Оглавление

Глава 1

Численные методы в линейной алгебре

§ 1. Введение

Рассмотрим систему линейных алгебраических уравнений:

$$Ax = f,\ |A| \neq 0, \tag{1.1}$$

где $A(m \times m)$, $x = (x_1, ..., x_m)^T$, $f = (f_1, ..., f_m)^T$.

В этом курсе мы будем рассматривать прямые и итерационные методы:

Посчитать определитель по определению: сложность $m!$
По Гауссу: $\frac{m^3}{3}$
Итерационные: $n \to \infty$, $x_n \to x$
$||x_n - x|| < \varepsilon\ n_0(\varepsilon)$

Задача на собственные значения: нормально решается только численно
$Ax = \lambda x, x \neq \vec{0}$
Частичная проблема собственных значений— найти хотя бы одно значение. Полная — все значения (QR-алгоритм). A^{-1} за m^3 действий за счёт хитрости в алгоритме.

Связь метода Гаусса с разложением матрицы на множители.

Метод Крамера — плохо с точки зрения округления и т.п.

Задача приведения матрицы A элементарными преобразованиямик треугольному виду.

Рассмотрим 1.1:

Прямой ход: (матрица A диагональная, c_{ij} сверху справа, 1 по диагонали, 0 внизу)

$\frac{(m^3-m)}{3}$ действий на нахождение коэффициентов матрицы

$\frac{m(m+1)}{2}$ на правые части

$\frac{m(m-1)}{2}$ обратный ход

Возможен случай, когда матрица A представима в виде

$$A = B \times C \tag{1.2}$$

— не каждую матрицу можно представить в таком виде.

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

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

B — нижняя диагональная матрица, C — верхняя диагональная матрица с единицами на диагонали $\Rightarrow$

$a_{ij} = b_{ii}c_{ij} + \sum_{l=1}^{i-1} b_{il}c_{ij}$, $b_{ii} \neq 0$

$$c_{ij} = \frac{a_{ij} - \sum_{l=1}^{i-1} b_{il}c_{ij}}{b_{ii}}, \ i \leq j \tag{1.3}$$

$a_{ij} = \sum_{l=1}^{i-1} b_{il}c_{lj} + b_{ii}c_{ij} + \sum_{l=i+1}^{m} b_{il}c_{lj} = 0 \Rightarrow$

$$b_{ij} = a_{ij} - \sum_{l=1}^{j-1} b_{i,l}c_{l,j}, \ i \geq j \tag{1.4}$$

Формулы 1.3, 1.4 — рекуррентные. Они связывают B и C.

Соотношение нелинейное, но при разумной организации алгоритма на это можно закрыть глаза.

Разумная организация: $b_{11} = a_{11}$

далее по формуле 1.3 $\Rightarrow c_{1j} = \frac{a_{1j}}{b_{11}}$, $j = 2...m$

по формуле 1.4 $\Rightarrow b_{i1} = a_{i1}$, $i = 2...m$

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

Во введении мы получили формулу при $b_{i,i} \neq 0$

$$c_{i,j} = \begin{cases} a_{i,j} - \sum\limits_{l=1}^{j-1} b_{i,l}c_{l,j}, \ i \geq j \\ \\ \frac{a_{i,j} - \sum\limits_{l=1}^{i-1} b_{i,l}c_{l,j}}{b_{i,i}}, \ i \leq j \end{cases}$$

Разложение возможно при определенных условиях. Сформулируем достаточные условия разложения матрицы A: $A = BC$.

Утверждение:

Пусть все главные угловые миноры матрицы A отличны от нуля:

$$\triangle_1 = a_{11} \neq 0$$

$$\triangle_2 = \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} \neq 0$$

$$\triangle_n = \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \dots & \dots & \dots & \dots \\ a_{n,1} & a_{n,2} & \dots & a_{n,n} \end{bmatrix} \neq 0 \ , i = 1, 2, \dots, n$$

Тогда факторизация матрицы A возможна и находится единственным образом.

Доказательство:

Для удобства введем

$$\triangle_0 = 1$$

Распишем матрицу A_i как произведение матриц B_i и C_i: $A_i = B_iC_i$. Определитель произведения матриц есть произведение определителей матриц. Тогда

$$\triangle_i = \mid A_i \mid = \mid B_i \mid\mid C_i \mid,$$

где $\mid C_i \mid = 1$.

Тогда

$$\triangle_i = b_{11}b_{22} \cdots b_{i-1,i-1}b_{i,i} \Rightarrow b_{i,i} = \frac{\triangle_i}{\triangle_{i-1}}, \ i = 1, 2, \dots, n$$

чтд.

Все элементы $b_{i,i}$ отличны от нуля.

