

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ФАКУЛЬТЕТ ВЫЧИСЛИТЕЛЬНОЙ МАТЕМАТИКИ И КИБЕРНЕТИКИ

Лекции по курсу “Численные методы”

Лектор: Н. И. Ионкин
3 курс, 3 поток

Москва, 2009

Оглавление

Введение	3
Колесо Самарского	3
Содержание курса	3
Список литературы	4
I Численные методы линейной алгебры	5
§1 Введение	5
§2 Разложение матрицы на множители. Связь этого разложения с методом Гаусса	6
Разложение матрицы на множители	6
Связь метода Гаусса с разложением матрицы на множители	8
§3 Обращение матриц методом Гаусса-Жордана	9
§4 Метод квадратного корня	11
§5 Примеры и канонический вид итерационных методов решения СЛАУ	13
§6 Теоремы о сходимости итерационных методов	16
§7 Оценка скорости сходимости итерационных методов	22
§8 Исследование сходимости попеременно треугольного итерационного метода	26
§9 Методы решения задач на собственные значения	28
Степенной метод решения частичной проблемы собственных значений	29
Метод обратных итераций	31
Метод обратных итераций со сдвигом	33
§10 Приведение матрицы к верхней почти треугольной форме (ВПТФ)	33
§11 Понятие о QR-алгоритме. Решение полной проблемы собственных значений.	38
QR-алгоритм	39
Свойства QR-алгоритма	40
II Интерполирование и приближение функций	42
§1 Постановка задачи интерполирования	42
Интерполирование алгебраическими полиномами	42
§2 Интерполяционная формула Лагранжа	43
§3 Интерполяционная формула Ньютона	44
§4 Интерполирование с кратными узлами. Интерполяционная формула Эрмита	45
Погрешность полинома Эрмита	47
§5 Использование полинома Эрмита третьей степени для получения точной оценки погрешности квадратурной формулы Симпсона	47
§6 Наилучшее среднеквадратичное приближение функции	50

III Численное решение нелинейных уравнений и систем нелинейных уравнений	53
§1 Введение	53
§2 Метод простой итерации	54
Метод Эйткена (ускорение сходимости)	55
§3 Метод Ньютона и метод секущих	55
Метод Ньютона для системы уравнений	56
Метод секущих	57
§4 Сходимость метода Ньютона и оценка сходимости	58
IV Разностные методы решения задач математической физики	60
§1 Разностные схемы для первой краевой задачи для уравнения теплопроводности	60
Явная разностная схема	61
Чисто неявная разностная схема (схема с опережением)	64
Симметричная разностная схема (схема Кранка-Никольсона)	66
Задача Штурма-Лиувилля	67
Разностная схема с весами. Погрешность аппроксимации.	70
Разностные методы для уравнения Пуассона	72
§2 Разностные схемы для уравнения Пуассона (задача Дирихле)	73
§3 Сходимость разностной задачи Дирихле	74
§4 Методы решения разностной задачи Дирихле	76
Простая итерация (метод Якоби)	76
Метод Зейделя	76
Попеременно-треугольный итерационный метод (метод Самарского)	77
§5 Основные понятия теории разностных схем. Аппроксимация. Устойчивость. Сходимость.	77
V Методы решения ОДУ и систем ОДУ	80
§1 Примеры численных методов решения задачи Коши	80
Общий m -этапный метод Рунге-Кутты	82
§2 Оценка точности на примере 2-х этапного метода Рунге-Кутты	83
§3 Многошаговые разностные методы	85
§4 Понятие устойчивости разностных методов	87
§5 Жесткие системы ОДУ	90
§6 Дальнейшее определение устойчивости и примеры разностных схем. Интегрирование жестких схем ДУ	91
Явная схема Эйлера	92
Неявная схема Эйлера	92
Симметричная схема	93

Введение

Колесо Самарского



При изучении объектов окружающего мира математическими методами используют принцип “колеса Самарского”, изображенный на рисунке. В данном курсе рассматривается фаза “разработка алгоритма” этого принципа.

Содержание курса

Глава I Численные методы линейной алгебры.

Глава II Интерполирование и приближение функций.

Глава III Решение нелинейных уравнений и систем.

Глава IV Разностные схемы для уравнений математической физики.

Глава V Решение задачи Коши для обыкновенных дифференциальных уравнений и систем обыкновенных дифференциальных уравнений.

Список литературы

1. Самарский А. А., Гулин А. В. “Численные методы” М. Наука 1983
2. Самарский А. А. “Теория разностных схем” М. Наука 1983
3. Бахвалов Н. С., Жидков Н. П., Кобельков Г. М. “Численные методы” М. Наука 1973
4. Самарский А. А. “Введение в численные методы” М. Наука 1982
5. Калиткин Н. Н. “Численные методы” М. Наука 1978
6. Самарский А. А., Николаев И. С. “Методы решения сеточных уравнений”
7. И. С. Березин, Н. П. Жидков “Методы вычислений”

Глава I

Численные методы линейной алгебры

§1 Введение

Рассмотрим систему линейных алгебраических уравнений (СЛАУ) в матричном виде

$$Ax = f, \quad (1)$$

где A — матрица размера $(m \times m)$, $|A| \neq 0$,

$$x = (x_1, \dots, x_m)^T,$$

$$f = (f_1, \dots, f_m)^T.$$

Из невырожденности матрицы A следует, что решение системы (1) существует и единственно. Выделяют две группы методов поиска решения СЛАУ:

1. Прямые (точные) методы. Примеры: метод Гаусса (требует $\sim m^3$ действий), формула Крамера (требует $\sim m!$ действий), метод квадратного корня. Эти методы позволяют за конечное число действий получить точное решение.
2. Итерационные (метод последовательных приближений).

x_0 — первое приближение,

$$x_n \xrightarrow{n \rightarrow \infty} x.$$

При работе с итерационными методами задача обычно ставится следующим образом: для данного $\epsilon > 0$ найти $n_0(\epsilon)$ такое, что

$$|x_n - x| < \epsilon \quad \forall n \geq n_0$$

Мы будем также рассматривать задачу на собственные значения. Она формулируется так: найти все такие числа λ и ненулевые векторы x , что для данной матрицы A выполняется

$$Ax = \lambda x. \quad (2)$$

λ называется собственным значением, x — собственным вектором матрицы A .

§2 Разложение матрицы на множители. Связь этого разложения с методом Гаусса

$$Ax = f, |A| \neq 0 \quad (1)$$

Подсчитаем число действий, необходимое для решения уравнения (1) методом Гаусса. Действием будем считать умножение или деление.

1. Прямой ход метода Гаусса:

$$\bullet A \Rightarrow \begin{pmatrix} 1 & \times & \cdots & \times \\ 0 & 1 & \cdots & \times \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \frac{m^3 - m}{3} \text{ действий}$$

Крестиками отмечены в общем случае ненулевые элементы.

$$\bullet \text{Преобразование правых частей: } \frac{m(m+1)}{2} \text{ действий}$$

2. Обратный ход

$$\bullet \frac{m(m-1)}{2} \text{ действий}$$

3. Всего $\frac{m^3}{3} + m^2 - \frac{m}{3}$ действий

Разложение матрицы на множители

Зададимся целью представить матрицу A в виде

$$A = B \cdot C, \quad (2)$$

где

$$B = \begin{pmatrix} b_{11} & 0 & \cdots & 0 \\ b_{21} & b_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mm} \end{pmatrix}$$

— нижнетреугольная матрица,

$$C = \begin{pmatrix} 1 & c_{12} & \cdots & c_{1m} \\ 0 & 1 & \cdots & c_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

— верхнетреугольная матрица с единицами на главной диагонали.

По формуле для элемента произведения матриц:

$$a_{ij} = \sum_{l=1}^m b_{il} c_{lj}$$

Перепишем предыдущее выражение, выделив слагаемое с c_{ij} :

$$a_{ij} = \sum_{l=1}^{i-1} b_{il}c_{lj} + b_{ii}c_{ij} + \sum_{l=i+1}^m b_{il}c_{lj}$$

Из вида матрицы B следует, что

$$\sum_{l=i+1}^m b_{il}c_{lj} = 0$$

Таким образом, предполагая, что $b_{ii} \neq 0$, получим

$$c_{ij} = \frac{a_{ij} - \sum_{l=1}^{i-1} b_{il}c_{lj}}{b_{ii}}, \quad i < j \quad (3)$$

Теперь из формулы элемента произведения матриц выделим слагаемое с b_{ij} :

$$a_{ij} = \sum_{l=1}^{j-1} b_{il}c_{lj} + b_{ij}c_{jj} + \sum_{l=j+1}^m b_{il}c_{lj}$$

Из вида матрицы C следует, что

$$\sum_{l=j+1}^m b_{il}c_{lj} = 0$$

Таким образом, можно записать:

$$b_{ij} = a_{ij} - \sum_{l=1}^{j-1} b_{il}c_{lj}, \quad i \geq j \quad (4)$$

Легко видеть, что формулы (3) и (4) позволяют вычислить все элементы матриц B и C .

Докажем существование и единственность факторизации матрицы A .

Утверждение. Пусть все главные миноры матрицы A отличны от 0:

$$A_1 = a_{11} \neq 0, A_2 = \begin{vmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{vmatrix} \neq 0, \dots, A_i \neq 0, \forall i = \overline{1, m}$$

Тогда разложение в форме (2) существует и единственно.

Доказательство. По свойству определителей:

$$\det A_i = \det B_i \cdot \det C_i$$

Так как все элементы главной диагонали матрицы C равны 1, то $\det C_i = 1, \forall i = \overline{1, m}$.

$$\det B_i = b_{11} \cdot b_{22} \cdot \dots \cdot b_{mm} \Rightarrow \det A_i = \det B_i$$

Исходя из этого утверждения, а также из вида определителя $\det B_i$, следует:

$$b_{ii} = \frac{\det A_i}{\det A_{i-1}}, \det A_0 = 1 \Rightarrow b_{11} = a_{11}$$

А так как все главные миноры (1) отличны от нуля, то все b_{ii} существуют и единственны по построению. \square

Связь метода Гаусса с разложением матрицы на множители

Рассмотрим метод Гаусса для решения СЛАУ с применением факторизации. Пусть $A = B \cdot C$, где B и C — матрицы в разложении в форме (2) (матрицы НПТФ и ВТФ соответственно). В этом случае исходная система будет выглядеть следующим образом:

$$B \cdot C \cdot x = f \quad (5)$$

Обозначим $C \cdot x = y$. Тогда система (5) распадается на две системы:

$$B \cdot y = f \quad (6)$$

$$C \cdot x = y, \quad (7)$$

Выпишем уравнения системы (6):

$$b_{i1}y_1 + b_{i2}y_2 + \dots + b_{ii}y_i = f_i, \forall i = \overline{1, m}$$

Поскольку $b_{ii} \neq 0$, мы можем разделить на b_{ii} , откуда получим:

$$y_i = \frac{f_i - \sum_{l=1}^{i-1} b_{il}y_l}{b_{ii}}, i = \overline{1, m} \quad (8)$$

Посчитаем число операций умножения и деления, требуемых для реализации полученной формулы. Для каждого уравнения этой системы получаем $(i - 1)$ операций умножения и 1 операцию деления. То есть при каждом фиксированном i получается i операций. А так как i может меняться от 1 до m , получаем формулу:

$$\sum_{i=1}^m i = 1 + 2 + \dots + (m - 1) + m = \frac{m \cdot (m + 1)}{2}$$

Мы получили в точности число шагов, требуемых для преобразования правых частей системы (6) при прямом ходе.

Для системы (7) оценка $\frac{m \cdot (m - 1)}{2}$ получается аналогично.

Задача. Показать, что факторизация матрицы A требует $\frac{m^3 - m}{3}$ операций умножения и деления.

Доказательство. Воспользуемся формулами для факторизации из предыдущего параграфа:

$$b_{ij} = a_{ij} - \sum_{l=1}^{j-1} b_{il}c_{lj}, \quad i \geq j$$

Для вычисления каждого b_{ij} потребуется $j - 1$ операция умножения. Отпустим индекс j :

$$\sum_{j=1}^i (j - 1) = \frac{i(i - 1)}{2}$$

Теперь отпустим индекс i :

$$\sum_{i=1}^m \frac{i(i - 1)}{2} = \frac{1}{2} \sum_{i=1}^m i^2 - \frac{1}{2} \sum_{i=1}^m i$$

Вторая сумма вычисляется элементарно, значение первой суммы нам известно из школьного курса – $\frac{m(m+1)(2m+1)}{6}$.

$$\frac{m(m+1)(2m+1)}{12} - \frac{m(m+1)}{4} = \frac{(m-1)m(m+1)}{6}$$

Далее следующая формула:

$$c_{ij} = \frac{a_{ij} - \sum_{l=1}^{i-1} b_{il}c_{lj}}{b_{ii}}, \quad i < j$$

Для вычисления каждого c_{ij} потребуется $j-1$ операция умножения и 1 операция деления. Отпустим индекс i :

$$\sum_{i=1}^{j-1} j = \frac{(j-1)j}{2}$$

Отпустим индекс j :

$$\begin{aligned} \frac{1}{2} \sum_{j=1}^m (j-1)j &= \frac{1}{2} \sum_{j=1}^m j^2 - \frac{1}{2} \sum_{j=1}^m j = \frac{m(m+1)(2m+1)}{12} - \frac{m(m+1)}{4} = \\ &= \frac{(m-1)m(m+1)}{6} \end{aligned}$$

Суммируя с предыдущим результатом, получаем окончательный ответ:

$$2 \frac{(m-1)(m+1)m}{6} = \frac{(m^3 - m)}{3}$$

□

Получается, что суммарная сложность метода совпадает со сложностью классического метода Гаусса.

Замечание. Факторизация дает существенный выигрыш в том случае, если решается СЛАУ с построяной матрицей A и меняющимися правыми частями f .

§3 Обращение матриц методом Гаусса-Жордана

Рассмотрим систему линейных алгебраических уравнений:

$$Ax = f, A \in R^{m \times m}, |A| \neq 0 \quad (1)$$

Определение. Матрица A^{-1} называется обратной к матрице A , если выполнено:

$$A \cdot A^{-1} = A^{-1} \cdot A = E$$

Обозначим $X = A^{-1}$:

$$A \cdot X = E, X = x_{ij} \quad i, j \in \overline{1, m}$$

Запишем СЛАУ порядка m (с m^2 неизвестными):

$$A \cdot X = E \quad (2)$$

Для решения подобной системы классическим методом Гаусса потребуется порядка m^6 операций. Покажем, что число действий можно снизить до m^3 . Для этого обозначим:

$$\delta_{ij} = \sum_{l=1}^m a_{il} X_{lj} \quad (3)$$

$$X^{(j)} = (x_{1j}, x_{2j}, \dots, x_{mj})^T, j = \overline{1, m} \quad (4)$$

$$\delta^{(j)} = (\underbrace{0}_1, 0, \dots, \underbrace{0}_{j-1}, \underbrace{1}_j, \underbrace{0}_{j+1}, \dots, \underbrace{0}_m) \quad (5)$$

Теперь система (2) сводится к m системам с m уравнениями в каждой. При этом матрица A одна и та же для всех систем:

$$A \cdot x^{(j)} = \delta^{(j)}, j = \overline{1, m} \quad (6)$$

Факторизуем матрицу A и перепишем полученные системы (6):

$$A = B \cdot C$$

$$B = \{b_{i,j}, i \geq j; i, j \in \overline{1, m}\} \text{ (Нижняя почти треугольная форма)}$$

$$C = \left\{ \begin{array}{l} 1, i = j; \\ c_{i,j}, i < j; \\ 0, i > j; \end{array} \right\} (i, j \in \overline{1, m}) \text{ (Верхняя треугольная форма)}$$

Обозначим $Cx^{(j)} = y^{(j)}$. Система уравнений примет вид:

$$By^{(j)} = \delta^{(j)} \quad (7)$$

$$Cx^{(j)} = y^{(j)} \quad (8)$$

Ещё раз отметим тот факт, что матрица A остается одинаковой для всех m систем. Соответственно, факторизацию матрицы A нужно проводить только один раз. При фиксированном j для решения формул (7) и (8) требуется m^2 операций. А поскольку общее количество систем равно m , общая сложность решения составляет m^3 . Суммируя этот показатель с числом операций, требуемых для факторизации матрицы A ($\frac{m^3-m}{3}$) получаем общую сложность $\frac{4m^3-m}{3}$. Полученная оценка — не предел. Для ее улучшения рассмотрим структуру матрицы B подробнее.

Распишем систему (7):

$$\begin{aligned} b_{11}y_1^{(j)} = 0 &\Rightarrow y_1^{(j)} = 0 \\ b_{21}y_1^{(j)} + b_{22}y_2^{(j)} = 0 &\Rightarrow y_2^{(j)} = 0 \\ &\dots \\ y_{i-1}^{(j)} &= 0 \\ y_i^{(j)} &= 1 \end{aligned}$$

Таким образом $y_i^{(j)} = 0, i < j; y_j^{(j)} = \frac{1}{b_{jj}}, i = j$.

Оставшиеся уравнения:

$$b_{i,j}y_j^{(j)} + b_{i,j+1}y_{j+1}^{(j)} + \dots + b_{i,i}y_i^{(j)} = 0, b_{ii} \neq 0 \Rightarrow$$

$$y_i^{(j)} = -\frac{\sum_{l=j}^{i-1} b_{i,l} y_l^{(j)}}{b_{ii}}, i = \overline{j+1, m}.$$

Оценим число операций для реализации указанного метода решения. Фиксируем индексы i и j : в этом случае нам потребуется $(i - j)$ умножений и 1 деление. Сначала отпустим индекс i ($i = \overline{j+1, m}$), тогда число операций будет равно:

$$(m - j + 1) + (m - j) + \dots + 2 + 1 = \frac{(m - j + 1)(m - j + 2)}{2}$$

Далее отпустим индекс j ($j = \overline{1, m}$):

$$\sum_{j=1}^m \frac{(m - j + 1)(m - j + 2)}{2} \quad (9)$$

Задача. Показать, что сумма (9) равна $\frac{m(m+1)(m+2)}{6}$.

Доказательство. Проведем замену: $k = m - j + 1$:

$$\sum_{k=1}^m \frac{k(k+1)}{2} = \sum_{k=1}^m \frac{k}{2} + \sum_{k=1}^m \frac{k^2}{2}$$

Откуда следует, что первая сумма равна $\frac{k(k+1)}{4}$, а вторая — $\frac{k(k+1)(2k+1)}{12}$, что в сумме и дает требуемую оценку. \square

Система (8) требует $\frac{m(m-1)}{2}$ действий, а, поскольку таких систем m штук (т.к. $j = \overline{1, m}$), то для их решения потребуется $\frac{m^2(m-1)}{2}$ операций. Таким образом, суммарное количество операций, требуемых на решение (1) равно:

$$\underbrace{\frac{m^3 - m}{3}}_{\text{факторизация (1)}} + \underbrace{\frac{m(m+1)(m+2)}{6}}_{\text{решение (7)}} + \underbrace{\frac{m^2(m-1)}{2}}_{\text{решение (8)}} = m^3$$

§4 Метод квадратного корня

Рассмотрим систему линейных алгебраических уравнений:

$$Ax = f, A \in R^{m \times m}, |A| \neq 0 \quad (1)$$

Определение. Матрица A называется эрмитовой (или самосопряженной) матрицей, если

$$A = A^*, a_{ij} = \overline{a_{ji}}$$

Пусть A - эрмитова матрица. Представим ее в виде:

$$A = S^* D S \quad (2)$$

где:

D - диагональная матрица $d_{ii} = \pm 1; i = \overline{1, m}$,

S - верхняя треугольная матрица $s_{ij} > 0, i < j, i, j = \overline{1, m}$,

S^* - матрица, сопряженная к S .

Рассмотрим метод нахождения матриц S и D на примере вещественных матриц второго порядка:

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \quad S = \begin{pmatrix} s_{11} & s_{12} \\ 0 & s_{22} \end{pmatrix}$$

$$S^* = \begin{pmatrix} s_{11} & 0 \\ s_{12} & s_{22} \end{pmatrix}, \quad D = \begin{pmatrix} d_{11} & 0 \\ 0 & d_{22} \end{pmatrix}$$

Перемножим матрицы S^*D и S :

$$DS = \begin{pmatrix} d_{11}s_{11} & d_{11}s_{12} \\ 0 & d_{22}s_{22} \end{pmatrix}$$

$$S^*DS = \begin{pmatrix} s_{11} & 0 \\ s_{12} & s_{22} \end{pmatrix} \begin{pmatrix} d_{11}s_{11} & d_{11}s_{12} \\ 0 & d_{22}s_{22} \end{pmatrix} = \begin{pmatrix} d_{11}s_{11}^2 & d_{11}s_{11}s_{12} \\ d_{11}s_{11}s_{12} & d_{11}s_{12}^2 + d_{11}s_{22}^2 \end{pmatrix}$$

Из (2) и условия равенства матриц получаем соотношения:

$$a_{11} = d_{11}s_{11}^2 \quad (3)$$

$$a_{12} = d_{11}s_{11}s_{12} \quad (4)$$

$$a_{22} = d_{11}s_{12}^2 + d_{11}s_{22}^2 \quad (5)$$

Из уравнения (3) имеем:

$$d_{11} = \text{sign}(a_{11})$$

$$s_{11} = \sqrt{|a_{11}|}$$

Из уравнения (4):

$$s_{12} = \frac{a_{12}}{s_{11}d_{11}}$$

Из уравнения (5):

$$d_{22} = \text{sign}(a_{22} - d_{11}s_{12}^2)$$

$$s_{22} = \sqrt{|a_{22} - d_{11}s_{12}^2|}$$

Таким образом, все элементы матриц S и D однозначно определены.

Рассмотрим теперь невырожденную эрмитову (или симметрическую) матрицу A с m строками и m столбцами и найдем для нее разложение в виде (2). Из того, что D является диагональной матрицей, получаем:

$$(DS)_{ij} = \sum_{l=1}^m d_{il}s_{lj} = d_{ii}s_{ij}, \quad s_{ii} > 0$$

Далее запишем выражение для a_{ij} :

$$a_{ij} = (S^*DS)_{ij} = \sum_{l=1}^m \bar{s}_{li}d_{ll}s_{lj}, \quad i \leq j \quad (6)$$

Выделим из суммы i -ый элемент:

$$a_{ij} = (S^*DS)_{ij} = \sum_{l=1}^{i-1} \bar{s}_{li}d_{ll}s_{lj} + \bar{s}_{ii}d_{ii}s_{ij} + \sum_{l=i+1}^m \bar{s}_{li}d_{ll}s_{lj} \quad (7)$$

При $j = i$ будем иметь:

$$a_{ii} = (S^*DS)_{ii} = \sum_{l=1}^{i-1} \bar{s}_{li} d_{ll} s_{li} + \bar{s}_{ii} d_{ii} s_{ii} + \sum_{l=i+1}^m \bar{s}_{li} d_{ll} s_{li}$$

В силу вида матрицы S $s_{li} = 0$, $l > i$, поэтому последняя из сумм будет равна 0. Учитывая равенство $\bar{s}_{li} s_{li} = |s_{li}|^2$, перепишем получившееся выражение в виде:

$$s_{ii}^2 d_{ii} = a_{ii} - \sum_{l=1}^{i-1} |s_{li}|^2 d_{ll}$$

Теперь можно записать формулы для элементов матриц S и D :

$$d_{ii} = \text{sign}\left(a_{ii} - \sum_{l=1}^{i-1} |s_{li}|^2 d_{ll}\right)$$

$$s_{ii} = \sqrt{\left|a_{ii} - \sum_{l=1}^{i-1} |s_{li}|^2 d_{ll}\right|}$$

Из (7) получим:

$$s_{ij} = \frac{a_{ij} - \sum_{l=1}^{i-1} \bar{s}_{li} d_{ll} s_{lj} - \sum_{l=i+1}^m \bar{s}_{li} d_{ll} s_{lj}}{\bar{s}_{ii} d_{ii}}$$

По данным формулам однозначно находятся все элементы матриц S и D .

Рассмотрим применение данного разложения к решению систем линейных алгебраических уравнений:

$$Ax = f$$

$$S^*DSx = f$$

Обозначим $Sx = y$, получим две системы линейных алгебраических уравнений:

$$\begin{cases} S^*Dy = f, \\ Sx = y; \end{cases}$$

Для решения этих двух систем потребуется порядка $\frac{m^3}{6}$ умножений и делений, а также m извлечений квадратного корня.

§5 Примеры и канонический вид итерационных методов решения СЛАУ

Рассмотрим СЛАУ:

$$Ax = f, \tag{1}$$

где A — матрица размера $(m \times m)$, $|A| \neq 0$,

$$x = (x_1, \dots, x_m)^T,$$

$$f = (f_1, \dots, f_m)^T.$$

Из невырожденности матрицы A следует, что решение системы (1) существует и единственно. Перепишем систему (1) в виде:

$$\sum_{j=1}^m a_{ij}x_j = f_i, \quad i = 1, \dots, m$$

Выделим из суммы i -ое слагаемое:

$$\sum_{j=1}^{i-1} a_{ij}x_j + a_{ii}x_i + \sum_{j=i+1}^m a_{ij}x_j = f_i, \quad i = 1, \dots, m$$

Пусть $a_{ii} \neq 0$, тогда можно выразить x_i :

$$x_i = \frac{f_i - \sum_{j=1}^{i-1} a_{ij}x_j - \sum_{j=i+1}^m a_{ij}x_j}{a_{ii}} \quad (2)$$

Обозначим через x_i^n i -ую компоненту n -ой итерации. Запишем метод Якоби (МЯ):

$$x_i^{n+1} = \frac{f_i}{a_{ii}} - \sum_{j=1}^{i-1} \frac{a_{ij}}{a_{ii}} x_j^n - \sum_{j=i+1}^m \frac{a_{ij}}{a_{ii}} x_j^n, \quad n = 0, 1, \dots; i = 1, \dots, m$$

Задан вектор x^0 - начальное приближение. Запишем метод Зейделя (МЗ):

$$x_i^{n+1} = \frac{f_i}{a_{ii}} - \sum_{j=1}^{i-1} \frac{a_{ij}}{a_{ii}} x_j^{n+1} - \sum_{j=i+1}^m \frac{a_{ij}}{a_{ii}} x_j^n, \quad n = 0, 1, \dots; i = 1, \dots, m;$$

Вектор x^0 также изначально задан.

Представим матрицу A в виде:

$$A = R_1 + D + R_2, \quad (3)$$

где

$$R_1 = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ a_{21} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & 0 \end{pmatrix}$$

— нижнетреугольная матрица с нулями на главной диагонали,

$$D = \begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{mm} \end{pmatrix}$$

— диагональная матрица,

$$R_2 = \begin{pmatrix} 0 & a_{12} & \cdots & a_{1m} \\ 0 & 0 & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$

— верхнетреугольная матрица с нулями на главной диагонали. Очевидно, такое разложение всегда осуществимо. Подставим представление (3) в (1):

$$(R_1 + D + R_2)x = f$$

$$Dx = f - R_1x - R_2x$$

Предположим теперь, что $\exists D^{-1}$. Тогда:

$$x = D^{-1}f - D^{-1}R_1x - D^{-1}R_2x$$

Метод Якоби можно записать следующим образом:

$$x^{n+1} = D^{-1}f - D^{-1}(R_1 + R_2)x^n$$

или

$$D(x^{n+1} - x^n) + Ax^n = f$$

Метод Зейделя:

$$x^{n+1} = D^{-1}f - D^{-1}R_1x^{n+1} - D^{-1}R_2x^n$$

или

$$(D + R_1)(x^{n+1} - x^n) + Ax^n = f$$

Из приведенных записей видно, что итерационные методы можно записать в различном виде. Поэтому целесообразно иметь единую форму записи итерационного метода.

Определение. Канонической формой записи двухслойного итерационного метода решения СЛАУ (1) называется его запись в виде:

$$B_{n+1} \frac{x^{n+1} - x^n}{\tau_{n+1}} + Ax^n = f, \quad n = 0, 1, \dots; x^0 - \text{задан} \quad (4)$$

$\tau_{n+1} > 0$ - итерационный параметр,

B_{n+1} - обратимая матрица.

Если $B_{n+1} = E$, то метод (4) называется явным. Если $B_{n+1} = B$, $\tau_{n+1} = \tau$, то метод (4) называется стационарным.

