Psi_0(\delta, u), \quad |g_{i\delta}(u) - g_i(u)| \leqslant \Psi_i(\delta, u) \quad \forall u \in U_0, \quad i = 1, \ldots, s, \tag{3}$$

где $\delta = (\delta_1, \delta_2, \ldots, \delta_q) > 0$ — параметры погрешности функции $\Psi_i(\delta, u)$, $i = 0, \ldots, s$, характеризующие уровень погрешности *входных данных* $J(u)$, $g_i(u)$, $i = 1, \ldots, s$, определены и неотрицательны при всех $\delta > 0$, $u \in U_0$, $\lim\limits_{\delta \to 0} \Psi_i(\delta, u) = 0$, $i = 0, \ldots, s$, $\forall u \in U_0$. Например, возможно, $\Psi_i(\delta, u) = |\delta| = \bar{\Psi}_i(\delta, u) = \delta_1 + \ldots + \delta_q$ или $\Psi_i(\delta, u) = |\delta|(1 + \Omega(u))$, $u \in U_0$, $i = 0, \ldots, s$, где $\Omega(u) \geqslant 0$ — известная функция. Зависимость функции Ψ_i от переменной u отражает нередко возникающую на практике ситуацию, когда в разных точках $u \in U_0$ измерения величин проводятся с разной точностью. При необходимости можем считать, что $0 < \delta \leqslant \delta_0 = (\delta_{01}, \ldots, \delta_{0q}) < \infty$ (т. е. $0 < \delta_i \leqslant \delta_{0i}$), где δ_{0i} — верхняя оценка для разумно мыслимых значений погрешности δ_i, $i = 1, \ldots, q$.

Пользуясь приближениями $J_\delta(u)$, $g_{i\delta}(u)$ из (3), по аналогии с (1), (2) мы можем составить задачу

$$J_\delta(u) \to \inf, \quad u \in U_\delta, \tag{4}$$

$$U_\delta = \{u \in U_0 \colon g_{i\delta}(u) \leqslant 0, \ i = 1, \ldots, m; \ g_{i\delta}(u) = 0, \ i = m+1, \ldots, s\}, \tag{5}$$

называемую *возмущенной задачей*. Допустим, что множество U_δ непусто и нам удалось определить величину $J_{\delta *}$, а в случае $J_{\delta *} > -\infty$ еще и какую-либо точку $u_{\delta *} \in U_{\delta *} = \{u \in U_\delta \colon J_\delta(u) = J_{\delta *}\}$. Возникает естественный вопрос: можно ли при малых δ использовать величину $J_{\delta *}$ и точку $u_{\delta *} \in U_{\delta *}$ в качестве приближения для J_* и U_* соответственно, зная, что в силу (3) погрешность задания входных данных стремится к нулю при $\delta \to 0$, т. е. можно ли ожидать, что $\lim\limits_{\delta \to 0} J_{\delta *} = J_*$ и $\lim\limits_{\delta \to 0} \rho(u_{\delta *}, U_*) = 0$? Для ответа на этот вопрос рассмотрим простенький пример.

Пример 1. Задача:

$$J(u) \equiv 0 \to \inf, \quad u \in U = E^1.$$

Очевидно, здесь $J_* = 0$, $U_* = E^1$. Пусть приближения $J_\delta(u)$ таковы, что

$$|J_\delta(u) - J(u)| \leqslant \Psi_0(\delta, u) \equiv \delta, \quad \delta > 0 \quad \forall u \in E^1. \tag{6}$$

Тогда возмущенная задача имеет вид:

$$J_\delta(u) \to \inf, \quad u \in U_\delta = U = E^1.$$

Нетрудно видеть, что $|J_{\delta *}(u) - J_*(u)| = |J_{\delta *}(u)| \leqslant \delta \ \forall \delta > 0$. Это означает, что при достаточно малом $\delta > 0$ величина $J_{\delta *}$ вполне может служить приближением для $J_* = 0$ независимо от выбора $J_\delta(u)$ из (6). Посмотрим, можно ли при условиях (6) использовать решение $u_{\delta *}$ возмущенной задачи в качестве приближения ко множеству U_*. Для этого мы прежде всего должны быть уверены, что множество $U_{\delta *}$ непусто. Однако нетрудно видеть, что выполнение такого требования можно гарантировать далеко не для всех реализаций $J_\delta(u)$ из (6). Так, например, если $J_\delta(u) = \delta e^{-|u|}$, то $U_{\delta *} = \varnothing \ \forall \delta > 0$. Заметим, что эта функция непрерывна на E^1. А если реализации $J_\delta(u)$ из (6) будут

разрывными, то ясно, что ситуация лишь ухудшится. Рассмотрим случай, когда в рассматриваемой задаче погрешность зависит от u. А именно, пусть

$$|J_\delta(u) - J(u)| \leqslant \Psi_0(\delta, u) \equiv \delta(1 + |u|), \quad \delta > 0 \quad \forall u \in E^1. \tag{7}$$

Условию (7) удовлетворяет функция $J_\delta(u) = -\delta(1 + |u|)$. Очевидно, тогда $J_{\delta *} = -\infty \ \forall \delta > 0$. Полученное значение $J_{\delta *}$ трудно признать высокоточным приближением для $J_* = 0$, какими бы малыми ни были числа $\delta > 0$. Нетрудно указать реализации приближения $J_\delta(u)$ из условия (7), для которых $J_{\delta *} > -\infty$. Однако и в этом случае возможно, что $J_{\delta *}$ не будет близким к $J_* = 0$. Так, например, если $J_\delta(u) = -\delta |u| e^{-\delta |u|}$, то условие (7) выполняется и $J_{\delta *} = J_\delta\left(\frac{1}{\delta}\right) = -e^{-1} \ \forall \delta > 0$, но $\lim\limits_{\delta \to 0} J_{\delta *} = e^{-1} \neq 0 = J_*$. Конечно, из (7) можно извлечь и такие реализации $J_\delta(u)$ (например, $J_\delta(u) \equiv 0$ или любые $J_\delta(u)$ из (6)), когда $\lim\limits_{\delta \to 0} J_{\delta *} = J_*$. Однако выборка реализаций $J_\delta(u)$ носит, как правило, случайный характер, и у нас нет оснований полагать, что попавшая в наше распоряжение реализация $J_\delta(u)$ из (7) непременно будет хорошей. Таким образом, при условиях (7) величина $J_{\delta *}$, полученная из возмущенной задачи, вообще говоря, не может служить хорошим приближением для J_*.

Из анализа рассмотренного примера 1 уже можно сделать некоторые выводы, касающиеся общей задачи (1), (2). Выяснилось, что конкретные реализации приближений $J_\delta(u)$, $g_{i\delta}(u)$ из (3), используемые для формирования возмущенной задачи (4), (5), могут быть «хорошими» или «плохими», и в зависимости от этого задача (4), (5) будет «хорошей» или «плохой». Решение «хорошей» возмущенной задачи может быть использовано для получения приближенного решения задачи (1), (2), а «плохая» задача для этих целей не годится. Однако у нас нет никаких критериев для оценки качества приближения, кроме условий (3), и мы не в состоянии отличить «хорошее» приближение от «плохого». Поэтому при попытке использовать возмущенную задачу (4), (5) для построения приближенного решения исходной задачи (1), (2) приходится иметь в виду всевозможные (включая самые «худшие») реализации приближений входных данных.

Кроме того, заметим, что использование элементов множества $U_{\delta *}$ в качестве приближений ко множеству U_* предполагает, что $U_{\delta *} \neq \varnothing$. Однако, как видно из примера 1, множество $U_{\delta *}$ может оказаться пустым даже в простейших задачах минимизации с точно заданным множеством. Имея в виду это обстоятельство, вместо множества $U_{\delta *}$ разумнее пользоваться множеством

$$U_{\delta\varepsilon} = \{u \in U_\delta \colon J_\delta(u) \leqslant J_{\delta *} + \varepsilon\}, \quad \varepsilon > 0, \tag{8}$$

так как, во-первых, это множество непусто по определению нижней грани при всех реализациях $J_\delta(u)$, $g_{i\delta}(u)$ из (3), лишь бы $U_\delta \neq \varnothing$, $J_{\delta *} > -\infty$, и, во-вторых, если все же $U_{\delta *} \neq \varnothing$, то $U_{\delta *} \subset U_{\delta\varepsilon} \ \forall \varepsilon > 0$. Для иллюстрации этих соображений рассмотрим

Пример 2. Задача: $J(u) = u^2 \to \inf$, $u \in U = E^1$. Здесь $J_* = 0$, $U_* = \{u_* = 0\}$. Пусть приближение $J_\delta(u)$ таково, что

$$|J_\delta(u) - J(u)| \leqslant \delta(1 + u^2), \quad 0 < \delta < 1, \quad \forall u \in E^1.$$

Это означает, что $u^2 - \delta(1 + u^2) = J_-(u) \leqslant J_\delta(u) \leqslant J_+(u) = u^2 + \delta(1 + u^2)$ $\forall u \in E^1$. Отсюда переходя к нижней грани по $u \in E^1$, получим $-\delta \leqslant J_{\delta *} = \inf\limits_{E^1} J_\delta(u) \leqslant \delta$, так что $\lim\limits_{\delta \to 0} J_{\delta *} = J_* = 0$. Кроме того, пользуясь графиками

функций $J_-(u)$, $J_+(u)$, нетрудно показать, что $U_{\delta\varepsilon} \in \left[-\sqrt{\frac{2\delta+\varepsilon}{1-\delta}}; \sqrt{\frac{2\delta+\varepsilon}{1-\delta}}\right]$ при всех $0<\delta<1$, $\varepsilon>0$, где множество $U_{\delta\varepsilon}$ определено согласно (8). Отсюда следует, что $\sup\limits_{u\in U_{\delta\varepsilon}} |u-u_*| \leqslant \sqrt{\frac{2\delta+\varepsilon}{1-\delta}} \to 0$ при $(\delta,\varepsilon)\to 0$. Это означает, что при достаточно малых δ, ε любая точка $u_{\delta\varepsilon}\in U_{\delta\varepsilon}$ может быть использована в качестве приближения к точке $u_*=0$. В то же время множество $U_{\delta *}=\{u\in E^1\colon J_\delta(u)=J_{\delta *}\}$ вполне может оказаться пустым.

Заметим также, что если в рассматриваемом примере приближения $J_\delta(u)$ удовлетворяет условию $|J_\delta(u)-J(u)|\leqslant\delta(1+|u|^\gamma)$, $u\in E^1$, $\gamma>2$, то в худшем случае может реализоваться выборка $J_\delta(u)=u^2-\delta(1+|u|^\gamma)$. Тогда $J_{\delta *}=-\infty$ $\forall\delta>0$, и о близости $J_{\delta *}$ к $J_*=0$ говорить не приходится, а множество (8) не имеет смысла.

Приведенные примеры показывают, что возмущенная задача (4), (5) в каких-то случаях вполне может быть использована для получения приближенного решения задачи (1), (2), в каких-то случаях — нет. Это зависит от того, устойчива задача (1), (2) к возмущениям входных данных $J(u)$, $g_i(u)$, $i=1,\ldots,s$, или неустойчива. Перейдем к строгим определениям, предполагая, что в (3) погрешности $\Psi_i(\delta,u)$, $i=0,\ldots,s$, фиксированы.

Определение 1. Задачу (1), (2) называют *устойчивой по функции*, если при любом выборе $J_\delta(u)$, $g_{i\delta}(u)$, $i=1,\ldots,s$, из условий (3) в возмущенной задаче (4), (5), множество $U_\delta\neq\varnothing$, $\forall\delta>0$ и справедливо равенство

$$\lim_{\delta\to 0} J_{\delta *}=J_* \tag{9}$$

(возможно $J_*=-\infty$ здесь не исключается). Если существуют $J_\delta(u)$, $g_{i\delta}(u)$, $i=1,\ldots,s$, удовлетворяющие условиям (3), для которых либо $U_\delta=\varnothing$ при некоторых сколь угодно малых $\delta>0$, либо $U_\delta\neq\varnothing$ при всех δ, $0<\delta<\delta_0$, но не выполняется равенство (9), то задача (1), (2) называется *неустойчивой по функции*.

Определение 2. Пусть в задаче (1), (2) $J_*>-\infty$, $U_*\neq\varnothing$. Эту задачу называют *устойчивой по аргументу*, если она устойчива по функции и

$$\lim_{(\delta,\varepsilon)\to 0}\ \sup_{u\in U_{\delta\varepsilon}}\ \inf_{v\in U_*}\rho(u,v)=0, \tag{10}$$

где множество $U_{\delta\varepsilon}$ определено согласно (8). Задача (1), (2) называется *неустойчивой по аргументу*, если она неустойчива по функции или не соблюдается равенство (10).

Поясним геометрический смысл равенства (10). Величина $\inf\limits_{v\in U_*}\rho(u,v)=\rho(u,U_*)$ — расстояние от точки u до множества U_* — нам уже знакома. Равенство означает, что для любого числа $\gamma>0$ найдутся такие числа $\delta_0>0$, $\varepsilon_0>0$, что при всех δ, ε, $0<|\delta|<\delta_0$, $0<\varepsilon<\varepsilon_0$ множество $U_{\delta\varepsilon}$ принадлежит множеству $\{u\in\mathcal{M}\colon \rho(u,U_*)\leqslant\gamma\}$, называемому *$\gamma$-раздутием* множества U_* (см. ниже упражнения 6, 7).

Если задача (1), (2) устойчива по функции, то, как следует из (9), при любых реализациях $J_\delta(u)$, $g_{i\delta}(u)$ из (3) величину $J_{\delta *}$ при достаточно малых $\delta>0$ можно взять в качестве приближенного решения этой задачи, рассматриваемой как задача первого типа. Если задача (1), (2) устойчива по аргументу, то, как следует из равенства (10), любая точка $u_{\delta\varepsilon}\in U_{\delta\varepsilon}$ при достаточно малых δ, ε будет близкой ко множеству U_* и может быть взята в качестве приближенного решения задачи (1), (2) второго типа.

На практике точно вычислить значение $J_{\delta *}$ удается лишь в редких случаях. Обычно приходится довольствоваться тем, что с помощью того или иного метода минимизации, примененного к задаче (4), (5), определяем какую-либо точку $u_\delta\in U_\delta$, для которой

$$J_{\delta *}\leqslant J_\delta(u_\delta)\leqslant J_{\delta *}+\mu(\delta),\quad \mu(\delta)>0,$$

где предполагается, что $J_{\delta *}>-\infty$, $\lim\limits_{\delta\to 0}\mu(\delta)=0$. Тогда, если задача (1), (2) устойчива по функции [по аргументу], то при малых $\delta>0$ величина $J_\delta(u_\delta)$ близка к J_* [точка $u_{\delta\varepsilon}\in U_{\delta\varepsilon}$ близка ко множеству U_*].

Проблеме устойчивости задач оптимизации посвящена обширная литература, см., например, [45; 52; 70; 84; 87; 88; 179; 181; 233; 277; 363; 448; 501; 502; 720; 789; 791; 817]. Здесь мы приведем две теоремы, выделяющие подклассы устойчивых задач (1), (2). А именно, рассмотрим задачу

$$J(u)\to\inf,\quad u\in U, \tag{11}$$

считая множество U известным точно (в (2) $m=s=0$, $U=U_0$), а приближения $J_\delta(u)$ функции $J(u)$ удовлетворяет условию

$$|J_\delta(u)-J(u)|\leqslant\delta\quad\forall u\in U,\quad \delta>0. \tag{12}$$

Теорема 1. *Задача* (11) *при выполнении условий* (12) *устойчива по функции.*

Доказательство. По условию $U_\delta=U$ $\forall\delta>0$. Неравенство (12) перепишем в виде

$$J(u)-\delta\leqslant J_\delta(u)\leqslant J(u)+\delta\quad\forall u\in U=U_\delta,\quad \delta>0. \tag{13}$$

Если $J_*=-\infty$, то из (13) следует, что $J_{\delta *}=\inf\limits_U J_\delta(u)=-\infty$, так что равенство (9) соблюдается. Если $J_*>-\infty$, то, переходя в неравенствах (13) почленно к нижней грани при $u\in U_\delta=U$, получим: $J_*-\delta\leqslant J_{\delta *}\leqslant J_*+\delta$ $\forall\delta>0$. Отсюда при $\delta\to 0$ приходим к равенству (9) и в этом случае.

Теорема 2. *Пусть U — компактное множество из $\mathcal{M}$, функция $J(u)$ полунепрерывна снизу на U, приближения $J_\delta(u)$ удовлетворяют условию* (12), *$U_\delta=U$ $\forall\delta>0$. Тогда задача* (11) *устойчива по аргументу.*

Доказательство. В силу теоремы 8.2.1 $J_*>-\infty$, $U_*\neq\varnothing$. Устойчивость задачи (11) по функции установлена в теореме 1. Из $J_*>-\infty$ и равенства (9) следует, что $J_{\delta *}>-\infty$ при всех достаточно малых $\delta>0$. Тогда множество $U_{\delta\varepsilon}\neq\varnothing$ $\forall\varepsilon>0$ по определению нижней грани. Возьмем произвольную последовательность $\{(\delta_k,\varepsilon_k)\}\to 0$, $\delta_k>0$, $\varepsilon_k>0$, $k=1,2,\ldots$ Для краткости положим $U_{\delta_k\varepsilon_k}=U_k$. По определению верхней грани найдется точка $u_k\in U_k$ такая, что

$$\sup_{u\in U_k}\rho(u,U_*)-\frac{1}{k}\leqslant\rho(u_k,U_*),\quad k=1,2,\ldots \tag{14}$$

Включение $u_k\in U_k$ согласно (8) означает, что $u_k\in U_{\delta_k\varepsilon_k}$ и для некоторой реализации $J_{\delta_k}(u)$ из (12) выполняются неравенства $J_{\delta_k *}=\inf\limits_U J_{\delta_k}(u)\leqslant J_{\delta_k}(u_k)\leqslant J_{\delta_k *}+\varepsilon_k$, $k=1,2,\ldots$ Отсюда и из теоремы 1 следует, что $\lim\limits_{k\to\infty} J_{\delta_k}(u_k)=J_*$. Тогда в силу (12) $\lim\limits_{k\to\infty} J(u_k)=J_*$. Это означает, что $\{u_k\}$ — минимизирующая последовательность задачи (11). По теореме 8.2.1 тогда $\lim\limits_{k\to\infty}\rho(u_k,U_*)=0$. Отсюда и из (14) имеем $\lim\limits_{k\to\infty}\sup\limits_{u\in U_k}\rho(u_k,U_*)=\lim\limits_{k\to\infty}\sup\limits_{u\in U_{\delta_k\varepsilon_k}}\rho(u,U_*)=0$. В

силу произвольности последовательности $(\delta_k, \varepsilon_k) \to 0$ отсюда непосредственно вытекает равенство (10). Теорема 2 доказана.

Для иллюстрации вышеизложенного рассмотрим несколько примеров.

Пример 3. Задача: $J(u) = \frac{u^2}{1+u^4} \to \inf$, $u \in U = E^1$. Здесь $J_* = 0$, $U_* = \{u_* = 0\}$. Пусть приближения $J_\delta(u)$ удовлетворяют условию (12), $U_\delta = E^1$ $\forall \delta > 0$. По теореме 1 эта задача устойчива по функции. Покажем, что она неустойчива по аргументу. Возьмем реализацию $J_\delta(u) = J(u)$, $u \in E^1$. Тогда $J_{\delta *} = J_* = 0$. Точка $u_\varepsilon = \frac{1}{\sqrt{\varepsilon}}$ принадлежит множеству (8), так как $J_\delta(u_\varepsilon) = \frac{\varepsilon}{1+\varepsilon^2} \leqslant \varepsilon = J_{\delta *} + \varepsilon$. Однако $\sup\limits_{u \in U_{\delta\varepsilon}} \rho(u, U_*) \geqslant \rho(u_\varepsilon, U_*) = |u_\varepsilon| \to \infty$ при $\varepsilon \to 0$. Равенство (10) не выполняется. Отметим, что в этой задаче множество U некомпактно.

Пример 4. Рассмотрим задачу оптимального управления: $J(u) = \int\limits_0^1 |x(t; u)|dt \to \inf$, $\dot{x}(t) = u(t)$, $0 \leqslant t \leqslant 1$, $x(0) = 0$, $u = u(t) \in U = \{u(t) \in L_2[0, 1]: |u(t)| \leqslant 1$ почти всюду на $[0, 1]\}$. Здесь $J_* = 0$, $U_* = \{u_*(t) \equiv 0\}$. Пусть приближение $J_\delta(u)$ удовлетворяет условию (12), $U_\delta = U$ $\forall \delta > 0$. В силу теоремы 1 задача устойчива по функции. Рассмотрим возмущенную задачу: $J_\delta(u) = \int\limits_0^1 |\delta \sin \frac{t}{\delta} - x(t; u)|dt \to \inf$, $u \in U_\delta = U$. Тогда $J_{\delta *} = 0$, $U_{\delta *} = \{u_{\delta *} = \cos \frac{t}{\delta}, t \in [0, 1]\}$. Однако $\sup\limits_{u \in U_{\delta\varepsilon}} \|u - u_*\|^2_{L_2} \geqslant \|u_{\delta *} - u_*\|^2_{L_2} = \int\limits_0^1 \cos^2 \frac{t}{\delta} dt = \frac{1}{2}\left(1 + \frac{\delta}{2} \sin \frac{2}{\delta}\right) \to \frac{1}{2}$ при $\delta \to 0$.

Задача неустойчива по аргументу в метрике $L_2[0, 1]$.

Пример 5. Задача: $J(u) = \int\limits_0^1 u^2(t)dt \to \inf$, $u = u(t) \in U = \{u \in C[0, 1]: |u(t)| \leqslant 1 \ \forall t \in [0, 1]\}$. Очевидно, $J_* = 0$, $U_* = \{u_* \equiv 0\}$. Пусть приближения $J_\delta(u)$ удовлетворяют условию (12), $U_\delta = U$. Задача устойчива по функции в силу теоремы 1. В возмущенной задаче: $J_\delta(u) = J(u) \to \inf$, $u \in U_\delta$, последовательность

$$u_k = u_k(t) = \begin{cases} 1 - |2kt - 1|, & 0 \leqslant t \leqslant \frac{1}{k}, \\ 0, & \frac{1}{k} \leqslant t \leqslant 1, \ k = 1, 2, \ldots \end{cases}$$

является минимизирующей. Отсюда следует, что $u_k \in U_{\delta\varepsilon}$ при $\forall \varepsilon > 0$, если номер k достаточно велик. Однако $\lim\limits_{(\delta, \varepsilon) \to 0} \sup\limits_{u \in U_{\delta\varepsilon}} \|u - u_*\|_C \geqslant \lim\limits_{k \to \infty} \|u_k - u_*\|_C = 1$. Поэтому задача неустойчива по аргументу в метрике $C[0, 1]$. Нетрудно проверить, что при тех же условиях эта задача будет устойчива по аргументу в метрике $L_2[0, 1]$.

Заметим, что в теоремах 1, 2 и примерах 1–5 предполагалось, что множество (2) известно точно. Если же множество также задается с погрешностями, то задача (1), (2) может оказаться неустойчивой (по функции или по аргументу) даже при дополнительном требовании компактности множества. Более того, множество (5) может оказаться пустым при всех сколь угодно малых δ, и возмущенная задача (4), (5) не будет иметь смысла. Покажем это на примерах.

Пример 6. Задача: $J(u) = u \to \inf$, $u \in U = \{u \in U_0 = E^1: g_1(u) = \frac{u^2}{1+u^4} \leqslant 0\}$. Здесь $U = \{0\}$. $J_* = 0$, $U_* = \{0\}$. Пусть функции $J_\delta(u)$, $g_{1\delta}(u)$ удовлетворяют условиям (3) с функциями $\Psi_0(\delta, u) = \Psi_1(\delta, u) = \delta$. В частности, если $J_\delta(u) = u$, $g_{1\delta}(u) = g_1(u) - \delta$, то $U_\delta \neq \varnothing$, $J_{\delta *} = -\infty$ $\forall \delta > 0$. Если же $g_{1\delta}(u) = g_1(u) + \delta$, то $U_\delta = \{u \in E^1: g_{1\delta}(u) \leqslant 0\} = \varnothing$ $\forall \delta > 0$, и возмущенная задача теряет смысл. Задача неустойчива по функции и тем более по аргументу.

Пример 7. Задача: $J(u) = u \to \inf$, $u \in U = \{u \in U_0 = E^1: g_1(u) = |u - 1| + |u + 1| - 2 = 0\}$. Здесь $U = [-1, +1]$, $J_* = -1$, $U_* = \{u_* = -1\}$. Пусть условия (3) выполняются с функциями $\Psi_0(\delta, u) = \Psi_1(\delta, u) = 2\delta(1 + |u|)$. В возмущенной задаче: $J_\delta(u) = u \to \inf$, $u \in U_\delta = \{u \in E^1: g_{1\delta}(u) = |(1+\delta)u - 1| + |(1-\delta)u + 1| - 2 = 0\}$ множество U_δ при всех δ, $0 < \delta < 1$, состоит из двух точек $u = 0$ и $u = 1$, так что $J_{\delta *} = 0$, $U_{\delta *} = \{u_{\delta *} = 0\}$. Равенство (9) нарушается, задача неустойчива даже по функции. Заметим, что если $g_{1\delta}(u) = |u-1| + |u+1| - 2 + \delta$, то множество $U_\delta = \{u \in E^1: g_{1\delta}(u) = 0\} = \varnothing$ $\forall \delta > 0$, возмущенная задача не имеет смысла.

Пример 8. Рассмотрим задачу линейного программирования: $J(u) = -x - y \to \inf$, $u \in U = \{u = (x, y) \in U_0: g_1(u) = x - y \leqslant 0, g_2(u) = -x + y \leqslant 0\}$, где $U_0 = \{u = (x, y) \in E^2: 0 \leqslant x \leqslant 1, 0 \leqslant y \leqslant 2\}$ — компактное множество. Здесь $J_* = -2$, $U_* = \{u_* = (1; 1)\}$. Пусть условия (3) выполняются с функциями $\Psi_i(\delta, u) = \delta(1 + |u|)$, $i = 0, 1, 2$. Возмущенная задача может иметь вид: $J_\delta(u) = -x - y \to \inf$, $u \in U_\delta = \{u \in U_0: g_{1\delta}(u) = (1+\delta)x - y \leqslant 0, g_{2\delta}(u) = (-1+\delta)x + y \leqslant 0\}$, где $0 < \delta < 1$. Тогда $U_\delta = \{u = (0, 0)\}$, $J_{\delta *} = 0$, $U_{\delta *} = U_\delta$. Равенство (9) не выполняется. Задача неустойчива по функции. Более того, если $g_{1\delta}(u) = g_1(u) + \delta$, $g_{2\delta}(u) = g_2(u) + \delta$ или $g_{1\delta}(u) = (1+\delta)x - y + \delta$, $g_{2\delta}(u) = -(1-\delta)x + y + \delta$, то множество $U_\delta = \{u \in U_0: g_{1\delta}(u) \leqslant 0, g_{2\delta}(u) \leqslant 0\}$ пусто при всех δ, $0 < \delta < 1$, и возмущенная задача не имеет смысла.

Пример 9. Рассмотрим задачу квадратичного программирования:

$$J(u) = x^2 + y^2 + z^2 \to \inf, \quad u \in U = \{u = (x, y, z) \in E^3:$$
$$g_1(u) = 1 - x - y - z \leqslant 0, \ g_2(u) = y + z - 1 \leqslant 0, \ g_3(x) = x \leqslant 0\}.$$

Если $u = (x, y, z) \in U$, то $y + z \geqslant 1 - x = 1 + |x| \geqslant 1 \geqslant y + z$, что возможно только при $x = 0$, $y + z = 1$. Это значит, что $U \subseteq U_1 = \{u = (x, y, z) \in E^3: y + z = 1, x = 0\}$. Очевидно, $U_1 \subseteq U$. Следовательно, $U = U_1$. Нетрудно проверить, что решением рассматриваемой задачи является точка $u_* = (x_* = 0, y_* = \frac{1}{2}, z_* = \frac{1}{2})$, причем $J(u_*) = J_* = \frac{1}{2}$. Так как множество U выпукло, а функция $J(u)$ сильно выпукла на U, то задача других решений не имеет. Пусть возмущенная задача имеет вид: $J_\delta(u) = J(u) = x^2 + y^2 + z^2 \to \inf$, $u \in U_\delta = \{u \in E^3: g_{1\delta}(u) = g_1(u) = 1 - x - y - z \leqslant 0, g_{2\delta}(u) = y + (1+\delta)z - 1 \leqslant 0, g_{3\delta}(u) = g_3(u) = x \leqslant 0\}$, $\delta > 0$. Покажем, что $U_\delta = U_{\delta 1} = \{u \in E^3: 1 - x - y - z \leqslant 0, y + (1+\delta)z - 1 \leqslant 0, x \leqslant 0, y \geqslant 1, z \leqslant 0\}$ $\forall \delta > 0$. Очевидно, $U_{\delta 1} \subseteq U_\delta$. Обратно, если $u = (x, y, z) \in U_\delta$, то $y + z \geqslant 1 - x = 1 + |x| \geqslant 1 \geqslant y + (1+\delta)z$, что возможно только при $0 \geqslant \delta z$ или $0 \geqslant z$. А тогда $y \geqslant 1 - x - z = 1 + |x| + |z| \geqslant 1$. Следовательно, $U_\delta \subseteq U_{\delta 1}$, так что $U_\delta = U_{\delta 1}$ $\forall \delta > 0$. Нетрудно проверить, что решением возмущенной задачи является точка $u_{\delta *} = (x_{\delta *} = 0, y_{\delta *} = 1, z_{\delta *} = 0)$. В самом деле, $J'(u_{\delta *}) = (0, 2, 0)$ и $\langle J'(u_{\delta *}), u - u_{\delta *} \rangle = 2(y - 1) \geqslant 0$ $\forall u \in U_{\delta 1} = U_\delta$ (теорема 4.2.3). Кроме того, $J_{\delta *} = J(u_{\delta *}) = 1$. Таким образом, $|J_{\delta *} - J_*| = \frac{1}{2}$, $\rho(u_{\delta *}, u_*) = \frac{1}{\sqrt{2}}$ $\forall \delta > 0$. Задача неустойчива ни по функции, ни по аргументу.

Приведенные примеры показывают, что неустойчивые задачи нередки в линейном, выпуклом и невыпуклом программировании, в области оптимального управления, в теории приближения. Если задача неустойчива по функции [или по аргументу], то при решении задачи первого [второго] типа пользоваться возмущенной задачей (4), (5) нужно с большой осторожностью, помня, что здесь многое зависит от того, какие реализации входных данных $J_\delta(u)$, $g_{i\delta}(u)$ из (3) имеются в нашем распоряжении, и что в общем случае на этом пути получение хороших приближений решения исходной задачи (1), (2) не гарантировано.

Возникает естественный вопрос: можно ли разумно распорядиться имеющимися у нас входными данными $J_\delta(u)$, $g_{i\delta}(u)$ из (3) и с их помощью сконструировать новую вспомогательную экстремальную задачу, решив которую мы могли бы при достаточно малых погрешностях получить решение исходной задачи с высокой точностью? Ответ на этот важный для практики вопрос дается в следующих параграфах, в которых излагаются различные устойчивые методы решения неустойчивых задач минимизации, разработанные в рамках общей теории и методов неустойчивых (некорректных) задач [184; 439; 695; 697 и др.].

Упражнения

1. Доказать, что задача: $J(u)=-u\to\inf$, $u\in U=\{u\in E^1_+=U_0 : g_1(u)=-u\leqslant 0,\ g_2(u)=u-1\leqslant 0\}$ устойчива по функции и по аргументу, если погрешности задания входных данных $J(u)$, $g_1(u)$, $g_2(u)$ удовлетворяют условиям (3) с функциями $\Psi_i(\delta,u)=\delta(1+|u|)$, $i=0,1,2$.

2. Исследовать на устойчивость задачу: $J(u)=(x-1)^2+(0\cdot y-1)^2\to\inf$, $u=(x,y)\in U=E^2$, считая, что погрешность задания функции $J(u)$ удовлетворяет условию: $|J_\delta(u)-J(u)|\leqslant\delta(1+|u|^2)$. Указание: рассмотреть приближение $J_\delta(u)=(x-1)^2+(\delta y-1)^2$.

3. Доказать, что задача: $J(u)=x-y\to\inf$, $u\in U=\{u=(x,y)\geqslant 0:\ g_1(u)=x+y-1\leqslant 0,\ g_2(u)=-x-y+1\leqslant 0\}$ неустойчива по функции, если погрешности задания входных данных $J(u)$, $g_1(u)$, $g_2(u)$ удовлетворяют условиям (3) при $\Psi_i(\delta,u)=\delta(1+|x|+|y|)$, $u\in E^2_+$. Указание: рассмотреть возмущенную задачу (4), (5) при $J_\delta(u)=x-y$, $g_{1\delta}(u)=(1+\delta)x+(1+\delta)y-1+\delta$, $g_{2\delta}(u)=(-1+\delta)x+(-1+\delta)y+1-\delta$, $0<\delta<1$.

4. Исследовать на устойчивость задачу: $J(u)=u\to\inf$, $u\in U=\{u\in E^1:\ g_1(u)=|u-1|+|u+1|-2\leqslant 0\}$, считая, что входные данные удовлетворяют условиям (3) с $\Psi_i(\delta,u)=2\delta(1+|u|)$, $i=0,1$. Указание: рассмотреть реализации $J_\delta(u)$, $g_{1\delta}(u)$ из примера 7.

5. Исследовать на устойчивость по функции и по аргументу задачу: $J(u)=x^2+y^2\to\inf$, $u=(x,y)\in E^2$, когда погрешность $|J_\delta(u)-J(u)|\leqslant\Psi(\delta,u)=\delta(1+|u|^\gamma)$, $\gamma\geqslant 0$. Рассмотреть случаи $\gamma=0,1,2,3$.

6. Пусть A, B — два множества из некоторого метрического пространства $\mathcal{M}$. *Уклонением* множества A от множества B называется величина $\beta(A,B)=\sup\limits_{a\in A}\inf\limits_{b\in B}\rho(a,b)$. Приведите пример двух множеств A, B таких, что $\beta(A,B)\neq\beta(B,A)$. Указание: рассмотрите случай, когда $B\subset A$, $B\neq A$ (см. равенство (10)).

7. Пусть A, B — два множества из метрического пространства $\mathcal{M}$. Величина $h(A,B)=\max\{\beta(A,B);\beta(B,A)\}$ (см. упражнение 6) называется *хаусдорфовым расстоянием* между множествами A и B. Докажите, что на множестве всех ограниченных замкнутых множеств из $\mathcal{M}$ величина $h(A,B)$ удовлетворяет всем аксиомам метрического пространства [192; 350; 393; 428]. Выясните геометрический смысл величины $h(A,B)$ (см. § 10.5).

§ 2. Методы регуляризации для решения неустойчивых задач первого типа

Рассмотрим задачу

$$J(u)\to\inf,\quad u\in U, \tag{1}$$

$$u\in U=\{u\in U_0:\ g_i(u)\leqslant 0,\ i=1,\ldots,m;\ g_i(u)=0,\ i=m+1,\ldots,s\}, \tag{2}$$

как задачу первого типа, когда ищется величина $J_*=\inf\limits_U J(u)$. Будем предполагать, что множество (2) непусто, причем множество U_0 известно точно, а вместо функций $J(u)$, $g_i(u)$ известны их приближения, удовлетворяющие условиям (1.3). Выше было показано, что задача (1), (2) может оказаться неустойчивой по функции (определение 1.1), и тогда величина $J_{\delta*}=\inf\limits_{U_\delta}J_\delta(u)$, полученная как решение возмущенной задачи (1.4), (1.5) первого типа, вообще говоря, не может служить хорошим приближением для величины J_*, какими бы малыми ни были погрешности задания входных данных. Для надежного получения решения неустойчивой по функции задачи (1), (2) первого типа нужно иметь специальные методы, называемые *методами регуляризации*. Такие методы для решения задач первого типа разрабатывались, например, в [445; 501; 507; 697; 817]. Изложим некоторые из них.

1. Сначала рассмотрим задачу:

$$J(u)\to\inf,\quad u\in U, \tag{3}$$

предполагая, что множество U известно точно, а приближения $J_\delta(u)$ функции $J(u)$ удовлетворяют условию

$$|J_\delta(u)-J(u)|\leqslant\Psi(\delta,u)\quad\forall u\in U;\quad\delta>0, \tag{4}$$

где функция $\Psi(\delta,u)\geqslant 0$ $\forall u\in U$, $\delta>0$, $\lim\limits_{\delta\to 0}\Psi(\delta,u)=0$ $\forall u\in U$. Следуя [445; 697; 817], рассмотрим вспомогательную задачу:

$$\chi_\delta(u)=J_\delta(u)+\Psi(\delta,u)\to\inf,\quad u\in U, \tag{5}$$

где $J_\delta(u)$ — произвольная функция, удовлетворяющая условию (4). Предположим, что каким-либо методом нам удалось найти величину $\chi_{\delta*}(u)=\inf\limits_{u\in U}\chi_\delta(u)$.

Теорема 1. *Пусть множество U в задаче* (3) *известно точно, пусть выполнено условие* (4). *Тогда при любом выборе $J_\delta(u)$ из* (4) *справедливо равенство*

$$\lim_{\delta\to 0}\chi_{\delta*}=J_*. \tag{6}$$

Доказательство. Из (4), (5) имеем: $\chi_\delta(u)\geqslant J(u)$ $\forall u\in U$. Следовательно, $\chi_{\delta*}\geqslant J_*\geqslant-\infty$. Тогда $J_*\leqslant\chi_{\delta*}\leqslant\chi_\delta(u)=J_\delta(u)+\Psi(\delta,u)\leqslant J(u)+2\Psi(\delta,u)$ $\forall u\in U$, $\forall\delta>0$. При $\delta\to 0$ отсюда получаем: $J_*\leqslant\varliminf\limits_{\delta\to 0}\chi_{\delta*}\leqslant\varlimsup\limits_{\delta\to 0}\chi_{\delta*}\leqslant J(u)+\lim\limits_{\delta\to 0}2\Psi(\delta,u)=J(u)$ $\forall u\in U$. Перейдем в этих неравенствах к нижней грани по $u\in U$. Будем иметь: $-\infty\leqslant J_*\leqslant\varliminf\limits_{\delta\to 0}\chi_{\delta*}\leqslant\varlimsup\limits_{\delta\to 0}\chi_{\delta*}\leqslant J_*$. Следовательно, $\lim\limits_{\delta\to 0}\chi_{\delta*}=\varlimsup\limits_{\delta\to 0}\chi_{\delta*}=J_*$, что равносильно (6). Теорема 1 доказана. □

Равенство (6) означает, что величина $\chi_{\delta *}$, полученная как решение задачи (5) при достаточно малых $\delta > 0$, будет близка к J_δ даже в том случае, когда исходная задача (3) была неустойчива по функции. Однако точное значение $\chi_{\delta *}$ можно найти лишь в редких случаях, и на практике мы можем иметь лишь какое-то приближение к $\chi_{\delta *}$. В свою очередь такое приближение можно получить, решая задачу (3), лишь тогда, когда задача (5) сама устойчива по функции. Будет ли она таковой? Оказывается, что при малых погрешностях, допускаемых при вычислении значений функции $\chi_\delta(u)$, задача (5) будет устойчивой по функции. А именно, пусть вместо точного значения $\chi_\delta(u)$ мы можем вычислить его приближение $\chi_{\delta\mu}(u)$ такое, что

$$|\chi_{\delta\mu}(u) - \chi_\delta(u)| \leqslant \mu \quad \forall u \in U. \tag{7}$$

Тогда, как следует из теоремы 1.1, задача (5) устойчива по функции, т. е. $\lim\limits_{\mu \to 0} \chi_{\delta\mu *} = \chi_{\delta *}$, где $\chi_{\delta\mu *} = \inf\limits_U \chi_{\delta\mu}(u)$. Более того, поскольку в (7) величина μ не зависит от δ, нетрудно установить более сильное утверждение:

$$\lim_{(\delta, \mu) \to 0} \chi_{\delta\mu *} = J_*. \tag{8}$$

Докажем его. Переходя к нижней грани по $u \in U$, из (7) имеем: $|\chi_{\delta\mu *} - \chi_{\delta *}| \leqslant \mu$. Тогда

$$|\chi_{\delta\mu *} - J_*| \leqslant |\chi_{\delta\mu *} - \chi_{\delta *}| + |\chi_{\delta *} - J_*| \leqslant \mu + |\chi_{\delta *} - J_*|.$$

Отсюда и из (6) следует равенство (8).

Допустим, что нам удалось найти величину $\widetilde{\chi}_{\delta\mu}$, для которой

$$|\widetilde{\chi}_{\delta\mu} - \chi_{\delta\mu *}| \leqslant \varepsilon(\mu), \quad \lim_{\mu \to 0} \varepsilon(\mu) = 0. \tag{9}$$

Из (8), (9) непосредственно следует, что при достаточно малых (δ, μ) величину $\widetilde{\chi}_{\delta\mu}$ можно взять в качестве приближения для искомой величины J_*. Для поиска $\widetilde{\chi}_{\delta\mu}$ может быть использована устойчивая по функции при условиях (7) задача (5).

2. Рассмотрим задачу (1), (2), когда множество U известно неточно и погрешность задания входных данных $J(u)$, $g_i(u)$, $i = 1, \ldots, s$, удовлетворяет условиям (1.3):

$$|J_\delta(u) - J(u)| \leqslant \Psi_0(\delta, u) \quad \forall u \in U_0 \tag{10}$$

$$|g_{i\delta}(u) - g_i(u)| \leqslant \Psi_i(\delta, u) \quad \forall u \in U_0, \quad i = 1, \ldots, s, \tag{11}$$

где функции $\Psi_i(\delta, u) \geqslant 0$ $\forall \delta > 0$, $\forall u \in U_0$, $\lim\limits_{\delta \to 0} \Psi_i(\delta, u) = 0$ $\forall u \in U_0$, $i = 0, \ldots, s$. Для учета ограничений типа равенств и неравенств из (2) воспользуемся штрафными функциями (см. § 5.15). Ограничимся рассмотрением простейшей штрафной функции:

$$P(u) = \sum_{i=1}^{s} (g_i^+(u))^p, \quad u \in U_0, \quad p > 0, \tag{12}$$

где $g_i^+(u) = \max\{0; g_i(u)\}$ при $i = 1, \ldots, m$, $g_i^+(u) = |g_i(u)|$ при $i = m+1, \ldots, s$. При выполнении условий (11) в качестве приближения для функции $P(u)$ можно взять функцию

$$P_\delta(u) = \sum_{i=1}^{s} (g_{i\delta}^+(u))^p, \quad u \in U_0, \quad p > 0. \tag{13}$$

Оценим разность $|P_\delta(u) - P(u)|$. Заметим, что из неравенства (5.14.15): $|a^+ - b^+| \leqslant |a - b|$ следует

$$|g_{i\delta}^+(u) - g_i^+(u)| \leqslant |g_{i\delta}(u) - g_i(u)| \leqslant \Psi_i(\delta, u) \quad \forall u \in U_0, \quad \delta > 0, \quad i = 1, \ldots, s. \tag{14}$$

Поскольку

$$|a^\nu - b^\nu| \leqslant |a - b|^\nu \quad \forall a \geqslant 0, \quad b \geqslant 0, \quad 0 < \nu \leqslant 1, \tag{15}$$

то с учетом (12)–(14) имеем

$$|P_\delta(u) - P(u)| \leqslant \sum_{i=1}^{s} \Psi_i(\delta, u) \quad \forall u \in U_0, \quad \delta > 0, \quad 0 < p \leqslant 1. \tag{16}$$

В случае $p > 1$ воспользуемся формулой конечных приращений Лагранжа для функции $\varphi(x) = x^p$, $x \in E_+^1$. Получим $|a^p - b^p| = p|b + \theta(a-b)|^{p-1}|a-b| \leqslant p(|b| + |a-b|)^{p-1})|a-b|$. $\forall a \geqslant 0$, $b \geqslant 0$. Отсюда и из (11)–(13) вытекает:

$$|P_\delta(u) - P(u)| \leqslant \sum_{i=1}^{s} p(g_i^+(u) + \Psi_i(\delta, u))^{p-1} \Psi_i(\delta, u) \quad \forall u \in U_0, \quad \delta > 0, \quad p > 1. \tag{17}$$

Из (16), (17) следует, что для штрафных функций (12), (13) при условиях (11) справедлива оценка

$$|P_\delta(u) - P(u)| \leqslant \widetilde{\Psi}(\delta, u) \quad \forall u \in U_0, \quad \delta > 0, \tag{18}$$

где

$$\widetilde{\Psi}(\delta, u) = \begin{cases} \sum\limits_{i=1}^{s} \Psi_i(\delta, u) & \text{при } 0 < p \leqslant 1, \\ \sum\limits_{i=1}^{s} p(g_i^+(u) + \Psi_i(\delta, u))^{p-1} \Psi_i(\delta, u) & \text{при } p > 1. \end{cases} \tag{19}$$

Заметим, что в дальнейшем мы не будем пользоваться явным видом (13), (19) функций $P_\delta(u)$, $\widetilde{\Psi}(\delta, u)$, нам будет достаточно того, что некоторая функция $P_\delta(u)$ удовлетворяет неравенству (18), где $\widetilde{\Psi}(\delta, u) \geqslant 0$ $\forall u \in U_0$, $\delta > 0$, $\lim\limits_{\delta \to 0} \widetilde{\Psi}(\delta, u) = 0$ $\forall u \in U_0$. Введем функцию

$$\chi_\delta(u) = J_\delta(u) + A(\delta) P_\delta(u) + \Psi_0(\delta, u) + A(\delta) \widetilde{\Psi}(\delta, u), \quad u \in U_0, \quad \delta > 0, \tag{20}$$

где $J_\delta(u)$, $P_\delta(u)$ какие-либо функции, удовлетворяющие условиям (10), (18), $A(\delta) > 0$ — штрафной коэффициент. Рассмотрим задачу

$$\chi_\delta(u) \to \inf, \quad u \in U_0. \tag{21}$$

Приведем достаточные условия, гарантирующие равенство

$$\lim_{\delta \to 0} \chi_{\delta *} = J_*, \quad \chi_{\delta *} = \inf_{u \in U_0} \chi_\delta(u). \tag{22}$$

Мы здесь не будем стремиться к формулировкам возможно общих условий, при которых справедливо равенство (22), а ограничимся рассмотрением уже

знакомого нам класса задач (1), (2) с сильно согласованной постановкой (см. определение 5.15.3). А именно, пусть существуют постоянные $c_1 \geqslant 0, \ldots, c_s \geqslant 0$, $\nu > 0$, такие, что

$$J_* \leqslant J(u) + \sum_{i=1}^{s} c_i (g_i^+(u))^\nu, \quad u \in U_0. \tag{23}$$

Условия, обеспечивающие выполнение неравенства (23), приведены в леммах 5.15.1, 5.15.5 (см. также пример 5.15.8), которые сохраняют силу и в метрических пространствах.

Теорема 2. *Пусть выполнены условия* (10), (18), (23), *параметры* p, ν *из* (12), (23) *таковы, что* $p \geqslant \nu$. *Пусть* $A(\delta) > 0$, $\lim\limits_{\delta \to 0} A(\delta) = +\infty$, $\lim\limits_{\delta \to 0} A(\delta)\widetilde{\Psi}(\delta, u) = 0 \ \forall u \in U_0$. *Тогда справедливо равенство* (22).

Доказательство. Из условия (23) вытекает неравенство

$$J_* \leqslant J(u) + A(\delta)P(u) + BA^{-\frac{\nu}{p-\nu}} \quad \forall u \in U_0, \quad \delta > 0, \quad p \geqslant \nu, \tag{24}$$

где при $p = \nu$ по определению полагается $A^{-\frac{\nu}{p-\nu}} = 0$, $B = 0$ и считается, что $A(\delta) \geqslant |c| = \max\limits_{1 \leqslant i \leqslant s} |c_i|$, а при $p > \nu$ здесь $B = (p-\nu)\nu^{\frac{\nu}{p-\nu}} p^{-\frac{p}{p-\nu}} |c|^{\frac{p}{p-\nu}}$, $|c| = \left(\sum\limits_{i=1}^{s} |c_i|^{\frac{p}{p-\nu}}\right)^{\frac{p-\nu}{p}}$. Для доказательства (24) при $p > \nu$ надо воспользоваться рассуждениями, использованными при переходе от неравенства (5.15.25) к (5.15.26), с заменой A_k на $A(\delta)$, функции $\Phi_k(u) = J(u) + A_k P(u)$ на $\Phi_\delta(u) = J(u) + A(\delta)P(u)$. При $p = \nu$ из (12) и (23) непосредственно имеем $J_* \leqslant J(u) + \max\limits_{1 \leqslant i \leqslant s} |c_i| P(u) \leqslant J(u) + A(\delta)P(u) \ \forall u \in U_0$. Из (24) с учетом неравенств (10), (18) и определения (20) функции $\chi_\delta(u)$ следует

$$J_* \leqslant J_\delta(u) + A(\delta)P_\delta(u) + \Psi_0(\delta, u) + A(\delta)\widetilde{\Psi}(\delta, u) + B(A(\delta))^{-\frac{\nu}{p-\nu}} = \chi_\delta(u) + B(A(\delta))^{-\frac{\nu}{p-\nu}} \quad \forall u \in U_0.$$

Тогда $J_* \leqslant \chi_{\delta *} + B(A(\delta))^{-\frac{\nu}{p-\nu}}$. Отсюда, снова обращаясь к (10), (18), (20), получаем

$$J_* - B(A(\delta))^{-\frac{\nu}{p-\nu}} \leqslant \chi_{\delta *} \leqslant \chi_\delta(u) \leqslant J(u) + A(\delta)P(u) + 2\Psi_0(\delta, u) + 2A(\delta)\widetilde{\Psi}(\delta, u) \quad \forall u \in U_0, \quad \delta > 0. \tag{25}$$

Эти неравенства справедливы и для точек $u \in U$. Так как $P(u) = 0$ при $u \in U$, то из (25) имеем: $J_* - B(A(\delta))^{-\frac{\nu}{p-\nu}} \leqslant \chi_{\delta *} \leqslant J(u) + 2\Psi_0(\delta, u) + 2A(\delta)\widetilde{\Psi}(\delta, u)$ $\forall u \in U$, $\forall \delta > 0$. Отсюда и из свойств функций $\Psi_0(\delta, u)$, $\widetilde{\Psi}(\delta, u)$ при $\delta \to 0$ следует: $J_* \leqslant \underline{\lim\limits_{\delta \to 0}} \chi_{\delta *} \leqslant \overline{\lim\limits_{\delta \to 0}} \chi_{\delta *} \leqslant J(u) \ \forall u \in U$. Перейдем в этих неравенствах к нижней грани по $u \in U$. Получим $\underline{\lim\limits_{\delta \to 0}} \chi_{\delta *} = \overline{\lim\limits_{\delta \to 0}} \chi_{\delta *} = J_*$, что равносильно (22). Теорема 2 доказана. Случай $J_* = -\infty$ в (22) не исключается. □

Отметим, что если погрешности при вычислении значений функции (20) таковы, что $|\chi_{\delta\mu}(u) - \chi_\delta(u)| \leqslant \mu \ \forall u \in U_0$ (ср. с (7)), то из теоремы 1.1 и равенств (22) вытекает, что задача (21) устойчива по функции и вполне может быть использована для получения приближенного значения искомой величины J_*.

3. Рассмотрим метод регуляризации для решения неустойчивых задач (1),(2) первого типа, не использующий штрафные функции и основанный на идее расширения множества. Предположим, что погрешности задания входных данных удовлетворяют условиям (10),(11). Введем множество

$$W(\delta, \theta) = \{u \in U_0: \ g_{i\delta}(u) \leqslant \Psi_i(\delta, u) + \theta, \ i = 1, \ldots, m, \ |g_{i\delta}(u)| \leqslant \Psi_i(\delta, u) + \theta, \ i = m+1, \ldots, s\} = \{u \in U_0: \ g_{i\delta}^+(u) \leqslant \Psi_i(\delta, u) + \theta, \ i = 1, \ldots, s\}, \ \delta > 0, \ \theta \geqslant 0. \tag{26}$$

Убедимся, что $W(\delta, \theta) \neq \varnothing \ \forall \delta > 0, \ \theta \geqslant 0$. Возьмем произвольную точку $u \in U$. Тогда $g_i^+(u) = 0$ и из (14) имеем $g_{i\delta}^+(u) \leqslant g_i^+(u) + \Psi_i(\delta, u) \leqslant \Psi_i(\delta, u) + \theta$, $i = 1, \ldots, s$. Отсюда следует, что $U \subset W(\delta, \theta) \ \forall \delta > 0, \ \theta \geqslant 0$, т. е. $W(\delta, \theta)$ — расширение множества U. Далее, предполагая, что задача (1), (2) имеет сильно согласованную постановку и нам известны величины c_i, ν из (23) (или их оценки сверху), введем функцию

$$\chi_\delta(u) = J_\delta(u) + \Psi_0(\delta, u) + \sum_{i=1}^{s} c_i (2\Psi_i(\delta, u))^\nu, \quad u \in U_0, \tag{27}$$

и рассмотрим задачу

$$\chi_\delta(u) \to \inf, \quad u \in W(\delta, \theta). \tag{28}$$

Теорема 3. *Пусть выполнены условия* (10), (11), (23), *пусть* $0 < \nu \leqslant 1$. *Тогда*

$$\lim_{(\delta, \theta) \to 0} \chi_{\delta *}(\theta) = J_*, \quad \lim_{\delta \to 0} \chi_{\delta *}(0) = J_*, \tag{29}$$

где $\chi_{\delta *}(\theta) = \inf\limits_{u \in W(\delta, \theta)} \chi_\delta(u) \ \forall \delta > 0, \ \theta \geqslant 0$.

Доказательство. Из (14), (26) следует

$$g_i^+(u) \leqslant g_{i\delta}^+(u) + \Psi_i(\delta, u) \leqslant 2\Psi_i(\delta, u) + \theta \quad \forall u \in W(\delta, \theta), \quad i = 1, \ldots, s.$$

Отсюда с учетом условий (10), (23), определения (27) функции $\chi_\delta(u)$ и неравенства (15) имеем

$$J_* \leqslant J(u) + \sum_{i=1}^{s} c_i (2\Psi_i(\delta, u) + \theta)^\nu \leqslant J_\delta(u) + \Psi_0(\delta, u) + \sum_{i=1}^{s} c_i (2\Psi_i(\delta, u))^\nu + \sum_{i=1}^{s} c_i |(2\Psi_i(\delta, u) + \theta)^\nu - (2\Psi_i(\delta, u))^\nu| \leqslant \chi_\delta(u) + |c|_1 \theta^\nu \quad \forall u \in W(\delta, \theta),$$

где $|c|_1 = c_1 + \ldots + c_s$. Переходя в этих неравенствах к нижней грани по $u \in W(\delta, \theta)$, получим $J_* \leqslant \chi_{\delta *}(\theta) + |c|_1 \theta^\nu \ \forall \delta > 0, \ \theta \geqslant 0$. Тогда

$$J_* \leqslant \chi_{\delta *}(\theta) + |c|_1 \theta^\nu \leqslant \chi_\delta(u) + |c|_1 \theta^\nu \leqslant J(u) + 2\Psi_0(\delta, u) + \sum_{i=1}^{s} c_i (\Psi_i(\delta, u))^\nu + |c|_1 \theta^\nu \quad \forall u \in W(\delta, \theta). \tag{30}$$

В частности, эти неравенства справедливы для всех $u \in U \subseteq W(\delta, \theta)$. При $(\delta, \theta) \to 0$ с учетом свойств функций $\Psi_i(\delta, u)$, $i = 0, \ldots, s$, из (30) имеем

$$J_* \leqslant \underline{\lim_{(\delta, \theta) \to 0}} \chi_{\delta *}(\theta) \leqslant \overline{\lim_{(\delta, \theta) \to 0}} \chi_{\delta *}(\theta) \leqslant J(u) \quad \forall u \in U.$$

В силу произвола в выборе $u \in U$ это возможно лишь тогда, когда $\underline{\lim}_{(\delta,\theta)\to 0} \chi_{\delta *}(\theta) = \overline{\lim}_{(\delta,\theta)\to 0} \chi_{\delta *}(\theta) = J_*$, что равносильно первому равенству (29). Опираясь на неравенства (30) при $\theta = 0$, аналогичными рассуждениями приходим ко второму равенству (29). Теорема 3 доказана. □

Выясним, будет ли сама задача (26)–(28) устойчивой по функции. Сразу заметим, что при $\theta = 0$ эта задача может оказаться неустойчивой. В самом деле, например, может случиться, что $g_{i\delta}(u) = g_i(u) + \Psi_i(\delta, u)$, $\delta > 0$, $i = 1, \ldots, m$, $m = s$. Тогда $W(\delta, 0) = U$ и, как мы видели в примерах 1.7, 1.8, множество $W(\delta, 0)$ может оказаться пустым при малейших возмущениях входных данных $g_{i\delta}(u)$, и задача (26)–(28) потеряет смысл. Рассмотрим случай $\theta > 0$. Предположим, что вместо точных значений $J_\delta(u)$, $g^+_{i\delta}(u)$, $\Psi_i(\delta, u)$ мы можем вычислить лишь их приближения $J_{\delta\mu}(u)$, $g^+_{i\delta\mu}(u)$, $\Psi_{i\mu}(\delta, u)$, для которых

$$|J_{\delta\mu}(u) - J_\delta(u)| \leqslant \mu, \quad |g^+_{i\delta\mu}(u) - g^+_{i\delta}(u)| \leqslant \mu, \quad i = 1, \ldots, s,$$
$$|\Psi_{i\mu}(\delta, u) - \Psi_i(\delta, u)| \leqslant \mu, \ i = 0, \ldots, s, \ \forall u \in U_0, \ \delta > 0, \ \theta > 0, \ \mu > 0. \tag{31}$$

Тогда вместо функции $\chi_\delta(u)$ из (27) будем иметь

$$\chi_{\delta\mu}(u) = J_{\delta\mu}(u) + \Psi_{0\mu}(\delta, u) + \sum_{i=1}^{s} c_i (2\Psi_{i\mu}(\delta, u))^\nu, \quad u \in U_0. \tag{32}$$

Отсюда с учетом (31) и неравенства (15) получаем

$$|\chi_{\delta\mu}(u) - \chi_\delta(u)| \leqslant 2\mu + \sum_{i=1}^{s} c_i |(2\Psi_{i\mu}(\delta, u))^\nu - (2\Psi_i(\delta, u))^\nu| \leqslant 2\mu + |c|_1 (2\mu)^\nu \quad \forall u \in U_0$$

или

$$\chi_\delta(u) - 2\mu - |c|_1 (2\mu)^\nu \leqslant \chi_{\delta\mu}(u) \leqslant \chi_\delta(u) + 2\mu + |c|_1 (2\mu)^\nu \quad \forall u \in U_0. \tag{33}$$

Далее, в определении (26) множества $W(\delta, \theta)$ положим $\theta = 2\mu$ и заменим $g^+_{i\delta}(u)$, $\Psi_i(\delta, u)$ функциями $g_{i\delta\mu}(u)$, $\Psi_{i\mu}(\delta, u)$, взятыми из (31). Тогда получим множество

$$W_\mu(\delta, 2\mu) = \{u \in U_0 : g^+_{i\delta\mu}(u) \leqslant \Psi_{i\mu}(\delta, u) + 2\mu, \ i = 1, \ldots, s\}. \tag{34}$$

Покажем, что

$$W(\delta, 0) \subseteq W_\mu(\delta, 2\mu) \subseteq W(\delta, \theta) \quad \forall \theta \geqslant 3\mu. \tag{35}$$

В самом деле, если $u \in W(\delta, 0)$, то $g^+_{i\delta}(u) \leqslant \Psi_i(\delta, u)$ и с учетом (31) имеем

$$g^+_{i\delta\mu}(u) \leqslant g^+_{i\delta}(u) + \mu \leqslant \Psi_i(\delta, u) + \mu \leqslant \Psi_{i\mu}(\delta, u) + 2\mu \quad \forall u \in W(\delta, 0),$$
$$i = 1, \ldots, s.$$

Это означает, что $W(\delta, 0) \subseteq W_\mu(\delta, 2\mu)$. Аналогично, если $u \in W_\mu(\delta, 2\mu)$, то

$$g^+_{i\delta}(u) \leqslant g^+_{i\delta\mu}(u) + \mu \leqslant \Psi_{i\mu}(\delta, u) + 2\mu \leqslant \Psi_i(\delta, u) + 3\mu \leqslant \Psi_i(\delta, u) + \theta,$$
$$i = 1, \ldots, s,$$

так что $W_\mu(\delta, 2\mu) \subseteq W(\delta, \theta)$ $\forall \theta \geqslant 3\mu$. Включения (35) установлены. Из них следует, что

$$\chi_{\delta *}(\theta) = \inf_{u \in W(\delta,\theta)} \chi_\delta(u) \leqslant \inf_{u \in W_\mu(\delta, 2\mu)} \chi_\delta(u) \leqslant \inf_{u \in W(\delta, 0)} \chi_\delta(u) = \chi_{\delta *}(0) \quad \forall \theta \geqslant 3\mu. \tag{36}$$

В неравенствах (33) почленно перейдем к нижней грани по $u \in W_\mu(\delta, 2\mu)$. С учетом (36) получим

$$\chi_{\delta *}(\theta) - 2\mu - |c|_1 (2\mu)^\nu \leqslant \inf_{u \in W_\mu(\delta, 2\mu)} \chi_{\delta\mu}(u) = \chi_{\delta\mu *} \leqslant \chi_{\delta *}(0) + 2\mu + |c|_1 (2\mu)^\nu, \quad \forall \theta \geqslant 3\mu.$$

Отсюда и из (29) следует, что

$$\lim_{(\delta,\theta)\to 0, \, \theta \geqslant 3\mu} \chi_{\delta\mu *} = J_*.$$

Это означает, что при условиях (31), $\theta \geqslant 3\mu > 0$, задача (26)–(28) устойчива по функции, и она может быть использована для вычисления величины $\chi_{\delta\mu *}$, которая согласно теореме 3 при достаточно малых (δ, θ) является приближенным решением задачи (1), (2) первого типа.

Отметим, что для задач линейного программирования в [179] предложен и исследован метод регуляризации, основанный на идее расширения множества и приводящий к устойчивой вспомогательной задаче вида (26)–(28), также являющейся задачей линейного программирования (см. ниже § 7, п. 4). Другой вариант метода регуляризации задач минимизации первого типа, основанный на сочетании идей расширения множества и метода покрытий, можно извлечь из работы [592], в которой рассмотрены более сложные по сравнению с (1), (2) многокритериальные задачи и устойчивые методы их решения.

4. В заключение параграфа сделаем несколько замечаний.

Замечание 1. В теории и методах неустойчивых (некорректных) задач фундаментальную роль играет понятие регуляризирующего оператора (регуляризирующего алгоритма) [63; 695]. Сформулируем это понятие применительно к задаче (1), (2) первого типа.

Определение 1. Оператор R_δ, который каждому набору приближенных данных $(J_\delta(u), \Psi_0(\delta, u), g_{i\delta}(u), \Psi_i(\delta, u), i = 1, \ldots, s)$, удовлетворяющих условиям (10), (11), ставит в соответствие число χ_δ, называется *регуляризирующим оператором задачи* (1), (2) *первого типа*, если $\lim_{\delta\to 0} \chi_\delta = J_*$.

Изложенные выше три метода регуляризации (метод (5), методы (20), (21) и (26)–(28)) определяют оператор, который при выполнении условий теорем 1–3 является регуляризирующим.

Замечание 2. Перечисленные методы регуляризации отличаются друг от друга тем, что для своей реализации требуют различную априорную информацию. В методе (5) мы должны заранее знать, что множество U непусто и известно точно и, кроме того, должны иметь приближения $J_\delta(u)$ для целевой функции, удовлетворяющие условию (4) с известной функцией $\Psi(\delta, u)$. Для реализации метода (20), (21) нужно иметь приближения $J_\delta(u)$, $P_\delta(u)$ и функции погрешности $\Psi_0(\delta, u)$ $\Psi(\delta, u)$ из (10), (18), множество U_0, а также знать информацию о том, что задача (1), (2) имеет сильно согласованную постановку и оценку сверху показателя ν из (23) для правильного выбора параметра p в штрафной функции (12); в методе (26)–(28) нужно еще дополнительно иметь оценки сверху для величин c_i из (23).

Подчеркнем, что все методы решения неустойчивых задач так или иначе предполагают наличие какой-либо априорной информации о рассматриваемой задаче, о ее входных данных и их погрешностях и т. д. Различные аспекты проблемы использования априорной информации в методах решения неустойчивых задач обсуждаются, например, в [184; 450; 509; 557; 693; 695; 697; 757; 817].

Замечание 3. В описанных методах регуляризации мы не касались конкретных методов определения величин $\chi_{\delta *}$ в задачах (5), (21) и (26)–(28). Для вычисления $\chi_{\delta *}$ могут быть использованы любые подходящие методы минимизации (например, методы из гл. 5) в зависимости от свойств функции $\chi_\delta(u)$, структуры множеств U, U_0, $W(\delta, \theta)$. К сожалению, здесь нет общих рецептов, и тонкую работу по согласованию параметров используемого метода минимизации с параметрами изложенных выше методов регуляризации каждый раз приходится проводить отдельно с учетом конкретных особенностей решаемой задачи. Требуют дальнейшего исследования также проблемы, связанные с задачами (1), (2) первого типа, когда входными данными являются не значения функций $J(u)$, $g_i(u)$, а какие-то другие элементы этих задач, например, производные упомянутых функций.

Упражнения

1. К задачам из примеров 1.1–1.9 применить изложенные выше методы регуляризации. Проверить, имеют ли эти задачи сильно согласованную постановку в смысле неравенства (23).

2. Применимы ли изложенные методы регуляризации к задачам из упражнений 1.1–1.4?

§ 3. Стабилизатор. Леммы о регуляризации

1. Задачу

$$J(u) \to \inf, \quad u \in U, \tag{1}$$

$$U = \{u \in U_0\colon g_i(u) \leqslant 0,\ i = 1, \ldots, m;\ g_i(u) = 0,\ i = m+1, \ldots, s\} \tag{2}$$

далее будем рассматривать, как задачу второго типа, считая, что U_0 — заданное множество из некоторого метрического пространства $\mathcal{M}$ с метрикой $\rho(u, v)$, функции $J(u)$, $g_i(u)$, $i = 1, \ldots, s$, определены на U_0 и принимают на этом множестве конечные значения. Напоминаем, что в задачах второго типа наряду с величиной, близкой к $J_* = \inf\limits_{u \in U} J(u)$, ищется точка $u \in U_0$, близкая в метрике $\mathcal{M}$ ко множеству $U_* = \{u \in U\colon J(u) = J_*\}$. Здесь и далее предполагается, что

$$J_* > -\infty, \quad U_* \neq \varnothing. \tag{3}$$

В § 1 было показано, что задача (1), (2) может оказаться неустойчивой по аргументу при возмущениях входных данных $J(u)$, $g_i(u)$, $i = 1, \ldots, s$, и попытка использовать решение (см. множество (1.8)) возмущенной задачи (1.4), (1.5) в качестве приближенного решения задачи (1), (2) второго типа не всегда оправдана. Для гарантированного получения хорошего приближения, как и в случае задачи (1), (2) первого типа, здесь нужно пользоваться методами регуляризации [695]. Для описания этих методов нам понадобится понятие стабилизатора.

Определение 1. Функция $\Omega(u)$, определенная на непустом множестве $U_\Omega \subseteq U_0$, называется *стабилизатором* задачи (1), (2) в метрике $\mathcal{M}$, если:

1) $\Omega(u) \geqslant 0\ \forall u \in U_\Omega$;

2) множество $\Omega_c = \{u \in U_\Omega\colon \Omega(u) \leqslant c\}$ при всех c, для которых $\Omega_c \neq \varnothing$, относительно компактно, т. е. из любой последовательности $\{u_k\} \in \Omega_c$ можно выбрать подпоследовательность, которая в метрике $\mathcal{M}$ сходится к некоторой точке $u \in \mathcal{M}$;

3) множество

$$U_{\Omega *} = U_* \cap U_\Omega \neq \varnothing. \tag{4}$$

Приведем примеры функций, которые могут служить стабилизатором для задачи (1), (2) в некоторых часто встречающихся функциональных пространствах.

Пример 1. Пусть $\mathcal{M} = E^n$ — n-мерное евклидово пространство. Тогда в качестве стабилизатора задачи (1), (2) можно взять функции

$$\Omega(u) = |u - \overline{u}|^l, \quad \Omega(u) = (\langle D(u - \overline{u}), u - \overline{u}\rangle)^l,\ u \in E^n, \tag{5}$$

где $\overline{u}$ — заданная точка из E^n, D — заданная положительно определенная матрица n-го порядка, l — любое положительное число. Относительная компактность множества Ω_c здесь вытекает из классической теоремы Больцано — Вейерштрасса [350; 352]. Требование (4) в задаче (1), (2) при $\mathcal{M} = E^n$ равносильно условию $U_* \neq \varnothing$, так как функции (5) определены на E^n и поэтому можем считать, что $U_\Omega = U_0$, $U = U \cap U_\Omega$, $U_{\Omega *} = U_* \cap U_\Omega = U_*$.

Пример 2. Пусть $\mathcal{M} = C_r[a, b]$ — банахово пространство r-мерных непрерывных вектор-функций $u = u(t) = (u^1(t), \ldots, u^r(t))$, $a \leqslant t \leqslant b$, с нормой $\|u\|_{C_r} = \max\limits_{a \leqslant t \leqslant b} |u(t)|$. Рассмотрим функцию

$$\Omega(u) = \|u\|^2_{H_r^1} = \int_a^b (|u(t)|^2 + |\dot{u}(t)|^2)\,dt, \tag{6}$$

определенную на множестве $H_r^1[a, b]$ (обозначения см в § 8.1). В силу известных теорем вложения (см., например, [492; 557; 648]) $H_r^1[a, b] \subset C_r[a, b]$. Убедимся, что множество $\Omega_c = \{u \in H_r^1[a, b]\colon \Omega(u) \leqslant c\}$ относительно компактно в метрике $C_r[a, b]$. В самом деле, из неравенства $\Omega(u) \leqslant c$ следует существование хотя бы одной точки $t_1 \in [a, b]$ такой, что $|u(t_1)| \leqslant (c(b-a)^{-1})^{1/2}$. Тогда $|u(t)| = \left|\int\limits_{t_1}^{t} \dot{u}(\tau)\,d\tau + u(t_1)\right| \leqslant (c(b-a)^{-1})^{1/2} + (b-a)^{1/2}\left(\int\limits_a^b |\dot{u}(\tau)|^2\,d\tau\right)^{1/2} \leqslant (c(b-a)^{-1})^{1/2} + (c(b-a))^{1/2} = \text{const}\ \ \forall u \in \Omega_c$. Далее, имеем $|u(t) - u(\tau)| = \left|\int\limits_\tau^t \dot{u}(s)\,ds\right| \leqslant |t-\tau|^{1/2}\left(\int\limits_a^b |\dot{u}(s)|^2\,ds\right)^{1/2} \leqslant (c|t-\tau|)^{1/2}$ $\forall u \in \Omega_c$. Таким образом, множество функций Ω_c равномерно ограничено и равностепенно непрерывно на отрезке $[a, b]$. Из теоремы [393, стр. 110] тогда следует, что из любой последовательности $\{u_k(t)\} \in \Omega_c$ можно выбрать подпоследовательность, сходящуюся к некоторой непрерывной функции $u(t)$ равномерно на отрезке $[a, b]$. Относительная компактность множества Ω_c в метрике $C_r[a, b]$ установлена. Для задачи (1), (2) со множеством $U_0 \subseteq C_r[a, b]$ требование (4) означает, что эта задача имеет хотя бы одно решение $u_* = u_*(t) \in H_r^1[a, b]$. При выполнении этого условия функция (6) с областью определения $U_\Omega = U_0 \cap H_r^1[a, b]$ является стабилизатором задачи (1), (2). Стабилизатором может служить также функция $\Omega(u) = (\|u\|_{H_r^1})^l$ при любом $l > 0$.

Покажем, что функция

$$\Omega(u) = \left(\max_{a \leqslant t \leqslant b} |u(t)| + \sup_{\substack{t, \tau \in [a, b] \\ t \neq \tau}} \frac{|u(t) - u(\tau)|}{|t - \tau|^\gamma}\right)^l, \quad 0 < \gamma \leqslant 1, \quad l > 0 \tag{7}$$

также может быть использована в качестве стабилизатора в задаче (1), (2) при $U_0 \subseteq C_r[a, b]$, если эта задача имеет хотя бы одно решение $u_* = u_*(t)$, удовлетворяющее условию Гельдера: $|u_*(t)-u_*(\tau)| \leqslant L|t-\tau|^\gamma$, $t, \tau \in [a, b]$. В самом деле, если $\Omega(u) \leqslant c$, то $\max\limits_{a \leqslant t \leqslant b} |u(t)| \leqslant c_1 = c^{1/l}$, $|u(t)-u(\tau)| \leqslant c_1|t-\tau|^\gamma$, $t, \tau \in [a, b]$ $\forall u \in \Omega_c$. Это значит, что множество функций Ω_c равномерно ограничено и равностепенно непрерывно на отрезке $[a, b]$. В силу теоремы Арцела множество Ω_c относительно компактно в метрике $C_r[a, b]$.

Функции (6), (7) могут служить стабилизатором и в задаче (1), (2) при $\mathcal{M} = L_p^r[a, b]$, $1 \leqslant p \leqslant \infty$, так как из равномерной сходимости последовательности непрерывных функций следует ее сходимость в норме $L_p[a, b]$.

Пример 3. Пусть $\mathcal{M} = C_r^m[a, b]$ — банахово пространство m раз непрерывно дифференцируемых r-мерных вектор-функций с нормой $\|u\|_{C_r^m} = \sum\limits_{i=0}^{m} \max\limits_{a \leqslant t \leqslant b} \left|\frac{d^i u(t)}{dt^i}\right|$. Если в задаче (1), (2) множество $U_0 \subseteq C_r^m[a, b]$, и эта задача имеет хотя бы одно решение $u(t) \in H_r^{m+1}[a, b]$ (см. обозначение в § 8.1), то стабилизатором в ней можно взять функцию

$$\Omega(u) = \|u\|^2_{H_r^{m+1}} = \int\limits_a^b \left(|u(t)|^2 + \sum_{i=1}^{m+1} \left|\frac{d^i u(t)}{dt^i}\right|^2\right) dt.$$

Если же такая задача (1), (2) имеет решение $u(t) \in C_r^m[a, b]$, m-я производная которого удовлетворяет условию Гельдера $\left|\frac{d^m u(t)}{dt^m} - \frac{d^m u(\tau)}{dt^m}\right| \leqslant L|t-\tau|^\gamma$, $\forall t, \tau \in [a, b]$, $0 < \gamma \leqslant 1$, то стабилизатором может служить функция

$$\Omega(u) = \|u\|_{C_r^m} + \sup_{\substack{t, \tau \in [a, b] \\ t \neq \tau}} \left|\frac{d^m u(t)}{dt^m} - \frac{d^m u(\tau)}{dt^m}\right| |t-\tau|^{-\gamma}.$$

Эти утверждения доказываются с помощью рассуждений, аналогичных приведенным в примере 2.

Пример 4. Пусть в задаче (1), (2) $U_0 \subseteq \mathcal{M} = L_1^r[a, b]$, и эта задача имеет хотя бы одно решение $u_* = u_*(t) \in V^r[a, b]$. Здесь под $V^r[a, b]$ понимается банахово пространство вектор-функций $u(t) = (u^1(t), \ldots, u^r(t))$ с ограниченным изменением, в котором норма равна $\|u\|_{V^r} = |u(a)| + V_a^b(u)$, $V_a^b(u)$ — полное изменение функции $u = u(t)$ на отрезке $[a, b]$ ([393], см. также § 6.4). В такой задаче стабилизатором можно взять функцию $\Omega(u) = \|u\|_{V^r}$. Покажем, что множество $\Omega_c = \{u \in V^r[a, b]: \Omega(u) \leqslant c\}$ относительно компактно в метрике $L_1^r[a, b]$. В самом деле, из неравенства $\Omega(u) \leqslant c$ следует, что $|u(t)| \leqslant |u(a)| + |u(t)-u(a)| \leqslant |u(a)| + V_a^b(u) \leqslant c$ $\forall t \in [a, b]$ $\forall u \in \Omega_c$. Далее, считая для определенности $u(t) \equiv u(b)$ $\forall t > b$, будем иметь

$$\int\limits_a^b |u(t+\tau)-u(t)|dt \leqslant \int\limits_a^b V_t^{t+\tau}(u)dt = \int\limits_a^b (V_a^{t+\tau}(u) - V_a^t(u))dt =$$
$$= \int\limits_b^{b+\tau} V_a^t(u)dt - \int\limits_a^{a+\tau} V_a^t(u)dt \leqslant 2c\tau \quad \forall \tau > 0, \quad \forall u \in \Omega_c.$$

Следовательно, множество функций Ω_c равномерно ограничено и равностепенно непрерывно в норме $L_1^r[a, b]$. Это значит (см. теорему 8.2.2), что множество Ω_c относительно компактно в метрике $L_1^r[a, b]$.

При исследовании задач (1), (2), когда $U_0 \subseteq \mathcal{M} = B$ — банахово пространство, ниже нам понадобится понятие слабого стабилизатора.

Определение 2. Функция $\Omega(u)$, определенная на непустом множестве $U_\Omega \subseteq U_0 \subseteq B$, называется *слабым стабилизатором* задачи (1), (2), если:

1) $\Omega(u) \geqslant 0$ $\forall u \in U_\Omega$;

2) множество $\Omega_c = \{u \in U_\Omega: \Omega(u) \leqslant c\}$ при всех c, для которых $\Omega_c \neq \varnothing$, относительно слабо компактно, т. е. из любой последовательности $\{u_k\} \in \Omega_c$ можно выбрать подпоследовательность, которая слабо в B сходится к некоторой точке $u \in B$;

3) множество $U_{\Omega *} = U_* \cap U_\Omega$ непусто.

Пример 5. Пусть $\mathcal{M} = B$ — рефлексивное банахово пространство. Тогда в качестве слабого стабилизатора задачи (1), (2) возьмем функцию

$$\Omega(u) = \|u - \overline{u}\|_B^\gamma, \quad u \in B; \quad \gamma > 0,$$

где $\overline{u}$ — заданный элемент из B. Относительно слабая компактность множества Ω_c следует из теоремы 8.2.3; требование $U_{\Omega *} \neq \varnothing$ равносильно условию $U_* \neq \varnothing$, так как здесь можем считать, что $U_\Omega = U_0$, и, следовательно, $U_{\Omega *} = U_*$.

Если U_0 — выпуклое множество из рефлексивного банахова пространства, то в качестве слабого стабилизатора можно взять равномерно выпуклую функцию $\Omega(u)$ на U_0. Тогда множество $\Omega_c = \{u \in U_0: \Omega(u) \leqslant c\}$ ограничено (теорема 4.7.1) и относительно слабо компактно (теорема 8.2.3).

Примеры 1–5 показывают, что при построении стабилизаторов для задачи (1), (2) весьма полезны знания условий относительной компактности [относительно слабой компактности] множеств в конкретных метрических [банаховых] пространствах, теорем вложения одного функционального пространства в другое [95; 492; 535; 557; 648; 649].

2. В трех основных методах регуляризации для решения задачи (1), (2) второго типа (методы стабилизации, невязки, квазирешений), которые будут изложены ниже, строится однопараметрическое семейство точек $\{u_\delta, \delta > 0\} \in U_\Omega$, удовлетворяющее условиям

$$J(u_\delta) \leqslant J_* + \beta(\delta), \quad J_* = \inf_U J(u); \tag{8}$$

$$g_i(u_\delta) \leqslant \rho(\delta), \ i = 1, \ldots, m, \ |g_i(u_\delta)| \leqslant \rho(\delta), \ i = m+1, \ldots, s; \tag{9}$$

$$\Omega(u_\delta) \leqslant \Omega_* + \gamma(\delta), \quad \Omega_* = \inf_{U_{0*}} \Omega(u), \tag{10}$$

где $\beta(\delta), \gamma(\delta), \rho(\delta)$ — некоторые функции параметра $\delta > 0$. При исследовании сходимости упомянутых методов регуляризации важную роль играют следующие две леммы, которые будем называть *леммами о регуляризации*.

Лемма 1. *Пусть*

1) *множество U_0 замкнуто в метрике $\mathcal{M}$, функции $J(u)$, $g_i(u)$, $i = 1, \ldots, m$; $|g_i(u)|$, $i = m+1, \ldots, s$, полунепрерывны снизу на U_0, множество* (2) *непусто и выполнены условия* (3);

2) *$\Omega(u)$ — стабилизатор задачи* (1), (2) *в метрике $\mathcal{M}$;*

3) *точки $u_\delta \in U_\Omega$, $\delta > 0$ таковы, что выполнены условия* (8)–(10), *где*

$$\lim_{\delta \to 0} \beta(\delta) = 0, \quad \lim_{\delta \to 0} \rho(\delta) = 0, \quad \sup_{\delta > 0} |\gamma(\delta)| < \infty. \tag{11}$$

Тогда

$$\lim_{\delta \to 0} J(u_\delta) = J_*, \quad \overline{\lim_{\delta \to 0}} g_i(u_\delta) \leqslant 0, \quad i = 1, \ldots, m;$$
$$\lim_{\delta \to 0} g_i(u_\delta) = 0, \quad i = m+1, \ldots, s, \quad \lim_{\delta \to 0} \rho(u_\delta, U_*) = 0. \tag{12}$$

Если наряду с условиями 1)–3) *выполнено еще условие*

4) $U_\Omega = U_0$, *функция* $\Omega(u)$ *полунепрерывна снизу на* U_0, $\lim\limits_{\delta\to 0}\gamma(\delta)=0$, *то наряду с* (12)

$$\lim_{\delta\to 0}\Omega(u_\delta)=\Omega_*,\quad \lim_{\delta\to 0}\rho(u_\delta, U_{**})=0, \tag{13}$$

где $U_{**}=\{u\in U_*\colon \Omega(u)=\Omega_*\}$.

Доказательство. Из (10), (11) следует, что $\Omega(u_\delta)\leqslant\Omega_*+\sup\limits_{\delta>0}|\gamma(\delta)|=c<\infty$, т. е. $u_\delta\in\Omega_c=\{u\in U_\Omega\colon \Omega(u)\leqslant c\}$. Поскольку множество Ω_c относительно компактно, то из любой последовательности $\{u_k=u_{\delta_k}\}$, где $\{\delta_k\}\to 0$, можно выбрать подпоследовательность $\{u_{k_r}\}$, сходящуюся к некоторой точке $v_*\in\mathcal{M}$. Из замкнутости U_0 и из $\{u_k\}\subset U_\Omega\subseteq U_0$ следует, что $v_*\in U_0$. Далее, пользуясь полунепрерывностью снизу функций $g_i(u)$, $i=1,\ldots,m$, $|g_i(u)|$, $i=m+1,\ldots,s$, и включением $v_*\in U_0$, из (9), (11) имеем

$$\begin{gathered} g_i(v_*)\leqslant\varliminf_{r\to\infty} g_i(u_{k_r})\leqslant\varlimsup_{r\to\infty} g_i(u_{k_r})\leqslant 0,\ i=1,\ldots,m,\\ 0\leqslant|g_i(v_*)|\leqslant\varliminf_{r\to\infty}|g_i(u_{k_r})|\leqslant\varlimsup_{r\to\infty}|g_i(u_{k_r})|=0,\ i=m+1,\ldots,s. \end{gathered} \tag{14}$$

Это означает, что $v_*\in U$. Из полунепрерывности снизу функции $J(u)$, условий (8), (11) вытекает $J_*\leqslant J(v_*)\leqslant\varliminf\limits_{r\to\infty}J(u_{k_r})\leqslant\varlimsup\limits_{r\to\infty}J(u_{k_r})\leqslant J_*$, т. е. $\lim\limits_{r\to\infty}J(u_{k_r})=J(v_*)=J_*$. Это означает, что $v_*\in U_*$. Тогда $0\leqslant\rho(u_{k_r}, U_*)\leqslant\rho(u_{k_r}, v_*)$, $r=1,2,\ldots$, поэтому $\lim\limits_{r\to\infty}\rho(u_{k_r}, U_*)=0=\rho(v_*, U_*)$. Тем самым показано, что для любой точки v_*, являющейся пределом какой-либо подпоследовательности $\{u_{k_r}\}$ произвольной последовательности $\{u_k=u_{\delta_k}\}$, где $\{\delta_k\}\to 0$, справедливы равенства $\lim\limits_{r\to\infty}J(u_{k_r})=J(v_*)=J_*$, $\lim\limits_{r\to\infty}\rho(u_{k_r}, U_*)=0=\rho(v_*, U_*)$. Отсюда следует, что каждое из числовых множеств $\{J(u_\delta),\ \delta>0\}$, $\{\rho(u_\delta, U_*),\ \delta>0\}$ имеет единственную предельную точку, равную J_* и 0 соответственно, что равносильно утверждению (12).

Пусть теперь выполнены все условия 1)–4) леммы. Снова возьмем произвольную последовательность $\{u_k=u_{\delta_k}\}$, где $\{\delta_k\}\to 0$, и произвольную точку v_*, являющуюся пределом какой-либо подпоследовательности $\{u_{k_r}\}$ этой последовательности. По доказанному $v_*\in U_*$. Отсюда с учетом равенства $U_{\Omega*}=U_*$, полунепрерывности снизу стабилизатора $\Omega(u)$ на U_0, $\lim\limits_{\delta\to 0}\gamma(\delta)=0$ и условия (10) имеем

$$\Omega_*=\inf_{U_*}\Omega(u)\leqslant\Omega(v_*)\leqslant\varliminf_{r\to\infty}\Omega(u_{k_r})\leqslant\varlimsup_{r\to\infty}\Omega(u_{k_r})\leqslant\Omega_*, \tag{15}$$

т. е. $\lim\limits_{r\to\infty}\Omega(u_{k_r})=\Omega(v_*)=\Omega_*$, $v_*\in U_{**}$. Тогда $0\leqslant\rho(u_{k_r}, U_{**})\leqslant\rho(u_{k_r}, v_*)$, $r=1,2,\ldots$, и, следовательно, $\lim\limits_{r\to\infty}\rho(u_{k_r}, U_{**})=0$. Это означает, что каждое из числовых множеств $\{\Omega(u_\delta),\ \delta>0\}$, $\{\rho(u_\delta, U_{**}),\ \delta>0\}$ имеет единственную предельную точку, равную Ω_* и 0 соответственно, что равносильно утверждению (13). Лемма 1 доказана. □

Заметим, что примерами стабилизатора $\Omega(u)$, удовлетворяющего всем условиям леммы 1, являются функции (5) на множестве $U_\Omega=U_0\subseteq\mathcal{M}=E^n$.

3. В выпуклых задачах (1), (2), когда $\mathcal{M}=B$ — банахово пространство, сходимость ко множеству U_* в метрике B семейства $\{u_\delta,\ \delta>0\}$, удовлетворяющего условиям (8)–(10), иногда удается получить, когда функция $\Omega(u)$

является слабым стабилизатором. Такая возможность обсуждается в следующей лемме.

Лемма 2. *Пусть*

1) B *— рефлексивное банахово пространство,* U_0 *— выпуклое замкнутое множество из* B, *функции* $J(u)$, $g_i(u)$, $i=1,\ldots,m$, $|g_i(u)|$, $i=m+1,\ldots,s$, *выпуклы и полунепрерывны снизу на* U_0, *множество* (2) *непусто и выполнены условия* (3);

2) *функция* $\Omega(u)$ *определена, полунепрерывна снизу, строго равномерно выпукла на* U_0;

3) *точки* $u_\delta\in U_0$ *удовлетворяют условиям* (8)–(10), *где*

$$\lim_{\delta\to 0}\beta(\delta)=0,\quad \lim_{\delta\to 0}\rho(\delta)=0,\quad \lim_{\delta\to 0}\gamma(\delta)=0.$$

Тогда

$$\begin{gathered} \lim_{\delta\to 0}J(u_\delta)=J(v_*)=J_*,\quad \varlimsup_{\delta\to 0}g_i(u_\delta)\leqslant 0,\quad i=1,\ldots,m;\\ \lim_{\delta\to 0}g_i(u_\delta)=0,\quad i=m+1,\ldots,s,\\ \lim_{\delta\to 0}\Omega(u_\delta)=\Omega(v_*)=\Omega_*,\quad \lim_{\delta\to 0}\|u_\delta-v_*\|=0, \end{gathered} \tag{16}$$

где v_* *— точка минимума функции* $\Omega(u)$ *на множестве* U_*.

Доказательство. Из (10) следует, что $u_\delta\in\Omega_c=\{u\in U_0\colon \Omega(u)\leqslant c\}$ $\forall\delta>0$, $c=\Omega_*+\sup\limits_{\delta>0}|\gamma(\delta)|<\infty$. Это множество ограничено в норме B, что доказывается также, как в теореме 4.7.1. Отсюда и из теоремы 8.2.3 вытекает, что Ω_c относительно слабо компактно в B. Поэтому из любой последовательности $\{u_k=u_{\delta_k}\}$, где $\{\delta_k\}\to 0$, можно выбрать подпоследовательность $\{u_{k_r}\}$, слабо сходящуюся к некоторой точке $v_*\in B$. Так как множество U_0 выпукло, замкнуто, $\{u_{k_r}\}\in U_0$, $r=1,2,\ldots$, то $v_*\in U_0$ (теорема 8.2.5). Функции $J(u)$, $g_i(u)$, $i=1,\ldots,m$; $|g_i(u)|$, $i=m+1,\ldots,s$, $\Omega(u)$ выпуклы и полунепрерывны снизу на U_0, следовательно, они слабо полунепрерывны снизу (теорема 8.2.7). Отсюда и из (9), (11) следуют неравенства (14), т. е. $v_*\in U$. Далее, как и в лемме 1 доказываем, что $\lim\limits_{r\to\infty}J(u_{k_r})=J(v_*)=J_*$, так что $v_*\in U_*$. Тогда справедливы неравенства (15), откуда получаем, что $\lim\limits_{r\to\infty}\Omega(u_{k_r})=\Omega(v_*)=\Omega_*$, $v_*\in U_{**}=\{u\in U_*\colon \Omega(u)=\Omega_*\}$, т. е. v_* — точка минимума строго равномерной выпуклой функции на выпуклом множестве U_*. Следовательно, $U_{**}=\{v_*\}$ (теорема 4.2.1). Кроме этого из теоремы 4.7.1 имеем: $\delta(\|u_{k_r}-v_*\|)\leqslant\Omega(u_{k_r})-\Omega(v_*)$, $r=1,2,\ldots$ Поэтому $\lim\limits_{r\to\infty}\delta(\|u_{k_r}-v_*\|)=0$. По свойству модуля строго равномерно выпуклой функции это возможно только при $\lim\limits_{r\to\infty}\|u_{k_r}-v_*\|=0$. Из проведенных рассуждений следует, что каждое из числовых множеств $\{J(u_\delta),\ \delta>0\}$, $\{\Omega(u_\delta),\ \delta>0\}$, $\{\|u_\delta-v_*\|,\ \delta>0\}$ при $\delta\to 0$ имеет единственную предельную точку, равную J_*, Ω_*, 0 соответственно, что доказывает равенства (16). □

Примеры строго равномерно выпуклых функций были приведены в § 8.2. Если $B=H$ — гильбертово пространство, то в лемме 2 можно взять сильно выпуклую функцию $\Omega(u)$, например, $\Omega(u)=\|u\|_H^2$.

4. При исследовании методов регуляризации нам еще понадобятся некоторые результаты из § 5.15 о штрафных функциях, которые мы здесь переформулируем в удобной для дальнейших ссылок форме.

Лемма 3. *Пусть задача* (1), (2) *имеет сильно согласованную постановку* (*определение* 5.15.3), *т. е.*

$$-\infty < J_* \leqslant J(u) + \sum_{i=1}^{s} c_i (g_i^+(u))^\nu, \quad u \in U_0 \tag{17}$$

при некоторых $c_i \geqslant 0$, $i = 1, \ldots, s$, $\nu > 0$; *точки* $u_\delta \in U_0$, $\delta > 0$, *удовлетворяют неравенству*

$$J(u_\delta) + A(\delta) P(u_\delta) \leqslant J_* + \beta(\delta), \tag{18}$$

где $P(u) = \sum_{i=1}^{s} (g_i^+(u))^p$, $u \in U_0$, — *штрафная функция множества* (2), $p \geqslant \nu$,

$$\beta(\delta) \geqslant 0, \quad A(\delta) > 0, \quad \lim_{\delta \to 0} \beta(\delta) = 0, \quad \lim_{\delta \to 0} A(\delta) = +\infty.$$

Тогда

$$0 \leqslant g_i^+(u_\delta) \leqslant (P(u_\delta))^{1/p} \leqslant \rho(\delta), \quad i = 1, \ldots, s, \quad \delta > 0, \tag{19}$$

$$-|c|(\rho(\delta))^\nu \leqslant J(u_\delta) - J_* \leqslant \beta(\delta), \quad \delta > 0, \tag{20}$$

где

$$\rho(\delta) = \begin{cases} \left(\dfrac{\beta(\delta)}{A(\delta) - |c|}\right)^{1/\nu}, \ A(\delta) > |c| = \max |c_i| \ \text{ при } \ p = \nu; \\ \left[\left(\dfrac{|c|}{A(\delta)}\right)^{\frac{p}{p-\nu}} + \dfrac{p}{p-\nu} \dfrac{\beta(\delta)}{A(\delta)}\right]^{1/p}, \ |c| = \left(\sum\limits_{i=1}^{s} |c_i|^{\frac{p}{p-\nu}}\right)^{\frac{p-\nu}{p}} \ \text{ при } \ p > \nu. \end{cases} \tag{21}$$

Доказательство проводится также, как в теореме 5.15.5 при $p > \nu$ и теореме 5.15.6 при $p = \nu$ с заменой в них величины A_k на $A(\delta)$, ε_k — на $\beta(\delta)$, u_k — на u_δ. Поэтому мы здесь ограничимся приведением лишь краткой схемы доказательства.

Из (17), (18) имеем $J(u_\delta) + A(\delta) P(u_\delta) \leqslant J(u_\delta) + \sum_{i=1}^{s} c_i (g_i^+(u_\delta))^\nu + \beta(\delta)$ или

$$0 \leqslant A(\delta) P(u_\delta) \leqslant \sum_{i=1}^{s} c_i (g_i^+(u_\delta))^\nu + \beta(\delta) \quad \forall \delta > 0. \tag{22}$$

Пусть $p > \nu$. Тогда с помощью неравенства Гельдера получаем

$$0 \leqslant \sum_{i=1}^{s} c_i (g_i^+(u_\delta))^\nu \leqslant |c| (P(u_\delta))^{\nu/p} \quad \forall \delta > 0. \tag{23}$$

Из (22), (23) следует, что $0 \leqslant A(\delta) P(u_\delta) \leqslant |c| (P(u_\delta))^{\nu/p} + \beta(\delta) \leqslant \frac{|c|}{(A(\delta))^{\nu/p}} (A(\delta)(P(u_\delta))^{\nu/p} + \beta(\delta)$. Отсюда и из леммы 2.6.11 при $z = (A(\delta) P(u_\delta))^{\nu/p}$ получаем

$$(A(\delta) P(u_\delta)) \leqslant \left(|c| A(\delta)^{-\nu/p}\right)^{\frac{p}{p-\nu}} + \frac{p}{p-\nu} \beta(\delta), \quad \forall \delta > 0.$$

После простых преобразований полученного неравенства, учитывая, что $g_i^+(u_\delta) \leqslant (P(u_\delta))^{1/p}$ приходим к оценке (19) при $p > \nu$.

При $p = \nu$ из (22) сразу имеем $A(\delta) P(u_\delta) \leqslant |c| P(u_\delta) + \beta(\delta)$. Отсюда при $A(\delta) > |c|$ следует $P(u_\delta) \leqslant \frac{\beta(\delta)}{A(\delta) - |c|}$, что равносильно оценке (19) при $p = \nu$. Наконец, из (17), (18), (23) вытекают неравенства $-|c|(P(u_\delta))^{\nu/p} \leqslant -\sum_{i=1}^{s} c_i (g_i^+(u))^\nu \leqslant J(u_\delta) - J_* \leqslant \beta(\delta)$. Отсюда и из (19) получаем оценку (20). □

Упражнения

1. Можно ли функцию $\Omega(u) = \int_0^1 (u^2(t) + \dot{u}^2(t)) dt$ взять в качестве стабилизатора в метрике $C[0, 1]$ в задаче минимизации функции $J(u) = \int_0^1 u^2(t) dt$ на множестве $U = L_2[0, 1]$?

2. Задача: $J(u) = \max\limits_{0 \leqslant t \leqslant 1} |u(t)| + \max\limits_{0 \leqslant t \leqslant 1} |\dot{u}(t)| \to \inf$, $u \in U = C^2[0, 1]$. При каких m, p функция $\Omega(u) = \int_0^1 \sum_{i=0}^{m} \left| \frac{d^i u(t)}{dt^i} \right|^p dt$ может служить стабилизатором этой задачи в метрике $C^2[0, 1]$?

3. Задача: $J(u) = \int_0^1 |u(t)| dt \to \inf$, $u \in U = L_1[0, 1]$. Какие из функций $\Omega(u) = \int_0^1 |u(t)|^p dt$, $\Omega(u) = \max\limits_{0 \leqslant t \leqslant 1} |u(t)|$, $\Omega(u) = \int_0^1 (|u(t)| + |\dot{u}(t)|) dt$ является стабилизатором этой задачи в метрике $L_1[0, 1]$?

4. Доказать, что если задача (1), (2) при $\mathcal{M} = C_r[a, b]$ имеет хотя бы одно решение, принадлежащее пространству $H_2^1[a, b]$, то функция $\Omega(u) = \int_a^b [k(t)(\dot{u}(t))^2 + q(t) u(t)] dt$, где функции $k(t), q(t)$ непрерывны и положительны на $[a, b]$, является стабилизатором в метрике $C_r[a, b]$.

5. Для задачи минимизации функций на выпуклом замкнутом множестве из $L_p[a, b]$, $1 < p < \infty$, привести примеры слабых стабилизаторов.

6. Доказать, что в задаче минимизации функций на множестве U из $C^m(G)$, $G \in E^n$, при $n < 2m$ в качестве стабилизатора в метрике $C^m(G)$ можно взять $\Omega(u) = \|u\|_{H^{m+1}}$, если $U_* \cap H^{m+1}(G) \neq \varnothing$. Указание: воспользоваться теоремой вложения пространства $H^{m+1}(G)$ в $C^m(G)$ при $n < 2m$ [492; 648; 649] (обозначения см. в § 8.1).

7. Доказать, что если $\Omega_1(u)$, $\Omega_2(u)$ — стабилизаторы [слабые стабилизаторы] с одной и той же областью определения U_Ω, то $\Omega(u) = \alpha_1 \Omega_1(u) + \alpha_2 \Omega_2(u)$, $\alpha_1 \geqslant 0$, $\alpha_2 \geqslant 0$, $\alpha_1 + \alpha_2 > 0$, также являются стабилизатором [слабым стабилизатором].

8. Доказать, что в равенствах (12), (13), (16) предел равномерен относительно выбора точек u_δ из множества $Z(\delta) = \{u \in U_\Omega: J(u_\delta) \leqslant J_* + \beta(\delta), g_i^+(u_\delta) \leqslant \rho(\delta), i = 1, \ldots, s, \Omega(u_\delta) \leqslant \Omega_* + \gamma(\delta)\}$.

§ 4. Метод стабилизации

1. Изложение методов регуляризации для решения неустойчивых задач второго типа начнем с *метода стабилизации*, разработанного А. Н. Тихоновым [695]. Этот метод в литературе часто называют *методом стабилизирующих функционалов*. Опишем этот метод применительно к следующей задаче второго типа:

$$J(u) \to \inf, \quad u \in U, \tag{1}$$

$$U = \{u \in U_0, \ g_i(u) \leqslant 0, \ i = 1, \ldots, m; \ g_i(u) = 0, \ i = m + 1, \ldots, s\}, \tag{2}$$

где U_0 — заданное множество из некоторого метрического пространства $\mathcal{M}$, функции $J(u)$, $g_i(u)$, $i = 1, \ldots, s$, определены на U_0 и принимают на нем конечное значение. Будем предполагать, что

$$J_* = \inf_U J(u) > -\infty, \quad U_* = \{u \in U: J(u) = J_*\} \neq \varnothing. \tag{3}$$

Для учета ограничений типа равенств и неравенств, задающих множество (2), воспользуемся штрафной функцией

$$P(u) = \sum_{i=1}^{s} (g_i^+(u))^p, \quad u \in U_0, \quad p > 0, \tag{4}$$

где $g_i^+ = \max\{g_i(u); 0\}$ при $i = 1, \ldots, m$, $g_i^+ = |g_i|$ при $i = m+1, \ldots, s$. Пусть $\Omega(u)$ — какой-либо стабилизатор задачи (1), (2) с областью определения $U_\Omega \subseteq U_0$. Предполагается, что множество U_0 известно точно, а вместо функций $J(u)$, $P(u)$, которые наряду с $J(u)$, $g_i(u)$, $i = 1, \ldots, s$, также будем называть входными данными задачи (1), (2), известны их приближения $J_\delta(u)$, $P_\delta(u)$, причем погрешности согласованы со стабилизатором в следующем смысле:

$$|J_\delta(u) - J(u)| \leqslant \delta(1+\Omega(u)), \quad |P_\delta(u) - P(u)| \leqslant \delta(1+\Omega(u)), \quad u \in U_\Omega, \tag{5}$$

$\delta > 0$ — скалярный параметр погрешности. Заметим, что возможный способ перехода от приближений $g_{i\delta}(u)$, $i = 1, \ldots, s$, к $P_\delta(u)$ обсуждался в § 2 (см. формулы (2.10)–(2.19)).

Составим функцию

$$t_\delta(u) = J_\delta(u) + A(\delta)P_\delta(u) + \alpha(\delta)\Omega(u), \quad u \in U_\Omega, \tag{6}$$

называемую *функцией Тихонова*. Здесь $J_\delta(u)$, $P_\delta(u)$ — какие-либо конкретные реализации приближений входных данных из (5), $A(\delta) > 0$ — штрафной коэффициент, $\alpha(\delta) > 0$ — *параметр регуляризации*.

Рассмотрим задачу первого типа

$$t_\delta(u) \to \inf, \quad u \in U_\Omega. \tag{7}$$

Для решения задачи (7) могут быть использованы методы, описанные в § 2. Дальнейшее изложение не зависит от метода решения этой задачи. Мы будем предполагать, что нам известна какая-либо точка u_δ, удовлетворяющая условиям

$$u_\delta \in U_\Omega, \quad t_\delta(u_\delta) \leqslant t_{\delta *} + \varepsilon(\delta), \tag{8}$$

где подразумевается, что $t_{\delta *} = \inf\limits_{u \in U_\Omega} t_\delta(u) > -\infty$, $\varepsilon(\delta) > 0$. На формальном уровне метод стабилизации описан. Осталось обсудить условия сходимости метода (8) при $\delta \to 0$, указать условия согласования параметров $\alpha(\delta)$, $A(\delta)$, $\varepsilon(\delta)$ этого метода.

Сначала приведем достаточные условия, гарантирующие неравенство $t_{\delta *} > -\infty$, $\forall \delta > 0$. Как и в § 2, ограничимся рассмотрением класса задач (1), (2), имеющих сильно согласованную постановку (определение 5.15.3), когда

$$-\infty < J_* \leqslant J(u) + \sum_{i=1}^{s} c_i (g_i^+(u))^\nu, \quad u \in U_0, \tag{9}$$

при некоторых $c_1 \geqslant 0, \ldots, c_s \geqslant 0$, $\nu > 0$. Тогда справедливо неравенство (2.24):

$$-\infty < J_* \leqslant J(u) + A(\delta)P(u) + BA^{-\frac{\nu}{p-\nu}} \quad \forall u \in U_0, \quad \delta > 0, \quad p \geqslant \nu, \tag{10}$$

где при $p = \nu$ по определению полагается $A^{-\frac{\nu}{p-\nu}} = 0$, $B = 0$ и при этом считается, что $A(\delta) \geqslant |c| = \max\limits_{1 \leqslant i \leqslant s} c_i$, а при $p > \nu$ здесь $B = (p-\nu)\nu^{\frac{\nu}{p-\nu}} p^{-\frac{p}{p-\nu}} |c|^{\frac{p}{p-\nu}}$, $|c| = \left(\sum\limits_{i=1}^{s} c_i^{\frac{p}{p-\nu}}\right)^{\frac{p-\nu}{p}}$. Предположим, что параметры $\alpha(\delta)$, $A(\delta)$ метода (8) таковы, что

$$\delta + \delta A(\delta) \leqslant \alpha(\delta) \quad \forall \delta > 0. \tag{11}$$

Тогда из (5), (6), (10), (11) имеем

$$t_\delta(u) \geqslant J(u) - \delta(1+\Omega(u)) + A(\delta)P(u) - A(\delta)\delta(1+\Omega(u)) + \alpha(\delta)\Omega(u) \geqslant J_* - BA^{-\frac{\nu}{p-\nu}} + (\alpha(\delta) - \delta - \delta A(\delta))\Omega(u) - (\delta + \delta A(\delta)) \geqslant J_* - BA^{-\frac{\nu}{p-\nu}} - (\delta + \delta A(\delta)) \quad \forall u \in U_\Omega, \quad \delta \geqslant 0.$$

Это означает, что $t_{\delta *} > -\infty$ $\forall \delta > 0$. Тогда точка u_δ из (8) существует при любом выборе $\varepsilon(\delta) > 0$ по определению нижней грани.

2. Покажем, что при соответствующем согласовании параметров $\alpha(\delta)$, $A(\delta)$, $\varepsilon(\delta)$ метода (8) и некоторых дополнительных требованиях к задаче (1), (2) точка u_δ из (8) при малых $\delta > 0$ будет близка в метрике $\mathcal{M}$ ко множеству U_*, а величина $J_\delta(u_\delta)$ близка к J_*, т. е. пара $(J_\delta(u_\delta), u_\delta)$ является приближенным решением задачи (1), (2) второго типа. А именно, справедлива

Теорема 1. *Пусть*

1) *множество U_0 замкнуто в метрике $\mathcal{M}$, функции $J(u), g_1(u), \ldots, g_m(u), |g_{m+1}(u)|, \ldots, |g_s(u)|$ полунепрерывны снизу на U_0, выполнены условия* (3), (9);

2) *$\Omega(u)$ — стабилизатор задачи* (1), (2), *$P(u)$ — штрафная функция, определенная формулой* (4) *с параметром $p \geqslant \nu$;*

3) *приближения $J_\delta(u)$, $P_\delta(u)$ функций $J(u)$, $P(u)$ удовлетворяют условиям* (5);

4) *параметры $\alpha(\delta)$, $A(\delta)$, $\varepsilon(\delta)$ таковы, что*

$$\alpha(\delta) > 0, \quad A(\delta) > 0, \quad \varepsilon(\delta) \geqslant 0, \quad \forall \delta > 0, \quad \lim_{\delta \to 0}(\alpha(\delta) + \varepsilon(\delta)) = 0, \quad \lim_{\delta \to 0} A(\delta) = +\infty, \tag{12}$$

$$\sup_{\delta > 0}\left(\frac{\delta + \delta A(\delta)}{\alpha(\delta)}\right) < 1, \quad \sup_{\delta > 0} \frac{\varepsilon(\delta)}{\alpha(\delta)} < \infty, \quad \inf_{\delta > 0} \alpha(\delta)(A(\delta))^{\frac{\nu}{p-\nu}} > 0 \tag{13}$$

(*при $p = \nu$ последнее из условий* (13) *не нужно, а условие $\lim\limits_{\delta \to 0} A(\delta) = +\infty$ можно заменить на $A(\delta) > \max\limits_{1 \leqslant i \leqslant s} c_i$*).

Тогда семейство $\{u_\delta, \delta > 0\}$, определяемое методом (8), *таково, что*

$$\lim_{\delta \to 0} J(u_\delta) = J_*, \quad \lim_{\delta \to 0} g_i^+(u_\delta) = 0, \quad i = 1, \ldots, s; \quad \lim_{\delta \to 0} \rho(u_\delta, U_*) = 0, \tag{14}$$

причем пределы в (14) *равномерны относительно выбора $J_\delta(u)$, $P_\delta(u)$ из* (5) *и выбора точки u_δ из* (8), *точнее,*

$$\lim_{\delta \to 0} \sup_{u \in X(\delta)} |J(u) - J_*| = 0, \quad \lim_{\delta \to 0} \sup_{u \in X(\delta)} |g_i^+(u)| = 0, \quad \lim_{\delta \to 0} \sup_{u \in X(\delta)} \rho(u, U_*) = 0, \tag{15}$$

где $X(\delta)$ — множество, представляющее собой объединение множеств $x(\delta) = \{u \in U_\Omega: t_\delta(u) = J_\delta(u) + A(\delta)P_\delta(u) + \alpha(\delta)\Omega(u) \leqslant t_{\delta *} + \varepsilon(\delta)\}$ по всевозможным реализациям $J_\delta(u)$, $P_\delta(u)$ из (5).