Замечание:

Условие отличия от нуля всех угловых миноров является достаточным. На практике это не является жестким требованием. Например в физике и химии присутствуют самосопряженные операторы. Мы не ставили перед собой задачу доказывать утверждение с минимальными требованиями.

Для чего нужна факторизация? Ответ: для удобства решения алгебраических уравнений.

$$Ax = f, \ | \ A \ | \neq 0, \ A = (m, m) \tag{1.5}$$

Рассмотрим связь Метода Гаусса с разложением матрицы A на множители $A = BC$, затем покажем эффективность метода.
$BCx = f \Rightarrow$

$$BY = f \tag{1.6}$$

$$CX = Y \tag{1.7}$$

Решение системы 1.5 распалось на два. Из 1.6 находим Y и подставляем в 1.7, откуда находим X

Задача 1:

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

Решение:

По формулам факторизации

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

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

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

.

Отпустим индекс i:

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

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

Имеем

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

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

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

Отпустим индекс j:

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

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

Суммируем с предыдущим результатом: $2\frac{(m-1)(m+1)m}{6} = \frac{m^3-m}{3}$.

Ясно, что основная работа идет на факторизацию. Сопоставим это с методом Гаусса.

Систему A сводим к верхней треугольной матрице (прямой ход):

$$A = \begin{bmatrix} 1 & c_{1,2} & c_{1,3} & \dots & c_{i,j} \\ 0 & 1 & c_{2,2} & \dots & c_{2,j} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

1-я связь: распишем системы 1.6 и 1.7 по координатам

$$1.6: \ b_{i,1}y_1 + b_{i,2}y_2 + \dots + b_{i,i}y_i = f_i, \ i = 1, 2, \dots, n$$

$$1.7: \ x_i + c_{i,i+1}x_{i+1} + \dots + c_{i,m}x_m = y_i, \ i = 1, 2, \dots, m$$

Выразим из 1.6 вектор y: считая, что $b_{i,i} \neq 0$

$$y_i = \frac{f_i - \sum_{l=1}^{i-1} ab_{i,l}y_l}{b_{i,i}}, \ i = 1, 2, \cdots, m$$

Из 1.7 выразим вектор x:

$$x_i = y_i - \sum_{l=i+1}^{m} c_{i,l}x_l$$

Посчитаем количество действий. В y_i находится $i-1$ умножений и 1 деление. Итого i действий.

Т.к. $i = 1, 2, \ldots, m \Rightarrow$

$$\sum_{i=1}^{m} i = 1 + 2 + \cdots + m = \{ \text{ арифметическая прогрессия} \} = \frac{m(m-1)}{2}$$

В методе Гаусса на преобразование правой части (в прямом ходе метода Гаусса) уходит $\frac{m(m-1)}{2}$ действий.

Рассмотрим 1.7:
$x_i = y_i - \sum\sum c_{i,l}x_l \to (m-i)$ умножений при фиксированном i.

Теперь отпускаем $i \Rightarrow (m-1) + (m-2) + \cdots + 1 = \frac{m(m-1)}{2}$. Столько же в обратном ходе метода Гаусса.

Таким образом метод Гаусса требует $\frac{m^3}{3} + m^2 - \frac{m}{3}$. Большая часть операций уходит на факторизацию.

§ 3. Обращение матриц методом Гаусса-Жордана

Эта тема является очень важной.

Алгебраическая постановка задачи: пусть дана квадратная невырожденная матрица $A(m,m), \mid A \mid \neq 0$.

По определению матрица A называется обратной, если $A^{-1}A = AA^{-1} = E$, где E-единичная матрица.

Существует 2 метода обращения:
1)

Через алгебраические дополнение. Это неэкономичный метод $\Rightarrow (m-1)!$ действий. А мы решаем задачи большого порядка. Например бывают трехмерные нестационарные задачи для уравнения теплопроводности порядка 10^6.

Пусть $A^{-1} = X$. В матрице X m^2 элементов. Тогда имеем

$$AX = E \tag{1.8}$$

СЛАУ $\to \sum\limits_{l=1}^{m} a_{i,l}x_{l,j} = \sigma_{i,j}$ по координатам. Самый экономичный метод. Докажем, что можно затратить m^3 операций на работу данного метода.

Система $AX = f$ может распасться на m систем. Таким образом мы сократим количество действий.