Метод простой итерации (ПИ, или метод релаксации) имеет следующий вид:

$$\frac{x^{n+1} - x^n}{\tau} + Ax^n = f, \quad \tau > 0$$

Рассмотрим более подробно попеременно-треугольный итерационный метод (ПТИМ):

$$(E + \omega R_1)(E + \omega R_2) \frac{x^{n+1} - x^n}{\tau_{n+1}} + Ax^n = f, \quad n = 0, 1, \dots; x^0 - \text{задан} \quad (5)$$

Здесь $\tau_{n+1} > 0$, $\omega > 0$ — итерационные параметры, $R_1 + R_2 = A$, где

$$R_1 = \begin{pmatrix} \frac{a_{11}}{2} & 0 & \cdots & 0 \\ a_{21} & \frac{a_{22}}{2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & \frac{a_{mm}}{2} \end{pmatrix},$$

$$R_2 = \begin{pmatrix} \frac{a_{11}}{2} & a_{12} & \cdots & a_{1m} \\ 0 & \frac{a_{22}}{2} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{a_{mm}}{2} \end{pmatrix}.$$

Реализация попеременно-треугольного итерационного метода может быть осуществлена по явным формулам. Пусть:

$$\begin{aligned} v^{n+1} &= \frac{x^{n+1} - x^n}{\tau_{n+1}} \\ w^{n+1} &= (E + \omega R_2)v^{n+1} \\ r^n &= f - Ax^n \end{aligned}$$

Тогда:

$$(E + \omega R_1)w^{n+1} = r^n, \text{ где } (E + \omega R_1) \text{— нижнетреугольная матрица,}$$

Из этого уравнения, путем обращения нижнетреугольной матрицы по явным формулам выписывается вектор w^{n+1} .

$$(E + \omega R_2)v^{n+1} = w^{n+1}, \text{ где } (E + \omega R_2) \text{— верхнетреугольная матрица,}$$

По известному вектору w^{n+1} , обращая верхнетреугольную матрицу по явным формулам можно найти вектор v^{n+1} , и далее:

$$x^{n+1} = x^n + \tau_{n+1}v^{n+1}.$$

Таким образом, несмотря на то, что ПТИМ - неявный итерационный метод, его реализация проста и сводится к попеременному обращению нижнетреугольной и верхнетреугольной матриц (отсюда - название метода).

§6 Теоремы о сходимости итерационных методов

Рассмотрим матричное уравнение вида

$$Ax = f, \tag{1}$$

где A — матрица размера $(m \times m)$, $|A| \neq 0$

Рассмотрим также матричное уравнение вида

$$B \frac{x^{n+1} - x^n}{\tau} + Ax^n = f, \tag{2}$$

где $n = 0, 1, 2, \dots$, $\exists B^{-1}$, и задан вектор начального приближения x^0

Рассмотрим линейное пространство H , такое что $\dim H = m$

Возьмем 2 произвольных вектора x и y из этого пространства:

$$x \in H, x = (x_1, x_2, \dots, x_m)^T$$

$$y \in H, y = (y_1, y_2, \dots, y_m)^T$$

Введем скалярное произведение векторов (x, y) по формуле:

$$(x, y) = \sum_{i=1}^m x_i y_i$$

Введем норму вектора x :

$$\|x\| = (x, x)^{\frac{1}{2}}$$

Замечание. Есть “слабые нормы”, которые обладают не поточечной близостью. Решения систем стараются брать в сильной норме, входные данные - в слабой, чтобы максимально расширить область применения метода.

Рассмотрим самосопряженный оператор $D = D^* > 0$

Определение. Будем говорить о скалярном произведении векторов x и y “в смысле D ”, если $(x, y)_D = (Dx, y)$

Это позволяет нам ввести энергетическую норму:

Определение. Энергетическая норма - норма, которая задается соотношением:

$$\|x\|_D = (Dx, x)^{\frac{1}{2}}$$

Вспомним некоторые свойства самосопряженного положительно определенного оператора D , известные из теории операторов:

1. $\exists D^{-1} = (D^{-1})^* > 0$
2. $\exists D^{\frac{1}{2}} = (D^{\frac{1}{2}})^* > 0$
3. $\exists D^{-\frac{1}{2}} = (D^{-\frac{1}{2}})^* > 0$

из этих свойств следует, что существует такое $\delta > 0$:

$$(Dx, x) \geq \delta \|x\|^2$$

В дальнейшем нам потребуется понятие положительной или неотрицательной определенности оператора.

Определение.

$$C > 0 \Leftrightarrow (Cx, x) > 0, \quad \forall x \neq 0$$

$$C \geq 0 \Leftrightarrow (Cx, x) \geq 0, \quad \forall x \in H$$

Задача. Дано: Оператор $C > 0$, H - вещественное линейное пространство с заданным скалярным произведением. Доказать, что:

$$(Cx, x) = \left(\frac{C + C^*}{2} x, x \right)$$

Доказательство. Для решения задачи воспользуемся следующими равенствами, верными для вещественного пространства H :

$$(C^*x, x) = (x, Cx) = (Cx, x), \quad \forall x \in H$$

Представим оператор C в виде суммы: $C = \frac{C+C^*}{2} + \frac{C-C^*}{2}$. Тогда:

$$\begin{aligned} (Cx, x) &= \left(\frac{C + C^*}{2} x, x \right) + \left(\frac{C - C^*}{2} x, x \right) = \\ &= \left(\frac{C + C^*}{2} x, x \right) + \frac{1}{2} \underbrace{\left((C^*x, x) - (Cx, x) \right)}_{=0} = \left(\frac{C + C^*}{2} x, x \right), \quad \forall x \in H \end{aligned}$$

□

Определение. Погрешность итерационного метода

$$V^n = x^n - x \quad (3)$$

Определение. Метод (2) сходится, если

$$\|V^n\| \rightarrow 0 (n \rightarrow \infty)$$

Из определения погрешности ясно, что решению матричного уравнения (2) на n -ой итерации соответствует вектор

$$x^n = V^n + x,$$

где x – точное решение системы.

Используя это соотношение, перепишем уравнение (2) через вектор погрешности:

$$B \frac{V^{n+1} - V^n}{\tau} + AV^n = 0, \quad (4)$$

где $n = 0, 1, 2, \dots$

Умножим уравнение (4) на B^{-1} слева:

$$\frac{V^{n+1} - V^n}{\tau} + B^{-1}AV^n = 0$$

Следовательно,

$$V^{n+1} = V^n - \tau B^{-1}AV^n = (E - \tau B^{-1}A)V^n = SV^n$$

Таким образом, получим матрицу S :

$$S = E - \tau B^{-1}A \quad (5)$$

Определение. S называется матрицей перехода от n -й итерации к $(n+1)$ -й

Теорема 1 (о сходимости итерационных методов). Итерационный метод (2) решения задачи (1) сходится при любом начальном приближении тогда и только тогда, когда все собственные значения матрицы S по модулю меньше единицы.

Замечание. Эта теорема хороша, но редко применима, т.к. в большинстве случаев искать собственные значения трудно.

Замечание. Далее всюду будем рассматривать только вещественные пространства.

Теорема 2 (Самарского). Пусть оператор $A = A^* > 0$ ($A^* = A^T$)

Пусть выполнено неравенство

$$B - 0,5\tau A > 0, \quad (\tau > 0) \quad (6)$$

Тогда итерационный метод (2) решения системы (1) сходится в среднеквадратичной норме при любом начальном приближении, то есть

$$\|x^n - x\| = \left(\sum_{i=1}^m (x_i^n - x_i)^2 \right)^{\frac{1}{2}} \rightarrow 0, n \rightarrow \infty$$

Доказательство. Введем $y_n = (AV^n, V^n)$

Рассмотрим y_{n+1} :

$$\begin{aligned} y_{n+1} &= (AV^{n+1}, V^{n+1}) = (ASV^n, SV^n) = \\ &= (A(E - \tau B^{-1}A)V^n, (E - \tau B^{-1}A)V^n) = ((A - \tau AB^{-1}A)V^n, (E - \tau B^{-1}A)V^n) = \\ &= (AV^n, V^n) - \tau [(AB^{-1}AV^n, V^n) + (AV^n, B^{-1}AV^n) - \tau(AB^{-1}AV^n, B^{-1}AV^n)] = \\ &\text{если учесть, что } \{(AB^{-1}AV^n, V^n) = (B^{-1}AV^n, A^*V^n) = (AV^n, B^{-1}AV^n)\}, \text{ получим} \\ &= y_n - \tau [2(AV^n, B^{-1}AV^n) - \tau(AB^{-1}AV^n, B^{-1}AV^n)] = \\ &= y^n + 2\tau \left((B - \frac{\tau}{2}A)B^{-1}AV^n, B^{-1}AV^n \right) \end{aligned}$$

Итак:

$$\frac{y^{n+1} - y^n}{\tau} + 2 \left(\underbrace{(B - 0, 5\tau A)}_{>0 \text{ по условию}} B^{-1}AV^n, B^{-1}AV^n \right) = 0$$

Следовательно, и все скалярное произведение больше либо равно нулю. А стало быть,

$$y^{n+1} \leq y^n,$$

и последовательность $\{y^n\}$ не возрастает и имеет предел.

Воспользуемся свойством положительно определенного оператора: если оператор $C > 0$, то $\exists \delta > 0 : (Cx, x) \geq \delta \|x\|^2, \forall x \in H$

Из этого свойства следует неравенство:

$$((B - 0, 5\tau A)B^{-1}AV^n, B^{-1}AV^n) \geq \delta \|B^{-1}AV^n\|^2,$$

где $\delta > 0$

$$\frac{y^{n+1} - y^n}{\tau} + 2\delta \|B^{-1}AV^n\|^2 \leq 0$$

При $n \rightarrow \infty$ получим

$$\lim_{n \rightarrow \infty} \|B^{-1}AV^n\| = 0$$

Введем $W^n = B^{-1}AV^n$. Отсюда

$$V^n = A^{-1}BW^n$$

Оценим норму погрешности:

$$\|V^n\| \leq \|A^{-1}B\| * \|W^n\|$$

В силу независимости $A^{-1}B$ от n и стремлению к нулю нормы $\|W^n\|$ при $n \rightarrow \infty$ получим, что

$$\lim_{n \rightarrow \infty} \|V^n\| = 0$$

Так как мы нигде не использовали начальное приближение x^0 , то формулировка теоремы остается верной для любого начального приближения x^0 \square

Следствие. Пусть $A = A^* > 0$. (Напомним, что $A = R_1 + D + R_2$, где R_1 и R_2 - нижнетреугольная и верхнетреугольная матрицы, а $D = \text{diag}(a_{11}, a_{22}, \dots, a_{mm})$)

Тогда метод Якоби сходится в среднеквадратичной норме при любом начальном приближении x^0 , если $2D > A$.

Доказательство.

$$D(x^{n+1} - x^n) + Ax^n = f$$

т.е. $B=D$, $\tau = 1$

По условию: $2D > A \Rightarrow D - 0,5A > 0 \Rightarrow$ выполнено условие теоремы, а значит метод сходится. □

Следствие. Пусть

$$\begin{aligned} A &= A^* > 0 \\ a_{ii} &> \sum_{j=1, j \neq i}^m |a_{ij}|, i = \overline{1, m} \end{aligned} \quad (7)$$

Тогда метод Якоби сходится при любом начальном приближении x_0

Доказательство. Возьмем произвольный вектор x , и распишем для него скалярное произведение (Ax, x) , используя известное неравенство $ab \leq \frac{a^2+b^2}{2}$:

$$\begin{aligned} (Ax, x) &= \sum_{i,j=1}^m a_{ij}x_i x_j \leq \sum_{i,j=1}^m |a_{ij}| * |x_i| * |x_j| \leq \\ &= \frac{1}{2} \sum_{i,j=1}^m |a_{ij}|x_i^2 + \frac{1}{2} \sum_{i,j=1}^m |a_{ij}|x_j^2 = \\ &= \frac{1}{2} \sum_{i,j=1}^m |a_{ij}|x_i^2 + \frac{1}{2} \sum_{i,j=1}^m |a_{ji}|x_i^2 = \\ &= \{a_{ij} = a_{ji}\} = \end{aligned}$$

$$\sum_{i,j=1}^m |a_{ij}|x_i^2 = \sum_{i=1}^m x_i^2 (a_{ii} + \sum_{j=1, j \neq i}^m |a_{ij}|)$$

Воспользуемся свойством диагонального преобладания (7)

$$(Ax, x) < 2 \sum_{i=1}^m a_{ii}x_i^2 = (2Dx, x) \Rightarrow 2D > A$$

а значит, по следствию 1 метод Якоби сходится при любом x_0 □

Следствие. Пусть

$$A = A^* > 0$$

Тогда метод Зейделя сходится при любом начальном приближении x^0

Доказательство. По определению метода Зейделя имеем:

$$B = R_1 + D, \tau = 1$$

Для доказательства утверждения, в силу теоремы Самарского, достаточно доказать, что

$$B - 0,5A > 0$$

Поскольку $A = R_1 + D + R_2$, то это соотношение преобразуется к следующему виду:

$$2(R_1 + D) > R_1 + D + R_2$$

$$R_1 + D - R_2 > 0$$

Следовательно,

$$((R_1 + D - R_2)x, x) > 0, \forall x \neq 0, x \in H$$

$$(R_1x, x) + (Dx, x) - (R_2x, x) > 0 \Rightarrow (Dx, x) > 0$$

Последнее следствие верно, так как $A = A^*$, а значит

$$R_1^* = R_2$$

$$(R_1x, x) = (x, R_1^*x) = (x, R_2x) = (R_2x, x)$$

Стало быть, для любого ненулевого вектора из H требуется выполнения неравенства $(Dx, x) > 0$. В силу самосопряженности оператора A это соотношение выполняется, кроме того, все вышеприведенные переходы равносильны, а значит выполнено условие теоремы Самарского. \square

Следствие. Пусть

$$A^T = A > 0,$$

$$\gamma_2 = \max \lambda_k^A,$$

$$0 < \tau < \frac{2}{\gamma_2}.$$

Тогда метод простой итерации (релаксации) сходится.

Доказательство. В нашем случае, $B = E$. Докажем, что

$$E - 0,5\tau > 0,$$

тогда утверждение будет следовать из теоремы Самарского. Запишем цепочку неравенств:

$$\tau < \frac{2}{\gamma_2},$$

$$0,5\tau\gamma_2 < 1,$$

что означает, что для любого λ_k^A – собственного значения матрицы A – выполнено

$$0,5\tau\lambda_k^A < 1,$$

$$1 - 0,5\tau\lambda_k^A > 0,$$

то есть

$$E - 0,5\tau > 0.$$

\square

§7 Оценка скорости сходимости итерационных методов

Рассмотрим СЛАУ

$$Ax = f, \quad (1)$$

где A – матрица размера $m \times m$, $|A| \neq 0$.

Запишем общий вид итерационного метода решения СЛАУ:

$$B \frac{x^{n+1} - x^n}{\tau} + Ax^n = f, \quad (2)$$

где B – обратимая матрица, $\tau > 0$, x^0 – задано, $n = 0, 1, \dots$

Введем обозначение:

$$v^n = x^n - x.$$

Тогда для v можно записать:

$$B \frac{v^{n+1} - v^n}{\tau} + Av^n = 0 \quad (3)$$

Для оценки скорости сходимости итерационных методов мы будем стремиться для некоторого ρ и некоторой нормы доказать т.н. ρ - оценку:

$$\|v^{n+1}\| \leq \rho \|v^n\|, \quad 0 < \rho < 1. \quad (4)$$

Тогда

$$\begin{aligned} \|v^n\| &\leq \rho^n \|v^0\|, \\ n \rightarrow \infty &\Rightarrow \|v^n\| \rightarrow 0. \end{aligned}$$

Пусть H – линейное пространство размерности m . $\forall x, y \in H$ определим:

$$(x, y) = \sum_{i=1}^m x_i y_i,$$

$$\|x\| = \sqrt{(x, x)}.$$

Пусть $D = D^* > 0$. Определим:

$$(x, y)_D = (Dx, y),$$

$$\|x\|_D = \sqrt{(x, x)_D} \text{ – энергетическая норма вектора } x$$

Найдем число итераций $n_0(\epsilon)$, необходимое для того, чтобы $\forall n > n_0(\epsilon)$ выполнялось:

$$\|x^n - x\| < \epsilon \|x^0 - x\|.$$

Из (4) следует, что

$$\|x^n - x\| \leq \rho^n \|x^0 - x\|.$$

Потребуем, чтобы $\rho^n \leq \epsilon$. Тогда

$$\begin{aligned} \frac{1}{\epsilon} &\leq \left(\frac{1}{\rho}\right)^n, \\ n \ln \frac{1}{\rho} &\geq \ln \frac{1}{\epsilon}, \end{aligned}$$

$$n_0(\epsilon) = \left[\frac{\ln \frac{1}{\epsilon}}{\ln \frac{1}{\rho}} \right].$$

Число $\ln \frac{1}{\rho}$ называется скоростью сходимости итерационного метода.

Пусть $D = D^* > 0$. Тогда $\exists \{e_k\}$ – ортонормированный базис (ОНБ) из собственных векторов D . Разложим вектор x по этому базису:

$$x = \sum_{k=1}^m c_k e_k.$$

Для вектора x имеет место равенство Парсеваля:

$$\|x\|^2 = \sum_{k=1}^m c_k^2.$$

Теорема. Пусть $A^* = A > 0$, $B^* = B > 0$,

$$\exists \rho : 0 < \rho < 1,$$

$$\frac{1-\rho}{\tau} B \leq A \leq \frac{1+\rho}{\tau} B. \quad (5)$$

Тогда итерационный метод (2) сходится к решению (1) и выполнена оценка

$$\|v^{n+1}\|_B \leq \rho \|v^n\|_B. \quad (6)$$

Замечание. Из того, что $A \leq \frac{1+\rho}{\tau} B$ и $\rho < 1$, следует, что

$$A < \frac{2}{\tau} B,$$

т.е.

$$B - 0.5\tau A > 0.$$

Таким образом, по теореме Самарского, сходимость имеет место.

Доказательство теоремы. Из положительной определенности матрицы B следует, что

$$\exists B^{\frac{1}{2}} = (B^{\frac{1}{2}})^* > 0,$$

$$\exists B^{-\frac{1}{2}} = (B^{-\frac{1}{2}})^* > 0.$$

Домножим обе части (3) слева на $B^{-\frac{1}{2}}$:

$$B^{\frac{1}{2}} \frac{v^{n+1} - v^n}{\tau} + B^{-\frac{1}{2}} A v^n = 0.$$

Обозначим $B^{\frac{1}{2}} v^n = z^n$.

Тогда

$$\frac{z^{n+1} - z^n}{\tau} + B^{-\frac{1}{2}} A B^{-\frac{1}{2}} z^n = 0.$$

Выразим z^{n+1} :

$$z^{n+1} = z^n - \tau B^{-\frac{1}{2}} A B^{-\frac{1}{2}} z^n.$$

Обозначим

$$S = E - \tau B^{-\frac{1}{2}} A B^{-\frac{1}{2}}. \quad (7)$$

Тогда $z^{n+1} = S z^n$. Назовем матрицу S матрицей перехода для z^n .

Докажем, что из того, что

$$\|z^{n+1}\| \leq \rho \|z^n\|,$$

следует, что

$$\|v^{n+1}\|_B \leq \rho \|v^n\|_B.$$

Действительно,

$$\|z^n\|^2 = (z^n, z^n) = (B^{\frac{1}{2}} v^n, B^{\frac{1}{2}} v^n) = (B v^n, v^n) = \|v^n\|_B^2.$$

Таким образом, осталось доказать, что

$$\|z^{n+1}\| \leq \rho \|z^n\|. \quad (8)$$

Пусть $s_k, k = 1, \dots, m$ – собственные значения матрицы S . Зафиксируем k , пусть x – собственный вектор, соответствующий собственному значению s_k :

$$Sx = s_k x \quad (x \neq 0).$$

Заметим, что

$$B^{\frac{1}{2}} Sx = (B^{\frac{1}{2}} - \tau A B^{-\frac{1}{2}})x = s_k B^{\frac{1}{2}} x.$$

Обозначим $y = B^{-\frac{1}{2}} x$. Тогда предыдущее выражение можно переписать в виде:

$$(B - \tau A)y = s_k B y,$$

$$(1 - s_k)B y = \tau A y,$$

$$A y = \frac{1 - s_k}{\tau} B y.$$

Из условия (5) теоремы следует, что

$$\frac{1 - \rho}{\tau} (B y, y) \leq (A y, y) = \frac{1 - s_k}{\tau} (B y, y) \leq \frac{1 + \rho}{\tau} (B y, y).$$

Поскольку $(B y, y) > 0$, то предыдущее неравенство влечёт

$$\frac{1 - \rho}{\tau} \leq \frac{1 - s_k}{\tau} \leq \frac{1 + \rho}{\tau}.$$

Следовательно, $|s_k| \leq \rho, k = 1, \dots, m$.

Поскольку все матрицы, входящие в правую часть выражения (7), являются самосопряженными, то и матрица S является самосопряженной. Следовательно, существует ортонормированный базис $\{e_k\}_1^m$, состоящий из собственных векторов матрицы S :

$$S e_k = s_k e_k, \quad k = 1, \dots, m.$$

Разложим вектор z^n по базису $\{e_k\}$:

$$z^n = \sum_{k=1}^n c_k^{(n)} e_k.$$

Тогда

$$z^{n+1} = Sz^n = \sum_{k=1}^n c_k^{(n)} S e_k = \sum_{k=1}^n c_k^{(n)} s_k e_k.$$

Пользуясь равенством Парсеваля и полученной оценкой для $|s_k|$, имеем:

$$\|z^{n+1}\|^2 = \sum_{k=1}^n (c_k^{(n)})^2 s_k^2 \leq \rho^2 \sum_{k=1}^n (c_k^{(n)})^2 = \rho^2 \|z^n\|^2.$$

Мы доказали (8) и (6). □

Следствие 1. Пусть $A^* = A > 0$, $B^* = B > 0$, $\exists 0 < \gamma_1 < \gamma_2$:

$$\gamma_1 B \leq A \leq \gamma_2 B. \quad (9)$$

Тогда, если $\tau = \frac{2}{\gamma_1 + \gamma_2} = \tau_0$, то $\|v_{n+1}\|_B \leq \rho \|v_n\|_B$, где $\rho = \frac{1-\xi}{1+\xi}$, $\xi = \frac{\gamma_1}{\gamma_2}$.

Доказательство. Найдем γ_1 и γ_2 :

$$\begin{cases} \tau = \frac{2}{\gamma_1 + \gamma_2}, \\ \rho = \frac{1-\xi}{1+\xi}; \end{cases} \Leftrightarrow \begin{cases} \gamma_1 + \gamma_2 = \frac{2}{\tau}, \\ \gamma_2 - \gamma_1 = \rho(\gamma_1 + \gamma_2); \end{cases} \Leftrightarrow \begin{cases} \gamma_1 + \gamma_2 = \frac{2}{\tau}, \\ \gamma_2 - \gamma_1 = \frac{2\rho}{\tau}; \end{cases} \Leftrightarrow \begin{cases} \gamma_1 = \frac{1-\rho}{\tau}, \\ \gamma_2 = \frac{1+\rho}{\tau}. \end{cases}$$

Таким образом, мы находимся в условиях доказанной теоремы. □

Следствие 2. Пусть $A^* = A > 0$, $B = E$,

γ^k – собственные значения матрицы A , $k = 1, \dots, m$,

$$\gamma_1 = \min_{k=1, \dots, m} \gamma^k, \quad \gamma_2 = \max_{k=1, \dots, m} \gamma^k.$$

Тогда итерационный метод имеет вид

$$B \frac{x^{n+1} - x^n}{\tau} + Ax^n = f$$

и имеет место ρ -оценка

$$\|v_{n+1}\| \leq \rho \|v_n\|, \quad \text{где } \rho = \frac{1-\xi}{1+\xi}, \quad \xi = \frac{\gamma_1}{\gamma_2}$$

Доказательство. Утверждение данного следствия вытекает из утверждения предыдущего следствия. □

§8 Исследование сходимости попеременно треугольного итерационного метода

Рассмотрим СЛАУ:

$$Ax = f, |A| \neq 0 \quad (1)$$

Запишем попеременно треугольный итерационный метод (ПТИМ):

$$(E + \omega R_1)(E + \omega R_2) \frac{x^{n+1} - x^n}{\tau} + Ax^n = f, \quad (2)$$

$$\omega > 0, \quad \tau > 0, \quad n = 0, 1, \dots, \quad x^0 \text{ задано,}$$

$$A = R_1 + R_2,$$

$$R_1 = \begin{pmatrix} 0,5a_{11} & 0 & \cdots & 0 \\ a_{21} & 0,5a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & 0,5a_{mm} \end{pmatrix} \text{ — нижнетреугольная матрица,}$$

$$R_2 = \begin{pmatrix} 0,5a_{11} & a_{12} & \cdots & a_{1m} \\ 0 & 0,5a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0,5a_{mm} \end{pmatrix} \text{ — верхнетреугольная матрица.}$$

Обозначим

$$B = (E + \omega R_1)(E + \omega R_2).$$

Это обозначение согласуется с обозначением для итерационного метода общего вида, рассматриваемого в предыдущих параграфах.

Теорема 1. Пусть $A^* = A > 0$, $\omega > \frac{\tau}{4}$. Тогда ПТИМ (2) сходится при любом начальном приближении x^0 в среднеквадратичной норме.

Доказательство. Распишем B :

$$B = (E + \omega R_2^*)(E + \omega R_2) = E + \omega A + \omega^2 R_2^* R_2 = (E - \omega R_2^*)(E - \omega R_2) + 2\omega A$$

Обозначим $C = E - \omega R_2$. Тогда $C^* = (E - \omega R_2^*)$, $C^*C > 0$, т.к. $(C^*Cx, x) = (Cx, Cx) > 0$ при $x \neq 0$,

$$B - 0,5\tau A > B - 2\omega A = C^*C > 0.$$

Таким образом, по теореме Самарского, имеет место сходимость в среднеквадратичной норме. \square

Теорема 2 (об оценке скорости сходимости ПТИМ). Пусть $A = A^* > 0$, $\exists \delta, \Delta > 0$, т.ч.

$$A \geq \delta E, \quad R_2^* R_2 \leq \frac{\Delta}{4} A. \quad (3)$$

Пусть

$$\omega = \frac{2}{\sqrt{\delta\Delta}}, \quad \tau = \frac{2}{\gamma_1 + \gamma_2}, \quad (4)$$

где

$$\gamma_1 = \frac{\sqrt{\delta}(\sqrt{\delta\Delta})}{2(\sqrt{\delta} + \sqrt{\Delta})}, \quad \gamma_2 = \frac{\sqrt{\delta\Delta}}{4}. \quad (5)$$

Тогда итерационный метод (2) решения (1) сходится и имеет место оценка

$$\|x^{n+1} - x\|_B \leq \rho \|x^n - x\|_B, \quad (6)$$

где

$$\rho = \frac{1 - \sqrt{\eta}}{1 + 3\sqrt{\eta}}, \quad \eta = \frac{\delta}{\Delta}, \quad (7)$$

$$B = (E + \omega R_2^*)(E + \omega R_2).$$

Доказательство. Докажем, что $\delta \leq \Delta$.

Из условия (3) следует, что $\forall x \in H : x \neq 0$ имеем

$$(Ax, x) \geq \delta \|x\|^2,$$

$$\|R_2x\|^2 = (R_2x, R_2x) = (R_2^*R_2x, x) \leq \frac{\Delta}{4}(Ax, x)$$

Поскольку $A = R_1 + R_2$, $R_1 = R_2^*$, то

$$(Ax, x) = (R_2^*x, x) + (R_2x, x) = 2(R_2x, x).$$

Таким образом,

$$\delta \|x\|^2 \leq (Ax, x) = \frac{(Ax, x)^2}{(Ax, x)} = \frac{4(R_2x, x)^2}{(Ax, x)}.$$

Из неравенства Коши-Буняковского:

$$\delta \|x\|^2 \leq \frac{4\|R_2x\|^2 \cdot \|x\|^2}{(Ax, x)} \leq 4\frac{\Delta}{4}\|x\|^2 = \Delta\|x\|^2$$

Сократив на $\|x\|^2$, получим $\delta \leq \Delta$.

В соответствии со следствием 1 параграфа 7, подберем коэффициенты γ_1 и γ_2 так, чтобы выполнялось $\gamma_1 B \leq A \leq \gamma_2 B$.

Из доказательства теоремы 1 данного параграфа $B \geq 2\omega A$. Таким образом, $A \leq \frac{1}{2\omega}B$, $\gamma_2(\omega) = \frac{1}{2\omega}$.

$$B = E + \omega A + \omega^2 R_2 R_2^* \leq \frac{1}{\delta}A + \omega A + \omega^2 \frac{\delta}{4}A = \left(\frac{1}{\delta} + \omega + \omega^2 \frac{\delta}{4}\right)A,$$

$$\gamma_1(\omega) = \left(\frac{1}{\delta} + \omega + \omega^2 \frac{\delta}{4}\right)^{-1}.$$

Таким образом, из следствия 1 параграфа 7 получаем, что для ПТИМ имеет место ρ -оценка (6), где $\rho(\omega) = \frac{1-\xi(\omega)}{1+\xi(\omega)}$, $\xi(\omega) = \frac{\gamma_1(\omega)}{\gamma_2(\omega)}$.