Пусть наряду с условиями 1)–4) *еще выполнено условие*

5) *$U_\Omega = U_0$, функция $\Omega(u)$ полунепрерывна снизу на U_0,*

$$\lim_{\delta \to 0} \frac{\delta + \delta A(\delta)}{\alpha(\delta)} = 0, \quad \lim_{\delta \to 0} \frac{\varepsilon(\delta)}{\alpha(\delta)} = 0, \quad \lim_{\delta \to 0} \alpha(\delta)(A(\delta))^{\frac{\nu}{p-\nu}} = +\infty \tag{16}$$

(*при* $p=\nu$ *последнее из условий* (15) *не нужно*). *Тогда справедливы равенства* (14) *и, кроме того,*

$$\lim_{\delta\to 0}\Omega(u_\delta)=\Omega_*,\quad \lim_{\delta\to 0}\rho(u_\delta, U_{**})=0, \tag{17}$$

где $\Omega_*=\inf\limits_{u\in U_*}\Omega(u)$, $U_{**}=\{u\in U_*\colon \Omega(u)=\Omega_*\}$, *причем пределы в* (17) *равномерны относительно выбора* $J_\delta(u)$, $P_\delta(u)$ *из* (5) *и выбора точки* u_δ *из* (8), *т. е.*

$$\lim_{\delta\to 0}\sup_{u\in X(\delta)}|\Omega(u)-\Omega_*|=0,\quad \lim_{\delta\to 0}\sup_{u\in X(\delta)}\rho(u, U_{**})=0. \tag{18}$$

Доказательство. Первое из неравенств (13) влечет за собой условие (11). Отсюда и из (10) следует, что множество $x(\delta)=\{u\in U_\Omega\colon t_\delta(u)\leqslant t_{\delta *}+\varepsilon(\delta)\}$ непусто при всех $\varepsilon(\delta)>0$ (если $\varepsilon(\delta)=0$, то условие $x(\delta)\neq\varnothing$ предполагается). Кроме того, по определению стабилизатора множество $U_{\Omega *}=U_*\cap U_\Omega\neq\varnothing$. Возьмем произвольные точки $u_\delta\in x(\delta)$, $u_*\in U_{\Omega *}$. Учитывая, что $\Omega(u)\geqslant 0$, $P(u)\geqslant 0$ на U_Ω, $P(u_*)=0$ и неравенства (5), (8), (10), можем написать следующую цепочку неравенств

$$\begin{aligned}
J(u_\delta)+A(\delta)P(u_\delta)&\leqslant J(u_\delta)+A(\delta)P(u_\delta)+\alpha(\delta)\Omega(u_\delta)\leqslant\\
&\leqslant J_\delta(u_\delta)+A(\delta)P_\delta(u_\delta)+\alpha(\delta)\Omega(u_\delta)+(\delta+\delta A(\delta))(1+\Omega(u_\delta))=\\
&=t_\delta(u_\delta)+(\delta+\delta A(\delta))(1+\Omega(u_\delta))\leqslant t_{\delta *}+\varepsilon(\delta)+(\delta+\delta A(\delta))(1+\Omega(u_\delta))\leqslant\\
&\leqslant J(u_*)+A(\delta)P(u_*)+\alpha(\delta)\Omega(u_*)+\varepsilon(\delta)+\\
&+(\delta+\delta A(\delta))(1+\Omega(u_\delta))+(\delta+\delta A(\delta))(1+\Omega(u_*))\leqslant\\
&\leqslant J(u_\delta)+A(\delta)P(u_\delta)+B(A(\delta))^{-\frac{\nu}{p-\nu}}+\alpha(\delta)\Omega(u_*)+\varepsilon(\delta)+\\
&+(\delta+\delta A(\delta))(1+\Omega(u_\delta))+(\delta+\delta A(\delta))(1+\Omega(u_*))\quad\forall\delta>0.
\end{aligned}\tag{19}$$

Выделив второе и последнее звенья этой цепочки, имеем

$$\begin{aligned}
\alpha(\delta)\Omega(u_\delta)&\leqslant B(A(\delta))^{-\frac{\nu}{p-\nu}}+\alpha(\delta)\Omega(u_*)+\varepsilon(\delta)+\\
&+(\delta+\delta A(\delta))(1+\Omega(u_\delta))+(\delta+\delta A(\delta))(1+\Omega(u_*)),\ \delta>0.
\end{aligned}$$

Пользуясь произволом в выборе точки $u_*\in U_{\Omega *}$ в этом неравенстве величину $\Omega(u_*)$ можем заменить на $\Omega_*=\inf\limits_{U_\Omega}\Omega(u)$ и переписать его в виде:

$$\begin{aligned}
(\alpha(\delta)-\delta-\delta A(\delta))\Omega(u_\delta)&\leqslant(\alpha(\delta)+\delta+\delta A(\delta))\Omega_*+\varepsilon(\delta)+\\
&+2(\delta+\delta A(\delta))+B(A(\delta))^{-\frac{\nu}{p-\nu}}
\end{aligned}$$

или

$$\Omega(u_\delta)\leqslant\Omega_*+\gamma(\delta),\quad \delta>0, \tag{20}$$

где

$$\gamma(\delta)=\frac{2\left(\frac{\delta+\delta A(\delta)}{\alpha(\delta)}\right)(1+\Omega_*)+\frac{\varepsilon(\delta)}{\alpha(\delta)}+\frac{B}{\alpha(\delta)(A(\delta))^{\frac{\nu}{p-\nu}}}}{1-\frac{\delta+\delta A(\delta)}{\alpha(\delta)}}. \tag{21}$$

Из выражения (21) следует, что $\gamma(\delta)\geqslant 0$, $\sup\limits_{\delta>0}\gamma(\delta)<\infty$ при выполнении условий (12), (13) и $\lim\limits_{\delta\to 0}\gamma(\delta)=0$ при условиях (12), (16). Далее, взяв первое

и предпоследнее звенья цепочки неравенств (19), с учетом уже доказанного неравенства (20), имеем

$$J(u_\delta)+A(\delta)P(u_\delta)\leqslant J_*+\alpha(\delta)\Omega(u_*)+\varepsilon(\delta)+(\delta+\delta A(\delta))(2+\Omega_*+\gamma(\delta)+\Omega(u_*)).$$

Пользуясь произволом в выборе $u_*\in U_{\Omega *}$, отсюда получаем

$$J(u_\delta)+A(\delta)P(u_\delta)\leqslant J_*+\beta(\delta), \tag{22}$$

где

$$\beta(\delta)=\alpha(\delta)\Omega_*+\varepsilon(\delta)+\alpha(\delta)\cdot\left(\frac{\delta+\delta A(\delta)}{\alpha(\delta)}\right)(2+2\Omega_*+\gamma(\delta))\geqslant 0,\quad \delta>0, \tag{23}$$

причем $\lim\limits_{\delta\to 0}\beta(\delta)=0$ при условиях (12), (13). Из (22), (23) и леммы 3.3 следуют оценки

$$0\leqslant g_i^+(u_\delta)\leqslant\rho(\delta),\quad i=1,\ldots,s; \tag{24}$$

$$-|c|(\rho(\delta))^\nu\leqslant J(u_\delta)-J_*\leqslant\beta(\delta),\quad\forall\delta>0, \tag{25}$$

где величина $\beta(\delta)$ определена формулой (23), величина $\rho(\delta)$ — формулой (3.21), причем $\lim\limits_{\delta\to 0}\rho(\delta)=0$. Оценки (20), (24), (25) означают, что семейство точек $\{u_\delta,\ \delta>0\}$ удовлетворяет условиям (3.8)–(3.10). Отсюда и из леммы 3.1 следуют равенства (14), (17). Остается доказать равенства (15), (18). Первые два равенства (15) вытекают из оценок (24), (25), так как величины $\rho(\delta)$, $\beta(\delta)$, как видно из формул (3.21), (23), не зависят от конкретной реализации $J_\delta(u)$, $P_\delta(u)$ из (5) и выбора точки u_δ из (8). Докажем третье из равенств (15). Пусть $\overline{\lim\limits_{\delta\to 0}}\sup\limits_{u\in X(\delta)}\rho(u, U_*)=\lim\limits_{k\to\infty}\sup\limits_{u\in X(\delta_k)}\rho(u, U_*)$, где $\{\delta_k\}\to 0$. По определению верхней грани для каждого номера k найдется точка $u_k\in X(\delta_k)$ такая, что

$$\sup_{u\in X(\delta_k)}\rho(u, U_*)\leqslant\rho(u_k, U_*)+\frac{1}{k},\quad k=1,2,\ldots \tag{26}$$

Включение $u_k\in X(\delta_k)$ означает, что $u_k\in U_\Omega$, $t_{\delta_k}(u_k)\leqslant t_{\delta_k *}+\varepsilon(\delta_k)$, где $t_{\delta_k}(u)$ составлена по формуле (6) для каких-то реализаций $J_{\delta_k}(u)$, $P_{\delta_k}(u)$ из (5), т. е. $u_k=u_{\delta_k}\in x(\delta_k)$. Из третьего равенства (14) имеем $\lim\limits_{k\to\infty}\rho(u_k, U_*)=0$. Отсюда и из (26) следует: $0\leqslant\overline{\lim\limits_{\delta\to 0}}\sup\limits_{u\in X(\delta)}\rho(u, U_*)\leqslant\overline{\lim\limits_{\delta\to 0}}\sup\limits_{u\in X(\delta)}\rho(u, U_*)=\lim\limits_{k\to\infty}\sup\limits_{u\in X(\delta_k)}\rho(u_k, U_*)\leqslant\lim\limits_{k\to\infty}\rho(u_k, U_*)=0$, что равносильно третьему равенству (15). Аналогично, выделяя последовательности $\{\delta_k\}\to 0$, для которых

$$\overline{\lim_{\delta\to 0}}\sup_{u\in X(\delta)}|\Omega(u)-\Omega_*|=\lim_{k\to\infty}\sup_{u\in X(\delta_k)}|\Omega(u)-\Omega_*|,$$

$$\overline{\lim_{\delta\to 0}}\rho(u, U_{**})=\lim_{k\to\infty}\sup_{u\in X(\delta_k)}\rho(u, U_{**}),$$

и последовательности $\{v_k\}$, $\{w_k\}$ со свойствами

$$v_k\in X(\delta_k),\quad \sup_{u\in X(\delta_k)}|\Omega(u)-\Omega_*|\leqslant|\Omega(v_k)-\Omega_*|+\frac{1}{k},\quad k=1,2,\ldots,$$

$$w_k\in X(\delta_k),\quad \sup_{u\in X(\delta_k)}\rho(u, U_{**})\leqslant\rho(w_k, U_{**})+\frac{1}{k},\quad k=1,2,\ldots,$$

с помощью уже доказанных равенств (17) приходим к (18). Теорема 1 доказана. □

3. Отдельно остановимся на выпуклых задачах (1), (2), когда $\mathcal{M}=B$ — банахово пространство. Покажем, что в этом случае для семейства $\{u_\delta, \delta>0\}$, определяемого методом стабилизации (8), сходимость к решению в норме B удается получить и тогда, когда функция $\Omega(u)$ является слабым стабилизатором (определение 3.2).

Теорема 2. *Пусть*

1) *B — рефлексивное банахово пространство, U_0 — выпуклое замкнутое множество из B, функции $J(u), g_i(u)$, $i=1,\ldots,m$, $|g_i(u)|$, $i=m+1,\ldots,s$, выпуклы и полунепрерывны снизу на U_0, множество* (2) *непусто, выполняются условия* (3), (9);

2) *функция $\Omega(u)$ полунепрерывна снизу, строго равномерно выпукла на U_0; $P(u)$ — штрафная функция, определенная формулой* (4) *с параметром $p\geqslant\nu$;*

3) *приближения $J_\delta(u), P_\delta(u)$ функций $J(u), P(u)$ удовлетворяют условиям* (5);

4) *параметры $\alpha(\delta), A(\delta), \varepsilon(\delta)$ удовлетворяют условиям* (12), (16).

Тогда семейство $\{u_\delta, \delta>0\}$, определенное методом (8), *таково, что*

$$\lim_{\delta\to0} J(u_\delta)=J_*,\quad \lim_{\delta\to0} g_i^+(u_\delta)=0,\quad i=1,\ldots,s,$$
$$\lim_{\delta\to0}\Omega(u_\delta)=\Omega_*,\quad \lim_{\delta\to0}\|u_\delta-v_*\|_B=0, \tag{27}$$

где v_* — точка минимума функции $\Omega(u)$ на множестве U_*, причем пределы в (27) *равномерны относительно выбора $J_\delta(u), P_\delta(u)$ из* (5) *и точки u_δ из* (8).

Доказательство. Повторяя доказательство теоремы 1 при $U_\Omega=U_0$, убеждаемся, что оценки (20)–(25) сохраняют силу при всех достаточно малых $\delta>0$. Это значит, что при сделанных предположениях семейство $\{u_\delta, \delta>0\}$ из (8) удовлетворяет условиям леммы 3.2, из которой следуют существование и единственность точки v_* и равенства (27). Равномерность пределов (27) доказывается также, как равенства (15), (18). □

4. Сделаем несколько замечаний, комментирующих теоремы 1, 2.

Замечание 1. В качестве параметров $\alpha(\delta), A(\delta), \varepsilon(\delta)$ метода (8), удовлетворяющих условиям (12), (13), (16), можно, например, принять

$$\alpha(\delta)=a_1\delta^\alpha,\quad A(\delta)=a_2\delta^{-A},\quad \varepsilon(\delta)=a_3\delta^\varepsilon,\quad \delta>0, \tag{28}$$

где $a_1>0$, $a_2>0$, $a_3\geqslant0$, $0<\alpha\leqslant\varepsilon$, $A>0$, $A+\alpha\leqslant1$, $\alpha\leqslant A\frac{p}{p-\nu}$ (при $p=\nu$ последнее неравенство не нужно). Поскольку условия (12), (13), (16) интересуют нас лишь при малых $\delta>0$, то от параметров $\alpha(\delta), A(\delta), \varepsilon(\delta)$ достаточно требовать соблюдения этих условий на каком-либо полуинтервале $0<\delta\leqslant\delta_0$, содержащем допускаемые неравенствами (5) погрешности измерений.

Замечание 2. Примером стабилизатора, который обладает всеми свойствами, требуемыми в теореме 1, являются функции (3.5), когда $\mathcal{M}=E^n$ и $U_\Omega=U_0$.

Замечание 3. Правые части условий (5) на погрешности можно заменить на $\delta c_0(1+\Omega(u))$, где $c_0=\mathrm{const}>0$. Теоремы 1, 2 и их доказательства при этом сохраняются, нужно лишь первое из неравенств (13) заменить на $\sup\limits_{\delta>0}\frac{c_0(\delta+\delta A(\delta))}{\alpha(\delta)}<1$, а условие (11) — на $c_0(\delta+\delta A(\delta))\leqslant\alpha(\delta)$, $\delta>0$.

Замечание 4. Кратко остановимся на методе стабилизации для задачи второго типа

$$J(u)\to\inf,\quad u\in U, \tag{29}$$

когда множество U известно точно (например, $U=\mathcal{M}$ или $U=B$). Эта задача является частным случаем задачи (1), (2) при $m=s=0$, $U=U_0$. Пусть для приближений $J_\delta(u)$ функции $J(u)$ выполнено первое из равенств (5), где $\Omega(u)$ — стабилизатор задачи (29) (см. определение 3.1 при $U=U_0$). Тогда метод (8) применим и к задаче (29), если в качестве функции Тихонова взять $t_\delta(u)=J_\delta(u)+\alpha(\delta)\Omega(u)$. Теоремы сходимости метода (8) в этом случае получатся из теорем 1, 2, если в них и в их доказательствах принять $U_0=U$, $P(u)=P_\delta(u)=0$, а в (12), (13), (16) исключить элементы, содержащие параметр $A(\delta)$ (например, равенство $\lim\limits_{\delta\to0}\frac{\delta+\delta A(\delta)}{\alpha(\delta)}=0$ здесь должно быть заменено на $\lim\limits_{\delta\to0}\frac{\delta}{\alpha(\delta)}=0$).

Замечание 5. Равенство $\lim\limits_{\delta\to0}\sup\limits_{x\in X(\delta)}\rho(u,U_*)=0$ из (15) выражает стремление к нулю при $\delta\to0$ уклонение множества $X(\delta)$ от множества U_* (см. (1.10), упражнение 1.6). Равенство $\lim\limits_{\delta\to0}\sup\limits_{x\in X(\delta)}|J(\delta)-J_*|=0$ означает, что уклонение множества $\{J(u), u\in X(\delta), \delta>0\}$ от множества, состоящего из одной точки J_*, стремится к нулю при $\delta\to0$. Аналогичный смысл имеют остальные равенства (15), (18).

Замечание 6. Выше мы считаем, что стабилизатор $\Omega(u)$ известен точно. Однако такое предположение не всегда оправдано. Например, если используется стабилизатор $\Omega(u)=\int\limits_a^b(u^2(t)+(\dot u(t))^2)dt$, то для вычисления интегралов здесь придется применять какие-либо квадратурные формулы и неизбежно появятся погрешности. Для охвата подобных ситуаций можно считать, что вместо точных значений $\Omega(u)$ известно их приближение $\Omega_\chi(u)$, причем

$$|\Omega_\chi(u)-\Omega(u)|\leqslant\chi(1+\Omega(u)),\quad u\in U_\Omega,\quad 0<\chi\leqslant1. \tag{30}$$

Условие (30) при возможных больших значениях $\Omega(u)$ характеризует относительную погрешность в задании стабилизатора, так как тогда $\left|\frac{\Omega_\chi(u)}{\Omega(u)}-1\right|\leqslant\chi\left(1+\frac{1}{\Omega(u)}\right)\simeq\chi$. При выполнении условий (30) в функции (6) вместо $\Omega(u)$ надо взять $\Omega_\chi(u)$, а параметр $\chi=\chi(\delta)$ надо согласовать с погрешностью δ, взяв $\sup\limits_{\delta>0}\chi(\delta)<\infty$ в дополнение к условиям (13) и $\lim\limits_{\delta\to0}\chi(\delta)=0$ — к условиям (16). Наличие в правых частях неравенств (5), (30) слагаемого 1 в сумме $1+\Omega(u)$ оправдано тем, что возможны малые значения $\Omega(u)$ или даже $\Omega(u)=0$ в некоторых точках u, и в этом случае замена $1+\Omega(u)$ на $\Omega(u)$ в (5), (30) привела бы к неоправданно жестким требованиям на точность задания функций $J(u), P(u), \Omega(u)$.

Замечание 7. Обсудим содержательный смысл элементов множества U_{**}, введенного в (17).

Определение 1. Точку $u_*\in U_*$ назовем *нормальным решением* задачи (1), (2) по функции $\Omega(u)$ или, короче, Ω-нормальным решением, если

$\Omega(u_*)=\inf\limits_{u\in U_*}\Omega(u)=\Omega_*$. Если $\Omega(u)=\|u\|$ — норма в банаховом пространстве, то Ω-нормальное решение называют просто нормальным решением задачи.

В задачах оптимального планирования функция $\Omega(u)$ может выражать собой стоимость затрат на организационные и технологические перестройки при переходе от существующего состояния производства к его новому состоянию, соответствующему оптимальному плану u. Тогда среди всех оптимальных планов $u\in U_*$ естественно выбрать тот план u_*, для которого стоимость затрат $\Omega(u)$ минимальна. Это означает, что наилучший оптимальный план u_* представляет собой Ω-нормальное решение задачи (1), (2) [374; 695]. Как утверждается в теореме 1, при определенных условиях метод стабилизации (8) может быть использован для получения приближений к Ω-нормальному решению задачи (1), (2).

Задача поиска Ω-нормального решения относится к так называемым *лексикографическим* задачам минимизации, когда имеется некоторое упорядоченное множество функций $J_1(u), J_2(u), \ldots, J_l(u)$, определенных на некотором множестве U и требуется последовательно решить задачи минимизации в следующем порядке. Сначала решается задача: $J_1(u)\to\inf$, $u\in U$, находится множество $U_{1*}=\{u\in U\colon J_1(u)=\inf\limits_U J_1(u)\}$. Затем рассматривается задача: $J_2(u)\to\inf$, $u\in U_{1*}$, определяется множество $U_{2*}=\{u\in U_{1*}\colon J_2(u)=\inf\limits_{U_{1*}} J_2(u)\}$ и т. д., пока не будет найдено множество $U_{l*}=\{u\in U_{l-1*}\colon J_l(u)=\inf\limits_{U_{l-1*}} J_l(u)\}$. Отсюда ясно, что Ω-нормальное решение задачи (1), (2) является решением лексикографической задачи, когда $J_1(u)=J(u)$, $J_2(u)=\Omega(u)$, $U_{1*}=U_*$, $U_{2*}=U_{**}$. Нетрудно понять, что лексикографические задачи относятся к неустойчивым задачам, так как множества точек минимума $U_{1*},\ldots,U_{l*}$ при небольших возмущениях входных данных могут сильно изменяться, более того, могут оказаться пустыми. Теорема 1 показывает, что для решения некоторых таких задач может быть использован метод стабилизации [169].

5. Для иллюстрации изложенного метода стабилизации рассмотрим несколько примеров.

Пример 1. Задача:

$$J(u)=\int\limits_c^d\Big(\int\limits_a^b K(s,t,u(t))dt\Big)^2 ds\to\inf,\quad u\in U=C[a,b],\qquad(31)$$

где $K(s,t,u)$ — непрерывная функция по совокупности аргументов $(s,t,u)\in[c,d]\times[a,b]\times E^1$. Эта задача тесно связана с интегральным уравнением $\int\limits_a^b K(s,t,u(t))dt=0$ и, вообще говоря, неустойчива в метрике $C[a,b]$ [695; 697]. Пусть задача (31) имеет хотя бы одно решение $u=u_*(t)\in H^1[a,b]$. Тогда функция $\Omega(u)=\|u\|^2_{H^1}=\int\limits_a^b(|u(t)|^2+|\dot u(t)|^2)dt$, определенная на множестве $U_\Omega=H^1[a,b]$, является стабилизатором задачи (31) в метрике $C[a,b]$ (пример 3.2). Пусть $|K(s,t,u)|\leqslant c_0(1+|u|)$, $c_0=\mathrm{const}>0$, и пусть вместо точной функции $K(s,t,u)$ нам известно приближение $K_\delta(s,t,u)$ такое, что

$$|K_\delta(s,t,u)-K(s,t,u)|\leqslant\delta(1+|u|),\quad \delta>0,\quad \forall(s,t,u)\in[c,d]\times[a,b]\times E^1.$$

Если в качестве приближенного значения функции $J(u)$ возьмем $J_\delta(u)=\int\limits_c^d\Big(\int\limits_a^b K_\delta(s,t,u(t))dt\Big)^2 ds$, получим

$$\begin{aligned}|J_\delta(u)-J(u)|&=\Big|\int\limits_c^d\Big(\int\limits_a^b(K_\delta(s,t,u(t))-K(s,t,u(t)))dt\Big)\times\\&\times\Big(\int\limits_a^b(K_\delta(s,t,u(t))+K(s,t,u(t)))dt\Big)ds\Big|\leqslant\\&\leqslant\int\limits_c^d\Big(\int\limits_a^b\delta(1+|u(t)|)dt\cdot\int\limits_a^b(\delta+2c_0)(1+|u(t)|)dt\Big)ds\leqslant\\&\leqslant\delta(\delta+2c_0)(d-c)(b-a+1)\Big(1+\int\limits_a^b|u(t)|^2dt\Big)\leqslant\\&\leqslant\delta(\delta+2c_0)(d-c)(b-a+1)^2(1+\|u\|^2_{H^1}),\end{aligned}$$

т. е. погрешность согласована со стабилизатором (см. замечание 6). Функция Тихонова

$$t_\delta(u)=\int\limits_c^d\Big(\int\limits_a^b K_\delta(s,t,u(t))dt\Big)^2 ds+\alpha(\delta)\int\limits_a^b(|u(t)|^2+|\dot u(t)|^2)dt$$

определена на $U_\Omega=H^1[a,b]$. Если $u_\delta=u_\delta(t)\in H^1[a,b]$, $t_\delta(u_\delta)\leqslant\inf\limits_{H^1[a,b]} t_\delta(u)+\varepsilon(\delta)$, параметры $\alpha(\delta)$, $\varepsilon(\delta)$ выбраны так, как требуется в теореме 2, то семейство $\{u_\delta(t),\delta>0\}$ сходится ко множеству U_* в метрике $C[a,b]$.

Если задача (31) имеет хотя бы одно решение $u(t)$, удовлетворяющее условию Гельдера, то в качестве стабилизатора этой задачи в метрике $C[a,b]$ можно использовать стабилизатор 3.7.

Пример 2. Рассмотрим задачу минимизации квадратичной функции

$$J(u)=\|Au-b\|^2_F\to\inf,\quad u\in U\qquad(32)$$

где $A\in\mathcal{L}(H\to F)$, H, F — гильбертовы пространства, $b\in H$, U — выпуклое замкнутое множество из H (например, $U=H$). Пусть эта задача имеет хотя бы одно решение. Условие разрешимости задачи (32) рассматривались в § 8.2 (теоремы 8.2.12, 8.2.13). Будем искать точку $u_*\in U_*$, ближе всего расположенную к заданному элементу $\overline{u}\in H$. Такая точка является Ω-нормальным решением задачи (32), если $\Omega(u)=\|u-\overline{u}\|^2_H$. Пусть оператор A и элемент b заданы своими приближениями $A_\delta\in\mathcal{L}(H\to F)$, $b_\delta\in F$, причем

$$\|A_\delta-A\|\leqslant\delta,\quad \|b_\delta-b\|\leqslant\delta,\qquad(33)$$

пусть множество U известно точно. В качестве приближенной функции тогда естественно взять $J_\delta(u)=\|A_\delta u-b_\delta\|^2$. Допускаемая при этом погрешность оценивается так:

$$\begin{aligned}|J_\delta(u)-J(u)|&=|\langle(A_\delta u-b_\delta)-(Au-b),(A_\delta u-b_\delta)+(Au-b)\rangle|\leqslant\\&\leqslant(\delta\|u\|+\delta)((2\|A\|+\delta)\|u\|+2\|b\|+\delta)\leqslant\\&\leqslant\delta(1+\|u\|)^2 2(\|A\|+\|b\|+\delta).\end{aligned}$$

Но $(1+\|u\|)^2\leqslant(1+\|\overline{u}\|+\|u-\overline{u}\|)^2\leqslant 2(1+\|\overline{u}\|)^2(1+\|u-\overline{u}\|^2)$, поэтому

$$|J_\delta(u)-J(u)|\leqslant 4\delta(1+\|\overline{u}\|)^2(\|A\|+\|b\|+\delta)(1+\Omega(u)),\quad u\in H.\qquad(34)$$

Отсюда видно, что погрешность с учетом замечания 3 удовлетворяет условию (5). Задача (32) при условиях (33), вообще говоря, неустойчива [334; 695] (см. упражнение 1.2). Для определения ее Ω-нормального решения можно воспользоваться методом стабилизации (8). Составим функцию Тихонова $t_\delta(u)=\|A_\delta u-b_\delta\|^2+\alpha\|u-\overline{u}\|^2_H$, $u\in H$, и определим точку u_δ из условий:

$$u_\delta\in U,\quad t_\delta(u_\delta)\leqslant\inf_{u\in U}t_\delta(u)+\varepsilon(\delta) \tag{35}$$

В частности, если $U=H$, то сильно выпуклая дифференцируемая функция $t_\delta(u)$ достигает минимума в единственной точке u_δ, которая является решением операторного уравнения

$$t'_\delta(u)=2A^*_\delta(A_\delta u-b_\delta)+2\alpha(\delta)(u-\overline{u})=0.$$

Если

$$\lim_{\delta\to0}\alpha(\delta)=\lim_{\delta\to0}\varepsilon(\delta)=\lim_{\delta\to0}\frac{\delta+\varepsilon(\delta)}{\alpha(\delta)}=0, \tag{36}$$

то из теоремы 2 следует, что $\lim\limits_{\delta\to0}\|u_\delta-u_*\|_H=0$, где u_* — Ω-нормальное решение задачи (32).

Пример 3. Задача:

$$J(u)=\int_c^d\Big(\int_a^b K(s,t)u(t)dt-b(s)\Big)^2ds\to\inf,\quad u\in U=L_2[a,b], \tag{37}$$

где $b(s)\in L_2[c,d]$, $K(s,t)\in L_2(Q)$, $Q=[c,d]\times[a,b]$. Эта задача тесно связана с интегральным уравнением Фредгольма первого рода

$$\int_a^b K(s,t)u(t)dt=b(s),\quad c\leqslant s\leqslant d.$$

Задача (37) является частным случаем задачи (32) при $H=L_2[a,b]$, $F=L_2[c,d]$

$$Au=\int_a^b K(t,s)u(s)ds \tag{38}$$

— оператор Фредгольма (пример 8.3.10). Пусть $U_*\neq\varnothing$ и пусть вместо функций $K(s,t)$, $b(s)$ известны их приближения $K_\delta(s,t)\in L_2(Q)$, $b_\delta(s)\in L_2[c,d]$ такие, что

$$\iint_Q|A_\delta(s,t)-A(s,t)|^2dsdt\leqslant\delta^2,\quad\int_c^d|b_\delta(s)-b(s)|^2ds\leqslant\delta^2, \tag{39}$$

Тогда для приближенной функции $J_\delta(u)=\int\limits_c^d\Big(\int\limits_a^b K_\delta(s,t)u(t)dt-b_\delta(s)\Big)^2ds$ справедлива оценка (34), где $\Omega(u)=\int\limits_a^b|u(t)-\overline{u}(t)|^2dt$. Функция Тихонова имеет вид

$$t_\delta(u)=\int_c^d\Big(\int_a^b A_\delta(s,t)u(t)dt-b_\delta(s)\Big)^2ds+\alpha(\delta)\int_a^b|u(t)-\overline{u}(t)|^2dt,\quad u\in L_2[a,b]. \tag{40}$$

Если выполнены условия (36), то точки u_δ, определяемые из (35), (40) при $U=L_2[a,b]$ таковы, что $\lim\limits_{\delta\to0}\int\limits_a^b|u_\delta(t)-u_*(t)|^2dt=0$, где $u_*=u_*(t)$ — Ω-нормальное решение задачи (37).

Пусть задача (37) имеет хотя бы одно решение $v_*=v_*(t)\in H^1[a,b]$. Покажем, как тогда можно построить семейство $\{v_\delta\in U,\ \delta>0\}$, сходящееся при $\delta\to0$ к одному из решений этой задачи равномерно на $[a,b]$. Заметим, что оператор (38) определен и на $H_1=H^1[a,b]\subset L_2[a,b]$, более того, $A\in\mathcal{L}(H_1\to F)$, так как

$$\|A\|^2_{\mathcal{L}(H_1\to F)}=\sup_{\|u\|_{H^1[a,b]}\leqslant1}\int_c^d((Au)(s))^2ds\leqslant\sup_{\|u\|_{L_2[a,b]}\leqslant1}\int_c^d((Au)(s))^2ds=$$
$$=\|A\|_{\mathcal{L}(H\to F)}=\iint_Q K^2(s,t)dtds. \tag{41}$$

Поэтому имеет смысл задача:

$$J(u)\to\inf,\quad u\in U_1=H^1[a,b], \tag{42}$$

где функция $J(u)$ взята из (37). Множество решений задачи (42) обозначим через U_{1*}. По предположению, существует точка $v_*\in U_*\cap H^1[a,b]$. Нетрудно видеть, что $J(v_*)\geqslant\inf\limits_{U_1}J(u)=J_{1*}\geqslant\inf\limits_{U}J(u)=J_*=J(v_*)$. Отсюда следует, что $J_{1*}=J_*$ и $U_{1*}=U_*\cap H^1[a,b]\neq\varnothing$. Рассуждая так же, как при выводе оценки (34) с учетом (39), (41), получаем

$$|J_\delta(u)-J(u)|\leqslant4\delta(\|A\|_{\mathcal{L}(H\to F)}+\|b\|_F+\delta)(1+\|u\|^2_{H^1[a,b]})\quad\forall u\in H^1[a,b].$$

Составим функцию Тихонова задачи (42): $t_\delta(u)=J_\delta(u)+\alpha(\delta)\|u\|^2_{H^1[a,b]}$ и определим точку $v_\delta\in U_1=H^1[a,b]$, $t_\delta(v_\delta)\leqslant\inf\limits_{U_1}t_\delta(u)+\varepsilon(\delta)$. При выполнении условий (36) согласно теореме 2 получим $\lim\limits_{\delta\to0}\|v_\delta-w_*\|_{H^1[a,b]}=0$, где $w_*\in U_{1*}$, $\|w_*\|_{H^1[a,b]}=\inf\limits_{U_{1*}}\|u\|_{H^1[a,b]}$. Так как $U_{1*}=U_*\cap H^1[a,b]$, то $w_*\in U_*$, так что построенное семейство $\{v_\delta,\ \delta>0\}$ сходится к решению w_* задачи (37) в норме $H^1[a,b]$ и, тем более, в норме $C[a,b]$.

Пример 4. Рассмотрим задачу оптимального управления:

$$J(u)=|x(T,u)-b|^2_{E^n}\to\inf,\quad u\in U \tag{43}$$

где $x(t;u)=x(t)$ — решение системы

$$\dot x(t)=D(t)x(t)+B(t)u(t),\quad t_0\leqslant t\leqslant T;\quad x(t_0)=0, \tag{44}$$

U — выпуклое замкнутое множество из $L^r_2[t_0,T]$ (обозначения см. в примере 8.2.14). Введем оператор $Au=x(T,u)$, действующий из $H=L^r_2[t_0,T]$ в $F=E^n$. В примере 8.2.14 было доказано, что $A\in\mathcal{L}(H\to F)$, и задача (43), (44) сводится к задаче (32). Пусть в задаче (43), (44) множество U_* решений непусто (см. теоремы 8.2.12, 8.2.13). Пусть вместо матриц $D(t)$, $B(t)$ и вектора b известны кусочно-непрерывные приближения $D_\delta(t)$, $B_\delta(t)$ и вектор b_δ такие, что

$$\|D_\delta(t)-D(t)\|\leqslant\delta,\quad\|B_\delta(t)-B(t)\|\leqslant\delta\quad\forall t\in[t_0,T],\quad|b_\delta-b|\leqslant\delta. \tag{45}$$

Определим приближение A_δ к оператору A следующим образом: $A_\delta u=x_\delta(T,u)$, где $x_\delta=x_\delta(t;u)$ — решение системы:

$$\dot x_\delta(t)=D_\delta(t)x_\delta(t)+B_\delta(t)u(t),\quad t_0\leqslant t\leqslant T,\quad x(t_0)=0. \tag{46}$$

Как и в примере 8.2.14 нетрудно установить, что $A_\delta \in \mathcal{L}(H \to F)$ $\forall \delta > 0$. Оценим $\|A_\delta - A\|$. Из (44), (46) имеем: $x_\delta(t;u) - x(t;u) = \int_{t_0}^{t} [D_\delta(\tau)x_\delta(\tau) + B_\delta(\tau)\ u(\tau) - D(\tau)x(\tau) - B(\tau)u(\tau)]d\tau = \int_{t_0}^{t} (D_\delta(\tau)[x_\delta(\tau;u) - x(\tau;u)] + [D_\delta(\tau) - D(\tau)]x(\tau;u) + [B_\delta(\tau) - B(\tau)]u(\tau))d\tau$. Отсюда и из (45) следует: $|x_\delta(t;u) - x(t;u)| \leqslant \sup_{\tau \in [t_0, T]} \|D_\delta(\tau)\| \int_{t_0}^{t} |x_\delta(\tau;u) - x(\tau;u)|d\tau + \delta \int_{t_0}^{T} |x(\tau;u)|d\tau + \delta \int_{t_0}^{T} |u(t)|\ dt$ $\forall t \in [t_0, T]$. Далее воспользуемся леммой 6.3.1. Учитывая неравенство $\sup_\tau \|D_\delta(\tau)\| \leqslant \sup_t \|D(t)\| + \delta = D_{\max} + \delta$ и оценку (8.2.8), получим

$$\|x_\delta(t;u) - x(t;u)\| \leqslant e^{(D_{\max}+\delta)(T-t_0)} \cdot \delta[(T-t_0)c_0 + \sqrt{T-t_0}]\|u\|_{L_2^r} \quad \forall t \in [t_0, T]. \tag{47}$$

Взяв в (47) $t = T$, придем к нужной оценке:

$$\|A_\delta - A\| = \sup_{\|u\|_{L_2^r} \leqslant 1} \|A_\delta u - Au\| = \sup_{\|u\|_{L_2^r} \leqslant 1} |x_\delta(T;u) - x(T;u)| \leqslant \delta c_1,$$
$$c_1 = e^{(D_{\max}+\delta)(T-t_0)} \cdot [(T-t_0)c_0 + \sqrt{T-t_0}]. \tag{48}$$

Таким образом, условия (33) на погрешности в этой задаче выполнены. Далее, предлагаем читателю взять стабилизатор $\Omega(u) = \|u\|_{L_2^r}^2$ и, следуя рассуждениям из примера 2, применить метод стабилизации для задачи (43), (44) при условиях (45), сформулировать условия сходимости метода к Ω-нормальному решению в норме $L_2^r[t_0, T]$.

Пример 5. Рассмотрим каноническую задачу линейного программирования:

$$J(u) = \langle c, u\rangle \to \inf, \quad u \in U = \{u \geqslant 0,\ Au = b\}, \tag{49}$$

где A — матрица размера $m \times n$, $b \in E^m$, $u \in E^n$. Пусть $J_* > -\infty$. Тогда $U_* \neq \varnothing$ (теорема 3.5.1), функция Лагранжа задачи (49) имеет седловую точку (теорема 3.5.5), и, следовательно, в этой задаче выполнено условие (9) сильно согласованной постановки (лемма 5.15.1). Пусть вместо точных c, A, b известны их приближения c_δ, A_δ, b_δ с погрешностями

$$|c_\delta - c| \leqslant \delta, \quad \|A_\delta - A\| \leqslant \delta, \quad |b_\delta - b| \leqslant \delta. \tag{50}$$

Опишем метод стабилизации для задачи (49) при условиях (50). Ограничения $Au = b$ будем учитывать с помощью штрафной функции $P(u) = |Au - b|^2$. Положим $J_\delta(u) = \langle c_\delta, u\rangle$, $P_\delta(u) = |A_\delta u - b_\delta|^2$. Стабилизатор возьмем $\Omega(u) = |u|^2$. Имеем:

$$|J_\delta(u) - J(u)| \leqslant \delta|u| \leqslant \tfrac{1}{2}\delta(1 + |u|^2),$$
$$|P_\delta(u) - P(u)| = |\langle (A_\delta u - b_\delta) - (Au - b), (A_\delta u - b_\delta) + (Au - b)\rangle| \leqslant$$
$$\leqslant \delta(1 + |u|)((2\|A\| + \delta)|u| + 2|b| + \delta) \leqslant 2\delta(1 + |u|)^2(\|A\| + |b| + \delta) \leqslant$$
$$\leqslant 4\delta(1 + |u|^2)(\|A\| + |b| + \delta), \quad u \in E^n.$$

Как видим, погрешности согласованы со стабилизатором (замечание 3). Задача (7) здесь представляет собой задачу квадратичного программирования для сильно выпуклой функции (§ 5.8):

$$t_\delta(u) = \langle c_\delta, u\rangle + A(\delta)|A_\delta u - b_\delta|^2 + \alpha(\delta)|u|^2 \to \inf,$$
$$u \in U_0 = \{u \in E^n:\ u \geqslant 0\}.$$

Определим точку u_δ из условий: $u_\delta \in U_0$, $t_\delta(u_\delta) \leqslant \inf_{u \in U_0} t_\delta(u) + \varepsilon(\delta)$. Если параметры $\alpha(\delta)$, $A(\delta)$, $\varepsilon(\delta)$ удовлетворяют условиям (12), (16), то, согласно теоремам 1, 2, семейство $\{u_\delta, \delta > 0\}$ при $\delta \to 0$ сходится к Ω-нормальному решению в задачи (49).

Другой вариант метода стабилизации для задачи (49) при условиях (50), заключающийся в переходе к новой вспомогательной задаче, также являющейся задачей линейного программирования, будет рассмотрен в § 7.

Приведем еще несколько простых примеров, показывающих, что условия (12), (13), (16) на параметры $\alpha(\delta)$, $A(\delta)$, $\varepsilon(\delta)$ на классе задач (1), (2) не являются грубыми.

Пример 6. Задача: $J(u) \equiv 0 \to \inf$, $u \in U = E^1$. Тогда $J_* = 0$, $U_* = U = E^1$. Возьмем стабилизатор $\Omega(u) = u^2$. Тогда множество U_{**} Ω-нормальных решений состоит из одной точки $u_* = 0$. Пусть $|J_\delta(u) - J(u)| \leqslant \delta(1 + u^2)$. Тогда может случиться, что $J_\delta(u) = -\delta(1 + u^2)$. Функция Тихонова равна $t_\delta(u) = -\delta(1+u^2) + \alpha u^2 = (\alpha - \delta)u^2 - \delta$. Если $\alpha = \alpha(\delta) < \delta$, то $t_{\delta *} = \inf_{E^1} t_\delta(u) = -\infty$. Поэтому для реализуемости метода (8) необходимо условие $\delta \leqslant \alpha(\delta)$ (ср. с (11)). Тогда $t_{\delta *} = -\delta = t_\delta(u_\delta)$, где $u_\delta = 0 = u_*$ $\forall \delta > 0$ без каких-либо дополнительных согласований параметров. Условию $t_\delta(u_\delta) \leqslant t_{\delta *} + \varepsilon(\delta) = -\delta + \varepsilon(\delta)$ удовлетворяет точка $u_\delta = \sqrt{\frac{\varepsilon(\delta)}{\alpha - \delta}} = \sqrt{\frac{\varepsilon(\delta)}{\alpha(\delta)}\left(1 - \frac{\delta}{\alpha(\delta)}\right)^{-1}}$, $\delta > 0$. Отсюда видно, что семейство $\{u_\delta, \delta > 0\}$ ограничено, если $\sup_{\delta > 0} \frac{\delta}{\alpha(\delta)} < 1$, $\sup_{\delta > 0} \frac{\varepsilon(\delta)}{\alpha(\delta)} < \infty$, а для выполнения равенства $\lim_{\delta \to 0} u_\delta = u_* = 0$ нужно еще требовать $\lim_{\delta \to 0} \frac{\varepsilon(\delta)}{\alpha(\delta)} = 0$.

Если оказалось, что $J_\delta(u) = \frac{1}{2}\delta(1 + u)^2$, то $t_\delta(u) = \frac{1}{2}\delta(1 + u)^2 + \alpha u^2$, $t_{\delta *} = \inf_{E^1} t_\delta(u) = \frac{\alpha\delta}{2\alpha + \delta} = t_\delta(u_\delta)$, где $u_\delta = \frac{-\delta}{2\alpha + \delta} = -\frac{\delta}{\alpha(\delta)}\left(2 + \frac{\delta}{\alpha(\delta)}\right)$. Отсюда видно, что сходимость u_δ к Ω-нормальному решению возможна лишь при $\lim_{\delta \to 0} \frac{\delta}{\alpha(\delta)} = 0$.

Пример 7. Задача: минимизировать функцию

$$J(u) = \begin{cases} \frac{(u-1)^2}{1+u^4}, & u > 1, \\ 0, & u \leqslant 1, \end{cases}$$

на множестве $U = \{u \in E^1:\ u \geqslant -1\}$. Здесь $J_* = 0$, $U_* = \{u \in E^1:\ |u| \leqslant 1\}$. Пусть $\Omega(u) = u^2$, $|J_\delta(u) - J(u)| \leqslant \delta(1 + u^2)$ $\forall u \in E^1$, множество U известно точно. Тогда $\Omega_* = \inf_{|u| \leqslant 1} u^2 = 0$, $U_{**} = \{u_* = 0\}$. Допустим, что $J_\delta(u) = J(u)$. Тогда $t_\delta(u) = J(u) + \alpha u^2$, $t_{\delta *}(u) = 0 = t_\delta(0)$. Пусть $\alpha = \alpha(\delta) = \delta^2$, $\varepsilon = \varepsilon(\delta) = 2\delta$ при

$0 \leqslant \delta \leqslant 1$, $\varepsilon(\delta)=2$ при $\delta>1$. Тогда условию $t_\delta(u) \leqslant t_{\delta *}+\varepsilon(\delta)$ удовлетворяет точка $u_\delta=\frac{1}{\sqrt{\delta}}$. Хотя и $\lim\limits_{\delta \to 0} J(u_\delta)=0=J_*$, но $\lim\limits_{\delta \to 0} \rho(u_\delta, U_{**})=+\infty$. Здесь нарушено условие $\sup\limits_{\delta>0} \frac{\varepsilon(\delta)}{\alpha(\delta)}<\infty$.

Рассмотрим другой набор параметров: $\alpha(\delta)=\delta^2$, $\varepsilon(\delta)=2\delta^2(1+\delta)$ при $0<\delta \leqslant 1$, $\varepsilon(\delta)=2$ при $\delta>1$. Пусть снова $J_\delta(u)=J(u)$. Тогда точка $u_\delta=1+\delta$ удовлетворяет условию $t_\delta(u) \leqslant t_{\delta *}+\varepsilon(\delta)$ при $0<\delta \leqslant 1$. Ясно, что $\lim\limits_{\delta \to 0} J(u_\delta)=0=J_*$, $\lim\limits_{\delta \to 0} \rho(u_\delta, U_{**})=0$, но $\lim\limits_{\delta \to 0} \rho(u_\delta, U_{**})=1$. Здесь $\sup\limits_{1 \geqslant \delta>0} \frac{\varepsilon(\delta)}{\alpha(\delta)}<\infty$, однако условие $\lim\limits_{\delta \to 0} \frac{\varepsilon(\delta)}{\alpha(\delta)}=0$ нарушено.

Пример 8. Задача: $J(u) \equiv 0 \to \inf$, $u \in U=\{u \in E^1=U_0: g(u)=u \leqslant 0\}$. Здесь $J_*=0$, $U_*=U$. Функция Лагранжа этой задачи $L(u, \lambda)=\lambda u$, $u \in E^1$, $\lambda \geqslant 0$ имеет седловую точку $(u_*, \lambda^*=0)$ при любом выборе $u_* \in U_*=U$, так что условие (9) сильно согласованной постановки здесь выполнено при $s=1$, $c=\lambda^*=0$, $\nu=1$ (лемма 5.15.1). Пусть приближения $J_\delta(u)$, $g_\delta(u)$ таковы, что $|J_\delta(u)-J(u)| \leqslant \delta(1+u^2)$, $u \in E^1$. Эта задача неустойчива (пример 1.8). Возьмем стабилизатор $\Omega(u)=u^2$, штрафную функцию $P(u)=\max\{0; u\}$ и ее приближение $P_\delta(u)=\max\{u; 0\}-\delta u^2$. Нетрудно видеть, что $|P_\delta(u)-P(u)| \leqslant \delta(1+\Omega(u))$ $\forall u \in E^1$. Пусть $J_\delta(u)=-\delta u^2$. Тогда функция Тихонова

$$t_\delta(u)=-\delta u^2+A(\delta)[\max\{u; 0\}-\delta u^2+\alpha u^2]=$$
$$=\begin{cases} A(\delta)u+u^2(\alpha-\delta-\delta A(\delta)), & u \geqslant 0; \\ u^2(\alpha-\delta-\delta A(\delta)), & u<0. \end{cases}$$

Если $\alpha-\delta-\delta A(\delta) \geqslant 0$, то $t_{\delta *}=\inf\limits_{E^1} t_\delta(u)=0$; если $\alpha-\delta-\delta A(\delta)<0$, то $t_{\delta *}=-\infty$. Как видим, нарушение условий (11) может привести к тому, что метод (8) потеряет смысл. Пусть $\alpha-\delta-\delta A(\delta) \geqslant 0$. Тогда условию (8) $t_\delta(u_\delta) \leqslant \varepsilon(\delta)$ удовлетворяет точка

$$u_\delta=\frac{2\frac{\varepsilon(\delta)}{\alpha(\delta)}}{\sqrt{\frac{A(\delta)}{\alpha(\delta)}+4\frac{\varepsilon(\delta)}{\alpha(\delta)}\left(1-\frac{\delta+\delta A(\delta)}{\alpha(\delta)}\right)}+\frac{A(\delta)}{\alpha(\delta)}}, \quad \delta>0.$$

Отсюда видно, что для справедливости равенства $\lim\limits_{\delta \to 0} \rho(u_\delta, U_*)=0$ необходимо выполнение соотношений $\sup\limits_{\delta>0} \frac{\varepsilon(\delta)}{\alpha(\delta)}<\infty$, $A(\delta)>0$, $\alpha(\delta)>0$, $\lim\limits_{\delta \to 0} A(\delta)=+\infty$, $\lim\limits_{\delta \to 0} \alpha(\delta)=0$ из (12), (13), причем первое из условий (13) строгого неравенства заменено нестрогим, вытекающим из (11). Заметим, что здесь $\lim\limits_{\delta \to 0} \rho(u_\delta, U_{**})=0$, $U_{**}=\{u_*=0\}$ и без требования условий (16).

6. Для удобства дальнейших ссылок в §§ 8–12 отдельно сформулируем и докажем специальный случай теоремы 2, когда семейство $\{u_\delta, \delta>0\}$ определяется из условий (8) при $\varepsilon(\delta)=0$, а в (6) используются точные входные данные.

Теорема 3. *Пусть*

1) *U_0 — выпуклое замкнутое множество из рефлексивного банахова пространства B, функции $J(u)$, $g_i(u)$, $i=1, \ldots, m$, $|g_i(u)|$, $i=m+1, \ldots, s$, выпуклы и полунепрерывны снизу на U_0, множество* (2) *непусто, выполняются условия* (3), (9);

2) *функция $\Omega(u)$ полунепрерывна снизу, строго равномерно выпукла на U_0, $P(u)$ — штрафная функция, определяемая формулой* (4) *с параметром $p \geqslant \max\{\nu; 1\}$;*

3) *параметры $\alpha(\delta)$, $A(\delta)$ удовлетворяют условиям*

$$\alpha(\delta)>0, \quad A(\delta)>0, \quad \forall \delta>0,$$
$$\lim_{\delta \to 0} \alpha(\delta)=0, \quad \lim_{\delta \to 0} A(\delta)=+\infty, \quad \lim_{\delta \to 0} \alpha(\delta)(A(\delta))^{\frac{\nu}{p-\nu}}=\infty \tag{51}$$

(при $p=\nu \geqslant 1$ последнее условие не нужно);

4) *точка u_δ является решением задачи*

$$T_\delta(u)=J(u)+A(\delta)P(u)+\alpha(\delta)\Omega(u) \to \inf, \quad u \in U_0, \tag{52}$$

т. е.

$$u_\delta \in U_0, \quad T_\delta(u_\delta)=\inf_{u \in U_0} T_\delta(u), \quad \delta>0. \tag{53}$$

Тогда

$$\lim_{\delta \to 0} J(u_\delta)=J_*, \quad \lim_{\delta \to 0} g_i^+(u_\delta)=0, \quad i=1, \ldots, s;$$
$$\lim_{\delta \to 0} \Omega(u_\delta)=\Omega_*, \quad \lim_{\delta \to 0} \|u_\delta-u_*\|=0, \tag{54}$$

где

$$\Omega_*=\inf_{u \in U_*} \Omega(u)=\Omega(u_*), \quad u_* \in U_*. \tag{55}$$

Доказательство. При сделанных предположениях функция $T_\delta(u)$ строго равномерно выпукла на выпуклом замкнутом множестве U_0, полунепрерывна снизу на U_0 и достигает своей нижней грани на U_0 в единственной точке (теорема 8.2.11), так что задача (52) разрешима и точка u_δ, удовлетворяющая условиям (53), определяется однозначно. Так как множество U_* выпукло, замкнуто, то по той же теореме 8.2.11 точка u_*, удовлетворяющая условиям (55), существует и единственна. Далее, с учетом (10), (53), соотношений $\Omega(u) \geqslant 0$, $P(u) \geqslant 0$ и $P(u_*)=0$ получаем цепочку неравенств

$$J(u_\delta)+A(\delta)P(u_\delta) \leqslant J(u_\delta)+A(\delta)P(u_\delta)+\alpha(\delta)\Omega(u_\delta)=T_\delta(u_\delta) \leqslant$$
$$\leqslant T_\delta(v_*)=J(v_*)+A(\delta)P(v_*)+\alpha(\delta)\Omega(v_*)=J_*+\alpha(\delta)\Omega_* \leqslant$$
$$\leqslant J(u_\delta)+A(\delta)P(u_\delta)+\alpha(\delta)\Omega_*+B(A(\delta))^{-\frac{\nu}{p-\nu}}, \quad \delta>0,$$

аналогичную цепочке (19). Отсюда имеем

$$\Omega(u_\delta) \leqslant \Omega_*+B/\left(\alpha(\delta)(A(\delta))^{\frac{\nu}{p-\nu}}\right), \quad \forall \delta>0 \tag{56}$$
$$J(u_\delta)+A(\delta)P(u_\delta) \leqslant J_*+\alpha(\delta)\Omega_*, \quad \forall \delta>0. \tag{57}$$

Из неравенства (57) и леммы 3.3 получаем оценки

$$0 \leqslant g_i^+(u_\delta) \leqslant \rho(\delta), \quad -|c|(\rho(\delta))^\nu \leqslant J(u_\delta)-J_* \leqslant \beta(\delta), \quad \forall \delta>0, \tag{58}$$

где величина $\rho(\delta)$, определяемая формулой (3.21), $\beta(\delta)=\alpha(\delta)\Omega_*$. Из (56), (57) и из леммы 3.2 следует утверждение теоремы 3. □

Переформулируем теорему 3 для случая, когда в (1), (2) $m=s=0$, $U=U_0$.

Теорема 4. *Пусть U — выпуклое замкнутое множество из рефлексивного банахова пространства B, функция $J(u)$ выпукла и полунепрерывна снизу на U; $J_* > -\infty$, $U_* \neq \varnothing$, $\Omega(u) = \|u\|^2$, $\alpha(\delta) > 0$, $\lim\limits_{\delta \to 0} \alpha(\delta) = 0$, точка u_δ определена из условий*

$$u_\delta \in U, \quad T_\delta(u_\delta) = \inf_U T_\delta(u), \quad T_\delta(u) = J(u) + \alpha(\delta)\|u\|^2 \quad \forall \delta > 0.$$

Тогда

$$J_* \leqslant J(u_\delta) \leqslant J_* + \alpha(\delta)\|u_*\|^2, \quad \|u_\delta\| \leqslant \|u_*\| = \inf_{U_*} \|u\|, \quad u_* \in U_*, \quad \lim_{\delta \to 0} \|u_\delta - u_*\| = 0.$$

Упражнения

1. Нарисовать график функции Тихонова $t_\delta(u) = u^2(1+u^4)^{-1} + \alpha(\delta)u^2$ для минимизации функции $J(u) = u^2(1+u^4)^{-1}$ на множестве $U = E^1$. Показать на этом графике, что для сходимости точек u_δ из (8) ко множеству $U_* = \{0\}$ необходимо согласованное стремление к нулю параметров $\alpha(\delta)$, $\varepsilon(\delta)$. Указать условие согласования этих параметров.

2. Применить метод стабилизации к задачам из примеров 1.1–1.9, указать условия согласования параметров метода.

3. Применить метод стабилизации (8) к общей задаче линейного программирования.

4. Применить метод стабилизации к задаче оптимального управления:

$$J(u) = |x(T; u) - b|^2_{E^n} \to \inf,$$
$$\dot{x}(t) = D(t)x(t) + B(t)u(t) + f(t), \quad t_0 \leqslant t \leqslant T; \quad x(t_0) = x_0,$$

$u = u(t) \in U$ — выпуклое замкнутое множество из $L_2^r[t_0, T]$, считая, что матрицы $D(t)$, $B(t)$, вектора $f(t)$, b, x_0 известны неточно (обозначения см. в примере 8.2.15) и выполнены условия (45), $|x_{0\delta} - x_0| \leqslant \delta$, $|f_\delta(t) - f(t)| \leqslant \delta$ $\forall t \in [t_0, T]$. Указание: воспользоваться стабилизатором $\Omega(u) = \|u\|^2_{L_2^r[t_0, T]}$.

5. В упражнении (4) функцию $J(u)$ заменить на $J(u) = \int\limits_a^b |x(t) - y(t)|^2 dt$, считая, что $|y_\delta(t) - y(t)| \leqslant \delta$.

6. В задачах из упражнений 4, 5 использовать стабилизатор $\Omega(u) = \|u\|^2_{H_r^1[t_0, T]}$, считая, что $U_* \cap H_r^1[t_0, T] \neq \varnothing$.

7. Показать, что задача $J(u) = \int\limits_a^b \left(\int\limits_a^t u(s)ds - f(t) \right)^2 dt \to \inf$, $u \in U = L_2[a, b]$, неустойчива к возмущениям функции $f(t) \in C^1[a, b]$ в метрике $L_2[a, b]$ или $C[a, b]$ (задача дифференцирования функции $f(t)$). Указание: взять $f_\delta(t) = f(t) + \delta \sin \frac{t}{\delta}$, $t \in [a, b]$. Применить к этой задаче метод стабилизации, взяв стабилизатор $\Omega(u) = \|u\|^2_{L_2[a, b]}$.

8. Сформулировать и доказать аналоги теорем 1, 2 для задачи: $J(u) \to \inf$, $u \in U$, когда множество U известно точно, а приближения $J_\delta(u)$ функции $J(u)$ удовлетворяют первому из неравенств (5) (см. замечание 3).

9. В задаче (32) предположить, что вместо A, b известны их приближения A_h, b_δ такие, что $\|A_h - A\| \leqslant h$, $\|b_\delta - b\| \leqslant \delta$, множество U известно точно. Описать метод стабилизации со стабилизатором $\Omega(u) = \|u\|^2$. Указать условия согласования параметров $\alpha = \alpha(h, \delta)$, $\varepsilon = \varepsilon(h, \delta)$, обеспечивающие сходимость метода к Ω-нормальному решению задачи в метрике H. Указание: воспользоваться схемой доказательства теорем 1, 2.

10. Исследовать сходимость при $\alpha \to +0$ метода стабилизации

$$\langle c, u \rangle + \frac{1}{2}\alpha|u|^2 \to \inf, \quad u \in U = \{u \geqslant 0: Au = b\} \tag{59}$$

для канонической задачи линейного программирования

$$\langle c, u \rangle \to \inf, \quad u \in U, \tag{60}$$

где A — матрица размера $m \times n$, $b \in E^m$, $c \in E^n$. Показать, что двойственная к (59) задача (§ 4.9) имеет вид

$$-\langle b, \lambda \rangle - \frac{1}{2\alpha}|\max\{-(c + A^\top \lambda); 0\}|^2 \to \sup, \quad \lambda \in E^m. \tag{61}$$

Убедиться, что задача (61) при $\alpha \to +0$ равносильна методу штрафных функций для двойственной к (60) задачи линейного программирования: $-\langle b, \lambda \rangle \to \sup$, $\lambda \in \Lambda = \{\lambda \in E^m: A^\top \lambda + c \geqslant 0\}$.

Указание: показать, что $\inf\limits_{x \geqslant 0} \left(\frac{1}{2}\alpha x^2 + \beta x \right) = -\frac{1}{2\alpha}|\max\{-\beta; 0\}|^2$ $\forall \alpha > 0$, $\forall \beta$.

§ 5. Метод невязки

1. Этот метод опишем применительно к уже рассмотренной задаче минимизации (4.1), (4.2) второго типа при тех же предположениях (4.3)–(4.5). Введем множество

$$V(\delta) = \{u \in U_\Omega: J_\delta(u) + A(\delta)P_\delta(u) \leqslant \overline{\varphi}_{\delta *} + \sigma(\delta)\}, \tag{1}$$

где $J_\delta(u)$, $P_\delta(u)$ — какие-либо конкретные реализации приближений $J(u)$, $P(u)$ из (4.5), $\Omega(u)$, $u \in U_\Omega$ — стабилизатор задачи (4.1), (4.2), $A(\delta) > 0$, $\sigma(\delta) > 0$ — параметры метода, $\lim\limits_{\delta \to 0} A(\delta) = +\infty$, $\lim\limits_{\delta \to 0} \sigma(\delta) = 0$, $\overline{\varphi}_{\delta *} = \inf\limits_{u \in U_\Omega} (J_\delta(u) + A(\delta)P_\delta(u) + (\delta + \delta A(\delta))\Omega(u))$. Предположим, что $V(\delta) \neq \varnothing$ и, приближенно решая задачу первого типа

$$\Omega(u) \to \inf, \quad u \in V(\delta),$$

нам удалось найти точку u_δ, удовлетворяющую условиям:

$$u_\delta \in V(\delta), \quad \Omega(u_\delta) \leqslant \Omega_{\delta *} + \mu(\delta), \tag{2}$$

где $\Omega_{\delta *} = \inf\limits_{u \in V(\delta)} \Omega(u)$, $\mu(\delta) > 0$. На формальном уровне метод невязки описан. Осталось указать условия, гарантирующие непустоту множества $V(\delta)$, условия согласования параметров метода, условия сходимости метода.

Но сначала кратко поясним название метода невязки. Этот метод первоначально применялся для минимизации функции $J(u) = \rho_F(Au, f)$, возникающей при исследовании уравнений $Au = f$, где A — оператор, действующий из некоторого метрического пространства $\mathcal{M}$ в метрическое пространство F, $f \in F$, $\rho_F(f_1, f_2)$ — расстояние между точками $f_1, f_2 \in F$. Величину $J(u) = \rho_F(Au, f)$ принято называть *невязкой* уравнения $Au = f$. Если это уравнение имеет решение, то $\inf\limits_{u \in \mathcal{M}} J(u) = J_* = 0$, и в этом случае метод невязки [334; 508; 697] сводится к задаче минимизации стабилизатора $\Omega(u)$ на множестве $V(\delta) = \{u \in U_\Omega: J(u) \leqslant \sigma(\delta)\}$, состоящем из точек, для которых невязки уравнения не превосходят некоторой малой величины $\sigma(\delta)$. В задаче минимизации: $J(u) \to \inf$, $u \in U$ невязкой можно назвать величину $J(u) - J_*$ и в качестве множества, на котором невязка мала, взять $V = \{u \in U: J(u) - J_* \leqslant \sigma\}$; в задаче (4.1), (4.2) при использовании штрафной функции $V = \{u \in U_0: \Phi(u) = J(u) + AP(u) \leqslant \inf\limits_{u \in U_0} \Phi(u) + \sigma\}$ и при неточном задании исходных данных аналогом множества V является множество (1). Таким образом, метод невязки заключается в минимизации стабилизатора $\Omega(u)$ на множестве точек из U_Ω, для которых невязка мала. Это обстоятельство нашло отражение в названии метода невязки.

Метод невязки (1), (2) удобно применять, когда величина $\overline{\varphi}_{\delta *}+\sigma(\delta)$, входящая в определение множества $V(\delta)$, известна. Если эта величина заранее неизвестна, то для ее определения предварительно нужно, решая задачу первого типа

$$\overline{\varphi}_\delta(u)=J_\delta(u)+A(\delta)P_\delta(u)+(\delta+\delta A(\delta))\Omega(u)\to\inf,\quad u\in U_\Omega, \tag{3}$$

найти величину $\overline{\varphi}_{\delta *}$ или ее оценку $\overline{\varphi}_{\delta *}+\sigma(\delta)$. При решении задач (3) и (1), (2) возможно использование методов из § 2. Дальнейшее изложение не зависит от метода решения перечисленных задач.

Теорема 1. *Пусть выполнены условия* 1)–3) *теоремы* 4.1 *и параметры* $A(\delta)$, $\sigma(\delta)$, $\mu(\delta)$ *таковы, что*

$$\begin{gathered}A(\delta)>0,\quad \sigma(\delta)>0,\quad \mu(\delta)>0,\\ \lim_{\delta\to 0}(\sigma(\delta)+\delta A(\delta))=0,\quad \lim_{\delta\to 0}A(\delta)=+\infty,\end{gathered} \tag{4}$$

$$\sup_{\delta>0}\mu(\delta)<\infty,\quad (\delta+\delta A(\delta))(3+\Omega_*)+B(A(\delta))^{-\frac{\nu}{p-\nu}}\leqslant\sigma(\delta),\quad \delta>0 \tag{5}$$

(*при* $p=\nu$ *последнее слагаемое в* (5) *отсутствует, а условие* $\lim_{\delta\to 0}A(\delta)=+\infty$ *можно заменить на* $A(\delta)>\max_{1\leqslant i\leqslant s}c_i$), *где* $B=(p-\nu)\nu^{\frac{\nu}{p-\nu}}p^{-\frac{p}{p-\nu}}|c|^{\frac{p}{p-\nu}}$, $|c|=\left(\sum_{i=1}^{s}|c_i|^{\frac{p}{p-\nu}}\right)^{\frac{p-\nu}{p}}$ *при* $p>\nu$, $\Omega_*=\inf_{U_\Omega}\Omega(u)$, $U_{\Omega *}=U_*\cap U_\Omega$. *Тогда множество* $V(\delta)$, *определенное равенством* (1), *непусто при всех* $\delta>0$, *семейство* $\{u_\delta,\ \delta>0\}$ *из* (2) *таково, что*

$$\lim_{\delta\to 0}J(u_\delta)=J_*,\quad \lim_{\delta\to 0}g_i^+(u_\delta)=0,\quad i=1,\ldots,s;\quad \lim_{\delta\to 0}\rho(u_\delta,U_*)=0. \tag{6}$$

Если кроме перечисленных условий еще известно, что $U_\Omega=U_0$, *функция* $\Omega(u)$ *— полунепрерывна снизу на* U_0, $\lim_{\delta\to 0}\mu(\delta)=0$, *то наряду с* (6) *справедливы равенства*

$$\lim_{\delta\to 0}\Omega(u_\delta)=\Omega_*,\quad \lim_{\delta\to 0}\rho(u_\delta,U_{**})=0, \tag{7}$$

где $U_{**}=\{u\in U_*\colon \Omega(u)=\Omega_*=\inf_{U_*}\Omega(u)\}$ *— множество* Ω*-нормальных решений задачи* (4.1), (4.2). *Пределы в* (6), (7) *равномерны относительно выбора* $J_\delta(u)$, $P_\delta(u)$ *из* (4.5) *и выбора точки* u_δ *из* (2), *т. е. справедливы равенства* (4.15), (4.18), *где* $X(\delta)$ *— объединение множеств* $x(\delta)=\{u\in V(\delta)\colon \Omega(u_\delta)\leqslant\Omega_{\delta *}+\mu(\delta)\}$ *по всевозможным реализациям* $J_\delta(u)$, $P_\delta(u)$ *из* (4.5).

Доказательство. Сначала покажем, что множество (1) непусто при всех $\delta>0$. Воспользуемся неравенством (4.10):

$$J_*\leqslant J(u)+A(\delta)P(u)+B(A(\delta))^{-\frac{\nu}{p-\nu}}\quad\forall u\in U_0,\quad \delta>0,\quad p\geqslant\nu \tag{8}$$

(при $p=\nu$ последнее слагаемое в (8) отсутствует и $A(\delta)\geqslant|c|=\max_{1\leqslant i\leqslant s}|c_i|$). Из (3), (8), (4.5) имеем

$$\begin{aligned}\overline{\varphi}_\delta(u)&\geqslant J(u)+A(\delta)P(u)-(\delta+\delta A(\delta))(1+\Omega(u))+\\ &+(\delta+\delta A(\delta))\Omega(u)\geqslant J_*-B(A(\delta))^{-\frac{\nu}{p-\nu}}-(\delta+\delta A(\delta))\quad\forall u\in U_\Omega,\end{aligned}$$

так что $\overline{\varphi}_{\delta *}>-\infty$. Далее, по определению Ω_* для любого ε, $0<\varepsilon<1$, найдется точка w_ε такая, что

$$w_\varepsilon\in U_{\Omega *},\quad \Omega(w_\varepsilon)\leqslant\Omega_*+\varepsilon. \tag{9}$$

Тогда с учетом (4.5), (8) получим

$$\begin{aligned}J_\delta(w_\varepsilon)+A(\delta)P_\delta(w_\varepsilon)&\leqslant J(w_\varepsilon)+A(\delta)P(w_\varepsilon)+\\ &+(\delta+\delta A(\delta))(1+\Omega(w_\varepsilon))\leqslant J_*+(\delta+\delta A(\delta))(1+\Omega_*+\varepsilon)\leqslant\\ &\leqslant J(u)+A(\delta)P(u)+B(A(\delta))^{-\frac{\nu}{p-\nu}}+(\delta+\delta A(\delta))(2+\Omega_*)\leqslant\\ &\leqslant J_\delta(u)+A(\delta)P_\delta(u)+(\delta+\delta A(\delta))(1+\Omega(u))+\\ &+B(A(\delta))^{-\frac{\nu}{p-\nu}}+(\delta+\delta A(\delta))(2+\Omega_*)=\\ &=\overline{\varphi}_\delta(u)+(\delta+\delta A(\delta))(3+\Omega_*)+B(A(\delta))^{-\frac{\nu}{p-\nu}}\quad\forall u\in U_\Omega.\end{aligned}$$

Переходя к нижней грани по $u\in U_\Omega$, отсюда с учетом второго неравенства (5) имеем: $J_\delta(w_\varepsilon)+A(\delta)P_\delta(w_\varepsilon)\leqslant\overline{\varphi}_{\delta *}+\sigma(\delta)$, т. е. $w_\varepsilon\in V(\delta)$ $\forall\delta>0$. Таким образом, $V(\delta)\neq\varnothing$ $\forall\delta>0$. Так как $\Omega(u)\geqslant 0$ $\forall u\in U_\Omega$, то $\Omega_{\delta *}=\inf_{u\in V(\delta)}\Omega(u)\geqslant 0$ и точка u_δ, удовлетворяющая условиям (2), существует по определению нижней грани. Для любой такой точки u_δ с учетом включения $w_\varepsilon\in V(\delta)$ имеем: $\Omega(u_\delta)\leqslant\Omega(w_\varepsilon)+\mu(\delta)\leqslant\Omega_*+\varepsilon+\mu(\delta)$ $\forall\varepsilon$, $0<\varepsilon<1$. Переходя к пределу при $\varepsilon\to+0$ отсюда получим

$$\Omega(u_\delta)\leqslant\Omega_*+\gamma(\delta),\quad \gamma(\delta)=\mu(\delta)\quad\forall\delta>0. \tag{10}$$

Кроме того, из (4.5), (1), (2), (9), (10) следует

$$\begin{aligned}J(u_\delta)+A(\delta)P(u_\delta)&\leqslant J_\delta(u_\delta)+A(\delta)P_\delta(u_\delta)+\\ &+(\delta+\delta A(\delta))(1+\Omega(u_\delta))\leqslant\overline{\varphi}_{\delta *}+\sigma(\delta)+(\delta+\delta A(\delta))(1+\Omega_*+\mu(\delta))\leqslant\\ &\leqslant\overline{\varphi}_\delta(w_\varepsilon)+\sigma(\delta)+(\delta+\delta A(\delta))(1+\Omega_*+\mu(\delta))\leqslant\\ &\leqslant J(w_\varepsilon)+A(\delta)P(w_\varepsilon)+(\delta+\delta A(\delta))(1+\Omega(w_\varepsilon))+\\ &+(\delta+\delta A(\delta))\Omega(w_\varepsilon)+\sigma(\delta)+(\delta+\delta A(\delta))(1+\Omega_*+\mu(\delta))\leqslant\\ &\leqslant J_*+(\delta+\delta A(\delta))(2+3\Omega_*+2\varepsilon+\mu(\delta))+\sigma(\delta)\quad\forall\varepsilon,\quad 0<\varepsilon<1.\end{aligned}$$

Отсюда при $\varepsilon\to 0$ имеем:

$$\begin{gathered}J(u_\delta)+A(\delta)P(u_\delta)\leqslant J_*+\beta(\delta),\\ \beta(\delta)=(\delta+\delta A(\delta))(2+3\Omega_*+\mu(\delta))+\sigma(\delta),\quad\forall\delta>0.\end{gathered} \tag{11}$$

Отсюда и из леммы 3.3 следуют оценки

$$0\leqslant g_i^+(u_\delta)\leqslant\rho(\delta),\quad i=1,\ldots,s;\quad -|c|(\rho(\delta))^\nu\leqslant J(u_\delta)-J_*\leqslant\beta(\delta),\quad\forall\delta>0, \tag{12}$$

где величина $\beta(\delta)$ взята из (11), а $\rho(\delta)$ определяется формулой (3.21). Из оценок (10)–(12) и леммы 3.1 получаем равенства (6), (7). Равенства (4.15), (4.18) здесь доказываются также, как в теореме 4.1. □

Замечание 1. Вместо второго из неравенств (5) на практике удобнее требовать:

$$\lim_{\delta\to 0}\frac{\delta+\delta A(\delta)}{\sigma(\delta)}=0,\quad \lim_{\delta\to 0}\sigma(\delta)(A(\delta))^{\frac{\nu}{p-\nu}}=\infty \tag{13}$$

(при $p=\nu$ последнее равенство не нужно). В качестве параметров $A(\delta)$, $\sigma(\delta)$, $\mu(\delta)$, удовлетворяющих условиям (4), (5), (13), можно, например, взять

$$A(\delta)=a_1\delta^{-A}, \quad \sigma(\delta)=a_2\delta^{\sigma}, \quad \mu(\delta)=a_3\delta^{\mu}, \quad \forall\delta>0 \tag{14}$$

где A, σ, μ, a_1, a_2, a_3 — положительные постоянные.

Замечание 4.2 сохраняет силу и для метода невязки.

2. В выпуклых задачах (4.1), (4.2) в банаховых пространствах в методе (1), (2) возможно использование слабого стабилизатора.

Теорема 2. *Пусть выполнены условия* 1)–3) *теоремы* 4.2 *и параметры* $A(\delta)$, $\sigma(\delta)$, $\mu(\delta)$ *удовлетворяют условиям* (4), (13), $\lim\limits_{\delta\to 0}\mu(\delta)=0$. *Тогда семейство* $\{u_\delta,\ \delta>0\}$, *определенное методом* (1), (2) *таково, что*

$$\lim_{\delta\to 0} J(u_\delta)=J_*, \quad \lim_{\delta\to 0} g_i^+(u_\delta)=0, \quad i=1,\ldots,s;$$
$$\lim_{\delta\to 0}\Omega(u_\delta)=\Omega_*, \quad \lim_{\delta\to 0}\|u_\delta-v_*\|_B=0, \tag{15}$$

где v_* — *точка минимума функции* $\Omega(u)$ *на множестве* U_*. *Пределы в* (15) *равномерны относительно выбора* $J_\delta(u)$, $P_\delta(u)$ *из* (4.5) *и точки* u_δ *из* (2).

Доказательство. Повторяя доказательство теоремы 1 при $U_\Omega=U_0$, убеждаемся, что оценки (10)–(12) сохраняют силу при всех достаточно малых $\delta>0$. Отсюда и из леммы 3.2 следуют существование и единственность точки v_* и равенства (15). Равномерность пределов (15) доказывается также, как равенства (4.15), (4.18). □

Замечания 4.3–4.7 с очевидными изменениями сохраняют силу и для метода (1), (2).

Упражнения

1. Описать метод невязки для задач из примеров 1.1–1.9, 4.1–4.5.

2. Привести примеры задач, показывающих, что условия (4), (5), (13) на параметры метода невязки не являются грубыми на классе задач (4.1), (4.2).

3. Применить метод невязки к задачам из упражнений 4.4–4.7.

4. Сформулировать и доказать аналоги теорем 1, 2 для задачи: $J(u)\to\inf$, $u\in U$, когда множество U известно точно, а приближения $J_\delta(u)$ функции $J(u)$ удовлетворяют первому из неравенств (4.5).

5. К задаче (4.32) применить метод невязки в предположении, что $\|A_h-A\|\leqslant h$, $\|b_\delta-b\|\leqslant\delta$, множество U известно точно, $\Omega(u)=\|u\|^2$. Указать условия согласования параметров метода с погрешностями (h,δ).

§ 6. Метод квазирешений

1. Этот метод также опишем применительно к задаче минимизации (4.1), (4.2) второго типа при тех же предположениях (4.3)–(4.5). Будем также предполагать, что нам известно число r такое, что

$$U_{\Omega*}\cap Q_r\neq\varnothing, \quad Q_r=\{u\in U_\Omega\colon \Omega(u)\leqslant r\}, \tag{1}$$

где $\Omega(u)$ стабилизатор задачи (1), (2) в метрике $\mathcal{M}$; напоминаем, что множество $U_{\Omega*}=U_\Omega\cap U_*$ непусто по определению стабилизатора. Метод квазирешений сводится к приближенному решению задачи первого типа

$$\varphi_\delta(u)=J_\delta(u)+A(\delta)P_\delta(u)\to\inf, \quad u\in Q_r,$$

и определению точки u_δ из условий

$$u_\delta\in Q_r, \quad \varphi_\delta(u_\delta)\leqslant\varphi_{\delta*}+\xi(\delta), \tag{2}$$

где $\varphi_{\delta*}=\inf\limits_{Q_r}\varphi_\delta(u)$, $\xi(\delta)>0$ $\forall\delta>0$.

Название метода квазирешений также связано с уравнением $Au=f$, где A — оператор, действующий из некоторого метрического пространства $\mathcal{M}$ в метрическое пространство F, $f\in F$. Пусть $\rho_F(Au,f)$ — невязка этого уравнения. *Квазирешением* уравнения $Au=f$ называют точку $u_*\in\mathcal{M}$, для которого $\rho_F(Au_*,f)=\inf\limits_{u\in\mathcal{M}}\rho_F(Au,f)$. Если $\rho_F(Au_*,f)=0$, то квазирешение u_* превращается в обычное решение уравнения $Au=f$. Однако квазирешение может существовать и тогда, когда это уравнение не имеет решения. Предположим, что известно компактное в $\mathcal{M}$ множество $Q\subseteq\mathcal{M}$, содержащее хотя бы одно квазирешение уравнения $Au=f$. Тогда метод квазирешения [334; 695] сводится к задаче минимизации: $\rho(Au,f)\to\inf$, $u\in Q$. Обобщением этого метода применительно к задаче (4.1), (4.2) является метод (2), в котором роль Q играет множество Q_r, роль невязки — функция $\varphi_\delta(u)$.

Теорема 1. *Пусть выполнены условия* 1)–3) *теоремы* 4.1, *условие* (1) *при некотором* r *и параметры* $A(\delta)$, $\xi(\delta)$ *метода* (2) *таковы, что*

$$A(\delta)>0, \quad \xi(\delta)>0, \quad \lim_{\delta\to 0}A(\delta)=+\infty, \quad \lim_{\delta\to 0}\xi(\delta)=0 \tag{3}$$

(*при* $p=\nu$ *условие* $\lim\limits_{\delta\to 0}A(\delta)=+\infty$ *можно заменить на* $A(\delta)>\max\limits_{1\leqslant i\leqslant s}c_i$).

Тогда точка u_δ, *определяемая условиями* (2), *существует при всех* $\delta>0$ *и*

$$\lim_{\delta\to 0}J(u_\delta)=J_*; \quad \lim_{\delta\to 0}g_i^+(u_\delta)=0, \quad i=1,\ldots,s; \quad \lim_{\delta\to 0}\rho(u_\delta,U_*)=0. \tag{4}$$

Пределы в (4) *равномерны относительно выбора* $J_\delta(u)$, $P_\delta(u)$ *из* (4.5) *и выбора точек* u_δ *из* (2), *т. е. справедливы равенства* (4.15), *где* $X(\delta)$ — *объединение множеств* $x(\delta)=\{u\in Q_r\colon \varphi_\delta(u)\leqslant\varphi_{\delta*}+\xi(\delta)\}$ *по всевозможным* $J_\delta(u)$, $P_\delta(u)$ *из* (4.5).

Доказательство. Прежде всего из (4.5), неравенства (4.10) и определения множества Q_r в (1) имеем:

$$\varphi_\delta(u)=J_\delta(u)+A(\delta)P_\delta(u)\geqslant J(u)+A(\delta)P(u)-(\delta+\delta A(\delta))(1+\Omega(u))\geqslant$$
$$\geqslant J_*-B(A(\delta))^{-\frac{\nu}{p-\nu}}-(\delta+\delta A(\delta))(1+r) \quad \forall u\in Q_r.$$

Следовательно, $\varphi_{\delta*}=\inf\limits_{Q_r}\varphi_\delta(u)>-\infty$ и точка u_δ из условий (2) существует по определению нижней грани. Возьмем произвольные точки u_δ из (2), $u_*\in$

$\in U_{\Omega_*} \cap Q_r$. С учетом (4.5), (2), неравенств $\Omega(u_\delta) \leqslant r$, $\Omega(u_*) \leqslant r$ получаем

$$J(u_\delta) + A(\delta)P(u_\delta) \leqslant \varphi_\delta(u_\delta) + (\delta + \delta A(\delta))(1 + \Omega(u_\delta)) \leqslant$$
$$\leqslant \varphi_{\delta *} + \xi(\delta) + (\delta + \delta A(\delta))(1 + r) \leqslant \varphi_\delta(u_*) + \xi(\delta) + (\delta + \delta A(\delta))(1 + r) \leqslant$$
$$\leqslant J(u_*) + A(\delta)P(u_*) + \xi(\delta) + 2(\delta + \delta A(\delta))(1 + r) = J_* + \beta(\delta),$$
$$\beta(\delta) = \xi(\delta) + 2(\delta + \delta A(\delta))(1 + r) \quad \forall \delta > 0. \tag{5}$$

Из (5) и леммы 3.3 следуют оценки

$$0 \leqslant g_i^+(u_\delta) \leqslant \rho(\delta), \quad i = 1, \ldots, s, \quad -|c|(\rho(\delta))^\nu \leqslant J(u_\delta) - J_* \leqslant \beta(\delta) \quad \forall \delta > 0, \tag{6}$$

где величина $\rho(\delta)$ определяется формулой (3.21), а $\beta(\delta)$ взята из (5). Неравенство $\Omega(u_\delta) \leqslant r$ можем записать в виде (3.10):

$$\Omega(u_\delta) \leqslant \Omega_* + \gamma(\delta), \quad \gamma(\delta) \equiv r - \Omega_*, \quad \delta > 0. \tag{7}$$

Из оценок (5)–(7) и леммы 3.1 получаем равенства (4). Равенства (4.15) доказываются также, как в теореме 4.1. □

Для того, чтобы семейство точек $\{u_\delta, \delta > 0\}$ из (2) при $\delta \to 0$ сходилось ко множеству Ω-нормальных решений задачи (4.1), (4.5), нужно, чтобы в неравенстве (3.10) величина $\gamma(\delta)$ обладала свойством $\lim\limits_{\delta \to 0} \gamma(\delta) = 0$. Однако, как видно из (7), описанный вариант метода квазирешений таким свойством не обладает. Существует более тонкий и конструктивно более сложный вариант метода квазирешений [113; 151; 165], в котором неравенство (3.10) выполняется с $\lim\limits_{\delta \to 0} \gamma(\delta) = 0$ и для которого имеют место равенства (4.18), а также справедлив аналог теоремы 4.2.

2. Сделаем несколько общих замечаний по поводу изложенных выше трех методов регуляризации.

Замечание 1. Каждый из этих методов (как и методы § 2, см. замечание 2.2) для своей реализации требует наличия некоторой априорной информации. Прежде всего во всех трех методах предполагается, что в задаче (4.1), (4.2) нам уже известно, что $U \neq \varnothing$, $J_* = \inf\limits_{u \in U} J(u) > -\infty$, $U_* = \{u \in U: J(u) = J_*\} \neq \varnothing$ и, кроме того, подразумевается, что существует стабилизатор в нужной метрике и выполнены условия (4.5). На практике перечисленная информация обычно добывается на стадии формирования математической модели задачи, предварительного исследования ее свойств, и, как правило, эта работа требует немалых усилий. Часть из этих проблем, относящихся к условиям $J_* > -\infty$, $U_* \neq \varnothing$, мы немного обсуждали в § 8.2. При выборе стабилизатора $\Omega(u)$ наиболее труднопроверяемым является условие принадлежности хотя бы одного решения u_* задачи (4.1), (4.2) области его определения U_Ω. Включение $u_* \in U_\Omega$ обычно означает, что это решение u_* обладает некоторыми дополнительными свойствами «гладкости» (непрерывность, гельдеровость, наличие производных и т. п.), которые в исходном определении решения задачи, возможно, и не предусматривались, и выяснение таких свойств не всегда простое дело.

В задаче (4.1), (4.2), когда множество U известно неточно, в §§ 4–6 мы еще предполагали, что задача имеет сильно согласованную постановку в смысле выполнения неравенства (4.9), что позволило нам правильно

выбрать параметры методов и единообразно исследовать их сходимость. По-видимому, условие (4.9), характеризующее условие роста целевой функции и ограничений на множестве U_0, может быть заменено другой более удобной априорной информацией о задаче.

Наличие перечисленной априорной информации дало нам возможность реализовать метод стабилизации. Для метода невязки к этой информации понадобилось добавить еще знание оценки J_* (или, точнее, оценки величины $\overline{\varphi}_{\delta *}$ из (5.1), а в методе квазирешений — оценки значения стабилизатора на каком-либо элементе $u_* \in U_{\Omega *}$ (см. условие (6.1)). Предупреждаем читателя, что вспомогательные задачи (4.8), (5.2), (6.2), лежащие в основе описанных методов регуляризации, могут иметь смысл и в том случае, когда какая-то часть упомянутой априорной информации отсутствует. Однако тогда решения u_δ этих вспомогательных задач могут и не иметь какого-либо отношения к исходной задаче (4.1), (4.2), что видно из следующего примера.

Пример 1. Задача: $J(u) = u \to \inf$, $u \in U = \{u \in E^1: u \geqslant 0, g_1(u) = u - 1 \leqslant 0, g_2(u) = -u + 2 \leqslant 0\}$. Очевидно, $U = \varnothing$, и задача не имеет решения. Тем не менее, если здесь мы формально применим метод стабилизации (4.8) со стабилизатором $\Omega(u) = u^2$, то придем к задаче: $t_\delta(u) = u + A(\delta)(\max\{g_1(u); 0\} + \max\{g_2(u); 0\}) + \alpha(\delta)u^2 \to \inf$, $u \geqslant 0$, которая при любых $\alpha(\delta) > 0$ будет иметь единственное решение u_δ. Однако в этой задаче $U_* = \varnothing$, и мы не можем утверждать, что $\rho(u_\delta, U_*) \to 0$ при $\delta \to 0$.

Этот простой пример показывает, что перед применением того или иного метода регуляризации нужно предварительно убедиться, что необходимая для реализации метода информация уже имеется. О роли априорной информации, важности ее учета при построении более эффективных методов регуляризации см., например, в [184; 450; 509; 557; 693; 695; 697; 757; 817].

Замечание 2. Изложенные в §§ 4–6 методы регуляризации содержат параметры, которые в каждом методе удовлетворяют своим условиям согласования и прямо или косвенно учитывают имеющуюся информацию о задаче (4.1), (4.2) (см. условие (4.12), (4.13), (4.16) в методе стабилизации, (5.4), (5.5) — в методе невязки, (6.3) — в методе квазирешений). Эти параметры не зависят от того, какая реализация входных данных $J_\delta(u)$, $P_\delta(u)$ из (4.5) имеется в нашем распоряжении и могут быть выбраны заранее, до начала вычислений. Поэтому такой выбор параметров принято называть *априорным*. Условия согласования параметров оставляют достаточно большой произвол в их выборе (см., например, формулы (4.28), (5.14)). Следует сказать, что выбор параметров является тонким делом, от этого зависит скорость сходимости методов, их трудоемкость, точность. Для некоторых классов задач проблема оптимального априорного выбора параметров исследована, например, в [334; 462; 659; 679; 695].

Существует и другой, так называемый *апостериорный* способ выбора параметров методов регуляризации, в котором учитываются свойства конкретных приближений входных данных, имеющихся в нашем распоряжении, свойства точек u_δ, определяемых из вспомогательных задач и т. д. Апостериорный выбор несколько усложняет реализацию методов, но обладает немалыми достоинствами в смысле точности, лучшего учета конкретных особенностей задачи и ее входных данных. Не имея возможности подробнее останавливаться на проблеме апостериорного выбора параметров методов регуляризации, отсылаем читателя к специальной литературе [130; 334; 507; 679; 693; 696; 697] и др.

Замечание 3. Каждый из методов, описанных в §§ 4–6, определяет оператор R_δ, который каждому набору приближенных входных данных $(J_\delta(u), P_\delta(u))$ из (4.5) и набору своих параметров ставит в соответствие точку u_δ, определяемых из вспомогательных задач (4.8), (5.2), (6.2) минимизации первого типа и при выполнении условий теорем сходимости (теоремы 4.1, 4.2, 5.1, 5.2, 6.1) обеспечивает выполнение равенств (4). Это означает, что такой оператор R_δ является регуляризирующим.

Определение 1. Оператор R_δ, который каждому набору приближенных входных данных задачи (4.1), (4.2) второго типа ставит в соответствие точку $u_\delta \in U_0$, называется *регуляризирующим*, если $\lim\limits_{\delta\to 0} J(u_\delta) = J_*$, $\lim\limits_{\delta\to 0} \rho(u, U_*) = 0$, $\lim\limits_{\delta\to 0} g_i^+(u_\delta) = 0$, $i = 1, \ldots, s$ (ср. с определением 2.1).

Проблема существования регуляризирующих операторов, построение оптимальных в том или ином смысле операторов R_δ для различных классов неустойчивых задач исследована, например, в [62; 63; 188; 334; 501; 693; 695; 697] и др.

Замечание 4. Неравенства (4.24), (4.25), (5.12), (6.6) представляют собой оценки скорости сходимости изложенных методов регуляризации по функции. При дополнительных ограничениях на задачу (4.1), (4.2) удается получить оценки скорости сходимости этих методов по аргументу, т. е. оценки для $\rho(u_\delta, U_*)$, $\rho(u_\delta, U_{**})$ [162; 173]. Оценки скорости сходимости методов регуляризации для различных классов неустойчивых задач исследовались в [179; 192; 334; 507; 508; 679; 697; 740] и др.

Упражнения

1. Для задачи: $J(u) \to \inf$, $u \in U$ с точно заданным множеством U описать метод квазирешений, сформулировать и доказать аналог теоремы 1.

2. Применить метод квазирешений для задач из примеров 1.1–1.9, 4.1–4.5, из упражнений 4.4–4.7.

3. Пусть u_α — точка минимума функции $T_\alpha(u) = |Au - b|^2 + \alpha|u|^2$ при $u \in E^n$; здесь A — матрица размера $m \times n$, $b \in E^m$, $\alpha > 0$. Доказать, что тогда $|u_\alpha - u_*| \leqslant \alpha|v|$, где v — решение уравнения $(A^*A)^2 u = A^* b$ с минимальной нормой [192, стр. 301].

§ 7. Методы регуляризации с расширением множества

В методах регуляризации, описанных и исследованных в §§ 4–6, ограничения типа равенств и неравенств в задании множества (4.2) учитывались с помощью штрафных функций. Ниже излагаются методы регуляризации, в которых вместо штрафных функций используется некоторое расширение множества, согласованное со стабилизатором и с погрешностями в задании исходных данных.

Как и выше, будем рассматривать задачу

$$J(u) \to \inf, \quad u \in U, \tag{1}$$

$$U = \{u \in U_0:\ g_i(u) \leqslant 0,\ i = 1, \ldots, m;\ g_i(u) = 0,\ i = m+1, \ldots, s\}, \tag{2}$$

где U_0 — заданное множество из метрического пространства $\mathcal{M}$ с метрикой $\rho = \rho(u, v)$, функции $J(u)$, $g_i(u)$, $i = 1, \ldots, s$, определены и конечны

на U_0. Пусть $J_* > -\infty$, $U_* \neq \varnothing$. Пусть $\Omega(u)$, $u \in U_\Omega \subseteq U_0$ — какой-либо стабилизатор задачи (1), (2) в метрике $\mathcal{M}$. Как и выше, будем предполагать, что множество U_0 известно точно, а вместо функций $J(u)$, $g_i(u)$ известны их приближения $J_\delta(u)$, $g_{i\delta}(u)$ такие, что

$$|J_\delta(u) - J(u)| \leqslant \delta(1 + \Omega(u)),\ |g_{i\delta}(u) - g_i(u)| \leqslant \delta(1 + \Omega(u)),$$
$$i = 1, \ldots, s,\ u \in U_\Omega,\ \delta > 0. \tag{3}$$

Рассмотрим следующее множество

$$W(\delta) = \{u \in U_\Omega:\ g_{i\delta}(u) \leqslant \theta(\delta)(1 + \Omega(u)),\ i = 1, \ldots, m;$$
$$|g_{i\delta}(u)| \leqslant \theta(\delta)(1 + \Omega(u)),\ i = m+1, \ldots, s\} =$$
$$= \{u \in U_\Omega:\ g_{i\delta}^+(u) \leqslant \theta(\delta)(1 + \Omega(u)),\ i = 1, \ldots, s\}, \quad \theta(\delta) \geqslant \delta. \tag{4}$$

Нетрудно убедиться, что множество $W(\delta)$ заведомо непусто и, более того, представляет собой расширение множества $U \cap U_\Omega$. В самом деле, для любой точки $u \in U \cap U_\Omega$ с учетом (3) имеем:

$$g_{i\delta}^+(u) \leqslant g_i^+(u) + \delta(1 + \Omega(u)) \leqslant \delta(1 + \Omega(u)) \leqslant \theta(\delta)(1 + \Omega(u)), \quad i = 1, \ldots, s.$$

Это означает, что $U \cap U_\Omega \subseteq W(\delta)$, так что $W(\delta)$ действительно является расширением множества $U \cap U_\Omega$.

Далее будем предполагать, что задача (1), (2) имеет сильно согласованную постановку, т. е. выполняется неравенство

$$J_* \leqslant J(u) + \sum_{i=1}^{s} c_i (g_i^+(u))^\nu, \quad u \in U_0, \quad c_i \geqslant 0, \quad \nu > 0. \tag{5}$$

Так как для любой точки $u \in W(\delta)$ справедливы неравенства

$$g_i^+(u) \leqslant g_{i\delta}^+(u) + \delta(1 + \Omega(u)) \leqslant \theta(1 + \Omega(u)) + \delta(1 + \Omega(u)) \leqslant$$
$$\leqslant 2\theta(1 + \Omega(u)) \quad \forall u \in W(\delta), \quad i = 1, \ldots, s, \tag{6}$$

то из (5) следует, что

$$J_* \leqslant J(u) + |c|_1 (2\theta)^\nu (1 + \Omega(u))^\nu \quad \forall u \in W(\delta), \quad \delta > 0, \quad |c|_1 = \sum_{i=1}^{s} |c_i|. \tag{7}$$

Перейдем к описанию методов регуляризации, представляющих собой модификации методов стабилизации, невязки, квазирешений, реализованные на расширенном множестве (4). Как и в § 2 будем считать, что $0 < \nu \leqslant 1$.

1. Начнем с *метода стабилизации*. Возьмем функцию Тихонова

$$t_\delta(u) = J_\delta(u) + \alpha(\delta)\Omega(u), \quad u \in U_\Omega.$$

Решая задачу минимизации первого типа: $t_\delta(u) \to \inf$, $u \in W(\delta)$, определим точку u_δ из условий

$$u_\delta \in W(\delta); \quad t_\delta(u_\delta) \leqslant t_{\delta *} + \varepsilon(\delta), \tag{8}$$

где $\varepsilon(\delta) > 0$, $t_{\delta *} = \inf\limits_{W(\delta)} t_\delta(u)$.

Теорема 1. *Пусть*
1) *множество* U_0 *замкнуто в метрике* M, *функции* $J(u)$, $g_i(u)$, $i=1,\ldots,m$, $|g_i(u)|$, $i=m+1,\ldots,s$, *полунепрерывны снизу на* U_0; $J_*>-\infty$, $U_*\neq\varnothing$, *выполнено условие* (5) *с* $0<\nu\leqslant 1$;
2) $\Omega(u)$ *— стабилизатор задачи* (1), (2) *в метрике* M;
3) *приближения* $J_\delta(u)$, $g_{i\delta}(u)$, $i=1,\ldots,s$, *удовлетворяют условиям* (3);
4) *параметры* $\alpha(\delta)$, $\theta(\delta)$, $\varepsilon(\delta)$ *таковы, что*

$$\alpha(\delta)>0,\quad \varepsilon(\delta)>0,\quad \theta(\delta)>\delta,\quad \lim_{\delta\to 0}(\alpha(\delta)+\varepsilon(\delta)+\theta(\delta))=0, \tag{9}$$

$$\sup_{\delta>0}\frac{\delta}{\alpha(\delta)}+|c|_1\frac{(2\theta(\delta))^\nu}{\alpha(\delta)}<1,\quad \sup_{\delta>0}\frac{\varepsilon(\delta)}{\alpha(\delta)}<\infty. \tag{10}$$

Тогда точка u_δ, *определяемая методом* (8), *существует при всех* $\delta>0$ *и*

$$\lim_{\delta\to 0}J(u_\delta)=J_*,\quad \lim_{\delta\to 0}g_i^+(u_\delta)=0,\quad i=1,\ldots,s,\quad \lim_{\delta\to 0}\rho(u_\delta,U_*)=0. \tag{11}$$

Пусть наряду с условиями 1)–4) *еще выполнено условие*
5) $U_\Omega=U_0$, *функция* $\Omega(u)$ *полунепрерывна снизу на* U_0,

$$\lim_{\delta\to 0}\frac{\delta+(\theta(\delta))^\nu}{\alpha(\delta)}=0,\quad \lim_{\delta\to 0}\frac{\varepsilon(\delta)}{\alpha(\delta)}=0. \tag{12}$$

Тогда кроме (11)

$$\lim_{\delta\to 0}\Omega(u_\delta)=\Omega_*,\quad \lim_{\delta\to 0}\rho(u_\delta,U_{**})=0, \tag{13}$$

где $\Omega_*=\inf\limits_{u\in U_*}\Omega(u)$, $U_{**}=\{u\in U_*\colon \Omega(u)=\Omega_*\}$ *— множество* Ω*-нормальных решений задачи* (1), (2). *Пределы* (11), (13) *равномерны относительно выбора* $J_\delta(u)$, $g_{i\delta}(u)$ *из* (3) *и выбора точки* u_δ *из* (8).

Доказательство. Из (3), (7), (10), с учетом неравенства $(1+\Omega(u))^\nu\leqslant(1+\Omega(u))$ имеем

$$\begin{aligned}t_\delta(u)&\geqslant J(u)+\alpha(\delta)\Omega(u)-\delta(1+\Omega(u))\geqslant\\ &\geqslant J_*-|c|_1(2\theta(\delta))^\nu(1+\Omega(u))+\alpha(\delta)\Omega(u)-\delta(1+\Omega(u))=\\ &=J_*+\Omega(u)(\alpha(\delta)-\delta-|c|_1(2\theta)^\nu)-\delta-|c|_1(2\theta(\delta))^\nu\geqslant\\ &\geqslant J_*-\delta-|c|_1(2\theta(\delta))^\nu>-\infty\quad\forall u\in W(\delta),\end{aligned}$$

так что $t_{\delta*}=\inf\limits_{W(\delta)}t_\delta(u)>-\infty$ и точка u_δ из (8) существует при всех $\delta>0$ по определению нижней грани. Далее с помощью условий (3), (7), (8) можем написать следующую цепочку неравенств:

$$\begin{aligned}J(u_\delta)&\leqslant J(u_\delta)+\alpha(\delta)\Omega(u_\delta)\leqslant J_\delta(u_\delta)+\alpha(\delta)\Omega(u_\delta)+\delta(1+\Omega(u_\delta))=\\ &=t_\delta(u_\delta)+\delta(1+\Omega(u_\delta))\leqslant t_{\delta*}+\varepsilon(\delta)+\delta(1+\Omega(u_\delta))\leqslant\\ &\leqslant J(u_*)+\alpha(\delta)\Omega(u_*)+\delta(1+\Omega(u_*))+\varepsilon(\delta)+\delta(1+\Omega(u_\delta))\leqslant\\ &\leqslant J(u_\delta)+|c|_1(2\theta(\delta))^\nu(1+\Omega(u_\delta))+\alpha(\delta)\Omega(u_*)+\\ &\quad+\delta(1+\Omega(u_*))+\varepsilon(\delta)+\delta(1+\Omega(u_\delta))\quad\forall u_*\in U_{\Omega*}.\end{aligned} \tag{14}$$

Выделим второе и последнее звенья этой цепочки:

$$\begin{aligned}\alpha(\delta)\Omega(u_\delta)\leqslant|c|_1(2\theta(\delta))^\nu(1+\Omega(u_\delta))+\alpha(\delta)\Omega(u_*)+\\ +\delta(1+\Omega(u_*))+\varepsilon(\delta)+\delta(1+\Omega(u_\delta)).\end{aligned}$$

Пользуясь произволом в выборе точки $u_*\in U_{\Omega*}$, в полученном неравенстве величину $\Omega(u_*)$ можем заменить на $\Omega_*=\inf\limits_{U_{\Omega*}}\Omega(u)$ и переписать его в виде

$$\Omega(u_\delta)\leqslant\Omega_*+\gamma(\delta),\quad \gamma(\delta)=\frac{\frac{2\delta+|c|_1(2\theta(\delta))^\nu}{\alpha(\delta)}(\Omega_*+1)+\frac{\varepsilon(\delta)}{\alpha(\delta)}}{1-\frac{|c|_1(2\theta(\delta))^\nu+\delta}{\alpha(\delta)}},\quad \delta>0. \tag{15}$$

Выделим первое и предпоследнее звенья цепочки неравенств (14). С учетом уже доказанной оценки (15) получим

$$\begin{aligned}&J(u_\delta)\leqslant J_*+\beta(\delta),\\ &\beta(\delta)=\alpha(\delta)\Omega_*+\delta(1+\Omega_*)+\varepsilon(\delta)+\delta(1+\Omega_*+\gamma(\delta)),\quad\delta>0.\end{aligned} \tag{16}$$

Кроме того, из (6), (15) имеем

$$g_i^+(u_\delta)\leqslant 2\theta(\delta)(1+\Omega_*+\gamma(\delta))=\rho(\delta),\quad i=1,\ldots,s,\quad\delta>0. \tag{17}$$

Заметим также, что из (5), (16), (17) следует

$$-|c|_1(\rho(\delta))^\nu\leqslant J(u_\delta)-J_*\leqslant\beta(\delta),\quad\delta>0. \tag{18}$$

Дальнейшее доказательство опирается на лемму 3.1 и проводится также, как в теореме 4.1. Замечания 4.1–4.7 к теореме 4.1 с очевидными изменениями сохраняют силу и здесь. □

В выпуклых задачах (1), (2) в банаховых пространствах возможно использование слабых стабилизаторов. Справедлива

Теорема 2. *Пусть*
1) B *— рефлексивное банахово пространство,* U_0 *— выпуклое замкнутое множество из* B, *функции* $J(u)$, $g_i(u)$, $i=1,\ldots,m$, $|g_i(u)|$, $i=m+1,\ldots,s$, *выпуклы и полунепрерывны снизу на* U_0; $J_*>-\infty$, $U_*\neq\varnothing$, *выполнено условие* (5) *с* $0<\nu\leqslant 1$;
2) *функция* $\Omega(u)$ *полунепрерывна снизу, строго равномерно выпукла на* U_0;
3) *приближения* $J_\delta(u)$, $g_{i\delta}(u)$, $i=1,\ldots,s$, *удовлетворяют условиям* (3);
4) *параметры* $\alpha(\delta)$, $\varepsilon(\delta)$, $\theta(\delta)$ *удовлетворяют условиям* (9), (12).
Тогда семейство точек $\{u_\delta,\ \delta>0\}$, *определенное методом* (8) *таково, что*

$$\begin{aligned}&\lim_{\delta\to 0}J(u_\delta)=J_*,\quad \lim_{\delta\to 0}g_i^+(u_\delta)=0,\quad i=1,\ldots,s,\\ &\lim_{\delta\to 0}\Omega(u_\delta)=\Omega_*,\quad \lim_{\delta\to 0}\|u_\delta-v_*\|=0,\end{aligned} \tag{19}$$

где v_* *— точка минимума функции* $\Omega(u)$ *на множестве* U_*. *Пределы в* (19) *равномерны относительно выбора* $J_\delta(u)$, $g_{i\delta}(u)$, $i=1,\ldots,s$, *из* (3) *и выбора точки* u_δ *из* (8).

Доказательство опирается на уже доказанные оценки (15)–(18), на лемму 3.2 и проводится так же, как в теореме 4.2. □

2. *Метод невязки* с расширением множества заключается в определении точки u_δ из условий

$$u_\delta \in V(\delta) = \{u \in W(\delta): J_\delta(u) \leqslant \overline{J}_{\delta *} + \sigma(\delta)\}, \tag{20}$$

$$\Omega(u_\delta) \leqslant \Omega_{\delta *} + \mu(\delta), \quad \delta > 0, \tag{21}$$

где $\overline{J}_{\delta *} = \inf\limits_{W(\delta)} (J_\delta(u) + \delta\Omega(u) + |c|_1 (2\theta(\delta))^\nu (1+\Omega(u))^\nu)$, $\Omega_{\delta *} = \inf\limits_{V(\delta)} \Omega(u)$, $\sigma(\delta) > 0$, $\mu(\delta) > 0$. Если величина $\overline{J}_{\delta *}$ или ее оценка $\overline{J}_{\delta *} + \sigma(\delta)$, характеризующие множество $V(\delta)$, заранее неизвестны, то предварительно нужно решить задачу первого типа

$$\overline{J}_\delta(u) = J_\delta(u) + \delta\Omega(u) + |c|_1 (2\theta(\delta))^\nu (1+\Omega(u))^\nu \to \inf, \quad u \in W(\delta),$$

и лишь затем приступить к поиску точки u_δ, как приближенного решения задачи первого типа: $\Omega(u) \to \inf$, $u \in V(\delta)$.

Теорема 3. *Пусть выполнены условия* 1)–3) *теоремы* 1 *и параметры* $\sigma(\delta)$, $\mu(\delta)$, $\theta(\delta)$ *таковы, что*

$$\sigma(\delta) > 0, \quad \mu(\delta) > 0, \quad \theta(\delta) > \delta, \quad \lim_{\delta \to 0} (\sigma(\delta) + \mu(\delta) + \theta(\delta)) = 0, \tag{22}$$

$$\sup_{\delta > 0} \mu(\delta) < \infty, \quad \delta(3 + \Omega_*) \leqslant \sigma(\delta), \quad \delta > 0, \tag{23}$$

где $\Omega_* = \inf\limits_{U_\Omega} \Omega(u)$. *Тогда множество* $V(\delta)$, *определенное согласно* (20), *непусто при всех* $\delta > 0$, *семейство* $\{u_\delta, \delta > 0\}$ *из* (21) *таково, что*

$$\lim_{\delta \to 0} J(u_\delta) = J_*, \quad \lim_{\delta \to 0} g_i^+(u_\delta) = 0, \quad i = 1, \ldots, s; \quad \lim_{\delta \to 0} \rho(u_\delta, U_*) = 0. \tag{24}$$

Если, кроме того, $U_\Omega = U_0$, *функция* $\Omega(u)$ *полунепрерывна снизу на* U_0, $\lim\limits_{\delta \to 0} \mu(\delta) = 0$, *то наряду с* (24) *справедливы равенства*

$$\lim_{\delta \to 0} \Omega(u_\delta) = \Omega_*, \quad \lim_{\delta \to 0} \rho(u_\delta, U_{**}) = 0, \tag{25}$$

где $U_{**} = \{u \in U_*: \Omega(u) = \Omega_* = \inf\limits_{U_*} \Omega(u)\}$. *Пределы в* (21), (25) *равномерны относительно выбора* $J_\delta(u)$, $g_{i\delta}(u)$ *из* (3) *и выбора точки* u_δ *из* (20), (21).

Доказательство. Из (3), (7) имеем: $\overline{J}_\delta(u) \geqslant J(u) - \delta(1 + \Omega(u)) + \delta\Omega(u) + |c|_1 (2\theta(\delta))^\nu (1+\Omega(u))^\nu \geqslant J_* - \delta$ $\forall u \in W(\delta)$, т. е. $\overline{J}_{\delta *} > -\infty$. Далее, по определению Ω_* для любого ε, $0 < \varepsilon \leqslant 1$, найдется точка $w_\varepsilon \in U_{\Omega *} \subset W(\delta)$, $\Omega(w_\varepsilon) \leqslant \Omega_* + \varepsilon$. Тогда с учетом (3), (7) получаем $J_\delta(w_\varepsilon) \leqslant J(w_\varepsilon) + \delta(1 + \Omega(w_\varepsilon)) \leqslant J(u) + |c|_1 (2\theta(\delta))^\nu (1+\Omega(u))^\nu + \delta(1 + \Omega_* + \varepsilon) \leqslant \overline{J}_\delta(u) + \delta(3 + \Omega_*)$ $\forall u \in W(\delta)$. Отсюда и из (23) следует, что $J_\delta(w_\varepsilon) \leqslant \overline{J}_{\delta *} + \sigma(\delta)$. Это значит, что $w_\varepsilon \in V(\delta)$, т. е. $V(\delta) \neq \varnothing$ $\forall \delta > 0$. Тогда точка u_δ из (20), (21) существует по определению нижней грани. Для любой точки u_δ с учетом включения $w_\varepsilon \in V(\delta)$ имеем: $\Omega(u_\delta) \leqslant \Omega(w_\varepsilon) + \mu(\delta) \leqslant \Omega_* + \varepsilon + \mu(\delta)$ $\forall \varepsilon$, $0 < \varepsilon \leqslant 1$. Отсюда при $\varepsilon \to 0$ получим

$$\Omega(u_\delta) \leqslant \Omega_* + \gamma(\delta), \quad \gamma(\delta) = \mu(\delta), \quad \delta > 0. \tag{26}$$

Далее, из (3), (20), (21), (26), $w_\varepsilon \in V(\delta)$ следует:

$$J(u_\delta) \leqslant J_\delta(u_\delta) + \delta(1 + \Omega(u_\delta)) \leqslant \overline{J}_{\delta *} + \sigma(\delta) + \delta(1 + \Omega_* + \mu(\delta)) \leqslant$$
$$\leqslant \overline{J}_\delta(w_\varepsilon) + \sigma(\delta) + \delta(1 + \Omega_* + \mu(\delta)) \leqslant$$
$$\leqslant J(w_\varepsilon) + \delta\Omega(w_\varepsilon) + |c|_1 (2\theta(\delta))^\nu (1 + \Omega(w_\varepsilon))^\nu +$$
$$+ \delta(1 + \Omega(w_\varepsilon)) + \sigma(\delta) + \delta(1 + \Omega_* + \mu(\delta)) \leqslant J_* + \delta(1 + 2\Omega_* + 2\varepsilon) +$$
$$+ |c|_1 (2\theta(\delta))^\nu (1 + \Omega_* + \varepsilon)^\nu + \sigma(\delta) + \delta(1 + \Omega_* + \mu(\delta)).$$

При $\varepsilon \to 0$ отсюда приходим к оценке

$$J(u_\delta) \leqslant J_* + \beta(\delta),$$
$$\beta(\delta) = \delta(2 + 3\Omega_* + \mu(\delta)) + |c|_1 (2\theta(\delta))^\nu (1 + \Omega_*)^\nu + \sigma(\delta), \quad \delta > 0. \tag{27}$$

Оценки (17), (18), где величина $\beta(\delta)$ взята из (27), остаются справедливыми и для метода невязки. Дальнейшее доказательство опирается на лемму 3.1 и проводится также, как в теореме 4.1. □

Теорема 4. *Пусть выполнены условия* 1)–3) *теоремы* 2, *параметры* $\sigma(\delta)$, $\mu(\delta)$, $\theta(\delta)$ *удовлетворяют условиям* (22), (23), $\lim\limits_{\delta \to 0} \mu(\delta) = 0$. *Тогда для семейства точек* $\{u_\delta, \delta > 0\}$, *определенных методом* (20), (21), *справедливы утверждения теоремы* 2.

Доказательство опирается на уже доказанные оценки (26), (27), (17), на лемму 3.2 и проводится так же, как в теореме 4.1. □

3. Наконец, кратко остановимся на *методе квазирешений* с расширением множества. Этот вариант метода квазирешений также применяется при выполнении предположения (6.1), когда известно число r такое, что

$$U_{\Omega *} \cap Q_r \neq \varnothing, \quad Q_r = \{u \in U_\Omega: \Omega(u) \leqslant r\}. \tag{28}$$

Тогда $U_{\Omega *} \cap Q_r \subseteq Q_{r\delta} = \{u \in W(\delta): \Omega(u) \leqslant r\} \neq \varnothing$ $\forall \delta > 0$. Суть метода: ищется точка u_δ из условий:

$$u_\delta \in Q_{r\delta}, \quad J_\delta(u_\delta) \leqslant J_{\delta *} + \xi(\delta), \tag{29}$$

где $J_{\delta *} = \inf\limits_{Q_{r\delta}} J_\delta(u)$, $\xi(\delta) > 0$ $\forall \delta > 0$.

Теорема 5. *Пусть выполнены условия* 1)–3) *теоремы* 1 *и параметры* $\xi(\delta)$, $\theta(\delta)$ *таковы, что*

$$\xi(\delta) > 0, \quad \theta(\delta) > \delta, \quad \lim_{\delta \to 0} (\xi(\delta) + \theta(\delta)) = 0. \tag{30}$$

Тогда точка u_δ, *определяемая условиями* (29), *существует при всех* $\delta > 0$ *и*

$$\lim_{\delta \to 0} J(u_\delta) = J_*; \quad \lim_{\delta \to 0} g_i^+(u_\delta) = 0, \quad i = 1, \ldots, s; \quad \lim_{\delta \to 0} \rho(u_\delta, U_*) = 0. \tag{31}$$

Пределы в (31) *равномерны относительно выбора* $J_\delta(u)$, $g_{i\delta}(u)$, $i = 1, \ldots, s$, *из* (3) *и выбора точек* u_δ *из* (29).

Доказательство. Из (3), (7) и неравенства $\Omega(u) \leqslant r$ $\forall u \in Q_{r\delta}$ имеем

$$J_\delta(u) \geqslant J(u) - \delta(1 + \Omega(u)) \geqslant J_* - |c|_1 (2\theta(\delta))^\nu (1 + \Omega(u))^\nu - \delta(1 + r) \geqslant$$
$$\geqslant J_* - |c|_1 (2\theta(\delta))^\nu (1 + r)^\nu - \delta(1 + r) \quad \forall u \in Q_{r\delta}.$$

Следовательно, $J_{\delta *} > -\infty$ и точка u_δ из (29) существует по определению нижней грани. Возьмем произвольные точки u_δ из (29), $u_* \in U_{\Omega *} \cap Q_r$. Тогда

$$\Omega(u_\delta) \leqslant \Omega_* + \gamma(\delta), \quad \gamma(\delta) = r - \Omega_* \quad \forall \delta > 0. \tag{32}$$

Кроме того с учетом (3), (29) получаем

$$J(u_\delta) \leqslant J_\delta(u_\delta) + \delta(1 + \Omega(u_\delta)) \leqslant J_{\delta *} + \xi(\delta) + \delta(1 + r) \leqslant$$
$$\leqslant J(u_*) + \delta(1 + \Omega(u_*)) + \xi(\delta) + \delta(1 + r) \leqslant J_* + \beta(\delta),$$
$$\beta(\delta) = \xi(\delta) + 2\delta(1 + r) \quad \forall \delta > 0. \tag{33}$$

Оценки (17), (18) с $\gamma(\delta)$ из (32), $\beta(\delta)$ из (33) справедливы и для метода (29). Дальнейшее доказательство опирается на лемму 3.1 и проводится так же, как в теореме 4.1. □

Требуемое в теоремах 1–5 условие $\theta(\delta) > \delta$ формально можно заменить на $\theta(\delta) \geqslant \delta$, однако при $\theta(\delta) = \delta$, как было замечено в § 2, входящие в методы регуляризации вспомогательные задачи первого типа могут оказаться неустойчивыми. Другие модификации методов регуляризации задач минимизации второго типа, основанных на сочетании идей расширения множества и методов покрытий, можно извлечь из работы [592], в которой рассмотрены более сложные по сравнению с (1), (2) многокритериальные задачи и устойчивые методы их решения.