2)

Рассмотрим обращение Гаусса-Жордана. Введем вектор столбец $X_j = (x_{1,j}, \ldots, x_{m,j})^T$

$$\sigma^j = (0, 0, \ldots, 0, 1, 0, \ldots, 0, 0)^T$$

Для обращения необходимо решить m систем

$$AX^j = \sigma^j,\ j = 1, 2, \ldots, m \tag{1.9}$$

Применим факторизацию. Пусть угловые миноры отличны от нуля $\Rightarrow$ $BCX^{(j)} = \sigma^{(j)},\ CX^{(j)} = Y^{(j)}$

$$BY^{(j)} = \sigma^{(j)} \tag{1.10}$$

$$CX^{(j)} = Y^{(j)},\ j = 1, 2, \ldots, m \tag{1.11}$$

В совокупности m^2 действий. Т.к. m систем $\Rightarrow m^3$ действий.

Один раз проведем факторизацию $\Rightarrow m^3 + \frac{m^3-m}{3}$. Итого $\frac{4}{3}m^3 - \frac{m}{3}$.

Если воспользоваться спецификой вида матриц, то возможно получить m^3 действий. Покажем это.

Пусть B – нижняя треугольная матрица. Посмотрим решение системы 1.10:

$$b_{1,1}y_1^{(j)} = 0 \Rightarrow y_1^{(j)} = 0$$

$$b_{2,1}y_1^{(j)} + b_{2,2}y_2^{(j)} = 0 \Rightarrow y_2^j = 0$$

аналогично до $(j-1)$

$$b_{j-1,1}y_1^{(j)} + b_{j-1,2}y_2^{(j)} + \cdots + b_{(j-1),i-1}y_{j-1}^{(j)} = 0 \Rightarrow$$

$$y_j^{(j)} = 0,\ i \le j-1$$

$$b_{j,j}y_j^{(j)} = 1 \Rightarrow y_j^{(j)} = \frac{1}{b_{j,j}},\ i = j$$

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

$$y_i^{(j)} = -\frac{\sum\limits_{l=j}^{i-1} b_{i,l}y_l^{(j)}}{b_{i,i}},\ i = j+1, j+2, \ldots, m$$

Сначала фиксируем i и j и считаем количество действий. Итого 1 деление и (i-j) умножений.

Отпускаем индекс i:

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

умножений при фиксированном i. Еще одно деление при $(i = j)$ и $m - j$ делений. Итого при фиксированном $j \Rightarrow \frac{(m-j+1)(m-j+2)}{2}$ действий.

Отпускаем индекс j, т.к. $j = 1, 2, \ldots, m \Rightarrow \sum_{j=1}^{m} \frac{(m-j+1)(m-j+2)}{2}$ действий.

Задача 2:

Доказать, что для решения системы 1.10 необходимо $\frac{m(m+1)(m+2)}{6}$ и для решения системы 1.11: $\frac{m(m-1)}{2}$, $j = 1, 2, \ldots, m \Rightarrow$ всего $\frac{m^2(m-1)}{2}$ умножений и делений.

Решение:

Пусть $k = m - j + 1 \Rightarrow$

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

$$\frac{k(k+1)}{4} + \frac{k(k+1)(2k+1)}{12} = \frac{3k(k+1) + k(k+1)(2k+1)}{12} = \frac{k(k+1)(k+2)}{6}.$$

Решение системы 1.11 требует $\frac{m(m-1)}{2}$ действий (прямой ход метода Гаусса). Поскольку всего m систем потребуется $\Rightarrow \frac{m^2(m-1)}{2}$ действий умножения и деления для их решения $\Rightarrow$

Общее количество действий на обращение матрицы $\frac{m^3-m}{3} + \frac{m(m+1)(m+2)}{6} + \frac{m^3-m^2}{6} = \frac{2m^3-2m+m^3+3m^2+2m+3m-3}{6} = m^3$.

§ 4. Метод квадратного корня

Рассмотрим систему

$$Ax = f, \tag{1.12}$$

где A-эрмитова невырожденная матрица.

По определению: $a_{i,j} = \bar{a}_{j,i}$, $A = A^*$, $\mid A \mid \neq 0$, $A(m, m)$.

Сузив класс, мы должны получить более сильный результат. Факторизуем матрицу A более хитрым способом в виде $A = S^* D S$

$$D = \begin{bmatrix} d_{1,1} & 0 & 0 & \ldots & 0 & 0 & 0 \\ 0 & d_{2,2} & 0 & \ldots & 0 & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & 0 & 0 & d_{m,m} \end{bmatrix}$$

, где $d_{i,i} = \pm 1$, $i = 1, \ldots, m$.

$$S = \begin{bmatrix} s_{1,1} & s_{2,2} & \ldots & s_{1,m} \\ 0 & s_{2,1} & \ldots & s_{2,m} \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & s_{m,m} \end{bmatrix},$$

где $s_{i,i} > 0,\ i = 1, 2, \ldots, m$.

