Московский государственный университет имени М. В. Ломоносова

Факультет Вычислительной Математики и Кибернетики

Неофициальные конспекты лекций численных методов 3-го курса 3-го потока (версия 8.0)

лектор Н.И. Ионкин

Москва

2012

Данные конспекты отражают содержание лекционного курса «Численные методы», читаемого студентам факультета вычислительной математики и кибернетики МГУ им. М. В. Ломоносова. Конспекты составили студенты 3–го курса кафедры СП А.С. Колганов, И.Т. Ядгаров, О.В. Горемыкин и Д.М. Биренбаум. Конспекты не являются официальной литературой и не могут быть использованы в качестве основного материала на экзамене. За все ошибки, допущенные в данных конспектах, составители ответственности не несут.

Отдельная благодарность за исправление ошибок: Байбородову А., Федорову А.,Белышову М.,Каганову В.,Моисееву Б.,Бабакову А., Пушину К.

Оглавление

Глава 1

Численные методы линейной алгебры

§1 Введение

Рассмотрим матричное уравнение вида

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

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

$$x = (x_1, x_2, \ldots, x_m)^T,$$

$$f = (f_1, f_2, \ldots, f_m)^T.$$

Так как матрица A невырождена, то решение системы (1) существует и единственно. Существует две группы методов поиска решения системы линейных алгебраических уравнений:

1. Прямые методы.
2. Итерационные (приближенные) методы.

Также будем рассматривать задачу на собственные значения:

$$Ax = \lambda x, \tag{2}$$

где $x \neq 0$ — собственные векторы, λ — собственные значения. При численном решении задачи на собственные значения обычно рассматривают две проблемы:

1. Частичная проблема собственных значений.
2. Полная проблема собственных значений – QR алгоритм.

В этой главе также будет решена задача нахождения обратной матрицы.

§2 Связь метода Гаусса с разложением матрицы на множители

Рассмотрим матричное уравнение вида

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

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

С помощью элементарных преобразований в прямом ходе метода Гаусса получаем верхнетреугольную матрицу с диагональными элементами, равными 1:

$$\begin{pmatrix} 1 & a_{12} & \dots & a_{1m} \\ 0 & 1 & \dots & a_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

Сведение матрицы A к данному виду требует $\dfrac{m^3 - m}{3}$ действий (умножений и делений). Кроме того требуется $\dfrac{m(m+1)}{2}$ — для преобразования правой части, а для обратного хода метода Гаусса потребуется $\dfrac{m(m-1)}{2}$ действий.

Представление матрицы A в виде

$$A = BC \tag{2}$$

называется факторизацией матрицы A. Матрицы B и C имеют вид:

$$B = \begin{pmatrix} b_{11} & 0 & \dots & 0 \\ b_{21} & b_{22} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ b_{m1} & b_{m2} & \dots & b_{mm} \end{pmatrix} C = \begin{pmatrix} 1 & c_{12} & \dots & c_{1m} \\ 0 & 1 & \dots & c_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

Покажем, что нахождение элементов матриц B и C возможно при определенном ограничении на матрицу A. Запишем представление (2) поэлементно:

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

Представим эту сумму в виде:

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

Так как $b_{il} = 0,\ l > i,\ l = \overline{1, m}$ то:

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

Поделим обе части на b_{ii} и выразим c_{ij} (для этого требуется $b_{ii} \neq 0$):

$$c_{ij} = \frac{a_{ij} - \sum\limits_{l=1}^{i-1} b_{il}c_{lj}}{b_{ii}} \quad , \; i < j \tag{3}$$

Перепишем разложение матрицы $A = BC$ в виде :

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

Учитывая, что $\sum\limits_{l=i+1}^{m} b_{il}c_{lj} = 0$ в силу того, что $c_{lj} = 0, \quad j < l$ и $c_{jj} = 1$, получим:

$$b_{ij} = a_{ij} - \sum_{l=1}^{j-1} b_{il}c_{lj} \quad i \geqslant j \tag{4}$$

Алгоритм нахождения матриц B и C в представлении (2):

1. $b_{11} = a_{11}, c_{1j} = \frac{a_{1j}}{b_{11}} = \frac{a_{1j}}{a_{11}}, j = \overline{2, m}$, то есть нашли все элементы c_{1j}.
2. $b_{i1} = a_{i1} \Rightarrow$ нашли все элементы b_{i1}.
3. Исходя из первых двух пунктов найдем диагональный элемент b_{22}.
4. И так далее.

Утверждение. *Пусть все главные угловые миноры матрицы A отличны от нуля. Тогда факторизация матрицы A, представленная в виде (2), возможна единственным образом.*

Доказательство: Введем $\Delta_0 = 1$. В силу того, что $A_i = B_iC_i \Rightarrow |A_i| = |B_i||C_i|$. Так как главная диагональ матрицы C состоит из единиц, то $|A_i| = \Delta_i = b_{11}b_{22}\dots b_{ii} \;\Rightarrow$ $b_{ii} = \frac{\Delta_i}{\Delta_{i-1}}, \quad i = \overline{1, m} \quad \Rightarrow \quad b_{ii} \neq 0.$ ∎

Связь метода Гаусса с разложением матрицы A на множители

Рассматриваем систему (1) порядка m и факторизацию матрицы A (2):

$$BCx = f \;\Rightarrow\; \begin{cases} BY = f & (5) \\ Cx = Y & (6) \end{cases}$$

Задача. *Доказать, что нахождение матриц B и C требует $\frac{m^3-m}{3}$ умножений и делений.*

Решение: Воспользуемся формулами для факторизации матрицы:

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

Для вычисления каждого b_{ij} потребуется $j-1$ умножение. Тогда зафиксировав индекс i, получим:

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

Далее посчитаем для индекса i:

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

Нетрудно получить значение первой суммы: $\dfrac{m(m+1)(2m+1)}{6}$. Тогда результирующая будет равна:

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

Далее для вычисления элементов c_{ij} воспользуемся формулой

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

Для каждого c_{ij} потребуется $i-1$ умножений и одно деление. Зафиксируем индекс j:

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

И для индекса j получаем аналогичную формулу:

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

Просуммируем два полученных результата для окончательно ответа:

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

□

В прямом методе Гаусса при сведении матрицы к верхнетреугольному виду с единицами на главной диагонали требуется $\dfrac{m^3-m}{3}$ действий, то есть в точности такое число, что и для факторизации матрицы A.

Распишем покоординатно системы (5) и (6):

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

$$x_i + c_{ii+1}x_{i+1} + \cdots + c_{im}x_m = y_i, i = \overline{1,m}$$

Считая, что $b_{ii} \neq 0$, выразим y_i и x_i:

$$y_i = \frac{f_i - \sum\limits_{l=1}^{i-1} b_{il}y_l}{b_{ii}} \tag{5*}$$

$$x_i = y_i - \sum_{l=i+1}^{m} c_{il}x_l \tag{6*}$$

Эти две формулы требуют $\frac{m(m+1)}{2} + \frac{m(m-1)}{2}$ действий. Для преобразования правых частей в методе Гаусса требуется $\frac{m(m+1)}{2}$, что совпадает с числом действий в системе (5*). Для обратного хода метода Гаусса требуется $\frac{m(m-1)}{2}$ действий, что совпадает с числом действий в системе (6*). Тогда общее количество действия для метода Гаусса равно:

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

§3 Обращение матриц методом Гаусса-Жордана

Рассмотрим матричное уравнение вида

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

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

Так как определитель матрицы A не равен нулю, то существует обратная к ней матрица и, по определению

$$A^{-1}A = AA^{-1} = E$$

Обозначим $A^{-1} = X$, тогда получим уравнение

$$AX = E,$$

где $X = x_{ij}, \quad i,j = \overline{1,m}$.

Для решения системы

$$AX = E$$

методом Гаусса требуется число операций порядка m^6. Покажем, что число действий можно снизить до m^3.

Распишем покоординатно:

$$\sum_{l=1}^{m} a_{il}x_{lj} = \delta_{ij} \quad \text{– в этой системе } m^2 \text{ неизвестных}$$

Введем вектор-столбец

$$x^{(j)} = (x_{1j}, x_{2j}, \ldots, x_{mj})^T, \quad i = \overline{1,m}.$$

Введем вектор правых частей единичной матрицы и обозначим его

$$\delta^{(j)} = (0, 0, \ldots, 0, 1, 0, \ldots, 0, 0),$$

где на j позиции стоит единица.

Тогда видно, что для того чтобы обратить матрицу, необходимо решить m систем

$$Ax^{(j)} = \delta^{(j)}, \quad j = \overline{1,m} \tag{2}$$

Таким образом, решение системы с m^2 неизвестными сведено к решению m систем с m неизвестными и фиксированной матрицей A. Теперь применим факторизацию (предполагая, что необходимое условие выполнено, т.е. все угловые миноры отличны от нуля): $A = BC$.

Вновь расписав уравнение, получим

$$BCx^{(j)} = \delta^{(j)}.$$

Обозначая вектор $Cx^{(j)} = Y^{(j)}$, видим, что для решения системы (2) нужно решить две системы:

$$BY^{(j)} = \delta^{(j)}, \quad j = \overline{1,m} \tag{3}$$

$$Cx^{(j)} = Y^{(j)}, \quad j = \overline{1,m} \tag{4}$$

Эти системы имеют треугольные матрицы, поэтому все неизвестные находятся по явным формулам. Следовательно,по совокупности эти две системы требуют m^2 действий. А так как необходимо решить m систем, то на решение всех систем понадобится m^3 действий. Факторизация, проведенная один раз, требует $\dfrac{m^3 - m}{3}$ действий. Откуда получаем, что общее количество действий равно $\dfrac{4}{3}m^3 - \dfrac{m}{3}$.

Однако, если воспользоваться спецификой видов матриц, то число действий можно уменьшить до m^3. Докажем данное утверждение.

Рассмотрим систему (3). Учитывая нижнетреугольную форму матрицы B, получим:

$$\begin{aligned}
b_{11}y_1^{(j)} = 0 \quad &\Longrightarrow \quad y_1^{(j)} = 0 \\
b_{21}y_1^{(j)} + b_{22}y_2^{(j)} = 0 \quad &\Longrightarrow \quad y_2^{(j)} = 0 \\
&\cdots \\
b_{j-1,1}y_1^{(j)} + b_{j-1,2}y_2^{(j)} + \cdots + b_{j-1,j-1}y_{j-1}^{j} = 0 \quad &\Longrightarrow \quad y_{j-1}^{(j)} = 0
\end{aligned}$$

Откуда получаем, что $y_i^{(j)} = 0, \quad$ где $i \leqslant j-1$

Следующее j-ое уравнение даст нам (учитывая, что $b_{jj} \neq 0$):

$$b_{jj}y_j^{(j)} = 1 \quad \Longrightarrow \quad y_j^{(j)} = \frac{1}{b_{jj}} \tag{$*$}$$

Оставшиеся уравнения системы имеют вид:

$$b_{i,j}y_j^{(j)} + b_{i,j+1}y_{j+1}^{(j)} + \cdots + b_{i,i}y_i^j = 0,\ i = \overline{j+1, m},\ b_{ii} \neq 0 \quad \Longrightarrow \quad y_i^{(j)} = \frac{-\sum\limits_{l=j}^{i-1} b_{il}y_{lj}}{b_{ii}}$$

Фиксируем все индексы, т.е. i и j и считаем число действий:

$$1 \text{ деление} \quad + \quad (i-j) \text{ умножений.}$$

Теперь отпускаем один из индексов, например, i. Тогда получим

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

умножении при фиксированном j.

Вдобавок к этому $(m-j)$ делений $+$ 1 деление от (*), откуда получаем, что при фиксированном j всего

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

умножений и делений.

Теперь, отпустив j, получим

$$\sum_{j=1}^{m} \frac{(m-j+1)(m-j+2)}{2} \tag{5}$$

делений и умножений при решении системы вида $B \cdot Y = f$.

Задача. *Доказать, что для реализации (5) необходимо* $\dfrac{m(m+1)(m+2)}{6}$ *операций.*

Решение:

$$\sum_{j=1}^{m} \frac{(m-j+1)(m-j+2)}{2} = \{ \text{ проведем замену } k = m-j+1\} = \sum_{k=1}^{m} \frac{k(k+1)}{2} =$$

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

Откуда следует, что первая сумма равна $\dfrac{k(k+1)}{4}$, а вторая — $\dfrac{k(k+1)(2k+1)}{12}$. Сложив полученные результаты, получим требуемое значение $\dfrac{m(m+1)(m+2)}{6}$. $\square$

Для реализации (4) необходимо $\frac{m(m-1)}{2}$ действий (обратный ход метода Гаусса), а учитывая, что нам надо решить m уравнений, получим $\frac{m^2(m-1)}{2}$ умножений и делений.

Таким образом, весь процесс Гаусса-Жордана обращения матрицы требует m^3 действий (факторизация + решение системы (3) + решение системы (4)):

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

§4 Метод квадратного корня

Рассмотрим матричное уравнение вида

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

где A — матрица размера $(m\times m)$, $|A|\neq 0$, $A=A^*$ (эрмитова, т.е. $a_{ij}=\overline{a_{ji}}$).

Представим матрицу A в виде $A=S^*DS$, где

$$D=\begin{pmatrix} d_{11} & 0 & \dots & 0 & 0\\ 0 & d_{22} & \dots & 0 & 0\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ 0 & 0 & \dots & 0 & d_{mm}\end{pmatrix},$$

$$S=\begin{pmatrix} s_{11} & s_{21} & \dots & s_{1m}\\ 0 & s_{22} & \dots & s_{2m}\\ \vdots & \vdots & \vdots & \vdots\\ 0 & 0 & \dots & s_{mm}\end{pmatrix},$$

где $s_{ii}>0,\ d_{ii}=\pm 1,\ i=\overline{1,m}$.

Следовательно S^* будет нижнетреугольная.

Покажем для симметричной матрицы второго порядка, что такое разложение возможно. Для простоты потребуем, чтобы она была вещественной.

$$A=\begin{pmatrix} a_{11} & a_{12}\\ a_{21} & a_{22}\end{pmatrix},\ S=\begin{pmatrix} s_{11} & s_{12}\\ 0 & s_{22}\end{pmatrix},\ D=\begin{pmatrix} d_{11} & 0\\ 0 & d_{22}\end{pmatrix},\ S^*=S^T=\begin{pmatrix} s_{11} & 0\\ s_{12} & s_{22}\end{pmatrix}$$

Найдем в начале произведение DS.

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

Теперь домножим на S^* слева.

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

Откуда получим, что

$$\begin{cases} a_{11} = d_{11}s_{11}^2 \\ a_{12} = s_{11}d_{11}s_{12} \\ a_{21} = a_{12} \\ a_{22} = d_{22}s_{22}^2 + s_{12}^2 d_{11} \end{cases}$$

Для того чтобы факторизация была возможна, необходимо, чтобы система была разрешима. Ее решение находится по формулам:

$$\begin{cases} d_{11} = sign(a_{11}) \\ s_{11} = \sqrt{|a_{11}|} \\ s_{12} = \dfrac{a_{12}}{s_{11}d_{11}} \\ d_{22} = sign(a_{22} - s_{12}^2 d_{11}) \\ s_{22} = \sqrt{|a_{22} - s_{12}^2 d_{11}|} \end{cases}$$

Таким образом показано, что для матрицы A (вещественной и симметричной) возможна факторизация и все находится по явным формулам. Рассмотрим общий случай. Элементы матрицы DS находятся по формуле:

$$(DS)_{ij} = \sum_{l=1}^{m} d_{il}s_{lj} = \Big\{\text{с учетом свойств матриц}\Big\} = d_{ii}s_{ij}$$

Заметим, что

$$(S^*)_{ij} = \overline{S}_{ji}$$

Умножим слева матрицу S^* на DS:

$$(S^*DS)_{ij} = \sum_{l=1}^{m} \overline{s}_{li}d_{ll}s_{lj}$$

Распишем сумму, приравняв ее к a_{ij} элементу

$$\sum_{l=1}^{i-1} \overline{s}_{li}d_{ll}s_{lj} + \overline{s}_{ii}d_{ii}s_{ij} + \sum_{l=i+1}^{m} \overline{s}_{li}d_{ll}s_{lj} = a_{ij}, \quad i,j = \overline{1,m}$$

Учтем специфику матрицы S: $\overline{s}_{li} = 0, \quad l > i$. Следовательно,

$$\sum_{l=i+1}^{m} \overline{s}_{li}d_{ll}s_{lj} = 0$$

Тогда получим

$$\overline{s}_{ii}d_{ii}s_{ij} = a_{ij} - \sum_{l=1}^{i-1} \overline{s}_{li}d_{ll}s_{lj}, \quad i \leqslant j \tag{3}$$

Рассмотрим два случая ($i = j$ и $i < j$).

Положим $i = j$. Получим:

$$\overline{s}_{ii} s_{ii} d_{ii} = a_{ii} - \sum_{l=1}^{i-1} \overline{s}_{ll} s_{li} d_{ll}$$

Так как $z\overline{z} = |z|^2$, то

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

Таким образом, мы получили формулы, из которых можем найти диагональные элементы матрицы D:

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

Далее понятно, что

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

Осталось вновь вернуться к исходной формуле, частный случай которой мы рассмотрели. Из формулы (3) можно найти элементы s_{ij}, считая, что $d_{ii}\overline{s}_{ii} \neq 0$:

$$s_{ij} = \frac{a_{ij} - \sum_{l=1}^{i-1} \overline{s}_{li} d_{ll} s_{lj}}{\overline{s}_{ii} d_{ii}}, \quad i < j.$$

Система перепишется в виде:

$$S^* D S x = f$$

Обозначая $DSx = Y$, получим:

$$\begin{cases} S^* Y = f & (4) \\ DSx = Y & (5) \end{cases}$$

Матрица S^* нижнетреугольная и из (4) легко находится вектор Y. Затем, решая уравнение (5), находим вектор x.

Был рассмотрен метод квадратного корня. Этот метод имеет преимущество по количеству арифметических действий (умножений и делений) перед методом Гаусса, который пропорционален $\frac{m^3}{3}$ действий, а рассмотренный нами способ - $\frac{m^3}{6}$, так как расчеты проводились только для части матрицы (но здесь есть еще m извлечений корня).

Следовательно, если матрица эрмитова или самосопряженная, то один из наиболее эффективных прямых методов - метод квадратного корня.

§5 Примеры и канонический вид итерационных методов решения СЛАУ

Рассмотрим матричное уравнение вида

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

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

$$x = (x_1, x_2, \ldots, x_m)^T,$$

$$f = (f_1, f_2, \ldots, f_m)^T.$$

Когда решается линейная система (1), то в правой части, как правило, стоит функция наблюдения. Ясно, что прямой метод дает точное (числовое) решение, абстрагируясь от округления - в силу конечной разрядной сетки.

Также, если m достаточно велико, то количество действий даже для суперкомпьютера может оказаться очень большим. А использование итерационных методов позволит решить систему (1) быстрее. Более того, в некоторых случаях возможно оценить число итераций, необходимых для достижения заданной точности.

В качестве примера рассмотрим методы Якоби и Зейделя. Запишем систему (1) покоординатно:

$$\sum_{j=1}^{m} a_{ij}x_j = f_i, \quad i = \overline{1, m} \tag{2}$$

Рассмотрим метод Якоби (МЯ). Перепишем равенство (2) следующим образом:

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

Тогда, отсюда можно выразить x_i:

$$x_i = \frac{f_i - \sum\limits_{j=1}^{i-1} a_{ij}x_j - \sum\limits_{j=i+1}^{m} a_{ij}x_j}{a_{ii}},$$

с предположением, что $a_{ii} \neq 0$.

Обозначим x_i^n — n-я итерация i-й координаты.

Чтобы организовать метод Якоби, слева «навешивают» итерацию $n+1$, а справа собирают все с n-ой итерацией.

$$x_i^{n+1} = \frac{f_i - \sum\limits_{j=1}^{i-1} a_{ij}x_j^n - \sum\limits_{j=i+1}^{m} a_{ij}x_j^n}{a_{ii}}, \tag{3}$$

где $n = 0, 1, 2, \ldots,$ $\quad x^0$ — задано.

В каждом конкретном итерационном методе выбор начального приближения является самостоятельной задачей.

Понятно, что никакой процесс не может продолжаться бесконечно долго, нужно будет где-то оборвать это действие. Заканчиваем итерационный процесс тогда, когда достигается оценка:

$$\|x^n - x\| < \varepsilon$$

Запишем метод Зейделя:

$$x_i^{n+1} = \frac{f_i - \sum\limits_{j=1}^{i-1} a_{ij} x_j^{n+1} - \sum\limits_{j=i+1}^{m} a_{ij} x_j^n}{a_{ii}}, \tag{4}$$

где $n = 0, 1, 2, \ldots,$ x^0 - задано.

Формулу (4) можно так же реализовать по явным методам. Рассмотрим этот способ.

Положим $i = 1$. Найдем первую координату $(n+1)$-ой итерации:

$$x_1^{n+1} = \frac{f_1 - \sum\limits_{j=2}^{m} a_{1j} x_j^n}{a_{11}}$$

Найдем вторую координату:

$$x_2^{n+1} = \frac{f_2 - a_{21} x_1^{n+1} - \sum\limits_{j=3}^{m} a_{2j} x_j^n}{a_{22}}$$

И так далее. То есть, организуя вычислительный процесс с первой координаты, можно найти все значения вектора x по явным формулам для $(n+1)$-й итерации.

Таким образом, метод Зеделя реализуется по явным формулам. При изучении итерационных методов важно не упускать из виду два вопроса: первый - сходимость метода и условия этой сходимости, второй - получение оценки скорости сходимости метода.

Для исследования этих вопросов удобно рассматривать системы в матричном виде. Любую матрицу мы можем представить в виде суммы трех матриц $R_1 + D + R_2$, где

$$R_1 = \begin{pmatrix} 0 & & 0 \\ & \ddots & \\ a_{ij} & & 0 \end{pmatrix} \text{ — нижнетреугольная матрица с нулевой диагональю,}$$

$$D = \begin{pmatrix} a_{11} & & 0 \\ & \ddots & \\ 0 & & a_{mm} \end{pmatrix} \text{ — диагональная матрица,}$$

$$R_2 = \begin{pmatrix} 0 & & a_{ij} \\ & \ddots & \\ 0 & & 0 \end{pmatrix} \text{ — верхнетреугольная матрица с нулевой диагональю.}$$

Тогда перепишем уравнение (1):

$$R_1x + Dx + R_2x = f$$

$$\Downarrow$$

$$Dx = f - R_1x - R_2x$$

Если предположить, что матрица D имеет обратную, а это как раз и требует выполнения условия $a_{ii} \neq 0$, то тогда, домножив слева на матрицу D^{-1}, получаем

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

Итак, организуем итерационный процесс для метода Якоби и Зейделя.
МЯ:

$$x^{n+1} = D^{-1}f - D^{-1}R_1x^n - D^{-1}R_2x^n \tag{5}$$

МЗ:

$$x^{n+1} = D^{-1}f - D^{-1}R_1x^{n+1} - D^{-1}R_2x^n \tag{6}$$

МЯ:

$$Dx^{n+1} = f - R_1x^n - R_2x^n \tag{7}$$

МЗ:

$$(D + R_1)x^{n+1} = f - R_2x^n \tag{8}$$

МЯ:

$$D(x^{n+1} - x^n) + Ax^n = f \tag{9}$$

МЗ:

$$(D + R_1)(x^{n+1} - x^n) + Ax^n = f \tag{10}$$

где $n = 0, 1, 2, \ldots,$ $\quad x^0$ — задано.

Видно, что если есть сходимость, то сходится к решению нашей системы. Так математики и пришли к каноническому виду двухслойного итерационного процесса.

Определение. *Канонической формой записи двухслойного итерационного метода решения системы* (1) *называется запись вида*

$$B_{n+1}\frac{x^{n+1} - x^n}{\tau_{n+1}} + Ax^n = f, \tag{11}$$

где $n = 0, 1, 2, \ldots,$ $\quad x^0$ *— задано,* $\tau_{n+1} > 0$ *(итерационный параметр) и* $\exists B_{n+1}^{-1}$. *Если* $B_{n+1} = E$, *то метод (11) называется явным, в противном случае – неявным.*

Замечание. *Если параметр* τ *и матрица* B *зависят от итерации, то метод называется нестационарным, иначе стационарным. Далее мы будем рассматривать только стационарные методы.*

Отметим, что если B_{n+1} – диагональная матрица, то метод реализуется по явным формулам, хотя формально метод неявный.

Рассмотрим еще один метод — метод простой итерации (метод релаксации).

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

$n=0,1,2,\ldots,\ x^0$ – задано.
Если τ зависит от n, то получающийся метод

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

называется методом Ричардсона.

Попеременно треугольный итерационный метод (метод Самарского)

Представим матрицу A в виде $A=R_1+R_2$, где

$$R_1=\begin{pmatrix}0.5a_{11} & & 0\\ & \ddots & \\ a_{ij} & & 0.5a_{mm}\end{pmatrix}$$ — нижнетреугольная матрица,

$$R_2=\begin{pmatrix}0.5a_{11} & & a_{ij}\\ & \ddots & \\ 0 & & 0.5a_{mm}\end{pmatrix}$$ — верхнетреугольная матрица.

Тогда попеременно треугольный итерационный метод имеет следующий вид:

$$(E+wR_1)(E+wR_2)\frac{x^{n+1}-x^n}{\tau}+Ax^n=f, \tag{12}$$

где $n=0,1,\ldots,\quad x^0$ — задано, $\tau>0$, $w>0$ – итерационные переметры.

В данном методе имеется два итерационных параметра. С точки зрения алгоритма реализации, этот метод не сложнее, чем предыдущие, а по сходимости - на порядок лучше.

Введем вектор

$$(E+wR_2)\frac{x^{n+1}-x^n}{\tau}=W^{n+1}$$

Определение. *Разность между правой и левой частями решаемой системы называется невязкой.*

В нашем случае невязка имеет вид

$$f-Ax^n=r^n$$

Тогда видно, что на первом этапе можем решать уравнение

$$(E + wR_1)W^{n+1} = r^n$$

Правая часть известна, а $(E + wR_1)$ — нижнетреугольная матрица. Нахождение вектора системы с нижнетреугольной матрицей осуществляется по явным формулам, начиная с первой компоненты вектора W.

Теперь, на втором этапе решаем уравнение

$$(E + wR_2)V^{n+1} = W^{n+1},$$

где

$$V^{n+1} = \frac{x^{n+1} - x^n}{\tau}.$$

На третьем этапе $(n+1)$-ю итерацию находим по формуле:

$$x^n + \tau V^{n+1} = x^{n+1}.$$

Таким образом, несмотря на то, что метод Самарского неявный с двумя итерационными параметрами, его реализация не представляет никакой трудности.

§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 двухслойный стационарный метод решения уравнения (1):

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

где $\tau > 0$, существует обратная матрица B^{-1}, $n = 0, 1, 2, \ldots$, и задано начальное условие x^0.

Когда мы говорим, что начальное условие задано, то это значит, что либо его надо выбирать исходя из каких-то жестких условий, либо есть свобода выбора.

Говоря о сходимости нужно четко понимать, по какой норме эта сходимость получается, поэтому нам необходимо ввести линейное нормированное пространство.

Пусть H - линейное пространство размерности m:

$$\dim H = m$$

Возьмем два произвольных вектора x и y из этого пространства:

$$x \in H, \quad x = (x_1, x_2, \ldots, x_m);$$

$$y \in H, \quad y = (y_1, y_2, \ldots, y_m);$$

Для того, чтобы нормировать это пространство, нужно ввести скалярное произведение и норму (пространство H может быть как вещественным, так и комплексным).

Введем скалярное произведение векторов x и y:

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

Если допускаем и унитарное пространство, то

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

Тогда введем норму: (которая известна из курса линейной алгебры как евклидова норма):

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

Эту норму математики также часто называют среднеквадратичной нормой. Заметим, что это не сильная[1] норма.

Рассмотрим самосопряженный линейный оператор[2] $D = D^* > 0$. Введем новую норму в вещественном пространстве.

Определение. *Энергетическая норма - это норма, задаваемая соотношением*

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

Замечание. *Заметим, что требование самосопряженности здесь очень важно. Если бы матрица D не была самосопряженной, то скалярное произведение (Dx, x) было бы комплексным числом, а значит и связывать с нормой это произведение мы бы не имели права.*

Вспомним несколько принципиально важных понятий, с которыми связаны очень непростые переходы, которые будут использоваться в доказательствах последующих теорем:

1. $D > 0 \Longleftrightarrow (Dx, x) > 0, \quad \forall x \neq 0;$
2. $D \geqslant 0 \Longleftrightarrow (Dx, x) \geqslant 0, \quad \forall x \in H;$

[1] Более сильную норму принято считать такую, в которой близость двух векторов будет более жесткой.Например, норма в C будет более сильной, чем в L_2, так как в C близость векторов будет покоординатная(поточечная).

[2] Как у Маяковского «Партия и Ленин — близнецы-братья», так и у нас слова линейный оператор и матрица отныне будут нести один и тот же смысл.

3. $D = D^* > 0 \Longrightarrow \exists\, \delta > 0 \colon \quad (Dx, x) \geqslant \delta \|x\|^2$;

   Здесь легко понять, что δ будет связана с минимальным собственным значением, так как, если у нас самосопряженный и положительно определенный оператор, то у него все собственные значения положительны и есть базис из собственных векторов. Если разложить вектор x по базису из собственных векторов и заменить собственное значение на минимальное, то δ как раз им и будет.

4. Если $D = D^* > 0$, то

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

Задача. *Пусть H — вещественное пространство, C — положительно определенный линейный оператор. Доказать, что*

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

Решение: Для решения данной задачи воспользуемся следующими равенствами, верными для вещественного пространства H:

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

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

$$(Cx, x) = (\frac{C + C^*}{2} x, x) + (\frac{C - C^*}{2} x, x) = (\frac{C + C^*}{2} x, x) + \frac{1}{2}((C^* x, x) - (Cx, x)) = (\frac{C + C^*}{2} x, x), \ \forall x \in H$$

□

Для того, чтобы мы могли говорить о сходимости, введем понятие погрешности.

Определение. *Вектор, вида*

$$V^n = x^n - x$$

называется погрешностью на n-ой итерации

Таким образом, для того, чтобы доказать, что итерационный метод сходится, нам необходимо показать, что в соответствующей норме погрешность на n-ой итерации будет стремиться к нулю при n стремящемся к бесконечности

$$\lim_{n \to \infty} \|V^n\| = 0, \tag{3}$$

что и будет означать, что наше приближение x^n будет стремиться к точному решению x.

Тогда, воспользовавшись тем, что $x^n = V^n + x$, перепишем уравнение (2) через вектор погрешности

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

где $n = 0, 1, \ldots$, $V^0 = x^0 - x$.

Таким образом, приступим к изучению уравнения (4). Для этого обычно выражают $n+1$ итерацию через n, с условием того, что существует обратная к B матрица.

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

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

Выразим отсюда погрешность на $n + 1$ итерации

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

Таким образом, мы получили матрицу S, которая связывает предыдущую итерацию с последующей

$$S = E - \tau B^{-1}A \tag{5}$$

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

Нетрудно заметить, что сходимость и скорость сходимости вектора V^n всецело зависит от свойств матрицы S, а именно от ее спектра. Именно эти свойства спектра матрицы S и изучает первая теорема, которую мы сейчас сформулируем.

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

$$|\lambda^S| < 1, \quad \forall\, x^0$$

Замечание. *Теорема, конечно замечательная, и казалось бы, владея техникой нахождения собственных значений оператора, мы бы с легкостью все решали. Однако алгебраические многочлены до четвертой степени математики решать умеют, а выше, как доказано Абелем и Галуа, вообще в радикалах не разрешимы. Поэтому, на самом деле, эта замечательная теорема для применения практически не годна - мы ей просто не сможем воспользоваться, кроме каких-нибудь простеньких случаев.*

Рассмотрим теорему Самарского о достаточных условиях сходимости итерационного метода. Понятно, что это не критерий, и, если эти условия не выполнены, то сходимость все-равно может иметь место. Зато условия, которые будут указаны, являются проверяемыми, и, применяя их к конкретной задаче, можно с уверенность сказать, что метод сходится.