4. Для иллюстрации вышеизложенного рассмотрим каноническую задачу линейного программирования:

$$J(u) = \langle c, u \rangle \to \inf, \quad u \in U = \{u \geqslant 0\colon Au = b\} =$$
$$= \{u \geqslant 0\colon g_i(u) = \langle a_i, u \rangle - b^i = 0,\ i = 1, \ldots, m\}, \tag{34}$$

где $A = \{a_{ij}\}$ — матрица размера $m \times n$, $a_i = (a_{i1}, \ldots, a_{in})$ — i-я строка этой матрицы, $b = (b^1, \ldots, b^m)^\top$, $c = (c^1, \ldots, c^n)^\top$. Предположим, что $U \neq \varnothing$, $J_* > -\infty$; тогда $U_* \neq \varnothing$ (теорема 3.5.1). Пусть вместо точных a_{ij}, b^i, c^j известны их приближения $a_{ij\delta}$, b^i_δ, c^j_δ такие, что

$$|a_{ij\delta} - a_{ij}| \leqslant \delta, \quad |b^i_\delta - b^i| \leqslant \delta, \quad |c^j_\delta - c^j| \leqslant \delta, \quad i = 1, \ldots, m; \quad j = 1, \ldots, n. \tag{35}$$

Положим тогда

$$J_\delta(u) = \langle c_\delta, u \rangle, \quad g_{i\delta}(u) = \langle a_{i\delta}, u \rangle - b^i_\delta, \quad i = 1, \ldots, m; \quad a_{i\delta} = (a_{i1\delta}, \ldots, a_{in\delta}). \tag{36}$$

Поскольку $u \geqslant 0$, то стабилизатором в этой задаче можно взять функцию $\Omega(u) = u^1 + \ldots + u^n$. Погрешности задания входных данных $J(u)$, $g_i(u)$ согласованы с этим стабилизатором:

$$|J_\delta(u) - J(u)| \leqslant \delta |u| \leqslant \delta(1 + \Omega(u)),$$
$$|g_{i\delta}(u) - g_i(u)| \leqslant \delta(1 + \Omega(u)), \quad \forall u \geqslant 0, \quad i = 1, \ldots, m. \tag{37}$$

Ограничимся здесь рассмотрением лишь приближений вида (36), хотя условия (37), конечно, допускают и другие функции $J_\delta(u)$, $g_{i\delta}(u)$, $i = 1, \ldots, m$. Тогда расширенное множество $W(\delta)$ может быть представлено в виде

$$W(\delta) = \{u \geqslant 0\colon -\theta(\delta)(1 + u^1 + \ldots + u^n) \leqslant \langle a_{i\delta}, u \rangle - b_{i\delta} \leqslant$$
$$\leqslant \theta(\delta)(1 + u^1 + \ldots + u^n),\ i = 1, \ldots, m\}. \tag{38}$$

Метод стабилизации сводится к задаче

$$t_\delta(u) = \langle c_\delta, u \rangle + \alpha(\delta)(u^1 + \ldots + u^n) \to \inf, \quad u \in W(\delta), \tag{39}$$

которая, как нетрудно видеть, также является задачей линейного программирования. Задачу (39) можно решать приближенно в смысле неравенства (8). Пусть параметры $\alpha(\delta)$, $\theta(\delta)$, $\varepsilon(\delta)$ удовлетворяют условиям (9) и $\lim\limits_{\delta \to 0} \frac{\delta + \theta(\delta) + \varepsilon(\delta)}{\alpha(\delta)} = 0$. Функция Лагранжа задачи (34) имеет седловую точку (теорема 3.5.5), поэтому эта задача имеет сильно согласованную постановку, и неравенство (5) может быть записано в виде: $J_* \leqslant \langle c, u \rangle + \sum\limits_{i=1}^{m} |\lambda_i^*| |\langle a_i, u \rangle - b^i| \ \forall u \geqslant 0$. Таким образом, все условия теоремы 1 выполнены, и для точек u_δ из (8), (39) справедливы все равенства (11), (13).

Нетрудно убедиться, что методы невязки и квазирешений для задачи (34) приводят к задачам (20), (21) и (29) соответственно, являющиеся также задачами линейного программирования. Более подробно методы регуляризации для общей задачи линейного программирования рассмотрены в [179], там же приведены оценки скорости сходимости методов, показано, что эти оценки неулучшаемы по порядку на классе задач линейного программирования и их порядок совпадает с порядком погрешности задания входных данных. Интересно отметить, что в задачах линейного программирования понятия устойчивости по функции и устойчивости по аргументу в предположениях (35), (36) равносильны [179], поэтому в линейном программировании нет большого смысла различать задачи первого и второго типов, строить для них отдельные методы регуляризации.

Замечания 6.1–6.4 с очевидными уточнениями остаются справедливыми и для методов регуляризации, рассмотренных в этом параграфе. Отметим также, что методы (8), (21), (29) могут применяться, например, в тех случаях, когда использование методов из §§ 4–6 затруднительно из-за возможной овражности вспомогательных задач, связанной с наличием в этих методах неограниченно возрастающего штрафного коэффициента.

Перейдем к изложению других методов регуляризации, которые представляют собой сочетание метода стабилизации из § 4 и уже известных нам методов минимизации из гл. 5, § 8.4.

Упражнения

1. Применить методы невязки и квазирешений к канонической задаче линейного программирования (34).

2. Применить методы (8), (21), (29) к задачам из примеров 1.6–1.9.

§ 8. Регуляризованный метод проекции градиента

Методы решения неустойчивых задач второго типа можно также строить на основе обычных методов минимизации, подвергнув их некоторой процедуре регуляризации. Для построения регуляризованных методов часто поступают следующим образом: в общей схеме конкретного метода миними-

зации вместо целевой функции используют функцию Тихонова. Так, например, если в итерационный процесс градиентного метода $u_{k+1}=u_k-\beta_k J'(u_k)$, $k=0,1,\ldots$, вместо $J'(u_k)$ формально подставить градиент функции Тихонова $T_\delta(u)=J(u)+\alpha(\delta)\|u\|^2$ в точке u_k, то придем к регуляризованному градиентному методу

$$u_{k+1}=u_k-\beta_k(J'(u_k)+2\alpha_k u_k),\quad \alpha_k=\alpha(\delta_k),\quad k=0,1,\ldots,$$

в котором надо согласованно выбрать параметры α_k,β_k. Именно по этой схеме мы дальше будем проводить регуляризацию метода проекции градиента, условного градиента и некоторых других методов минимизации из гл. 5. Для исследования возникающих здесь трудных проблем согласования параметров исходного метода минимизации с параметрами, входящими в функцию Тихонова, для доказательства сходимости получившегося метода мы будем пользоваться принципом итеративной регуляризации из [62; 63].

1. Начнем с метода проекции градиента для задачи

$$J(u)\to\inf,\quad u\in U, \tag{1}$$

$$U=\{u\in U_0:\ g_i(u)\leqslant 0,\ i=1,\ldots,m;\ g_i(u)=0,\ i=m+1,\ldots,s\}, \tag{2}$$

где U_0 — заданное выпуклое замкнутое множество из гильбертова пространства H (возможно, $U_0=H$), считая, что функции $J(u)$, $g_i(u)$ дифференцируемы по Фреше на U_0 (определение 8.3.3). Ограничения типа равенств и неравенств из (2) будем учитывать с помощью штрафной функции (5.15.8):

$$P(u)=\sum_{i=1}^{s}(g_i^+(u))^p,\quad u\in U_0,\quad p>1. \tag{3}$$

Тогда точная функция Тихонова

$$T_k(u)=J(u)+A_kP(u)+\alpha_k\|u\|^2,\quad u\in U_0,\quad A_k>0,\quad \alpha_k>0,\quad k=0,1,\ldots \tag{4}$$

при каждом k дифференцируема по Фреше на U_0, причем ее градиент равен

$$T_k'(u)=J'(u)+A_kP'(u)+2\alpha_k u,\quad u\in U_0,\quad k=0,1,\ldots. \tag{5}$$

Пусть вместо точных значений градиентов $J'(u)$, $P'(u)$ известны их приближения $J_k'(u)$, $P_k'(u)$. Тогда в качестве приближения для точного значения градиента (5) можем взять

$$t_k'(u)=J_k'(u)+A_kP_k'(u)+2\alpha_k u,\quad u\in U_0,\quad k=0,1,\ldots. \tag{6}$$

Рассмотрим итерационный метод

$$u_{k+1}=\mathcal{P}_{U_0}(u_k-\beta_k t_k'(u_k)),\quad \beta_k>0,\quad k=0,1,\ldots, \tag{7}$$

где $\mathcal{P}_{U_0}(z)$ — проекция точки $z\in H$ на множество U_0. Метод (7) будем называть регуляризованным методом проекции градиента. Приведем условия на задачу (1), (2), условия согласования параметров α_k, A_k, β_k с погрешностью задания производных $J'(u)$, $P'(u)$, обеспечивающие сходимость последовательности $\{u_k\}$, порожденной методом (7), к нормальному решению задачи (1), (2) в метрике H.

Теорема 1. *Пусть*

1) *U_0 — выпуклое замкнутое множество из гильбертова пространства H, функции $J(u)$, $g_i(u)$, $i=1,\ldots,s$, дифференцируемы по Фреше на U_0, функции $J(u)$, $g_i(u)$, $i=1,\ldots,m$, $|g_i(u)|$, $i=m+1,\ldots,s$, выпуклы на U_0; $J_*>-\infty$, $U_*\neq\infty$ и справедливы неравенства*

$$\begin{aligned}&\|J'(u)-J'(v)\|^2\leqslant L\langle J'(u)-J'(v),u-v\rangle\quad\forall u,v\in U_0,\\ &\|P'(u)-P'(v)\|^2\leqslant L\langle P'(u)-P'(v),u-v\rangle\quad\forall u,v\in U_0,\end{aligned} \tag{8}$$

где $L=\mathrm{const}>0$; задача (1), (2) имеет сильно согласованную постановку, т. е. выполняется неравенство

$$J_*\leqslant J(u)+\sum_{i=1}^{s}c_i(g_i^+(u))^\nu\quad\forall u\in U_0.$$

при некоторых $c_i\geqslant 0$, $\nu>0$; параметр p штрафной функции (3) таков, что $p>1$, $p\geqslant\nu$; 2) при каждом фиксированном $u\in U_0$ и каждом $k\geqslant 0$ вместо точных $J'(u)$, $P'(u)$ имеются их приближения $J_k'(u)$, $P_k'(u)$ с погрешностями:

$$\begin{aligned}&\|J_k'(u)-J'(u)\|\leqslant\delta_k(1+\|u\|),\quad \|P_k'(u)-P'(u)\|\leqslant\delta_k(1+\|u\|)\\ &\forall u\in U_0,\quad \delta_k>0,\quad k=0,1,\ldots;\end{aligned} \tag{9}$$

3) *числовые последовательности $\{\alpha_k\}$, $\{A_k\}$, $\{\beta_k\}$, $\{\delta_k\}$ удовлетворяют условиям*

$$\alpha_k>0,\quad A_{k+1}\geqslant A_k>0,\quad \delta_k>0,\quad 0<\beta_k\leqslant\frac{2}{4\alpha_k+L+LA_k},\quad k=0,1,\ldots, \tag{10}$$

$$\lim_{k\to\infty}(\alpha_k+\delta_k)=0,\quad \lim_{k\to\infty}A_k=+\infty,\quad \lim_{k\to\infty}\alpha_kA_k^{\frac{\nu}{p-\nu}}=+\infty \tag{11}$$

(при $p=\nu>1$ последнее условие не нужно),

$$\lim_{k\to\infty}\frac{\alpha_k-\alpha_{k+1}}{\alpha_k^2\beta_k}=0,\quad \lim_{k\to\infty}\frac{A_{k+1}-A_k}{\alpha_k^2\beta_k}=0,\quad \lim_{k\to\infty}\frac{\delta_k+\delta_kA_k}{\alpha_k}=0. \tag{12}$$

Тогда при любом начальном приближении $u_0\in U_0$ последовательность $\{u_k\}$, определяемая методом (7), такова, что

$$\lim_{k\to\infty}J(u_k)=J(u_*)=J_*,\quad \lim_{k\to\infty}g_i^+(u_k)=0,\quad i=1,\ldots,s,\quad \lim_{k\to\infty}\|u_k-u_*\|=0, \tag{13}$$

где u_* — нормальное решение задачи (1), (2). Сходимость в (13) равномерная относительно выбора реализаций $J_k'(u)$, $P_k'(u)$ из условия (9).

Замечание 1. Из условия (8) следует, что $J(u)$, $P(u)$ — выпуклые функции и их градиенты удовлетворяют условию Липшица:

$$\|J'(u)-J'(v)\|\leqslant L\|u-v\|,\quad \|P'(u)-P'(v)\|\leqslant L\|u-v\|,\quad \forall u,v\in U \tag{14}$$

(теорема 4.2.16, достаточность). Заметим, что теорема 4.2.16 остается верной и в гильбертовом пространстве, если $\mathrm{int}\,U_0\neq\varnothing$, $J(u)\in C^2(U_0)$.

Замечание 2. В качестве последовательностей $\{\alpha_k\}$, $\{A_k\}$, $\{\beta_k\}$, $\{\delta_k\}$, удовлетворяющих условиям (10)–(12), можно, например, взять

$$\begin{aligned}&\alpha_k=a(k+1)^{-\alpha},\quad A_k=(k+1)^A,\quad \delta_k=b(k+1)^{-\gamma},\\ &\beta_k=\frac{2}{L+4a+L(k+1)^A},\quad k=0,1,\ldots,\end{aligned} \tag{15}$$

где постоянные $a>0$, $b>0$, $\alpha>0$, $A>0$, $\gamma>0$, $\alpha+A<\min\{\frac{1}{2};\gamma\}$, $\alpha<A\frac{\nu}{p-\nu}$.

Доказательство теоремы 1. При сделанных предположениях множества U, U_* выпуклы, замкнуты, и сильно выпуклая $\Omega(u)=\|u\|^2$ достигает нижней грани на U_* в единственной точке u_* (теорема 8.2.10), т. е. нормальное решение u_* задачи (1), (2) существует и единственно. Кроме того, функция $T_k(u)$, определенная формулой (4), сильно выпукла на выпуклом замкнутом множестве U_0, поэтому условия

$$v_k\in U_0,\quad T_k(v_k)=\inf_{u\in U_0}T_k(u) \tag{16}$$

однозначно определяют точку v_k. Покажем, что

$$\lim_{k\to\infty}J(v_k)=J_*,\quad \lim_{k\to\infty}g_i^+(v_k)=0,\quad \lim_{k\to\infty}\|v_k-u_*\|=0. \tag{17}$$

Воспользуемся теоремой 4.3. Положим

$$\alpha(\delta)=\alpha_k,\quad A(\delta)=A_k,\quad u_\delta=v_k\quad\forall\delta,\ \delta_k\leqslant\delta<\delta_{k+1},\ k=0,1,\dots$$

Из (10), (11) следует, что так определенные функции $\alpha(\delta)$, $A(\delta)$ удовлетворяют условиям (4.51), а для такой точки u_δ в силу (4), (16) выполнены условия (4.53). Отсюда и из теоремы 4.3 вытекают равенства (17); из (4.56) следует оценка

$$\|v_k\|\leqslant R=\begin{cases}\left(\|u_*\|^2+\dfrac{B}{\alpha_k A_k^{\frac{\nu}{p-\nu}}}\right)^{1/2} & \text{при } p>\nu,\\ \|u_*\| & \text{при } p=\nu>1.\end{cases} \tag{18}$$

Поскольку

$$\|u_k-u_*\|\leqslant\|u_k-v_k\|+\|v_k-u_*\|,\quad k=0,1,\dots, \tag{19}$$

то из (17), (19) следует, что для доказательства равенства $\lim\limits_{k\to\infty}\|u_k-u_*\|=0$ достаточно установить, что $\lim\limits_{k\to\infty}\|u_k-v_k\|=0$. Обозначим $w_k=\|u_k-v_k\|^2$. Справедливо элементарное неравенство

$$(a+b)^2\leqslant(1+\alpha_k\beta_k)a^2+(1+\alpha_k\beta_k)\frac{1}{\alpha_k\beta_k}b^2\quad\forall a,b,\quad\alpha_k>0,\quad\beta_k>0. \tag{20}$$

Из (20) имеем

$$w_{k+1}=\|u_{k+1}-v_{k+1}\|^2\leqslant(\|u_{k+1}-v_k\|+\|v_k-v_{k+1}\|)^2\leqslant$$
$$\leqslant(1+\alpha_k\beta_k)\|u_{k+1}-v_k\|^2+(1+\alpha_k\beta_k)\frac{1}{\alpha_k\beta_k}\|v_k-v_{k+1}\|^2,\quad k=0,1,\dots \tag{21}$$

Сначала оценим слагаемое $\|v_k-v_{k+1}\|^2$. Из (16) следует

$$\langle T_k'(v_k),u-v_k\rangle\geqslant 0\quad\forall u\in U_0,\quad k=0,1,\dots \tag{22}$$

(теорема 8.3.3). Взяв в (22) $u=v_{k+1}$, получим

$$\langle T_k'(v_k),v_{k+1}-v_k\rangle\geqslant 0,\quad k=0,1,\dots$$

Аналогично имеем

$$\langle T_{k+1}'(v_{k+1}),v_k-v_{k+1}\rangle\geqslant 0,\quad k=0,1,\dots$$

Из сильной выпуклости функции (4) с постоянной сильной выпуклости $\varkappa=\mu=2\alpha_k$ и теоремы 4.3.3 следует, что

$$2\alpha_k\|v_k-v_{k+1}\|^2\leqslant\langle T_k'(v_{k+1})-T_k'(v_k),v_{k+1}-v_k\rangle,\quad k=0,1,\dots$$

Сложим последние три неравенства:

$$2\alpha_k\|v_k-v_{k+1}\|^2\leqslant\langle T_k'(v_{k+1})-T_{k+1}'(v_{k+1}),v_{k+1}-v_k\rangle,\quad k=0,1,\dots \tag{23}$$

Из (5) и (23) получаем

$$2\alpha_k\|v_k-v_{k+1}\|^2\leqslant(A_k-A_{k+1})\langle P'(v_{k+1}),v_{k+1}-v_k\rangle+$$
$$+2(\alpha_k-\alpha_{k+1})\langle v_{k+1},v_{k+1}-v_k\rangle,\quad k=0,1,\dots \tag{24}$$

Зафиксируем произвольную точку $\overline{u}\in U_0$. Тогда из (14), (18) имеем

$$\|P'(v_{k+1})\|\leqslant L\|v_{k+1}-\overline{u}\|+\|P'(\overline{u})\|\leqslant L(R+\|\overline{u}\|)+\|P'(\overline{u})\|.$$

Отсюда и из (18), (24) вытекает оценка

$$\|v_{k+1}-v_k\|\leqslant\frac{|\alpha_k-\alpha_{k+1}|}{\alpha_k}R+\frac{A_{k+1}-A_k}{2\alpha_k}[L(R+\|\overline{u}\|)+\|P'(\overline{u})\|],\quad k=0,1,\dots \tag{25}$$

Перейдем к оценке величины $\|u_{k+1}-v_k\|$ из первого слагаемого правой части (21). Примем в (22) $u=u_{k+1}$, получим

$$\langle T_k'(v_k),u_{k+1}-v_k\rangle\geqslant 0,\quad k=0,1,\dots \tag{26}$$

Далее, из (7) следует (теорема 4.4.1)

$$\langle u_{k+1}-u_k+\beta_k t_k'(u_k),u-u_{k+1}\rangle\geqslant 0\quad\forall u\in U_0,\quad k=0,1,\dots$$

Отсюда при $u=v_k$ имеем

$$\langle u_{k+1}-u_k+\beta_k t_k'(u_k),v_k-u_{k+1}\rangle\geqslant 0,\quad k=0,1,\dots$$

Сложим это неравенство с неравенством (26), умноженным на $\beta_k>0$, получим

$$\langle u_{k+1}-u_k,v_k-u_{k+1}\rangle+\beta_k\langle t_k'(u_k)-T_k'(v_k),v_k-u_{k+1}\rangle\geqslant 0,\quad k=0,1,\dots \tag{27}$$

С учетом формул (5), (6) для $T_k'(u)$, $t_k'(u)$ перепишем (27) в виде

$$0\leqslant\langle u_{k+1}-u_k,v_k-u_{k+1}\rangle+2\alpha_k\beta_k\langle u_k-v_k,v_k-u_{k+1}\rangle+$$
$$+\beta_k\langle J_k'(u_k)-J'(v_k),v_k-u_{k+1}\rangle+$$
$$+\beta_k A_k\langle P_k'(u_k)-P'(v_k),v_k-u_{k+1}\rangle,\quad k=0,1,\dots \tag{28}$$

Оценим каждое слагаемое из правой части (28). Для первого и второго слагаемых имеем

$$\langle u_{k+1}-u_k,v_k-u_{k+1}\rangle=\frac{1}{2}\|u_k-v_k\|^2-\frac{1}{2}\|u_{k+1}-u_k\|^2-\frac{1}{2}\|v_k-u_{k+1}\|^2,\quad k=0,1,\dots \tag{29}$$

$$\langle u_k-v_k,v_k-u_{k+1}\rangle=\frac{1}{2}\|u_{k+1}-u_k\|^2-\frac{1}{2}\|u_k-v_k\|^2-\frac{1}{2}\|v_k-u_{k+1}\|^2,\quad k=0,1,\dots \tag{30}$$

Для оценки третьего и четвертого слагаемых воспользуемся неравенствами:

$$\begin{aligned}&\langle J'(u)-J'(v),\,v-w\rangle\leqslant\tfrac{1}{4}L\|u-w\|^2,\\&\langle P'(u)-P'(v),\,v-w\rangle\leqslant\tfrac{1}{4}L\|u-w\|^2\quad\forall u,v,w\in U_0,\end{aligned}\tag{31}$$

вытекающими из условий (8) (теорема 4.2.16). Из первого неравенства (31) при $u=u_k$, $v=v_k$, $w=u_{k+1}$ и условий (9), (14) следует

$$\begin{aligned}\langle J_k'(u_k)-J'(v_k),v_k-u_{k+1}\rangle&=\langle(J_k'(u_k)-J'(u_k))+(J'(u_k)-J'(v_k)),\,v_k-u_{k+1}\rangle\leqslant\\&\leqslant\delta_k(1+\|u_k\|)\|v_k-u_{k+1}\|+\tfrac{1}{4}L\|u_k-u_{k+1}\|^2.\end{aligned}\tag{32}$$

С помощью элементарных неравенств $|ab|\leqslant a^2+\frac{1}{4}b^2$, $(a+b)^2\leqslant 2(a^2+b^2)$ и оценки (18) из (32) имеем

$$\begin{aligned}\langle J_k'(u_k)-J'(v_k),\,v_k-u_{k+1}\rangle&\leqslant\tfrac{1}{4}L\|u_k-u_{k+1}\|^2+\delta_k\|u_{k+1}-v_k\|^2+\\&+\tfrac{1}{2}\delta_k\|u_k-v_k\|^2+\tfrac{1}{2}\delta_k(1+R)^2,\quad k=0,1,\ldots\end{aligned}\tag{33}$$

Рассуждая также с учетом второго неравенства (31), получаем оценку

$$\begin{aligned}\langle P_k'(u_k)-P'(v_k),\,v_k-u_{k+1}\rangle&\leqslant\tfrac{1}{4}L\|u_k-u_{k+1}\|^2+\delta_k\|u_{k+1}-v_k\|^2+\\&+\tfrac{1}{2}\delta_k\|u_k-v_k\|^2+\tfrac{1}{2}\delta_k(1+R)^2,\quad k=0,1,\ldots\end{aligned}\tag{34}$$

Подставим оценки (29), (30), (33), (34) в (28). После простых преобразований будем иметь

$$\begin{aligned}\|u_{k+1}-v_k\|^2(1+2\alpha_k\beta_k-2\beta_k\delta_k-2\beta_k\delta_kA_k)&\leqslant\\\leqslant\|u_k-v_k\|^2(1-2\alpha_k\beta_k+\beta_k\delta_k+\beta_k\delta_kA_k)&+\\+\|u_{k+1}-u_k\|^2(-1+2\alpha_k\beta_k+\tfrac{1}{2}L\beta_k+\tfrac{1}{2}L\beta_kA_k)&+\\+(\beta_k\delta_k+\beta_k\delta_kA_k)(1+R)^2,\quad k=0,1,\ldots\end{aligned}\tag{35}$$

В силу условий (10)–(12) справедливы неравенства:

$$\begin{aligned}-1+2\alpha_k\beta_k+\tfrac{1}{2}L\beta_k+\tfrac{1}{2}L\beta_kA_k&=-1+\tfrac{\beta_k}{2}(4\alpha_k+L+LA_k)\leqslant 0,\quad k=0,1,\ldots,\\1+2\alpha_k\beta_k-2\beta_k\delta_k-2\beta_k\delta_kA_k&=1+2\alpha_k\beta_k\Big(1-\tfrac{\delta_k+\delta_kA_k}{\alpha_k}\Big)\geqslant 1\quad\forall k\geqslant k_0,\end{aligned}\tag{36}$$

где k_0 — достаточно большой номер. Поэтому из (35) следует

$$\begin{aligned}\|u_{k+1}-v_k\|^2\leqslant\|u_k-v_k\|^2(1-2\alpha_k\beta_k+\beta_k\delta_k+\beta_k\delta_kA_k)&+\\+\beta_k(\delta_k+\delta_kA_k)(1+R)^2\quad\forall k\geqslant k_0.&\end{aligned}\tag{37}$$

Подставим оценки (25), (37) в (21). Получим

$$w_{k+1}\leqslant(1-s_k)w_k+d_k,\quad\forall k\geqslant k_0,\tag{38}$$

где

$$s_k=\alpha_k\beta_k\Big[1+2\alpha_k\beta_k-(1+\alpha_k\beta_k)\tfrac{\delta_k+\delta_kA_k}{\alpha_k}\Big],\tag{39}$$

$$\begin{aligned}d_k=\alpha_k\beta_k(1+\alpha_k\beta_k)\Big(&\tfrac{\delta_k+\delta_kA_k}{\alpha_k}(1+R)^2+\Big[\tfrac{|\alpha_k-\alpha_{k+1}|}{\alpha_k^2\beta_k}R+\\&+\tfrac{A_{k+1}-A_k}{2\alpha_k^2\beta_k}L(R+\|\overline{u}\|+\|P'(\overline{u})\|)\Big]^2\Big).\end{aligned}\tag{40}$$

С учетом условий (10)–(12), взяв при необходимости номер k_0 из (36) еще большим, можем считать, что

$$0<s_k\leqslant 1\quad\forall k\geqslant k_0.\tag{41}$$

Кроме того,

$$0\leqslant A_{k+1}-A_k=\tfrac{A_{k+1}-A_k}{\alpha_k^2\beta_k}\alpha_k^2\beta_k\leqslant\alpha_k\beta_k\quad\forall k\geqslant k_0.$$

Суммируя эти неравенства по k от k_0 до N, получим, что $\sum\limits_{k=k_0}^{N}\alpha_k\beta_k\geqslant A_{N+1}-A_{k_0}\to+\infty$ при $N\to\infty$, т. е. $\sum\limits_{k=0}^{\infty}\alpha_k\beta_k=+\infty$. Как следует из (39), тогда $\sum\limits_{k=0}^{\infty}s_k=+\infty$. Наконец, из условий (10)–(12) и выражений (39), (40) для s_k, d_k вытекает, что $\lim\limits_{k\to\infty}\frac{d_k}{s_k}=0$. Отсюда и из (38), (41) с помощью леммы 2.6.6 имеем $w_k=\|u_k-v_k\|\to 0$ при $k\to\infty$. Тогда, как следует из (17), (19), $\lim\limits_{k\to\infty}\|u_k-u_*\|=0$. Отсюда и из непрерывности функций $J(u)$, $g_i^+(u)$ вытекают и остальные равенства (13). Так как величины s_k, d_k из (38)–(40) не зависят от выбора реализаций $J_k'(u)$, $P_k'(u)$ из (9), то предел $\lim\limits_{k\to\infty}w_k=0$, а также все пределы из (13) равномерны относительно этих реализаций. Теорема 1 доказана. □

2. В практических задачах исходные данные как правило задаются с какой-то фиксированной погрешностью. В частности, в рассматриваемой задаче (1), (2) вместо условия (9), где $\{\delta_k\}\to 0$, практически более реальным представляется следующее условие: при каждом фиксированном $u\in U_0$ вместо точного значения градиентов $J'(u)$, $P'(u)$ известны их приближения $J_\delta'(u)$, $P_\delta'(u)$ такие, что

$$\|J_\delta'(u)-J'(u)\|\leqslant\delta(1+\|u\|),\quad\|P_\delta'(u)-P'(u)\|\leqslant\delta(1+\|u\|),\quad u\in U_0,\tag{42}$$

где $\delta>0$ — заданное число. Тогда в методе (7) приближения $J_k'(u)$, $P_k'(u)$ естественно заменить на $J_\delta'(u)$, $P_\delta'(u)$ и вместо $t_k'(u)$ из (6) использовать

$$t_k'(u)=J_\delta'(u)+A_kP_\delta'(u)+2\alpha_ku,\quad k=0,1,\ldots\tag{43}$$

Однако при фиксированном уровне погрешности $\delta>0$ условия (10)–(12) согласованного изменения параметров $\{\alpha_k\}$, $\{A_k\}$, $\{\beta_k\}$ и $\{\delta_k=\delta\}$, $k=0,1,\ldots$, заведомо будут нарушены, и поэтому процесс (7), (43) может расходиться. Возникает интересный вопрос: до какого разумного номера

$k = k(\delta)$ следует продолжать итерационный процесс (7), (43), чтобы получившуюся точку $u_{k(\delta)} = u(\delta)$ можно было принять в качестве приближения u_*, соответствующего погрешности $\delta > 0$? Оказывается, опираясь на теорему 1, можно дать ответ на этот важный для практики вопрос.

А именно, зафиксируем какую-либо начальную точку $u_0 \in U_0$ и последовательности $\{\alpha_k\}$, $\{A_k\}$, $\{\beta_k\}$, $\{\delta_k\}$, удовлетворяющие условиям (10)–(12); можем считать, что $\delta_0 \geqslant \delta$ (например, можно взять последовательности (15) при $\delta_0 = b \geqslant \delta$). Подчеркнем, что поскольку выполнение условия (9) теперь не предполагается (вместо (9) у нас имеется лишь условие (42) с фиксированным $\delta > 0$), то последовательность $\{\delta_k\}$ будем считать параметром метода (7), (43) (наряду с другими параметрами α_k, A_k, β_k), напрямую не связывая $\{\delta_k\}$ с каким-либо условием вида (9). Можно предложить следующее простое *правило останова* процесса (7), (43): при каждом δ, $0 < \delta \leqslant \delta_0$, итерации нужно продолжать до такого наибольшего номера $k = k(\delta)$, при котором выполняются неравенства

$$\delta_k \geqslant \delta, \quad k = 0, 1, \ldots, k(\delta). \tag{44}$$

Поскольку $\{\delta_k\} \to 0$, $\delta_0 \geqslant \delta$, то при любом $\delta > 0$ такой номер $k(\delta)$ непременно существует. Обоснование сформулированного правила останова (44) процесса (7), (43) дается в следующей теореме.

Теорема 2. *Пусть выполнены все условия теоремы* 1, *кроме условия* (9) *и приближения* $J_\delta'(u)$, $P_\delta'(u)$ *для* $J'(u)$, $P'(u)$ *удовлетворяют условию* (42). *Пусть точки* $u_1, u_2, \ldots, u_{k(\delta)}$ *получены методом* (7), (43), *где номер* $k(\delta)$ *определен в соответствии с правилом останова* (44). *Тогда точка* $u(\delta) = u_{k(\delta)}$ *такова, что*

$$\lim_{\delta \to 0} J(u(\delta)) = J_*, \quad \lim_{\delta \to 0} g_i^+(u(\delta)) = 0, \quad i = 1, \ldots, s, \quad \lim_{\delta \to 0} \|u(\delta) - u_*\| = 0, \tag{45}$$

причем сходимость в (45) *равномерная относительно выбора* $J_\delta'(u)$, $P_\delta'(u)$ *из* (42).

Доказательство. Из (42), (44) следует, что

$$\|J_\delta'(u) - J'(u)\| \leqslant \delta_k(1 + \|u\|), \quad \|P_\delta'(u) - P'(u)\| \leqslant \delta_k(1 + \|u\|)$$
$$\forall u \in U_0, \quad k = 0, 1, \ldots, k(\delta), \tag{46}$$

так что функции $J_k'(u) \equiv J_\delta'(u)$, $P_k'(u) = P_\delta'(u)$ удовлетворяют условию (9) при всех $k = 0, 1, \ldots, k(\delta)$. Далее, согласно правилу останова номер $k(\delta)$ является наибольшим номером, для которого выполнено условие (44). Отсюда и из $\{\delta_k\} \to 0$ следует, что $k(\delta) \to +\infty$ при $\delta \to 0$. Это значит, что при всех малых $\delta > 0$ номер $k(\delta)$ в (46) можно сделать как угодно большим.

Согласно теореме 1 при выполнении всех ее условий, включая условие (9), последовательность $\{u_k\}$, получаемая методом (6), (7), сходится в норме H к нормальному решению u_*, т. е. для любого числа $\varepsilon > 0$ найдется номер $N = N(\varepsilon)$ такой, что

$$\|u_k - u_*\| \leqslant \varepsilon \quad \forall k \geqslant N(\varepsilon), \tag{47}$$

причем номер $N(\varepsilon)$ не зависит от выбора реализаций $J_k'(u)$, $P_k'(u)$ из (9). Так как $\lim\limits_{\delta \to 0} k(\delta) = +\infty$, то существует число $\delta(\varepsilon) > 0$ такое, что $k(\delta) > N(\varepsilon)$

при всех δ, $0 < \delta < \delta(\varepsilon)$. Это значит, что для всех δ, $0 < \delta < \delta(\varepsilon)$, метод (7), (43), (44) порождает точки $z_1, \ldots, z_{k(\delta)}$, которые могут быть получены также и методом (6), (7) с реализациями $J_k'(u) = J_\delta'(u)$, $P_k'(u) = P_\delta'(u)$, $k = 0, 1, \ldots, k(\delta)$, удовлетворяющими в силу (46) условию (9). Поскольку $k(\delta) > N(\varepsilon)$, то можем воспользоваться неравенством (47) при $k = k(\delta)$ и утверждать, что $\|u_{k(\delta)} - u_*\| < \varepsilon$ при всех δ, $0 < \delta < \delta(\varepsilon)$. В силу произвольности $\varepsilon > 0$ отсюда приходим к равенству $\lim\limits_{\delta \to 0} \|u_{k(\delta)} - u_*\| = 0$. Приняв $u(\delta) = u_{k(\delta)}$, с учетом непрерывности функций $J(u)$, $P(u)$ получаем равенства (45). Теорема 2 доказана. □

Равенства (45) оправдывают сформулированное выше правило останова (44) процесса (7), (43) при фиксированном уровне погрешностей $\delta > 0$ в (42). Тем самым построен оператор R_δ, который каждому набору входных данных $(J_\delta'(u), P_\delta'(u), \delta)$ из (42) ставит в соответствие точку $u_{k(\delta)} = u(\delta)$, определяемую методом (7), (43), (44). Равенства (45) означают, что такой оператор R_δ является регуляризирующим (определение 6.1). Подчеркнем, что в этом определении оператора R_δ параметры α_k, A_k, β_k, δ_k из (10)–(12) и начальная точка u_0 предполагаются фиксированными и не меняются от изменения δ, $0 < \delta < \delta_0$.

Замечание 3. В методе (7) операцию проектирования можно выполнять с погрешностью, определяя точку u_{k+1} из условий

$$u_{k+1} \in U_0, \quad \varphi_k(u_{k+1}) \leqslant \inf_{U_0} \varphi_k(u) + \varepsilon_k^2, \quad \varepsilon_k > 0, \quad k = 0, 1, \ldots, \tag{48}$$

где $\varphi_k(u) = \|u - (u_k - \beta_k t_k'(u_k))\|^2$. Тогда получим $\|u_{k+1} - \mathcal{P}_{U_0}(u_k - \beta_k t_k'(u_k))\|^2 \leqslant \varphi_k(u_{k+1}) - \varphi_k(\mathcal{P}_{U_0}(u_k - \beta_k t_k'(u_k))) \leqslant \varepsilon_k^2$ (ср. с (5.2.8)). Для последовательности $\{u_k\}$, определяемой из (48), справедливы все утверждения теорем 1, 2, нужно лишь к (10)–(12) добавить условия: $\lim\limits_{k \to \infty} \varepsilon_k = 0$, $\lim\limits_{k \to \infty} \frac{\varepsilon_k}{\alpha_k \beta_k} = 0$.

3. Кратко остановимся на задаче

$$J(u) \to \inf, \quad u \in U, \tag{49}$$

где U — выпуклое замкнутое множество из H, известное точно (например, $U = H$). Регуляризованный метод проекции градиента (7) здесь имеет вид:

$$u_{k+1} = \mathcal{P}_U(u_k - \beta_k t_k'(u_k)), \quad t_k'(u) = J_k'(u) + 2\alpha_k u, \quad k = 0, 1, \ldots \tag{50}$$

Теорема 3. *Пусть*

1) U *— выпуклое замкнутое множество из* H, *функция* $J(u)$ *выпукла и дифференцируема по Фреше на* U; $J_* > -\infty$, $U_* \neq \infty$ *и справедливо первое из неравенств* (8) *при* $U_0 = U$;

2) *вместо точного градиента* $J'(u)$ *известно его приближение* $J_k'(u)$, *удовлетворяющее первому из неравенств* (9);

3) *числовые последовательности* $\{\alpha_k\}$, $\{\beta_k\}$, $\{\delta_k\}$ *таковы, что*

$$\alpha_k > 0, \quad \delta_k > 0, \quad 0 < \beta_k \leqslant \frac{2}{L + 4\alpha_k}, \quad k = 0, 1, \ldots, \quad \lim_{k \to \infty} (\alpha_k + \delta_k) = 0,$$
$$\sum_{k=0}^{\infty} \alpha_k \beta_k = +\infty, \quad \lim_{k \to \infty} \frac{|\alpha_k - \alpha_{k+1}|}{\alpha_k^2 \beta_k} = 0, \quad \lim_{k \to \infty} \frac{\delta_k}{\alpha_k} = 0.$$

Тогда при любом начальном приближении $u_0 \in U$ *последовательность* $\{u_k\}$, *определяемая методом* (50) *такова, что*

$$\lim_{k\to\infty} J(u_k) = J_*, \quad \lim_{k\to\infty} \|u_k - u_*\| = 0, \tag{51}$$

где u_* *— нормальное решение задачи* (49). *Сходимость в* (51) *равномерна относительно выбора реализаций* $J_k'(u)$ *из* (9).

Доказательство проводится по той же схеме, как и в теореме 1, надо лишь учесть, что здесь $m = s = 0$, $U_0 = U$, $P(u) \equiv 0$, $A_k = 0$, $R = \|u_*\|$. Из (23) вместо (25) здесь получаем: $\|v_{k+1} - v_k\| \leqslant \frac{|\alpha_k - \alpha_{k+1}|}{\alpha_k}\|u_*\|$. Равенство $\lim\limits_{k\to\infty}\|v_k - u_*\| = 0$ следует из теоремы 4.4. Интересно отметить, что в отличие от теоремы 1, где $\lim\limits_{k\to\infty}\beta_k = 0$, здесь $\lim\limits_{k\to\infty}\beta_k = \frac{2}{L} > 0$. □

Замечания 1–3 и правило останова (44) сохраняют силу и для метода (50).

Упражнения

1. Применить метод (50) к задаче минимизации квадратичного функционала $J(u) = \|Au - b\|_F^2$ при условиях (4.33).

2. Применить метод (7) или (50) к задачам из примеров 4.3–4.5, из упражнений 4.3–4.5.

3. Пусть $\{\alpha_k\}$ — последовательность положительных чисел, $\alpha_k \geqslant \alpha_{k+1}$, $k = 0, 1, \ldots$, $\lim\limits_{k\to\infty}\frac{\alpha_k - \alpha_{k+1}}{\alpha_k^2} = 0$. Доказать, что тогда $\sum\limits_{k=0}^{\infty}\alpha_k = +\infty$.

§ 9. Регуляризованный метод условного градиента

1. Будем рассматривать задачу

$$J(u) \to \inf, \quad u \in U,$$
$$U = \{u \in U_0:\ g_i(u) \leqslant 0,\ i = 1, \ldots, m;\ g_i(u) = 0,\ i = m+1, \ldots, s\}, \tag{1}$$

где U_0 — выпуклое замкнутое ограниченное множество из гильбертова пространства H, функции $J(u)$, $g_i(u)$ дифференцируемы по Фреше на U_0. Для описания и исследования регуляризованного метода условного градиента будет удобнее взять функцию Тихонова в следующем виде:

$$T_k(u) = \gamma_k J(u) + P(u) + \gamma_k\alpha_k\|u\|^2, \quad \gamma_k > 0, \quad \alpha_k > 0, \quad u \in U_0,$$

где $P(u) = \sum\limits_{i=1}^{s}(g_i^+(u))^p$ — штрафная функция множества U, $p > 1$. Если разделить функцию $T_k(u)$ на γ_k и обозначить $A_k = \frac{1}{\gamma_k}$, то придем к функции Тихонова из § 8. Пусть вместо точных функций $J(u)$, $P(u)$ и их градиентов $J'(u)$, $P'(u)$ известны их приближения $J_k(u)$, $P_k(u)$, $J_k'(u)$, $P_k'(u)$, тогда вместо функции $T_k(u)$ и ее производной

$$T_k'(u) = \gamma_k J'(u) + P'(u) + 2\gamma_k\alpha_k u, \quad u \in U_0,$$

будем иметь их приближения

$$t_k(u) = \gamma_k J_k(u) + P_k(u) + \gamma_k\alpha_k\|u\|^2,$$
$$t_k'(u) = \gamma_k J_k'(u) + P_k'(u) + 2\gamma_k\alpha_k u, \quad k = 0, 1, \ldots$$

Подчеркнем, что здесь элементы $J_k'(u)$, $P_k'(u) \in H$ необязательно являются производными функций $J(u)$, $P(u)$, так как они могут быть получены на практике в результате отдельных измерений, возможно, напрямую не связанных с измерениями самих функций $J(u)$, $P(u)$. По этой же причине $t_k'(u) \in H$ необязательно равна производной приближенной функции $t_k(u)$.

Пусть начальная точка u_0 известна. Если k-е приближение $u_k \in U_0$ при некотором $k \geqslant 0$ уже известно, то сначала определим вспомогательное приближение $\overline{u}_k$ из условий

$$\overline{u}_k \in U_0:\ \langle t_k'(u_k), \overline{u}_k - u_k\rangle \leqslant \inf_{u\in U_0}\langle t_k'(u_k), u - u_k\rangle + \varepsilon_k, \quad \varepsilon_k > 0, \tag{2}$$

и в качестве следующего приближения возьмем точку

$$u_{k+1} = u_k + \beta_k(\overline{u}_k - u_k), \quad 0 < \beta_k \leqslant 1, \tag{3}$$

где число β_k таково, что

$$t_k(u_{k+1}) \leqslant \inf_{0\leqslant\beta\leqslant 1} t_k(u_k + \beta(\overline{u}_k - u_k)) + \delta_k, \quad \delta_k > 0. \tag{4}$$

Регуляризованный метод условного градиента для задачи (1) описан. Нетрудно видеть, что метод (2)–(4) получен из одного из вариантов метода условного градиента (§ 5.4) формальной заменой $J(u)$, $J'(u)$ на $t_k(u)$, $t_k'(u)$ соответственно.

Теорема 1. *Пусть выполнены условия:*

1) U_0 *— выпуклое замкнутое ограниченное множество из гильбертова пространства* H, *функции* $J(u)$, $g_i(u)$, $i = 1, \ldots, m$, $|g_i(u)|$, $i = m+1, \ldots, s$, *выпуклы на* U_0; $J(u)$, $g_i(u)$, $i = 1, \ldots, s$, *непрерывно дифференцируемы по Фреше,* $U_* \neq \infty$, *производные* $J'(u)$, $P'(u)$ *удовлетворяют условию Липшица:*

$$\|J'(u) - J'(v)\| \leqslant L\|u - v\|, \quad \|P'(u) - P'(v)\| \leqslant L\|u - v\| \quad \forall u, v \in U_0; \tag{5}$$

задача (1) *имеет сильно согласованную постановку, т. е.*

$$J_* \leqslant J(u) + \sum_{i=1}^{s} c_i(g_i^+(u))^\nu, \quad u \in U_0, \quad c_i \geqslant 0, \quad \nu > 0;$$

параметр p *штрафной функции таков, что* $p \geqslant \nu$, $p > 1$;

2) *вместо точных* $J(u)$, $P(u)$, $J'(u)$, $P'(u)$ *известны их приближения* $J_k(u)$, $P_k(u)$, $J_k'(u)$, $P_k'(u)$ *такие, что*

$$|J_k(u) - J(u)| \leqslant \eta_k, \quad |P_k(u) - P(u)| \leqslant \eta_k, \quad u \in U_0,$$
$$\|J_k'(u) - J'(u)\| \leqslant \xi_k, \quad \|P_k'(u) - P'(u)\| \leqslant \xi_k, \quad u \in U_0, \quad k = 0, 1, \ldots; \tag{6}$$

3) *числовые последовательности* $\{\alpha_k\}$, $\{\gamma_k\}$, $\{\delta_k\}$, $\{\varepsilon_k\}$, $\{\eta_k\}$, $\{\xi_k\}$ *таковы, что*

$$\alpha_k \geqslant \alpha_{k+1} > 0, \ \gamma_k \geqslant \gamma_{k+1} > 0, \ \delta_k > 0, \ \varepsilon_k > 0, \ \eta_k > 0, \ \xi_k > 0, \ k = 0, 1, 2, \tag{7}$$

$$\lim_{k\to\infty}(\alpha_k+\gamma_k+\delta_k+\varepsilon_k+\eta_k+\xi_k)=0,\quad \lim_{k\to\infty}\alpha_k\gamma_k^{-\frac{\nu}{p-\nu}}=+\infty \tag{8}$$

(при $p=\nu>1$ последнее условие не нужно),

$$\alpha_k\gamma_k-\alpha_{k+1}\gamma_{k+1}\leqslant c_0(\gamma_k-\gamma_{k+1}),\quad \alpha_k\gamma_k\geqslant c_1k^{-\mu},\quad \sum_{k=0}^{\infty}(\delta_k+\eta_k)<\infty, \tag{9}$$

$$\gamma_k-\gamma_{k+1}+\delta_k+\varepsilon_k+\eta_k+\xi_k\leqslant c_2k^{-2\rho},\quad k=1,2,\ldots$$
$$0<\mu<p<1,\quad c_i=\text{const}>0,\quad i=0,1,2. \tag{10}$$

Тогда при любом выборе начальной точки $u_0\in U_0$ последовательность $\{u_k\}$, определяемая методом (2)–(4) такова, что

$$\lim_{k\to\infty}J(u_k)=J(u_*)=J_*,\quad \lim_{k\to\infty}g_i^+(u_k)=0,\quad i=1,\ldots,s,\quad \lim_{k\to\infty}\|u_k-u_*\|=0, \tag{11}$$

где u_* — нормальное решение задачи (1). Сходимость в (11) равномерная относительно выбора реализаций $J_k(u)$, $P_k(u)$, $J_k'(u)$, $P_k'(u)$ из условий (6).

Доказательство. При сделанных предположениях множество U выпукло, замкнуто, ограничено, $J_*>-\infty$, множество U_* непусто, выпукло, замкнуто, ограничено, нормальное решение u_* задачи (1) существует и единственно (теоремы 8.2.8, 8.2.10). Функция Тихонова сильно выпукла на U_0, поэтому условия

$$v_k\in U_0,\quad T_k(v_k)=\inf_{U_0}T_k(u) \tag{12}$$

однозначно определяют точку v_k. В теореме 4.3 положим

$$\alpha(\delta)=\alpha_k,\quad A(\delta)=\gamma_k^{-1},\quad u_\delta=v_k,\quad \delta_k\leqslant\delta<\delta_{k+1},\quad k=0,1,\ldots$$

Из (7), (8) следует, что такие функции $\alpha(\delta)$, $A(\delta)$ удовлетворяют условиям (4.51), а для точки u_δ в силу (12) выполнены условия (4.53). Отсюда и из теоремы 4.3 следует

$$\lim_{k\to\infty}J(v_k)=J(u_*)=J_*,\quad \lim_{k\to\infty}g_i^+(v_k)=0,\quad i=1,\ldots,s;\quad \lim_{k\to\infty}\|v_k-u_*\|=0. \tag{13}$$

По условию множество U_0 ограничено, так что

$$\sup_{U_0}\|u\|\leqslant D<\infty,\quad \sup_{u,\,v\in U_0}\|u-v\|\leqslant d<\infty. \tag{14}$$

Отсюда, из (5), леммы 2.6.1 имеем: $|J(u)|\leqslant|J(w)|+\|J'(w)\|\|u-w\|+\frac{1}{2}L\|u-w\|^2\leqslant|J(w)|+d\|J'(w)\|+\frac{1}{2}Ld^2$, $|P(u)|\leqslant|P(w)|+d\|P'(w)\|+\frac{1}{2}Ld^2$ $\forall u\in U_0$, где w — какая-либо фиксированная точка из U_0. Следовательно,

$$\sup_{U_0}|J(u)|=J^{**}<\infty,\quad \sup_{U_0}|P(u)|=P^{**}<\infty,$$
$$\sup_{U_0}|T_k(u)|\leqslant c_3<\infty,\quad k=1,2,\ldots, \tag{15}$$

здесь и далее через c_i, как и в (9), (10), обозначаются положительные константы, не зависящие от u, k. Далее, из условий (6) следует, что

$$|t_k(u)-T_k(u)|\leqslant\eta_k+\gamma_k\eta_k,\quad |t_k'(u)-T_k'(u)|\leqslant\xi_k+\gamma_k\xi_k$$
$$\forall u\in U_0,\quad k=0,1,\ldots \tag{16}$$

Из (9), (14), (15) получаем

$$|T_k(u)-T_{k+1}(u)|\leqslant(\gamma_k-\gamma_{k+1})|J(u)|+(\gamma_k\alpha_k-\gamma_{k+1}\alpha_{k+1})\|u\|^2\leqslant$$
$$\leqslant c_4(\gamma_k-\gamma_{k+1})\quad\forall u\in U_0 \tag{17}$$

В силу (5), (7) имеем

$$\|T_k'(u)-T_k'(v)\|\leqslant(\gamma_kL+L+2\gamma_k\alpha_k)\|u-v\|\leqslant c_5\|u-v\|$$
$$\forall u,v\in U_0,\quad k=0,1,\ldots \tag{18}$$

Введем числовую последовательность

$$a_k=T_k(u_k)-T_k(v_k),\quad k=0,1,\ldots$$

и получим для нее некоторые рекуррентные неравенства. Имеем

$$a_{k+1}=T_{k+1}(u_{k+1})-T_{k+1}(v_{k+1})=(T_{k+1}(u_{k+1})-T_k(u_{k+1}))+(T_k(u_{k+1})-T_k(u_k))+$$
$$+(T_k(u_k)-T_k(v_k))+(T_k(v_k)-T_k(v_{k+1}))+(T_k(v_{k+1})-T_{k+1}(v_{k+1})).$$

Заметим, что третье слагаемое из правой части этого равенства есть a_k, четвертое слагаемое неположительно в силу (12), а первое и пятое слагаемое можно оценить с помощью неравенства (17). Поэтому

$$0\leqslant a_{k+1}\leqslant 2c_4(\gamma_k-\gamma_{k+1})+a_k+(T_k(u_{k+1})-T_k(u_k)),\quad k=1,2,\ldots \tag{19}$$

Оценим последнее слагаемое из правой части (19). Из условия (4) и неравенств (16) имеем

$$T_k(u_{k+1})-T_k(u_k)\leqslant t_k(u_{k+1})-t_k(u_k)+2(\eta_k+\gamma_k\eta_k)\leqslant\delta_k+2(\eta_k+\gamma_k\eta_k). \tag{20}$$

Наряду с (20) нам далее понадобится также и более тонкая оценка. Для ее получения сначала установим, что

$$\lim_{k\to\infty}(T_k(u_{k+1})-T_k(u_k))=0, \tag{21}$$
$$\lim_{k\to\infty}\langle T_k'(u_k),\overline{u}_k-u_k\rangle=0. \tag{22}$$

С учетом (17), (20) имеем

$$T_{k+1}(u_{k+1})\leqslant T_k(u_{k+1})+c_4(\gamma_k-\gamma_{k+1})\leqslant$$
$$\leqslant T_k(u_k)+\delta_k+2(\eta_k+\gamma_k\eta_k)+c_4(\gamma_k-\gamma_{k+1}). \tag{23}$$

В силу (15) последовательность $\{T_k(u_k)\}$ ограничена сверху, а в силу (7), (9) получим

$$\sum_{k=1}^{\infty}[\delta_k+2(\eta_k+\gamma_k\eta_k)+c_4(\gamma_k-\gamma_{k+1})]\leqslant c_6\sum_{k=1}^{\infty}(\delta_k+\eta_k)+c_4\gamma_0<\infty.$$

Отсюда и из (23) с помощью леммы 2.6.2 заключаем, что существует конечный предел $\lim_{k\to\infty}T_k(u_k)$. Тогда $\lim_{k\to\infty}[T_k(u_k)-T_{k+1}(u_{k+1})]=0$. Отсюда и из (8), (17) имеем

$$|T_k(u_{k+1})-T_k(u_k)|\leqslant|T_k(u_{k+1})-T_{k+1}(u_{k+1})|+$$
$$+|T_{k+1}(u_{k+1})-T_k(u_k)|\to 0\text{ при }k\to\infty$$

Равенство (21) установлено.

Далее, с учетом (4), (14), (16), (18) получаем

$$T_k(u_{k+1}) - T_k(u_k) \leqslant t_k(u_{k+1}) - t_k(u_k) + 2(\eta_k + \gamma_k\eta_k) \leqslant$$
$$\leqslant t_k(u_k + \beta(\overline{u}_k - u_k)) - t_k(u_k) + \delta_k + 2(\eta_k + \gamma_k\eta_k) \leqslant T_k(u_k + \beta(\overline{u}_k - u_k)) -$$
$$- T_k(u_k) + \delta_k + 4(\eta_k + \gamma_k\eta_k) = \int_0^1 \langle T_k'(u_k + t\beta(\overline{u}_k - u_k)), \beta(\overline{u}_k - u_k)\rangle dt +$$
$$+ \delta_k + 4(\eta_k + \gamma_k\eta_k) = \int_0^1 \langle T_k'(u_k), \beta(\overline{u}_k - u_k)\rangle dt + \int_0^1 \langle T_k'(u_k + t\beta(\overline{u}_k - u_k)) -$$
$$- T_k'(u_k), \beta(\overline{u}_k - u_k)\rangle dt + \delta_k + 4(\eta_k + \gamma_k\eta_k) \leqslant$$
$$\leqslant \beta\langle T_k'(u_k), \overline{u}_k - u_k\rangle + \frac{1}{2}\beta^2 c_5 d^2 + \delta_k + 4(\eta_k + \gamma_k\eta_k)$$
$$\forall\beta,\ 0 \leqslant \beta \leqslant 1, \quad k = 0, 1, \ldots \qquad (24)$$

Из (24) следует неравенство

$$\langle T_k'(u_k), \overline{u}_k - u_k\rangle \geqslant \frac{1}{\beta}(T_k(u_{k+1}) - T_k(u_k)) - \frac{1}{2}\beta c_5 d^2 -$$
$$- \frac{1}{\beta}(\delta_k + 4(\eta_k + \gamma_k\eta_k)) \quad \forall\beta,\ 0 \leqslant \beta \leqslant 1,\ k = 0, 1, \ldots$$

Переходя здесь к пределу сначала при $k \to \infty$, затем при $\beta \to +0$, с учетом уже доказанного равенства (21) получим

$$\lim_{k\to\infty} \langle T_k'(u_k), \overline{u}_k - u_k\rangle \geqslant 0. \qquad (25)$$

С другой стороны, в силу условия (2) и неравенств (14), (16) имеем

$$\langle T_k'(u_k), \overline{u}_k - u_k\rangle \leqslant \langle t_k'(u_k), \overline{u}_k - u_k\rangle + (\xi_k + \gamma_k\xi_k)d \leqslant \varepsilon_k + (\xi_k + \gamma_k\xi_k)d;$$

здесь учтено, что $\inf\limits_{u\in U_0} \langle t_k'(u_k), u - u_k\rangle \leqslant \langle t_k'(u_k), u_k - u_k\rangle = 0$. Отсюда и из (8) следует, что

$$\overline{\lim_{k\to\infty}} \langle T_k'(u_k), \overline{u}_k - u_k\rangle \leqslant 0.$$

Совмещая это неравенство с (25), приходим к равенству (22).

Далее, из выпуклости $T_k(u)$ (теорема 4.2.2), из (2), (14), (16) имеем

$$0 \leqslant a_k = T_k(u_k) - T_k(v_k) \leqslant \langle T_k'(u_k), u_k - v_k\rangle \leqslant \langle t_k'(u_k), u_k - v_k\rangle +$$
$$+ (\xi_k + \gamma_k\xi_k)d \leqslant -\langle t_k'(u_k), \overline{u}_k - u_k\rangle + \varepsilon_k + (\xi_k + \gamma_k\xi_k)d \leqslant$$
$$\leqslant -\langle T_k'(u_k), \overline{u}_k - u_k\rangle + \varepsilon_k + 2(\xi_k + \gamma_k\xi_k)d$$

или

$$a_k - \varepsilon_k - 2(\xi_k + \gamma_k\xi_k)d \leqslant -\langle T_k'(u_k), \overline{u}_k - u_k\rangle, \quad k = 0, 1, \ldots \qquad (26)$$

Множество натуральных чисел разобьем на два класса:

$$I_0 = \{k\colon a_k > \varepsilon_k + 2(\xi_k + \gamma_k\xi_k)d\}, \quad I_1 = \{k\colon a_k \leqslant \varepsilon_k + 2(\xi_k + \gamma_k\xi_k)d\}.$$

Пусть сначала $k \in I_0$. Тогда в силу (26) получаем $0 < -\langle T_k'(u_k), \overline{u}_k - u_k\rangle = |\langle T_k'(u_k), \overline{u}_k - u_k\rangle|$, и из неравенства (24) имеем

$$T_k(u_{k+1}) - T_k(u_k) \leqslant -\beta|\langle T_k'(u_k), \overline{u}_k - u_k\rangle| + \frac{1}{2}\beta^2 c_5 d^2 + \delta_k + 4(\eta_k + \gamma_k\eta_k)$$
$$\forall\beta, \quad 0 \leqslant \beta \leqslant 1, \quad \forall k \in I_0. \qquad (27)$$

Заметим, что функция $\varphi_k(\beta) = -\beta|\langle T_k'(u_k), \overline{u}_k - u_k\rangle| + \frac{1}{2}\beta^2 c_5 d^2$ достигает на числовой оси своего минимума в точке $\beta_{k*} = \frac{1}{c_5 d^2}|\langle T_k'(u_k), \overline{u}_k - u_k\rangle|$. В силу (22) $\lim\limits_{k\to\infty} \beta_{k*} = 0$. Поэтому, беря при необходимости постоянную c_5 еще большей, можем считать, что $0 \leqslant \beta_{k*} \leqslant 1 \ \ \forall k \geqslant 0$. Полагая $\beta = \beta_{k*}$, из (27) тогда получим

$$T_k(u_{k+1}) - T_k(u_k) \leqslant -\frac{1}{2c_5 d^2}|\langle T_k'(u_k), \overline{u}_k - u_k\rangle|^2 + \delta_k + 4(\eta_k + \gamma_k\eta_k) \quad \forall k \in I_0. \qquad (28)$$

Так как в силу (15) $0 \leqslant a_k \leqslant 2c_3 = c_7$, то из (26) для $k \in I_0$ имеем

$$|\langle T_k'(u_k), \overline{u}_k - u_k\rangle|^2 \geqslant (a_k - \varepsilon_k - 2d(\xi_k + \gamma_k\xi_k))^2 \geqslant$$
$$\geqslant a_k^2 - 2c_7[2d(\xi_k + \gamma_k\xi_k) + \varepsilon_k] \geqslant a_k^2 - c_8(\varepsilon_k + \xi_k), \quad k \in I_0.$$

Подставив эту оценку в (28), получим

$$T_k(u_{k+1}) - T_k(u_k) \leqslant -\frac{1}{2c_5 d^2}a_k^2 + \frac{c_8}{2c_5 d^2}(\varepsilon_k + \xi_k) + \delta_k + 4(\eta_k + \gamma_k\eta_k), \quad k \in I_0.$$

Отсюда с учетом условия (10) из (19) имеем

$$0 \leqslant a_{k+1} \leqslant a_k - a_k^2\frac{1}{2c_5 d^2} + c_9(\varepsilon_k + \xi_k + \delta_k + \eta_k + \gamma_k - \gamma_{k+1}) \leqslant$$
$$\leqslant a_k - a_k^2\frac{1}{c_{10}} + c_{11}k^{-2\rho} \leqslant a_k - a_k^2\frac{1}{A} + Ak^{-2\rho}, \quad A = \max\{c_{10}; c_{11}\}, \quad k \in I_0. \qquad (29)$$

Пусть теперь $k \in I_1$. Тогда подставив в (19) оценку (20), с учетом (10) получим

$$a_{k+1} \leqslant a_k + c_{12}(\gamma_k - \gamma_{k+1} + \delta_k + \eta_k) \leqslant a_k + c_{13}k^{-2\rho}, \quad k \in I_1. \qquad (30)$$

Кроме того, по определению I_1

$$0 \leqslant a_k \leqslant \varepsilon_k + 2(\xi_k + \gamma_k\xi_k)d \leqslant c_{14}k^{-2\rho} \quad \forall k \in I_1. \qquad (31)$$

Таким образом, последовательность $\{a_k\}$ удовлетворяет условиям (29)–(31). Отсюда и из леммы 2.6.5 следует, что

$$0 \leqslant a_k = T_k(u_k) - T_k(v_k) \leqslant c_{15}k^{-\rho}, \quad k = 1, 2, \ldots$$

Из (12), теоремы 8.2.10 тогда имеем

$$\alpha_k\gamma_k\|u_k - v_k\|^2 \leqslant T_k(u_k) - T_k(v_k) = a_k \leqslant c_{15}k^{-\rho}, \quad k = 1, 2, \ldots$$

Отсюда с учетом второго неравенства (9) получим

$$\|u_k - v_k\|^2 \leqslant c_{15}k^{-\rho}\frac{1}{\alpha_k\gamma_k} \leqslant c_{16}k^{-\rho+\mu}, \quad k = 1, 2, \ldots, \quad 0 < \mu < \rho < 1. \qquad (32)$$

Из (13), (32) следуют все утверждения теоремы 1. □

В качестве последовательностей, удовлетворяющих условиям (7)–(10), можно, например, взять

$$\begin{gathered}\alpha_k = a_1(k+1)^{-\alpha}, \quad \gamma_k = a_2(k+1)^{-\gamma}, \quad \delta_k = a_3(k+1)^{-\delta}, \\ \varepsilon_k = a_4(k+1)^{-\varepsilon}, \quad \eta_k = a_5(k+1)^{-\chi_1}, \quad \xi_k = a_6(k+1)^{-\chi_2}, \quad k=0,1,\ldots \\ a_i>0, \quad i=1,\ldots,6; \quad 0<\alpha<\gamma\frac{\nu}{p-1}, \quad \gamma>0, \quad \alpha+\gamma<\mu<\rho<1/2, \\ 0<\delta\leqslant 1, \quad 0<\chi_1\leqslant 1, \quad 2\rho<\min\{1+\gamma;\,\delta;\,\varepsilon;\,\chi_1;\,\chi_2\}.\end{gathered} \tag{33}$$

2. Сформулируем правило останова в методе (2)–(4) для случая, когда уровень погрешности в задании исходных данных фиксирован. А именно, пусть при каждом $u\in U_0$ вместо точных значений $J(u)$, $P(u)$, $J'(u)$, $P'(u)$ известны их приближения $J_\eta(u)$, $P_\eta(u)$, $J'_\xi(u)$, $P'_\xi(u)$ такие, что

$$\begin{gathered}|J_\eta(u)-J(u)|\leqslant\eta, \quad |P_\eta(u)-P(u)|\leqslant\eta, \quad u\in U_0, \\ |J'_\xi(u)-J'(u)|\leqslant\xi, \quad |P'_\xi(u)-P'(u)|\leqslant\xi, \quad u\in U_0,\end{gathered} \tag{34}$$

где $\eta>0$, $\xi>0$ — фиксированные числа. Тогда в методе (2)–(4) можно взять

$$t_k(u)=\gamma_k J_\eta(u)+P_\eta(u)+\gamma_k\alpha_k\|u\|^2, \quad t'_k(u)=\gamma_k J'_\xi(u)+P'_\xi(u)+2\gamma_k\alpha_k u. \tag{35}$$

Возникает вопрос: до какого разумного номера $k=k(\eta,\xi)$ следует продолжать процесс (2)–(4), (35), чтобы получившуюся точку $u(\eta,\xi)=u_{k(\eta,\xi)}$ можно было принять в качестве приближения к нормальному решению u_* задачи (1)? Для ответа на этот вопрос зафиксируем какую-либо начальную точку $u_0\in U_0$ и последовательности, удовлетворяющие условиям (7)–(10), считая, что $\eta_0\geqslant\eta$, $\xi_0\geqslant\xi$ (например, можно взять последовательности (33) при $a_5\geqslant\eta$, $a_6\geqslant\xi$). Рассмотрим следующее правило останова процесса (2)–(4), (35): при каждом фиксированном η, ξ, $0<\eta\leqslant\eta_0$, $0<\xi\leqslant\xi_0$, итерации будем продолжать до такого наибольшего номера $k=k(\eta,\xi)$, при котором выполняются неравенства

$$\eta_k\geqslant\eta, \quad \xi_k\geqslant\xi, \quad k=0,1,\ldots,k(\eta,\xi). \tag{36}$$

Теорема 2. *Пусть выполнены все условия теоремы* 1, *кроме условия* (6); *пусть выполнены условия* (34). *Пусть точки* $u_1,u_2,\ldots,u_k$, $k=k(\eta,\xi)$, *получены методом* (2)–(4), (35), (36). *Тогда точка* $u(\delta)=u_{k(\delta)}$, $\delta=(\eta,\xi)$ *такова, что*

$$\lim_{\delta\to 0}J(u(\delta))=J_*, \quad \lim_{\delta\to 0}g_i^+(u(\delta))=0, \quad i=1,\ldots,s; \quad \lim_{\delta\to 0}\|(u(\delta)-u_*\|=0.$$

Доказательство опирается на теорему 1 и проводится также, как и аналогичная теорема 2 из § 8. □

Из теоремы 2 следует, что оператор R_δ, который каждому набору $(J_\eta(u), P_\eta(u), J'_\xi(u), P'_\xi(u), \delta=(\eta,\xi))$ входных данных из (34) ставит в соответствие точку $u(\delta)=(\eta,\xi)=u_{k(\eta,\xi)}$, определяемую методом (2)–(4), (35), (36), является регуляризирующим оператором задачи (1) в метрике H (определение 6.1).

Упражнения

1. Для задачи: $J(u)\to\inf$, $u\in U$, считая, что множество U известно точно, описать регуляризованный метод условного градиента, сформулировать и доказать аналоги теорем 1, 2.

2. Применить регуляризованный метод условного градиента к задаче минимизации квадратичного функционала $J(u)=\|Au-b\|_F^2$ при условиях (4.33) и к задачам из примеров 4.3–4.5, из упражнений 4.3–4.5.

§ 10. Регуляризованный проксимальный метод

1. Рассмотрим задачу минимизации

$$J(u)\to\inf, \quad u\in U, \tag{1}$$

где U — выпуклое замкнутое множество из некоторого гильбертова пространства H, функция $J(u)$ выпукла и полунепрерывна снизу на U. Введем функцию Тихонова

$$T_k(u)=J(u)+\alpha_k\|u\|^2, \quad \alpha_k>0, \quad k=0,1,\ldots$$

В качестве начальной возьмем какую-либо точку $u_0\in U$. Пусть известно k-е приближение $u_k\in U$ при некотором $k\geqslant 0$. Составим функцию

$$\Phi_k(u)=\tfrac{1}{2}\|u-u_k\|^2+\beta_k T_k(u) \tag{2}$$

и определим следующее приближение u_{k+1} из условия

$$u_{k+1}\in U, \quad \Phi_k(u_{k+1})=\inf_{u\in U}\Phi_k(u). \tag{3}$$

Нетрудно видеть, что метод (2), (3) получен из проксимального метода (см. § 5.6) заменой целевой функции $J(u)$ на функцию Тихонова $T_k(u)$. Так как функция $\Phi_k(u)$ сильно выпукла и полунепрерывна снизу на выпуклом и замкнутом множестве U, то точка u_{k+1} условием (3) определяется однозначно (теорема 8.2.10).

Теорема 1. *Пусть* U *— выпуклое замкнутое множество из некоторого гильбертова пространства* H, *функция* $J(u)$ *выпукла и полунепрерывна снизу на* U; $J_*>-\infty$, $U_*\neq\varnothing$. *Параметры* α_k, β_k *таковы, что*

$$\begin{gathered}\alpha_k>0, \quad \beta_k>0, \quad \sup_{k\geqslant 0}\beta_k<\infty, \quad \sum_{k=0}^{\infty}\alpha_k\beta_k=+\infty, \\ \lim_{k\to\infty}\alpha_k=0, \quad \lim_{k\to\infty}\frac{|\alpha_k-\alpha_{k+1}|}{\alpha_k^3\beta_k^2}=0, \quad \lim_{k\to\infty}\frac{\alpha_k}{\alpha_{k+1}}=c_0>0.\end{gathered} \tag{4}$$

Тогда последовательность $\{u_k\}$, *определяемая методом* (2), (3) *при любом выборе* $u_0\in U$ *обладает свойством*

$$\lim_{k\to\infty}J(u_k)=J_*, \quad \lim_{k\to\infty}\|u_k-u_*\|=0, \tag{5}$$

где u_* *— нормальное решение задачи* (1).

Доказательство. При сделанных предположениях нормальное решение u_* задачи (1) существует и определяется однозначно (теорема 8.2.10). Рассмотрим последовательность $\{v_k\}$, определяемую условием

$$v_k \in U, \quad T_k(v_k) = \inf_U T_k(u), \quad k = 0, 1, \ldots \tag{6}$$

Согласно теореме 4.4

$$\|v_k\| \leqslant \|u_*\|, \quad k = 0, 1, \ldots, \quad \lim_{k\to\infty} J(v_k) = J_*, \quad \lim_{k\to\infty} \|v_k - u_*\| = 0. \tag{7}$$

Поскольку

$$\|u_k - u_*\| \leqslant \|u_k - v_k\| + \|v_k - u_*\|, \quad k = 0, 1, \ldots, \tag{8}$$

то для доказательства теоремы остается доказать, что $\omega_k = \|u_k - v_k\| \to 0$ при $k \to \infty$.

Убедимся, что величины ω_k удовлетворяют неравенствам

$$0 \leqslant \omega_{k+1} \leqslant (1 - s_k)\omega_k + d_k, \quad k = 0, 1, \ldots, \tag{9}$$

где

$$s_k = 1 - (1 + 2\alpha_k\beta_k)^{-1/2}, \quad d_k = \left(\frac{|\alpha_k - \alpha_{k+1}|}{\alpha_k + \alpha_{k+1}}\right)^{1/2} \|u_*\|. \tag{10}$$

Заметим, что

$$\omega_{k+1} = \|u_{k+1} - v_{k+1}\| \leqslant \|v_{k+1} - v_k\| + \|v_k - u_{k+1}\|, \quad k = 0, 1, \ldots \tag{11}$$

Оценим сверху первое слагаемое из правой части неравенства (11). Так как функция $T_k(u)$ сильно выпукла с константой сильной выпуклости $\varkappa = 2\alpha_k > 0$, то из (6) и теоремы 8.2.10 получаем

$$\alpha_k \|u - v_k\|^2 \leqslant T_k(u) - T_k(v_k) \quad \forall u \in U.$$

В частности, при $u = v_{k+1}$ отсюда имеем

$$\alpha_k \|v_{k+1} - v_k\|^2 \leqslant T_k(v_{k+1}) - T_k(v_k).$$

Аналогично устанавливается неравенство

$$\alpha_{k+1} \|v_k - v_{k+1}\|^2 \leqslant T_{k+1}(v_k) - T_{k+1}(v_{k+1}).$$

Сложим два последних неравенства и получим

$$(\alpha_k + \alpha_{k+1})\|v_k - v_{k+1}\|^2 \leqslant (\alpha_k - \alpha_{k+1})(\|v_{k+1}\|^2 - \|v_k\|^2) \leqslant$$
$$\leqslant |\alpha_k - \alpha_{k+1}| \max\{\|v_k\|^2, \|v_{k+1}\|^2\}.$$

Отсюда с учетом (7) имеем

$$\|v_k - v_{k+1}\| \leqslant \left(\frac{|\alpha_k - \alpha_{k+1}|}{\alpha_k + \alpha_{k+1}}\right)^{1/2} \|u_*\|, \quad k = 0, 1, \ldots \tag{12}$$

Оценим второе слагаемое из правой части (11). Функция (2) сильно выпукла с константой сильной выпуклости $\varkappa = 1 + 2\alpha_k\beta_k$, то из (3) и теоремы 8.2.10 следует, что

$$\left(\frac{1}{2} + \alpha_k\beta_k\right)\|u - u_{k+1}\|^2 \leqslant \Phi_k(u) - \Phi_k(u_{k+1}) \quad \forall u \in U. \tag{13}$$

Положим в (13) $u = v_k$. С учетом (6) имеем

$$\left(\frac{1}{2} + \alpha_k\beta_k\right)\|v_k - u_{k+1}\|^2 \leqslant \Phi_k(v_k) - \Phi_k(u_{k+1}) \leqslant$$
$$\leqslant \tfrac{1}{2}\|v_k - u_k\|^2 - \tfrac{1}{2}\|u_{k+1} - u_k\|^2 + \beta_k(T_k(v_k) - T_k(u_{k+1})) \leqslant \tfrac{1}{2}\|v_k - u_k\|^2$$

или

$$\|v_k - u_{k+1}\| \leqslant (1 + 2\alpha_k\beta_k)^{-\frac{1}{2}}\|v_k - u_k\| \quad k = 0, 1, \ldots \tag{14}$$

Подставив оценки (12), (14) в (11), получаем неравенства (9) с величинами s_k, d_k из (10). Из условий (4) и формул (10) для s_k, d_k вытекает, что $0 < s_k < 1$, $\lim_{k\to\infty} \frac{s_k}{\alpha_k\beta_k} = 1$, $\sum_{k=0}^{\infty} s_k = +\infty$, $\lim_{k\to\infty} \frac{d_k}{s_k} = 0$. Тогда $\omega_k = \|v_k - u_k\| \to 0$ при $k \to \infty$ (лемма 2.6.6). Отсюда и из (7), (8) следуют равенства (5). Теорема 1 доказана. □

2. Теперь изложим регуляризованный проксимальный метод в предположении, что вместо точных значений функции $J(u)$ нам известны приближения $J_k(u)$, удовлетворяющие неравенству

$$|J_k(u) - J(u)| \leqslant \delta_k(1 + \|u\|^2) \quad \forall u \in U, \quad \delta_k > 0, \quad k = 0, 1, \ldots \tag{15}$$

Пусть при некотором $k \geqslant 0$ известна точка $z_k \in U$. Введем функцию

$$\varphi_k(u) = \tfrac{1}{2}\|u - z_k\|^2 + \beta_k(J_k(u) + \alpha_k\|u\|^2), \quad u \in U. \tag{16}$$

Допустим, что приближенно решая задачу минимизации: $\varphi_k(u) \to \inf$, $u \in U$ каким-либо методом нам удалось определить точку z_{k+1} из условия

$$z_{k+1} \in U, \quad \varphi_k(z_{k+1}) \leqslant \inf_U \varphi_k(u) + \varepsilon_k, \quad \varepsilon_k > 0. \tag{17}$$

Теорема 2. *Пусть выполнены условия теоремы* 1, *неравенства* (15), *параметры* α_k, β_k, δ_k, ε_k *положительны и наряду с* (4) *удовлетворяют условиям*

$$\lim_{k\to\infty} \frac{\delta_k}{\beta_k\alpha_k^2} = 0, \quad \lim_{k\to\infty} \frac{\varepsilon_k}{\beta_k^2\alpha_k^2} = 0, \quad \lim_{k\to\infty} \frac{\beta_k}{\beta_{k+1}} = c_1 > 0, \quad \lim_{k\to\infty} \frac{\delta_k}{\delta_{k+1}} = c_2 > 0,$$
$$\lim_{k\to\infty} \frac{\varepsilon_k}{\varepsilon_{k+1}} = c_3 > 0, \quad \alpha_k \geqslant 2\delta_k, \quad k = 0, 1, \ldots, \quad \lim_{k\to\infty} \delta_k = \lim_{k\to\infty} \varepsilon_k = 0. \tag{18}$$

Тогда последовательность $\{z_k\}$, *определяемая методом* (16), (17) *при любом выборе начального приближения* $z_0 \in U$, *сходится в норме* H *к нормальному решению* u_* *задачи* (1), $\lim_{k\to\infty} J(z_k) = J_*$, *равномерно относительно выбора реализаций* $J_k(u)$ *из* (15), *точек* z_{k+1} *из* (17).

Доказательство. Из (15), (16), (18) имеем

$$\varphi_k(u) \geqslant \tfrac{1}{2}\|u-z_k\|^2+\beta_k(J(u)+\alpha_k\|u\|^2)-\beta_k\delta_k(1+\|u\|^2) \equiv h_k(u), \quad u \in U,$$

где функция $h_k(u)$ сильно выпукла на U с константой сильной выпуклости $\varkappa_k = 1 + 2\beta_k(\alpha_k - \delta_k) \geqslant 1$. Поэтому $\inf\limits_U \varphi_k(u) \geqslant \inf\limits_U h_k(u) > -\infty$ и при любом $\varepsilon_k > 0$ существует точка z_{k+1}, удовлетворяющая условиям (17). Пусть $\{z_k\}$ — произвольная последовательность, полученная методом (16), (17) при каком-угодно выборе функции $J_k(u)$ из условия (15). Введем вспомогательную последовательность $\{w_k\}$, определяемую условием

$$w_{k+1} \in U, \quad \psi_k(w_{k+1}) = \inf_U \psi_k(u), \tag{19}$$

где $\psi_k(u) = \frac{1}{2}\|u - z_k\|^2 + \beta_k(J(u) + \alpha_k\|u\|^2)$. Так как функция $\psi_k(u)$ сильно выпукла на U с константой сильной выпуклости $\varkappa_k = 1 + 2\alpha_k\beta_k$, то условие (19) однозначно определяет точку w_{k+1}, причем (теорема 8.2.10)

$$\left(\tfrac{1}{2} + \alpha_k\beta_k\right)\|u - w_{k+1}\|^2 \leqslant \psi_k(u) - \psi_k(w_{k+1}) \quad \forall u \in U. \tag{20}$$

Кроме того, к рассуждениям привлечем также последовательность $\{u_k\}$, определяемую условиями (2), (3) при $u_0 = z_0$. Тогда

$$\|z_k - u_*\| \leqslant \|z_k - w_k\| + \|w_k - u_k\| + \|u_k - u_*\|, \quad k = 0, 1, \ldots \tag{21}$$

В силу (5) $\|u_k - u_*\| \to 0$ при $k \to \infty$. Остается убедиться, что первое и второе слагаемые из правой части (21) также стремятся к нулю. С этой целью положим в (20) $u = u_{k+1}$, в (13) — $u = w_{k+1}$ и сложим получившиеся неравенства:

$$\begin{aligned}(1+2\alpha_k\beta_k)\|u_{k+1}-w_{k+1}\|^2 &\leqslant \psi_k(u_{k+1})-\psi_k(w_{k+1})+\Phi_k(w_{k+1})-\Phi_k(u_{k+1}) = \\ &= \tfrac{1}{2}(\|u_{k+1} - z_k\|^2 - \|w_{k+1} - z_k\|^2 + \|w_{k+1} - u_k\|^2 - \|u_{k+1} - u_k\|^2) = \\ &= \langle u_{k+1} - w_{k+1}, u_k - z_k\rangle, \quad k = 0, 1, \ldots\end{aligned}$$

Отсюда, пользуясь элементарными неравенствами $|ab| \leqslant \frac{1}{2}(a^2+b^2)$, $(a+b)^2 \leqslant (1+\varepsilon)\left(a^2 + \frac{b^2}{\varepsilon}\right)$, $\forall a, b, \varepsilon > 0$, имеем

$$\begin{aligned}(1+2\alpha_k\beta_k)\|u_{k+1}-w_{k+1}\|^2 &\leqslant \tfrac{1}{2}\|u_{k+1}-w_{k+1}\|^2+\tfrac{1}{2}(\|u_k-w_k\|+\|w_k-z_k\|)^2 \leqslant \\ &\leqslant \tfrac{1}{2}\|u_{k+1}-w_{k+1}\|^2+\tfrac{1}{2}(1+2\alpha_k\beta_k)(\|u_k-w_k\|^2+\tfrac{1}{\alpha_k\beta_k}\|w_k-z_k\|^2)\end{aligned}$$

или

$$\begin{gathered}(1+4\alpha_k\beta_k)\|u_{k+1}-w_{k+1}\|^2 \leqslant (1+\alpha_k\beta_k)(\|u_k-w_k\|^2+\tfrac{1}{\alpha_k\beta_k}\|w_k-z_k\|^2), \\ k = 0, 1, \ldots\end{gathered} \tag{22}$$

Далее, из (20) при $u = z_{k+1}$ с учетом (15)–(17) получим

$$\begin{aligned}\left(\tfrac{1}{2} + \alpha_k\beta_k\right)\|z_{k+1} - w_{k+1}\|^2 &\leqslant \psi_k(z_{k+1}) - \psi_k(w_{k+1}) \leqslant \\ &\leqslant \varphi_k(z_{k+1}) - \varphi_k(w_{k+1}) + \beta_k\delta_k(2 + \|z_{k+1}\|^2 + \|w_{k+1}\|^2) \leqslant \\ &\leqslant \varepsilon_k+2\beta_k\delta_k+\beta_k\delta_k(2\|z_{k+1}-w_{k+1}\|^2+3\|w_{k+1}\|^2)\leqslant\varepsilon_k+2\beta_k\delta_k+ \\ +2\beta_k\delta_k\|z_{k+1}-w_{k+1}\|^2+6\beta_k\delta_k\|w_{k+1}-u_{k+1}\|^2+6\beta_k\delta_k\|u_{k+1}\|^2, &\quad k = 0, 1, \ldots\end{aligned} \tag{23}$$

Так как в силу теоремы 1 имеем $\lim\limits_{k\to\infty}\|u_k - u_*\| = 0$, то $\sup\limits_{k\geqslant 0}\|u_k\| = R < \infty$. Тогда из неравенства (23) с учетом условия $\alpha_k \geqslant 2\delta_k$ получим

$$\|z_{k+1}-w_{k+1}\|^2 \leqslant 2\varepsilon_k+4\beta_k\delta_k(1+3R^2)+12\beta_k\delta_k\|w_{k+1}-u_{k+1}\|^2, \quad k=0,1,\ldots$$

или

$$\|z_k-w_k\|^2 \leqslant 2\varepsilon_{k-1}+4\beta_{k-1}\delta_{k-1}(1+3R^2)+12\beta_{k-1}\delta_{k-1}\|w_k-u_k\|^2, \quad k=0,1,\ldots \tag{24}$$

Подставим оценку (24) в правую часть (22). Получим

$$\begin{aligned}(1+4\alpha_k\beta_k)\|u_{k+1}-w_{k+1}\|^2 \leqslant &\left(1+\alpha_k\beta_k+12\beta_{k-1}\delta_{k-1}(1+\alpha_k\beta_k)\tfrac{1}{\alpha_k\beta_k}\right)\|u_k-w_k\|^2+ \\ &+(2\varepsilon_{k-1}+4\beta_{k-1}\delta_{k-1}(1+3R^2))(1+\alpha_k\beta_k)\tfrac{1}{\alpha_k\beta_k}, \quad k=1,2,\ldots\end{aligned}$$

Отсюда заключаем, что величина $\omega_k = \|u_k - w_k\|^2$ также удовлетворяет неравенству (9), где

$$\begin{aligned}s_k &= \frac{3\alpha_k\beta_k}{1+4\alpha_k\beta_k}\left(1-\frac{4\beta_{k-1}\delta_{k-1}}{\alpha_k^2\beta_k^2}(1+\alpha_k\beta_k)\right), \\ d_k &= \alpha_k\beta_k(1+\alpha_k\beta_k)\left(\frac{2\varepsilon_{k-1}}{\alpha_k^2\beta_k^2}+\frac{4\beta_{k-1}}{\beta_k}\frac{\delta_{k-1}}{\beta_k\alpha_k^2}(1+3R^2)\right), \ k=1,2,\ldots\end{aligned} \tag{25}$$

Из условий (4), (18) следует, что $0<s_k\leqslant 1$ для всех достаточно больших номеров k. Поскольку $\lim\limits_{k\to\infty}\frac{s_k}{\alpha_k\beta_k}=3$, то из $\sum\limits_{k=0}^{\infty}\alpha_k\beta_k=+\infty$ вытекает, что $\sum\limits_{k=0}^{\infty}s_k=+\infty$. Кроме того, $\lim\limits_{k\to\infty}\frac{d_k}{s_k}=\lim\limits_{k\to\infty}\frac{d_k}{\alpha_k\beta_k}=0$. Тогда $\omega_k = \|u_k - w_k\|^2 \to 0$ при $k \to \infty$ (лемма 2.6.6), а из (24) имеем $\|z_k - w_k\| \to 0$ при $k \to \infty$. Поэтому из (5), (21) получим

$$\lim_{k\to\infty}\|z_k - u_*\| = 0, \quad \lim_{k\to\infty} J(z_k) = J_*.$$

Заметим, что сходимость здесь равномерная относительно выбора реализаций $J_k(u)$ из (15), точек z_{k+1} из (17), так как величины α_k, β_k, δ_k, ε_k в (9), (10), (24), (25) от перечисленных реализаций не зависят. Таким образом, теорема 2 доказана. □

В качестве последовательностей, удовлетворяющих условиям (4), (18), можно, например, взять

$$\beta_k=1, \quad \alpha_k=2c(k+1)^{-\alpha}, \quad \delta_k=c(k+1)^{-\delta}, \quad \varepsilon_k=(k+1)^{-\varepsilon}, \quad k=0,1,\ldots \tag{26}$$

где $\delta > 0$, $\varepsilon > 0$, $0 < 2\alpha < \min\{\delta; \varepsilon; 1\}$, $c > 0$.

3. В практических задачах (1) более реальным, чем (15), представляется следующее условие: при каждом фиксированном $u \in U$ вместо точного значения $J(u)$ может быть вычислено его приближение $J_\delta(u)$ такое, что

$$|J_\delta(u) - J(u)| \leqslant \delta(1 + \|u\|^2), \quad u \in U, \tag{27}$$

где $\delta > 0$ известное число. Тогда в методе (16), (17) приближение $J_k(u)$ естественно заменить на $J_\delta(u)$ и вместо функции $\varphi_k(u)$ из (16) пользоваться функцией

$$\varphi_k(u) = \tfrac{1}{2}\|u - z_k\| + \beta_k(J_\delta(u) + \alpha_k\|u\|^2), \quad u \in U, \quad k = 0, 1, \ldots \tag{28}$$

Сформулируем правило останова процесса (17), (28) в зависимости от уровня погрешности δ в (27). Будем считать, что зафиксированы какая-либо начальная точка и параметры α_k, β_k, δ_k, ε_k, удовлетворяющие условиям (4), (18) (например, как в (26)) и пусть $\delta < \delta_0$. Тогда процесс (17), (28) можно продолжать до такого наибольшего номера $k = k(\delta)$, при котором выполняются неравенства

$$\delta_k \geqslant \delta, \quad k = 0, 1, \ldots, k(\delta). \tag{29}$$

Поскольку $\{\delta_k\} \to 0$, $\delta_0 > \delta$, то такой номер $k(\delta)$ непременно найдется. Оправданием сформулированного правила останова (29) процесса (17), (28) служит

Теорема 3. *Пусть выполнены все условия теорем* 1, 2, *кроме условия* (15), *пусть приближение* $J_\delta(u)$ *функции* $J(u)$ *удовлетворяет условию* (27). *Тогда точка* $u(\delta) = z_{k(\delta)}$, *полученная методом* (17), (28), (29), *обладает свойством*

$$\lim_{\delta \to 0} J(u(\delta)) = J_*, \quad \lim_{\delta \to 0} \|u(\delta) - u_*\| = 0,$$

где u_* — *нормальное решение задачи* (1).

Доказательство опирается на теорему 2 и проводится также, как и аналогичная теорема 2 из § 8. □

Из теоремы 3 следует, что оператор R_δ, который каждому набору $(J_\delta(u), \delta)$ входных данных из (27) ставит в соответствие точку $u(\delta) = z_{k(\delta)}$, определяемую методом (17), (28), (29), является регуляризирующим оператором задачи (1) в метрике H (определение 6.1).

Заметим, что для задачи (1), когда множество U имеет вид (8.2) и задано с погрешностью, аналогичный регуляризованный проксимальный метод в сочетании со штрафными функциями был исследован в [168]. Другие более тонкие варианты регуляризованного проксимального метода, в которых для решения задач вида (16) используются те или иные конкретные методы минимизации, изучались в [799].

§ 11. Регуляризованный метод Ньютона

1. Рассмотрим задачу:

$$J(u) \to \inf, \quad u \in U = H, \tag{1}$$

где H — гильбертово пространство. Введем функцию Тихонова

$$T_k(u) = J(u) + \alpha_k\|u\|^2, \; u \in U.$$

Если $J(u) \in C^2(H)$, то $T_k(u) \in C^2(H)$, причем

$$T_k'(u) = J'(u) + 2\alpha_k u, \quad T_k''(u) = J''(u) + 2\alpha_k I,$$

где I — единичный оператор в H. Применяя k-й шаг метода (5.10.8) к функции $T_k(u)$, получим регуляризованный метод Ньютона для задачи (1) с точными данными

$$u_{k+1} = u_k - (T_k''(u_k))^{-1} T_k'(u_k), \quad k = 0, 1, \ldots \tag{2}$$

Пусть вместо точных $J'(u)$, $J''(u)$ известны их приближения $J_k'(u) \in H$, $J_k''(u) \in \mathcal{L}(H \to H)$ такие, что

$$\|J_k'(u) - J'(u)\| \leqslant \delta_{1k}(1 + \|u\|), \quad \|J_k''(u) - J''(u)\| \leqslant \delta_{2k}, \quad u \in H, \quad k = 0, 1, \ldots \tag{3}$$

Подчеркнем, что в (3) не предполагается, что $J_k''(u)$ обязательно получен непосредственным дифференцированием $J_k'(u)$ — здесь не исключается возможность, что приближения $J_k'(u)$, $J_k''(u)$ найдены в результате отдельных независимых вычислений или измерений. В качестве приближений для $T_k'(u)$, $T_k''(u)$ возьмем

$$t_k'(u) = J_k'(u) + 2\alpha_k u, \quad t_k''(u) = J_k''(u) + 2\alpha_k I, \quad u \in H, \quad k = 0, 1, \ldots$$

Пусть существует обратный оператор $(t_k''(u))^{-1}$ и известно его приближение $D_k(u) \in \mathcal{L}(H \to H)$ такое, что

$$\|D_k(u) - (t_k''(u))^{-1}\| \leqslant \delta_{3k}, \quad u \in H, \quad k = 0, 1, \ldots \tag{4}$$

Теперь вместо (2) можем рассмотреть метод

$$u_{k+1} = u_k - D_k(u_k) t_k'(u_k), \quad k = 0, 1, \ldots \tag{5}$$

Теорема 1. *Пусть*

1) *функция* $J(u) \in C^2(H)$, *выпукла на* H; $J_* > -\infty$, $U_* \neq \varnothing$, u_* — *нормальное решение задачи* (1)

$$\|J''(u) - J''(v)\| \leqslant L\|u - v\| \quad \forall u, v \in H, \quad L = \text{const} > 0; \tag{6}$$

2) *приближения* $J_k'(u)$, $J_k''(u)$ *производных* $J'(u)$, $J''(u)$ *удовлетворяют условию* (3), *а для обратного оператора* $(t_k''(u))^{-1}$ *известно его приближение* $D_k(u)$ *с погрешностью* (4);

3) *параметры* α_k, δ_{1k}, δ_{2k}, δ_{3k} *таковы, что*

$$\alpha_k > 0, \quad 1 \leqslant \frac{\alpha_k}{\alpha_{k+1}} \leqslant 2, \quad (\alpha_k - \alpha_{k+1})\|u_*\| \leqslant D\alpha_k^2, \quad D = \frac{\gamma^2}{192L}, \quad \lim_{k\to\infty} \alpha_k = 0, \tag{7}$$

$$\delta_{1k} > 0, \quad 0 < \delta_{2k} \leqslant \alpha_k, \quad \delta_{3k} > 0, \quad \lim_{k\to\infty} (\delta_{1k} + \delta_{2k}) = 0, \tag{8}$$

$$12\left(\frac{\gamma}{12} c_{1k} + c_{2k} + \frac{12}{\gamma} c_{3k}\right) \leqslant 1 - \gamma, \quad k = 0, 1, \ldots, \tag{9}$$

где γ — *фиксированное число*, $0 < \gamma < 1$ (*например*, $\gamma = 1/2$),

$$c_{1k} = \tfrac{1}{2}\eta_k + \tfrac{1}{2}(1 + \alpha_k\delta_{3k})^2\left(1 + \frac{\delta_{1k}}{2\alpha_k}\right)^2 - \tfrac{1}{2},$$

$$c_{2k} = \frac{\delta_{1k}}{2\alpha_k^2}\left[(1 + \alpha_k\delta_{3k}) + 2(1 + \alpha_k\delta_{3k})^2\left(1 + \frac{\delta_{1k}}{2\alpha_k}\right)\right] L(1 + \|u_*\|) +$$

$$+ \frac{\eta_k}{\alpha_k}(2L\|u_*\| + \|J''(u_*)\| + 2\alpha_k),$$

$$c_{3k} = \frac{\delta_{1k}}{\alpha_k^3}(1+\alpha_k\delta_{3k})(2L\|u_*\|+\|J''(u_*)\|)+2\alpha_k)L(1+\|u_*\|)+$$
$$+\frac{1}{2}\frac{\delta_{1k}^2}{\alpha_k^4}(1+\alpha_k\delta_{3k})^2L^2(1+\|u_*\|)^2,$$
$$\eta_k = \frac{\delta_{1k}}{2\alpha_k}+\frac{\delta_{2k}}{2\alpha_k}+\alpha_k\delta_{3k}+\frac{1}{2}\delta_{1k}\delta_{3k}; \quad (10)$$

4) *начальные* u_0, α_0, δ_{10} *таковы, что*

$$\|J_0'(u_0)\|+2\alpha_0\|u_0\|+\delta_{10}(1+\|u_0\|)\leqslant\frac{\gamma}{12L}\alpha_0^2. \quad (11)$$

Тогда последовательность $\{u_k\}$, *определяемая методом* (5), *существует и*

$$\lim_{k\to\infty}J(u_k)=J_*, \quad \lim_{k\to\infty}\|u_k-u_*\|=0, \quad (12)$$

причем сходимость равномерная относительно выбора реализаций $J_k'(u)$, $J_k''(u)$, $D_k(u)$ *из* (3), (4).

Доказательство. Из выпуклости функции $J(u)\in C^2(H)$ следует, что $\langle J''(u)\xi,\xi\rangle\geqslant 0\ \forall u,\xi\in H$ (теорема 4.2.5), поэтому

$$\langle T_k''(u)\xi,\xi\rangle\geqslant 2\alpha_k\|\xi\|^2 \quad \forall u,\xi\in H, \quad k=0,1,\ldots \quad (13)$$

Это значит, что функция $T_k(u)$ сильно выпукла на H. Тогда существует, притом единственная, точка $v_k\in H$ такая, что (теорема 8.2.10)

$$T_k(v_k)=\inf_H T_k(u), \quad k=0,1,\ldots \quad (14)$$

Согласно теореме 4.4

$$\|v_k\|\leqslant\|u_*\|, \quad k=0,1,\ldots, \quad \lim_{k\to\infty}J(v_k)=J_*, \quad \lim_{k\to\infty}\|v_k-u_*\|=0. \quad (15)$$

Кроме того, из (14) следует, что $T_k'(v_k)=0$. Отсюда и из сильной выпуклости $T_k(u)$ имеем (теорема 4.3.3)

$$2\alpha_k\|u-v_k\|^2\leqslant\langle T_k'(u)-T_k'(v_k),u-v_k\rangle=\langle T_k'(u),u-v_k\rangle\leqslant\|T_k'(u)\|\|u-v_k\|$$

или

$$\|u-v_k\|\leqslant\frac{1}{2\alpha_k}\|T_k'(u)\| \quad \forall u\in H, \quad k=0,1,\ldots \quad (16)$$

Из условий теоремы вытекает

$$\|T_k''(u)-T_k''(v)\|\leqslant L\|u-v\| \quad \forall u,v\in H, \quad k=0,1,\ldots, \quad (17)$$

$$\|T_k'(u)-t_k'(u)\|\leqslant\delta_{1k}(1+\|u\|), \quad \|T_k''(u)-t_k''(u)\|\leqslant\delta_{2k} \quad \forall u\in H, \quad k=0,1,\ldots \quad (18)$$

С помощью индукции докажем, что при сделанных предположениях метод (5) действительно определяет некоторую последовательность $\{u_k\}$, причем справедлива оценка

$$a_k=\|T_k'(u_k)\|\leqslant A\alpha_k^2, \quad A=\frac{\gamma}{12L}, \quad k=0,1,\ldots \quad (19)$$

Из (11), (18) следует

$$a_0=\|T_0'(u_0)\|\leqslant\|t_0'(u_0)\|+\delta_{10}(1+\|u_0\|)\leqslant A\alpha_0^2.$$

Пусть при некотором $k\geqslant 0$ известна точка u_k и верна оценка (19). В силу (15), (16) имеем

$$\|u_k\|\leqslant\|u_k-v_k\|+\|v_k\|\leqslant\frac{1}{2}\alpha_k a_k+\|u_*\|. \quad (20)$$

Из (18), (20) вытекает

$$\|t_k'(u_k)\|\leqslant\delta_{1k}(1+\|u_k\|)+\|T_k'(u_k)\|\leqslant$$
$$\leqslant\delta_{1k}(1+\frac{1}{2\alpha_k}a_k+\|u_*\|)+a_k=\left(1+\frac{\delta_{1k}}{2\alpha_k}\right)a_k+\delta_{1k}(1+\|u_*\|). \quad (21)$$

Далее, из неравенств (13), (18) с учетом (8) получим

$$\langle t_k''(u)\xi,\xi\rangle=\langle T_k''(u)\xi,\xi\rangle+\langle(t_k''(u)-T_k''(u))\xi,\xi\rangle\geqslant$$
$$\geqslant 2\alpha_k\|\xi\|^2-\delta_{2k}\|\xi\|^2=(2\alpha_k-\delta_{2k})\|\xi\|^2\geqslant\alpha_k\|\xi\|^2 \quad \forall\xi\in H. \quad (22)$$

Из неравенств (13), (22) следует существование обратных операторов $(T_k''(u))^{-1}$, $(t_k''(u))^{-1}$ и оценка их норм (см. упражнение 8.3.27):

$$\|(T_k''(u))^{-1}\|\leqslant\frac{1}{2\alpha_k}, \quad \|(t_k''(u))^{-1}\|\leqslant\frac{1}{\alpha_k}. \quad (23)$$

Таким образом, имеет смысл говорить об операторе $D_k(u)$, являющемся приближением к $(t_k''(u))^{-1}$ в смысле неравенства (4), и о точке u_{k+1}, определяемой формулой (5).

Оценим $a_{k+1}=\|T_{k+1}'(u_{k+1})\|$. С учетом формулы $T_k'(u)=J'(u)+2\alpha_k u$ имеем

$$a_{k+1}\leqslant\|T_{k+1}'(u_{k+1})-T_k'(u_{k+1})\|+\|T_k'(u_{k+1})\|\leqslant$$
$$\leqslant\|T_k'(u_{k+1})\|+2(\alpha_k-\alpha_{k+1})\|u_{k+1}\|. \quad (24)$$

Но в силу (15), (16)

$$\|u_{k+1}\|\leqslant\|u_{k+1}-v_k\|+\|v_k\|\leqslant\frac{1}{2\alpha_k}\|T_k'(u_{k+1})\|+\|u_*\|. \quad (25)$$

Подставим оценку (25) в (24). С учетом условий (7) получим

$$a_{k+1}\leqslant\|T_k'(u_{k+1})\|\left(1+\frac{\alpha_k-\alpha_{k+1}}{\alpha_k}\right)+2(\alpha_k-\alpha_{k+1})\|u_*\|\leqslant 3\|T_k'(u_{k+1})\|+2D\alpha_k^2. \quad (26)$$

Для оценки $\|T_k'(u_{k+1})\|$ воспользуемся равенством

$$T_k'(u_{k+1})=T_k'(u_k)+\int_0^1 T_k''(u_k+\theta(u_{k+1}-u_k))(u_{k+1}-u_k)d\theta=$$
$$=T_k''(u_k)\{[(T_k''(u_k))^{-1}-(t_k''(u_k))^{-1}]T_k'(u_k)+$$
$$+(t_k''(u_k))^{-1}[T_k'(u_k)-t_k'(u_k)]+[(t_k''(u_k))^{-1}-D_k(u_k)]t_k'(u_k)\}+$$
$$+\int_0^1[T_k''(u_k)-T_k''(u_k+\theta(u_{k+1}-u_k))]D_k(u_k)t_k'(u_k)d\theta \quad (27)$$

(определение интеграла от функций со значениями в H и формулу Ньютона — Лейбница для них см. в [393; 557]).

С учетом оценок (18), (23) имеем

$$\|(T_k''(u_k))^{-1}-(t_k''(u_k))^{-1}\|=\|T_k''(u_k)^{-1}(t_k''(u_k)-T_k''(u_k))(t_k''(u_k))^{-1}\|\leqslant\frac{\delta_{2k}}{2\alpha_k^2}. \quad (28)$$

Далее, из (4), (23) следует

$$\|D_k(u_k)\|\leqslant\|(t_k''(u_k))^{-1}\|+\delta_{3k}\leqslant\frac{1}{\alpha_k}(1+\alpha_k\delta_{3k}). \quad (29)$$

Кроме того, с учетом (15)–(17) получаем

$$\|T_k''(u_k)\|\leqslant\|T_k''(u_k)-T_k''(u_*)\|+\|T_k''(u_*)\|\leqslant L\|u_k-u_*\|+\|T_k''(u_*)\|\leqslant$$
$$\leqslant L\|u_k-v_k\|+L\|v_k-u_*\|+\|T_k''(u_*)\|\leqslant$$
$$\leqslant\frac{L}{2\alpha_k}a_k+2L\|u_*\|+\|J''(u_*)\|+2\alpha_k. \quad (30)$$

Пользуясь (4), (5), (17), (18), (20), (21), (23), (28)–(30), из (27) имеем

$$\|T_k'(u_{k+1})\|\leqslant\|T_k''(u_k)\|\left(\frac{\delta_{2k}}{2\alpha_k^2}a_k+\frac{\delta_{1k}}{\alpha_k}(1+\|u_k\|)+\delta_{3k}\|t_k'(u_k)\|\right)+$$
$$+\frac{1}{2}L\|D_k(u_k)\|^2\|t_k'(u_k)\|^2\leqslant\left[\frac{L}{2\alpha_k}a_k+2L\|u_*\|+\|J''(u_*)\|+2\alpha_k\right]\times$$
$$\times\left[\frac{\delta_{2k}}{2\alpha_k^2}a_k+\frac{\delta_{1k}}{\alpha_k}(1+\frac{1}{2\alpha_k}a_k+\|u_*\|)+\delta_{3k}\left(\left(1+\frac{\delta_{1k}}{\alpha_k}\right)a_k+\delta_{1k}\delta_{3k}(1+\|u_*\|)\right]+\right.$$
$$\left.+\frac{1}{2}L\frac{1}{\alpha_k^2}(1+\alpha_k\delta_{3k})^2\left[\left(1+\frac{\delta_{1k}}{2\alpha_k}\right)a_k+\delta_{1k}(1+\|u_*\|)\right]^2.\right.$$

Перемножим выражения в квадратных скобках и получившийся результат расположим по степеням a_k^2, a_k^1, a_k^0. С использованием обозначений $c_{1k}, c_{2k}, c_{3k}, \eta_k$ из (10) будем иметь

$$\|T_k'(u_{k+1})\|\leqslant\frac{1}{2}L\frac{a_k^2}{\alpha_k^2}+La_k^2\frac{1}{\alpha_k^2}c_{1k}+a_kc_{2k}+\frac{1}{L}\alpha_k^2c_{3k}. \quad (31)$$

Поскольку $c_{1k}=c_{2k}=c_{3k}=0$ при $\delta_{1k}=\delta_{2k}=\delta_{3k}=0$, то первое слагаемое в правой части (31), не зависящее от δ_{ik}, представляет собой главный член в этой оценке и поэтому выделен отдельно. Подставим оценку (31) в (26). С учетом индуктивного предположения (19) получим

$$a_{k+1}\leqslant3\left(\frac{1}{2}L\frac{a_k^2}{\alpha_k^2}+L\frac{a_k^2}{\alpha_k^2}c_{1k}+a_kc_{2k}+\frac{1}{L}\alpha_k^2c_{3k}\right)+2D\alpha_k^2\leqslant$$
$$\leqslant\frac{3}{2}(LA^2\alpha_k^2+2D\alpha_k^2)+3(LA^2\alpha_k^2c_{1k}+A\alpha_k^2c_{2k}+\frac{1}{L}\alpha_k^2c_{3k})=$$
$$=A\alpha_{k+1}^2\left(\frac{\alpha_k}{\alpha_{k+1}}\right)^2\left[\left(\frac{3}{2}LA+\frac{2D}{A}\right)+3(LAc_{1k}+c_{2k}+\frac{1}{LA}c_{3k})\right]\leqslant A\alpha_{k+1}^2;$$

здесь мы учли, что $\frac{\alpha_k}{\alpha_{k+1}}\leqslant2$ в силу (7), $\frac{3}{2}LA+\frac{2D}{A}=\frac{\gamma}{4}$ по определению A, D из (7), (19), а $3(LAc_{1k}+c_{2k}+\frac{1}{LA}c_{3k})\leqslant3\left(\frac{\gamma}{12}c_{1k}+c_{2k}+\frac{12}{\gamma}c_{3k}\right)\leqslant\frac{1-\gamma}{4}$ в силу условия (9). Индуктивные рассуждения закончены, оценка (19) доказана.

Наконец, из неравенства (16) при $u=u_k$ с помощью оценки (19) имеем

$$\|u_k-v_k\|\leqslant\frac{1}{2\alpha_k}a_k\leqslant\frac{1}{2\alpha_k}A\alpha_k^2=\frac{1}{2}A\alpha_k,\quad k=0,1,\ldots, \quad (32)$$

так что $\lim\limits_{k\to\infty}\|u_k-v_k\|=0$. Отсюда и из (15) следуют равенства (12). Заметим, что правая часть оценки (32) не зависит от выбора реализаций $J_k'(u)$, $J_k''(u)$, $D_k(u_k)$ из (3), (4). Это значит, что пределы (12) равномерны относительно выбора указанных реализаций. Теорема 1 доказана. □

Замечание 1. Покажем, что последовательности $\{\alpha_k\}$, $\{\delta_{1k}\}$, $\{\delta_{2k}\}$, $\{\delta_{3k}\}$, удовлетворяющие условиям (7)–(9) существуют. В качестве $\{\alpha_k\}$, например, можно взять

$$\alpha_k=\frac{a}{(k+1)^\alpha},\quad k=0,1,\ldots,\quad 0<\alpha\leqslant1,\quad a>0. \quad (33)$$

Тогда $1\leqslant\frac{\alpha_k}{\alpha_{k+1}}=\left(1+\frac{1}{k+1}\right)^\alpha\leqslant2^\alpha\leqslant2$. Далее, $\|u_*\|\left(\frac{\alpha_k-\alpha_{k+1}}{\alpha_k^2}\right)=\frac{\|u_*\|}{a}\alpha(k+1+\theta)^{-3-\alpha}\leqslant\frac{\|u_*\|}{a}\alpha\leqslant\frac{\gamma^2}{192L}$, если число a взять достаточно большим. Как видно из формул (10), условия (8), (9) заведомо будут выполнены, если при каждом $k\geqslant0$ числа δ_{1k}, δ_{2k}, δ_{3k} взять достаточно малыми. Более того, при любом выборе начального приближения u_0, взяв в (33) число a достаточно большим, нетрудно добиться выполнения неравенства (11). Это означает, что регуляризованный метод Ньютона свободен от известного недостатка обычного метода Ньютона (см. § 5.10), требующего выбора хорошего начального приближения, расширяет возможности этого метода. При выборе параметров метода (5) в условиях (7)–(10) вместо трудновычисляемых величин $\|u_*\|$, $\|J''(u_*)\|$ можно пользоваться их оценками

$$\|u_*\|\leqslant d,\quad \|J''(u_*)\|\leqslant L\|u_*-w\|+\|J''(w)\|\leqslant L(d+\|w\|)+\|J''(w)\|, \quad (34)$$

где w — произвольная точка из H, например, $w=0$. Теорема 1 и ее доказательство полностью сохраняются, если в них величины $\|u_*\|$, $\|J''(u_*)\|$ заменить их оценками (34).

2. В практических задачах (1) более реальной, чем (3), представляется следующая ситуация, когда вместо точных $J'(u)$, $J''(u)$ известны приближения $J_\delta'(u)$, $J_\delta''(u)$, для которых

$$\|J_\delta'(u)-J'(u)\|\leqslant\delta_1(1+\|u\|),\quad \|J_\delta''(u)-J''(u)\|\leqslant\delta_2\quad\forall u\in H, \quad (35)$$

где δ_1, δ_2 — известные положительные числа. Тогда в методе (4), (5) можно взять

$$t_k'(u)=J_\delta'(u)+2\alpha_ku,\quad t_k''(u)=J_\delta''(u)+2\alpha_kI. \quad (36)$$

Сформулируем правило останова процесса (4), (5), (36) в зависимости от уровня погрешностей δ_1, δ_2 в (35). Будем считать, что зафиксированы какие-либо начальная точка u_0 и параметры α_k, δ_{1k}, δ_{2k}, δ_{3k}, удовлетворяющие условиям (7)–(11), и пусть

$$\delta_{10}\geqslant\delta_1,\quad \delta_{20}\geqslant\delta_2. \quad (37)$$

Тогда процесс (4), (5), (36) можно продолжить до такого наибольшего номера $k=k(\delta)$, при котором выполняются неравенства

$$\delta_{1k}\geqslant\delta_1,\quad \delta_{2k}\geqslant\delta_2,\quad k=0,1,\ldots,k(\delta). \quad (38)$$

Поскольку $\{\delta_{1k}\}\to 0$, $\{\delta_{2k}\}\to 0$ по условию, и выполнены неравенства $\delta_{10}\geqslant \delta_1$, $\delta_{20}\geqslant\delta_2$, то такой номер $k(\delta)$ обязательно найдется. Обоснованием правила останова (38) процесса (4), (5), (36) служит

Теорема 2. *Пусть выполнены все условия теоремы* 1, *кроме условия* (3), *пусть приближения* $J_\delta'(u)$, $J_\delta''(u)$ *для* $J'(u)$, $J''(u)$ *удовлетворяют условию* (35). *Тогда точка* $u(\delta)=u_{k(\delta)}$, *полученная методом* (4), (5), (36), (38), *обладает свойством*

$$\lim_{\delta\to 0} J(u(\delta))=J_*,\quad \lim_{\delta\to 0}\|u(\delta)-u_*\|=0,$$

где u_* *— нормальное решение задачи* (1).