Минимизируем $\rho(\omega)$. Для этого минимизируем $f(\omega) = \frac{\gamma_2(\omega)}{\gamma_1(\omega)}$.

$$f'(\omega) = \frac{1}{2} \left(\frac{\Delta}{4} - \frac{1}{\delta\omega^2} \right),$$

$$f'(\omega) = 0 \text{ при } \omega = \omega_0 = \frac{2}{\sqrt{\delta\Delta}}.$$

Поскольку $f''(\omega_0) > 0$, то при $\omega = \omega_0$ достигается минимум $f(\omega)$, следовательно, на ω_0 достигается минимум и $\rho(\omega)$

Подставив полученное значение ω_0 в выражения для $\gamma_1(\omega)$, $\gamma_2(\omega)$ и $\rho(\omega)$, получим (5) и (7). \square

Напомним, что число итераций, необходимое для достижения точности ϵ , можно вычислить по формуле:

$$n_0(\epsilon) = \left\lceil \frac{\ln \frac{1}{\epsilon}}{\ln \frac{1}{\rho}} \right\rceil.$$

Величина $\ln \frac{1}{\rho}$ называется скоростью сходимости итерационного метода.

Сравним ПТИМ и метод простой итерации (ПИ) по скорости сходимости.

Напомним, что метод ПИ имеет вид:

$$\frac{x^{n+1} - x^n}{\tau} + Ax^n = f,$$

$$\tau > 0, \quad n = 0, 1, \dots, \quad x^0 - \text{задано.}$$

В реальных задачах $\eta = O(m^{-2})$. В соответствии с этим, оценим скорость сходимости ПТИМ:

$$\ln \frac{1}{\rho} = \frac{1 + 3\sqrt{\eta}}{1 - \sqrt{\eta}} = \Theta(\sqrt{\eta}) \Rightarrow n_0(\epsilon) = \Theta(m)$$

Теперь оценим скорость сходимости ПИ:

$$\ln \frac{1}{\rho} = \ln \frac{1 + \xi}{1 - \xi} = \ln \frac{(1 + \xi)^2}{1 - \xi^2} \cong \ln(1 + 2\xi) \cong \ln(1 + 2\eta) \cong \eta \Rightarrow n_0(\epsilon) = \Theta(m^2).$$

Таким образом, ПТИМ сходится на порядок быстрее ПИ.

§9 Методы решения задач на собственные значения

Пусть матрица A имеет размерность $m \times m$. Рассмотрим задачу на собственные значения матрицы A :

$$Ax = \lambda x, \quad x \neq 0. \tag{1}$$

Если число λ и вектор x удовлетворяют (1), то λ называется собственным значением матрицы (оператора) A , а x называется собственным вектором матрицы (оператора) A .

Для нахождения собственных значений A нужно решить уравнение

$$f(\lambda) = |A - \lambda E| = 0.$$

При этом, $f(\lambda)$ – многочлен степени m . При $m \geq 5$ данная задача аналитически не разрешима в общем случае.

Заметим, что в общем случае $\lambda \in \mathbb{C}$, даже если $A \in \mathbb{R}^{m \times m}$

Различают две проблемы собственных значений:

1. *Частичная проблема собственных значений.* Требуется найти некоторое подмножество спектра матрицы A (как правило, минимальное и максимальное по модулю собственные значения).
2. *Полная проблема собственных значений.* Требуется найти весь спектр матрицы A .

Для простоты будем рассматривать только собственные вектора, имеющие норму 1: $\|x\| = 1$.

Степенной метод решения частичной проблемы собственных значений

Этот метод имеет вид

$$x_{n+1} = Ax_n, \quad (2)$$

$$n = 0, 1, \dots, \quad x_0 - \text{задано.}$$

Пусть собственные значения $\lambda_1, \dots, \lambda_m$ матрицы A пронумерованы так, что $|\lambda_1| \leq |\lambda_2| \leq \dots \leq |\lambda_m|$.

Для доказательства сходимости данного метода потребуем выполнение следующих условий:

- А) Существует базис $\{e_k\}_{k=1}^m$ из собственных векторов A : $Ae_k = \lambda_k e_k$, $k = 1, \dots, m$.
- В) $\left| \frac{\lambda_{m-1}}{\lambda_m} \right| < 1$.
- С) При разложении начального приближения по базису $\{e_k\}$: $x_0 = c_1 e_1 + c_2 e_2 + \dots + c_m e_m$ выполнено $c_m \neq 0$.

Запишем x_n :

$$x_n = c_1 \lambda_1^n e_1 + c_2 \lambda_2^n e_2 + \dots + c_m \lambda_m^n e_m,$$

$$\frac{x_n}{\lambda_m^n} = c_1 \left(\frac{\lambda_1}{\lambda_m} \right)^n e_1 + c_2 \left(\frac{\lambda_2}{\lambda_m} \right)^n e_2 + \dots + c_m e_m.$$

Таким образом, при $n \rightarrow \infty$ x_n стремится по направлению к собственному вектору, отвечающему максимальному по модулю собственному значению.

Обозначим через $x_{n+1}^{(i)}$ i -ую координату вектора x_{n+1} . Тогда:

$$x_{n+1}^{(i)} = c_1 \lambda_1^{n+1} e_1^{(i)} + c_2 \lambda_2^{n+1} e_2^{(i)} + \dots + c_m \lambda_m^{n+1} e_m^{(i)}$$

$$x_n^{(i)} = c_1 \lambda_1^n e_1^{(i)} + c_2 \lambda_2^n e_2^{(i)} + \dots + c_m \lambda_m^n e_m^{(i)}$$

Поделим $x_{n+1}^{(i)}$ на $x_n^{(i)}$:

$$\frac{x_{n+1}^{(i)}}{x_n^{(i)}} = \frac{c_m \lambda_m^{n+1} e_m^{(i)} \left(1 + \frac{c_{m-1}}{c_m} \frac{e_{m-1}^{(i)}}{e_m^{(i)}} \left(\frac{\lambda_{m-1}}{\lambda_m} \right)^{n+1} + \dots + \frac{c_1}{c_m} \frac{e_1^{(i)}}{e_m^{(i)}} \left(\frac{\lambda_1}{\lambda_m} \right)^{n+1} \right)}{c_m \lambda_m^n e_m^{(i)} \left(1 + \frac{c_{m-1}}{c_m} \frac{e_{m-1}^{(i)}}{e_m^{(i)}} \left(\frac{\lambda_{m-1}}{\lambda_m} \right)^n + \dots + \frac{c_1}{c_m} \frac{e_1^{(i)}}{e_m^{(i)}} \left(\frac{\lambda_1}{\lambda_m} \right)^n \right)} =$$

$$= \lambda_m + O\left(\left(\frac{\lambda_{m-1}}{\lambda_m}\right)^n\right) = \lambda_m^{(n)}$$

Таким образом, $\lambda_m^{(n)} - \lambda_m = O\left(\left(\frac{\lambda_{m-1}}{\lambda_m}\right)^n\right)$, то есть мы решили задачу нахождения максимального по модулю собственного значения. Сформулируем соответствующее утверждение:

Утверждение. Пусть выполнены следующие предположения:

1. (A) Матрица A имеет базис из собственных векторов $\{e_i\}_{i=1}^{i=m}$
2. (B) $|\frac{\lambda_{m-1}}{\lambda_m}| < 1$
3. (C) $x_0 = c_1 e_1 + c_2 e_2 + \dots + c_m e_m$, где $c_m \neq 0$

Тогда $x_n \rightarrow e_m$ (по направлению) при $n \rightarrow \infty$, где e_m - собственный вектор, отвечающий наибольшему по модулю собственному значению λ_m , а $\lambda_m^{(n)} = \frac{x_{n+1}^{(i)}}{x_n^{(i)}} = \lambda_m + O\left(\left(\frac{\lambda_{m-1}}{\lambda_m}\right)^n\right)$.

Замечание. Условия (A) и (B) несколько ограничивают класс задач, к которым применим этот метод, хотя он все равно остается достаточно широким.

Замечание. Найти $\lambda_m^{(n)}$ можно также по формуле:

$$\lambda_m^{(n)} = \frac{(x_{n+1}, x_n)}{(x_n, x_n)} = \frac{(Ax_n, x_n)}{(x_n, x_n)}$$

Рассмотрим два случая:

1. Пусть $A = A^*$. Тогда $\exists \{e_i\}_{i=1}^{i=m}$ - ортонормированный базис из собственных векторов матрицы A :

$$Ae_k = \lambda_k e_k, \quad k = 1, \dots, m, \quad e_k \neq 0$$

$$(e_i, e_j) = \delta_{ij}$$

$$x_{n+1} = c_1 \lambda_1^{n+1} e_1 + c_2 \lambda_2^{n+1} e_2 + \dots + c_m \lambda_m^{n+1} e_m$$

$$x_n = c_1 \lambda_1^n e_1 + c_2 \lambda_2^n e_2 + \dots + c_m \lambda_m^n e_m$$

Найдем $\lambda_m^{(n)}$:

$$\begin{aligned} \lambda_m^{(n)} &= \frac{(x_{n+1}, x_n)}{(x_n, x_n)} = \frac{c_1^2 \lambda_1^{2n+1} + c_2^2 \lambda_2^{2n+1} + \dots + c_m^2 \lambda_m^{2n+1}}{c_1^2 \lambda_1^{2n} + c_2^2 \lambda_2^{2n} + \dots + c_m^2 \lambda_m^{2n}} = \\ &= \frac{c_m^2 \lambda_m^{2n+1} \left(1 + \left(\frac{c_{m-1}}{c_m}\right)^2 \left(\frac{\lambda_{m-1}}{\lambda_m}\right)^{2n+1} + \dots + \left(\frac{c_1}{c_m}\right)^2 \left(\frac{\lambda_1}{\lambda_m}\right)^{2n+1}\right)}{c_m^2 \lambda_m^{2n} \left(1 + \left(\frac{c_{m-1}}{c_m}\right)^2 \left(\frac{\lambda_{m-1}}{\lambda_m}\right)^{2n} + \dots + \left(\frac{c_1}{c_m}\right)^2 \left(\frac{\lambda_1}{\lambda_m}\right)^{2n}\right)} = \\ &= \lambda_m + O\left(\left(\frac{\lambda_{m-1}}{\lambda_m}\right)^{2n}\right) \end{aligned}$$

Таким образом, при $A = A^*$ получили более быструю сходимость.

2. Пусть $\exists \{e_i\}_{i=1}^{i=m}$ - базис из собственных векторов (ортонормированность не предполагается). Тогда:

$$\begin{aligned} \lambda_m^{(n)} &= \frac{(x_{n+1}, x_n)}{(x_n, x_n)} = \frac{\sum_{i,j=1}^m c_i c_j \lambda_i^{n+1} \lambda_j^n (e_i, e_j)}{\sum_{i,j=1}^m c_i c_j \lambda_i^n \lambda_j^n (e_i, e_j)} = \\ &= \frac{c_m^2 \lambda_m^{2n+1} (e_m, e_m) \left(1 + \frac{c_{m-1}}{c_m} \frac{(e_m, e_{m-1})}{(e_m, e_m)} \left(\frac{\lambda_{m-1}}{\lambda_m} \right)^n + \dots + \left(\frac{c_1}{c_m} \right)^2 \frac{(e_1, e_1)}{(e_m, e_m)} \left(\frac{\lambda_1}{\lambda_m} \right)^{2n+1} \right)}{c_m^2 \lambda_m^{2n} (e_m, e_m) \left(1 + \frac{c_{m-1}}{c_m} \frac{(e_m, e_{m-1})}{(e_m, e_m)} \left(\frac{\lambda_{m-1}}{\lambda_m} \right)^n + \dots + \left(\frac{c_1}{c_m} \right)^2 \frac{(e_1, e_1)}{(e_m, e_m)} \left(\frac{\lambda_1}{\lambda_m} \right)^{2n} \right)} = \\ &= \lambda_m + O \left(\left(\frac{\lambda_{m-1}}{\lambda_m} \right)^n \right) \\ \lambda_m^{(n)} - \lambda_m &= O \left(\left(\frac{\lambda_{m-1}}{\lambda_m} \right)^n \right) \end{aligned}$$

Метод обратных итераций

Пусть матрица A ($m \times m$) такова, что $\exists A^{-1}$. Рассмотрим итерационный степенной метод решения частичной проблемы собственных значений:

$$Ax_{n+1} = x_n, \quad n = 0, 1, \dots, \quad x_0 - \text{задан.}$$

Домножим обе части слева на A^{-1} :

$$x_{n+1} = A^{-1}x_n, \quad n = 0, 1, \dots, \quad x_0 - \text{задан.}$$

Получили степенной метод для обратной матрицы. Пусть верны следующие предположения:

1. (A) Матрица A имеет базис из собственных векторов $\{e_i\}_{i=1}^{i=m}$
2. (B) $|\frac{\lambda_1}{\lambda_2}| < 1$
3. (C) $x_0 = c_1 e_1 + c_2 e_2 + \dots + c_m e_m$, где $c_1 \neq 0$

Тогда:

$$\begin{aligned} x_n &= c_1 \lambda_1^{-n} e_1 + c_2 \lambda_2^{-n} e_2 + \dots + c_m \lambda_m^{-n} e_m \\ \lambda_1^n x_n &= c_1 e_1 + c_2 \left(\frac{\lambda_1}{\lambda_2} \right)^n e_2 + \dots + c_m \left(\frac{\lambda_1}{\lambda_m} \right)^n e_m \end{aligned}$$

Таким образом, $x_n \rightarrow e_1$ (по направлению) при $n \rightarrow \infty$.

Задача. Пусть выполнены условия (A), (B) и (C). Тогда метод обратных итераций позволяет найти собственное значение $\lambda_1^{(n)} = \lambda_1 + O \left(\left(\frac{\lambda_1}{\lambda_2} \right)^n \right)$, где $\lambda_1^{(n)} = \frac{x_n^{(i)}}{x_{n+1}^{(i)}}$.

Доказательство. Выпишем выражения для $x_n^{(i)}$ и $x_{n+1}^{(i)}$:

$$\begin{aligned} x_n &= c_1 \lambda_1^{-n} e_1 + c_2 \lambda_2^{-n} e_2 + \dots + c_m \lambda_m^{-n} e_m \\ x_{n+1} &= c_1 \lambda_1^{-n-1} e_1 + c_2 \lambda_2^{-n-1} e_2 + \dots + c_m \lambda_m^{-n-1} e_m \end{aligned}$$

Теперь поделим $x_n^{(i)}$ на $x_{n+1}^{(i)}$:

$$\begin{aligned} \frac{x_n^{(i)}}{x_{n+1}^{(i)}} &= \frac{c_1 \lambda_1^{-n} e_1^{(i)} \left(1 + \frac{c_2}{c_1} \frac{e_2^{(i)}}{e_1^{(i)}} \left(\frac{\lambda_2}{\lambda_1} \right)^{-n} + \dots + \frac{c_m}{c_1} \frac{e_m^{(i)}}{e_1^{(i)}} \left(\frac{\lambda_m}{\lambda_1} \right)^{-n} \right)}{c_1 \lambda_1^{-n-1} e_1^{(i)} \left(1 + \frac{c_2}{c_1} \frac{e_2^{(i)}}{e_1^{(i)}} \left(\frac{\lambda_2}{\lambda_1} \right)^{-n-1} + \dots + \frac{c_m}{c_1} \frac{e_m^{(i)}}{e_1^{(i)}} \left(\frac{\lambda_m}{\lambda_1} \right)^{-n-1} \right)} = \\ &= \lambda_1 + O \left(\left(\frac{\lambda_1}{\lambda_2} \right)^n \right) = \lambda_1^{(n)} \end{aligned}$$

□

Пусть теперь $A = A^*$. Тогда $\exists \{e_i\}_{i=1}^{i=m}$ - ортонормированный базис из собственных векторов матрицы A . Найдем $\lambda_1^{(n)}$:

$$\begin{aligned} \lambda_m^{(n)} &= \frac{(x_n, x_n)}{(x_{n+1}, x_n)} = \frac{c_1^2 \lambda_1^{-2n} + c_2^2 \lambda_2^{-2n} + \dots + c_m^2 \lambda_m^{-2n}}{c_1^2 \lambda_1^{-2n+1} + c_2^2 \lambda_2^{-2n-1} + \dots + c_m^2 \lambda_m^{-2n-1}} = \\ &= \frac{c_1^2 \lambda_1^{-2n} \left(1 + \left(\frac{c_2}{c_1} \right)^2 \left(\frac{\lambda_2}{\lambda_1} \right)^{-2n} + \dots + \left(\frac{c_m}{c_1} \right)^2 \left(\frac{\lambda_m}{\lambda_1} \right)^{-2n} \right)}{c_1^2 \lambda_1^{-2n-1} \left(1 + \left(\frac{c_2}{c_1} \right)^2 \left(\frac{\lambda_2}{\lambda_1} \right)^{-2n-1} + \dots + \left(\frac{c_m}{c_1} \right)^2 \left(\frac{\lambda_m}{\lambda_1} \right)^{-2n-1} \right)} = \\ &= \lambda_1 + O \left(\left(\frac{\lambda_1}{\lambda_2} \right)^{2n} \right) \end{aligned}$$

Таким образом, при $A = A^*$ снова имеем более быструю сходимость.

Задача. Пусть $\exists \{e_i\}_{i=1}^{i=m}$ - базис из собственных векторов матрицы A . Тогда $\lambda_1^{(n)} = \lambda_1 + O \left(\left(\frac{\lambda_1}{\lambda_2} \right)^n \right)$.

Доказательство.

$$\begin{aligned} \lambda_1^{(n)} &= \frac{(x_n, x_n)}{(x_{n+1}, x_n)} = \frac{\sum_{i,j=1}^m c_i c_j \lambda_i^{-n} \lambda_j^{-n} (e_i, e_j)}{\sum_{i,j=1}^m c_i c_j \lambda_i^{-n-1} \lambda_j^{-n} (e_i, e_j)} = \\ &= \frac{c_1^2 \lambda_1^{-2n} (e_1, e_1) \left(1 + \frac{c_2}{c_1} \frac{(e_1, e_2)}{(e_1, e_1)} \left(\frac{\lambda_2}{\lambda_1} \right)^{-n} + \dots + \left(\frac{c_m}{c_1} \right)^2 \frac{(e_m, e_m)}{(e_1, e_1)} \left(\frac{\lambda_m}{\lambda_1} \right)^{-2n} \right)}{c_1^2 \lambda_1^{-2n-1} (e_1, e_1) \left(1 + \frac{c_2}{c_1} \frac{(e_1, e_2)}{(e_1, e_1)} \left(\frac{\lambda_2}{\lambda_1} \right)^{-n} + \dots + \left(\frac{c_m}{c_1} \right)^2 \frac{(e_m, e_m)}{(e_1, e_1)} \left(\frac{\lambda_m}{\lambda_1} \right)^{-2n-1} \right)} = \\ &= \lambda_1 + O \left(\left(\frac{\lambda_1}{\lambda_2} \right)^n \right) \\ \lambda_1^{(n)} - \lambda_1 &= O \left(\left(\frac{\lambda_1}{\lambda_2} \right)^n \right) \end{aligned}$$

□

Сформулируем еще одно утверждение:

Утверждение. Если есть хотя бы одно комплексное собственное значение $\lambda_k = \lambda_0 + i\lambda_1$, $\lambda_1 \neq 0$, то и отвечающий ему собственный вектор должен быть комплексным, и начальное приближение для него должно быть комплексным.

Доказательство. Пусть $x_k = \mu_0 + i\mu_1$, $\mu_1 \neq 0$ - собственный вектор матрицы A , отвечающий собственному значению $\lambda_k = \lambda_0 + i\lambda_1$. Тогда:

$$Ax_k = A(\mu_0 + i\mu_1) = (\lambda_0 + i\lambda_1)(\mu_0 + i\mu_1) = \lambda_0\mu_0 - \lambda_1\mu_1 + i(\lambda_0\mu_1 + \lambda_1\mu_0)$$

В силу линейности:

$$A\mu_0 = \lambda_0\mu_0 - \lambda_1\mu_1$$

$$A\mu_1 = \lambda_0\mu_1 + \lambda_1\mu_0$$

Предположим, что $\mu_1 = 0$. Тогда $\lambda_1\mu_0 = 0$, $\mu_0 = 0$, откуда следует, что $x_k = 0$, а это противоречит тому, что x - ненулевой вектор. \square

Метод обратных итераций со сдвигом

Иногда бывает нужно найти собственное значение из внутренней части спектра. Рассмотрим метод обратных итераций со сдвигом:

$$(A - \alpha E)x_{n+1} = x_n, \alpha \in \mathbb{R}, n \in \mathbb{N}$$

$$n = 0, 1, \dots, \quad x_0 - \text{задан.}$$

Пусть существует $(A - \alpha E)^{-1} = B$, тогда получим степенной метод для матрицы B :

$$x_{n+1} = Bx_n, \quad n = 0, 1, \dots, \quad x_0 - \text{задан.}$$

Собственные значения матрицы B :

$$\lambda_k^B = \frac{1}{\lambda_k^A - \alpha}$$

Тогда $x_n \rightarrow e_l$ (по направлению), где l таково, что:

$$\lambda_l^B = \max_{k=1, \dots, m} \frac{1}{\lambda_k^A - \alpha} = \frac{1}{\lambda_l^A - \alpha}$$

Замечание. Если известно приближенное значение какого-то собственного значения, а мы хотим его уточнить, то можно использовать этот метод. Найти весь спектр этим методом практически невозможно.

§10 Приведение матрицы к верхней почти треугольной форме (ВПТФ)

Легче всего найти собственные значения у диагональной или треугольной матрицы. Наша задача - привести матрицу A ($m \times m$) к треугольной. Однако, приведение матрицы A к треугольной форме методом Гаусса не сохраняет спектр матрицы. Спектр матрицы сохраняется при преобразовании подобия:

$$C = Q^{-1}AQ$$

Если матрица Q - ортогональна (унитарна), то сохраняется симметрия.

Определение. Матрица находится в верхней почти треугольной форме (ВПТФ), если она имеет вид (ненулевые элементы обозначены через x):

$$A = \begin{pmatrix} x & x & x & \cdots & x & x & x \\ x & x & x & \cdots & x & x & x \\ 0 & x & x & \cdots & x & x & x \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & x & x & x \\ 0 & 0 & 0 & \cdots & 0 & x & x \end{pmatrix}$$

Рассмотрим вектор-столбец ν :

$$\nu = (\nu_1, \nu_2, \dots, \nu_m)^T$$

$$\nu^T = (\nu_1, \nu_2, \dots, \nu_m)$$

Определение. Элементарным отражением, соответствующим вектору ν , называется преобразование, задаваемое матрицей:

$$H = E - 2 \frac{\nu\nu^T}{\|\nu\|^2}$$

Замечание.

$$\nu^T \nu = \nu_1^2 + \nu_2^2 + \cdots + \nu_m^2 = \|\nu\|^2$$

$$\nu\nu^T = \begin{pmatrix} \nu_1^2 & \nu_1\nu_2 & \cdots & \nu_1\nu_m \\ \nu_2\nu_1 & \nu_2^2 & \cdots & \nu_2\nu_m \\ \vdots & \vdots & \ddots & \vdots \\ \nu_m\nu_1 & \nu_m\nu_2 & \cdots & \nu_m^2 \end{pmatrix}$$

Матрица $\nu\nu^T$ - симметричная.

Свойства оператора H :

1. $H^T = H$
2. $H^{-1} = H^T$

Докажем второе свойство, то есть что H является ортогональной матрицей.

Доказательство.

$$H^T H = H^2 = \left(E - 2 \frac{VV^T}{\|V\|^2} \right) \left(E - 2 \frac{VV^T}{\|V\|^2} \right) =$$

$$E - 4 \frac{VV^T}{\|V\|^2} + 4 \frac{V(V^T V)V^T}{\|V\|^4}$$

Используя свойство $V^T V = \|V\|^2$, сократим в последнем соотношении дроби и получим, что

$$H^T H = E$$

Значит, матрица H является ортогональной. □

Сформулируем и докажем свойство 3.

Теорема. Для любого вектора x :

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_m \end{pmatrix}$$

можно выбрать такой вектор $V = (v_1, v_2, \dots, v_m)$, что

$$Hx = \begin{pmatrix} -\sigma \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

где $\sigma = \|x\|$, то есть преобразование H подавляет все координаты вектора кроме первой.

Доказательство. Выберем V в виде:

$$V = x + \sigma z, \quad \sigma \in \mathbb{R}, \quad z = (1, 0, 0, \dots, 0)^T$$

$$\begin{aligned} Hx &= x - 2 \frac{(x + \sigma z)(x + \sigma z)^T}{(x + \sigma z)^T(x + \sigma z)} x = \\ &= x - (x + \sigma z) \frac{2(x + \sigma z)^T x}{(x + \sigma z)^T(x + \sigma z)} \end{aligned}$$

Для дальнейшего преобразования воспользуемся свойствами 1 и 2 :

$$(x + \sigma z)^T x = \|x\|^2 + \sigma x_1$$

$$\begin{aligned} (x + \sigma z)^T(x + \sigma z) &= \|x\|^2 + \sigma x_1 + \sigma x_1 + \sigma^2 = \\ &= \|x\|^2 + 2\sigma x_1 + \sigma^2 \end{aligned}$$

Положим $\sigma = \|x\|$. Тогда получим:

$$\begin{aligned} x - (x + \sigma z) \frac{2(x + \sigma z)^T x}{(x + \sigma z)^T(x + \sigma z)} &= \\ x - (x + \sigma z) \frac{(\|x\|^2 + 2\sigma x_1 + \sigma^2) + \|x\|^2 - \sigma^2}{\|x\|^2 + 2\sigma x_1 + \sigma^2} &= x - x - \sigma z = \begin{pmatrix} -\sigma \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \end{aligned}$$

□

Полученные 3 свойства мы будем применять при приведении матрицы к верхней почти треугольной форме.

Пусть дана произвольная матрица A порядка $m \times m$:

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mm} \end{pmatrix}$$

Представим ее в виде блочной матрицы следующего вида:

$$A = \begin{pmatrix} a_{11} & y_{m-1} \\ x_{m-1} & A_{m-1} \end{pmatrix}$$

где

$$y_{m-1} = (a_{12}, a_{13}, \dots, a_{1m})$$

$$x_{m-1} = (a_{21}, a_{31}, \dots, a_{m1})$$

а матрица A_{m-1} получается из матрицы A удалением первого столбца и первой строки.

Воспользуемся свойством 3:

$$H_{m-1}x_{m-1} = \begin{pmatrix} -\|x_{m-1}\| \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Рассмотрим матрицу U_1 порядка $m \times m$ вида:

$$U_1 = \begin{pmatrix} 1 & 0_{12} \\ 0_{21} & H_{m-1} \end{pmatrix}$$

Очевидно, $U_1 = U_1^T$, значит:

$$U_1^2 = \begin{pmatrix} 1 & 0_{12} \\ 0_{21} & H_{m-1} \end{pmatrix} \begin{pmatrix} 1 & 0_{12} \\ 0_{21} & H_{m-1} \end{pmatrix} = \begin{pmatrix} 1 & 0_{12} \\ 0_{21} & H_{m-1}^2 \end{pmatrix} = E$$

Последнее равенство верно в силу того, что $H_{m-1}^2 = E$. Таким образом, мы получили что матрица U_1 - ортогональная.

Обозначим $C_1 = U_1^{-1}AU_1$

$$C_1 = \begin{pmatrix} a_{11} & c_{12}^{(1)} & \dots \\ -\sigma_1 z_{m-1} & c_{22}^{(2)} & \dots \\ 0 & c_{32}^{(1)} & c_{ij}^{(1)} \\ \vdots & \vdots & \dots \\ 0 & c_{m2}^{(1)} & \dots \end{pmatrix}$$

Таким образом, структуру матрицы можно представить так:

$$\begin{pmatrix} \times & \times & \dots & \times \\ \times & \times & \dots & \times \\ 0 & \times & \dots & \times \\ 0 & \times & \dots & \times \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \times & \dots & \times \end{pmatrix}$$

Возьмем вектор $x_{m-2} = \begin{pmatrix} c_{32}^{(1)} \\ \vdots \\ c_{m2}^{(1)} \end{pmatrix}$

По свойству 3, можно построить такой оператор H_{m-2} , что:

$$H_{m-2}x_{m-2} = \begin{pmatrix} -\|x_{m-2}\| \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Рассмотрим матрицу U_2 , построенную аналогично матрице U_1

$$U_2 = \begin{pmatrix} E_2 & 0_{12} \\ 0_{21} & H_{m-2} \end{pmatrix}$$

где $E_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

Свойства матрицы U_2 :

1. $U_2 = U_2^T$
2. $U_2 = U_2^{-1}$ - ортогональная матрица

Аналогично, обозначим C_2 :

$$C_2 = U_2^{-1}C_1U_2 = U_2^{-1}U_1^{-1}AU_1U_2$$

Посмотрим на структуру матрицы C_2 :

$$C_2 = \begin{pmatrix} \times & \times & \times & \dots & \times \\ \times & \times & \times & \dots & \times \\ 0 & \times & \times & \dots & \times \\ 0 & 0 & \times & \dots & \times \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \times & \dots & \times \end{pmatrix}$$

Очевидно, что сделав таким образом $m-2$ шага, мы придем к верхней почти треугольной форме. В итоге мы получим:

$$C = U_{m-2}^{-1}U_{m-3}^{-1} \dots U_2^{-1}U_1^{-1}AU_1U_2 \dots U_{m-3}U_{m-2} =$$

$$= \begin{pmatrix} \times & \times & \times & \dots & \times \\ \times & \times & \times & \dots & \times \\ 0 & \times & \times & \dots & \times \\ 0 & 0 & \times & \dots & \times \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \times & \times \end{pmatrix} \text{-ВПТФ}$$

Обозначим $U = U_1U_2 \dots U_{m-2}$

$$U^T = U_{m-2}^T U_{m-3}^T \dots U_2^T U_1^T = U_{m-2}^{-1} U_{m-3}^{-1} \dots U_2^{-1} U_1^{-1} = U^{-1}$$

Следовательно, матрица U - ортогональна.

$$C = U^{-1}AU \Rightarrow C \sim A$$

Причем подобие выполняется с помощью ортогональной матрицы U .
Элементы матрицы C имеют вид:

$$c_{ij} = 0, \quad i \geq j + 2, \quad j = 1, 2, \dots, m - 2$$

Замечание. Собственные значения матрицы A совпадают с собственными значениями матрицы C , т.е.:

$$\lambda_k^A = \lambda_k^C, \quad k = \overline{1, m}$$

Доказательство.

$$Ax = \lambda x, \quad x \neq 0$$

$$U^{-1}Ax = \lambda U^{-1}x, \text{ обозначим } U^{-1}x = y \neq 0 \Rightarrow x = Uy$$

$$U_{-1}AUy = \lambda y, \quad y \neq 0$$

$$Cy = \lambda y$$

□

Замечание. Если $A = A^T$, то $C = C^T$

Доказательство.

$$C^T = (U^{-1}AU)^T = U^T A^T (U^{-1})^T = U^{-1}AU = C$$

□

§11 Понятие о QR-алгоритме. Решение полной проблемы собственных значений.