Теорема 2 (Самарского)**.** *Пусть H — вещественное пространство, $A = A^* > 0$, где A — матрица системы* (1)*, $\tau > 0$ и выполнено матричное неравенство:*

$$B - 0.5\,\tau A > 0 \tag{6}$$

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

$$\|x^n - x\| = \left(\sum_{j=1}^{m}(x_j^n - x_j)^2\right)^{\frac{1}{2}} \xrightarrow[n\to\infty]{} 0\,, \quad \forall x^0$$

Доказательство: Введем числовую последовательность $y_n = (AV^n, V^n) \geqslant 0$. Для начала докажем, что она монотонная. Для этого рассмотрим y_{n+1}:

$$\begin{aligned}
y_{n+1} &= (AV^{n+1},\, V^{n+1}) = \Big\{V^{n+1} = SV^n\Big\} = \\
&= (ASV^n, SV^n) = \Big\{S = (E - \tau B^{-1}A)\Big\} = \\
&= (\,A(E - \tau B^{-1}A)V^n,\, (E - \tau B^{-1}A)V^n\,) = \\
&= (AV^n,\, V^n) - \tau\big[(AV^n,\, B^{-1}AV^n) + (AB^{-1}AV^n,\, V^n) - \tau(AB^{-1}AV^n,\, B^{-1}AV^n)\big] = *
\end{aligned}$$

Итак, первое слагаемое по определению есть y_n, рассмотрим вторую часть равенства. Так как пространство H вещественное, то скалярное произведение коммутативно, поэтому

$$(AB^{-1}AV^n,\, V^n) = (B^{-1}AV^n,\, A^*V^n) = (B^{-1}AV^n,\, AV^n) = (AV^n,\, B^{-1}AV^n)$$

Тогда получим

$$\begin{aligned}
* &= y_n - \tau\big[\,2(AV^n, B^{-1}AV^n) - \tau(AB^{-1}AV^n, B^{-1}AV^n)\,\big] = \\
&= \Big\{(a,\, b) - \alpha(c,\, b) = (a - \alpha c,\, b)\Big\} = y^n - 2\tau\big((B - 0.5\tau A)B^{-1}AV^n,\, B^{-1}AV^n\big)
\end{aligned}$$

Таким образом мы получили тождество

$$\frac{y_{n+1} - y_n}{\tau} + 2\big((B - 0.5\tau A)B^{-1}AV^n,\, B^{-1}AV^n\big) = 0,$$

в котором оператор $(B - 0.5\tau A)$ положительный по условию, а следовательно и все скалярное произведение неотрицательно:

$$\big((B - 0.5\tau A)B^{-1}AV^n,\, B^{-1}AV^n\big) \geqslant 0$$

Отсюда следует, что

$$y_{n+1} \leqslant y_n,$$

что и означает монотонность последовательности y_n.

А значит, согласно теореме Вейерштрасса, у последовательности существует предел y:

$$\exists \lim_{n\to\infty} y_n = y$$

Мы доказали, что введенная числовая последовательность y_n является монотонной и ограниченной снизу, а значит, имеет предел. Первая часть теоремы доказана.

Чтобы приступить ко второй части доказательства, нам понадобится доказать свойство положительно определенного линейного оператора, которое сформулируем в виде задачи.

Задача. *Пусть H — вещественное пространство, C — положительный линейный оператор. Доказать, что $\exists\,\delta$, такое что:*

$$(Cx,\, x) \geqslant \delta \|x\|^2$$

Итак, воспользовавшись данным свойством, можем сказать, что существует такая константа δ, что

$$\big((B - 0.5\tau A)B^{-1}AV^n,\, B^{-1}AV^n)\big) \geqslant \delta\, \|B^{-1}AV^n\|^2$$

Теперь, вспомнив ранее полученное тождество, можем записать

$$\frac{y_{n+1} - y_n}{\tau} + 2\delta\, \|B^{-1}AV^n\|^2 \leqslant \frac{y_{n+1} - y_n}{\tau} + 2\big((B - 0.5\tau A)B^{-1}AV^n,\, B^{-1}AV^n\big) = 0$$

Обозначим для удобства

$$W^n = B^{-1}AV^n \tag{7}$$

Отсюда видно, что если устремить n к бесконечности, то норма вектора W^n устремится к нулю

$$\lim_{n\to\infty} \|W^n\| = 0$$

А теперь выразим погрешность через введенный нами вектор. Для этого домножим равенство (7) слева сначала на B, затем на A^{-1}

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

Так как норма произведения операторов не превосходит произведения их норм, то, оценив погрешность, получим, что она стремится к нулю при n стремящемся к бесконечности

$$\|V^n\| \leqslant \|A^{-1}B\| \cdot \|W^n\| \xrightarrow[n\to\infty]{} 0$$

Так как нигде в доказательстве на начальное приближение мы не опирались, то оно произвольное. Следовательно

$$\|V^n\| = \|x^n - x\| = \left(\sum_{j=1}^{m} (x_j^n - x_j)^2\right)^{\frac{1}{2}} \xrightarrow[n\to\infty]{} 0\,, \quad \forall x^0$$

■

Следствие 1. *Пусть* $A = A^* > 0$.

Тогда метод Якоби сходится в среднеквадратичной норме при любом начальном приближении, если выполнено операторное неравенство

$$2D > A,$$

где $A = R_1 + D + R_2, \quad D = diag(a_{11}, a_{22}, \ldots, a_{nn})$.

Доказательство: Положим, $B = D$ и $\tau = 1$. Тогда перепишем уравнение (2) в виде

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

По теореме Самарского метод сходится, если

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

или в нашем случае

$$D - 0.5A > 0,$$

а это выполняется в силу условия. Следовательно, метод Якоби сходится. ∎

Следствие 2. *Пусть самосопряженная положительно определенная матрица* $A = A^* > 0$ *является матрицей со строгим диагональным преобладанием*

$$a_{ii} > \sum_{\substack{j=1 \\ i \neq j}}^{m} |a_{ji}| \tag{8}$$

Тогда метод Якоби сходится при любом начальном приближении в среднеквадратичной норме.

Доказательство: Покажем, что если есть строгое диагональное преобладание, то верно матричное неравенство

$$2D > A,$$

которое доказывает сходимость метода (согласно следствию 1). Рассмотрим квадратичную форму $(Ax,\, x)$:

$$(Ax,\, x) = \sum_{i,j=1}^{m} a_{ij} x_i x_j \leqslant \sum_{i,j=1}^{m} |a_{ij}| \cdot |x_i| \cdot |x_j| \leqslant \frac{1}{2} \sum_{i,j=1}^{m} |a_{ij}| \cdot |x_i|^2 + \frac{1}{2} \sum_{i,j=1}^{m} |a_{ij}| \cdot |x_j|^2 =$$
$$= \Big\{ a_{ij} = a_{ji} \Big\} = (\frac{1}{2} \sum_{i,j=1}^{m} |a_{ij}| \cdot |x_i|^2) \cdot 2 = \sum_{i,j=1}^{m} |a_{ij}| \cdot |x_i|^2$$

Выделим отдельно суммирование по индексу i:

$$(Ax, x) \leqslant \sum_{i=1}^{m} x_i^2 \left(a_{ii} + \sum_{\substack{j=1 \\ i \neq j}}^{m} |a_{ij}| \right) < \Big\{ \sum_{\substack{j=1 \\ i \neq j}}^{m} |a_{ij}| < a_{ii} \Big\} < \sum_{i=1}^{m} 2a_{ii} x_i^2 = (2Dx,\, x).$$

Таким образом, мы получили, что

$$(Ax,\, x) < (2Dx,\, x),$$

откуда следует, что

$$2D > A$$

Следовательно выполняется условие следствия 1, и итерационный метод Якоби сходится при любом начальном приближении.

■

Задача. *Пусть* $A = A^* > 0$.
Доказать, что $a_{ii} > 0, \quad i = \overline{1, m}$.

Следствие 3. *Пусть* $A = A^* > 0$.
Тогда метод Зейделя сходится в среднеквадратичной норме при любом начальном приближении.

Доказательство: По определению метода Зейделя имеем:

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

Исходя из условия теоремы Самарского, для того, чтобы сходился метод Зейделя, достаточно выполнения неравенства

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

Докажем это.
Распишем матрицы A и B:

$$D + R_1 - \frac{1}{2}(R_1 + D + R_2) > 0$$

$$\Downarrow$$

$$D + R_1 - R_2 > 0$$

$$\Downarrow$$

$$\big((D + R_1 - R_2)x,\, x\big) > 0$$

$$\Downarrow$$

$$0 < (Dx,\, x) + (R_1 x,\, x) - (R_2 x,\, x) = (Dx,\, x) + (R_1 x,\, x) - (R_1^* x,\, x) = (Dx,\, x)$$

Откуда получаем, что

$$(Dx,\, x) > 0$$

Если матрица самосопряженная и положительно определенная, то все ее диагональные элементы больше нуля. Следовательно, матрица D тоже является положительно определенной, откуда следует верность неравенства.

■

Следствие 4. *Пусть* $A = A^*,\ A > 0$. *Рассмотрим метод простой итерации:*

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

Тогда если выберем $\tau\ 0 < \tau < \dfrac{2}{\gamma_2},\ \gamma_2 = \max \lambda_k^A$ *по всем* k, *то метод сходится в среднеквадратичной норме при любом начальном приближении.*

Доказательство: Достаточное условие сходимости: $B - 0.5\tau A > 0$. Следовательно $E - 0.5\tau A > 0$ и $1 - 0.5\tau\lambda_k^A > 0$. Откуда получаем исходное условие на τ с учетом $\tau > 0$ ∎

§7 Оценка скорости сходимости итерационных методов

Рассматриваем систему уравнений с квадратной матрицей порядка m:

$$Ax = f\ ,\ |A| \neq 0 \tag{1}$$

Решаем итерационным методом:

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

где $\tau > 0,\ \exists B^{-1},\ n = 1, 2, 3, \dots$, начально приближение x^0 — задано.
Введем погрешность $V^n = x^n - x$. Тогда из (2) получаем:

$$B\frac{V^{n+1} - V^n}{\tau} + AV^n = 0\ , \tag{3}$$

где $\tau > 0,\ n = 1, 2, 3, \dots$, начально приближение $V^0 = x^0 - x$ — задано.
Поставим задачу получения оценки вида:

$$\|V^{n+1}\| \leqslant \rho\|V^n\|\ , \tag{4}$$

где $n = 1, 2, 3, \dots$, $\rho \in (0, 1)$ и рассматриваемая норма пока не фиксирована.
Применив эту оценку как рекуррентную, получим: $\|V^n\| \leqslant \rho^n\|V^0\|$ и при $n \to \infty$ — $\rho^n \to 0 \Rightarrow$ норма погрешности $\|V^n\| \to 0$. Значит имеет место сходимость. Более быстрая сходимость будет при ρ близко к нулю. ρ называется константой затухания погрешности.
Если удастся получить данную оценку (4), то можно будет посчитать количество итераций исходя из нужной точности:

$$\|x^n - x\| \leqslant \varepsilon\|x^0 - x\|$$

$$\|x^n - x\| \leqslant \rho^n\|x^0 - x\|$$

Из свойства транзитивности неравенств получим: $\rho^n \leqslant \varepsilon \Rightarrow n ln\frac{1}{\rho} \geqslant ln\frac{1}{\varepsilon}$. Тогда проделав число итераций $n_0(\varepsilon) = \left[ln\frac{1}{\varepsilon} \;/\; ln\frac{1}{\rho}\right]$ будет верна оценка $n \geqslant n_0(\varepsilon)$, где ε — точность приближения, $ln\frac{1}{\rho}$ — скорость сходимости.

Введем вещественное линейное пространство H , размерность которого равна m. В нем вводим скалярное произведение

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

и среднеквадратичную норму:

$$\|x\| = \sqrt{(x, x)} = \sqrt{\sum_{i=1}^{m} {x_i}^2}$$

Остальные нормы энергетические:

$$\forall B = B^* \,, B > 0 \quad \|x\|_B = \sqrt{(Bx, x)}$$

Теорема 3 (об оценке скорости сходимости)**.** *Пусть даны матрицы $A = A^* \,, A > 0$ и $B = B^* \,, B > 0$, $\exists \rho \in (0, 1)$ такое , что выполнено операторное неравенство:*

$$\frac{1-\rho}{\tau} B \leqslant A \leqslant \frac{1+\rho}{\tau} B \tag{5}$$

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

$$\|V^{n+1}\|_B \leqslant \rho \|V^n\|_B \quad n = 1, 2, 3, \ldots \tag{6}$$

Доказательство:

Замечание. *Рассмотрим правую часть операторного неравенства (5) : $A \leqslant \frac{1+\rho}{\tau} B$. Откуда получаем $B - 0.5\tau A \geqslant 0$ при $\rho = 1$. Значит сходимость будет при любом начальном приближении к среднеквадратичной норме. Оценку (6) можно получить и в норме оператора A.*

Если $B = B^* \,, B > 0$, то тогда $\exists \sqrt{B} = \left(\sqrt{B}\right)^*$, $\sqrt{B} > 0$ и $B^{-\frac{1}{2}} = \left(B^{-\frac{1}{2}}\right)^*$, $B^{-\frac{1}{2}} > 0$. Применим $B^{-\frac{1}{2}}$ к (3):

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

Введем вектор $z^n = B^{\frac{1}{2}} V^n$. Чтобы получить (6), достаточно получить

$$\|z^{n+1}\| \leqslant \rho \|z^n\| \tag{7}$$

Это верно в силу самосопряженности оператора B : $\|z^n\|^2 = (z^n, z^n) = (B^{\frac{1}{2}}V^n, B^{\frac{1}{2}}V^n) = (BV^n, V^n) = \|V^n\|_B$
Подставим $V^n = B^{-\frac{1}{2}}z^n$ в полученное выше:

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

и выразим z^{n+1} через z^n: $z^{n+1} = z^n - \tau B^{-\frac{1}{2}}AB^{-\frac{1}{2}}z^n = Sz^n$,где

$$S = E - \tau B^{-\frac{1}{2}}AB^{-\frac{1}{2}} \tag{8}$$

— матрица перехода от n-й итерации к $(n+1)$-й итерации вектора z.

Докажем ,что все собственные значения матрицы S не превосходят по модулю ρ , $Se_k = s_k e_k,\ k = 1, \dots, m,\ e_k \not\equiv 0,\ s_k$ — собственные значения матрицы S.
Покажем, что $S = S^*$:

$$S^* = (E - \tau B^{-\frac{1}{2}}AB^{-\frac{1}{2}})^* = E^* - \tau(B^{-\frac{1}{2}})^*A^*(B^{-\frac{1}{2}})^* = E - \tau B^{-\frac{1}{2}}AB^{-\frac{1}{2}} = S$$

Покажем , что $|s_k| \leqslant \rho$:

$$(E - \tau B^{-\frac{1}{2}}AB^{-\frac{1}{2}})e_k = s_k e_k,\ e_k \neq 0,\ k = 1, \dots, m$$

Подействуем оператором $B^{\frac{1}{2}}$ слева : $(B^{\frac{1}{2}} - \tau AB^{-\frac{1}{2}})e_k = s_k B^{\frac{1}{2}}e_k$. Обозначая $B^{-\frac{1}{2}}e_k = y$, перепишем задачу в виде: $(B - \tau A)y = s_k By$.
$\tau Ay = (1 - s_k)By$ или $Ay = \dfrac{1 - s_k}{\tau}By$
Свяжем полученное с условием (5) :

$$(Ay, y) = \frac{1 - s_k}{\tau}(By, y)\ ,\ \frac{1 - \rho}{\tau}(By, y) \leqslant \frac{1 - s_k}{\tau}(By, y) \leqslant \frac{1 + \rho}{\tau}(By, y), \quad (By, y) > 0$$

Откуда следует, что $1 - \rho \leqslant 1 - s_k \leqslant 1 + \rho$ и $|s_k| \leqslant \rho\ \ \forall k = \overline{1, m}$. Значит все значения спектра матрицы S не превосходят по модулю ρ.

Замечание. *Если $D = D^*$, то существует ортонормированный базис из собственных векторов: $De_k = d_k e_k, \quad \forall k = \overline{1, m}, \quad (e_k, e_l) = \delta_{kl} = \begin{cases} 0, & \text{если } k \neq l; \\ 1, & \text{если } k = l. \end{cases}$*

Тогда любой вектор $x \in H$ разложим по базису однозначно: $x = \sum\limits_{i=1}^{m} c_i e_i$ и верно равенство Парсеваля : $\|x\|^2 = \sum\limits_{i=1}^{m} {c_i}^2$ — сумма коэффициентов Фурье.

Найдем оценку для z^{n+1}.
$z^{n+1} = Sz^n \ \Rightarrow\ z^n = \sum\limits_{k=1}^{m} c_k^{(n)} e_k \ \Rightarrow\ z^{n+1} = \sum\limits_{k=1}^{m} c_k^{(n)} Se_k = \sum\limits_{k=1}^{m} c_k^{(n)} s_k e_k.$
Согласно равенству Парсеваля:

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

И по доказанному выше, получаем:

$$\|z^{n+1}\|^2 \leqslant \rho^2 \sum_{k=1}^{m} (c_k^{(n)})^2 = \rho^2 \|z^n\|^2 \;\Rightarrow\; \|z^{n+1}\| \leqslant \rho \|z^n\|$$

■

Следствие 1. *Пусть* $A = A^*,\ A > 0,\ B = B^*,\ B > 0, u\ \exists \gamma_2 > \gamma_1 > 0 : \gamma_1 B < A < \gamma_2 B$. *Тогда если* $\tau = \tau_0 = \dfrac{2}{\gamma_1 + \gamma_2}$, *то выполняется оценка (6)* $\|V^{n+1}\|_B \leqslant \rho \|V^n\|_B$, *где* $\rho = \dfrac{1-\xi}{1+\xi}$, $\xi = \dfrac{\gamma_1}{\gamma_2}$

Доказательство: Если $\dfrac{1+\rho}{\tau} = \gamma_2$ и $\dfrac{1-\rho}{\tau} = \gamma_1$ — условия соответствуют теореме. Вычитая и складывая ,получим :

$$\begin{cases} \frac{2}{\tau} = \gamma_1 + \gamma_2 \\ \frac{2\rho}{\tau} = \gamma_2 - \gamma_1 \end{cases} \Rightarrow \rho = \frac{\gamma_2 - \gamma_1}{\gamma_1 + \gamma_2}$$

Так как $\gamma_2 \neq 0$

$$\rho = \frac{1 - \dfrac{\gamma_1}{\gamma_2}}{1 + \dfrac{\gamma_1}{\gamma_2}} \quad \text{и полагая} \quad \xi = \frac{\gamma_1}{\gamma_2}\ , \quad \text{получаем условия следствия.}$$

■

Следствие 2. *Для метода простой итерации*

$$\frac{x^{n+1} - x^n}{\tau} + Ax^n = f\ ,\ n = 1, 2, 3, \dots\ ,\ x^0 - \textit{задано}$$

Пусть $A = A^*,\ A > 0$ *и* $\gamma_1 = \min(\lambda_k^A)$, *а* $\gamma_2 = \max(\lambda_k^A)$, $\forall k = \overline{1, m}$. *Тогда если* $\tau = \dfrac{2}{\gamma_2 + \gamma_1}$, *то* $\|V^{n+1}\| \leqslant \rho \|V^n\|$, *положив* $\rho = \dfrac{1-\xi}{1+\xi}$, $\xi = \dfrac{\gamma_1}{\gamma_2}$

Доказательство: Воспользуемся следствием 1: операторное неравенство перейдет в $\gamma_1 E \leqslant A \leqslant \gamma_2 E$. Далее аналогично следствию 1 (при $B = E$). ■

§8 Исследование сходимости попеременно треугольного итерационного метода

Рассмотрим матричное уравнение вида

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

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

Матрица A имеет структуру $A = R_1 + R_2$, где

$$R_1 = \begin{pmatrix} 0.5a_{11} & 0 & \dots & 0 & 0 \\ a_{21} & 0.5a_{22} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{m-1,m} & 0.5a_{mm} \end{pmatrix},$$

$$R_2 = \begin{pmatrix} 0.5a_{11} & a_{12} & \dots & a_{1,m-1} & a_{1m} \\ 0 & 0.5a_{22} & \dots & a_{2,m-1} & a_{2m} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0.5a_{mm} \end{pmatrix}$$

Тогда ПТИМ записывается в виде:

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

где $n = 0, 1, 2, 3, \dots,\ \ \tau > 0,\ \ \omega > 0,\ \ x^0$ - задано.(τ и ω — итерационные параметры, а x^0 — начальное условие)

Метод неявный, так как $B = (E + \omega R_1)(E + \omega R_2)$.

Теорема 4 (о достаточных условиях сходимости ПТИМ)**.** *Пусть $A = A^* > 0,\ w > \frac{\tau}{4}$, тогда итерационный метод (2), решение задачи (1), сходится при $\forall x^0$ в среднеквадратичной норме.*

Доказательство: Из самой формулировки теоремы видно, что необходимо использовать теорему Самарского. Из того, что $R_1 = R_2^*$ получаем

$$B = (E+\omega R_2^*)(E+\omega R_2) = E+\omega(R_2^*+R_2)+w^2R_2^*R_2 = \{A = R_2^*+R_2\} = E+\omega A+\omega^2 R_2^*R_2$$

А с другой стороны, мы можем записать B так:

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

Теперь осталось показать, что $(E - \omega R_2^*)(E - \omega R_2)$ неотрицательная величина, и тогда мы по теореме Самарсокого сможем связать B и A.

Обозначим $E - \omega R_2 = C \Longrightarrow C^* = E - \omega R_2^*$. Откуда получаем, что скалярное произведение $(C^*Cx, x) = (Cx, Cx) \geqslant 0$. Из этой оценки можем записать, что $B \geqslant 2\omega A$, а учитывая $B - 0.5\tau A > 0$, получим, что $B \geqslant 2\omega A > 0.5\tau A$.

Из этого получаем оценку

$$2\omega > 0.5\tau \Longrightarrow w > \frac{\tau}{4}$$

Значит при выполнении этого условия, справедлива теорема Самарского. ■

Теорема 5 (об оценке скорости сходимости ПТИМ)**.** *Путь* $A = A^* > 0$, *пусть* $\exists\ \delta > 0,\ \Delta > 0$, *такие что выполняется операторные неравенства:*

$$A \geqslant \delta E,\ R_2^* R_2 \leqslant \frac{\Delta}{4} A \tag{3}$$

Положим $\omega = \dfrac{2}{\sqrt{\delta\Delta}}$, $\tau = \dfrac{2}{\gamma_1 + \gamma_2}$, *где*

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

Тогда для ПТИМ (2) имеет место оценка в энергетической норме B

$$\|V^{n+1}\|_B \leqslant \rho \|V^n\|_B, \tag{5}$$

где

$$\rho = \frac{1 - \sqrt{\eta}}{1 + 3\sqrt{\eta}}, \quad \eta = \frac{\delta}{\Delta}, \tag{6}$$

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

Доказательство: Для сходимости необходимо, чтобы ρ была меньше единицы, а для этого необходимо, чтобы η была меньше единицы, в противном случае оценка (5) теряет общий смысл. Откуда получаем, что должно выполняться:

$$\delta \leqslant \Delta.$$

Убедимся в этом.

Из условия (5) получим, что

$$\|x^{n+1} - x\|_B \ \leqslant \rho(\omega) \|x^n - x\|_B,$$

где

$$\rho(\omega) = \frac{1 - \xi(\omega)}{1 + \xi(\omega)}, \quad \xi(\omega) = \frac{\gamma_1(\omega)}{\gamma_2(\omega)}.$$

Из (3) видно, что

$$(Ax, x) \geqslant \delta(x, x) = \delta\|x\|^2,$$

$$(R_2^* R_2 x, x) \leqslant \frac{\Delta}{4}(Ax, x).$$

А так как $(R_2^* R_2 x, x) = (R_2 x, R_2 x) = \|R_2 x\|^2$, то $\|R_2 x\|^2 \leq \dfrac{\Delta}{4}(Ax, x)$. Противоречий не возникает, так как A положительно определена.

Поскольку $A = R_1 + R_2$, то с учетом того, что мы работаем в вещественном пространстве, получим что

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

Таким образом,

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

Теперь воспользуемся неравенством Коши-Буняковского:

$$\frac{4(R_2x,x)^2}{(Ax,x)} \leqslant \frac{4\dfrac{\Delta}{4}(Ax,x)}{(Ax,x)}\|x\|^2 = \Delta\|x\|^2.$$

Откуда получаем, что $\delta \leqslant \Delta$. Исходя из следствия (1) §7, будем подбирать параметр ω такой, чтобы минимизировать $\rho(\omega)$. Рассмотрим функцию $f(\omega) = \dfrac{\gamma_2(\omega)}{\gamma_1(\omega)}$. Из ранее доказанной теоремы $B \geqslant 2\omega A$, следовательно $A \leqslant \dfrac{1}{2\omega}B$. Откуда ясно, что $\gamma_2(\omega) = \dfrac{1}{2\omega}$.

Так как,

$$B = E + \omega A + \omega^2 R_2^* R_2 \leqslant \frac{1}{\delta}A + \omega A + \omega^2\frac{\Delta}{4}A = (\frac{1}{\delta} + \omega + \frac{\omega^2\Delta}{4})A.$$

Получаем, что $\gamma_1(\omega) = \left(\dfrac{1}{\delta} + \omega + \dfrac{\omega^2\Delta}{4}\right)^{-1}$. Перейдем к минимизации $\rho(\omega)$. Она достигает минимум, тогда достигает минимум функции $f(\omega)$:

$$f(\omega) = \frac{\dfrac{1}{\delta} + \omega + \dfrac{\Delta}{4}\omega^2}{2\omega} = \frac{1}{2}\left(1 + \frac{1}{\delta\omega} + \frac{\Delta}{4}\omega\right).$$

Найдем производную:

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

Видно, что $f'(\omega) = 0$ при $\dfrac{\Delta}{4} = \dfrac{1}{\delta\omega^2}$. И отсюда $\omega_0 = \omega = \dfrac{2}{\sqrt{\delta\Delta}}$.

$$f''(\omega) = \frac{1}{2}\frac{2}{\delta\omega^3} > 0$$

Поскольку $f''(\omega) > 0$, то при $\omega = \omega_0$ достигается минимум $f(\omega)$, следовательно и $\rho(\omega)$. Подставляя найденное ω в $\gamma_2(\omega)$ получим, что

$$\gamma_2(\omega) = \frac{\sqrt{\delta\Delta}}{4}.$$

Теперь вычислим $\gamma_1(\omega)$ и $\rho(\omega)$:

$$\gamma_1(\omega) = \frac{1}{\dfrac{1}{\delta} + \omega + \dfrac{\Delta}{4}\omega^2} = \frac{1}{\dfrac{1}{\delta} + \dfrac{2}{\sqrt{\delta\Delta}} + \dfrac{\Delta}{4}\dfrac{4}{\delta\Delta}} = \frac{\sqrt{\delta}}{2}\frac{\sqrt{\delta\Delta}}{\sqrt{\delta} + \sqrt{\Delta}},$$

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

$$\rho(\omega) = \frac{1-\xi(\omega)}{1+\xi(\omega)} = \frac{1-\sqrt{\dfrac{\delta}{\Delta}}}{1+3\sqrt{\dfrac{\delta}{\Delta}}} = \left\{\eta = \frac{\delta}{\Delta}\right\} = \frac{1-\sqrt{\eta}}{1+3\sqrt{\eta}}.$$

■

$n_0(\varepsilon) = \left[ln\frac{1}{\varepsilon}/ln\frac{1}{\rho}\right]$ — число итераций, необходимое для достижения точности ε.

В реальных задачах часто $\eta = O(m^{-2})$.

Теперь посчитаем, сколько необходимо итераций, чтобы получить решение с точностью ε, для этого оценим величину $\dfrac{1}{\rho}$:

$$\frac{1}{\rho} = \frac{1+3\sqrt{\eta}}{1-\sqrt{\eta}} = \frac{(1+3\sqrt{\eta})(1+\sqrt{\eta})}{1-\eta} \approx 1+4\sqrt{\eta}.$$

Тогда $ln\dfrac{1}{\rho} \approx \sqrt{\eta}$. Итак, получим, что $n_0(\varepsilon) \approx \dfrac{1}{\sqrt{\eta}} = O(m)$.

Для сравнения рассмотрим метод простой итерации(МПИ):

$$\frac{x^{n+1}-x^n}{\tau} + Ax^n = f, \quad \tau > 0, \quad n = 0,1,2,3,\ldots$$

Из следствия 2 §7 справедлива оценка скорости сходимости

$$\|x^{n+1}-x\| \leqslant \rho\|x^n - x\|,$$

где

$$\rho = \frac{1-\xi}{1+\xi}, \quad \xi = \frac{\gamma_1}{\gamma_2}, \quad \gamma_1 = \min_k \lambda_k^A, \quad \gamma_2 = \max_k \lambda_k^A.$$

Здесь $\xi = \eta$, а $\eta = O(m^{-2})$.

Найдем число итераций.

$$\frac{1}{\rho} = \frac{1+\xi}{1-\xi} = \frac{1+\eta^2}{1-\eta^2} \approx 1+2\eta.$$

Тогда $ln\dfrac{1}{\rho} \approx \eta$, следовательно $n_0(\varepsilon) \approx \dfrac{1}{\eta} = O(m^2)$.

§9 Методы решения задач на собственные значения

Пусть A — произвольная матрица размера $(m \times m)$. Рассмотрим задачу на собственные значения матрицы A: найти такие λ и нетривиальные x, удовлетворяющие

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

где λ называется собственным значением, а x — собственным вектором.

Для нахождения собственных значений матрицы A необходимо решить характеристическое уравнение $f(\lambda) = |A - \lambda E| = 0$ — многочлен от λ степени m. При $m \geqslant 5$ задача не разрешима в общем случае.

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.

Степенной метод

Это итерационный метод решения частичной проблемы собственных значений. Обозначим x_n — n-я итерация С.В., x_0 — начальное приближение.

$$x_{n+1} = Ax_n, \tag{1}$$

где $n = 0, 1, 2 \ldots,$ x_0 - задано.

Рассматривая формулу (1) как рекуррентную, получим:

$$x_n = A^n x_0 \tag{2}$$

Теперь сделаем допущения при доказательстве. Докажем, что степенной метод будет давать вектор, сходящийся при $n \to \infty$ по направлению к вектору, отвечающему максимальному по модулю собственному значению. Упорядочим собственные значения в порядке неубывания модулей.

$$|\lambda_1| \leqslant |\lambda_2| \leqslant \ldots \leqslant |\lambda_m|$$

Для доказательства сходимости степенного метода сделаем следующие предположения:

1. Пусть у матрицы A есть базис из собственных векторов (не обязательно ортонормированный):

   $$\{e_i\}^m,$$

   или

   $$Ae_i = \lambda_i e_i, \quad e_i \neq 0 \tag{3}$$

   Следовательно, любое начальное приближение можно представить в виде разложения по этому базису:

   $$x_0 = c_1 e_1 + c_2 e_2 + \ldots + c_m e_m$$

2. Пусть в представлении начального приближения $c_m \neq 0$

3. Пусть $\left|\frac{\lambda_{m-1}}{\lambda_m}\right| < 1$

Утверждение. *Пусть для матрицы A выполнены условия 1, 2 и 3, тогда степенной метод сходится по направлению к собственному вектору, отвечающему максимальному собственному значению:*

$$x_n \longrightarrow e_m, \quad n \to \infty$$

Доказательство: Из формулы (2) следует:

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

Далее поделим обе части на $\lambda_m^n c_m \neq 0$:

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

Устремим n к бесконечности. Ясно, что тогда все отношения, вида $\dfrac{\lambda_i}{\lambda_m}$ будут равны нулю, откуда получим, что

$$x_n \longrightarrow e_m \quad \text{по направлению}, \quad n \to \infty$$

■

Степенной метод позволяет найти максимальное по модулю собственное значение.

Докажем, что

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

Возьмем две соседние итерации:

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

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

Поделим их друг на друга:

$$\lambda_m^{(n)} = \frac{x_{n+1}^{(i)}}{x_n^{(i)}} = \frac{c_1\lambda_1^{n+1}e_1^{(i)} + c_2\lambda_2^{n+1}e_2^{(i)} + \ldots + c_m\lambda_m^{n+1}e_m^{(i)}}{c_1\lambda_1^n e_1^{(i)} + c_2\lambda_2^n e_2^{(i)} + \ldots + c_m\lambda_m^n e_m^{(i)}} =$$
$$= \frac{\lambda_m^{n+1}c_m e_m^{(i)}\left(1 + \frac{c_{m-1}}{c_m}\left(\frac{\lambda_{m-1}}{\lambda_m}\right)^{n+1}\frac{e_{m-1}^{(i)}}{e_m^{(i)}} + \ldots + \frac{c_1}{c_m}\left(\frac{\lambda_1}{\lambda_m}\right)^{n+1}\frac{e_1^{(i)}}{e_m^{(i)}}\right)}{\lambda_m^n c_m e_m^{(i)}\left(1 + \frac{c_{m-1}}{c_m}\left(\frac{\lambda_{m-1}}{\lambda_m}\right)^{n}\frac{e_{m-1}^{(i)}}{e_m^{(i)}} + \ldots + \frac{c_1}{c_m}\left(\frac{\lambda_1}{\lambda_m}\right)^{n}\frac{e_1^{(i)}}{e_m^{(i)}}\right)} =$$
$$= \left\{\text{поделим числитель и знаменатель на } \lambda_m^n\right\} = \lambda_m + O\left(\frac{\lambda_{m-1}}{\lambda_m}\right)^n.$$

Откуда получаем:

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

или

$$\lim_{n\to\infty} \frac{x_{n+1}^{(i)}}{x_n^{(i)}} = \lambda_m, \quad \forall i = \overline{1,m}$$

Таким образом доказано, что для любой матрицы, удовлетворяющей условиям 1, 2 и 3, степенной метод сходится. Этот же метод позволяет найти максимальное по модулю собственное значение.