Доказательство опирается на теорему 1 и проводится также, как и аналогичная теорема 2 из § 8. □

Замечание 2. Покажем, что при любых $\delta_1>0$, $\delta_2>0$ из (35) можно выбрать параметры α_k, δ_{1k}, δ_{2k}, δ_{3k}, удовлетворяющие всем условиям (7)–(9), (11), (37), причем можно сделать так, что $\delta_{3k}=\delta_3$ для всех $k=0,1,\ldots$ Ограничимся рассмотрением последовательности $\{\alpha_k\}$, определенной согласно (33). Введем функцию $\varphi(\alpha,\delta_1,\delta_2,\delta_3)$, которая получена из левой части неравенства (9) при $\alpha_k=\alpha$, $\delta_{1k}=\delta_1$, $\delta_{2k}=\delta_2$, $\delta_{3k}=\delta_3$, т. е. $\varphi(\alpha_k,\delta_{1k},\delta_{2k},\delta_{3k})\equiv 12\left(\frac{\gamma}{12}c_{1k}+c_{2k}+\frac{12}{\gamma}c_{3k}\right)$. Зафиксируем любые δ_{10}, δ_{20} из условий (37). Заметим, что $\varphi(\alpha,\delta_{10},\delta_{20},\delta_3)\to 0$ при $\alpha\to\infty$, $\delta_3\to 0$, $\alpha\delta_3\to 0$. Поэтому можем зафиксировать $\alpha=\alpha_0=a$ столь большим, а δ_{30} столь малым, чтобы выполнялись условия (7), (11), $0<\delta_{20}\leqslant a$, и, кроме того,

$$\varphi(\alpha_0,\delta_{10},\delta_{20},\delta_{30})\leqslant\frac{1-\gamma}{2}. \tag{39}$$

Таким образом, условие (9) при $k=0$ выполнено. Далее заметим, что при любых фиксированных $\alpha>0$, $\delta_3>0$ функция $\varphi(\alpha,\delta_1,\delta_2,\delta_3)$ монотонно растет по каждой переменной $\delta_1>0$, $\delta_2>0$, а функция $\varphi(\alpha,0,0,\delta_3)=12\left(\frac{\gamma}{12}\alpha\delta_3+\frac{1}{2}(1+\alpha\delta_3)^2-\frac{1}{2}\right)$ монотонно растет по α при каждом фиксированном $\delta_3>0$. Отсюда и из (39) имеем

$$\varphi(\alpha_k,0,0,\delta_{30})\leqslant\varphi(\alpha_0=a,0,0,\delta_{30})\leqslant\varphi(a,\delta_{10},\delta_{20},\delta_{30})\leqslant\frac{1}{2}(1-\gamma). \tag{40}$$

Далее, фиксируя $\alpha=\alpha_k$, $\delta_3=\delta_{30}>0$ и учитывая, что $\varphi(\alpha_k,\delta_1,\delta_2,\delta_{30})-\varphi(\alpha_k,0,0,\delta_{30})\to 0$ при $(\delta_1,\delta_2)\to 0$, при каждом $k\geqslant 1$ выберем δ_{1k}, δ_{2k} столь малыми, чтобы

$$\varphi(\alpha_k,\delta_{1k},\delta_{2k},\delta_{30})-\varphi(\alpha_k,0,0,\delta_{30})\leqslant\frac{1}{2}(1-\gamma).$$

Отсюда и из (40) следует, что

$$\varphi(\alpha_k,\delta_{1k},\delta_{2k},\delta_{30})\leqslant 1-\gamma,$$

т. е. условие (9) выполняется. Тем самым доказано, что параметры α_k, δ_{1k}, δ_{2k}, $\delta_{3k}=\delta_{30}$, $k=0,1,\ldots$, с требуемыми свойствами (7)–(9), (11), (37) существуют. Следует, конечно, оговориться, что приведенные выше рассуждения относительно выбора параметров метода (5), (36), (38) на практике реализовать можно, лишь имея оценки (34) и оценку постоянной L из (6). Но даже при отсутствии этих оценок мы можем быть уверены в том, что для широких классов задач (1) «хорошие» параметры метода существуют и их имеет смысл искать, опираясь на эвристические и иные соображения, на численный эксперимент.

Заметим, что для задачи (1), когда множество U имеет вид (8.2), $U_0\equiv H$, и задано с погрешностью, аналогичный регуляризованный метод Ньютона в сочетании со штрафными функциями исследовался в [158].

§ 12. Регуляризованный непрерывный метод проекции градиента

1. Рассмотрим задачу

$$J(u)\to\inf,\quad u\in U, \tag{1}$$

где U — выпуклое замкнутое множество из гильбертова пространства H, функция $J(u)$ определена и дифференцируема по Фреше на H. Введем функцию Тихонова

$$T(u,t)=J(u)+\tfrac{1}{2}\alpha(t)\|u\|^2,\quad \alpha(t)>0,\quad t\geqslant 0,\quad u\in H, \tag{2}$$

которая также будет дифференцируемой по Фреше, и ее градиент равен

$$T'(u,t)=J'(u)+\alpha(t)u,\quad u\in H.$$

Будем предполагать, что множество U известно точно, а вместо градиента $J'(u)$ известно его приближение $J'(u,t)$, зависящее от параметра $t\geqslant 0$ и удовлетворяющее неравенству

$$\|J'(u,t)-J'(u)\|\leqslant\delta(t)(1+\|u\|),\quad \delta(t)>0,\quad t\geqslant 0,\quad u\in H. \tag{3}$$

Пусть $u=u(t)$, $t\geqslant 0$ — траектория дифференциального уравнения

$$u'(t)=\mathcal{P}_U(u(t)-\beta(t)(J'(u,t)+\alpha(t)u(t)))-u(t),\quad t>0;\quad u(0)=u_0, \tag{4}$$

где u_0 — произвольная фиксированная точка из H, $\mathcal{P}_U(z)$ — проекция точки $z\in H$ на множество U, производная $u'(t)=\lim\limits_{\Delta t\to 0}\frac{u(t+\Delta t)-u(t)}{\Delta t}$, где сходимость понимается в норме H. Далее нас будет интересовать поведение траектории системы (4) при $t\to\infty$. Регуляризованный непрерывный метод проекции градиента описан. Нетрудно видеть, что метод (4) получен из метода (5.2.34) формальной заменой градиента $J'(u)$ на приближенное значение градиента функции (2).

Теорема 1. *Пусть*

1) U *— выпуклое замкнутое множество из* H, *функция* $J(u)$ *выпукла на* H; $J(u)\in C^1(H)$, $J_*>-\infty$, $U_*\neq\varnothing$, *справедливо неравенство*

$$\|J'(u)-J'(v)\|^2\leqslant L\langle J'(u)-J'(v),u-v\rangle\quad\forall u,v\in H;\quad L=\mathrm{const}>0; \tag{5}$$

2) *приближение* $J'(u,t)$ *градиента* $J'(u)$ *непрерывно по* u *при всех* $t\geqslant 0$, *измеримо по* t *при всех* $u\in H$ *(об измеримых функциях со значениями из* H *см., например,* [557; 357]) *и удовлетворяет условию* (3);

3) *параметры* $\alpha(t)$, $\beta(t)$, $\delta(t)$ *метода* (4) *таковы, что*

$$\begin{gathered}\alpha(t),\beta(t)\in C^1[0,+\infty),\quad \delta(t)\in C[0,+\infty),\quad \alpha(t)\ \text{выпукла на}\ [0,+\infty),\\ \alpha(t)>0,\quad \beta(t)>0,\quad \delta(t)>0,\quad \alpha'(t)\leqslant 0,\quad \beta'(t)\leqslant 0\quad\forall t\geqslant 0,\\ 8\tfrac{\delta(t)}{\alpha(t)}+2\alpha(0)\beta(0)\leqslant 1,\quad (1+L)(\beta(0)+4\beta(0)\delta(t))\leqslant 3\quad\forall t\geqslant 0,\\ \lim_{t\to\infty}\left(\alpha(t)+\beta(t)+\tfrac{\delta(t)}{\alpha(t)}+\tfrac{|\alpha'(t)|}{\alpha^2(t)\beta(t)}+\tfrac{|\beta'(t)|}{\alpha(t)\beta^2(t)}\right)=0.\end{gathered} \tag{6}$$

Тогда

$$\lim_{t\to\infty} J(u(t)) = J_*, \quad \lim_{t\to\infty} (\|u(t)-u_*\| + \|u'(t)\|) = 0, \tag{7}$$

где $u(t)$ — решение задачи Коши (4), u_* — нормальное решение задачи (1), причем сходимость в (7) равномерная относительно выбора реализаций $J'(u,t)$ из (3).

В качестве функций $\alpha(t), \beta(t), \delta(t)$, удовлетворяющих условиям (6), можно, например, взять

$$\alpha(t) = a_1(1+t)^{-1/8}, \quad \beta(t) = a_2(1+t)^{-1/4}, \quad \delta(t) = a_3(1+t)^{-1}, \quad t \geqslant 0, \tag{8}$$

где $a_1 > 0$, $a_2 > 0$, $a_3 > 0$, $8\frac{a_3}{a_1} + 2a_1a_2 \leqslant 1$, $4a_2a_3 + (1+L)a_2 \leqslant 3$.

Доказательство. Из условий теоремы следует, что функция $f(u,t) = P_U(u - \beta(t)(J'(u,t) + \alpha(t)u)) - u$ измерима по t при всех $u \in H$ и непрерывна по u на H для всех $t \geqslant 0$, $\|f(u,t)\| \leqslant c_0(1+\|u\|)$. Отсюда следует, что задача Коши (4) при любом начальном условии $u_0 \in E^n$ имеет хотя бы одно решение, определенное при всех t, $0 \leqslant t < \infty$ [132; 208]. Так как функция $T(u,\tau)$ сильно выпукла на U, то найдется единственная точка $v(\tau)$ такая, что

$$v(\tau) \in U, \quad T(v(\tau),\tau) = \inf_U T(u,\tau), \quad \tau \geqslant 0, \tag{9}$$

(теорема 8.2.10). Положим $\tau = \frac{1}{\delta}$, $t_\delta(u) = T\left(u, \frac{1}{\delta}\right)$, $\delta > 0$. Тогда точка $u_\delta = v\left(\frac{1}{\delta}\right)$, $\delta > 0$, является решением задачи

$$t_\delta(u) = J(u) + \alpha\left(\frac{1}{\delta}\right)\|u\|^2 \to \inf, \quad u \in U,$$

так как $t_\delta(u_\delta) = T\left(v\left(\frac{1}{\delta}\right), \frac{1}{\delta}\right) = T(v(\tau),\tau) = \inf_U T(u,\tau) = \inf_U t_\delta(u)$. Поскольку $\alpha\left(\frac{1}{\delta}\right) > 0$, $\lim_{\delta\to 0} \alpha\left(\frac{1}{\delta}\right) = 0$, то из теоремы 4.4 имеем

$$\|v(\tau)\| = \|u_\delta\| \leqslant \|u_*\| \quad \forall \tau > 0, \quad \lim_{\tau\to\infty} J(v(\tau)) = \lim_{\delta\to 0} J(u_\delta) = J_*,$$
$$\lim_{\tau\to\infty} \|v(\tau) - u_*\| = \lim_{\delta\to 0} \|u_\delta - u_*\| = 0. \tag{10}$$

Далее, из условия (9) следует (теорема 8.3.3):

$$\langle T'(v(\tau),\tau), v - v(\tau)\rangle = \langle J'(v(\tau)) + \alpha(\tau)v(\tau), v - v(\tau)\rangle \geqslant 0 \quad \forall v \in U, \ \forall \tau \geqslant 0. \tag{11}$$

Из свойства оператора проектирования (неравенство (4.4.1)) с учетом (4) имеем:

$$\langle u'(t) + \beta(t)(J'(u(t),t) + \alpha(t)u(t), w - u'(t) - u(t)\rangle \geqslant 0 \quad \forall w \in U, \ t \geqslant 0. \tag{12}$$

В неравенстве (11) положим $v = u'(t) + u(t) \in U$, умножим его на $\beta(t)$ и сложим с (12) при $w = v(\tau) \in U$. После простых преобразований получим

$$\|u'(t)\|^2 + (1+\alpha(t)\beta(t))\langle u'(t), u(t) - v(\tau)\rangle + \alpha(t)\beta(t)\|u(t) - v(\tau)\|^2 \leqslant$$
$$\leqslant \beta(t)\langle (J'(u(t),t) - J'(u(t))) + (J'(u(t)) - J'(v(\tau))) +$$
$$+ (\alpha(t) - \alpha(\tau))v(\tau), v(\tau) - u(t) - u'(t)\rangle \quad \forall t, \tau \geqslant 0. \tag{13}$$

Из условия (5) следует, что $\langle J'(u(t)) - J'(v(\tau)), v(\tau) - u(t) - u'(t)\rangle \leqslant \frac{1}{4}L\|u'(t)\|^2$ (теорема 4.2.16). Учитывая это неравенство и условие (3), из (13) имеем

$$\|u'(t)\|^2 + (1+\alpha(t)\beta(t))\langle u'(t), u(t) - v(\tau)\rangle + \alpha(t)\beta(t)\|u(t) - v(\tau)\|^2 \leqslant$$
$$\leqslant \tfrac{1}{4}L\beta(t)\|u'(t)\|^2 + \beta(t)\delta(t)(1+\|u(t)\|)(\|u'(t)\| + \|u(t) - v(\tau)\|) +$$
$$+ \beta(t)|\alpha(t) - \alpha(\tau)|\|v(\tau)\|(\|u'(t)\| + \|u(t) - v(\tau)\|) \quad \forall t, \tau \geqslant 0. \tag{14}$$

Пользуясь элементарными неравенствами $|ab| \leqslant \frac{1}{2}\varepsilon a^2 + \frac{1}{2\varepsilon}b^2$, $(a+b)^2 \leqslant 2a^2 + 2b^2$ $\forall a, b, \varepsilon > 0$, и, учитывая, что $\|v(\tau)\| \leqslant \|u_*\|$, $\alpha(0) \geqslant \alpha(t) > 0$ $\forall t \geqslant 0$, получим:

$$(1+\|u(t)\|)(\|u'(t)\| + \|u(t) - v(\tau)\|) \leqslant (1 + \|v(\tau)\| + \|u(t) - v(\tau)\|) \times$$
$$\times (\|u'(t)\| + \|u(t) - v(\tau)\|) = (1 + \|v(\tau)\|)(\|u'(t)\| + \|u(t) - v(\tau)\|) +$$
$$+ \|u(t) - v(\tau)\|\|u'(t)\| + \|u(t) - v(\tau)\|^2 \leqslant (1 + \|v(\tau)\|)^2 + \tfrac{1}{4}(\|u'(t)\| +$$
$$+ \|u(t) - v(\tau)\|)^2 + \tfrac{1}{2}\|u'(t)\|^2 + \tfrac{1}{2}\|u(t) - v(\tau)\|^2 + \|u(t) - v(\tau)\|^2 \leqslant$$
$$\leqslant (1 + \|u_*\|)^2 + \|u'(t)\|^2 + 2\|u(t) - v(\tau)\|^2 \quad \forall t, \tau \geqslant 0;$$
$$|\alpha(t) - \alpha(\tau)|\|v(\tau)\|(\|u'(t)\| + \|u(t) - v(\tau)\|) \leqslant$$
$$\leqslant |\alpha(t) - \alpha(\tau)|\|u_*\|\|u'(t)\| + |\alpha(t) - \alpha(\tau)|\|u_*\|\|u(t) - v(\tau)\| \leqslant$$
$$\leqslant |\alpha(t) - \alpha(\tau)|^2\|u_*\|^2 + \tfrac{1}{4}\|u'(t)\|^2 + \frac{|\alpha(t) - \alpha(\tau)|^2}{\alpha(t)}\|u_*\|^2 +$$
$$+ \tfrac{1}{4}\alpha(t)\|u(t) - v(\tau)\|^2 \leqslant |\alpha(t) - \alpha(\tau)|^2\|u_*\|^2\frac{1+\alpha(0)}{\alpha(t)} +$$
$$+ \tfrac{1}{4}\|u'(t)\|^2 + \tfrac{1}{4}\alpha(t)\|u(t) - v(\tau)\|^2 \quad \forall t, \tau \geqslant 0.$$

Подставим эти оценки в правую часть (14). После простых преобразований будем иметь

$$\tfrac{1}{2}(1+\alpha(t)\beta(t))\tfrac{d}{dt}\|u(t) - v(\tau)\|^2 + (1 - \beta(t)\delta(t) - \tfrac{1}{4}(L+1)\beta(t))\|u'(t)\|^2 +$$
$$+ \alpha(t)\beta(t)\left(\tfrac{3}{4} - \frac{2\delta(t)}{\alpha(t)}\right)\|u(t) - v(\tau)\|^2 \leqslant (1 + \|u_*\|)^2\beta(t)\delta(t) +$$
$$+ \beta(t)\frac{(\alpha(t) - \alpha(\tau))^2}{\alpha(t)}\|u_*\|^2(1 + \alpha(0)) \leqslant c_1 b(t,\tau) \quad \forall t, \tau \geqslant 0, \tag{15}$$

где $c_1 = \max\{(1 + \|u_*\|^2; (1 + \alpha(0))\|u_*\|^2\}$,

$$b(t,\tau) = \beta(t)\delta(t) + \frac{\beta(t)}{\alpha(t)}|\alpha(t) - \alpha(\tau)|^2, \quad t, \tau \geqslant 0. \tag{16}$$

Умножим неравенство (15) на функцию $\mu(t) = \exp\left(\int_0^t \alpha(\theta)\beta(\theta)d\theta\right) > 0$ и

проинтегрируем на отрезке $[0, t]$. Первый интеграл преобразуем по частям. Получим

$$\frac{1}{2}(1+\alpha(t)\beta(t))\|u(t)-v(\tau)\|^2\mu(t)+\int_0^t(1-\beta(s)\delta(s)-\frac{1}{4}(L+1)\times$$
$$\times\beta(s))\mu(s)\|u'(s)\|^2ds+\int_0^t[\alpha(s)\beta(s)\Big(\frac{3}{4}-2\frac{\delta(s)}{\alpha(s)}\Big)\mu(s)-$$
$$-\frac{1}{2}\frac{d}{dt}((1+\alpha(s)\beta(s))\mu(s))]\|u(s)-v(\tau)\|^2ds\leqslant c_1\int_0^t b(s,\tau)\mu(s)ds+c_2,$$
$$c_2=\frac{1}{2}(1+\alpha(0)\beta(0))(\|u_0\|+\|u_*\|)^2. \qquad (17)$$

Из условий (6) вытекает, что подынтегральные функции в левой части неравенства (17) неотрицательны, поэтому из (17) следует

$$\frac{1}{2}(1+\alpha(t)\beta(t))\mu(t)\|u(t)-v(\tau)\|^2\leqslant c_1\int_0^t b(s,\tau)\mu(s)ds+c_2\quad\forall t,\tau\geqslant 0.$$

Полагая $\tau=t$, отсюда имеем

$$\frac{1}{2}(1+\alpha(t)\beta(t))\mu(t)\|u(t)-v(t)\|^2\leqslant c_1\int_0^t b(s,t)\mu(s)ds+c_2\quad\forall t\geqslant 0. \qquad (18)$$

Так как функция $\alpha(t)$ выпукла и монотонно убывает, то

$$0\leqslant\alpha(s)-\alpha(t)\leqslant\alpha'(s)(s-t)\quad\forall t,s,\quad 0\leqslant s\leqslant t$$

(теорема 4.2.2), поэтому $|\alpha(s)-\alpha(t)|\leqslant|\alpha'(s)||s-t|$, $0\leqslant s\leqslant t$. Отсюда и из определения (16) функции $b(s,t)$ получаем

$$b(s,t)\leqslant\beta(s)\delta(s)+\frac{\beta(s)}{\alpha(s)}(\alpha'(s))^2(s-t)^2\quad\forall t,s,\quad 0\leqslant s\leqslant t.$$

Тогда из (18) с учетом неравенства $1+\alpha(t)\beta(t)\geqslant 1$ имеем

$$\|u(t)-v(t)\|^2\leqslant 2c_1\frac{1}{\mu(t)}\int_0^t(\beta(s)\delta(s)+\frac{\beta(s)}{\alpha(s)}(\alpha'(s))^2(s-t)^2)\mu(s)ds+\frac{c_2}{\mu(t)} \qquad (19)$$
$$\forall t\geqslant 0.$$

Совершим в (19) предельный переход при $t\to\infty$. Сначала докажем две леммы.

Лемма 1. *Пусть функция* $\nu(t)\in C^1[0,+\infty)$ *такова, что*

$$\nu(t)>0,\quad\nu'(t)\leqslant 0\quad\forall t\geqslant 0,\quad\lim_{t\to\infty}\frac{\nu'(t)}{\nu^2(t)}=0.$$

Тогда

$$\lim_{t\to\infty}\nu^n(t)\exp\Big(\int_0^t\nu(\theta)d\theta\Big)=+\infty \qquad (20)$$

при каждом $n=0,1,\ldots$

Доказательство. Зафиксируем произвольное целое число $n\geqslant 0$. По условию леммы $\frac{|\nu'(t)|}{\nu^2(t)}=-\frac{\nu'(t)}{\nu^2(t)}\to 0$ при $t\to\infty$, поэтому найдется $t_n\geqslant 0$, что что $-\frac{\nu'(s)}{\nu^2(t)}\leqslant\frac{1}{n+1}$ $\forall s\geqslant t_n$. Интегрируя это неравенство на отрезке $[t_n, t]$, $t\geqslant t_n$, получим

$$\int_{t_n}^t\Big(-\frac{\nu'(s)}{\nu^2(t)}\Big)ds=\int_{t_n}^t\frac{d}{ds}\Big(\frac{1}{\nu(s)}\Big)ds=\frac{1}{\nu(t)}-\frac{1}{\nu(t_n)}\leqslant\frac{t-t_n}{n+1}\quad\forall t\geqslant t_n.$$

Следовательно,

$$\nu(t)\geqslant\frac{n+1}{t-t_n+c_n}\quad\forall t\geqslant t_n,\quad c_n=\frac{n+1}{\nu(t_n)}>0. \qquad (21)$$

Отсюда имеем

$$\int_0^t\nu(\theta)d\theta\geqslant\int_{t_n}^t\frac{n+1}{\theta-t_n+c_n}d\theta=(n+1)\ln\frac{t-t_n+c_n}{c_n}=\ln\Big(\frac{t-t_n+c_n}{c_n}\Big)^{n+1}\quad\forall t\geqslant t_n$$

Из (21) и последнего неравенства следует

$$\nu^n(t)\exp\Big(\int_0^t\nu(\theta)d\theta\Big)\geqslant\frac{(n+1)^n}{c_n^{n+1}}(t-t_n+c_n)\quad\forall t\geqslant t_n.$$

Отсюда при $t\to\infty$ приходим к равенству (20). Лемма 1 доказана. □

Лемма 2. *Пусть функция* $\nu(t)$ *удовлетворяет условиям леммы* 1 *и* $\mu(t)=\exp\Big(\int_0^t\nu(\theta)d\theta\Big)$. *Тогда*

$$\lim_{t\to\infty}\frac{\frac{d}{dt}(\nu^n(t)\mu(t))}{\nu^{n+1}(t)\mu(t)}=1\quad\forall n=0,1,\ldots \qquad (22)$$

Доказательство. Зафиксируем любое целое число $n\geqslant 0$. Заметим, что $\frac{d}{dt}(\nu^n(t)\mu(t))=n\nu'(t)\nu^{n-1}(t)\mu(t)+\nu^{n+1}(t)\mu(t)=(n\nu'(t)\nu^{n-1}(t)+$ $+\nu^{n+1}(t))\mu(t)$, поэтому $\frac{\frac{d}{dt}(\nu^n(t)\mu(t))}{\nu^{n+1}(t)\mu(t)}=n\frac{\nu'(t)}{\nu^2(t)}+1$. Учитывая условие $\lim_{t\to\infty}\frac{\nu'(t)}{\nu^2(t)}=0$, отсюда получаем равенство (22). Лемма 2 доказана. □

Продолжим доказательство теоремы 1. Положим $\nu(t)=\alpha(t)\beta(t)$. Из условий (6) следует, что $\nu'(t)=\alpha'(t)\beta(t)+\alpha(t)\beta'(t)\leqslant 0$, $\lim_{t\to\infty}\frac{\nu'(t)}{\nu^2(t)}=$ $=\lim_{t\to\infty}\Big(\frac{\alpha'(t)}{\alpha^2(t)\beta(t)}+\frac{\beta'(t)}{\alpha(t)\beta^2(t)}\Big)=0$. Согласно лемме 1 тогда

$$\lim_{t\to\infty}\alpha^n(t)\beta^n(t)\mu(t)=+\infty\quad\forall n=0,1,\ldots \qquad (23)$$

Из леммы 2 для функции $\nu(t)=\alpha(t)\beta(t)$ имеем

$$\lim_{t\to\infty}\frac{(\alpha(t)\beta(t))^{n+1}\mu(t)}{\frac{d}{dt}(\alpha^n(t)\beta^n(t)\mu(t))}=1. \qquad (24)$$

Пользуясь равенством (23) при $n=0$, заключаем, что второе слагаемое в правой части неравенства (19) стремится к нулю при $t\to\infty$. Далее, трижды применяя правило Лопиталя с учетом (6), (23), (24), получим:

$$\lim_{t\to\infty}\frac{\int_0^t(\beta(s)\delta(s)+\frac{\beta(s)}{\alpha(s)}(\alpha'(s))^2(t-s)^2)\mu(s)ds}{\mu(s)}=\lim_{t\to\infty}\Big(\int_0^t\frac{\beta(s)}{\alpha(s)}(\alpha'(s))^2(t-s)\mu(s)ds+$$
$$+\beta(t)\delta(t)\mu(t)\Big)\frac{1}{\alpha(t)\beta(t)\mu(t)}=\lim_{t\to\infty}\frac{2\int_0^t\frac{\beta(s)}{\alpha(s)}(\alpha'(s))^2\mu(s)ds}{\frac{d}{dt}(\alpha(t)\beta(t)\mu(t))}+\lim_{t\to\infty}\frac{\delta(t)}{\alpha(t)}=$$

$$= \lim_{t\to\infty} \frac{2\int_0^t \frac{\beta(s)}{\alpha(s)}(\alpha'(s))^2\mu(s)ds}{\alpha^2(t)\beta^2(t)\mu(t)} \cdot \lim_{t\to\infty} \frac{\alpha^2(t)\beta^2(t)\mu(t)}{\frac{d}{dt}(\alpha(t)\beta(t)\mu(t))} = \lim_{t\to\infty} \frac{2\frac{\beta(t)}{\alpha(t)}(\alpha'(t))^2\mu(t)}{\frac{d}{dt}(\alpha^2(t)\beta^2(t)\mu(t))} =$$

$$= \lim_{t\to\infty} \frac{2\frac{\beta(t)}{\alpha(t)}(\alpha'(t))^2\mu(t)}{\alpha^3(t)\beta^3(t)\mu(t)} \cdot \lim_{t\to\infty} \frac{\alpha^3(t)\beta^3(t)\mu(t)}{\frac{d}{dt}(\alpha^2(t)\beta^2(t)\mu(t))} = \lim_{t\to\infty} 2\left(\frac{\alpha'(t)}{\alpha^2(t)\beta(t)}\right)^2 = 0.$$

Отсюда и из (19) следует, что $\lim\limits_{t\to\infty} \|u(t)-v(t)\| = 0$. Тогда с учетом (10) имеем $\|u(t)-u_*\| \leqslant \|u(t)-v(t)\| + \|v(t)-u_*\| \to 0$, $J(u(t)) \to J_*$ при $t\to\infty$.

Наконец, из (6), (15), (16) при $\tau = t$ получаем

$$\tfrac{1}{4}\|u'(t)\|^2 \leqslant c_1 b(t,t) + (1+\alpha(t)\beta(t))\|u'(t)\|\|u(t)-v(t)\| \leqslant$$

$$\leqslant c_1\beta(t)\alpha(t) + \tfrac{\varepsilon}{2}\|u'(t)\|^2 + \tfrac{1}{2\varepsilon}(1+\alpha(t)\beta(t))^2\|u(t)-v(t)\|^2 \quad \forall\varepsilon>0, \quad \forall t\geqslant 0.$$

Примем здесь $\varepsilon = \frac{1}{4}$. Тогда $\frac{1}{8}\|u'(t)\|^2 \leqslant c_1\beta(t)\alpha(t) + 2(1+\alpha(t)\beta(t))^2\|u(t)-v(t)\|^2$ $\forall t \geqslant 0$. Отсюда с учетом уже доказанного имеем: $\lim\limits_{t\to\infty} \|u'(t)\| = 0$. Равенства (7) доказаны. Равномерная сходимость в (7) относительно выбора $J'(u,t)$ из (3) следует из того, что в (15), (19) коэффициенты при степенях $\|u(t)-v(\tau)\|^2$, $\|u'(t)\|^2$ не зависят от конкретных реализаций $J'(u,t)$ из (3). Теорема 1 доказана. □

2. Сформулируем правило останова процесса (4) для случая, когда уровень погрешности в задании градиента фиксирован и вместо $J'(u,t)$ из (3) имеем приближение $J'_\delta(u) \in H$ такое, что

$$\|J'_\delta(u) - J'(u)\| \leqslant \delta(1+\|u\|) \quad \forall u \in H, \tag{25}$$

где $\delta > 0$ — известное число. Заменяя в уравнении (4) $J'(u,t)$ на $J_\delta(u)$ получим процесс

$$z'(t) = P_U(z(t) - \beta(t)(J'_\delta(z(t)) + \alpha(t)z(t))) - z(t), \quad t\geqslant 0, \quad z(0) = u_0. \tag{26}$$

Будем предполагать, что параметры $\alpha(t)$, $\beta(t)$, $\delta(t)$, удовлетворяющие условиям (6), как-то зафиксированы, пусть $\delta(0) > \delta$. При каждом фиксированном δ, $0 < \delta < \delta(0)$, процесс (26) будем продолжать до момента $t = t(\delta)$, определяемого условием

$$t(\delta) = \sup\{t:\ \delta(s) > \delta, 0 \leqslant s \leqslant t\}. \tag{27}$$

Поскольку $\delta(t) \to 0$ при $t\to\infty$, $\delta(0) > \delta$, то такой момент $t(\delta)$ будет конечным при каждом $\delta > 0$ и, зная $\delta(t)$, его можно заранее вычислить с нужной точностью. Обоснованием сформулированного правила останова (27) процесса (26) служит

Теорема 2. *Пусть выполнены все условия теоремы* 1, *кроме условия* (3), *пусть приближение* $J'_\delta(u)$ *градиента* $J'(u)$ *удовлетворяет условию* (25). *Пусть* $z(t)$, $0 \leqslant t \leqslant t(\delta)$, *траектория процесса* (26), *момент* $t(\delta)$ *определен согласно* (27). *Тогда*

$$\lim_{\delta\to 0} J(z(t(\delta))) = J_*, \quad \lim_{\delta\to 0} \|z(t(\delta)) - u_*\| = 0. \tag{28}$$

Доказательство. Из (25), (27) следует, что

$$\|J'_\delta(u) - J'(u)\| \leqslant \delta(t)(1+\|u\|) \quad \forall u \in H, \quad 0 \leqslant t \leqslant t(\delta), \tag{29}$$

так что функция $J'(u,t) = J'_\delta(u)$ удовлетворяет условию (3) при всех t, $0 \leqslant t \leqslant t(\delta)$. Далее, согласно правилу останова (27), из $\delta(t) \to 0$ при $t\to\infty$ следует, что $t(\delta) \to \infty$ при $\delta \to 0$. Это значит, что при всех малых $\delta > 0$ момент $t(\delta)$ в (29) можно сделать сколь угодно большим.

Согласно теореме 1 при выполнении всех ее условий, включая условие (3), траектория $u(t)$, порождаемая методом (4), сходится в норме H к точке u_*, т. е. для $\forall\varepsilon > 0$ $\exists$ момент $T = T(\varepsilon)$ такой, что

$$\|u(t) - u_*\| < \varepsilon \quad \forall t \geqslant T(\varepsilon), \tag{30}$$

причем момент $T(\varepsilon)$ не зависит от выбора реализаций $J'(u,t)$ из (3). Так как $\lim\limits_{\delta\to 0} t(\delta) = \infty$, то $\exists\delta(\varepsilon) > 0$, что $t(\delta) \geqslant T(\varepsilon)$ при всех δ, $0 < \delta < \delta(\varepsilon)$. Это значит, что для всех δ, $0 < \delta < \delta(\varepsilon)$, метод (26) при $0 \leqslant t \leqslant t(\delta)$, где $t(\delta)$ определен согласно (27), порождает траекторию $z(t)$, $0 \leqslant t \leqslant t(\delta)$, которую можно получить также и методом (4) с реализациями $J'(\delta,t) = J'_\delta(u)$ при $0 \leqslant t \leqslant t(\delta)$, удовлетворяющими, в силу (29), условию (3). Поскольку $t(\delta) \geqslant T(\varepsilon)$, то, используя неравенство (30) при $t = t(\delta)$б получаем неравенство $\|z(t(\delta)) - u_*\| < \varepsilon$, справедливое при всех δ, $0 < \delta < \delta(\varepsilon)$. В силу произвольности $\varepsilon > 0$ приходим ко второму равенству (28). Отсюда и из непрерывности $J(u)$ следует первое равенство (28). Теорема 2 доказана. □

Из теоремы 2 следует, что оператор R_δ, который каждому набору $(J'_\delta(u), \delta)$ из (25) ставит в соответствие точку $u(\delta) = z(t(\delta))$, определяемую методом (26), (27), является регуляризующим оператором (определение 6.1).

Для задачи (1), когда множество U имеет вид (8.2) и задано с погрешностью, аналогичный регуляризованный непрерывный метод проекции градиента в сочетании со штрафными функциями исследован в [522].

В заключение отметим, что при отборе материала в §§ 8–12 мы руководствовались желанием пошире продемонстрировать разнообразие технических приемов, используемых при исследовании регуляризованных методов минимизации.

Другие регуляризованные методы вида (4), использующие дифференциальные уравнения более высоких порядков и их разностные аналоги, а также регуляризованные варианты других методов для задач минимизации и их обобщений, см., например, в [62–64; 147; 149–153; 155–159; 168; 171–174; 176–178; 185; 389; 522; 537].

§ 13. Метод динамической регуляризации

Рассмотрим процесс, динамика которого описывается системой

$$\dot{x}(t) = Ax(t) + Bu(t), \quad 0 \leqslant t \leqslant T; \quad x(0) = x_0, \tag{1}$$

где t — время, $x = x(t) = (x^1(t), \ldots, x^n(t))$ — фазовая траектория, $x_0 \in E^n$ — начальная точка, $u = u(t) = (u^1(t), \ldots, u^r(t))$ — управление, A, B — заданные матрицы размера $n\times n$, $n\times r$ соответственно, момент T задан. Управление $u = u(t)$ будем называть *допустимым*, если

$$u \in U = \{u = u(t) \in L_2^r[0,T]:\ u(t) \in V \text{ почти всюду } [0,T]\}, \tag{2}$$

где V — заданное множество из E^r.

При исследовании задач оптимального управления, связанных с системами вида (1), нам до сих пор приходилось определять траекторию (решение) $x(t)=x(t;u)$, $0\leqslant t\leqslant T$, задачи (1) по известному управлению $u\in U$. В этом параграфе мы займемся исследованием обратной задачи, когда траектория $x(t;u)$ задана и ищется допустимое управление (или одно из таких управлений), которое порождает эту траекторию. Эту задачу, постановка которой ниже будет уточнена, мы кратко будем именовать обратной задачей (1). Подобные задачи, когда по наблюдаемой траектории (или ее элементам) требуется восстановить какие-либо характеристики (коэффициенты, параметры) динамической системы, имеют широкие приложения и относятся к *обратным задачам*, теория и методы которых составляют интенсивно развивающуюся область математики (см., например, [17; 60; 127; 230; 236; 268–270; 300; 365; 403; 405; 438; 439; 474; 482; 556; 557; 618; 619; 621; 651; 708; 771; 782; 788; 812]). Обратные задачи, вообще говоря, неустойчивы к ошибкам наблюдения и для их решения нужно пользоваться методами регуляризации.

Пример 1. Рассмотрим задачу: $\dot{x}(t)=u(t)$, $0\leqslant t\leqslant T$, $x(0)=0$; $u=u(t)\in V=\{u\in E^1\colon |u|\leqslant 1\}$. Предположим, что ведется наблюдение за траекторией $x(t)\equiv 0$, соответствующей допустимому управлению $u_*=u_*(t)\equiv 0$. Пусть из-за погрешностей наблюдения вместо точной траектории $x(t)\equiv 0$ получено ее приближение $x_\delta(t)=\delta\sin\frac{t}{\delta}$, $0\leqslant t\leqslant T$; $0<\delta\leqslant 1$. Так как $u(t)=\dot{x}(t)$, то по аналогии в качестве приближения к искомому управлению $u_*(t)\equiv 0$ можно попытаться взять $u_\delta=u_\delta(t)=\dot{x}(t)=\cos\frac{t}{\delta}$, $0\leqslant t\leqslant T$. Однако $\|u_\delta-u_*\|^2_{L_2}=\int\limits_0^T\left(\cos\frac{t}{\delta}\right)^2dt=\frac{1}{2}\left(T+\frac{\delta}{2}\left(\sin\frac{2T}{\delta}\right)\right)\to\frac{T}{2}\neq 0$ при $\delta\to 0$, т. е. обратная задача неустойчива в метрике $L_2[0,T]$ (ср. пример 1.4).

Для решения обратной задачи (1) воспользуемся методом динамической регуляризации, разработанным Ю. С. Осиповым и его учениками [421; 556; 557; 808]. Уточним постановку задачи. Будем предполагать, что наблюдения за траекторией $x(t)=x(t;u)$ проводятся в некоторые дискретные моменты времени t_i, $i=0,\ldots,N-1$, $0=t_0<t_1<\ldots<t_{N-1}<t_N=T$, причем из-за погрешностей измерений вместо точных значений $x(t_i)$ известны их приближения $x_{\delta i}$ такие, что

$$|x_{\delta i}-x(t_i)|\leqslant\delta,\quad i=0,\ldots,N-1,\quad 0<\delta\leqslant\delta_0. \tag{3}$$

Пусть величины $x_{\delta 0},\ldots,x_{\delta N-1}$ измеряются и поступают в наше распоряжение последовательно во времени. Управление $u_\delta(t)$, являющееся приближением к искомому управлению, будем строить последовательно на отрезках $[t_0,t_1]$, $[t_1,t_2],\ldots,[T_{N-1},T_N]$ по мере поступления информации об измерениях, используя при построении $u_\delta(t)$ на частичном отрезке $[t_i,t_{i+1}]$ лишь значения $x_{\delta 0},\ldots,x_{\delta i}$ и не предполагая знания остальных значений $x_{\delta i+1},\ldots,x_{\delta N-1}$, которые в момент t_i, возможно, еще не измерены. Приведем индуктивное описание метода динамической регуляризации. В этом методе наряду с управлением $u_\delta(t)$ на каждом шаге еще строится вспомогательная функция $z_\delta(t)$, помогающая отслеживать наблюдаемую траекторию $x(t)$ по ее приближенным значениям $x_{\delta i}$ из (3).

Пусть при $i=0$ известно наблюдаемое значение $x_{\delta 0}$ начальной точки $x(0)=x_0$. Положим $z_\delta(0)=x_{\delta 0}$. Решая вспомогательную задачу минимизации

$$t_{\delta 0}(u)=2\langle z_\delta(0)-x_{\delta 0},Bu\rangle+\alpha(\delta)|u|^2\to\inf,\quad u\in V;\quad \alpha(\delta)>0,$$

определим точку $u_0\in V$ такую, что $t_{\delta 0}(u_0)=\inf\limits_{u\in V}t_{\delta 0}(u)$. Затем полагаем

$$u_\delta(t)=u_0,\quad z_\delta(t)=z_\delta(0)+(Ax_{\delta 0}+Bu_0)t\quad\forall t\in[t_0=0,t_1].$$

Пусть для некоторого i, $0<i<N-1$, уже определены $u_\delta(t)$, $z_\delta(t)$, $0\leqslant t\leqslant t_i$, и пусть нам стало известно измерение $x_{\delta i}$ наблюдаемой траектории $x(t)$ в момент t_i. Тогда решаем вспомогательную задачу минимизации

$$t_{\delta i}(u)=2\langle z_\delta(t_i)-x_{\delta i},Bu\rangle+\alpha(\delta)|u|^2\to\inf,\quad u\in V \tag{4}$$

и находим точку $u_i\in V$, $t_{\delta i}(u_i)=\inf\limits_{u\in V}t_{\delta i}(u)$. Затем полагаем

$$u_\delta(t)=u_i,\quad z_\delta(t)=z_\delta(t_i)+(Ax_{\delta i}+Bu_i)(t-t_i),\quad t\in[t_i,t_{i+1}]. \tag{5}$$

Далее, по мере поступления информации $x_{\delta i+1},\ldots,x_{\delta N-1}$ последовательно определяются $u_\delta(t)$, $z_\delta(t)$ на следующих отрезках $[t_{i+1},t_{i+2}],\ldots$ $\ldots,[t_{N-1},t_N=T]$.

Описанный метод представляет собой сочетание принципа экстремального прицеливания Н. Н. Красовского [414] и метода регуляризации А. Н. Тихонова (см. метод стабилизации, § 4) [695]. Вспомогательную траекторию $z_\delta(t)$ из (5) принято называть *поводырем*, а условие (4) выбора u_i называется *правилом экстремального прицеливания*. Для построения поводыря использован аналог разностного метода Эйлера для решения задачи Коши (1), отличающийся от классического метода Эйлера [74; 89] тем, что в (5) правая часть $Ax+Bu$ вычислена в точке $x=x_{\delta i}$, управление $u=u_i$ взято из (4). Заметим, что задача (4) является конечномерной задачей минимизации и для ее решения могут быть использованы, например, методы из гл. 5. Функция $t_{\delta i}(u)$ в (4) сильно выпукла, квадратична и на выпуклом замкнутом множестве V достигает своей нижней грани в единственной точке u_i. В частности, если V — многогранное множество, то задача (4) превращается в задачу квадратичного программирования (см. § 5.8). Отметим, что при $i=0$ функция $t_{\delta 0}(u)$ имеет особенно простой вид: $t_{\delta 0}(u)=\alpha(\delta)|u|^2$. Важно также заметить, что в точном определении точки минимума u_i в (4) нет необходимости — достаточно найти u_i из условий

$$u_i\in V,\quad t_{\delta i}(u_i)\leqslant\inf_{u\in V}t_{\delta i}(u)+\varepsilon(\delta),\quad \varepsilon(\delta)>0. \tag{6}$$

В дальнейшем будем предполагать, что функции $u_\delta(t)$, $z_\delta(t)$ построены по формулам (5) с использованием точек u_i из (6).

Покажем, что если параметры $\alpha(\delta)$, $\varepsilon(\delta)$, $h(\delta)=\max\limits_{0\leqslant i\leqslant N-1}\{t_{i+1}-t_i\}$ метода (4)–(6) согласованно стремятся к нулю при $\delta\to 0$, то построенное по формуле (5) управление $u_\delta(t)$, $0\leqslant t\leqslant T$, при $\delta\to 0$ сходится в метрике $L_2^r[0,T]$ к нормальному решению обратной задачи (1). Поясним, что здесь понимается под нормальным решением этой задачи. Введем множество $U_*=\{u=u(t)\in U\colon x(t)=x(t;u),0\leqslant t\leqslant T\}$, где $x(t)$, $0\leqslant t\leqslant T$ — наблюдаемая траектория, U — определено согласно (2). Таким образом, множество U_* состоит из всех допустимых управлений, порождающих одну и ту же траекторию $x(t)$ системы (1). Из самой постановки обратной задачи (1) следует, что $U_*\neq\varnothing$, так как предполагается, что наблюдаемая траектория $x(t)$ действительно является траекторией системы (1), порожденной хотя бы одним управлением $v=v(t)\in U$, т. е. $x(t)=x(t;v)$, $0\leqslant t\leqslant T$. Однако

множество U_* может состоять более, чем из одного элемента v. Управление $u_* = u_*(t)$, $0 \leqslant t \leqslant T$, называется *нормальным* решением обратной задачи (1), если

$$u_* \in U_*, \quad \|u_*\|_{L_2^r} = \inf_{u \in U_*} \|u\|_{L_2^r}.$$

Если V — выпуклое замкнутое множество, то нормальное решение этой задачи существует и единственно. В самом деле, тогда U — выпуклое замкнутое множество в $L_2^r[0, T]$ (пример 8.2.9). Нетрудно проверить, что множество U_* также выпукло. Далее, если $u_k = u_k(t) \to u_* = u_*(t)$ при $k \to \infty$ в норме $L_2^r[0, T]$, то $\lim\limits_{k \to \infty} x(t; u_k) = x(t; u_*)$ $\forall t \in [0, T]$ (теорема 6.3.1). Поэтому если $u_k \in U_*$, то $u_* \in U_*$, т. е. U_* замкнуто в $L_2^r[0, T]$. Применяя теорему 8.2.10 к задаче: $J(u) = \|u\|_{L_2^r}^2 \to \inf$, $u \in U_*$, убеждаемся, что нормальное решение обратной задачи (1) существует и определяется однозначно.

Теорема 1. *Пусть V — выпуклое замкнутое ограниченное множество из E^r, приближенные значения $x_{\delta i}$ наблюдаемой траектории $x(t)$ в момент t_i, $i=0, \ldots, N-1$, удовлетворяют условию* (3), *параметры $\alpha=\alpha(\delta)$, $\varepsilon=\varepsilon(\delta)$, $h=h(\delta)$ положительны, стремятся к нулю при $\delta \to 0$ и*

$$\lim_{\delta \to 0} \frac{\delta + h(\delta) + \varepsilon(\delta)}{\alpha(\delta)} = 0. \tag{7}$$

Тогда функции $u_\delta(t)$, $z_\delta(t)$, $0 \leqslant t \leqslant T$, определяемые методом (4)–(6) таковы, что

$$\lim_{\delta \to 0} \|u_\delta - u_*\|_{L_2^r} = 0, \quad \lim_{\delta \to 0} \max_{0 \leqslant t \leqslant T} |z_\delta(t) - x(t)|_{E^n} = 0, \quad \lim_{\delta \to 0} \|\dot{z}_\delta - \dot{x}\|_{L_2^r[0,T]} = 0, \tag{8}$$

где u_* — нормальное решение обратной задачи (1), *$x = x(t) = x(t, u_*)$ — наблюдаемая траектория системы* (1).

Доказательство. Прежде всего убедимся, что

$$|x(t; u)|_{E^n} \leqslant c_0 < \infty \quad \forall t \in [0, T] \quad \forall u \in U, \tag{9}$$

$$|\dot{x}(t; u)| \leqslant c_1 < \infty, \quad |x(t; u) - x(\tau; u)| \leqslant c_1 |t - \tau| \quad \forall t, \tau \in [0, T], \quad \forall u \in U. \tag{10}$$

Неравенство (9) нетрудно доказать, пользуясь леммой Гронуолла 6.3.1 (см. аналогичную оценку (8.2.8)). Отсюда и из уравнения (1), а также из ограниченности V следует: $|\dot{x}(t; u)| \leqslant \|A\| c_0 + \|B\| \cdot |V| = c_1 < \infty$, где $|V| = \sup\limits_{u \in V} |u|$. Тогда $|x(t; u) - x(\tau; u)| = \left| \int\limits_\tau^t \dot{x}(\xi; u) d\xi \right| \leqslant c_1 |t - \tau|$ $\forall t, \tau \in [0, T]$. Из (3), (9) получаем

$$|x_{\delta i}| \leqslant \delta + \max_{0 \leqslant i \leqslant N-1} |x(t_i)| \leqslant \delta_0 + c_0 = c_2 < \infty \quad \forall i = 0, \ldots, N-1. \tag{11}$$

Аналогичные оценки верны и для поводыря $z_\delta(t)$. Из (5) следует, что функция $z_\delta(t)$ непрерывна на $[0, T]$ и кусочно дифференцируема: $\dot{z}_\delta(t) = A x_{\delta i} + B u_i$ $\forall t \in [t_i, t_{i+1}]$, $i = 0, \ldots, N-1$. Отсюда и из (11) имеем

$$|\dot{z}_\delta(t)| \leqslant \|A\| c_2 + \|B\| |V| = c_3 \quad \forall t \in [0, T]. \tag{12}$$

Тогда

$$|z_\delta(t)| = \left| \int_0^t \dot{z}_\delta(\xi) d\xi + x_{\delta 0} \right| \leqslant c_3 T + c_2,$$
$$|z_\delta(t) - z_\delta(\tau)| = \left| \int_0^t \dot{z}_\delta(\xi) d\xi \right| \leqslant c_3 |t - \tau| \quad \forall t, \tau \in [0, T]. \tag{13}$$

Введем теперь функцию (функция Ляпунова):

$$w(t) = |z_\delta(t) - x(t)|_{E^n}^2 + \alpha(\delta) \int_0^t |u_\delta(\tau)|^2 d\tau, \quad 0 \leqslant t \leqslant T. \tag{14}$$

Зафиксируем произвольный момент времени $t \in (0, T]$ и вычислим производную $\dot{w}(t)$ этой функции. Пусть для определенности $t \in (t_i, t_{i+1}]$. Тогда

$$\dot{w}(t) = 2\langle z_\delta(t) - x(t), \dot{z}_\delta(t) - \dot{x}(t) \rangle_{E^n} + \alpha(\delta) |u_\delta(t)|_{E^r}^2 =$$
$$= 2\langle z_\delta(t_i) - x_{\delta i}, \dot{z}_\delta(t) - \dot{x}(t) \rangle + 2\langle (z_\delta(t) - z_\delta(t_i)) + (x_{\delta i} - x(t_i)) +$$
$$+ (x(t_i) - x(t)), \dot{z}_\delta(t) - \dot{x}(t) \rangle + \alpha(\delta) |u_i|_{E^r}^2. \tag{15}$$

Второе слагаемое правой части (15) в силу неравенств (3), (9)–(13) оценивается величиной $2(c_3 h(\delta) + \delta + c_1 h(\delta))(c_1 + c_3) \leqslant c_4(\delta + h(\delta))$. Поэтому из (15) с учетом уравнений (1), $\dot{z}_\delta(t) = A x_{\delta i} + B u_i$ имеем:

$$\dot{w}(t) \leqslant 2\langle z_\delta(t_i) - x_{\delta i}, A x_{\delta i} + B u_i - A x(t) - B u_*(t) \rangle + c_4(\delta + h(\delta)) + \alpha(\delta) |u_i|_{E^r}^2 =$$
$$= 2\langle z_\delta(t_i) - x_{\delta i}, B u_i - B u_*(t) \rangle + \alpha(\delta) |u_i|_{E^r}^2 + 2\langle z_\delta(t_i) - x_{\delta i}, A(x_{\delta i} - x(t_i)) +$$
$$+ A(x(t_i, u_*) - x(t, u_*)) \rangle + c_4(\delta + h(\delta)), \quad t \in (t_i, t_{i+1}].$$

Отсюда и из (3), (4), (6), (10), (11), (13) следует

$$\dot{w}(t) \leqslant t_{\delta i}(u_i) - t_{\delta i}(u_*(t)) + \alpha(\delta) |u_*(t)|_{E^r}^2 + c_5(\delta + h(\delta)) \leqslant$$
$$\leqslant \alpha(\delta) |u_*(t)|_{E^r}^2 + c_6(\delta + h(\delta) + \varepsilon(\delta))$$

почти для всех $t \in [0, T]$. Интегрируя это неравенство на отрезке $[0, t]$ с учетом равенства $w(0) = 0$ получим

$$w(t) \leqslant \alpha(\delta) \int_0^t |u_*(\tau)|^2 d\tau + c_6 T(\delta + h(\delta) + \varepsilon(\delta)) \quad \forall t \in [0, T].$$

Отсюда с учетом определения (14) функции $w(t)$ имеем

$$|z_\delta(t) - x(t)|_{E^n}^2 + \alpha(\delta) \int_0^t |u_\delta(\tau)|^2 d\tau \leqslant$$
$$\leqslant \alpha(\delta) \int_0^T |u_*(\tau)|^2 d\tau + c_6 T(\delta + h(\delta) + \varepsilon(\delta)) \quad \forall t \in [0, T]. \tag{16}$$

Из оценки (16) следуют два важных неравенства:

$$\max_{0 \leqslant t \leqslant T} |z_\delta(t) - x(t)|_{E^n}^2 \leqslant \alpha(\delta) \|u_*\|_{L_2^r}^2 + c_6 T(\delta + h(\delta) + \varepsilon(\delta)) \quad \forall \delta, \quad 0 < \delta \leqslant \delta_0, \tag{17}$$

$$\max_{0 \leqslant t \leqslant T} \int_0^t |u_\delta(\tau)|^2 d\tau = \int_0^T |u_\delta(\tau)|^2 d\tau = \|u_\delta\|_{L_2^r}^2 \leqslant \|u_*\|_{L_2^r}^2 + c_6 T \frac{\delta + h(\delta) + \varepsilon(\delta)}{\alpha(\delta)}$$
$$\forall \delta, \quad 0 < \delta \leqslant \delta_0. \tag{18}$$

Второе из равенств (8) вытекает из (17) при $\delta \to 0$. Из (18) с учетом условия (7) имеем $\|u_\delta\|_{L_2^r}^2 \leqslant \|u_*\|_{L_2^r}^2 + \sup\limits_{0 < \delta \leqslant \delta_0} c_6 T \frac{\delta + h(\delta) + \varepsilon(\delta)}{\alpha(\delta)} = c_7 < \infty$. Тогда существует последовательность $\{\delta_k\} \to 0$ такая, что $u_k = u_k(t) = u_{\delta_k}(t)$

слабо в $L_2^r[0, T]$ сходится к некоторому управлению $v_* = v_*(t)$, причем $v_* \in U$ в силу слабой компактности U (теорема 8.2.3, 8.2.6). Кроме того, $\lim\limits_{k\to\infty} x(t, u_k) = x(t, v_*)$ $\forall t \in [0, T]$ (примеры 8.2.14–8.2.17). Покажем, что $v_* \in U_*$. Из (5) при $t = t_{j+1}$ имеем $z_\delta(t_{j+1}) = z_\delta(t_j) + \int\limits_{t_j}^{t_{j+1}} (Ax_{\delta j} + Bu_j)dt$, $j = 0, \ldots, N-1$. Суммируя это равенство по j от $j = 0$ до некоторого $j = i-1 \leqslant N-1$ с учетом начального условия $z_\delta(0) = x_{\delta 0}$ получим

$$z_\delta(t_i) = x_{\delta 0} + \sum_{j=0}^{i-1} \int_{t_j}^{t_{j+1}} (Ax_{\delta j} + Bu_j)d\tau. \tag{19}$$

Введем кусочно постоянную функцию $\widetilde{x}_\delta(\tau) = x_{\delta j}$, $\tau \in [t_j, t_{j+1}]$. Учитывая определение (5) функции $u_\delta(\tau)$, равенство (19) теперь можем переписать в виде

$$z_\delta(t_i) = x_{\delta 0} + \int_0^{t_i} (A\widetilde{x}_\delta(\tau) + Bu_\delta(\tau))d\tau.$$

Сложим это равенство с $z_\delta(t)$ из (5). Получим

$$z_\delta(t) = x_{\delta 0} + \int_0^{t} (A\widetilde{x}_\delta(\tau) + Bu_\delta(\tau))d\tau. \tag{20}$$

В силу (3), (10) при $\tau \in [t_j, t_{j+1}]$ имеем: $|\widetilde{x}_\delta(\tau) - x(\tau)| \leqslant |x_{\delta j} - x(t_j)| + |x(t_j) - x(\tau)| \leqslant \delta + c_1 h(\delta) \to 0$ при $\delta \to 0$. Это значит, что $\widetilde{x}_\delta(t) \to x(t)$ при $\delta \to 0$ равномерно на $[0, T]$. Кроме того, $\lim\limits_{k\to\infty} \int\limits_0^t Bu_{\delta_k}(\tau)d\tau = \int\limits_0^t Bv_*(\tau)d\tau$ $\forall t \in [0, T]$, так как $\{u_{\delta_k}\} \to v_*$ слабо в $L_2^r[0, T]$. Тогда из (20) при $\delta = \delta_k \to 0$ с учетом (3) и уже доказанного второго равенства (8) получаем

$$x(t) = x_0 + \int_0^t (Ax(\tau) + Bv_*(\tau))d\tau \quad \forall t \in [0, T].$$

Это означает, что $x(t) = x(t, v_*)$, $0 \leqslant t \leqslant T$, т. е. $v_* \in U_* = \{u = u(t) \in U: x(t, u) = x(t), 0 \leqslant t \leqslant T\}$. Так как $u_* = u_*(t) \in U_*$ — нормальное решение обратной задачи (1), то из оценки (18) при $\delta = \delta_k \to 0$ с учетом слабой полунепрерывности снизу функции $\|u\|^2$ и равенства (7) имеем:

$$\|u_*\|^2_{L_2^r} \leqslant \|v_*\|^2_{L_2^r} \leqslant \varliminf_{k\to\infty} \|u_{\delta_k}\|^2_{L_2^r} \leqslant \varlimsup_{k\to\infty} \|u_{\delta_k}\|^2_{L_2^r} \leqslant \|u_*\|^2_{L_2^r}.$$

Следовательно, существует предел $\lim\limits_{k\to\infty} \|u_{\delta_k}\|^2_{L_2^r} = \|u_*\|^2_{L_2^r} = \|v_*\|^2_{L_2^r}$. Отсюда вытекает, что v_* также является нормальным решением обратной задачи (1). Однако нормальное решение единственно, поэтому $v_* = u_*$. Тем самым доказано, что семейство $\{u_\delta(t), \delta > 0\}$ при $\delta \to 0$ слабо в $L_2^r[0, T]$ сходится к u_* и $\lim\limits_{\delta\to 0} \|u_\delta\|^2_{L_2^r} = \|u_*\|^2_{L_2^r}$. Тогда $\|u_\delta - u_*\|^2_{L_2^r} = \|u_\delta\|^2_{L_2^r} - 2\langle u_*, u_\delta\rangle_{L_2^r} + \|u_*\|^2_{L_2^r} \to 0$ при $\delta \to 0$. Первое из равенств (8) доказано.

Наконец, дифференцируя равенство (20), имеем: $\dot{z}_\delta(t) = A\widetilde{x}_\delta(t) + Bu_\delta(t)$ $\forall t \in [0, T]$, $t \neq t_i$, $i = 1, \ldots, N-1$. Правая часть этого равенства в силу уже доказанных соотношений при $\delta \to 0$ сходится в норме $L_2^r[0, T]$ к $Ax(t) + Bu_*(t) = \dot{x}(t)$. Это значит, что $\lim\limits_{\delta\to 0} \|\dot{z}_\delta - \dot{x}\|_{L_2^r} = 0$. Теорема 1 доказана. □

Из этой теоремы следует, что оператор R_δ, который каждому набору измерений $\{x_{\delta i}, i = 0, \ldots, N-1; \delta\}$ ставит в соответствие управление $u_\delta(t)$, $0 \leqslant t \leqslant T$, определенное методом (4)–(6), является регуляризирующим оператором обратной задачи (1).

Возвращаясь к описанию метода (4)–(6), приведем соображения, поясняющие откуда возникла задача (4). Роль поводыря, как было замечено выше, заключается в том, чтобы отслеживать наблюдаемую траекторию $x(t)$ по ее приближенным значениям $x_{\delta i}$ из (3) с тем, чтобы затем строить управление $u_\delta(t) = u_i$, $t \in [t_i, t_{i+1}]$, $i = 0, \ldots, N-1$, выбирая u_i наилучшим образом в смысле (4). Эту идею наилучшего выбора u_i можно реализовать несколько иначе, сделав его, возможно, более понятным, если u_i выбирать из условия минимума уклонения значения поводыря $z_\delta(t_{i+1}) = z_\delta(t_{i+1}, u)$ от наблюдаемого значения $x_{\delta i}$, считая u_i решением задачи

$$|z_\delta(t_{i+1}, u) - x_{\delta i}|^2 \to \inf, \quad u \in V. \tag{21}$$

С учетом (5) имеем:

$$\begin{aligned} |z_\delta(t_{i+1}, u) - x_{\delta i}|^2 &= |z_\delta(t_i) + (Ax_{\delta i} + Bu)(t_{i+1} - t_i) - x_{\delta i}|^2 = \\ &= |z_\delta(t_i) - x_{\delta i}|^2 + 2\langle z_\delta(t_i) - x_{\delta i}, Bu\rangle(t_{i+1} - t_i) + \\ &+ 2\langle z_\delta(t_i) - x_{\delta i}, Ax_{\delta i}\rangle(t_{i+1} - t_i) + |Ax_{\delta i} + Bu|^2(t_{i+1} - t_i)^2. \end{aligned} \tag{22}$$

Как видим, первое и третье слагаемые из правой части (22) от u не зависят и не влияют на выбор точки минимума u_i в задаче (21), поэтому они могут быть опущены без ущерба для задачи (21). Если теперь пренебречь и четвертым слагаемым из правой части (22), имеющим порядок $O(h^2)$, то от задачи (21) придем к задаче: $2\langle z_\delta(t_i) - x_{\delta i}, Bu\rangle \to \inf$, $u \in V$. Применив к этой задаче метод стабилизации (§ 4), получаем задачу (4).

Метод динамической регуляризации выше мы изложили для простейшей обратной задачи (1). Применение этого метода к более сложным задачам, включая обратные задачи для уравнений с частными производными см., например, в [403; 421; 474; 556; 557; 808].

ГЛАВА 10

Аппроксимация экстремальных задач

Численная реализация многих методов решения задач оптимального управления процессами, описываемыми системами обыкновенных дифференциальных уравнений или уравнений с частными производными, невозможна без использования тех или иных методов приближенного решения возникающих здесь начально-краевых задач, приближенного вычисления встречающихся интегралов.

Для решения начально-краевых задач часто применяют такие методы, как разностный метод, метод конечных элементов, метод прямых, метод характеристик, методы Ритца или Галёркина и т. д., для приближенного вычисления интегралов используют формулы численного интегрирования [74; 89; 480; 481; 630–635]. В результате исходная задача оптимального управления заменяется некоторой последовательностью вспомогательных аппроксимирующих экстремальных задач. Здесь возникают естественные вопросы: будут ли сходиться решения вспомогательных экстремальных задач к решению исходной задачи, каким условиям должны удовлетворять аппроксимирующие экстремальные задачи для обеспечения сходимости? Аналогичные вопросы возникают, когда исходные данные — целевая функция и множество — известны с погрешностью.

В этой главе мы ограничимся рассмотрением разностных аппроксимаций для простейших задач оптимального управления и, кроме того, приведем общие условия аппроксимации экстремальных задач. Вопросам аппроксимации различных классов экстремальных задач посвящены, например, работы [2–4; 111; 114; 120; 129; 140; 141; 152; 153; 170; 184; 303; 339; 340; 360; 362; 363; 402; 461; 462; 464; 467; 501; 502; 504; 593–599; 607; 700; 720].

§ 1. Разностная аппроксимация квадратичной задачи оптимального управления

Рассмотрим следующую задачу оптимального управления:

$$J(u)=|x(T,u)-y|^2\to\inf, \tag{1}$$

$$\dot{x}(t)=A(t)x(t)+B(t)u(t)+f(t),\quad t_0\leqslant t\leqslant T,\quad x(t_0)=x_0, \tag{2}$$

$$u(t)\in U=\{u=u(t)\in L_2^r[t_0,T]\colon u(t)\in V \text{ почти всюду на } [t_0,T]\}, \tag{3}$$

где $A(t)=\{a_{ij}(t)\}$ — матрица порядка $n\times n$, $B(t)=\{b_{ij}(t)\}$ — матрица порядка $n\times r$, $f(t)=\{f^i(t)\}$ матрица порядка $n\times 1$, т. е. вектор-столбец; моменты времени t_0, T, а также точки x_0, $y\in E^n$ заданы; V — заданное множество из E^r, $x(t,u)=x(t)=(x^1(t),\ldots,x^n(t))$ — решение (траектория) задачи (2), соответствующее управлению $u=u(t)=(u^1(t),\ldots,u^r(t))\in L_2^r[t_0,T]$. Будем предполагать, что элементы $a_{ij}(t)$, $b_{ij}(t)$, $f^i(t)$ матриц $A(t)$, $B(t)$, $f(t)$ кусочно непрерывны на отрезке $t_0\leqslant t\leqslant T$.

Напоминаем, что задача (1)–(3) уже рассматривалась в §§ 8.2–8.4. Здесь мы займемся исследованием разностных аппроксимаций этой задачи. Разобьем отрезок $t_0\leqslant t\leqslant T$ на N частей точками $\{t_i,\ i=0,\ldots,N\}$:

$t_0<t_1<\ldots<t_{N-1}<t_N=T$, приняв эти точки в качестве узловых, уравнения (2) заменим разностными уравнениями с помощью простейшей явной схемы Эйлера. В результате придем к следующей задаче:

$$I_N([u]_N)=|x_N([u]_N)-y|^2\to\inf, \tag{4}$$

$$x_{i+1}=x_i+\Delta t_i(A_ix_i+B_iu_i+f_i),\quad i=0,\ldots,N-1, \tag{5}$$

$$[u]_N\in U_N=\{[u]_N=(u_0,u_1,\ldots,u_{N-1})\colon u_i\in V,\ i=0,\ldots,N-1\}, \tag{6}$$

где $\Delta t_i=t_{i+1}-t_i$, $A_i=A(t_i+0)$, $B_i=B(t_i+0)$, $f_i=f(t_i+0)$, $i=0,\ldots,N-1$; $[x([u]_N)]_N=(x_1([u]_N),\ldots,x_N([u]_N))$ — решение задачи (5), соответствующее управлению $[u]_N$.

Введем пространство L_{2N}^r дискретных функций — управлений $[u]_N=(u_0,u_1,\ldots,u_{N-1})$, $[v]_N=(v_0,\ldots,v_{N-1}),\ldots$ — со скалярным произведением

$$\langle[u]_N,[v]_N\rangle_{L_{2N}}=\sum_{i=0}^{N-1}\Delta t_i\langle u_i,v_i\rangle_{E^r}$$

и с нормой

$$\|[u]_N\|_{L_{2N}}=(\langle[u]_N,[u]_N\rangle)^{1/2}=\left(\sum_{i=0}^{N-1}\Delta t_i|u_i|^2_{E^r}\right)^{1/2}.$$

Пространство L_{2N}^r является разностным аналогом пространства $L_2^r[t_0,T]$, соответствующим разбиению $\{t_i,\ i=0,\ldots,N\}$ отрезка $[t_0,T]$.

Таким образом, задаче (1)–(3), рассматриваемой в пространстве $L_2^r[t_0,T]$ при каждом целом $N\geqslant 1$ и разбиении $\{t_i,\ i=0,\ldots,N\}$ отрезка $[t_0,T]$, соответствует дискретная задача оптимального управления (4)–(6), рассматриваемая в пространстве L_{2N}^r.

При каждом фиксированном $N\geqslant 1$ и разбиении $\{t_i,\ i=0,\ldots,N\}$ задачу (4)–(6) можно решать с помощью разностного аналога методов проекции градиента, условного градиента и других методов из § 8.4; при вычислении градиента функции (4) можно пользоваться результатами § 8.6; здесь возможно также использование метода динамического программирования (гл. 7).

Предположим, что при каждом $N\geqslant 1$ и заданном разбиении $\{t_i,\ i=0,\ldots,N\}$ с помощью какого-либо метода минимизации получены приближенное значение $I_{N_*}+\varepsilon_N$ нижней грани I_{N_*} функции (4) при условиях (5), (6) и дискретное управление $[u]_{N\varepsilon}=(u_{0\varepsilon},\ldots,u_{N-1,\varepsilon})\colon u_{i\varepsilon}\in V,\ i=0,\ldots,N-1$, такие, что

$$I_{N_*}\leqslant I_N([u]_{N\varepsilon})\leqslant I_{N_*}+\varepsilon_N, \tag{7}$$

где $\{\varepsilon_N\}$ — положительная последовательность, сходящаяся к нулю.

Возникают вопросы, будет ли сходиться последовательность $\{I_{N_*}\}$ к J_* нижней грани функции (1) при условиях (2), (3), если неограниченно измельчать шаг разбиения $\{t_i,\ i=0,\ldots,N\}$, т. е.

$$\lim_{N\to\infty}\max_{0\leqslant i\leqslant N-1}\Delta t_i=0,$$

и можно ли принять дискретное управление $[u]_{N\varepsilon}$ из (7) в качестве некоторого приближения оптимального управления задачи (1)–(3)?

Для ответа на эти вопросы нам понадобятся некоторые свойства решений задач (2) и (5). Приведем эти свойства. Будем пользоваться обозначениями

$$A_{\max}=\sup_{t_0\leqslant t\leqslant T}\|A(t)\|,\quad B_{\max}=\sup_{t_0\leqslant t\leqslant T}\|B(t)\|,\quad f_{\max}=\max_{t_0\leqslant t\leqslant T}|f(t)|.$$

Если W — произвольное ограниченное множество из $L_2^r[t_0, T]$, т. е. $\sup\limits_{u\in W}\|u\|_{L_2}\leqslant R<\infty$, то тогда

$$\sup_{u\in W}\max_{t_0\leqslant t\leqslant T}|x(t,u)|\leqslant C_0, \tag{8}$$

где

$$C_0=e^{A_{\max}(T-t_0)}[|x_0|+B_{\max}(T-t_0)^{1/2}R+f_{\max}(T-t_0)].$$

В самом деле, по определению решения задачи (2) имеем

$$x(t,u)=\int_{t_0}^{t}[A(\tau)x(\tau,u)+B(\tau)u(\tau)+f(\tau)]d\tau+x_0. \tag{9}$$

При всех t, $t_0\leqslant t\leqslant T$, тогда справедливо неравенство

$$|x(t,u)|\leqslant A_{\max}\int_{t_0}^{t}|x(\tau,u)|d\tau+B_{\max}\int_{t_0}^{T}|u(\tau)|d\tau+f_{\max}(T-t_0)+|x_0|.$$

Отсюда с помощью леммы 6.3.1 получаем оценку (8).

Далее, если W — произвольное ограниченное множество из $L_2^r[t_0, T]$, то

$$\sup_{u\in W}|x(t,u)-x(\tau,u)|\leqslant C_1|t-\tau|^{1/2},\quad t_0\leqslant t,\tau\leqslant T, \tag{10}$$

где

$$C_1=A_{\max}C_0(T-t_0)^{1/2}+B_{\max}R+(T-t_0)^{1/2}f_{\max},$$

постоянная C_0 взята из (8). Действительно, из (9) с помощью оценки (8) имеем

$$|x(t,u)-x(\tau,u)|=\left|\int_{\tau}^{t}[A(\xi)x(\xi,u)+B(\xi)u(\xi)+f(\xi)]d\xi\right|\leqslant$$
$$\leqslant A_{\max}C_0|t-\tau|+B_{\max}|t-\tau|^{1/2}R+|t-\tau|f_{\max}\leqslant C_1(t-\tau)^{1/2}$$

при всех $t, \tau\in[t_0, T]$ и $u\in W$.

Если W — ограниченное множество из $L_\infty^r[t_0, T]$, то вместо (10) можно аналогично получить более лучшее неравенство

$$\sup_{u\in W}|x(t,u)-x(\tau,u)|\leqslant C_1|t-\tau|,\quad t_0\leqslant t,\tau\leqslant T, \tag{11}$$

где $C_1=A_{\max}C_0+B_{\max}\sup\limits_{u\in W}\|u\|_{L_\infty}+f_{\max}$.

Далее, докажем, что если последовательность $\{u_k=u_k(t)\}$ сходится к $u=u(t)$ слабо в $L_2^r[t_0, T]$, то $\{x(t, u_k)\}$ сходится к $x(t, u)$ равномерно на отрезке $[t_0, T]$, т. е.

$$\lim_{k\to\infty}\sup_{t_0\leqslant t\leqslant T}|x(t,u_k)-x(t,u)|=0. \tag{12}$$

Сначала покажем, что $\{x(t, u_k)\}$ сходится к $x(t, u)$ при каждом $t\in[0, T]$. Обозначим $z_k(t)=x(t, u_k)-x(t, u)$, $t_0\leqslant t\leqslant T$. Из равенства (9) имеем

$$z_k(t)=\int_{t_0}^{t}A(\tau)z_k(\tau)d\tau+\int_{t_0}^{t}B(\tau)[u_k(\tau)-u(\tau)]d\tau,\quad t_0\leqslant t\leqslant T.$$

Тогда

$$|z_k(t)|\leqslant A_{\max}\int_{t_0}^{t}|z_k(\tau)|d\tau+\left|\int_{t_0}^{t}B(\tau)(u_k(\tau)-u(\tau))d\tau\right|,\quad t_0\leqslant t\leqslant T.$$

Отсюда с помощью леммы 6.3.1 получим

$$|z_k(t)|\leqslant A_{\max}\int_{t_0}^{t}b_k(\tau)e^{A_{\max}(t-\tau)}d\tau+b(t),\quad t_0\leqslant t\leqslant T, \tag{A.12}$$

где

$$b_k(t)=\left|\int_{t_0}^{t}B(\tau)(u_k(\tau)-u(\tau))d\tau\right|=\left(\sum_{i=1}^{n}\left|\int_{t_0}^{t}\langle b_i(\tau),u_k(\tau)-u(\tau)\rangle d\tau\right|^2\right)^{1/2},\quad t_0\leqslant t\leqslant T,$$

$b_i(\tau)$ — i-я строка матрицы $B(t)$. В силу слабой сходимости $\{u_k(t)\}$ к $u(t)$ имеем

$$\int_{t_0}^{t}\langle b_i(\tau),u_k(\tau)-u(\tau)\rangle d\tau=\int_{t_0}^{T}\langle c_i(\tau),u_k(\tau)-u(\tau)\rangle d\tau\to 0\quad\forall t\in[t_0,T],\quad k\to\infty,$$

где $c_i(\tau)=b_i(\tau)$ при $t_0\leqslant\tau\leqslant t$, $c_i(\tau)=0$ при $t<\tau\leqslant T$. Поэтому $\lim\limits_{k\to\infty}b_k(t)=0$ $\forall t\in[t_0, T]$. Кроме того, $\lim\limits_{k\to\infty}\int_{t_0}^{t}b_k(\tau)e^{A_{\max}(t-\tau)}d\tau=0$ в силу теоремы Лебега о предельном переходе под знаком интеграла [695]. Отсюда и из (A.12) получаем $\lim\limits_{k\to\infty}z_k(t)=0$ или $\lim\limits_{k\to\infty}x(t, u_k)=x(t, u)$ $\forall t\in[t_0, T]$.

Допустим, что $\{x(t, u_k)\}$ не сходится к $x(t, u)$ равномерно на $[t_0, T]$. Это значит, что существует число $\varepsilon_0>0$ такое, что для любого номера $m\geqslant 1$ найдутся номер $k_m>m$ и точка $t_{k_m}\in[t_0, T]$, для которых $|x(t_{k_m}, u_{k_m})-x(t_{k_m}, u)|\geqslant\varepsilon$. Можем считать, что $k_1<k_2<\ldots<k_m<\ldots$. Заметим также, что слабо сходящаяся последовательность $\{u_k\}$ ограничена по норме $L_2^r[t_0, T]$, т. е. $\sup\limits_{k\geqslant 1}\|u_k\|_{L_2}\leqslant R<\infty$. Согласно оценкам (8), (10) тогда семейство функций $\{x(t, u_{k_m})\}$ равномерно ограничено и равностепенно непрерывно на отрезке $[t_0, T]$. В силу теоремы Арцела [695] из $\{x(t, u_{k_m})\}$ можно выбрать подпоследовательность, которая равномерно на $[t_0, T]$ сходится к $x(t, u)$. Без умаления общности можем считать, что сама подпоследовательность $\{x(t, u_{k_m})\}$ равномерно сходится к $x(t, u)$. Это означает, что для любого $\varepsilon>0$, в частности, для $\varepsilon=\varepsilon_0$, найдется номер m_0 такой, что $|x(t, u_{k_m})-x(t, u)|<\varepsilon_0$ для всех $m\geqslant m_0$ и всех $t\in[t_0, T]$. В то же время по определению подпоследовательности $\{x(t, u_{k_m})\}$ имеем $|x(t_{k_m}, u_{k_m})-x(t_{k_m}, u)|\geqslant\varepsilon_0$. Противоречие. Равенство (12) доказано.

Далее, для любых $u, v\in L_2^r[t_0, T]$ справедлива оценка

$$\sup_{t_0\leqslant t\leqslant T}|x(t,u)-x(t,v)|\leqslant C_2\|u-v\|_{L_2}, \tag{13}$$

где

$$C_2=e^{A_{\max}(T-t_0)}(T-t_0)^{1/2}B_{\max}.$$

В самом деле, из (9) следует, что

$$|x(t,u)-x(t,v)|=\left|\int_{t_0}^{t}[A(\tau)(x(\tau,u)-x(\tau,v))+B(\tau)(u(\tau)-v(\tau))]d\tau\right|\leqslant$$
$$\leqslant A_{\max}\int_{t_0}^{t}|x(\tau,u)-x(\tau,v)|d\tau+B_{\max}(T-t_0)^{1/2}\left(\int_{t_0}^{T}|u(\tau)-v(\tau)|^2d\tau\right)^{1/2}.$$

Отсюда с помощью леммы 6.3.1 получаем оценку (13).

На любом ограниченном множестве W из $L_2^r[t_0, T]$ функция (1) удовлетворяет условию Липшица

$$|J(u) - J(v)| \leqslant C_3 \|u - v\|_{L_2}, \quad u, v \in W, \tag{14}$$

где $C_3 = (4C_0 + 2|y|)C_2$, постоянные C_0, C_2 взяты из (11), (13). Действительно,

$$|J(u) - J(v)| = \big||x(T, u) - y|^2 - |x(T, v) - y|^2\big| =$$

$$= |2\langle x(T, v) + \theta(x(T, u) - x(T, v)) - y,\ x(T, u) - x(T, v)\rangle|, \quad 0 < \theta < 1,$$

так что

$$|J(u) - J(v)| \leqslant 2(|x(T, u)| + |x(T, v)| + |y|)|x(T, u) - x(T, v)|.$$

Отсюда и из оценок (8), (13) следует неравенство (14).

Если W_N — произвольное ограниченное множество из L_{2N}^r, т. е. $\sup\limits_{W_N} \|[u]_N\| \leqslant R < \infty$, и, кроме того,

$$d_N = \max_{0 \leqslant i \leqslant N-1} \Delta t_i \leqslant \frac{(T - t_0)M}{N}, \quad M = \text{const} > 0, \tag{15}$$

то

$$\sup_{[u]_N \in W_N} \max_{0 \leqslant i \leqslant N} |x_i([u]_N)| \leqslant C_4, \tag{16}$$

где

$$C_4 = e^{A_{\max}(T - t_0)M}(|x_0| + B_{\max}(T - t_0)^{1/2}R + f_{\max}(T - t_0)).$$

В самом деле, из (5) имеем

$$x_{i+1} = \sum_{j=0}^{i} \Delta t_j (A_j x_j + B_j u_j + f_j) + x_0, \quad i = 0, \ldots, N. \tag{17}$$

Следовательно,

$$|x_{i+1}| \leqslant A_{\max} d_N \sum_{j=0}^{i} |x_j| + B_{\max}(T - t_0)^{1/2} \|[u]_N\|_{L_{2N}} + (T - t_0) f_{\max} + |x_0|,$$
$$i = 0, \ldots, N-1.$$

С помощью леммы 8.6.1 отсюда получаем

$$|x_i| \leqslant (1 + A_{\max} d_N)^i (|x_0| + B_{\max}(T - t_0)^{1/2}R + f_{\max}(T - t_0)), \quad i = 0, \ldots, N. \tag{18}$$

Поскольку $1 + x \leqslant e^x$ при всех действительных x, то с учетом условия (15) имеем

$$(1 + A_{\max} d_N)^i \leqslant \left(e^{A_{\max} d_N}\right)^i \leqslant e^{A_{\max} d_N N} \leqslant e^{A_{\max}(T - t_0)M}. \tag{19}$$

Отсюда и из (18) следует оценка (16).

Для исследования связи между задачами (1)–(3) и (4)–(6) нам ниже будут полезны следующие два отображения Q_N и P_N. Отображение Q_N, действующее из пространства $L_2^r[t_0, T]$ в L_{2N}^r определяется так:

$$Q_N(u) = (u_0, u_1, \ldots, u_{N-1}) : u_i = \frac{1}{\Delta t_i} \int_{t_i}^{t_{i+1}} u(t)\,dt, \quad i = 0, \ldots, N-1, \tag{20}$$

отображение P_N, действующее из L_{2N}^r в $L_2^r[t_0, T]$ определяется следующим образом:

$$P_N([u]_N) = u_i \quad \text{при} \quad t_i < t \leqslant t_{i+1}, \quad i = 0, \ldots, N-1. \tag{21}$$

Из (20), (21) непосредственно следует, что

$$\|Q_N(u)\|_{L_{2N}}^2 = \sum_{i=0}^{N-1} \left(\frac{1}{\Delta t_i} \int_{t_i}^{t_{i+1}} u(t)\,dt\right)^2 \Delta t_i = \sum_{i=0}^{N-1} \frac{1}{\Delta t_i} \left(\int_{t_i}^{t_{i+1}} u(t)\,dt\right)^2 \leqslant$$
$$\leqslant \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} |u(t)|^2\,dt = \int_{t_0}^{T} |u(t)|^2\,dt = \|u\|_{L_2}^2, \tag{22}$$

$$\|P_N([u]_N)\|_{L_2}^2 = \int_{t_0}^{T} |P_N([u]_N)|^2\,dt = \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} |u_i|^2\,dt = \sum_{i=0}^{N-1} \Delta t_i |u_i|^2 = \|[u]_N\|_{L_{2N}}^2. \tag{23}$$

Ниже докажем несколько лемм, связанных со свойствами отображений Q_N и P_N.

Лемма 1. *Пусть V — выпуклое замкнутое множество из E^r, а управление $u = u(t)$ принадлежит $L_2^r[t_0, T]$ и $u(t) \in V$ почти всюду на отрезке $[t_0, T]$. Тогда*

$$u_i = \frac{1}{\Delta t_i} \int_{t_i}^{t_{i+1}} u(t)\,dt \in V$$

при всех $t_i, t_{i+1} \in [t_0, T]$, $t_i < t_{i+1}$.

Доказательство. Сначала установим, что всякое выпуклое замкнутое множество V из E^r является пересечением полупространств, образованных всевозможными опорными гиперплоскостями к множеству V и содержащих V (см. определение 4.5.2). Возьмем произвольную граничную точку v множества V. Пусть c_v — опорный вектор множества V в точке v, т. е. $c_v \neq 0$ и $\langle c_v, u - v\rangle \geqslant 0$ при всех $u \in V$. Обозначим

$$W = \bigcap_{v \in \Gamma p V} \{u: \langle c_v, u - v\rangle \geqslant 0\}.$$

Надо показать, что $W = V$. Если $u \in V$, то $\langle c_v, u - v\rangle \geqslant 0$ для всех $v \in \Gamma p V$, поэтому $u \in W$. Это значит, что $V \subseteq W$.

Покажем, что справедливо включение $W \subseteq V$. Допустим противное: пусть существует точка $w \in W$, но $w \notin V$. Тогда по теореме 4.5.1 множество V и точка w сильно отделимы, и гиперплоскость $\langle c_v, u - v\rangle = 0$, где $v = P_V(w)$ — проекция точки w на множество V, обладает свойствами: $\langle c_v, u - v\rangle \geqslant 0$ при всех $u \in V$, а $\langle c_v, w - v\rangle < 0$. Но поскольку $v \in \Gamma p V$ и $w \in W$, то $\langle c_v, w - v\rangle \geqslant 0$ по определению множества W. Противоречие. Следовательно, $W \subseteq V$. Требуемое равенство $W = V$ доказано.

Возьмем произвольное управление $u = u(t)$, удовлетворяющее условию леммы. Пусть v — любая точка из $\Gamma p V$, c_v — опорный вектор к V в точке v. Тогда $\langle c_v, u(t) - v\rangle \geqslant 0$ почти всюду на $[t_0, T]$. Интегрируя это неравенство на отрезке $[t_i, t_{i+1}]$, получим

$$\frac{1}{\Delta t_i} \int_{t_i}^{t_{i+1}} \langle c_v, u(t)\rangle\,dt = \left\langle c_v, \frac{1}{\Delta t_i} \int_{t_i}^{t_{i+1}} u(t)\,dt \right\rangle = \langle c_v, u_i\rangle \geqslant \langle c_v, v\rangle$$

или $\langle c_v, u_i - v\rangle \geqslant 0$ для всех $v \in \Gamma p V$. Следовательно, $u_i \in W = V$. Лемма 1 доказана. □

Лемма 2. *Пусть матрицы $A(t), B(t), f(t)$ кусочно непрерывны на отрезке $t_0 \leqslant t \leqslant T$, разбиения $\{t_i, i = 0, \ldots, N\}$ отрезка $[t_0, T]$ удовлетворяют условию (15). Пусть W и W_N — произвольные ограниченные*

множества из $L_2^r[t_0, T]$ *и* L_{2N}^r *соответственно, т. е.* $\sup\limits_{W}\|u\|_{L_2} \leqslant R < \infty$, $\sup\limits_{W_N}\|[u]_N\|_{L_{2N}} \leqslant R < \infty$. *Тогда*

$$\sup_{u\in W}\max_{0\leqslant i\leqslant N}|x(t_i, u) - x_i(Q_N(u))| \leqslant \delta_N, \tag{24}$$

$$\sup_{[u]_N\in W_N}\max_{0\leqslant i\leqslant N}|x(t_i, P_N([u]_N)) - x_i([u]_N)| \leqslant \delta_N, \quad N = 1, 2, \ldots, \tag{25}$$

где

$$\delta_N = e^{A_{\max}(T-t_0)M}\Big[\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\Big(\|A(\tau)-A_i\|\sup_{\|u\|_{L_2}}\max_{t_0\leqslant t\leqslant T}|x(t, u)|+$$

$$+A_{\max}\sup_{\|u\|_{L_2}}|x(\tau, u)-x(t_i, u)|+|f(\tau)-f_i|\Big)d\tau\Big]+$$

$$+R\Big(\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\|B(\tau)-B_i\|^2 d\tau\Big)^{1/2}\to 0 \quad при \quad N\to\infty. \tag{26}$$

Доказательство. Возьмем произвольные $u\in L_2^r[t_0, T]$ и $[u]_N\in L_{2N}^r$, для которых $\|u\|_{L_2}\leqslant R$, $\|[u]_N\|_{L_{2N}}\leqslant R$, и соответствующие им решения $x(t, u)$ и $[x([u]_N)]_N$ задач (2) и (5). Из равенств (9) и (17) следует

$$|x(t_{i+1},u)-x_{i+1}([u]_N)| = \Big|\sum_{j=0}^{i}\int_{t_j}^{t_{j+1}}[(A(\tau)-A_j)x(\tau,u)+A_j(x(\tau,u)-x(t_j,u))+$$

$$+A_j(x(t_j, u)-x_j([u]_N))+(B(\tau)-B_j)u(\tau)+B_j(u(\tau)-u_j)+(f(\tau)-f_j)]d\tau\Big| \leqslant$$

$$\leqslant A_{\max}d_N\sum_{j=0}^{i}|x(t_j, u) - x_j([u]_N)| + \sum_{j=0}^{N-1}\int_{t_j}^{t_{j+1}}(\|A(\tau) - A_j\||x(\tau, u)| +$$

$$+A_{\max}|x(\tau, u)-x(t_j, u)|+|f(\tau)-f_j|)d\tau+\Big(\sum_{j=0}^{N-1}\int_{t_j}^{t_{j+1}}\|B(\tau)-B_j\|^2 d\tau\Big)^{1/2}\|u\|_{L_2}+$$

$$+\sum_{j=0}^{N-1}\|B_i\|\Big|\int_{t_j}^{t_{j+1}}(u(\tau) - u_j)d\tau\Big|, \quad i = 0, \ldots, N-1.$$

Отсюда с помощью леммы 8.6.1 и неравенств (18) получаем

$$|x(t_i, u)-x_i([u]_N)| \leqslant \delta_N+e^{A_{\max}(T-t_0)M}B_{\max}\sum_{j=0}^{N-1}\Big|\int_{t_j}^{t_{j+1}}(u(\tau)-u_j)d\tau\Big|, \tag{27}$$

$$i = 0, \ldots, N,$$

для всех $u, [u]_N$, $\|u\|_{L_2}\leqslant R$, $\|[u]_N\|_{L_{2N}}\leqslant R$. Однако, если $\|u\|_{L_2}\leqslant R$, то в силу (22) имеем $\|Q_N(u)\|_{L_{2N}}\leqslant R$. Поэтому, учитывая, что $\int_{t_j}^{t_{j+1}}(u(\tau) - Q_N(u))d\tau = 0$, $j = 0, \ldots, N-1$, из (27) при $[u]_N = Q_N(u)$ получим оценку (24). Аналогично, если $\|[u]_N\|_{L_{2N}}\leqslant R$, то согласно (23) тогда $\|P_N([u]_N)\|_{L_2}\leqslant R$ и поэтому, учитывая, что $\int_{t_j}^{t_{j+1}}(P_N([u]_N) - u_j)d\tau = 0$, $j = 0, \ldots, N-1$, из (27) при $u = P_N([u]_N)$ получим оценку (25).

Остается доказать, что величина δ_N, определяемая равенством (26), стремится к нулю при $N\to\infty$. Пусть $\theta_1, \ldots, \theta_s$ — точки разрыва элементов матрицы $A(t)$ на отрезке $[t_0, T]$: $t_0 = \theta_0 < \theta_1 < \ldots < \theta_s < T = \theta_{s+1}$. Доопределим

матрицу $A(t)$, $\theta_i < t < \theta_{i+1}$, при $t = \theta_i$ и $t = \theta_{i+1}$ предельными значениями при $t\to\theta_i+0$ и $t\to\theta_{i+1}-0$; тогда $A(t)$ будет непрерывной и, следовательно, равномерно непрерывной на отрезке $[\theta_i, \theta_{i+1}]$, $i = 0, \ldots, s$. Отсюда следует, что для любого $\varepsilon > 0$ можно указать номер $N_0 = N_0(\varepsilon)$ такой, что как только $N\geqslant N_0$, $0 < t_{i+1} - t_i = \Delta t_i \leqslant d_N$ и отрезок $[t_i, t_{i+1}]$ не содержит ни одной точки $\{\theta_j\}$ разрыва $A(t)$, то $\|A(\tau)-A_i\| < \varepsilon$ для всех τ, $t_i\leqslant\tau\leqslant t_{i+1}$. Имеем

$$\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\|A(\tau)-A_i\|d\tau = \sum_i{}'\int_{t_i}^{t_{i+1}}\|A(\tau)-A_i\|d\tau+\sum_i{}''\int_{t_i}^{t_{i+1}}\|A(\tau)-A_i\|d\tau, \tag{28}$$

где $\sum_i{}'$ означает суммирование по тем номерам i, для которых отрезок $[t_i, t_{i+1}]$ не содержит ни одной точки разрыва $A(t)$, а $\sum_i{}''$ — суммирование по остальным номерам i. Тогда

$$\sum_i{}'\int_{t_i}^{t_{i+1}}\|A(\tau) - A_i\|d\tau \leqslant \sum_i{}'\int_{t_i}^{t_{i+1}}\varepsilon d\tau \leqslant (T - t_0)\varepsilon \quad \text{для всех} \quad N\geqslant N_0.$$

Взяв при необходимости номер N_0 еще большим, можем сделать

$$\sum_i{}''\int_{t_i}^{t_{i+1}}\|A(\tau) - A_i\|d\tau \leqslant 2A_{\max}\sum_i{}''\Delta t_i \leqslant 2A_{\max}(s+1)d_N < \varepsilon$$

для всех $N\geqslant N_0$. Отсюда и из (28) следует, что

$$\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\|A(\tau) - A_i\|d\tau\to 0 \quad \text{при} \quad N\to\infty.$$

Аналогично доказывается, что

$$\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\|B(\tau) - B_i\|^2 d\tau\to 0, \quad \sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}|f(\tau) - f_i|d\tau\to 0$$

при $N\to\infty$. Отсюда и из неравенств (8), (10) следует, что $\delta_N\to 0$ при $N\to\infty$. Лемма 2 доказана. □

Лемма 3. *Пусть выполнены все условия леммы* 2, *пусть* $u = u(t)$ — *произвольное управление из* W. *Тогда*

$$|I_N(Q_N(u)) - J(u)| \leqslant C_5\delta_N, \quad C_5 = 2C_0 + 2C_4 + 2|y|, \quad N = 1, 2, \ldots \tag{29}$$

где C_0, C_4 — *постоянные из* (8), (16), *а величина* δ_N *определена формулой* (26).

Доказательство. Заметим, что

$$|I_N(Q_N(u)) - J(u)| = ||x_N(Q_N(u)) - y|^2 - |x(T, u) - y|^2| =$$

$$= |2\langle x(T, u) + \theta(x_N(Q_N(u)) - x(T, u)) - y, x_N(Q_N(u)) - x(T, u)\rangle|,$$

$0 < \theta < 1$. Отсюда и из оценок (8), (16), (24) следует утверждение леммы 3. □

Лемма 4. *Пусть выполнены все условия леммы* 2 *и пусть* $[u]_N$ — *произвольное управление из* W_N. *Тогда*

$$|J(P_N([u]_N)) - I_N([u]_N)| \leqslant C_6\delta_N, \quad C_6 = 2C_4 + 2C_0 + 2|y|, \quad N = 1, 2, \ldots, \tag{30}$$

где C_0, C_4 — *постоянные из* (8), (16), δ_N *определена формулой* (26).

Доказательство. Имеем

$$|J(P_N([u]_N)) - I_N([u]_N)| = |2\langle x_N([u]_N) + \theta(x(T, P_N([u]_N)) - x_N([u]_N)) - y, x(T, P_N([u]_N)) - x_N([u]_N)\rangle|, \quad 0 < \theta < 1.$$

Отсюда и из оценок (8), (16), (25) следует утверждение леммы 4. □

Теорема 1. *Пусть матрицы $A(t), B(t), f(t)$ кусочно непрерывны на отрезке $t_0 \leqslant t \leqslant T$, V — выпуклое замкнутое ограниченное множество из E^r, разбиения $\{t_i, i = 0, \ldots, N\}$ отрезка $[t_0, T]$ удовлетворяют условию* (15). *Пусть J_* — нижняя грань функции* (1) *при условиях* (2), (3), I_{N*} — *нижняя грань функции* (4) *при условиях* (5), (6), *последовательность $\{[u]_{N\varepsilon}\}$ определена из условий* (7). *Тогда $\lim\limits_{N\to\infty} I_{N*} = J_*$ и справедливы оценки*

$$-C_6\delta_N \leqslant I_{N*} - J_* \leqslant C_5\delta_N, \quad N = 1, 2, \ldots, \tag{31}$$

$$0 \leqslant J(P_N([u]_{N\varepsilon})) - J_* \leqslant (C_5 + C_6)\delta_N + \varepsilon_N, \quad N = 1, 2, \ldots, \tag{32}$$

где постоянные C_5, C_6 взяты из оценок (29), (30) *при $W = U$ и $W_N = U_N$ соответственно, а величина δ_N определена формулой* (26).

Доказательство. При сделанных предположениях $U_* \neq \varnothing$ (пример 8.2.15) возьмем какое-либо управление $u_* \in U_*$. Согласно лемме 1 имеем $Q_N(u_*) \in U_N$. Отсюда и из леммы 3 следует

$$I_{N*} \leqslant I_N(Q_N(u_*)) \leqslant J(u_*) + C_5\delta_N = J_* + C_5\delta_N, \quad N = 1, 2, \ldots \tag{33}$$

Далее, функция (4) конечного числа переменных $[u]_N = (u_0, \ldots, u_{N-1})$ на компактном множестве U_N достигает своей нижней грани, т. е. $I_{N*} > -\infty$, $U_{N*} \neq \varnothing$. Возьмем какое-нибудь управление $[u]_{N*} \in U_{N*}$. Из (21) непосредственно следует, что $P_N([u]_{N*}) \in U$. Из леммы 4 тогда получим

$$J_* \leqslant J(P_N([u]_{N*})) \leqslant I_N([u]_{N*}) + C_6\delta_N = I_{N*} + C_6\delta_N, \quad N = 1, 2, \ldots \tag{34}$$

Из неравенств (33), (34) следует оценка (31). Так как согласно лемме 2 имеем $\lim\limits_{N\to\infty} \delta_N = 0$, то $\lim\limits_{N\to\infty} I_{N*} = J_*$.

Рассмотрим последовательность $\{[u]_{N*}\}$ из (7). Тогда $P_N([u]_{N\varepsilon}) \in U$ и $0 \leqslant J(P_N([u]_{N\varepsilon})) - J_* = [J(P_N([u]_{N\varepsilon})) - I_N([u]_{N\varepsilon})] + [I_N([u]_{N\varepsilon}) - I_{N*}] + [I_{N*} - J_*]$, $N = 1, 2, \ldots$ Отсюда и из неравенств (7), (30), (31) следует оценка (32). Тем самым установлено, что $\{P_N([u]_{N\varepsilon})\}$ — минимизирующая последовательность задачи (1)–(3). Рассуждая так же, как при доказательстве теоремы 8.2.4, можно показать, что любая точка v_*, являющаяся слабым пределом какой-либо подпоследовательности из $\{P_N([u]_{N\varepsilon})\}$, принадлежит U_*. □

Заметим, что множество U, определяемое условиями (3), при выполнении условий теоремы 1 ограничено в метрике $L^r_\infty[t_0, T]$ и, следовательно, справедливо неравенство (11). Поэтому, если матрицы $A(t), B(t), f(t)$ на интервалах непрерывности удовлетворяют условию Липшица (например, если эти матрицы не зависят от времени), то из (8), (11), (26) и (31) следует оценка

$$|I_{N*} - J_*| \leqslant C_7 d_N = C_7 \max_{0 \leqslant i \leqslant N-1} \Delta t_i. \tag{35}$$

Заметим также, что при доказательстве теоремы 1 неравенства (29), (30) были использованы не полностью: для получения оценки (31) оказалось

достаточно справедливости неравенств (33), (34). В следующем параграфе будет выяснено, что неравенства (33), (34) в некотором смысле являются необходимыми для справедливости равенства $\lim\limits_{N\to\infty} I_{N*} = J_*$.

В заключение предлагаем читателю в качестве упражнения доказать оценки типа (31), (32) в предположении, что в задаче (1)–(3) элементы матриц $A(t), B(t), f(t)$ принадлежат $L_\infty[t_0, T]$, а в (4)–(6) принято

$$A_i = \frac{1}{\Delta t_i}\int_{t_i}^{t_{i+1}} A(t)dt, \quad B_i = \frac{1}{\Delta t_i}\int_{t_i}^{t_{i+1}} B(t)dt, \quad f_i = \frac{1}{\Delta t_i}\int_{t_i}^{t_{i+1}} f(t)dt, \quad i = 0, \ldots, N-1.$$

§ 2. Общие условия аппроксимации

Перейдем к рассмотрению общей задачи минимизации и сформулируем критерий аппроксимации по функции. Пусть $X, X_1, X_2, \ldots, X_N, \ldots$ — некоторые множества произвольной природы. Элементы множества X будем обозначать через u, а элементы X_N — через $[u]_N$. Пусть U — некоторое непустое множество из X, U_N — непустое множество из X_N, $N = 1, 2, \ldots$ Пусть функции $J(u)$, $I_1([u]_1), \ldots, I_N([u]_N), \ldots$ определены соответственно на множествах $U, U_1, \ldots, U_N, \ldots$ Рассмотрим задачу

$$J(u) \to \inf, \quad u \in U, \tag{1}$$

и последовательность «аппроксимирующих» ее задач

$$I_N([u]_N) \to \inf, \quad [u]_N \in U_N, \quad N = 1, 2, \ldots \tag{2}$$

Определение 1. Обозначим

$$J_* = \inf_U J(u), \quad I_{N*} = \inf_{U_N} I_N([u]_N), \quad N = 1, 2, \ldots$$

Скажем, что последовательность задач (2) аппроксимирует задачу (1) по функции, если

$$\lim_{N\to\infty} I_{N*} = J_*. \tag{3}$$

Нетрудно видеть, что в схему задач (1), (2) укладываются задачи (1.1)–(1.3) и (1.4)–(1.6), в которых роль множеств X и X_N играют пространства $L^r_2[t_0, T]$ и L^r_{2N} соответственно, множества U и U_N описываются условиями (1.3) и (1.6), причем в теореме 1.1 сформулированы условия, гарантирующие аппроксимацию по функции. Заметим, что в задачах (1.1)–(1.3) и (1.4)–(1.6) множества X и X_N имеют различную природу: в задаче (1.1)–(1.3) управления и траектории зависят от непрерывного времени, а в (1.4)–(1.6) — от дискретного времени.

В § 1 для задачи (1.1)–(1.3) аппроксимирующая последовательность задач (1.4)–(1.6) была получена с помощью разностной аппроксимации уравнений (1.2) и множества (1.3). В этой и других задачах оптимального управления, рассмотренных в гл. 6–9, при конструировании аппроксимирующих задач наряду с разностными методами могут быть использованы и другие методы, такие, как например, метод конечных элементов, метод прямых, возможна аппроксимация управления с помощью частичных сумм ряда, представляющего собой разложение по каким-нибудь базисным функциям

или по степеням какого-либо параметра и т. п. Все эти методы аппроксимации экстремальных задач также укладываются в схему задач (1), (2). Разумеется, в (1), (2) не исключается и такая возможность, когда множества X и X_N имеют одну и ту же природу, а множества U_N и функции $I_N([u]_N) = I_N(u)$, $N = 1, 2, \ldots$, представляют собой приближенно заданные множество U и функцию $J(u)$ соответственно.

Таким образом, задачи (1), (2) позволяют охватить широкие классы экстремальных задач и их аппроксимаций. Возникает важный вопрос: каким условиям должна удовлетворять последовательность задач (2) для того, чтобы она аппроксимировала задачу (1) по функции, т. е. чтобы выполнялось равенство (3)?

1. Следующая теорема дает один из возможных ответов на этот вопрос, указывает подход к построению последовательности аппроксимирующих задач в конкретных экстремальных задачах, к исследованию сходимости в смысле равенства (3).

Теорема 1. *Для того чтобы последовательность задач* (2) *аппроксимировала задачу* (1) *по функции, необходимо и достаточно, чтобы существовали отображения* $Q_N\colon X \to X_N$ *и* $P_N\colon X_N \to X$ *такие, что*

1) *для любой точки* $u \in U$ *справедливо включение* $Q_N(u) \in U_N$, $N = 1, 2, \ldots$, *и*

$$\overline{\lim_{N\to\infty}}\,(I_N(Q_N(u)) - J(u)) \leqslant 0; \tag{4}$$

2) *для любой точки* $[u]_N \in U_N$ *справедливо включение* $P_N([u]_N) \in U$, $N = 1, 2, \ldots$, *и*

$$\overline{\lim_{N\to\infty}}\,(J(P_N([u]_N)) - I_N([u]_N)) \leqslant 0. \tag{5}$$

Если выполнены условия 1), 2) *и, кроме того, последовательности* $\{\beta_N\}$, $\{\gamma_N\}$ *неотрицательны, стремятся к нулю и таковы, что*

$$I_N(Q_N(u)) - J(u) \leqslant \beta_N, \quad u \in U, \quad N = 1, 2, \ldots, \tag{6}$$

$$J(P_N([u]_N)) - I_N([u]_N) \leqslant \gamma_N, \quad [u]_N \in U_N, \quad N = 1, 2, \ldots, \tag{7}$$

то справедлива оценка

$$-\gamma_N \leqslant I_{N*} - J_* \leqslant \beta_N, \quad N = 1, 2, \ldots \tag{8}$$

Наконец, если последовательность $\{[u]_{N\varepsilon}\}$ *такова, что*

$$[u]_{N\varepsilon} \in U_N, \quad I_{N*} \leqslant I_N([u]_{N\varepsilon}) \leqslant I_{N*} + \varepsilon_N, \quad \lim_{N\to\infty} \varepsilon_N = 0, \tag{9}$$

то при выполнении условий 1), 2) $\lim\limits_{N\to\infty} J(P_N([u]_{N\varepsilon})) = J_*$, *а из условий* (6), (7) *следует оценка*

$$0 \leqslant J(P_N([u]_{N\varepsilon})) - J_* \leqslant \beta_N + \gamma_N + \varepsilon_N, \quad N = 1, 2, \ldots \tag{10}$$

Доказательство. Необходимость. Пусть последовательность задач (2) аппроксимирует задачу (1) по функции, т. е. справедливо равенство (3). Возьмем произвольные последовательности $\{v_N\}$, $\{[v]_N\}$ такие, что

$$v_N \in U, \quad N = 1, 2, \ldots; \quad \lim_{N\to\infty} (J(v_N) - J_*) = 0, \tag{11}$$

$$[v]_N \in U_N, \quad N = 1, 2, \ldots; \quad \lim_{N\to\infty} (I_N([v]_N) - I_{N*}) = 0. \tag{12}$$

Существование таких последовательностей вытекает из определения нижней грани. Определим отображения Q_N и P_N следующим образом:

$$Q_N(u) = [v]_N \quad \text{при всех} \quad u \in X, \quad N = 1, 2, \ldots$$

$$P_N([u]_N) = v_N \quad \text{при всех} \quad [u]_N \in X_N, \quad N = 1, 2, \ldots$$

Ясно, что $Q_N(u) \in U_N$ при всех $u \in U$, $P_N([u]_N) \in U$ при всех $[u]_N \in U_N$, $N = 1, 2, \ldots$ Далее, так как $J_* \leqslant J(u)$ при $u \in U$, то

$$I_N(Q_N(u)) - J(u) = (I_N(Q_N(u)) - I_{N*}) + (I_{N*} - J_*) + (J_* - J(u)) \leqslant (I_N([v]_N) - I_{N*}) + (I_{N*} - J_*)$$

при всех $u \in U$, $N = 1, 2, \ldots$ Переходя в этом неравенстве к верхнему пределу при $N \to \infty$, с учетом условий (3), (12) придем к неравенству (4). Наконец, поскольку $I_{N*} \leqslant I_N([u]_N)$ при $[u]_N \in U_N$, то

$$J(P_N([u]_N)) - I_N([u]_N) = (J(P_N([u]_N)) - J_*) + (J_* - I_{N*}) + (I_{N*} - I_N([u]_N)) \leqslant (J(v_N) - J_*) + (J_* - I_{N*})$$

при всех $[u]_N \in U_N$, $N = 1, 2, \ldots$ Переходя в этом неравенстве к верхнему пределу при $N \to \infty$, с учетом условий (3), (11) получим неравенство (5). Необходимость доказана.

Достаточность. Пусть для некоторых отображений Q_N, P_N выполнены условия 1), 2). Докажем, что тогда справедливо равенство (3). Поскольку $Q_N(u) \in U_N$ при всех $u \in U$, то $I_{N*} \leqslant I_N(Q_N(u))$ и

$$I_{N*} - J_* = (I_{N*} - I_N(Q_N(u))) + (I_N(Q_N(u)) - J(u)) + (J(u) - J_*) \leqslant (I_N(Q_N(u)) - J(u)) + (J(u) - J_*)$$

при всех $u \in U$ и $N = 1, 2, \ldots$ Переходя в этом неравенстве к верхнему пределу при $N \to \infty$, с учетом условия (4) получим

$$\overline{\lim_{N\to\infty}}\,(I_{N*} - J_*) \leqslant J(u) - J_*, \quad u \in U.$$

Левая часть этого неравенства не зависит от u, поэтому, переходя в правой части к нижней грани по $u \in U$, будем иметь $\overline{\lim\limits_{N\to\infty}}\,(I_{N*} - J_*) \leqslant J_* - J_* = 0$. С другой стороны, поскольку $P_N([u]_N) \in U$ при всех $[u]_N \in U_N$, то $J_* \leqslant J(P_N([u]_N))$ и

$$J_* - I_{N*} = (J_* - J(P_N([u]_N))) + (J(P_N([u]_N)) - I_N([u]_N)) + (I_N([u]_N) - I_{N*}) \leqslant (J(P_N([u]_N)) - I_N([u]_N)) + (I_N([u]_N) - I_{N*})$$

для всех $[u]_N \in U_N$, $N = 1, 2, \ldots$ При $N \to \infty$ отсюда с учетом условия (5) получим

$$\overline{\lim_{N\to\infty}}\,(J_* - I_{N*}) \leqslant \overline{\lim_{N\to\infty}}\,(I_N([u]_N) - I_{N*})$$

при любом выборе $[u]_N \in U_N$, $N = 1, 2, \ldots$ В частности, если, пользуясь определением нижней грани I_{N*}, при каждом $N = 1, 2, \ldots$ взять $[u]_N \in U_N$ так, чтобы $I_N([u]_N) \leqslant I_{N*} + 1/N$, то из предыдущего неравенства будем иметь

$$\overline{\lim_{N\to\infty}}\,(J_* - I_{N*}) \leqslant \lim_{N\to\infty} 1/N = 0.$$

Итак, из вышеизложенного следует, что

$$0 \leqslant -\overline{\lim_{N\to\infty}}(J_* - I_{N*}) = \underline{\lim_{N\to\infty}}(I_{N*} - J_*) \leqslant \overline{\lim_{N\to\infty}}(I_{N*} - J_*) \leqslant 0,$$

т. е.

$$\lim_{N\to\infty}(I_{N*} - J_*) = \overline{\lim_{N\to\infty}}(I_{N*} - J_*) = 0.$$

Отсюда следует, что предел $\lim\limits_{N\to\infty} J_{N*}$ существует и равен J_*. Достаточность доказана.

Докажем оценку (8). Так как $Q_N(u) \in U_N$ при всех $u \in U$, то с учетом оценки (6) имеем $I_{N*} \leqslant I_N(Q_N(u)) \leqslant J(u) + \beta_N$, или $I_{N*} \leqslant J(u) + \beta_N$ для всех $u \in U$, $N = 1, 2, \ldots$ Переходя в правой части последнего неравенства к нижней грани по $u \in U$, получим $I_{N*} \leqslant J_* + \beta_N$, или $I_{N*} - J_* \leqslant \beta_N$, $N = 1, 2, \ldots$ Правое неравенство (8) доказано. Далее, так как $P_N([u]_N) \in U$ при всех $[u]_N \in U_N$, то с учетом оценки (7) имеем $J_* \leqslant J(P_N([u]_N)) \leqslant J_N([u]_N) + \gamma_N$, или $J_* \leqslant I_N([u]_N) + \gamma_N$ для всех $[u]_N \in U_N$, $N = 1, 2, \ldots$ Переходя в правой части этого неравенства к нижней грани по $[u]_N \in U_N$, получим $J_* \leqslant I_{N*} + \gamma_N$, или $-\gamma_N \leqslant I_{N*} - J_*$, $N = 1, 2, \ldots$ Левая оценка (8) также доказана.

Итак, пусть последовательность $[u]_{N\varepsilon}$ удовлетворяет условиям (9). Тогда

$$0 \leqslant J(P_N([u]_{N\varepsilon})) - J_* = (J(P_N([u]_{N\varepsilon})) - I_N([u]_{N\varepsilon})) + (I_N([u]_{N\varepsilon}) - I_{N*}) +$$
$$+ (I_{N*} - J_*) \leqslant (J(P_N([u]_{N\varepsilon})) - I_N([u]_{N\varepsilon})) + \varepsilon_N + (I_{N*} - J_*), \quad N = 1, 2, \ldots$$

Отсюда и из (3)–(5) следует, что последовательность $\{P_N([u]_{N\varepsilon})\}$ является минимизирующей для задачи (1), а из оценок (6)–(8) вытекает оценка (10). Теорема 1 доказана. □

Нетрудно видеть, что проведенное в § 1 исследование поведения разностных аппроксимаций (1.4)–(1.6) задачи (1.1)–(1.3) укладывается в схему теоремы 1. В самом деле, отображения Q_N и P_N в этой задаче были определены формулами (1.20), (1.21), в леммах 1.3 и 1.4 были установлены неравенства (4), (5) и оценки (6), (7), а оценки (1.31), (1.32) являются следствиями оценок (8), (10).

2. Для иллюстрации теоремы 1 кратко остановимся здесь еще на задаче (1.1), (1.2) при условии

$$u = u(t) \in U = \left\{u(t) \in L_2^r[t_0, T]: \int_{t_0}^{T} |u(t)|^2 dt \leqslant R^2\right\}, \tag{13}$$

где $R > 0$ — заданное число. В качестве аппроксимирующей последовательности для задачи (1.1), (1.2), (13) возьмем разностные аппроксимации (1.4), (1.5) при условии

$$[u]_N \in U_N = \left\{[u]_N = (u_0, u_1, \ldots, u_{N-1}) \in L_{2N}^r: \sum_{i=0}^{N-1} \Delta t_i |u_i|^2 \leqslant R^2\right\}. \tag{14}$$

В качестве отображений Q_N и P_N возьмем отображения, определяемые формулами (1.20), (1.21). Справедлива

Теорема 2. *Пусть матрицы $A(t), B(t), f(t)$ кусочно непрерывны на отрезке $[t_0, T]$, разбиения $\{t_i, i = 0, \ldots, N\}$ отрезка $[t_0, T]$ таковы, что $d_N = \max\limits_{0 \leqslant i \leqslant N-1} \Delta t_i \leqslant (T - t_0)M/N$, $N = 1, 2, \ldots$, $M = \text{const} > 0$. Пусть J_* — нижняя грань функции (1.1) при условиях (1.2), (13), I_{N*} — нижняя грань функции (1.4) при условиях (1.5), (14), а последовательность $\{[u]_{N\varepsilon}\}$ определена из условий (9). Тогда справедливы оценки*

$$-C_6\delta_N \leqslant I_{N*} - J_* \leqslant C_5\delta_N, \quad N = 1, 2, \ldots,$$
$$0 \leqslant J(P_N([u]_{N\varepsilon})) - J_* \leqslant (C_5 + C_6)\delta_N + \varepsilon_N, \quad N = 1, 2, \ldots,$$

где постоянные C_5, C_6 взяты из оценок (1.29), (1.30) при $W = U$ и $W_N = U_N$ соответственно, величина δ_N определена формулой (1.26).