Пусть $A = \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{1,2} & a_{2,2} \end{bmatrix}$, $D = \begin{bmatrix} d_{1,1} & 0 \\ 0 & d_{2,2} \end{bmatrix}$, $S = \begin{bmatrix} s_{1,1} & s_{1,2} \\ 0 & s_{2,2} \end{bmatrix}$.

$S^* = S^T = \begin{bmatrix} s_{1,1} & 0 \\ s_{1,2} & s_{2,2} \end{bmatrix}$

Тогда $DS = \begin{bmatrix} d_{1,1} & 0 \\ 0 & d_{2,2} \end{bmatrix} \begin{bmatrix} s_{1,1} & s1,2 \\ 0 & s_{2,2} \end{bmatrix} = \begin{bmatrix} d_{1,1}s_{1,1} & d_{1,1}s_{1,2} \\ 0 & d_{2,2}s_{2,2} \end{bmatrix}$

$S^*DS = \begin{bmatrix} s_{1,1} & 0 \\ s_{1,2} & s_{2,2} \end{bmatrix} \begin{bmatrix} d_{1,1}s_{1,1} & d_{1,1}s_{1,2} \\ 0 & d_{2,2}s_{2,2} \end{bmatrix} = \begin{bmatrix} d_{1,1}s_{1,1}^2 & s_{1,1}d_{1,1}s_{1,2} \\ s_{1,1}d_{1,1}s_{1,2} & d_{2,2}s_{2,2}^2 \end{bmatrix} \Rightarrow$

$$\begin{cases} a_{1,1} = d_1 s_{1,1}^2 \\ a_{1,2} = s_{1,1} d_{1,1} s_{1,2} \\ a_{2,2} = d_{2,2} s_2^2 \end{cases}$$

$$\Rightarrow d_{1,1} = sign\ a_{1,1} \rightarrow s_{1,1} = \sqrt{|\ a_{1,1}\ |}$$

$$\Rightarrow s_{1,2} = \frac{a_{1,2}}{s_{1,1} d_{1,1}} \rightarrow d_{2,2} = sign\ a_{2,2}$$

$$\Rightarrow s_{2,2} = \sqrt{|\ a_{2,2}\ |}$$

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

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

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

Для факторизации матрицы система должна быть разрешима.

Таким образом, для вещественной самосопряженной матрицы разложение возможно и находится по явным формулам.

Примечание:

Для корректности действий необходимо наложить соответствующие ограничения на исходную матрицу (так как в процессе преобразований мы выполняли деление). В данном случае, ограничений, накладываемых для обеспечения возможности факторизации матрицы, вполне достаточно.

Рассмотрим теперь общий случай: эрмитову матрицу A в комплексном пространстве, и соответствующую систему линейных алгебраических уравнений:

$$Ax = f,\ |A| \neq 0,\ A(m \times m),\ A = A^*$$

Будем искать представление матрицы A в виде

$$A = S^* D S, \tag{1.13}$$

где S — верхняя треугольная положительная матрица ($s_{ij} > 0$),
$D = diag(d_{11}, \ldots, d_{mm}),\ d_{ii} = \pm 1$
S^* — нижняя треугольная матрица с элементами $\bar{s_{ji}}$.

Так как D - диагональная матрица, то:

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

Учтём, что: $s_{ij} = \bar{s}_{ji}$.

Тогда

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

Выделим из суммы i-ый элемент:

$$a_{ij} = (S^* DS)_{ij} = \sum_{l=1}^{i-1} \bar{s}_{li} d_{ll} s_{lj} + \bar{s}_{ii} d_{ii} s_{ij} + \sum_{l=i+1}^{m} \bar{s}_{li} d_{ll} s_{lj} \tag{1.14}$$

В силу вида матрицы S $s_{li} = 0,\ l > i$, последняя сумма равна 0.

Запишем выражение для a_{ii}:

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

.

Так как $\bar{s}_{li} s_{li} = |s_{li}|^2$, то

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

Отсюда

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

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

(с учётом положительности s_{ij}).

Из исходной формулы 1.14 окончательно находим:

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

Примечание:
"Черту" над s_{ii} ставить необязательно, так диагональные элементы матрицы S вещественны (см. выше).

Таким образом, при определенных условиях исходную матрицу можно преобразовать к виду 1.13.

Рассмотрим применение данного разложения к решению системы линейных алгебраических уравнений:

$$Ax = f$$

$$S^*DSx = f \tag{1.15}$$

Обозначим

$$DSx = y. \tag{1.16}$$

Тогда получим две системы линейных алгебраических уравнений:

$$\begin{cases} S^*y = f \\ DSx = y \end{cases}$$

Тогда обращая S, можем легко посчитать y, а решая уравнение 1.16 - посчитать x.