Но это не единственный способ нахождения максимального по модулю собственного значения с использованием этого метода. Покажем, что его можно найти из других соображений, причем, если взять самосопряженную матрицу, то сходимость метода будет более быстрой.

Утверждение. *Максимальное по модулю собственное значение может быть найдено следующим образом:*

$$\lambda_m^{(n)} = \frac{(Ax_n, x_n)}{(x_n, x_n)} \tag{4}$$

Доказательство:

Рассмотрим задачу для двух видов базисов из собственных векторов: ортонормированного и не ортонормированного.

Случай первый.

Пусть

$$A = A^*.$$

Тогда

$$Ae_k = \lambda_k e_k, \quad e_k \neq 0, \quad \forall k = \overline{1,m}.$$

Причем, $\{e_i\}_1^m$ — ортонормированный базис из собственных векторов, то есть $(e_i, e_j) = \delta_{ij}$. Тогда

$$\lambda_m^{(n)} = \frac{(x_{n+1}, x_n)}{(x_n, x_n)} = \frac{c_m^2\lambda_m^{2n+1} + c_{m-1}^2\lambda_{m-1}^{2n+1} + \ldots + c_1^2\lambda_1^{2n+1}}{c_m^2\lambda_m^{2n} + c_{m-1}^2\lambda_{m-1}^{2n} + \ldots + c_1^2\lambda_1^{2n}} = \Big\{\text{воспользуемся условием 1}\Big\} =$$

$$= \frac{\lambda_m^{2n+1}c_m^2\left(1 + \left(\frac{c_{m-1}}{c_m}\right)^2\left(\frac{\lambda_{m-1}}{\lambda_m}\right)^{2n+1} + \ldots + \left(\frac{c_1}{c_m}\right)^2\left(\frac{\lambda_1}{\lambda_m}\right)^{2n+1}\right)}{\lambda_m^{2n}c_m^2\left(1 + \left(\frac{c_{m-1}}{c_m}\right)^2\left(\frac{\lambda_{m-1}}{\lambda_m}\right)^{2n} + \ldots + \left(\frac{c_1}{c_m}\right)^2\left(\frac{\lambda_1}{\lambda_m}\right)^{2n}\right)} = \lambda_m + O\left(\frac{\lambda_{m-1}}{\lambda_m}\right)^{2n}$$

$$\Downarrow$$

$$\lambda_m^{(n)} - \lambda_m = O\left(\frac{\lambda_{m-1}}{\lambda_m}\right)^{2n}$$

Второй случай.

Пусть $\{e_i\}_1^m$ — базис из собственных векторов (не ортонормированных).

$$\lambda_m^{(n)} = \frac{(Ax_n, x_n)}{(x_n, x_n)} = \frac{\sum\limits_{i,j=1}^{m} c_i c_j \lambda_i^{n+1} \lambda_j^n (e_i, e_j)}{\sum\limits_{i,j=1}^{m} c_i c_j \lambda_i^n \lambda_j^n (e_i, e_j)} =$$

$$= \frac{\lambda_m^{2n+1} c_m^2 (e_m, e_m) + \lambda_m^{n+1} \lambda_{m-1}^n c_{m-1} c_m (e_{m-1}, e_m) + \ldots + c_1^2 \lambda_1^{2n+1} (e_1, e_1)}{\lambda_m^{2n} c_m^2 (e_m, e_m) + \lambda_m^n \lambda_{m-1}^n c_{m-1} c_m (e_{m-1}, e_m) + \ldots + c_1^2 \lambda_1^{2n} (e_1, e_1)} =$$

$$= \lambda_m \frac{1 + \left(\frac{\lambda_{m-1}}{\lambda_m}\right)^{n+1} \left(\frac{c_{m-1}}{c_m}\right) \frac{(e_{m-1}, e_{m-1})}{(e_m, e_m)} + \ldots + \left(\frac{\lambda_1}{\lambda_m}\right)^{2n+1} \left(\frac{c_1}{c_m}\right)^2 \frac{(e_1, e_1)}{(e_m, e_m)}}{1 + \left(\frac{\lambda_{m-1}}{\lambda_m}\right)^{n} \left(\frac{c_{m-1}}{c_m}\right) \frac{(e_{m-1}, e_{m-1})}{(e_m, e_m)} + \ldots + \left(\frac{\lambda_1}{\lambda_m}\right)^{2n} \left(\frac{c_1}{c_m}\right)^2 \frac{(e_1, e_1)}{(e_m, e_m)}}$$

Тогда

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

■

Замечание. *Покажем, что если матрица A вещественная, а собственные значения комплексные, то собственный вектор, отвечающий этому собственному значению, должен быть комплексным:*

$$\lambda = \lambda_0 + i\lambda_1, \quad \lambda_1 \neq 0.$$

Тогда в качестве начального приближения нужно брать комплексный вектор.

Пусть $x = \mu_0 + i\mu_1$ — собственный вектор, где μ_0, μ_1 — вещественные. Покажем, что $\mu_1 \neq 0$:

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

$$\Downarrow$$

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

Пусть $\mu_1 = 0$, тогда

$$A\mu_1 = \lambda_1\mu_0 + \lambda_0\mu_1 \Longrightarrow \mu_0 = 0 \Longrightarrow x \equiv 0.$$

Получили противоречие, а значит $\mu_1 \neq 0$.

Метод обратных итераций

Метод заключается в нахождении минимального по модулю собственного значения.

Изначально нужно наложить ограничения на матрицу. Рассмотрим матрицу, для которой существует обратная. Тогда организуем итерационный метод по следующей формуле

$$Ax_{n+1} = x_n, \tag{5}$$

где $n = 0, 1, 2, \ldots,\quad x_0$ — выбираем.

Метод, задаваемый формулой (5) — неявный. Так как $\exists A^{-1}$, то

$$x_{n+1} = A^{-1}x_n \tag{6}$$

Таким образом метод превратился в степенной, но для матрицы A^{-1}.

Мы знаем, что у невырожденной матрицы собственные значения связаны с собственными значениями обратной матрицы как обратные числа. А собственные вектора у прямой и обратной матрицы совпадают.

$$\lambda_k^{A^{-1}} = \frac{1}{\lambda_k^A}, \quad k = \overline{1,m}, \quad |\lambda_1| \leqslant |\lambda_2| \leqslant \cdots \leqslant |\lambda_m|,\ \lambda_k = \lambda_k^A$$

Пусть выполнены условия:

1. $\{e_i\}_1^m$ — базис из собственных векторов
2. $x_0 = c_1e_1 + c_2e_2 + \cdots + c_me_m, \quad$ где $c_1 \neq 0$
3. $\left|\frac{\lambda_1}{\lambda_2}\right| < 1$

Так как

$$A^{-1}e_k = \frac{1}{\lambda_k}e_k, \quad e_k \neq 0, \quad \forall k = \overline{1,m}\ ,$$

то, записав вектор x_0 по базису, применим A^{-1}:

$$x_n = (A^{-1})^n x_0 = \sum_{k=1}^{m} c_k\lambda_k^{-n}e_k = c_1\lambda_1^{-n}e_1 + \ldots + c_m\lambda_m^{-n}e_m$$

$$\Downarrow$$

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

Получим, что

$$x_n \longrightarrow e_1, \quad n \to \infty$$

Задача. *Доказать, что* $\lambda_1^{(n)} - \lambda_1 = O\left(\frac{\lambda_1}{\lambda_2}\right)^n$, *где* $\lambda_1^{(n)} = \frac{x_n^{(i)}}{x_{n+1}^{(i)}},\ i = \overline{1,m}$

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

$$x_n^{(i)} = c_1\lambda_1^{-n}e_1^{(i)} + c_2\lambda_2^{-n}e_2^{(i)} + \cdots + c_m\lambda_m^{-n}e_m^{(i)} \tag{1}$$

$$x_{n+1}^{(i)} = c_1\lambda_1^{-n-1}e_1^{(i)} + c_2\lambda_2^{-n-1}e_2^{(i)} + \cdots + c_m\lambda_m^{-n-1}e_m^{(i)} \tag{2}$$

Теперь поделим (1) на (2):

$$\frac{x_n^{(i)}}{x_{n+1}^{(i)}} = \frac{c_1\lambda_1^{-n}e_1^{(i)}\left(1 + \frac{c_2e_2^{(i)}}{c_1e_1^{(i)}}\left(\frac{\lambda_2}{\lambda_1}\right)^{-n} + \cdots + \frac{c_me_m^{(i)}}{c_1e_1^{(i)}}\left(\frac{\lambda_m}{\lambda_1}\right)^{-n}\right)}{c_1\lambda_1^{-n-1}e_1^{(i)}\left(1 + \frac{c_2e_2^{(i)}}{c_1e_1^{(i)}}\left(\frac{\lambda_2}{\lambda_1}\right)^{-n-1} + \cdots + \frac{c_me_m^{(i)}}{c_1e_1^{(i)}}\left(\frac{\lambda_m}{\lambda_1}\right)^{-n-1}\right)} =$$

$$= \lambda_1 + O\left(\frac{\lambda_1}{\lambda_2}\right)^n = \lambda_1^n$$

Что и требовалось показать. □

Задача. *Пусть* $A = A^*$, *существует* A^{-1},

$$\lambda_1^{(n)} = \frac{(x_n, x_n)}{(x_{n+1}, x_n)}$$

Доказать, что

$$\lambda_1^{(n)} - \lambda_1 = O\left(\frac{\lambda_1}{\lambda_2}\right)^{2n}$$

Решение: Найдем $\lambda_1^{(n)}$, учитывая ортонормированность базиса из собственных векторов:

$$\lambda_1^{(n)} = \frac{(x_n, x_n)}{(x_{n+1}, x_n)} = \frac{c_1^2\lambda_1^{-2n} + c_2^2\lambda_2^{-2n} + \cdots + c_m^2\lambda_m^{-2n}}{c_1^2\lambda_1^{-2n-1} + c_2^2\lambda_2^{-2n-1} + \cdots + c_m^2\lambda_m^{-2n-1}} =$$

$$= \frac{c_1^2\lambda_1^{-2n}\left(1 + \left(\frac{c_2}{c_1}\right)^2\left(\frac{\lambda_2}{\lambda_1}\right)^{-2n} + \cdots + \left(\frac{c_m}{c_1}\right)^2\left(\frac{\lambda_m}{\lambda_1}\right)^{-2n}\right)}{c_1^2\lambda_1^{-2n-1}\left(1 + \left(\frac{c_2}{c_1}\right)^2\left(\frac{\lambda_2}{\lambda_1}\right)^{-2n-1} + \cdots + \left(\frac{c_m}{c_1}\right)^2\left(\frac{\lambda_m}{\lambda_1}\right)^{-2n-1}\right)} =$$

$$= \lambda_1 + O\left(\frac{\lambda_1}{\lambda_2}\right)^{2n}$$

Что и требовалось показать. □

Метод обратных итераций со сдвигом

Рассмотрим метод обратных итераций со сдвигом:

$$(A - \alpha E)x_{n+1} = x_n, \tag{6}$$

где $n = 0, 1, \ldots,$ x_0 — задано и α такое число, что

$$\exists (A - \alpha E)^{-1}$$

Тогда ясно, что

$$x_{n+1} = (A - \alpha E)^{-1}x_n \tag{7}$$

Обозначим $B = (A - \alpha E)^{-1}$.

Так как для этой матрицы B метод стал степенным, то мы можем сделать вывод о том, куда сходятся итерации.

$$x_n \to e_l$$

$$\max_{1 \leqslant k \leqslant m} \frac{1}{|\lambda_k - \alpha|} = \frac{1}{|\lambda_l - \alpha|}$$

Здесь точно так же, как и в степенном методе, можно найти собственное значение

$$\lambda_l = \alpha + \lim_{n\to\infty} \frac{x_n^{(i)}}{x_{n+1}^{(i)}}$$

Казалось бы, что тогда этим методом можно найти все собственные значения матрицы A, но он для этого не применяется. Метод обратных итераций со сдвигом применяется при уточнении собственных векторов и собственных значений.

§10 Приведение матрицы к верхней почти треугольной форме

Берем произвольную матрицу A порядка m. Задача поиска всего спектра — очень важная и сложная задача. Очевидно, что лучше всего матрицу A свести к матрице C, называемой предельной, у которой все собственные значения сразу можно найти (это матрица треугольного вида, относительно главной диагонали или диагональная). Для этого можно было бы воспользоваться методом Гаусса, однако приведение матрицы A к треугольной форме методом Гаусса не сохраняет спектр матрицы. Поэтому лучше воспользоваться преобразованием подобия. Если

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

тогда у матриц C и A совпадают собственные значения.

Под верхней почти треугольной матрицей (ВПТМ) понимается матрица вида

$$\begin{pmatrix} \times & \times & \dots & \times & \times \\ \times & \times & \dots & \times & \times \\ 0 & \times & \dots & \times & \times \\ 0 & 0 & \dots & \times & \times \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \times & \times \end{pmatrix}, \tag{$*$}$$

где под $\times$ подразумевается, вообще говоря, ненулевой элемент.

Первым делом покажем, что любую матрицу с помощью преобразований элементарных отражений можно привести к виду $(*)$. А далее можно организовать итерационный процесс, приводящий матрицу к трехдиагональной структуре.

Напомним, что ортогональная матрица — это такая матрица S, для которой выполнено:

$$S^{-1} = S^T$$

Рассмотрим вектор $V = (V_1, V_2, \dots, V_m)^T$

Определение. *Элементарным отражением, который соответствует вектор-столбцу V, называется преобразование, задаваемое матрицей*

$$H = E - \frac{2VV^T}{\|V\|^2} \tag{2}$$

У этого преобразования есть три очень важных свойства. Во-первых, оно симметричное, во-вторых, ортогональное и, наконец, имеет совершенно уникальное свойство заключающееся в том, что действуя на произвольный вектор, преобразование обнуляет все его координаты, кроме первой.

Во-первых,

$$\|V\|^2 = V^T V = V_1^2 + V_2^2 + \cdots + V_m^2$$

$$VV^T = \begin{pmatrix} V_1^2 & V_1V_2 & \dots & V_1V_m \\ V_2V_1 & V_2^2 & \dots & V_2V_m \\ \dots & \dots & \dots & \\ V_mV_1 & V_mV_2 & \dots & V_m^2 \end{pmatrix}$$

Видно, что полученная матрица симметричная. Так как второе слагаемое E в равенстве (2) тоже симметричное, то:

$$H^T = H$$

Теперь докажем второе свойство, утверждающее, что матрица H является ортогональной:

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

Для этого посчитаем произведение $H^T H$. Если мы покажем, что это единичная матрица, то тогда, по определению обратной матрицы, H^T будет совпадать с H^{-1}.

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

$$= E - 4\frac{VV^T}{\|V\|^2} + 4\frac{VV^TVV^T}{\|V\|^4} = \left\{\text{в силу ассоциативности} \quad V^TV = \|V\|^2\right\} = E$$

Утверждение. *Пусть задан произвольный вектор $x = (x_1, x_2, \dots, x_m)^T$, тогда можно выбрать вектор V так, что*

$$Hx = (-\sigma, 0, 0, \dots, 0)^T, \quad \sigma = \|x\| = \left(\sum_{i=1}^{m} x_i^2\right)^{\frac{1}{2}}.$$

Доказательство: Возьмем вектор

$$V = x + \sigma z,$$

где $z = (1, 0, 0, \dots, 0)^T$.

Тогда образ вектора x будет равен

$$Hx = x - \frac{2(x+\sigma z)(x+\sigma z)^T}{(x+\sigma z)^T(x+\sigma z)}\, x = x - (x+\sigma z)\,\frac{2(x+\sigma z)^T x}{(x+\sigma z)^T(x+\sigma z)}$$

Посчитаем числитель и знаменатель дроби по отдельности.

Числитель:

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

Знаменатель:

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

Достаточно убедиться, что, если мы положим σ равной

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

то получим, что числитель равен

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

а знаменатель

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

Получаем, что числитель совпадает со знаменателем, а значит вся дробь обращается в единицу.

Тогда

$$Hx = x - x - \sigma z = (-\sigma, 0, 0, \ldots, 0)^T$$

■

Последнее свойство позволяет привести совершенно произвольную матрицу к верхней почти треугольной форме.

Запишем матрицу A блочно

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

где

$$y_{m-1} = (a_{12}, a_{13}, \ldots, a_{1m}),$$
$$x_{m-1} = (a_{21}, a_{31}, \ldots, a_{m1})^T.$$

По доказанному утверждению, для вектора x_{m-1} выбираем вектор V:

$$V = x_{m-1} + \sigma z_{m-1},$$

$$\|x\| = \sigma, \quad z_{m-1} = \underbrace{(1,0,\ldots,0)^T}_{m-1}$$

Выбрав вектор V таким образом, построим матрицу элементарных отражений:

$$H_{m-1}x_{m-1} = -\sigma z_{m-1} = (-\sigma, 0, 0, \ldots, 0)^T$$

Но матрица H имеет порядок $(m-1)$. Её умножать на матрицу A порядка m мы не можем. Поэтому построим матрицу U_1:

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

где

$$o_{12} = (0, \ldots, 0),$$
$$o_{21} = (0, \ldots, 0)^T.$$

Проверим, сохранила ли матрица U_1 те три свойства, которыми обладало преобразование элементарного отражения.

Очевидно, что

$$U_1 = U_1^T.$$

Проверим, является ли она ортогональной.

$$U_1^2 = U_1 U_1 = \begin{pmatrix} 1 & o_{12} \\ o_{21} & H_{m-1} \end{pmatrix} \cdot \begin{pmatrix} 1 & o_{12} \\ o_{21} & H_{m-1} \end{pmatrix} = \begin{pmatrix} 1 & o_{12} \\ o_{21} & H_{m-1}^2 \end{pmatrix}$$

Так как матрица H_{m-1} ортогональная, то тогда понятно, что полученная матрица является единичной, откуда получим, что

$$U_1^{-1} = U_1^T = U_1.$$

Докажем, что умножение этой матрицы на матрицу A слева позволит обнулить в первом столбце все элементы, кроме первых двух. Домножим матрицу A слева на матрицу U_1^{-1}:

$$U_1^{-1} A = U_1 A = \begin{pmatrix} 1 & o_{12} \\ o_{21} & H_{m-1} \end{pmatrix} \cdot \begin{pmatrix} a_{11} & y_{m-1} \\ x_{m-1} & A_{m-1} \end{pmatrix} = \begin{pmatrix} a_{11} & y_{m-1} \\ H_{m-1}x_{m-1} & H_{m-1}A_{m-1} \end{pmatrix}$$

Далее домножим справа на U_1:

$$U_1^{-1} A U_1 = \begin{pmatrix} a_{11} & y_{m-1} \\ H_{m-1}x_{m-1} & H_{m-1}A_{m-1} \end{pmatrix} \cdot \begin{pmatrix} 1 & o_{12} \\ o_{21} & H_{m-1} \end{pmatrix} =$$

$$= \begin{pmatrix} a_{11} & y_{m-1}H_{m-1} \\ -\sigma_1 z_{m-1} & H_{m-1}A_{m-1}H_{m-1} \end{pmatrix} = \left\{ \text{из-за ортогональности} \quad H_{m-1} = H_{m-1}^{-1} \right\} =$$

$$= \begin{pmatrix} a_{11} & y_{m-1}H_{m-1} \\ -\sigma_1 z_{m-1} & H_{m-1}^{-1}A_{m-1}H_{m-1} \end{pmatrix} = \left\{ \sigma_1 = \|x_{m-1}\| \right\} = \begin{pmatrix} \times & \times & \times & \ldots & \times \\ \times & \times & \times & \ldots & \times \\ 0 & \times & \times & \ldots & \times \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & \times & \times & \ldots & \times \end{pmatrix}$$

Итак, мы получили матрицу C_1, такую что

$$C_1 = U_1^{-1} A U_1$$

или, другими словами, подобную матрице A, причем матрица подобия U_1 ортогональна.

Далее возьмем другой вектор размерности $(m-2)$:

$$x_{m-2} = (c_{32}^{(1)}, c_{42}^{(1)}, \dots, c_{m2}^{(1)})^T$$

по которому можно так же выбрать вектор V и построить преобразование H_{m-2}:

$$H_{m-2}x_{m-2} = -\sigma_2 z_{m-2},$$

где $\sigma_2 = \|x_{m-2}\|$

Теперь можно сказать, что вновь по H_{m-2} восстановим матрицу порядка m, добавив элементы:

$$U_2 = \begin{pmatrix} E_2 & \vdots & o_{12} \\ \dots\dots\dots\dots & & \\ o_{21} & \vdots & H_{m-2} \end{pmatrix},$$

где

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

Легко убедиться, матрица U_2 симметрична и ортогональна.

После всего этого, применим U_2 к C_1:

$$C_2 = U_2^{-1}U_1^{-1}AU_1U_2 = \begin{pmatrix} \times & \times & \dots & \times \\ \times & \times & \dots & \times \\ 0 & \times & \dots & \times \\ 0 & 0 & \dots & \times \\ \dots\dots\dots\dots & & & \\ 0 & 0 & \dots & \times \end{pmatrix}$$

Спустя $(m-2)$ шага, аналогично выстраивая U_k, получим матрицу C:

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

которая будет верхней почти треугольной формы:

$$C_{ji} = 0, \quad \forall i \geqslant j+2.$$

Обозначим

$$U = U_1U_2 \dots U_{m-2}$$

Найдем U^{-1}:

$$U^{-1} = (U_1, U_2, \dots, U_{m-2})^{-1} = U_{m-2}^{-1}U_{m-3}^{-1} \dots U_2^{-1}U_1^{-1} =$$
$$= \Big\{\text{каждый множитель ортогонален}\Big\} = U_{m-2}^T \dots U_2^T U_1^T = U^T$$

Получили, что

$$U^{-1} = U^T.$$

А значит, итоговое произведение будет иметь вид

$$C = U^{-1}AU.$$

Показано, что любую матрицу можно свести к верхней почти треугольной форме с помощью преобразования подобия с преобразующей ортогональной матрицей.

Замечание. *Собственные значения матрицы C равны собственным значениям матрицы A:*

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

Доказательство: Запишем задачу на собственные значения для одной из матриц:

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

Домножим слева на U^{-1}:

$$U^{-1}Ax = \lambda U^{-1}x.$$

Обозначим $U^{-1}x = y$ или $x = Uy$, тогда

$$\underbrace{U^{-1}AU}_{C} y = \lambda y, \quad y \neq 0$$

$$\Downarrow$$

$$Cy = \lambda y, \quad y \neq 0$$

■

Замечание. *Если матрица A симметрична (вещественный случай), то C тоже будет симметричной:*

$$A = A^T \Rightarrow C = C^T$$

Доказательство:

$$C = U^{-1}AU$$

Найдем транспонированную:

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

■

§11 Понятие QR–алгоритма. Решение полной проблемы собственных значений

Рассмотрим произвольную квадратную матрицу A порядка m. Поставим перед собой задачу: факторизовать матрицу A в виде произведения:

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

где $Q^{-1} = Q^T$ (ортогональная), R — верхнетреугольной формы.

Обозначим

$$x = (a_{11}, a_{21}, \ldots, a_{m1})^T.$$

Согласно доказанному свойству выше, возьмем вектор V:

$$V = x + \|x\| z.$$

По этому вектору построим матрицу H_1

$$H_1 = E - \frac{2VV^T}{\|V\|^2}, \quad H_1^{-1} = H_1 = H_1^T.$$

Тогда получим

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

Ясно, что проделав $(m-1)$ аналогичных шагов, можно получить верхнетреугольную матрицу.

В итоге получаем матрицу R:

$$R = H_{m-1} H_{m-2} \cdot \ldots \cdot H_2 H_1 A$$

Обозначим

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

Найдем обратную

$$Q^{-1} = (H_1 \cdot \ldots \cdot H_{m-1})^{-1} = H_{m-1}^{-1} \cdot \ldots \cdot H_1^{-1} = H_{m-1}^T \cdot \ldots \cdot H_2^T H_1^T = Q^T$$

Итак, $Q^{-1} = Q^T$ ортогональная матрица, а A представима в виде:

$$A = QR$$

Получили, что любую матрицу можно разложить в виде QR, где Q — ортогональная матрица, а R — матрица верхнетреугольной формы.

Если A — полная матрица, то разложение (1) пропорционально m^3 действий. Если матрица A имеет верхнюю почти треугольную форму, то разложение требует порядка m^2 действий. А если A — трёхдиагональная, то необходимо всего порядка m действий.

На первом этапе в QR алгоритме обозначим

$$A = A_0.$$

Вначале делается QR разложение этой матрицы:

$$A_0 = Q_0 R_0, \tag{2}$$

s где $Q_0^{-1} = Q_0^T$, а R - ВТФ.

Далее находим A_1:

$$A_1 = R_0 Q_0. \tag{3}$$

Утверждается, что она сохраняет спектр A_0, а значит и спектр A.

Тогда из (2) получаем, что

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

Затем, подставляя в (3), получим

$$A_1 = Q_0^{-1} A Q_0.$$

Откуда видно, что матрицы A_1 и A_0 подобны, следовательно

$$\lambda_k^{A_1} = \lambda_k^{A_0}, \quad k = \overline{1, m}.$$

В результате этих двух шагов получаем матрицу A_1, которая сохраняет спектр и симметрию(если она была) матрицы A_0, так как A_1 получается из неё с помощью преобразования, использующего ортогональную матрицу. Следующим шагом разложим A_1 в виде

$$A_1 = Q_1 R_1.$$

Тогда на втором шаге получим матрицы A_2:

$$A_2 = R_1 Q_1.$$

Ясно, что мы записали частный случай рекуррентного итерационного метода:

$$A_k = Q_k R_k, \tag{4}$$

где $Q_k^{-1} = Q_k^T$, R_k — ВТФ, а

$$A_{k+1} = R_k Q_k, \quad k = 0, 1, 2, \ldots \tag{5}$$

$$\lambda_n^{A_{k+1}} = \lambda_n^{A_k}, \quad n = \overline{1, m}$$

Кроме того

$$A_{k+1} = Q_k^{-1} A_k Q_k.$$

Мы построили итерационный процесс (4), (5).

В подавляющем большинстве случаев, если λ_k — вещественные, A_k сходится к матрице верхней треугольной формы:

$$\lim_{k \to \infty} A_k = \begin{pmatrix} \times & \times & \ldots & \times & \times \\ 0 & \times & \ldots & \times & \times \\ 0 & 0 & \ldots & \times & \times \\ \vdots & \vdots & \ldots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 & \times \end{pmatrix}.$$

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 (так называемый машинный ноль), определяющего точность вычислений. А на главной диагонали будут находиться собственный значения матрицы.

У вещественной матрицы могут быть и комплексные собственные значения. Тогда, в случае с комплексными с/з матрица примет вид:

$$\lim_{k\to\infty} A_k = \begin{pmatrix} \times & \dots & \dots & \dots & \dots & \dots \\ 0 & \times & \dots & \dots & \dots & \dots \\ 0 & 0 & x & y & \dots & \dots \\ 0 & 0 & -y & x & \dots & \dots \\ \vdots & \vdots & \vdots & \dots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 & \times \end{pmatrix}.$$

Теперь на диагонали появились клетки второго порядка такие, что на главной диагонали стоят вещественные части собственных значений, а под ней — мнимые.

Таким образом QR алгоритм (4)-(5) сходится в случае вещественных собственных значений к предельной верхнетреугольной матрице, а в случае комплексных собственных значений появятся на главной диагонали клетки второго порядка.

Подчеркнем тот факт, что QR алгоритм применим к любой матрице и не ухудшает верхнюю почти треугольную форму матрицы. Таким образом, для уменьшения числа итераций, матрицу сначала можно свести к верхней почти треугольной форме.

К минусам QR алгоритма можно отнести следующее:

1. надо помнить промежуточные матрицы
2. большое число действий, необходимых для получения предельных матриц.

§12 Предварительное преобразование матрицы к верхней почти треугольной форме

$$A_k = Q_k R_k \tag{1}$$

$$A_{k+1} = R_k Q_k, \quad k = 0, 1, 2, \dots \tag{2}$$

Для того, чтобы доказать, что QR алгоритм не будет портить верхнюю почти треугольную форму, докажем 2 леммы.

Лемма 1. *Пусть матрицы A, B, C — матрицы одного порядка, такие что B имеет верхнюю треугольную форму, A – верхнюю почти треугольную форму, тогда*

$$C = BA \tag{3}$$

имеет верхнюю почти треугольную форму.

Доказательство: Элементы матрицы C имеют вид

$$c_{ij} = \sum_{\alpha=1}^{m} b_{i\alpha} a_{\alpha j}$$

Теперь будем учитывать, что B имеет ВТФ, тогда $b_{i\alpha} = 0 \quad \forall i > \alpha$. Откуда получим, что

$$c_{ij} = \sum_{\alpha=i}^{m} b_{i\alpha} a_{\alpha j}.$$

Теперь учтем, что A — ВПТФ $\Rightarrow a_{\alpha j} = 0 \quad \forall \alpha > j+1$, откуда

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

Рассмотрим c_{ij}: $i > j+1$. Такие элементы, очевидно, будут равны нулю, а это и означает, что матрица C - ВПТФ. ■

Лемма 2. *Пусть матрицы A, B, C — матрицы одного порядка, такие что B имеет верхнюю треугольную форму, A – верхнюю почти треугольную форму, тогда*

$$C = AB \tag{3}$$

имеет верхнюю почти треугольную форму.

Доказательство: Элементы матрицы C имеют вид

$$c_{ij} = \sum_{\alpha=1}^{m} a_{i\alpha} b_{\alpha j}$$

Теперь будем учитывать, что B имеет ВТФ, тогда $b_{\alpha j} = 0 \quad \forall \alpha > j$. Откуда получим, что

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

Теперь учтем, что A — ВПТФ $\Rightarrow a_{i\alpha} = 0 \quad \forall i > \alpha + 1$, откуда

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

Рассмотрим c_{ij}: $i > j+1$. Такие элементы, очевидно, будут равны нулю, а это и означает, что матрица C - ВПТФ. ■

Вернемся к QR алгоритму. R_k имеет верхнюю треугольную форму. Из (1) $Q_k = A_k R_k^{-1}$. Изначально мы A_0 привели к верхней почти треугольной форме $\Rightarrow$ A_k имеет верхнюю почти треугольную форму, тогда мы получаем, что матрица Q_k имеет верхнюю почти треугольную форму. Из (2) получаем, что A_{k+1} тоже имеет верхнюю почти треугольную форму.

Глава 2

Интерполирование и приближение функций

§1 Постановка задачи интерполирования

Пусть есть функция $f(x)$ и известны ее значения в узлах $x \in [a, b]$,

$$a \leqslant x_0 < x_1 < \ldots < x_n \leqslant b$$

$\{x_i\}_0^n$ — узлы интерполирования. В этих узлах заданы значения функции

$$f(x_i) = f_i, \quad i = \overline{0, n} \tag{1}$$

Надо построить такую непрерывную функцию, которая в узлах совпадает с f_i. Эта задача имеет бесконечное число решений. Мы рассмотрим случай, когда функция $f(x)$ будет приближаться алгебраически полиномами. Пусть есть полином n-ой степени

$$P_n(x) = a_0 + a_1 x + \cdots + a_n x^n, \quad a_i \text{ – вещественные числа, } a_n \neq 0 \tag{2}$$

Нужно построить такой полином, который в узлах будет совпадать с интерполированной функцией. Он будет называться интерполяционным полиномом, если:

$$P_n(x_i) = f_i, \quad i = \overline{0, n} \tag{3}$$

Эта задача имеет всегда единственное решение. Для нахождения коэффициентов a_i из равенства (2) получаем из (3) систему:

$$\begin{cases} a_0 + a_1 x_0 + a_2 x_0^2 + \cdots + a_n x_0^n = f_0 \\ a_0 + a_1 x_1 + a_2 x_1^2 + \cdots + a_n x_1^n = f_1 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_0 + a_1 x_n + a_2 x_1^2 + \cdots + a_n x_n^n = f_n \end{cases} \tag{4}$$

Выпишем определитель матрицы из коэффициентом этой системы(определитель Вандермонда):

$$\begin{vmatrix} 1 & x_0 & x_0^2 & \cdots & x_0^n \\ 1 & x_2 & x_2^2 & \cdots & x_2^n \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^n \end{vmatrix} = \prod_{n \geqslant i > j \geqslant 0} (x_i - x_j)$$

Если узлы разные, то определитель отличен от нуля и, согласно критерию, система (4) имеет единственное решение.