Доказательство. Из соотношений (1.22), (1.23) следует, что отображения (1.20), (1.21) таковы, что $Q_N(u) \in U_N$ при всех $u \in U$ и $P_N([u]_N) \in U$ при всех $[u]_N \in U_n$, $N = 1, 2, \ldots$ Далее, оценки (1.8), (1.10), (1.16) и леммы 1.2–1.4 остаются справедливыми и для множеств (13), (14). Поэтому из лемм 1.3 и 1.4 имеем неравенства (4), (5) и оценки (6), (7). Отсюда и из теоремы 1 следуют все утверждения теоремы 2. □

3. Следует оговориться, что в более сложных задачах минимизации построение аппроксимирующих задач, удовлетворяющих условиям теоремы 1, может оказаться непростым делом. Так, например, даже в простейших задачах оптимального управления вида (1.1)–(1.3) или (1.1), (1.2), (1.3) при наличии фазовых ограничений не всегда бывает ясно, как строить разностные аппроксимации, обеспечивающие непустоту множества U_N и удовлетворяющие условиям теоремы 1. Для определения возникающих здесь трудностей иногда полезно работать с расширениями U^ε и сужениями $U^{-\varepsilon}$ множества, на котором имеется минимум. Приведем два критерия аппроксимации по функции, использующие упомянутые расширения и сужения множества U.

Теорема 3. *Для того чтобы последовательность задач (2) аппроксимировала задачу (1) по функции, необходимо и достаточно, чтобы при некотором $\varepsilon_0 > 0$ существовали семейства непустых множеств $U^\varepsilon \subseteq X$, $U^{-\varepsilon} \subseteq X$, $0 < \varepsilon < \varepsilon_0$, и отображения Q_N: $X \to X_N$, P_N: $X_N \to X$ такие, что функция $J(u)$ определена на объединении множеств U^ε, $U^{-\varepsilon}$ по всем ε, $0 < \varepsilon < \varepsilon_0$, и, кроме того:*

1) для любого ε, $0 < \varepsilon < \varepsilon_0$, найдется номер $N_1 = N_1(\varepsilon)$ такой, что $Q_N(u) \in U_N$ при всех $u \in U^{-\varepsilon}$ и $N \geqslant N_1$ и при каждом фиксированном ε, $0 < \varepsilon < \varepsilon_0$, для всех $u \in U^{-\varepsilon}$ выполняется неравенство

$$\overline{\lim_{N\to\infty}}(I_N(Q_N(u)) - J(u)) \leqslant 0; \tag{15}$$

2) для любого ε, $0 < \varepsilon < \varepsilon_0$, найдется номер $N_2 = N_2(\varepsilon)$ такой, что $P_N([u]_N) \in U^\varepsilon$ для всех $[u]_N \in U_N$ и $N \geqslant N_2$ при любом выборе $[u]_N \in U_N$, $N \geqslant 1$, выполняется неравенство

$$\overline{\lim_{N\to\infty}}(J(P_N([u]_N)) - I_N([u]_N)) \leqslant 0; \tag{16}$$

3) справедливы неравенства

$$\overline{\lim_{\varepsilon\to 0}} J_*(\varepsilon) \geqslant J_*, \tag{17}$$
$$\underline{\lim_{\varepsilon\to 0}} J_*(-\varepsilon) \leqslant J_*, \tag{18}$$

где $J_*(\varepsilon) = \inf\limits_{U^\varepsilon} J(u)$, $J_*(-\varepsilon) = \inf\limits_{U^{-\varepsilon}} J(u)$.

Доказательство. Необходимость. Пусть последовательность задач (2) аппроксимирует задачу (1) по функции, т. е. справедливо

равенство (3). Положим $\varepsilon_0 = 1$, $U^\varepsilon = U^{-\varepsilon} = U$, $0 < \varepsilon < \varepsilon_0$. Тогда $J_*(\varepsilon) = J_*(-\varepsilon) = J_*$, $0 < \varepsilon < \varepsilon_0$, и условия (17), (18) тривиально выполняются. Выберем произвольные последовательности $\{v_N\}$, $\{[v]_N\}$ такие, что

$$v_N \in U, \quad N = 1, 2, \ldots, \quad \lim_{N\to\infty} (J(v_n) - J_*) = 0, \tag{19}$$

$$[v]_N \in U_N, \quad N = 1, 2, \ldots, \quad \lim_{N\to\infty} (I_N([v]_n) - I_{N*}) = 0. \tag{20}$$

Определим отображения Q_N и P_N следующим образом:

$$Q_N(u) = [v]_N, \quad u \in X; \quad P_N([u]_N) = v_N, \quad [u]_N \in X_N, \quad N = 1, 2, \ldots$$

Ясно, что $Q_N(u) \in U_N$ при всех $u \in U^{-\varepsilon} = U$ и всех ε, $0 < \varepsilon < \varepsilon_0$, $N \geqslant 1 = N_1$, $P_N([u]_N) \in U^\varepsilon = U$ при всех $[u]_N \in U_N$ и всех ε, $0 < \varepsilon < \varepsilon_0$, $N \geqslant 1 = N_2$.

Далее, так как $J_* \leqslant J(u)$ при $u \in U = U^{-\varepsilon}$, то

$$I_N(Q_N(u)) - J(u) = (I_N([v]_N) - I_{N*}) + (I_{N*} - J_*) + (J_* - J(u)) \leqslant (I_N([v]_N) - I_{N*}) + (I_{N*} - J_*)$$

при всех $u \in U = U^{-\varepsilon}$, $0 < \varepsilon < \varepsilon_0$, $N = 1, 2, \ldots$ Переходя в этом неравенстве к верхнему пределу при $N \to \infty$ с учетом условий (3), (20), придем к неравенству (15).

Наконец, поскольку $I_{N*} \leqslant I_N([u]_N)$ при $[u]_N \in U_N$, то

$$J(P_N([u]_N)) - I_N([u]_N) = (J(v_N) - J_*) + (J_* - I_{N*}) + (I_{N*} - I_N([u]_N)) \leqslant (J(v_N) - J_*) + (J_* - I_{N*})$$

при всех $[u]_N \in U_N$, $N = 1, 2, \ldots$ Отсюда с учетом условий (3), (19) при $N \to \infty$ получим неравенство (16). Необходимость доказана.

Достаточность. Пусть выполнены условия 1)–3). Докажем, что тогда справедливо равенство (3). Зафиксируем произвольное число ε, $0 < \varepsilon < \varepsilon_0$. Так как $Q_N(u) \in U_N$ при всех $u \in U^{-\varepsilon}$ и $N \geqslant N_1$, то $I_{N*} \leqslant I_N(Q_N(u))$ и

$$I_{N*} - J_* = (I_{N*} - I_N(Q_N(u))) + (I_N(Q_N(u)) - J(u)) + (J(u) - J_*) \leqslant (I_N(Q_N(u)) - J(u)) + (J(u)) - J_*), \quad u \in U^{-\varepsilon}, \quad N \geqslant N_1.$$

Отсюда при $N \to \infty$ с учетом условия (15) получим

$$\overline{\lim_{N\to\infty}} (I_{N*} - J_*) \leqslant J(u) - J_*, \quad u \in U^{-\varepsilon}, \quad 0 < \varepsilon < \varepsilon_0.$$

Левая часть этого неравенства не зависит от $u \in U^{-\varepsilon}$, поэтому, переходя в правой части к нижней грани по $u \in U^{-\varepsilon}$, будем иметь

$$\overline{\lim_{N\to\infty}} (I_{N*} - J_*) \leqslant J(-\varepsilon) - J_*, \quad 0 < \varepsilon < \varepsilon_0.$$

При $\varepsilon \to 0$ отсюда с помощью условия (18) получим $\overline{\lim_{N\to\infty}} (I_{N*} - J_*) \leqslant 0$. С другой стороны, поскольку $P_N([u]_N) \in U^\varepsilon$ при всех $[u]_N \in U_N$, $N \geqslant N_2$, то $J_*(\varepsilon) \leqslant J(P_N([u]_N))$ и

$$J_*(\varepsilon) - I_{N*} = (J_*(\varepsilon) - J(P_N([u]_N))) + (J(P_N([u]_N)) - I_N([u]_N)) + (I_N([u]_N) - I_{N*}) \leqslant (J(P_N([u]_N)) - I_N([u]_N)) + (I_N([u]_N) - I_{N*}), \quad [u]_N \in U_N, \quad N \geqslant N_2.$$

Отсюда с учетом условия (16) при $N \to \infty$ непосредственно получим, что

$$\overline{\lim_{N\to\infty}} (J_*(\varepsilon) - I_{N*}) \leqslant \overline{\lim_{N\to\infty}} (I_N([u]_N) - I_{N*})$$

при любом выборе $[u]_N \in U_N$, $N \geqslant 1$, и любом фиксированном ε, $0 < \varepsilon < \varepsilon_0$. В частности, если, пользуясь определением I_{N*}, при каждом $N \geqslant 1$ взять $[u]_N \in U_N$ так, чтобы $I_N([u]_N) \leqslant I_{N*} + 1/N$, то из предыдущего неравенства будем иметь

$$\overline{\lim_{N\to\infty}} (J_*(\varepsilon) - I_{N*}) = J_*(\varepsilon) + \overline{\lim_{N\to\infty}} (-I_{N*}) \leqslant \overline{\lim_{N\to\infty}} 1/N = 0$$

при любом ε, $0 < \varepsilon < \varepsilon_0$. Отсюда с помощью условия (17) при $\varepsilon \to 0$ получим

$$\overline{\lim_{N\to\infty}} (J_* - I_{N*}) = J_* + \overline{\lim_{N\to\infty}} (-I_{N*}) \leqslant 0.$$

Итак, показано, что

$$0 \leqslant -\overline{\lim_{N\to\infty}} (J_* - I_{N*}) = \underline{\lim_{N\to\infty}} (I_{N*} - J_*) \leqslant \overline{\lim_{N\to\infty}} (I_{N*} - J_*) \leqslant 0,$$

$$\text{т. е.} \quad \lim_{N\to\infty} I_{N*} = \overline{\lim_{N\to\infty}} I_{N*} = J_*.$$

Отсюда следует равенство (3). Достаточность доказана. □

Теорема 4. *Для того чтобы последовательность задач* (2) *аппроксимировала задачу* (1) *по функции, необходимо и достаточно, чтобы существовала последовательность непустых множеств* $U^{\varepsilon_N} \subseteq X$, $N = 1, 2, \ldots$, *и отображения* $Q_N: X \to X_N$, $P_N: X_N \to X$ *такие, что функция* $J(u)$ *определена на объединении множеств* $\left(\bigcup_{N=1}^{\infty} U^{\varepsilon_N}\right) \cup U$ *и, кроме того:*

1) *при всех* $u \in U$ *справедливо включение* $Q_N(u) \in U_N$, $N = 1, 2, \ldots$, *и выполняется неравенство*

$$\overline{\lim_{N\to\infty}} (I_N(Q_N(u)) - J(u)) \leqslant 0;$$

2) *при всех* $[u]_N \in U_N$ *справедливо включение* $P_N([u]_N) \in U^{\varepsilon_N}$, $N = 1, 2, \ldots$, *и выполняется неравенство*

$$\overline{\lim_{N\to\infty}} (J(P_N([u]_N)) - I_N([u]_N)) \leqslant 0;$$

3) *справедливо неравенство*

$$\underline{\lim_{N\to\infty}} J_*(\varepsilon_N) \geqslant J_*, \quad \text{где} \quad J_*(\varepsilon_N) = \inf_{U^{\varepsilon_N}} J(u).$$

Если же выполнены условия 1)–3) *и, кроме того, имеются неотрицательные последовательности* $\{\beta_N\}$, $\{\gamma_N\}$, $\{\nu_N\}$, *сходящиеся к нулю и такие, что*

$$I_N(Q_N(u)) - J(u) \leqslant \beta_N, \quad u \in U, \quad N = 1, 2, \ldots, \tag{21}$$

$$J(P_N([u]_N)) - I_N([u]_N) \leqslant \gamma_N, \quad [u]_N \in U_N, \quad N = 1, 2, \ldots, \tag{22}$$

$$J_* - J_*(\varepsilon_N) \leqslant \nu_N, \quad N = 1, 2, \ldots, \tag{23}$$

то справедлива оценка

$$-\gamma_N - \nu_N \leqslant I_{N*} - J_* \leqslant \beta_N, \quad N = 1, 2, \ldots \tag{24}$$

Доказательство того, что условия 1)–3) необходимы и достаточны для выполнения равенства (3), проводится так же, как в теореме 3, нужно лишь

в этих рассуждениях заменить $U^{-\varepsilon}$ на U, U^{ε} на $\{U^{\varepsilon_N}\}$. Докажем оценку (24). Из того, что $Q_N(u) \in U_N$ при всех $u \in U$, $N = 1, 2, \ldots$, и из условия (21) следует, что $I_{N*} \leqslant I_N(Q_N(u)) \leqslant J(u) + \beta_N$, или $I_{N*} \leqslant J(u) + \beta_N$ при всех $u \in U$, $N = 1, 2, \ldots$

Переходя к нижней грани по $u \in U$, отсюда получим $I_{N*} \leqslant J_* + \beta_N$, $N = 1, 2, \ldots$. Правое неравенство (24) доказано. Далее, так как $P_N([u]_N) \in U^{\varepsilon_N}$ при всех $[u]_N \in U_N$, $N = 1, 2, \ldots$, то с учетом условий (22), (23) имеем

$$J_* - \nu_N \leqslant J_*(\varepsilon_N) \leqslant J(P_N([u]_N)) \leqslant I_N([u]_N) + \gamma_N,$$

или $J_* - \nu_N \leqslant I_N([u]_N) + \gamma_N$ при всех $[u]_N \in U_N$, $N = 1, 2, \ldots$. Отсюда следует, что $J_* - \nu_N \leqslant I_{N*} + \gamma_N$, $N = 1, 2, \ldots$. Оценка (24) доказана. □

Приложения критериев аппроксимации, приведенных в теоремах 1, 3, 4, рассмотрим ниже. Наряду с теоремами 1, 3, 4 существуют и другие варианты критериев аппроксимации по функции [114; 303; 363; 501]; некоторые такие критерии сформулированы далее в виде упражнений.

Упражнения

1. Равенство (3) имеет место тогда и только тогда, когда выполнены следующие два условия: 1) для любого числа $\delta > 0$ существует номер $N_1 = N_1(\delta)$ такой, что для всех $N \geqslant N_2$ и $u \in U$ найдется точка $[u]_{N\delta} \in U_N$, удовлетворяющая неравенству $I_N([u]_{N\delta}) \leqslant J(u) + \delta$; 2) для любого числа $\delta > 0$ существует номер $N_2 = N_2(\delta)$ такой, что для всех $N \geqslant N_2$ и $[u]_N \in U_N$ найдется точка $u_{N\delta} \in U$, для которой $J(u_{N\delta}) \leqslant I_N([u]_N) + \delta$. Доказать.

2. Для того чтобы имело место равенство (3), необходимо и достаточно, чтобы существовали отображения $Q_N\colon X \to X_N$ и $P_N\colon X_N \to X$ такие, что: 1) для некоторой последовательности $\{u_N\} \in U$, $\lim\limits_{N\to\infty}(J(u_N) - J_*) = 0$, выполняются условия: $Q_N(u_N) \in U_N$, $N = 1, 2, \ldots$, и $\overline{\lim\limits_{N\to\infty}}(I_N(Q_N(u_N)) - J(u_N)) \leqslant 0$; 2) для некоторой последовательности $[u]_N \in U_N$, $N = 1, 2, \ldots$, $\lim\limits_{N\to\infty}(I_N([u]_N) - I_{N*}) = 0$, выполняются условия: $P_N([u]_N) \in U$, $N = 1, 2, \ldots$, и $\overline{\lim\limits_{N\to\infty}}(J(P_N([u]_N)) - I_N([u]_N)) \leqslant 0$. Доказать.

3. Равенство (3) имеет место тогда и только тогда, когда при некотором $\varepsilon_0 > 0$ существуют семейства непустых множеств $U^{\varepsilon} \subseteq X$, $U^{-\varepsilon} \subseteq X$, $0 < \varepsilon < \varepsilon_0$, и отображения $Q_N\colon X \to X_N$ и $P_N\colon X_N \to X$ такие, что функция $J(u)$ определена на объединении множеств U^{ε}, $U^{-\varepsilon}$ по всем ε, $0 < \varepsilon < \varepsilon_0$, и, кроме того: 1) для некоторого семейства $\{u_\varepsilon\}$, $u_\varepsilon \in U^{-\varepsilon}$, $0 < \varepsilon < \varepsilon_0$, $\lim\limits_{\varepsilon\to 0}(J(u_\varepsilon) - J_*(-\varepsilon)) = 0$, при каждом ε найдется номер $N_1 = N_1(\varepsilon)$ такой, что $Q_N(u_\varepsilon) \in U_N$ при всех $N \geqslant N_1$ и $\overline{\lim\limits_{N\to\infty}}(I_N(Q_N(u_\varepsilon)) - J(u_\varepsilon)) \leqslant 0$; 2) для некоторой последовательности $[u]_N \in U_N$, $N = 1, 2, \ldots$, $\lim\limits_{N\to\infty}(I_N([u]_N) - I_{N*}) = 0$, и любого ε, $0 < \varepsilon < \varepsilon_0$, найдется номер $N_2 = N_2(\varepsilon)$ такой, что $P_N([u]_N) \in U^{\varepsilon}$ при всех $N \geqslant N_2$ и $\overline{\lim\limits_{N\to\infty}}(J(P_N([u]_N)) - I_N([u]_N)) \leqslant 0$; 3) выполнены условия (17), (18). Доказать.

§ 3. Разностная аппроксимация для квадратичной задачи с фазовыми ограничениями

1. Рассмотрим следующую задачу оптимального управления:

$$J(u) = |x(T, u) - y|^2 \to \inf, \tag{1}$$

$$\dot{x}(t) = A(t)x(t) + B(t)u(t) + f(t), \quad t_0 \leqslant t \leqslant T, \quad x(t_0) = x_0, \tag{2}$$

$$u = u(t) \in U = \{u(t) \in L_2^r[t_0, T]\colon u(t) \in V \text{ почти всюду на } [t_0, T],\ x(t, u) \in G(t),\ t_0 \leqslant t \leqslant T\}, \tag{3}$$

где $G(t)$ — заданные множества из E^n при каждом $t \in [t_0, T]$, $x_0 \in G(t_0)$. В качестве аппроксимирующих задач возьмем последовательность задач:

$$I_N([u]_N) = |x_N([u]_N) - y|^2 \to \inf, \tag{4}$$

$$x_{i+1} = x_i + \Delta t_i(A_i x_i + B_i u_i + f_i), \quad i = 0, \ldots, N-1, \tag{5}$$

$$[u]_N = (u_0, u_1, \ldots, u_{N-1}) \in U_N = \{[u]_N \in L_{2N}^r\colon u_i \in V,\ i = 0, \ldots, N-1;\ x_i([u]_N) \in G_i = G(t_i),\ i = 0, \ldots, N\} \tag{6}$$

(все обозначения здесь взяты из § 1).

Для того чтобы задача (1)–(3) имела смысл, естественно требовать, чтобы $U \neq \varnothing$. Возникают вопросы: можно ли тогда гарантировать, что при достаточно мелком разбиении отрезка $[t_0, T]$ множество U_N также будет непустым, и при каких условиях последовательность задач (4)–(6) будет аппроксимировать задачу (1)–(3) по функции?

Как показывает следующий пример, из того, что $U \neq \varnothing$, вообще говоря, не следует, что $U_N \neq \varnothing$.

Пример 1. Пусть требуется минимизировать функцию $J(u) = x^2(1)$ при условиях $\dot{x}(t) = x(t) - u(t)$, $0 \leqslant t \leqslant 1$, $x(0) = 1$, $u = u(t) \in U = \{u(t) \in L_2[0, 1]\colon 0 \leqslant u(t) \leqslant 1$ почти всюду на $[0, 1]$; $x(t, u) \geqslant e^t$ при всех $t \in [0, 1]\}$.

Поскольку $x(t, u) = e^t\left(1 - \int\limits_0^1 e^{\tau} u(\tau)\, d\tau\right) \leqslant e^t$, $0 \leqslant t \leqslant 1$, при всех $u(t) \in L_2[0, 1]$, $0 \leqslant u(t) \leqslant 1$, то условие $x(t, u) \geqslant e^t$, $0 \leqslant t \leqslant 1$, выполняется лишь при $u = u(t) \equiv 0$. Следовательно, множество U непусто и состоит из единственного управления $u(t) \equiv 0$.

Рассмотрим следующую разностную аппроксимацию этой задачи:

$$I_N([u]_N) = x_N^2 \to \inf, \quad x_{i+1} = x_i + \Delta t(x_i - u_i), \quad i = 0, \ldots, N-1, \quad x_0 = 1,$$
$$[u]_N = (u_0, u_1, \ldots, u_{N-1}) \in U_N = \{[u]_N \in L_{2N}\colon 0 \leqslant u_i \leqslant 1,\ i = 0, \ldots, N-1;$$
$$x_i \geqslant e^{t_i},\ i = 0, \ldots, N\}, \quad \Delta t = 1/N, \quad N = 1, 2, \ldots$$

Из $0 \leqslant u_i \leqslant 1$ следует, что $(1 + \Delta t)x_i - \Delta t \leqslant x_{i+1} = (1 + \Delta t)x_i - \Delta t u_i \leqslant (1 + \Delta t)x_i$, $i = 0, \ldots, N-1$, $x_0 = 1$. Отсюда по индукции нетрудно получить, что $1 \leqslant x_i \leqslant (1 + \Delta t)^i$, $i = 0, \ldots, N$. Так как $1 + \Delta t < e^{\Delta t}$ при всех $\Delta t > 0$, то $x_i < e^{i\Delta t} = e^{t_i}$ для всех $i = 1, \ldots, N$ при любом выборе $[u]_N = (u_0, u_1, \ldots, u_{N-1})$, $0 \leqslant u_i \leqslant 1$, $i = 0, \ldots, N-1$. Это значит, что множество U_N пусто при всех $N = 1, 2, \ldots$

2. Таким образом, важно выяснить, при каких условиях из непустоты множества (3) следует, что множество (6) также не будет пустым, а также указать способы аппроксимации задачи (1)–(3) для случаев, когда условие $U \neq \varnothing$ не гарантирует, что $U_N \neq \varnothing$. Здесь мы ограничимся изложением результатов, принадлежащих М. М. Потапову (теорема 1) и Е. Р. Авакову (теорема 2). Для упрощения выкладок задачу (1)–(3) будем рассматривать при дополнительном предположении, что множество $G(t)$ не зависит от t, т. е. $G(t) \equiv G$, $t_0 \leqslant t \leqslant T$.

Ниже нам понадобятся так называемые *ε-расширения* G^{ε} и *ε-сужения* $G^{-\varepsilon}$ множества G, определяемые так:

$$G^{\varepsilon} = \{x \in E^n\colon \rho(x, G) = \inf_{z \in G} |x - z| \leqslant \varepsilon\}, \quad \varepsilon \geqslant 0,$$
$$G^{-\varepsilon} = \{x \in G\colon \inf_{z \in \Gamma pG} |x - z| \geqslant \varepsilon\}, \quad \varepsilon \geqslant 0, \tag{7}$$

где $\mathrm{Гр}G$ — совокупность граничных точек множества G. Кроме того, мы будем широко пользоваться оценками и соотношениями (1.8)–(1.26), считая, что встречающиеся в них константы $C_0, C_1, \ldots$ отвечают множествам управлений

$$W=\{u(t)\in L_2^r[t_0,T]:\ u(t)\in V \quad \text{почти всюду на} \quad [t_0,T]\},$$

$$W_N=\{[u]_N=(u_0,u_1,\ldots,u_{N-1})\in L_{2N}^r:\ u_i\in V,\quad i=0,\ldots,N-1\}. \tag{8}$$

Теорема 1. *Пусть матрицы $A(t), B(t), f(t)$ кусочно непрерывны на отрезке $t_0\leqslant t\leqslant T$, V — выпуклое замкнутое ограниченное множество из E^r, G — выпуклое замкнутое множество из E^n с непустой внутренностью, причем существуют число $\varepsilon_0>0$ и управление $\overline{u}=\overline{u}(t)\in U$ такие, что*

$$x(t,\overline{u})\in G^{-\varepsilon_0},\quad t_0\leqslant t\leqslant T. \tag{9}$$

Пусть, кроме того, разбиения $\{t_i, i=0,\ldots,N\}$ отрезка $[t_0,T]$ удовлетворяют условию

$$d_N=\max_{0\leqslant i\leqslant N-1}\Delta t_i\leqslant (T-t_0)M_0/N,\quad M_0=\mathrm{const}>0,\quad N=1,2,\ldots$$

Тогда множество U_N, определяемое условиями (6), при всех достаточно больших N непусто и $\lim\limits_{N\to\infty} I_{N*}=J_*$, где J_* — нижняя грань функции (1) при условиях (2), (3), I_{N*} — нижняя грань функции (4) при условиях (5), (6).

Доказательство. Положим $X=L_2^r[t_0,T]$, $X_N=L_{2N}^r$, $N=1,2,\ldots$,

$$U^{\varepsilon}=\{u=u(t)\in W:\ x(t,u)\in G^{\varepsilon},\quad t_0\leqslant t\leqslant T\},$$
$$U^{-\varepsilon}=\{u=u(t)\in W:\ x(t,u)\in G^{-\varepsilon},\quad t_0\leqslant t\leqslant T\},$$

где $0<\varepsilon<\varepsilon_0$, число ε_0 взято из условия теоремы, множество W — из (8). Определим отображения $Q_N: X\to X_N$ и $P_N: X_N\to X$ формулами (1.20), (1.21):

$$Q_N(u)=(u_0,u_1,\ldots,u_{N-1}):\ u_i=\frac{1}{\Delta t_i}\int\limits_{t_i}^{t_{i+1}}u(\tau)d\tau,\quad i=0,\ldots,N-1,$$
$$P_N([u]_N)=u_i\quad \text{при}\quad t_i<t\leqslant t_{i+1},\quad i=0,\ldots,N-1. \tag{10}$$

Проверим, что для введенных множеств U^{ε}, $U^{-\varepsilon}$ и отображений Q_N, P_N выполнены все условия теоремы 2.3. Зафиксируем произвольное число ε, $0<\varepsilon<\varepsilon_0$. Возьмем какое-либо управление $u\in U^{-\varepsilon}$. Согласно оценке (1.24)

$$\max_{0\leqslant i\leqslant N}|x(t_i,u)-x_i(Q_N(u))|\leqslant\delta_N,\quad N=1,2,\ldots \tag{11}$$

Так как $\delta_N\to 0$ при $N\to\infty$, то найдется номер $N_1=N_1(\varepsilon)$ такой, что $\delta_N\leqslant\varepsilon$ при всех $N\geqslant N_1$. По определению (7) множества $G^{-\varepsilon}$ шар $|x-z|\leqslant\varepsilon$ принадлежит множеству G при всех $z\in G^{-\varepsilon}$. Тогда для любого $u\in U^{-\varepsilon}$ шар $|x-x(t_i,u)|\leqslant\varepsilon$ принадлежит множеству G при всех $i=0,\ldots,N$ и $N\geqslant N_1$. Отсюда и из оценки (11) следует, что $x_i(Q_N(u))\in G$ при всех $i=0,\ldots,N$, $u\in U^{-\varepsilon}$, $N\geqslant N_1$. Кроме того, $Q_N(u)$ принадлежит множеству W_N из (8) при всех $u\in W$ согласно лемме 1.1. Тем самым показано, что $Q_N(u)\in U_N$ при всех $u\in U^{-\varepsilon}$ и $N\geqslant N_1$. Кроме того, из леммы 1.3 имеем

$$\lim_{N\to\infty}(I_N(Q_N(u))-J(u))=0,\quad u\in U^{-\varepsilon}.$$

Таким образом, условие 1) теоремы 2.3 выполнено.

Далее, согласно ранее приведенной оценке (1.25), получаем

$$\max_{0\leqslant i\leqslant N}|x(t_i,P_N([u]_N))-x_i([u]_N)|\leqslant\delta_N,\quad N=1,2,\ldots,$$

а из оценки (1.11) имеем

$$|x(t,P_N([u]_N))-x(t_i,P_N([u]_N))|\leqslant C_1 d_N$$

при всех t, $t_i\leqslant t\leqslant t_{i+1}$, $i=0,\ldots,N-1$, $[u]_N\in U_N$, $N=1,2,\ldots$ Тогда

$$|x(t,P_N([u]_N))-x_i([u]_N)|\leqslant\delta_N+C_1 d_N \tag{12}$$

для всех t, $t_i\leqslant t\leqslant t_{i+1}$, $i=0,\ldots,N-1$, $[u]_N\in U_N$, $N=1,2,\ldots$ Однако $\delta_N\to 0$, $d_N\to 0$ при $N\to\infty$, поэтому найдется номер $N_2=N_2(\varepsilon)$ такой, что $\delta_N+C_1 d_N\leqslant\varepsilon$ при всех $N\geqslant N_2$. Отсюда, из оценки (12) и определения (6) множества U_N следует, что $x(t,P_N([u]_N))\in G^{\varepsilon}$ при всех t, $t_0\leqslant t\leqslant T$, т. е. $P_N([u]_N)\in U^{\varepsilon}$ при всех $N\geqslant N_2$. Кроме того, из леммы 1.4 имеем

$$\lim_{N\to\infty}(J(P_N([u]_N))-I_N([u]_N))=0.$$

Таким образом, условие 2) теоремы 2.3 также выполнено.

Наконец, проверим, выполняется ли условие 3) теоремы 2.3. Сначала установим, что $\lim\limits_{\varepsilon\to 0}J_*(\varepsilon)=J_*$, где $J_*(\varepsilon)=\inf\limits_{U^{\varepsilon}}J(u)$. Заметим, что если $0<\varepsilon_1<\varepsilon_2<\varepsilon_0$, то $U\subseteq U^{\varepsilon_1}\subseteq U^{\varepsilon_2}$ и, следовательно, $J_*(\varepsilon_2)\leqslant J_*(\varepsilon_1)\leqslant J_*$. Отсюда следует, что существует предел $\lim\limits_{\varepsilon\to 0}J_*(\varepsilon)$, причем

$$\lim_{\varepsilon\to 0}J_*(\varepsilon)\leqslant J_*. \tag{13}$$

При каждом ε, $0<\varepsilon<\varepsilon_0$, выберем $u_{\varepsilon}\in U^{\varepsilon}$ так, чтобы $\lim\limits_{\varepsilon\to 0}(J_*(\varepsilon)-J(u_{\varepsilon}))=0$. Так как множество W из (8) слабо компактно в $L_2^r[t_0,T]$ и $\{u_{\varepsilon}\}\in W$, $0<\varepsilon<\varepsilon_0$, то существует последовательность $\{\varepsilon_k\}$, сходящаяся к нулю, $0<\varepsilon_k<\varepsilon_0$, и такая, что последовательность управлений $\{u_k=u_{\varepsilon_k}\}$ будет слабо сходиться к некоторому управлению $v=v(t)\in W$. Согласно (1.12) последовательность $\{x(t,u_k)\}$ сходится к $x(t,v)$ равномерно на отрезке $[t_0,T]$, т. е. $\sup\limits_{t_0\leqslant t\leqslant T}|x(t,u_k)-x(t,v)|=\xi_k\to 0$ при $k\to\infty$. Поскольку $x(t,u_k)\in G^{\varepsilon_k}$, то

$$x(t,v)\in G^{\varepsilon_k+\xi_k},\quad t_0\leqslant t\leqslant T,\quad k=1,2,\ldots \tag{14}$$

Однако $\varepsilon_k+\xi_k\to 0$ при $k\to\infty$, траектория $x(t,v)$ от номера k не зависит, множество G замкнуто. Отсюда и из (14) следует, что $x(t,v)\in G$ при всех t, $t_0\leqslant t\leqslant T$. Это значит, что $v\in U$ и $J(v)\geqslant J_*$. Поскольку $\{u_k\}$ сходится к v слабо в $L_2^r[t_0,T]$, а функция $J(u)$ слабо непрерывна, то $\lim\limits_{k\to\infty}J(u_k)=J(v)\geqslant J_*$. С учетом определения последовательности $\{u_k=u_{\varepsilon_k}\}$ тогда получим

$$\lim_{\varepsilon\to 0}J_*(\varepsilon)=\lim_{k\to\infty}J_*(\varepsilon_k)=J(v)\geqslant J_*.$$

Отсюда и из (13) следует, что $\lim\limits_{\varepsilon\to 0}J_*(\varepsilon)=J_*$.

Теперь покажем, что $\lim\limits_{\varepsilon\to 0}J_*(-\varepsilon)=J_*$, где $J_*(-\varepsilon)=\inf\limits_{U^{-\varepsilon}}J(u)$. Заметим, что если $0<\varepsilon_1<\varepsilon_2<\varepsilon_0$, то $U^{-\varepsilon_2}\subseteq U^{-\varepsilon_1}\subseteq U$ и, следовательно, $J_*\leqslant J_*(\varepsilon_1)\leqslant J_*(\varepsilon_2)$. Отсюда следует, что существует предел $\lim\limits_{\varepsilon\to 0}J_*(-\varepsilon)$, причем

$$\lim_{\varepsilon\to 0}J_*(-\varepsilon)\geqslant J_*. \tag{15}$$

Возьмем любую последовательность управлений $\{u_k\} \in U$, $\lim\limits_{k\to\infty} J(u_k) = J_*$, и числовую последовательность $\{\alpha_k\}$, $0 < \alpha_k < 1$, $\lim\limits_{k\to\infty} \alpha_k = 0$. По условию теоремы существует управление $\overline{u} = \overline{u}(t) \in U$, для которого справедливо включение (9). Составим последовательность $v_k = \alpha_k \overline{u} + (1-\alpha_k)u_k$, $k = 1, 2, \ldots$, и покажем, что $v_k \in U^{-\varepsilon_0\alpha_k}$, $k = 1, 2, \ldots$ Так как $\overline{u}, u_k \in W$, то $v_k \in W$, $k = 1, 2, \ldots$, в силу выпуклости W. Остается показать, что

$$x(t, v_k) \in G^{-\varepsilon_0\alpha_k}, \quad t_0 \leqslant t \leqslant T, \quad k = 1, 2, \ldots \tag{16}$$

Заметим теперь, что

$$x(t, v_k) = \alpha_k x(t, \overline{u}) + (1-\alpha_k)x(t, u_k), \quad t_0 \leqslant t \leqslant T, \quad k = 1, 2, \ldots \tag{17}$$

Из условия (9) следует, что при каждом $t \in [t_0, T]$ и $k = 1, 2, \ldots$ шар $S(t) = \{x\colon |x - x(t, \overline{u})| \leqslant \varepsilon_0\}$ принадлежит множеству G. Покажем, что тогда шар $S_k(t) = \{x\colon |x - x(t, v_k)| \leqslant \varepsilon_0\alpha_k\}$ также принадлежит G при всех $t \in [t_0, T]$ и $k = 1, 2, \ldots$ Возьмем произвольную точку $x \in S_k(t)$ и положим $z = x(t, \overline{u}) + (x - x(t, v_k))/\alpha_k$. Так как $|z - x(t, \overline{u})| = |x - x(t, v_k)|/\alpha_k \leqslant \varepsilon_0$, то $z \in S(t) \in G$. Из определения точки z и равенства (17) тогда имеем $x = x(t, v_k) + \alpha_k(z - x(t, \overline{u})) = \alpha_k z + (1-\alpha_k)x(t, u_k)$, где $z \in G$, $x(t, u_k) \in G$, $0 < \alpha_k < 1$. В силу выпуклости множества G отсюда следует, что $x \in G$. Таким образом, $S_k(t) \in G$ при всех $t \in [t_0, T]$, $k = 1, 2, \ldots$ Отсюда и из определения шара $S_k(t)$ вытекает включение (16). Тем самым показано, что $v_k \in U^{-\varepsilon_0\alpha_k}$, $k = 1, 2, \ldots$

Зафиксируем некоторый номер k и возьмем число ε таким, чтобы $0 < \varepsilon < \varepsilon_0\alpha_k$. Тогда $v_k \in U^{-\alpha_k\varepsilon_0} \subset U^{-\varepsilon}$ и $J_*(-\varepsilon) \leqslant J(v_k)$ при всех ε, $0 < \varepsilon < \varepsilon_0\alpha_k$. Отсюда при $\varepsilon \to 0$ получим

$$\lim_{\varepsilon\to 0} J_*(-\varepsilon) \leqslant J(v_k) \quad \text{при всех} \quad k = 1, 2, \ldots \tag{18}$$

Так как $\lim\limits_{k\to\infty} \alpha_k = 0$, $\sup\limits_W \|u\| \leqslant R < \infty$, то с учетом оценки (1.14) имеем $|J(u_k) - J(v_k)| \leqslant C_3\|u_k - v_k\| = C_3\alpha_k\|u_k - \overline{u}\| \leqslant 2RC_3\alpha_k \to 0$ при $k \to \infty$. Следовательно, $\lim\limits_{k\to\infty} J(v_k) = \lim\limits_{k\to\infty} J(u_k) = J_*$. Отсюда и из неравенства (18) вытекает, что $\lim\limits_{\varepsilon\to 0} J_*(-\varepsilon) \leqslant J_*$. Сравнивая полученное неравенство с (15), заключаем, что $\lim\limits_{\varepsilon\to 0} J_*(-\varepsilon) = J_*$. Таким образом, условие 3) теоремы 2.3 также выполнено. Из теоремы 2.3 следует, что $\lim\limits_{N\to\infty} J_{N*} = J_*$. Теорема 1 доказана. □

3. При доказательстве того, что множество (6) непусто, в теореме 1 было существенно использовано условие (9). Однако это условие не всегда легко проверяемо и не всегда оно выполняется. Поэтому при аппроксимации задачи (1)–(3) вместо задачи (4)–(6) можно попытаться рассмотреть задачу минимизации функции (4) при условии (5) на несколько расширенном по сравнению с (6) множестве

$$U_N = \{[u]_N \in W\colon x_i([u]_N) \in G^{\xi_N},\ i = 0, \ldots, N\}. \tag{19}$$

Оказывается, если исходное множество (3) непусто, то при достаточно большом ξ_N и множество (19) не будет пустым, и, кроме того, если $\xi_N \to 0$ при $N \to \infty$ согласованно с $d_N = \max\limits_{0\leqslant i\leqslant N-1} \Delta t_i$, то последовательность задач (4), (5), (19) будет аппроксимировать задачу (1)–(3) по функции. А именно, справедлива

Теорема 2. *Пусть матрицы $A(t), B(f), f(t)$ кусочно непрерывны на отрезке $t_0 \leqslant t \leqslant T$, V — выпуклое замкнутое ограниченное множество из E^r, G — замкнутое множество из E^n, множество (3) непусто. Пусть, кроме того, разбиения отрезка $[t_0, T]$ удовлетворяют условию*

$$d_N = \max_{0\leqslant i\leqslant N-1} \Delta t_i \leqslant (T - t_0)M_0/N,\ M_0 = \text{const} > 0,\ N = 1, 2, \ldots,$$

и $\lim\limits_{N\to\infty} \xi_N = 0$, $\xi_N \geqslant \delta_N$, $N = 1, 2, \ldots$, где величина δ_N определена формулой (1.26). Тогда множество (19) непусто при всех $N = 1, 2, \ldots$ и $\lim\limits_{N\to\infty} I_{N*} = J_*$, где J_* — нижняя грань функции (1) при условиях (2), (3), I_{N*} — нижняя грань функции (4) при условиях (5), (19). Если, кроме того, множество G выпукло, имеет непустую внутренность и выполнено условие (9), а также $\xi_N \leqslant M_1\delta_N$, $M_1 = \text{const} \geqslant 1$, $N = 1, 2, \ldots$, то справедлива оценка

$$2RC_3\frac{(M_1+1)\delta_N + C_1 d_N^{1/2}}{\varepsilon_0 + 2\delta_N + C_1 d_N^{1/2}} - C_6\delta_N \leqslant I_{N*} - J_* \leqslant C_5\delta_N, \quad N = 1, 2, \ldots, \tag{20}$$

где постоянные C_1, C_3, C_5, C_6 взяты из (1.10), (1.14), (1.29), (1.30) соответственно, $\sup\limits_W \|u\| \leqslant R$.

Доказательство. Положим $X = L_2^r[t_0, T]$, $X_N = L_{2N}^r$, $N = 1, 2, \ldots$, $U^{\varepsilon_N} = \{u(t) \in W\colon x(t, u) \in G^{\varepsilon_N}\}$, где $\varepsilon_N = \xi_N + \delta_N + C_1 d_N^{1/2}$ (кстати, здесь можно было бы воспользоваться оценкой (1.11) и всюду ниже принять $\varepsilon_N = \xi_N + \delta_N + C_1 d_N$). Определим отображения Q_N, P_N формулами (10). Проверим выполнение условий теоремы 2.4.

Возьмем произвольное управление $u \in U$. Согласно оценке (1.24)

$$\max_{0\leqslant i\leqslant N} |x(t_i, u) - x_i(Q_N(u))| \leqslant \delta_N, \quad N = 1, 2, \ldots$$

Отсюда, учитывая, что $x(t, u) \in G$, $t_0 \leqslant t \leqslant T$, и $\delta_N \leqslant \xi_N$, $N = 1, 2, \ldots$, имеем включение $x_i(Q_N(u)) \in G^{\xi_N}$ при всех $i = 0, \ldots, N$, $N = 1, 2, \ldots$ Это значит, что $Q_N(u) \in U_N$ при всех $u \in U$ и $N = 1, 2, \ldots$ Кроме того, из леммы 1.3 имеем оценку

$$I_N(Q_N(u)) - J(u) \leqslant C_5\delta_N = \beta_N, \quad u \in U, \quad N = 1, 2, \ldots$$

Таким образом, выполнение условия 1) теоремы 2.4 проверено и, кроме того, установлена оценка (2.21).

Далее, согласно оценке (1.25), имеем

$$\max_{0\leqslant i\leqslant N} |x(t_i, P_N([u]_N)) - x_i([u]_N)| \leqslant \delta_N, \quad [u]_N \in U_N, \quad N = 1, 2, \ldots$$

Из оценки (1.10) следует

$$|x(t, P_N([u]_N)) - x(t_i, P_N([u]_N))| \leqslant C_1 d_N^{1/2}$$

при всех t, $t_i \leqslant t \leqslant t_{i+1}$, $i = 0, \ldots, N-1$, $[u]_N \in U_N$, $N = 1, 2, \ldots$ Тогда

$$|x(t, P_N([u]_N)) - x_i([u]_N)| \leqslant \delta_N + C_1 d_N^{1/2}$$

для всех t, $t_i \leqslant t \leqslant t_{i+1}$, $i=0,\ldots,N-1$, $[u]_N \in U_N$, $N=1,2,\ldots$ Отсюда, учитывая, что по определению множества U_N точки $x_i([u]_N)$ принадлежат G^{ξ_N}, $i=0,\ldots,N$, имеем

$$x(t, P_N([u]_N)) \in G^{\varepsilon_N} \quad \text{при} \quad \varepsilon_N = \xi_N + \delta_N + C_1 d_N^{1/2}$$

для всех $[u]_N \in U_N$, $N=1,2,\ldots$ Тем самым показано, что $P_N([u]_N) \in U^{\varepsilon_N}$ при всех $[u]_N \in U_N$, $N=1,2,\ldots$ Кроме того, из леммы 1.4 следует

$$J(P_N([u]_N)) - I_N([u]_N) \leqslant C_6 \delta_N = \gamma_N, \quad [u]_N \in U_N, \quad N=1,2,\ldots$$

Таким образом, выполнение условия 2) теоремы 2.4 также проверено и, кроме того, установлена оценка (2.22).

Далее, рассуждая так же, как при доказательстве равенства $\lim\limits_{\varepsilon \to 0} J_*(\varepsilon) = J_*$ в теореме 1, можно показать, что $\lim\limits_{N \to \infty} J_*(\varepsilon_N) = J_*$. Это значит, что условие 3) теоремы 2.4 также выполнено. Из теоремы 2.4 следует, что $\lim\limits_{N \to \infty} I_{N*} = J_*$.

Дополнительно предполагая, что G — выпуклое множество и существует управление $\overline{u} = \overline{u}(t) \in U$ такое, что $x(t, \overline{u}) \in G^{-\varepsilon_0}$, $t_0 \leqslant t \leqslant T$, $\varepsilon_0 > 0$, докажем оценку (20). Так как множество U^{ε_N} слабо компактно в $L_2^r[t_0, T]$, а функция (1) слабо непрерывна на U^{ε_N}, то в силу теоремы 8.2.4 при каждом фиксированном $N=1,2,\ldots$ существует управление $u_{N*} \in U^{\varepsilon_N}$ такое, что $J_*(\varepsilon_N) = J(u_{N*})$. Составим управление $v_N = \alpha_N \overline{u} + (1-\alpha_N) u_{N*}$, где $\alpha_N = \varepsilon_N/(\varepsilon_0+\varepsilon_N)$, $N=1,2,\ldots$ Так как W — выпуклое множество, $0<\alpha_N<1$, а $\overline{u}$, $u_{N*} \in W$, то $v_N \in W$ при всех $N=1,2,\ldots$ Далее, в силу определения (7) сужения множества $G^{-\varepsilon_0} = (G^{\varepsilon_N})^{-(\varepsilon_0+\varepsilon_N)}$ — это $(\varepsilon_0+\varepsilon_N)$-сужение множества G^{ε_N}. Тогда $x(t, \overline{u}) \in G^{-\varepsilon_0} = (G^{\varepsilon_N})^{-(\varepsilon_0+\varepsilon_N)}$, $x(t, u_{N*}) \in G^{\varepsilon_N}$. Повторив рассуждения, проведенные при доказательстве включения (16), получаем, что

$$x(t, v_N) \in (G^{\varepsilon_N})^{-(\varepsilon_0+\varepsilon_N)\alpha_N} = G^{\varepsilon_N - \alpha_N(\varepsilon_0+\varepsilon_N)} = G^0 = G, \quad t_0 \leqslant t \leqslant T, \quad N=1,2,\ldots,$$

поскольку $\varepsilon_N - \alpha_N(\varepsilon_0+\varepsilon_N) = 0$ в силу определения α_N. Следовательно, $v_N \in U$ и $J(v_N) \geqslant J_*$, $N=1,2,\ldots$ С учетом оценки (1.14) тогда имеем $0 \leqslant J_* - J_*(\varepsilon_N) = J_* - J(u_{N*}) = (J_* - J(v_N)) + (J(v_N) - J(u_{N*})) \leqslant J(v_N) - J(u_{N*}) \leqslant C_3 \|v_N - v_{N*}\|$. Из определения v_N следует $\|v_N - u_{N*}\| = \alpha_N \|\overline{u} - u_{N*}\| \leqslant \alpha_N \cdot 2R$, так что

$$0 \leqslant J_* - J_*(\varepsilon_N) \leqslant 2RC_3\alpha_N = 2RC_3\varepsilon_N/(\varepsilon_0+\varepsilon_N) =$$
$$= 2RC_3(\xi_N + \delta_N + C_1 d_N^{1/2})/(\varepsilon_0 + \xi_N + \delta_N + C_1 d_N^{1/2}), \quad N=1,2,\ldots$$

Поскольку $\delta_N \leqslant \xi_N \leqslant M_1 \delta_N$, то $0 \leqslant J_* - J_*(\varepsilon_N) \leqslant \nu_N = 2RC_3((M_1+1)\delta_N + C_1 d_N^{1/2})/(\varepsilon_0 + 2\delta_N + C_1 d_N^{1/2})$, $N=1,2,\ldots$ Оценка (2.23) получена. Теперь ясно, что оценка (20) является следствием оценки (2.24). Теорема 2 доказана. □

Предлагаем читателю самостоятельно убедиться в справедливости теорем 1, 2, когда вместо множеств (3), (6), (19) берутся соответственно множества

$$U = \{u(t) \in L_2^r[t_0, T]: \|u\|_{L_2} \leqslant R, \quad x(t,u) \in G, \quad t_0 \leqslant t \leqslant T\},$$
$$U_N = \{[u]_N \in L_{2N}^r[t_0, T]: \|[u]_N\|_{L_{2N}} \leqslant R, \quad x_i([u]_N) \in G, \quad i=0,\ldots,N\},$$
$$U_N = \{[u]_N \in L_{2N}^r[t_0, T]: \|[u]_N\|_{L_{2N}} \leqslant R, \quad x_i([u]_N) \in G^{\xi_N}, \quad i=0,\ldots,N\}.$$

§ 4. Регуляризация аппроксимаций экстремальных задач

Пусть X — заданное множество, на котором введена некоторая топология τ (см. определения в 8.2, п. 7), функция $J(u)$ определена на X. Рассмотрим задачу

$$J(u) \to \inf, \quad u \in U, \tag{1}$$

где U — заданное подмножество из X. Предположим, что

$$J_* = \inf_U J(u) > -\infty, \quad U_* = \{u \in U: J(u) = J_*\} \neq \varnothing.$$

Пусть последовательность задач

$$I_N([u]_N) \to \inf, \quad u \in U_N \subseteq X_N, \quad N=1,2,\ldots, \tag{2}$$

где X_N, U_N — заданные множества, $I_N([u]_N)$ — заданные функции на U_N, $N=1,2,\ldots$, аппроксимирует задачу (1) по функции.

Возникает вопрос: нельзя ли провести регуляризацию последовательности задач (2) или, иными словами, нельзя ли с помощью задач (2) построить минимизирующую последовательность для задачи (1), сходящуюся к множеству U_* в топологии τ?

1. В следующей теореме, принадлежащей Е. Р. Авакову, указываются достаточные условия для регуляризации последовательности задач (2).

Теорема 1. *Пусть выполнены следующие условия:*

1) $J_* > -\infty$, $U_* \neq \varnothing$, *функция* $\Omega(u)$ *определена и неотрицательна на* X;

2) *последовательность* $\{[v]_N\}$ *определена из условий*

$$[v]_n \in U_N, \quad T_N([v]_N) \leqslant T_{N*} + \mu_N, \quad N=1,2,\ldots, \tag{3}$$

где

$$T_N([u]_N) = I_N([u]_N) + \alpha_N \Omega_N([u]_N), \quad [u]_N \in U_N,$$

— функция Тихонова задачи (2), $\Omega_N([u]_N)$ *— заданная на* U_N *функция,* $T_{N*} = \inf\limits_{U_N} T_N([u]_N)$; $\{\mu_N\}$, $\{\alpha_N\}$ *— положительные последовательности, сходящиеся к нулю (если нижняя грань* T_{N*} *достигается в какой-либо точке из* U_N, *то в* (3) *не исключается возможность* $\mu_N = 0$);

3) *существуют последовательность множеств* $U^{\varepsilon_N} \subseteq X$ *и отображения* P_N: $X_N \to X$ *такие, что*

$$P_N([v]_N) \in U^{\varepsilon_N}, \quad N=1,2,\ldots, \tag{4}$$
$$J(P_N([v]_N)) - I_N([v]_N) \leqslant \gamma_N, \quad N=1,2,\ldots, \tag{5}$$
$$\Omega(P_N([v]_N)) \leqslant \Omega_N([v]_N) + \xi_N, \quad N=1,2,\ldots, \tag{6}$$

где $\{\gamma_N\}$, $\{\xi_N\}$ *— последовательности, сходящиеся к нулю;*

4) *существуют отображения* Q_N: $X \to X_N$ *такие, что для всех* $u_* \in U_*$ *выполняются соотношения*

$$Q_N(u_*) \in U_N, \quad N=1,2,\ldots, \tag{7}$$
$$I_N(Q_N(u_*)) - J(u_*) \leqslant \beta_N, \quad N=1,2,\ldots, \tag{8}$$
$$\Omega_N(Q_N(u_*)) \leqslant \Omega(u_*) + \eta_N, \quad N=1,2,\ldots, \tag{9}$$

где $\{\beta_N\}$, $\{\eta_N\}$ — *последовательности, сходящиеся к нулю;*

5) *имеются оценки*

$$J_* - J_*(\varepsilon_N) \leqslant \nu_N, \quad J_*(\varepsilon_N) = \inf_{U^{\varepsilon_N}} J(u), \quad N = 1, 2, \ldots, \tag{10}$$

где $\{\nu_N\}$ — *последовательность, сходящаяся к нулю.*

Тогда

$$\lim_{N\to\infty} J(P_N([v]_N)) = J_*. \tag{11}$$

Пусть наряду с условиями 1)–5) *выполняются еще следующие условия:*

6) *функции* $J(u), \Omega(u)$ τ-*секвенциально полунепрерывны снизу на* X; *функция* $\Omega(u)$ *является* τ-*стабилизатором, т. е. множество* $\Omega_C = \{u \in X, \Omega(u) \leqslant C\}$ τ-*секвенциально компактно при любом* $C \geqslant 0$; *всякая точка* $v \in X$, *являющаяся* τ-*пределом какой-либо последовательности* $\{u_N\}$, $u_N \in U^{\varepsilon_N}$, $N = 1, 2, \ldots$, *принадлежит* U; *(см. определения 8.2.9, 8.2.10, 8.2.13);*

7) *справедливо равенство*

$$\lim_{N\to\infty} (\beta_N + \gamma_N + \mu_N + \nu_N)/\alpha_N = 0. \tag{12}$$

Тогда последовательность $\{P_N([v]_N)\}$ τ-*сходится к множеству* $U_{**} = \{w \in U_*: \Omega(w) = \inf_{U_*} \Omega(u) = \Omega_*\}$ *и* $\lim_{N\to\infty} \Omega(P_N([v]_N)) = \Omega_*$.

Доказательство. Возьмем произвольную точку $u_* \in U_*$. Из условий (3)–(10) следует цепочка неравенств

$$\begin{aligned}
J_* &\leqslant J_*(\varepsilon_N) + \nu_N \leqslant J(P_N([v]_N)) + \nu_N \leqslant J(P_N([v]_N)) + \alpha_N \Omega(P_N([v]_N)) + \nu_N \leqslant \\
&\leqslant I_N([v]_N) + \gamma_N + \alpha_N \Omega_N([v]_N) + \alpha_N \xi_N + \nu_N = T_N([v]_N) + \gamma_N + \nu_N + \alpha_N \xi_N \leqslant \\
&\leqslant T_{N*} + \mu_N + \gamma_N + \nu_N + \alpha_N \xi_N \leqslant T_N(Q_N(u_*)) + \mu_N + \gamma_N + \nu_N + \alpha_N \xi_N = \\
&= I_N(Q_N(u_*)) + \alpha_N \Omega_N(Q_N(u_*)) + \mu_N + \gamma_N + \nu_N + \alpha_N \xi_N \leqslant \\
&\leqslant J(u_*) + \beta_N + \alpha_N \Omega(u_*) + \alpha_N \eta_N + \mu_N + \gamma_N + \nu_N + \alpha_N \xi_N \leqslant \\
&\leqslant J_*(\varepsilon_N) + 2\nu_N + \beta_N + \alpha_N \Omega(u_*) + \mu_N + \gamma_N + \nu_N + \alpha_N(\xi_N + \eta_N) \leqslant \\
&\leqslant J(P_N([v]_N)) + \alpha_N \Omega(u_*) + 2\nu_N + \beta_N + \mu_N + \gamma_N + \alpha_N(\eta_N + \xi_N), \quad N = 1, 2, \ldots
\end{aligned}$$

Отсюда с учетом произвола в выборе $u_* \in U_*$ имеем

$$J_* - \nu_N \leqslant J(P_N([v]_N)) \leqslant J_* + \alpha_N \Omega_* + \beta_N + \gamma_N + \mu_N + \alpha_N(\xi_N + \eta_N), \quad N = 1, 2, \ldots \tag{13}$$

$$\Omega(P_N([v]_N)) \leqslant \Omega_* + (\beta_N + \gamma_N + \mu_N + \nu_N)/\alpha_N + \xi_N + \eta_N, \quad N = 1, 2, \ldots \tag{14}$$

Из (13) следует равенство (11), причем неравенства (13) представляют собой оценку скорости сходимости в (11).

Далее предположим, что выполнены все условия 1)–7) теоремы. Для краткости обозначим $v_N = P_N([v]_N)$, $N = 1, 2, \ldots$. Из (14) имеем $\Omega(v_N) \leqslant \Omega_* + \mathrm{const} = C < \infty$. Кроме того, согласно (4) имеем $v_N \in U^{\varepsilon_N} \subseteq X$, $N = 1, 2, \ldots$. Это значит, что $\{v_N\} \in \Omega_C$. По условию, множество Ω_C τ-секвенциально компактно. Поэтому последовательность $\{v_N\}$ имеет хотя бы одну τ-сходящуюся подпоследовательность. Возьмем произвольную точку v_*, являющуюся

τ-пределом какой-либо подпоследовательности $\{v_{N_k}\}$. По условию теоремы $v_* \in U$. Тогда с учетом τ-секвенциальной полунепрерывности снизу функции $J(u)$ и равенства (11) имеем $J_* \leqslant J(v_*) \leqslant \varliminf_{k\to\infty} J(v_{N_k}) = \lim_{N\to\infty} J(v_N) = J_*$, т. е. $J(v_*) = J_*$, или $v_* \in U_*$. Далее, из τ-секвенциальной полунепрерывности снизу функции $\Omega(u)$ и неравенства (14) вытекает, что $\Omega_* \leqslant \Omega(v_*) \leqslant \varliminf_{k\to\infty} \Omega(v_{N_k}) \leqslant \varlimsup_{k\to\infty} \Omega(v_{N_k}) \leqslant \Omega_*$, т. е. $\lim_{k\to\infty} \Omega(v_{N_k}) = \Omega_* = \Omega(v_*)$, или $v_* \in U_{**}$. Тем самым показано, что любая точка v_*, являющаяся τ-пределом какой-либо подпоследовательности $\{v_{N_k}\}$, принадлежит множеству U_{**} и $\{\Omega(v_{N_k})\} \to \Omega_*$. Отсюда получаем, что последовательность $\{v_N\}$ τ-сходится к множеству U_{**} и $\{\Omega(v_N)\} \to \Omega_*$ при $N \to \infty$. Теорема доказана. □

2. Для иллюстрации теоремы 1 рассмотрим задачу

$$J(u) = |x(T, u) - y|^2 \to \inf, \tag{15}$$

$$\dot{x}(t) = A(t)x(t) + B(t)u(t) + f(t), \quad t_0 \leqslant t \leqslant T, \quad x(t_0) = x_0, \tag{16}$$

$$u = u(t) \in U = \{u(t) \in L_2^r[t_0, T]: u(t) \in V$$

$$\text{почти всюду на} \quad [t_0, T];\ x(t, u) \in G,\ t_0 \leqslant t \leqslant T\}; \tag{17}$$

все обозначения здесь взяты из § § 1, 3. Так же, как в § 3, для аппроксимации этой задачи рассмотрим последовательность разностных задач

$$I_N([u]_N) = |x_N([u]_N) - y|^2 \to \inf, \tag{18}$$

$$x_{i+1} = x_i + \Delta t_i(A_i x_i + B_i u_i + f_i), \quad i = 0, \ldots, N-1, \tag{19}$$

$$[u]_N \in U_N = \{[u]_N = (u_0, u_1, \ldots, u_{N-1}): u_i \in V,$$

$$i = 0, \ldots, N-1; \quad x_i([u]_N) \in G^{\xi_N}, \quad i = 0, \ldots, N\}, \quad N = 1, 2, \ldots \tag{20}$$

(обозначения см. в § § 1, 3). Если в задаче (15)–(17) фазовых ограничений нет, т. е. $G \equiv E^n$, то в (20) нет необходимости рассматривать расширения G^{ξ_N} и можно принять

$$\xi_N = 0, \quad G^{\xi_N} = E^n, \quad N = 1, 2, \ldots$$

Положим $X = L_2^r[t_0, T]$, $X_N = L_{2N}^r$, $N = 1, 2, \ldots$, в качестве стабилизатора возьмем $\Omega(u) = \int_{t_0}^{T} |u(t)|^2 dt = \|u\|_{L_2}^2$, $\Omega_N([u]_N) = \sum_{i=0}^{N-1} \Delta t_i |u_i|^2 = \|[u]_N\|_{L_{2N}}^2$ и определим отображения $Q_N: X \to X_N$, $P_N: X_N \to X$ формулами

$$Q_N(u) = (u_0, u_1, \ldots, u_{N-1}): u_i = \frac{1}{\Delta t_i} \int_{t_i}^{t_{i+1}} u(t)dt, \quad i = 0, \ldots, N-1,$$

$$P_N([u]_N) = u_i, \quad t_i < t \leqslant t_{i+1}, \quad i = 0, \ldots, N-1.$$

Из соотношений (1.22), (1.23) следует, что

$$\Omega_N(Q_N(u)) \leqslant \Omega(u), \quad \Omega(P_N([u]_N)) = \Omega_N([u]_N) \tag{21}$$

при всех $u \in L_2^r[t_0, T]$, $[u]_N \in L_{2N}^r$.

Заметим также, что ниже мы будем пользоваться соотношениями (1.8)–(1.26), считая, что встречающиеся в них константы $C_0, C_1, \ldots$ соответствуют множествам

$$W = \{u(t) \in L_2^r[t_0, T]: u(t) \in V \quad \text{почти всюду на} \quad [t_0, T]\},$$

$$W_N = \{[u]_N \in L_{2N}^r: u_i \in V,\ i = 0, \ldots, N-1\}, \quad N = 1, 2, \ldots \tag{22}$$

Теорема 2. *Пусть матрицы $A(t), B(t), f(t)$ кусочно непрерывны на отрезке $[t_0, T]$; V — выпуклое замкнутое ограниченное множество из E^r; G — выпуклое замкнутое множество из E^n, $\operatorname{int} G \neq \varnothing$; существуют число $\varepsilon_0 > 0$ и управление $\overline{u} = \overline{u}(t) \in U$ такие, что $x(t, \overline{u}) \in G^{-\varepsilon_0}$, $t_0 \leqslant t \leqslant T$. Пусть разбиение $\{t_i, i = 0, \ldots, N\}$ отрезка $[t_0, T]$ удовлетворяет условию*

$$d_N = \max_{0 \leqslant i \leqslant N-1} \Delta t_i \leqslant (T - t_0) M_0 / N, \quad N = 1, 2, \ldots, M_0 = \text{const} > 0,$$

и, кроме того,

$$\delta_N \leqslant \xi_N \leqslant M_1 \delta_N, \quad M_1 = \text{const} \geqslant 1, \quad N = 1, 2, \ldots,$$

где величина δ_N взята из (1.26), а последовательности $\{\alpha_N\}$, $\{\mu_N\}$ таковы, что

$$\lim_{N \to \infty} \frac{\delta_N + \mu_N + d_N}{\alpha_N} = 0. \tag{23}$$

Наконец, пусть последовательность $\{[v]_N\}$ определена условиями

$$[v]_N \in U_N, \quad T_N([v]_N) \leqslant T_{N*} + \mu_N, \quad N = 1, 2, \ldots, \tag{24}$$

где

$$T_N([v]_N) = I_N([v]_N) + \alpha_N \sum_{i=0}^{N-1} \Delta t_i |u_i|^2, \quad T_{N*} = \inf_{U_N} T_N([u]_N).$$

Тогда последовательность $\{P_N([v]_N)\}$ сходится к Ω-нормальному решению $u_* = u_*(t)$ задачи (15)–(17) по норме $L_2^r[t_0, T]$.

Доказательство. Из условий теоремы следует, что множество (17) непусто, выпукло, замкнуто и ограничено, а функция (15) выпукла и непрерывна в метрике $L_2^r[t_0, T]$ (она даже слабо непрерывна в $L_2^r[t_0, T]$). В силу теоремы 8.2.8 множество U_* непусто, выпукло, замкнуто и ограничено. Тогда сильно выпуклая функция $\Omega(u) = \|u\|_{L_2}^2$ на множестве U_* согласно теореме 8.2.10 достигает своей нижней грани $\inf_{U_*} \Omega(u) = \Omega_*$ в единственной точке $u_* = u_*(t) \in U_*$. Таким образом, Ω-нормальное решение задачи (15)–(17) существует и единственно.

Определим множество

$$U^{\varepsilon_N} = \{u(t) \in W: x(t, u) \in G^{\varepsilon_N},\ t_0 \leqslant t \leqslant T\},$$

где $\varepsilon_N = \xi_N + \delta_N + C_1 d_N$ (если $G \equiv E^n$, то $\varepsilon_N = 0$), $N = 1, 2, \ldots$. Заметим, что поскольку множество W из (22) ограничено в метрике $L_\infty^r[t_0, T]$, то можем пользоваться оценкой (1.11). Рассуждая так же, как при доказательстве теоремы 3.2, можно показать, что в рассматриваемом случае для введенных множеств U^{ε_N}, отображений Q_N, P_N справедливы включения (4), (7) и неравенства (5), (8), (10) при $\beta_N = C_5 \delta_N$, $\gamma_N = C_6 \delta_N$, $\nu_N = 2RC_3 \varepsilon_0^{-1}((M_1 + 1)\delta_N + C_1 d_N)$, $N = 1, 2, \ldots$. Тогда из условия (23) вытекает условие (12). Из (21) следуют неравенства (6), (9) при $\xi_N = \eta_N = 0$, $N = 1, 2, \ldots$.

Далее, функция $J(u)$ слабо непрерывна, $\Omega(u)$ неотрицательна и слабо полунепрерывна снизу на $L_2^r[t_0, T]$, множество $\Omega_C = \{u \in L_2^r[t_0, T]: \Omega(u) \leqslant C\}$ слабо компактно при любом $C \geqslant 0$.

Наконец, всякая точка $v = v(t) \in L_2^r[t_0, T]$, являющаяся слабым пределом какой-либо последовательности $\{u_N\}$, $u_N \in U^{\varepsilon_N}$, $N = 1, 2, \ldots$, принадлежит множеству U. В самом деле, условие $u_N \in U^{\varepsilon_N}$ означает, что $u_N \in W$, $x(t, u_N) \in G^{\varepsilon_N}$, $t_0 \leqslant t \leqslant T$, $N = 1, 2, \ldots$ Из слабой замкнутости множества W следует, что $v \in W$. Далее, согласно (1.12),

$$\sup_{t_0 \leqslant t \leqslant T} |x(t, u_N) - x(t, v)| = \chi_N \to 0$$

при $N \to \infty$, поэтому $x(t, v) \in G^{\varepsilon_N + \chi_N}$, $t_0 \leqslant t \leqslant T$, $N = 1, 2, \ldots$ Так как $\varepsilon_N + \chi_N \to 0$, множество G замкнуто, а $x(t, v)$ не зависит от N, то последнее включение возможно только при $x(t, v) \in G$, $t_0 \leqslant t \leqslant T$. Следовательно, $v \in U$.

Таким образом, в рассматриваемом случае все условия теоремы 1 выполнены, если в качестве топологии τ в ней взять слабую топологию пространства $L_2^r[t_0, T]$. Из теоремы 1 тогда следует, что последовательность $\{P_N([v]_N)\}$ сходится к u_* слабо в $L_2^r[t_0, T]$ и, кроме того,

$$\|P_N([v]_N)\|_{L_2}^2 = \Omega(P_N([v]_N)) \to \|u_*\|_{L_2}^2 = \Omega(u_*)$$

при $N \to \infty$. Но

$$\|P_N([v]_N) - u_*\|_{L_2}^2 = 2(\|P_N([v]_N)\|_{L_2}^2 + \|u_*\|_{L_2}^2) - \|P_N([v]_N) + u_*\|_{L_2}^2, \quad N = 1, 2, \ldots$$

Отсюда, пользуясь слабой полунепрерывностью снизу нормы в $L_2^r[t_0, T]$, получаем, что последовательность $\{P_N([v]_N)\}$ сходится к u_* по норме $L_2^r[t_0, T]$. Теорема 2 доказана. □

3. Прокомментируем условие (23). Оно означает, что шаг $d_N = \max_{0 \leqslant i \leqslant N-1} \Delta t_i$ разбиений отрезка $[t_0, T]$, а также точность решения задачи минимизации функции Тихонова $T_N([u]_N)$ на множестве U_N в смысле неравенства (24) должны быть согласованы с параметром регуляризации α_N. В частном случае, когда матрицы $A(t), B(t), f(t)$ на интервалах непрерывности удовлетворяют условию Липшица (например, когда эти матрицы не зависят от t), пользуясь оценкой (1.11) из (1.26), имеем $\delta_N \leqslant C_8 d_N$, $C_8 = \text{const} > 0$. Тогда условие (23) запишется в виде

$$\lim_{N \to \infty} \frac{d_N + \mu_N}{\alpha_N} = 0. \tag{25}$$

Заметим, что теорема 2 остается верной и в случае, когда в задачах (15)–(17) и (18)–(20) множества U и U_N отличны от множеств (17), (20) и имеют, например, вид

$$U = \{u(t) \in L_2^r[t_0, T]: \|u\|_{L_2} \leqslant R, \quad x(t, u) \in G,\ t_0 \leqslant t \leqslant T\}, \tag{26}$$

$$U_N = \{[u]_N \in L_{2N}^r: \|[u]_N\|_{L_{2N}} \leqslant R, \quad x_i([u]_N) \in G^{\xi_N},\ i = 0, \ldots, N\},\ N = 1, 2, \ldots$$

Однако для множеств (26) условие (23) нужно заменить на

$$\lim_{N \to \infty} \frac{\delta_N + \mu_N + d_N^{1/2}}{\alpha_N} = 0. \tag{27}$$

Подчеркнем, что если при регуляризации разностных аппроксимаций задач оптимального управления параметр регуляризации α_N и шаг d_N разбиений отрезка $[t_0, T]$ стремятся к нулю несогласованно, то получающаяся при этом последовательность $\{P_N([v]_N)\}$ может не сходиться к множеству Ω-нормальных решений в нужной метрике. Это видно из следующих примеров [359].

Пример 1. Рассмотрим задачу

$$J(u)=x(1,u)-\int_0^1(x(t,u)+tu(t))dt\to\inf,$$

$$\dot x(t)=u(t),\quad 0\leqslant t\leqslant 1,\quad x(0)=0,$$

$$u=u(t)\in U=\{u(t)\in L_p[0,1],\ |u(t)|\leqslant 1 \text{ почти всюду на } [0,1]\},$$

где p — фиксированное число, $1<p<\infty$.

Поскольку $\int_0^1 tu(t)dt=\int_0^1 t\dot x(t)dt=x(1)-\int_0^1 x(t)dt$, то $J(u)\equiv 0$ при всех $u\in U$. Следовательно, $J_*=0$, $U_*=U$. Возьмем стабилизатор $\Omega(u)=\int_0^1|u(t)|^p dt$. Тогда $U_{**}=\{v\in U_*:\ \Omega(v)=\inf_{U_*}\Omega(u)\}=\{0\}$, т. е. Ω-нормальное решение единственно и равно $u_*=0$.

Рассмотрим следующую разностную аппроксимацию этой задачи:

$$I_N([u]_N)=x_N-\sum_{i=0}^{N-1}(x_i+t_iu_i)\Delta t\to\inf,$$

$$x_{i+1}=x_i+\Delta tu_i,\quad i=0,\dots,N-1,\quad x_0=0,$$

$$[u]_N\in U_N=\{[u]_N=(u_0,u_1,\dots,u_{N-1}):\ |u_i|\leqslant 1,\quad i=0,\dots,N-1\},$$

$$\Delta t=1/N=d_N,\quad N=1,2,\dots$$

Введем разностную функцию Тихонова

$$T_N([u]_N)=x_N-\sum_{i=0}^{N-1}(x_i+t_iu_i)\Delta t+\alpha_N\sum_{i=0}^{N-1}\Delta t|u_i|^p,\ [u]_N\in U_N.$$

Так как $x_N-\sum_{i=0}^{N-1}\Delta tx_i=x_N-\sum_{i=0}^{N-1}(t_{i+1}-t_i)x_i=x_N-\sum_{i=0}^{N-1}t_ix_{i-1}+\sum_{i=0}^{N-1}t_ix_i=x_N+\sum_{i=1}^{N-1}t_i(x_i-x_{i-1})-x_{N-1}=\sum_{i=1}^{N}t_i(x_i-x_{i-1})=\sum_{i=0}^{N-1}t_{i+1}\Delta tu_i$, то $T_N([u]_N)=\sum_{i=0}^{N-1}\Delta t\times(\Delta tu_i+\alpha_N|u_i|^p)$. Следовательно,

$$\min_{U_N}T_N([u]_N)=T_{N*}=\sum_{i=0}^{N-1}\Delta t\min_{|u|\leqslant 1}(\Delta tu+\alpha_N|u|^p)=\min_{|u|\leqslant 1}(\Delta tu+\alpha_N|u|^p).$$

Отсюда, исследуя поведение функции $\varphi(u)=\Delta tu+\alpha_N|u|^p$ на отрезке $[-1,1]$, нетрудно установить, что

$$T_{N*}=\begin{cases}-\frac{\Delta t}{q}\left(\frac{\Delta t}{\alpha_N p}\right)^{1/(p-1)} & \text{при}\quad \frac{\Delta t}{p\alpha_N}\leqslant 1,\ q=(p-1)^{-1}p,\\ -\Delta t+\alpha_N & \text{при}\quad \frac{\Delta t}{p\alpha_N}>1,\end{cases}$$

причем нижняя грань достигается при $[v]_N=(v_0,v_1,\dots,v_{N-1})$, где

$$v_i=\begin{cases}-\left(\frac{\Delta t}{p\alpha_N}\right)^{1/(p-1)} & \text{при}\quad \frac{\Delta t}{p\alpha_N}\leqslant 1,\\ -1 & \text{при}\quad \frac{\Delta t}{p\alpha_N}>1,\end{cases}\qquad i=0,\dots,N-1.$$

Таким образом, если $\Delta t=d_N=1/N\to 0$, $\alpha_N\to 0$, но $\frac{\Delta t}{p\alpha_N}>1$, то $P_N([v]_N)=-1$ при всех t, $0\leqslant t\leqslant 1$, $N=1,2,\dots$, и последовательность $\{P_N([v]_N)\}$ не может сходиться к нормальному решению $u_*=0$ в метрике $L_p[0,1]$ ни при каких p, $1<p<\infty$. Остается рассмотреть случай, когда $\Delta t=d_N\to 0$, $\alpha_N\to 0$, $\frac{\Delta t}{p\alpha_N}\leqslant 1$. Тогда $P_N([v]_N)=-\left(\frac{\Delta t}{p\alpha_N}\right)^{-1/(1-p)}$ при всех t, $0\leqslant t\leqslant 1$, и

$$\|P_N([v]_N)-0\|_{L_p}^p=\int_0^1\left(\frac{\Delta t}{p\alpha_N}\right)^{p/(1-p)}dt=\left(\frac{d_N}{p\alpha_N}\right)^{1/q},\quad N=1,2,\dots$$

Отсюда ясно, что для сходимости последовательности $\{P_N([v]_N)\}$ к $u_*=0$ в метрике $L_p[0,1]$, $1<p<\infty$, необходимо и достаточно, чтобы $\frac{d_N}{\alpha_N}=\frac{1}{N\alpha_N}\to 0$ при $N\to\infty$.

Пример 2. Рассмотрим задачу

$$J(u)=4x(1,u)-3\int_0^1 u(t)dt\to\inf,\quad \dot x(t)=u(t),\quad 0\leqslant t\leqslant 1,\quad x(0)=0,$$

$$u=u(t)\in U=\{u(t)\in L_p[0,1]:\ \|u\|_{L_p}^p=\int_0^1|u(t)|^pdt\leqslant 1,$$

$$u(t)\geqslant 0 \text{ почти всюду на } [0,1]\},\quad 1<p<\infty.$$

Поскольку $x(t,u)=\int_0^1 u(\tau)d\tau$, то

$$J(u)=4\int_0^1 u(t)dt-3\int_0^1 u(t)dt=\int_0^1 u(t)dt\geqslant 0$$

при $u\in U$. Отсюда ясно, что $J_*=0$, $U_*=\{0\}$. В качестве стабилизатора возьмем $\Omega(u)=\|u\|_{L_p}^p$.

Рассмотрим следующую разностную аппроксимацию этой задачи:

$$I_N([u]_N)=4\left(\frac{x_N+x_{N-1}}{2}\right)-3\sum_{i=0}^{N-1}\Delta tu_i\to\inf,$$

$$x_{i+1}=x_i+\Delta tu_i,\quad i=0,\dots,N,\quad x_0=0,$$

$$[u]_N\in U_N=\{[u]_N:\ \sum_{i=0}^{N-1}\Delta t|u_i|^p\leqslant 1,\quad u_i\geqslant 0,\quad i=0,\dots,N-1\},$$

$$\Delta t=1/N,\quad N=1,2,\dots$$

Введем разностную функцию Тихонова

$$T_N([u]_N)=2(x_N+x_{N-1})-3\sum_{i=0}^{N-1}\Delta tu_i+\alpha_N\sum_{i=0}^{N-1}\Delta t|u_i|^p.$$

Так как $x_{j+1}=\sum_{i=0}^{j}\Delta t u_i$, $j=0,\ldots,N-1$, то получаем

$$T_N([u]_N)=\sum_{i=0}^{N-2}\Delta t(u_i+\alpha_N|u_i|^p)+\Delta t(-u_{N-1}+\alpha_N|u_{N-1}|^p)\geqslant$$
$$\geqslant\Delta t(-u_{N-1}+\alpha_N|u_{N-1}|^p),\quad [u]_N\in U_N.$$

Отсюда, исследуя поведение функции $\varphi(u)=-u+\alpha_N|u|^p$ при $|u|^p\Delta t\leqslant 1$, $u\geqslant 0$, или при $0\leqslant u\leqslant(\Delta t)^{-1/p}$, нетрудно получить, что

$$T_{N*}=\inf_{U_N}T_N([u]_N)=\begin{cases}-\frac{1}{q}\frac{1}{(p\alpha_N)^{1/(p-1)}} & \text{при}\quad \Delta t\leqslant(p\alpha_N)^q,\\ -(\Delta t)^{1/q}+\alpha_N & \text{при}\quad \Delta t>(p\alpha_N)^q,\end{cases}\qquad q=p(p-1)^{-1},$$

причем нижняя грань T_{N*} достигается в точке

$$[v]_N=\begin{cases}(0,0,\ldots,0,(\Delta t)^{-1/p}) & \text{при}\quad \Delta t>(p\alpha_N)^q,\\ (0,0,\ldots,0,(p\alpha_N)^{-1/(p-1)}) & \text{при}\quad \Delta t\leqslant(p\alpha_N)^q.\end{cases}$$

Таким образом, если $\Delta t=1/N$, $\alpha_N\to\infty$, $N\to\infty$, но $\Delta t>(p\alpha_N)^q$, то $P_N([v]_N)\equiv 0$ при $0\leqslant t\leqslant 1-\Delta t$, $P_N([v]_N)=(\Delta t)^{-1/p}$ при $1-\Delta t<t\leqslant 1$ и $\|P_N([v]_N)\|_{L_p}^p=1$, $N=1,2,\ldots$, так что $\{P_N([v]_N)\}$ не будет сходиться к $u_*=0$ в метрике $L_p[0,1]$ ни при каких p, $1<p<\infty$.

Остается рассмотреть случай $\Delta t\leqslant(p\alpha_N)^q$. Тогда $P_N([v]_N)=0$ при $0\leqslant t\leqslant 1-\Delta t$, $P_N([v]_N)=(p\alpha_N)^{1/(1-p)}$ при $1-\Delta t<t\leqslant 1$ и $\|P_N([v]_N)\|_{L_p}^p=\frac{\Delta t}{(p\alpha_N)^q}=\frac{d_N}{(p\alpha_N)^q}$, $N=1,2,\ldots$ Следовательно, для сходимости $\{P_N([v]_N)\}$ к $u_*=0$ в метрике $L_p[0,1]$, $1<p<\infty$, необходимо и достаточно, чтобы $\frac{d_N}{\alpha_N^q}=\frac{1}{N\alpha_N^q}\to 0$ при $N\to\infty$.

Любопытно, что в примере 2 условие сходимости $\{P_N([v]_N)\}$ в метрике $L_p[0,1]$ зависит от p, а в примере 1 такое условие $\frac{1}{N\alpha_N}\to 0$ не зависит от p, но зато $U\subset L_p[0,1]$ при всех p, $1\leqslant p\leqslant+\infty$.

Приведенные примеры показывают, что при регуляризации разностных аппроксимаций задач оптимального управления условия типа (23), (25) или (27) являются существенными.

4. При построении разностной задачи (18)–(20), аппроксимирующей задачу (15)–(17), были использованы расширения множества G, согласованные с шагом d_N разбиения $\{t_i\}$. Следуя Е. Р. Авакову, покажем, что при выполнении условий теоремы 2 регуляризованную аппроксимирующую задачу можно построить и без расширений множества G. А именно, в качестве множества U_N вместо (20) введем множество

$$U_N=\{[u]_N=(u_0,u_1,\ldots,u_{N-1}):\ u_i\in V,$$
$$i=0,\ldots,N-1,\quad x_i([u]_N)\in G,\quad i=0,\ldots,N\}\qquad(28)$$

и возьмем число N_0 столь большим, чтобы

$$\delta_N\leqslant\varepsilon_0\quad\text{при всех}\quad N\geqslant N_0,\qquad(29)$$

где величина δ_N взята из (1.26), а ε_0 — из условий теоремы 2.

Прежде всего покажем, что тогда множество (28) непусто при всех $N\geqslant N_0$. Будем пользоваться теми же отображениями Q_N, P_N, а также множествами (22), которые были введены в п. 2. Возьмем Ω-нормальное решение u_* задачи (15)–(17) и положим

$$w_N=\zeta_N\overline{u}+(1-\zeta_N)u_*,\quad \zeta_N=\delta_N/\varepsilon_0,\quad N\geqslant N_0.$$

В силу (29) имеем $0<\zeta_N\leqslant 1$. Отсюда и из выпуклости множества W следует, что $w_N\in W$. Так как $x(t,\overline{u})\in G^{-\varepsilon_0}$, $x(t,u_*)\in G$, $t_0\leqslant t\leqslant T$, то, рассуждая так же, как при доказательстве включения (3.16), получим, что $x(t,w_N)\in G^{-\zeta_N\varepsilon_0}$, $t_0\leqslant t\leqslant T$. Отсюда и из оценки (1.24) следует, что $x_i(Q_N(w_N))\in G^{-\zeta_N\varepsilon_0+\delta_N}=G$, $i=0,\ldots,N$. Это значит, что

$$Q_N(w_N)\in U_N\neq\varnothing,\quad N\geqslant N_0.\qquad(30)$$

Далее, $\|w_N-u_*\|=\zeta_N\|\overline{u}-u_*\|\leqslant 2R\zeta_N=2R\delta_N/\varepsilon_0$, где $R=\sup_W\|u\|<\infty$. Поэтому

$$|\Omega(w_N)-\Omega(u_*)|=|\|w_N\|^2-\|u_*\|^2|\leqslant 4R^2\delta_N/\varepsilon_0\qquad(31)$$

и, кроме того, из оценки (1.14) имеем

$$|J(w_N)-J(u_*)|\leqslant 2RC_3\delta_N/\varepsilon_0,\quad N\geqslant N_0.\qquad(32)$$

Наконец, согласно оценке (1.29)

$$I_N(Q_N(w_N))-J(w_N)\leqslant C_5\delta_N=\beta_N,\quad N\geqslant N_0.\qquad(33)$$

Теперь возьмем последовательность $\{[v]_N\}$, определенную условиями (24) для множества (28). Тогда $x_i([v]_N)\in G$, $i=\overline{0,N}$. Из оценок (1.11), (1.25) получим $x(t,P_N([v]_N))\in G^{\varepsilon_N}$, $t_0\leqslant t\leqslant T$, где $\varepsilon_N=\delta_N+C_1d_N$. Положим $v_N=\chi_N\overline{u}+(1-\chi_N)P_N([v]_N)$, $\chi_N=\varepsilon_N/(\varepsilon_0+\varepsilon_N)$, $N\geqslant N_0$. Так как $\overline{u}\in W$, $P_N([v]_N)\in W$, $0<\chi_N<1$, то $v_N\in W$. По условию теоремы 2 имеем $x(t,\overline{u})\in G^{-\varepsilon_0}$, $t_0\leqslant t\leqslant T$. Однако $G^{-\varepsilon_0}=G^{-\varepsilon_N-(\varepsilon_0+\varepsilon_N)}=(G^{\varepsilon_N})^{-(\varepsilon_0+\varepsilon_N)}$ — это $(\varepsilon_0+\varepsilon_N)$-сужение множества $G^{-\varepsilon_N}$. Поэтому, рассуждая так же, как при доказательстве включения (3.16), получим $x(t,v_N)\in(G^{\varepsilon_N})^{-\chi_N(\varepsilon_0+\varepsilon_N)}=G$, $t_0\leqslant t\leqslant T$. Это значит, что $v_N\in U$, $N\geqslant N_0$.

Далее, $\|v_N-P_N([v]_N)\|=\chi_N\|\overline{u}-P_N([v]_N)\|\leqslant 2R\chi_N$. Из оценки (1.14) тогда имеем

$$|J(v_N)-J(P_N([v]_N))|\leqslant 2RC_3\chi_N\leqslant 2RC_3(\delta_N+C_1d_N)/\varepsilon_0=\nu_N.$$

Следовательно,

$$J_*\leqslant J(v_N)\leqslant J(P_N([v]_N))+\nu_N,\quad N\geqslant N_0.\qquad(34)$$

Наконец, согласно оценке (1.30)

$$J(P_N([v]_N))-I_N([v]_N)\leqslant C_6\delta_N=\gamma_N,\quad N\geqslant N_0.\qquad(35)$$

Из соотношений (21), (24), (30)–(35) следует цепочка неравенств

$$J_*\leqslant J(P_N([v]_N))+\nu_N\leqslant J(P_N([v]_N))+\alpha_N\Omega(P_N([v]_N))+\nu_N\leqslant$$
$$\leqslant I_N([v]_N)+\gamma_N+\alpha_N\Omega_N([v]_N)+\nu_N\leqslant T_{N*}+\mu_N+\gamma_N+\nu_N\leqslant$$

$$\leqslant T_N(Q_N(w_N))+\mu_N+\gamma_N+\nu_N=I_N(Q_N(w_N))+\alpha_N\Omega_N(Q_N(w_N))+\mu_N+\gamma_N+\nu_N\leqslant$$
$$\leqslant J(w_N)+\alpha_N\Omega_N(w_N)+\beta_N+\mu_N+\gamma_N+\nu_N\leqslant$$
$$\leqslant J(u_*)+\alpha_N\Omega(u_*)+2R(C_3+2R\alpha_N)\delta_N/\varepsilon_0+$$
$$+\beta_N+\mu_N+\gamma_N+\nu_N\leqslant J(P_N([v]_N))+\alpha_N\Omega(u_*)+$$
$$+2R(C_3+2R\alpha_N)\delta_N/\varepsilon_0+\beta_N+\mu_N+\gamma_N+2\nu_N,\quad N\geqslant N_0.$$

Отсюда имеем

$$J_*-\nu_N\leqslant J(P_N([v]_N))\leqslant J_*+\alpha_N\Omega(u_*)+$$
$$+2R(C_3+2R\alpha_N)\delta_N/\varepsilon_0+\beta_N+\mu_N+\gamma_N,\quad N\geqslant N_0,$$
$$\Omega(P_N([v]_N))\leqslant\Omega_*+(2R(C_3+2R\alpha_N)\delta_N/\varepsilon_0+\beta_N+\mu_N+\gamma_N+\nu_N)/\alpha_N\leqslant$$
$$\leqslant\Omega_*+\text{const}\,(\delta_N+d_N+\mu_N)/\alpha_N,\quad N\geqslant N_0.$$

Следовательно, $\lim\limits_{N\to\infty}J(P_N([v]_N))=J_*$. Кроме того, повторив рассуждения, проведенные выше, получаем, что последовательность $\{P_N([v]_N)\}$ сходится к u_* слабо в $L_2^r[t_0, T]$ и $\{\Omega(P_N([v]_N))\}\to\Omega_*=\Omega(u_*)$, что равносильно сходимости $\{P_N([v]_N)\}$ к u_* в метрике $L_2^r[t_0, T]$.

§ 5. Разностная аппроксимация квадратичной задачи с переменной областью управления

1. Рассмотрим задачу

$$J(u)=|x(T, u)-y|^2\to\inf,\tag{1}$$
$$\dot{x}(t)=A(t)x(t)+B(t)u(t)+f(t),\quad t_0\leqslant t\leqslant T;\quad x(t_0)=x_0,\tag{2}$$
$$u=u(t)\in U=\{u(t)\in L_2[t_0, T]:\ u(t)\in V(t)\ \text{почти всюду на}\ [t_0, T]\},\tag{3}$$

где $V(t)$ — заданное семейство множеств, зависящих от $t\in[t_0, T]$. Для аппроксимации этой задачи введем последовательность разностных задач

$$I_N([u]_N)=|x_N([u]_N)-y|^2\to\inf,\tag{4}$$
$$x_{i+1}=x_i+\Delta t_i(A_ix_i+B_iu_i+f_i),\quad i=0, \ldots, N-1,\tag{5}$$
$$[u]_N\in U_N=\{[u]_N=(u_0, u_1, \ldots, u_{N-1}):\ u_i\in$$
$$\in V_i=V(t_i),\ i=0, \ldots, N-1\},\quad N=1, 2, \ldots\tag{6}$$

(обозначения см. в § 1).

Для исследования поведения решений задачи (4)–(6) при $N\to\infty$ ниже будут использованы теорема 2.1 и схема рассуждений из § 1. Однако из-за зависимости области управления от времени в рассматриваемой задаче не все результаты из § 1 сохраняют силу. Например, для отображения $Q_N(u)=(u_0, u_1, \ldots, u_{N-1})$, где $u_i=\frac{1}{\Delta t_i}\int\limits_{t_i}^{t_{i+1}}u(t)dt$, $i=0, \ldots, N-1$ (см. формулы (1.20)), включения $u_i\in V_i=V(t_i)$ могут нарушаться, несмотря на то, что $u=u(t)\in V(t)$ почти всюду на $[t_0, T]$ и $V(t)$ выпукло при каждом $t\in[t_0, T]$. Аналогично, для отображения $P_N([u]_N)$, определяемого формулами (1.21), включение $P_N([u]_N)\in U$ может соблюдаться не при всех $[u]_N$ из множества (6). В то же время, кажется, что если множества $V(t)$ непрерывно зависят от t, то упомянутые включения, по-видимому, будут нарушаться незначительно. Однако пока неясно, что значит, что множества $V(t)$ непрерывно зависят от t, как понимать близость между множествами, что такое расстояние между множествами. Перейдем к обсуждению этих вопросов.

2. Из различных возможных подходов к определению понятия расстояния между множествами здесь мы остановимся на понятии расстояния в смысле Хаусдорфа или, короче, хаусдорфова расстояния.

Определение 1. Пусть M — метрическое пространство с расстоянием $\rho(a, b)$ между точками $a, b\in M$, и пусть A и B — два множества из M. *Хаусдорфовым расстоянием* между множествами A и B называется величина

$$h(A, B)=\max\{\sup_{a\in A}\inf_{b\in B}\rho(a, b);\ \sup_{b\in B}\inf_{a\in A}\rho(a, b)\}.\tag{7}$$

Поясним геометрический смысл хаусдорфова расстояния, считая, что множества A и B замкнуты в метрике M. Напомним, что величина

$$\rho(a, Z)=\inf_{z\in Z}\rho(a, z)$$

называется расстоянием от точки $a\in M$ до множества $Z\subset M$. Кроме того, как и в § § 3, 4, введем ε-расширение множества Z так:

$$Z^\varepsilon=\{z\in M:\ \rho(z, Z)\leqslant\varepsilon\},\quad \varepsilon\geqslant 0.$$

Тогда величина

$$\sup_{a\in A}\inf_{b\in B}\rho(a, b)=\sup_{a\in A}\rho(a, B)=\delta(A, B)=\beta,$$

называемая *уклонением множества A от множества B*, равна минимальному числу, на которое надо расширить множество B, для того чтобы получившееся после расширения множество содержало множество A, т. е. $A\subseteq B^\varepsilon$ при всех $\varepsilon\geqslant\beta$ и $A\not\subset B^\varepsilon$ при $0\leqslant\varepsilon<\beta$. Аналогично, величина

$$\sup_{b\in B}\inf_{a\in A}\rho(a, b)=\sup_{b\in B}\rho(b, A)=\delta(B, A)=\gamma,$$

называемая *уклонением множества B от множества A*, такова, что $B\subseteq A^\varepsilon$ при всех $\varepsilon\geqslant\gamma$ и $B\not\subset A^\varepsilon$ при $0\leqslant\varepsilon<\gamma$. Таким образом, хаусдорфово расстояние $h(A, B)$ между множествами A и B равно нижней грани всех тех чисел $\varepsilon>0$ таких, что $A\subseteq B^\varepsilon$ и $B\subseteq A^\varepsilon$.

Отсюда следует, что если $h(A, B)\leqslant\varepsilon$, то справедливы следующие два включения:

$$A\subseteq B^\varepsilon,\quad B\subseteq A^\varepsilon,\quad \varepsilon>0.\tag{8}$$

Рассмотрим несколько примеров.

Пример 1. Пусть $M=E^1$, $A=\{u\in E^1:\ a\leqslant u\leqslant b\}$, $B=\{u\in E^1:\ c\leqslant u\leqslant d\}$. Пользуясь приведенной выше геометрической интерпретацией хаусдорфова расстояния, нетрудно вычислить, что

$$h(A, B)=\max\{|a-c|, |b-d|\}.$$

Пример 2. Пусть $M=R_\infty^r$ — r-мерное линейное пространство с метрикой $\rho_\infty(u, v)=\max\limits_{1\leqslant i\leqslant r}|u^i-v^i|$, соответствующей норме $\|u\|_\infty=\max\limits_{1\leqslant i\leqslant r}|u^i|$, и

$$A=\{u=(u^1, u^2, \ldots, u^r):\ a^i\leqslant u^i\leqslant b^i,\quad i=1, \ldots, r\},$$
$$B=\{u=(u^1, u^2, \ldots, u^r):\ c^i\leqslant u^i\leqslant d^i,\quad i=1, \ldots, r\}.$$

В рассматриваемой метрике A^ε — ε-расширение множества A — имеет вид $A^\varepsilon = \{u = (u^1, \ldots, u^r)\colon a^i - \varepsilon \leqslant u^i \leqslant b^i + \varepsilon,\ i = 1, \ldots, r\}$. Тогда ясно, что

$$h_\infty(A, B) = \max\{|a-c|_\infty, |b-d|_\infty\} = \max\{\max_{1\leqslant i\leqslant r} |a^i - c^i|;\ \max_{1\leqslant i\leqslant r} |b^i - d^i|\},$$

где $a = (a^1, \ldots, a^r)$, $b = (b^1, \ldots, b^r)$, $c = (c^1, \ldots, c^r)$, $d = (d^1, \ldots, d^r)$.

Если те же множества A и B рассматривать в евклидовом пространстве $M = E^r$ с метрикой $\rho(u, v) = \left(\sum_{i=1}^r |u^i - v^i|^2\right)^{1/2}$, то для соответствующего хаусдорфова расстояния имеем оценку

$$h_\infty(A, B) \leqslant h(A, B) \leqslant h_\infty(A, B)\sqrt{r},$$

или

$$\max\{|a-c|_\infty, |b-d|_\infty\} \leqslant h(A, B) \leqslant \sqrt{r}\max\{|a-c|_\infty, |b-d|_\infty\}$$

Эти оценки следуют из неравенств $|u|_\infty \leqslant |u|_{E^r} \leqslant \sqrt{r}|u|_\infty$.

Пример 3. Пусть $M = E^r$, $V(t) = \{u = (u^1, \ldots, u^r) \in E^r\colon \alpha_i(t) \leqslant u^i \leqslant \beta_i(t),\ i = 1, \ldots, r\}$, $t_0 \leqslant t \leqslant T$, где $\alpha(t) = (\alpha_1(t), \ldots, \alpha_r(t))$, $\beta(t) = (\beta_1(t), \ldots, \beta_r(t))$ — заданные функции, $\alpha_i(t) \leqslant \beta_i(t)$, $t_0 \leqslant t \leqslant T$. Пользуясь результатами примера 2, имеем

$$\frac{1}{\sqrt{r}}\max\{|\alpha(t) - \alpha(\tau)|_{E^r}; |\beta(t) - \beta(\tau)|_{E^r}\} \leqslant h(V(t); V(\tau)) \leqslant$$
$$\leqslant \sqrt{r}\max\{|\alpha(t) - \alpha(\tau)|_{E^r};\quad |\beta(t) - \beta(\tau)|_{E^r}\},\quad t_0 \leqslant t,\quad \tau \leqslant T.$$

Пример 4. Пусть $M = E^2$, $A = \{u = (x, y)\colon x^2 + y^2 \leqslant 1\}$, $B = B(t) = \{u = (x, y)\colon (x - 2t)^2 + y^2 \leqslant t^2\}$, $t \geqslant 0$. Из геометрических соображений нетрудно получить, что $\delta(A, B) = t + 1$ при всех $t \geqslant 0$;

$$\delta(B, A) = \begin{cases} 0 & \text{при } 0 \leqslant t \leqslant 1/3, \\ 3t - 1 & \text{при } t > 1/3, \end{cases} \qquad h(A, B) = \begin{cases} t+1 & \text{при } 0 \leqslant t \leqslant 1, \\ 3t - 1 & \text{при } t > 1. \end{cases}$$

Заметим, что здесь $h(B(t), B(\tau)) = 3|t - \tau|$ при всех $t, \tau \geqslant 0$.

Покажем, что хаусдорфово расстояние обладает следующими тремя замечательными свойствами [428]:

1) Если A и B — замкнутые множества из метрического пространства M, то $h(A, B) = 0$ тогда и только тогда, когда $A = B$. В самом деле, если замкнутые множества A и B не совпадают, то либо $\delta(A, B) > 0$, либо $\delta(B, A) > 0$, поэтому $h(A, B) = \max\{\delta(A, B); \delta(B, A)\} > 0$. Если же $A = B$, то, очевидно, $h(A, B) = 0$.

2) Хаусдорфово расстояние симметрично, т. е. $h(A, B) = h(B, A)$. Это свойство следует из определения (7) и симметричности расстояния $\rho(a, b)$ в исходном метрическом пространстве M.

3) Справедливо неравенство треугольника

$$h(A, B) \leqslant h(A, C) + h(C, B), \qquad A, B, C \in M. \tag{9}$$

В самом деле, из неравенства треугольника для исходного пространства M имеем $\rho(a, b) \leqslant \rho(a, c) + \rho(c, b)$ при всех $a \in A$, $b \in B$, $c \in C$. Тогда

$$\rho(a, B) = \inf_{b\in B} \rho(a, b) \leqslant \rho(a, c) + \inf_{b\in B} \rho(c, b) = \rho(a, c) + \rho(c, B) \leqslant$$
$$\leqslant \rho(a, c) + \sup_{c\in C} \rho(c, B) = \rho(a, c) + \delta(C, B) \leqslant \rho(a, c) + h(C, B)$$

для всех $a \in A$, $c \in C$. В силу произвольности $c \in C$ отсюда получаем

$$\rho(a, B) \leqslant \inf_{c\in C} \rho(a, c) + h(C, B) = \rho(a, C) + h(C, B) \leqslant$$
$$\leqslant \delta(A, C) + h(C, B) \leqslant h(A, C) + h(C, B), \quad a \in A.$$

Следовательно,

$$\delta(A, B) = \sup_{a\in A} \rho(a, B) \leqslant h(A, C) + h(C, B).$$

Поменяв в предыдущем рассуждении множества A и B ролями, будем иметь

$$\delta(B, A) \leqslant h(A, C) + h(C, B).$$

Из последних двух неравенств следует неравенство (9).

Приведенные свойства 1)–3) хаусдорфова расстояния показывают, что множество всех ограниченных замкнутых подмножеств метрического пространства в свою очередь также образует метрическое пространство с метрикой $h(A, B)$.

3. Изучим некоторые свойства множеств, зависящих от времени.

Определение 2. Пусть $V(t)$, $t_0 \leqslant t \leqslant T$, — некоторое семейство множеств из метрического пространства M. Говорят, что это семейство множеств *непрерывно по Хаусдорфу в точке* t, если для любого $\varepsilon > 0$ найдется число $\delta > 0$ такое, что $h(V(t), V(\tau)) < \varepsilon$ для всех τ, для которых $|t - \tau| < \delta$.

Лемма 1. *Пусть $V(t)$, $t_0 \leqslant t \leqslant T$, — семейство множеств из E^r, причем в каждой точке $t \in [t_0, T]$ множество $V(t)$ замкнуто, ограничено и непрерывно по Хаусдорфу. Тогда множества $V(t)$ ограничены равномерно по $t \in [t_0, T]$, т. е. найдется постоянная $R > 0$ такая, что*

$$\sup_{t_0\leqslant t\leqslant T}\ \sup_{u\in V(t)} |u| \leqslant R.$$

Доказательство. Пусть, вопреки утверждению, множества $V(t)$ не являются равномерно ограниченными на $[t_0, T]$. Это значит, что для любого натурального числа n найдутся $t_n \in [t_0, T]$ и $u_n \in V(t_n)$ такие, что $|u_n| \geqslant n$, $n = 1, 2, \ldots$ Так как отрезок $[t_0, T]$ — ограниченное замкнутое множество на числовой оси, то из последовательности $\{t_n\}$ можно выбрать подпоследовательность $\{t_{n_k}\}$, сходящуюся при $n_k \to \infty$ к некоторой точке $\tau \in [t_0, T]$. Без ограничения общности можем считать, что сама последовательность $\{t_n\}$ стремится к τ. Так как семейство $V(t)$, $t_0 \leqslant t \leqslant T$, непрерывно по Хаусдорфу в точке τ, то для любого $\varepsilon > 0$ найдется номер n_0 такой, что $h(V(t_n), V(\tau)) < \varepsilon$ при всех $n \geqslant n_0$. Согласно (8) это означает, что $V(t_n) \subset (V(\tau))^\varepsilon$, $n \geqslant n_0$. Но $V(\tau)$ — ограниченное множество, поэтому его ε-расширение $(V(\tau))^\varepsilon$ также ограничено. Тогда последовательность $\{u_n\}\colon u_n \in V(t_n) \subset (V(\tau))^\varepsilon$, $n = 1, 2, \ldots$, будет ограниченной. В то же время по построению $|u_n| \geqslant n$, $n = 1, 2, \ldots$ Полученное противоречие доказывает лемму 1. □

Лемма 2. *Пусть $V(t)$, $t_0 \leqslant t \leqslant T$, — семейство множеств из E^r, причем в каждой точке $t \in [t_0, T]$ множество $V(t)$ замкнуто, ограничено и непрерывно по Хаусдорфу. Тогда это семейство равномерно непрерывно на отрезке $[t_0, T]$, т. е. для любого $\varepsilon > 0$ найдется $\delta > 0$ такое, что $h(V(t), V(\tau)) < \varepsilon$ для всех $t, \tau \in [t_0, T]$, лишь бы $|t - \tau| < \delta$.*

Доказательство. Пусть множества $V(t)$ не являются равномерно непрерывными на $[t_0, T]$. Это значит, что существует число $\varepsilon_0 > 0$ такое, что для любого натурального числа n найдутся точки t_n, $\tau_n \in [t_0, T]$, для которых хотя $|t_n - \tau_n| < 1/n$, но $h(V(t_n), V(\tau_n)) \geqslant \varepsilon_0$, $n = 1, 2, \ldots$ Из последовательности $\{t_n\}$ выберем подпоследовательность $\{t_{n_k}\}$, сходящуюся к некоторой точке $t \in [t_0, T]$. Так как $|t_n - \tau_n| < 1/n$, то $\{\tau_{n_k}\}$ также сходится к t. Из непрерывности $V(t)$ по Хаусдорфу следует, что $h(V(t_{n_k}), V(t)) \to 0$, $h(V(\tau_{n_k}), V(t)) \to 0$ при $k \to \infty$. Тогда, пользуясь неравенством треугольника (9), получим

$$h(V(t_{n_k}), V(\tau_{n_k})) \leqslant h(V(t_{n_k}), V(t)) + h(V(\tau_{n_k}), V(t)) \to 0$$

при $k \to \infty$. В то же время по построению $h(V(t_{n_k}), V(\tau_{n_k})) \geqslant \varepsilon_0 > 0$, $k = 1, 2, \ldots$ Полученное противоречие доказывает лемму 2. □

Определение 3. *Хаусдорфовым модулем непрерывности семейства множеств* $V(t)$, $t_0 \leqslant t \leqslant T$, называется функция $\omega_V(d) = \sup h(v(t), V(\tau))$, где верхняя грань берется по всем t, $\tau \in [t_0, T]$, для которых $|t - \tau| \leqslant d$.

Нетрудно видеть, что $\omega_V(d)$ не убывает при возрастании d. Если выполнены условия леммы 2, то $\omega_V(d) \to 0 = \omega_V(0)$ при $d \to 0$.

Заметим, что для множеств $B(t)$ из примера 4 модуль непрерывности равен $\omega_B(d) = 3d$. Для множеств $V(t)$ из примера 3 для модуля непрерывности $V(t)$ справедлива оценка $\omega_V(d) \leqslant \sqrt{r} \max\{\omega_\alpha(d); \omega_\beta(d)\}$, где $\omega_\alpha(d)$, $\omega_\beta(d)$ — модули непрерывности функций $\alpha(t)$, $\beta(t)$, $t_0 \leqslant t \leqslant T$. В частности, если $\alpha(t)$, $\beta(t)$ удовлетворяют условию Гельдера $|\alpha(t) - \alpha(\tau)| \leqslant L|t - \tau|^\alpha$, $|\beta(t) - \beta(\tau)| \leqslant L|t - \tau|^\alpha$, $0 < \alpha \leqslant 1$, то $\omega_V(d) \leqslant \sqrt{r} L d^\alpha$.

Лемма 3. *Пусть* $V(t)$, $t_0 \leqslant t \leqslant T$, *— семейство множеств из* E^r, *причем в каждой точке* $t \in [t_0, T]$ *множество* $V(t)$ *выпукло, замкнуто, ограничено и непрерывно по Хаусдорфу. Если функция* $u(t) = (u^1(t), \ldots, u^r(t))$, $t_0 \leqslant t \leqslant T$, *непрерывна в точке* t, *то функция*

$$v(t) = P_{V(t)} u(t), \quad t_0 \leqslant t \leqslant T, \tag{10}$$

где $P_{V(t)} z$ *— проекция точки* $z \in E^r$ *на множество* $V(t)$, *также непрерывна в точке* t. *Если функция* $u(t)$ *кусочно непрерывна на отрезке* $[t_0, T]$, *то функция* (10) *тоже кусочно непрерывна на этом отрезке.*

Доказательство. Пусть $t \in [t_0, T]$ — какая-либо точка непрерывности функции $u(t)$. С учетом (10) имеем

$$|v(\tau) - v(t)| = |P_{V(\tau)} u(\tau) - v(t)| \leqslant$$
$$\leqslant |P_{V(\tau)} u(\tau) - P_{V(\tau)} u(t)| + |P_{V(\tau)} u(t) - v(t)|, \quad \tau \in [t_0, T]. \tag{11}$$

Операция проектирования обладает сжимающим свойством (теорема 1.4.2):

$$|P_{V(\tau)} u(\tau) - P_{V(\tau)} u(t)| \leqslant |u(\tau) - u(t)|.$$

Отсюда и из непрерывности $u(t)$ в точке t следует, что первое слагаемое в правой части неравенства (11) стремится к нулю при $\tau \to t$. Покажем, что и второе слагаемое стремится к нулю при $\tau \to t$.

По лемме 1 семейство множеств $V(\tau)$ ограничено равномерно по $\tau \in [t_0, T]$. Следовательно, множество точек $\{P_{V(\tau)} u(t)\} \in V(\tau)$, $t_0 \leqslant \tau \leqslant T$, ограничено и имеет хотя бы одну предельную точку ω при $\tau \to t$. Это значит, что существует последовательность $\{\tau_k\} \to t$ такая, что $\omega_k = P_{V(\tau_k)} u(t) \to \omega$ при $k \to \infty$. Покажем, что $\omega = P_{V(t)} u(t) = v(t)$. Сначала убедимся в том, что $\omega \in V(t)$. В силу непрерывности семейства множеств $V(t)$ по Хаусдорфу для любого $\varepsilon > 0$ найдется номер k_0 такой, что $h(V(\tau_k), V(t)) \leqslant \varepsilon$ при всех $k \geqslant k_0$. Согласно (8) тогда $V(\tau_k) \subset (V(t))^\varepsilon$ при всех $k \geqslant k_0$. Поэтому $\omega_k \in (V(t))^\varepsilon$, $k \geqslant k_0$. Так как $V(t)$ замкнуто, то $(V(t))^\varepsilon$ также замкнуто. Отсюда и из того, что $\{\omega_k\} \to \omega$, следует, что $\omega \in (V(t))^\varepsilon$. В силу произвольности $\varepsilon > 0$ и замкнутости $V(t)$ последнее включение возможно лишь в случае $\omega \in V(t)$.

Далее, имеем

$$|\omega_k - u(t)| = |P_{V(\tau_k)} u(t) - u(t)| = \inf_{u \in V(\tau_k)} |u - u(t)| \leqslant |P_{V(\tau_k)} v(t) - u(t)| \leqslant$$
$$\leqslant |v(t) - u(t)| + |P_{V(\tau_k)} v(t) - v(t)|, \quad k = 1, 2, \ldots$$

Поскольку

$$|P_{V(\tau_k)} v(t) - v(t)| = \inf_{u \in V(\tau_k)} |u - v(t)| = \rho(v(t), V(\tau_k)) \leqslant$$
$$\leqslant \sup_{v \in V(t)} \rho(v, V(\tau_k)) \leqslant h(V(t), V(\tau_k)) \to 0$$

при $k \to \infty$, то, переходя к пределу при $k \to \infty$, из предыдущего неравенства получим $|\omega - u(t)| \leqslant |v(t) - u(t)|$. С другой стороны, по определению проекции точки и условию (10) с учетом $\omega \in V(t)$ имеем

$$|v(t) - u(t)| = \inf_{u \in V(t)} |u - u(t)| \leqslant |\omega - u(t)|.$$

Следовательно,

$$|\omega - u(t)| = |v(t) - u(t)| = |P_{V(t)} u(t) - u(t)|.$$

Однако проекция точки $u(t)$ на выпуклое замкнутое множество $V(t)$ определяется однозначно (см. теорему 1.4.2), т. е. $\omega = v(t) = P_{V(t)} u(t)$.

Тем самым показано, что любая точка ω, являющаяся предельной для семейства точек $\{P_{V(\tau)} u(t)\}$ при $\tau \to t$, совпадает с $v(t)$. Это значит, что указанное семейство сходится к $v(t)$, т. е. $|P_{V(\tau)} u(t) - v(t)| \to 0$ при $\tau \to t$.

Таким образом, второе слагаемое в правой части неравенства (11) также стремится к нулю при $\tau \to t$. Следовательно, если $u(t)$ непрерывна в точке t, то и функция $v(t)$ из (10) также непрерывна в этой точке. В частности, если $u(t)$ непрерывна на отрезке $[t_0, T]$, то $v(t)$ также непрерывна на этом отрезке.

Наконец, из приведенного доказательства видно, что если существует предел $\lim_{\tau \to t \pm 0} u(\tau) = u(t \pm 0)$, то существует $\lim_{\tau \to t \pm 0} v(\tau) = P_{V(t)} u(t \pm 0)$. Это значит, что если $u(t)$ кусочно непрерывна на отрезке $[t_0, T]$, то $v(t)$ также кусочно непрерывна на этом отрезке. Лемма 3 доказана. □

4. Вернемся к задачам (1)–(3) и (4)–(6). Через J_* обозначим нижнюю грань функции (1) при условиях (2), (3), через I_{N*} — нижнюю грань функции (4) при условиях (5), (6).