Об эффективности этого метода:
Нам придётся решать уравнения только при $i \leq j$, следовательно, для решения этих двух систем потребуется, грубо говоря, $\frac{m^3}{6}$ умножений и делений (выигрыш по сравнению с методом Гаусса - в два раза!), а также m извлечений квадратного корня (отсюда, кстати, метод называется методом квадратного корня).

Таким образом, если матрица эрмитова, то один из эффективных прямых методов решения СЛАУ - метод квадратного корня, который требует меньше действий, чем метод Гаусса.

На практике, мы легко можем алгоритмически реализовать решение СЛАУ прямыми методами, например, методом Гаусса. Учитывая специальные виды матриц (диагональные, блочные и т.д.), которые часто встречаются на практике, мы сможем сократить количество действий.

Зачем же возникла необходимость применять итерационные методы решения? Об этом - в следущем параграфе.

§ 5. Примеры и канонический вид итерационных методов решения СЛАУ

Рассмотрим СЛАУ:

$$Ax = f, \tag{1.17}$$

где A - матрица размера $(m \times m)$, $|A| \neq 0$, $x = (x_1, \ldots, x_m)^T$, $f = (f_1, \ldots, f_m)^T$ (Матрица уже необязательно эрмитова).

Зачем нужны итерационные методы решения? Во-первых, на практике правые части системы обычно заданы с некоторой точностью. Прямой метод даёт точное решение этой системы. Нам же достаточно решения, верного с той же точностью, что и правая часть.

Во-вторых, количество действий прямых методов решения пропорционально m^3. Рассмотренные ниже методы позволяют найти решение всего за m итераций.

Рассмотрим итерационные методы Якоби (который также называют методом простой итерации) и Зейделя.

Из невырожденности матрицы следует, что решение СЛАУ существует и единственно. Перепишем систему 1.17 покоординатно:

$$\sum_{j=1}^{m} a_{ij} x_j = f_i, \ i = 1, \ldots, m \tag{1.18}$$

Начнем с метода Якоби (МЯ). Выделим из суммы i-ое слагаемое:

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

Пусть $a_{ii} \neq 0$. Тогда можно выразить x_i:

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

Обозначим через $x_i^n i$ n-ую итерацию i-ой координаты.

Запишем метод Якоби (МЯ). Чтобы его организовать, "навешиваем" $n+1$ в левой части:

$$x_i^{n+1} = \frac{f_i}{a_{ii}} - \sum_{j=1}^{i-1} \frac{a_{ij}}{a_{ii}} x_j^n - \sum_{j=i+1}^{m} \frac{a_{ij}}{a_{ii}} x_j^n, \ n = 0, 1, \ldots, \ i = 1, \ldots, m \tag{1.19}$$

Для того, чтобы начать вычисление, нужна некоторая нулевая итерация. Мы считаем, что она задана (т.е. x^0) - начальное приближение). Позже мы покажем, что начальное приближение может быть любым.

Итак, получаем итерационный метод, причем вычисление каждой итерации мы производим по *явным* формулам.

Метод, конечно, нужно оборвать на каком-то этапе, он должен быть конечным. Мы прекращаем вычисления, когда в некоторой норме достигнем нужной точности, т.е. когда $||x^n - x|| < eps$ при некотором n.

Запишем теперь метод Зейделя (МЗ):

$$x_i^{n+1} = \frac{f_i}{a_{ii}} - \sum_{j=1}^{i-1} \frac{a_{ij}}{a_{ii}} x_j^{n+1} - \sum_{j=i+1}^{m} \frac{a_{ij}}{a_{ii}} x_j^n, n = 0, 1, \ldots; i = 1, \ldots, m \tag{1.20}$$

Начальное приближение x^0 также считаем заданным изначально.

Метод Якоби является явным, так как вычисляется по явным формулам. Метод Зейделя - неявный. Но если разумно организовать процесс вычислений, а именно — начинать процесс вычислений с первой координаты, то каждую итерацию мы вычислим в явном виде:

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

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

и т.д.

Позже, когда мы докажем теорему о достаточных условиях сходимости итерационных методов, то увидим, что проверка для МЗ проще, чем для МЯ. В этом преимущество метода Зейделя. При этом станет видно, что скорость сходимости обоих методов медленная по сравнению с другими итерационными методами.

При рассмотрении итерационных методов решения СЛАУ обычно возникает 2 вопроса:

1. Вопрос сходимости: есть ли условия, при которых метод будет сходиться?

   При этом важно понимать, что как только мы говорим о сходимости некоторого метода, нужно указывать, в какой норме доказана сходимость. Из сходимости в одной норме, вообще говоря, не следует сходимость по другой норме.