Замечание. *Поскольку показано существование и единственность интерполирующего полинома, то при его поиске в любой форме он будет тождественно равен всем своим представлениям в иных формах, полученных с помощью других методов.*

§2 Интерполяционная форма Лагранжа $L_n(x)$

Пусть существуют $(n+1)$ несовпадающий узел, функция интерполирования $f(x)$, такая что

$$f(x_i) = f_i, \quad i = \overline{1, n}.$$

Мы хотим построить полином $L_n(x)$ так, что

$$L_n(x_i) = f_i. \tag{1}$$

Тогда говорим, что мы построили интерполяционный полином для функции $f(x)$ по узлам x_i.

Если мы его ищем в виде

$$L_n(x) = \sum_{k=0}^{n} c_n(x) f(x_k), \tag{2}$$

где $c_n(x)$ — полином n-ой степени, тогда говорят, что строят полином Лагранжа.

Введем полином $(n+1)$-ой степени:

$$\omega(x) = (x - x_0)(x - x_1) \cdots (x - x_n).$$

Обозначим через $[\cdot]$ все члены произведения в ω, за исключением $(x - x_k)$. Тогда

$$\omega'(x) = [\cdot] + (x - x_k)[\cdot]'. \tag{3}$$

Следовательно,

$$\omega'(x_k) = \prod_{k=0, k \neq i}^{n} (x_k - x_i).$$

Тогда из (1)

$$c_k(x) = \frac{\omega(x)}{(x - x_k)\omega'(x_k)}, \quad L_n(x) = \sum_{k=0}^{n} \frac{\omega(x) f(x_k)}{(x - x_k)\omega'(x_k)}. \tag{4}$$

Легко можно получить оценку погрешности $L_n(x)$:

$$\psi_{L_n(x)} = f(x) - L_n(x).$$

Введем функцию:

$$g(s) = f(s) - L_n(s) - k\omega(s), \quad k = const.$$

Выберем постоянную k из условия $g(x) = 0$ в данной точке $x \neq x_i$. Константа $k(x)$ получается равной $k = \dfrac{f(x) - L_n(x)}{\omega(x)}$. Затем, полагая $f(x) \in C^{n+1}[a,b]$ и применяя (n+1) раз теорему Ролля, получаем, что $\exists \xi \in [a,b]$, в которой $g^{n+1}(\xi) = 0$. Тогда очевидно что:

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

Возьмем это выражение по модулю, тогда

$$|\psi_{L_n(x)}| \leqslant \frac{M_{n+1}}{(n+1)!}|\omega(x)|,$$

где $M_{n+1} = \sup\limits_x |f^{n+1}(x)|$

Замечание. *Если интерполируемая функция $f(x)$ полином степени не выше n, то полином Лагранжа точно её приближает, так как $M_{n+1} = 0$.*

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

§3 Разделенные разности

Пусть у нас есть отрезок $[a,b]$, на котором выбраны узлы $\{x_i\}_0^n$, в которых заданы значения некоторой функции $f(x_i) = f_i,\ i = \overline{0,n}$:

$$a \leqslant x_0 < x_1 < \cdots < x_n \leqslant b$$

Определение. *Разделенной разностью 1-го порядка, построенной по узлам x_i, x_j называется отношение*

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

Узлы x_i и x_j могут быть и не соседними. Далее для простоты будем рассматривать разделенные разности, построенные по соседним узлам. По разделенной разности первого порядка, можно построить разделенную разность 2-го порядка, например:

$$f(x_0, x_1, x_2) = \frac{f(x_1, x_2) - f(x_0, x_1)}{x_2 - x_0}$$

$$f(x_1, x_2, x_3) = \frac{f(x_2, x_3) - f(x_1, x_2)}{x_3 - x_1}$$

Можно ввести разделенную разность любого порядка. Пусть задана разделенная разность k-го порядка $f_1(x_j, x_{j+1}, \ldots, x_{j+k})$ по узлам $(x_j, x_{j+1}, \ldots, x_{j+k})$ и разделенная разность k-го порядка $f_2(x_{j+1}, x_{j+2}, \ldots, x_{j+k+1})$ по узлам $(x_{j+1}, x_{j+2}, \ldots, x_{j+k+1})$, тогда разделенная разность $k+1$ порядка равна:

$$f(x_j, x_{j+1}, x_{j+2}, \ldots, x_{j+k+1}) = \frac{f_2 - f_1}{x_{j+k+1} - x_j}$$

Утверждение. *Разделенная разность $k-$ го порядка представляется в виде*

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

где $w(x) = (x - x_0)(x - x_1)\ldots(x - x_n)$. Будем обозначать через $w_{\alpha,\beta}(x)$:

$$w_{\alpha,\beta}(x) = (x - x_\alpha)(x - x_{\alpha+1})\ldots(x - x_\beta)$$

Тогда, используя равенство $w'(x_i) = w'_{0,k}(x_i) = \prod\limits_{j=0, j\neq i}^{k} (x_i - x_j)$ можно записать условие утверждения следующем образом:

$$f(x_0, x_1, \ldots, x_k) = \sum_{i=0}^{k} \frac{f(x_i)}{w'_{0,k}(x_i)} \tag{1}$$

Доказательство: Доказательство проведем методом полной математической индукции. По определению $f(x_0, x_1) = \dfrac{f(x_1) - f(x_0)}{x_1 - x_0}$. Тогда $f(x_0, x_1) = \dfrac{f(x_0)}{x_0 - x_1} + \dfrac{f(x_1)}{x_1 - x_0}$, то есть для $k = 1$ формула справедлива. Предположим, что она справедлива и для $k = l$:

$$f(x_0, x_1, \ldots, x_l) = \sum_{i=0}^{l} \frac{f(x_i)}{w'_{0,l}(x_i)}, \quad f(x_1, \ldots, x_l, x_{l+1}) = \sum_{i=1}^{l+1} \frac{f(x_i)}{w'_{1,l+1}(x_i)} \tag{2}$$

По определению составим $l+1$ разделенную разность и представим её в виде трех слагаемых:

$$f(x_0, x_1, \ldots, x_{l+1}) = \frac{f(x_0)}{(x_0 - x_{l+1})w'_{0,l}(x_0)} + \frac{f(x_{l+1})}{(x_{l+1} - x_0)w'_{1,l+1}(x_{l+1})} + \\ + \sum_{i=1}^{l} \frac{f(x_i)}{x_{l+1} - x_0} \left(\frac{1}{w'_{1,l+1}(x_i)} - \frac{1}{w'_{0,l}(x_i)} \right)$$

Очевидно, что: $(x_{l+1} - x_0)w'_{1,l+1}(x_{l+1}) = w'_{0,l+1}(x_{l+1})$ и $(x_0 - x_{l+1})w'_{0,l}(x_0) = w'_{0,l+1}(x_0)$. Далее рассмотрим выражение , стоящее под знаком суммы:

$$\frac{1}{x_{l+1} - x_0} \left(\frac{1}{w'_{1,l+1}(x_i)} - \frac{1}{w'_{0,l}(x_i)} \right)$$

Домножая числитель и знаменатель первой дроби, стоящий в скобках на $x_i - x_0$, а второй — на $x_i - x_{l+1}$, получим:

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

Очевидно, что $w'_{1,l+1}(x_{l+1})(x_i - x_0) = w'_{0,l+1}(x_i)$, $\quad w'_{0,l}(x_i)(x_i - x_{l+1}) = w'_{0,l+1}(x_i)$. Подставляя полученные выражения в исходную сумму, получаем формулу (1) для $k = l + 1$:

$$f(x_0, x_1, \ldots, x_{l+1}) = \sum_{i=0}^{l+1} \frac{f(x_i)}{w'_{0,l+1}(x_i)}$$

■

Выразим значение функции в $k-$м узле через f_0 и разделенные разности до $k-$го порядка:

$$\text{при } k = 1: \;\; f(x_0, x_1) = \frac{f(x_1) - f(x_0)}{x_1 - x_0} = \frac{f(x_0)}{x_0 - x_1} + \frac{f(x_1)}{x_1 - x_0}$$

Домножив обе части на $x_1 - x_0$ получим:

$$(x_1 - x_0)f(x_0, x_1) = f(x_1) - f(x_0) \text{ и } \quad f(x_1) = f(x_0) + (x_1 - x_0)f(x_0, x_1) \tag{3}$$

$$\text{при } \; k = 2: \;\; f(x_0, x_1, x_2) = \frac{f(x_0)}{(x_0 - x_1)(x_0 - x_2)} + \frac{f(x_1)}{(x_1 - x_0)(x_1 - x_2)} + \frac{f(x_2)}{(x_2 - x_1)(x_2 - x_0)}$$

Следовательно:

$$(x_2 - x_1)(x_2 - x_0)f(x_0, x_1, x_2) = \frac{f(x_0)(x_2 - x_1)}{x_1 - x_0} + \frac{f(x_1)(x_2 - x_0)}{x_0 - x_1} + f(x_2)$$

Подставим ранее полученное для $f(x_1)$ и объединим полученные слагаемые:

$$f(x_2) = f(x_0) + f(x_0, x_1)(x_2 - x_0) + (x_2 - x_0)(x_2 - x_1)f(x_0, x_1, x_2) \tag{4}$$

Тогда можно сказать, что

$$f(x_k) = f(x_0) + (x_k - x_0)f(x_0, x_1) + \ldots + (x_k - x_0)\ldots(x_k - x_{k-1})f(x_0, x_1, \ldots, x_k) \tag{5}$$

Замечание. *Безусловно, формулу (5) можно аккуратно доказать методом полной математической индукции. Мы показали лишь переход от $k = 1$ к $k = 2$.*

§4 Интерполяционная формула Ньютона

Получим в явном виде интерполяционный полином Ньютона $N_n(x)$ для функции $f(x)$ по узлам $\{x_i\}_0^n$. Для этого воспользуемся формулой (5). Полином $N_n(x)$ получается из (5) заменой x_n на x:

$$N_n(x) = f(x_0) + (x - x_0)f(x_0, x_1) + \ldots + (x - x_0)\ldots(x - x_{n-1})f(x_0, x_1, \ldots, x_n) \tag{6}$$

По определению интерполяционного полинома нужно показать,что $N_n(x_i) = f(x_i)\ \forall i = \overline{0,n}$. Ясно, что

$$N_n(x_i) = f(x_0) + (x_i - x_0)f(x_0, x_1) + \ldots + (x_i - x_0)\ldots(x_i - x_{i-1})f(x_0, x_1, \ldots, x_i)$$

По формуле (5) это равно $f(x_i)$. А значит полином (6) является интерполяционным полиномом и носит название интерполяционного полинома Ньютона. Для вычисления погрешности интерполяционного полинома Ньютона можно воспользоваться формулой:

$$\psi_{N_n(x)} = \frac{f^{(n+1)}(\xi)}{(n+1)!}w(x)$$

Или в более привычной форме, положив $M_{n+1} = \sup|f^{n+1}(x)|$ по всем x :

$$|\psi_{N_n(x)}| \leqslant \frac{M_{n+1}}{(n+1)!}|w(x)|$$

Замечание. *Если узлы фиксированы, то удобен полином Лагранжа, а если фиксирована функция, а количество узлов увеличивается на каждой итерации, то удобен полином Ньютона.*

§5 Интерполирование с кратными узлами. Полином Эрмита

Постановка задачи: Пусть заданы $m+1$ узел $\{x_i\}_0^m$ и значения функции в этих узлах $f(x_0)\ldots f(x_m)$. Кроме того, в каждом узле заданы производные $f^{(a_0-1)}(x_0)\ldots f^{(a_m-1)}(x_m)$ где $a_i -$ кратность для узла x_i. Задача заключается в построении полинома $n-$й степени, значения в узлах которого совпадают со значениями заданной функции, а его производные - со значением соответствующих производных заданной функции:

$$H_n^{(i)}(x_k) = f^{(i)}(x_k) \tag{1}$$

Тогда ясно, что $a_0+a_1+\cdots+a_m = n+1$. Если это выполнено, то полином Эрмита ищется и представляется в виде:

$$H_n(x) = \sum_{k=0}^{m}\sum_{i=0}^{a_k-1} c_{k,i}(x)f^{(i)}(x_k) \tag{2}$$

где $c_{k,i}(x)-$ полином $n-$й степени от x.
Построим полином Эрмита с кратными узлами $H_3(x)$, где узел x_1-кратный (в нем заданы значения $f(x_1), f'(x_1)$), а узлы x_0, x_2- простые:

$$\begin{cases} H_3(x_0) = f(x_0) \\ H_3(x_1) = f(x_1) \\ H_3(x_2) = f(x_2) \\ H_3'(x_1) = f'(x_1) \end{cases}$$

Будем искать его в виде:

$$H_3(x) = c_0(x)f(x_0) + c_1(x)f(x_1) + c_2(x)f(x_2) + b_1(x)f'(x_1) \tag{3}$$

Ясно,что:

$$\begin{cases} c_0(x_0) = 1 \\ c_0(x_1) = 0 \\ c_0(x_2) = 0 \\ c_0'(x_1) = 0 \end{cases} \quad \begin{cases} c_1(x_0) = 0 \\ c_1(x_1) = 1 \\ c_1(x_2) = 0 \\ c_1'(x_1) = 0 \end{cases} \quad \begin{cases} c_2(x_0) = 0 \\ c_2(x_1) = 0 \\ c_2(x_2) = 1 \\ c_2'(x_1) = 0 \end{cases} \quad \begin{cases} b_1(x_0) = 0 \\ b_1(x_1) = 0 \\ b_1(x_2) = 0 \\ b_1'(x_1) = 1 \end{cases}$$

Исходя из этой таблицы, выпишем последовательно все коэффициенты полинома $H_3(x)$.Ищем $c_0(x)$ в виде: $c_0(x) = k(x - x_2)(x - x_1)^2$, а k найдем из условия $c_0(x_0) = 1$. Значит $1 = k(x_0 - x_2)(x_0 - x_1)^2$ и :

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

Аналогичными рассуждениями можно получить коэффициент $c_2(x)$:

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

$b_1(x)$ будем искать в виде $b_1(x) = k_1(x - x_0)(x - x_1)(x - x_2)$. Перепишем $b_1(x)$ в виде: $b_1(x) = (x - x_1)[k_1(x - x_0)(x - x_2)]$, $b_1'(x) = [\dots] + (x - x_1)[\dots]'$. И в точке x_1 производная будет равна : $b_1'(x_1) = k_1(x_1 - x_0)(x_1 - x_2) = 1$.. Откуда получаем, что

$$b_1(x) = \frac{(x - x_0)(x - x_1)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)}$$

Коэффициент $c_1(x)$ будем искать в виде x_1 : $c_1(x) = (x - x_0)(x - x_2)(ax + b)$. Так как:

$$\begin{cases} c_1(x_1) = 1 = (x_1 - x_0)(x_1 - x_2)(ax_1 + b) \\ c_1'(x_1) = 0 = a(x_1 - x_0)(x_1 - x_2) + (ax_1 + b)(2x - x_0 - x_2) \end{cases}$$

то

$$a = -\frac{2x_1 - x_0 - x_2}{(x_1 - x_0)^2(x_1 - x_2)^2} \quad b = \frac{1}{(x_1 - x_0)(x_1 - x_2)}\left(1 + \frac{x_1(2x_1 - x_0 - x_2)}{(x_1 - x_0)(x_1 - x_2)}\right)$$

Окончательный вид $c_1(x)$:

$$c_1(x) = \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)}\left(1 - \frac{(2x_1 - x_0 - x_2)(x - x_1)}{(x_1 - x_0)(x_1 - x_2)}\right)$$

Таким образом, показано, что коэффициенты $H_3(x)$ находятся в явном виде и однозначно.

Оценка погрешности полинома Эрмита $H_3(x)$, построенного по узлам x_0, x_1, x_2

Введем функцию

$$q(s) = f(s) - H_3(s) - Kw(s), \tag{4}$$

где $s \in [x_0, x_2]$, $x \in [x_0, x_2]$, и

$$w(s) = (x - x_0)(x - x_1)^2(x - x_2) \quad \text{— полином 4-й степени.}$$

Постоянную $K(x)$ выбираем из условия:

$$q(x) = 0 = f(x) - H_3(x) - Kw(x)$$

Тогда ясно, что константу K нужно брать равной

$$K = \frac{f(x) - H_3(x)}{w(x)}$$

Функция $q(x)$ имеет 4 нуля на отрезке $[x_0, x_2]$. Предположим, что функция $f(x)$ гладкая. Значит к ней можно применить теорему Ролля. У полученной функции $q'(x)$ — не менее 3х нулей. Так как узел x_1 – кратный, то на данном этапе добавляется $q'(x_1) = 0$, откуда следует, что $q'(x)$ имеет не менее 4х нулей, а значит, $q''(x)$ – не менее 3х нулей, $q'''(x)$ – не менее 2х нулей, и тогда $q^{IV}(x)$ – не менее одного нуля. Таким образом существует точка ξ, в которой $q^{IV}(\xi) = 0$.

Тогда продифференцируем 4 раза функцию $q(s)$:

$$q^{IV}(s) = f^{IV}(s) - k \cdot 4!$$

Согласно теореме Ролля $\exists \xi \in [a, b]$, в которой $q^{IV}(\xi) = 0$. Если обозначить

$$M_4 = \sup_{x_0 \leqslant x \leqslant x_2} |f^{IV}(x)|,$$

то получим

$$|f(x) - H_3(x)| = |\psi_{H_3}(x)| \leqslant \frac{M_4}{4!}|w(x)| = \frac{M_4}{4!}|(x - x_0)(x - x_1)^2(x - x_2)|$$

Получили оценку для ПЭ $H_3(x)$.

Используя оценку Ψ_{H_3} получим, что $\forall n \in N$:

$$\Psi_{H_n}(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}(x - x_0)^{a_0}(x - x_1)^{a_1} \dots (x - x_m)^{a_m} \tag{5}$$

И если обозначить

$$M_{n+1} = \sup_x |f^{(n+1)}(x)|,$$

то

$$|\Psi_{H_n}(x)| \leqslant \frac{M_{n+1}}{(n+1)!}|w(x)|$$

Полином $H_3(x)$ может быть получен из многочлена $L_3(x)$ с помощью предельного перехода. Пусть есть узлы x_0, x_1, x_2, x_3, где x_3 — фиктивный узел. По этим 4м узлам можно построить $L_3(x)$

Первое слагаемое

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

При сведении к $H_3(x)$ коэффициенты $(x-x_3) \to (x-x_1)$ и $(x_0-x_3) \to (x_0-x_1)$.

Задача. *Доказать, что*

$$\lim_{x_3 \to x_1} L_3(x) = H_3(x)$$

Решение: Рассмотри полином Лагранжа для функции f:

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

Тогда

$$\lim_{x_3 \to x_1} L_3(x) = \frac{(x-x_1)^2(x-x_2)}{(x_0-x_1)^2(x_0-x_2)} f(x_0) + \dots = H_3(x)$$

Подробные выкладки предлагаем провести читателю. □

§6 Использование полинома $H_3(x)$ для оценки погрешности квадратурной формулы Симпсона

Рассмотрим интеграл $\int\limits_a^b f(x)dx$. Будем вычислять его с помощью квадратурной формулы Симпсона.

Разобьем отрезок на n частичных сегментов

$$a \leqslant x_0 < x_1 < x_2 < \dots < x_N \leqslant b$$

так, что

$$x_i - x_{i-1} = h = const.$$

Тогда, квадратурная формула Симпсона на частичном сегменте имеет вид:

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

где $f(x_i) = f_i$, а $f_{i-\frac{1}{2}}$ — значение функции в середине отрезка $[x_{i-1}, x_i]$:

$$f_{i-\frac{1}{2}} = f(x_{i-1} + 0.5h)$$

Если подынтегральная функция имеет вид

$$f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3,$$

то квадратурная формула Симпсона просто точна (для второй степени — точна по построению).

Докажем, что формула точна для x^3:

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

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

Тогда

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

$$= \frac{h}{6}\left((x_{i-1} + x_i)(x_{i-1}^2 - x_i x_{i-1} + x_i^2) + \frac{(x_i + x_{i-1})(x_i^2 + 2x_i x_{i-1} + x_{i-1}^2)}{2}\right) =$$

$$= \frac{h}{6}(x_i + x_{i-1})\left(\frac{2x_{i-1}^2 - 2x_i x_{i-1} + 2x_i^2 + x_i^2 + 2x_i x_{i-1} + x_{i-1}^2}{2}\right) =$$

$$= \frac{h}{12}(x_i + x_{i-1})3(x_{i-1}^2 + x_i^2) = \frac{h}{4}(x_i + x_{i-1})(x_i^2 + x_{i-1}^2) = \int\limits_{x_{i-1}}^{x_i} x^3 dx.$$

Это и есть аналитическое выражение интеграла. Если подынтегральная функция степени не выше, чем 3, то квадратурная формула Симпсона (КФС) точна. Следовательно, для ПЭ $H_3(x)$ она будет точна.

Теперь получим погрешность для КФС: Построим $H_3(x)$ по узлам $x_{i-1}, x_{i-\frac{1}{2}}, x_i$, где

$$H_3(x_{i-1}) = f(x_{i-1});$$

$$H_3(x_{i-\frac{1}{2}}) = f_{i-\frac{1}{2}};$$

$$H_3(x_i) = f_i;$$

$$H_3'(x_{i-\frac{1}{2}}) = f'_{i-\frac{1}{2}}$$

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

Теперь подынтегральную функцию заменим на

$$f(x) = H_3(x) + \Psi_{H_3}(x)$$

Получим

$$\int\limits_{x_{i-1}}^{x_i} f(x)dx = \{\text{в силу линейности}\} = \int\limits_{x_{i-1}}^{x_i} H_3(x)dx + \int\limits_{x_{i-1}}^{x_i} \Psi_{H_3}(x)dx =$$

$$= \frac{h}{6}(H_3(x_{i-1}) + 4H_3(x_{i-\frac{1}{2}}) + H_3(x_i)) + \int\limits_{x_{i-1}}^{x_i} \Psi_{H_3}(x)dx =$$

$$= \frac{h}{6}(f_{i-1} + 4f_{i-\frac{1}{2}} + f_i) + \Psi_i(f)$$

Тогда ясно, что погрешность КФС на частичном сегменте:

$$\Psi_i(f) = \int\limits_{x_{i-1}}^{x_i} f(x)dx - \frac{h}{6}(f_{i-1} + 4f_{i-\frac{1}{2}} + f_i) = \int\limits_{x_{i-1}}^{x_i} \Psi_{H_3}(x)dx,$$

Обозначим

$$M_4 = \sup_{x_{i-1} \leqslant x \leqslant x_i} |f^{IV}(x)|.$$

Тогда можно сказать, что

$$|\Psi_i(f)| \leqslant \frac{M_4}{4!} \int\limits_{x_{i-1}}^{x_i} (x - x_{i-1})(x - x_{i-\frac{1}{2}})^2(x_i - x)dx$$

Задача.

Показать, что $\int\limits_{x_{i-1}}^{x_i} (x - x_{i-1})(x - x_{i-\frac{1}{2}})^2(x_i - x)dx = \frac{h^5}{120}$

Решение: Проведем замену в подынтегральном выражении: $x = x_{i-1} + th,\ t \in [0, 1]$. Тогда $dx = hdt$ и $x - x_{i-1} = th,\ \ x_i - x = h(1 - t),\ \ \left(x - x_{i-\frac{1}{2}}\right)^2 = h^2\left(t - \frac{1}{2}\right)^2$.
Таким образом подставляя эти выражения в интеграл, получаем:

$$\int\limits_{x_{i-1}}^{x_i} (x - x_{i-1})(x_i - x)\left(x - x_{i-\frac{1}{2}}\right)^2 dx =$$

$$h^5 \int\limits_0^1 (t)(1 - t)\left(t - \frac{1}{2}\right)^2 dt = h^5 \int\limits_0^1 \left(2t^3 - \frac{5}{4}t^2 - t^4 + \frac{1}{4}t\right) dt = \frac{h^5}{120}$$

□

Таким образом $|\psi_i(f)| \leqslant \frac{M_4}{4!}\frac{h^5}{120}$. Следовательно, погрешность КФС на частичном отрезке имеет 5й порядок по h. Запишем погрешность на всем отрезке:

$$\Psi_h(f) = \int\limits_a^b f(x)dx - \sum_{i=1}^{N} \Psi_i(f),$$

$$\left|\Psi_h(f) \leqslant \left(\frac{h}{2}\right)^4 \frac{M_4(b-a)}{180}\right|,$$

$$hN = b - a,$$

Таким образом квадратурная формула Симпсона на всем отрезке интегрирования имеет 4й порядок точности по h.

§7 Наилучшее среднеквадратичное приближение функций

Введем для начала пространство функций, интегрируемых с квадратом:
$H = L_2$ – гильбертово пространство — пространство функций таких, что

$$\forall f \in L_2 \quad \int_a^b f^2(x)dx < \infty$$

Чтобы сделать пространство нормированным, введем скалярное произведение:

$$\forall f, g \in L_2 \quad (f, g) = \int_a^b f(x)g(x)dx$$

и норму

$$\|f\|_{L_2} = \left(\int_a^b f^2(x)dx\right)^{\frac{1}{2}}$$

Переходим к постановке задачи. Рассмотрим линейно независимые функции

$$\varphi_0(x), \varphi_1(x), \ldots, \varphi_n(x) \in L_2$$

или

$$\int_a^b \varphi_i^2 dx < \infty, \quad i = \overline{0, n}$$

Составим многочлен

$$\varphi(x) = c_0\varphi_0(x) + c_1\varphi_1(x) + \ldots + c_n\varphi_n(x) = \sum_{x=0}^{n} c_k\varphi_k(x), \tag{1}$$

где $\varphi(x)$ — обобщенный многочлен, построенный по функциям $\varphi_i(x)$.

Среди всех многочленов $\varphi(x)$ нужно найти многочлен $\overline{\varphi}(x)$:

$$\overline{\varphi}(x) = \sum_{k=0}^{n} \overline{c_k}\varphi_k(x),$$

который минимизирует интеграл (норму)

$$\|f(x)-\overline{\varphi}(x)\|_{L_2}^2=\int\limits_a^b(f(x)-\overline{\varphi}(x))^2dx=\min_{\varphi(x)}\int\limits_a^b(f(x)-\varphi(x))^2dx.$$

В этом случае говорят, что $\overline{\varphi}(x)$ — наилучшее среднеквадратичное приближение (НСКП) f по системе $\{\varphi_i(x)\}_0^n$. Построим НСКП в случае, когда базисная функция $\varphi_0(x)$ одна, то есть при $n=0$:

$$\int\limits_a^b\varphi_0^2(x)dx<\infty,\quad \varphi_0(x)\in L_2$$

Ясно, что

$$F(c_0)=\int\limits_a^b(f(x)-c_0\varphi_0(x))^2dx=\int\limits_a^b f^2(x)dx-2c_0\int\limits_a^b f(x)\varphi_0(x)dx+\int\limits_a^b c_0^2\varphi_0^2(x)dx$$

— квадратичная функция относительно c_0. Коэффициент c_0 в наилучшем среднеквадратичном приближении находится из условия $F'(c_0)=0$:

$$2\int\limits_a^b f(x)\varphi_0(x)dx=2c_0\int\limits_a^b\varphi_0^2(x)dx$$

$$\Downarrow$$

$$\overline{c_0}=\frac{\int\limits_a^b f(x)\varphi_0(x)dx}{\int\limits_a^b\varphi_0^2(x)dx}=\frac{(f,\varphi_0)}{(\varphi_0,\varphi_0)}$$

Таким образом НСКП $\overline{\varphi}(x)=\overline{c_0}\varphi_0(x)=\dfrac{(f,\varphi_0)}{(\varphi_0,\varphi_0)}\varphi_0$. Если

$$\varphi_0(x)=1,\quad \overline{c_0}=\frac{\int\limits_a^b f(x)dx}{(b-a)},$$

то

$$\overline{\varphi}(x)=\overline{c_0}\cdot 1=\frac{\int\limits_a^b f(x)dx}{(b-a)},$$

будет являться средним значением интеграла.

Перейдем к общему случаю. Пусть $\{\varphi_i(x)\}_0^n$ — система линейно независимых (базисных) функций из L_2, то есть

$$\int\limits_a^b \varphi_i^2(x)dx < \infty, \quad \varphi_i(x) \in L_2[a,b].$$

Тогда составим функцию

$$F(c_0, c_1, \ldots, c_n) = \int\limits_a^b (f(x) - \sum_{k=0}^{n} c_k\varphi_k(x))^2 dx.$$

Найдем минимум этой функции. Заметим, что он достигается, когда

$$\frac{\partial F}{\partial c_k} = 0, \quad k = \overline{0,n},$$

Итак,

$$\begin{aligned} &F(c_0, c_1, \ldots, c_n) = \\ &= \int\limits_a^b f^2(x)dx - 2\sum_{k=0}^{n} c_k \int\limits_a^b f(x)\varphi_k(x)dx + \sum_{k=0}^{n} c_k \sum_{l=0}^{n} c_l \int\limits_a^b \varphi_k(x)\varphi_l(x)dx = \\ &= (f,f) - 2\sum_{k=0}^{n} c_k(f,\varphi_k) + \sum_{k=0}^{n} c_k \sum_{l=0}^{n} c_l(\varphi_k,\varphi_l) \end{aligned}$$

Минимум функции $F(c_0, c_1, \ldots, c_n)$ достигается в точке, где

$$\frac{\partial F(c_0, \ldots, c_n)}{\partial c_k} = 0, \quad k = \overline{0,n}.$$

В итоге получаем систему уравнений:

$$\sum_{l=0}^{n} c_l(\varphi_k, \varphi_l) = (f, \varphi_k), \quad k = \overline{0,n}.$$

Или в координатном виде:

$$\begin{cases} c_0(\varphi_0,\varphi_0) + c_1(\varphi_0,\varphi_1) + \ldots + c_n(\varphi_0,\varphi_n) = (f,\varphi_0) \\ c_1(\varphi_1,\varphi_0) + c_1(\varphi_1,\varphi_1) + \ldots + c_n(\varphi_1,\varphi_n) = (f,\varphi_1) \\ \ldots \\ c_0(\varphi_n,\varphi_0) + c_1(\varphi_n,\varphi_1) + \ldots + c_n(\varphi_n,\varphi_n) = (f,\varphi_n) \end{cases} \tag{1}$$

Правые части системы известны. Выпишем матрицу из коэффициентов:

$$\begin{pmatrix} (\varphi_0,\varphi_0) & (\varphi_0,\varphi_1) & \ldots & (\varphi_0,\varphi_n) \\ (\varphi_1,\varphi_0) & (\varphi_1,\varphi_1) & \ldots & (\varphi_1,\varphi_n) \\ \ldots & \ldots & \ldots & \ldots \\ (\varphi_n,\varphi_0) & (\varphi_n,\varphi_1) & \ldots & (\varphi_n,\varphi_n) \end{pmatrix} = G(\varphi_0, \ldots, \varphi_n), \quad |G| > 0$$

Полученная матрица является матрицей Грама. Если система линейно независима, то матрица Грама невырожденная. Следовательно, по критерию определенности СЛАУ существует единственное решение:

$$\overline{c_0}, \overline{c_1}, \ldots, \overline{c_n}$$

Эти коэффициенты позволяют выписать обобщенный многочлен

$$\overline{\varphi}(x) = \sum_{i=0}^{n} \overline{c_i}\varphi_i(x),$$

который является НСКП для функции f.

Если система исходных базисных функций $\{\varphi_i(x)\}_0^n$ ортонормирована, то матрица Грама превратится в единичную ($G = E$) и

$$\overline{c}_i = (f, \varphi_i), \quad i = \overline{0, n}, \tag{2}$$

где $\overline{e}_i$ — коэффициенты Фурье.

Рассмотрим систему линейно независимых функций:

$$1, x, x^2, \ldots, x^n, \quad \varphi_i(x) = x^i$$

исходя из нее можно построить ортогональную систему. Для этого рассматривается скалярное произведение:

$$\int_{\alpha}^{\beta} \rho(x)\varphi_k(x)\varphi_l(x)dx = (\varphi_k, \varphi_l),$$

где α и β – выбираемые границы, $\rho(x) > 0$ – весовая функция. Если выбирать различные α, β и $\rho(x)$, то можно получить полиномы Якоби, Лежандра, Чебышева и другие. Эти полиномы — ортогональные полиномы.(по ним строить НСКП наиболее удобно).

Замечание. *Если система $\{\varphi_i(x)\}_0^n$ ортонормирована, то нетрудно вычислить отклонение от НСКП:*

$$F(c_0, c_1, \ldots, c_n) = \int_a^b [f(x) - \sum_{k=0}^{n} c_k\varphi_k(x)]^2 dx = (f, f) - \sum_{k=0}^{n} c_k^2 \geqslant 0.$$

Тогда

$$(f, f) \geqslant \sum_{k=0}^{n} c_k^2, \quad \|f\|^2 \geqslant \sum_{k=0}^{n} c_k^2$$

Это есть неравенство Бесселя.