Изученные в предыдущем параграфе свойства позволят нам представить матрицу A в виде:

$$A = QR \tag{1}$$

где $Q^{-1} = Q^T$ - ортогональная матрица, а R имеет верхнетреугольную форму.

Возьмем вектор $x = (a_{11}, a_{21}, \dots, a_{m1})$. Для него существует такая ортогональная матрица $H_1 = E - 2 \frac{VV^T}{\|V\|^2}$, которая “подавляет” все координаты вектора x , кроме первой. Матрица $H_1 A$ имеет вид:

$$H_1 A = \begin{pmatrix} \times & \times & \dots & \times \\ 0 & \times & \dots & \times \\ 0 & \times & \dots & \times \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \times & \dots & \times \end{pmatrix}$$

Построим матрицу $H_2 = \begin{pmatrix} 1 & 0 \\ 0 & H \end{pmatrix}$ порядка $m \times m$, такую, что

$$H_2 H_1 A = \begin{pmatrix} \times & \times & \times & \dots & \times \\ 0 & \times & \times & \dots & \times \\ 0 & 0 & \times & \dots & \times \\ \vdots & \vdots & \vdots & \ddots & \times \\ 0 & 0 & \times & \dots & \times \end{pmatrix}$$

Очевидно, что за $(m-1)$ шаг мы обнулим все элементы под главной диагональю:

$$H_{m-1} H_{m-2} \dots H_2 H_1 A = R = \begin{pmatrix} \times & \times & \dots & \times \\ 0 & \times & \dots & \times \\ \vdots & \vdots & \ddots & \times \\ 0 & 0 & \dots & \times \end{pmatrix} - \text{ВТФ}$$

Построим матрицу Q следующим образом:

$$Q = H_1 H_2 \dots H_{m-1}$$

Найдем Q^T :

$$Q^T = H_{m-1}^T H_{m-2}^T \dots H_2^T H_1^T = H_{m-1}^{-1} H_{m-2}^{-1} \dots H_1^{-1} = Q^{-1}$$

Следовательно, матрица Q ортогональна.

Таким образом, мы получили разложение матрицы A .

Замечание. При факторизации в виде QR :

1. Для произвольной матрицы A требуется $O(m^3)$ действий.
2. Для матрицы A вида ВПТФ требуется $O(m^2)$ действий.
3. Для трехдиагональной матрицы A требуется $O(m)$ действий.

QR-алгоритм

Возьмем матрицу A_0 . Представим ее в виде $A_0 = Q_0 R_0$, где $Q_0^T = Q_0^{-1}$, R - матрица ВТФ.

Положим

$$A_1 = R_0 Q_0 \tag{2}$$

$$R_0 = Q_0^{-1} A_0$$

$$A_1 = Q_0^{-1} A_0 Q_0$$

Таким образом, матрицы A_1 и A_0 подобны с ортогональной матрицей.

Аналогично, сделаем следующие шаги:

$$A_1 = Q_1 R_1, \quad Q_1^T = Q_1^{-1}, \quad R_1 - \text{ВТФ}$$

...

$$A_k = Q_k R_k, \quad k = 0, 1, \dots \tag{3}$$

$$A_{k+1} = R_k Q_k \tag{4}$$

Устремим $k \rightarrow \infty$, тогда:

$$A_k \rightarrow \begin{pmatrix} \times & \times & \dots & \times \\ 0 & \times & \dots & \times \\ \vdots & \vdots & \ddots & \times \\ 0 & 0 & \dots & \times \end{pmatrix}$$

Где на главной диагонали будут стоять собственные значения матриц A_0, A_1, \dots (спектры этих матриц совпадают).

Заметим, что под главной диагональю могут и не получаться нули в математическом смысле. Достаточно, чтобы значения под главной диагональю были по модулю меньше некоторого числа (т.н. машинный ноль), определяющего точность вычислений.

Если у A_k комплексные собственные значения, то в матрице на главной диагонали будут появляться квадраты 2×2 , она будет иметь вид:

$$A_k \rightarrow \begin{pmatrix} \times & & & X \\ & \times & & \\ & & \lambda_0 & \lambda_1 \\ & & -\lambda_1 & \lambda_0 \\ 0 & & & \ddots \end{pmatrix}$$

Перечислим основные плюсы и минусы QR-алгоритма:

1. (+) Для любой матрицы можно найти весь спектр.
2. (-) Во время вычислений нужно держать всю матрицу в памяти.
3. (-) Если собственные значения комплексны, то появляются клетки 2го порядка, которые при последующих итерациях не будут сходиться к 0.

Свойства QR-алгоритма

Утверждение. Пусть матрица B – ВТФ, а матрица A – ВПТФ. Тогда $Q = BA$ – ВПТФ.

Доказательство. По формуле умножения матриц:

$$c_{ij} = \sum_{\alpha=1}^n b_{i\alpha} c_{\alpha j}.$$

Так как B – ВТФ, то все $b_{i\alpha}$ при $i > \alpha$ равны нулю, так как A – ВПТФ, то все $a_{\alpha j}$ при $\alpha > j + 1$ равны нулю. Модифицируем формулу согласно этим утверждениям:

$$c_{ij} = \sum_{\alpha=i}^{j+1} b_{i\alpha} c_{\alpha j}.$$

То есть если $i > j + 1$, то $c_{ij} = 0$. А это и значит, что C – верхняя почти треугольная матрица. \square

Утверждение. Пусть матрица B – ВПТФ, а матрица A – ВТФ. Тогда $Q = BA$ – ВПТФ.

Доказательство. Доказательство полностью аналогично предыдущему утверждению. \square

Используя данные утверждения, можно значительно ускорить QR-разложение матрицы. QR-алгоритм преобразует матрицу $A \rightarrow A_0$ – ВПТФ:

$$A_0 = Q_0 R_0$$

$$Q_0 = A_0 R_0^{-1} \text{ – ВПТФ по доказанному утверждению}$$

$$A_1 = R_0 Q_0 \text{ – ВПТФ по доказанному утверждению}$$

То есть форма матрицы $A_n (n \in \mathbf{N})$ не ухудшается, следовательно, очередная матрица может быть выполнена не более чем за n^2 действий.

Если же матрица A_0 - симметричная, то один шаг потребует всего n действий.

Глава II

Интерполирование и приближение функций

§1 Постановка задачи интерполирования

Пусть $f(x)$ – дискретная функция аргумента $x, x \in [a, b], a, b \in \mathbf{R}$. Функция $f(x)$ определена в точках $x_0, x_1, \dots, x_n, n \in \mathbf{N}; a \leq x_0 < x_1 < \dots < x_n \leq b = \{X_i\}_0^n$ – узлах функции. Во всех узлах заданы значения $f(x_i) = y_i, \forall i = \overline{0, n}$. Требуется найти значение функции $f(x)$ в произвольной точке.

Замечание. В указанной формулировке решений задачи бесконечно много. Для уточнения дополнительно указывают класс функций, которые будут использоваться для построения значений $f(x)$ в произвольной точке.

Интерполирование алгебраическими полиномами

Определение. Назовем интерполяционным полиномом Лагранжа функции $f(x)$ по узлам $\{X_i\}_0^n$ полином степени n :

$$P_n(x) = a_0 + a_1x + \dots + a_nx^n, \quad (1)$$

при этом значения коэффициентов $a_0 \dots a_n$ выбираются таким образом, чтобы при любых значениях $i = \overline{1, n}$ было выполнено:

$$P_n(x_i) = f(x_i) \quad (2)$$

Утверждение. Покажем, что интерполяционный полином $P_n(x)$ для функции $f(x)$ по узлам $\{X_i\}_0^n$ существует и единственен.

Доказательство. Распишем $n + 1$ уравнение из условия (2). Получим систему линейных уравнений:

$$a_0 + a_1x_0 + \dots + a_nx_0^n = f_0, a_0 + a_1x_1 + \dots + a_nx_1^n = f_1, \dots, a_0 + a_1x_n + \dots + a_nx_n^n = f_n,$$

Теперь посмотрим на определитель этой системы:

$$\Delta = \begin{bmatrix} 1 & x_0 & x_0^2 & \dots & x_0^n \\ 1 & x_1 & x_1^2 & \dots & x_1^n \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_n & x_n^2 & \dots & x_n^n \end{bmatrix}$$

Из курса линейной алгебры известно, что данный определитель (определитель Вандермонда) равен произведению разности всех пар $(x_i, x_j), i \neq j$. По условию никакие два различных узла не могут дать нам нулевую разность, следовательно, определитель системы не равен нулю. А это и означает, что решение (т.е. $P_n(x)$) существует и единственно. \square

Замечание. Поскольку мы доказали существование и единственность интерполирующего полинома, то при его поиске, в какой бы форме мы его не получили, он будет тождественно равен всем своим представлениям в иных формах, полученных с помощью других методов.

§2 Интерполяционная формула Лагранжа

Будем искать интерполяционный полином в виде

$$L_n(x) = \sum_{k=0}^n c_k(x)f(x_k), \text{ где:} \quad (1)$$

$c_k(x)$ – полином n -й степени, $f(x_k)$ – известные значения функции в узлах.

Замечание. По определению $L_n(x_i) = f(x_i), \forall i = \overline{1, n}$.

Будем строить полином следующим образом:

Пусть $\omega(x) = (x - x_0)(x - x_1) \cdot \dots \cdot (x - x_n) = \prod_{i=0}^n (x - x_i)$.

Тогда:

$$\omega'(k) = ([\dots](x - x_k)) = [\dots] + [\dots]'(x - x_k) = \prod_{\substack{i=0 \\ i \neq k}}^n (x - x_i)$$

. ($\omega(k)$ – значение функции в точке x_k).

Полиномы $c_k(x)$ возьмем равными $\frac{\omega(x)}{(x-x_k)\omega'(x)}$.

Определим погрешность метода как разность между значением полинома Лагранжа и значением функции:

$$\psi_n(x) = f(x) - L_n(x) \quad (2)$$

Замечание. Для оценки погрешности метода мы требуем $f(x) \in C^{n+1}[a, b]$.

Утверждение.

$$\forall x^* \in [a, b] : r_n(x^*) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \cdot \omega_{n+1}(x^*), \xi \in (a, b)$$

Доказательство. Пусть $g(s) = f(s) + L_n(s) - k\omega(s)$, где k – константа.

Очевидно, что $g(s)$ имеет $n+2$ нуля: $n+1$ за счет обращения в ноль в узлах и последний ноль за счет совпадения $f(s) + L_n(s) = k\omega$. В этом случае k и есть искомая оценка. По теореме Ролля $g^{(n+1)}(\xi) = 0$. Найдем эту производную:

$$g^{(n+1)}(s) = (f(s) + L_n(s) - k\omega(s))^{(n+1)} = f^{(n+1)}(\xi) + 0 - k \cdot n!$$

Откуда и получаем:

$$f(x) + L_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \omega(x)$$

\square

Замечание. Полином Лагранжа, вообще говоря, не сходится к $f(x)$.

§3 Интерполяционная формула Ньютона

Определение. Назовем разделенной разностью первого порядка, построенной по узлам x_i и x_j , следующее соотношение:

$$f(x_i, x_j) = \frac{f(x_j) - f(x_i)}{x_j - x_i} \quad (1)$$

Разделенной разностью второго порядка по узлам x_{i-1}, x_i, x_{i+1} называется соотношение:

$$f(x_{i-1}, x_i, x_{i+1}) = \frac{f(x_{i-1}x_i) - f(x_i x_{i+1})}{x_{i-1} - x_{i+1}}$$

Аналогично определяем распределенную разность больших порядков.

Утверждение. Распределенную разность k -го порядка можно представить в виде:

$$f(x_0, x_1, \dots, x_k) = \sum_{i=0}^k \frac{f(x_i)}{\omega'_{0,k}(x_i)}$$

Причем запись $\omega_{a,b}(x)$ означает:

$$\omega_{a,b}(x) = (x - x_a)(x - x_{a+1}) \cdot \dots \cdot (x - x_b), a < b$$

Доказательство. Не ограничивая общности, будем рассматривать узлы с индексами $0..k$, $k \in \mathbb{N}$. Докажем утверждение по индукции.

База:

$$\begin{aligned} k = 1 : f(x_0, x_1) &= \frac{f(x_0)}{x_0 - x_1} + \frac{f(x_1)}{x_1 - x_0} = \\ &= \frac{f(x_0)}{\omega'_{0,1}(x_0)} + \frac{f(x_1)}{\omega'_{0,1}(x_1)} \end{aligned}$$

Переход:

$$\begin{aligned} k = l : \\ f(x_0, \dots, x_l) &= \sum_{i=0}^l \frac{f(x_i)}{\omega'_{0,l}(x_i)} \end{aligned}$$

Покажем что $f(x_0, \dots, x_{l+1}) = \sum_{i=0}^{l+1} \frac{f(x_i)}{\omega'_{0,l+1}(x_i)}$:

$$\begin{aligned} f(x_0, \dots, x_{l+1}) &= \frac{f(x_1, \dots, x_{l+1}) - f(x_0, \dots, x_l)}{x_{l+1} - x_0} = \\ &= \frac{1}{x_{l+1} - x_0} \left(\sum_{i=1}^{l+1} \frac{f(x_i)}{\omega'_{1,l+1}(x_i)} - \sum_{i=0}^l \frac{f(x_i)}{\omega'_{0,l}(x_i)} \right) = \\ &= \frac{1}{x_{l+1} - x_0} \left(\frac{f(x_{l+1})}{\omega'_{1,l+1}(x_{l+1})} - \frac{f(x_0)}{\omega'_{0,l}(x_0)} + \sum_{i=1}^l \frac{1}{\omega'_{1,l+1}(x_i)} - \frac{1}{\omega'_{0,l}(x_i)} \right) \end{aligned}$$

Рассмотрим слагаемые отдельно:

$$(x_{l+1} - x_0)\omega'_{1,l+1}(x_{l+1}) = \omega'_{0,l+1}(x_{l+1})$$

$$(x_{l+1} - x_0)\omega'_{0,l}(x_0) = \omega'_{0,l+1}(x_0)$$

$$\frac{1}{\omega'_{1,l+1}(x_i)} - \frac{1}{\omega'_{0,l}(x_i)} = \frac{(x_i - x_0)}{\omega'_{0,l+1}(x_i)} - \frac{(x_i - x_{l+1})}{\omega'_{0,l+1}(x_i)} = (x_{l+1} - x_0)$$

Подставив преобразованные слагаемые, получим:

$$\frac{f(x_0)}{\omega'_{0,l+1}(x_0)} + \frac{f(x_{l+1})}{\omega'_{0,l+1}(x_{l+1})} + \sum_{i=1}^l \frac{f(x_i)}{\omega'_{0,l+1}(x_i)}$$

Что и требовалось доказать. \square

Определение. Назовем интерполяционным полиномом Ньютона функции $f(x)$ по узлам $\{X_i\}_0^n$ полином степени n :

$$N_n(x) = f(x_0) + (x - x_0)f(x_0, x_1) + \dots + \prod_{i=0}^{n-1} f(x_0, x_1, \dots, x_n) \quad (2)$$

Покажем, что $N_n(x)$ интерполяционный полином:

$$N_n(x_i) = f(x_0) + (x_i - x_0)f(x_0, x_1) + \dots + \prod_{i=0}^{i-1} f(x_0, x_1, \dots, x_i) + 0$$

Эта сумма представляет собой разделенную разность порядка i , равную как раз $f(x_i)$.

Замечание. Полученный полином – тот же полином Лагранжа, только записанный в другой форме.

Соответственно, его погрешность та же, что и у полинома Лагранжа.

Отличие полинома Ньютона от Лагранжа в том, что для увеличения точности $N_n(x)$ надо только добавить информацию о новых узлах и не надо пересчитывать значения для старых.

§4 Интерполирование с кратными узлами. Интерполяционная формула Эрмита

Пусть имеется m узлов: x_0, x_1, \dots, x_m , при этом $a_k \in \mathbf{N}$, $k = \overline{1, m}$ – кратность каждого узла ($a_0 + \dots + a_m = n + 1$, где n – степень интерполирующего полинома).

Определение. Назовем интерполяционным полиномом Эрмита полином:

$$H_n(x) = \sum_{k=0}^m \sum_{i=0}^{a_k-1} c_{k,i}(x) f^{(i)}(x_k), \quad (1)$$

где $c_{k,i}(x)$ – полином n -й степени, коэффициенты которого находятся из условия:

$$H_n^{(i)}(x_k) = F^{(i)}(x_k), k = \overline{0, n}, i = \overline{0, a_k - 1}$$

Существование и единственность данного полинома очевидны, перейдем сразу к построению $H_n(x)$. В общем случае выражение для полинома Эрмита достаточно громоздко, поэтому ограничимся рассмотрением конкретной задачи:

Построить $H_3(x) = c_0(x)f(x_0) + c_1(x)f(x_1) + c_2(x)f(x_2) + b(x)f'(x_1)$.

Запишем условия, при которых данный полином будет интерполяционным:

$$c_0(x_0) = 1, c_1(x_0) = 0, c_2(x_0) = 0, b(x_0) = 0,$$

$$c_0(x_1) = 0, c_1(x_1) = 1, c_2(x_1) = 0, b(x_1) = 0,$$

$$c_0(x_2) = 0, c_1(x_2) = 0, c_2(x_2) = 1, b(x_2) = 0,$$

$$c'_0(x_1) = 0, c'_1(x_1) = 0, c'_2(x_1) = 0, b'(x_1) = 1.$$

Будем искать $c_0(x)$ в виде $k(x - x_1)^2(x - x_2)$, k выбираем из условия $c_0(x_0) = 1$:

$$1 = k(x_0 - x_1)^2(x_0 - x_2) \implies c_0(x) = \frac{(x - x_1)^2(x - x_2)}{(x_0 - x_1)^2(x_0 - x_2)}$$

Аналогично получаем выражение для $c_2(x)$:

$$c_2(x) = \frac{(x - x_1)^2(x - x_0)}{(x_2 - x_1)^2(x_2 - x_0)}$$

Теперь вычислим $b(x)$:

$$b_1(x) = k(x - x_0)(x - x_1)(x - x_2)$$

$$b'_1(x_1) = k(x - x_0)(x - x_2) = 1$$

Откуда получаем $b_1(x) = \frac{(x-x_0)(x-x_2)}{(x-x_2)^2(x_1-x_2)}(x-x_1)$.

Далее пусть $c_1(x) = (x - x_0)(x - x_2)(ax + b)$:

$$c_1(x_1) = 1 = (x_1 - x_0)(x_1 - x_2)(ax_1 + b)$$

$$c'_1(x_1) = 0 = a(x_1 - x_0)(x_1 - x_2) + (ax_1 + b)(2x_1 - x_0 - x_2)$$

Из этих уравнений получаем:

$$a = -\frac{(2x_1 - x_0 - x_2)}{(x_1 - x_0)^2(x_1 - x_2)^2}$$

$$b = \frac{1}{(x_1 - x_0)(x_1 - x_2)} \left[1 + \frac{x_1(2x_1 - x_0 - x_2)}{(x_1 - x_0)(x_1 - x_2)} \right]$$

Подставляя найденные выражения, имеем:

$$H_3(x) = f(x_0) \cdot \frac{(x - x_1)^2(x - x_2)}{(x_0 - x_1)^2(x_0 - x_2)} + f(x_1) \cdot \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)} \cdot$$

$$\left(1 + \frac{(x_1 - x)(2x_1 - x_0 - x_2)}{(x_1 - x_0)(x_1 - x_2)} \right) + f(x_2) \cdot \frac{(x - x_1)^2(x - x_0)}{(x_2 - x_1)^2(x_2 - x_0)} +$$

$$f'(x_1) \cdot \frac{(x - x_0)(x - x_2)}{(x - x_2)^2(x_1 - x_2)}(x - x_1)$$

Погрешность полинома Эрмита

$$\psi_{H_3(x)} = f(x) - H_3(x)$$

Введем функцию $g(s) = f(s) + H_3(s) - k\omega(s)$, где k – константа, $\omega(s) = (x - x_0)(x - x_1)^2(x - x_2)$. Константу k получаем из условия $g(x) = 0$:

$$k = \frac{f(x) - H_3(x)}{\omega(x)}$$

Далее для применения теоремы Ролля требуем $\exists f^{(4)}(x) \in [a, b]$. Применив несколько раз теорему Ролля к $g(s)$, получим:

$$\exists \xi \in (a, b) : g^{(4)}(\xi) = 0,$$

откуда и получим окончательную оценку

$$f(x) - H_3(x) = \frac{f^{(4)}(\xi)}{4!}\omega(x)$$

Погрешность полинома Эрмита n -ой степени равна:

$$f(x) - H_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x-x_0)^{a_0}(x-x_1)^{a_1}\dots(x-x_n)^{a_n}$$

$$a_0 + a_1 + \dots + a_n = n + 1$$

Задача. Пусть заданы узлы x_0, x_1, x_2, x_3 , причем $x_3 \neq x_i, i = 0, 1, 2$, и значения функции f в этих узлах. Доказать, что $\lim_{x_3 \rightarrow x_1} L_3(x) = H_3(x)$.

Доказательство. Рассмотрим полином Лагранжа для функции f :

$$L_3(x) = \frac{(x-x_1)(x-x_2)(x-x_3)}{(x_0-x_1)(x_0-x_2)(x_0-x_3)}f(x_0) + \dots$$

При $x_3 \rightarrow x_1$:

$$L_3(x) \rightarrow \frac{(x-x_1)^2(x-x_2)}{(x_0-x_1)^2(x_0-x_2)}f(x_0) + \dots = H_3(x)$$

□

§5 Использование полинома Эрмита третьей степени для получения точной оценки погрешности квадратурной формулы Симпсона

Рассмотрим квадратурную формулу Симпсона для $\int_a^b f(x)dx$ на отрезке $[a, b]$ с разбиением на частичные отрезки $[x_{i-1}, x_i]$, $x_i - x_{i-1} = h$, объединение которых дает $[a, b]$. На i -ом частичном отрезке имеем:

$$\int_{x_{i-1}}^{x_i} f(x)dx = \frac{h}{6}(f_{i-1} + 4f_{i-\frac{1}{2}} + f_i)$$

$$f_i = f(x_i), f_{i-\frac{1}{2}} = f\left(x_i - \frac{h}{2}\right)$$

Если $f(x) = a_0 + a_1x + a_2x^2$, то квадратурная формула Симпсона точна (по построению). Формула Симпсона будет точна и для кубических многочленов ($f(x) = a_0 + a_1x + a_2x^2 + a_3x^3$). Чтобы показать это, найдем $\int_{x_{i-1}}^{x_i} x^3 dx$:

$$\int_{x_{i-1}}^{x_i} x^3 dx = \frac{1}{4}(x_i^4 - x_{i-1}^4) = \frac{1}{4}(x_i^2 - x_{i-1}^2)(x_i^2 + x_{i-1}^2) = \frac{h}{4}(x_i + x_{i-1})(x_i^2 + x_{i-1}^2)$$

Теперь запишем формулу Симпсона для $\int_{x_{i-1}}^{x_i} x^3 dx$ и преобразуем ее:

$$\begin{aligned} \frac{h}{6} \left(x_{i-1}^3 + 4x_{i-\frac{1}{2}}^3 + x_i^3 \right) &= \frac{h}{6} \left(x_i^3 + x_{i-1}^3 + 4 \frac{(x_i + x_{i-1})^3}{2} \right) = \\ &= \frac{h}{6} (x_i + x_{i-1}) \left(x_i^2 - x_i x_{i-1} + x_{i-1}^2 + \frac{x_i^2 + 2x_i x_{i-1} + x_{i-1}^2}{2} \right) \\ &= \frac{h}{6} (x_i + x_{i-1}) \left(\frac{3x_i^2 + 3x_{i-1}^2}{2} \right) = \frac{h}{4} (x_i + x_{i-1}) (x_i^2 + x_{i-1}^2) \end{aligned}$$

Таким образом, мы показали, что формула Симпсона точна и для многочленов третьей степени. Приближим подынтегральную функцию $f(x)$ полиномом Эрмита $H_3(x)$:

$$\begin{aligned} H_3(x_{i-1}) &= f_{i-1} \\ H_3(x_{i-\frac{1}{2}}) &= f_{i-\frac{1}{2}} \\ H_3(x_i) &= f_i \\ H_3'(x_{i-\frac{1}{2}}) &= f'_{i-\frac{1}{2}} \\ f(x) &= H_3(x) + \psi_{H_3}(x) \\ \int_{x_{i-1}}^{x_i} f(x) dx &= \int_{x_{i-1}}^{x_i} H_3(x) dx + \int_{x_{i-1}}^{x_i} \psi_{H_3}(x) dx = \\ &= \frac{h}{6} (H_3(x_{i-1}) + 4H_3(x_{i-\frac{1}{2}}) + H_3(x_i)) + \int_{x_{i-1}}^{x_i} \psi_{H_3}(x) dx = \\ &= \frac{h}{6} (f_{i-1} + 4f_{i-\frac{1}{2}} + f_i) + \int_{x_{i-1}}^{x_i} \psi_{H_3}(x) dx \end{aligned}$$

Найдем погрешность на i -ом частичном отрезке:

$$\psi_i = \int_{x_{i-1}}^{x_i} f(x) dx - \frac{h}{6} (f_{i-1} + 4f_{i-\frac{1}{2}} + f_i) = \int_{x_{i-1}}^{x_i} \psi_{H_3}(x) dx$$

Погрешность для полинома Эрмита имеет вид:

$$\psi_{H_3}(x) = \frac{f^{(4)}(\xi)}{4!} (x - x_{i-1})(x - x_{i-\frac{1}{2}})^2(x - x_i)$$

Пусть $M_4 = \sup_{\xi \in [x_{i-1}, x_i]} |f^{(4)}(\xi)|$, тогда справедлива оценка:

$$|\psi_{H_3}(x)| \leq \frac{M_4}{4!} \int_{x_{i-1}}^{x_i} (x - x_{i-1})(x - x_{i-\frac{1}{2}})^2(x_i - x) dx = O(h^5)$$

Задача. Доказать, что $\int_{x_{i-1}}^{x_i} (x - x_{i-1})(x - x_{i-\frac{1}{2}})^2(x_i - x) dx = \frac{h^5}{120}$.

Доказательство. Проведем замену в подынтегральном выражении:

$$x = x_{i-1} + th, \quad 0 \leq t \leq 1$$

Тогда:

$$\begin{aligned} dx &= h dt \\ x - x_{i-1} &= th \\ x_i - x &= h(1 - t) \\ (x - x_{i-\frac{1}{2}})^2 &= h^2 \left(t - \frac{1}{2}\right)^2 \end{aligned}$$

Таким образом, требуемый интеграл легко вычислить:

$$\begin{aligned} &\int_{x_{i-1}}^{x_i} (x - x_{i-1})(x - x_{i-\frac{1}{2}})^2(x_i - x) dx = \\ &= h^5 \int_0^1 t(1-t) \left(t - \frac{1}{2}\right)^2 dt = h^5 \int_0^1 \left(2t^3 - \frac{5}{4}t^2 - t^4 + \frac{t}{4}\right) dt = \frac{h^5}{120} \end{aligned}$$

□

Теперь мы можем оценить погрешность на всем отрезке $[a, b]$:

$$\begin{aligned} \Psi &= \sum_{i=1}^n \psi_i(h) \\ |\Psi| &\leq \frac{M_4 h^5 n}{4! \cdot 120} \end{aligned}$$

Учтем, что $hn = b - a$:

$$|\Psi| \leq \frac{M_4 h^4 (b - a)}{4! \cdot 120} = \left(\frac{h}{2}\right)^4 \frac{M_4 (b - a)}{180}$$

Замечание. Если подынтегральную функцию заменить соответствующим полиномом Лагранжа, то погрешность квадратурной формулы Симпсона увеличится.