   (Это иногда вытекает из некоторых теорем (из 4 курса) - из более сильной метрики можно получить сходимость в более слабой, но не всегда).

2. Скорость сходимости: а надо ли получить наиболее быструю скорость сходимости, и при каких условиях сходимость будет наиболее быстрая? ЗЫ: конечно надо!

Для исследования этих двух вопросов удобно системы рассматривать в матричном виде.

Представим матрицу A в виде

$$A = R_1 + D + R_2,$$

где

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

— нижнетреугольная матрица. Относительно первой побочной диагонали, с нулями на главной диагонали

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

— диагональная матрица,

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

—верхнетреугольная матрица с нулями на главной диагонали.

Можем переписать систему в матричном виде. (Позже мы увидим, что тогда все итерационные методы, которые мы изучаем, исследуются единым образом - в этом заслуга Самарского).

$$R_1x + Dx + R_2x = f$$
$$Dx = f - R_1x - R_2x$$

Предположим, что матрица D имеет обратную (т.е. a_{ii} отличны от нуля). Тогда:

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

Запишем методы Якоби и Зейделя (вектор начального приближения x^0 так же, как и раньше, считаем заданным):

$$Dx^{n+1} = f - R_1x^n - R_2x^n \tag{1.21}$$

- метод Якоби,

$$(D + R_1)x^{n+1} = f - R_2x^n \tag{1.22}$$

- метод Зейделя.

Можем их переписать и в более удобном виде:

$$D(x^{n+1} - x^n) + Ax^n = f \tag{1.23}$$

$$(D + R_1)(x^{n+1} - x^n) + Ax^n = f \tag{1.24}$$

Из последнего представления хорошо видно, что при наличии сходимости метода мы приходим (в пределе) к точному решению системы.

Так мы приходим к каноническому виду итерационных методов.

Определение:

Канонической формой записи двухслойного итерационного метода решения системы 1.17 называется его запись в виде:

$$B_{n+1}\frac{x^{n+1} - x^n}{\tau} + Ax^n = f,\ n = 0, 1, \ldots; \tag{1.25}$$

x^0 - задан, $\tau_{n+1} > 0$ - итерационный параметр, B_{n+1} - обратимая матрица.

Примечание:

Зачем нужен положительный итерационный параметр? Это позволяет ускорить скорость сходимости. Выбор итерационного параметра - это целая теория. Мы коснемся этого вопроса далее.

Если $B_{n+1} = E$, то метод 1.26 называется явным. Если $B_{n+1} = B,\ \tau_{n+1} = \tau$, то метод называют стационарным.

Примечание:

Здесь имеется некоторая терминологическая неточность: конечно, если даже матрица B_{n+1} не является единичной, мы можем найти явные формулы решения. Но мы всё-таки придерживаемся формального определения.

Рассмотрим и другие методы:

Метод простой итерации (ПИ) задается следующим образом:

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

Если заменить в последней формуле $\tau = \tau_n$ - переменный параметр, то приходим к методу Ричардсона (в котором для быстрой сходимости рекомендуется выбирать так называемый чебышевский набор параметров).

Рассмотрим попеременно-треугольный итерационный метод (его так же называют методом Самарского).

Пусть $A = R_1 + R_2$, где

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

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

Попеременно-треугольный итерационный метод (ПТИМ) задаётся формулой:

$$(E + \omega R_1)(E + \omega R_2)\frac{x^{n+1} - x^n}{\tau_{n+1}} + Ax^n = f,\ n = 0, 1, \dots; \tag{1.26}$$

x^0 - заданное начальное приближение.

Метод, конечно, не является явным. Но мы организуем процесс так, чтобы вычислять каждую итерацию по явным формулам.

Введем обозначения:

$$w^{n+1} = (E + \omega R_2)v^{n+1},$$

где $v^{n+1} = \frac{x^{n+1} - x^n}{\tau}$

$$r_n = f - Ax^n$$

— величина, называемая *невязкой*.

Тогда реализацию метода выполняем в 3 этапа:

- 1-ый этап:

  $(E + \omega R_1)w^{n+1} = r^n \Rightarrow$ обращая нижнюю треугольную матрицу, находим w^{n+1}.

- 2-ой этап:

  $(E + \omega R_2)v^{n+1} = w^{n+1} \Rightarrow$ обращая верхнюю треугольную матрицу, находим v^{n+1}.

- 3-ий этап:

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

§ 6. Теоремы о сходимости итерационных методов

Рассмотрим еще один метод решения СЛАУ:

$$Ax = f, \tag{1.27}$$

где A - невырожденная матрица размера $m \times m$

Рассмотрим метод:

$$B\frac{x^{n+1} - x^n}{\tau} + Ax^n = f \tag{1.28}$$

где $\tau > 0$, B - обратимая, $n = 0, 1, 2..$; x^0 - заданное начальное условие.