Если $\{\varphi_i\}_0^n$ — ортонормированный базис, то неравенство переходит в равенство Парсеваля:

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

Глава 3

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 $[a, b]$.

$$f(x) = 0, \quad x \in [a, b] \tag{1}$$

Решая это уравнение нужно пройти 2 этапа:

1. Локализация корней уравнения.

   Пусть x^* — корень
   $$f(x^*) = 0.$$

   Нужно указать окрестность корня:

   $$U_a(x^*) = \{x : |x^* - x| < a\}.$$

2. Построение итерационного процесса

   $$x_n \longrightarrow x^*.$$

Аналогичная задача будет ставиться и для системы :

$$\begin{cases} f_1(x_1, \ldots, x_m) = 0 \\ f_2(x_1, \ldots, x_m) = 0 \\ \ldots \\ f_m(x_1, \ldots, x_m) = 0 \end{cases} \tag{2}$$

Введя вектор $\overline{x} = (x_1, x_2, \ldots, x_m)$, $\overline{f} = (f_1, \ldots, f_m)^\tau$, можем систему (2) переписать в виде $\overline{f}(\overline{x}) = 0$.

Можно $\overline{f}$ трактовать как отображение $R_m \longrightarrow R_m$ (нелинейное). Укажем способы локализации корня:

1. Если f — непрерывная функция, то разобьем отрезок на узлы x_i и вычислим значения в узлах. Если $f(x_i)f(x_{i-1}) < 0$, то ясно, что на этом частичном отрезке есть по крайней мере один корень. Далее этой процедуре подвергаем отрезок $[x_i, x_{i-1}]$. Тогда найдем a–окрестность $U_a(x_*)$.

2. Метод бисекции. Пусть есть отрезок $[a, b]$, на котором есть корень. И пусть для определенности $f(a) < 0$, $f(b) > 0$. Сначала рассмотрим середину отрезка $x_0 = \dfrac{(a+b)}{2}$. Вычислив значение в этой точке (пусть оно > 0), определим, что корень принадлежит $[a, x_0]$. Зная это проделываем то же самое и получаем точку $x_1 = \dfrac{(a+x_0)}{2}$. Пусть это значение тоже > 0, тогда значение x_* находится в интервале $[a, x_1]$. Далее продолжаем процесс до нужной точности.

Предположим, что корней в x_* много, тогда на первом этапе выразим

$$f = (x - x_*)g(x).$$

Далее будем проводить процесс для $g(x)$.

§2 Метод простой итерации

Рассмотрим нелинейное уравнение:

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

Пусть задана окрестность корня $U_a(x_*)$. Тогда метод простой итерации строится исходя из уравнения:

$$x = S(x) \tag{2}$$

по формуле:

$$x_{n+1} = S(x_n), \tag{3}$$

где $n = 0, 1, \ldots, \quad x_0 \in U_a(x_*)$. Функция $S(x)$ выбирается в виде

$$S(x) = x + r(x)f(x), \tag{4}$$

где $f(x)$ — исходная функция, $r(x)$ — функция: $\operatorname{sign}(r(x)) \neq 0 \quad \forall x \in U_a(x_*)$.

Определение. *$S(x)$ удовлетворяет условию Липшица с константой q, если $\forall x_1, x_2 \in U_a(x_*)$ выполняется неравенство:*

$$|S(x_1) - S(x_2)| \leqslant q|x_1 - x_2| \tag{5}$$

Утверждение. *Пусть $S(x)$ — удовлетворяет условию Липшица с константой $q \in (0,1)$ и пусть начальное приближение x_0 берется из $U_a(x_*)$. Тогда метод простой итерации сходится со скоростью геометрической прогрессии со знаменателем q.*

Доказательство: Докажем по индукции. $n = 0 : x_0 \in U_a(x_*)$. Покажем, что если $x_n \in U_a$, то и $x_{n+1} \in U_a$. Это будет означать, что при итерационном процессе мы не выйдем из окрестности U_a.

Оценим $|x_{n+1} - x_*|$: так как $x_{n+1} = S(x_n)$, то

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

Так как $q < 1$, то это означает, что

$$q|x_n - x_*| < a.$$

А значит $x_{n+1} \in U_a(x_*)$.

Это соотношение можно применять как рекуррентное:

$$|x_n - x_*| \leqslant q^n |x_0 - x_*|$$

и при $n \to \infty$:

$$\lim_{n\to\infty} q^n = 0.$$

Это означает, что в пределе при $n \to \infty$ x_n даст нам x_*. Данный итерационный метод имеет медленную сходимость, так как связь x_{n+1} и x_n — линейная. ■

Замечание. *Пусть*

$$\max_{x\in U_a} |S'(x)| = q < 1,$$

тогда МПИ сходится, если $x_0 \in U_a(x_*)$.

Замечание. *Рассмотрим следующий итерационный процесс:*

$$\frac{x_{n+1} - x_n}{\tau} + f(x_n) = 0, \tag{6}$$

где $n \in \mathbb{N}_0$, $x_0 \in U_a(x_*)$, $\tau > 0$.

Пусть для определенности $f'(x) > 0$. Запишем метод (6), как метод простой итерации:

$$S(x) = x - \tau f(x).$$

В силу первого замечания, если $|S'(x)| < 1$, то сходимость есть.

$S'(x) = 1 - \tau f'(x)$, в предположении, что f — гладкая.

Обозначим

$$M_1 = \max_{x\in U_a(x_*)} |f'(x)|.$$

Тогда

$$|1 - \tau M_1| < 1 \quad \Longrightarrow \quad -1 < 1 - \tau M_1 < 1 \quad \Longrightarrow \quad 0 < \tau < \frac{2}{M_1}.$$

Ускорение сходимости итерационного метода (метод Эйткена)

Пусть известно, что две соседние итерации удовлетворяют следующему условию:

$$x_n - x_* \approx Aq^n, \quad n = 1, 2, 3, \dots$$

Запишем 3 соседних итерации:

$$x_{n-1} - x_* \approx Aq^{n-1} \tag{7}$$

$$x_n - x_* \approx Aq^n \tag{8}$$

$$x_{n+1} - x_* \approx Aq^{n+1} \tag{9}$$

Поставим задачу выразить корень через эти итерации:

$$x_* \approx x_{n+1} - Aq^{n+1},$$

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

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

В итоге

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

Так как все равенства приближенные, то x_* берется за значение x_{n+1} итерации.

§3 Метод Ньютона и метод секущих

Рассматриваем нелинейное уравнение

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

Корень уравнения локализован, то есть указана окрестность $U_a(x_*)$. Обозначим n – ю итерацию через x^n. Тогда запишем:

$$x^{n+1} = x^n - \frac{f(x^n)}{f'(x^n)}, \ n = 0, 1, 2, 3, \dots \quad x^0 \in U_a(x_*) \tag{2}$$

Методом Ньютона решения нелинейного уравнения $f(x) = 0$ называется метод, записанный в виде (2). Для доказательства сходимости метода Ньютона будем требовать, чтобы функция была гладкой до третей производной. (так же этот метод называется методом касательных)

Если дана функция $f(x)$ то уравнение касательной в точке $(x^n, f(x^n))$ будет иметь вид: $y - f(x^n) = f'(x^n)(x - x^n)$. Тогда за x^{n+1} принимается абсцисса точки, в которой касательная пересекает ось ординат. Формулу (2) легко получить из следующих соображений. Разложим по формуле Тейлора в окрестности корня: $f(x_*) = 0 =$

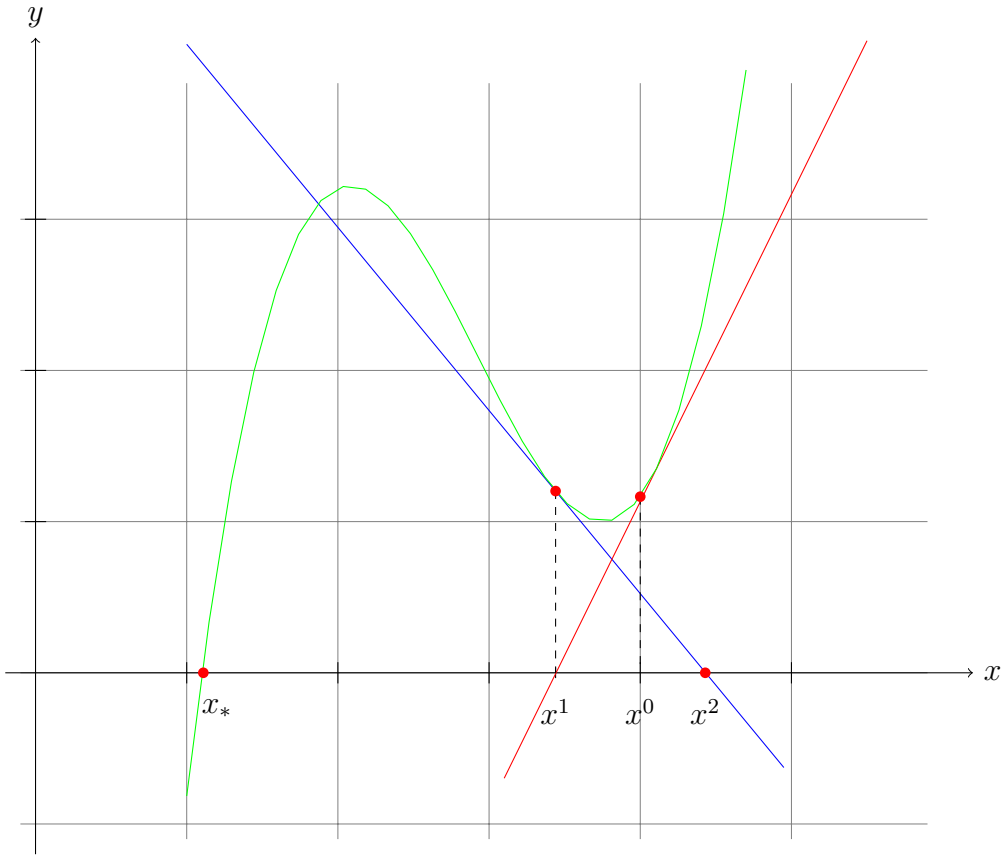
$f(x) + (x_* - x)f'(x^n) + \dots$. Тогда если вместо x_* подставить x^{n+1} и вместо $(x - x^n)$, то получим:

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

Разрешим уравнение относительно x^{n+1} при условии, что $f'(x) \neq 0$:

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

Приведем графически пример, когда при неудачном выборе начального приближения метод Ньютона зациклится:



Рассмотрим модифицированный метод Ньютона:

$$x^{n+1} = x^n - \frac{f(x^n)}{f'(x^0)}, n = 0, 1, 2, 3, \dots \quad x^0 \in U_a(x_*) \tag{3}$$

Один из плюсов этого метода - не нужно считать производную на каждой итерации.Но сходимость этого метода (в смысле скорости) будет хуже чем (2). Запишем модифицированный метод Ньютона для нелинейных систем. Пусть существует два нелинейных уравнения :

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

В качестве решения системы понимается точка (x_1^*, x_2^*), в которой $f_i(x_1^*, x_2^*) = 0,\ i =$

1, 2. Разложим обе функции в ряд Тейлора:

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

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

Сделаем замену x_i на x_i^n , x_i^* на x_i^{n+1}

$$\begin{cases} f_1(x_1^n, x_2^n) + (x_1^{n+1} - x_1^n)\dfrac{\partial f_1(x_1^n, x_2^n)}{\partial x_1^n} + (x_2^{n+1} - x_2^n)\dfrac{\partial f_1(x_1^n, x_2^n)}{\partial x_2^n} = 0 \\ f_2(x_1^n, x_2^n) + (x_1^{n+1} - x_1^n)\dfrac{\partial f_2(x_1^n, x_2^n)}{\partial x_1^n} + (x_2^{n+1} - x_2^n)\dfrac{\partial f_2(x_1^n, x_2^n)}{\partial x_2^n} = 0 \end{cases}$$

Чтобы записать данную систему в компактном виде, введем векторы $f = (f_1, f_2)^T,\ x = (x_1, x_2)^T$ и матрицу

$$J(x) = \begin{pmatrix} \dfrac{\partial f_1}{\partial x_1} & \dfrac{\partial f_1}{\partial x_2} \\ \dfrac{\partial f_2}{\partial x_1} & \dfrac{\partial f_2}{\partial x_2} \end{pmatrix} \tag{5}$$

Заменим x_i на x_i^n , x_i^* на x_i^{n+1} : $f(x^n) = J(x^n)(x^{n+1} - x^n) = 0$. Если предположить, что $J(x^n)$ обратима, то

$$x^{n+1} = x^n - J^{-1}(x^n)f(x^n) \tag{6}$$

Чтобы на каждой итерации не обращать матрицу J, вводят вектор $V^{n+1} = x^{n+1} - x^n$: $f(x^n) + J(x^n)V^n = 0$. Найдя этот вектор, можно найти x^{n+1} по формуле : $x^{n+1} = x^n + V^{n+1}$

Запишем метод для любого количества нелинейных уравнений:

$$\begin{cases} f_1(x_1, x_2, \dots, x_n) = 0 \\ f_2(x_1, x_2, \dots, x_n) = 0 \\ \dots \\ f_m(x_1, x_2, \dots, x_n) = 0 \end{cases}$$

Метод Ньютона для этой системы получается аналогичным образом. Пусть $J = (f_{ij}), f_{ij} = \dfrac{\partial f_i}{\partial x_j}$, $i = \overline{1, m}$ $j = \overline{1, m}$, квадратная матрица порядка m и векторы $f = (f_1, f_2, \dots, f_m)^T$ $x = (x_1, x_2, \dots, x_m)^T$. Тогда метод Ньютона будет иметь вид:

$$x^{n+1} = x^n - J^{-1}(x^n)f(x^n) \tag{7}$$

где $n = 0, 1, 2, 3, \dots$, и x^0 – задано из окрестности корня.

Метод секущих

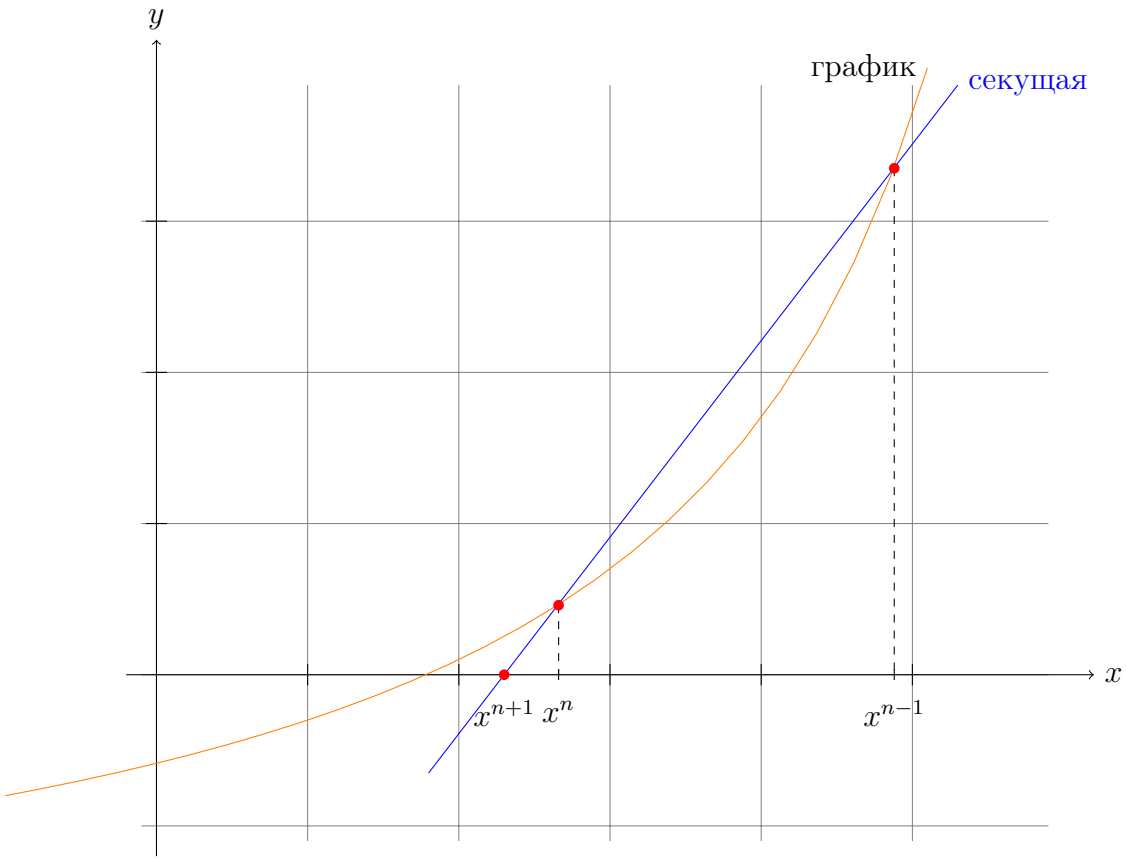
Запишем метод Ньютона

$$x^{n+1} = x^n - \frac{f(x^n)}{f'(x^n)}, \quad n = 1, 2, 3, \ldots, \quad x_0 \in U_a(x_*)$$

решения уравнения $f(x) = 0$. Исходя из этой формулы, заменим производную на ее дискретный аналог и подставим в формулу: $f'(x^n) = \dfrac{f(x^n) - f(x^{n-1})}{x^n - x^{n-1}}$. Теперь получим итерационный метод, называемый методом секущих:

$$x^{n+1} = x^n - \frac{(x^n - x^{n-1})f(x^n)}{f(x^n) - f(x^{n-1})} \tag{8}$$

где $n = 0, 1, 2, 3, \ldots \quad x^0, x^1 -$ заданы. Этот метод является двухшаговым. Значение x^{n+1} находится с использованием x^n и x^{n-1}. На рисунке изображена графическая иллюстрация описанного метода:



§4 Сходимость метода Ньютона. Оценка скорости сходимости

Рассматриваем метод Ньютона для нелинейного уравнения

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

$x^{n+1} = x^n - \dfrac{f(x^n)}{f'(x^n)}$, где $n = 0, 1, 2, 3, \ldots$ и $x^0 \in U_a(x_*)$

Метод Ньютона можно трактовать как метод простой итерации:

$$x^{n+1} = S(x^n) \tag{3}$$

где $S(x) = x - \dfrac{f(x)}{f'(x)}$. При изучении сходимости метода простой итерации было замечено, что простая итерация (3) сходится, если $|S'(x)| = q < 1$ для $x \in U_a(x_*)$. Продифференцируем $S(x)$:

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

Следовательно, $S'(x_*) = 0$ так как $f(x_*) = 0$. Введем погрешность $z_n = x^n - x_*$. Покажем, что эта погрешность на двух соседних итерациях будет связана квадратичным образом. Рассмотрим

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

Разложим это по формуле Тейлора: $z_{n+1} = S(x_*) + S'(x_*)z_n + 0.5S''(x)z_n^2 - S(x_*) = 0.5S''(x)z_n^2$, где $x = x_* + \theta z_n$, $|\theta| < 1$. Предположим, что существует постоянная $M > 0$ такая, что для $\forall x \in U_a(x_*)$ выполняется неравенство:

$$|0.5S''(x)| \leqslant M \text{ или, что то же самое } \frac{1}{2}\left|\left(\frac{f(x)f''(x)}{(f'(x))^2}\right)'\right| \leqslant M \tag{4}$$

Оценим z_{n+1} : $|z_{n+1}| \leqslant M|z_n^2|$. Домножим это неравенство на M и введем функцию $V_n = M|z_n|$. Тогда получим $V_{n+1} \leqslant V_n^2$. То есть две соседние погрешности связаны квадратично. Применяя эту формулу как рекуррентную, получим: $V_n \leqslant V_0^{2^n}$ и вернувшись к z :

$$M|z_n| \leqslant (M|z_0|)^{2^n} \text{ и } |z_n| \leqslant \frac{1}{M}(M|z_0|)^{2^n} \tag{5}$$

Обозначим через q величину $M|z_0|$ и если $q < 1$, то метод Ньютона сходится.Следовательно, начальное приближение x_0 нужно выбирать так, чтобы выполнялось неравенство:

$$|x^0 - x_*| < \frac{1}{M} \tag{6}$$

Тогда итерационный метод Ньютона сходится и имеет место оценка:

$$|x^n - x_*| < \frac{1}{M}(M|x^0 - x_*|)^{2^n} \tag{7}$$

Утверждение. *Пусть существует константа M, удовлетворяющая условию (4) и пусть в некоторой окрестности корня $U_a(x_*)$, $a = a(M)$ начальное приближение выбирается в соответствии с условием (6). Тогда итерационный метод Ньютона решения уравнения (1) сходится и имеет место оценка (7).*

Доказательство: Данное утверждение было доказано выше. ∎

Глава 4

Разностные методы решения задач математической физики

§1 Явная разностная схема для первой краевой задачи уравнения теплопроводности

Постановка задачи:

$$\frac{\partial U}{\partial t} = \frac{\partial^2 U}{\partial x^2} + f(x,t) \tag{1}$$

где $x \in (0,1)$ $t \in (0,T]$. Уравнение решается внутри области. Выпишем краевые условия первого рода:

$$\begin{cases} U(0,t) = \mu_1(t) \\ U(1,t) = \mu_2(t) \end{cases} \qquad t \in [0,T] \tag{2}$$

А так же начальное условие:

$$U_0(x) = U(x,0), x \in [0,1] \tag{3}$$

Задача состоит в нахождении непрерывной в замкнутой области функции $U(x,t)$, удовлетворяющей внутри области уравнению (1), на границе – условию (2) и начальному условию (3).

Известно, что решение данной задачи существует, единственно и устойчиво.Введем сетку:

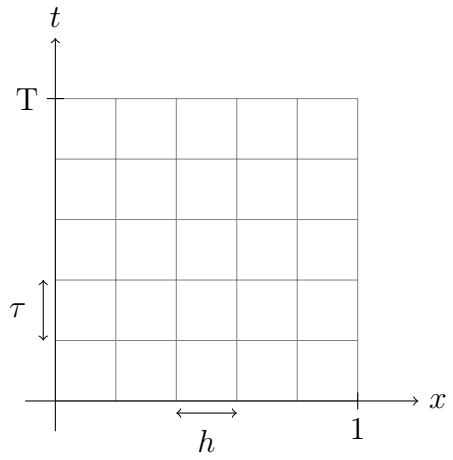
$w_h = \{x_i = ih, i = \overline{1, N-1},\ h = \frac{1}{N}\}$ где $h > 0$ – шаг по переменной x.

$\overline{w}_h = \{x_i = ih, i = \overline{0, N},\ h = \frac{1}{N}\}$

$w_\tau = \{t_j = j\tau\ ,\ j = \overline{1, j_0},\ \tau j_0 = T\}$

$\overline{w}_\tau = \{t_j = j\tau\ ,\ j = \overline{0, j_0},\ \tau j_0 = T\}$ где $\tau > 0$ – шаг по времени.

Тогда множество внутренних узлов $w_{\tau h} = w_\tau \times w_h$ и множество всех узлов: $\overline{w}_{\tau h} = \overline{w}_\tau \times \overline{w}_h$



Совокупность всех узлов в определенный момент времени будем называть слоем.

$$y_i^n = y(x_i, t_n), \quad f_i^n = f(x_i, t_n), \quad (x_i, t_n) \in \overline{\omega}_{\tau,h}.$$

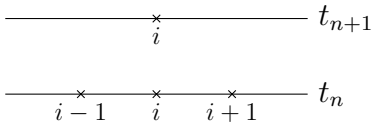
Выпишем явную разностную схему:

$$\frac{y_i^{n+1} - y_i^n}{\tau} = \frac{y_{i-1}^n - 2y_i^n + y_{i+1}^n}{h^2} + f(x_i, t_n), \quad (x_i, f_n) \in \omega_{\tau h} \tag{4}$$

$$\begin{cases} y_0^{n+1} = \mu_1(t_{n+1}) \\ y_N^{n+1} = \mu_2(t_{n+1}) \end{cases} \qquad t_{n+1} \in \overline{\omega}_{\tau,h} \tag{5}$$

$$y_i^0 = U_0(x_i), \quad x_i \in \overline{\omega} \tag{6}$$

Совокупность узлов, в которых записывается разностная схема, называют шаблоном. Здесь использован четырехточечный шаблон вида:



Таким образом задаче (1)-(3) мы поставили в соответствие её дискретный аналог (4)-(6). Возникает система линейных алгебраических уравнений, которая и называется разностной схемой.

Записанная разностная схема использует два слоя : t_{n+1} и t_n. В слое t_n берется три узла : $(i-1)$, (i), $(i+1)$, а в слое t_{n+1} только (i)-ый узел. То есть разностная схема записана на четырехточечном шаблоне.

Решение на (n+1) слое находится по явной формуле:

$$y_i^{n+1} = y_i^n + \frac{\tau}{h^2}(y_{i-1}^n - 2y_i^n + y_{i+1}^n) + \tau f_i^n \quad i = \overline{1, N-1}$$

$$\begin{cases} y_0^{n+1} = \mu_1(t_{n+1}) \\ y_N^{n+1} = \mu_2(t_{n+1}) \end{cases}$$

При изучении разностных схем возникают следующие вопросы:

1. Погрешность аппроксимации.

2. Существование и единственность решения разностных схем.

3. Алгоритм нахождения численного решения.

4. Сходимость разностный схемы к решению исходной задачи.

5. Устойчивость решения по начальному условию и правой части.

Явная разностная схема является условно сходящейся, т.е. если

$$\gamma = \frac{\tau}{h^2} \leqslant 0.5, \tag{7}$$

тогда разностная схема сходится, иначе она сходиться не будет. Также явная разностная схема является условно устойчивой.

Теперь введем погрешность разностной схемы:

$$z_i^n = y_i^n - U(x_i, t^n) \tag{8}$$

$$y_i^n = z_i^n + U_i^n$$

Подставим y_i^n в разностное уравнение.

$$\frac{z_i^{n+1} - z_i^n}{\tau} = \frac{z_{i-1}^n - 2z_i^n + z_{i+1}^n}{h^2} + \psi_i^n \tag{9}$$

$$z_0^{n+1} = z_N^{n+1} = 0, \quad z_i^0 = 0 \tag{10}$$

Выпишем погрешность аппроксимации на решение:

$$\psi_i^n = \frac{U_{i-1}^n - 2U_i^n + U_{i+1}^n}{h^2} - \frac{U_i^{n+1} - U_i^n}{\tau} + f_i^n \tag{11}$$

$$(x_i, t_n) \in \omega_{\tau h}$$

Определение. *ψ_i^n называется погрешностью аппроксимации разностной схемы на решение исходной задачи.*

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:

$$u(x_{i-1}, t_n) = u_{i-1}^n = u(x_i, t_n) - u_x(x_i, t_n)h + \frac{1}{2}u_{xx}(x_i, t_n)h^2 - \frac{1}{6}u_{xxx}(x_i, t_n)h^3 + O(h^4)$$

Полученные разложения подставим в формулу:

$$\psi_i^n = \frac{u_{i-1}^n - 2u_i^n + u_{i+1}^n}{h^2} - \frac{u_i^{n+1} - u_i^n}{\tau} + f_i^n$$

Приведя подобные слагаемые, получаем оценку $\psi_i^n = O(\tau + h^2)$ □

Докажем, что условия (7) достаточно для сходимости в норме $\|\cdot\|_c$.

$$\|y^n\|_c = \max_{0 \leqslant i \leqslant N} |y_i^n|$$

Пусть $\gamma = \dfrac{\tau}{h^2} \leqslant 0.5$ Тогда разностная схема (4)-(6) сходится в норме $\|\cdot\|_c$. Разрешим задачу относительно (n+1) слоя.

$$z_i^{n+1} = z_i^n + \gamma(z_{i-1}^n - 2z_i^n + z_{i+1}^n) + \tau\psi_i^n$$

Соберем подобные члены.

$$z_i^{n+1} = (1 - 2\gamma)z_i^n + \gamma(z_{i-1}^n + z_{i+1}^n) + \tau\psi_i^n$$

$$(1 - 2\gamma) \geqslant 0$$

Следовательно,

$$|z_i^{n+1}| \leqslant (1 - 2\gamma)|z_i^n| + \gamma(|z_{i-1}^n| + |z_{i+1}^n|) + \tau|\psi_i^n|$$
$$|z_i^{n+1}| \leqslant (1 - 2\gamma)\|z^n\|_c + 2\gamma\|z^n\| + \tau\|\psi^n\|_c = \|z^n\|_c + \tau\|\psi^n\|_c$$

Так как это справедливо для всех i, то

$$\|z^{n+1}\|_c \leqslant \|z^n\|_c + \tau\|\psi^n\|_c$$

Рассматривая формулу как рекуррентную, получим

$$\|z^{n+1}\|_c \leqslant \|z^0\|_c + \tau \sum_{k=0}^{n} \|\psi^k\|_c \tag{12}$$

Так как

$$\|\psi^k\| \leqslant M(\tau + h^2), \quad M > 0,$$

где M не зависит от τ и h. Тогда мы можем записать, что

$$\|z^{n+1}\|_c \leqslant M t_{n+1}(\tau + h^2)$$

$$t_{n+1} \leqslant T.$$

Положим $MT = M_1$, которая не зависит от шагов τ и h. Теперь получаем окончательную оценку:

$$\|z^{n+1}\| \leqslant M(\tau + h^2).$$

Если $\tau \to 0, \quad h \to 0$, то $\|z^{n+1}\|_c \to 0$.

Т.е. $\|y_i^{n+1} - U_i^{n+1}\| \to 0$. Таким образом, имеет место сходимость численного решения к решению исходной задачи с первым порядком по τ и вторым по h.

Если мы рассмотрим разностную схему, но при нулевых краевых условиях,то оценка будет для y:

$$\|y^{n+1}\|_c \leqslant \|U_0\|_c + \sum_{k=0}^{n} \tau \|f^k\|_c \tag{14}$$

Говорят, что разностная схема устойчива в $\|\cdot\|_c$ по начальному условию и по правой части. Оценка (14) называется априорной оценкой. Таким образом мы показали, что условие (7) является достаточным. Теперь покажем, что оно также является и необходимым.

Выпишем однородное уравнение, соответствующее неоднородному уравнению (4):

$$\frac{y_i^{n+1} - y_i^n}{\tau} = \frac{y_{i-1}^n - 2y_i^n + y_{n+1}^n}{h^2} \tag{15}$$

Будем искать некоторые частные решение в следующем виде:

$$y_j^n = q^n e^{ijh\varphi}. \tag{16}$$

где

$$i^2 = -1, \quad \varphi \in R.$$

Подставим эту формулу в уравнение

$$y_i^{n+1} = y_i^n + \gamma(y_{i-1}^n - 2y_i^n + y_{i+1}^n).$$

Тогда мы получим выражение для q:

$$q = 1 + \gamma(e^{ih\varphi} - 2 + e^{-ih\varphi}) = 1 + 2\gamma(\cos(h\varphi) - 1) = 1 - 4\gamma \sin^2 \frac{h\varphi}{2}$$

Видно, что $|q| > 1$ означает неустойчивость.

$$1 - 4\gamma \sin^2 \frac{h\varphi}{2} < -1,$$

$$2 < 4\gamma \sin^2 \frac{h\varphi}{2},$$

Откуда получаем, что

$$\gamma > \frac{1}{2}$$

Если $|q| > 1$, т.е. $\gamma > \dfrac{1}{2}$, то гармоники будут неограниченно возрастать и разностная схема будет расходиться.

Таким образом, условие $\gamma \leqslant \dfrac{1}{2}$ является необходимым и достаточны для сходимости и устойчивости явной разностной схемы.