§6 Наилучшее среднеквадратичное приближение функции

Определение. Функция $f(x)$ называется интегрируемой с квадратом на отрезке $[a, b]$, если $\int_a^b f^2(x)dx < \infty$.

Рассмотрим линейное пространство $H = L_2$ функций, интегрируемых с квадратом на отрезке $[a, b]$. Введем скалярное произведение функций $f(x)$ и $g(x)$ ($f, g \in L_2$):

$$(f, g) = \int_a^b f(x)g(x)dx$$

Теперь определим норму:

$$\|f\| = (f, f)^{\frac{1}{2}} = \left(\int_a^b f^2(x)dx \right)^{\frac{1}{2}}$$

Рассмотрим совокупность функций:

$$\phi_0(x), \phi_1(x), \dots, \phi_n(x) \text{ — ЛНЗ и интегрируемые с квадратом, } \phi_i \in L_2 \quad (1)$$

Рассмотрим обобщенный многочлен:

$$\phi(x) = \sum_{k=0}^n C_k \phi_k(x), \quad C_k \text{ — числа} \quad (2)$$

Среди всех обобщенных многочленов нам необходимо найти обобщенный многочлен $\overline{\phi(x)}$, такой что:

$$\begin{aligned} \|f(x) - \overline{\phi(x)}\| &= \min_{\phi \in L_2} \|f(x) - \phi(x)\| = \\ &= \min_{\phi \in L_2} \left(\int_a^b \left(f(x) - \sum_{k=0}^n C_k \phi_k(x) \right)^2 dx \right)^{\frac{1}{2}} \end{aligned}$$

Обобщенный многочлен $\overline{\phi(x)}$ называется наилучшим среднеквадратичным приближением функции $f(x)$. Покажем, что оно существует и единственно.

Утверждение. Наилучшее среднеквадратичное приближение существует и единственно.

Доказательство. Рассмотрим сначала случай $n = 0$:

$$\phi_0(x) \in L_2, \phi(x) = C_0 \phi_0(x)$$

Введем функцию $F(C_0)$:

$$F(C_0) = \|f - \phi(x)\|^2 = \int_a^b (f(x) - \phi(x))^2 dx =$$

$$\begin{aligned} &= \int_a^b f^2(x)dx - 2C_0 \int_a^b f(x)\phi_0(x)dx + C_0^2 \int_a^b \phi_0^2(x)dx = \\ &= (f, f) - 2C_0(f, \phi_0) + C_0^2(\phi_0, \phi_0) \end{aligned}$$

Необходимое условие минимума для функции F:

$$\frac{dF}{dC_0} = 0$$

Минимум этой функции по переменной C_0 находится в вершине параболы:

$$\overline{C_0} = \frac{(f, \phi_0)}{(\phi_0, \phi_0)}$$

Таким образом:

$$\overline{\phi(x)} = \overline{C_0}\phi_0(x)$$

Рассмотрим пример. Пусть $\phi_0(x) = 1$, тогда:

$$\overline{C_0} = \frac{\int_a^b f(x)dx}{\int_a^b dx} = \frac{1}{b-a} \int_a^b f(x)dx$$

$$\overline{\phi(x)} = \frac{1}{b-a} \int_a^b f(x)dx$$

Получили среднее значение функции $f(x)$ на $[a, b]$.

Рассмотрим теперь общий случай. Пусть задана система функций (1). Введем функцию $F(C_0, C_1, \dots, C_n)$:

$$\begin{aligned} F(C_0, C_1, \dots, C_n) &= \|f - \phi(x)\|^2 = \int_a^b \left(f(x) - \sum_{k=0}^n C_k \phi_k(x) \right)^2 dx = \\ &= \int_a^b f^2(x)dx - 2 \sum_{k=0}^n C_k \int_a^b f(x)\phi_k(x)dx + \sum_{k=0}^n C_k \sum_{l=0}^n C_l \int_a^b \phi_k^2(x)dx = \\ &= (f, f) - 2 \sum_{k=0}^n C_k (f, \phi_k) + \sum_{k=0}^n C_k \sum_{l=0}^n C_l (\phi_k, \phi_l) \end{aligned}$$

Запишем необходимое условие минимума для функции F:

$$\frac{\partial F(C_0, C_1, \dots, C_n)}{\partial C_k} = 0, \quad k = 0, 1, \dots, n$$

Тогда получим систему уравнений для нахождения коэффициентов C_i

$$\sum_{l=0}^n C_l (\phi_k, \phi_l) = (f, \phi_k), \quad k = 0, 1, \dots, n \quad (3)$$

Матрицей этой системы является матрица Грама:

$$G = \begin{pmatrix} (\phi_0, \phi_0) & (\phi_0, \phi_1) & \cdots & (\phi_0, \phi_n) \\ (\phi_1, \phi_0) & (\phi_1, \phi_1) & \cdots & (\phi_1, \phi_n) \\ \vdots & \vdots & \ddots & \vdots \\ (\phi_n, \phi_0) & (\phi_n, \phi_1) & \cdots & (\phi_n, \phi_n) \end{pmatrix}$$

Так как система функций (1) линейно независима, то определитель Грама $\det G \neq 0$, а значит наилучшее среднеквадратичное приближение существует и единственно (можно однозначно найти коэффициенты C_i). \square

Замечание. Если система функций (1) - ортонормированная, то есть $(\phi_k, \phi_l) = \delta_{kl}$, то $C_k = (f, \phi_k)$ - коэффициенты Фурье.

Замечание. Пусть задана система функций:

$$1, x, x^2, \dots, x^n$$

Пусть $\rho(x) > 0$ - весовая функция.

$$\int_{\alpha}^{\beta} \rho(x) \phi_k(x) \phi_l(x) dx = 0$$

Выбирая $\rho(x)$, α , β можно получить ортогональные многочлены.

Пусть $\{\phi_i\}_0^n$ - ортонормированная система. Тогда наименьшее отклонение:

$$\begin{aligned} 0 &\leq \int_a^b \left(f(x) - \sum_{k=0}^n C_k \phi_k(x) \right)^2 dx = \\ &= \int_a^b f^2(x) dx - 2 \sum_{k=0}^n C_k (\phi_k, f) + \sum_{k=0}^n C_k^2 = \\ &= (f, f) - \sum_{k=0}^n C_k^2 \geq 0 \end{aligned}$$

Таким образом, мы получили неравенство Бесселя:

$$\sum_{k=0}^n C_k^2 \leq \|f\|^2$$

Если система $\{\phi_k\}_0^n$ - ортонормированный базис, и $(\phi_k, \phi_l) = \delta_{kl}$, то полученное неравенство Бесселя станет равенством Парсеваля:

$$\|f\|^2 = \sum_{k=0}^{\infty} C_k^2$$

Глава III

Численное решение нелинейных уравнений и систем нелинейных уравнений

§1 Введение

Пусть задана функция $f(x)$, $x \in \mathbb{R}$, причем функция f непрерывна. Будем решать уравнение на отрезке $[a, b]$

$$f(x) = 0, x \in [a, b]$$

Процесс решения разбивают на 2 этапа:

1. Локализуем корни (при этом корни могут быть комплексные)
2. Строим итерационный метод нахождения корня

Определение. a -окрестностью корня x_* называется множество точек

$$U_a(x_*) = \{x : |x - x_*| \leq a\}$$

Рассмотрим способы локализации корня:

1. Разобьем отрезок $[a, b]$ множеством точек $\{x_i\}_1^N$

$$a \leq x_0 < x_1 < \dots < x_N \leq b$$

Тогда можно утверждать, что если $f(x_{i-1})f(x_i) < 0$, то на отрезке $[x_{i-1}, x_i]$ есть по крайней мере один корень (также их может быть нечетное число). Если же $f(x_{i-1})f(x_i) > 0$, то сказать ничего нельзя, так как на этом отрезке либо четное число корней, либо корней нет вообще.

2. Метод бисекции (деления пополам) Пусть

$$f(x) \in C[a, b]; f(a) < 0, f(b) > 0$$

Возьмем $x_0 = \frac{a+b}{2}$.

Если $f(x_0) > 0$, то корень уравнения $x_* \in (a, x_0)$
 Если $f(x_0) < 0$, то корень уравнения $x_* \in (x_0, b)$
 Если $f(x_0) = 0$, то мы нашли корень уравнения.

Во первом случае возьмем $x_1 = \frac{a+x_0}{2}$, во втором $x_1 = \frac{x_0+b}{2}$, и аналогично повотрим процедуру локализации корня, и так далее.

В случае, если дана система уравнений

$$\begin{cases} f_1(x_1, \dots, x_m) = 0, \\ \dots \\ f_m(x_1, \dots, x_m) = 0, \end{cases}$$

ее можно представить в виде $\vec{f}(\vec{x}) = 0$, где $\vec{f} = (f_1, f_2, \dots, f_m)^T$, $\vec{x} = (x_1, x_2, \dots, x_m)^T$

§2 Метод простой итерации

Итак, мы решаем уравнение

$$f(x) = 0 \quad (1)$$

x_* - корень уравнения, локализованный на $U_a(x_*)$

Заменим уравнение на эквивалентное

$$x = S(x) \quad (2)$$

$$S(x) = x + r(x)f(x) \quad (3)$$

где функция $r(x)$ не меняет знак на $U_a(x_*)$

Построим последовательность $\{x_n\}$ следующим образом:

$$x_0 \in U_a(x_*)$$

$$x_{n+1} = S(x_n), \quad n = 0, 1, \dots \quad (4)$$

Определение. Функция $S(x)$ Липшиц-непрерывна(удовлетворяет условию Липшица) с константой $q > 0$, если

$$|S(x_1) - S(x_2)| \leq q|x_1 - x_2|, \quad \forall x_1, x_2 \in (a, b)$$

Утверждение. Если $S(x)$ удовлетворяет условию Липшица с $0 < q < 1$ на $U_a(x_*)$ и $|x - x_0| < a$, то метод простой итерации (4) решения уравнения (1) сходится, причем со скоростью геометрической прогрессии со знаменателем q .

Доказательство. По построению $|x_0 - x_*| < a$, значит

$$|x_{n+1} - x_*| = |S(x_n) - S(x_*)| \leq q|x_n - x_*| \Rightarrow$$

$$|x_n - x_*| \leq q^n a$$

$\lim_{n \rightarrow \infty} q_n = 0$, так как $0 < q < 1$.

Следовательно, метод сходится, причем со скоростью геометрической прогрессии со знаменателем q . \square

Замечание. Если $S(x)$ дифференцируема на $U_a(x_*)$, то $q = \sup_{x \in U_a(x_*)} |S'(x)|$

Замечание. Пусть $f(x)$ дифференцируема, $f'(x) > 0$ на $U_a(x_*)$ и $\exists M_1 = \sup_{x \in U_a(x_*)} |f'(x)|$. Тогда запишем метод простой итерации в виде:

$$\frac{x_{n+1} - x_n}{\tau} + f(x_n) = 0, \quad \tau > 0$$

$$x_{n+1} = S(x_n), \quad S(x) = x - \tau f(x)$$

Следовательно, $\exists S'(x) = 1 - \tau f'(x)$ на $U_a(x_*)$. Для сходимости метода необходимо, чтобы $q = \sup_{x \in U_a(x_*)} |1 - \tau f'(x)| < 1$, т.е. чтобы $0 < \tau < \frac{2}{M_1}$

Метод Эйткена (ускорение сходимости)

Метод Эйткена не является теоретически обоснованным, но при приближенных значениях параметров позволяет увеличить скорость сходимости.

Пусть $x_n - x_* \simeq Aq^n$, где A и q - некоторые константы. Тогда:

$$x_{n-1} - x_* = Aq^{n-1}$$

$$x_n - x_* = Aq^n$$

$$x_{n+1} - x_* = Aq^{n+1}$$

следовательно,

$$(x_{n+1} - x_n)^2 = A^2 q^{2n} (q - 1)^2$$

$$(x_{n+1} - 2x_n + x_{n-1}) = Aq^{n-1} (q - 1)^2$$

Откуда получаем:

$$\frac{(x_{n+1} - x_n)^2}{x_{n+1} - 2x_n + x_{n-1}} = Aq^{n+1} = x_{n+1} - x_*$$

Стало быть:

$$x_* \simeq x_{n+1} - \frac{(x_{n+1} - x_n)^2}{x_{n+1} - 2x_n + x_{n-1}}$$

Из-за неточности в качестве следующей итерации мы должны взять значение, близкое к x_*

§3 Метод Ньютона и метод секущих

Мы решаем уравнение

$$f(x) = 0 \tag{1}$$

Пусть корень локализован на $U_a(x_*)$, $f(x) \in C^1(U_a(x_*))$, при этом $f'(x) \neq 0$ на $U_a(x_*)$. Разложим $f(x_*)$ по Тейлору:

$$0 = f(x_*) = f(x) + f'(x)(x_* - x) + o(x_* - x) \approx f(x) + f'(x)(x_* - x)$$

Положим в этой формуле $x = x_n$, $x_* = x_{n+1}$, тогда получим:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

Взяв $x_0 \in U_a(x_*)$, получаем метод Ньютона:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \quad n = 0, 1, 2, \dots$$

На каждой итерации считать производную затратно, в то же время на небольшом интервале она, как правило, меняется не сильно. Следовательно, можно использовать производную, один раз вычисленную на первой итерации. Получаем модифицированный метод Ньютона:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_0)}, \quad n = 0, 1, 2, \dots; x_0 \in U_a(x_*)$$

Модифицированный метод Ньютона сходится медленнее обычного метода Ньютона, но быстрее метода простой итерации.

Метод Ньютона для системы уравнений

Рассмотрим систему:

$$\begin{cases} f_1(x_1, x_2) = 0, \\ f_2(x_1, x_2) = 0, \end{cases} \quad (2)$$

Пусть (x_1^*, x_2^*) - ее решение. Разложим f_1 и f_2 в окрестности корня:

$$0 = f_1(x_1^*, x_2^*) = f_1(x_1, x_2) + \frac{\partial f_1(x_1, x_2)}{\partial x_1}(x_1^* - x_1) + \frac{\partial f_1(x_1, x_2)}{\partial x_2}(x_2^* - x_2) + \dots$$

$$0 = f_2(x_1^*, x_2^*) = f_2(x_1, x_2) + \frac{\partial f_2(x_1, x_2)}{\partial x_1}(x_1^* - x_1) + \frac{\partial f_2(x_1, x_2)}{\partial x_2}(x_2^* - x_2) + \dots$$

Заменяя x_i на x_i^n и x_i^* на x_i^{n+1} , получим:

$$f_1(x_1^n, x_2^n) + \frac{\partial f_1(x_1^n, x_2^n)}{\partial x_1}(x_1^{n+1} - x_1^n) + \frac{\partial f_1(x_1^n, x_2^n)}{\partial x_2}(x_2^{n+1} - x_2^n) = 0$$

$$f_2(x_1^n, x_2^n) + \frac{\partial f_2(x_1^n, x_2^n)}{\partial x_1}(x_1^{n+1} - x_1^n) + \frac{\partial f_2(x_1^n, x_2^n)}{\partial x_2}(x_2^{n+1} - x_2^n) = 0$$

Обозначим $x^n = (x_1^n, x_2^n)^T$, $f^n = (f_1^n, f_2^n)^T$, а также

$$I(x^n) = \begin{bmatrix} \frac{\partial f_1(x_1^n, x_2^n)}{\partial x_1} & \frac{\partial f_1(x_1^n, x_2^n)}{\partial x_2} \\ \frac{\partial f_2(x_1^n, x_2^n)}{\partial x_1} & \frac{\partial f_2(x_1^n, x_2^n)}{\partial x_2} \end{bmatrix} \quad (3)$$

Тогда уравнение можно записать в виде:

$$f(x^n) + I(x^n)(x^{n+1} - x^n) = 0 \quad (4)$$

Если $\forall n \exists I^{-1}(x^n)$, то

$$x_{n+1} = x^n - I^{-1}(x^n)f(x^n), \quad n = 0, 1, 2, \dots; \quad x_0 - \text{задано} \quad (5)$$

Замечание. Считать $I^{-1}(x^n)$ не очень удобно, поэтому обычно вводят погрешность

$$v^{n+1} = x^{n+1} - x^n$$

и решают на каждой итерации уравнение:

$$I(x^n)v^{n+1} = -f(x^n)$$

Замечание. В случае системы можно применить модифицированный метод Ньютона:

$$x^{n+1} = x^n - I^{-1}(x^n)f(x^n)$$

Но в этом случае скорость сходимости будет значительно меньше.

Если дана система из m уравнений:

$$\begin{cases} f_1(x_1, \dots, x_m) = 0, \\ f_2(x_1, \dots, x_m) = 0, \\ \dots \\ f_m(x_1, \dots, x_m) = 0, \end{cases}$$

то также можно использовать метод Ньютона, в этом случае

$$I(x^n)_{ij} = \frac{\partial f_i(x_n)}{\partial x_j}, \quad i, j = \overline{1, m}$$

Система в этом случае имеет тот же вид:

$$f(x^n) + I(x^n)(x^{n+1} - x^n) = 0$$

Метод секущих

Запишем метод Ньютона:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \quad x_0 \in U_a(x_*), \quad n = 0, 1, 2, \dots$$

Заменим в нем $f'(x^n)$ на $\frac{f(x^n) - f(x^{n-1})}{x^n - x^{n-1}}$.

Получим

$$x^{n+1} = x^n - \frac{x^n - x^{n-1}}{f(x^n) - f(x^{n-1})} f(x^n) \quad (6)$$

Поскольку в записи данного метода участвуют три последовательные итерации $(x^{n+1}, x^n$ и $x^{n-1})$, то он называется двухшаговым методом. Для того, чтобы воспользоваться им, требуется задать два начальных приближения $(x^0$ и $x^1)$. Их можно получить методом простой итерации или методом Ньютона.

Заметим, что, используя метод секущих, мы получаем x^{n+1} при помощи интерполяции функции f полиномом первой степени (линейной функцией), используя ее значение в узлах x^n и x^{n-1} .

§4 Сходимость метода Ньютона и оценка сходимости

Рассматривается нелинейное уравнение

$$f(x) = 0. \quad (1)$$

Запишем для него метод Ньютона:

$$x^{n+1} = x^n - \frac{f(x^n)}{f'(x^n)}, \quad n = 0, 1, \dots; \quad x^0 \in U_a(x_*). \quad (2)$$

Запишем это метод в более общем виде:

$$x^{n+1} = S(x^n), \quad \text{где } S(x) = x - \frac{f(x)}{f'(x)}.$$

Тогда

$$S'(x) = 1 - \frac{(f'(x))^2 - f(x)f''(x)}{(f'(x))^2} = \frac{f(x)f''(x)}{(f'(x))^2}.$$

Заметим, что $S'(x_*) = 0$.

Пусть $z_n = x^n - x_*$ — погрешность. Тогда

$$z^{n+1} = x^{n+1} - x_* = S(x^n) - S(x_*) = S(z_n + x_*) - S(x_*).$$

Воспользуемся формулой Тейлора с остаточным членом в форме Лагранжа:

$$z^{n+1} = S(x_*) + S'(x_*)z_n + \frac{1}{2}S''(\tilde{x}^n)z_n^2 - S(x_*) = \frac{1}{2}S''(\tilde{x}^n)z_n^2,$$

где $\tilde{x}^n = x^n + \theta z_n$, $|\theta| < 1$.

Пусть $\exists M > 0$ такое, что

$$\frac{1}{2}|S''(x)| \leq M, \quad x \in U_a(x_*). \quad (3)$$

Тогда

$$\begin{aligned} |z_{n+1}| &\leq M|z_n|^2, \\ M|z_{n+1}| &\leq (M|z_n|)^2. \end{aligned}$$

Применим это неравенство рекурсивно, получим

$$\begin{aligned} M|z_n| &\leq (M|z_0|)^{2^n}, \\ |z_n| &\leq \frac{1}{M}(M|z_0|)^{2^n}. \end{aligned}$$

Если $M|z_0| < 1$, то при $n \rightarrow \infty$ получаем $|z_n| \rightarrow 0 \Rightarrow x^n \rightarrow x_*$.

Таким образом, для сходимости данного метода достаточно потребовать

$$|z_0| = |x_0 - x_*| \leq \frac{1}{M}. \quad (4)$$

Для z_n имеем оценку

$$|z_n| = |x_n - x_*| \leq \frac{1}{M}(M|x_0 - x_*|)^{2^n}. \quad (5)$$

Мы доказали следующую теорему.

Теорема (об оценке скорости сходимости метода Ньютона). Пусть $\exists M > 0$ такое, что

$$\frac{1}{2} \left| \left(\frac{f(x)f'(x)}{(f'(x))^2} \right)' \right| \leq M \quad \forall x \in U_a(x_*),$$
$$|x_0 - x_*| \leq \frac{1}{M}.$$

Тогда метод Ньютона сходится и имеет место оценка

$$|x_n - x_*| \leq \frac{1}{M} (M|x_0 - x_*|)^{2^n}.$$

Замечание. Если метод Ньютона сходится, то он сходится очень быстро.

Замечание. Начальное приближение должно быть близко к корню (в соответствии с условием (4)).

Напомним, что модифицированный метод Ньютона имеет вид:

$$x^{n+1} = x^n - \frac{f(x^n)}{f'(x^0)}.$$

Для этого метода $S(x)$ имеет вид

$$S(x) = x - \frac{f(x)}{f'(x^0)}.$$

Для этого метода аналогичное утверждение не имеет места, ибо $S'(x_*) \neq 0$ в общем случае.

Глава IV

Разностные методы решения задач математической физики

§1 Разностные схемы для первой краевой задачи для уравнения теплопроводности

Рассмотрим область $D = \{(x, y) \in \mathbb{R}^2 : 0 < x < 1, 0 < t \leq T\}$ (T – заданное положительное число).

Запишем первую краевую задачу для уравнения теплопроводности в этой области:

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + f(x, t), \quad (x, t) \in D, \quad (1)$$

краевые условия:

$$\begin{cases} u(0, t) = \mu_1(t), \\ u(1, t) = \mu_2(t), \end{cases} \quad (2)$$

начальное условие:

$$u(x, 0) = u_0(x). \quad (3)$$

Введем следующие обозначения:

$$\begin{aligned} \omega_h &= \{x_i = ih, i = 1, \dots, N-1, hN = 1\}, \\ \bar{\omega}_h &= \{x_i = ih, i = 0, \dots, N, hN = 1\}, \\ \omega_\tau &= \{t_j = j\tau, j = 1, \dots, j_0, \tau j_0 = T\}, \\ \bar{\omega}_\tau &= \{t_j = j\tau, j = 0, \dots, j_0, \tau j_0 = T\}, \\ \omega_{\tau h} &= \omega_\tau \times \omega_h, \\ \bar{\omega}_{\tau h} &= \bar{\omega}_\tau \times \bar{\omega}_h, \\ u_i^n &= u(x_i, t_n), \\ f_i^n &= f(x_i, t^n). \end{aligned}$$

Множества ω_* и $\bar{\omega}_*$ называются сетками, элементы этих множеств – узлами. Значения τ и h называются шагами сетки. Внутренними узлами назовем узлы сетки $\omega_{\tau h}$.

Будем обозначать численное решение поставленной задачи через $y(x, t)$. Пусть

$$y_i^n = y(x_i, t_n).$$

Явная разностная схема

Запишем рассматриваемую задачу:

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial t^2} + f(x, t), \quad 0 < x < 1, \quad 0 < t \leq T, \quad (4)$$

краевые условия:

$$\begin{cases} u(0, t) = \mu_1(t), \\ u(1, t) = \mu_2(t), \end{cases} \quad (5)$$

начальное условие:

$$u(x, 0) = u_0(x). \quad (6)$$

Разностный аналог задачи (4) – (6) имеет вид:

$$\frac{y_i^{n+1} - y_i^n}{\tau} = \frac{y_{i-1}^n - 2y_i^n + y_{i+1}^n}{h^2} + f(x_i, t_n), \quad (x_i, t_n) \in \omega_{\tau h}, \quad (7)$$

$$\begin{cases} y_0^{n+1} = \mu_1(t_{n+1}), & t_{n+1} \in \bar{\omega}_\tau, \\ y_N^{n+1} = \mu_2(t_{n+1}), & t_{n+1} \in \bar{\omega}_\tau, \end{cases} \quad (8)$$

$$y_i^0 = u_0(x_i), \quad x_i \in \bar{\omega}_h. \quad (9)$$

Множество узлов $\{(x_i, t_n), i = 0, \dots, N\}$ называется n -м слоем.

При изучении разностных схем возникают следующие вопросы:

1. Существование и единственность решения
2. Погрешность аппроксимации разностной схемы
3. Алгоритм нахождения численного решения
4. Исследование устойчивости разностной схемы
5. Оценка скорости сходимости разностной схемы

Ответим на вопросы 1 и 3 для явной разностной схемы. Перепишем (7) в виде

$$y_i^{n+1} = y_i^n + \frac{\tau}{h^2}(y_{i-1}^n - 2y_i^n + y_{i+1}^n) + \tau f_i^n, \quad i = 1, \dots, N-1. \quad (10)$$

Значения y в граничных узлах ($i = 0, i = N$) заданы формулами (8). Значения y при $n = 0$ – формулой (9). Таким образом, решение явной разностной схемы существует и единственно и выписан алгоритм его нахождения. Задача решается по слоям, т.е. значения на $(n+1)$ -м слое находятся по явной формуле по известным значениям на n -м слое.

Определим погрешность разностной схемы x_i^n так:

$$x_i^n = y_i^n - u_i^n.$$

Введем функцию ψ_i^n так:

$$\psi_i^n = \frac{u_{i-1}^n - 2u_i^n + u_{i+1}^n}{h^2} - \frac{u_i^{n+1} - u_i^n}{\tau} + f_i^n. \quad (11)$$

Тогда (7) можно переписать следующим образом:

$$\frac{z_i^{n+1} - z_i^n}{\tau} = \frac{z_{i-1}^n - 2z_i^n + z_{i+1}^n}{h^2} + \psi_i^n, \quad (x_i, t_n) \in \omega_{\tau h}. \quad (12)$$

Определение. Функция ψ_i^n , определяемая равенством (11), называется погрешностью аппроксимации разностной схемы (7) – (9) на решение задачи (4) – (6).

Задача. Доказать, что $\psi_i^n = O(\tau + h^2)$.

Решение. Разложим $u(x_i, t_{n+1})$ в узле (x_i, t_n) по формуле Тейлора:

$$u(x_i, t_{n+1}) = u_i^{n+1} = u(x_i, t_n) + u_t(x_i, t_n)\tau + O(\tau^2).$$

Разложим $u(x_{i+1}, t_n)$ в узле (x_i, t_n) по формуле Тейлора:

$$u(x_{i+1}, t_n) = u_{i+1}^n = u(x_i, t_n) + u_x(x_i, t_n)h + \frac{1}{2}u_{xx}(x_i, t_n)h^2 + \frac{1}{6}u_{xxx}(x_i, t_n)h^3 + O(h^4).$$

Разложим $u(x_{i-1}, t_n)$ в узле (x_i, t_n) по формуле Тейлора:

$$u(x_{i-1}, t_n) = u_{i-1}^n = u(x_i, t_n) - u_x(x_i, t_n)h + \frac{1}{2}u_{xx}(x_i, t_n)h^2 - \frac{1}{6}u_{xxx}(x_i, t_n)h^3 + O(h^4).$$

Подставив выписанные разложения в (11), приведя подобные члены и воспользовавшись (4), получим

$$\psi_i^n = O(\tau + h^2).$$

□

Краевые условия для z имеют вид:

$$z_0^{n+1} = z_N^{n+1} = 0, \quad t_{n+1} = \bar{\omega}_\tau. \quad (13)$$

А начальное условие для z :

$$z_i^0 = 0, \quad x_i = \bar{\omega}_h. \quad (14)$$

Введем норму на слое:

$$\|y^n\|_C = \max_{0 \leq i \leq N} |y_i^n|.$$

Введенная таким образом норма называется равномерной (сильной).

Выразим z_i^{n+1} в формуле (12):

$$z_i^{n+1} = z_i^n + \frac{\tau}{h^2}(z_{i-1}^n - 2z_i^n + z_{i+1}^n) + \tau\psi_i^n.$$

Потребуем выполнения следующего условия:

$$\frac{\tau}{h^2} = \gamma \leq \frac{1}{2}. \quad (15)$$

Если разностная схема сходится при ограничении на шаги сетки, то такая разностная схема называется условно сходящейся. Если сходимость разностной схемы не зависит от шагов сетки, то разностная схема называется абсолютно сходящейся.

Докажем, что условие (15) является необходимым и достаточным для сходимости (и устойчивости) явной разностной схемы.