Говоря о сходимости, надо понимать о какой норме для сходимости идет речь. Введем нормы.

Пусть H - линейное пространство размерности m, т.е. $\forall x \in H \quad x = (x_1, ..x_m)$

Пока что мы не будем говорить о том, считаем ли мы H вещественными или комплексными числами.

Введем скалярное произведение: $(x, y) = \sum\limits_{i=1}^{m} x_i \bar{y}_i$

Тогда за норму можно взять $\|x\| = \sqrt{(x, x)}$ - среднеквадратичная норма.

Рассмотрим самосопряженный оператор $D = D^*$.
Примечание:
Понятия матрица и оператор для нас будут синонимами.

Самосопряженный оператор D в $\mathbb{C}$ обладает таким свойством, что

$$\forall x \in \mathbb{C} \quad (Dx, x) \in \mathbb{R}$$

Тогда можно связать норму с самосопряженным оператором. Также можно ввести энергетическую норму:

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

где D - самосопряженный оператор.

К примеру, если D - единичная матрица, то мы получим обычную среднеквадратичную норму.

Определение: Оператор D - положительно определен, если $(Dx, x) > 0 \quad \forall x \neq 0$

Определение:

Оператор D - неотрицательный, если $(Dx, x) \geq 0 \quad \forall x$. Таким образом может существовать $x_0 : (Dx_0, x_0) = 0$, т.е. неотрицательный оператор может быть не положительно определенным.

Свойства положительных самосопряженных операторов

1. $\exists \delta > 0 \quad : (Dx, x) \geq \delta \|x\|^2$ - здесь δ будет связано с min собственным значением оператора D.

2. Следующий факт предлагается в виде задачи:

   Задача 1:

   Пусть H - вещественное пространство, а оператор C - положительно определенный, но не самосопряженный. Доказать, что $(Cx, x) = (\frac{C+C^*}{2}x, x)$

   Решение:

   Поскольку H- вещественное пространство, справедливо

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

$$C = \frac{C + C^*}{2} + \frac{C - C^*}{2} \Rightarrow$$

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

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

3. Если $D = D^*$ и $D > 0$, тогда $\exists D^{-1}$ - обратный оператор, такой что $D^{-1} = (D^{-1})^*$ и $D^{-1} > 0$

   Более того $\exists D^{\frac{1}{2}} = (D^{\frac{1}{2}})^* \quad$ и $D^{\frac{1}{2}} > 0$
   Более того $\exists D^{-\frac{1}{2}} = (D^{-\frac{1}{2}})^* \quad$ и $D^{-\frac{1}{2}} > 0$

4. Введем вектор v^n:

$$v^n = x^n - x \tag{1.29}$$

   Этот вектор - погрешность на n-ой итерации. Для того, чтобы итерационный метод сходился, необходимо, чтобы $\|v^n\| \to 0$ при $n \to \infty$.

   Таким образом, мы получили задачу для v^n.

   Выразив x^n из 1.29 и подставив в 1.28, получаем:

$$B\frac{v^{n+1}-v^n}{\tau}+Av^n=0, \tag{1.30}$$

где $n=0,1,2,...,\quad v^0=x^0-x.$

Получили аналогичное, но однородное уравнение .

Пусть B - обратимый оператор, тогда, домножив на B^{-1}, получим:

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

Разрешим относительно первого слоя:

$$v^{n+1}=v^n-\tau B^{-1}Av^n=(E-\tau B^{-1}A)v^n$$

Обозначим через S :

$$S=E-\tau B^{-1}A \tag{1.31}$$

Матрица S называется матрицей перехода от n-ой итерации к $(n+1)$-ой итерации, т.е. $v^{n+1}=Sv^n$.

Все свойства процесса зависят от S, а именно от спектра матрицы S.

Теорема 1 (без доказательства):
Итерационный метод 1.28 решения задачи 1.27 сходится при любом начальном приближении x^0 тогда и только тогда, когда все собственные значения матрицы S по модулю меньше единицы: $\mid \lambda^s \mid < 1$.
Примечание: Эта теорема очень редко применима в чистом виде.

Опять рассматриваем H - как вещественное пространство. Очевидно, матрицы, самосопряженные в $\mathbb{R}$, будут симметрическими.