§2 Чисто неявная разностная схема (схема с опережением) для первой краевой задачи уравнения теплопроводности

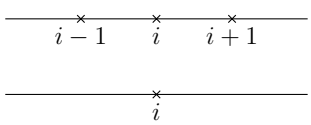
Задача (1) - (3) остается прежней.

$$\frac{y_i^{n+1}-y_i^n}{\tau}=\frac{y_{i-1}^{n+1}-2y_i^{n+1}+y_{i+1}^{n+1}}{h^2}+f(x_i,t_{n+1}),\quad (x_i,t_{n+1})\in\omega_{\tau h} \tag{4}$$

$$\begin{cases} y_0^{n+1}=\mu_1(t_{n+1}) \\ y_n^{n+1}=\mu_2(t_{n+1}) \end{cases} \qquad t_{n+1}\in\overline{\omega}_\tau \tag{5}$$

$$y_i^0=U_0(x_i),\quad x_i\in\overline{\omega}_h \tag{6}$$

Мы поставили в соответствие задаче (1) - (3) разностную схему (4) - (6). Записанная разностная схема использует два слоя : t_{n+1} и t_n. В слое t_{n+1} берется три узла : $(i-1)$, (i), $(i+1)$, а в слое t_n только (i)-ый узел. Данная разностная схема также является четырехточечной:



Уравнение (4) запишем в виде:

$$-\gamma y_{i+1}^{n+1}+(1+2\gamma)y_i^{n+1}-\gamma y_{i-1}^{n+1}=F_i(y^n),\quad i=\overline{1,N-1}$$

$$-\gamma y_{i+1}^{n+1}+(1+2\gamma)y_i^{n+1}-\gamma y_{i-1}^{n+1}=F_i(y^n)$$

$$F_i(y^n)=y_i^n+\tau f_i^{n+1}$$

Эта система решается методом прогонки. Выпишем трехдиагональную матрицу.

$$A=\begin{pmatrix} 1+2\gamma & -\gamma & 0 & \dots & 0 & 0 \\ -\gamma & 1+2\gamma & -\gamma & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1+2\gamma & -\gamma \\ 0 & 0 & 0 & \dots & -\gamma & 1+2\gamma \end{pmatrix}$$

Матрица A с диагональным преобладанием, следовательно решение разностной задачи существует и единственно и оно находится методов прогонки. Для того, чтобы показать сходимость, нужно знать, в какой норме доказывают сходимость. Введем сеточную функцию , которая называется погрешностью решения:

$$z_i^n=y_i^n-u(x_i,t_n)=y_i^n-u_i^n.$$

Запишем задачу для погрешности (уравнение для нахождения погрешности будет иметь такой же вид, что и (4), за исключением того, что вместо аппроксимации правой части будет погрешность аппроксимации на решение, а краевые и начальные условия будут нулевыми):

$$\frac{z_i^{n+1}-z_i^n}{\tau}=\frac{z_{i-1}^{n+1}-2z_i^{n+1}+z_{i+1}^{n+1}}{h^2}+\psi_i^n, \tag{7}$$

$$z_0^{n+1} = z_N^{n+1} = 0, \tag{8}$$

$$z_i^0 = 0. \tag{9}$$

Получим априорную оценку решения, используя, фактически, принцип максимума. Этот подход позволяет получать оценку в сильной норме — в норме $\|\cdot\|_c$.

Для начала выпишем ψ_i^n:

$$\psi_i^n = -\frac{u_i^{n+1} - u_i^n}{\tau} + \frac{u_{i-1}^{n+1} - 2u_i^{n+1} + u_{i+1}^{n+1}}{h^2} + f_i^{n+1} \tag{10}$$

По определению это есть погрешность аппроксимации разностной схемы (4)-(6) на решение исходной задачи (1) - (3). Теперь нужно получить оценку нормы z через погрешность аппроксимации ψ, из которой будет следовать сходимость. Найдем на $(n+1)$ слое узел x_{i_0}, где достигается максимум:

$$\max_{0 \leqslant i \leqslant N} |z_i^{n+1}| = |z_{i_0}^{n+1}| = \|z^{n+1}\|_c$$

В этом узле запишем уравнение (7) в виде:

$$\left(1 + \frac{\tau}{h^2}\right) z_{i_0}^{n+1} = \frac{\tau}{h^2}\left(z_{i_0-1}^{n+1} + z_{i_0+1}^{n+1}\right) + z_{i_0}^n + \tau\psi_{i_0}^n.$$

Обозначая $\dfrac{\tau}{h^2} = \gamma$, получаем очевидное неравенство:

$$(1+2\gamma)|z_{i_0}^{n+1}| \leqslant \gamma(|z_{i_0-1}^{n+1}| + |z_{i_0+1}^{n+1}|) + |z_{i_0}^n| + \tau|\psi_{i_0}^n|$$

Если справа вместо модуля поставить норму, то правая часть только усилится:

$$(1+2\gamma)\,|z_{i_0}^{n+1}| \leqslant \gamma\left(\|z^{n+1}\|_c + \|z^{n+1}\|_c\right) + \|z^n\|_c + \tau\|\psi^n\|_c$$

Так как, по предположению, в узле x_{i_0} на $(n+1)$ слое достигается максимум, неравенство принимает вид:

$$(1+2\gamma)\|z^{n+1}\|_c \leqslant 2\gamma\|z^{n+1}\|_c + \|z^n\|_c + \tau\|\psi^n\|_c$$

Отсюда следует:

$$\|z^{n+1}\|_c \leqslant \|z^n\|_c + \tau\|\psi^n\|_c$$

Заметим, что мы получили абсолютно ту же самую оценку, что и была у нас для явной разностной схемы. Но там было предположение, что $\gamma \leqslant 0.5$, а здесь она получена безо всяких ограничений.

Полученное неравенство можно рассматривать, как рекуррентную формулу:

$$\|z^{n+1}\|_c \leqslant \|z^0\|_c + \sum_{k=0}^{n} \tau\|\psi^k\|_c,$$

Учитывая, что $\|z^0\|_c = 0$, так как начальная погрешность нулевая, получим:

$$\|z^{n+1}\|_c \leqslant \sum_{k=0}^{n} \tau\|\psi^k\|_c$$

Известно, что

$$\|\psi^n\|_c \leqslant M(\tau + h^2),$$

где M не зависит от τ и h.

Тогда окончательно получим:

$$\|z^n\|_c \leqslant M_1(\tau + h^2),$$

где $M_1 = TM$ не зависит от τ и h.

Следовательно

$$\|z^{n+1}\|_c \longrightarrow 0, \text{ при } \quad \tau, h \to 0,$$

что означает сходимость разностной схемы к решению исходной задачи с первым порядком точности по τ и вторым по h. Попутно заметим, что если в разностной задаче (4) - (6) взять нулевые краевые условия

$$y_0^{n+1} = y_N^{n+1} = 0,$$

то получим оценку аналогично той, которую получили выше

$$\|y^{n+1}\|_c \leqslant \|u_0\|_c + \tau \sum_{k=0}^{n} \|f^k\|_c$$

Эта оценка означает устойчивость решения по начальному условию и по правой части уравнения.

§3 Симметричная разностная схема (схема Кранка–Никольсона) для первой краевой задачи уравнения теплопроводности

Поставим задачу:

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + f(x,t), \quad 0 < x < 1, \quad 0 < t \leqslant T, \tag{1}$$

$$y(0,t) = \mu_1(t), \quad y(1,t) = \mu_2(t), \quad 0 \leqslant t \leqslant T, \tag{2}$$

$$y(x,0) = u_0(x), \quad 0 \leqslant x \leqslant 1. \tag{3}$$

Введем обозначение:

$$y^n_{\overline{x}x,i} = \frac{y^n_{i+1} - 2y^n_i + y^n_{i-1}}{h^2}$$

и рассмотрим симметричную разностную схему, имеющую вид:

$$\frac{y_i^{n+1} - y_i^n}{\tau} = \frac{1}{2}(y^{n+1}_{\overline{x}x,i} + y^n_{\overline{x}x,i}) + f(x_i, t_{n+\frac{1}{2}})$$

здесь

$$(x_i, t_{n+\frac{1}{2}}) = (x_i, t_n + \frac{1}{2}\tau) \in w_{\tau_n} \tag{4}$$

Добавим краевые и начальное условия:

$$y_0^{n+1} = \mu_1(t_{n+1}), \quad y_N^{n+1} = \mu_2(t_{n+1}), \quad t_{n+1} \in \overline{w_\tau} \tag{5}$$

$$y_i^0 = u_0(x_i), \quad x_i \in \overline{w_n} \tag{6}$$

Итак, мы поставили в соответствие задаче (1)-(3) разностную схему (4)-(6).

Покажем, что эта задача будет абсолютно сходящейся, но в среднеквадратичной норме.

Введем погрешность решения:

$$z_i^n = y_i^n - u_i^n$$

Выразим y_i^n из этой погрешности и подставим в уравнение (4):

$$\frac{z_i^{n+1} - z_i^n}{\tau} = \frac{1}{2}(z_{\overline{x}x,i}^{n+1} + z_{\overline{x}x,i}^n) + \psi_i^n,$$

$$\psi_i^n = -\frac{u_i^{n+1} - u_i^n}{\tau} + \frac{1}{2}(u_{\overline{x}x,i}^{n+1} + u_{\overline{x}x,i}^n) + f(x_i, t_{n+\frac{1}{2}}) \tag{7}$$

Задача. *Показать, что*

$$\psi_i^n = O(\tau^2 + h^2)$$

Решение: Разложим $u_{i\pm1}^{n+1}$ и u_i^n в ряд Тейлора в окрестности точки $(x_i, t_{n+\frac{1}{2}})$:

$$u_i^{n+1} = u_i^{n+\frac{1}{2}} + u_{t,i}^{n+\frac{1}{2}}\frac{\tau}{2} + \frac{1}{2}u_{tt,i}^{n+\frac{1}{2}}\left(\frac{\tau}{2}\right)^2 + O(\tau^3)$$

$$u_i^n = u_i^{n+\frac{1}{2}} - u_{t,i}^{n+\frac{1}{2}}\frac{\tau}{2} + \frac{1}{2}u_{tt,i}^{n+\frac{1}{2}}\left(\frac{\tau}{2}\right)^2 + O(\tau^3)$$

Подставим эти разложения в формулу для ψ_i^n:

$$\psi_i^n = 0.5(u_{\overline{x}x,i}^{n+1} + u_{\overline{x}x,i}^n) - \frac{u_i^{n+1} - u_i^n}{\tau} + f(x_i, t_n + 0.5\tau) =$$

$$= -u_{t,i}^{n+\frac{1}{2}} + O(\tau^2) + 0.5(u_{\overline{x}x,i}^{n+1} + u_{\overline{x}x,i}^n) + f_i^{n+\frac{1}{2}}$$

Теперь в представлении второй разностной производной разложим все вхождения функции в ряд Тейлора. Приводя подобные слагаемые, получим:

$$u_{\overline{x}x,i}^n = u_{xx,i}^n + u_{xxxx,i}^n\frac{h^2}{12} + O(h^4)$$

Применим это разложение для $u_{\overline{x}x,i}^{n+1}$, а затем проведем еще одно разложение в ряд Тейлора в точке $(x_i, t_{n+\frac{1}{2}})$:

$$u_{\overline{x}x,i}^{n+1} = u_{xx,i}^{n+1} + u_{xxxx,i}^{n+1}\frac{h^2}{12} + O(h^4) =$$

$$= u_{xx,i}^{n+\frac{1}{2}} + u_{xxt,i}^{n+\frac{1}{2}}\frac{\tau}{2} + u_{xxxx,i}^{n+1}\frac{h^2}{12} + u_{xxxxt,i}^{n+1}\frac{h^2}{12}\frac{\tau}{2} + O(\tau^2 + h^4)$$

Тоже самое проделаем для $u_{\overline{x}x,i}^n$:

$$u_{\overline{x}x,i}^{n+1} = u_{xx,i}^{n+1} + u_{xxxx,i}^{n+1}\frac{h^2}{12} + O(h^4) =$$

$$= u_{xx,i}^{n+\frac{1}{2}} - u_{xxt,i}^{n+\frac{1}{2}}\frac{\tau}{2} + u_{xxxx,i}^{n+1}\frac{h^2}{12} - u_{xxxxt,i}^{n+1}\frac{h^2}{12}\frac{\tau}{2} + O(\tau^2 + h^4)$$

Подставим эти разложения в формулу для ψ_i^n и учтем уравнение теплопроводности:

$$\psi_i^n = (-u_{t,i}^{n+\frac{1}{2}} + u_{xx,i}^{n+\frac{1}{2}} + f_i^{n+\frac{1}{2}}) + u_{xxxx,i}^{n+\frac{1}{2}}\frac{h^2}{12} + O(\tau^2 + h^4) = O(\tau^2 + h^2).$$

□

§4 Задача Штурма-Лиувилля

$$\begin{cases} \dfrac{d^2u}{dx^2} + \lambda u(x) = 0, \text{ где} \quad u(x) \not\equiv 0, \quad x \in (0,1) \\ u(0) = u(1) = 0 \end{cases} \tag{8}$$

λ — собственные значения, $u(x)$ — собственные функции.

Решением данной задачи являются собственные значения λ_k и собственные функции $u_k(x)$:

$$\lambda_k = \pi^2 k^2, \quad k = 1, 2, \ldots,$$

$$0 < \lambda_1 < \lambda_2 < \lambda_3 < \ldots < \lambda_n < \ldots,$$

$$u_k(x) = c \cdot \sin(\pi k x),$$

где $c = \text{const} \neq 0$.

Если положить $c = \sqrt{2}$, то получим

$$u_k(x) = \sqrt{2} \cdot sin(\pi k x)$$

и тогда $u_k(x)$ будет образовывать ортонормированный базис в нормированном пространстве $L_2[0,1]$:

$$L_2 : (u_k, u_l) = \delta_{kl}; \quad \{u_k\}_1^\infty.$$

Тогда $\forall f \in L_2$ получаем, что

$$f(x) = \sum_{k=1}^{\infty} c_k u_k(x),$$

где $c_k = (f, u_k)$ – коэффициенты Фурье. В этом случае справедливо равенство Парсеваля:

$$\|f\|_{L_2}^2 = \sum_{k=1}^{\infty} c_k^2$$

Рассмотрим разностную задачу Штурма–Лиувилля:

$$y_{\overline{x}x_i} + \lambda y(x_i) = 0, \quad x_i \in w_h; \quad y(x_i) \not\equiv 0 \tag{9}$$

$$y_0 = y_N = 0. \tag{10}$$

Перепишем задачу в виде:

$$y_{i+1} - 2y_i + y_{i-1} + h^2\lambda y_i = 0$$

Решение будем искать в виде:

$$y(x_i) = \sin(\alpha x_i),$$

где $\alpha \in R$, $i = \overline{1, N-1}$.

Перепишем уравнение в виде $y_{i+1} + y_{i-1} = (2 - h^2\lambda)y_i \quad i = \overline{1, N-1}$. Очевидно, что:

$$y_{i+1} + y_{i-1} = y(x_i + h) + y(x_i - h) = \sin\alpha(x_i + h) + \sin\alpha(x_i - h) = 2\sin(\alpha x_i)\cos(\alpha h)$$

Следовательно:

$$2\cos(\alpha h)\sin(\alpha x_i) = (2 - h^2\lambda)\sin\alpha x_i$$

Так как $\sin(\alpha x_i) \not\equiv 0$:

$$2\cos\alpha h = (2 - h^2\lambda) \quad \Longrightarrow \quad h^2\lambda = 2(1 - \cos\alpha h) = 4\sin^2\left(\frac{\alpha h}{2}\right).$$

Откуда следует, что

$$\lambda = \frac{4}{h^2}\sin^2\left(\frac{\alpha h}{2}\right).$$

Для того, чтобы найти α воспользуемся краевым условием для y, а именно

$$y_N = 0 = \sin\alpha,$$

откуда следует, что $\alpha_k = \pi k$.

Тогда собственные значения λ_k равны

$$\lambda_k = \frac{4}{h^2} \cdot \frac{\sin^2(\pi k h)}{2}, \quad k = \overline{1, N-1}. \tag{11}$$

Теперь можно выписать собственные функции:

$$y(x_i) = C\sin(\pi k x_i), \quad x_i \in \overline{w}_h. \tag{12}$$

Рассмотрим линейное пространство H сеточных функций размерности $N - 1$

$$\dim H_{N-1} = N - 1,$$

со скалярным произведением

$$\forall f, g \in H_{N-1}: \quad (f, g) = \sum_{i=1}^{N-1} f_i g_i h$$

и нормой:

$$\|f\|_{L_2(w_h)} = \|f\| = \left(\sum_{i=1}^{N-1} f_i^2 h \right)^{\frac{1}{2}}.$$

Если взять $C = \sqrt{2}$, то система функций $\{y_k\}_{k=1}^{N-1}$ будет образовывать ортонормированный базис в смысле этого скалярного произведения.

$$y_k(x_i) = \sqrt{2} \sin \pi k x_i, \quad i, k = \overline{1, N-1}$$

А это означает, что

$$(y_k, y_e) = \delta_{k,e} \quad \forall f \in H_{N-1} \quad f = \sum_{k=1}^{N-1} C_k y_k$$

Для удобства обозначим

$$y_k(x_i) = \mu_k(x_i);$$

$\{\mu\}_1^{N-1}$ — ортонормированный базис из собственных векторов.

Вновь справедливо равенство Парсеваля:

$$\|f\|^2_{L_2(w_h)} = \sum_{k=1}^{N-1} C_k^2.$$

Теперь можем переходить к получению априорной оценки. Решение задачи будем искать в виде:

$$z_i^n = \sum_{k=1}^{N-1} c_k(t_n) \mu_k(x_i), \qquad (*)$$

где $x_i \in \overline{w}_h$, $\{\mu_k\}_1^{N-1}$ — ортонормированный базис из собственных функций.

Тогда погрешность аппроксимации тоже будет представляться в виде:

$$\psi_i^n = \sum_{k=1}^{N} \psi^{(k)}(t_n) \mu_k(x_i) \qquad (**)$$

Подставляя (*) и (**) в исходное уравнение, получим:

$$\frac{\sum_{k=1}^{N} (c_k(t_{n+1}) - c_k(t_n))}{\tau} \mu_k(x_i) = \sum_{k=1}^{N-1} \frac{1}{2} (c_k(t_{n+1}) + c_k(t_n)) \mu^{(k)}_{\overline{x}x,i} + \sum_{k=1}^{N-1} \psi^{(k)} \mu_k(x_i)$$

Учитывая, что $\mu^{(k)}_{\overline{x}x,i} = -\lambda_k \mu^{(k)}_{x_i}$ и приравнивая все коэффициенты при $\mu^{(k)}(x)$, получим:

$$\frac{c_k(t_{n+1}) - c_k(t_n)}{\tau} + \frac{1}{2} \lambda_k (c_k(t_{n+1}) + c_k(t_n)) = \psi^{(k)}(t_n), \quad k = \overline{1, N-1}$$

Получили задачу для нахождения функции c_k. Домножим все на τ и соберем подобные члены:

$$(1 + 0.5\tau\lambda_k) \cdot c_k(t_{n+1}) = (1 - 0.5\tau\lambda_k) \cdot c_k(t_n) + \tau\psi^{(k)}(t_n)$$

Так как $(1 + 0.5\tau\lambda_k) \neq 0$, то окончательно получим:

$$c_k(t_{n+1}) = \frac{(1 - 0.5\tau\lambda_k)}{(1 + 0.5\tau\lambda_k)} c_k(t_n) + \frac{\tau}{(1 + 0.5\tau\lambda_k)} \psi^{(k)}(t_n)$$

Обозначим $q_k = \frac{(1-0.5\tau\lambda_k)}{(1+0.5\tau\lambda_k)}$, $|q_k| < 1$.

Задача. *Показать, что*

$$|q_k| = \left| \frac{1 - 0.5\tau\lambda_k}{1 + 0.5\tau\lambda_k} \right| \leqslant 1$$

Теперь можно подставить значение c_k в формулу решения (*):

$$z_i^{(n+1)} = \sum_{k=1}^{N-1} c_k(t_n + 1)\mu_k(x_i) = \sum_{k=1}^{N-1} q_k c_k(t_n)\mu_k(x_i) + \sum_{k=1}^{N-1} \frac{\tau}{(1 + 0.5\tau\lambda_k)} \psi^{(k)}(t_n)\mu_k(x_i)$$

Обозначим первую сумму через V_i^{n+1}, а вторую — W_i^{n+1}. Ясно, что если оценить эти функции в среднеквадратичной норме, то получим:

$$\|z^{n+1}\|_{L_2(w_h)} = \|z^{n+1}\| \leqslant \|V^{n+1}\| + \|W^{n+1}\|$$

Оценим квадрат нормы V_i^{n+1}, используя равенство Парсеваля:

$$\|V^{n+1}\|^2 = \sum_{k=1}^{N-1} q_k^2 c_k^2(t_n) \leqslant \sum_{k=1}^{N-1} c_k^2(t_n) = \|z^n\|^2$$

Аналогичным образом поступим с W_i^{n+1}:

$$\|W^{n+1}\|^2 = \sum_{k=1}^{N-1} \frac{\tau^2}{(1 + 0.5\tau\lambda_k)} \left(\psi^{(k)}(t_n)\right)^2 \leqslant \tau^2 \sum_{k=1}^{N-1} \left(\psi^{(k)}(t_n)\right)^2 = \tau^2 \|\psi^n\|^2$$

То есть получим оценку

$$\|W^{n+1}\| \leqslant \tau\|\psi^n\|.$$

Следовательно, учитывая эти неравенства, основное неравенство примет вид:

$$\|z^{n+1}\| \leqslant \|z^n\| + \tau\|\psi^n\|.$$

Применим эту оценку как рекуррентную:

$$\|z^{n+1}\| \leqslant \|z^0\| + \sum_{j=0}^{n} \tau\|\psi(t_j)\|,$$

где z^0 — нулевое слагаемое.

А так как $\psi^n = O(\tau^2 + h^2)$, то $\exists M = \text{const} > 0$ не зависящая от τ и h, такая, что

$$\|\psi(t_j)\| \leqslant M(\tau^2 + h^2) \Rightarrow \|z^{n+1}\|_{L_2(w_h)} \leqslant M_1(\tau^2 + h^2),$$

где $M_1 = MT$, не зависит от τ и h.

Это означает, что разностная схема сходится в норме $L_2(w_h)$ со вторым порядком точности по τ и h.

С граничными условиями $y_0^{n+1} = y_N^{n+1} = 0$ получим:

$$\|y^{n+1}\|_{L_2(w_h)} \leqslant \|U_0\|_{L_2(w_h)} + \sum_{j=0}^{n}(\|f^j\|_{L_2(w_h)}\tau)$$

Эта оценка означает устойчивость по начальному условию и правой части уравнения.

§5 Разностная схема с весами. Погрешность аппроксимации на решение

Рассматриваем все ту же задачу (1), (2), (3). Заменим уравнение (1) его разностным аналогом:

$$\frac{y_i^{n+1} - y_i^n}{\tau} = \sigma y_{\overline{x}x,i}^{n+1} + (1-\sigma)y_{\overline{x}x,i}^n + \varphi_i^n \tag{4}$$

а также краевые и начальные условия (5) и (6), где σ - весовой множитель, $\sigma \in \mathbb{R}$, $0 \leqslant \sigma \leqslant 1$.

Ясно что:

1. Если $\sigma = 0$, $\varphi_i^n = f_i^n$, то получим явную разностную схему.
2. Если $\sigma = 1$, $\varphi_i^n = f_i^{n+1}$, то получим чисто неявную разностную схему.
3. Если $\sigma = 0.5$, $\varphi_i^n = f_i^{n+\frac{1}{2}}$, то получим симметричную разностную схему.

Введем функцию погрешности $z_i^n = y_i^n - u_i^n$. Перепишем задачу для погрешности аппроксимации:

$$\frac{z_i^{n+1} - z_i^n}{\tau} = \sigma z_{\overline{x}x,i}^{n+1} + (1-\sigma)z_{\overline{x}x,i}^n + \psi_i^n, \tag{7}$$

где

$$\psi_i^n = \sigma u_{\overline{x}x,i}^{n+1} + (1-\sigma)u_{\overline{x}x,i}^n - \frac{u_i^{n+1} - u_i^n}{\tau} + \varphi_i^n$$

Разложим функцию $u(x,t)$ по формуле Тейлора, предполагая наличие нужной гладкости (до 6-й непрерывной производной по x и до 3-й по t включительно). Разложение будем проводить в узле $(x_i, t_{n+\frac{1}{2}})$ (обозначим $u'_t = \dot{u}$, $u'_x = u'$):

$$u_{i+1} = u_i + hu'_i + \frac{h^2}{2}u''_i + \frac{h^3}{6}u'''_i + \frac{h^4}{24}u_i^{IV} + \dots,$$

$$u_{i-1} = u_i - hu_i' + \frac{h^2}{2}u_i'' - \frac{h^3}{6}u_i''' + \frac{h^4}{24}u_i^{IV} + \dots,$$

$$u_i^{n+1} = u_i(t_{n+\frac{1}{2}}) + \frac{\tau}{2}\dot{u}_i(t_{n+\frac{1}{2}}) + \frac{\tau^2}{8}\ddot{u}_i(t_{n+\frac{1}{2}}) + \frac{\tau^3}{48}\dddot{u}_i(t_{n+\frac{1}{2}}) + \dots,$$

$$u_i^n = u_i(t_{n+\frac{1}{2}}) - \frac{\tau}{2}\dot{u}_i(t_{n+\frac{1}{2}}) + \frac{\tau^2}{8}\ddot{u}_i(t_{n+\frac{1}{2}}) - \frac{\tau^3}{48}\dddot{u}_i(t_{n+\frac{1}{2}}) + \dots,$$

Используя эти формулы, получаем:

$$\frac{u_i^{n+1} - u_i^n}{\tau} = \dot{u}_i(t_{n+\frac{1}{2}}) + O(\tau^2)$$

Учитывая выписанные выше разложения, легко получаем

$$u_{\overline{x}x,i} = u_i'' + \frac{h^2}{12}u_i^{IV} + O(h^4)$$

Откуда следует, что

$$u_{\overline{x}x,i} - u_i'' = \frac{h^2}{12}u_i^{IV} = O(h^2)$$

Теперь можем оценить погрешность аппроксимации на решение:

$$\psi_i^n = \sigma\left(u'' + \frac{\tau}{2}\dot{u}'' + \frac{h^2}{12}u^{IV} + O(\tau h^2) + O(h^4)\right) +$$

$$+(1-\sigma)\left(u'' - \frac{\tau}{2}\dot{u}'' + \frac{h^2}{12}u^{IV} + O(\tau h^2) + O(h^4)\right) - \dot{u} + \varphi_i^n + O(\tau^2 + h^4) \qquad (*)$$

Для получения 4–го порядка погрешности аппроксимации по h необходимо убрать все члены порядка h^2, то есть $\frac{h^2}{12}u^{IV}$. Для этого рассмотрим уравнение

$$u'' = \dot{u} - f,$$

продифференцировав его 2 раза по x, получим:

$$u^{IV} = \dot{u}'' - f''.$$

Подставляя это в (*), будем иметь:

$$\psi_i^n = u'' - \dot{u} + \varphi_i^n + \tau(\sigma - 0.5)\dot{u}'' + \frac{h^2}{12}u^{IV} + O(\tau^2 + h^4) =$$

$$= \underbrace{u'' - \dot{u} + f(x_i, t_{n+\frac{1}{2}})}_{0} + \varphi_i^n - f(x_i, t_{n+\frac{1}{2}}) + \left[(\sigma - 0.5) + \frac{h^2}{12}\right]\tau\dot{u}^{IV} - \frac{h^2}{12}f''(x_i, t_{n+\frac{1}{2}}) + O(\tau^2 + h^4)$$

Выберем σ так, чтобы коэффициент $\left[(\sigma - 0.5) + \frac{h^2}{12}\right]$ обратился в ноль:

$$\sigma_* = \frac{1}{2} - \frac{h^2}{12\tau}.$$

Теперь, если положить:

$$\varphi_i^n = f(x_i, t_{n+\frac{1}{2}}) + \frac{h^2}{12} f''(x_i, t_{n+\frac{1}{2}})$$

то погрешность аппроксимации имеет порядок $O(\tau^2 + h^4)$. Эта схема называется разностной схемой повышенного порядка точности.

Следовательно, если:

$$\begin{aligned}
&\sigma = 0: \quad \varphi_i^n = f_i^n, \text{ то } \psi_i^n = O(\tau + h^2) \\
&\sigma = 1: \quad \varphi_i^n = f_i^{n+1}, \text{ то } \psi_i^n = O(\tau + h^2) \\
&\sigma = \frac{1}{2}: \quad \varphi_i^n = f_i(t_{n+\frac{1}{2}}), \text{ то } \psi_i^n = O(\tau^2 + h^2)
\end{aligned}$$

При всех остальных σ (не равных 1-3 и σ_*) будем получать $\psi_i^n = O(\tau + h^2)$.

§6 Разностные схемы для уравнения Пуассона (задача Дирихле)

Рассмотрим уравнение Пуассона в области D:

$$\frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2} = f(x_1, x_2), \tag{1}$$

где $(x_1, x_2) \in D$, а $D = \{(x_1, x_2) : 0 < x_1 < l_1; \quad 0 < x_2 < l_2\}$. На границе Γ заданы условия первого рода.

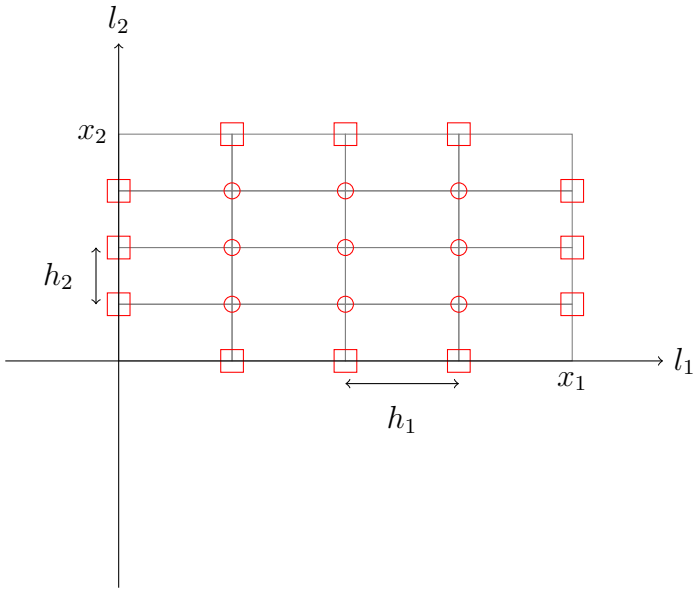
$$u(x_1, x_2)\Big|_\Gamma = \mu(x_1, x_2). \tag{2}$$

Нужно найти функцию, непрерывную в замкнутой области $D \cup \Gamma$, которая удовлетворяет уравнению (1) в D и условиям (2) на границе Γ.

Вводим сетку

$$w_h = \{(x_1^{(i)}, x_2^{(i)}) = x_{ij}, x_1^{(i)} = ih_1, \ i = \overline{1, N_1 - 1}, \ N_1 h_1 = l_1; x_2^{(j)} = jh_2, \ i = \overline{1, N_2 - 1}, \ N_2 h_2 = l_2\}.$$

Ввели множество внутренних узлов. На рисунке они обозначены кружком, граничные узлы - квадраты.



$\Gamma_h = \{x_{0j}\}_1^{N_2-1} \cup \{x_{N_1 j}\}_1^{N_2-1} \cup \{x_{i0}\}_1^{N_i-1} \cup \{x_{iN_2}\}_1^{N_i-1}$. Тогда под $\overline{\omega_h}$ будем понимать $\omega_h \cup \Gamma_h$ - все узлы сетки.