Теорема 1. *Пусть матрицы* $A(t)$, $B(t)$, $f(t)$ *кусочно непрерывны на отрезке* $[t_0, T]$, *множества* $V(t)$ *выпуклы, замкнуты, ограничены и*

непрерывны по Хаусдорфу при всех $t \in [t_0, T]$. *Пусть разбиения* $\{t_i, i = 0, \ldots, N\}$ *отрезка* $[t_0, T]$ *таковы, что*

$$d_N = \max_{0 \leqslant i \leqslant N-1} \Delta t_i \leqslant (T - t_0) M_0 / N, \quad N = 1, 2, \ldots$$

Тогда $\lim\limits_{N \to \infty} I_{N*} = J_*$ *и справедлива оценка*

$$-C_6 \delta_N \leqslant I_{N*} - J_* \leqslant C_5 \delta_N, \quad N = 1, 2, \ldots, \tag{12}$$

где $C_5 = 2C_0 + 2C_4 + 2|y| = C_6$, *постоянные* C_0, C_4 *взяты из* (1.8), (1.16) *при* $W = U$, $W_N = U_N$ *соответственно, а величина* δ_N *определяется формулой*

$$\delta_N = e^{A_{\max}(T - t_0)M_0} \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} (\|A(\tau) - A_i\| C_0 + A_{\max} C_1 d_N + |f(\tau) - f_i| + \sup_{t_0 \leqslant t \leqslant T} \sup_{u \in V(t)} |u| \|B(\tau) - B_i\| + B_{\max} \omega(d_N)) d\tau; \tag{13}$$

здесь $\omega(d)$ *— хаусдорфов модуль непрерывности множеств* $V(t)$, $t_0 \leqslant t \leqslant T$, $A_{\max} = \sup\limits_{t_0 \leqslant t \leqslant T} \|A(t)\|$, $B_{\max} = \sup\limits_{t_0 \leqslant t \leqslant T} \|B(t)\|$, *постоянная* C_1 *взята из оценки* (1.11) *при* $W = U$.

Доказательство. Положим $X = L_2^r[t_0, T]$, $X_N = L_{2N}^r$. Отображения Q_N: $X \to X_N$, P_N: $X_N \to X$ определим так:

$$Q_N(u) = (u_0, u_1, \ldots, u_{N-1}): \; u_i = P_{V_i}(v_i), \quad v_i = \frac{1}{\Delta t_i} \int_{t_i}^{t_{i+1}} u(t) dt, \tag{14}$$

$$P_N([u]_N) = P_{V(t)} u_i \quad \text{при} \quad t_i \leqslant t < t_{i+1}, \quad i = 0, \ldots, N-1, \tag{15}$$

где $P_V(z)$ — проекция точки $z \in E^r$ на множество V. Как видим, в рассматриваемом случае отображения Q_N, P_N имеют несколько более сложный вид, чем в предыдущих параграфах, в которых задача вида (1)–(3) исследовалась при $V(t) \equiv V$, $t_0 \leqslant t \leqslant T$. При $V(t) \equiv V$, $t_0 \leqslant t \leqslant T$, формулы (14), (15) превращаются в формулы (1.20) и (1.21) соответственно.

Так как множества $V(t)$ выпуклы и замкнуты при каждом $t \in [t_0, T]$, то отображения Q_N, P_N из (14), (15) определяются однозначно. Функция $v(t) = P_N([u]_N) = P_{V(t)} u_N(t)$, полученная проектированием кусочно постоянной функции $u_N(t) = u_i$, $t_i \leqslant t < t_{i+1}$, $i = 0, \ldots, N-1$, согласно лемме 3, кусочно непрерывна на $[t_0, T]$ и, следовательно, $P_N([u]_N) \in L_2^r[t_0, T]$ при всех $[u]_N \in L_{2N}^r$. Отсюда и из (14), (15) вытекает, что

$$Q_N(u) \in U_N, \quad P_N([u]_N) \in U \quad \text{при всех} \quad u \in U, \quad [u]_N \in U_N, \quad N = 1, 2, \ldots$$

Заметим также, что согласно лемме 1

$$\sup_{t_0 \leqslant t \leqslant T} \sup_{u \in V(t)} |u| \leqslant R < \infty,$$

откуда следует, что множество (3) ограничено в метрике $L_\infty^r[t_0, T]$, и мы можем пользоваться оценкой (1.11) при $W = U$, а также оценками (1.8), (1.16) при $W = U$, $W_N = U_N$.

Покажем теперь, что справедливы оценки

$$\sup_{u \in U} \max_{0 \leqslant i \leqslant N} |x(t_i, u) - x_i(Q_N(u))| \leqslant \delta_N, \quad u \in U, \quad N = 1, 2, \ldots, \tag{16}$$

$$\sup_{[u]_N \in U_N} \max_{0 \leqslant i \leqslant N} |x(t_i, P_N([u]_N)) - x_i([u]_N)| \leqslant \delta_N, \quad [u]_N \in U_N, \quad N = 1, 2, \ldots, \tag{17}$$

где величина δ_N определяется формулой (13). Рассуждая так же, как при получении неравенства (1.27), с помощью оценок (1.8) и (1.11) имеем

$$|x(t_i, u) - x_i([u]_N)| \leqslant e^{A_{\max}(T - t_0)M_0} \sum_{i=0}^{N-1} \left[\int_{t_i}^{t_{i+1}} (\|A(\tau) - A_i\| C_0 + A_{\max} C_1 d_N + \|B(\tau) - B_i\| R + |f(\tau) - f_i|) d\tau + B_{\max} \left| \int_{t_i}^{t_{i+1}} (u(\tau) - u_i) d\tau \right| \right],$$
$$u \in U, \quad [u]_N \in U_N, \quad i = 0, \ldots, N, \quad N = 1, 2, \ldots \tag{18}$$

Зафиксируем какое-либо управление u и в (18) примем $[u]_N = Q_N(u)$. Поскольку по условию $h(V(\tau), V_i) \leqslant \omega(|\tau - t_i|) \leqslant \omega(d_N)$ при $t_i \leqslant \tau \leqslant t_{i+1}$, то согласно (8) имеем $V(\tau) \subset V_i^{\omega(d_N)}$ при всех $\tau \in [t_i, t_{i+1}]$, $i = 0, \ldots, N-1$. Тогда из включения $u(\tau) \subset V(\tau)$ следует, что $u(\tau) \in V_i^{\omega(d_N)}$ почти всюду на $[t_i, t_{i+1}]$. Отсюда, замечая, что множество $V_i^{\omega(d_N)}$ выпукло, замкнуто и не зависит от τ, с помощью леммы 1.1 получаем $v_i = \frac{1}{\Delta t_i} \int_{t_i}^{t_{i+1}} u(\tau) d\tau \in V_i^{\omega(d_N)}$, $i = 0, \ldots, N-1$. Следовательно,

$$|P_{V_i}(v_i) - v_i| = \rho(v_i, V_i) \leqslant h(V_i^{\omega(d_N)}, V_i) \leqslant \omega(d_N), \quad i = 0, \ldots, N-1.$$

Отсюда и из (14) имеем

$$\left| \int_{t_i}^{t_{i+1}} (u(\tau) - u_i) d\tau \right| = \Delta t_i |v_i - u_i| \leqslant \Delta t_i \omega(d_N), \quad i = 0, \ldots, N-1.$$

Тогда из (18) при $[u]_N = Q_N(u)$, $u \in U$, получим оценку (16).

Далее, возьмем какое-либо $[u]_N \in U_N$ и в (18) примем $u = P_N([u]_N)$. Из того, что $h(V(\tau), V_i) \leqslant \omega(|\tau - t_i|) \leqslant \omega(d_N)$ при $t_i \leqslant \tau \leqslant t_{i+1}$, согласно (8) следует, что $V_i \subset (V(\tau))^{\omega(d_N)}$ при всех $\tau \in [t_i, t_{i+1}]$. Поскольку $[u]_N = (u_0, u_1, \ldots, u_{N-1}) \in U_N$ означает, что $u_i \in V_i$, то $u_i \in (V(\tau))^{\omega(d_N)}$ при всех $\tau \in [t_i, t_{i+1}]$, $i = 0, \ldots, N-1$. Следовательно, $|P_{V(\tau)} u_i - u_i| = \rho(u_i, V(\tau)) \leqslant h((V(\tau))^{\omega(d_N)}, V_i) \leqslant \omega(d_N)$; отсюда и из (15) имеем $\left| \int_{t_i}^{t_{i+1}} (P_N([u]_N) - u_i) d\tau \right| \leqslant \Delta t_i \omega(d_N)$, $i = 0, \ldots, N-1$. Тогда из (18) при $u = P_N([u]_N)$, $[u]_N \in U_N$ следует оценка (17).

Далее, рассуждая так же, как в леммах 1.3 и 1.4, из оценок (16), (17) получаем, что

$$|I_N(Q_N(u)) - J(u)| \leqslant C_5 \delta_N = \beta_N, \quad u \in U, \quad N = 1, 2, \ldots,$$

$$|J(P_N([u]_N)) - I_N([u]_N)| \leqslant C_6 \delta_N = \gamma_N, \quad [u]_N \in U_N, \quad N = 1, 2, \ldots$$

Таким образом, выполнены все условия теоремы 2.1, из которой следует оценка (12). Остается заметить, что величина δ_N из оценки (12), определяемая формулой (13), стремится к нулю при $N \to \infty$; это доказывается так же, как аналогичное утверждение в лемме 1.2. Теорема 1 доказана. □

Предлагаем читателям самостоятельно рассмотреть задачи (1)–(3) и (4)–(6) при наличии фазовых ограничений

$$x(t, u) \in G(t), \quad t_0 \leqslant t \leqslant T, \quad x_i([u]_N) \in G(t_i), \quad i = 0, \ldots, N,$$

где $G(t)$ — замкнутые множества, непрерывные по Хаусдорфу на отрезке $[t_0, T]$, и доказать аналоги теорем из § § 3, 4.

§ 6. Аппроксимация задачи быстродействия

1. Пусть процесс описывается условиями

$$\dot{x}(t)=A(t)x(t)+B(t)u(t)+f(t), \quad t\geqslant t_0, \quad x(t_0)=x_0, \tag{1}$$

$$u=u(t)\in L_2^r[t_0,a] \text{ при любом } a>t_0,\ u(t)\in V(t) \text{ для почти всех } t\geqslant t_0, \tag{2}$$

где $A(t)$, $B(t)$, $f(t)$ — матрицы порядка $n\times n$, $n\times r$, $n\times 1$ соответственно, определенные при всех $t\geqslant t_0$ и кусочно непрерывные на любом конечном отрезке $[t_0, a]$; начальный момент t_0 и точка $x_0\in E^n$ известны; $V(t)$ — заданное при $t\geqslant t_0$ семейство множеств из E^r. Через $x(t,u)$, $t\geqslant t_0$, как обычно, будем обозначать траекторию задачи (1), соответствующую управлению $u=u(t)$, $t\geqslant t_0$.

Пусть Y и $G(t)$, $t\geqslant t_0$, — заданные множества из E^n.

Определение 1. Скажем, что система (1), (2) $(T, G(t), Y)$-*управляема*, если существуют хотя бы одно управление $u=u(t)$, $t\geqslant t_0$, удовлетворяющее условиям (2), и момент $t=t(u)$, $t_0\leqslant t(u)\leqslant T$, такие, что

$$x(t,u)\in G(t), \quad t_0\leqslant t\leqslant t(u), \tag{3}$$

$$x(t(u),u)\in Y, \quad x(t,u)\notin Y \quad \text{при} \quad t_0\leqslant t<t(u). \tag{4}$$

Момент времени $t(u)$, удовлетворяющий условиям (4), назовем *временем первой встречи* траектории $x(t,u)$ с множеством Y.

Таким образом, $(T, G(t), Y)$-управляемость системы (1), (2) при некотором $T\geqslant t_0$ означает, что множество $U(T)$, которое состоит из всех управлений $u=u(t)$, удовлетворяющих условиям (2)–(4), непусто. Напоминаем, что задача управления системой (1) при $G(t)\equiv E^n$, $Y=\{x_1\}$ и специальном выборе множества $V(t)$ исследовалась в §§ 8.11, 8.12.

Будем рассматривать задачу

$$t(u)\to\inf, \quad u\in U(T), \tag{5}$$

представляющую собой задачу быстродействия, в которой требуется за наименьшее время попасть из точки x_0 на множество Y, двигаясь по траекториям системы (1), (2) с соблюдением фазовых ограничений (3).

Величину $t_*=\inf t(u)$ называют *оптимальным временем* задачи (1)–(5); управление $u_*=u_*(t)\in U(T)$, для которого $t(u_*)=t_*$, называют *оптимальным управлением*, а $x(t,u_*)$ — *оптимальной траекторией* задачи (1)–(5).

2. Приведем достаточные условия, при которых в (5) нижняя грань достигается.

Теорема 1. *Пусть матрицы $A(t)$, $B(t)$, $f(t)$, $t\geqslant t_0$, кусочно непрерывны на любом конечном отрезке, множество $V(t)$ при каждом t выпукло, замкнуто и* $\sup\limits_{t_0\leqslant t\leqslant a}\sup\limits_{u\in V(t)}|u|\leqslant R=R(a)<\infty$ *при всех $a>t_0$; множества $G(t)$, $t\geqslant t_0$, и Y замкнуты; система* (1), (2) $(T, G(t), Y)$-*управляема при некотором $T>t_0$. Тогда $t_*=\inf\limits_{U(T)} t(u)\leqslant T$ и в задаче* (1)–(5) *существует оптимальное управление.*

Доказательство. По условию множество $U(T)\neq\varnothing$, и при любых $u\in U(T)$ справедливо неравенство $t_0\leqslant t(u)\leqslant T$. Поэтому $t_0\leqslant t_*\leqslant T<\infty$. По определению нижней грани (5) существуют последовательность $\{t_k\}\to t_*$, $T\geqslant t_k\geqslant t_*$, $k=1,2,\ldots$, и управления $u_k=u_k(t)$, $t_0\leqslant t\leqslant T$, $k=1,2,\ldots$, такие, что

$$x(t,u_k)\in G(t), \quad t_0\leqslant t\leqslant t_k, \quad x(t_k,u_k)\in Y,$$

$$x(t,u_k)\notin Y, \quad t_0\leqslant t<t_k, \quad k=1,2,\ldots$$

Заметим, что множество

$$W(T)=\{u=u(t)\in L_2^r[t_0,T]\colon u(t)\in V(t) \text{ почти всюду на } [t_0,T]\} \tag{6}$$

выпукло, замкнуто и ограничено в метрике $L_2^r[t_0, T]$ (пример 8.2.9) и, следовательно, слабо компактно в $L_2^r[t_0, T]$ (теорема 8.2.6). Поэтому, выбирая при необходимости подпоследовательность из $\{u_k\}$, можем считать, что сама последовательность $\{u_k\}$ сходится к некоторому управлению $u_*=u_*(t)\in W(T)$ слабо в $L_2^r[t_0, T]$. Тогда $\{x(t,u_k)\}$ сходится к $x(t,u_*)$ равномерно на $[t_0, T]$. Однако $x(t,u_k)\in G(t)$ при всех t, $t_0\leqslant t\leqslant t_*\leqslant t_k\leqslant T$, поэтому в силу замкнутости $G(t)$ справедливо включение $x(t,u_*)\in G(t)$, $t_0\leqslant t\leqslant t_*$. Далее, в силу оценки (1.11) имеем

$$\sup_{u\in W(T)}|x(t_k,u)-x(t_*,u)|\leqslant C_1|t_k-t_*|\to 0 \quad \text{при} \quad k\to\infty.$$

Тогда $|x(t_k,u_k)-x(t_*,u_*)|\leqslant|x(t_k,u_k)-x(t_*,u_k)|+|x(t_*,u_k)-x(t_*,u_*)|\to 0$ при $k\to\infty$, т. е. $\{x(t_k,u_k)\}\to x(t_*,u_*)$. Однако $x(t_k,u_k)\in Y$, $k=1,2,\ldots$, причем Y — замкнутое множество. Следовательно, $x(t_*,u_*)\in Y$. Таким образом, $u_*\in U(T)$ и время первой встречи $t(u_*)$ траектории $x(t,u_*)$ с множеством Y таково, что $t(u_*)\leqslant t_*$. С другой стороны, $t_*\leqslant t(u_*)$ в силу определения t_*. Это значит, что $t(u_*)=t_*$, т. е. u_* — оптимальное управление в задаче (1)–(5). Теорема 1 доказана. □

3. Для аппроксимации задачи (1)–(5) для каждого натурального числа $N\geqslant 1$ на полуоси $t\geqslant t_0$ введем точки $t_0=t_{0N}<t_{1N}<\ldots<t_{iN}<\ldots$, $\lim\limits_{i\to\infty} t_{iN}=\infty$. Положим $A_{iN}=A(t_{iN}+0)$, $B_{iN}=B(t_{iN}+0)$, $f_{iN}=f(t_{iN}+0)$, $V_{iN}=V(t_{iN})$, $G_{iN}=G(t_{iN})$, $\Delta t_{iN}=t_{i+1N}-t_{iN}$, и рассмотрим следующую аппроксимацию системы (1), (2):

$$x_{i+1N}=x_{iN}+\Delta t_{iN}(A_{iN}x_{iN}+B_{iN}u_{iN}+f_{iN}), \quad i=0,1,\ldots, \quad x_{0N}=x_0, \tag{7}$$

$$[u]_N=(u_{0N},u_{1N},\ldots,u_{iN},\ldots)\colon u_{iN}\in V_{iN}, \quad i=0,1,\ldots \tag{8}$$

Через $[x([u]_N)]_N=(x_0([u]_N)=x_0,\ldots,x_i([u]_N)=x_{iN},\ldots)$ будем обозначать траекторию дискретной задачи (7), соответствующую управлению $[u]_N$.

Введем расширение множеств G^{ξ_N}, Y^{ν_N}, где $\{\xi_N\}$, $\{\nu_N\}$ — положительные последовательности, стремящиеся к нулю. Напоминаем, что согласно (3.7) ε-расширением множества $Z\subset E^n$ называется множество

$$Z^\varepsilon=\{x\in E^n\colon \rho(x,Z)=\inf_{z\in Z}|x-z|\leqslant\varepsilon\}.$$

Определение 2. Систему (7), (8) назовем $(T, G_{iN}^{\xi_N}, Y^{\nu_N})$-*управляемой*, если существуют хотя бы одно управление $[u]_N$, удовлетворяющее условиям (8), и точка $t_{i_NN}\in\{t_{iN}\}$, $t_0\leqslant t_{i_NN}\leqslant T$, такие, что

$$x_i([u]_N)\in G_{iN}^{\xi_N}, \quad i=0,\ldots,i_N, \tag{9}$$

$$x_{i_NN}([u]_N)\in Y^{\nu_N}, \quad x_i([u]_N)\notin Y^{\nu_N}, \quad \text{при} \quad 0\leqslant i\leqslant i_N-1. \tag{10}$$

Момент времени $t_{i_N N} = t_N([u]_N)$, удовлетворяющий условиям (10), назовем *временем первой встречи дискретной траектории* $[x([u]_N)]_N$ *с множеством* Y^{ν_N}.

Таким образом, $(T, G_{iN}^{\xi_N}, Y^{\nu_N})$-управляемость системы (7), (8) означает, что множество $U_N(T)$, которое состоит из всех управлений $[u]_N$, удовлетворяющих условиям (8)–(10), непусто.

Рассмотрим дискретную задачу быстродействия

$$t_N([u]_N) \to \inf, \quad [u]_N \in U_N(T). \tag{11}$$

Величину $t_{N*} = \inf\limits_{U_N(T)} t_N([u]_N)$ будем называть *оптимальным временем задачи* (7)–(11).

Приведем достаточные условия, при которых из $(T, G(t), Y)$-управляемости системы (1), (2) следует $(T, G_{iN}^{\xi_N}, Y^{\nu_N})$-управляемость системы (7), (8) и последовательность задач (7)–(11) аппроксимирует задачу (1)–(5) по функции, т. е. $\lim\limits_{N\to\infty} t_{N*} = t_*$.

Теорема 2. *Пусть матрицы* $A(t), B(t), f(t)$ *определены при* $t \geqslant t_0$, *кусочно непрерывны на любом конечном отрезке* $[t_0, a]$; *множество* $V(t)$ *при всех* $t \geqslant t_0$ *выпукло, замкнуто, ограничено и непрерывно по Хаусдорфу, множество* $G(t)$ *при всех* $t \geqslant t_0$ *замкнуто и непрерывно по Хаусдорфу; система* (1), (2) $(T, G(t), Y)$-*управляема для некоторого* T, $t_0 \leqslant T < \infty$. *Пусть разбиения* $\{t_{iN}\} = \{t_{0N} = t_0 < t_{1N} < \ldots < t_{iN} < \ldots\}$ *таковы, что*

$$d_N = d_N(T) = \max_{0\leqslant i\leqslant m_N} \Delta t_{iN} = (T - t_0)M(T)/(m_N + 1), \quad N = 1, 2, \ldots,$$

где m_N — *число, определяемое условием* $t_{m_N N} < T \leqslant t_{m_N+1, N}$, $M_0(T) = \mathrm{const} > 0$; *последовательности* $\{\xi_N\}$, $\{\nu_N\}$ *из* (9), (10) *стремятся к нулю и таковы, что*

$$\xi_N \geqslant \delta_N, \quad \nu_N \geqslant \delta_N + C_1 d_N + \omega_G(d_N), \quad N = 1, 2, \ldots,$$

где величина δ_N *определяется формулой* (5.13), *постоянная* C_1 *взята из* (1.11) *при* $W = W(T)$, *а* $\omega_G(d)$ — *хаусдорфов модуль непрерывности множеств* $G(t)$ *на* $[t_0, T]$. *Тогда дискретная система* (7), (8) $(T, G_{iN}^{\xi_N}, Y^{\nu_N})$-*управляема при* $N = 1, 2, \ldots$ *и* $\lim\limits_{N\to\infty} t_{N*} = t_*$.

Доказательство. Положим $X = L_2^r[t_0, T]$, $X_N = L_{2N}^r = \{[u]_N = (u_{0N}, u_{1N}, \ldots, u_{m_N N})$: $\sum\limits_{i=0}^{m_N} u_{iN}^2 \Delta t_{iN} = \|[u]_N\|_{L_{2N}}^2 < \infty\}$. Напоминаем, что $t_{m_N N} < T \leqslant t_{m_N+1, N}$. Ниже будем считать, что $t_{m_N+1, N} = T$; узловые точки $t_{iN} > T$ нам не понадобятся, так как все рассуждения будут проводиться на отрезке $[t_0, T]$. Введем отображения Q_N: $X \to X_N$, P_N: $X_N \to X$ следующим образом:

$$Q_N(u) = (u_{0N}, u_{1N}, \ldots, u_{m_N N})\colon u_{iN} = P_{V_{iN}}(v_{iN}),$$

$$v_{iN} = \frac{1}{\Delta t_{iN}} \int\limits_{t_{iN}}^{t_{i+1N}} u(t)\,dt, \quad i = 0, \ldots, m_N, \tag{12}$$

$$P_N([u]_N) = P_{V(t)}(u_i), \quad t_{iN} < t \leqslant t_{i+1N}, \quad i = 0, \ldots, m_N, \tag{13}$$

где $P_V(z)$ — проекция точки $z \in E^r$ на множество V.

Дальнейшие рассуждения, представляющие собой проверку условий 1)–3) теоремы 2.4, оформим в виде трех лемм.

Лемма 1. *Пусть выполнены условия теоремы* 2. *Тогда*

$$Q_N(u) \in U_N(T), \quad t_N(Q_N(u)) \leqslant t(u) \quad \textit{при всех} \quad u \in U(T), \quad N = 1, 2, \ldots$$

Доказательство. Возьмем произвольное управление $u \in U(T)$. Тогда существует момент $t(u)$ — время первой встречи траектории $x(t, u)$ со множеством Y, определяемый условиями (3), (4). Возьмем точку $t_{s_N N} \in \{t_{iN}\}$ такую, что $t_{s_N N} < t(u) \leqslant t_{s_N+1, N}$. Рассмотрим задачу (7) при $[u]_N = Q_N(u)$. Покажем, что соответствующая дискретная траектория $x_i(Q_N(u))$, $i = 0, \ldots, m_N + 1$, такова, что

$$x_i(Q_N(u)) \in G_{iN}^{\xi_N}, \quad i = 0, \ldots, s_N, \quad x_{s_N}(Q_N(u)) \in Y^{\nu_N}. \tag{14}$$

В самом деле, согласно оценке (5.16)

$$|x(t_i, u) - x_i(Q_N(u))| \leqslant \delta_N, \quad i = 0, \ldots, m_N + 1. \tag{15}$$

Так как $x(t, u) \in G(t)$ при $t_0 \leqslant t \leqslant t(u)$, то $x(t_i, u) \in G_{iN}$ и $x_i(Q_N(u)) \in G_{iN}^{\delta_N} \subset G_{iN}^{\xi_N}$ при всех $i = 0, \ldots, s_N$. Далее, согласно оценке (1.11) и выбору узла $t_{s_N N}$ имеем

$$|x(t(u), u) - x(t_{s_N N}, u)| \leqslant C_1(t_{s_N+1, N} - t_{s_N N}) \leqslant C_1 d_N.$$

Отсюда и из (15) с учетом включения $x(t(u), u) \in Y$ получаем $x_{s_N}(Q_N(u)) \in Y^{\delta_N + C_1 d_N} \subset Y^{\nu_N}$. Включения (14) доказаны. Отсюда следует, что время первой встречи $t_N(Q_N(u))$ траектории $[x(Q_N(u))]_N$ со множеством Y^{ν_N} удовлетворяет неравенствам $t_N(Q_N(u)) \leqslant t_{s_N N} < t(u) \leqslant T$. Кроме того, $Q_N(u) \in U_N(T)$, так что система (7), (8) $(T, G_{iN}^{\xi_N}, Y^{\nu_N})$-управляема и дискретная задача быстродействия (7)–(11) имеет смысл. Лемма 1 доказана. □

Лемма 2. *Пусть выполнены все условия теоремы* 2, *и пусть* $U^{\varepsilon_N}(T)$ — *множество всех управлений* $u = u(t)$, *которые удовлетворяют условиям* (2) *и для которых существует момент* $t(u)$, $t_0 \leqslant t(u) \leqslant T$, *такой, что*

$$x(t, u) \in G^{\varepsilon_N}(t), \quad t_0 \leqslant t \leqslant t(u),$$

$$x(t(u), u) \in Y^{\varepsilon_N}, \quad x(t, u) \notin Y^{\varepsilon_N} \quad \textit{при} \quad t_0 \leqslant t < t(u).$$

Пусть $\varepsilon_N = \xi_N + \nu_N$, $N = 1, 2, \ldots$ *Тогда*

$$P_N([u]_N) \in U^{\varepsilon_N}(T), \quad t(P_N([u]_N)) \leqslant t_N([u]_N)$$

при всех $[u]_N \in U_N(T)$, $N = 1, 2, \ldots$

Доказательство. Возьмем произвольное управление $[u]_N \in U_N(T)$. Тогда существует момент $t_N([u]_N) = t_{i_N N} \leqslant T$, определяемый условиями (9), (10). Рассмотрим задачу (1) при $u = P_N([u]_N)$ и покажем, что

$$x(t, P_N([u]_N)) \in G^{\varepsilon_N}(t), \quad t_0 \leqslant t \leqslant t_{i_N N}, \quad x(t_{i_N N}, P_N([u]_N)) \in Y^{\varepsilon_N}. \tag{16}$$

Согласно оценке (5.17)

$$|x(t_i, P_N([u]_N)) - x_i([u]_N)| \leqslant \delta_N, \quad i = 0, \ldots, i_N. \tag{17}$$

Отсюда и из (9) следует, что $x(t_i, P_N([u]_N)) \in G_{iN}^{\xi_N + \delta_N}$, $i = 0, \ldots, i_N$. Далее из оценки (1.11) имеем

$$|x(t, P_N([u]_N)) - x(t_i, P_N([u]_N))| \leqslant C_1 |t - t_{iN}| \leqslant C_1 d_N$$

при всех t, $t_{iN} \leqslant t \leqslant t_{i+1N}$, $i = 0, \ldots, i_N$. Это значит, что $x(t, P_N([u]_N)) \in G_{iN}^{\xi_N + \delta_N + C_1 d_N}$, $t_{iN} \leqslant t \leqslant t_{i+1N}$, $i = 0, \ldots, i_N$. Наконец, из $h(G(t), G(t_{iN})) \leqslant \omega_G(d_N)$, $t_{iN} \leqslant t \leqslant t_{i+1N}$, следует, что $G(t_{iN}) \subset G^{\omega_G(d_N)}(t)$, или $G_{iN}^{\xi_N + \delta_N + C_1 d_N} \subset G^{\varepsilon_N}(t)$, $t_{iN} \leqslant t \leqslant t_{i+1N}$, $i = 0, \ldots, i_N$. Тогда $x(t, P_N([u]_N)) \in G^{\varepsilon_N}(t)$, $t_0 \leqslant t \leqslant t_{i_N N}$. Кроме того, из (10) и (17) имеем $x(t_{i_N N}, P_N([u]_N)) \in Y^{\nu_N + \delta_N} \subset Y^{\varepsilon_N}$. Включения (16) доказаны. Отсюда следует, что $P_N([u]_N) \in U^{\varepsilon_N}(T)$ и время $t(P_N([u]_N))$ первой встречи траектории $x(t, P_N([u]_N))$ со множеством Y^{ε_N} не превышает $t_N([u]_N)$. Лемма 2 доказана. □

Наряду с задачей (1)–(5) рассмотрим расширенную задачу быстродействия. Для этого при каждом $\varepsilon > 0$ введем множество $U^\varepsilon(T)$ всех управлений $u = u(t)$, $t \geqslant t_0$, которые удовлетворяют условиям (2) и для которых существует момент $t(u)$, $t_0 \leqslant t(u) \leqslant T$, такой, что

$$x(t, u) \in G^\varepsilon(t), \quad t_0 \leqslant t \leqslant t(u), \tag{18}$$

$$x(t(u), u) \in Y^\varepsilon, \quad x(t, u) \notin Y^\varepsilon \quad \text{при} \quad t_0 \leqslant t < t(u). \tag{19}$$

Рассмотрим задачу

$$t(u) \to \inf, \quad u \in U^\varepsilon(T). \tag{20}$$

Обозначим через $t_*(\varepsilon) = \inf\limits_{U^\varepsilon(T)} t(u)$ — оптимальное время задачи быстродействия (1), (2), (18)–(20).

Лемма 3. *Пусть выполнены все условия теоремы* 2. *Тогда* $\lim\limits_{\varepsilon \to +0} t_*(\varepsilon) = t_*$.

Доказательство. Поскольку $G(t) \subset G^\varepsilon(t) \subset G^\gamma(t)$, $Y \subset Y^\varepsilon(t) \subset Y^\gamma$ при всех $0 < \varepsilon < \gamma$, то $U(T) \subseteq U^\varepsilon(T) \subseteq U^\gamma(T)$. Отсюда следует, что система (1), (2), $(T, G^\varepsilon(t), Y^\varepsilon)$-управляема при всех $\varepsilon > 0$ и $t_*(\gamma) \leqslant t_*(\varepsilon) \leqslant t_*$ при $0 < \varepsilon < \gamma$. Тогда существует $\lim\limits_{\varepsilon \to 0} t_*(\varepsilon) = T_0 \leqslant t_*$. Покажем, что $T_0 = t_*$. Возьмем какую-либо последовательность $\{\varepsilon_k\} \to 0$, $\varepsilon_k > 0$, $k = 1, 2, \ldots$. Заметим, что из замкнутости множеств $G(t)$, Y следует замкнутость $G^\varepsilon(t)$, Y^ε. Отсюда и из теоремы 1 с учетом уже установленной $(T, G^\varepsilon(t), Y^\varepsilon)$-управляемости системы (1), (2) вытекает, что в задаче (1), (2), (18)–(20) существует хотя бы одно оптимальное управление $u_{\varepsilon *} = u_{\varepsilon *}(t) \in U^\varepsilon(T)$. Положим $u_k = u_{\varepsilon_k *}(t)$, $t_0 \leqslant t \leqslant T$, $k = 1, 2, \ldots$ Таким образом,

$$\begin{gathered} u_k \in W(T), \quad x(t, u_k) \in G^{\varepsilon_k}(t), \quad t_0 \leqslant t \leqslant t_k = t_*(\varepsilon_k), \\ x(t_k, u_k) \in Y^{\varepsilon_k}, \quad x(t, u_k) \notin Y^{\varepsilon_k}, \\ t_0 \leqslant t < t_k, \quad k = 1, 2, \ldots \end{gathered} \tag{21}$$

Поскольку множество $W(T)$, определяемое условиями (6), слабо компактно в $L_2^r[t_0, T]$, то, выбирая при необходимости подпоследовательность из $\{u_k\}$, можем считать, что сама последовательность $\{u_k\}$ слабо в $L_2^r[t_0, T]$ сходится к некоторому управлению $v_* \in W(T)$. Согласно (1.12) тогда

$$\sup_{t_0 \leqslant t \leqslant T} |x(t, u_k) - x(t, v_*)| = \mu_k \to 0 \quad \text{при} \quad k \to \infty.$$

Отсюда и из (21) следует, что

$$x(t, v_*) \in G^{\varepsilon_k + \mu_k}(t), \quad t_0 \leqslant t \leqslant t_k, \quad x(t_k, v_*) \in Y^{\varepsilon_k + \mu_k}, \quad k = 1, 2, \ldots \tag{22}$$

Кроме того, согласно оценке (1.11)

$$|x(t, v_*) - x(T_0, v_*)| \leqslant C_1 (T_0 - t_k) = \beta_k, \quad t_k \leqslant t \leqslant T_0,$$

поэтому

$$x(t, v_*) \in G^{\varepsilon_k + \mu_k + \beta_k}(t_k), \quad t_k \leqslant t \leqslant T_0, \tag{23}$$

$$x(T_0, v_*) \in Y^{\varepsilon_k + \mu_k + \beta_k}, \quad k = 1, 2, \ldots \tag{24}$$

Далее, учитывая, что множество $G(t)$ непрерывно по Хаусдорфу и $h(G(t), G(T_0)) \leqslant \omega_G(T_0 - t) \leqslant \omega_G(T_0 - t_k) = \gamma_k$ при всех t, $t_k \leqslant t \leqslant T_0$, имеем $G(t_k) \in G^{\gamma_k}(t)$ при t, $t_k \leqslant t \leqslant T_0$. Отсюда и из (23) получаем включение $x(t, v_*) \in G^{\varepsilon_k + \mu_k + \beta_k + \gamma_k}(t)$ при $t_k \leqslant t \leqslant T_0$. С учетом первого из включений (22) тогда будем иметь, что

$$x(t, v_*) \in G^{\varepsilon_k + \mu_k + \beta_k + \gamma_k}(t) \quad \text{при всех} \quad t, \quad t_0 \leqslant t \leqslant T_0, \quad k = 1, 2, \ldots \tag{25}$$

Поскольку $\varepsilon_k + \mu_k + \beta_k + \gamma_k \to 0$ при $k \to \infty$, а множества $G(t)$, Y замкнуты, то из (24), (25) при $k \to \infty$ получим, что $x(t, v_*) \in G(t)$, $t_0 \leqslant t \leqslant T_0$, $x(T_0, v_*) \in Y$. Таким образом, $v_* \in U(T)$ и время первой встречи $t(v_*)$ траектории $x(t, v_*)$ со множеством Y таково, что $t_* \leqslant t(v_*) \leqslant T_0$. Отсюда имеем неравенство $t_* \leqslant T_0$. Выше было установлено, что $T_0 \leqslant t_*$. Следовательно, $t_* = T_0 = \lim\limits_{\varepsilon \to 0} t_*(\varepsilon)$. Лемма 3 доказана. □

Таким образом, если принять $J(u) = t(u)$, $u \in U = U(T)$, $I_N([u]_N) = t_N([u]_N)$, $[u]_N \in U_N = U_N(T)$, то тогда из лемм 1–3 вытекает выполнение условий 1)–3) теоремы 2.4. Отсюда следует справедливость утверждения теоремы 2. □

4. Таким образом, показано, что при выполнении условий теоремы 2 для приближенного решения задачи быстродействия (1)–(5) может быть использована последовательность разностных аппроксимирующих задач (7)–(11). В свою очередь, для решения разностной задачи (7)–(11) при каждом фиксированном N можно рассмотреть следующее семейство задач: минимизировать функцию

$$I_N([u]_N, y) = |x_j([u]_N) - y|^2 \tag{26}$$

при условиях (7), (8) и

$$x_i([u]_N) \in G_i^{\xi_N}, \quad i = 0, \ldots, j, \quad y \in Y, \tag{27}$$

где j — фиксированный номер, последовательно пробегающий значения $j = 0, 1, \ldots, m_N$; здесь номер m_N определяется условием $t_{m_N N} < T \leqslant t_{m_N + 1, N}$, момент T взят из теоремы 2. Для решения задачи (26), (27), (7), (8) при каждом фиксированном j могут быть использованы известные методы минимизации функций конечного числа переменных или дискретные аналоги методов из гл. 8.

Обозначим $\rho_{jN} = \inf I_N([u]_N, y)$, где нижняя грань берется по всем $([u]_N, y)$, удовлетворяющим условиям (7), (8), (27). Может случиться, что $\rho_{jN} > 0$ при всех j, $0 \leqslant j < k$, а $\rho_{kN} = 0$ — это значит, что $t_{N*} = t_{kN}$. Если же $\rho_{jN} > 0$ при всех $j = 0, \ldots, m_N$, то ясно, что $t_{N*} \geqslant T > t_{m_N N}$. Отсюда следует,

что, взяв номер N достаточно большим, согласно теореме 2 в принципе можно получить достаточно точное значение оптимального времени задачи (1)–(5). Однако нужно заметить, что такой подход к решению задачи быстродействия на практике может оказаться не очень удобным, поскольку с ростом N растет число задач вида (26), (27), (7), (8) и, следовательно, вообще говоря, растет и объем вычислений. Поэтому желательно иметь другие более удобные методы решения задачи (1)–(5), не требующие перебора всех задач вида (26), (27), (7), (8).

5. Остановимся на одном из таких методов. Для простоты ограничимся рассмотрением задачи быстродействия (1)–(5) при дополнительных предположениях, когда фазовые ограничения отсутствуют, множество Y состоит из одной точки, а множество $V(t)$ не зависит от времени, т. е.

$$G(t)\equiv E^n, \quad V(t)\equiv V \quad \text{при} \quad t_0\leqslant t\leqslant T, \quad Y=\{y\}. \tag{28}$$

Как и выше, будем предполагать, что матрицы $A(t), B(t), f(t)$ кусочно непрерывны на любом конечном отрезке $[t_0, a]$, множество V выпукло, замкнуто и ограничено.

Возьмем некоторое достаточно большое число $T > t_0$, зафиксируем t, $t_0 < t \leqslant T$, и рассмотрим задачу

$$J(u, t)=|x(t, u)-y|^2\to\inf, \tag{29}$$

$$\dot{x}(\tau)=A(\tau)x(\tau)+B(\tau)u(\tau)+f(\tau), \quad t_0\leqslant\tau\leqslant t, \quad x(t_0)=x_0, \tag{30}$$

$$u=u(\tau)\in W=W(T)=\{u(\tau)\in L_2^r[t_0, T]: \\ u(\tau)\in V \text{ почти всюду на } [t_0, T]\}. \tag{31}$$

Заметим, что значения управлений $u(\tau)$ при $t\leqslant\tau\leqslant T$ на задачу (29)–(31) не влияют. Но тем не менее мы здесь рассматриваем множество (31), так как в дальнейшем нам будет удобно считать, что управления доопределены на всем отрезке $[t_0, T]$. Обозначим

$$\rho(t)=\inf_{u\in W} J(u, t), \quad t_0 < t\leqslant T; \tag{32}$$

при $t = t_0$ положим $\rho(t_0)=|x_0-y|^2$. Будем считать, что $\rho(t_0) > 0$, так как при $\rho(t_0)=0=|x_0-y|^2$ задача (1)–(5), (28) становится тривиальной: $t_*=t_0$.

Так как множество W слабо компактно в $L_2^r[t_0, T]$ и функция $J(u, t)$ слабо непрерывна на W, то в (32) нижняя грань достигается, т. е. существует управление $u=u_t\in W$ такое, что $\rho(t)=J(u_t, t)=|x(t, u_t)-y|^2$ (пример 8.2.15). Отсюда ясно, что для того, чтобы момент t_* был оптимальным временем задачи (1)–(5), (28), необходимо и достаточно, чтобы

$$\rho(t_*)=0, \quad \rho(t) > 0 \quad \text{при} \quad t_0\leqslant t < t_*,$$

т. е. t_* — минимальный корень уравнения $\rho(t)=0$. Это значит, что для поиска t_* могут быть использованы известные методы решения уравнений. В частности, здесь может быть использован метод, описанный в § 5.18 (см. процесс (5.18.20) и пояснения к нему). Этот метод был описан в предположении, что функция $\rho(t)$ удовлетворяет условию Липшица. Покажем, что в рассматриваемой задаче это условие выполняется. Пусть

$$\rho(t)=|x(t, u_t)-y|^2, \quad \rho(\tau)=|x(\tau, u_\tau)-y|^2, \quad u_t, u_\tau\in W.$$

Тогда из определения (32) функции $\rho(t)$ с учетом оценок (1.8), (1.11) имеем

$$\rho(t)-\rho(\tau)\leqslant|x(t, u_\tau)-y|^2-|x(\tau, u_\tau)-y|^2\leqslant \\ \leqslant 2(C_0+|y|)|x(t, u_\tau)-x(\tau, u_\tau)|\leqslant 2(C_0+|y|)C_1|t-\tau|,$$

$$\rho(t)-\rho(\tau)\geqslant|x(t, u_t)-y|^2-|x(\tau, u_t)-y|^2\geqslant \\ \geqslant -2(C_0+|y|)|x(t, u_t)-x(\tau, u_t)|\geqslant -2(C_0+|y|)C_1|t-\tau|.$$

Следовательно,

$$|\rho(t)-\rho(\tau)|\leqslant L|t-\tau|, \quad t_0\leqslant t, \tau\leqslant T, \quad L=2(C_0+|y|)C_1. \tag{33}$$

Для вычисления приближенного значения $\rho(t)$, удовлетворяющего условию (5.18.20) можно воспользоваться разностными аппроксимациями задачи (29)–(31), описанными в § 1.

Заметим, что метод (5.18.20) поиска минимального корня уравнения $\rho(t)$ может быть модифицирован на случай функций $\rho(t)$, удовлетворяющих более общим, чем (33), условиям

$$|\rho(t)-\rho(\tau)|\leqslant\omega(|t-\tau|), \quad t_0\leqslant t, \tau\leqslant T,$$

где $\omega(d)$ — неубывающая функция переменной $d\geqslant 0$, $\omega(0)=0$, и применен для решения нелинейных задач быстродействия [141].

О других аспектах задач быстродействия, различных приложениях, о дифференциальных играх преследования — уклонения, обобщающих задачи быстродействия на случай конфликтных ситуаций, см., например, в [10; 39; 82; 97; 98; 100; 138; 140; 141; 212; 241; 244; 245; 287; 288; 310; 312; 333; 336–338; 380; 382–384; 400; 440; 504; 505; 530; 551–554; 569–571; 589; 637; 647; 719; 755; 801; 809; 815].

§ 7. Разностная аппроксимация задачи об оптимальном нагреве стержня

1. Рассмотрим задачу

$$J(u)=\int_0^l|x(s, T; u)-b(s)|^2ds\to\inf, \qquad u\in U, \tag{1}$$

где $x=x(s, t)=x(s, t; u)$ — решение краевой задачи

$$\frac{\partial x}{\partial t}=\frac{\partial^2 x}{\partial s^2}+u(s, t), \quad (s, t)\in Q=(0, l)\times(0, T), \tag{2}$$

$$\left.\frac{\partial x}{\partial s}\right|_{s=0}=\left.\frac{\partial x}{\partial s}\right|_{s=l}=0, \quad 0 < t < T; \quad x|_{t=0}=0, \quad 0\leqslant s\leqslant l, \tag{3}$$

управление $u=u(s, t)$ принадлежит множеству

$$U=\{u(s, t)\in L_2(Q): \iint_Q u^2(s, t)dsdt\leqslant R^2\}, \tag{4}$$

$b=b(s)\in L_2(0, l)$ — заданная функция, $R=\text{const} > 0$.

Напоминаем, что близкую задачу о нагреве стержня мы уже рассматривали в § 8.7.

Определение 1. Обобщенным решением задачи (2), (3), соответствующим управлению $u \in L_2(Q)$ будем называть функцию $x = x(s, t; u) \in \in H^{1,0}(Q)$, имеющую следы $x(\cdot, t) \in L_2(0, l)$ при всех $t \in (0, T)$, непрерывные в метрике $L_2(0, l)$, и удовлетворяющую интегральному тождеству

$$\int_0^l x(s, T)\psi(s, T)ds - \iint_Q \left(x\frac{\partial\psi}{\partial t} - \frac{\partial x}{\partial s}\frac{\partial\psi}{\partial s}\right)dsdt = \iint_Q f\psi dsdt$$
$$\forall\psi = \psi(s, t) \in H^1(Q)$$

(ср. с определением 8.7.1).

Существование и единственность обобщенного решения задачи (2), (3) при каждом фиксированном $u \in L_2(Q)$ доказана, например, в [441; 492]. Отметим, что эта задача на самом деле имеет более гладкое решение, чем это указано в определении 1 — этот вопрос кратко будет обсуждаться ниже. Так же, как это делалось в § 8.7, можно доказать, что множество U_* решений задачи минимизации (1)–(4) непусто.

2. Сформулируем разностную задачу минимизации, аппроксимирующую задачу (1)–(4). На прямоугольнике $\overline{Q} = [0, l] \times [0, T]$ зададим сетку $\omega_{h\tau} = \{(s_i, t_j)\colon s_i = ih,\ t_j = j\tau,\ i = 0, \ldots, M,\ j = 0, \ldots, N\}$, где h, τ — шаги сетки, $hM = l$, $\tau N = T$. Для функции $y_{h\tau} = \{y_{ij},\ i = 0, \ldots, M,\ j = 0, \ldots, N\}$, заданной на сетке $\omega_{h\tau}$, введем разделенные разности

$$y_{\bar{t}ij} = \frac{1}{\tau}(y_{ij} - y_{ij-1}), \quad y_{\bar{s}ij} = \frac{1}{h}(y_{ij} - y_{i-1j}), \quad y_{sij} = \frac{1}{h}(y_{i+1j} - y_{ij}),$$
$$y_{\bar{s}sij} = \frac{1}{h}(y_{sij} - y_{\bar{s}ij}) = \frac{1}{h^2}(y_{i+1j} - 2y_{ij} + y_{i-1j}).$$

Рассмотрим разностную задачу минимизации:

$$I_{h\tau}(u_{h\tau}) = \sum_{i=1}^{M-1} h|y_{iN} - b_i|^2 \to \inf, \quad u_{h\tau} \in U_{h\tau}, \tag{5}$$

где $y_{h\tau} = y_{h\tau}(u_{h\tau})$ — решение разностной краевой задачи

$$y_{\bar{t}ij} = y_{\bar{s}sij} + u_{ij}, \quad i = 1, \ldots, M-1, \quad j = 1, \ldots, N, \tag{6}$$
$$y_{\bar{s}1j} = 0, \quad y_{\bar{s}Mj} = 0, \quad j = 1, \ldots, N, \quad y_{i0} = 0, \quad i = 0, \ldots, M; \tag{7}$$

разностное управление $u_{h\tau} = \{u_{ij},\ i = 1, \ldots, M-1, j = 1, \ldots, N\}$ принадлежит множеству

$$U_{h\tau} = \{u_{ij}\colon \sum_{j=1}^{N}\sum_{i=1}^{M-1} h\tau u_{ij}^2 \leqslant R^2\}. \tag{8}$$

Задача (6), (7) при каждом фиксированном $u_{h\tau} = \{u_{ij}\}$ представляет собой систему линейных алгебраических уравнений относительно неизвестных $y_{h\tau} = \{y_{ij}\}$, для нахождения ее решения можно воспользоваться методом прогонки [89; 635]. При каждом фиксированном h, τ задача (5)–(8) является задачей минимизации непрерывной функции конечного числа переменных $\{u_{ij}\}$ на компактном множестве $U_{h\tau}$, множество $U_{h\tau*}$ ее решений непусто. Для поиска решений задачи (5)–(8) могут быть использованы методы гл. 5.

Обозначим $J_* = \inf_U J(u)$, $I_{h\tau*} = \inf_{U_{h\tau}} I_{h\tau}(u_{h\tau})$. Следуя [362], ниже покажем, что при специальном выборе $b_h = (b_1, \ldots, b_{M-1})$ и $(h, \tau) \to 0$ решение разностной задачи (5)–(8) сходится к решению задачи (1)–(4) по функции,

т. е. $\lim_{(h,\tau)\to 0} I_{h\tau*} = J_*$. С этой целью сначала нам нужно будет получить априорные энергетические оценки решений краевых задач (2), (3) и (6), (7), а также оценить близость решений этих задач.

3. Начнем с доказательства следующих двух оценок для достаточно гладких классических решений задачи (2), (3):

$$\max_{0\leqslant t\leqslant T}\int_0^l x^2(s, t; u)ds + \iint_Q x_s^2(s, t; u)dsdt \leqslant C\iint_Q u^2(s, t)dsdt, \tag{9}$$

$$\max_{0\leqslant t\leqslant T}\int_0^l \left|\frac{\partial x(s,t;u)}{\partial s}\right|^2 ds + \iint_Q \left(\left|\frac{\partial x(s,t;u)}{\partial t}\right|^2 + \left|\frac{\partial^2 x(s,t;u)}{\partial s^2}\right|^2\right)dsdt \leqslant C\iint_Q u^2(s,t)dsdt. \tag{10}$$

В оценках (9), (10) постоянные C, зависящие лишь от l, T, но не зависящие от выбора $u \in L_2(Q)$, конечно, разные. Однако конкретные значения этих постоянных в дальнейшем для нас несущественны, поэтому здесь и далее подобные константы мы будем обозначать одной и той же буквой C. Кроме того, далее без дополнительных оговорок будем пользоваться неравенством Коши — Буняковского для сумм и интегралов и элементарными неравенствами

$$|ab| \leqslant \frac{\varepsilon}{2}a^2 + \frac{1}{2\varepsilon}b^2, \quad (a+b)^2 \leqslant 2a^2 + 2b^2,$$
$$(a+b+c)^2 \leqslant 3(a^2+b^2+c^2) \quad \forall a, b, c \in \mathbb{R} \quad \forall\varepsilon > 0.$$

Заметим, что оценка (9) аналогична оценке (8.7.15) и доказывается аналогично. А именно, умножим уравнение (2) на $x(s, t; u)$ и полученное равенство проинтегрируем по прямоугольнику $Q_\tau = \{(s, t)\colon 0 \leqslant s \leqslant l, 0 \leqslant t \leqslant \tau\}$, где τ — произвольный фиксированный момент времени, $0 \leqslant \tau \leqslant T$:

$$\iint_{Q_\tau}\frac{\partial x}{\partial t}xdsdt - \iint_{Q_\tau}\frac{\partial^2 x}{\partial s^2}xdsdt = \iint_{Q_\tau} uxdsdt. \tag{11}$$

С учетом условий (3) имеем

$$\iint_{Q_\tau}\frac{\partial x}{\partial t}xdsdt = \int_0^l\left(\int_0^\tau \frac{1}{2}\frac{\partial}{\partial t}(x^2)dt\right)ds = \frac{1}{2}\int_0^l x^2(s, \tau)ds,$$

$$\iint_{Q_\tau}\frac{\partial^2 x}{\partial s^2}xdsdt = \int_0^\tau\left(\frac{\partial x}{\partial s}x\Big|_{s=0}^{l} - \int_0^l\left(\frac{\partial x(s,t)}{\partial s}\right)^2 ds\right)dt = -\iint_{Q_\tau}\left(\frac{\partial x}{\partial s}\right)^2 dsdt$$

Подставим эти равенства в (11). Получим:

$$\frac{1}{2}\int_0^l x^2(s, \tau)ds + \iint_{Q_\tau}\left(\frac{\partial x}{\partial s}\right)^2 dsdt = \iint_{Q_\tau} uxdsdt \leqslant$$
$$\leqslant \int_0^\tau\left(\int_0^l x^2(s, t)ds\right)^{1/2}\left(\int_0^l u^2(s, t)ds\right)^{1/2}dt \leqslant \max_{0\leqslant t\leqslant\tau}\left(\int_0^l x^2(s, t)ds\right)^{1/2} \times$$
$$\times\int_0^\tau\left(\int_0^l u^2(s, t)ds\right)^{1/2}dt \leqslant \max_{0\leqslant t\leqslant\tau}\left(\int_0^l x^2(s, t)ds\right)^{1/2}\sqrt{T}\|u\|_{L_2(Q)}. \tag{12}$$

Отсюда следует

$$\int_0^l x^2(s, \tau)ds \leqslant \max_{0\leqslant t\leqslant T}\left(\int_0^l x^2(s, t)ds\right)^{1/2}\cdot 2\sqrt{T}\|u\|_{L_2(Q)} \qquad \forall\tau \in [0, T],$$

поэтому

$$\max_{0\leqslant\tau\leqslant T}\int_0^l x^2(s,\tau)ds = \left(\max_{0\leqslant t\leqslant T}\left(\int_0^l x^2(s,t)dt\right)^{1/2}\right)^2 \leqslant \max_{0\leqslant t\leqslant T}\left(\int_0^l x^2(s,t)ds\right)^{1/2}\cdot 2\sqrt{T}\|u\|_{L_2(Q)},$$

или

$$\max_{0\leqslant t\leqslant T}\int_0^l x^2(s,t)ds \leqslant 4T\|u\|^2_{L_2(Q)}. \tag{13}$$

Далее, из (12), (13) имеем

$$\iint\limits_{Q_\tau}\left(\frac{\partial x}{\partial s}\right)^2 dsdt \leqslant \left(\max_{0\leqslant t\leqslant T}\int_0^l x^2(s,t)ds\right)^{1/2}\sqrt{T}\|u\|_{L_2(Q)} \leqslant 2T\|u\|^2_{L_2(Q)} \quad \forall\tau\in[0,T].$$

Сложив это неравенство при $\tau=T$ с (13), придем к оценке (9) с постоянной $C=6T$.

Докажем оценку (10). Умножим уравнение (2) на $\frac{\partial x(s,t)}{\partial t}$ и проинтегрируем по области Q_τ:

$$\iint\limits_{Q_\tau}\left(\frac{\partial x}{\partial t}\right)^2 dsdt = \iint\limits_{Q_\tau}\frac{\partial^2 x}{\partial s^2}\frac{\partial x}{\partial t}dsdt = \iint\limits_{Q_\tau} u\frac{\partial x}{\partial t}dsdt. \tag{14}$$

Поскольку в силу условий (3)

$$\iint\limits_{Q_\tau}\frac{\partial^2 x}{\partial s^2}\frac{\partial x}{\partial t}dsdt = \int_0^\tau\left(\frac{\partial x}{\partial s}\frac{\partial x}{\partial t}\Big|_{s=0}^{l} - \int_0^l \frac{\partial x}{\partial s}\frac{\partial^2 x}{\partial s\partial t}ds\right)dt = -\int_0^l\left(\int_0^\tau \frac{1}{2}\frac{\partial}{\partial t}\left(\left(\frac{\partial x}{\partial s}\right)^2\right)dt\right)ds = -\frac{1}{2}\int_0^l\left(\frac{\partial x(s,\tau)}{\partial s}\right)^2 ds,$$

то из (14) имеем

$$\iint\limits_{Q_\tau}\left(\frac{\partial x}{\partial t}\right)^2 dsdt + \frac{1}{2}\int_0^l\left(\frac{\partial x(s,\tau)}{\partial s}\right)^2 ds = \iint\limits_{Q_\tau} u\frac{\partial x}{\partial t}dsdt \leqslant \frac{1}{2}\|u\|^2_{L_2(Q)} + \frac{1}{2}\iint\limits_{Q_\tau}\left(\frac{\partial x}{\partial t}\right)^2 dsdt$$

или

$$\iint\limits_{Q_\tau}\left(\frac{\partial x}{\partial t}\right)^2 dsdt + \int_0^l\left(\frac{\partial x(s,\tau)}{\partial s}\right)^2 ds \leqslant \|u\|^2_{L_2(Q)} \quad \forall\tau\in[0,T].$$

Отсюда имеем два неравенства

$$\iint\limits_{Q_\tau}\left(\frac{\partial x}{\partial t}\right)^2 dsdt \leqslant \|u\|^2_{L_2(Q)}, \quad \int_0^l\left(\frac{\partial x(s,\tau)}{\partial s}\right)^2 ds \leqslant \|u\|^2_{L_2(Q)} \quad \forall\tau\in[0,T].$$

Пользуясь произволом в выборе $\tau\in[0,T]$, получаем

$$\iint\limits_{Q}\left(\frac{\partial x}{\partial t}\right)^2 dsdt \leqslant \|u\|^2_{L_2(Q)}, \quad \max_{0\leqslant\tau\leqslant T}\int_0^l\left(\frac{\partial x(s,\tau)}{\partial s}\right)^2 ds \leqslant \|u\|^2_{L_2(Q)}. \tag{15}$$

Кроме того, из уравнения (2) с учетом (15) имеем

$$\iint\limits_{Q}\left(\frac{\partial^2 x}{\partial s^2}\right)^2 dsdt = \iint\limits_{Q}\left(\frac{\partial x}{\partial t}-u\right)^2 dsdt \leqslant 2\iint\limits_{Q}\left(\frac{\partial x}{\partial t}\right)^2 dsdt + 2\iint\limits_{Q}u^2 dsdt \leqslant 4\|u\|^2_{L_2(Q)}. \tag{16}$$

Сложив неравенства (15), (16), придем к оценке (10) с константой $C=6$. Приведем два следствия из оценок (9), (10):

$$\max_{(s,t)\in\overline{Q}}|x(s,t;u)|^2 \leqslant C\|u\|^2_{L_2(Q)}, \tag{17}$$

$$\max_{0\leqslant s\leqslant l}\int_0^T\left(\frac{\partial x(s,t;u)}{\partial s}\right)^2 dt \leqslant C\|u\|^2_{L_2(Q)}. \tag{18}$$

Для доказательства оценки (17) заметим, что

$$x^2(s,t) = \left(\int_\xi^s \frac{\partial x(\eta,t)}{\partial s}d\eta + x(\xi,t)\right)^2 \leqslant 2\left(\int_\xi^s \frac{\partial x(\eta,t)}{\partial s}d\eta\right)^2 + 2x^2(\xi,t) \leqslant$$
$$\leqslant 2l\int_0^l\left(\frac{\partial x(s,t)}{\partial s}\right)^2 ds + 2x^2(\xi,t) \quad \forall(s,t)\in Q, \quad \forall\xi\in[0,l].$$

Интегрируя это неравенство по ξ на $[0,l]$, с учетом оценок (13), (15) получим

$$lx^2(s,t) \leqslant 2l\max_{0\leqslant t\leqslant T}\int_0^l\left(\frac{\partial x(s,t)}{\partial s}\right)^2 ds + 2\max_{0\leqslant t\leqslant T}\int_0^l x^2(s,t)ds \leqslant$$
$$\leqslant (2l+8T)\|u\|^2_{L_2(Q)} \quad \forall(s,t)\in Q.$$

Отсюда получаем оценку (17) с $C=\left(2+8\frac{T}{l}\right)$. Оценка (18) следует из (16) и неравенства

$$\int_0^T\left(\frac{\partial x(s,t)}{\partial s}\right)^2 dt = \int_0^T\left(\int_0^s \frac{\partial^2 x(\xi,t)}{\partial s^2}d\xi + \frac{\partial x(0,t)}{\partial s}\right)^2 dt \leqslant l\iint\limits_{Q}\left(\frac{\partial^2 x}{\partial s^2}\right)^2 dsdt \quad \forall s\in[0,l].$$

4. Опираясь на оценки (9), (10), нетрудно доказать, что решение задачи (2), (3) представимо в виде ряда Фурье

$$x(s,t;u) = \sum_{k=1}^\infty e_k(s)\int_0^t u_k(\tau)e^{-\lambda_k(t-\tau)}d\tau, \tag{19}$$

где $e_k(s)=\sqrt{\frac{2}{l}}\cos\lambda_k s$ — собственные функции оператора $-\frac{d^2\varphi}{ds^2}$ при $\varphi'(0)=\varphi'(l)=0$, соответствующие собственному числу $\lambda_k=\left(\frac{\pi k}{l}\right)^2$, $k=1,2,\ldots$; $u_k(t)=\int_0^l u(s,t)e_k(s)ds\in L_2[0,T]$. Функции $\{e_k(s)\}$ образуют полную ортонормированную систему в $L_2[0,T]$ (см., например, [492; 557], и ряд $u(s,t)=\sum_{k=1}^\infty u_k(t)e_k(s)$ сходится в метрике $L_2(Q)$ и в метрике $L_2[0,l]$ при почти всех $t\in[0,T]$. Частичные суммы $x_m(s,t)=\sum_{k=1}^m e_k(s)\int_0^t u_k(\tau)e^{-\lambda_k(t-\tau)}d\tau$, $m=1,2,\ldots$, ряда (19) являются классическим решением задачи (2), (3), и для них справедливы априорные оценки (9), (10), (17), (18). Из этих оценок следует, что последовательность $\{x_m(s,t)\}$ фундаментальна в пространствах $C(\overline{Q})$, $H^{2,1}(Q)$, ряд $\frac{\partial x(s,t)}{\partial s}$, полученный формальным дифференцированием ряда (19), сходится в пространствах $C([0,T],L_2(0,l))$, $C([0,l],L_2(0,T))$ с нормами $\max_{0\leqslant t\leqslant T}\left(\int_0^l f^2(s,t)ds\right)^{1/2}$, $\max_{0\leqslant s\leqslant l}\left(\int_0^T f^2(s,t)dt\right)^{1/2}$ со-

ответственно. В силу полноты перечисленных пространств сумма ряда (19) при каждом фиксированном $u = u(s, t) \in L_2(Q)$ является непрерывной функцией на $\overline{Q}$, принадлежит $H^{2,1}(Q)$, почти всюду на Q удовлетворяет уравнению (2), почти всюду на $[0, T]$ следы $\frac{\partial x(\cdot, t)}{\partial s}$ при $s = 0$, $s = l$ и след $x(s, \cdot)$ при $t = 0$ удовлетворяют условиям (3), т. е. ряд (19) представляет собой почти классическое решение краевой задачи (2), (3); для него справедливы оценки (9), (10), (17), (18).

5. Покажем, что для решения разностной задачи (6), (7) справедливы следующие оценки

$$\max_{1 \leqslant j \leqslant N} \sum_{i=1}^{M-1} h y_{ij}^2 + \sum_{j=1}^{N} \sum_{i=1}^{M-1} h\tau y_{\bar{s}ij}^2 \leqslant C \sum_{j=1}^{N} \sum_{i=1}^{M-1} h\tau u_{ij}^2, \tag{20}$$

$$\max_{1 \leqslant j \leqslant N} \sum_{i=1}^{M-1} h y_{\bar{s}ij}^2 + \sum_{j=1}^{N} \sum_{i=1}^{M-1} h\tau (y_{\bar{t}ij}^2 + y_{\bar{s}sij}^2) \leqslant C \sum_{j=1}^{N} \sum_{i=1}^{M-1} h\tau u_{ij}^2, \tag{21}$$

являющиеся разностными аналогами энергетических оценок (9), (10) соответственно. Оценки (20), (21) доказываются по той же схеме, что и оценки (9), (10). Для доказательства (20) умножим уравнение (6) на $h\tau y_{ij}$ и просуммируем по i от 1 до $M-1$:

$$\sum_{i=1}^{M-1} \tau h y_{\bar{t}ij} y_{ij} - \sum_{i=1}^{M-1} \tau h y_{\bar{s}sij} y_{ij} = \sum_{i=1}^{M-1} \tau h u_{ij} y_{ij}, \quad j = 1, \ldots, N \tag{22}$$

Нетрудно проверить, что

$$\tau y_{\bar{t}ij} y_{ij} = (y_{ij} - y_{ij-1}) y_{ij} = \frac{1}{2}(y_{ij}^2 - y_{ij-1}^2) + \frac{1}{2}(y_{ij} - y_{ij-1})^2 \geqslant$$

$$\geqslant \frac{1}{2}(y_{ij}^2 - y_{ij-1}^2) \quad i = 1, \ldots, M-1, \quad j = 1, \ldots, N. \tag{23}$$

Поэтому

$$\sum_{i=1}^{M-1} \tau h y_{\bar{t}ij} y_{ij} \geqslant \frac{1}{2} \sum_{i=1}^{M-1} h y_{ij}^2 - \frac{1}{2} \sum_{i=1}^{M-1} h y_{ij-1}^2, \quad j = 1, \ldots, N. \tag{24}$$

Для преобразования второго слагаемого из левой части равенства (22) воспользуемся известной формулой суммирования по частям:

$$\sum_{i=1}^{M-1} h a_{si} b_i = -\sum_{i=1}^{M-1} h a_i b_{\bar{s}i} + a_M b_{M-1} - a_1 b_0, \tag{25}$$

где $a_{si} = \frac{a_{i+1} - a_i}{h}$, $b_{\bar{s}i} = \frac{b_i - b_{i-1}}{h}$. Отсюда при $a_i = y_{\bar{s}ij}$, $b_i = y_{ij}$ с учетом граничных условий (7) имеем

$$\sum_{i=1}^{M-1} \tau h y_{\bar{s}sij} y_{ij} = -\sum_{i=1}^{M-1} \tau h y_{\bar{s}ij}^2, \quad j = 1, \ldots, N. \tag{26}$$

Подставим (24), (26) в (22). Получим

$$\frac{1}{2} \sum_{i=1}^{M-1} h y_{ij}^2 - \frac{1}{2} \sum_{i=1}^{M-1} h y_{ij-1}^2 + \sum_{i=1}^{M-1} h\tau y_{\bar{s}ij}^2 \leqslant \sum_{i=1}^{M-1} h\tau u_{ij} y_{ij}, \quad j = 1, \ldots, N.$$

Просуммируем эти неравенства по j от 1 до некоторого n, $1 \leqslant n \leqslant N$. С учетом начального условия $y_{i0} = 0$, $i = 0, \ldots, M$, получим разностный аналог неравенств (12):

$$\frac{1}{2} \sum_{i=1}^{M-1} h y_{in}^2 + \sum_{j=1}^{n} \sum_{i=1}^{M-1} \tau h y_{\bar{s}ij}^2 \leqslant \sum_{j=1}^{n} \sum_{i=1}^{M-1} h\tau u_{ij} y_{ij} \leqslant$$

$$\leqslant \sum_{j=1}^{n} \tau \Big(\sum_{i=1}^{M-1} h u_{ij}^2 \Big)^{1/2} \max_{1 \leqslant j \leqslant n} \Big(\sum_{i=1}^{M-1} h y_{ij}^2 \Big)^{1/2} \leqslant$$

$$\leqslant \max_{1 \leqslant j \leqslant N} \Big(\sum_{i=1}^{M-1} h y_{ij}^2 \Big)^{1/2} \sqrt{T} \Big(\sum_{j=1}^{N} \sum_{i=1}^{M-1} h\tau u_{ij}^2 \Big)^{1/2} \quad \forall n, \quad 1 \leqslant n \leqslant N. \tag{27}$$

Тогда

$$\sum_{i=1}^{M-1} h y_{in}^2 \leqslant \max_{1 \leqslant j \leqslant N} \Big(\sum_{i=1}^{M-1} h y_{ij}^2 \Big)^{1/2} 2\sqrt{T} \Big(\sum_{j=1}^{N} \sum_{i=1}^{M-1} h\tau u_{ij}^2 \Big)^{1/2} \quad \forall n, \quad 1 \leqslant n \leqslant N.$$

Отсюда имеем

$$\max_{1 \leqslant n \leqslant N} \sum_{i=1}^{M-1} h y_{in}^2 = \Big(\max_{1 \leqslant j \leqslant N} \Big(\sum_{i=1}^{M-1} h y_{ij}^2 \Big)^{1/2} \Big)^2 \leqslant$$

$$\leqslant \max_{1 \leqslant j \leqslant N} \Big(\sum_{i=1}^{M-1} h y_{ij}^2 \Big)^{1/2} 2\sqrt{T} \Big(\sum_{j=1}^{N} \sum_{i=1}^{M-1} h\tau u_{ij}^2 \Big)^{1/2}$$

или

$$\max_{1 \leqslant j \leqslant N} \sum_{i=1}^{M-1} h y_{ij}^2 \leqslant 4T \sum_{j=1}^{N} \sum_{i=1}^{M-1} h\tau u_{ij}^2. \tag{28}$$

Кроме того, из (27), (28) получаем

$$\sum_{j=1}^{N} \sum_{i=1}^{M-1} h\tau y_{\bar{s}ij}^2 \leqslant 2T \sum_{j=1}^{N} \sum_{i=1}^{M-1} h\tau u_{ij}^2.$$

Сложив это неравенство с (28) придем к оценке (20) с $C = 6T$.

Докажем оценку (21). Умножим уравнение (6) на $h\tau y_{\bar{t}ij}$ и просуммируем по i от 1 до $M-1$:

$$\sum_{i=1}^{M-1} \tau h y_{\bar{t}ij}^2 - \sum_{i=1}^{M-1} \tau h y_{\bar{s}sij} y_{\bar{t}ij} = \sum_{i=1}^{M-1} h\tau u_{ij} y_{\bar{t}ij}, \quad j = 1, \ldots, N. \tag{29}$$

Второе слагаемое из левой части (29) преобразуем с помощью формулы (25) при $a_i = y_{\bar{s}ij}$, $b_i = y_{\bar{t}ij}$. С учетом граничных условий (7) имеем

$$-\sum_{i=1}^{M-1} \tau h y_{\bar{s}sij} y_{\bar{t}ij} = \sum_{i=1}^{M-1} \tau h y_{\bar{s}ij} y_{\bar{t}\bar{s}ij} - (y_{\bar{s}Mj} y_{\bar{t}M-1j} - y_{\bar{s}1j} y_{\bar{t}0j})\tau = \sum_{i=1}^{M-1} \tau h y_{\bar{s}ij} y_{s\bar{t}ij}.$$

Правую часть этого равенства оценим теперь снизу с помощью неравенства вида (23):

$$\tau y_{s\bar{t}ij} y_{\bar{s}ij} = (y_{\bar{s}ij} - y_{\bar{s}ij-1}) y_{\bar{s}ij} \geqslant \frac{1}{2}(y_{\bar{s}ij}^2 - y_{\bar{s}ij-1}^2).$$

Получим

$$-\sum_{i=1}^{M-1} \tau h y_{\bar{s}sij} y_{\bar{t}ij} \geqslant \frac{1}{2} \sum_{i=1}^{M-1} h y_{\bar{s}ij}^2 - \frac{1}{2} \sum_{i=1}^{M-1} h y_{\bar{s}ij-1}^2.$$

Подставим эту оценку в (29):

$$\sum_{i=1}^{M-1} \tau h y_{\bar{t}ij}^2 + \frac{1}{2}\sum_{i=1}^{M-1} h y_{\bar{s}ij}^2 - \frac{1}{2}\sum_{i=1}^{M-1} h y_{\bar{s}ij-1}^2 \leqslant \sum_{i=1}^{M-1} \tau h u_{ij} y_{\bar{t}ij}, \quad j = 1, \ldots, N.$$

Отсюда, суммируя по j от 1 до некоторого n, $1 \leqslant n \leqslant N$, с учетом равенств $y_{\bar{s}i0} = 0$, $i = 1, \ldots, M$, получим

$$\sum_{j=1}^{n}\sum_{i=1}^{M-1} \tau h y_{\bar{t}ij}^2 + \frac{1}{2}\sum_{i=1}^{M-1} h y_{\bar{s}in}^2 \leqslant \sum_{j=1}^{N}\sum_{i=1}^{M-1} \tau h u_{ij} y_{\bar{t}ij} \leqslant$$

$$\leqslant \frac{1}{2}\sum_{j=1}^{n}\sum_{i=1}^{M-1} h\tau y_{\bar{t}ij}^2 + \frac{1}{2}\sum_{j=1}^{N}\sum_{i=1}^{M-1} h\tau u_{ij}^2$$

или

$$\sum_{j=1}^{n}\sum_{i=1}^{M-1} \tau h y_{\bar{t}ij}^2 + \sum_{i=1}^{M-1} h y_{\bar{s}in}^2 \leqslant \sum_{j=1}^{N}\sum_{i=1}^{M-1} h\tau u_{ij}^2 \quad \forall n, \quad 1 \leqslant n \leqslant N. \tag{30}$$

Отсюда следует, что $\sum_{i=1}^{M-1} h y_{\bar{s}in}^2 \leqslant \sum_{j=1}^{N}\sum_{i=1}^{M-1} h\tau u_{ij}^2$ $\forall n$, $1 \leqslant n \leqslant N$, поэтому

$$\max_{1 \leqslant n \leqslant N} \sum_{i=1}^{M-1} h y_{\bar{s}in}^2 \leqslant \sum_{j=1}^{N}\sum_{i=1}^{M-1} \tau h u_{ij}^2. \tag{31}$$

Кроме того, из (30) имеем

$$\sum_{j=1}^{N}\sum_{i=1}^{M-1} h\tau y_{\bar{t}ij}^2 \leqslant \sum_{j=1}^{N}\sum_{i=1}^{M-1} h\tau u_{ij}^2. \tag{32}$$

Наконец, из уравнения (6) с учетом (32) получаем

$$\sum_{j=1}^{N}\sum_{i=1}^{M-1} h\tau y_{\bar{s}ij}^2 = \sum_{j=1}^{N}\sum_{i=1}^{M-1} h\tau (y_{\bar{t}ij} - u_{ij})^2 \leqslant$$

$$\leqslant 2\sum_{j=1}^{N}\sum_{i=1}^{M-1} h\tau y_{\bar{t}ij}^2 + 2\sum_{j=1}^{N}\sum_{i=1}^{M-1} h\tau u_{ij}^2 \leqslant 4\sum_{j=1}^{N}\sum_{i=1}^{M-1} h\tau u_{ij}^2. \tag{33}$$

Сложив неравенства (31)–(33), придем к оценке (21). Отметим, что справедливы и разностные аналоги оценок (17), (18), однако они нам ниже явно не понадобятся.

6. Оценим разность между решениями задач (2), (3) и (6), (7). Введем гильбертово пространство $L_{2h\tau}$, являющееся разностным аналогом пространства $L_2(Q)$. Элементами пространства $L_{2h\tau}$ являются сеточные функции $f_{h\tau} = \{f_{ij},\ i = 1, \ldots, M-1,\ j = 1, \ldots, N\}$, скалярное произведение и норма в $L_{2h\tau}$ равны

$$\langle f_{h\tau}, g_{h\tau}\rangle_{L_{2h\tau}} = \sum_{j=1}^{N}\sum_{i=1}^{M-1} h\tau f_{ij} g_{ij}, \quad \|f_{h\tau}\|_{L_{2h\tau}} = \Big(\sum_{j=1}^{N}\sum_{i=1}^{M-1} h\tau f_{ij}^2\Big)^{1/2}.$$

Через $p_{h\tau} f_{h\tau}$ будем обозначать кусочно-постоянное продолжение сеточной функции $f_{h\tau}$ по правилу

$$p_{h\tau} f_{h\tau} = (p_{h\tau} f_{h\tau})(s, t) = f_{ij},$$
$$(s,t) \in Q_{ij} = \{(s,t):\ s_i \leqslant s < s_{i+1},\ t_{j-1} < t \leqslant t_j\},\quad i = 1, \ldots, M-1,\quad j = 1, \ldots, N.$$

Область определения функции $p_{h\tau} f_{h\tau}$ обозначим $Q_h = \{(s, t):\ h \leqslant s < l,\ 0 < t \leqslant T\}$. Нетрудно видеть, что

$$\iint_{Q_h} p_{h\tau} f_{h\tau}\, ds dt = \sum_{j=1}^{N}\sum_{i=1}^{M-1} h\tau f_{ij},$$

$$\langle p_{h\tau} f_{h\tau}, p_{h\tau} g_{h\tau}\rangle_{L_2(Q_h)} = \sum_{j=1}^{N}\sum_{i=1}^{M-1} h\tau f_{ij} g_{ij} = \langle f_{h\tau}, g_{h\tau}\rangle_{L_{2h\tau}}, \quad \|p_{h\tau} f_{h\tau}\|_{L_2(Q_h)} = \|f_{h\tau}\|_{L_{2h\tau}}.$$

Разностное уравнение (6) теперь можем записать в виде

$$p_{h\tau} y_{\bar{t}h\tau} - p_{h\tau} y_{\bar{s}\bar{s}h\tau} = p_{h\tau} u_{h\tau}, \quad (s, t) \in Q_h. \tag{34}$$

Из уравнения (2) вычтем (34), полученное равенство умножим на $x(s, t) - p_{h\tau} y_{h\tau}$ и проинтегрируем по области Q_h:

$$\iint_{Q_h} \Big(\frac{\partial x(s,t)}{\partial t} - p_{h\tau} y_{\bar{t}h\tau}\Big)(x(s,t) - p_{h\tau} y_{h\tau})\, ds dt -$$

$$- \iint_{Q_h} \Big(\frac{\partial^2 x(s,t)}{\partial s^2} - p_{h\tau} y_{\bar{s}\bar{s}h\tau}\Big)(x(s,t) - p_{h\tau} y_{h\tau})\, ds dt =$$

$$= \iint_{Q_h} (u(s,t) - p_{h\tau} u_{h\tau})(x(s,t) - p_{h\tau} y_{h\tau})\, ds dt. \tag{35}$$

Первое слагаемое из левой части равенства (35) оценим тогда следующим образом:

$$\iint_{Q_h} \Big(\frac{\partial x(s,t)}{\partial t} - p_{h\tau} y_{\bar{t}h\tau}\Big)(x(s,t) - p_{h\tau} y_{h\tau})\, ds dt =$$

$$= \sum_{j=1}^{N}\sum_{i=1}^{M-1} \int_{t_{j-1}}^{t_j}\int_{s_i}^{s_{i+1}} \Big(\frac{\partial x(s,t)}{\partial t} - y_{\bar{t}ij}\Big)(x(s,t) - y_{ij})\, ds dt =$$

$$= \sum_{j=1}^{N}\sum_{i=1}^{M-1} \int_{t_{j-1}}^{t_j}\int_{s_i}^{s_{i+1}} \Big[\frac{1}{2}\frac{\partial}{\partial t}(x(s,t) - (t-t_j)y_{\bar{t}ij} - y_{ij})^2 + \Big(\frac{\partial x(s,t)}{\partial t} - y_{\bar{t}ij}\Big)(t-t_j)y_{\bar{t}ij}\Big]\, ds dt =$$

$$= \sum_{i=1}^{M-1} \int_{s_i}^{s_{i+1}} \frac{1}{2}\sum_{j=1}^{N}(x(s,t) - (t-t_j)y_{\bar{t}ij} - y_{ij})^2\Big|_{t=t_{j-1}}^{t_j} ds +$$

$$+ \sum_{j=1}^{N}\sum_{i=1}^{M-1} \int_{t_{j-1}}^{t_j}\int_{s_i}^{s_{i+1}} \Big[\frac{\partial x(s,t)}{\partial t}(t-t_j)y_{\bar{t}ij} - (t-t_j)y_{\bar{t}ij}^2\Big]\, ds dt =$$

$$= \sum_{i=1}^{M-1} \int_{s_i}^{s_{i+1}} \frac{1}{2}\sum_{j=1}^{N}[(x(s,t_j) - y_{ij})^2 - (x(s,t_{j-1}) - y_{ij-1})^2]\, ds +$$

$$+ \sum_{j=1}^{N}\sum_{i=1}^{M-1} \int_{s_i}^{s_{i+1}} [x(s,t)(t-t_j)y_{\bar{t}ij}\big|_{t=t_{j-1}}^{t_j} - \int_{t_{j-1}}^{t_j} x(s,t)y_{\bar{t}ij}\, dt - y_{\bar{t}ij}^2 \int_{t_{j-1}}^{t_j} (t-t_j)\, dt]\, ds =$$

$$= \frac{1}{2}\sum_{i=1}^{M-1} \int_{s_i}^{s_{i+1}} |x(s,t_N) - y_{iN}|^2\, ds + \sum_{j=1}^{N}\sum_{i=1}^{M-1} \int_{s_i}^{s_{i+1}} [\tau x(s,t_{j-1})y_{\bar{t}ij} -$$

$$- \int_{t_{j-1}}^{t_j} x(s,t)y_{\bar{t}ij}\, dt + \frac{\tau^2}{2}y_{\bar{t}ij}^2]\, ds \geqslant \frac{1}{2}\sum_{i=1}^{M-1} \int_{s_i}^{s_{i+1}} (x(s,t_N) - y_{iN})^2\, ds +$$

$$+ \sum_{j=1}^{N}\sum_{i=1}^{M-1} \int_{t_{j-1}}^{t_j}\int_{s_i}^{s_{i+1}} (x(s,t_{j-1}) - x(s,t))y_{\bar{t}ij}\, ds dt. \tag{36}$$

Преобразуем теперь второе слагаемое из левой части (35). Предварительно заметим, что

$$\sum_{i=1}^{M-1}\int_{s_i}^{s_{i+1}}\left(\frac{\partial^2 x(s,t)}{\partial s^2}-y_{\bar{s}sij}\right)(x(s,t)-y_{ij})ds=$$

$$=\sum_{i=1}^{M-1}\int_{s_i}^{s_{i+1}}\left[\left(\frac{\partial^2 x(s,t)}{\partial s^2}-x_{\bar{s}s}(s_i,t)\right)(x(s,t)-y_{ij})+(x_{\bar{s}s}(s_i,t)-y_{\bar{s}sij})(x(s,t)-x(s_i,t))+\right.$$

$$\left.+(x_{\bar{s}s}(s_i,t)-y_{\bar{s}sij})(x(s_i,t)-y_{ij})\right]ds\quad\forall t,\quad t_{j-1}<t\leqslant t_j,\quad j=1,\ldots,N,\qquad(37)$$

где использованы обозначения: $x_{\bar{s}s}(s_i,t)=\frac{1}{h}(x_s(s_i,t)-x_{\bar{s}}(s_i,t))$, $x_s(s_i,t)=\frac{1}{h}(x(s_{i+1},t)-x(s_i,t))$, $x_{\bar{s}}(s_i,t)=\frac{1}{h}(x(s_i,t)-x(s_{i-1},t))$. Первое слагаемое из правой части (37) можно представить в виде

$$\sum_{i=1}^{M-1}\int_{s_i}^{s_{i+1}}\frac{\partial}{\partial s}\left[\left(\frac{\partial x(s,t)}{\partial s}-(s-s_i)x_{\bar{s}s}(s_i,t)-x_{\bar{s}}(s_i,t)\right)(x(s,t)-y_{ij})\right]ds=$$

$$=\sum_{i=1}^{M-1}\left[\left(\frac{\partial x(s,t)}{\partial s}-(s-s_i)x_{\bar{s}s}(s_i,t)-x_{\bar{s}}(s_i,t)\right)(x(s,t)-y_{ij})\Big|_{s=s_i}^{s_{i+1}}-\right.$$

$$\left.-\int_{s_i}^{s_{i+1}}\left(\frac{\partial x(s,t)}{\partial s}-(s-s_i)x_{\bar{s}s}(s_i,t)-x_{\bar{s}}(s_i,t)\right)\frac{\partial x(s,t)}{\partial s}ds\right]=$$

$$=\sum_{i=1}^{M-1}\left[\left(\frac{\partial x(s_{i+1},t)}{\partial s}-x_{\bar{s}}(s_{i+1},t)\right)(x(s_{i+1},t)-y_{i+1j})-\right.$$

$$\left.-\left(\frac{\partial x(s_i,t)}{\partial s}-x_{\bar{s}}(s_i,t)\right)(x(s_i,t)-y_{ij})+\left(\frac{\partial x(s_{i+1},t)}{\partial s}-x_{\bar{s}}(s_{i+1},t)\right)(y_{i+1j}-y_{ij})\right]-$$

$$-\sum_{i=1}^{M-1}\int_{s_i}^{s_{i+1}}\left(\frac{\partial x(s,t)}{\partial s}-(s-s_i)x_{\bar{s}s}(s_i,t)-x_{\bar{s}}(s_i,t)\right)\frac{\partial x(s,t)}{\partial s}ds=$$

$$=\left(\frac{\partial x(s_M,t)}{\partial s}-x_{\bar{s}}(s_M,t)\right)(x(s_M,t)-y_{Mj})-\left(\frac{\partial x(s_1,t)}{\partial s}-x_{\bar{s}}(s_1,t)\right)(x(s_1,t)-y_{1j})+$$

$$+\sum_{i=1}^{M-1}\left(\frac{\partial x(s_{i+1},t)}{\partial s}-x_{\bar{s}}(s_{i+1},t)\right)hy_{sij}-\sum_{i=1}^{M-1}\int_{s_i}^{s_{i+1}}\left(\frac{\partial x(s,t)}{\partial s}-x_{\bar{s}}(s_i,t)\right)\frac{\partial x(s,t)}{\partial s}ds+$$

$$+\sum_{i=1}^{M-1}\int_{s_i}^{s_{i+1}}x_{\bar{s}s}(s_i,t)(s-s_i)\frac{\partial x(s,t)}{\partial s}ds\quad\forall t,\quad t_{j-1}<t\leqslant t_j,\quad j=1,\ldots,N.\qquad(38)$$

Третье слагаемое из правой части равенства (37) преобразуем с помощью формулы (25) при $a_i=x_{\bar{s}}(s_i,t)-y_{\bar{s}ij}$, $b_i=x(s_i,t)-y_{ij}$:

$$\sum_{i=1}^{M-1}\int_{s_i}^{s_{i+1}}(x_{\bar{s}s}(s_i,t)-y_{\bar{s}sij})(x(s_i,t)-y_{ij})ds=$$

$$=\sum_{i=1}^{M-1}h(x(s_i,t)-y_{ij})_{\bar{s}s}(x(s_i,t)-y_{ij})=$$

$$=-\sum_{i=1}^{M-1}h(x_{\bar{s}}(s_i,t)-y_{\bar{s}ij})^2+(x_{\bar{s}}(s_M,t)-y_{\bar{s}Mj})(x(s_{M-1},t)-y_{M-1j})-$$

$$-(x_{\bar{s}}(s_1,t)-y_{\bar{s}1j})(x(0,t)-y_{0j})\quad\forall t,\quad t_{j-1}<t\leqslant t_j,\quad j=1,\ldots,N.\qquad(39)$$

Неравенство (36) и равенство (37) с учетом (38), (39) подставим в левую часть (35). Получим:

$$\frac{1}{2}\sum_{i=1}^{M-1}\int_{s_i}^{s_{i+1}}(x(s,t_N)-y_{iN})^2ds+\sum_{j=1}^{N}\sum_{i=1}^{M-1}h\int_{t_{j-1}}^{t_j}(x_{\bar{s}}(s_i,t)-y_{\bar{s}ij})^2dt\leqslant\sum_{i=1}^{10}R_i,\qquad(40)$$

где

$$R_1=\sum_{j=1}^{N}\sum_{i=1}^{M-1}\int_{t_{j-1}}^{t_j}\int_{s_i}^{s_{i+1}}(x(s,t)-x(s,t_{j-1}))y_{\bar{t}ij}dsdt,$$

$$R_2=\sum_{j=1}^{N}\int_{t_{j-1}}^{t_j}\left(\frac{\partial x(s_M,t)}{\partial s}-x_{\bar{s}}(s_M,t)\right)(x(s_M,t)-y_{Mj})dt,$$

$$R_3=-\sum_{j=1}^{N}\int_{t_{j-1}}^{t_j}\left(\frac{\partial x(s_1,t)}{\partial s}-x_{\bar{s}}(s_1,t)\right)(x(s_1,t)-y_{1j})dt,$$

$$R_4=\sum_{j=1}^{N}\sum_{i=1}^{M-1}\int_{t_{j-1}}^{t_j}\left(\frac{\partial x(s_{i+1},t)}{\partial s}-x_{\bar{s}}(s_{i+1},t)\right)hy_{sij}dt,$$

$$R_5=-\sum_{j=1}^{N}\sum_{i=1}^{M-1}\int_{t_{j-1}}^{t_j}\int_{s_i}^{s_{i+1}}\left(\frac{\partial x(s,t)}{\partial s}-x_{\bar{s}}(s_i,t)\right)\frac{\partial x(s,t)}{\partial s}dsdt,$$

$$R_6=\sum_{j=1}^{N}\sum_{i=1}^{M-1}\int_{t_{j-1}}^{t_j}\int_{s_i}^{s_{i+1}}x_{\bar{s}s}(s_i,t)(s-s_i)\frac{\partial x(s,t)}{\partial s}dsdt,$$

$$R_7=\sum_{j=1}^{N}\sum_{i=1}^{M-1}\int_{t_{j-1}}^{t_j}\int_{s_i}^{s_{i+1}}(x_{\bar{s}s}(s_i,t)-y_{\bar{s}sij})(x(s,t)-x(s_i,t))dsdt,$$

$$R_8=\sum_{j=1}^{N}\int_{t_{j-1}}^{t_j}(x_{\bar{s}}(s_M,t)-y_{\bar{s}Mj})(x(s_{M-1},t)-y_{M-1j})dt,$$

$$R_9=-\sum_{j=1}^{N}\int_{t_{j-1}}^{t_j}(x_{\bar{s}}(s_1,t)-y_{\bar{s}1j})(x(0,t)-y_{0j})dt,$$

$$R_{10}=\sum_{j=1}^{N}\sum_{i=1}^{M-1}\int_{t_{j-1}}^{t_j}\int_{s_i}^{s_{i+1}}(u(s,t)-u_{ij})(x(s,t)-y_{ij})dsdt.$$

Оценим сверху каждую из величин $|R_i|$, $i=1,\ldots,10$. Для этого наряду с оценками (9), (10), (20), (21) нам понадобятся еще несколько вспомогательных неравенств. Приведем их. Начнем с неравенства:

$$(x(s_i,t)-y_{ij})^2\leqslant C\left(\sum_{m=1}^{M-1}h[(x_{\bar{s}}(s_m,t)-y_{\bar{s}mj})^2+(x(s_m,t)-y_{mj})^2]\right)\qquad(41)$$

$$\forall i=1,\ldots,M-1,\quad j=1,\ldots,N,\quad t\in[0,T].$$

Напоминаем, что через C, как и в оценках (9), (10), (20), (21), мы здесь и далее обозначаем постоянные, которые не зависят от h, τ, u, $u_{h\tau}$. Будем считать, что $h<1$, $h<l/2$, $\tau<1$. Нетрудно проверить, что

$$x(s_i,t)=\sum_{m=p+1}^{i}hx_{\bar{s}}(s_m,t)+x(s_p,t)\quad\forall t\in[0,T],$$

$$y_{ij}=\sum_{m=p+1}^{i}hy_{\bar{s}mj}+y_{pj}\quad\forall j=1,\ldots,N,$$

где $1 \leqslant p \leqslant i \leqslant M-1$; при $i=p$ по определению считаем $\sum\limits_{m=p+1}^{p} a_m = 0$. Тогда

$$(x(s_i, t) - y_{ij})^2 = \Big(\sum_{m=p+1}^{i} h(x_{\bar{s}}(s_m, t) - y_{\bar{s}mj}) + (x(s_p, t) - y_{pj}) \Big)^2 \leqslant$$
$$\leqslant 2 \sum_{m=1}^{M-1} h(x_{\bar{s}}(s_m, t) - y_{\bar{s}mj})^2 + 2(x(s_p, t) - y_{pj})^2.$$

Умножим это неравенство на h и просуммируем его по p от 1 до $M-1$. Получим

$$(l-h)(x(s_i, t) - y_{ij})^2 \leqslant 2(l-h) \sum_{m=1}^{M-1} h(x_{\bar{s}}(s_m, t) - y_{\bar{s}mj})^2 +$$
$$+ 2 \sum_{p=1}^{M-1} h(x(s_p, t) - y_{pj}) \quad \forall i = 1, \ldots, M-1, \quad j = 1, \ldots, N, \quad t \in [0, T],$$

что равносильно неравенству (41) с $C = \max\left\{2; \frac{4}{l}\right\}$. Для $i=0$, $i=M$ с учетом граничных условий (7) имеем:

$$(x(0, t) - y_{0j})^2 = ((x(0, t) - x(s_1, t)) + (x(s_1, t) - y_{1j}))^2 \leqslant$$
$$\leqslant 2h \int_0^{s_1} \left| \frac{\partial x(\xi, t)}{\partial s} \right|^2 d\xi + 2(x(s_1, t) - y_{1j})^2,$$
$$(x(s_M, t) - y_{Mj})^2 = ((x(s_M, t) - x(s_{M-1}, t)) + (x(s_{M-1}, t) - y_{M-1\,j}))^2 \leqslant$$
$$\leqslant 2h \int_{s_{M-1}}^{s_M} \left| \frac{\partial x(\xi, t)}{\partial s} \right|^2 d\xi + 2(x(s_{M-1}, t) - y_{M-1\,j})^2. \quad (42)$$

Заметим также, что

$$\sum_{j=1}^{N} \int_{t_{j-1}}^{t_j} \sum_{m=1}^{M-1} h(x(s_m, t) - y_{mj})^2 dt = \sum_{j=1}^{N} \sum_{m=1}^{M-1} \int_{t_{j-1}}^{t_j} \int_{s_m}^{s_{m+1}} [(x(s_m, t) - x(s, t)) +$$
$$+ (x(s, t) - x(s, t_j)) + (x(s, t_j) - y_{mj})]^2 ds dt \leqslant$$
$$\leqslant 3 \sum_{j=1}^{N} \sum_{m=1}^{M-1} \int_{t_{j-1}}^{t_j} \int_{s_m}^{s_{m+1}} \left[\left(\int_{s_m}^{s} \frac{\partial x(\xi, t)}{\partial s} d\xi \right)^2 + \left(\int_{t}^{t_j} \frac{\partial x(s, \tau)}{\partial t} d\tau \right)^2 + (x(s, t_j) - y_{mj})^2 \right] ds dt \leqslant$$
$$\leqslant 3 \sum_{j=1}^{N} \sum_{m=1}^{M-1} \int_{t_{j-1}}^{t_j} \int_{s_m}^{s_{m+1}} \left[h \int_{s_m}^{s_{m+1}} \left| \frac{\partial x(\xi, t)}{\partial s} \right|^2 d\xi + \tau \int_{t_{j-1}}^{t_j} \left| \frac{\partial x(s, \tau)}{\partial t} \right|^2 d\tau + (x(s, t_j) - y_{mj})^2 \right] ds dt \leqslant$$
$$\leqslant 3h^2 \left\| \frac{\partial x}{\partial s} \right\|^2_{L_2(Q)} + 3\tau^2 \left\| \frac{\partial x}{\partial t} \right\|^2_{L_2(Q)} + 3 \sum_{j=1}^{N} \sum_{m=1}^{M-1} \tau \int_{s_m}^{s_{m+1}} (x(s, t_j) - y_{mj})^2 ds \leqslant$$
$$\leqslant C(h^2 + \tau^2) \|u\|^2_{L_2(Q)} + 3 \sum_{j=1}^{N} \sum_{m=1}^{M-1} \tau \int_{s_m}^{s_{m+1}} (x(s, t_j) - y_{mj})^2 ds \quad (43)$$

в силу (9), (10). Из (41)–(43) следует, что

$$\sum_{j=1}^{N} \int_{t_{j-1}}^{t_j} (x(s_i, t) - y_{ij})^2 dt \leqslant C \sum_{j=1}^{N} \sum_{m=1}^{M-1} \Big[h \int_{t_{j-1}}^{t_j} (x_{\bar{s}}(s_m, t) - y_{\bar{s}mj})^2 dt +$$
$$+ \tau \int_{s_m}^{s_{m+1}} (x(s, t_j) - y_{mj})^2 ds \Big] + C(h + \tau) \|u\|^2_{L_2(Q)} \quad i = 0, \ldots, M. \quad (44)$$

Далее заметим, что

$$\frac{\partial x(s, t)}{\partial s} - x_{\bar{s}}(s_i, t) = \frac{1}{h} \int_{s_{i-1}}^{s_i} \left(\frac{\partial x(s, t)}{\partial s} - \frac{\partial x(\xi, t)}{\partial s} \right) d\xi = \frac{1}{h} \int_{s_{i-1}}^{s_i} \left(\int_{\xi}^{s} \frac{\partial^2 x(\eta, t)}{\partial s^2} d\eta \right) d\xi$$
$$\forall s \in [0, l], \quad t \in [0, T], \quad i = 1, \ldots, M.$$

Отсюда для всех s, $s_{i-1} \leqslant s \leqslant s_{i+1}$, $i = 1, \ldots, M-1$ имеем

$$\sum_{j=1}^{N} \int_{t_{j-1}}^{t_j} \left(\frac{\partial x(s, t)}{\partial s} - x_{\bar{s}}(s_i, t) \right)^2 dt \leqslant \sum_{j=1}^{N} \int_{t_{j-1}}^{t_j} \left(\frac{1}{h} \int_{s_{i-1}}^{s_i} \left(\int_{s_{i-1}}^{s_{i+1}} \left| \frac{\partial^2 x(\eta, t)}{\partial s^2} \right| d\eta \right) d\xi \right)^2 dt \leqslant$$
$$\leqslant h \sum_{j=1}^{N} \int_{t_{j-1}}^{t_j} \int_{s_{i-1}}^{s_{i+1}} \left| \frac{\partial^2 x(\eta, t)}{\partial s^2} \right|^2 d\eta. \quad (45)$$

Из (45) и оценки (10) получаем

$$\sum_{j=1}^{N} \int_{t_{j-1}}^{t_j} \left(\frac{\partial x(s, t)}{\partial s} - x_{\bar{s}}(s_i, t) \right)^2 dt \leqslant 2h \sum_{j=1}^{N} \int_{t_{j-1}}^{t_j} \int_{0}^{l} \left| \frac{\partial^2 x(\eta, t)}{\partial s^2} \right|^2 d\eta =$$
$$= 2h \left\| \frac{\partial^2 x}{\partial s^2} \right\|^2_{L_2(Q)} \leqslant hC \|u\|^2_{L_2(Q)} \quad \forall s, \ s_{i-1} \leqslant s \leqslant s_{i+1}, \ i = 1, \ldots, M-1. \quad (46)$$

Кроме того, с учетом (45), (10) имеем

$$\sum_{j=1}^{N} \sum_{i=1}^{M-1} \int_{t_{j-1}}^{t_j} \int_{s_i}^{s_{i+1}} \left(\frac{\partial x(s, t)}{\partial s} - x_{\bar{s}}(s_i, t) \right)^2 ds dt \leqslant$$
$$\leqslant \sum_{j=1}^{N} \sum_{i=1}^{M-1} \int_{t_{j-1}}^{t_j} \int_{s_i}^{s_{i+1}} \left(h \int_{s_{i-1}}^{s_{i+1}} \left| \frac{\partial^2 x(\eta, t)}{\partial s^2} \right|^2 d\eta \right) ds dt \leqslant$$
$$\leqslant 2h^2 \left\| \frac{\partial^2 x}{\partial s^2} \right\|^2_{L_2(Q)} \leqslant h^2 C \|u\|^2_{L_2(Q)}. \quad (47)$$

Наконец, из равенства

$$x_{\bar{s}\bar{s}}(s_i, t) = \frac{1}{h^2} \left(\int_{s_i}^{s_{i+1}} \frac{\partial x(\xi, t)}{\partial s} d\xi - \int_{s_{i-1}}^{s_i} \frac{\partial x(\eta, t)}{\partial s} d\eta \right) = \frac{1}{h^2} \left(\int_{s_i}^{s_{i+1}} \frac{\partial x(\xi, t)}{\partial s} d\xi - \right.$$
$$\left. - \int_{s_i}^{s_{i+1}} \frac{\partial x(\xi - h, t)}{\partial s} d\xi \right) = \frac{1}{h^2} \int_{s_i}^{s_{i+1}} \left(\int_{\xi - h}^{\xi} \frac{\partial^2 x(\eta, t)}{\partial s^2} d\eta \right) d\xi, \quad i = 1, \ldots, M-1, \quad t \in [0, T]$$

и оценки (10) следует

$$\sum_{j=1}^{N} \sum_{i=1}^{M-1} \int_{t_{j-1}}^{t_j} \int_{s_i}^{s_{i+1}} (x_{\bar{s}\bar{s}}(s_i, t))^2 ds dt = \sum_{j=1}^{N} \sum_{i=1}^{M-1} \int_{t_{j-1}}^{t_j} \frac{1}{h^3} \left(\int_{s_i}^{s_{i+1}} \left(\int_{\xi - h}^{\xi} \frac{\partial^2 x(\eta, t)}{\partial s^2} d\eta \right) d\xi \right)^2 \leqslant$$
$$\leqslant \sum_{j=1}^{N} \sum_{i=1}^{M-1} \int_{t_{j-1}}^{t_j} \frac{1}{h^2} \int_{s_i}^{s_{i+1}} \left(\int_{\xi - h}^{\xi} \frac{\partial^2 x(\eta, t)}{\partial s^2} d\eta \right)^2 d\xi dt \leqslant$$
$$\leqslant \sum_{j=1}^{N} \sum_{i=1}^{M-1} \int_{t_{j-1}}^{t_j} \frac{1}{h^2} \int_{s_i}^{s_{i+1}} h \int_{s_{i-1}}^{s_{i+1}} \left| \frac{\partial^2 x(\eta, t)}{\partial s^2} \right|^2 d\eta d\xi dt =$$
$$= \sum_{j=1}^{N} \sum_{i=1}^{M-1} \int_{t_{j-1}}^{t_j} \int_{s_{i-1}}^{s_{i+1}} \left| \frac{\partial^2 x(\eta, t)}{\partial s^2} \right|^2 d\eta dt \leqslant 2 \left\| \frac{\partial^2 x}{\partial s^2} \right\|^2_{L_2(Q)} \leqslant C \|u\|^2_{L_2(Q)}. \quad (48)$$

Теперь можем перейти к оценкам величин R_i, $i=1,\dots,10$, из (40). Начнем с оценки R_1:

$$|R_1| = \Big|\sum_{j=1}^{N}\sum_{i=1}^{M-1}\int_{t_{j-1}}^{t_j}\int_{s_i}^{s_{i+1}}\Big(\int_{t_{j-1}}^{t}\frac{\partial x(s,\tau)}{\partial t}d\tau\Big)y_{\bar{t}ij}\,dsdt\Big| \leqslant$$

$$\leqslant \sum_{j=1}^{N}\sum_{i=1}^{M-1}\int_{t_{j-1}}^{t_j}\int_{s_i}^{s_{i+1}}\Big(\int_{t_{j-1}}^{t_j}\Big|\frac{\partial x(s,\tau)}{\partial t}\Big|d\tau|y_{\bar{t}ij}|\Big)dsdt \leqslant$$

$$\leqslant \frac{1}{2}\tau\Big(\Big\|\frac{\partial x}{\partial t}\Big\|^2_{L_2(Q)} + \|y_{\bar{t}h\tau}\|^2_{L_{2h\tau}}\Big) \leqslant \tau C(\|u\|^2_{L_2(Q)} + \|u_{h\tau}\|^2_{L_{2h\tau}}); \qquad (49)$$

здесь мы учли оценки (10), (21). Далее, с помощью оценок (44) при $i=M$, (46) при $i=M$, $s=s_M$ имеем

$$|R_2| \leqslant \Big(\sum_{j=1}^{N}\int_{t_{j-1}}^{t_j}(x(s_M,t)-y_{Mj})^2dt\Big)^{1/2}\Big(\sum_{j=1}^{N}\int_{t_{j-1}}^{t_j}\Big(\frac{\partial x(s_M,t)}{\partial s}-x_{\bar{s}}(s_M,t)\Big)^2dt\Big)^{1/2} \leqslant$$

$$\leqslant \frac{\varepsilon}{2}C\Big[\sum_{j=1}^{N}\sum_{m=1}^{M-1}\int_{t_{j-1}}^{t_j}\int_{s_m}^{s_{m+1}}((x_{\bar{s}}(s_m,t)-y_{\bar{s}mj})^2+(x(s,t_j)-y_{mj})^2)dsdt+$$

$$+(\tau+h)\|u\|^2_{L_2(Q)}\Big] + \frac{h}{2\varepsilon}C\|u\|^2_{L_2(Q)} \quad \forall\varepsilon>0. \qquad (50)$$

Аналогично с помощью тех же оценок (44) при $i=1$, (46) при $i=1$, $s=s_1$ получаем

$$|R_3| \leqslant \frac{\varepsilon}{2}C\Big[\sum_{j=1}^{N}\sum_{m=1}^{M-1}\int_{t_{j-1}}^{t_j}\int_{s_m}^{s_{m+1}}((x_{\bar{s}}(s_m,t)-y_{\bar{s}mj})^2+$$

$$+(x(s,t_j)-y_{mj})^2)dsdt+(\tau+h)\|u\|^2_{L_2(Q)}\Big] + \frac{h}{2\varepsilon}C\|u\|^2_{L_2(Q)} \quad \forall\varepsilon>0. \qquad (51)$$

Из оценок (10), (20), (21), (47) следует

$$|R_4| \leqslant \Big(\sum_{i=1}^{M-1}\int_{s_i}^{s_{i+1}}\Big(\sum_{j=1}^{N}\int_{t_{j-1}}^{t_j}\Big(\Big(\frac{\partial x(s_{i+1},t)}{\partial s}-\frac{\partial x(s,t)}{\partial s}\Big)+$$

$$+\Big(\frac{\partial x(s,t)}{\partial s}-x_{\bar{s}}(s_i,t)\Big)-hx_{\bar{s}\bar{s}}(s_i,t)\Big)^2dt\Big)ds\Big)^{1/2}\|y_{sh\tau}\|_{L_{2h\tau}} \leqslant$$

$$\leqslant \Big(3\sum_{j=1}^{N}\sum_{i=1}^{M-1}\int_{t_{j-1}}^{t_j}\int_{s_i}^{s_{i+1}}\Big(\Big(\int_{s}^{s_{i+1}}\frac{\partial^2x(\xi,t)}{\partial s^2}d\xi\Big)^2+$$

$$+\Big(\frac{\partial x(s,t)}{\partial s}-x_{\bar{s}}(s_i,t)\Big)^2+h^2(x_{\bar{s}\bar{s}}(s_i,t))^2\Big)dsdt\Big)^{1/2}\|y_{\bar{s}h\tau}\|_{L_{2h\tau}} \leqslant$$

$$\leqslant hC\|u\|_{L_2(Q)}\|u_{h\tau}\|_{L_{2h\tau}} \leqslant hC(\|u\|^2_{L_2(Q)}+\|u_{h\tau}\|^2_{L_{2h\tau}}). \qquad (52)$$

Из (9), (47) имеем

$$|R_5| \leqslant \Big(\sum_{j=1}^{N}\sum_{i=1}^{M-1}\int_{t_{j-1}}^{t_j}\int_{s_i}^{s_{i+1}}\Big(\frac{\partial x(s,t)}{\partial s}-x_{\bar{s}}(s_i,t)\Big)^2dsdt\Big)^{1/2}\Big\|\frac{\partial x}{\partial s}\Big\|_{L_2(Q)} \leqslant hC\|u\|^2_{L_2(Q)}. \qquad (53)$$

Далее, с учетом оценок (9), (10), (48) получаем

$$|R_6| \leqslant \Big(\sum_{j=1}^{N}\sum_{i=1}^{M-1}\int_{t_{j-1}}^{t_j}\int_{s_i}^{s_{i+1}}|x_{\bar{s}\bar{s}}(s_i,t)|^2(s-s_i)^2dsdt\Big)^{1/2}\Big\|\frac{\partial x}{\partial s}\Big\|^2_{L_2(Q)} =$$

$$= \Big(\sum_{j=1}^{N}\sum_{i=1}^{M-1}\int_{t_{j-1}}^{t_j}|x_{\bar{s}\bar{s}}(s_i,t)|^2\frac{h^3}{3}dt\Big)^{1/2}\Big\|\frac{\partial x}{\partial s}\Big\|_{L_2(Q)} \leqslant$$

$$\leqslant 2h\Big\|\frac{\partial^2x}{\partial s^2}\Big\|_{L_2(Q)}\cdot\Big\|\frac{\partial x}{\partial s}\Big\|_{L_2(Q)} \leqslant hC\|u\|^2_{L_2(Q)}. \qquad (54)$$

Из (9), (10), (21), (48) следует

$$|R_7| \leqslant \Big(\sum_{j=1}^{N}\sum_{i=1}^{M-1}\int_{t_{j-1}}^{t_j}\int_{s_i}^{s_{i+1}}(|x_{\bar{s}\bar{s}}(s_i,t)|+|y_{\bar{s}\bar{s}ij}|)^2dsdt\Big)^{1/2}\times$$

$$\times\Big(\sum_{j=1}^{N}\sum_{i=1}^{M-1}\int_{t_{j-1}}^{t_j}\int_{s_i}^{s_{i+1}}\Big(\int_{s_i}^{s}\frac{\partial x(\xi,t)}{\partial s}d\xi\Big)^2dsdt\Big)^{1/2} \leqslant$$

$$\leqslant \Big(2\Big\|\frac{\partial^2x}{\partial s^2}\Big\|_{L_2(Q)}+\|y_{\bar{s}\bar{s}h\tau}\|_{L_{2h\tau}}\Big)\Big(\sum_{j=1}^{N}\sum_{i=1}^{M-1}\int_{t_{j-1}}^{t_j}\int_{s_i}^{s_{i+1}}(s-s_i)\int_{s_i}^{s_{i+1}}\Big|\frac{\partial x(\xi,t)}{\partial s}\Big|^2d\xi dsdt\Big)^{1/2} \leqslant$$

$$\leqslant C(\|u\|_{L_2(Q)}+\|u_{h\tau}\|_{L_{2h\tau}})h\Big\|\frac{\partial x}{\partial s}\Big\|_{L_2(Q)} \leqslant hC(\|u\|^2_{L_2(Q)}+\|u_{h\tau}\|^2_{L_{2h\tau}}). \qquad (55)$$

При оценке $|R_8|$ воспользуемся тем, что $x_{\bar{s}}(s_M,t)-y_{\bar{s}Mj}=x_{\bar{s}}(s_M,t)-\frac{\partial x(s_M,t)}{\partial s}$ в силу граничных условий $\frac{\partial x(s_M,t)}{\partial s}=0$, $y_{\bar{s}Mj}=0$, неравенствами (44) при $i=M-1$, оценками (46) при $i=M$, $s=s_M$:

$$|R_8| \leqslant \Big(\sum_{j=1}^{N}\int_{t_{j-1}}^{t_j}(x(s_{M-1},t)-y_{M-1j})^2dt\Big)^{1/2}\Big(\sum_{j=1}^{N}\int_{t_{j-1}}^{t_j}\Big(\frac{\partial x(s_M,t)}{\partial s}-x_{\bar{s}}(s_M,t)\Big)^2dt\Big)^{1/2} \leqslant$$

$$\leqslant \frac{\varepsilon}{2}C\Big[\sum_{j=1}^{N}\sum_{m=1}^{M-1}\int_{t_{j-1}}^{t_j}\int_{s_m}^{s_{m+1}}((x_{\bar{s}}(s_m,t)-y_{\bar{s}mj})^2+(x(s,t_j)-y_{mj})^2)dsdt+$$

$$+(\tau+h)\|u\|^2_{L_2(Q)}\Big] + \frac{h}{2\varepsilon}C\|u\|^2_{L_2(Q)} \quad \forall\varepsilon>0. \qquad (56)$$

Аналогично с учетом равенства $x_{\bar{s}}(s_1,t)-y_{\bar{s}1j}=x_{\bar{s}}(s_1,t)-\frac{\partial x(0,t)}{\partial s}$, вытекающего из граничных условий (3), (7), с помощью тех же оценок (44) при $i=0$, неравенств (46) при $i=1$, $s=0$ получаем

$$|R_9| \leqslant \frac{\varepsilon}{2}C\Big[\sum_{j=1}^{N}\sum_{m=1}^{M-1}\int_{t_{j-1}}^{t_j}\int_{s_m}^{s_{m+1}}((x_{\bar{s}}(s_m,t)-y_{\bar{s}mj})^2+(x(s,t_j)-y_{mj})^2)dsdt+$$

$$+(\tau+h)\|u\|^2_{L_2(Q)}\Big] + \frac{h}{2\varepsilon}C\|u\|^2_{L_2(Q)} \quad \forall\varepsilon>0. \qquad (57)$$

Наконец, для $|R_{10}|$ с учетом оценки (10) имеем

$$|R_{10}| \leqslant \frac{1}{2}\sum_{j=1}^{N}\sum_{i=1}^{M-1}\int_{t_{j-1}}^{t_j}\int_{s_i}^{s_{i+1}}(u(s,t)-u_{ij})^2dsdt+\frac{1}{2}\sum_{j=1}^{N}\sum_{i=1}^{M-1}\int_{t_{j-1}}^{t_j}\int_{s_i}^{s_{i+1}}((x(s,t)-x(s,t_j))+$$

$$+(x(s,t_j)-y_{ij}))^2dsdt \leqslant \frac{1}{2}\|u-p_{h\tau}u_{h\tau}\|^2_{L_2(Q_h)}+\sum_{j=1}^{N}\sum_{i=1}^{M-1}\int_{t_{j-1}}^{t_j}\int_{s_i}^{s_{i+1}}\Big(\Big(\int_{t}^{t_j}\frac{\partial x(s,\tau)}{\partial t}d\tau\Big)^2+$$

$$+(x(s,t_j)-y_{ij})^2\Big)\,dsdt \leqslant \frac{1}{2}\|u-p_{h\tau}u_{h\tau}\|^2_{L_2(Q_h)}+$$
$$+\tau^2C\|u\|^2_{L_2(Q)}+\sum_{j=1}^{N}\tau\sum_{i=1}^{M-1}\int_{s_i}^{s_{i+1}}(x(s,t_j)-y_{ij})^2ds. \quad (58)$$

Подставим оценки (49)–(58) в (40). Получим

$$\frac{1}{2}\sum_{i=1}^{M-1}\int_{s_i}^{s_{i+1}}(x(s,t_N-y_{iN}))^2ds+\sum_{j=1}^{N}\sum_{i=1}^{M-1}h\int_{t_{j-1}}^{t_j}(x_{\bar s}(s_i,t)-y_{\bar s ij})^2dt\leqslant$$
$$\leqslant\varepsilon C\sum_{j=1}^{N}\sum_{i=1}^{M-1}\int_{t_{j-1}}^{t_j}h(x_{\bar s}(s_i,t)-y_{\bar s ij})^2dt+\|u-p_{h\tau}u_{h\tau}\|^2_{L_2(Q_h)}+$$
$$+\tau(1+\varepsilon C)\sum_{j=1}^{N}\Big(\sum_{i=1}^{M-1}\int_{s_i}^{s_{i+1}}(x(s,t_j)-y_{ij})^2ds\Big)+$$
$$+C(\tau+h)\Big(1+\varepsilon+\frac{1}{\varepsilon}\Big)(\|u\|^2_{L_2(Q)}+\|u_{h\tau}\|^2_{L_{2h\tau}})$$

или

$$\Big(\frac{1}{2}-\tau(1+\varepsilon C)\Big)\sum_{i=1}^{M-1}\int_{s_i}^{s_{i+1}}(x(s,t_N)-y_{iN})^2ds+$$
$$+(1-\varepsilon C)\sum_{j=1}^{N}\sum_{i=1}^{M-1}h\int_{t_{j-1}}^{t_j}(x_{\bar s}(s_i,t)-y_{\bar s ij})^2dt\leqslant$$
$$\leqslant\tau(1+\varepsilon C)\sum_{j=1}^{N-1}\Big(\sum_{i=1}^{M-1}\int_{s_i}^{s_{i+1}}(x(s,t_j)-y_{ij})^2ds\Big)+$$
$$+C(\tau+h)\Big(1+\varepsilon+\frac{1}{\varepsilon}\Big)(\|u\|^2_{L_2(Q)}+\|u_{h\tau}\|^2_{L_{2h\tau}})+\|u-p_{h\tau}u_{h\tau}\|^2_{L_2(Q_h)}\quad\forall\varepsilon>0. \quad (59)$$

Зафиксируем $\varepsilon>0$ столь малым, чтобы $1-\varepsilon C>0$. Кроме того, будем считать τ столь малым, что $\frac{1}{2}-\tau(1+\varepsilon C)\geqslant\frac{1}{4}$. Тогда из (59) имеем

$$\sum_{i=1}^{M-1}\int_{s_i}^{s_{i+1}}(x(s,t_N)-y_{iN})^2ds\leqslant 4\tau(1+\varepsilon C)\sum_{j=1}^{N-1}\Big(\sum_{i=1}^{M-1}\int_{s_i}^{s_{i+1}}(x(s,t_j)-y_{ij})^2ds\Big)+$$
$$+C(\tau+h)\Big(1+\varepsilon+\frac{1}{\varepsilon}\Big)(\|u\|^2_{L_2(Q)}+\|u_{h\tau}\|^2_{L_{2h\tau}})+\|u-p_{h\tau}u_{h\tau}\|^2_{L_2(Q_h)}.$$

Отсюда, пользуясь леммой 8.6.1, получаем

$$\sum_{i=1}^{M-1}\int_{s_i}^{s_{i+1}}(x(s,t_N)-y_{iN})^2ds\leqslant$$
$$\leqslant(1+\tau C_1)^N(\|u-p_{h\tau}u_{h\tau}\|^2_{L_2(Q_h)}+(\tau+h)C_2(\|u\|^2_{L_2(Q)}+\|u_{h\tau}\|^2_{L_{2h\tau}})),$$

где $C_1=4(1+\varepsilon C)$, $C_2=C\Big(1+\varepsilon+\frac{1}{\varepsilon}\Big)$. Поскольку $(1+\tau C_1)\leqslant e^{\tau C_1}$, $(1+\tau C_1)^N\leqslant e^{TC_1}$, то

$$\sum_{i=1}^{M-1}\int_{s_i}^{s_{i+1}}(x(s,t_N)-y_{iN})^2ds\leqslant \quad (60)$$
$$\leqslant C(\|u-p_{h\tau}u_{h\tau}\|^2_{L_2(Q_h)}+(\tau+h)(\|u\|^2_{L_2(Q)}+\|u_{h\tau}\|^2_{L_{2h\tau}})).$$

Отметим, что из (59), (60) нетрудно вывести оценку [362]:

$$\sum_{j=1}^{N}\sum_{i=1}^{M-1}\int_{t_{j-1}}^{t_j}\int_{s_i}^{s_{i+1}}\Big(\frac{\partial x(s,t)}{\partial s}-y_{\bar s ij}\Big)^2dsdt\leqslant$$
$$\leqslant C(\|u-p_{h\tau}u_{h\tau}\|^2_{L_2(Q_h)}+(\tau+h)(\|u\|^2_{L_2(Q)}+\|u_{h\tau}\|^2_{L_{2h\tau}})).$$

7. Пользуясь оценкой (60), докажем, что задача (5)–(8) аппроксимирует задачу (1)–(4) по функции.