Докажем достаточность условия (15). Пусть это условие выполнено. Тогда

$$z_i^{n+1} = (1 - 2\gamma)z_i^n + \gamma(z_{i-1}^n + z_{i+1}^n) + \tau\psi_i^n,$$

$$\begin{aligned}
|z_i^{n+1}| &\leq (1 - 2\gamma)|z_i^n| + \gamma(|z_{i-1}^n| + |z_{i+1}^n|) + \tau\psi_i^n, \\
|z_i^{n+1}| &\leq (1 - 2\gamma)\|z^n\|_C + \gamma(\|z^n\|_c + \|z^n\|_C) + \tau\|\psi^n\|_C, \\
|z_i^{n+1}| &\leq \|z^n\|_C + \tau\|\psi^n\|_C,
\end{aligned}$$

поскольку это выполняется для всех i , то

$$\|z^{n+1}\|_C \leq \|z^n\|_C + \tau\|\psi^n\|_C. \quad (16)$$

Применяя формулу (16) как рекуррентную, получим

$$\|z^{n+1}\|_C \leq \|z^0\|_C + \tau \sum_{k=0}^n \|\psi^k\|_C,$$

поскольку $\|z^{n+1}\|_C = 0$, то

$$\|z^{n+1}\|_C \leq \tau \sum_{k=0}^n \|\psi^k\|_C.$$

Т.к. $\psi_i^n = O(\tau + h^2)$, то $\exists M > 0$: $\|\psi^n\|_C \leq M(\tau + h^2)$, M не зависит от τ и h .

Учитывая, что $\sum_{k=0}^n \tau = t^{n+1} \leq T$, имеем

$$\|z^{n+1}\|_C \leq MT(\tau + h^2) = M_1(\tau + h^2).$$

При этом, M_1 не зависит от τ и h .

Мы получили априорную оценку

$$\|z^{n+1}\|_C \leq M_1(\tau + h^2). \quad (17)$$

Из полученной оценки следует, что

$$\tau, h \rightarrow 0 \Rightarrow \|z^{n+1}\| \rightarrow 0, \text{ т.е. } \|y^{n+1} - u^{n-1}\| \rightarrow 0.$$

Таким образом, имеет место сходимость численного решения к решению исходной задачи.

Несколько слов об устойчивости.

Пусть $y(0, t) = y(1, t) = 0$. Тогда, проведя рассуждения, аналогичным описанным выше, имеем

$$\begin{aligned}
\|y^{n+1}\|_C &\leq \|y_0\|_C + \sum_{k=0}^n \tau \|f^k\|_C, \\
\|y^{n+1}\|_C &\leq \|y_0\|_C + \tau \sum_{k=0}^n \|f^k\|_C.
\end{aligned} \quad (18)$$

Разностную схему, в которой выполняется (18), называют устойчивой по начальному условию и правой части. Таким образом, явная разностная схема устойчива по начальному условию и правой части при выполнении условия (15).

Докажем, что условие (15) является необходимым для сходимости явной разностной схемы. Рассмотрим однородную систему

$$\frac{y_i^{n+1} - y_i^n}{\tau} = \frac{y_{i-1}^n - 2y_i^n + y_{i+1}^n}{h^2}, \quad (x_i, t_n) \in \omega_{\tau h}. \quad (19)$$

Будем искать ее решение в виде $y_j^n = q^n e^{ijh\phi}$, где $i^2 = -1$, $\phi \in \mathbb{R}$, $q \in \mathbb{C}$. Подставим это в уравнение (19). Получим

$$q = 1 + \gamma(e^{ih\phi} - 2 + e^{-ih\phi}) = 1 + \gamma(2 \cos h\phi - 2) = 1 - 4\gamma \sin^2 \frac{h\phi}{2}.$$

Если взять ϕ такое, что $|q| > 1$, т.е. $\gamma > \frac{1}{2}$, то гармоники будут неограниченно возрастать и разностная схема будет расходиться.

Таким образом, условие (15) является необходимым и достаточным для сходимости и устойчивости явной разностной схемы.

Чисто неявная разностная схема (схема с опережением)

Запишем рассматриваемую задачу:

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial t^2} + f(x, t), \quad 0 < x < 1, \quad 0 < t \leq T, \quad (20)$$

краевые условия:

$$\begin{cases} u(0, t) = \mu_1(t), \\ u(1, t) = \mu_2(t), \end{cases} \quad (21)$$

начальное условие:

$$u(x, 0) = u_0(x). \quad (22)$$

Разностный аналог задачи (20) – (22) имеет вид:

$$\frac{y_i^{n+1} - y_i^n}{\tau} = \frac{y_{i-1}^{n+1} - 2y_i^{n+1} + y_{i+1}^{n+1}}{h^2} + f(x_i, t_{n+1}), \quad (x_i, t_{n+1}) \in \omega_{\tau h}, \quad (23)$$

$$\begin{cases} y_0^{n+1} = \mu_1(t_{n+1}), & t_{n+1} \in \bar{\omega}_\tau, \\ y_N^{n+1} = \mu_2(t_{n+1}), & t_{n+1} \in \bar{\omega}_\tau, \end{cases} \quad (24)$$

$$y_i^0 = u_0(x_i), \quad x_i \in \bar{\omega}_h. \quad (25)$$

Перепишем (23) в виде:

$$\gamma y_{i-1}^{n+1} - (1 + 2\gamma)y_i^{n+1} + \gamma y_{i+1}^{n+1} = -(y_i^n + f_i^{n+1}), \quad i = 1, \dots, N-1.$$

Данная система уравнений состоит из трехточечных уравнений. Ее матрица A является трехдиагональной. Эта система решается методом прогонки. Можно доказать, что $|A| \neq 0$. Таким образом, решение данной системы существует и единственно, и находится методом прогонки.

Введем погрешность:

$$z_i^n = y_i^n - u(x_i, t_n) = y_i^n - u_i^n$$

Тогда для погрешности получим уравнение:

$$\frac{z_i^{n+1} - z_i^n}{\tau} = \frac{z_{i+1}^{n+1} - 2z_i^{n+1} + z_{i-1}^{n+1}}{h^2} + \psi_i^n, \quad (26)$$

где

$$\psi_i^n = \frac{u_{i+1}^{n+1} - 2u_i^{n+1} + u_{i-1}^{n+1}}{h^2} - \frac{u_i^{n+1} - u_i^n}{\tau} + f_i^{n+1} \quad (27)$$

Задача. Показать, что ψ_i^n из (27) есть $O(\tau + h^2)$.

Решение. Разложим $u_{i\pm 1}^{n+1}$ и u_i^n в ряд Тейлора:

$$u_{i\pm 1}^{n+1} = u_i^{n+1} \pm u_{x,i}^{n+1}h + u_{xx,i}^{n+1}\frac{h^2}{2} \pm u_{xxx,i}^{n+1}\frac{h^3}{6} + O(h^4)$$

$$u_i^n = u_i^{n+1} - u_{t,i}^{n+1}\tau + O(\tau^2)$$

Подставим эти разложения в формулу (27). Получим:

$$\psi_i^n = (-u_{t,i}^{n+1} + u_{xx,i}^{n+1} + f_i^{n+1}) + O(\tau + h^2) = O(\tau + h^2)$$

□

Заметим, что:

$$z_0^{n+1} = z_N^{n+1} = z_i^0 = 0, \quad i = 0, \dots, N \quad (28)$$

Пусть i_0 , такой что:

$$|z_{i_0}^{n+1}| = \max_{1 \leq i \leq N} |z_i^{n+1}| = \|z^{n+1}\|_C$$

$$z_i^{n+1} = z_i^n + \gamma(z_{i+1}^{n+1} - 2z_i^{n+1} + z_{i-1}^{n+1}) + \tau\psi_i^n, \quad \gamma = \frac{\tau}{h^2}$$

$$(1 + 2\gamma)z_i^{n+1} = z_i^n + \gamma(z_{i+1}^{n+1} + z_{i-1}^{n+1}) + \tau\psi_i^n$$

Запишем последнее равенство для узла i_0 :

$$(1 + 2\gamma)z_{i_0}^{n+1} = z_{i_0}^n + \gamma(z_{i_0+1}^{n+1} + z_{i_0-1}^{n+1}) + \tau\psi_{i_0}^n$$

$$(1 + 2\gamma)|z_{i_0}^{n+1}| \leq |z_{i_0}^n| + \gamma(|z_{i_0+1}^{n+1}| + |z_{i_0-1}^{n+1}|) + \tau|\psi_{i_0}^n|$$

$$(1 + 2\gamma)\|z^{n+1}\|_C \leq \|z^n\|_C + 2\gamma\|z^{n+1}\|_C + \tau\|\psi^n\|_C$$

$$\|z^{n+1}\|_C \leq \|z^n\|_C + \tau\|\psi^n\|_C$$

Последнее соотношение является рекуррентным. Применим его n раз:

$$\|z^{n+1}\|_C \leq \|z^0\|_C + \sum_{k=0}^n \tau\|\psi^k\|_C$$

Из (28) имеем: $\|z^0\|_C = 0$. Так как $\|\psi^k\| \leq M(\tau + h^2)$, где M не зависит от τ и h , то:

$$\|z^{n+1}\|_C \leq M \sum_{k=0}^n \tau(\tau + h^2)$$

Таким образом, окончательно получаем:

$$\|z^{n+1}\|_C \leq M_1(\tau + h^2), \quad M_1 = M\tau_{n+1} - \text{не зависит от } \tau \text{ и } h.$$

Из последнего соотношения следует, что чисто неявная разностная схема абсолютно сходится (имеем абсолютную сходимость первого порядка по τ и второго порядка по h). Если $y_0^{n+1} = y_N^{n+1} = 0$, то:

$$\|y^{n+1}\|_C \leq \|u_0\|_C + \sum_{k=0}^n \tau\|f^k\|_C$$

Таким образом, получаем устойчивость чисто неявной разностной схемы по начальному приближению и правой части.

Симметричная разностная схема (схема Кранка-Никольсона)

Обозначим через $y_{\bar{x},i}^m$ вторую разностную производную по пространственной переменной:

$$y_{\bar{x},i}^m = \frac{y_{i+1}^m - 2y_i^m - y_{i-1}^m}{h^2}$$

Разностная схема имеет вид:

$$\frac{y_i^{n+1} - y_i^n}{\tau} = 0.5(y_{\bar{x},i}^{n+1} + y_{\bar{x},i}^n) + f(x_i, t_n + 0.5\tau) \quad (29)$$

$$y_0^{n+1} = \mu_1(t_{n+1}), y_N^{n+1} = \mu_2(t_{n+1}), \quad t_{n+1} \in \bar{\omega}_t \quad (30)$$

$$y_i^0 = u_0(x_i), \quad x_i \in \bar{\omega}_h \quad (31)$$

Введем погрешность: $z_i^n = y_i^n - u_i^n$. Тогда для погрешности имеем:

$$\frac{z_i^{n+1} - z_i^n}{\tau} = 0.5(z_{\bar{x},i}^{n+1} + z_{\bar{x},i}^n) + \psi_i^n, \quad (x_i, t_{n+1}) \in \omega_{\tau h} \quad (32)$$

$$z_0^{n+1} = z_N^{n+1} = 0, \quad z_i^0 = 0, \quad i = 0, \dots, N \quad (33)$$

$$\psi_i^n = 0.5(u_{\bar{x},i}^{n+1} + u_{\bar{x},i}^n) - \frac{u_i^{n+1} - u_i^n}{\tau} + f(x_i, t_n + 0.5\tau) \quad (34)$$

Задача. Показать, что ψ_i^n из (34) есть $\underline{O}(\tau^2 + h^2)$.

Решение. Разложим $u_{i\pm 1}^{n+1}$ и u_i^n в ряд Тейлора в окрестности точки $(x_i, t_{n+\frac{1}{2}})$:

$$u_i^{n+1} = u_i^{n+\frac{1}{2}} + u_{t,i}^{n+\frac{1}{2}} \frac{\tau}{2} + \frac{1}{2} u_{tt,i}^{n+\frac{1}{2}} \left(\frac{\tau}{2}\right)^2 + \underline{O}(\tau^3)$$

$$u_i^n = u_i^{n+\frac{1}{2}} - u_{t,i}^{n+\frac{1}{2}} \frac{\tau}{2} + \frac{1}{2} u_{tt,i}^{n+\frac{1}{2}} \left(\frac{\tau}{2}\right)^2 + \underline{O}(\tau^3)$$

Подставим эти разложения в формулу (34):

$$\psi_i^n = -u_{t,i}^{n+\frac{1}{2}} + \underline{O}(\tau^2) + 0.5(u_{\bar{x},i}^{n+1} + u_{\bar{x},i}^n) + f_i^{n+\frac{1}{2}}$$

Теперь в представлении второй разностной производной разложим все вхождения функции в ряд Тейлора. Приводя подобные слагаемые, получим:

$$u_{\bar{x},i}^n = u_{xx,i}^n + u_{xxxx,i}^n \frac{h^2}{12} + \underline{O}(h^4)$$

Применим это разложение к $u_{\bar{x},i}^{n+1}$, а затем проведем еще одно разложение в ряд Тейлора в точке $(x_i, t_{n+\frac{1}{2}})$:

$$\begin{aligned} u_{\bar{x},i}^{n+1} &= u_{xx,i}^{n+1} + u_{xxxx,i}^{n+1} \frac{h^2}{12} + \underline{O}(h^4) = \\ &= u_{xx,i}^{n+\frac{1}{2}} + u_{xxt,i}^{n+\frac{1}{2}} \frac{\tau}{2} + u_{xxxx,i}^{n+\frac{1}{2}} \frac{h^2}{12} + u_{xxxxt,i}^{n+\frac{1}{2}} \frac{h^2}{12} \cdot \frac{\tau}{2} + \underline{O}(\tau^2 + h^4) \end{aligned}$$

То же самое сделаем и с $u_{\bar{x}x,i}^n$:

$$\begin{aligned} u_{\bar{x}x,i}^n &= u_{xx,i}^n + u_{xxxx,i}^n \frac{h^2}{12} + \underline{O}(h^4) = \\ &= u_{xx,i}^{n+\frac{1}{2}} - u_{xxt,i}^{n+\frac{1}{2}} \frac{\tau}{2} + u_{xxxx,i}^{n+\frac{1}{2}} \frac{h^2}{12} - u_{xxxxt,i}^{n+\frac{1}{2}} \frac{h^2}{12} \cdot \frac{\tau}{2} + \underline{O}(\tau^2 + h^4) \end{aligned}$$

Подставим эти разложения в выражение для ψ_i^n и учтем уравнение теплопроводности:

$$\psi_i^n = (-u_{t,i}^{n+\frac{1}{2}} + u_{xx,i}^{n+\frac{1}{2}} + f_i^{n+\frac{1}{2}}) + u_{xxxx,i}^{n+\frac{1}{2}} \frac{h^2}{12} + \underline{O}(\tau^2 + h^4) = \underline{O}(\tau^2 + h^2)$$

□

Задача Штурма-Лиувилля

Рассмотрим задачу Штурма-Лиувилля для дифференциального уравнения второго порядка:

$$\begin{cases} u''(x) + \lambda u(x) = 0, & 0 < x < 1, \\ u(0) = u(1) = 0; \end{cases}$$

$u(x)$, не равные тождественно нулю, - собственные функции ЗШЛ, а λ - собственные значения ЗШЛ. Решением данной задачи являются собственные значения λ_k и собственные функции $u_k(x)$:

$$\begin{aligned} \lambda_k &= (\pi k)^2, \quad k = 1, 2, \dots \\ 0 &< \lambda_1 < \lambda_2 < \dots < \lambda_n < \dots \\ u_k(x) &= C \sin(\pi k x) \end{aligned}$$

Рассмотрим линейное пространство L_2 функций, интегрируемых с квадратом на $[0, 1]$ ($\int_0^1 f^2(x) dx < \infty$). Введем скалярное произведение и норму в L_2 :

$$\forall f, g \in L_2 : (f, g) = \int_0^1 f(x)g(x) dx$$

$$\|f\|_{L_2} = \left(\int_0^1 f^2(x) dx \right)^{\frac{1}{2}}$$

Возьмем $C = \sqrt{2}$, тогда $(u_k, u_l) = \delta_{kl}$, то есть $\{u_k(x)\}_{k=1}^{\infty}$ - ортонормированный базис в L_2 .

Таким образом, любую функцию $f \in L_2$ можно представить в виде: $f(x) = \sum_{k=1}^{\infty} C_k u_k(x)$, где C_k - коэффициенты Фурье. Имеет место равенство Парсеваля:

$$\|f\|_{L_2}^2 = \sum_{k=1}^{\infty} C_k^2$$

Рассмотрим дискретный аналог задачи Штурма-Лиувилля:

$$\begin{cases} y_{\bar{x},i} + \lambda y_i = 0, & i = 1, \dots, N-1, \\ y_0 = y_N = 0, \\ y_i \text{ не равны тождественно } 0; \end{cases} \quad (35)$$

Подставим в первое уравнение представление второй разностной производной:

$$y_{i+1} - 2y_i + y_{i-1} + \lambda h^2 y_i = 0$$

$$y_{i+1} + y_{i-1} = (2 - \lambda h^2) y_i$$

Будем искать $y_i = y(x_i)$ в виде $\sin(\alpha x_i)$, $\alpha \in \mathbb{R}$. Тогда:

$$y_{i+1} = \sin(\alpha(x_i + h)), \quad y_{i-1} = \sin(\alpha(x_i - h))$$

$$y_{i+1} + y_{i-1} = \sin(\alpha(x_i + h)) + \sin(\alpha(x_i - h)) = 2 \sin(\alpha x_i) \cos(\alpha h)$$

$$2 \sin(\alpha x_i) \cos(\alpha h) = (2 - \lambda h^2) \sin(\alpha x_i)$$

$$2 \cos(\alpha h) = (2 - \lambda h^2)$$

$$\lambda = \frac{2(1 - \cos(\alpha h))}{h^2} = \frac{4 \sin^2 \frac{\alpha h}{2}}{h^2} \quad (36)$$

Из условия $y_0 = y_N = 0$ имеем:

$$\sin \alpha = 0, \quad \alpha = \pi k, \quad k \in \mathbb{Z}$$

Итак, мы нашли собственные значения и собственные функции дискретной задачи Штурма-Лиувилля:

$$\lambda_k = \frac{4}{h^2} \sin^2 \frac{\alpha h}{2}$$

$$y_k(x_i) = \sqrt{2} \sin(\pi k x_i), \quad i = 0, \dots, N, \quad k = 1, \dots, N-1$$

Константу выбираем равной $\sqrt{2}$ из соображений ортонормированности.

Введем H - линейное пространство сеточных функций: $\dim H = N-1, \forall u \in H : u_0 = u_N = 0$. Определим скалярное произведение и норму в H :

$$\forall u, v \in H : (u, v) = \sum_{i=1}^{N-1} u_i v_i h$$

$$\|u\|_H = \left(\sum_{i=1}^N u_i^2 h \right)^{\frac{1}{2}}$$

Пусть $(y_k, y_l) = \delta_{kl}$, тогда $\{y_k(x_i)\}_{k=1}^{N-1}$ - ортонормированный базис в H , и любую функцию $u \in H$ можно представить в виде:

$$u(x_i) = \sum_{k=1}^{N-1} C_k y_k(x_i), \quad i = 0, \dots, N$$

Имеет место также и равенство Парсеваля:

$$\|u\|_H^2 = \sum_{k=1}^{N-1} C_k^2$$

Пусть $\mu_k(x_i) \equiv y_k(x_i)$, $k = 1, \dots, N-1$, $i = 0, \dots, N$, - собственные функции дискретной ЗШЛ (35). Вернемся к (32) - (34). Разложим z_i^n и ψ_i^n по базису из μ_n :

$$z_i^n = \sum_1^{N-1} k(t_n) \mu_k(x_n)$$

$$\psi_i^n = \sum_1^{N-1} \psi^{(k)}(t_n) \mu_k(x_n)$$

Подставим эти разложения в (32):

$$\sum_{k=1}^{N-1} \mu_k(x_i) (c_k(t_{n+1}) - c_k(t_n)) =$$

$$= 0.5\tau \sum_{k=1}^{N-1} (\mu_k)_{\bar{x}x,i} (c_k(t_{n+1}) + c_k(t_n)) + \tau \sum_1^{N-1} \psi^{(k)}(t_n) \mu_k(x_n)$$

$$\frac{c_k(t_{n+1}) - c_k(t_n)}{\tau} + 0,5\lambda_k (c_k(t_{n+1}) + c_k(t_n)) = \psi^{(k)}(t_n)$$

$$n = 0, 1, \dots, k = \overline{1, N-1}, c_k(0) = (r(0), \mu_k) = 0$$

Разрешим уравнение относительно $(n+1)$ -го слоя.

$$c_k(t_{n+1}) = \frac{1 - 0,5\tau\lambda_k}{1 + 0,5\tau\lambda_k} c_k(t_n) + \frac{\tau}{1 + 0,5\tau\lambda_k} \psi^{(k)}(t_n)$$

Положим $q_k = \frac{1 - 0,5\tau\lambda_k}{1 + 0,5\tau\lambda_k}$

$$c_k(t_{n+1}) = q_k c_k(t_n) + \frac{\tau}{1 + 0,5\tau\lambda_k} \psi^{(k)}(t_n)$$

Тогда:

$$z_i^{n+1} = \sum_{k=1}^{N-1} c_k(t_{n+1}) \mu^{(k)}(x_i) =$$

$$\sum_{k=1}^{N-1} q_k c_k(t_n) \mu^k(x_i) + \sum_{k=1}^{N-1} \frac{\tau}{1 + 0,5\tau\lambda_k} \psi^k(t_n) \mu^k(x_i) = v_i^{n+1} + w_i^{n+1}$$

Очевидно,

$$\|z^{n+1}\| \leq \|v^{n+1}\| + \|w^{n+1}\| \quad (37)$$

Оценим $\|v^{n+1}\|$, используя равенство Парсеваля.

$$|q_k| < 1 \Rightarrow \|v^{n+1}\|^2 = \sum_{k=1}^{N-1} q_k^2 c_k^2(t_n) \leq \sum_{k=1}^{N-1} c_k^2(t_n) = \|z^n\|^2 \quad (38)$$

Аналогично,

$$\|w^{n+1}\|^2 \leq \tau^2 \|\psi^n\|^2 \quad (39)$$

Учитывая (38) и (39) неравенство (37) примет вид:

$$\|z^{n+1}\| \leq \|z^n\| + \tau \|\psi^n\| \leq \|z^0\| + \sum_{k=1}^{N-1} \tau \|\psi^k\|$$

Из ранее решенной задачи: $\|\psi^k\| \leq M(\tau^2 + h^2) \Rightarrow$

$$\|z^{n+1}\| \leq MT(\tau^2 + h^2) \rightarrow 0 \text{ при } \tau, h \rightarrow 0, \quad 0 < M = const \quad (40)$$

где M и T не зависят от τ и h .

Разностная схема с весами. Погрешность аппроксимации.

Построим для задачи (1) разностную схему:

$$\frac{y_i^{n+1} - y_i^n}{\tau} = \sigma y_{\bar{x}\bar{x},i}^{n+1} + (1 - \sigma) y_{\bar{x}\bar{x},i}^n + \phi_i^n \in \omega_{\tau h} \quad (41)$$

$$y_0^{n+1} = \mu_1(t_{n+1}), \quad t_{n+1} \in \bar{\omega}_\tau$$

$$y_N^{n+1} = \mu_2(t_{n+1}), \quad t_{n+1} \in \bar{\omega}_\tau$$

$$y_i^0 = u_0(x_i), \quad x_i \in \bar{\omega}_h$$

$$\sigma \in \mathbb{R}, \quad 0 \leq \sigma \leq 1$$

Для различных σ получаем:

1. $\sigma = 0$ - явная разностная схема.
2. $\sigma = 1$ - чисто неявная разностная схема.
3. $\sigma = 0.5$ - симметричная разностная схема.
4. $\sigma \neq 0, 1, 0.5$ - неявная разностная схема.

Введем погрешность $z_i^n = y_i^n - u_i^n$.

$$\frac{z_i^{n+1} - z_i^n}{\tau} = \sigma z_{\bar{x}\bar{x},i}^{n+1} + (1 - \sigma) z_{\bar{x}\bar{x},i}^n + \psi_i^n$$

$$z_0^{n+1} = z_N^{n+1} = z_i^0 = 0$$

Погрешность аппроксимации разностной схемы (41) на решении:

$$\psi_i^n = \sigma u_{\bar{x},i}^{n+1} + (1 - \sigma)u_{\bar{x},i}^n - \frac{u_i^{n+1} - u_i^n}{\tau} + \phi_i^n \quad (42)$$

Обозначим $u' = \frac{\partial u}{\partial x}$, $\dot{u} = \frac{\partial u}{\partial t}$. Пусть функция $u(x, t)$ шесть раз непрерывно дифференцируема по x и три раза по t . Разложим ее по формуле Тейлора в окрестности точки $(x_i, t_{n+\frac{1}{2}})$:

$$u_{i+1} = u_i + hu'_i + \frac{h^2}{2}u''_i + \frac{h^3}{6}u'''_i + \frac{h^4}{24}u''''_i + \dots$$

$$u_{i-1} = u_i - hu'_i + \frac{h^2}{2}u''_i - \frac{h^3}{6}u'''_i + \frac{h^4}{24}u''''_i + \dots$$

$$u_i^{n+1} = u_i(t_{n+\frac{1}{2}}) + \frac{\tau}{2}\dot{u}_i(t_{n+\frac{1}{2}}) + \frac{\tau^2}{8}\ddot{u}_i(t_{n+\frac{1}{2}}) + \frac{\tau^3}{48}\dddot{u}_i(t_{n+\frac{1}{2}}) + \dots$$

$$u_i^n = u_i(t_{n+\frac{1}{2}}) - \frac{\tau}{2}\dot{u}_i(t_{n+\frac{1}{2}}) + \frac{\tau^2}{8}\ddot{u}_i(t_{n+\frac{1}{2}}) - \frac{\tau^3}{48}\dddot{u}_i(t_{n+\frac{1}{2}}) + \dots$$

$$u_{\bar{x},i} = \frac{u_{i+1} - 2u_i + u_{i-1}}{h^2} = u''_i + \frac{h^2}{12}u''''_i + O(h^4)$$

$$\frac{u_i^{n+1} - u_i^n}{\tau} = \dot{u}_i(t_{n+\frac{1}{2}}) + O(\tau^2)$$

Воспользуемся неравенством $\tau h^2 \leq \frac{\tau^2 + h^4}{2}$:

$$\psi_i'' = \sigma(u''_i + \frac{\tau}{2}\dot{u}''_i + \frac{h^2}{12}u''''_i + O(h^4) + O(\tau^2)) +$$

$$(1 - \sigma)(u''_i - \frac{\tau}{2}\dot{u}''_i + \frac{h^2}{12}u''''_i + O(h^4) + O(\tau^2)) -$$

$$\dot{(u)}_i + \phi_i^n + O(\tau^2 + h^4) =$$

$$(u''_i - \dot{u}_i + \phi_i^n) + (\sigma - 0.5)\tau\dot{u}'' + \frac{h^2}{12}u''''_i + O(\tau^2 + h^4)$$

Продифференцировав уравнение $u'' - \dot{u} + f = 0$ дважды по x , получим:

$$u'''' - \dot{u}'' + f'' = 0 \Rightarrow u'''' = \dot{u}'' + f''$$

Подставим u''''_i в формулу погрешности аппроксимации:

$$\psi_i^n = \underbrace{u''_i - \dot{u}_i + f(x_i, t_{n+\frac{1}{2}})}_{=0} - \frac{h^2}{12}f''(x_i, t_{n+\frac{1}{2}}) + \phi_i - f(x_i, t_{n+\frac{1}{2}}) +$$

$$\left((\sigma - 0.5)\tau + \frac{h^2}{12} \right) \dot{u}'' + O(\tau^2 + h^4)$$

Таким образом, порядок погрешности аппроксимации зависит от параметра σ и аппроксимации функции f :

$$1. \sigma = \sigma^* = \frac{1}{2} - \frac{h^2}{12\tau} \Rightarrow$$

$$\begin{aligned} \phi_i^n &= f(x_i, t_{n+\frac{1}{2}}) + \frac{h^2}{12} f''(x_i, t_{n+\frac{1}{2}}) \\ \psi_i^n &= O(\tau^2 + h^4) \end{aligned}$$

$$2. \sigma = 0.5 \Rightarrow$$

$$\begin{aligned} \phi_i^n &= f(x_i, t_{n+\frac{1}{2}}) + O(h^2) + O(\tau^2) \\ \psi_i^n &= O(\tau^2 + h^2) \end{aligned}$$

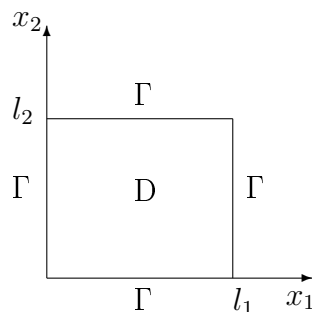
$$3. \sigma \neq \sigma^*, \sigma \neq 0.5 \Rightarrow \phi_i^n = f(x_i, t_n) + O(\tau + h^2) \Rightarrow \psi_i^n = O(\tau + h^2)$$

Разностные методы для уравнения Пуассона

Рассмотрим уравнение Пуассона в области D:

$$\frac{\partial^2 u}{\partial^2 x_1} + \frac{\partial^2 u}{\partial^2 x_2} = f(x_1, x_2)$$

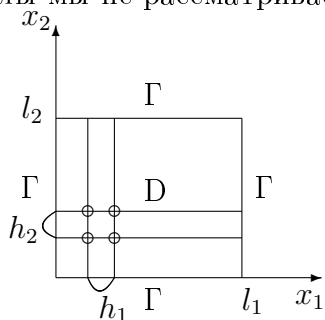
$$(x_1, x_2) \in D, D = \{(x_1, x_2) : 0 < x_1 < l_1; 0 < x_2 < l_2\}$$



Введем на области D сетку:

$$\begin{aligned} \omega_h = \{ & (x_1^{(i)}, x_2^{(j)}), x_1^{(i)} = ih, i = \overline{1, N_1 - 1}, h_1 N_1 = l_1 \\ & x_2^{(j)} = jh_2, j = \overline{1, N_2 - 1}, h_2 N_2 = l_2 \} \end{aligned}$$

Заметим, что при этом граничные узлы мы не рассматриваем.