Так же будем использовать обозначение:

$$y_{\overline{x_1}x_1,ij} = \frac{y_{i+1j} - 2y_{ij} + y_{i-1j}}{h_1^2}$$

Поставим в соответствие уравнению (1) разностное уравнение:

$$y_{\overline{x_1}x_1,ij} + y_{\overline{x_2}x_2,ij} = f_{ij}, \text{ где } f_{ij} = f(x_1^{(i)}, x_2^{(j)}) = f(x_{ij});\ x_{ij} = \omega_h \tag{3}$$

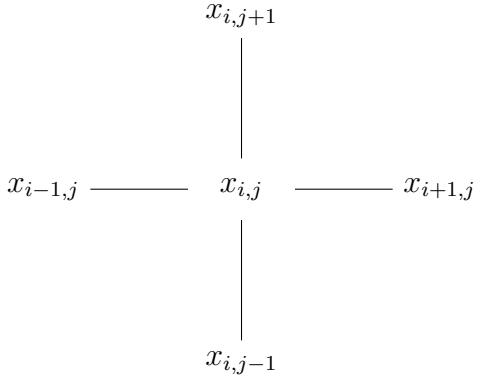
$$y_{ij} = y(x_1^{(i)}, x_2^{(j)}) = y(x_{ij})$$

$$y_{ij} = \mu_{ij} \text{ на границе } \Gamma_h;\ x_{ij} \in \Gamma_h \tag{4}$$

$0 < i < N_1, \quad 0 < j < N_2$

Получили разностную схему (3),(4). Уравнения (3),(4) являются СЛАУ. Для погрешности введем обозначение : $z_{ij} = y_{ij} - U_{ij}$, где $U_{ij} = U(x_{ij}), \quad x_{ij} \in \overline{\omega}_h$

Разностная схема записана на пятиточечном шаблоне:



Для погрешности z_{ij} имеем задачу:

$$z_{\overline{x_1}x_1,ij} + z_{\overline{x_2}x_2,ij} = -\psi_{ij} \tag{5}$$

$$z_{ij} = 0 \text{ на границе } \Gamma_h \tag{6}$$

где ψ_{ij} – погрешность аппроксимации на решение.

$$\psi_{ij} = U_{\overline{x_1}x_1,ij} + U_{\overline{x_2}x_2,ij} - f_{ij} - \text{ невязка} \tag{7}$$

Задача. *Показать, что* $\psi_{ij} = O(h_1^2 + h_2^2)$, *если* $U(x_1, x_2) \in C^4(\overline{D})$

§7 Разрешимость разностной задачи Дирихле. Сходимость разностной схемы

Распишем разностную схему относительно центрального узла x_{ij}:

$$\left(\frac{2}{h_1^2} + \frac{2}{h_2^2}\right) y_{ij} = \frac{y_{i-1j} + y_{i+1j}}{h_1^2} + \frac{y_{ij+1} + y_{ij-1}}{h_2^2} - f_{ij} \tag{8}$$

$y_{ij} = \mu_{ij}$ на границе Γ_h, $\quad 0 < i < N_1, \quad 0 < j < N_2$
Для записи однородной задачи заменим y на V:

$$\left(\frac{2}{h_1^2} + \frac{2}{h_2^2}\right) V_{ij} = \frac{V_{i-1j} + V_{i+1j}}{h_1^2} + \frac{V_{ij+1} + V_{ij-1}}{h_2^2}, \quad x_{ij} \in \omega_h \tag{9}$$

$V_{ij} = 0$ на границе Γ_h

Теорема 6. *Система линейных алгебраический уравнений (9) имеет только тривиальное решение* $V_{ij} = 0$, $x_{ij} \in \overline{\omega_h}$

Доказательство: Доказательство проведем от противного. Пусть существует узел $x_{i_0j_0}$, в котором достигается максимум, отличный от нуля:

$$|V_{i_0j_0}| = \max_{i,j} |V_{ij}| = \|V\|_C \text{ значит } V_{i_0j_0} \neq 0$$

Среди всех таких узлов выберем такой, в котором выполнены следующие два условия:
1) $|V_{i_0j_0}| = \|V\|_C$
2) Хотя бы в одном узле из четырех узлов $(i_0 + 1, j_0), (i_0 - 1, j_0), (i_0, j_0 + 1), (i_0, j_0 - 1)$ выполнено неравенство $|V_{ij}| < |V_{i_0j_0}|$.
Ясно, что такой узел существует. Запишем уравнение в этом узле:

$$\left(\frac{2}{h_1^2} + \frac{2}{h_2^2}\right) V_{i_0j_0} = \frac{V_{i_0-1j_0} + V_{i_0+1j_0}}{h_1^2} + \frac{V_{i_0j_0+1} + V_{i_0j_0-1}}{h_2^2}$$

Оценим по модулю:

$$\left(\frac{2}{h_1^2} + \frac{2}{h_2^2}\right) |V_{i_0j_0}| \leqslant \frac{|V_{i_0-1j_0}| + |V_{i_0+1j_0}|}{h_1^2} + \frac{|V_{i_0j_0+1}| + |V_{i_0j_0-1}|}{h_2^2}$$

Согласно второму условию выбора узла, получим:

$$\left(\frac{2}{h_1^2}+\frac{2}{h_2^2}\right)\|V\|_C<\left(\frac{2}{h_1^2}+\frac{2}{h_2^2}\right)\|V\|_C$$

Откуда получаем противоречие. Значит система имеет только тривиальное решение. ■

Следствие 3. *Разностная схема имеет и притом единственное решение при любых правых частях и граничных условиях.*

Если получим оценку $\|z\|_C \leqslant M(h_1^2+h_2^2)$, где M не зависит от $h_1,\ h_2$, то тогда видим, что при $h_1, h_2 \to 0$ то и $\|z\|_C \to 0$, что будет означать сходимость, так как $\|z\|_C = \|y-U\|_C$
Введем линейный оператор L_h:

$$L_hV_{ij}=\left(\frac{2}{h_1^2}+\frac{2}{h_2^2}\right)V_{ij}-\frac{V_{i-1j}+V_{i+1j}}{h_1^2}-\frac{V_{ij+1}+V_{ij-1}}{h_2^2},\quad x_{ij}\in\omega_h$$

Лемма 3. *Пусть $V_{ij}\geqslant 0$ на границе и пусть $L_hV_{ij}\geqslant 0,\quad x_{ij}\in\omega_h$. Тогда V_{ij} неотрицательна всюду.*

Доказательство: Доказательство проведем от противного. Предположим, что существует узел $x_{i_0j_0}$, в котором $V_{i_0j_0}<0$. Относительно этого узла выберем тот узел i_0j_0, в котором выполнено:
1)$|V_{i_0j_0}|=\min|V_{ij}|$
2) значение хотя бы в одном из оставшихся четырех узлов шаблона $V_{i_0j_0}<V_{ij}$, где $i,j-$ один из четырех узлов $(i_0+1,j_0),(i_0-1,j_0),(i_0,j_0+1),(i_0,j_0-1)$
Рассматриваем действие оператора $L_hV_{i_0j_0}=\left(\frac{2}{h_1^2}+\frac{2}{h_2^2}\right)V_{i_0j_0}-\frac{V_{i_0-1j_0}+V_{i_0+1j_0}}{h_1^2}-\frac{V_{i_0j_0+1}+V_{i_0j_0-1}}{h_2^2}=\frac{V_{i_0j_0}-V_{i_0-1j_0}}{h_1^2}+\frac{V_{i_0j_0}-V_{i_0+1j_0}}{h_1^2}+\frac{V_{i_0j_0}-V_{i_0j_0-1}}{h_2^2}+\frac{V_{i_0j_0}-V_{i_0j_0+1}}{h_2^2}$. Каждое слагаемое в правой части не положительно, но согласно второму условию существует слагаемое, которое отрицательно. Откуда следует, что $L_hV_{i_0j_0}<0$. Получили противоречие. ■

Утверждение. *Рассмотрим две разностные задачи:*

$$L_hy_{ij}=\varphi_{ij},\quad x_{ij}\in\omega_h, y_{ij}\ \text{на границе задано} \tag{10}$$

$$L_hY_{ij}=\Phi_{ij}\quad x_{ij}\in\omega_h, Y_{ij}\ \text{на границе задано} \tag{11}$$

Если $|y_{ij}|\leqslant Y_{ij}$ на границе $(x_{ij}\in\Gamma_h)$, а внутри области $(x_{ij}\in\omega_h)\ \ |\varphi_{ij}|\leqslant\Phi_{ij}$, то $|y_{ij}|\leqslant Y_{ij}$ всюду. $(x_{ij}\in\overline{\omega_h})$

Доказательство: Введем $V_{ij}=Y_{ij}-y_{ij},\quad W_{ij}=Y_{ij}+y_{ij}$. Тогда запишем операторы для каждой из функций: $L_hV_{ij}=\Phi_{ij}-\varphi_{ij}\geqslant 0;\quad V_{ij}\geqslant 0,\quad x_{ij}\in\Gamma_h$. Откуда следует , что $V_{ij}\geqslant 0\ \ x_{ij}\in\overline{\omega_h}$. Аналогично для функции W_{ij}. ■

Рассмотрим задачу относительно погрешности:

$$L_h z_{ij} = \psi_{ij}, \quad x_{ij} \in \omega_h \tag{12}$$

$z_{ij} = 0$ на границе
Возьмем в качестве Y_{ij} следующую функцию:

$$Y_{ij} = k(l_1^2 + l_2^2 - (x_1^{(i)})^2 - (x_2^{(j)})^2), \text{ где константа } k > 0 \tag{13}$$

Следовательно $Y_{ij} \geqslant 0$ при $x_{ij} \in \overline{\omega_h}$

Задача. *Показать,что* $L_h Y_{ij} = 4k$.

$$L_h Y_{ij} = \|\psi\|_C, \quad 4k = \|\psi\|_C \quad x_{ij} \in \omega_h \tag{14}$$

$Y_{ij} \geqslant 0$ на границе
Применим к (13) и (14) следствие:

$$|z_{ij}| \leqslant Y_{ij}, \quad x_{ij} \in \overline{\omega_h}$$

$0 \leqslant Y_{ij} \leqslant \dfrac{l_1^2 + l_2^2}{4}\|\psi\|_C$ следовательно получаем:

$$\|z\|_C \leqslant \frac{l_1^2 + l_2^2}{4}\|\psi\|_C \tag{15}$$

Из этой априорной оценки следует сходимость разностной схемы со вторым порядком по h_1 и h_2. Устойчивость: если рассматриваем разностную схему (3) с нулевым условием на границе, то получаем в точности такую же задачу, как и для z. А это означает, что выполняется оценка:

$$\|y\|_C \leqslant \frac{l_1^2 + l_2^2}{4}\|\varphi\|_C \tag{16}$$

означающая устойчивость разностной схемы по правой части уравнения.

Теорема 7. *Пусть решение $U(x_1, x_2)$ исходной задачи принадлежит классу $C^4(\overline{D})$. Тогда разностная схема (3) (4) сходится к исходной задаче (1) (2) со вторым порядком точности по h_1 и h_2*

Доказательство: Так как $\|\psi\|_C \leqslant M(h_1^2 + h_2^2)$, где $M > 0$ не зависит от $h_1,\ h_2$, то получаем $\|z\|_C \leqslant M_1(h_1^2 + h_2^2)$, где $M_1 = M\dfrac{l_1^2 + l_2^2}{4}$ не зависит от $h_1,\ h_2$. Следовательно, эта оценка говорит о том, что имеет место сходимость со вторым порядком точности по $h_1,\ h_2$ ∎

§8 Методы решения разностных схем для задачи Дирихле

Запишем разностную схему на шаблоне "крест"для $0 < i < N_1, \quad 0 < j < N_2$:

$$\frac{y_{i-1j} - 2y_{ij} + y_{i+1j}}{h_1^2} + \frac{y_{ij-1} - 2y_{ij} + y_{ij+1}}{h_2^2} = f_{ij} \tag{1}$$

$$y_{ij} = \mu_{ij}, \quad x_{ij} \in \Gamma_h \tag{2}$$

Так как нужна хорошая точность, нужно брать $h_1, \ h_2$ очень маленькими, но тогда число уравнений может стать очень большим. Но эти уравнения специального вида: в матрице коэффициентов много нулей. Самый распространенный способ решения таких систем алгебраических уравнений — итерационный метод. Для того чтобы построить итерационный метод, разрешим (1) относительно центрального узла:

$$\left(\frac{2}{h_1^2} + \frac{2}{h_2^2}\right) y_{ij} = \frac{y_{i-1j} + y_{i+1j}}{h_1^2} + \frac{y_{ij-1} + y_{ij+1}}{h_2^2} - f_{ij}$$

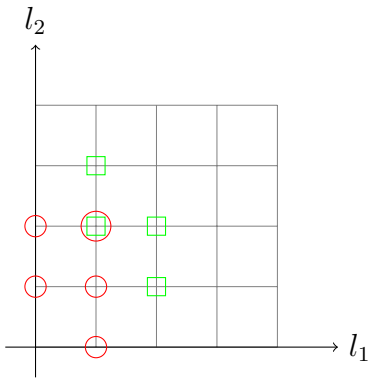
номер итерации будем писать в верхнем индексе. Тогда метод Якоби будем выглядеть следующим образом:

$$\left(\frac{2}{h_1^2} + \frac{2}{h_2^2}\right) y_{ij}^{(s+1)} = \frac{y_{i-1j}^{(s)} + y_{i+1j}^{(s)}}{h_1^2} + \frac{y_{ij-1}^{(s)} + y_{ij+1}^{(s)}}{h_2^2} - f_{ij} \tag{3}$$

начальное приближение $y_{ij}^{(0)}$ задано, $s = 0, 1, 2, \ldots$. Если $h = h_1 = h_2$(или $\max h_1, h_2$), то число итераций в методе Якоби будет пропорционально $O(h^{-2})$ – эта сходимость медленная (следовательно метод не эффективен). Метод Зейделя записывается следующим образом:

$$\left(\frac{2}{h_1^2} + \frac{2}{h_2^2}\right) y_{ij}^{(s+1)} = \frac{y_{i-1j}^{(s+1)} + y_{i+1j}^{(s)}}{h_1^2} + \frac{y_{ij-1}^{(s+1)} + y_{ij+1}^{(s)}}{h_2^2} - f_{ij} \tag{4}$$

начальное приближение $y_{ij}^{(0)}$ задано, $s = 0, 1, 2, \ldots$. Видим ,что здесь сразу разрешить относительно $s + 1$ итерации не удастся. Но все равно все можно найти по явным формулам.



Вначале находим y_{11} по явной формуле. Затем находим $y_{1j},\ j = \overline{1, N_2 - 1}$. Далее находим y_{2j} для $j = \overline{1, N_2 - 1}$ по явным формулам. Таким образом, начиная с узла (1,1) и заканчивая узлом $(N_1 - 1, N_2 - 1)$, все значения y_{ij} находятся по явным формулам. У данного алгоритма тоже медленная сходимость. Кружочком обозначена s–я итерация, квадратом — $(s+1)$–я итерация.

Рассмотрим попеременно треугольный итерационный метод. Запишем данную систему алгебраических уравнений в виде $Ay = \varphi$, где $A = A^* > 0$ (суждение о положительно определенности можно сделать из вида собственных значений). И представим матрицу в следующем виде: $A = R_1 + R_2$ (ниже и верхнетреугольная форма матриц,диагональные элементы которых равны $0.5a_{ij}$).

$$(E + \omega R_1)(E + \omega R_2)\frac{y^{(s+1)} - y^{(s)}}{\tau} + Ay^{(s)} = \varphi \tag{5}$$

где $\tau,\ \omega$ – положительные действительные числа - итерационные параметры, удовлетворяющие условию $\omega > \frac{\tau}{4}$, начальное приближение $y_{ij}^{(0)}$ задано, $s = 0, 1, 2, \ldots$. Этот метод является неявным ,так как $B = (E + \omega R_1)(E + \omega R_2)$, где E - единичная матрица. Этот метод тоже разрешим явными формулами:

$1)(E + \omega R_2)\frac{y^{(s+1)} - y^{(s)}}{\tau} = W^{(s+1)}$

$2)(E + \omega R_1)W^{(s+1)} = \varphi - Ay^{(s)}$

$3)V^{(s+1)} = \frac{y^{(s+1)} - y^{(s)}}{\tau}$

Получается, что из второго находится $W^{(s+1)}$, обращая $(E + \omega R_1)$. Далее из первого находим $V^{(s+1)}$, обращая верхнетреугольную форму, а из третьего досчитываем $V^{(s+1)}$. Если выбираем итерационные параметры правильно, то число итераций будет $O(h^{-1}) = n_0(\varepsilon)$. Попеременно – треугольный итерационный метод сходится на порядок быстрее, чем методы Якоби и Зейделя.

§9 Основные понятия теории разностных схем: аппроксимация, устойчивость, сходимость

Все последующие определения присущи любым линейным классическим задачам. Для нелинейных это будет не так. Пусть решается дифференциальная задача с линейным оператором:

$$LU(x) = f(x), \quad \text{где } x \in G \tag{1}$$

G - некоторая область.

Оператор L – линейный, включает в себя краевые и начальные условия. x – многомерный вектор. Например, для двумерного уравнения теплопроводности это будут переменные x_1, x_2, t. Области G поставим в соответствие множество узлов G_h, следовательно h – норма всех шагов, обобщающая характеристика. Если эта норма стремится к нулю, то количество узлов возрастает. Выбор сетки - серьезный вопрос на практике. В данном параграфе этот вопрос не рассматривается.

После введения сетки вводится сеточная функция y_h и разностный оператор L_h. В итоге получаем

$$L_h y_h(x) = \varphi_h(x), \text{ где } x \in G_h - \text{ узлы сетки} \tag{2}$$

Например, $h = \max h_1, h_2, \ldots, h_n$ Тогда (2) - разностная схема , где h любая норма, шагов столько, какова размерность задачи. Уравнение (2) представляет собой систему алгебраических уравнений, называется разностной схемой, аппроксимирующей исходную задачу. Пусть функция $U(x)$ из нормированного линейного пространства B_0, $x \in G$, а $y_h(x)$ из линейного нормированного пространства B_h, $x \in G_h$. Введем оператор проектирования: $P : B_0 \to B_h$. Тогда функция $P_h U = U_h(x)$, $x \in G_h$. Введем нормы: $\|C\|_0$ - в пространстве B_0, $\|C\|_h$ - в пространстве B_h. нормы должны быть согласованы:

$$\lim_{h \to 0} \|U_h\|_h = \|U\|_0$$

Рассмотрим конкретный пример: пусть область $G = x : 0 \leqslant x \leqslant 1$. Тогда ясно, что сетка будет иметь вид: $G_h = \{x_i = ih,\ i = \overline{0, N},\ hN = 1\}$, $U(x),\ x \in G$, $y_h(x) \in G_h$. Рассмотрим разные нормы: $\|U\|_0 = \max\limits_{0 \leqslant x \leqslant 1} |U(x)| = \|U\|_C$ - норма в B_0,

$$\|y_h\|_h = \max_{0 \leqslant i \leqslant N} |y_i| = \|y\|_C - \text{ норма в } B_h$$

И эти нормы согласованы:

$$\|U\|_0 = \left(\int_0^1 U^2(x) dx \right)^{\frac{1}{2}} - \text{ норма в } L_2, \quad \|y\|_h = \left(\sum_{i=0}^{N} y_i^2 h \right)^{\frac{1}{2}}$$

И эти нормы тоже согласованы. Приведем пример несогласованных норм:

$$\|U\|_0 = \left(\int_0^1 U^2(x) dx \right)^{\frac{1}{2}} \quad \|y\|_h = \left(\sum_{i=0}^{N} y_i^2 \right)^{\frac{1}{2}}$$

Возьмем функцию $U(x) = 1$. Тогда $\|U_h\|_h = \left(\sum_{i=0}^{N} 1 \right)^{\frac{1}{2}} = (N+1)^{\frac{1}{2}}$. Ясно, что если $h \to 0$:

$$\lim_{h \to 0} \|U_h\|_h = \infty.$$

Если нормы не согласованы, то нет гарантии, что при $h \to 0$ $y_h \to U$, где U – единственное решение исходной задачи. Оператор проектирования можно ввести в более общем виде следующим образом:

$$(P_h U)_i = \frac{1}{h} \int_{x_i - 0.5h}^{x_i + 0.5h} U(x) dx$$

во всех внутренних точках $i = \overline{1, N-1}$. На границе:

$$(P_h U)_0 = \frac{1}{0.5h} \int_0^{0.5h} U(x) dx \ , \quad (P_h U)_N = \frac{1}{0.5h} \int_{1-0.5h}^{1} U(x) dx$$

Будем исследовать функцию

$$z_h = y_h - U_h, \tag{3}$$

которая называется погрешностью решения.
Выразим из (3): $y_h = z_h + U_h$. Подставим полученное в начальное уравнение. В силу линейности операторов имеем:

$$L_h z_h + L_h U_h = \varphi_h, \;\; \Rightarrow L_h z_h = \psi_h, \text{ где } \psi_h = \varphi_h - L_h U_h \tag{4}$$

Определение. *Сеточная функция (4) называется погрешностью аппроксимации разностной схемы (2) на решение задачи (1).*

Определение. *Говорят, что разностная схема имеет $k-$й порядок аппроксимации, если существуют положительные константы M_1, k не зависящие от шагов h, для которых справедлива оценка: $\|\psi_h\|_h \leqslant M_1 h^k$, где k не обязательно натуральное (может быть и дробной, главное, чтобы не зависела от шагов).*

Задача (1) корректно поставлена, если выполнены 2 условия:
1) у данной задачи существует и притом единственное решение $U(x) \;\; \forall f(x), \;\; x \in G$
2) решение непрерывно зависит от правой части.

Определение. *Говорят, что разностная схема корректно поставлена, если выполнены два условия:*
1) *у данной задачи существует и притом единственное решение $y_h \in B_h$ при любых правых частях φ*
2) *существует положительная константа M_2 не зависящая от шагов сетки h такая, что выполнена оценка :*

$$\|y_h\|_h \leqslant M_2 \|\varphi_h\|_h \tag{5}$$

Оценка (5) называется априорной оценкой, и она означает устойчивость - это внутреннее свойство разностной схемы.

Определение. *Говорят, что решение разностной задачи сходится к решению исходной задачи, если $\|z_h\|_h = \|y_h - U_h\| \to 0$ при $h \to 0$. Может быть как быстрая сходимость, так и медленная.*

Определение. *Говорят, что разностная схема имеет $k-$й порядок точности, если существуют положительные константы M_3, k не зависящие от шагов h, для которых выполнена оценка $\|z_h\|_h \leqslant M_3 h^k$.*

Теорема 8. *(Филиппова) Пусть исходная задача (1) корректно поставлена и пусть разностная схема (2), аппроксимирующая задачу (1), корректна. Тогда решение разностной задачи сходится к решению исходной задач с порядком погрешности аппроксимации.*

Доказательство: Если разностная схема корректна, то $\|y_h\|_h \leqslant M_2 \|\varphi_h\|_h$ и M_2 не зависит от шагов h. Запишем задачу для погрешности: $\|z_h\|_h \leqslant M_2 \|\psi_h\|_h$
Так как у нас есть аппроксимация, то $\|\psi_h\|_h \leqslant M_1 h^k$, M_1 не зависит от шагов h. Из этих неравенств вытекает оценка:

$$\|z_h\|_h \leqslant M_3 h^k, \text{ где } M_3 = M_2 M_1 - \text{ не зависит от шагов } h \text{ и } \lim_{h \to 0} \|z_h\|_h = 0$$

Причем точность будет такого порядка, как и аппроксимация ∎

Глава 5

Методы решения обыкновенных дифференциальных уравнений и систем ОДУ

§1 Постановка задачи Коши и примеры численных методов интегрирования задачи Коши

$$\begin{cases} \dfrac{du}{dt} = f(t, u(t)), \quad t > 0, \\ u(0) = u_0; \end{cases} \tag{1}$$

$$u(t) = (u_1(t), u_2(t), \ldots, u_m(t))^T,$$
$$f(t, u(t)) = (f_1(t, u(t)), \ldots, f_m(t, u(t))^T,$$

$$|u|^2 = u_1^2 + u_2^2 + \ldots + u_m^2,$$

Рассмотрим параллелепипед

$$R = \{(t, u) : |t| \leqslant a, \quad |u - u_0| \leqslant b\}.$$

Определение. *Функция $f(t, u)$ удовлетворяет в R условию Липшица по второму аргументу, если $\exists L > 0$ и выполнено неравенство:*

$$|f(t, u) - f(t, v)| \leqslant L|u - v|$$

Пусть $f(t, u)$ из (1) удовлетворяет условию Липшица в R. Тогда решение (1) $u(t)$ существует и единственно при $0 < t < T$. Проинтегрируем первое уравнение из (1) и учтем начальное условие:

$$u(t) = u_0 + \int\limits_0^t f(t, u(x))dx$$

На этом представлении основан метод Пикара:

$$u_{n+1}(t) = u_0 + \int\limits_0^t f(t, u_n(x))dx$$

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-Кутта, вторая - многошаговые разностные методы (например, метод Адамса).

Введем сетку:

$$w_\tau = \{t_n = n\tau, \tau > 0, n = 0, 1, 2, \dots\}$$

Пример 1. Схема Эйлера.

Введем обозначения:

$$u_n = u(t_n) \quad f_n = f(t_n, u_n)$$

Тогда схема Эйлера имеет вид:

$$\begin{cases} \dfrac{y_{n+1} - y_n}{\tau} = f(t_n, y_n), \quad t_n \in w_\tau \\ y_0 = u_0, \quad n = 0, 1, \dots \end{cases} \tag{2}$$

Получили явную схему. Явная, так как

$$y_{n+1} = y_n + \tau f_n, \quad n = 0, 1, \dots$$

Запишем погрешность аппроксимации:

$$\psi_n = -\frac{u_{n+1} - u_n}{\tau} + f_n \tag{3}$$

Разложим u_{n+1} в ряд Тейлора в точке t_n:

$$\frac{u_{n+1} - u_n}{\tau} = u'_n + O(\tau)$$

Подставим последнее выражение в (3):

$$\psi_n = -u'_n + f_n + O(\tau)$$

Учитывая, что $-u'_n + f_n = 0$, окончательно получим:

$$\psi_n = O(\tau),$$

Это означает, что разностная схема (2) аппроксимирует исходную задачу.

Если будет доказана оценка $|y_n - u(t_n)| \leqslant M\tau$, где M не зависит от τ, то это будет означать сходимость.

Пример 2. Метод Рунге-Кутта (двухэтапный) или схема предиктор-корректор.

Поставим в соответствие задаче (1) разностную схему, введя полуцелый слой:

$$t_{n+\frac{1}{2}} = t_n + 0.5\tau$$

Метод является двухэтапным, так как для нахождения решения в точке t_{n+1} используются два этапа:

$$t_n \longrightarrow t_{n+\frac{1}{2}} \longrightarrow t_{n+1}$$

Выполним первый шаг по схеме Эйлера:

$$\frac{y_{n+\frac{1}{2}} - y_n}{0.5\tau} = f(t_n, y_n) \tag{4}$$

Второй шаг:

$$\frac{y_{n+1} - y_n}{\tau} = f(t_{n+\frac{1}{2}}, y_{n+\frac{1}{2}}), \tag{5}$$

где $y_0 = u_0$, $n = 0, 1, \ldots$

Далее видно, что из (4) следует:

$$y_{n+\frac{1}{2}} = y_n + 0.5\tau f_n,$$

а из (5) получаем:

$$y_{n+1} = y_n + \tau f(t_{n+\frac{1}{2}}, y_n + 0.5\tau f_n).$$

Позже будет показано, что погрешность аппроксимации этого метода имеет второй порядок по τ. Оценка погрешности общего двухэтапного метода Рунге–Кутта.

Рассмотрим общий вид метода:

$$\begin{aligned}
&\frac{y_{n+1} - y_n}{\tau} = \sigma_1 K_1 + \sigma_2 K_2, \\
&y_0 = u_0, \quad n = 0, 1, 2 \ldots, \quad \sigma_1, \sigma_2 \in \mathbb{R}, \\
&K_1 = f(t_n, y_n), \quad K_2 = f(t_n + a_2\tau, y_n + b_{21}\tau K_1),
\end{aligned}$$

где $\sigma_1, \sigma_2, a_2, b_{21}$ — вещественные числа.

Подставим значения K_1 и K_2 в уравнение:

$$\frac{y_{n+1} - y_n}{\tau} = \sigma_1 f(t_n, y_n) + \sigma_2 f(t_n + a_2\tau, y_n + b_{21}\tau f(t_n, y_n))$$

Тогда можем записать погрешность аппроксимации на решение (1):

$$\psi_n = -\frac{u_{n+1} - u_n}{\tau} + \sigma_1 f(t_n, u_n) + \sigma_2 f\left(t_n + a_2\tau, u_n + b_{21}\tau f(t_n, u_n)\right) \tag{6}$$

Разложим u_{n+1} в ряд Тейлора в окрестности точки (t_n, u_n):

$$\frac{u_{n+1} - u_n}{\tau} = u'_n + \frac{\tau}{2} u''_n + O(\tau^2)$$

Далее разложим $f(t_n + a_2\tau, u_n + b_{21}\tau f_n)$ в окрестности точки (t_n, u_n):

$$f(t_n + a_2\tau, u_n + b_{21}\tau f_n) = f(t_n, u_n) + a_2\tau\frac{\partial f_n}{\partial t} + b_{21}\tau f_n\frac{\partial f_n}{\partial u} + O(\tau^2)$$

Перепишем ψ_n с учетом проведенных преобразований:

$$\psi_n = -\left(u'_n + \frac{\tau}{2}u''_n + O(\tau^2)\right) + \sigma_1 f_n + \sigma_2\left(f_n + a_2\tau\frac{\partial f_n}{\partial t} + b_{21}\tau f_n\frac{\partial f_n}{\partial n}\right) + O(\tau^2).$$

Заметим, что

$$u''_n = \frac{d}{dt}\left(\frac{du}{dt}\right) = \frac{d}{dt}\left(f_n\right) = \frac{\partial f_n}{\partial t} + f_n\frac{\partial f_n}{\partial t}.$$

Тогда погрешность аппроксимации приобретет вид:

$$\psi_n = -\left(u'_n + 0.5\tau\left(\frac{\partial f_n}{\partial t} + f_n\frac{\partial f_n}{\partial n}\right)\right) + (\sigma_1 + \sigma_2)\, f_n + \sigma_2 a_2\tau\frac{\partial f_n}{\partial t} + \sigma_2 b_{21}\tau\frac{\partial f_n}{\partial u} + O(\tau^2) =$$
$$= -u'_n + (\sigma_1 + \sigma_2)f_n + (\sigma_2 a_2 - 0.5)\tau\frac{\partial f_n}{\partial t} + \tau(\sigma_2 b_{21} - 0.5)f_n\frac{\partial f_n}{\partial n} + O(\tau^2).$$

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 погрешности аппроксимации)

Если выполнено условие (1), то $\psi_n = O(\tau)$, а если выполнены оба условия, $\psi_n = O(\tau^2)$. Положим $\sigma_2 = \sigma$, а $\sigma_1 = 1 - \sigma$, тогда условие 1 автоматически выполнено и мы получим однопараметрические семейство:

$$\frac{y_{n+1} - y_n}{\tau} = (1 - \sigma)K_1 + \sigma K_2.$$

В примере предиктор-корректор параметры имели следующие значения:

$$a_2 = b_{21} = 0.5, \quad \sigma = 1$$

Если взять $\sigma = 0.5$, $b_{21} = a_2 = 1$, то получим симметричную схему:

$$\frac{y_{n+1} - y_n}{\tau} = 0.5(f(t_n, y_n) + f(t_{n+1}, y_{n+1})).$$

§2 Методы Рунге-Кутта

Решаем задачу Коши:

$$\begin{cases} \dfrac{du}{dt} = f(t, u(t)), \quad t > 0 \\ u(0) = u_0 \end{cases} \tag{1}$$

Поговорим об m–этапном методе Рунге–Кутта. Идея заключается в переходе $t_n \to t_{n+1}$, используя m промежуточных этапов.

$$\begin{aligned} &\frac{y_{n+1}-y_n}{\tau} = \sigma_1 K_1 + \sigma_2 K_2 + \ldots + \sigma_m K_m \\ &y_0 = u_0, \quad n = 0, 1, \ldots, \end{aligned} \tag{2}$$

где

$$\begin{aligned} &K_1 = f(t_n, y_n), \\ &K_2 = f(t_n + a_2\tau, y_n + b_{21}\tau K_1), \\ &K_3 = f(t_n + a_3\tau, y_n + b_{31}\tau K_1 + b_{32}\tau K_2), \\ &\ldots \\ &K_m = f(t_n + a_m\tau, y_n + b_{m1}\tau K_1 + b_{m2}\tau K_2 + \ldots + b_{mm-1}\tau K_{m-1}). \end{aligned}$$

Условие аппроксимации:

$$\sum_{i=1}^{m} \sigma_i = 1$$

На практике редко используются методы Рунге–Кутта для $m > 4$. Приведем примеры разностных методов Рунге–Кутта, имеющих третий и четвертый порядок погрешности аппроксимации.

Пример. Схема Рунге–Кутта 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: $\psi_n = O(\tau^3)$.

Пример. Схема Рунге–Кутта 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: $\psi_n = O(\tau^4)$.