Теорема 1. *Пусть в задаче* (5)–(8) $b_i=\frac{1}{h}\int_{s_i}^{s_{i+1}}b(\xi)d\xi$, $i=1,\ldots,M-1$. *Тогда* $\lim\limits_{(h,\tau)\to 0}I_{h\tau *}=J_*$.

Доказательство. Оценим разность $J(u)-I_{h\tau}(u_{h\tau})$, считая, что $u\in U$, $u_{h\tau}\in U_{h\tau}$. Учитывая оценки (9), (10), (17), (20), (21), (60), определение множеств (4), (8), неравенство

$$\sum_{i=1}^{M-1}hb_i^2=\sum_{i=1}^{M-1}h\Big(\frac{1}{h}\int_{s_i}^{s_{i+1}}b(\xi)d\xi\Big)^2\leqslant\sum_{i=1}^{M-1}\int_{s_i}^{s_{i+1}}b^2(\xi)d\xi\leqslant\|b\|^2_{L_2[0,l]},$$

имеем

$$|J(u)-I_{h\tau}(u_{h\tau})|=\int_0^h|x(s,T;u)-b(s)|^2ds+\sum_{i=1}^{M-1}\int_{s_i}^{s_{i+1}}((x(s,t_N;u)-y_{iN})+$$
$$+(b_i-b(s)))(x(s,t_N;u)-b(s)+y_{iN}-b_i)ds\leqslant\int_0^h 2(|x(s,T;u)|^2+b^2(s))ds+$$
$$+\Big[\Big(\sum_{i=1}^{M-1}\int_{s_i}^{s_{i+1}}|x(s,t_N;u)-y_{iN}|^2ds\Big)^{1/2}+\Big(\sum_{i=1}^{M-1}\int_{s_i}^{s_{i+1}}|b(s)-b_i|^2ds\Big)^{1/2}\Big]\times$$
$$\times\Big[\Big(\int_0^l x^2(s,t_N;u)ds\Big)^{1/2}+\Big(\sum_{i=1}^{M-1}hy_{iN}^2\Big)^{1/2}+\|b\|_{L_2[0,l]}+\Big(\sum_{i=1}^{M-1}hb_i^2\Big)^{1/2}\Big]\leqslant$$
$$\leqslant C\Big(\|u-p_{h\tau}u_{h\tau}\|_{L_2(Q_h)}+\sqrt{\tau+h}+\Big(\sum_{i=1}^{M-1}\int_{s_i}^{s_{i+1}}(b(s)-b_i)^2ds\Big)^{1/2}\Big)+$$
$$+hCR^2+2\int_0^h b^2(s)ds. \quad (61)$$

Заметим, что

$$\sum_{i=1}^{M-1}\int_{s_i}^{s_{i+1}}(b(s)-b_i)^2ds=\sum_{i=1}^{M-1}\int_{s_i}^{s_{i+1}}\Big(\frac{1}{h}\int_{s_i}^{s_{i+1}}(b(s)-b(\xi))d\xi\Big)^2ds\leqslant$$
$$\leqslant\sum_{i=1}^{M-1}\int_{s_i}^{s_{i+1}}\Big(\frac{1}{h}\int_{-h}^{h}(b(s+z)-b(s))^2dz\Big)ds=$$
$$=\frac{1}{h}\int_{-h}^{h}\Big(\sum_{i=1}^{M-1}\int_{s_i}^{s_{i+1}}(b(s+z)-b(s))^2ds\Big)dz\leqslant\max_{|z|\leqslant h}\int_0^l|b(s+z)-b(s)|^2ds\to 0 \quad (62)$$

при $h\to 0$ в силу непрерывности в среднем функции $b(s)\in L_2(0,l)$ [371; 393; 648]; здесь предполагается, что $b(s)\equiv 0$ вне $[0,l]$.

Выше было отмечено, что задачи (1)–(4), (5)–(8) имеют решение, т. е. $U_*\neq\varnothing$, $U_{h\tau *}\neq\varnothing$. Зафиксируем какие-либо $u_*\in U_*$, $u_{h\tau *}\in U_{h\tau *}$. Поскольку $\|p_{h\tau}u_{h\tau *}\|_{L_2(Q_h)}=\|u_{h\tau *}\|_{L_{2h}}\leqslant R$, то, приняв $p_{h\tau}u_{h\tau *}=0$ вне Q_h, можем считать, что $p_{h\tau}u_{h\tau *}\in U$. Для управления $u_*\in U_*$ построим его дискретный аналог

$Q_{h\tau}u_*=\{u_{*ij},\ i=1,\ldots,M-1,\ j=1,\ldots,N\}$ по правилу $u_{*ij}=\frac{1}{h\tau}\iint\limits_{Q_{ij}} u_*(s,t)dsdt$. Так как

$$\|Q_{h\tau}u_*\|^2_{L_{2h\tau}}=\sum_{j=1}^{N}\sum_{i=1}^{M-1}h\tau\left(\frac{1}{h\tau}\iint\limits_{Q_{ij}}u_*(s,t)dsdt\right)^2\leqslant$$
$$\leqslant\sum_{j=1}^{N}\sum_{i=1}^{M-1}\iint\limits_{Q_{ij}}u_*^2(s,t)dsdt=\|u_*\|^2_{L_2(Q_h)}\leqslant\|u_*\|^2_{L_2(Q)}\leqslant R^2,$$

то $Q_{h\tau}u_*\in U_{h\tau}$. Нетрудно проверить, что $p_{h\tau}Q_{h\tau}u_*\in U$. Кроме того,

$$\lim_{(h,\tau)\to 0}\|u_*-p_{h\tau}Q_{h\tau}u_*\|_{L_2(Q_h)}=0, \tag{63}$$

что доказывается также, как аналогичное утверждение (62). Из оценки (61) имеем

$$J_*-I_{h\tau*}\leqslant J(p_{h\tau}u_{h\tau*})-I_{h\tau}(u_{h\tau*})\leqslant$$
$$\leqslant C\left(\sqrt{\tau+h}+\left(\sum_{i=1}^{M-1}\int\limits_{s_i}^{s_{i+1}}|b(s)-b_i|^2ds\right)^{1/2}\right)+hCR^2+2\int\limits_0^h b^2(s)ds. \tag{64}$$

Из (64) с учетом (62) получаем

$$\overline{\lim_{(h,\tau)\to 0}}(J_*-I_{h\tau*})\leqslant 0. \tag{65}$$

С другой стороны, из той же оценки (61) следует

$$J_*-I_{h\tau*}\geqslant J(u_*)-I_{h\tau}(Q_{h\tau}u_*)\geqslant -C\left(\|u_*-p_{h\tau}Q_{h\tau}u_*\|_{L_2(Q_h)}+\sqrt{\tau+h}+\right.$$
$$\left.+\left(\sum_{i=1}^{M-1}\int\limits_{s_i}^{s_{i+1}}|b(s)-b_i|^2ds\right)^{1/2}\right)-hCR^2-2\int\limits_0^h b^2(s)ds. \tag{66}$$

Отсюда и из (62), (63) имеем

$$\underline{\lim_{(h,\tau)\to 0}}(J_*-I_{h\tau*})\geqslant 0. \tag{67}$$

Из (65), (67) вытекает утверждение теоремы 1. Неравенства (64), (66) можно считать оценкой скорости сходимости для $J_*-I_{h\tau*}$. Если $b(s)$, $u_*=u_*(s,t)$ достаточно гладкие, например, $b(s)\in H^1[0,l]$, $u_*\in H^1(Q)$, $\frac{\partial^2 u_*}{\partial s\partial t}\in L_2(Q)$, то из (64), (66) следует, что $J_*-I_{h\tau*}=O(\sqrt{\tau+h})$. □

Вопросы аппроксимации для задач оптимального управления процессами, описываемыми уравнениями с частными производными, изучались, например, в работах [120; 170; 339; 360–363; 464; 467; 504; 593; 596–599; 607; 700].

Упражнения

1. Рассмотреть задачу (1)–(3), считая, что

$$U=\{u(s,t)\in L_2(Q)\colon \alpha\leqslant u(s,t)\leqslant\beta \text{ почти всюду на } Q\}, \tag{68}$$

где α, β — заданные постоянные. В задаче (5)–(7) принять $U_{h\tau}=\{u_{ij}\colon \alpha\leqslant u_{ij}\leqslant\beta, i=1,\ldots,M-1;\ j=1,\ldots,N\}$. Доказать, что получившаяся разностная задача минимизации аппроксимирует исходную по функции. Указание: воспользоваться оценкой (60).

2. Рассмотреть задачу (1)–(3), считая, что $u(s,t)=d(s)u(t)$, $d(s)\in L_2[0,l]$ — заданная функция, управление $u=u(t)$ принадлежит множеству

$$U_1=\{u(t)\in L_2(0,T)\colon \int\limits_0^T u^2(t)dt\leqslant R^2\} \tag{69}$$

или

$$U_2=\{u(t)\in L_2(0,T)\colon \alpha\leqslant u(t)\leqslant\beta \text{ почти всюду на } [0,T]\}. \tag{70}$$

В задаче (5)–(8) взять $u_{ij}=d_iu_j$, $d_i=\frac{1}{h}\int\limits_{s_i}^{s_{i+1}}d(\xi)d\xi$, $i=1,\ldots,M-1$, $j=1,\ldots,N$; управление $u_\tau=(u_1,\ldots,u_N)\in U_{1\tau}=\{u_\tau\colon \sum\limits_{j=1}^{N}\tau u_j^2\leqslant R^2\}$ в случае множества (69) и $u_\tau\in U_{2\tau}=\{u_\tau\colon \alpha\leqslant u_j\leqslant\beta,\ j=1,\ldots,N\}$ — в случае (70). Доказать, что полученная разностная задача минимизации аппроксимирует исходную по функции. Указание: воспользоваться оценкой (60).

3. Доказать дифференцируемость функций (1), (5) в пространствах $L_2(Q)$, $L_{2h\tau}$ соответственно. Описать методы проекции градиента, условного градиента для задач (1)–(3), (5)–(7), считая, что множество U имеем вид (4) или (68). Те же проблемы рассмотреть для множеств (69), (70).

4. Опираясь на представление решения краевой задачи (2), (3) в форме ряда (19), описать метод моментов для задачи управления $x(s,T;u)=b(s)$; рассмотреть случаи, когда множество U имеет вид (4), (68)–(70); дать обоснование метода [287; 802].

5. Рассмотреть задачу (1)–(3), когда в (3) граничные условия заменены на $x|_{s=0}=x|_{s=l}=0$, и множество U определяется одним из равенств (4), (68)–(70). В задаче (5)–(7) граничные условия в (7) заменить на $y_{0j}=0$, $y_{Mj}=0$, $j=1,\ldots,N$, аппроксимирующие множества $U_{h\tau}$, U_τ взять соответственно из (8) или из упражнений 1, 2. Исследовать сходимость по функции полученных разностных задач минимизации. Указание: получить аналоги оценок (9), (10), (20), (21), (60).

§ 8. Об аппроксимации максиминных задач

1. Пусть X, Y — множества произвольной природы, U, V — заданные множества, $U\subseteq X$, $V\subseteq Y$, функция $J(u,v)$ переменных u, v определена при всех $(u,v)\in U\times V$.

Рассмотрим задачу: найти величину

$$\sup_{u\in U}\inf_{v\in V}J(u,v)=J_*. \tag{1}$$

Задачи такого типа возникают в теории игр и исследовании операций, в вопросах приближения функций, при исследовании влияния погрешности исходных данных на решение задачи минимизации и т. д. [2–4; 26; 30; 153; 218–220; 310–313; 340; 392; 399; 402; 419; 431; 432; 479; 501; 532; 537; 551–554; 569–571; 594; 595; 650; 687; 689; 720; 754; 755; 801].

Пусть X_N, Y_N, $N=1,2,\ldots$, — некоторые множества произвольной природы, U_N, V_N — заданные множества, $U_N\subseteq X_N$, $V_N\subseteq Y_N$, функции $I_N([u]_N,[v]_N)$ определены при всех $([u]_N,[v]_N)\in U_N\times V_N$, $N=1,2,\ldots$ Рассмотрим последовательность задач: найти

$$\sup_{[u]_N\in U_N}\inf_{[v]_N\in V_N}I_N([u]_N,[v]_N)=I_{N*},\quad N=1,2,\ldots \tag{2}$$

Возникает интересный для приложений вопрос: каким условиям должны удовлетворять множества U_N, V_N и функции $I_N([u]_N,[v]_N)$ для того, чтобы последовательность задач (2) аппроксимировала задачу (1) по функции, т. е.

$$\lim_{N\to\infty}I_{N*}=J_*? \tag{3}$$

Следующая теорема дает ответ на этот вопрос.

Теорема 1. *Для того чтобы последовательность задач* (2) *аппроксимировала задачу* (1) *по функции, необходимо и достаточно выполнения следующих двух условий:*

1) *для каждого натурального числа* $N\geqslant 1$ *существует отображение* P_N: $X_N\to X$ *и для любого* $[u]_N\in U_N$ *существует отображение* Q_N: $Y\to Y_N$ *такие, что* $P_N([u]_N)\in U$ *при* $[u]_N\in U_N$, $Q_N(v)\in V_N$ *при* $v\in V$ *и*

$$\overline{\lim_{N\to\infty}}[I_N([u]_N,Q_N(v_N))-J(P_N([u]_N),v_N)]\leqslant 0 \tag{4}$$

при любом выборе $[u]_N \in U_N$ *и* $v_N \in V$ *(подчеркнем, что отображение* Q_N *в* (4), *вообще говоря, зависит от* $[u]_N \in U_N$*);*

2) *для каждого натурального числа* $N \geqslant 1$ *существует отображение* $\overline{Q}_N\colon X \to X_N$ *и для любого* $u_N \in U$ *существует отображение* $\overline{P}_N\colon Y_N \to Y$ *такие, что* $\overline{Q}_N(u) \in U_N$ *при* $u \in U$, $\overline{P}_N([v]_N) \in V$ *при* $[v]_N \in V_N$,

$$\overline{\lim_{N\to\infty}}[J(u_N, \overline{P}_N([v]_N)) - I_N(\overline{Q}_N(u_N), [v]_N)] \leqslant 0 \tag{5}$$

при любом выборе $u_N \in U$ *и* $[v]_N \in V_N$ *(подчеркнем, что отображение* $\overline{P}_N$ *в (5), вообще говоря, зависит от* $u_N \in U$*).*

Доказательство. Из определений величин J_*, I_{N*} следует, что

1) существуют $u_{N*} \in U$, $N = 1, 2, \ldots$, такие, что

$$\overline{\lim_{N\to\infty}}(J_* - J(u_{N*}, v_N)) \leqslant 0 \tag{6}$$

при любом выборе $v_N \in V$;

2) существуют $[u]_{N*} \in U_N$, $N = 1, 2, \ldots$, такие, что

$$\overline{\lim_{N\to\infty}}(I_{N*} - I_N([u]_{N*}, [v]_N)) \leqslant 0 \tag{7}$$

при любом выборе $[v]_N \in V_N$;

3) для каждого фиксированного $u_N \in U$ найдется точка $v_{N*} \in V$ такая, что

$$\overline{\lim_{N\to\infty}}(J(u_N, v_{N*}) - J_*) \leqslant 0; \tag{8}$$

4) для каждого фиксированного $[u]_N \in U_N$ найдется точка $[v]_{N*} \in V_N$ такая, что

$$\overline{\lim_{N\to\infty}}(I_N([u]_N, [v]_{N*}) - I_{N*}) \leqslant 0. \tag{9}$$

В самом деле, из определения верхней грани следует существование таких $u_{N*} \in U$ и $u_{N*} \in U_N$, что

$$\inf_{v\in V} J(u_{N*}, v) \geqslant J_* - 1/N, \quad \inf_{[v]_N \in V_N} I_N([u]_{N*}, [v]_N) \geqslant I_{N*} - 1/N, \quad N = 1, 2, \ldots$$

Вспоминая определение нижней грани, отсюда при любом выборе $v_N \in V$, $[v]_N \in V_N$ имеем

$$J(u_{N*}, v_N) \geqslant J_* - 1/N, \quad I_N([u]_{N*}, [v]_N) \geqslant I_{N*} - 1/N,$$

или

$$J_* - J(u_{N*}, v_N) \leqslant 1/N, \quad I_{N*} - I_N([u]_{N*}, [v]_N) \leqslant 1/N, \quad N = 1, 2, \ldots$$

Отсюда при $N \to \infty$ получим неравенства (6) и (7).

Далее, зафиксируем произвольные $u_N \in U$ и $[u]_N \in U_N$. По определению величин $\inf_{v\in V} J(u_N, v)$, $\inf_{[v]_N \in V_N} I_N([u]_N, [v]_N)$ найдутся $v_{N*} \in V$, $[v]_{N*} \in V_N$ такие, что

$$J(u_N, v_{N*}) \leqslant \inf_{v\in V} J(u_N, v) + 1/N \leqslant J_* + 1/N,$$

$$I_N([u]_N, [v]_{N*}) \leqslant \inf_{[v]_N \in V_N} I_N([u]_N, [v]_N) + 1/N \leqslant I_{N*} + 1/N,$$

или

$$J(u_N, v_{N*}) - J_* \leqslant 1/N, \quad I_N([u]_N, [v]_{N*}) - I_{N*} \leqslant 1/N, \quad N = 1, 2, \ldots$$

Отсюда при $N \to \infty$ получим неравенства (8), (9).

Необходимость. Пусть задачи (1), (2) таковы, что выполнено равенство (3). Покажем, что тогда необходимо выполняются условия 1), 2). Определим отображение $P_N\colon X_N \to X$ так: $P_N([u]_N) = u_{N*}$ при всех $[u]_N \in X_N$, $N = 1, 2, \ldots$, где $u_{N*} \in U$ взяты из (6). Тогда из неравенства (6) следует, что

$$\overline{\lim_{N\to\infty}}(J_* - J(P_N([u]_N), v_N)) \leqslant 0, \quad [u]_N \in U_N, \quad v_N \in V.$$

Зафиксируем произвольный элемент $[u]_N \in U_N$, возьмем соответствующие ему $[v]_{N*} \in V_N$ из (9) и определим отображение $Q_N\colon Y \to Y_N$ так: $Q_N(v) = [v]_{N*}$ при всех $v \in Y$, $N = 1, 2, \ldots$ Отсюда и из (9) имеем

$$\overline{\lim_{N\to\infty}}(I_N([u]_N, Q_N(v_N)) - I_{N*}) \leqslant 0, \quad [u]_N \in U_N, \quad v_N \in V. \tag{11}$$

Из (3), (10), (11) тогда следует, что

$$\overline{\lim_{N\to\infty}}(I_N([u]_N, Q_N(v_N)) - J(P_N([u]_N), v_N)) \leqslant \overline{\lim_{N\to\infty}}(I_{N*} - J_*) + \overline{\lim_{N\to\infty}}(I_N([u]_N, Q_N(v_N)) - I_{N*}) + \overline{\lim_{N\to\infty}}(J_* - J(P_N([u]_N), v_N)) \leqslant 0$$

при любом выборе $[u]_N \in U_N$ и $v_N \in V$. Необходимость условия 1) установлена.

Далее, определим отображение $\overline{Q}_N\colon X \to X_N$ так: $\overline{Q}_N(u) = [u]_{N*}$ при всех $u \in X$, $N = 1, 2, \ldots$, где $[u]_{N*}$ взяты из (7). Тогда из неравенства (7) следует, что

$$\overline{\lim_{N\to\infty}}(I_{N*} - I_N(\overline{Q}_N(u), [v]_N)) \leqslant 0, \quad u \in U, \quad [v]_N \in V_N. \tag{12}$$

Зафиксируем произвольный элемент $u_N \in U$, возьмем соответствующий ему $v_{N*} \in V$ из (8) и определим отображение $\overline{P}_N\colon Y_N \to Y$ так: $\overline{P}_N([v]_N) = v_{N*}$ при всех $[v]_N \in Y_N$, $N = 1, 2, \ldots$ Отсюда и из (8) имеем

$$\overline{\lim_{N\to\infty}}(J(u_N, \overline{P}_N([v]_N)) - J_*) \leqslant 0, \quad u_N \in U, \quad [v]_N \in V_N. \tag{13}$$

Из (3), (12), (13) тогда следует, что

$$\overline{\lim_{N\to\infty}}(J(u_N, \overline{P}_N([v]_N)) - I_N(\overline{Q}_N(u_N), [v]_N)) \leqslant \overline{\lim_{N\to\infty}}(J_* - I_{N*}) + \overline{\lim_{N\to\infty}}(J(u_N, \overline{P}_N([v]_N)) - J_*) + \overline{\lim_{N\to\infty}}(I_{N*} - I_N(\overline{Q}_N(u_N), [v]_N)) \leqslant 0$$

при любом выборе $u_N \in U$, $[v]_N \in V_N$. Необходимость условия 2) также установлена.

Достаточность. Пусть выполнены условия 1), 2) теоремы. Покажем, что тогда имеет место равенство (3). Возьмем точки $[u]_{N*} \in U_N$ из (7) и положим $u_N = P_N([u]_{N*})$, а затем из (8) возьмем точки $v_{N*} \in V$, соответствующие именно точкам $u_N = P_N([u]_{N*})$, $N = 1, 2, \ldots$ Тогда, полагая в (4) $[u]_N = [u]_{N*}$, $v_N = v_{N*}$ и взяв в качестве отображения Q_N то, которое соответствует точке $[u]_{N*}$, с учетом неравенств (7), (8) получим

$$\overline{\lim_{N\to\infty}}(I_{N*} - J_*) \leqslant \overline{\lim_{N\to\infty}}(I_{N*} - I_N([u]_{N*}, Q_N(v_{N*}))) + \overline{\lim_{N\to\infty}}(I_N([u]_{N*}, Q_N(v_{N*})) - J(P_N([u]_{N*}), v_{N*})) + \overline{\lim_{N\to\infty}}(J(P_N([u]_{N*}), v_{N*}) - J_*) \leqslant 0.$$

Далее, возьмем точки $u_{N*} \in U$ из (6) и положим $[u]_N = \overline{Q}_N(u_{N*})$, а затем из (9) возьмем точки $[v]_{N*} \in V_N$, соответствующие именно точкам $[u]_N = \overline{Q}_N(u_{N*})$, $N = 1, 2, \ldots$ Тогда, полагая в (5) $u_N = u_{N*}$, $[v]_N = [v]_{N*}$ и взяв в качестве отображения $\overline{P}_N$ то, которое соответствует точке u_{N*}, с учетом неравенств (6), (9) получим

$$\overline{\lim_{N\to\infty}}(J_* - I_{N*}) \leqslant \overline{\lim_{N\to\infty}}(J_* - J(u_{N*}, \overline{P}_N([v]_{N*}))) + \overline{\lim_{N\to\infty}}(J(u_{N*}, \overline{P}_N([v]_{N*})) - I_N(\overline{Q}_N(u_{N*}), [v]_{N*})) + \overline{\lim_{N\to\infty}}(I_N(\overline{Q}_N(u_{N*}), [v]_{N*}) - I_{N*}) \leqslant 0.$$

Таким образом, имеем

$$0 \leqslant -\overline{\lim_{N\to\infty}}(J_* - I_{N*}) = \underline{\lim_{N\to\infty}}(I_{N*} - J_*) \leqslant \overline{\lim_{N\to\infty}}(I_{N*} - J_*) \leqslant 0,$$

или $\underline{\lim}_{N\to\infty} I_{N*} = \overline{\lim}_{N\to\infty} I_{N*} = J_*$, т. е. равенство (3). Теорема 1 доказана. □

2. Для иллюстрации теоремы 1 рассмотрим задачу: найти

$$\sup_{u\in U}\inf_{v\in V} J(u, v) = J_*, \tag{14}$$

где

$$J(u, v) = |x(T, u, v) - y|^2 \tag{15}$$

при следующих условиях:

$$\dot{x}(t)=A(t)x(t)+B(t)u(t)+C(t)v(t)+f(t),\quad t_0\leqslant t\leqslant T;\quad x(t_0)=x_0, \tag{16}$$

$$u=u(t)\in U=\{u(t)\in L_2^r[t_0,T]\colon u(t)\in P\quad\text{почти всюду на}\quad[t_0,T]\}, \tag{17}$$

$$v=v(t)\in V=\{v(t)\in L_2^q[t_0,T]\colon v(t)\in Q\quad\text{почти всюду на}\quad[t_0,T]\}, \tag{18}$$

где $A(t), B(t), C(t), f(t)$ — матрицы порядка $n\times n$, $n\times r$, $n\times q$, $n\times 1$ соответственно: моменты t_0, T, точки $x_0, y\in E^n$ заданы; P и Q — заданные множества из E^r и E^q соответственно; $x(t,u,v)$ — решение задачи (16), соответствующее управлениям $u=u(t)\in L_2^r[t_0,T]$, $v=v(t)\in L_2^q[t_0,T]$. Будем предполагать, что матрицы $A(t), B(t), C(t), f(t)$ кусочно непрерывны на отрезке $[t_0,T]$.

Разобьем отрезок $t_0\leqslant t\leqslant T$ на N частей точками $t_0<t_1<\ldots<t_N=T$ и, приняв эти точки в качестве узловых, уравнения (16) заменим разностными уравнениями с помощью схемы Эйлера. В результате придем к следующей разностной аппроксимации задачи (14)–(18): найти

$$\sup_{[u]_N\in U_N}\inf_{[v]_N\in V_N} I_N([u]_N,[v]_N)=I_{N*}, \tag{19}$$

где

$$I_N([u]_N,[v]_N)=|x_N([u]_N,[v]_N)-y|^2 \tag{20}$$

при условиях

$$x_{i+1}=x_i+\Delta t_i(A_ix_i+B_iu_i+C_iv_i+f_i),\quad i=0,\ldots,N-1, \tag{21}$$

$$[u]_N\in U_N=\{[u]_N=(u_0,\ldots,u_{N-1})\in L_{2N}^r\colon u_i\in P,\quad i=0,\ldots,N-1\}, \tag{22}$$

$$[v]_N\in V_N=\{[v]_N=(v_0,\ldots,v_{N-1})\in L_{2N}^q\colon v_i\in Q,\quad i=0,\ldots,N-1\}; \tag{23}$$

здесь $\Delta t_i=t_{i+1}-t_i$, $A_i=A(t_i+0)$, $B_i=B(t_i+0)$, $C_i=C(t_i+0)$, $f_i=f(t_i+0)$, $[x([u]_N,[v]_N)]_N=(x_0,x_1([u]_N,[v]_N),\ldots,x_N([u]_N,[v]_N))$ — решение задачи (21), соответствующее управлениям

$$[u]_N\in L_{2N}^r,\quad [v]_N\in L_{2N}^q,\quad N=1,2,\ldots$$

Опираясь на теорему 1, сформулируем условия, при которых последовательность задач (19)–(23) аппроксимирует задачу (14)–(18) по функции.

Теорема 2. *Пусть матрицы $A(t), B(t), C(t), f(t)$ кусочно непрерывны на отрезке $[t_0,T]$, множества $P\subset E^r$, $Q\in E^q$ выпуклы, замкнуты и ограничены, разбиения $\{t_i, i=0,\ldots,N\}$ отрезка $[t_0,T]$ таковы, что*

$$d_N=\max_{0\leqslant i\leqslant N-1}\Delta t_i\leqslant (T-t_0)M_0/N,\quad N=1,2,\ldots$$

Тогда

$$\lim_{N\to\infty} I_{N*}=J_*.$$

Доказательство. Заметим, что

$$\sup_{u\in U}\sup_{v\in V}\max_{t_0\leqslant t\leqslant T}|x(t,u,v)|\leqslant C_0<\infty, \tag{24}$$

$$\sup_{[u]_N\in U_N}\sup_{[v]_N\in V_N}\max_{0\leqslant i\leqslant N}|x_i([u]_N,[v]_N)|\leqslant C_4<\infty, \tag{25}$$

$$\sup_{u\in U}\sup_{v\in V}|x(t,u,v)-x(\tau,u,v)|\leqslant C_1|t-\tau|,\quad t_0\leqslant t,\tau\leqslant T, \tag{26}$$

где C_0, C_1, C_4 — положительные константы. Оценки (24)–(20) доказываются так же, как соответствующие оценки (1.8), (1.16), (1.11).

Положим $X=L_2^r[t_0,T]$, $X_N=L_{2N}^r$, $Y=L_2^q[t_0,T]$, $Y_N=L_{2N}^q$. Определим отображения $\overline{Q}_N\colon X\to X_N$, $Q_N\colon Y\to Y_N$, $P_N\colon X_N\to X$, $\overline{P}_N\colon Y_N\to Y$ следующим образом:

$$\overline{Q}_N(u)=(u_0,u_1,\ldots,u_{N-1})\colon u_i=\frac{1}{\Delta t_i}\int_{t_i}^{t_{i+1}}u(t)dt,\quad i=0,\ldots,N-1,$$

$$Q_N(v)=(v_0,v_1,\ldots,v_{N-1})\colon v_i=\frac{1}{\Delta t_i}\int_{t_i}^{t_{i+1}}v(t)dt,\quad i=0,\ldots,N-1,$$

$$P_N([u]_N)=u_i,\quad \overline{P}_N([v]_N)=v_i\quad\text{при}\quad t_i\leqslant t<t_{i+1},\quad i=0,\ldots,N-1.$$

С помощью леммы 1.1 получаем, что

$$\overline{Q}_N(u)\in U_N\text{ при всех }u\in U,\quad Q_N(v)\in V_N\text{ при всех }v\in V,$$

$$P_N([u]_N)\in U\text{ при всех }[u]_N\in U_N,\quad \overline{P}_N([v]_N)\in V\text{ при всех }[v]_N\in V_N,\quad N=1,2,\ldots$$

Справедливы оценки

$$\sup_{[u]_N\in U_N}\sup_{v\in V}\max_{0\leqslant i\leqslant N}|x(t_i,P_N([u]_N),v)-x_i([u]_N,Q_N(v))|\leqslant\delta_N, \tag{27}$$

$$\sup_{[v]_N\in V_N}\sup_{u\in U}\max_{0\leqslant i\leqslant N}|x(t_i,u,\overline{P}_N([v]_N))-x_i(\overline{Q}_N(u),[v]_N)|\leqslant\delta_N, \tag{28}$$

где

$$\delta_N=e^{A_{\max}(T-t_0)M_0}\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}(\|A(\tau)-A_i\|C_0+A_{\max}C_1d_N+\sup_P|u|\|B(\tau)-B_i\|+\sup_Q|v|\|C(\tau)-C_i\|+|f(\tau)-f_i|)d\tau\to 0$$

при $N\to\infty$. Оценки (27), (28) доказываются с использованием оценок (24), (26) совершенно так же, как аналогичные оценки (1.24), (1.25).

Рассуждая так же, как в леммах 1.3 и 1.4, с помощью оценок (24)–(28) получаем, что

$$|I_N([u]_N,Q_N(v))-J(P_N([u]_N,v))|\leqslant C_5\delta_N, \tag{29}$$

$$|J(u,\overline{P}_N([v]_N))-I_N(\overline{Q}_N(u),[v]_N)|\leqslant C_5\delta_N \tag{30}$$

при всех $[u]_N\in U_N$, $v\in V$, $u\in U$, $[v]_N\in V_N$, $N=1,2,\ldots$, где $C_5=2C_0+2C_4+2|y|$.

Из оценок (29), (30) следуют неравенства (4), (5). Таким образом, все условия теоремы 1 выполнены. Отсюда следует, что $\lim\limits_{N\to\infty} I_{N*}=J_*$. Теорема 2 доказана. □

3. В рассмотренной выше максиминной задаче (1), (2) множества U, V, а также их «аппроксимации» U_N, V_N не связаны между собой, поэтому такие задачи принято называть максиминными задачами с несвязанными множествами. В приложениях, например в теории игр с непротивоположными интересами [220; 650; 720], часто встречаются максиминные задачи, в которых множества U, V и их «аппроксимации» связаны друг с другом. Такие задачи называют максиминными задачами со связанными множествами.

Рассмотрим простейшую задачу такого типа. Пусть X, Y — множества произвольной природы, U — заданное подмножество из X, и пусть каждой точке $u\in U$ поставлено в соответствие множество $V(u)\subseteq Y$. Пусть функция $J(u,v)$ определена при всех $(u,v)\in U\times Y$. Рассмотрим задачу: найти величину

$$\sup_{u\in U}\inf_{v\in V(u)} J(u,v)=J_*. \tag{31}$$

Пусть X_N, Y_N, $N=1,2,\ldots$, — некоторые множества произвольной природы, U_N — заданное множество из X_N, и пусть каждому $[u]_N\in U_N$ поставлено в соответствие множество $V_N([u]_N)\subseteq Y_N$, функция $I_N([u]_N,[v]_N)$ определена при всех $([u]_N,[v]_N)\in U_N\times Y_N$. Рассмотрим последовательность задач: найти величины

$$\sup_{[u]_N\in U_N}\inf_{[v]_N\in V_N([u]_N)} I_N([u]_N,[v]_N)=I_{N*},\quad N=1,2,\ldots \tag{32}$$

В следующей теореме даются условия, при которых последовательность задач (32) аппроксимирует задачу (31) по функции, т. е. $\lim\limits_{N\to\infty} I_{N*}=J_*$.

Теорема 3. *Для того чтобы последовательность задач (32) аппроксимировала задачу (31) по функции, необходимо и достаточно выполнения следующих двух условий:*

1) для каждого натурального числа $N\geqslant 1$ существует отображение $P_N\colon X_N\to X$ и для произвольного фиксированного $[u]_N\in U_N$ существует отображение $Q_N\colon Y\to Y_N$ такие, что

$$P_N([u]_N)\in U\quad\textit{при всех}\quad[u]_N\in U_N, \tag{33}$$

$$Q_N(v)\in V_N([u]_N)\quad\textit{при всех}\quad v\in V(P_N([u]_N)), \tag{34}$$

$$\overline{\lim_{N\to\infty}}(I_N([u]_N,Q_N(v_N))-J(P_N([u]_N),v_N))\leqslant 0 \tag{35}$$

при всех $v_N\in V(P_N([u]_N))$, $[u]_N\in U_N$;

2) *для каждого натурального числа* $N \geqslant 1$ *существует отображение* $\overline{Q}_N\colon X \to X_N$ *и для любого фиксированного* $u_N \in U$ *существует* $\overline{P}_N\colon Y_N \to Y$ *такие, что*

$$\overline{Q}_N(u) \in U_N \quad \textit{при всех} \quad u \in U, \tag{36}$$

$$\overline{P}_N([v]_N) \in V(u_N) \quad \textit{при всех} \quad [v]_N \in V_N(\overline{Q}_N(u_N)), \tag{37}$$

$$\overline{\lim_{N\to\infty}}(J(u_N, \overline{P}_N([v]_N)) - I_N(\overline{Q}_N(u_N), [v]_N)) \leqslant 0 \tag{38}$$

при всех $[v]_N \in V_N(\overline{Q}_N(u_N))$, $u_N \in U$.

Доказательство. Из определений верхней и нижней граней, величин J_*, I_{N*} следует, что

1) для каждого натурального $N \geqslant 1$ существуют $u_{N*} \in U$ такие, что

$$\overline{\lim_{N\to\infty}}(J_* - J(u_{N*}, v_N)) \leqslant 0 \quad \text{при всех} \quad v_N \in V(u_{N*}); \tag{39}$$

2) для каждого $N \geqslant 1$ существуют $[u]_{N*} \in U_N$ такие, что

$$\overline{\lim_{N\to\infty}}(I_{N*} - I_N([u]_{N*}, [v]_N)) \leqslant 0 \tag{40}$$

при всех $[v]_N \in V_N([u]_{N*})$;

3) для любого $u_N \in U$ существует элемент $v_{N*} \in V(u_N)$ такой, что

$$\overline{\lim_{N\to\infty}}(J(u_N, v_{N*}) - J_*) \leqslant 0; \tag{41}$$

4) для любого $[u]_N \in U_N$ существует $[v]_{N*} \in V_N([u]_N)$ такой, что

$$\overline{\lim_{N\to\infty}}(I_N([u]_N, [v]_{N*}) - I_{N*}) \leqslant 0. \tag{42}$$

Справедливость соотношений (39)–(42) устанавливается так же, как и аналогичных соотношений (6)–(9).

Необходимость. Пусть $\lim\limits_{N\to\infty} I_{N*} = J_*$. Определим отображение $P_N\colon X_N \to X$ так: $P_N([u]_N) = u_{N*}$ при всех $[u]_N \in X_N$, $N = 1, 2, \ldots$, где $u_{N*} \in U$ взят из (39). Ясно, что тогда включение (33) выполнено и, кроме того, согласно (39)

$$\overline{\lim_{N\to\infty}}(J_* - J(P_N([u]_N), v_N)) \leqslant 0, \quad v_N \in V(u_{N*}).$$

Возьмем произвольный элемент $[u]_N \in U_N$, по нему из (42) найдем соответствующий $[v]_{N*} \in V_N([u]_N)$ и определим отображение $Q_N\colon Y \to Y_N$ так: $Q_N(v) = [v]_{N*}$ при всех $v \in Y$, $N = 1, 2, \ldots$ Тогда справедливо включение (34) и, кроме того, согласно (42)

$$\overline{\lim_{N\to\infty}}(I_N([u]_N, Q_N(v_N)) - I_{N*}) \leqslant 0.$$

Отсюда следует, что

$$\overline{\lim_{N\to\infty}}(I_N([u]_N, Q_N(v_N)) - J(P_N([u]_N), v_N)) \leqslant \overline{\lim_{N\to\infty}}(I_{N*} - J_*) + \overline{\lim_{N\to\infty}}(I_N([u]_N, Q_N(v_N)) - I_{N*}) + \overline{\lim_{N\to\infty}}(J_* - J(P_N([u]_N), v_N)) \leqslant 0$$

при всех $v_N \in V(P_N([u]_N))$, $[u]_N \in U_N$. Необходимость условия 1) доказана.

Далее, определим отображение $\overline{Q}_N\colon X \to X_N$ так: $\overline{Q}_N(u) = [u]_{N*}$ при всех $u \in X$, $N = 1, 2, \ldots$, где $[u]_{N*}$ взят из (40). Ясно, что тогда включение (36) выполнено и, кроме того, согласно (40)

$$\overline{\lim_{N\to\infty}}(I_{N*} - I_N(\overline{Q}_N(u_N), [v]_N)) \leqslant 0.$$

Возьмем произвольный элемент $u_N \in U$, по нему из (41) найдем соответствующий $v_{N*} \in V(u_N)$ и определим отображение $\overline{P}_N\colon Y_N \to Y$ так: $\overline{P}_N([v]_N) = v_{N*}$ при всех $[v]_N \in Y_N$, $N = 1, 2, \ldots$ Тогда справедливо включение (37) и, кроме того, согласно (41)

$$\overline{\lim_{N\to\infty}}(J(u_N, \overline{P}_N([v]_N)) - J_*) \leqslant 0.$$

Отсюда непосредственно следует, что

$$\overline{\lim_{N\to\infty}}(J(u_N, \overline{P}_N([v]_N)) - I_N(\overline{Q}_N(u_N), [v]_N)) \leqslant \overline{\lim_{N\to\infty}}(J_* - I_{N*}) + \overline{\lim_{N\to\infty}}(J(u_N, \overline{P}_N([v]_N)) - J_*) + \overline{\lim_{N\to\infty}}(I_{N*} - I_N(\overline{Q}_N(u_N), [v]_N)) \leqslant 0$$

при всех $[v]_N \in V_N(\overline{Q}_N(u_N))$, $u_N \in U$. Необходимость условия 2) также доказана.

Достаточность. Пусть выполнены условия 1), 2). Возьмем $[u]_{N*} \in U_N$ из (40), положим $u_N = P_N([u]_{N*})$ и из (41) возьмем $v_{N*} \in V(u_N) = V(P_N([u]_{N*}))$. Обозначим $[v]_N = Q_N(v_{N*})$. Из (34) тогда имеем $[v]_N \in V_N([u]_{N*})$, $N = 1, 2, \ldots$ Пользуясь условием (35) при $[u]_N = [u]_{N*}$, $v_N = v_{N*}$, с учетом (40), (41) получим

$$\overline{\lim_{N\to\infty}}(I_{N*} - J_*) \leqslant \overline{\lim_{N\to\infty}}(I_{N*} - I_N([u]_{N*}, [v]_N)) + \overline{\lim_{N\to\infty}}(I_N([u]_{N*}, Q_N(v_{N*})) - J(P_N([u]_{N*}), v_{N*})) + \overline{\lim_{N\to\infty}}(J(u_N, v_{N*}) - J_*) \leqslant 0.$$

Далее, возьмем $u_{N*} \in U$ из (41), положим $[u]_N = \overline{Q}_N(u_{N*})$ и из (42) возьмем соответствующий элемент $[v]_{N*} \in V_N([u]_N) = V_N(\overline{Q}_N(u_{N*}))$. Обозначим $v_N = \overline{P}_N([v]_{N*})$. Из (37) тогда следует, что $v_N \in V(u_{N*})$, $N = 1, 2, \ldots$ Пользуясь условием (38) при $u_N = u_{N*}$, $[v]_N = [v]_{N*}$, с учетом (39), (42) имеем

$$\overline{\lim_{N\to\infty}}(J_* - I_{N*}) \leqslant \overline{\lim_{N\to\infty}}(J_* - J(u_{N*}, v_N)) + \overline{\lim_{N\to\infty}}(J(u_{N*}, \overline{P}_N([v]_{N*})) - I_N(\overline{Q}_N(u_{N*}), [v]_{N*})) + \overline{\lim_{N\to\infty}}(I_N([u]_N, [v]_{N*}) - I_{N*}) \leqslant 0.$$

Таким образом, $\underline{\lim} I_{N*} = \overline{\lim} I_{N*} = J_*$, т. е. $\lim\limits_{N\to\infty} I_{N*} = J_*$. Теорема 3 доказана. □

Для иллюстрации теоремы 3 рассмотрим задачу (31) для случая, когда

$$J(u, v) = |x(T, u, v) - y|^2,$$

где $x(t, u, v)$ — решение задачи (16), $u \in X = L_2^r[t_0, T]$, $v \in Y = L_2^q[t_0, T]$, множества U и V имеют вид (17), (18), а множество $V(u)$ при каждом $u \in U$ определяется так:

$$V(u) = \Big\{v \in V\colon \int_{t_0}^{T} g(u(t), v(t))dt \leqslant 0\Big\}. \tag{43}$$

Эту задачу кратко будем называть задачей (31), (15)–(18), (43).

Для аппроксимации этой задачи рассмотрим последовательность задач (32), где $I_N([u]_N, [v]_N) = |x_N([u]_N, [v]_N) - y|^2$; $x_i([u]_N, [v]_N)$, $i = 0, \ldots, N$, — решение задачи (21), соответствующее управлениям $[u]_N \in X_N = L_{2N}^r$, $[v]_N \in Y_N = L_{2N}^q$; множества U_N, V_N имеют вид (22), (23), а множество $V_N([u]_N)$ при каждом $[u]_N = (u_0, u_1, \ldots, u_{N-1}) \in U_N$ строится так:

$$V_N([u]_N) = \Big\{[v]_N = (v_0, v_1, \ldots, v_{N-1}) \in V_N\colon \sum_{i=0}^{N-1} g(u_i, v_i)\Delta t_i \leqslant 0\Big\}. \tag{44}$$

Эти задачи кратко будем называть задачами (32), (20)–(23), (44).

Оказывается, если выполнены все условия теоремы 2 и, кроме того, функция $g(u, v)$ непрерывна по совокупности $(u, v) \in P \times Q$, выпукла по переменной $v \in Q$ при каждом фиксированном $u \in P$ и вогнута по переменной $u \in P$ при каждом фиксированном $v \in Q$, то последовательность задач (32), (20)–(23), (44) аппроксимирует задачу (31), (15)–(18), (43) по функции. Для того чтобы убедиться в этом, достаточно проверить выполнение условий теоремы 3.

Определим отображения P_N, $\overline{P}_N$, Q_N, $\overline{Q}_N$ так же, как в задачах (14)–(18) и (20)–(23). Тогда условия (33), (36) будут выполняться. Проверим справедливость включений (34), (37). С этой целью заметим, что из выпуклости $g(u, v)$ по переменной v следует неравенство

$$g\Big(u, \frac{1}{\Delta t_i}\int_{t_i}^{t_{i+1}} v(t)dt\Big) \leqslant \frac{1}{\Delta t_i}\int_{t_i}^{t_{i+1}} g(u, v(t))dt, \quad i = 0, \ldots, N-1, \quad u \in U, \tag{45}$$

обобщающее неравенство $g\Big(u, \sum\limits_{j=1}^{m} \alpha_j v_j\Big) \leqslant \sum\limits_{j=1}^{m} g(u, v_j)\alpha_j$, $\alpha_j \geqslant 0$, $\sum\limits_{j=1}^{m} \alpha_j = 1$ (см. неравенство (4.2.2)). Аналогично, из вогнутости $g(u, v)$ по переменной u получаем

$$g\Big(\frac{1}{\Delta t_i}\int\limits_{t_i}^{t_{i+1}} u(t)dt, v\Big) \geqslant \frac{1}{\Delta t_i}\int\limits_{t_i}^{t_{i+1}} g(u(t), v)dt, \quad i = 0, \ldots, N-1, \quad v \in V. \tag{46}$$

Зафиксируем некоторое $[u]_N = (u_0, \ldots, u_{N-1})$ и возьмем произвольное управление $v \in V(P_N([u]_N))$, т. е. $v \in V$, $\int\limits_{t_0}^{T} g(P_N([u]_N), v(t))\,dt \leqslant 0$. Тогда для $Q_N(v) = (v_0, v_1, \ldots, v_{N-1})$, $v_i = \frac{1}{\Delta t_i}\int\limits_{t_i}^{t_{i+1}} v(t)dt$, $i = 0, \ldots, N-1$, с учетом неравенства (45) имеем

$$\sum_{i=0}^{N-1} g(u_i, v_i)\Delta t_i = \sum_{i=0}^{N-1} g\Big(u_i, \frac{1}{\Delta t_i}\int\limits_{t_i}^{t_{i+1}} v(t)dt\Big)\Delta t_i \leqslant$$
$$\leqslant \sum_{i=0}^{N-1}\int\limits_{t_i}^{t_{i+1}} g(u_i, v(t))dt = \int\limits_{t_0}^{T} g(P_N([u]_N), v(t))dt \leqslant 0.$$

Это означает, что $Q_N(v) \in V_N([u]_N)$. Включение (34) доказано.

Далее, зафиксируем некоторое $u_N = u_N(t) \in U$, найдем $\overline{Q}_N(u_N) = (\overline{u}_0, \overline{u}_1, \ldots, \overline{u}_{N-1})$, $\overline{u}_i = \frac{1}{\Delta t_i}\int\limits_{t_i}^{t_{i+1}} u_N(t)dt$, $i = 0, \ldots, N-1$, и возьмем произвольное управление $[v]_N = (v_0, \ldots, v_{N-1}) \in V_N(\overline{Q}_N(u_N))$, т. е. $[v]_N \in V_N$, $\sum\limits_{i=0}^{N-1} g_i(\overline{u}_i, v_i)\Delta t_i \leqslant 0$. С учетом неравенства (46) тогда имеем

$$\int\limits_{t_0}^{T} g(u_N(t), \overline{P}_N([v]_N))dt = \sum_{i=0}^{N-1}\int\limits_{t_i}^{t_{i+1}} g(u_N(t), v_i)dt \leqslant$$
$$\leqslant \sum_{i=0}^{N-1} g\Big(\frac{1}{\Delta t_i}\int\limits_{t_i}^{t_{i+1}} u_N(t)dt, v_i\Big)\Delta t_i = \sum_{i=0}^{N-1} g(\overline{u}_i, v_i)\Delta t_i \leqslant 0.$$

Это означает, что $\overline{P}_N([v]_N) \in V(u_N)$. Включение (37) также доказано. Неравенства (35), (38) являются следствием неравенств (27), (28).

Таким образом, все условия теоремы 3 выполнены. Согласно этой теореме последовательность задач (32), (20)–(23), (44) аппроксимирует задачу (31), (15)–(18), (43) по функции.

4. В конкретных максиминных задачах построение аппроксимирующих задач, удовлетворяющих условиям сформулированных выше теорем, далеко не всегда является простым делом. Например, в задачах вида (31), (32) может встретить трудности построение отображений $P_N, Q_N, \overline{P}_N, \overline{Q}_N$, удовлетворяющих условиям (33), (34), (36), (37), обеспечение непустоты множеств U_N, $V_N([u]_N)$. Для преодоления указанных трудностей часто бывает полезно работать с расширениями множеств, встречающихся в исходной и аппроксимирующей максиминных задачах [3; 153; 594; 595]. Проиллюстрируем эту идею на примере задачи (31).

Теорема 4. *Для того чтобы последовательность задач* (32) *аппроксимировала задачу* (31) *по функции, необходимо и достаточно, чтобы существовали последовательности множеств* $\{U^{\varepsilon_N}\} \subset X$, $\{V^{\chi_N}(u)\}$, $\{V^{\gamma_N}(u)\} \in Y$ *таких, что* $U^{\varepsilon_N} \neq \varnothing$ *и* $V^{\chi_N}(u) \neq \varnothing$ *при* $u \in U$, $V^{\gamma_N}(u) \neq \varnothing$ *при* $u \in U^{\varepsilon_N}$, $N = 1, 2, \ldots$, *функция* $J(u, v)$ *определена при всех* $(u, v) \in \Big(\Big(\bigcup\limits_{N=1}^{\infty} U^{\varepsilon_N}\Big) \cup U\Big) \times Y$, *и, кроме того, выполнены следующие условия:*

1) *для каждого натурального числа* $N \geqslant 1$ *существует отображение* P_N: $X_N \to X$ *и для произвольного фиксированного* $[u]_N \in U_N$ *существует отображение* Q_N: $Y \to Y_N$ *такие, что*

$$P_N([u]_N) \in U^{\varepsilon_N}, \tag{47}$$

$$Q_N(v) \in V_N([u]_N) \quad \textit{при всех} \quad v \in V^{\gamma_N}(P_N([u]_N)), \tag{48}$$

$$\overline{\lim_{N\to\infty}}(I_N([u]_N, Q_N(v_N)) - J(P_N([u]_N, v_N)) \leqslant 0 \tag{49}$$

при всех $v_N \in V^{\gamma_N}(P_N([u]_N))$;

2) *для каждого натурального числа* $N \geqslant 1$ *существует отображение* $\overline{Q}_N$: $X \to X_N$ *и для любого фиксированного* $u_N \in U$ *существует отображение* $\overline{P}_N$: $Y_N \to Y$ *такие, что*

$$\overline{Q}_N(u_N) \in U_N, \tag{50}$$

$$\overline{P}_N([v]_N) \in V^{\chi_N}(u_N) \quad \textit{при всех} \quad [v]_N \in V_N(\overline{Q}_N(u_N)), \tag{51}$$

$$\overline{\lim_{N\to\infty}}(J(u_N, \overline{P}_N([v]_N)) - I_N(\overline{Q}_N(u_N), [v]_N)) \leqslant 0 \tag{52}$$

при всех $[v]_N \in V_N(\overline{Q}_N(u_N))$;

3) *справедливы неравенства*

$$\underline{\lim_{N\to\infty}} J_*(0, \chi_N) \geqslant J_*, \tag{53}$$

$$\overline{\lim_{N\to\infty}} J_*(\varepsilon_N, \gamma_N) \leqslant J_*, \tag{54}$$

где $J_*(\varepsilon_N, \gamma_N) = \sup\limits_{u\in U^{\varepsilon_N}} \inf\limits_{v\in V^{\gamma_N}(u)} J(u, v)$, $J_*(0, \chi_N) = \sup\limits_{u\in U} \inf\limits_{v\in V^{\chi_N}(u)} J(u, v)$.

Доказательство. Из определений величин $J_*(\varepsilon_N, \gamma_N)$, $J_*(0, \chi_N)$, I_{N*} следует, что

1) для каждого $N \geqslant 1$ существуют $u_{N*} \in U$ такие, что

$$\overline{\lim_{N\to\infty}}(J_*(0, \chi_N) - J(u_{N*}, v_N)) \leqslant 0 \tag{55}$$

при любом выборе $v_N \in V^{\chi_N}(u_{N*})$;

2) для каждого $N \geqslant 1$ существуют $[u]_{N*} \in U_N$ такие, что

$$\overline{\lim_{N\to\infty}}(I_{N*} - I_N([u]_{N*}, [v]_N)) \leqslant 0 \tag{56}$$

при любом выборе $[v]_N \in V_N([u]_{N*})$;

3) для каждого фиксированного элемента $u_N \in U^{\varepsilon_N}$ найдется элемент $v_{N*} \in V^{\gamma_N}(u_N)$ такой, что

$$\overline{\lim_{N\to\infty}}(J(u_N, v_{N*}) - J_*(\varepsilon_N, \gamma_N)) \leqslant 0; \tag{57}$$

4) для каждого фиксированного элемента $[u]_N \in U_N$ найдется элемент $[v]_{N*} \in V_N([u]_N)$ такой, что

$$\overline{\lim_{N\to\infty}}(I_N([u]_N, [v]_{N*}) - I_{N*}) \leqslant 0. \tag{58}$$

Справедливость соотношений (55)–(58) устанавливается так же, как и аналогичных соотношений (6)–(9) или (39)–(42).

Необходимость. Пусть $\lim\limits_{N\to\infty} I_{N*} = J_*$. Положим $U^{\varepsilon_N} = U$, $V^{\gamma_N}(u) = V^{\chi_N}(u) = V(u)$ при всех $u \in U$, $N = 1, 2, \ldots$ Тогда

$$J_*(\varepsilon_N, \gamma_N) = J_*(0, \chi_N) = J_*, \quad N = 1, 2, \ldots, \tag{59}$$

и условия (53), (54) тривиально выполняются. Определим отображение P_N: $X_N \to X$ так: $P_N([u]_N) = u_{N*}$ при всех $[u]_N \in X_N$, $N = 1, 2, \ldots$, где элемент $u_{N*} \in U = U^{\varepsilon_N}$ взят из (55). Ясно, что тогда включение (47) выполнено и, кроме того, согласно (55), (59) имеем $\overline{\lim\limits_{N\to\infty}}(J_* - J(P_N([u]_N), v_N)) \leqslant 0$ при всех $v_N \in V^{\chi_N}(u_{N*}) = V(u_{N*}) = V(P_N([u]_N)) = V^{\gamma_N}(P_N([u]_N))$, $[u]_N \in U_N$. Далее, зафиксируем любой элемент $[u]_N \in U_N$, по нему из (58) найдем соответствующий $[v]_{N*} \in V_N([u]_N)$ и определим отображение Q_N: $Y \to Y_N$ так: $Q_N(v) = [v]_{N*}$ при всех $v \in Y$, $N = 1, 2, \ldots$ Тогда справедливо включение (48) и, кроме того, согласно (58) имеем $\overline{\lim\limits_{N\to\infty}}(I_N([u]_N, Q_N(v_N)) - I_{N*}) \leqslant 0$ при всех $v_N \in V^{\gamma_N}(P_N([u]_N))$. Отсюда следует, что

$$\overline{\lim_{N\to\infty}}(I_N([u]_N, Q_N(v_N)) - J(P_N([u]_N), v_N)) \leqslant \overline{\lim_{N\to\infty}}(I_{N*} - J_*) +$$
$$+ \overline{\lim_{N\to\infty}}(I_N([u]_N, Q_N(v_N)) - I_{N*}) + \overline{\lim_{N\to\infty}}(J_* - J(P_N([u]_N), v_N)) \leqslant 0$$

при всех $v_N \in V^{\gamma_N}(P_N([u]_N))$, $[u]_N \in U_N$. Необходимость условия 1) доказана.

Далее, определим отображение $\overline{Q}_N\colon X \to X_N$ так: $\overline{Q}_N(u) = [u]_{N*}$ при всех $u \in X$, $N = 1, 2, \ldots$, где $[u]_{N*}$ взят из (56). Тогда включение (50), очевидно, выполнено и, кроме того, согласно (56) имеем $\overline{\lim\limits_{N\to\infty}}{}^{*}(I_{N*} - I_N(\overline{Q}_N(u_N), [v]_N)) \leqslant 0$ при всех $[v]_N \in V_N(\overline{Q}_N(u_N))$, $u_N \in U$.

Зафиксируем произвольный элемент $u_N \in U = U^{\varepsilon_N}$, по нему из (57) найдем соответствующий $v_{N*} \in V^{\gamma_N}(u_N) = V^{\chi_N}(u_N)$ и определим отображение $\overline{P}_N\colon Y_N \to Y$ так: $\overline{P}_N([v]_N) = v_{N*}$ при всех $[v]_N \in Y_N$, $N = 1, 2, \ldots$. Тогда справедливо включение (51) и, кроме того, согласно (57), (59) имеем $\lim\limits_{N\to\infty}(J(u_N, \overline{P}_N([v]_N)) - J_*) \leqslant 0$ при всех $[v]_N \in V_N(\overline{Q}_N(u_N))$. Отсюда следует, что

$$\overline{\lim_{N\to\infty}}(J(u_N, \overline{P}_N([v]_N)) - I_N(\overline{Q}_N(u_N), [v]_N)) \leqslant \overline{\lim_{N\to\infty}}(J_* - I_{N*}) + \overline{\lim_{N\to\infty}}(J(u_N, \overline{P}_N([v]_N)) - J_*) + \overline{\lim_{N\to\infty}}(I_{N*} - I_N(\overline{Q}_N(u_N), [v]_N)) \leqslant 0$$

при всех $[v]_N \in V_N(\overline{Q}_N(u_N))$, $u_N \in U$. Необходимость условия 2) также установлена.

Достаточность. Пусть выполнены условия 1)–3). Покажем, что тогда $\lim\limits_{N\to\infty} I_{N*} = J_*$. Возьмем $[u]_{N*} \in U_N$ из (56), положим $u_N = P_N([u]_{N*})$, а из (57) возьмем $v_{N*} \in V^{\gamma_N}(u_N) = V^{\gamma_N}(P_N([u]_{N*}))$ и обозначим $[v]_N = Q_N(v_{N*})$. Из (48) тогда имеем $[v]_N \in V_N([u]_{N*})$, $N = 1, 2, \ldots$. Пользуясь условиями (48), (49) при $[u]_N = [u]_{N*}$, $v_N = v_{N*}$, с учетом соотношений (54), (56), (57), получим

$$\overline{\lim_{N\to\infty}}(I_{N*} - J_*) \leqslant \overline{\lim_{N\to\infty}}(I_{N*} - I_N([u]_{N*}, [v]_N)) + \overline{\lim_{N\to\infty}}(I_N([u]_{N*}, Q_N(v_{N*})) - J(P_N([u]_{N*}), v_{N*})) + \overline{\lim_{N\to\infty}}(J(u_N, v_{N*}) - J_*(\varepsilon_N, \gamma_N)) + \overline{\lim_{N\to\infty}}(J_*(\varepsilon_N, \gamma_N) - J_*) \leqslant 0.$$

Далее, возьмем $u_{N*} \in U$ из (55), положим $[u]_N = \overline{Q}_N(u_{N*})$, а из (58) возьмем $[v]_{N*} \in V_N([u]_N) = V_N(\overline{Q}_N(u_{N*}))$ и обозначим $v_N = \overline{P}_N([v]_{N*})$. Из (51) тогда следует, что $v_N \in V^{\chi_N}(u_{N*})$, $N = 1, 2, \ldots$. Пользуясь условиями (51), (52) при $u_N = u_{N*}$, $[v]_N = [v]_{N*}$, с учетом соотношений (53), (55), (58) имеем

$$\overline{\lim_{N\to\infty}}(J_* - I_{N*}) \leqslant \overline{\lim_{N\to\infty}}(I_N([u]_N, [v]_{N*}) - I_{N*}) + \overline{\lim_{N\to\infty}}(J(u_{N*}, \overline{P}_N([v]_{N*})) - I_N(\overline{Q}_N(u_{N*}), [v]_{N*})) + \overline{\lim_{N\to\infty}}(J_*(0, \chi_N) - J(u_{N*}, v_N)) + \overline{\lim_{N\to\infty}}(J_* - J_*(0, \chi_N)) \leqslant 0.$$

Таким образом, $\lim\limits_{N\to\infty} I_{N*} = \overline{\lim\limits_{N\to\infty}} I_{N*} = J_*$, т. е. $\lim\limits_{N\to\infty} I_{N*} = J_*$. Теорема 4 доказана. □

Для иллюстрации теоремы 4 рассмотрим задачу (31) для случая, когда $J(u, v) = |x(T, u, v) - y|^2$, где $x(t, u, v)$, $t_0 \leqslant t \leqslant T$, — решение задачи (16), соответствующее управлениям $u \in X = L_2^r[t_0, T]$, $v \in Y = L_2^q[t_0, T]$.

$$W = \{u = u(t) \in L_2^r[t_0, T]\colon u(t) \in P \quad \text{почти всюду на} \quad [t_0, T]\}, \tag{60}$$

$$V = \{v = v(t) \in L_2^q[t_0, T]\colon v(t) \in Q \quad \text{почти всюду на} \quad [t_0, T]\}, \tag{61}$$

$$V(u) = \{v = v(t) \in V\colon x(t, u, v) \in G, \quad t_0 \leqslant t \leqslant T\}, \tag{62}$$

$$U = \{u \in W\colon V(u) \neq \varnothing\}; \tag{63}$$

здесь P, Q, G — заданные множества из евклидовых пространств E^r, E^q, E^n соответственно. Эту задачу кратко будем называть задачей (31), (15), (16), (60)–(63). Для аппроксимации этой задачи рассмотрим последовательность задач (32), где $I_N([u]_N, [v]_N) = |x_N([u]_N, [v]_N) - y|^2$, $x_i([u]_N, [v]_N)$, $i = 0, \ldots, N$ — решение задачи (21), соответствующее управлениям $[u]_N \in X_N = L_{2N}^r$, $[v]_N \in Y_N = L_{2N}^q$,

$$W_N = \{[u]_N = (u_0, \ldots, u_{N-1}) \in L_{2N}^r\colon u_i \in P, \ i = 0, \ldots, N-1\}, \tag{64}$$

$$V_N = \{[v]_N = (v_0, \ldots, v_{N-1}) \in L_{2N}^q\colon v_i \in Q, \ i = 0, \ldots, N-1\}, \tag{65}$$

$$V_N([u]_N) = \{[v]_N \in V_N\colon x_i([u]_N, [v]_N) \in G^{\mu_N}, \ i = 0, \ldots, N\}, \tag{66}$$

$$U_N = \{[u]_N \in W_N\colon \text{существует } [v]_N \in V_N \text{ такой, что } x_i([u]_N, [v]_N) \in G^{\xi_N}, \ i = 0, \ldots, N\}, \quad N = 1, 2, \ldots \tag{67}$$

Эти разностные задачи кратко будем называть задачами (32), (20), (21), (64)–(67).

Для исследования поведения аппроксимирующих разностных задач определим отображения P_N, Q_N, $\overline{P}_N$, $\overline{Q}_N$ так же, как в задачах (14)–(18), (19)–(23). Будем предполагать, что выполнены все условия теоремы 2 и, кроме того, G выпукло и замкнуто. Тогда справедливы включения $\overline{Q}_N(u) \in W_N$ при $u \in W$, $Q_N(v) \in V_N$ при $v \in V$, $P_N([u]_N) \in W$ при $[u]_N \in W_N$, $\overline{P}_N([v]_N) \in V$ при $[v]_N \in V_N$. Кроме того, нетрудно видеть, что здесь по-прежнему сохраняют силу оценки (24)–(30) с заменой в них U на W, U_N на W_N, где W, W_N взяты из (60), (64). Имеют место также неравенства

$$\sup_{[u]_N \in W_N} \sup_{[v]_N \in V_N} |x_i([u]_N, [v]_N) - x(t_i, P_N([u]_N), \overline{P}_N([v]_N))| \leqslant \delta_N, \tag{68}$$

$$\sup_{u \in W} \sup_{v \in V} \max_{0 \leqslant i \leqslant N} |x_i(\overline{Q}_N(u), Q_N(v)) - x(t_i, u, v)| \leqslant \delta_N, \tag{69}$$

где δ_N взято из (27), (28).

Для того чтобы задача (31), (15), (16), (60)–(63) имела смысл, естественно предполагать, что множество (63) непусто. Оказывается, что тогда при достаточно больших μ_N, ξ_N множества (66), (67) также непусты. А именно, выберем μ_N, ξ_N в (66) так, чтобы

$$\delta_N \leqslant \xi_N \leqslant \mu_N, \quad N = 1, 2, \ldots \tag{70}$$

Возьмем какие-либо $u \in U$, $v \in V(u)$. Тогда $x(t, u, v) \in G^\delta$, $t_0 \leqslant t \leqslant T$. Отсюда в силу (69), (70) имеем $x_i(\overline{Q}_N(u), Q_N(v)) \in G^N \subset G^{\xi_N}$, $i = 0, \ldots, N$. Это значит, что $Q_N(u) \in U_N$, т. е. $U_N \neq \varnothing$. Кроме того, из (70) следует, что $V_N([u]_N) \neq \varnothing$ при всех $[u]_N \in U_N$. Таким образом, задачи (32), (20), (21), (64)–(67) имеют смысл, и величины I_{N*}, $N = 1, 2, \ldots$, определены.

В дальнейшем будем предполагать, что выполнено условие: существуют число $\beta > 0$ и управления $u_0 \in W$, $v_0 \in V$ такие, что $x(t, u_0, v_0) \in G^{-\beta}$, $t_0 \leqslant t \leqslant T$. Введем множества

$$V^{\gamma_N}(u) = \{v \in V\colon x(t, u, v) \in G^{\gamma_N}, \quad t_0 \leqslant t \leqslant T\}, \tag{71}$$

$$U^{\varepsilon_N} = \{u \in U\colon V^{\varepsilon_N}(u) \neq \varnothing\}, \tag{72}$$

где

$$\varepsilon_N \geqslant \xi_N + \delta_N + C_1 d_N, \quad \gamma_N = M_2 \varepsilon_N / (\beta + \varepsilon_N), \quad N = 1, 2, \ldots \tag{73}$$

Здесь $M_2 = 2R(T - t_0)^{1/2} M_1$, R — диаметр множества P, а константа M_1 взята из неравенства (см. (1.13)) $\sup\limits_{v \in V} \sup\limits_{t_0 \leqslant t \leqslant T} |x(t, u, v) - x(t, w, v)| \leqslant M_1 \|u - w\|_{L_2}$.

Будем предполагать, что

$$\mu_N \geqslant \delta_N + \gamma_N, \quad \chi_N \geqslant \mu_N + \delta_N + C_1 d_N, \quad N = 1, 2, \ldots, \quad \lim_{N\to\infty} \chi_N = 0. \tag{74}$$

Проверим выполнение условий теоремы 4. Заметим, что $V(u) \subset V^{\varepsilon_N}(u)$ при всех $u \in U$. Это значит, что $U \subset U^{\varepsilon_N}$, т. е. $U^{\varepsilon_N} \neq \varnothing$. Непустота множеств $V^{\gamma_N}(u)$ при всех $u \in U^{\varepsilon_N}$ будет установлена ниже перед проверкой условия (54).

Зафиксируем $N \geqslant 1$ и $[u]_N \in U_N$. Согласно (67) это значит, что существует $[v]_N \in V_N$ такой, что $x_i([u]_N, [v]_N) \in G^{\xi_N}$, $i = 0, \ldots, N$. Отсюда с учетом соотношений (26), (68), (73) имеем $x(t, P_N([u]_N), \overline{P}_N([v]_N)) \in G^{\varepsilon_N}$, $t_0 \leqslant t \leqslant T$. Это значит, что $\overline{P}_N([v]_N) \in V^{\varepsilon_N}(P_N([u]_N))$ и $P_N([u]_N) \in U^{\varepsilon_N}$. Включение (47) установлено. Далее, возьмем любое $v \in V^{\gamma_N}(P_N([u]_N))$. Согласно (71), тогда $x(t, P_N([u]_N), v) \in G^{\gamma_N}$, $t_0 \leqslant t \leqslant T$. Из (27), (74) следует, что $x_i([u]_N, Q_N(v)) \in G^{\delta_N + \gamma_N} \subset G^{\mu_N}$, $i = 0, \ldots, N$. Отсюда и из (66) заключаем, что $Q_N(v) \in V_N([u]_N)$. Включение (48) доказано. Неравенство (49) является следствием неравенства (29). Условие 1) теоремы 4 выполнено.

С помощью соотношений (26), (28), (30), (60)–(67), (69)–(74) аналогично убеждаемся в справедливости условий 2) этой теоремы. В самом деле, возьмем произвольный элемент $u_N \in U$. Это значит, что существует $v \in V$ такой, что $x(t, u_N, v) \in G$, $t_0 \leqslant t \leqslant T$. Отсюда и из (69), (70) имеем $x_i(\overline{Q}_N(u_N), Q_N(v)) \in G^{\delta_N} \subset G^{\xi_N}$, $i = 0, \ldots, N$, т. е. $\overline{Q}_N(u_N) \in U_N$. Включение (50) установлено. Далее, возьмем произвольный элемент $[v]_N \in V_N(\overline{Q}_N(u_N))$. Согласно (66), это значит, что $x_i(\overline{Q}_N(u_N), [v]_N) \in G^{\mu_N}$, $i = 0, \ldots, N$. Из (26), (28), (74) следует, что $x(t, u_N, \overline{P}_N([v]_N)) \in G^{\chi_N}$, $t_0 \leqslant t \leqslant T$, т. е. $\overline{P}_N([v]_N) \in V^{\chi_N}(u_N)$. Включение (51) доказано. Неравенство (52) является следствием неравенства (30). Условие 2) теоремы 4 также выполнено.

Остается проверить выполнение условия 3). Сначала докажем неравенство (53). По определению величины J_* из (31) для любого $k \geqslant 1$ существует $u_k \in U$ такой, что $\inf\limits_{v \in V(u_k)} J(u_k, v) \geqslant J_* - 1/k$. Далее, по определению нижней грани найдутся элементы $v_{kN} \in V^{\chi_N}(u_k)$ такие, что $J(u_k, v_{kN}) \leqslant \inf\limits_{v \in V^{\chi_N}(u_k)} J(u_k, v) + 1/k \leqslant J_*(0, \chi_N) + 1/k$. Тогда

$$J_* - J_*(0, \chi_N) \leqslant \inf_{v \in V(u_k)} J(u_k, v) - J(u_k, v_{kN}) + 2/k, \tag{75}$$

$N, k = 1, 2, \ldots$ Выбирая при необходимости подпоследовательности, можем считать, что $\overline{\lim\limits_{N \to \infty}} J(u_k, v_{kN}) = \lim\limits_{N \to \infty} J(u_k, v_{kN})$, а $\{v_{kN}\}$ слабо сходится в $L_2^q[t_0, T]$ при $N \to \infty$ к некоторому элементу $v_k \in V$. Так как $v_{kN} \in V^{\chi_N}(u_k)$, то $x(t, u_k, v_{kN}) \in G^{\chi_N}$, $t_0 \leqslant t \leqslant T$, $N = 1, 2, \ldots$ Поскольку $x(t, u_k, v_{kN}) \to x(t, u_k, v_k)$ равномерно на $[t_0, T]$ при $N \to \infty$, а множество G замкнуто, то $x(t, u_k, v_k) \in G$, $t_0 \leqslant t \leqslant T$. Это значит, что $v_k \in V(u_k)$, и поэтому $\inf\limits_{v \in V(u_k)} J(u_k, v) \leqslant J(u_k, v_k)$. Тогда, учитывая, что $\lim\limits_{N \to \infty} J(u_k, v_{kN}) = J(u_k, v_k)$, из (79) при $N \to \infty$ имеем $\overline{\lim\limits_{N \to \infty}} (J_* - J_*(0, \chi_N)) \leqslant \inf\limits_{v \in V(u_k)} J(u_k, v) - J(u_k, v_k) + 2/k$ при всех $k = 1, 2, \ldots$ Наконец, устремляя здесь $k \to \infty$, получим $\overline{\lim\limits_{N \to \infty}} (J_* - J_*(0, \chi_N)) \leqslant 0$, что равносильно условию (53).

Заметим, что выше мы пользовались тем, что $V^{\gamma_N}(u) \neq \varnothing$ при всех $u \in U^{\varepsilon_N}$. Докажем это. Пусть $u \in U^{\varepsilon_N}$, $v \in V^{\varepsilon_N}(u)$. Положим $u_N = \beta_N u_0 + (1 - \beta_N)u$, $v_N = \beta_N v_0 + (1 - \beta_N)v$, где

$$0 < \beta_N = \varepsilon_N/(\beta + \varepsilon_N) < 1. \tag{76}$$

Так как $x(t, u, v) \in G^{\varepsilon_N}$, $x(t, u_0, v_0) \in G^{-\beta}$, то с учетом (76) имеем $x(t, u_N, v_N) \in G$. Но $|x(t, u, v_N) - x(t, u_N, v_N)| \leqslant M_1 \|u - u_N\|_{L_2} \leqslant M_2 \beta_N$, и поэтому с помощью (73), (76) получаем $x(t, u, v_N) \in G^{M_2 \beta_N} = G^{\gamma_N}$. Следовательно, $V^{\gamma_N}(u) \neq \varnothing$ при всех $u \in U^{\varepsilon_N}$. По определению $J_*(\varepsilon_N, \gamma_N)$ существует $u_{N*} \in U^{\varepsilon_N}$ такой, что $J(\varepsilon_N, \gamma_N) - 1/N \leqslant \inf\limits_{v \in V^{\gamma_N}(u_{N*})} J(u_{N*}, v)$. Возьмем некоторый элемент $w_N \in V^{\varepsilon_N}(u_{N*})$. Положим $u_N = \beta_N u_0 + (1 - \beta_N)u_{N*}$, $v_N = \beta_N v_0 + (1 - \beta_N)w_N$, где β_N взято из (76). Как и выше, получаем, что $x(t, u_N, v_N) \in G$, т. е. $V(u_N) \neq \varnothing$. Возьмем $v_{N*} \in V(u_N)$ таким, чтобы $\inf\limits_{v \in V(u_N)} J(u_N, v) + 1/N \geqslant J(u_N, v_{N*})$. Так как $|x(t, u_N, v_{N*}) - x(t, u_{N*}, v_{N*})| \leqslant M_2 \beta_N$ и $x(t, u_N, v_{N*}) \in G$, то $x(t, u_{N*}, v_{N*}) \in G^{\gamma_N}$, т. е. $v_{N*} \in V^{\gamma_N}(u_{N*})$. Таким образом,

$$J_*(\varepsilon_N, \gamma_N) \leqslant \inf_{v \in V^{\gamma_N}(u_{N*})} J(u_{N*}, v) + 1/N \leqslant J(u_{N*}, v_{N*}) + 1/N \leqslant J(u_N, v_{N*}) + M_3 \beta_N + 1/N \leqslant$$

$$\leqslant \inf_{v \in V(u_N)} J(u_N, v) + M_4 \varepsilon_N + 2/N \leqslant J_* + M_4 \varepsilon_N + 2/N, \quad N \geqslant 1.$$

Отсюда при $N \to \infty$ получаем условие (54). Таким образом, при сделанных выше предположениях согласно теореме 4 последовательность задач (32), (20), (21), (64)–(67) аппроксимирует задачу (31), (15), (16), (60)–(63) по функции.

Вопросы аппроксимации более сложных задач, содержащих кратные максимины с несвязанными и связанными множествами, исследовались в [2–4].

5. Заметим, что в настоящей книге мы обсуждали проблемы оптимизации (в основном, поиска минимума), подразумевая, что целевая функция у нас одна. Однако на практике чаще встречаются более сложные, так называемые многокритериальные задачи, когда целевых функций (критериев полезности) несколько, их аргументы между собой связаны, и нужно найти какой-то оптимум по этим критериям. Но что такое оптимум в таких задачах? Ведь критерии, как правило, противоречивы, отражают реально существующую конфликтную ситуацию, и ожидать, что все критерии будут достигать своего, скажем, минимума в одной и той же точке, не приходится. Понятие оптимума в многокритериальных задачах должно вводиться на принципиально новой основе, оно должно содержать в себе некий компромисс, как-то учитывать характер конфликта, который может быть весьма разнообразным (антагонистическим, коалиционным, иерархическим и т. п.). На сегодняшний день выработано несколько разумных принципов оптимальности в многокритериальных задачах (например, оптимальность по Парето, по Нэшу и т. д.), которые получили признание и находят широкие применения. В рамках данной книги мы не в состоянии даже мало-мальски останавливаться на возникающих здесь интересных и актуальных проблемах и отсылаем читателя к специальной литературе по исследованию операций, теории игр, теории принятия решений, теории и методов равновесного

программирования [30; 31; 182; 183; 194; 218–220; 240; 241; 244; 245; 255; 310–313; 373; 375; 410; 412–414; 423; 440; 466; 473; 482; 500; 511; 566; 567; 569–571; 580; 581; 589; 590–592; 608; 643; 647; 650; 663; 686; 749–751; 755; 762; 763; 794; 809; 810]. К сказанному добавим, что задачи оптимизации, составившие предмет настоящей книги, являются необходимой ступенькой к познанию многокритериальных задач, теория и методы которых пока еще переживают пору становления и, несомненно, будут бурно развиваться.

В заключение хочется сказать, что невозможно «объять необъятное» и создать более-менее полное пособие по вопросам оптимизации. Остается лишь процитировать следующие утешающие слова из книги [75]: «При выборе способа решения конкретной задачи всякое пособие играет роль лишь общего руководства, отталкиваясь от которого исследователь анализирует свои проблемы», надеясь, что настоящее пособие кому-то поможет в таком анализе.

Список литературы

1. Абрамов Л. М., Капустин В. Ф. Математическое программирование. — Л.: Изд-во ЛГУ, 1981.
2. Аваков Е. Р. Об условиях аппроксимации кратного максимина. — Вестник Московск. ун-та. Сер. вычислит. матем. и киберн., 1977, № 3, С. 37–43.
3. Аваков Е. Р. Об условиях аппроксимации максиминных задач со связанными множествами. — Ж. вычислит. матем. и матем. физики, 1978, **18**, № 3, С. 603–613.
4. Аваков Е. Р. Об условиях аппроксимации кратного максимина по связанным множествам. — Вестник Московск. ун-та. Сер. вычислит. матем. и киберн., 1979, № 2, С. 16–25.
5. Аваков Е. Р. Условия экстремума для гладких задач с ограничениями типа равенств // Ж. вычислит. матем. и матем. физики. — 1985, Т. 25, № 5, С. 680–693.
6. Аваков Е. Р. Необходимые условия минимума для нерегулярных задач в банаховых пространствах. Принцип максимума для анормальных задач оптимального управления. Труды матем. ин-та АН СССР. — 1988, Т. 185, С. 3–29.
7. Аваков Е. Р. Необходимые условия экстремума для гладких анормальных задач с ограничениями типа равенств и неравенств. Матем. заметки, 1989, Т. 45, № 6, С. 3–11.
8. Аваков Е. Р. Теоремы об оценках в окрестности особой точки отображения. Матем. заметки. 1990, Т. 47, вып. 5, С. 3–13.
9. Аваков Е. Р., Аграчев А. А., Арутюнов А. В. Множество уровня гладкого отображения в окрестности особой точки и нули квадратичного отображения. Матем. сб., 1991, Т. 182, № 8, С. 1091–1104.
10. Аввакумов С. Н., Киселев Ю. Н., Орлов М. В. Методы решения задач оптимального управления на основе принципа максимума Понтрягина. Труды матем. ин-та РАН, 1995, Т. 211, С. 3–31.
11. Авдонин С. А., Иванов С. А. Управляемость систем с распределенными параметрами и семейства экспонент. Киев: Изд-во Минвуза Украинской ССР, 1989.
12. Аграчев А. А. Топология квадратичных отображений и гессианы гладких отображений. Итоги науки и техники. Сер. Алгебра, Геометрия, Топология. — М.: ВИНИТИ. 1988, Т. 26, С. 85–124.
13. Акулич И. Л. Математическое программирование в примерах и задачах. — М.: Высшая школа, 1993.
14. Алексеев В. М., Тихомиров В. М., Фомин С. В. Оптимальное управление. — М.: Наука, 1979.
15. Алексеев В. М., Галеев Э. М., Тихомиров В. М. Сборник задач по оптимизации. Теория. Примеры. Задачи. — М.: Наука. 1984.
16. Алексеев О. Г. Комплексное применение методов дискретной оптимизации. — М.: Наука, 1987.
17. Алифанов О. М., Артюхин Е. А., Румянцев С. В. Экстремальные методы решения некорректных задач. М.: Наука, 1988.
18. Андронов В. Г., Белоусов В. Г. О сходимости по функционалу метода штрафных функций. Вестник Московск. ун-та. Серия 15. Вычисл. матем. и кибернетика, 1996, № 2, С. 59–61.
19. Андронов В. Г., Белоусов Е. Г. О слабой сходимости по аргументу метода штрафных функций. // Ж. вычисл. матем. и матем. физики. — 1997, Т. 37, № 4, С. 404–414.
20. Андронов В. Г., Белоусов Е. Г. О бержевой сходимости метода штрафных функций. // Ж. вычисл. матем. и матем. физики. — 1998, Т. 38, № 4, С. 575–591.
21. Анрион Р. Теория второй вариации и ее приложения в оптимальном управлении. — М.: Наука, 1979.
22. Антипин А. С. Об едином подходе к методам решения некорректных экстемальных задач. Вестник Московск. ун-та, серия 1, Математика и механика, 1973, № 2, С. 60–67.
23. Антипин А. С. Метод регуляризации в задачах выпуклого программирования. — Экономика и матем. методы, 1975, **11**, № 2, С. 336–342.
24. Антипин А. С. Методы нелинейного программирования, основанные на прямой и двойственной модификации функции Лагранжа. — М.: изд-во ВНИИСИ, 1979.
25. Антипин А. С. Непрерывные и итерационные процессы с оператором проектирования и типа проектирования. Сб. «Вопросы кибернетики. Вычислительные вопросы анализа больших систем». — М.: Изд-во АН СССР, 1989, С. 5–43.
26. Антипин А. С. Управляемые проксимальные дифференциальные системы для решения седловых задач. // Дифференц. уравнения, 1992, Т. 28, № 11, С.1846–1861.

27. Антипин А. С., Недич А., Ячимович М. Трехшаговый метод линеаризации для задач минимизации. Известия вузов, сер. матем., 1994, № 12, С. 3–7.
28. Антипин А. С., Васильев Ф. П. О непрерывном методе минимизации в пространствах с переменной метрикой. Известия вузов, сер. матем., 1995, № 12, С. 3–9.
29. Антипин А. С., Недич А. Непрерывный метод линеаризации второго порядка для задач выпуклого программирования. Вестник МГУ, сер. вычислит. матем. и кибернетики, 1996. № 2, С. 3–12.
30. Антипин А. С. Равновесное программирование: проксимальные методы. // Ж. вычисл. матем. и матем. физики, 1997, Т. 37, № 11, С. 1327–1339.
31. Антипин А. С. Методы решения вариационных неравенств со связанными ограничениями. // Ж. вычисл. матем. и матем. физики, 2000, Т. 40, № 9, С. 1291–1307.
32. Анциферов Е. Г., Ащепков Л. Т., Булатов В. П. Методы оптимизации и их приложения. Ч. 1. Математическое программирование. — Новосибирск: Наука, 1990.
33. Аоки М. Введение в методы оптимизации. — М.: Наука, 1977.
34. Аркин В. И., Кречетов Л. И. Стохастические множители Лагранжа в задачах управления и экономической динамики. В сборнике работ «Вероятностные проблемы управления в экономике», М.: Наука, 1977, С. 5–32.
35. Арман Ж.-Л. П. Приложения теории оптимального управления системами с распределенными параметрами к задачам оптимизации конструкций. — М.: Мир, 1977.
36. Арутюнов А. В., Тынянский Н. Т. К необходимым условиям локального минимума в теории оптимального управления. Докл. АН СССР, 1984, Т. 275, № 2, С. 268–272.
37. Арутюнов А. В., Тынянский Н. Т. О принципе максимума в задаче с фазовыми ограничениями. Изв. АН СССР. Сер. технич. кибернетика, 1984, № 4, С. 60–68.
38. Арутюнов А. В. К необходимым условиям оптимальности в задаче с фазовыми ограничениями. Докл. АН СССР, 1985, Т. 280, № 5, С. 1033–1037.
39. Арутюнов А. В. О необходимых условиях оптимальности в задаче с фазовыми ограничениями // ДАН СССР. — 1985. — Т. 280, № 5. — С. 1033–1037.
40. Арутюнов А. В. Возмущения экстремальных задач с ограничениями и необходимые условия оптимальности. Итоги науки и техники. ВИНИТИ. Сер. Математический анализ. — 1989, Т. 27, С. 147–235.
41. Арутюнов А. В. К теории принципа максимума в задачах оптимального управления с фазовыми ограничениями. Докл. АН СССР, 1989, Т. 304, № 1, С. 11–14.
42. Арутюнов А. В., Асеев С. М., Благодатских В. И. Необходимые условия первого порядка в задаче оптимального управления дифференциальным включением с фазовыми ограничениями. Матем. сб., 1993, Т. 184, № 6, С. 3–32.
43. Арутюнов А. В. Условие второго порядка в экстремальных задачах с конечномерным образом. 2-нормальные отображения. Изв. РАН, сер. матем., 1996, Т. 60, № 1, С. 37–62.
44. Арутюнов А. В. Условия экстремума. Анормальные и вырожденные задачи. — М.: Факториал, 1997.
45. Арутюнов А. В. Расширения и возмущения задач оптимального управления. Труды матем. ин-та РАН, 1998, Т. 220, С. 27-34.
46. Асеев С. М. Метод гладких аппроксимаций в теории необходимых условий оптимальности для дифференциальных включений. Изв. РАН, сер. матем., 1997, Т. 61, № 2, С. 3–26.
47. Асеев С. М. Задача оптимального управления для дифференциального включения с фазовым ограничением. Гладкие аппроксимации и необходимые условия оптимальности. Итоги науки и техники. Серия «Современная математика и ее приложения», Т. 64. М.: Изд-во ВИНИТИ, 1999, С. 57–81.
48. Астафьев Н. Н. Линейные неравенства и выпуклость. — М.: Наука, 1982.
49. Астафьев Н. Н. Бесконечные системы линейных неравенств в математическом программировании. — М.: Наука, 1991.
50. Афанасьев А. П., Дикусар В. В., Милютин А. А., Чуканов С. А. Необходимое условие в оптимальном управлении. — М.: Наука, 1990.
51. Ахмедов К. Т., Ахиев С. С. Необходимое условие оптимальности для некоторых задач теории оптимального управления. Докл. АН Азерб. ССР, 1972, Т. 28, № 5, С. 12–15.
52. Ашманов С. А. Линейное программирование. — М.: Наука. 1981.
53. Ашманов С. А. Введение в математическую экономику. — М.: Наука. 1984.
54. Ашманов С. А., Тимохов А. В. Теория оптимизации в задачах и упражнениях. — М.: Наука, 1991.
55. Ащепков Л. Т. Оптимальное управление линейными системами. — Иркутск: Изд-во ИГУ, 1982.

56. Ащепков Л. Т., Белов Б. И., Булатов В. П., Васильев О. В., Срочко В. А., Тарасенко Н. В. Методы решения задач математического программирования и оптимального управления. — Новосибирск: Наука, 1984.
57. Ащепков Л. Т. Оптимальное управление разрывными системами. — Новосибирск: Наука, 1987.
58. Ащепков Л. Т., Величенко В. В. Оптимальное управление. — Владивосток: Изд-во Дальневосточного ун-та, 1989.
59. Бабенко К. И. Основы численного анализа. — М.: Наука, 1986.
60. Баев А. В. О решении одной обратной задачи для волнового уравнения с помощью регуляризирующего алгоритма. Ж. вычисл. матем. и матем. физики, 1985, Т. 25, № 1, С. 140–146.
61. Базара М., Шетти К. Нелинейное программирование. Теория и алгоритмы. — М.: Мир, 1982.
62. Бакушинский А. Б., Гончарский А. В. Итеративные методы решения некорректных задач. — М.: Наука, 1989.
63. Бакушинский А. Б., Гончарский А. В. Некорректные задачи. Численные методы и приложения. М.: Изд-во Московск. ун-та, 1989.
64. Бакушинский А. Б. Итерационные методы для решения нелинейных операторных уравнений без свойства регулярности. Фундаментальная и прикладная матем., 1997, Т. 3, № 1.
65. Балакришнан А. Введение в теорию оптимизации в гильбертовом пространстве. — М.: Мир, 1974.
66. Банди Б. Методы оптимизации. Вводный курс. — М.: Радио и связь, 1988.
67. Баничук Н. В. Оптимизация форм упругих тел. — М.: Наука, 1980.
68. Баничук Н. В. Введение в оптимизацию конструкций. М.: Наука, 1986.
69. Баничук Н. В., Иванова С. Ю., Шаранюк А. В. Динамика конструкций. Анализ и оптимизация. М.: Наука, 1989.
70. Банк Б., Белоусов Е. Г., Мандель Р., Черемных Ю. Н., Широнин В. М. Математическая оптимизация: вопросы разрешимости и устойчивости. — М.: Изд-во МГУ, 1986.
71. Батищев Д. И. Методы оптимального проектирования. — М.: Радио и связь, 1984.
72. Батурин В. А., Урбанович Д. Е. Приближенные методы оптимального управления, основанные на принципе расширения. — Новосибирск: Наука, 1997.
73. Батухтин В. Д., Майборода Л. А. Оптимизация разрывных функций. — М.: Наука, 1984.
74. Бахвалов Н. С., Жидков Н. П., Кобельков Г. М. Численные методы. — М.: Наука. 1987.
75. Бахвалов Н. С., Лапин А. В., Чижонков Е. В. Численные методы в задачах и упражнениях. М.: Высшая школа, 2000.
76. Бейко И. В., Бублик Б. Н., Зинько П. Н. Методы и алгоритмы решения задач оптимизации. — Киев: Вища школа, 1983.
77. Беленький В. З., Волконский В. А., Иванков С. А., Поманский А. Б., Шапиро А. Д. Итеративные методы в теории игр и программировании. — М.: Наука, 1974.
78. Беллман Р. Динамическое программирование. — М.: Изд-во иностр. литер., 1960.
79. Беллман Р. Процессы регулирования с адаптацией. — М.: Наука, 1964.
80. Беллман Р., Дрейфус С. Прикладные задачи динамического программирования. — М.: Наука, 1965.
81. Беллман Р., Калаба Р. Динамическое программирование и современная теория управления. — М.: Наука, 1969.
82. Белолипецкий А. А., Рябов А. Ю. Асимптотические оценки решений задачи оптимального быстродействия вблизи точек излома изохронной поверхности. // Ж. вычисл. матем. и матем. физики, 1986, Т. 26, № 4, С. 521–535.
83. Белоусов Е. Г. Введение в выпуклый анализ и целочисленное программирование. — М.: Изд-во МГУ, 1977.
84. Белоусов Е. Г., Андронов В. Г. Разрешимость и устойчивость задач полиномиального программирования. — М.: Изд-во МГУ, 1993.
85. Белоусов Е. Г., Андронов В. Г. О сводимости общей задачи выпуклого программирования к задаче на безусловный экстремум. // Известия вузов, сер. матем., 1995, № 12, С.21–29.
86. Бенсусан А., Лионс Ж. — — Л. Импульсное управление и квазивариационные неравенства. М.: Наука, 1987.

87. Бердышев В. И. Устойчивость задачи минимизации при возмущении множества допустимых элементов. — Матем. сб., 1977, **103 (145)**, № 4 (8), С. 467–479.
88. Бердышев В. И. Непрерывность многозначного отображения, связанного с задачей минимизации функционала. // Изв. АН СССР, сер. матем., 1980, Т. 44, № 3, С. 483–509.
89. Березин И. С., Жидков Н. П. Методы вычислений. Т. 1. — М.: Наука, 1966. Т. 2. — Физматгиз, 1962.
90. Березнев В. А. Математические методы планирования производственной программы предприятий легкой промышленности. — М.: Легкая индустрия, 1980.
91. Березнев В. А., Карманов В. Г., Третьяков А. А. Устойчивые методы решения экстремальных задач с приближенной информацией. — М.: Изд-во Научного совета по комплексной проблеме «Кибернетика» АН СССР, 1987.
92. Береснев В. Л., Гимади Э. Х., Дементьев В. Т. Экстремальные задачи стандартизации. — Новосибирск: Наука, 1978.
93. Бертсекас Д., Шрив С. Стохастическое оптимальное управление: случай дискретного времени. — М.: Наука, 1985.
94. Бертсекас Д. Условная оптимизация и методы множителей Лагранжа. — М.: Радио и связь, 1987.
95. Бесов О. В., Ильин В. П., Никольский С. М. Интегральные представления функций и теоремы вложения. — М.: Наука, 1975.
96. Благодатских В. И. Задача управляемости для линейных систем. — Труды Матем. института АН СССР, 1977, **143**, С. 57–67.
97. Благодатских В. И. Линейная теория оптимального управления. — М.: Изд-во МГУ, 1978.
98. Благодатских В. И. Теория дифференциальных включений, ч. 1. — М.: Изд-во Московск. ун-та, 1979.
99. Благодатских В. И., Филиппов А. Ф. Дифференциальные включения и оптимальное управление. — Труды Матем. института АН СССР, 1985, Т. 169, С. 194–252.
100. Благодатских В. И., Григоренко Н. Л., Киселев Ю. Н. Практикум по оптимальному управлению. — М.: Изд-во МГУ, 1986.
101. Блисс Г. А. Лекции по вариационному исчислению. — М.: Изд-во иностр. литературы, 1950.
102. Бобылев Н. А., Климов В. С. Методы нелинейного анализа в негладкой оптимизации. — М.: Наука, 1992.
103. Бобылев Н. А., Коровин С. К. Топологические методы в вариационных задачах. М.: Изд-во РАЕН, 1997.
104. Бокмельдер Е. П., Дыхта В. А., Москаленко А. И., Овсянникова Н. А. Условия экстремума и конструктивные методы решения в задачах оптимизации гиперболических систем. Новосибирск: Наука, 1993.
105. Болтянский В. Г. Математические методы оптимального управления. — М.: Наука. 1969.
106. Болтянский В. Г. Оптимальное управление дискретными системами. — М.: Наука, 1973.
107. Борисович Ю. Г., Обуховский В. В. О задаче оптимизации для управляемых систем параболического типа. Труды Матем. института, РАН, 1995, Т. 211, С. 95–101.
108. Брайсон А., Хо Ю — Ши. Прикладная теория оптимального управления. — М.: Мир, 1972.
109. Брежнева О. А., Третьяков А. А. Новые методы решения существенно нелинейных проблем. М.: Вычислительный центр РАН, 2000.
110. Бублик Б. Н., Гаращенко Ф. Г., Кириченко Н. Ф. Структурно-параметрическая оптимизация и устойчивость динамики пучков. — Киев: Наукова думка, 1985.
111. Будак Б. М., Беркович Е. М., Соловьева Е. Н. О сходимости разностных аппроксимаций для задач оптимального управления. — Ж. вычислит. матем. и матем. физики, 1969, **9**, № 3, С. 522–547.
112. Будак Б. М., Васильев Ф. П. Приближенные методы решения задач оптимального управления (тезисы лекций), вып. 2. — М.: Изд-во Московск. ун-та, 1969.
113. Будак Б. М., Виньоли А., Гапоненко Ю. Л. Об одном способе регуляризации для непрерывного выпуклого функционала. — Ж. вычислит. матем. и матем. физики. 1969, **9**, № 5, С. 1046–1056.
114. Будак Б. М., Беркович Е. М. Об аппроксимации экстремальных задач, I, II. — Ж. вычислит. матем. и матем. физики, 1971, **11**, № 3, С. 580–596; № 4, С. 870–884.
115. Будак Б. М., Васильев Ф. П. Некоторые вычислительные аспекты задач оптимального управления. — М.: Изд-во МГУ, 1975.

116. Булавский В. А., Звягина Р. А., Яковлева М. А. Численные методы линейного программирования. — М.: Наука, 1977.
117. Булатов В. П. Методы погружения в задачах оптимизации. — Новосибирск: Наука. 1977.
118. Бурак Я. И., Зозуляк Ю. Д., Гера Б. В. Оптимизация переходных процессов в термоупругих оболочках. Киев: Наукова Думка, 1984.
119. Бурдуковская А. В., Васильев О. В. Об алгоритмах оптимизации в системах канонических гиперболических уравнений с частными производными. Иркутск: Изд-во Иркутск. ун-та, серия «Оптимизация и управление», вып. 2, 1998.
120. Бурковская В. Л., Макаров В. Л. О применимости метода сеток и метода прямых к решению одного класса задач теории оптимального управления. Ж. вычислит. матем. и матем. физики, 1983, Т. 23, № 4, С. 798–805.
121. Буслаев В. С. Вариационное исчисление. — Л.: Изд-во ЛГУ, 1980. — 288 с.
122. Бутковский А. Г. Теория оптимального управления системами с распределенными параметрами. — М.: Наука, 1965.
123. Бутковский А. Г. Методы управления системами с распределенными параметрами. — М.; Наука, 1975.
124. Бутковский А. Г. Управление системами с распределенными параметрами. — Автоматика и телемеханика, 1979, № 11, С. 16–65.
125. Бутковский А. Г., Пустыльников Л. М. Теория подвижного управления системами с распределенными параметрами. М.: Наука, 1980.
126. Бутковский А. Г., Самойленко Ю. И. Управление квантовомеханическими процессами. М.: Наука, 1984.
127. Бухгейм А. Л. Введение в теорию обратных задач. Новосибирск: Наука, 1988.
128. Вазан М. Стохастическая аппроксимация. — М.: Мир, 1972.
129. Вайникко Г. М. Анализ дискретизационных методов. Тарту: Изд-во Тартусск. ун-та, 1976.
130. Вайникко Г. М. Методы решения линейных некорректно поставленных задач в гильбертовых пространствах. Тарту: Изд-во Тартусск. ун-та, 1982.
131. Вайникко Г. М., Веретенников А. Ю. Итерационные процедуры в некорректных задачах. М.: Наука, 1986.
132. Варга Дж. Оптимальное управление дифференциальными и функциональными уравнениями. — М.: Наука, 1977.
133. Васильев Н. С. О численном решении экстремальных задач построения эллипсоидов и параллелепипедов. // Ж. вычислит. матем. и матем. физики, 1987, Т. 27, № 3, С. 340–348.
134. Васильев О. В., Срочко В. А., Терлецкий В. А. Методы оптимизации и их приложения. Часть 2. Оптимальное управление. Новосибирск: Наука, 1990.
135. Васильев О. В. Лекции по методам оптимизации. Иркутск: Изд-во Иркутск. ун-та, 1994.
136. Васильев О. В., Аргучинцев А. В. Методы оптимизации в задачах и упражнениях. М.: Физматлит, 1999.
137. Васильев Ф. П. Условия оптимальности для некоторых классов систем, не разрешенных относительно производной. — ДАН СССР, 1969, **184**, № 6, С. 1267–1270.
138. Васильев Ф. П. Об итерационных методах решения задач быстродействия, связанных с параболическими уравнениями. — Ж. вычислит. матем. и матем. физики, 1970, **10**, № 4, С. 942–957.
139. Васильев Ф. П., Иванов Р. П. Некоторые приближенные методы решения задач быстродействия в банаховых пространствах при наличии фазовых ограничений. — ДАН СССР, 1970, **195**, № 3, С. 526–529.
140. Васильев Ф. П. Лекции по методам решения экстремальных задач. — М.: Изд-во Московск. ун-та, 1974.
141. Васильев Ф. П. Численный метод решения задачи быстродействия при приближенном задании исходных данных. — Вестник Московск. ун-та. Сер. вычислит. матем. и киберн., 1977, № 3, С. 26–36.
142. Васильев Ф. П. О регуляризации некорректных экстремальных задач. — ДАН СССР, 1978, **241**, № 5, С. 1001–1004.
143. Васильев Ф. П. О методе нагруженных функционалов // Вестник МГУ. Сер. вычисл. матем. и киберн., 1978. — № 3. — С. 24 — 32.
144. Васильев Ф. П., Воронцов М. А., Литвинова О. А. Об оптимальном управлении процессом теплового самовоздействия. — Ж. вычислит. матем. и матем. физики, 1979, **19**, № 4, С. 1053–1058.

145. Васильев Ф. П. О регуляризации некорректных задач минимизации на множествах, заданных приближенно. — Ж. вычислит. матем. и матем. физики, 1980, **20**, № 1, С. 38–50.
146. Васильев Ф. П., Ковач М. О регуляризации некорректных экстремальных задач с использованием штрафных и барьерных функций. — Вестник Московск. ун-та. Сер. вычислит. матем. и киберн., 1980, № 2, 29–35.
147. Васильев Ф. П., Ячимович М. Д. Об итеративной регуляризации метода условного градиента и метода Ньютона при неточно заданных исходных данных. — ДАН СССР, 1980, **250**, № 2, С. 265–269.
148. Васильев Ф. П. Численные методы решения экстремальных задач. — М.: Наука, 1980 (1-е издание), 1988 (2-е издание).
149. Васильев Ф. П., Ячимович М. Д. Об итеративной регуляризации метода Ньютона. — Ж. вычислит. матем. и матем. физики, 1981, **21**, № 3, С. 775–778.
150. Васильев Ф. П., Хромова Л. Н., Ячимович М. Д. Итеративная регуляризация одного метода минимизации третьего порядка. — Вестник Московск. ун-та. Сер. вычислит. матем. и киберн., 1981, № 1, С. 31–36.
151. Васильев Ф. П. Методы решения экстремальных задач. М.: Наука, 1981.
152. Васильев Ф. П. Итеративная регуляризация разностных аппроксимаций одной задачи оптимального управления. Вестник Киевского ун-та, серия «Моделирование и оптимизация сложных систем». 1982, № 1, С. 40–45.
153. Васильев Ф. П., Аваков Е. Р. Разностная аппроксимация одной максиминной задачи оптимального управления с фазовыми ограничениями. Вестник Московск. ун-та, серия 15, Вычисл. матем. и киберн., 1982, № 2, С. 11–17.
154. Васильев Ф. П., Константинова Т. В. Об одном обобщении метода нагруженных функционалов // Вестник МГУ. Серия 15, Вычисл. матем. и киберн. — 1983, № 2. — С. 3–8.
155. Васильев Ф. П. Регуляризованный метод Стеффенсена с аппроксимацией обратного оператора. Вестник Московск. ун-та, серия 15, Вычисл. матем. и киберн., 1984, № 4, С. 3–7.
156. Васильев Ф. П., Ковач М. О регуляризации некорректно поставленных экстремальных задач при неточно заданных исходных данных. Computational mathematics, Banach Center Publication, Warszawa, 1984, V. 13, pp. 237–263.
157. Васильев Ф. П. Регуляризация некоторых методов минимизации высокого порядка при неточных исходных данных. Ж. вычисл. матем. и матем. физики, 1985, Т. 25, № 4, С. 492–499.
158. Васильев Ф. П. О регуляризации метода Ньютона при неточном задании исходных данных. Труды матем. ин-та АН СССР, 1985, Т. 167, С. 53–59.
159. Васильев Ф. П., Солодкая М. С., Ячимович М. Д. О регуляризованном методе линеаризации при наличии погрешностей в исходных данных. Вестник Московск. ун-та, серия 15, Вычисл. матем. и киберн., 1985, № 4, С. 3–8.
160. Васильев Ф. П. О методе невязки для решения неустойчивых задач минимизации. Вестник Московск. ун-та, серия 15, Вычисл. матем. и киберн., 1987, № 4, С. 6–10.
161. Васильев Ф. П. Применение негладких штрафных фуикций в методе регуляризации неустойчивых задач минимизации // Ж. вычислит. матем. и матем. физики. — 1987. — Т. 27, № 10. — С. 1444–1450.
162. Васильев Ф. П. Оценка скорости сходимости метода регуляризации А. Н. Тихонова для неустойчивых задач минимизации. Докл. АН СССР, 1988, Т. 299, № 4, С. 792–796.
163. Васильев Ф. П. О регуляризации неустойчивых задач минимизации. Труды матем. ин-та АН СССР, 1988, Т. 185, С. 60–65.
164. Васильев Ф. П., Ковач М., Фуллер Р. Об устойчивости нечеткого решения систем линейных алгебраических уравнений с нечеткими коэффициентами. Вестник Московск. ун-та, серия 15, Вычисл. матем. и киберн., 1989, № 1, С. 5–9.
165. Васильев Ф. П., Куржанский М. А. О методе квазирешений для неустойчивых задач минимизации с неточно заданными исходными данными. Вестник Московск. ун-та, серия 15, Вычисл. матем. и киберн., 1989, № 4, С. 13–18.
166. Васильев Ф. П., Ишмухаметов А. З., Потапов М. М. Обобщенный метод моментов в задачах оптимального управления. М.: Изд-во Московск. ун-та, 1989.
167. Васильев Ф. П. Методы регуляризации для неустойчивых задач минимизации, основанные на идее расширения множества. Вестник Московск. ун-та, серия 15, Вычисл. матем. и киберн., 1990, № 1, С. 3–16.
168. Васильев Ф. П., Обрадович О. Регуляризованный проксимальный метод для задач минимизации с неточными исходными данными. Ж. вычисл. матем. и матем. физики, 1993, Т. 33, № 2, С. 179–188.