Γ_h - граничные узлы.

$$\Gamma_h = \{x_{0,j}\}_{j=1}^{N_2-1} \cup \{x_{N_1,j}\}_{j=1}^{N_2-1} \cup \{x_{i,0}\}_{i=1}^{N_1-1} \cup \{x_{i,N_2}\}_{i=1}^{N_1-1}$$

$$\overline{\omega}_h = \omega_h \cup \Gamma_h$$

§2 Разностные схемы для уравнения Пуассона (задача Дирихле)

Здесь будет рисунок с формулировки задачи.

$$G = \{(x_1, x_2) : 0 < x_1 < l_1, 0 < x_2 < l_2\},$$

$$\bar{G} = G \cup \Gamma$$

$$\frac{d^2 u}{dx_1^2} + \frac{d^2 u}{dx_2^2} = f(x_1, x_2), (x_1, x_2) \in G \quad (1)$$

$$U|_r = \mu(x_1, x_2) \quad (2)$$

Введем сетку:

$$\omega_h = \{(x_1^{(i)}, x_2^{(j)}) : x_1^{(i)} = ih_1, i = \overline{1, N_1 - 1}, N_1 h_1 = l_1;$$

$$x_2^{(j)} = jh_2, j = \overline{1, N_2 - 1}, N_2 h_2 = l_2\},$$

$$\Gamma_h = \{x_{0,j}\}_{j=1}^{N_2-1} \cup \{x_{N_1,j}\}_{j=1}^{N_2-1} \cup \{x_{i,0}\}_{i=1}^{N_1-1} \cup \{x_{i,N_2}\}_{i=1}^{N_1-1}$$

$$\bar{\omega}_h = \omega_h \cup \Gamma_h$$

Пусть $y_{ij} = y(x_1^i, x_2^j)$, $f_{ij} = f(x_1^i, x_2^j)$.

Запишем разностную схему для задачи (1), (2):

$$y_{\bar{x}_1 x_1, ij} + y_{\bar{x}_2 x_2, ij} = f_{ij}, (x_1^i, x_2^j) \in \omega_h \quad (3)$$

$$y_{ij}|_{\Gamma_h} = \mu(x_1^i, x_2^j), (x_1^i, x_2^j) \in \Gamma_h \quad (4)$$

(3) и (4) представляют собой СЛАУ. Распишем (3):

$$\frac{y_{i+1,j} - 2y_{ij} + y_{i-1,j}}{h_1^2} + \frac{y_{i,j+2} - 2y_{ij} + y_{i,j-1}}{h_2^2} = f_{ij}$$

$$y_{ij}|_{\Gamma_h} = \mu_{ij}, \quad i = \overline{1, N_1 - 1}, j = \overline{1, N_2 - 1}$$

Погрешность разностного решения $Z_{ij} = y_{ij} - U_{ij}$ удовлетворяет задаче

$$Z_{\bar{x}_1 x_1, ij} + Z_{\bar{x}_2 x_2, ij} = -\psi_{ij} \quad (5)$$

$$Z_{ij}|_{\Gamma_h} = 0$$

Покажем существование и единственность решения системы (3). Докажем, что решение, соответствующее однородной системе, тривиально. Соответственно, решение неоднородной системы существует и единственно. Перепишем систему (3) в виде:

$$\left(\frac{2}{h_1^2} + \frac{2}{h_2^2}\right) * V_{i,j} = \frac{V_{i+1,j} + V_{i-1,j}}{h_1^2} + \frac{V_{i,j+1} + V_{i,j-1}}{h_2^2}, \quad (6)$$

$$0 < i < N_1,$$

$$0 < j < N_2.$$

Теорема. Система (6) имеет только тривиальное решение.

Доказательство. Предположим, что найдется такой узел x_{ij} , где $v_{ij} \neq 0$. Тогда $\exists i_0, j_0$, такие что:

а)

$$|V_{i_0, j_0}| = \max_{\substack{0 \leq i \leq N_1 \\ 0 \leq j \leq N_2}} |V_{ij}|$$

б) хотя бы в одном узле $(i_0, j_0 \pm 1)$, $(i_0 \pm 1, j_0)$ будет выполнено $|V_{ij}| < |V_{i_0, j_0}|$

Рассмотрим разностную схему в узле i_0, j_0 :

$$\left(\frac{2}{h_1^2} + \frac{2}{h_2^2}\right) * V_{i_0, j_0} = \frac{V_{i_0+1, j_0} + V_{i_0-1, j_0}}{h_1^2} + \frac{V_{i_0, j_0+1} + V_{i_0, j_0-1}}{h_2^2}$$

Оценим по модулю значение левой части уравнения:

$$\left(\frac{2}{h_1^2} + \frac{2}{h_2^2}\right) * |V_{i_0, j_0}| \leq \frac{2\|V_{i_0+1, j_0}\|_C}{h_1^2} + \frac{2\|V_{i_0, j_0+1}\|_C}{h_2^2}$$

Так как $|V_{i_0, j_0}| = \|V\|_C$:

$$\left(\frac{2}{h_1^2} + \frac{2}{h_2^2}\right) * \|V\|_C < \left(\frac{2}{h_1^2} + \frac{2}{h_2^2}\right) * \|V\|_C$$

Пришли к противоречию. Следовательно, предположение неверно и теорема доказана. \square

Следствие. Разностная задача имеет единственное решение для любых функций f и μ .

§3 Сходимость разностной задачи Дирихле

Рассмотрим задачу:

$$\begin{aligned} Z_{\bar{x}_1, x_1, ij} + Z_{\bar{x}_2, x_2, ij} &= -\psi, & x_{ij} \in \omega_h \\ Z_{ij}|_{\Gamma_h} &= 0, & x_{ij} \in \Gamma_h \end{aligned} \quad (1)$$

Введем разностный оператор:

$$L_h V_{ij} = \left(\frac{2}{h_1^2} + \frac{2}{h_2^2}\right) * V_{ij} - \frac{V_{i_0+1, j_0} + V_{i_0-1, j_0}}{h_1^2} + \frac{V_{i_0, j_0+1} + V_{i_0, j_0-1}}{h_2^2}, \quad x_{ij} \in \omega_h$$

Утверждение. Пусть $V_{ij} \geq 0$, $X_{ij} \in \Gamma_h$, $L_h V_{ij} \geq 0$, $x_{ij} \in \omega_h$. Тогда $V_{ij} \geq 0$.

Доказательство. Докажем от противного. Предположим, что $\exists(i_0, j_0)$ такие, что:

а)

$$|V_{i_0, j_0}| = \min_{\substack{0 \leq i \leq N_1 \\ 0 \leq j \leq N_2}} |V_{ij}|$$

б) хотя бы в одном узле $(i_0, j_0 \pm 1)$, $(i_0 \pm 1, j_0)$ будет выполнено $V_{i_0, j_0} < V_{ij}$

Тогда:

$$L_h V_{i_0, j_0} = \frac{V_{i_0, j_0} + V_{i_0+1, j_0}}{h_1^2} + \frac{V_{i_0, j_0} + V_{i_0-1, j_0}}{h_1^2} + \frac{V_{i_0, j_0} + V_{i_0, j_0+1}}{h_2^2} + \frac{V_{i_0, j_0} + V_{i_0, j_0-1}}{h_2^2}$$

Согласно условию, хотя бы одно из этих слагаемых меньше 0. Следовательно сумма тоже меньше нуля. Противоречие завершает доказательство. \square

Следствие. Пусть у нас есть две задачи:

$$L_h y_{ij} = \phi_{ij}, \quad x_{ij} \in \omega_h$$

$$L_h Y_{ij} = \Phi_{ij} \quad x_{ij} \in \omega_h$$

Пусть на границе выполняются условия:

$$y_{ij} \leq Y_{ij}, \quad x_{ij} \in \Gamma_h$$

$$|\phi_{ij}| \leq \Phi_{ij}, \quad x_{ij} \in \omega_h$$

Тогда всюду выполнено:

$$|y_{ij}| \leq Y_{ij}, \quad x_{ij} \in \overline{\omega_h}$$

Доказательство. В силу линейности задачи для V :

$$L_h V_{ij} = \Phi_{ij} + \phi_{ij}$$

$$L_h \omega_{ij} = \Phi_{ij} - \phi_{ij}$$

Правые части обоих уравнений не меньше нуля в силу вышеуказанных условий. А это, в силу доказанного утверждения, означает выполнение условия, которое требовалось доказать:

$$|y_{ij}| \leq Y_{ij}, \quad x_{ij} \in \overline{\omega_h}$$

□

Перепишем задачу для погрешности аппроксимации в виде:

$$L_h Z_{ij} = \psi_{ij} \quad x_{ij} \in \omega_h \quad (2)$$

$$Z_{ij} = 0, \quad x_{ij} \in \Gamma_h$$

Для доказательства сходимости разностной схемы необходимо подобрать мажоранту Y так, чтобы выполнялось условие:

$$L_h Y_{ij} = K_1, K_1 = const > 0$$

Y будем искать в виде:

$$Y_{ij} = (l_1^2 + l_2^2 - (x_1^{(i)})^2 - (x_2^{(j)})^2)K, \quad \text{где } K > 0$$

$$Y_{ij} L_h \geq 0, \quad x_{ij} \in \overline{\omega_h}$$

$$Y_{ij} L_h = 4K$$

Положим $4K = \|\psi\|_C$:

$$\left. \begin{array}{l} 0 = |z_{ij}|_{\Gamma_h} \leq Y_{ij}|_{\Gamma_h}, \\ 4K = \|\psi\|_C \geq |\psi_{ij}|, \quad x_{ij} \in \omega_h \end{array} \right\} (|z_{ij}| \leq Y_{ij}, \quad z_{ij} \in \overline{\omega_h})$$

$$\|z\|_C \leq Y_{ij} \leq (l_1^2 + l_2^2)K = \frac{l_1^2 + l_2^2}{4} \|\psi\|_C \Rightarrow \|z\|_C \leq M \|\psi\|_C$$

$$\psi = O(h_1^2 + h_2^2) \Rightarrow \|\psi\|_C \leq M(l_1^2 + l_2^2) \Rightarrow \|z\|_C \leq M^2(h_1^2 + h_2^2)$$

Тем самым, мы доказали следующую теорему:

Теорема 1. Пусть $U(x_1, x_2) \in C^4(\bar{D})$. Тогда разностная схема (3) - (4) сходится и имеет место оценка:

$$\|y_{ij} - U_{ij}\|_C \leq M_1(h_1^2 + h_2^2)$$

Где M_1 не зависит от h_1 и h_2 .

Доказательство. Из полученной оценки:

$$\|\psi\|_C \leq M_2(h_1^2 + h_2^2) \tag{3}$$

$$\|y_{ij} - U_{ij}\|_C \leq M\left(\frac{l_1^2 + l_2^2}{4}\right)(h_1^2 + h_2^2)$$

□

§4 Методы решения разностной задачи Дирихле

$$\frac{d^2u}{dx_1^2} + \frac{d^2u}{dx_2^2} = f(x_1, x_2) \in D \tag{1}$$

$$U|_{\Gamma_u} = \mu(x_1, x_2) \tag{2}$$

Разрешим систему относительно центрального узла:

$$\left(\frac{2}{h_1^2} + \frac{2}{h_2^2}\right)y_{ij} = \frac{y_{i+1,j} + y_{i-1,j}}{h_1^2} + \frac{y_{i,j+1} + y_{i,j-1}}{h_2^2} - f_{ij} \tag{3}$$

Будем обозначать итерацию под номером $s - y_{ij}^{(s)}$.

Простая итерация (метод Якоби)

Получаем следующий итерационный процесс:

$$\begin{aligned} \left(\frac{2}{h_1^2} + \frac{2}{h_2^2}\right)y_{ij}^{(s+1)} &= \frac{y_{i+1,j}^{(s)} + y_{i-1,j}^{(s)}}{h_1^2} + \frac{y_{i,j+1}^{(s)} + y_{i,j-1}^{(s)}}{h_2^2} - f_{ij} \\ s &= 0, 1, \dots \\ y_{ij}^{(s)} &= \mu_{ij} \\ y_{ij}^{(0)} &- \text{ задано} \end{aligned}$$

Для достижения заданной точности требуется порядка $n_0(\epsilon) \sim (h^{-2}) \sim (N^2)$, где $N = \max(N_1, N_2)$.

Метод Зейделя

$$\begin{aligned} \left(\frac{2}{h_1^2} + \frac{2}{h_2^2}\right)y_{ij}^{(s+1)} &= \frac{y_{i-1,j}^{(s+1)} + y_{i+1,j}^{(s)}}{h_1^2} + \frac{y_{i,j-1}^{(s+1)} + y_{i,j+1}^{(s)}}{h_2^2} - f_{ij} \\ y_{ij}^{(s+1)} &= \mu_{ij}, s = 0, 1, \dots \\ \text{при } s = 0, y_{ij}^0 &- \text{ задано} \end{aligned}$$

Покажем, как находить решение: Начнем с узла (1, 1), далее движемся вверх до (1, n), потом из точки (2, 1) движемся вверх и т.д. Эдак будет рисунок метода. С точки зрения организации алгоритма - незначительное усложнение. С точки зрения сходимости метод аналогичен методу Якоби: для получения требуемой точности требуется порядка $n_0(\epsilon) \sim O(N^2)$

Попеременно-треугольный итерационный метод (метод Самарского)

Перепишем нашу систему в виде СЛАУ:

$$Ay = \phi, \text{ где } A^* = A > 0, A = R_1 + R_2$$

$$R_1 = \begin{pmatrix} 0,5a_{11} & 0 & \dots & 0 \\ \dots & 0,5a_{22} & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & 0,5a_{mm} \end{pmatrix}$$

$$R_2 = \begin{pmatrix} 0,5a_{11} & \dots & \dots & \dots \\ 0 & 0,5a_{22} & a_{ij} & \dots \\ \dots & \dots & \dots & \dots \\ 0 & 0 & 0,5a_{mm} \end{pmatrix}$$

$$(E + \omega R_1)(E + \omega R_2) \frac{y^{(s+1)} + y^{(s)}}{\tau} + Ay^{(s)} = \phi$$

Где $\omega > 0, \tau > 0$ - итерационные параметры, y_0 задано. При правильной организации процесса $w^{(s+1)}$ вычисляется по явным формулам (так как $E + \omega R_1$ - нижняя треугольная матрица):

$$(E + \omega R_1)w^{(s+1)} = \phi - Ay^{(s)}(E + \omega R_2)v^{(s+1)} = w^{(s+1)}y^{(s+1)} = y^{(s)} + \tau v^{(s+1)}$$

При $\omega > \frac{\tau}{4}$ сходится для любых $y^{(0)}$.

При этом $n_0(\epsilon) \sim O(N)$.

§5 Основные понятия теории разностных схем. Аппроксимация. Устойчивость. Сходимость.

Рассмотрим произвольную линейную дифференциальную задачу:

$$Lu(x) = f(x), x \in G \tag{1}$$

Считаем, что краевые и начальные условия будут учитываться либо видом оператора L , либо видом правой части. Принципиально, что L - линейный оператор. Введем на множестве G сетку G_h , где h - некоторая норма шагов сетки. Тогда x из непрерывного превращается в дискретное: $x \in G_h$. Тем же образом ставим в соответствие функции $y(x)$ ее разностный аналог $y_h(x)$. Аналогично поступаем с оператором $L : L_h y_h = \phi(x), x \in G_h$. Рассмотрим линейное нормированное пространство непрерывных функций B_0 с нормой $\|u\|_0$ и $u(x) \in B_0$. Соответственно B_h - дискретное нормированное пространство с нормой $\|u\|_h$ и $u_h(x) \in B_h$.

Определение. Нормы B_0 и B_h согласованы, если

$$\lim_{h \rightarrow 0} \|u_h\|_h = \|u\|_0$$

Если нормы несогласованы, то решение разностной схемы может сходиться к решению, которое не является решением исходной задачи. Введем оператор проектирования $P_h : B_0 \rightarrow B_h$. Таким образом $\forall u \in B_0 : P_h(u) = u_h \in B_h$.

Например:

$$G = \{x : 0 \leq x \leq 1\}.$$

$$G_h = x_i : x_i = hi, i = \overline{0, N}, hN = 1, h = \frac{1}{N} > 0;$$

$$P_h(u|_{x_i}) : u_h(x_i) = u(x_i);$$

$$B_h = \{y = (y_0, y_1, \dots, y_N)\};$$

Рассмотрим примеры норм:

$$1. \|u\|_C = \max_{x \in G} |u_x| = \|u\|_0$$

Согласованная с ней норма в B_h :

$$\|y\|_C = \max_{0 \leq i \leq N} |y_i| = \|y\|_h;$$

$$2. \|u\|_0 = \|u\|_{L_2} = \left(\int_0^1 u^2(x) dx\right)^{1/2},$$

Согласованная с ней норма в B_h :

$$\|y\|_h = \|u\|_{L_2} = \left(\sum_{i=0}^N y_i^2 h\right)^{1/2}$$

3. Покажем, что норма $\left(\sum_{i=0}^N y_i^2\right)^{1/2}$ не согласована ни с одной из норм в B_0 .
От противного: пусть $u(x) \equiv 1$, тогда:

$$\|u_h\|_h = \left(\sum_{i=0}^N 1\right)^{1/2} = \sqrt{N+1}$$

Тогда, $\|u_h\|_h \xrightarrow{h \rightarrow 0} \infty$, чего быть не может.

Определение. Сеточная функция $z_h(x)$ называется погрешностью разностной схемы:

$$z_h(x) = y_h(x) - u_h(x), x \in G_h$$

Определение. Сеточная функция $\psi_h(x)$ называется погрешностью аппроксимации разностной схемы на решении исходной задачи:

$$\psi_h(x) = \phi_h(x) - L_h u_h(x), x \in G_h$$

Определение. Разностная схема аппроксимирует задачу (1), если:

$$\|\psi_h\|_h \rightarrow 0, h \rightarrow 0.$$

Определение. Разностная схема имеет порядок аппроксимации k , если $\exists M_1 > 0, k > 0$, которые не зависят от h и имеет место оценка:

$$\|\psi_h\|_h \leq M_1 h^k.$$

Определение. Дифференциальная задача называется поставленной корректно, если:

1. решение задачи существует и единственно,
2. решение задачи непрерывно зависит от $f(x)$.

Определение. Разностная схема называется корректной, если при всех достаточно малых h : 1. $\forall \phi(x)$ решение \exists , 2. $\forall M_2 = \text{const} > 0, M_2$ не зависящая от h , что:

$$\|u_h\|_h \leq M_2 \|\psi_h\|_h \quad (2)$$

Оценка (2) называется априорной оценкой и означает устойчивость разностной схемы.

Замечание. Слева и справа не обязательно одинаковые нормы.

Определение. Говорят, что разностная схема сходится к решению исходной задачи (1), если:

$$\|z_h\|_h = \|y_h - u_h\|_h \rightarrow 0, h \rightarrow 0$$

Определение. Говорят, что разностная схема имеет порядок точности k , если $\exists M_3 = \text{const} > 0$ и не зависящая от h , что:

$$\|z_h\|_h \leq M_3 h^k$$

Теорема (Теорема Филлипова). Пусть дифференциальная задача корректно поставлена и соответствующая ей разностная схема также корректна. Тогда решение разностной задачи сходится к решению дифференциальной задачи с порядком погрешности аппроксимации.

Доказательство.

$$\|y_h\|_h \leq M_2 \|\phi_h\|_h$$

$$\|z_h\|_h \leq M_2 \|\psi_h\|_h$$

M_2 не зависит от h

Далее:

$$\|\psi_h\|_h \leq M_1 h^k, \quad M_1 \text{ не зависит от } h$$

Получаем:

$$\|z_h\|_h \leq M_1 M_2 h^k = M_3 h^k, \quad M_3 \text{ не зависит от } h$$

$$\|z_h\|_h = \|y_h - u_h\|_h \rightarrow 0 \text{ при } h \rightarrow 0$$

□

Замечание. Пусть $\exists v : y_h \rightarrow v : \|y_h - v_h\|_h \rightarrow 0$ при $h \rightarrow 0$. Тогда:

$$\|u_h - v_h\|_h \leq \| -y_h + u_h \|_h + \|y_h - v_h\|_h \rightarrow 0 \text{ при } h \rightarrow 0$$

Если норма согласованная, то:

$$\lim_{h \rightarrow 0} \|u_h - v_h\|_h = \|u - v\|_0 = 0 \Rightarrow u \equiv v$$

Глава V

Методы решения ОДУ и систем ОДУ

§1 Примеры численных методов решения задачи Коши

$$\begin{cases} \frac{du}{dt} = f(t, u(t)), & t > 0, \\ u(0) = u_0; \end{cases} \quad (1)$$

$$u(t) = (u_1(t), u_2(t), \dots, u_m(t))^T$$

$$f(t, u(t)) = (f_1(t, u(t)), f_2(t, u(t)), \dots, f_m(t, u(t)))^T$$

Рассмотрим параллелепипед $R = \{|t| \leq a, |u - u_0| \leq b\}$

Определение. Функция $f(t, u)$ удовлетворяет в R условию Липшица по второму аргументу, если:

$$|f(t, u) - f(t, v)| \leq L|u - v|, \quad L = \text{const}$$

Пусть $f(t, u)$ из (1) удовлетворяет условию Липшица в R . Тогда решение (1) $u(t)$ существует и единственно при $0 < t < T$ ("в малом"). Проинтегрируем первое уравнение из (1) и учтем начальное условие:

$$u(t) = u(0) + \int_0^t f(x, u(x)) dx$$

На этом представлении основан метод Пикара:

$$u_{n+1}(t) = u(0) + \int_0^t f(x, u_n(x)) dx, \quad n = 0, 1, \dots$$

Этот метод не может быть эффективным методом решения задачи (1), так как интеграл не всегда можно посчитать аналитически, да и сходимость была бы медленной. Поэтому для решения систем ОДУ применяются разностные методы: первая группа методов - методы Рунге-Кутты, вторая - многошаговые разностные методы (например, метод Адамса). Введем последовательность ω_τ :

$$\omega_\tau = \{t_n = n\tau, \tau > 0, n = 0, 1, \dots\}$$

Пример. Явная схема Эйлера.

Введем обозначения: $y_n = y(t_n)$, $f(t_n, y(t_n)) = f_n$. Тогда явная схема Эйлера имеет вид:

$$\begin{cases} \frac{y_{n+1} - y_n}{\tau} = f_n, & t_n \in \omega_\tau, \\ y(0) = u_0; \end{cases} \quad (2)$$

Выразим y_{n+1} из первого уравнения:

$$y_{n+1} = y_n + \tau f_n$$

Все компоненты в правой части известны, то есть y_{n+1} можно найти в явном виде. Обозначим $u(t_n)$ через u_n . Введем погрешность $z_n = y_n - u_n$.

$$|z_n| \leq M\tau, \quad M \text{ не зависит от } \tau$$

Таким образом, имеем первый порядок точности по τ . Запишем погрешность аппроксимации ψ_n на решении исходной задачи:

$$\psi_n = -\frac{u_{n+1} - u_n}{\tau} + f(t_n, u_n) \quad (3)$$

Разложим u_{n+1} в ряд Тейлора в точке t_n . Тогда:

$$\frac{u_{n+1} - u_n}{\tau} = u'_n + \frac{\tau}{2} u''_n + \underline{O}(\tau^2)$$

Подставим последнее выражение в (3):

$$\psi_n = -u'_n + f(t_n, u_n) - \frac{\tau}{2} u''_n + \underline{O}(\tau^2)$$

Учитывая, что $-u'_n + f(t_n, u_n) = 0$, окончательно получаем:

$$\psi_n = \underline{O}(\tau)$$

Пример. Схема «предиктор-корректор» (схема Рунге-Кутты).

Обозначим $t_n + 0.5\tau$ через $t_{n+\frac{1}{2}}$.

$$\begin{cases} \frac{y_{n+\frac{1}{2}} - y_n}{0.5\tau} = f(t_n, y_n) - \text{«предиктор»}, \\ \frac{y_{n+1} - y_n}{0.5\tau} = f(t_{n+\frac{1}{2}}, y_{n+\frac{1}{2}}) - \text{«корректор»}, \\ y(0) = u_0; \end{cases} \quad (4)$$

$$y_{n+1} = y_n + \tau f(t_{n+\frac{1}{2}}, y_{n+\frac{1}{2}})$$

Для данной схемы имеем: $\psi_n = \underline{O}(\tau^2)$

Рассмотрим общий вид двухэтапного метода Рунге-Кутты:

$$\begin{cases} \frac{y_{n+1} - y_n}{\tau} = \sigma_1 K_1 + \sigma_2 K_2, \\ K_1 = f(t_n, y_n), \\ K_2 = f(t_n + a_2\tau, y_n + b_{21}\tau K_1); \end{cases} \quad (5)$$

Запишем погрешность аппроксимации (5) на решении (1):

$$\psi_n = -\frac{u_{n+1} - u_n}{\tau} + \sigma_1 f(t_n, u_n) + \sigma_2 f(t_n + a_2\tau, u_n + b_{21}\tau f(t_n, u_n)) \quad (6)$$

Разложим u_{n+1} в ряд Тейлора в точке t_n . Тогда:

$$\frac{u_{n+1} - u_n}{\tau} = u'_n + \frac{\tau}{2}u''_n + \underline{O}(\tau^2)$$

Далее разложим $f(t_n + a_2\tau, y_n + b_{21}\tau f(t_n, u_n))$ в окрестности точки (t_n, u_n) :

$$f(t_n + a_2\tau, y_n + b_{21}\tau f(t_n, u_n)) = f(t_n, u_n) + \frac{\partial f_n}{\partial t}a_2\tau + \frac{\partial f_n}{\partial u}b_{21}\tau f(t_n, u_n) + \underline{O}(\tau^2)$$

Далее:

$$\frac{d^2u_n}{dt^2} = \frac{d}{dt}(f(t, u_n(t))) = \frac{\partial f_n}{\partial t} + \frac{\partial f_n}{\partial u}f_n$$

Перепишем теперь ψ_n с учетом проведенных преобразований:

$$\begin{aligned} \psi_n &= - \left(u'_n + 0.5\tau \left(\frac{\partial f_n}{\partial t} + \frac{\partial f_n}{\partial u}f_n \right) \right) + \sigma_1 f(t_n, u_n) + \sigma_2 f(t_n, u_n) + \\ &\quad + \sigma_2 \frac{\partial f_n}{\partial t} a_2\tau + \sigma_2 \frac{\partial f_n}{\partial u} b_{21}\tau f(t_n, u_n) + \underline{O}(\tau^2) = \\ &\quad = -u'_n + (\sigma_1 + \sigma_2)f(t_n, u_n) + \\ &\quad + \tau \left((\sigma_2 a_2 - 0.5) \frac{\partial f_n}{\partial t} + ((\sigma_2 b_{21} - 0.5)) \frac{\partial f_n}{\partial u} b_{21}\tau f(t_n, u_n) \right) + \underline{O}(\tau^2) \end{aligned}$$

Потребуем, чтобы были выполнены следующие условия:

1. $\sigma_1 + \sigma_2 = 1$ (условие аппроксимации)
2. $\sigma_2 a_2 = \sigma_2 b_{21} = 0.5$ (для того, чтобы достичь второго порядка аппроксимации)

Если выполнено только условие 1, то $\psi_n = \underline{O}(\tau)$, а если выполнены оба условия, то $\psi_n = \underline{O}(\tau^2)$. Положим $\sigma_2 = \sigma$, а $\sigma_1 = 1 - \sigma$, тогда условие 1 автоматически выполнено. В последнем примере параметры имели следующие значения: $a_2 = b_{21} = 0.5, \sigma = 1$. Если взять $\sigma = 0.5, b_{21} = a_2 = 1$, то получим симметричную схему:

$$\frac{y_{n+1} - y_n}{\tau} = 0.5(f(t_n, y_n) + f(t_{n+1}, y_{n+1}))$$

Общий m -этапный метод Рунге-Кутты

Рассмотрим общий m -этапный метод Рунге-Кутты:

$$\frac{y_{n+1} - y_n}{\tau} = \sigma_1 K_1 + \sigma_2 K_2 + \dots + \sigma_m K_m$$

$$\sum_{i=1}^m \sigma_i = 1 - \text{условие аппроксимации}$$

$$K_1 = f(t_n, y_n)$$

$$K_2 = f(t_n + a_2\tau, y_n + b_{21}\tau K_1)$$

$$K_3 = f(t_n + a_3\tau, y_n + b_{31}\tau K_1 + b_{32}\tau K_2)$$

...

$$K_m = f(t_n + a_m\tau, y_n + b_{m1}\tau K_1 + b_{m2}\tau K_2 + \dots + b_{mm-1}\tau K_{m-1})$$

На практике редко используются методы Рунге-Кутты для $m > 4$. Приведем примеры разностных методов Рунге-Кутты, имеющих третий и четвертый порядок погрешности аппроксимации.