Оценка точности на примере двухэтапного метода Рунге-Кутта

Все сложности порождены нелинейностью задачи

$$\begin{aligned} &\frac{du}{dt} = f(t, u(t)), \quad t > 0, \\ &u(0) = u_0 \end{aligned} \tag{1}$$

Выше было показано, что если $\sigma a_2 = \sigma b_2 = \dfrac{1}{2}$, то будет второй порядок аппроксимации двухэтапного метода Рунге–Кутта.

$$\begin{aligned} &\frac{y_{n+1} - y_n}{\tau} = (1-\sigma) f_n + \sigma f(t_n + a\tau, y_n + a\tau f_n), \\ &y_0 = u_0, \quad n = 0, 1, \ldots \end{aligned} \tag{2}$$

Как правило σ неотрицательно. Для удобства положим $0 \leqslant \sigma \leqslant 1$ и $\sigma a = \dfrac{1}{2}$.
Введем функцию погрешности z_n:

$$z_n = y_n - u_n$$

Получаем задачу:

$$\begin{aligned} &\frac{z_{n+1} - z_n}{\tau} = -\frac{u_{n+1} - u_n}{\tau} + (1-\sigma) f(t_n, y_n) + \sigma f(t_n + a\tau, y_n + a\tau f(t_n, y_n)), \\ &z_0 = 0, n = 0, 1, 2, \ldots \end{aligned} \tag{3}$$

Для сходимости нужно показать, что

$$|z_n| \longrightarrow 0, \quad n \longrightarrow \infty.$$

Покажем, что $|z_n| \leqslant M|\psi|$, где M не зависит от τ. Перепишем (3) в эквивалентном виде, только сформировав погрешность аппроксимации:

$$\begin{aligned} \frac{z_{n+1} - z_n}{\tau} = -\frac{u_{n+1} - u_n}{\tau} + (1-\sigma) f(t_n, u_n) + \sigma f(t_n + a\tau, u_n + a\sigma f(t_n, u_n)) + \\ + (1-\sigma)(f(t_n, y_n) - f(t_n, u_n)) + \sigma(f(t_n + a\tau, y_n + a\tau f(t_n, y_n)) - f(t_n + a\tau, u_n + a\tau f(t_n, u_n))) = \\ = \psi_n + (1-\sigma)\varphi_n^{(1)} + \sigma\varphi_n^{(2)}, \end{aligned}$$

где ψ_n, $\varphi_n^{(1)}$, $\varphi_n^{(2)}$ — введенные функции.
То есть получим:

$$\frac{z_{n+1} - z_n}{\tau} = \psi_n + (1-\sigma)\varphi_n^{(1)} + \sigma\varphi_n^{(2)}$$

Введем допущение: функция f по второму аргументу удовлетворяет условию Липшица с константой L. Оценим, исходя из этого допущения, $\varphi_n^{(1)}$ и $\varphi_n^{(2)}$:

$$|\varphi_n^{(1)}| = |f(t_n, y_n) - f(t_n, u_n)| \leqslant L|y_n - u_n| = L|z_n|;$$

$$\begin{aligned} |\varphi_n^{(2)}| \leqslant |f(t_n + a\tau, y_n + a\tau f(t_n, y_n)) - f(t_n + a\tau, u_n + a\tau f(t_n, u_n))| \leqslant \\ \leqslant L|y_n + a\tau f(t_n, y_n) - u_n - a\tau f(t_n, u_n)| \leqslant L|y_n - u_n| + a\tau |f(t_n, y_n) - f(t_n, u_n)|) \leqslant L(|z_n| + a\tau L|z_n|) \end{aligned}$$

Теперь сделаем предположение следующего характера:

$$\sigma a \leqslant \frac{1}{2};$$

Тогда

$$z_{n+1} = z_n + \tau\psi_n + (1-\sigma)\tau\varphi_n^{(1)} + \sigma\varphi_n^{(2)}$$

Легко видеть, что

$$|z_{n+1}| \leqslant \tau|\psi_n| + (1-\sigma)\tau L|z_n| + \sigma\tau L(|z_n| + a\tau L|z_n|) = \\ = \tau L|z_n| + \tau L\left(\sigma + \sigma a\tau L\right)|z_n| + \tau|\psi_n| + |z_n|;$$

$$|z_{n+1}| \leqslant \tau|\psi_n| + (1 + \tau L + 0.5\tau^2 L^2)|z_n|;$$

Заметим, что $1+\tau L+0.5\tau^2 L^2$ являются первыми членами разложения по Тейлору функции $e^{\tau L}$. Следовательно:

$$|z_{n+1}| \leqslant e^{\tau L}|z_n| + \tau|\psi_n|;$$

обозначая $e^{\tau L} = \rho > 0$, получим оценку:

$$|z_{n+1}| \leqslant \rho|z_n| + \tau|\psi_n|;$$

Это соотношение можно рассматривать как рекуррентную формулу. Легко видеть, что:

$$|z_{n+1}| \leqslant \rho^n|z_0| + \sum_{j=0}^{n} \rho^{n-j}\tau|\psi_j|;$$

$$|z_{n+1}| \leqslant \sum_{j=0}^{n} \rho^{n-j}\tau|\psi_j| = \max_{0\leqslant j\leqslant n}|\psi| \sum_{j=0}^{n} \tau\rho^{n-j} = t_{n+1}\max_{0\leqslant j\leqslant n}|\psi|\rho^n = t_{n+1}e^{t_n L}\|\psi\|_c$$

Окончательно получаем

$$|z_{n+1}| \leqslant M\|\psi\|_c,$$

где $M > 0$ не зависит от τ.

Ясно, что

$$z_n \to 0, \quad \text{при } n \to \infty,$$

то есть имеет место сходимость с погрешностью аппроксимации.

Напомним, что схема предиктор-корректор удовлетворяет условиям:

1. $\sigma = 1, a = 0.5, \quad \psi_n = O(\tau^2) \Longrightarrow |z_{n+1}| \leqslant M\tau^2;$
2. $\sigma = 0.5, a = 1, \quad \psi_n = O(\tau^2) \Longrightarrow |z_{n+1}| \leqslant M\tau^2.$

Если $\sigma = 0, \quad \forall a \quad \psi_n = O(\tau)$, то

$$|z_{n+1}| \leqslant M\tau.$$

§3 Многошаговые разностные методы решения задачи Коши

$$\begin{cases} \dfrac{du}{dt} = f(t, u(t)), \quad t > 0 \\ u(0) = u_0 \quad u_n = u(t_n) \end{cases} \tag{1}$$

Определение. *Линейным m-шаговым разностным методом решения задачи* (1) *называется метод, записанный уравнением:*

$$\sum_{k=0}^{m} \frac{a_k}{\tau} y_{n-k} = \sum_{k=0}^{m} b_k f_{n-k} \tag{2}$$

$$y_{n-k} = y(t_n - k\tau),$$

$$f_{n-k} = f(t_n - k\tau, y_{n-k}),$$

где a_k, b_k, $(k = \overline{0, m})$ действительные числа, причем $a_0 \neq 0, b_m \neq 0$, $\tau > 0$.

Если $b_0 = 0$, то метод явный, в противном случае назовем его неявным.

Для начальных вычислений по формуле (2) необходимы значения $y_0, y_1, \ldots, y_{m-1}$ — так называемый "разгонный этап". Это приводит к некоторой сложности, так как нам известно только значение y_0. Остальные, как правило, получаются другими методами. Поэтому будем считать, что все они заданы.

Попробуем сопоставить многошаговые разностные методы с методом Рунге–Кутта.

Плюсы многошагового разностного метода.

1. Компактная красивая формула.
2. Можно легко получить более высокий порядок погрешности аппроксимации.

Минусы многошагового разностного метода.

1. Наличие разгонного этапа, так как $y_1, y_2, \ldots, y_m$ нужно найти, применяя другие методы.
2. При нахождении y_k используются значения $y_{n-1}, y_{n-2}, \ldots, y_{n-m}$ — их нужно помнить.

Если будет совсем высокий порядок, то мы позже убедимся, что это плохо — потеряется устойчивость.

Условие нормировки:

$$\sum_{k=0}^{m} b_k = 1 \tag{3}$$

Оценка погрешности m–шагового разностного метода

Переходим к вычислению оценки погрешности аппроксимации на решение:

$$\begin{aligned} \psi_n &= -\sum_{k=0}^{m} \frac{a_k}{\tau} u_{n-k} + \sum_{k=0}^{m} b_k f(t_n - k\tau, u_{n-k}) \\ u_{n-k} &= u(t_n - k\tau) \end{aligned} \tag{4}$$

Применим формулу Тейлора в окрестности точки t_n:

$$u_{n-k} = \sum_{l=0}^{p} \frac{(-k\tau)^l}{l!} u^{(l)}(t_n) + O(\tau^{p+1})$$

$$f(t_n - k\tau, u_{n-k}) = u'_{n-k} = \sum_{l=0}^{p-1} \frac{(-k\tau)^l}{l!} u^{(l+1)}(t_n) + O(\tau^p)$$

Подставим все это в погрешность аппроксимации и проведем очевидные преобразования:

$$\begin{aligned} \psi_n &= -\sum_{k=0}^{m} \frac{a_k}{\tau} \sum_{l=0}^{p} \frac{(-k\tau)^l}{l!} u^{(l)}(t_n) + \sum_{k=0}^{m} b_k \sum_{l=0}^{p-1} \frac{(-k\tau)^l}{l!} u^{(l+1)}(t_n) + O(\tau^p) = \\ &= -\sum_{l=0}^{p} \sum_{k=0}^{m} \frac{a_k}{\tau} \frac{(-k\tau)^l}{l!} u(t_n)^{(l)} + \sum_{l=1}^{p} \sum_{k=0}^{m} b_k \frac{(-k\tau)^{l-1}}{l(l-1)!} u^{(l)}(t_n) + O(\tau^p) = \\ &= -\sum_{k=0}^{m} \frac{a_k}{\tau} u(t_n) + \sum_{l=1}^{p} \left(-\sum_{k=1}^{m} \frac{a_k}{\tau} \frac{(-k\tau)^l}{l!} u^{(l)} + \sum_{k=1}^{m} l b_k \frac{(-k\tau)^{l-1}}{l!} u^l(t_n) \right) + O(\tau^p) = \\ &= -\sum_{k=0}^{m} \frac{a_k}{\tau} u(t_n) - \sum_{l=1}^{p} \left(\sum_{k=0}^{m} \frac{(-k\tau)^{l-1}}{l!} u^{(l)}(t_n)(k a_k + l b_k) \right) + O(\tau^p). \end{aligned}$$

Условие аппроксимации:

$$\sum_{k=0}^{m} a_k = 0 \tag{5}$$

Для достижения аппроксимации порядка p должно быть выполнено соотношение:

$$\sum_{k=0}^{m} (k a_k + l b_k) = 0, \tag{6}$$

где $l = \overline{1, p}$.

В многошаговом методе $2m+2$ неизвестных — $a_0, a_1, \ldots, a_m, b_0, b_1, \ldots, b_m$ и $p+2$ уравнений. Чтобы система не была переопределенной, должно выполняться неравенство

$$p + 2 \leqslant 2m + 2 \Longrightarrow p \leqslant 2m,$$

что означает, что порядок погрешности аппроксимации не выше, чем $2m$.

Среди всех многошаговых методов широко известен метод Адамса:

$$\frac{y_n - y_{n-1}}{\tau} = \sum_{k=0}^{m} b_n f_{n-k}$$

$$a_0 = 1, \quad a_1 = -1, \quad a_k = 0, \quad k \geqslant 2$$

Выбирая коэффициенты b_k в правой части, можно достичь любую погрешность.

§4 Понятие устойчивости многошаговых разностных методов

Никакой разностный метод с помощью компьютеров точно не реализуется, потому что:

1. входные данные могут быть не точными
2. будет происходить машинное округление

Рассмотрим для примера следующую схему:

$$y_{n+1} = qy_n \tag{1}$$

$$n = 0, 1, \ldots,$$

где y_0 задано, а q — любое число, возможно комплексное.

Если $|q| > 1$, то процесс вычисление по этой формуле будет неустойчивым, так как на каждом шаге мы находим приближенное значение $\tilde{y}_n = y_n + \delta_n$. Следовательно $\tilde{y}_{n+1} = qy_n + q\delta_n = y_{n+1} + \delta_{n+1}$, откуда видно, что $\delta_{n+1} = q\delta_n$. А так как $|q| > 1$, то $|\delta_{n+1}| \to \infty$.Если же $|q| \leqslant 1$, то $|\delta_{n+1}|$ не возрастает, и мы получим устойчивый метод.

Рассмотрим модельную задачу:

$$\begin{cases} U'(t) + \lambda U(t) = 0, \quad t > 0, \quad \lambda > 0 \\ U(0) = U_0 \end{cases} \tag{2}$$

Её решение имеет вид $U(t) = U_0 e^{-\lambda t}$. Если $\lambda > 0$, то $|U(t)| \leqslant U_0$, т.е. имеет место устойчивость по начальному условию.

Рассмотрим следующую задачу:

$$\begin{cases} \dfrac{dU}{dt} = f(t, U(t)), \quad t > 0 \\ U(0) = U_0 \end{cases} \tag{3}$$

Явная схема Эйлера для задачи (3) представляется в виде

$$\frac{y_n - y_{n-1}}{\tau} = f(t_n, y_n)$$

А применительно к модельной задаче, она будет выглядеть следующим образом:

$$\begin{cases} \dfrac{y_n - y_{n-1}}{\tau} + \lambda y_{n-1} = 0 \\ y_0 = U_0 \end{cases}$$

Мы можем разрешить относительно y_n:

$$y_n = y_{n-1} - \tau\lambda y_{n-1}$$

Следовательно, вводя обозначение $q = 1 - \tau\lambda$, получим

$$y_{n+1} = (1 - \tau\lambda)y_n = qy_n$$

Если $|q| \leqslant 1$, то эта разностная схема устойчива.
Получим, что

$$-1 \leqslant 1 - \tau\lambda \leqslant 1.$$

Правая часть автоматически выполняется, следовательно $\tau\lambda \leqslant 2$ и

$$0 < \tau \leqslant \frac{2}{\lambda} \tag{4}$$

Условие (4) означает устойчивость разностной схемы.
Рассмотрим неявную схему Эйлера для этой задачи:

$$\frac{y_{n+1} - y_n}{\tau} = f(t_{n+1}, y_{n+1})$$

А применительно к модельной задаче, она будет выглядеть следующим образом:

$$\frac{y_{n+1} - y_n}{\tau} + \lambda y_{n+1} = 0$$

Выразим y_{n+1} через y_n.

$$y_{n+1} - \tau\lambda y_{n+1} = y_n$$

$$(1 + \tau\lambda)y_{n+1} = y_n$$

$$y_{n+1} = \frac{1}{1 + \tau\lambda} y_n, \quad q = \frac{1}{1 + \tau\lambda}$$

Видно, что $0 < q < 1$. Следовательно, вне зависимости от шагов метод стал абсолютно устойчивым. Таким образом показано, что для устойчивости дифференциальной задачи (2) разностные методы могут быть как устойчивыми, так и неустойчивыми.

Общий m-шаговый разностный метод

$$\begin{cases} \dfrac{dU}{dt} = f(t, U(t)), t > 0 \\ U(0) = U_0 \end{cases} \tag{1}$$

Модельная задача примет следующий вид:

$$\begin{cases} \dfrac{dU}{dt} + \lambda U(t) = 0, \quad t > 0 \\ U(0) = U_0 \end{cases} \tag{2}$$

Тогда разностный метод для задачи (1) примет вид:

$$\sum_{k=0}^{m} \frac{a_k}{\tau} y_{n-k} = \sum_{k=0}^{m} b_k f_{n-k} \tag{3}$$

где $y_0, y_1, \ldots, y_{m-1}$ заданы, а a_k, b_k не зависят от τ.

Этот же разностный метод, применительно к модельной задаче запишется в виде:

$$\sum_{k=0}^{m} (\frac{a_k}{\tau} + b_k \lambda) y_{n-k} = 0 \tag{4}$$

Перепишем (4) как

$$\sum_{k=0}^{m} (a_k + \tau \lambda b_k) y_{n-k} = 0 \tag{5}$$

Решение данного уравнения ищется в виде $y_j = q^j$. Если эту формулу подставить в уравнение (5), то в силу однородности сократим на q^{n-m} и получим уравнеие

$$F(q, \tau) = \sum_{k=0}^{m} (a_k + \tau \lambda b_k) q^{m-k} = 0. \tag{6}$$

Уравнение (6) называется характеристическим для разностной схемы (4). Для устойчивости необходимо, чтобы его корни по модулю не превосходили 1. Поиск корней уравнения (6) представляет собой, как правило, сложнейшую задачу. Поэтому будем считать, что τ — мало и положим $\tau = 0$. Тогда получим

$$F(q, 0) = \sum_{k=0}^{m} a_k q^{m-k} = 0 \tag{7}.$$

Уравнение (7) также называется характеристическим.

Определение. *Говорят, что разностная схема (3) удовлетворяет условию (α), если все корни характеристического уравнения (7) лежат внутри или на границе единичного круга комплексной плоскости, причем на границе нет кратных корней.*

Теорема 9. *Пусть разностная схема (3) удовлетворяет условию (α) и $|f'_n| \leqslant L$ для $0 \leqslant t \leqslant T$, при $0 \leqslant t_n = n\tau \leqslant T$. Тогда для всех достаточно малых τ выполняется оценка*

$$|y(t_n) - U(t_n)| \leqslant M \left(\sum_{j=m}^{n} \tau |\psi| + \max_{0 \leqslant i \leqslant m-1} |y_i - U(t_i)| \right) ,$$

где M не зависит от τ ($M = M(L,T)$).

Замечание (1)**.** *Метод Адамса удовлетворяет условию (α):*

$$\frac{y_n - y_{n-1}}{\tau} = \sum_{k=0}^{m} b_k f_{n-k}$$

$$y_0 = U_0$$

$$a_0 = 1, \quad a_1 = -1$$

Тогда характеристическое уравнение имеет вид:

$$q^n - q^{n-1} = 0,$$

оно имеет корни $q_0 = 0$ и $q_1 = 1$, причем q_1 — некратный корень.

Замечание (2)**.** *Говорить об условной или безусловной устойчивости не имеет смысла. Она всегда условная, т.к. рассматриваются малые τ.*

Замечание (3)**.** *Для неявных схем наивысший порядок погрешности аппроксимации $p \leqslant 2m$. Для явных схем $p \leqslant 2m - 1$.*

Однако, схемы высокого порядка не удовлетворяют условию (α), т.е. не являются устойчивыми. Наивысший порядок аппроксимации для схем, удовлетворяющих условию (α), следующий:

1. *Для явных схем:*
   - *(a) Если m — четно, то $p \leqslant m + 2$*
   - *(b) Если m — нечетно, то $p \leqslant m + 1$*
2. *Для неявных схем $p \leqslant m$*

Задача. *Доказать, что для разностной схемы*

$$\frac{y_n + 4y_{n+2} - 5y_{n_2}}{6\tau} = \frac{2f_{n-1} + f_{n-2}}{3}$$

погрешность аппроксимации $O(\tau^3)$.

Доказательство.

$$\psi_n = -\frac{U_n + 4U_{n-1} - 5U_{n-2}}{6\tau} + \frac{2f_{n-1} + f_{n-2}}{3}$$

Запишем условия, налагаемые на многошаговый разностный метод для того, чтобы погрешность аппроксимации имела третий порядок:

$$\begin{cases} b_0 = 1 - \sum\limits_{k=1}^{m} b_k \\ a_0 = -\sum\limits_{k=1}^{m} a_k \\ \sum\limits_{k=1}^{m} a_k k = -1 \\ \sum\limits_{k=0}^{m} k^{l-1}(a_k k + b_k) = 0,\ l = 2, 3. \end{cases}$$

В данном случае $m = 2,\ a_0 = \dfrac{1}{6},\ a_1 = \dfrac{2}{3},\ a_2 = -\dfrac{5}{6},\ b_0 = 0,\ b_1 = \dfrac{2}{3},\ b_2 = \dfrac{1}{3}.$ Выписанные условия легко проверяются, следовательно, $\psi_n = O(\tau^3)$. □

§5 Жесткие системы ОДУ

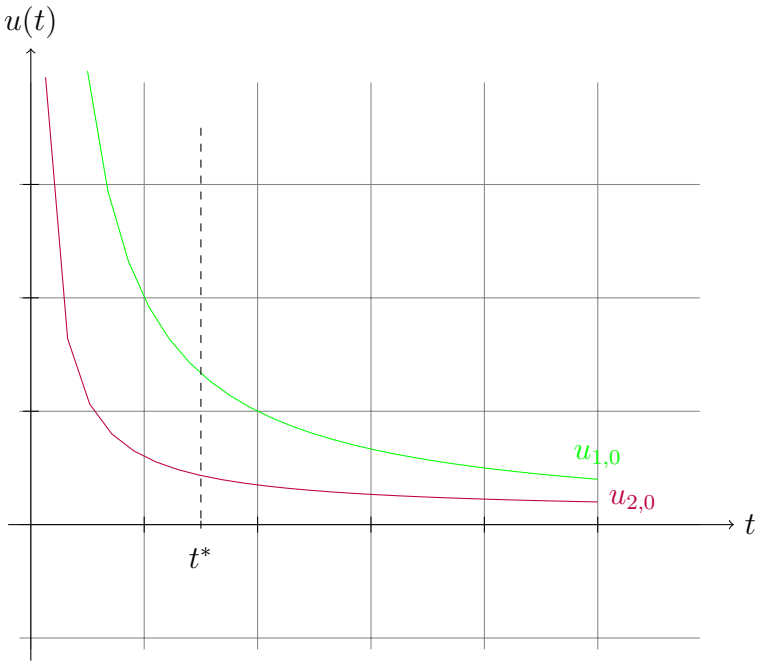
Рассмотрим систему ОДУ:

$$\begin{cases} \dfrac{dU_1(t)}{dt} + a_1 U(t) = 0, \quad t > 0 \\ U_1(0) = U_{1,0} \\ \dfrac{dU_2(t)}{dt} + a_2 U(t) = 0, \quad t > 0 \\ U_2(0) = U_{2,0} \\ a_1 > 0, \quad a_2 > 0, \quad a_2 \gg a_1 \text{ ,то есть } a_2 \text{ много больше } a_1 \end{cases} \tag{1}$$

Решение имеет вид:

$$U_1(t) = U_{1,0} e^{-a_1 t},$$
$$U_2(t) = U_{2,0} e^{-a_2 t}.$$

В приведенном случае одна функция убывает медленно, а другая быстро. Тогда получим,что при $t = t_*$ вторая компонента решения $U_2(t)$ близка к нулю, и если ее находить численно, то шаг сетки можно брать не обязательно маленьким:



1) Рассмотрим явную схему Эйлера:

$$\frac{y_1^{n+1} - y_1^n}{\tau} + a_1 y_1^n = 0 \tag{2}$$

$$\frac{y_2^{n+1} - y_2^n}{\tau} + a_2 y_2^n = 0 \tag{3}$$

Устойчивость схемы (2) достигается при $0 < \tau < \frac{2}{a_1}$, а для (3) $0 < \tau < \frac{2}{a_2}$. А чтобы обеспечить устойчивость системы (1) мы должны выбрать минимальное значение τ. Если учесть, что $a_2 \gg a_1$, то минимальный шаг будет у (3).

2) Рассмотрим неявную схему Эйлера:

$$\frac{y_1^{n+1} - y_1^n}{\tau} + a_1 y_1^{n+1} = 0$$

$$\frac{y_2^{n+1} - y_2^n}{\tau} + a_2 y_2^{n+1} = 0$$

Эти схемы абсолютно устойчивые. А шаг ограничен только условием точности.Введем понятие жесткости системы ОДУ

Задача Коши для линейных уравнений

$$\begin{cases} \dfrac{d\overline{U}}{dt} + A\overline{U}(t) = 0, \quad t > 0 \\ \overline{U}(0) = \overline{U}_0 \end{cases} \tag{5}$$

Где $A = A(m \times m)$ не зависит от t.

Определение. *Система дифференциальных уравнений (5) называется жесткой, если выполняются два условия:*
1) $Re\lambda_k^A > 0, \quad k = \overline{1,m}$
2) $s = \dfrac{\max_{1\leqslant k\leqslant m}|Re\lambda_k^A|}{\min_{1\leqslant k\leqslant m}|Re\lambda_k^A|} >> 1$

Задача Коши для нелинейных уравнений

$$\begin{cases} \dfrac{d\overline{U}}{dt} = \overline{f}(t,\overline{U}), \quad t > 0, \quad \overline{U}(0) = \overline{U}_0 \end{cases} \tag{6}$$

где $\overline{U}(t) = (U_1(t), U_2(t), \ldots, U_m(t))$, $f(t,\overline{U}(t)) = (f_1(t,\overline{U}(t)), f_2(t,\overline{U}(t)), \ldots, f_m(t,\overline{U}(t)))$. Пусть $\overline{V}(t)$ — известное решение. Проведем процесс линеаризации в окрестности известного решения $\overline{V}(t)$: $\overline{z}(t) = \overline{U}(t) - \overline{V}(t)$. Получаем k уравнений вида:

$$\frac{d\overline{z}_k}{dt} = f_k(t,\overline{z}(t) + \overline{V}(t) - f_k(t,\overline{V}(t))), \quad k = \overline{1,m}.$$

Раскладывая в окрестности известного решения относительно точки $(t,\overline{V}(t))$, получим

$$\frac{d\overline{z}_k}{dt} = f_k(t,\overline{V}(t)) + \frac{\partial f_k(t,\overline{V}(t))}{\partial U_1}z_1(t) + \frac{\partial f_k(t,\overline{V}(t))}{\partial U_2}z_2(t) + \cdots + \frac{\partial f_k(t,\overline{V}(t))}{\partial U_m}z_m(t) - f_k(t,\overline{V}(t)) + o(|z|).$$

Обозначим

$$J(t,\overline{V}(t)) = a_{ij} = \frac{\partial f_i(t,\overline{V}(t))}{\partial U_j}, \quad i,j = \overline{1,m}.$$

Следовательно:

$$\frac{d\overline{z}_k}{dt} = J(t,\overline{V}(t))\overline{z} \tag{7}$$

Линейная система (7) — система первого приближения. Теперь введем число жесткости, как отношение

$$s(t) = \frac{max_k|Re\lambda_k^J|}{min_k|Re\lambda_k^J|} \tag{8}$$

Определение. *Система (6) называется жесткой на решение* $\overline{V}(t)$, $0 \leqslant t \leqslant T$, *если выполнено 2 условия:*
1)$Re\lambda_k^J < 0, \quad k = \overline{1,m}$
2)$s(t) \gg 1$

§6 Дальнейшее определение устойчивости и примеры разностных схем, интегрирования жестких систем дифференциальных уравнений

Запишем задачу Коши:

$$\begin{cases} \dfrac{dU}{dt} = f(t, U(t)), \quad t > 0 \\ U(0) = U_0 \end{cases} \tag{1}$$

Запишем модельную задачу:

$$\begin{cases} \dfrac{dU}{dt} = \lambda U(t) \\ U(0) = U_0 \end{cases} \tag{2}$$

где λ — собственные значения матрицы J.

Введем комплексное число $\mu = \tau\lambda$ ($\mu = \mu_0 + i\mu_1$).

Определение. *Областью устойчивости разностного метода называется множество точек комплексной плоскости μ, для которых данный метод, примененный к уравнению (2), устойчив.*

Рассмотрим явную схему Эйлера :

$$\frac{y_{n+1} - y_n}{\tau} = f(t_n, y_n)$$

Применительно к модельной задачи (2) разностная схема примет вид:

$$\frac{y_{n+1} - y_n}{\tau} = \lambda y_n$$

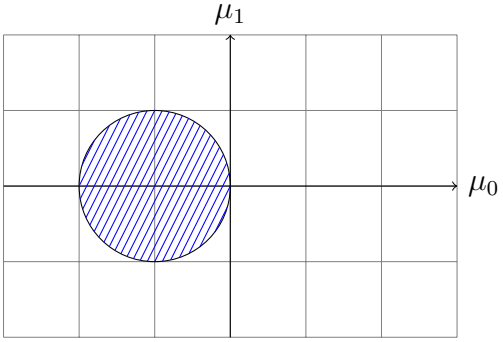
Выразим y_{n+1}:

$$y_{n+1} = y_n + \lambda\tau y_n = (1 + \mu)y_n$$

Метод является устойчивым, если $|q| \leqslant 1$ или, в нашей задаче, $|\mu + 1| \leqslant 1$. Тогда получаем, что

$$|1 + \mu_0 + i\mu_1| \leqslant 1,$$
$$(1 + \mu_0)^2 + \mu_1^2 \leqslant 1.$$

В данном случае область устойчивости представляет собой внутренность круга с центром в точке (0, -1) и радиусом 1 в системе координат (μ_0, μ_1):



Рассмотрим неявную схему Эйлера :

$$\frac{y_{n+1} - y_n}{\tau} = f(t_{n+1}, y_{n+1})$$

Применительно к модельной задачи (2) разностная схема примет вид:

$$\frac{y_{n+1} - y_n}{\tau} = \lambda y_{n+1}$$

Разрешим её относительно y_{n+1}:

$$y_{n+1} = y_n + \lambda\tau y_{n+1},$$

$$(1 - \mu)y_{n+1} = y_n,$$

$$y_{n+1} = \frac{1}{1 - \mu} y_n.$$

Для устойчивости необходимо, чтобы $q = \left|\frac{1}{1 - \mu}\right| \leqslant 1$.

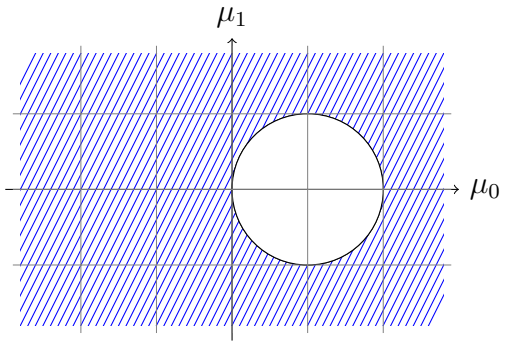
Получаем, что

$$|1 - \mu| \geqslant 1,$$

$$|1 - \mu_0 - i\mu_1| \geqslant 1,$$

$$(1 - \mu_0)^2 + \mu_1^2 \geqslant 1.$$

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 всю левую полуплоскость комплексной плоскости, т.е. $Re\mu < 0$.*

Утверждение. *Если разностный метод A - устойчив, то он абсолютно устойчив, т.е. $\forall\tau > 0$.*

Утверждение. *Явных A-устойчивых методов нет.*

Утверждение. *Среди неявных A-устойчивых методов существуют разностные методы не выше второго порядка.*

Рассмотрим симметричную схему:

$$\frac{y_{n-1} - y_n}{\tau} = 0.5(f(t_n, y_n) + f(t_{n+1}, y_{n+1}))$$

Применительно к модельной задачи (2) разностная схема примет вид:

$$\frac{y_{n-1} - y_n}{\tau} = 0.5\lambda(y_n + y_{n+1}).$$

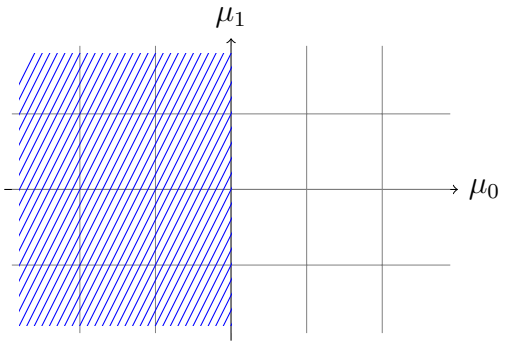
Выразим y_{n+1}:

$$y_{n+1} = y_n + 0.5\mu(y_n + y_{n+1}),$$
$$(1 - 0.5\mu)y_{n+1} = (1 + 0.5\mu)y_n,$$
$$y_{n+1} = \frac{1 + 0.5\mu}{1 - 0.5\mu}.$$

Получаем, что $q = \dfrac{1 + 0.5\mu}{1 - 0.5\mu}$. Найдем область устойчивости, т.е. область, где $|q| \leqslant 1$:

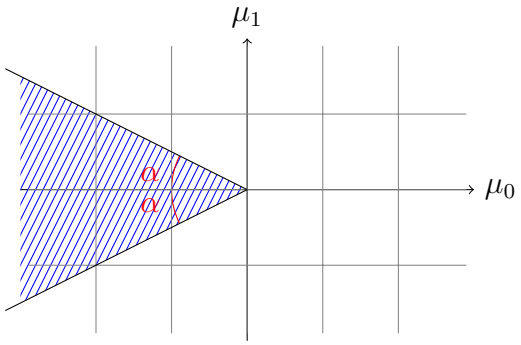
$$|1 + 0.5\mu| \leqslant |1 - 0.5\mu|,$$
$$|1 + 0.5\mu_0 + 0.5i\mu_1| \leqslant |1 - 0.5\mu_0 - 0.5i\mu_1|,$$
$$(1 + 0.5\mu_0)^2 + (0.5\mu_1)^2 \leqslant (1 - 0.5\mu_0)^2 + (0.5\mu_1)^2,$$
$$1 + \mu_0 + \frac{\mu_0^2}{4} + (0.5\mu_1)^2 \leqslant 1 - \mu_0 + \frac{\mu_0^2}{4} + (0.5\mu_1)^2,$$
$$\mu_0 \leqslant 0.$$

Получаем, что симметричная схема является A-устойчивой.



Симметричный разностный метод является одним из лучших разностных методов второго порядка для численного интегрирования жестких систем.

Определение. *Разностный метод называется $A(\alpha)$ устойчивым, если область его устойчивости содержит угол левой полуплоскости, так что $|arg(-\mu) < \alpha|$.*



Утверждение. *Явных $A(\alpha)$-устойчивых методов нет.*

Утверждение. *Среди неявных $A(\alpha)$-устойчивых методов были построены схемы третьего и четвертого порядка.*

Пример чисто неявной $A(\alpha)$-устойчивой разностной схемы четвертого порядка:

$$\frac{25y_{n+1} - 48y_{n+3} + 36y_{n+2} - 16y_{n+1} + 3y_n}{12\tau} = f(t_{n+4}, y_{n+4}).$$