169. Васильев Ф. П., Ячимович М. Метод стабилизации для решения лексикографических задач. Ж. вычисл. матем. и матем. физики, 1993, Т. 33, № 8, С. 1123–1134.
170. Васильев Ф. П., Куржанский М. А., Потапов М. М. Метод прямых в задачах граничного управления и наблюдения для уравнения колебаний струны. Вестник Московск. ун-та, серия 15, Вычисл. матем. и киберн., 1993, № 3, С. 8–15.
171. Васильев Ф. П., Недич А. Регуляризованный непрерывный метод проекции градиента третьего порядка. Дифференциальные уравнения, 1994, Т. 30, № 12, С. 2033–2042.
172. Васильев Ф. П., Недич А. Регуляризованный непрерывный метод проекции градиента второго порядка. Вестник Московск. ун-та, серия 15, Вычисл. матем. и киберн., 1994, № 2, С. 3–11.
173. Васильев Ф. П., Ковач М. Об оценке скорости сходимости методов регуляризации для неустойчивых задач минимизации. Numerical analysis and mathematical modelling. Banach Center Publication, Warszawa, 1994, V. 29, pp. 233–244.
174. Васильев Ф. П., Недич А., Ячимович М. Трехшаговый регуляризованный метод линеаризации для решения задач минимизации. Известия вузов, Математика, 1994, № 12, С. 1–8.
175. Васильев Ф. П. О двойственности в линейных задачах управления и наблюдения. Дифференциальные уравнения, 1995, Т. 31, № 11, С. 1893–1900.
176. Васильев Ф. П., Антипин А. С. О непрерывном методе минимизации с переменной метрикой. Известия вузов, Математика, 1995, № 12, С. 3–9.
177. Васильев Ф. П., Недич А., Ячимович М. Двухшаговый регуляризованный метод линеаризации для решения задач минимизации. Ж. вычисл. матем. и матем. физики, 1996, Т. 36, № 5, С. 9–19.
178. Васильев Ф. П., Антипин А. С., Амочкина Т. В. Регуляризованный непрерывный метод минимизации с переменной метрикой при неточно заданных исходных данных. Вестник Московск. ун-та, серия 15, Вычисл. матем. и киберн., 1996, № 4, С. 5–11.
179. Васильев Ф. П., Иваницкий А. Ю. Линейное программирование. М.: Факториал, 1998.
180. Васильев Ф. П. Критерии устойчивости общей задачи линейного программирования. Вестник Московск. ун-та, серия 15, Вычисл. матем. и киберн., 1998, № 2, С. 17–20.
181. Васильев Ф. П. К вопросу устойчивости методов регуляризации в линейном программировании. Вестник Московск. ун-та, серия 15, Вычисл. матем. и киберн., 1998, № 4, С. 19–23.
182. Васильев Ф. П., Антипин А. С. Метод стабилизации для решения задач равновесного программирования с неточно заданным множеством. Ж. вычисл. матем. и матем. физики, 1999, Т. 39, № 11, С. 1779–1786.
183. Васин А. А. Модели процессов с несколькими участниками. — М.: Изд-во МГУ, 1983.
184. Васин В. В., Агеев А. Л. Некорректные задачи с априорной информацией. — Екатеринбург: Наука, 1993.
185. Васин В. В. Методы итеративной регуляризации для некорректных задач. Известия вузов. Математика, 1995, N.11, С. 69–84.
186. Вахрамеев С. А. Замечание о выпуклости в гладких нелинейных системах. Итоги науки и техники. Серия «Современная математика и ее приложения», Т. 60. М.: Изд-во ВИНИТИ, 1999, С. 42–73.
187. Вигак В. М. Управление температурными напряжениями и перемещениями. Киев: Наукова думка, 1988.
188. Винокуров В. А. О понятии регуляризуемости разрывных отображений. Ж. вычисл. матем. и матем. физики, 1971, Т. 11, № 5, С. 1097–1112.
189. Винокуров В. А. Приближенный метод невязки в нерефлексивных пространствах. Ж. вычисл. матем. и матем. физики, 1972, Т. 12, № 1, С. 207–212.
190. Винокуров В. А. Два замечания о выборе параметра регуляризации. Ж. вычисл. матем. и матем. физики, 1972, Т. 12, № 2, С. 481–483.
191. Владимиров А. А., Нестеров Ю. Е., Чеканов Ю. Н. О равномерно выпуклых функционалах // Вестник МГУ. Серия 15, Вычисл. матем. и киберн. — 1978. № 3. — С. 12–23.
192. Воеводин В. В. Линейная алгебра. — М.: Наука, 1980.
193. Воробьев Н. Н. Числа Фибоначчи. — М.: Наука, 1978.
194. Воробьев Н. Н. Основы теории игр. Бескоалиционные игры. — М.: Наука, 1984.
195. Воронцов М. А., Корягин А. В., Шмальгаузен В. И. Управляемые оптические системы. М.: Наука, 1989.
196. Воронцов М. А., Железных Н. И., Потапов М. М. О градиентной процедуре внутрирезонаторного управления световыми пучками. Ж. вычислит. матем. и матем. физики, 1990, Т. 30, № 3, С. 449–456.

197. Вуколов Э. А., Ефимов А. В., Земсков В. Н., Каракулин А. Ф., Лесин В. В., Поспелов А. С., Терещенко А. М. Сборник задач по математике для втузов. Ч. 4. Методы оптимизации. Уравнения в частных производных. Интегральные уравнения. — М.: Наука, 1990.
198. Габасов Р. Ф., Кириллова Ф. М. Качественная теория оптимальных процессов. — М.: Наука, 1971.
199. Габасов Р. Ф., Кириллова Ф. М. Особые оптимальные управления. — М.: Наука, 1973.
200. Габасов Р. Ф., Кириллова Ф. М. Оптимизация линейных систем. — Минск: Изд-во БГУ, 1973.
201. Габасов Р. Ф., Кириллова Ф. М. Принцип максимума в теории оптимального управления. — Минск: Наука и техника, 1974.
202. Габасов Р. Ф., Кириллова Ф. М. Основы динамического программирования. — Минск: Изд-во БГУ, 1975.
203. Габасов Р. Ф., Кириллова Ф. М. Методы линейного программирования. — Минск: Изд-во БГУ. Часть 1, 1977, часть 2, 1978, часть 3, 1980.
204. Габасов Р. Ф., Кириллова Ф. М. Методы оптимизации. — Минск: Изд-во БГУ, 1981.
205. Габасов Р, Кириллова Ф. М., Тятюшкин А. И. Конструктивные методы оптимизации. Ч. 1. Линейные задачи. — Минск: Изд-во Белорусск. ун-та, 1983.
206. Габасов Р, Кириллова Ф. М. Конструктивные методы оптимизации. Ч. 2. Задачи управления. — Минск: Изд-во Белорусск. ун-та, 1984.
207. Габасов Р, Кириллова Ф. М., Костюкова О. И. Конструктивные методы оптимизации. Ч. 3. Сетевые задачи, 1986, Ч. 4. Выпуклые задачи, 1987, Минск: Изд-во Белорусск. ун-та.
208. Гаевский Х., Грегер К., Захариас К. Нелинейные операторные уравнения и операторные дифференциальные уравнения. — М.: Наука, 1978.
209. Галеев Э. М., Тихомиров В. М. Краткий курс теории экстремальных задач. — М.: Изд-во МГУ, 1989.
210. Галеев Э. М., Тихомиров В. М. Оптимизация: теория, примеры, задачи. — М.: Эдиториал УРСС, 2000.
211. Гамкрелидзе Р. В., Харатишвили Г. А. Экстремальные задачи в линейных топологических пространствах. Известия АН СССР, серия матем., 1969, Т. 33, № 4, С. 781–839.
212. Гамкрелидзе Р. В. Основы оптимального управления. — Тбилиси: Изд-во Тбилисского ун-та, 1977.
213. Гантмахер Ф. Р. Теории матриц. — М.: Наука, 1967.
214. Ганшин Г. С. Методы оптимизации и решение уравнений. — М.: Наука, 1987.
215. Гапоненко Ю. Л. Некорректные задачи на слабых компактах. М.: Изд-во Московск. ун-та, 1989.
216. Гасс С. Линейное программирование (методы и приложения). — М.: Физматгиз, 1961.
217. Гельфанд И. М., Фомин С. В. Вариационное исчисление. — М.: Физматгиз, 1961.
218. Гермейер Ю. Б. Введение в теорию исследования операций. — М.: Наука, 1971.
219. Гермейер Ю. Б. Игры с непротивоположными интересами. — М.: Наука, 1976.
220. Гермейер Ю. Б., Морозов В. В., Сухарев А. Г., Федоров В. В. Задачи по исследованию операций. — М.: Изд-во Московск. ун-та, 1979.
221. Гернет Н. Н. Об основной простейшей задаче вариационного исчисления. С. — Петербург, 1913.
222. Гилл Ф., Мюррей У., Райт М. Практическая оптимизация. — М.: Мир, 1985.
223. Гилязов С. Ф. Методы решения линейных некорректных задач. М.: Изд-во Московск. ун-та, 1987.
224. Гилязов С. Ф. Приближенное решение некорректных задач. М.: Изд-во Московск. ун-та, 1995.
225. Гирсанов И. В. Лекции по математической теории экстремальных задач. — М.: Изд-во МГУ, 1970.
226. Гладков Д. И. Оптимизация систем неградиентным случайным поиском. — М.: Энергоатомиздат, 1984.
227. Гловински Р., Лионс Ж. — Л., Тремольер Р. Численное исследование вариационных неравенств. — М.: Мир, 1979.

228. Голичев И. И. Метод функции от оператора и итерационные процессы в некоторых задачах оптимального управления. Известия АН СССР, серия матем., 1991, Т. 55, № 4, С. 815–837.
229. Голичев И. И. Итерационный метод решения некорректных граничных задач. Ж. вычисл. матем. и матем. физики, 1993, Т. 33, № 11, С. 1626–1637.
230. Гольдман Н. Л. Обратные задачи Стефана. Теория и методы решения, М.: Изд-во Московск. ун-та, 1999.
231. Гольштейн Е. Г., Юдин Д. Б. Новые направления в линейном программировании. — М.: Сов. радио 1966.
232. Гольштейн Е. Г., Юдин Д. Б. Задачи линейного программирования транспортного типа. — М.: Наука, 1969.
233. Гольштейн Е. Г. Теория двойственности в математическом программировании и ее приложения. — М.: Наука, 1971.
234. Гольштейн Е. Г., Третьяков Н. В. Модифицированные функции Лагранжа. Теория и методы оптимизации. — М.: Наука, 1989.
235. Гончарский А. В., Черепащук А. М., Ягола А. Г. Численные методы решения обратных задач астрофизики. — М.: Наука, 1978.
236. Гончарский А. В., Черепащук А. М., Ягола А. Г. Некорректные задачи астрофизики. М.: Наука, 1985.
237. Горбунов В. К. Методы редукции неустойчивых вычислительных задач. Фрунзе: Изд-во «Илим», 1984.
238. Горбунов В. К. Экстремальные задачи обработки результатов измерений. Фрунзе: Изд-во «Илим», 1990.
239. Горбунов В. К. Введение в теорию экстремума. Ульяновск: Изд-во Ульяновского гос. ун-та, 1999.
240. Горелик В. А., Кононенко А. Ф. Теоретико-игровые модели принятия решений в эколого-экономических системах — М.: Радио и связь, 1982.
241. Горелик В. А., Горелов М. А., Кононенко А. Ф. Анализ конфликтных ситуаций в системах управления. М.: Радио и связь, 1991.
242. Гребеников А. И. Метод сплайнов и решение некорректных задач теории приближений. М.: Изд-во Московск. ун-та, 1983.
243. Грешилов А. А. Прикладные задачи математического программирования. — М.: Изд-во Московск. технич. ун-та, 1990.
244. Григоренко Н. Л. Дифференциальные игры преследования несколькими объектами. — М.: Изд-во МГУ, 1983.
245. Григоренко Н. Л. Математические методы управления несколькими динамическими процессами. — М.: Изд-во МГУ, 1990.
246. Григорьев И. С., Григорьев К. Г., Петрикова Ю. Д. О наискорейших маневрах космического аппарата с реактивным двигателем большой ограниченной тяги в гравитационном поле в вакууме. Космические исследования, 2000, Т. 38, № 2, С. 171–192.
247. Григорьев К. Г., Григорьев И. С. Исследование оптимальных пространственных траекторий перелетов космического аппарата с реактивным двигателем большой ограниченной тяги между орбитами искусственных спутников Земли и Луны. Космические исследования, 1995, Т. 33, № 1, С. 52–75.
248. Григорьев К. Г., Заплетин М. П. О вертикальном старте в оптимизационных задачах ракетодинамики. Космические исследования, 1997, Т. 35, № 4, С. 363–377.
249. Гродзовский Г. Л., Иванов Ю. Н., Токарев В. В. Механика космического полета с малой тягой. — М.: Наука, 1966.
250. Гроссман К., Каплан А. А. Нелинейное программирование на основе безусловной оптимизации. — Новосибирск: Наука, 1981.
251. Гупал А. М. Стохастические методы решения негладких экстремальных задач. — Киев: Наукова думка, 1979.
252. Гуревич Т. Ф., Лущук В. О. Сборник задач по математическому программированию. — М.: Колос, 1977.
253. Гурман В. И. Вырожденные задачи оптимального управления. — М.: Наука, 1977.
254. Гурман В. И. Принцип расширения в задачах управления. — М.: Физматлит, 1997.
255. Давыдов Э. Г. Исследование операций. — М.: Высшая школа, 1990.
256. Дамбраускас А. П. Симплексный поиск. — М.: Энергия, 1979.
257. Данилин Ю. М., Пиявский С. А. Об одном алгоритме отыскания абсолютного минимума. Сб. работ «Теория оптимальных решений». Вып. 2. — Киев: Институт кибернетики, АН УССР, 1967, С. 25–37.

258. Данфорд Н., Шварц Дж. Т. Линейные операторы. Общая теория. — М.: ИЛ, 1962.
259. Данциг Дж. Линейное программирование, его применения и обобщения. — М.: Прогресс, 1966.
260. Даффин Р., Питерсон Э., Зенер К. Геометрическое программирование. — М.: Мир, 1972.
261. Дементьев В. Т., Ерзин А. И., Ларин Р. М., Шамордин Ю. В. Задачи оптимизации иерархических структур. Новосибирск: Изд-во Новосибирск. ун-та, 1996.
262. Демиденко Е. З. Оптимизация и регрессия. — М.: Наука, 1989.
263. Демьянов В. Ф., Малоземов В. Н. Введение в минимакс. — М.: Наука, 1972.
264. Демьянов В. Ф., Васильев Л. В. Недифференцируемая оптимизация. — М.: Наука, 1981.
265. Демьянов В. Ф., Рубинов А. М. Основы негладкого анализа и квазидифференциальное исчисление. — М.: Наука, 1990.
266. Демьянов В. Ф. Точные штрафные функции в задачах негладкой оптимизации. Вестник С. — Петербургск. ун-та. Серия 1, Матем., механика, астрономия, 1994, вып. 4 (№ 22), С. 21–27.
267. Демьянов В. Ф. Условия экстремума и вариационные задачи. Санкт-Петербург, Изд-во С.-П. ун-та, 2000.
268. Денисов А. М., Лукшин А. В. Математические модели однокомпонентной динамики сорбции. М.: Изд-во Московск. ун-та, 1989.
269. Денисов А. М. Введение в теорию обратных задач. — М.: Изд-во МГУ, 1994.
270. Денисов А. М. Единственность определения нелинейного коэффициента системы уравнений в частных производных в малом и в целом. Докл. АН РАН, 1994, Т. 338, № 4, С. 444–447.
271. Денисов Д. В. Метод итеративной регуляризации в задачах условной минимизации. — Ж. вычислит. матем. и матем. физики, 1978, **18**, № 6, С. 1405–1415.
272. Деннис Дж., Шнабель Р. Численные методы безусловной оптимизации и решения нелинейных уравнений. — М.: Мир, 1988.
273. Дикин И. И., Зоркальцев В. И. Итеративное решение задач математического программирования. — Новосибирск: Наука, 1980.
274. Дикусар В. В., Милютин А. А. Качественные и численные методы в принципе максимума. — М.: Наука, 1989.
275. Дикусар В. В. Регуляризация вырожденной задачи оптимального управления. Дифференциальные уравнения, 1998, Т. 34, № 11, С. 1856–1865.
276. Дмитрук А. В. Принцип максимума для общей задачи оптимального управления с фазовыми и регулярными смешанными ограничениями. Сб. «Оптимальность управляемых динамических систем», вып. 14. — М.: ВНИИСИ, 1990. — С. 26–42.
277. Дончев А. Системы оптимального управления. Возмущения, приближения и анализ чувствительности. — М.: Мир, 1987.
278. Дубовицкий А. Я., Милютин А. А. Задачи на экстремум при наличии ограничений. // Ж. вычисл. матем. и матем. физики, 1965, Т. 5, № 3, С.395–453.
279. Дубовицкий А. Я., Милютин А. А. Теория принципа максимума. // Методы теории экстремальных задач в экономике. — М.: Наука, 1981, С. 138–177.
280. Дубовицкий А. Я., Дубовицкий В. А. Необходимые условия сильного минимума в задачах оптимального управления с вырождением концевых и фазовых ограничений. Успехи матем. наук, 1985, Т. 40, № 2, С. 175–176.
281. Дудов С. И. Дифференцируемость по направлениям функции расстояния. Матем. сб., 1995, Т. 186, № 3, С. 29–52.
282. Дыхта В. А. Вариационный принцип максимума и квадратичные условия оптимальности импульсных и особых процессов. Сибирский матем. журн., 1994, Т. 35, № 1, С. 70–82.
283. Дыхта В. А., Самсонюк О. Н. Оптимальное импульсное управление с приложениями. М.: Физматлит, 2000.
284. Дьяченко М. И., Ульянов П. Л. Мера и интеграл. М.: Факториал, 1998.
285. Дюво Г., Лионс Ж.-Л. Неравенства в механике и физике. М.: Наука, 1980.
286. Евтушенко Ю. Г. Методы решения экстремальных задач и их применение в системах оптимизации. — М.: Наука, 1982.
287. Егоров А. И. Оптимальное управление тепловыми и диффузионными процессами. — М.: Наука, 1978.
288. Егоров А. И. Оптимальное управление линейными системами. Киев: Выща школа, 1998.

289. Егоров Ю. В. Некоторые задачи теории оптимального управления. — Ж. вычислит. матем. и матем. физики, 1963, **3**, № 5, С. 887–904.
290. Егоров Ю. В. Необходимые условия оптимальности в банаховых пространствах. — Матем. сборник, 1964, **64 (106)**, № 1, С. 79–101.
291. Елкин В. И. Редукция нелинейных управляемых систем: дифференциально-геометрический подход. М.: Физматлит, 1997.
292. Емеличев В. А., Комлик В. И. Метод построения последовательности планов для решения задач дискретной оптимизации. — М.: Наука, 1981.
293. Емельянов С. В., Коровин С. К. Новые типы обратной связи. Управление при неопределенности. — М.: Наука, Физматлит 1997.
294. Еремин И. И. О методе штрафов в выпуклом программировании. Кибернетика, 1967, № 4, С. 63–67.
295. Еремин И. И., Астафьев Н. Н. Введение в теорию линейного и выпуклого программирования. — М.: Наука, 1976.
296. Еремин И. И., Мазуров В. Д. Нестационарные процессы математического программирования. — М.: Наука, 1979.
297. Еремин И. И., Мазуров В. Д., Астафьев Н. Н. Несобственные задачи линейного и выпуклого программирования. — М.: Наука, 1983.
298. Еремин И. И. Противоречивые модели оптимального планирования. — М.: Наука, 1988.
299. Еремин И. И. Теория линейной оптимизации. — Екатеринбург: Изд-во Уральского отделения РАН, 1998.
300. Еремин Ю. А., Свешников А. Г. Задачи распознавания и синтеза в теории дифракции. Ж. вычислит. матем. и матем. физики, 1992, Т. 32, № 10, С. 1594–1607.
301. Ермаков С. М., Жиглявский А. А. Математическая теория оптимального эксперимента. — М.: Наука, 1987.
302. Ермольев Ю. М. Методы стохастического программирования. — М.: Наука, 1976.
303. Ермольев Ю. М., Гуленко В. П., Царенко Т. И. Конечно-разностный метод в задачах оптимального управления. — Киев: Наукова думка, 1978.
304. Ермольев Ю. М., Ляшко И. И., Михалевич В. С., Тюптя В. И. Математические методы исследования операций. — Киев: Вища школа, 1979.
305. Ермольев Ю. М., Ястремский А. И. Стохастические модели и методы в экономическом планировании. — М.: Наука, 1979.
306. Жадан В. Г. Об одном классе итеративных методов решения задач выпуклого программирования // Ж. вычислит. матем. и матем. физики, 1984, Т. 24, № 5, С. 665–676.
307. Жданов В. А. О методе покоординатного спуска. // Матем. заметки, 1977, Т. 22, вып. 1, С. 137–142.
308. Жиглявский А. А. Математическая теория глобального случайного поиска. — Л.: Изд-во ЛГУ, 1985.
309. Жиглявский А. А., Жилинскас А. Г. Методы поиска глобального экстремума. — М.: Наука, 1991.
310. Жуковский В. И., Чикрий А. А. Линейно-квадратичные дифференциальные игры. — Киев: Наукова думка, 1994.
311. Жуковский В. И., Салуквадзе М. Е. Оптимизация гарантий в многокритериальных задачах управления. — Тбилиси, Мецниереба, 1996.
312. Жуковский В. И. Введение в дифференциальные игры при неопределенности. Ч. 1,2. — М.: Изд-во международного НИИ проблем управления, 1997.
313. Жуковский В. И. Кооперативные игры при неопределенности и их приложения. М.: Эдиториал УРСС, 1999.
314. Заботин Я. И., Кораблев А. И., Хабибуллин Р. Ф. Условия экстремума функционала при наличии ограничений. — Кибернетика, 1973, № 6, С. 65–70.
315. Заботин Я. И., Кораблев А. И. Псевдовыпуклые функционалы и их экстремальные свойства. — Известия вузов. Сер. матем., 1974, № 4 (143), С. 27–31.
316. Заботин Я. И. Минимаксный метод решения задачи математического программирования. — Известия вузов. Сер. матем., 1975, № 6 (157), С. 36–43.
317. Заботин Я. И. Лекции по линейному программированию. — Казань: Изд-во Казанск. ун-та, 1985.
318. Завриев С. К. Стохастические градиентные методы решения минимаксных задач. — М.: Изд-во МГУ, 1984.
319. Зангвилл У. И. Нелинейное программирование. — М.: Советское радио, 1973.
320. Заславский Ю. Л. Сборник задач по линейному программированию. — М.: Наука, 1969.

321. Зеликин М. И. Оптимальное управление и вариационное исчисление. — М.: Изд-во МГУ, 1985.
322. Зеликин М. И., Борисов В. Ф. Режимы учащающихся переключений в задачах оптимального управления. Труды Матем. института им. В. А. Стеклова РАН, 1991, Т. 197, С. 85–167.
323. Зеликин М. И. Нерегулярность оптимального управления в регулярных экстремальных задачах. Фундаментальная и прикладная математика, 1995, Т. 1, № 2, С. 399–408.
324. Зеликин М. И. Однородные пространства и уравнение Риккати в вариационном исчислении. М.: Факториал, 1998.
325. Зеликина Л. Ф. Многомерный синтез и теоремы о магистрали в задачах оптимального управления. В сборнике работ «Вероятностные проблемы управления в экономике», М.: Наука, 1977, С. 33–114.
326. Зойтендейк Г. Методы возможных направлений. — М.: Изд-во иностран. литер., 1963.
327. Зорич В. А. Математический анализ. — М.: Фазис, ч. I, 1997, ч. II, 1998.
328. Зубов В. И. Лекции по теории управления. — М.: Наука, 1975.
329. Зубов В. И. Динамика управляемых систем. — М.: Высшая школа, 1982.
330. Зуховицкий С. И., Авдеева Л. И. Линейное и выпуклое программирование. — М.: Наука, 1967.
331. Иванов А. П., Кирин Н. Е. Сопряженные задачи теории управления. — .: Изд-во Ленингр. ун-та, 1988.
332. Иванов В. А., Фалдин Н. В. Теория оптимальных систем автоматического управления. — М.: Наука, 1981.
333. Иванов В. В., Березовский А. И., Задирака В.К., Здоренко Л. Д., Лепеха Н. П. Методы алгоритмизации непрерывных производственных процессов. — М.: Наука, 1975.
334. Иванов В. К., Васин В. В., Танана В. П. Теория линейных некорректных задач и ее приложения. — М.: Наука, 1978.
335. Иванов В. К., Мельникова И. В., Филинков А. И. Дифференциально-операторные уравнения и некорректные задачи. М.: Наука, Физматлит, 1995.
336. Иванов Г. Е., Половинкин Е. С. О сильно выпуклых линейных дифференциальных играх. Дифференциальные уравнения, 1995, Т. 31, № 10, С. 1641–1648.
337. Иванов Р. П. Об одном критерии оптимальности и связанном с ним итерационном методе решения задачи быстродействия. — Ж. вычислит. матем. и матем. физики, 1971, **11**, № 3, С. 597–610.
338. Иванов Р. П. Об одном итерационном методе решения задачи быстродействия. — Ж. вычислит. матем. и матем. физики, 1971, **11**, № 4, С. 1031–1037.
339. Иванович Л. Д. Разностная аппроксимация и регуляризация задачи об оптимальном нагреве стержня. Вестник Московск. ун-та, Серия 15. Вычислит. матем. и кибернетика, 1982, № 3, С. 10–15.
340. Иванович Л. Д. Разностная аппроксимация и регуляризация максиминной задачи о нагреве стержня. Вестник Московск. ун-та, Серия 15. Вычислит. матем. и кибернетика, 1984, № 2, С. 20–23.
341. Ижуткин В. С., Кокурин М. Ю. О гибридном методе нелинейного программирования, использующем криволинейный спуск // Известия вузов. Сер. матем. — 1986, № 2. — С. 61–64.
342. Ижуткин В. С., Кокурин М. Ю. Методы приведенных направлений с допустимыми точками для задачи нелинейного программирования. Ж. вычисл. матем. и матем. физики, 1990, Т. 30, № 2, С. 217–230.
343. Ижуткин В. С., Петропавловский М. В. Методы приведенных направлений на основе дифференцируемой штрафной функции для задачи нелинейного программирования. Известия вузов. Математика, 1994, № 12, С. 50–59.
344. Ижуткин В. С., Петропавловский М. В. Методы приведенных направлений на основе модифицированной функции Лагранжа для задачи нелинейного программирования. Известия вузов. Математика, 1995, № 12, С. 33–42.
345. Ижуткин В. С., Петропавловский М. В., Блинов А. В. Метод центров и барьерных функций с использованием приведенных направлений для задачи нелинейного программирования. Известия вузов, математика, 1996, № 2, С. 30–41.
346. Измаилов А. Ф. Условия оптимальности для вырожденных экстремальных задач с ограничениями типа неравенств. Ж. вычисл. матем. и матем. физики, 1994, Т. 34, № 6, С. 837–854.
347. Измаилов А. Ф., Третьяков А. А. Факторанализ нелинейных отображений. — М.: Наука, 1994.

348. Измаилов А. Ф., Третьяков А. А. 2-регулярные решения нелинейных задач. Теория и численные методы. — М.: Физматлит, 1999.
349. Икрамов Х. Д. Задачник по линейной алгебре. М.: Наука, 1975.
350. Ильин В. А., Садовничий В. А., Сендов Бл. Х. Математический анализ. Начальный курс. — М.: Изд-во МГУ, 1985.
351. Ильин В. А., Ким Г. Д. Линейная алгебра и аналитическая геометрия. — М: Изд-во Московск. ун-та, 1998.
352. Ильин В. А., Позняк Э. Г. Основы математического анализа. — М.: Физматлит, ч. I, 1998, ч. II, 1998.
353. Ильин В. А., Позняк Э. Г. Линейная алгебра. — М.: Физматлит, 1999.
354. Ильин В. А. Волновое уравнение с граничным управлением на двух концах за произвольный промежуток времени. Дифференциальные уравнения, 1999, Т. 35, № 11, С. 1517–1534.
355. Ильин В. А. Волновое уравнение с граничным управлением на одном конце при закрепленном втором конце. Дифференциальные уравнения, 1999, Т. 35, № 12, С. 1640–1659.
356. Интрилигатор М. Математические методы оптимизации и экономическая теория. — М.: Прогресс, 1975.
357. Иосида К. Функциональный анализ. М.: Мир, 1967.
358. Иоффе А. Д., Тихомиров В. М. Теория экстремальных задач. — М.: Наука, 1974.
359. Ишмухаметов А. З., Потапов М. М. О согласовании параметра регуляризации с шагом разностной сетки. Вестник Московск. ун-та, Серия 15. Вычислит. матем. и кибернетика, 1980, № 3, С. 66–68.
360. Ишмухаметов А. З. Условия аппроксимации и устойчивости в задачах оптимального управления гиперболическими системами. Ж. вычисл. матем. и матем. физики, 1994, Т. 34, № 1, С. 12–28.
361. Ишмухаметов А. З. Методы решения задач оптимизации. М.: Изд-во МЭИ, 1998.
362. Ишмухаметов А. З., Юлина А. В. Аппроксимация квадратичной задачи оптимального управления параболической системой. Вестник МЭИ, Математика, 1998, № 6, С. 73–84.
363. Ишмухаметов А. З. Вопросы устойчивости и аппроксимации задач оптимального управления. М.: Изд-во ВЦ РАН, 2000.
364. Йованович М. В. Замечание о сильно выпуклых и квазивыпуклых функциях. Матем. заметки, 1996, Т. 60, вып. 5, С. 778–779.
365. Кабанихин С. И. Проекционно-разностные методы определения коэффициентов гиперболических уравнений. Новосибирск: Наука, 1988.
366. Казимиров В. И., Плотников В. И., Старобинец И. М. Абстрактная схема метода вариаций и необходимые условия экстремума // Изв. АН СССР. Сер. матем. — 1985. — Т. 49, № 1. — С. 141–159.
367. Калашников А. Л. Порядковая регуляризация некорректной задачи оптимального управления. — В сб.: Дифференциальные и интегральные уравнения, вып. 2. Горький: Изд-во Горьковск. ун-та, 1978, С. 124–129.
368. Калихман И. Л. Сборник задач по математическому программированию. — М.: Высшая школа, 1975.
369. Калихман И. Л., Войтенко М. А. Динамическое программирование в примерах и задачах. — М.: Высшая школа, 1979.
370. Канторович Л. В. Экономический расчет наилучшего использования ресурсов. — М.: Изд-во АН СССР, 1960.
371. Канторович Л. В., Акилов Г. П. Функциональный анализ. — М.: Наука, 1977.
372. Капустин В. Ф. Практические занятия по курсу математического программирования. — Л.: Изд-во ЛГУ, 1976.
373. Карлин С. Математические методы в теории игр, программировании и экономике. — М.: Мир, 1964.
374. Карманов В. Г. Математическое программирование. — М.: Физматлит, 2000.
375. Карманов В. Г., Федоров В. В. Моделирование в исследовании операций. — М.: Твема, 1996.
376. Карташев А. П., Рождественский Б. Л. Обыкновенные дифференциальные уравнения и основы вариационного исчисления. — М.: Наука, 1986.
377. Катковник В. Я. Линейные оценки и стохастические задачи оптимизации. — М.: Наука, 1976.
378. Киндерлерер Д., Стампаккья Г. Введение в вариационные неравенства и их приложения. М.: Мир, 1983.

379. Кирин Н. Е. Методы последовательных оценок в задачах оптимизации управляемых систем. — Л.: Изд-во ЛГУ, 1975.
380. Кирин Н. Е., Морозкин Н. Д. Численные приближения экстремалей управляемых динамических систем. — Уфа: Изд-во Башкирск. ун-та, 1989.
381. Кирин Н. Е. Методы оценивания и управления в динамических системах. — Санкт-Петербург, Изд-во С.-Петербургск. ун-та, 1993.
382. Киселев Ю. Н. Линейная теория быстродействия с возмущениями. — М.: Изд-во МГУ, 1986.
383. Киселев Ю. Н. Оптимальное управление. — М.: Изд-во Московск. ун-та, 1988.
384. Киселев Ю. Н. Быстросходящиеся алгоритмы решения линейной задачи быстродействия. Кибернетика. Киев, 1990, № 6, С. 47–57, 62.
385. Киселев Ю. Н. Построение точных решений для нелинейной задачи быстродействия специального вида. Фундаментальная и прикладная математика. 1997, Т. 3, № 3, С. 847–868.
386. Кларк Ф. Оптимизация и негладкий анализ. — М.: Наука, 1988.
387. Коваленко А. Г. Элементы выпуклого векторного программирования. — Куйбышев: Изд-во Куйбышевск. ун-та, 1990.
388. Ковалев М. М. Дискретная оптимизация. — Минск: Изд-во БГУ, 1977.
389. Ковач М. Непрерывный аналог итеративной регуляризации градиентного типа. — Вестник Московск. ун-та. Серия 15, Вычислит. матем. и киберн., 1979, № 3, С. 36–42.
390. Ковач М. О сходимости метода обобщенных барьерных функций. Вестник Московск. ун-та. Серия 15. Вычисл. матем. и кибернетика, 1981, № 1, С. 40–45.
391. Кокурин М. Ю. Операторная регуляризация и исследование нелинейных монотонных задач. Йошкар-Ола, Изд-во Марийского ун-та, 1998.
392. Коллатц Л., Крабс В. Теория приближений. Чебышевские приближения и их приложения. — М.: Наука, 1978.
393. Колмогоров А. Н., Фомин С. В. Элементы теории функций и функционального анализа. — М.: Наука, 1976.
394. Колпакова Э. В., Колпаков В. И. Восстановление математических объектов по неполно заданной информации. Саратов: Изд-во Саратовск. технич. ун-та, 1995.
395. Комков В. Теория оптимального управления демпфированием колебаний простых упругих систем. — М.: Мир, 1975.
396. Коннов И. В. Методы недифференцируемой оптимизации. — Казань, Изд-во Казанского ун-та, 1993.
397. Коннов И. В. Методы решения конечномерных вариационных неравенств. Казань: Изд-во «ДАС», 1998.
398. Корбут А. А., Финкельштейн Ю. Ю. Дискретное программирование. — М.: Наука, 1969.
399. Корнейчук Н. П. Экстремальные задачи теории приближения. — М.: Наука, 1976.
400. Коробов В. И., Скляр Г. М. Мин-проблема моментов Маркова и быстродействие. Сибирский матем. журнал, 1991, Т. 32, № 1, С. 60–71.
401. Коростелев А. П. Стохастические рекуррентные процедуры (локальные свойства). — М.: Наука, 1984.
402. Короткий А. И., Осипов Ю. С. Аппроксимация в задачах позиционного управления параболическими системами. Прикладн. матем. и механика, 1978, **42**, № 4, С. 599–605.
403. Короткий А. И. Восстановление управлений и параметров динамических систем при неполной информации. Известия вузов. Математика, 1998, N.11, С. 47–55.
404. Костоусова Е. К. О параллельном алгоритме решения задачи наблюдения для одномерного волнового уравнения. Сб. работ «Алгоритмы и программные средства параллельных вычислений». Екатеринбург, Изд-во УрО РАН, 1995, С. 101–114.
405. Кочиков И. В., Курамшина Г. М., Пентин Ю. А., Ягола А. Г. Обратные задачи колебательной спектроскопии. М.: Изд-во Московск. ун-та, 1993.
406. Коша А. Вариационное исчисление. — М.: Высшая школа, 1983.
407. Кравчук А. С. Вариационные и квазивариационные неравенства в механике. М.: МГАПИ, 1997.
408. Крайко А. Н. Вариационные задачи газовой динамики. — М.: Наука, 1979.
409. Краснов М. Л., Макаренко Г. И., Киселев А. И. Вариационное исчисление. Задачи и упражнения. — М.: Наука, 1973.
410. Краснощеков П. С., Петров А. А. Принципы построения моделей. — М.: Фазис, ВЦ РАН, 2000.
411. Красовский Н. Н. Теория управления движением. — М.: Наука, 1968.
412. Красовский Н. Н. Игровые задачи о встрече движений. — М.: Наука, 1970.

413. Красовский Н. Н., Субботин А. И. Позиционные дифференциальные игры. — М.: Наука, 1974.
414. Красовский Н. Н. Управление динамической системой. Задача о минимуме гарантированного результата. — М.: Наука, 1985.
415. Крейн М. Г., Нудельман А. А. Проблема моментов Маркова и экстремальные задачи. М.: Наука, 1973.
416. Крейн С. Г. Линейные уравнения в банаховом пространстве. М.: Наука, 1971.
417. Кротов В. Ф., Букреев В. З., Гурман В. И. Новые методы вариационного исчисления в динамике полета. — М.: Машиностроение, 1969.
418. Кротов В. Ф., Гурман В. И. Методы и задачи оптимального управления. — М.: Наука, 1973.
419. Кружков С. Н. Нелинейные уравнения первого порядка и связанные с ними дифференциальные игры. — Успехи матем. наук, 1969, **24**, № 2, С. 227–228.
420. Крылов Н. В. Управляемые процессы диффузионного типа. — М.: Наука, 1977.
421. Кряжимский А. В., Осипов Ю. С. О моделировании управления в динамической системе. Известия АН СССР, сер. техн. киберн., 1983, № 2, С. 51–60.
422. Кузнецов Ю. Н., Кузубов В. И., Волощенко А. Б. Математическое программирование. — М.: Высшая школа, 1980.
423. Кукушкин Н. С., Морозов В. В. Теория неантагонистических игр. — М.: Изд-во МГУ, 1984.
424. Кулешов А. А. О задаче оптимального управления для одной смешанной системы. Вестник Московск. ун-та, Серия 15. Вычислит. матем. и кибернетика, 1982, № 3, С. 15–20.
425. Кулешов А. А. Разностная аппроксимация и регуляризация одной задачи оптимального управления процессом, описываемым эллиптическим уравнением. Докл. АН СССР, 1983, Т. 269, № 4, С. 809–813.
426. Куликов А. Н., Фазылов В. Р. Выпуклая оптимизация с заданной точностью. Ж. вычислит. матем. и матем. физики, 1990, Т. 30, № 5, С. 663–671.
427. Куликовский Р. Оптимальные и адаптивные процессы в системах автоматического регулирования. — М.: Наука, 1967.
428. Куратовский К. Топология, Т. 1, М.: Мир, 1966.
429. Куржанский А. Б., Осипов Ю. С. К задаче об управлении с ограниченными фазовыми координатами. Прикладная матем. и механика, 1968, Т. 32, вып. 2, С. 194–202.
430. Куржанский А. Б., Осипов Ю. С. К задачам об управлении при стесненных координатах. Прикладная матем. и механика, 1969, Т. 33, вып. 4, С. 705–719.
431. Куржанский А. Б. Управление и наблюдение в условиях неопределенности. — М.: Наука, 1977.
432. Куржанский А. Б., Сивергина И. Ф. Метод гарантированных оценок и задачи регуляризации для эволюционных систем. Ж. вычислит. матем. и матем. физики, 1992, Т. 32, № 11, С. 1720–1733.
433. Куржанский М. А., Потапов М. М., Разгулин А. В. Проекционная схема метода прямых в задачах зонного управления и наблюдения для уравнения колебания струны. Вестник Московск. ун-та, серия 15, Вычислит. Матем. и кибернетика, 1994, № 3, С. 29–35.
434. Кусраев А. Г., Кутателадзе С. С. Субдифференциальное исчисление. — Новосибирск: Наука, 1987.
435. Кутателадзе С. С., Рубинов А. М. Двойственность Минковского и ее приложения. — Новосибирск: Наука. Сибирское отделение, 1976.
436. Кухта К. Я., Кравченко В. П., Красношапка В. А. Качественная теория управляемых динамических систем с непрерывно-дискретными параметрами. Киев: Наукова Думка, 1986.
437. Кюнци Г. П., Крелле В. Нелинейное программирование. — М.: Советское радио, 1965.
438. Лаврентьев М. М. О некоторых некорректных задачах математической физики. Новосибирск, изд-во СО АН СССР, 1962.
439. Лаврентьев М. М., Романов В. Г., Шишатский С. П. Некорректные задачи математической физики и анализа. М.: Наука, 1980.
440. Лагунов В. Н. Введение в дифференциальные игры. — Вильнюс: Институт матем. и кибернетики АН Литовской ССР, 1979.
441. Ладыженская О. А. Краевые задачи математической физики. — М.: Наука, 1973.
442. Ларичев О. И., Горвиц Г. Г. Методы поиска локального экстремума овражных функций. М.: Наука, 1990.

443. Латтес Р., Лионс Ж. — Л. Метод квазиобращения и его приложения. М.: Мир, 1970.
444. Лебедев В. И. Функциональный анализ и вычислительная математика. М.: Изд-во Всероссийского ин-та научной и технической информации, 1994.
445. Левин А. М. О регуляризации вычисления нижних граней функционалов. Ж. вычисл. матем. и матем. физики, 1984, Т. 24, № 8, С. 1123–1128.
446. Левин В. Л. Выпуклый анализ в пространствах измеримых функций и его применение в математике и экономике. М.: Наука, 1985.
447. Левитин Е. С., Поляк Б. Т. Методы минимизации при наличии ограничений. — Ж. вычислит. матем. и матем. физики, 1966, **6**, № 5, С. 787–823.
448. Левитин Е. С. Теория возмущений в математическом программировании и ее приложения. — М.: Наука, 1992.
449. Лейхтвейс К. Выпуклые множества. — М.: Наука, 1985.
450. Леонов А. С., Ягола А. Г. Можно ли решить некорректно поставленную задачу без знания погрешностей данных? Вестник Московск. ун-та, серия 3, Физика, Астрономия, 1995, Т. 36, № 4, С. 28–33.
451. Леонов А. С. Функции нескольких переменных с ограниченной вариацией в некорректных задачах. Ж. вычисл. матем. и матем. физики, 1996, Т. 36, № 9, С. 35–49.
452. Леонов А. С., Ягола А. Г. Метод L-кривой всегда дает неустранимую систематическую ошибку. Вестник Московск. ун-та, серия 3, Физика, астрономия, 1997, № 6, С. 17–19.
453. Леонов А. С. Замечания о полной вариации функций нескольких переменных и многомерном аналоге принципа выбора Хелли. Математ. заметки, 1998, Т. 63, № 1, С. 69–80.
454. Леонов А. С., Ягола А. Г. Адаптивные регуляризующие алгоритмы для решения некорректных задач. Вестник Московск. ун-та, серия 3, Физика, астрономия, 1998, № 2, С. 62–63.
455. Леонов А. С. О многомерных некорректных задачах с разрывными решениями. Сибирск. матем. журн., 1998, Т. 39, № 1, С. 74–86.
456. Лесин В. В., Лисовец Ю. П. Основы методов оптимизации. М.: Изд-во Московск. авиацион. института, 1998.
457. Ли Э. Б., Маркус Л. Основы теории оптимального управления. — М.: Наука, 1972.
458. Лионс Ж. — Л., Мадженес Э. Неоднородные граничные задачи и их приложения. М.: Мир, 1971.
459. Лионс Ж. — Л. Оптимальное управление системами, описываемыми уравнениями с частными производными. — М.: Мир, 1972.
460. Лионс Ж. — Л. Управление сингулярными распределенными системами. М.: Наука, 1987.
461. Лисковец О. А. Дискретные схемы в методе регуляризации для некорректных экстремальных задач. — ДАН СССР, 1979, **248**, № 6, С. 1299–1303.
462. Лисковец О. А. Вариационные методы решения неустойчивых задач. Минск: Наука и техника, 1981.
463. Литвинов В. Г. Оптимизация в эллиптических граничных задачах с приложениями к механике. М.: Наука, 1987.
464. Лопес Саура И. Разностная аппроксимация и регуляризация одной параболической задачи оптимального управления. Вестник Московск. ун-та, Серия 15. Вычислит. матем. и кибернетика, 1987, № 3, С. 35–42.
465. Лоран П. — Ж. Аппроксимация и оптимизация. — М.: Мир, 1975.
466. Лотов А. В. Введение в экономико-математическое моделирование. — М.: Наука, 1984.
467. Лубышев Ф. В. Разностные аппроксимации задач оптимального управления системами, описываемыми уравнениями в частных производных. Уфа: Изд-во Башкирского ун-та, 1999.
468. Лукьянов А. Т., Неронов В. С. Об оптимальном управлении одной параболически-гиперболической системой. — Изв. АН Казахской ССР, 1976, № 3. С. 77–80.
469. Лурье К. А. Оптимальное управление в задачах математической физики. — М.: Наука, 1975.
470. Лэсдон Л. С. Оптимизация больших систем. — М.: Наука, 1975.
471. Ляшенко И. Н., Карагодова Е. А., Черникова Н. В., Шор Н. З. Линейное и нелинейное программирование — Киев: Вища школа, 1975.
472. Магарил — Ильяев Г. Г., Тихомиров В. М. Выпуклый анализ и его приложения. М.: Изд-во Эдиториал УРСС, 2000.
473. Макаров В. Л., Рубинов А. М. Математическая теория экономической динамики и равновесия. — М.: Наука, 1973.

474. Максимов В. И. Конечномерная аппроксимация входов в гиперболических вариационных неравенствах. Ж. вычислит. матем. и матем. физики, 1995, Т. 35, № 11, С. 1615–1629.
475. Малков В. П., Угодчиков А. Г. Оптимизация упругих систем. М.: Наука, 1981.
476. Мансимов К. Б. Особые управления в системах с запаздыванием. Баку, Изд-во «ЭЛМ», 1999.
477. Марданов М. Д. Некоторые вопросы математической теории оптимальных процессов в системах с запаздываниями. — Баку: Изд-во Азербайдж. ун-та, 1987.
478. Маркин Е. А., Стрекаловский А. С. О существовании, единственности и устойчивости решения для одного класса динамических систем, описывающих химические процессы. — Вестник Московск. ун-та. Серия 15, Вычислит. матем. и киберн., 1977, № 4, С. 3–11.
479. Маркин Е. А. О задаче уклонения для одного класса динамических систем, описываемых волновым уравнением. — Кибернетика, 1978, № 1, С. 132–133.
480. Марчук Г. И., Шайдуров В. В. Повышение точности решений разностных схем. — М.: Наука, 1979.
481. Марчук Г. И. Методы вычислительной математики. — М.: Наука, 1980.
482. Марчук Г. И. Математическое моделирование в проблеме окружающей среды. М.: Наука, 1982.
483. Марчук Г. И. Сопряженные уравнения и анализ сложных систем. М.: Наука, Физматлит, 1992.
484. Марчук Г. И., Агошков В. И., Шутяев В. П. Сопряженные уравнения и методы возмущений в нелинейных задачах математической физики. М.: Наука, Физматлит, 1993.
485. Матвеев А. С. Вариационный анализ в задачах оптимизации систем с распределенными параметрами и вектор-функции множества. Сибирск. матем. журнал, 1990, Т. 31, № 6, С. 127–141.
486. Матвеев А. С., Якубович В. А. Абстрактная теория оптимального управления. — Санкт — Петербург: Изд-во С.-Петербургск. ун-та, 1994.
487. Мееров М. В. Исследование и оптимизация многосвязных систем управления. — М.: Наука, 1986.
488. Мезенцев А. В. Сборник задач по теории оптимального управления. — М.: Изд-во МГУ, 1980.
489. Меркулов В. И. Управление движением жидкости. Новосибирск: Наука, 1981.
490. Мечёнов А. С. Регуляризованный метод наименьших квадратов. М.: Изд-во Московск. ун-та, 1988.
491. Мину М. Математическое программирование. — М.: Наука, 1990.
492. Михайлов В. П. Дифференциальные уравнения в частных производных. — М.: Наука, 1983.
493. Михалевич В. С., Кукса А. И. Методы последовательной оптимизации в дискретных сетевых задачах оптимального распределения ресурсов. — М.: Наука, 1983.
494. Михалевич В. С., Трубин В. А., Шор Н. З. Оптимизационные задачи производственно-транспортного планирования: модели, методы, алгоритмы. — М.: Наука, 1986.
495. Михалевич В. С., Гупал А. М., Норкин В. И. Методы невыпуклой оптимизации. — М.: Наука, 1987.
496. Михлин С. Г. Вариационные методы в математической физике. М.: ГИТТЛ, 1957.
497. Моисеев Н. Н. Численные методы в теории оптимальных систем. — М.: Наука, 1971.
498. Моисеев Н. Н. Элементы теории оптимальных систем. — М.: Наука. 1975.
499. Моисеев Н. Н., Иванилов Ю. П., Столярова Е. М. Методы оптимизации. — М.: Наука. 1978.
500. Моисеев Н. Н. Математические задачи системного анализа. — М.: Наука, 1981.
501. Молодцов Д. А. Устойчивость принципов оптимальности. М.: Наука, 1987.
502. Мордухович Б. Ш. Методы аппроксимаций в задачах оптимизации и управления. — М.: Наука, 1988.
503. Мороз А. И. Курс теории систем. — М.: Высшая школа, 1987.
504. Морозкин Н. Д. О сходимости конечномерных приближений в задаче оптимального одномерного нагрева с учетом фазовых ограничений. Ж. вычисл. матем. и матем. физики, 1996, Т. 36, № 10, С. 12–22.
505. Морозкин Н. Д. Оптимальное управление процессами нагрева с учетом фазовых ограничений. Уфа: Изд-во Башкирск. ун-та, 1997.

506. Морозов В. А., Медведев Н. В., Иваницкий А. Ю. Регуляризация задач алгебры и анализа. — М.: Изд-во МГУ, 1987.
507. Морозов В. А. Методы регуляризации неустойчивых задач. М.: Изд-во Московск. ун-та, 1987.
508. Морозов В. А. Регулярные методы решения некорректно поставленных задач. — М.: Наука, 1987.
509. Морозов В. А., Гребеников А. И. Методы решения некорректно поставленных задач. Алгоритмический аспект. М.: Изд-во Московск. ун-та, 1992.
510. Морозов В. А. Регуляризация при больших помехах. Ж. вычисл. матем. и матем. физики, 1996, Т. 36, № 9, С. 13–21.
511. Морозов В. В., Сухарев А. Г., Федоров В. В. Исследование операций в задачах и упражнениях. — М.: Высшая школа, 1986.
512. Морозов С. Ф., Сумин В. И. Оптимизация нелинейных систем теории переноса. Ж. вычисл. матем. и матем. физики, 1979, Т. 19, № 1, С. 99–111.
513. Москаленко А. И. Методы нелинейных отображений в оптимальном управлении. — Новосибирск: Наука, 1983.
514. Мосолов П. П., Мясников В. П. Механика жесткопластических сред. М.: Наука, 1981.
515. Муравей Л. А. Задача управления границей для эллиптических уравнений. Вестник Московск. ун-та, серия 15, Вычисл. матем. и киберн., 1998, № 3, С. 7–13.
516. Муртаф Б. Современное линейное программирование. — М.: Мир, 1984.
517. Мухачева Э. А., Рубинштейн Г. Ш. Математическое программирование. — Новосибирск: Наука, 1987.
518. Назин А. В., Позняк А. С. Адаптивный выбор вариантов. Рекуррентные алгоритмы. — М.: Наука, 1986.
519. Наттерер Ф. Математические аспекты компьютерной томографии. М.: Мир, 1990.
520. Недич А. Трехшаговый метод проекции градиента для задачи минимизации. Известия вузов. Математика, 1993, № 10, С. 32–37.
521. Недич А. Непрерывный метод проекции градиента третьего порядка для задач минимизации. Дифференц. уравнения, 1994, Т. 30, № 11, С. 1914–1922.
522. Недич А. Регуляризованный непрерывный метод проекции градиента для задач минимизации с неточными исходными данными. Вестник Московскю ун-та, Серия 15, Вычислит. матем. и кибернетика, 1994, № 1, С. 3–10.
523. Немировский А. С., Юдин Д. Б. Сложность задач и эффективность методов оптимизации. — М.: Наука, 1979.
524. Неронов В. С. Оптимальное управление системами с распределенными параметрами. Алма-Ата, Изд-во Казахск. ун-та, 1989.
525. Нестеров Ю. Е. Эффективные методы в нелинейном программировании. — М.: Радио и связь, 1989.
526. Нефедов В. Н. Методы регуляризации многокритериальных задач оптимизации. — М.: Изд-во Московск. авиацион. ин-та, 1984.
527. Нефедов В. Н. Полиномиальные задачи оптимизации. Ж. вычисл. матем. и матем. физики, 1987, Т. 27, № 5, С. 661–675.
528. Никольский М. С. Об одном способе убегания. — ДАН СССР, 1974, **214**, № 2, С. 287–290.
529. Никольский М. С. О квазилинейной задаче убегания. — ДАН СССР, 1975, **221**, № 3, С. 539–542.
530. Никольский М. С. Первый прямой метод Л. С. Понтрягина в дифференциальных играх. — М.: Изд-во МГУ, 1984.
531. Никольский М. С. О задаче управления линейной системой с нарушениями. Докл. АН СССР, 1986, Т. 287, № 6, С. 1317–1320.
532. Никольский М. С. Об одной минимаксной задаче управления. Труды матем. ин-та АН СССР, 1988, Т. 185, С. 187–191.
533. Никольский М. С., Силин Д. Б. О наилучшем приближении выпуклого компакта элементами аддиала. Труды матем. ин-та АН СССР, 1995, Т. 211, С. 338–354.
534. Никольский С. М. Курс математического анализа. — М.: Наука, 1973, Т. 1, 2.
535. Никольский С. М. Приближение функций многих переменных и теоремы вложения. — М.: Наука, 1977.
536. Новикова Н. М. Стохастические методы численного решения выпуклых вариационных неравенств. Ж. вычисл. матем. и матем. физики, 1988, Т. 28, № 2, С. 186–197.
537. Новикова Н. М. Итеративная регуляризация метода штрафов для бесконечномерной задачи поиска стохастической седловой точки. Ж. вычисл. матем. и матем. физики, 1991, Т. 31, № 9, С. 1289–1304.

538. Новикова Н. М. Некоторые методы численного решения непрерывных выпуклых стохастических задач оптимального управления. Ж. вычисл. матем. и матем. физики, 1991, Т. 31, № 11, С. 1605–1618.
539. Новикова Н. М. Дискретные и непрерывные задачи оптимизации. — М.: Изд-во вычисл. центра РАН, 1996.
540. Новоженов М. М., Сумин В. И., Сумин М. И. Методы оптимального управления системами математической физики. Горький: Изд-во Горьковск. ун-та, 1986.
541. Ногин В. Д., Протодьяконов И. О., Евлампиев И. И. Основы теории оптимизации. — М.: Высшая школа, 1986.
542. Нурминский Е. А. Численные методы решения детерминированных и стохастических минимаксных задач. — Киев: Наукова думка, 1979.
543. Обен Ж. — П., Экланд И. Прикладной нелинейный анализ. М.: Мир, 1988.
544. Овсянников А. А. Моделирование и оптимизация динамики пучков заряженных частиц. Ленинград, Изд-во Ленинградск. ун-та, 1990.
545. Олейников В. А., Зотов Н. С., Пришвин А. М. Основы оптимального и экстремального управления. — М.: Высшая школа, 1969.
546. Олейников В. А., Зотов Н. С., Пришвин А. М., Соловьев Н. В. Сборник задач и примеров по теории автоматического управления. — М.: Высшая школа, 1969.
547. Ольхофф Н. Оптимальное проектирование конструкций. М.: Мир, 1981.
548. Оревков Ю. П. Методы оптимальных решений. — М.: Изд-во МГУ, 1986.
549. Орловский С. А. Проблемы принятия решений при нечеткой исходной информации. М.: Наука, 1981.
550. Ортега Д., Рейнболдт В. Итерационные методы решения нелинейных систем уравнений со многими неизвестными. — М.: Мир, 1975.
551. Осипов Ю. С. К теории дифференциальных игр в системах с распределенными параметрами. — ДАН СССР, 1975, **223**, № 6, С. 1314–1317.
552. Осипов Ю. С., Охезин С. П. К теории дифференциальных игр в параболических системах. — ДАН СССР, 1976, **226**, № 6, С. 1267–1270.
553. Осипов Ю. С. Позиционное управление в параболических системах. — Прикладная матем. и механика, 1977, **41**, № 2, С. 195–201.
554. Осипов Ю. С., Охезин С. П. К теории позиционного управления в гиперболических системах. — ДАН СССР, 1977, **233**, № 4, С. 551–554.
555. Осипов Ю. С., Суетов А. П. Существование оптимальных форм эллиптических систем. Случай краевых условий Дирихле. Свердловск: Изд-во УрО АН СССР, 1990.
556. Осипов Ю. С., Кряжимский А. В., Максимов В. И. Задачи динамической регуляризации для систем с распределенными параметрами. Свердловск: Изд-во ИММ УрО АН СССР, 1991.
557. Осипов Ю. С., Васильев Ф. П., Потапов М. М. Основы метода динамической регуляризации. М.: Изд-во Московск. ун-та, 1999.
558. Островский Г. М., Волин Ю. М. Моделирование сложных химико-технологических схем. — М.: Химия, 1975.
559. Охезин С. П. Дифференциальная игра сближения - уклонения для параболической системы при интегральных ограничениях на управления игроков. — Прикладная матем. и механика, 1977, **41**, № 2, С. 202–209.
560. Охезин С. П. Об управлении гиперболической системой в условиях неопределенности. — Прикладная матем. и механика, 1978, **42**, № 4, С. 606–612.
561. Павловский Ю. Н., Смирнова Т. Г. Проблема декомпозиции в математическом моделировании. М.: Фазис, 1998.
562. Панин В. М. Методы конечных штрафов с линейной аппроксимацией ограничений. Кибернетика, 1984, ч. 1, № 2, С. 44–50, ч. 2, № 4, С. 73–81.
563. Пахомов В. Ф. Нелинейное программирование. — М.: Изд-во МГУ, 1982.
564. Первозванский А. А., Гайцгорн В. Г. Декомпозиция, агрегирование и приближенная оптимизация. — М.: Наука, 1979.
565. Перевозчиков А. Г. О сложности вычисления глобального экстремума в одном классе многоэкстремальных задач. // Ж. вычисл. матем. и матем. физики, 1990, Т. 30, № 3, С. 379–387.
566. Петров А. А., Поспелов И. Г., Шананин А. А. Опыт математического моделирования экономики. — М.: Энергоатомиздат, 1996.
567. Петров А. А., Поспелов И. Г., Шананин А. А. От госплана к неэффективному рынку: математический анализ российских экономических структур. The Edwin Mellen Press, Lewiston, New York, 1999, 400 p.
568. Петров Ю. П. Вариационные методы теории оптимального управления. — Ленинград Энергия, 1977.

569. Петросян Л. А. Дифференциальные игры преследования. — Л.: Изд-во Ленинградск. ун-та, 1977.
570. Петросян Л. А., Томский Г. В. Динамические игры и их приложения. — Л.: Изд-во ЛГУ, 1982.
571. Петросян Л. А., Зенкевич Н. А., Семина Е. А. Теория игр. М.: Высшая школа, 1998.
572. Пинягина О. В., Фазылов В. Р. Метод выпуклого программирования с заданной абсолютно-относительной погрешностью. Ж. вычислит. матем. и матем. физики, 1998, Т. 38, № 8, С. 1247–1254.
573. Плотников В. И. Теоремы существования оптимизирующих функций для оптимальных систем с распределенными параметрами. — Изв. АН СССР. Сер. матем., 1970, **34**, № 3, С. 689–711.
574. Плотников В. И. Необходимые и достаточные условия оптимальности и условия единственности оптимизирующих функций для управляемых систем общего вида. — Изв. АН СССР. Сер. матем., 1972, **36**, № 3, С. 652–679.
575. Плотников В. И., Сумин В. И. Оптимизация объектов с распределенными параметрами, описываемых системами Гурса — Дарбу. — Ж. вычислит. матем. и матем. физики, 1972, **12**, № 1, С. 61–77.
576. Плотников В. И., Сумин В. И. Оптимизация распределенных систем в лебеговом пространстве. Сибирск. матем. журнал, 1981, Т. 22, № 6, С. 142–161.
577. Плотников В. И., Сумин М. И. О построении минимизирующих последовательностей в задачах управления системами с распределенными параметрами. Ж. вычисл. матем. и матем. физики, 1982, Т. 22, № 1, С. 49–56.
578. Плотников В. И., Шашков В. М., Кузенков О. А. Оптимальное управление линейными сосредоточенными системами. Нижний Новгород: Изд-во Нижегородск. ун-та, 1993.
579. Погорелов А. Г. Обратные задачи нестационарной химической кинетики. М.: Наука, 1988.
580. Подиновский В. В., Гаврилов В. М. Оптимизация по последовательно применяемым критериям. — М.: Советское радио, 1975.
581. Подиновский В. В., Ногин В. Д. Парето-оптимальные решения многокритериальных задач. — М.: Наука, 1982.
582. Полак Э. Численные методы оптимизации. Единый подход. — М.: Мир, 1974.
583. Половинкин Е. С. Необходимые условия оптимальности с дифференциальными включениями. Труды МИРАН, 1995, Т. 211, С. 387–400.
584. Половинкин Е. С. Сильно выпуклый анализ. Матем. сб., 1996, Т. 187, № 2, С. 103–130.
585. Половинкин Е. С. О сильно выпуклых множествах и сильно выпуклых функциях. Итоги науки и техники, Серия «Современная математика и ее применения». М.: 1999, Т. 61, С. 66–138.
586. Поляк Б. Т. Введение в оптимизацию. — М.: Наука, 1983.
587. Понтрягин Л. С., Болтянский В. Г., Гамкрелидзе Р. В., Мищенко Е. Ф. Математическая теория оптимальных процессов. — М.: Наука, 1976.
588. Понтрягин Л. С. Обыкновенные дифференциальные уравнения. — М.: Наука, 1983.
589. Понтрягин Л. С. Избранные научные труды, Т. II, М.: Наука, 1988.
590. Попов Н. М. Приближенное решение многокритериальных задач с функциональными ограничениями. // Ж. вычисл. матем. и матем. физики, 1986, Т. 26, № 10, С. 1468–1481.
591. Попов Н. М. К оцениванию информационной сложности глобальной оптимизации и глобального решения уравнений. // Ж. вычисл. матем. и матем. физики, 1992, Т. 32, № 12, С. 1853–1868.
592. Попов Н. М. О некоторых принципах оптимальности в многокритериальных задачах. Проблемы математической физики. Сб. работ факультета ВМиК МГУ. М.: Изд-во «Диалог МГУ», 1998, С. 217–224.
593. Потапов М. М. Разностная аппроксимация и регуляризация задач оптимального управления системами Гурса — Дарбу. — Вестник Московск. ун-та. Серия 15, Вычислит. матем. и киберн., 1978, № 2, С. 17–26.
594. Потапов М. М. Разностная аппроксимация максиминных задач для систем Гурса — Дарбу при наличии фазовых ограничений. — Вестник Московск. ун-та. Сер. вычислит. матем. и киберн., 1978, № 4, С. 28–36.
595. Потапов М. М. Об аппроксимации по функционалу максиминных задач со связанными переменными. — Ж. вычислит. матем. и матем. физики, 1979, **19**, № 3, С. 610–621.

596. Потапов М. М. Аппроксимация экстремальных задач в математической физике (гиперболические уравнения). М.: Изд — во Московск. ун-та, 1985.
597. Потапов М. М., Разгулин А. В., Шамеева Т. Ю. Аппроксимация и регуляризация задачи оптимального управления для уравнения типа Шредингера. Вестник Московск. ун-та, Серия 15. Вычислит. матем. и кибернетика, 1987, № 1, С. 8–13.
598. Потапов М. М. Метод прямых в задачах граничного управления и наблюдения для гиперболического уравнения с краевыми условиями второго и третьего рода. Вестник Московск. ун-та, серия 15, Вычислит. матем. и кибернетика, 1996, № 2, С. 35–41.
599. Потапов М. М. О сильной сходимости разностных аппроксимаций для задач граничного управления и наблюдения для волнового уравнения. Ж. вычислит. матем. и матем. физики, 1998, Т. 38, № 3, С. 387–397.
600. Потапов М. М. Устойчивый метод решения линейных уравнений с неравномерно возмущенным оператором. Докл. РАН, 1999, Т. 365, № 5, С. 1–3.
601. Практикум по линейному программированию. Под ред. Черемных Ю. Н., Павловой Л. С., Сутормина Е. И. — М.: Изд-во МГУ, 1984.
602. Пропой А. И. Элементы теории оптимальных дискретных процессов. — М.: Наука, 1973.
603. Пшеничный Б. Н., Данилин Ю. М. Численные методы в экстремальных задачах. — М.: Наука, 1975.
604. Пшеничный Б. Н. Выпуклый анализ и экстремальные задачи. — М.: Наука, 1980.
605. Пшеничный Б. Н. Необходимые условия экстремума. — М.: Наука, 1982.
606. Пшеничный Б. Н. Метод линеаризации. — М.: Наука, 1983.
607. Разгулин А. В. Применение проекционно-разностного метода в задачах наблюдения и управления для уравнения типа Шредингера. Вестник Московск. ун-та, серия 15, Вычислит. матем. и кибернетика, 1996, № 1, С. 42–52.
608. Разумихин Б. С. Физические модели и методы теории равновесия в программировании и экономике. — М.: Наука, 1975.
609. Райтум У. Е. Задачи оптимального управления для эллиптических уравнений. Рига: Зинатне, 1989.
610. Растригин Л. А. Системы экстремального управления. — М.: Наука, 1974.
611. Раушенбах Б. В., Токарь Е. Н. Управление ориентацией космических аппаратов. — М.: Наука, 1974.
612. Рейклейтис Г., Рейвиндран А., Рэгсдел К. Оптимизация в технике. В двух книгах. — М.: Мир, 1986.
613. Ржевский С. В. Монотонные методы выпуклого программирования. — Киев: Наукова думка, 1993.
614. Рихтер К. Динамические задачи дискретной оптимизации. — М.: Радио и связь, 1985.
615. Рождественский Б. Л., Яненко Н. Н. Системы квазилинейных уравнений и их приложения к газовой динамике. — М.: Наука, 1978.
616. Ройтенберг Я. Н. Автоматическое управление. — М.: Наука, 1978.
617. Рокафеллар Р. Выпуклый анализ. — М.: Мир, 1973.
618. Романов В. Г. Обратные задачи математической физики. М.: Наука, 1984.
619. Романов В. Г., Кабанихин С. И. Обратные задачи геоэлектрики. М.: Наука, 1991.
620. Романовский И. В. Алгоритмы решения экстремальных задач. — М.: Наука, 1977.
621. Русов В. Д., Бибикова Ю. Ф., Ягола А. Г. Восстановление изображений в электронно-микроскопической авторадиографии поверхности. М.: Энергоатомиздат, 1991.
622. Рязанцева И. П. О выборе параметра регуляризации при решении выпуклых экстремальных задач. Ж. вычисл. матем. и матем. физики, 1997, Т. 37, № 7, С. 895–896.
623. Рязанцева И. П., Дунцева Е. А. Об одном непрерывном методе решения выпуклых экстремальных задач. Дифференциальные уравнения, 1998, Т. 34, № 4, С. 480–485.
624. Саати Т. Целочисленные методы оптимизации и связанные с ними экстремальные проблемы. — М.: Мир, 1973.
625. Савелова Т. И. Об устойчивом суммировании рядов Фурье. Ж. вычисл. матем. и матем. физики, 1979, Т. 19, № 4, С. 830–835.
626. Савелова Т. И. О связи метода регуляризации А. Н. Тихонова для некорректных уравнений типа свертки с решением краевых задач. Ж. вычисл. матем. и матем. физики, 1982, Т. 22, № 6, С. 1316–1322.
627. Савелова Т. И., Бухараева Т. И. Представления группы SU(2) и их применения. М.: Изд-во МИФИ, 1996.

628. Савелова Т. И. Примеры решеня некорректно поставленных задач. М.: Изд-во Московск. инженерно-физич. ин-та, 1999.
629. Садовничий В. А. Теория операторов. М.: Высшая школа, 1999.
630. Самарский А. А. Введение в теорию разностных схем. — М.: Наука, 1971.
631. Самарский А. А., Гулин А. В. Устойчивость разностных схем. — М.: Наука, 1973.
632. Самарский А. А., Попов Ю. П. Разностные схемы газовой динамики. — М.: Наука, 1975.
633. Самарский А. А., Андреев В. Б. Разностные методы для эллиптических уравнений. — М.: Наука, 1976.
634. Самарский А. А. Теория разностных схем. — М.: Наука, 1977.
635. Самарский А. А., Николаев Е. С. Методы решения сеточных уравнений. — М.: Наука, 1978.
636. Самсонов С. П. Восстановление выпуклого множества по его опорной функции с заданной точностью. // Вестник Московск. ун-та, серия 15, Вычислит. матем. и киберн., 1983, № 1, С. 68–71.
637. Сатимов Н. Ю. Об одном способе убегания в дифференциальных играх. — Матем. сборник, 1976, **99 (141)**, № 3, С. 380–393.
638. Сеа Ж. Оптимизация. Теория и алгоритмы. — М.: Мир, 1973.
639. Сергиенко И. В., Лебедева Т. Т., Рощин В. А. Приближенные методы решения дискретных задач оптимизации. — Киев: Наукова думка, 1980.
640. Серовайский С. Я. Вариационные неравенства в оптимизационных задачах. Алма-Ата, Изд-во Казахск. ун-та, 1981.
641. Сиразетдинов Т. К. Оптимизация систем с распределенными параметрами. — М.: Наука, 1977.
642. Сиразетдинов Т. К. Устойчивость систем с распределенными параметрами. Новосибирск: Наука, 1987.
643. Сиразетдинов Т. К. Методы решения многокритериальных задач синтеза технических систем. М.: Машиностроение, 1988.
644. Скарин В. Д. Об одном подходе к анализу несобственных задач линейного программирования. Ж. вычислит. матем. и матем. физики, 1986, Т. 26, № 3, С. 439–448.
645. Слугин С. Н., Шашков В. М., Миронов А. В. Топологический достаточный признак существования оптимального управления динамической системой. — Известия вузов. Математика, 1977, № 10 (185), С. 134–137.
646. Смирнов Е. Я. Стабилизация программных движений. С.-Петербург, Изд-во С.-Петербургск. ун-та, 1997.
647. Смольяков Э. Р. Теория антагонизмов и дифференциальные игры. М.: Эдиториал УРСС, 2000.
648. Соболев С. Л. Некоторые применения функционального анализа в математической физике. Изд-во СО АН СССР, 1962.
649. Соболев С. Л. Введение в теорию кубатурпых формул. — М.: Наука, 1974.
650. Современное состояние теории исследования операций. Сб. работ / Под ред. Н. Н. Моисеева. — М.: Наука, 1979.
651. Соловьева С. И. Итерационный метод решения обратной задачи для нелинейного дифференциального уравнения. Вестник Московск. ун-та, серия 15, Вычисл. матем. и киберн., 1998, № 4, С. 14–16.
652. Солодовников А. С. Введение в линейную алгебру и линейное программирование. — М.: Просвещение, 1966.
653. Срочко В. А. Вычислительные методы оптимального управления. — Иркутск: Изд-во Иркутск. ун-та, 1982.
654. Срочко В. А. Вариационный принцип максимума и методы линеаризации в задачах оптимального управления. — Иркутск: Изд-во Иркутск. ун-та, 1989.
655. Срочко В. А. Методы линейно-квадратичных аппроксимаций для решения задач оптимального управления. Оптимизация, управление, интеллект. Иркутск: Изд-во Иркутск. ун-та, ВЦ СО РАН, 1995, № 1, С. 110–135.
656. Срочко В. А. Итерационные методы решения задач оптимального управления. М.: Физматлит, 2000.
657. Старосельский Л. А., Шелудько Г. А., Кантор Б. Я. Об одной реализации метода оврагов с адаптацией величины овражного шага по экспоненциальному закону. Ж. вычисл. матем. и матем. физики, 1968, Т. 8, № 5, С. 1161–1167.
658. Старостенко В. И. Устойчивые численные методы в задачах гравиметрии. Киев: Наукова думка, 1978.
659. Страхов В. Н. О построении оптимальных по порядку приближенных решений линейных условно-корректных задач. Дифференц. уравнения. 1973, Т. 9, № 10, С. 1862–1874.

660. Стрекаловский А. С. К проблеме глобального экстремума // ДАН СССР. — 1987. — Т. 292, № 5. — С. 1062–1066.
661. Стрекаловский А. С. Условие глобальной оптимальности в задачах D. С. программирования. Иркутск: Изд-во Иркутск. ун-та, 1997. Серия «Оптимизация и управление», вып. 1.
662. Стронгин Р. Г. Численные методы в многоэкстремальных задачах управления. — М.: Наука, 1978.
663. Субботин А. И., Ченцов А. Г. Оптимизация гарантии в задачах управления. — М.: Наука, 1981.
664. Сумин В. И. Об обосновании градиентных методов для распределенных задач оптимального управления. Ж. вычисл. матем. и матем. физики, 1990, Т. 30, № 1, С. 3–21.
665. Сумин В. И. О достаточных условиях устойчивости существования глобальных решений управляемых краевых задач. Дифференц. уравнения, 1990, Т. 26, № 12, С. 2097–2109.
666. Сумин В. И. Функциональные вольтерровы уравнения в теории оптимального управления распределенными системами. Нижний Новгород: Изд-во Нижегородск. ун-та, 1992.
667. Сумин М. И. О первой вариации в теории оптимального управления системами с распределенными параметрами. Дифференц. уравнения, 1991, Т. 27, № 12, С. 2179–2181.
668. Сумин М. И. Субоптимальное управление системами с распределенными параметрами: минимизирующие последовательности, функция значений. Ж. вычисл. матем. и матем. физики, 1997, Т. 37, № 1, С. 23–41.
669. Сумин М. И. Субоптимальное управление системами с распределенными параметрами: свойства нормальности, субградиентный двойственный метод. Ж. вычисл. матем. и матем. физики, 1997, Т. 37, № 2, С. 162–178.
670. Сухарев А. Г., Тимохов А. В., Федоров В. В. Курс методов оптимизации. — М.: Наука, 1986.
671. Сухарев А. Г. Минимаксные алгоритмы в задачах численного анализа. — М.: Наука, 1989.
672. Сухинин М. Ф. Полутейлоровские снизу отображения и достаточные условия экстремума. Матем. сборник, 1991, № 6, С. 877–891.
673. Сухинин М. Ф. Избранные главы нелинейного анализа. М.: Изд-во ун-та Дружбы народов, 1992.
674. Сухинин М. Ф. К вопросу о беллмановском подходе в теории оптимального управления. Матем. заметки, 1999, Т. 66, вып. 5, С. 770–776.
675. Сухинин М. Ф. Численное решение некоторых экстремальных задач. М.: Изд-во ун-та дружбы народов, 2000.
676. Схрейвер А. Теория линейного и целочисленного программирования. В двух томах. — М.: Мир, 1991.
677. Тадумадзе Т. А. Некоторые вопросы качественной теории оптимального управления. — Тбилиси: Изд-во Тбилисского ун-та, 1983.
678. Танаев В. С., Шкурба В. В. Введение в теорию расписаний. — М.: Наука, 1975.
679. Танана В. П. Методы решения операторных уравнений. М.: Наука, 1981.
680. Танана В. П., Рекант М. А., Янченко С. И. Оптимизация методов решения операторных уравнений. Екатеринбург, Изд-во Уральского ун-та, 1987.
681. Тарасова В. П. Метод стратегии противника в задачах оптимального поиска. — М.: Изд-во Московск. ун-та, 1988.
682. Темам Р. Математические задачи теории пластичности. М.: Наука, 1991.
683. Тер-Крикоров А. М. Оптимальное управление и математическая экономика. — М.: Наука, 1977.
684. Тетерев А. Г. Методы одномерной оптимизации. — Куйбышев: Изд-во Куйбышевск. ун-та, 1983.
685. Тетерев А. Г. Линейные задачи оптимизации. — Куйбышев: Изд-во Куйбышевск. ун-та, 1983.
686. Тимохов А. В. Математические модели экономического воспроизводства. — М.: Изд-во МГУ, 1982.
687. Тихомиров В. М. Некоторые вопросы теории приближений. — М.: Изд-во МГУ, 1976.
688. Тихомиров В. М. Рассказы о максимумах и минимумах. — М.: Наука, 1986.
689. Тихомиров В. М. Выпуклый анализ. Теория приближений.Итоги науки и техники. Серия Современные проблемы математики, фундаментальные направления, том 14, М.: Изд-во ВИНИТИ, 1987.

690. Тихомиров В. М. Теория экстремума и экстремальные задачи классического анализа. Итоги науки и техники. Серия «Современная математика и ее приложения», Т. 65. М.: Изд-во ВИНИТИ, 1999, С. 188–258.
691. Тихонов А. Н., Васильев Ф.П., Потапов М. М., Юрий А. Д. О регуляризации задач минимизации на множествах, заданных приближенно. — Вестник Московск. ун-та. Серия 15, Вычислит. матем. и киберн., 1977, № 1, С. 4–19.
692. Тихонов А. Н., Васильев Ф. П. Методы решения некорректных экстремальных задач. — В кн.: Banach Center Publications. V. 3. Mathematical models and numerical methods. Warszawa, 1978, p. 297–342.
693. Тихонов А. Н., Гончарский А. В., Степанов В. В., Ягола А. Г. Регуляризующие алгоритмы и априорная информация. М.: Наука, 1983.
694. Тихонов А. Н., Васильева А. Б., Свешников А. Г. Дифференциальные уравнения. — М.: Наука, 1985.
695. Тихонов А. Н., Арсенин В. Я. Методы решения некорректных задач. — М.: Наука, 1986.
696. Тихонов А. Н., Гончарский А. В., Степанов В. В., Ягола А. Г. Численные методы решения некорректных задач. М.: Наука, 1990.
697. Тихонов А. Н., Леонов А. С., Ягола А. Г. Нелинейные некорректные задачи. М.: Наука, Физматлит, 1995.
698. Тихонов А. Н., Самарский А. А. Уравнения математической физики. — М.: Изд-во Московск. ун-та, 1999.
699. Толстоногов А. А. Дифференциальные включения в банаховом пространстве. Новосибирск: Наука, 1986.
700. Толстошеин А. Ю. О задаче оптимального управления процессами, описываемыми одной смешанной системой. Вестник Московск. ун-та, Серия 15. Вычислит. матем. и кибернетика, 1986, № 4, С. 19–24.
701. Толстых В. К. Прямой экстремальный подход для оптимизации систем с распределенными параметрами. Донецк: Изд-во «Юго — Восток», 1997.
702. Тонков Е. Л. О множестве управляемости линейного уравнения. Дифференциальные уравнения, 1983, Т. 19, № 2, С. 269–278.
703. Тонков Е. Л. Задачи управления показателями Ляпунова. Дифференц. уравнения, 1995, Т. 31, № 10, С. 1682–1686.
704. Трауб Дж., Вожьняковский Х. Общая теория оптимальных алгоритмов. — М.: Мир, 1983.
705. Треногин В. А. Функциональный анализ. М.: Наука, 1980.
706. Троицкий В. А. Оптимальные процессы колебаний механических систем. — Л.: Машиностроение, 1976.
707. Троицкий В. А., Петухов Л. В. Оптимизация упругих тел. М.: Наука, 1982.
708. Туйкина С. Р. Численные методы решения некоторых обратных задач динамики сорбции. Вестник Московск. ун-та, серия 15, Вычисл. матем. и киберн., 1998, № 4, С. 16–19.
709. Уайлд Д. Дж. Методы поиска экстремума. — М.: Наука, 1967.
710. Уздемир А. П. Динамические целочисленные задачи оптимизации в экономике. — М.: Наука, Физматлит, 1995.
711. Ульм С. Ю. Методы декомпозиции для решения задач оптимизации. — Таллин: Изд-во Валгус, 1979.
712. Уонэм М. Линейные многомерные системы управления. — М.: Наука, 1980.
713. Урясьев С. П. Адаптивные алгоритмы стохастической оптимизации и теории игр. М.: Наука, 1990.
714. Успенский А. Б., Федоров В. В. Вычислительные аспекты метода наименьших квадратов при анализе и планировании регрессионных экспериментов. — М.: Изд-во Московск. ун-та, 1975.
715. Усынин Г. Б., Карабасов А. С., Чирков В. А. Оптимизационные модели реакторов на быстрых нейтронах. М.: Атомиздат, 1981.
716. Уткин В. И. Скользящие режимы в задачах оптимизации и управления. — М.: Наука, 1981.
717. Уткин В. И., Орлов Ю. В. Теория бесконечномерных систем управления на скользящих режимах. М.: Наука, 1990.
718. Фазылов В. Р. Отыскание минимакса с заданной точностью. Ж. вычислит. матем. и матем. физики, 1994, Т. 34, № 5, С. 793–799.
719. Федоренко Р. П. Приближенное решение задач оптимального управления. — М.: Наука, 1978.
720. Федоров В. В. Численные методы максимина. — М.: Наука, 1979.

721. Фиакко А., Мак-Кормик Г. Нелинейное программирование. Методы последовательной безусловной минимизации. — М.: Мир, 1972.
722. Филиппов А. Ф. О некоторых вопросах оптимального регулирования. // Вестник МГУ, серия 1, Матем. и механика, 1959, № 2, С. 25–38.
723. Филиппов А. Ф. Дифференциальные уравнения с разрывной правой частью. М.: Наука, 1985.
724. Флеминг У., Ришел Р. Оптимальное управление детерминированными и стохастическими системами. — М.: Мир, 1978.
725. Форд Л., Фалкерсон Д. Потоки в сетях. — М.: Мир, 1966.
726. Формальский А. М. Управляемость и устойчивость систем с ограниченными ресурсами. — М.: Наука, 1974.
727. Фурасов В. Д. Устойчивость движения, оценки и стабилизация. — М.: Наука, 1977.
728. Фурасов В. Д. Устойчивость и стабилизация дискретных процессов. М.: Наука, 1982.
729. Фурасов В. Д. Моделирование плохоформализуемых процессов. М.: Изд-во «Academia», 1997.
730. Фурсиков А. В. Оптимальное управление распределенными системами. Теория и приложения. Новосибирск: Научная книга, 1999.
731. Хайлов Е. Н. Об экстремальных управлениях однородной билинейной системы, управляемой в положительном ортанте. Труды матем. ин-та РАН, 1998, Т. 220, С. 217–235.
732. Хапаев М. М. Дифференциальные уравнения, содержащие сингулярные многообразия в задачах управления и минимизации. Дифференц. уравнения, 1995, Т. 31, № 11, С. 1886–1892.
733. Харатишвили Г. Л., Мачаидзе З. А., Маркозашвили Н. И., Тадумадзе Т. А. Абстрактная вариационная теория и ее применения к оптимальным задачам с запаздываниями. — Тбилиси: Мецниереба, 1973.
734. Харатишвили Г. Л., Тадумадзе Т. А. Нелинейные оптимальные системы управления с переменными запаздываниями. Матем. сб., 1978, Т. 107, вып. 4, С. 613–633.
735. Харди Г. Г., Литтльвуд Дж. Е., Полиа Г. Неравенства. — М.: Изд-во иностран. лит-ры, 1948.
736. Хачиян Л. Г. Полиномиальные алгоритмы в линейном программировании. // Ж. вычисл. матем. и матем. физики, 1980, Т. 20, № 1, С. 51–68.
737. Хедли Дж. Нелинейное и динамическое программирование. — М.: Мир, 1967.
738. Химмельблау Д. Прикладное нелинейное программирование. — М.: Мир, 1975.
739. Хоменюк В. В. Оптимальные системы управления. — М.: Наука, 1977.
740. Хромова Г. В. Об оценке погрешности метода регуляризации Тихонова для интегральных уравнений с ядром Грина. Вестник Московск. ун-та, серия 15, Вычисл. матем. и киберн., 1992, № 4, С. 22–27.
741. Хромова Г. В. Приближающие свойства резольвент дифференциальных операторов в задаче приближения функций и их производных. Ж. вычисл. матем. и матем. физики, 1998, Т. 38, № 7, С. 1036–1043.
742. Хромова Л. Н. Об одном методе минимизации с кубической скоростью сходимости. — Вестник Московск. ун-та. Серия 15, Вычислит. матем. и киберн., 1980, № 3, С. 52–56.
743. Ху Т. Целочисленное программирование и потоки в сетях. — М.: Мир, 1974.
744. Цирлин А. М., Балакирев В. С., Дудников Г. Е. Вариационные методы оптимизации управляемых объектов. — М.: Энергия, 1976.
745. Цирлин А. М. Методы усредненной оптимизации и их приложения. М.: Физматлит, 1997.
746. Цурков В. И. Декомпозиция в задачах большой размерности. — М.: Наука, 1981.
747. Цурков В. И., Литвинчев И. С. Декомпозиция в динамических задачах с перекрестными связями. — М.: Физматлит, 1994.
748. Ченцов А. Г. Конечно-аддитивные меры и релаксации экстремальных задач. Екатеринбург: Наука, 1993.
749. Черемных Ю. Н. Качественное исследование оптимальных траекторий динамических моделей экономики. — М.: Изд-во Московск. ун-та, 1975.
750. Черемных Ю. Н. Анализ поведения траекторий динамики народнохозяйственных моделей. — М.: Наука, 1982.
751. Черемных Ю. Н. Математические модели развития народного хозяйства. — М.: Изд-во МГУ, 1986.
752. Черников С. Н. Линейные неравенства. — М.: Наука, 1968.
753. Черноусько Ф. Л., Баничук Н. В. Вариационные задачи механики и управления. — М.: Наука, 1973.

754. Черноусько Ф. Л., Колмановский В. Б. Оптимальное управление при случайных возмущениях. — М.: Наука, 1978.
755. Черноусько Ф. Л., Меликян А. А. Игровые задачи управления и поиска. — М.: Наука, 1978.
756. Черноусько Ф. Л., Акуленко Л. Д., Соколов Б. Н. Управление колебаниями. — М.: Наука, 1980.
757. Чечкин А. В. Математическая информатика. М.: Наука, 1991.
758. Чирич Н. Т. О регуляризованном методе линеаризации выпуклой функции на многогранном множестве при наличии погрешностей в исходных данных. // Вестник Московск. ун-та, серия 15, Вычислит. матем. и кибернетика, 1987, № 2, С. 20–25.
759. Численные методы условной оптимизации // Сб. работ под ред. Гилл Ф., Мюррэй У. — М.: Мир, 1977.
760. Чичинадзе В. К. Решение невыпуклых нелинейных задач оптимизации. — М.: Наука, 1983.
761. Шамеева Т. Ю. Об оптимизации в задаче о распространении светового пучка в неоднородной среде. Вестник Московск. ун-та, Серия 15. Вычислит. матем. и кибернетика, 1985, № 1, С. 12–19.
762. Шананин А. А. Об агрегации функций спроса. Экономика и матем методы. 1989, № 2.
763. Шананин А. А. Двойственность для задач обобщенного программирования и вариационные принципы в моделях экономического равновесия. Докл. РАН, 1999, Т. 366, № 4, С. 462–464.
764. Шафиев Р. А. Псевдообращение операторов и некоторые применения. Баку: Изд-во «ЭЛМ», 1989.
765. Шевченко В. Н. Качественные вопросы целочисленного программирования. М.: Физматлит, 1995.
766. Шепилов М. А. О методе обобщенного градиента для экстремальных задач. // Ж. вычислит. матем. и матем. физики, 1976, Т. 16, № 1, С. 242–247.
767. Шикин Е. В. Линейные пространства и отображения. — М: Изд-во Московск. ун-та, 1987.
768. Шилов Г. Е. Математический анализ (функции нескольких вещественных переменных). — М.: Наука, 1972.
769. Шор Н. З. Методы минимизации недифференцируемых функций и их приложения. — Киев: Наукова думка, 1979.
770. Шор Н. З., Стеценко С. И. Квадратичные экстремальные задачи и недифференцируемая оптимизация. Киев: Наукова думка, 1989.
771. Щеглов А. Ю. О равномерном приближении решения одной обратной задачи методом типа квазиобращения. Матем. заметки, 1993, Т. 53, вып. 2, С. 168–174.
772. Эдвардс Р. Функциональный анализ. М.: Мир, 1969.
773. Экланд И., Темам Р. Выпуклый анализ и вариационные проблемы. — М.: Мир, 1979.
774. Эльстер К.-Х., Рейнгардт Р., Шойбле М., Донат Г. Введение в нелинейное программирование. — М.: Наука, 1985.
775. Юдин Д. Б., Гольштейн Е. Г. Линейное программирование. М.: Физматгиз, 1963.
776. Юдин Д. Б., Гольштейн Е. Г. Линейное программирование. Теория, методы и приложения. — М.: Наука, 1969.
777. Юдин Д. Б. Задачи и методы стохастического программирования. — М.: Советское радио, 1979.
778. Юнусов М. К. Оптимальное управление системами в некоторых процессах тепло-массопереноса. Душанбе: Дониш, 1987.
779. Юрий А. Д. Об одной оптимальной задаче типа Стефана. Докл. АН СССР, 1980, Т. 251, № 6, С. 1317–1321.
780. Якубович В. А. К абстрактной теории оптимального управления. — Сибирск. матем. журн., I, 1977, **18**, № 3, С. 685–707; II, 1978, **19**, № 2, С. 436–460; III, 1979, **20**, № 4, С. 385–410; IV, 1979, **20**, № 5, С. 1131–1159.
781. Янг Л. Лекции по вариационному исчислению и теории оптимального управления. — М.: Мир, 1974.
782. Яхно В. Г. Обратные задачи для дифференциальных уравнений упругости. Новосибирск: Наука, 1990.
783. Ячимович М. Итеративная регуляризация одного варианта метода условного градиента. Вестник Московск. ун-та, Серия 15, Вычисл. матем. и кибернетика, 1980, № 4, С. 13–19.

784. Avdonin S. A., Ivanov S. A. Families of Exponentials. The Method of Moments in Controllability Problems for Distributed Parameter Systems. Cambridge University Press. 1995.
785. Bertsekas D. P. Nonlinear Programming. Athena Scientific, 1999.
786. Boukari D., Fiacco A.V. Survey of penalty, exact-penalty and multiplier methods from 1968 to 1993. Optimization, 1995, V. 32, pp. 301–334.
787. Bulirsch R., Montrone F., Pesch H. J. Abort landing in the presence of windshear as a minimax optimal control problem. I Necessary conditions, II Multiple shooting and homotopy. J. Optim. Theory Appl., 1991, V. 70, № 1, pp. 1–23, № 2, pp. 223–254.
788. Denisov A. M., Lamos H. An inverse problem for a nonlinear mathematical model of sorption dynamics with mixed-diffusional kinetics. J. Inverse and Ill-posed Problems. 1996, V. 4, № 3, pp. 191–202.
789. Dontchev A. L., Zolezzi T. Well-Posed Optimization Problems. Springer — Verlag, Berlin — Heidelberg, 1993.
790. Fattorini H. O. Optimal control of nonlinear systems: convergence of suboptimal controls. I. Lecture Notes in Pure and Applied Mathematics, 108, Marsel Dekker, New York, 1987.
791. Fiacco A. V. Introduction to sensitivity and stability analysis in nonlinear programming. New York, Academic Press, 1983.
792. Grossmann C., Kaplan A. A. Strafmethoden und modifizierte Lagrangefunktionen in der nichtlinearen Optimierung. Teubner, Leipzig, 1979.
793. Grossmann C., Temo J. Numerik der nichtlinearen Optimierung. Teubner, Stuttgart, 1997.
794. Henkin G., Shananin A. Bernstein theorems and Radon transform. Application to the theory of production functions. Translation of mathematical monographs, 1990, V. 81.
795. Hiriart — Urruty J. B., Lemareshal C. Convex Analysis and Minimization Algorithms, V. I, II. Springer, 1993.
796. Hoffman A. J. On approximate solutions of systems of linear inequalities. Journ. of Research of Nat. Bureau of Standarts, 1952, V. 49, pp. 263–265.
797. Jai A. El, Prichard A. J. Sensors and controls in the analysis of distributed systems NY: J. Wiley and Sons, 1988.
798. Janković V., Jovanović M. On contractibility of the operator $I - t\nabla f$. Matematićki Vesnik, 1997, V. 49, № 3–4, pp. 245–248.
799. Kaplan A., Tichatschke R. Stable Methods for Ill-Posed Variational Problems. Berlin, Academie Verlag, 1994.
800. Kaplan A., Tichatschke R. Multi-step-prox-regularization method for solving convex variational problems. Optimization, 1995, V. 33, pp. 287–319.
801. Krabs W. Optimization and Approximation. John Wiley & Sons, 1979.
802. Krabs W. On moment theory and controllability of one-dimensional vibrating systems and heating processes. Springer — Verlag, Berlin, Heidelberg, 1992.
803. Lemaire B. The proximal algorithm. Intern. Series of Numer. Mathem., 1989, vol. 87, Basel; Birkhäuser Verlag, 1989, pp. 73–87.
804. Lions J. — L. Exact controllability, stabilization and perturbations for distributed systems. SIAM Rev., 1988, V. 30, № 2, p. 1–68.
805. Malanowski K. Regularity of solutions in stability analysis of optimization and optimal control problems. Control and Cybernetics, 1994, V. 23, № 1–2, pp. 61–86.
806. Mangasarian O. L. Nonlinear Programming, SIAM, Philadelphia, 1994.
807. Nesterov Y., Nemirovskii A. Interior-Point Polynomial Algorithms in Convex Programming. Philadelphia, 1994.
808. Osipov Yu. S., Kryazhimskii A. V. Inverse problems for ordinary differential equation: dynamical solutions. Gordon and Breach, London, 1995.
809. Petrosjan L. A. Differential Games of Pursuit. World Scientific, Singapore, London, 1993.
810. Petrov A. A., Shananin A. A. Integrability conditions, income distribution and social structures. Lecture notes in economics and mathematic systems, V. 453, 1998.
811. Polak E. Optimization: Algorithms and consistent approximations. Springer, New York, 1997.
812. Prilepko A. I., Orlovsky D. G., Vasin I. A. Methods for solving inverse problems in mathematical physics. Marsel Dekker, 2000.
813. Rockafellar R. T. Monotone Operators and the proximal point algorithm, SIAM J. Control and Optimisation, 1976, p. 877–898.
814. Rockafellar R. T., Wets R. J. — — B. Variational Analysis. Springer. 1998.
815. Silin D. B. On Discontinuous Strategies in Optimal Control Problems. Journ. of Mathematical Systems, Estimation and Control, 1994, V. 4, № 2, pp. 205–217.
816. Strekalovsky A. S. Global optimality conditions for nonconvex optimization. Journ. of Global Opt., 1998, V. 12, pp. 415–434.

817. Tikhonov A. N., Leonov A. S., Yagola A. G. Nonlinear ill-posed problems. Chapman and Hall, London, 1998.
818. Vasiliev O. V. Optimization methods. World Federation Publishers Company, ICN, Atlanta, USA, 1996.
819. Zelikin M. I., Borisov V. F. Theory of Chattering Control with applications to Astronautics. Robotics, Economics and Engineering. Birkhauser, Boston, 1994.

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

УКАЗАТЕЛЬ ОБОЗНАЧЕНИЙ

$\mathbb{R}$ — числовая ось

$[a, b] = \{x \in \mathbb{R}: a \leqslant x \leqslant b\}$ — отрезок

$(a, b) = \{x \in \mathbb{R}: a < x < b\}$ — интервал

$(a, b] = \{x \in \mathbb{R}: a < x \leqslant b\}$, $[a, b) = \{x \in \mathbb{R}: a \leqslant x < b\}$ — полуинтервалы

$\mathbb{R}^n$ — n-мерное линейное пространство векторов-столбцов $x = \begin{pmatrix} x^1 \\ \dots \\ x^n \end{pmatrix}$

x^i — i-ая координата вектора $x \in \mathbb{R}^n$, $i = 1, \dots, n$

$x^\top = (x^1, \dots, x^n)$ — вектор-строка, полученная транспонированием вектора-столбца x

$|x|_p = \left(\sum\limits_{i=1}^{n} |x^i|^p \right)^{1/p}$, $1 \leqslant p < \infty$, $|x|_\infty = \max\limits_{1 \leqslant i \leqslant n} |x^i|$ — нормы вектора x в $\mathbb{R}^n$

E^n — n-мерное евклидово пространство, представляющее собой пространство $\mathbb{R}^n$, оснащенное скалярным произведением $\langle x, y \rangle = \sum\limits_{i=1}^{n} x^i y^i$

$|x| = |x|_2 = |x|_{E^n} = \left(\sum\limits_{i=1}^{n} |x^i|^2 \right)^{1/2}$ — евклидова норма вектора

$e_i = (0, \dots, 0, 1, 0, \dots, 0)^\top$ — вектор-столбец, i-я координата которого равна 1, остальные координаты равны нулю, $i = 1, \dots, n$

$\{e_1, \dots, e_n\}$ — ортонормированный базис пространства E^n

$A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \dots & \dots & \dots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} = \{a_{ij}, i = 1, \dots, m, j = 1, \dots, n\} = \begin{pmatrix} a^1 \\ \dots \\ a^n \end{pmatrix} = (A_1, \dots, A_m)$ — матрица размера $m \times n$ с элементами a_{ij}, где

$a_i = (a_{i1}, \dots, a_{in})$ — i-я строка матрицы A,

$A_j = (a_{1j}, \dots, a_{mj})^\top$ — j-й столбец матрицы A

A — квадратная матрица n-го порядка, если $m = n$

$A^\top = \begin{pmatrix} a_{11} & \dots & a_{m1} \\ \dots & \dots & \dots \\ a_{1n} & \dots & a_{mn} \end{pmatrix} = (a_1^\top, \dots, a_n^\top) = \begin{pmatrix} A_1^\top \\ \dots \\ A_m^\top \end{pmatrix}$ — матрица размера $n \times m$, полученная транспонированием матрицы A размера $m \times n$

$A = A^\top$ — квадратная симметричная матрица

$I_n = (e_1, \dots, e_n) = \begin{pmatrix} e_1^\top \\ \dots \\ e_n^\top \end{pmatrix}$ — единичная матрица n-го порядка со столбцами $e_1, \dots, e_n$

$\det A$ — определитель квадратной матрицы A

A^{-1} — обратная матрица для квадратной матрицы A с $\det A \neq 0$

$\operatorname{rang} A$ — ранг матрицы A

$Ax = ((Ax)^1, \dots, (Ax)^m)^\top$, где $(Ax)^i = \sum\limits_{j=1}^{n} a_{ij} x^j$, $i = 1, \dots, m$ — произведение матрицы размера $m \times n$ на вектор $x \in E^n$

$\|A\| = \max\limits_{|x|_{E^n} \leqslant 1} |Ax|_{E^m}$ — норма матрицы A размера $m \times n$

AB — произведение матрицы $A = \{a_{ij}\}$ размера $m \times n$ на матрицу $B = \{b_{ij}\}$ размера $n \times q$ является матрицей $C = \{c_{ij}\}$ размера $m \times q$ с элементами $c_{ij} = \sum\limits_{k=1}^{n} a_{ik} b_{kj}$, $i = 1, \dots, m$, $j = 1, \dots, q$

$A \geqslant 0$ — неотрицательно определенная симметричная матрица, если $\langle Ax, x \rangle \geqslant 0$ $\forall x \in E^n$

$A > 0$ — положительно определенная симметричная матрица, если $\langle Ax, x \rangle > 0$ $\forall x \in E^n$, $x \neq 0$

$A \leqslant 0$ — неположительно определенная матрица, если $-A \geqslant 0$

$A < 0$ — отрицательно определенная матрица, если $-A > 0$

$x = (x^1, \dots, x^n)^\top \geqslant 0$, если $x^i \geqslant 0$ $\forall i = 1, \dots, n$

$x = (x^1, \dots, x^n)^\top > 0$, если $x^i > 0$ $\forall i = 1, \dots, n$

$x = (x^1, \dots, x^n)^\top \geqslant y = (y^1, \dots, y^n)^\top$, если $x - y \geqslant 0$

$x = (x^1, \dots, x^n)^\top > y = (y^1, \dots, y^n)^\top$, если $x - y > 0$

$E_+^n = \{x \in E^n: x \geqslant 0\}$ — неотрицательный ортант пространства E^n

$x_+ = (x_+^1, \dots, x_+^n)^\top$, где $x_+^i = \max\{0; x^i\}$, $i = 1, \dots, n$

$x = (x^1, \dots, x^n)^\top \succ 0$ — лексикографически положительный вектор, если $x \neq 0$ и первая ненулевая координата этого вектора положительна

$x = (x^1, \dots, x^n)^\top \succ y = (y^1, \dots, y^n)^\top$, если $x - y \succ 0$

$x_* = \operatorname{lex}\min\limits_{i \in M} x_i$ — лексикографический минимум множества векторов $\{x_i, i \in M\}$, если для каждого номера $i \in M$ либо $x_i \succ x_*$, либо $x_i = x_*$

$S = S(v, B) = \begin{pmatrix} \Gamma \\ \Delta \end{pmatrix}$ — симплекс-таблица угловой точки v с базисом B, где $\Gamma = (\Gamma_1, \dots, \Gamma_r)^\top$, $\Gamma_i = (\gamma_{i0}, \gamma_{i1}, \dots, \gamma_{in})$, $i = 1, \dots, r$; $\Delta = (\Delta_0, \Delta_1, \dots, \Delta_n)$ (подробности см. в § 3 гл. 3)

$S = \begin{pmatrix} \Gamma \\ \Delta \end{pmatrix} \overset{\Gamma}{\succ} 0$ — лексикографически положительная симплекс-таблица, если $\Gamma_i \succ 0$ $\forall i = 1, \dots, r$

$S_1 = \begin{pmatrix} \Gamma_1 \\ \Delta_1 \end{pmatrix} \succ S_2 = \begin{pmatrix} \Gamma_2 \\ \Delta_2 \end{pmatrix}$ — симплекс-таблица S_1 лексикографически больше симплекс-таблицы S_2, если $\Delta_1 \succ \Delta_2$

$\overline{X}$ — замыкание множества X

$\operatorname{Гр} X$ — граничные точки множества X

$\operatorname{int} X$ — внутренние точки множества X

$\operatorname{diam} X = \sup\limits_{x, y \in X} |x - y|$ — диаметр множества X

$\operatorname{aff} X$ — аффинная оболочка множества X

$\operatorname{Lin} X$ — несущее подпространство множества X

$\operatorname{co} X$ — выпуклая оболочка множества X

$\operatorname{ri} X$ — относительная внутренность множества X

$\dim X$ — размерность множества X

$\operatorname{mes} X$ — лебегова мера множества X

$|X|$ — количество элементов множества X

$\rho(x, X) = \inf\limits_{y \in X} |x - y|$ — расстояние от точки x до множества X

$X \cup Y$ — объединение двух множеств X и Y

$X \cap Y$ — пересечение двух множеств X и Y

$X + Y$ — сумма двух множеств X и Y

$X - Y$ — разность двух множеств X и Y

αX — произведение числа α на множество X

$X \times Y$ — прямое (декартово) произведение двух множеств X и Y

$\varnothing$ — пустое множество

$O(v, \varepsilon) = \{x \in E^n: |x - v| < \varepsilon\}$ — ε-окрестность точки v, или открытый шар радиуса $\varepsilon > 0$ с центром в точке v

$S(v, R) = \{x \in E^n: |x - v| \leqslant R\}$ — замкнутый шар радиуса R с центром в точке v

$\Gamma = \Gamma(c, \gamma) = \{x \in E^n: \langle c, x \rangle = \gamma\}$ — гиперплоскость с нормальным вектором $c \neq 0$, γ — заданное число

$\Gamma^+ = \{x \in E^n: \langle c, x \rangle > \gamma\}$ — открытое положительное полупространство гиперплоскости Γ

$\Gamma^- = \{x \in E^n: \langle c, x \rangle < \gamma\}$ — открытое отрицательное полупространство гиперплоскости Γ

$\overline{\Gamma^+} = \{x \in E^n: \langle c, x \rangle \geqslant \gamma\}$ — замкнутое положительное полупространство гиперплоскости Γ

$\overline{\Gamma^-} = \{x \in E^n: \langle c, x \rangle \leqslant \gamma\}$ — замкнутое отрицательное полупространство гиперплоскости Γ

K — конус с вершиной в нуле

$L^\perp$ — ортогональное дополнение подпространства $L \subseteq E^n$

$\Pi(\overline{\lambda})$ — сопровождающее подпространство точки $\overline{\lambda}$

$\mathcal{P}_X(z)$ — проекция точки $z \in E^n$ на множество X

$\mathcal{P}_X^G(z)$ — проекция точки $z \in E^n$ на множество X в метрике G

$f(x) \to \inf$, $x \in X$ — краткая символическая запись задачи минимизации функции $f(x)$ на множестве X

$f_* = \inf\limits_{x \in X} f(x)$ — нижняя грань функции $f(x)$ на множестве X

$X_* = \{x \in X: f(x) = f_* > -\infty\}$ — множество точек минимума функции $f(x)$ на X

$f(x) \to \sup$, $x \in X$ — краткая символическая запись задачи максимизации функции $f(x)$ на множестве X

$f^* = \sup\limits_{x \in X} f(x)$ — верхняя грань функции $f(x)$ на множестве X

$X^* = \{x \in X: f(x) = f^* < +\infty\}$ — множество точек максимума функции $f(x)$ на X

$\frac{\partial f(x)}{\partial x^i} = f_{x^i}(x)$ — частная производная функции $f(x) = f(x^1, \ldots, x^n)$ в точке x по переменной x^i, $i = 1, \ldots, n$

$f'(x) = \left\{ \frac{\partial f(x)}{\partial x^1}, \ldots, \frac{\partial f(x)}{\partial x^n} \right\}$ — градиент функции $f(x)$ в точке x

$\frac{\partial^2 f(x)}{\partial x^i \partial x^j} = f_{x^i x^j}(x)$ — частная производная функции $f(x) = f(x^1, \ldots, x^n)$ в точке x по переменным x^i, x^j, $i, j = 1, \ldots, n$

$f''(x) = \left\{ \frac{\partial^2 f(x)}{\partial x^i \partial x^j}, i, j = 1, \ldots, n \right\}$ — вторая производная функции $f(x)$ в точке x, она является симметричной квадратной матрицей n-го порядка

$\partial f(x)$ — субдифференциал функции $f(x)$ в точке x

$\frac{df(x)}{de} = \lim\limits_{t \to +0} \frac{f(x+te) - f(x)}{t}$ — производная функции $f(x)$ в точке x по направлению e, $|e| = 1$

$\dot{x}(t) = \frac{dx(t)}{dt}$ — первая производная функции $x(t)$ по времени t

$\ddot{x}(t) = \frac{d^2 x(t)}{dt^2}$ — вторая производная функции $x(t)$ по времени t

$\mathrm{pr}\,(z)$ — значение проксимального оператора в точке z

$\lim\limits_{k \to \infty} x_k$ — предел последовательности $\{x_k\} = (x_1, x_2, \ldots < x_k, \ldots)$

$\overline{\lim\limits_{k \to \infty}} x_k$ — верхний предел последовательности $\{x_k\}$

$\underline{\lim\limits_{k \to \infty}} x_k$ — нижний предел последовательности $\{x_k\}$

$\lim\limits_{x \to a} f(x)$ — предел функции $f(x)$ при $x \to a$

$\overline{\lim\limits_{x \to a}} f(x)$ — верхний предел функции $f(x)$ при $x \to a$

$\underline{\lim\limits_{x \to a}} f(x)$ — нижний предел функции $f(x)$ при $x \to a$

$f(a+0) = \lim\limits_{t \to a+0} f(t)$ — предел функции $f(x)$ одной переменной t при стремлении t к точке a справа

$f(a-0) = \lim\limits_{t \to a-0} f(t)$ — предел функции $f(x)$ при стремлении t к точке a слева

$\mathcal{L}s\limits_{k \to \infty} \Pi_k$ — верхний предел последовательности множеств $\{\Pi_k\}$

$O(t)$ — величина, определенная в окрестности точки $t = 0$ и такая, что $\left|\frac{O(t)}{t}\right| \leqslant c$, где c — некоторая неотрицательная постоянная

$o(t)$ — величина, определенная в окрестности точки $t = 0$ и такая, что $\lim\limits_{t \to 0} \frac{o(t)}{t} = 0$

$C(X)$ — пространство непрерывных функций $f(x)$ на замкнутом ограниченном множестве X с нормой $\|f\|_C = \max\limits_{x \in X} |f(x)|$

$C'(X)$ — пространство непрерывно дифференцируемых функций на множестве X

$C^2(X)$ — пространство дважды непрерывно дифференцируемых функций на множестве X

$C^{1,1}(X)$ — пространство непрерывно дифференцируемых функций на множестве X, градиент $f'(x)$ которых удовлетворяет условию Липшица на X (определение 1 § 6 гл. 2)

$Q(L)$ — множество функций, удовлетворяющих условию Липшица на множестве X с константой L

$L_1[a, b]$ — пространство функций $f = f(t)$, интегрируемых по Лебегу на отрезке $[a, b]$, с нормой $\|f\|_{L_1[a,b]} = \|f\|_{L_1} = \int\limits_a^b |f(t)|dt$

$L_2[a, b]$ — пространство функций $f = f(t) \in L_1[a, b]$, $|f(t)|^2 \in L_1[a, b]$ со скалярным произведением $\langle f, g \rangle_{L_2} = \int\limits_a^b f(t)g(t)dt$ и нормой $\|f\|_{L_2} = \sqrt{\langle f, f \rangle_{L_2}}$

$L_p[a, b]$, $1 < p < \infty$ — пространство функций $f = f(t) \in L_1[a, b]$, $|f(t)|^p \in L_1[a, b]$ с нормой $\|f\|_{L_p} = \left(\int\limits_a^b |f(t)|^p dt \right)^{1/p}$

$L_\infty[a, b]$ — пространство ограниченных измеримых функций $f = f(t)$ с нормой $\|f\|_{L_\infty}$ (подробности см. в § 1 гл. 6)

$L_2^r[a, b]$ — пространство r-мерных вектор-функций $f = f(t) = (f^1(t), \ldots, f^r(t))$, $f^i(t) \in L_2[a, b]$, $i = 1, \ldots, r$, со скалярным произведением $\langle f, g \rangle_{L_2^r} = \int\limits_a^b \langle f(t), g(t) \rangle_{E^r} dt$ и нормой $\|f\|_{L_2^r} = \sqrt{\langle f, f \rangle_{L_2^r}}$

$L_p^r[a, b]$, $1 \leqslant p \leqslant +\infty$ — пространство r-мерных вектор-функций $f = f(t) = (f^1(t), \ldots, f^r(t))$, $f^i(t) \in L_p[a, b]$, $i = 1, \ldots, r$, с нормой $\|f\|_{L_p^r} = \left(\int\limits_a^b |f(t)|_{E^r}^p dt \right)^{1/p}$ при $1 \leqslant p < \infty$ и $\|f\|_{L_\infty^r} = \| |f(t)|_{E^r} \|_{L_\infty}$

$H = H(x, u, t, \psi, a_0)$ — функция Гамильтона — Понтрягина

$l = l(x, y, t, T, a)$ — малый Лагранжиан

$\mathcal{L} = \mathcal{L}(x, \lambda)$ — функция Лагранжа

$L = L(x, \lambda)$ — функция Лагранжа в нормальной форме

$\Lambda(v)$ — конус Лагранжа точки v локального минимума

$\Lambda^-(v)$ — конус Лагранжа точки v локального максимума

$\Lambda_a(v)$ — конус Арутюнова точки v локального минимума

$\Lambda_a^-(v)$ — конус Арутюнова точки v локального максимума

$K(v)$ — конус критических направлений в точке v

$\ker G'(v) = \{h \in E^n : \langle g_i'(v), h \rangle = 0, i \in M\}$ — ядро отображения $G'(v) = \{g_i'(v), i \in M\}$, где $G(x) = \{g_i(x), i \in M\}$

$\forall$ — квантор общности

$\exists$ — квантор существования

Издательство «ФАКТОРИАЛ Пресс»

117449, Москва, а/я 331 factorial @mail.compnet.ru www.compnet.ru/factorial

ИМЕЮТСЯ В ПРОДАЖЕ:

Мищенко А. С., Фоменко А. Т. Курс дифференциальной геометрии и топологии. 448 с., 2000
Винберг Э.Б. Курс алгебры. 530 с., 2001
Рыжков В. В. Лекции по аналитической геометрии. 208 с., 2000
Каток А. Б., Хасселблат Б. Введение в современную теорию динамических систем. 768 с., 1999.
Стоянов Й. Контрпримеры в теории вероятностей. 288 с., 1999.
Гильберт Д. Избранные труды. Т. 1, 2. 575 + 698 с., 1998.
Демидов Е. Е. Квантовые группы. 128 с., 1998.
Постников М. М. Лекции по геометрии. Риманова геометрия. 496 с., 1998.
Зеликин М. И. Однородные пространства и уравнения Риккати в вариационном исчислении. 351 с., 1998.
Дьяченко М. И., Ульянов П. Л. Мера и интеграл. 160 с., 1998.
Прасолов В. В., Соловьев Ю. П. Эллиптические функции и алгебраические уравнения. 288 с., 1998.
Сосинский А. Б. Как написать математическую статью по-английски? 112 с., 1998.
Топологические методы в теории гамильтоновых систем. (Сборник работ под ред. Болсинова А. В., Фоменко А. Т., Шафаревича А. И.) 320 с., 1999.
Виноградова И. А., Олехник С. Н., Садовничий В. А. Математический анализ в задачах и упражнениях (несобственные интегралы и ряды Фурье). 488 с., 1998.
Виноградова И. А., Олехник С. Н., Садовничий В. А. Математический анализ в задачах и упражнениях (числовые и функциональные ряды). 478 с., 1996.
Сэвидж Дж. Сложность вычислений. 368 с., 1998.
Мерфи Дж. C^*-алгебры и теория операторов. 356 с., 1997.
Соловьев Ю. П., Троицкий Е. В. C^*-алгебры и эллиптические операторы в дифференциальной топологии. 352 с., 1997.
Трофимов В. В., Фоменко А. Т. Алгебра и геометрия интегрируемых гамильтоновых дифференциальных уравнений. 448 с., 1995.
Зайцев В. Ф., Полянин А. Д. Справочник по линейным обыкновенным дифференциальным уравнениям. 304 с., 1997.
Зайцев В. Ф., Полянин А. Д. Справочник по нелинейным обыкновенным дифференциальным уравнениям. 512 с., 1997.
Полянин А. Д., Манжиров А. В. Справочник по интегральным уравнениям. Точные решения. 431 с., 1998.
Желобенко Д. П. Введение в теорию представлений. 136 с., 2002.
Мануйлов В. М., Троицкий Е. В. C^*-гильбертовы модули. 224 с., 2002.
Тертычный-Даури В. Ю. Стохастическая механика. 464 с., 2002.
Смирнов В. А. Симплициальные и операдные методы в теории гомотопий. 272 с., 2002.

ГОТОВЯТСЯ К ПЕЧАТИ:

Кнепп А. Эллиптические кривые
Морган Дж. Инварианты Зайберга—Виттена и их применения к топологии гладких четырехмерных многообразий.

Все книги можно приобрести в интернет-магазине: **http://www.bolero.ru/**

Аннотации, оглавления, введения к книгам можно найти на www-странице издательства: http://www.compnet.ru/factorial.