Пример. Схема Рунге-Кутты четвертого порядка.

$$\begin{aligned}\frac{y_{n+1} - y_n}{\tau} &= \frac{1}{6} (K_1 + 2K_2 + 2K_3 + K_4) \\ K_1 &= f(t_n, y_n) \\ K_2 &= f(t_n + 0.5\tau, y_n + 0.5\tau K_1) \\ K_3 &= f(t_n + 0.5\tau, y_n + 0.5\tau K_2) \\ K_4 &= f(t_n + \tau, y_n + \tau K_3)\end{aligned}$$

Данная схема имеет четвертый порядок аппроксимации по τ : $\psi_n = \underline{O}(\tau^4)$.

Пример. Схема Рунге-Кутты третьего порядка.

$$\begin{aligned}\frac{y_{n+1} - y_n}{\tau} &= \frac{1}{6} (K_1 + 4K_2 + K_3) \\ K_1 &= f(t_n, y_n) \\ K_2 &= f(t_n + 0.5\tau, y_n + 0.5\tau K_1) \\ K_3 &= f(t_n + \tau, y_n - \tau K_1 - 2\tau K_2)\end{aligned}$$

Данная схема имеет третий порядок аппроксимации по τ : $\psi_n = \underline{O}(\tau^3)$.

§2 Оценка точности на примере 2-х этапного метода Рунге-Кутты

$$\begin{cases} \frac{du}{dt} = f(t, u(t)), & t > 0 \\ u(0) = u_0 \end{cases} \quad (1)$$

$$\begin{aligned}\frac{y_{n+1} - y_n}{\tau} &= (1 - \sigma)f(t_n, y_n) + \sigma f(t_n + a\tau, y_n + a\tau f(t_n, y_n)) \\ y_0 &= u_0 \\ t_n &\in \omega_\tau\end{aligned}$$

σ - параметр, в качестве которого можно выбирать любое число, лишь бы выполнялось условие второй погрешности аппроксимации. Обычно выбирают $\sigma \in [0, 1]$.

a - некоторая константа. Будем рассматривать $a \geq 0$, но, вообще говоря, это необязательно. Введем функцию погрешности z_n :

$$z_n = y_n - u(t_n) = y_n - u_n \Rightarrow \quad (2)$$

$$\frac{z_{n+1} - z_n}{\tau} = -\frac{u_{n+1} - u_n}{\tau} + (1 - \sigma)f(t_n, y_n) + \sigma f(t_n + a\tau, y_n + a\tau f(t_n, y_n)) \quad (3)$$

Для сходимости нужно показать, что:

$$|z_n| \rightarrow 0, \quad n \rightarrow \infty$$

Покажем, что $|z_n| \leq M\tau^2$, где M не зависит от τ

$$\begin{aligned} \frac{z_{n+1} - z_n}{\tau} &= -\frac{u_{n+1} - u_n}{\tau} + (1 - \sigma)f(t_n, u_n) + \\ &\sigma f(t_n + a\tau, u_n + a\tau f(t_n, u_n)) - (1 - \sigma)f(t_n, u_n) + \\ &(1 - \sigma)f(t_n, y_n) - \sigma f(t_n + a\tau, u_n + a\tau f(t_n, u_n)) + \\ &\sigma f(t_n + a\tau, y_n + a\tau f(t_n, y_n)) = \psi_n + \phi_n^{(1)} + \phi_n^{(2)} \end{aligned}$$

где $\psi_n, \phi_n^{(1)}, \phi_n^{(2)}$ обозначены слагаемые:

$$\begin{aligned} \psi_n &= -\frac{u_{n+1} - u_n}{\tau} + (1 - \sigma)f(t_n, u_n) + \sigma f(t_n + a\tau, u_n + a\tau f(t_n, u_n)), \\ \phi_n^{(1)} &= (1 - \sigma)(f(t_n, y_n) - f(t_n, u_n)), \end{aligned} \quad (4)$$

$$\phi_n^{(2)} = \sigma \left[f(t_n + a\tau, y_n + a\tau f(t_n, y_n)) - f(t_n + a\tau, u_n + a\tau f(t_n, u_n)) \right].$$

Введем допущение: функция f по второму аргументу удовлетворяет условию Липшица с константой L . Оценим, исходя из этого допущения, $\phi_n^{(1)}$ и $\phi_n^{(2)}$:

$$\begin{aligned} |\phi_n^{(1)}| &\leq (1 - \sigma)|f(t_n, y_n) - f(t_n, u_n)| \leq (1 - \sigma)L|y_n - u_n| = (1 - \sigma)L|z_n|, \\ |\phi_n^{(2)}| &\leq \sigma L|y_n + a\tau f(t_n, y_n) - u_n + a\tau f(t_n, u_n)| \leq \\ &\leq \sigma L(|y_n - u_n| + \underbrace{a}_{\geq 0} \tau L|y_n - u_n|) = \sigma L(1 + a\tau L) \underbrace{|y_n - u_n|}_{z_n} \end{aligned}$$

Из (3) \Rightarrow

$$\begin{aligned} z_{n+1} &= z_n + \tau\psi_n + \tau\phi_n^{(1)} + \tau\phi_n^{(2)} \\ |z_{n+1}| &\leq |z_n| + \tau|\psi_n| + \tau[(1 - \sigma)L|z_n| + \sigma L|z_n| + \sigma a\tau L^2|z_n|] = \\ &\tau|\psi_n| + (1 + \tau L + \tau^2 a\sigma L^2)|z_n| \end{aligned}$$

Рассмотрим $\sigma a \leq 0,5$, заметив, что $1 + \tau L + 0,5\tau^2 L^2$ являются первыми членами разложения по Тейлору функции $e^{\tau L}$:

$$|z_{n+1}| \leq \tau|\psi_n| + (1 + \tau L + 0,5\tau^2 L^2)|z_n| \leq e^{\tau L}|z_n| + \tau|\psi_n| \quad (5)$$

Обозначим $e^{\tau L} = \rho$. Получим оценку:

$$|z_{n+1}| \leq \rho|z_n| + \tau|\psi_n| \quad (6)$$

Соотношение (6) можно рассмотреть как рекуррентную формулу. Легко видеть, что:

$$\begin{aligned} z_{n+1} &\leq \rho^{n+1}|z_0| + \sum_{j=0}^n \rho^{n-j}\tau|\psi_j| \\ |z_{n+1}| &\leq \max_{0 \leq j \leq n} |\psi_j| \sum_{j=0}^n \rho^{n-j}\tau \leq t_{n+1} e^{L t_{n+1}} \max_{0 \leq j \leq n} |\psi_j| \end{aligned}$$

Окончательно, получаем:

$$|z_{n+1}| \leq M \max_{0 \leq j \leq n} |\psi_j|, \quad M \text{ не зависит от } \tau \quad (7)$$

Видно, что точность будет совпадать с порядком погрешности аппроксимации, а именно:

1. $\sigma a = 0, 5 \Rightarrow \psi = O(\tau^2) \Rightarrow |z_n| = O(\tau^2)$, т.е. имеем второй порядок погрешности.
2. $\sigma = 0, \forall a \Rightarrow \psi = O(\tau), \quad |z_n| \leq M_1 \tau, \quad M_1$ не зависит от τ , получаем первый порядок точности.

§3 Многошаговые разностные методы

$$\begin{cases} \frac{du}{dt} = f(t, u(t)), & t > 0 \\ u(0) = u_0 \end{cases} \quad (1)$$

Введем сетку $\omega_\tau = \{t_n = n\tau, \tau > 0, n = 0, 1, \dots\}$.
 Обозначим $y_k = y(t_k), f_k = f(t, y_k)$.

Определение. *Линейным m -шаговым разностным методом решения задачи (1) называется метод, записанный уравнением:*

$$\sum_{k=0}^m \frac{a_k}{\tau} y_{n-k} = \sum_{k=0}^m b_k f_{n-k}, \quad (2)$$

где a_k, b_k - числа, $\tau > 0$. При этом $a_0 \neq 0, b_m \neq 0, n = m, m+1, \dots$

Если $b_0 = 0$, то (2) - явный метод. Если $b_0 \neq 0$, то (2) - неявный метод.

Для начала вычислений по формуле (2) необходимы значения y_0, \dots, y_{m-1} - т.н. "Разгонный этап". Так как формула (2) однородна по a_k и b_k , то полагают $\sum_{k=0}^m b_k = 1$ (условие нормировки).

Неявный m -шаговый разностный метод записывается в виде

$$\frac{a_0}{\tau} y_n - b_0 f(t_n, y_k) = F(y_{n-1}, y_{n-2}, \dots, y_{n-m}) \quad (3)$$

$$F = \sum_{k=1}^m b_k f_{n-k} - \sum_{k=1}^m \frac{a_k}{\tau} y_{n-k}$$

Уравнение (3) решается чаще всего методом Ньютона, причем в качестве $y_n^{(0)}$ берется y_{n-1} .
 В явном разностном методе значения y_n находятся по явной формуле

$$y_n = \frac{\tau}{a_0} \left(\sum_{k=1}^m b_k f_{n-k} - \sum_{k=1}^m \frac{a_k}{\tau} y_{n-k} \right)$$

Оценим погрешность аппроксимации на решении

$$\psi_n = - \sum_{k=0}^m \frac{a_k}{\tau} u_{n-k} + \sum_{k=0}^m b_k f(t_{n-k}, u_{n-k}) \quad (4)$$

$$\begin{aligned}
 u_n &= u(t_n - k\tau) = \sum_{l=0}^p \frac{(-k\tau)^l}{l!} u^{(l)}(t_n) + O(\tau^{p+1}) \\
 f(t_{n-k}, u_{n-k}) &= u'_{n-k} = \sum_{l=0}^{p-1} \frac{(-k\tau)^l}{l!} u^{(l+1)}(t_n) + O(\tau^p) \\
 \psi_n &= - \sum_{k=0}^m \frac{a_k}{\tau} \sum_{l=0}^p \frac{(-k\tau)^l}{l!} u^{(l)}(t_n) + \sum_{k=0}^m b_k \sum_{l=0}^{p-1} u^{(l+1)}(t_n) = O(\tau^p) = \\
 &= \left\{ \text{сдвиг индексов} \right\} = - \sum_{l=0}^p \sum_{k=0}^m \frac{a_k}{\tau} \frac{(-k\tau)^l}{l!} u_n^{(l)} + \\
 &\quad \sum_{l=1}^p \sum_{k=0}^m b_k \frac{(-k\tau)^{l-1}}{(l-1)!} u_n^{(l)} + O(\tau^p) = \\
 &= - \sum_{k=0}^m \frac{a_k}{\tau} u_n + \sum_{k=0}^m \left[\sum_{l=1}^p (-k\tau)^{l-1} (a_k k + l b_k) \frac{u_n^{(l)}}{l(l-1)!} \right] + O(\tau^p)
 \end{aligned}$$

Условие аппроксимации:

$$\sum_{k=0}^m a_k = 0$$

Для достижения аппроксимации порядка p должно быть выполнено соотношение:

$$\sum_{k=0}^m k^{l-1} (a_k k + l b_k) = 0, \quad l = 1, 2, \dots, p$$

В многошаговом методе $2m+2$ неизвестных - $a_0, a_1, \dots, a_m, b_0, \dots, b_m$, и $p+2$ уравнений. Чтобы система не была переопределенной, должно выполняться $p \leq 2m \Rightarrow$ наивысший порядок аппроксимации равен $2m$.

Таким образом, для достижения порядка погрешности аппроксимации p должны выполняться следующие соотношения:

$$\begin{aligned}
 a_0 &= - \sum_{k=1}^m a_k \\
 b_0 &= 1 - \sum_{k=1}^m b_k
 \end{aligned}$$

$$\sum_{k=0}^m k^{l-1} (a_k k + l b_k) = 0, \quad l = 1, 2, \dots, p$$

§4 Понятие устойчивости разностных методов

Рассмотрим задачу Коши:

$$\begin{cases} \frac{du}{dt} = f(t, u(t)), & t > 0, \\ u(0) = u_0. \end{cases} \quad (1)$$

Рассмотрим для примера такую схему:

$$\begin{aligned} y_n &= qy_{n-1}, & q \in \mathbb{C}, & \quad q = \text{const}, \\ n &= 0, 1, \dots; & y_0 & \text{ задан.} \end{aligned}$$

Придадим y_n возмущение δ_n :

$$\tilde{y}_n = y_n + \delta_n.$$

Тогда $\tilde{y}_{n+1} = q\tilde{y}_n = qy_n + q\delta_n = y_{n+1} + \delta_{n+1}$, где $\delta_{n+1} = q\delta_n$.

Если $|q| > 1$, то δ_n нарастает, следовательно, об устойчивости говорить нельзя.

Рассмотрим модельную задачу:

$$\begin{cases} \frac{du}{dt} + \lambda u(t) = 0, & t > 0, \\ u(0) = u_0. \end{cases} \quad (2)$$

Ее решение имеет вид $u(t) = u_0 e^{-\lambda t}$. Если $\lambda > 0$, то $|u(t)| \leq |u_0|$, т.е. имеет место устойчивость по начальному условию.

Устойчивость — внутреннее свойство разностной схемы. Разностная схема не обязательно сохраняет устойчивость исходной задачи.

Рассмотрим явную схему Эйлера:

$$\begin{cases} \frac{y_{n+1} - y_n}{\tau} = f(t_n, y_n), \\ y_0 = u_0. \end{cases} \quad (3)$$

Запишем ее для модельной задачи:

$$\frac{y_{n+1} - y_n}{\tau} + \lambda y_n = 0.$$

Выразим y_{n+1} :

$$y_{n+1} = y_n - \tau \lambda y_n = (1 - \tau \lambda) y_n.$$

Обозначим $q = 1 - \tau \lambda$. Тогда $y_{n+1} = qy_n$. Таким образом, для устойчивости необходимо, чтобы выполнялось $|q| \leq 1$, т.е.

$$\begin{aligned} 1 - \tau \lambda &\geq -1, \\ 0 < \tau \lambda &\leq 2. \end{aligned}$$

Таким образом, для того, чтобы явная схема Эйлера была устойчивой (для модельной задачи), необходимо выполнение условия

$$0 < \tau \leq \frac{2}{\lambda}. \quad (4)$$

Это означает, что явная схема Эйлера является условно устойчивой (для модельной задачи).

Рассмотрим неявную схему Эйлера:

$$\frac{y_{n+1} - y_n}{\tau} = f(t_{n+1}, y_{n+1}).$$

Перепишем ее:

$$y_{n+1} + \tau f(t_{n+1}, y_{n+1}) = y_n. \quad (5)$$

Для решения нелинейного уравнения (5) обычно применяется метод Ньютона, в качестве начального приближения для нахождения y_{n+1} используют y_n .

Перепишем (5) для модельной задачи:

$$\begin{aligned} y_{n+1} + \tau \lambda y_{n+1} &= y_n, \\ (1 + \tau \lambda) y_n &= t_{n+1}, \\ y_{n+1} &= q y_n, \quad q = \frac{1}{1 + \tau \lambda}. \end{aligned}$$

Заметим, что $|q| < 1$ при $\tau > 0, \lambda > 0$. Это значит, что неявная схема Эйлера является абсолютно устойчивой (для модельной задачи).

Таким образом, для устойчивой дифференциальной задачи существуют как устойчивые, так и неустойчивые схемы.

Рассмотрим произвольный m -шаговый разностный метод:

$$\sum_{k=0}^m \frac{a_k}{\tau} y_{n-k} = \sum_{k=0}^m b_k f_{n-k}, \quad y_0, \dots, y_{m-1} \text{ заданы.} \quad (6)$$

Запишем его для модельной задачи:

$$\begin{aligned} \sum_{k=0}^m \frac{a_k}{\tau} y_{n-k} + \lambda \sum_{k=0}^m b_k y_{n-k} &= 0, \\ \sum_{k=0}^m (a_k + \tau \lambda b_k) y_{n-k} &= 0. \end{aligned} \quad (7)$$

Будем искать решение этого уравнения в виде $y_j = q^j$.

Подставим это (7):

$$\sum_{k=0}^m (a_k + \tau \lambda b_k) q^{n-k} = 0.$$

Разделим обе части этого уравнения на q^{m-n} , получим

$$F_m(\tau, q) = \sum_{k=0}^m (a_k + \tau \lambda b_k) q^{m-k} = 0. \quad (8)$$

Уравнение (8) называется характеристическим уравнением. Для устойчивости необходимо, чтобы его корни по модулю не превосходили 1 (иначе решение будет неограниченно нарастать). Однако, нахождение корней уравнения (8) – трудная задача, и обычно рассматривают более простое уравнение:

$$F_m(0, q) = \sum_{k=0}^m a_k q^{m-k} = 0. \quad (9)$$

Уравнение (9), также как и уравнение (8), называют характеристическим уравнением.

Определение. Говорят, что разностная схема (6) удовлетворяет условию (α) , если все корни характеристического уравнения (9) лежат внутри или на границе единичного круга комплексной плоскости, причем на границе нет кратных корней.

Теорема. Пусть разностная схема (6) удовлетворяет условию (α) и $|f'_n| \leq L$ при $0 \leq t_n = \tau n \leq T$. Тогда для любого достаточно малого τ справедливо

$$|y_n - u_n| \leq M \left(\sum_{j=m}^n \tau |\psi_j| + \max_{0 \leq i \leq m-1} |y_i - u(t_i)| \right), \quad (10)$$

где M не зависит от τ , ψ_j – погрешность аппроксимации разностного метода (6) на решение задачи (1).

Замечание. Метод Адамса удовлетворяет условию (α) :

$$a_0 = -a_1 = 1,$$

$$\frac{y_n - y_{n-1}}{\tau} = \sum_{k=0}^m b_k f_{n-k}.$$

Характеристическое уравнение имеет вид:

$$q^n - q^{n-1} = 0,$$

оно имеет корни $q = 0$ и $q = 1$, причем $q = 1$ – не кратный корень.

Замечание. Для неявных схем наивысший порядок погрешности аппроксимации $p \leq 2m$. Для явных схем $p \leq 2m - 1$.

Однако, схемы высокого порядка не удовлетворяют условию (α) , т.е. не являются устойчивыми. Наивысший порядок погрешности аппроксимации для схем, удовлетворяющих условию (α) , следующий:

1. Для неявных схем:

(a) Если m – четно, то $p \leq m + 2$.

(b) Если m – нечетно, то $p \leq m + 1$.

2. Для явных схем $p \leq m$.

Замечание. Говорить об условной или безусловной устойчивости не имеет смысла. Она всегда условная, т.к. рассматриваются малые τ .

Задача. Доказать, что для схемы

$$\frac{y_n + 4y_{n-1} - 5y_{n-2}}{6\tau} = \frac{2f_{n-1} + f_{n-2}}{3}$$

имеет место $\psi_n = O(\tau^3)$.

Решение.

$$\psi_n = -\frac{u_n + 4u_{n-1} - 5u_{n-2}}{6\tau} + \frac{2f_{n-1} + f_{n-2}}{3}.$$

Запишем условия, налагаемые на многошаговый разностный метод для того, чтобы погрешность аппроксимации имела порядок 3:

$$\begin{cases} b_0 = 1 - \sum_{k=1}^m b_k, \\ a_0 = -\sum_{k=1}^m a_k, \\ \sum_{k=1}^m a_k k = -1, \\ \sum_{k=0}^m k^{l-1}(a_k k + b_k) = 0, \quad l = 2, 3. \end{cases}$$

В нашем случае, $m = 2$, $a_0 = \frac{1}{6}$, $a_1 = \frac{2}{3}$, $a_2 = -\frac{5}{6}$, $b_0 = 0$, $b_1 = \frac{2}{3}$, $b_2 = \frac{1}{3}$. Выписанные условия, как легко проверить, выполняются. Таким образом, $\psi_n = O(\tau^3)$. \square

Рассмотренная в предыдущей задаче схема неустойчива. Действительно, для нее характеристическое уравнение имеет вид:

$$q^2 + 4q - 5 = 0.$$

Это уравнение имеет корни $q_1 = 1, q_2 = -5$. Т.к. $|q_2| > 1$, то данный разностный метод не удовлетворяет условию (α) .

§5 Жесткие системы ОДУ

Рассмотрим систему ОДУ

$$\begin{cases} \frac{du_1}{dt} + a_1 u_1(t) = 0, & t > 0, \\ \frac{du_2}{dt} + a_2 u_2(t) = 0, & t > 0, \\ u_1(0) = u_{10}, u_2(0) = u_{20}, a_1 > 0, a_2 > 0. \end{cases} \quad (1)$$

Решение имеет вид:

$$\begin{aligned} u_1(t) &= u_{10} e^{-a_1 t}, \\ u_2(t) &= u_{20} e^{-a_2 t}. \end{aligned}$$

Пусть $a_1 \gg a_2$. Тогда такая система ОДУ называется жесткой. С некоторого момента t^* решение $u_2(t)$ мало отличается от 0. Однако, если мы решаем эту систему при помощи явной схемы Эйлера, то нам нужно использовать шаг $\tau \leq \min\{\frac{2}{a_1}, \frac{2}{a_2}\} = \frac{2}{a_2}$. Это будет весьма маленький шаг, ибо $a_1 \gg a_2$. Но с некоторого момента t^* u_2 можно “не считать”, т.е. использование маленького шага излишне. Таким образом, явные схемы для жестких систем ОДУ не годятся. Если использовать неявную схему, то можно взять более крупный шаг.

$$\frac{d\bar{u}}{dt} + A\bar{u}(t) = 0, t > 0 \quad (2)$$

$A(m \cdot m)$ с постоянными числами, $\bar{u}^{(t)} = (u_1(t), u_2(t), \dots, u_m(t))^T$.

Определение. Система линейных уравнений называется жесткой, если:

1. $\text{Re}\lambda_k^A > 0$ (устойчивость по Ляпунову),
2. $s = \frac{\max_{1 \leq k \leq m} |\text{Re}\lambda_k^A|}{\min_{1 \leq k \leq m} |\text{Re}\lambda_k^A|} \gg 1$ (s – число жесткости).

Введем понятие жесткости для нелинейной системы:

$$\frac{d\bar{u}}{dt} = f(t, \bar{u}(t)), t > 0, \bar{u}(0) = \bar{u}_0 \quad (3)$$

Пусть $v(t)$ – некоторое решение задачи (3), тогда рассмотрим в окрестности данного решения разность:

$$\begin{aligned} \bar{z}(t) &= \bar{u}(t) - \bar{v}(t) \\ \frac{d\bar{z}_k}{dt} &= f_k(t, \bar{v}(t) + \bar{z}(t)) - f_k(t, \bar{v}(t)), k = \overline{1, m} \end{aligned}$$

Разложим $f_k(t, \bar{v}(t) + \bar{z}(t))$ в окрестности точки $(t, \bar{v}(t))$, удерживая только первую производную:

$$f_k(t, \bar{v}(t) + \bar{z}(t)) = f_k(t, \bar{v}(t)) + \frac{\partial f_k}{\partial u_1}(t, \bar{v}(t))z_1(t) + \dots + \frac{\partial f_k}{\partial u_m}(t, \bar{v}(t))z_m(t) + o(|z|)$$

Обозначим

$$\frac{\partial \bar{z}}{\partial t} = J(t, \bar{v}(t))\bar{z} \quad (4)$$

По определению,

$$J(t, \bar{v}(t))\bar{z} = \left(\frac{\partial f_i(t, \bar{v}(t))}{\partial u_j} \right)_{ij}, i, j = \overline{1, n}.$$

Теперь введем число жесткости s как отношение: $s = \frac{\max \text{Re}\lambda_k^J}{\min \text{Re}\lambda_k^J}$.

Определение. Система (3) называется жесткой, если:

1. $\text{Re}\lambda_k^J < 0$,
2. $s(t) \gg 1$.

§6 Дальнейшее определение устойчивости и примеры разностных схем. Интегрирование жестких схем ДУ

Рассмотрим линейную задачу:

$$\begin{aligned} \frac{d\bar{u}}{dt} &= \Lambda \bar{u}, \quad t > 0 \\ \bar{u}(0) &= \bar{u}_0 \end{aligned} \quad (1)$$

При этом Λ – собственные значения матрицы первого приближения J системы (4).

Определение. Областью устойчивости разностного метода для задачи (3) называется множество всех точек комплексно плоскости $\mu = \tau\Lambda \in \mathbf{C}$, для которых метод устойчив.

Пример:

Явная схема Эйлера

$$\frac{y_{n+1} - y_n}{\tau} = \lambda y_n, \quad y_{n+1} = (1 + \mu)y_n$$

Метод является устойчивым, если $|q| < 1$, $q = 1 + \mu$. Тогда область устойчивости определяется неравенством:

$$\begin{aligned} |1 + \mu| &\leq 1 \\ |1 + \mu_0 + i\mu_1| &\leq 1 \\ (1 + \mu_0)^2 + \mu_1^2 &\leq 1 \end{aligned}$$

В данном случае область устойчивости представляет собой внутренность круга с центром в точке $(0, -1)$ и радиусом 1 в координатах μ_0, μ_1 .

Неявная схема Эйлера

$$\frac{y_{n+1} - y_n}{\tau} = f(t_{n+1}, y_{n+1})$$

$$\frac{y_{n+1} - y_n}{\tau} - \lambda y_{n+1} = 0$$

$$y_{n+1}(1 + \tau\lambda) = y_n$$

Обозначим $q = \frac{1}{1 - \tau\lambda}$. Найдем область устойчивости: $|q| \leq 1$

$$\begin{aligned} \left| \frac{1}{1 - \tau\lambda} \right| &\leq 1 \\ |1 - \mu| &\geq 1 \\ (1 - \mu_0)^2 + \mu_1^2 &\geq 1 \end{aligned}$$

Областью устойчивости неявной схемы Эйлера является внешность круга радиуса 1 с центром в точке $(1, 0)$.

Определение. Разностный метод A -устойчив, если область его устойчивости содержит левую полуплоскость.

Замечание. Если разностный метод A -устойчив, то он абсолютно устойчив.

Утверждение. Доказано, что абсолютно устойчивых многошаговых разностных методов не существует.

Утверждение. Доказано, что не существует абсолютно устойчивых многошаговых неявных разностных методов, точность которых выше 2 порядка.

Рассмотрим пример разностного метода, имеющего второй порядок точности, который является A -устойчивым.

Симметричная схема

$$\frac{y_{n+1} - y_n}{\tau} = 0.5(f(t_n, y_n) + f(t_{n+1}, y_{n+1}))$$

Проверим, будет ли данная схема абсолютно устойчивой:

$$\begin{aligned} & \frac{y_{n+1} - y_n}{\tau} - 0.5\lambda(y_n + y_{n+1}) \\ & (y_{n+1} - y_n) - \tau - 0.5\mu(y_n + y_{n+1}) = 0 \\ & (1 - 0.5\mu)y_{n+1} = (1 + 0.5\mu)y_n \\ & y_{n+1} = qy_n, q = \frac{(1 + 0.5\mu)}{(1 - 0.5\mu)} \end{aligned}$$

Устойчивость

$$\begin{aligned} & |q| \leq 1 \\ & |1 + 0.5\mu| \leq |1 - 0.5\mu| \\ & (1 + 0.5\mu_0)^2 + \mu_1^2 \leq (1 - 0.5\mu_0)^2 + \mu_1^2 \\ & 1 + \mu_0 + 0.25\mu_0^2 \leq 1 - \mu_0 + 0.25\mu_0^2 \end{aligned}$$

А это возможно только в том случае, если $\mu_0 \leq 0$.

Определение. Разностный метод называется $A(\alpha)$ -устойчивым ($\alpha > 0$), если область устойчивости этого метода содержит угол в левой полуплоскости ($\mu_0 \leq 0$).

Замечание. Явных $A(\alpha)$ -устойчивых методов не существует. Были найдены $A(\alpha)$ -устойчивые методы 3го и 4го порядка.

Рассмотрим пример разностной схемы 4го порядка, которая является $A(\alpha)$ -устойчивая для некоторого $\alpha > 0$:

$$\frac{25y_{n+4} - 48y_{n+3} + 36y_{n+2} - 16y_{n+1} + 3y_n}{12\tau} = f(t_{n+4}, y_{n+4})$$