Теорема 2 (теорема Самарского о достаточных условиях сходимости двухслойных итерационных методов):
Пусть $A=A^*>0$, $\tau>0$ - положительный параметр, и выполнено:

$$B-0.5\tau A>0 \tag{1.32}$$

Тогда итерационный метод 1.28 решения системы 1.27 сходится по среднеквадратичной норме при любом начальном приближении x^0 :

$$\|x^n-x\|=(\sum_{i=1}^{m}(x_i^n-x_i)^2)^{\frac{1}{2}}\to 0,\quad n\to\infty\quad \forall x^0 \tag{1.33}$$

Условие 1.32 называется условием Самарского.

Доказательство:

Введем числовую последовательность $y_n : y_n = (Av^n, v^n)$; y_n - ограничена снизу: $y_n \geq 0$;

Рассмотрим y_{n+1}:

$$y_{n+1} = (Av^{n+1}, v^{n+1}) = (ASv^n, Sv^n) = (A(E - \tau B^{-1}A)v^n, (E - \tau B^{-1}A)v^n) =$$

$$= (Av^n, v^n) - \tau[(Av^n, B^{-1}Av^n) + (AB^{-1}Av^n, v^n) - \tau(AB^{-1}Av^n, B^{-1}Av^n)] =$$

$$= y_n - \tau[2(Av^n, B^{-1}Av^n) - \tau(AB^{-1}Av^n, B^{-1}Av^n)] =$$

$$= y_n - 2\tau((B - 0.5\tau A)B^{-1}Av^n, B^{-1}Av^n))$$

Полученное тождество

$$y_{n+1} = y_n - 2\tau((B - 0.5\tau A)B^{-1}Av^n, B^{-1}Av^n));$$

обычно переписывают в следующем виде:

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

Здесь $(B - 0.5\tau A)B^{-1}Av^n, B^{-1}Av^n \geq 0$;

Значит $y_{n+1} \leq y_n$, следовательно y_n - монотонно убывает и имеет место нижнее ограничение: $\exists \lim\limits_{n\to\infty} y_n = y$

Пусть $H \in \mathbb{R}; C > 0$, тогда $\exists\delta > 0 : (Cx, x) \geq \delta\|x\|^2$

Это следует из свойства положительного оператора C.

т.к. $B = 0.5\tau A > 0$, то:

$$\exists\delta > 0 : ((B - 0.5\tau A)B^{-1}Av^n, B{-}1Av^n) \geq \delta\|B^{-1}Av^n\|^2 \qquad (1.34)$$

Применяя данное неравенство, можно завершить доказательство теоремы:

$$\frac{y_{n+1} - y_n}{\tau} + 2\delta\|B^{-1}Av^n\|^2 \leq 0;$$

Для удобства введем $W^n = B^{-1}Av^n$.

Отсюда видно, что если $n \to \infty$, то $\|W^n\| \to 0$

Теперь можно выразить погрешность:

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

т.е. $\|v^n\| \le \|A^{-1}B\|\|W^n\| \quad \Rightarrow \ \|v^n\| \overset{n\to\infty}{\longrightarrow} 0$

чтд.

Следствие 1:
Пусть $A = A^* > 0$ (матрица A самосопряженная и положительная)
Тогда метод Якоби сходится в среднеквадратичной норме при любом начальном приближении, если выполнено $2D > A$ (тут D - диагональная матрица в методе Якоби, где $A = R_1 + D + R_2$)
Доказательство:
Метод Якоби: $D\frac{x^{n+1}-x^n}{\tau} + Ax^n = f$, где $B = D, \tau = 1$

$B - 0.5\tau A > 0 \quad \Rightarrow D - 0.5A > 0 \quad \Rightarrow 2D > A$
$\Rightarrow$ выполнены условия теоремы Самарского, тогда следствие выполняется.
чтд.

Следствие 2:
Пусть $A = A^* > 0$. Пусть также A - матрица со строгим диагональным преобладанием, т.е. элемент диагонали больше суммы остальных в строке:

$$a_{ii} > \sum_{\substack{j=1 \\ i\neq j}}^{m} |a_{ij}| \tag{1.35}$$

Тогда метод Якоби сходится при любом начальном приближении x_0 в среднеквадратичной норме.

Доказательство:
Рассмотрим квадратичную форму:

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

По свойству, аналогичному $2ab \le a^2 + b^2$, это равно

$$(\frac{1}{2}\sum_{i,j=1}^{m} |a_{ij}||x_i|^2) \cdot 2 = \sum_{i,j=1}^{m} |a_{ij}||x_i|^2$$

Тогда

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

чтд.

Задача 3:
Пусть $A = A^* > 0$. Доказать, что $a_{ii} > 0$;

Решение:
Из 1.35 следует, что:

$$a_{ii} + \sum_{i,j=1}^{m} |a_{ij}| < 2a_{ii}$$

Тогда

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

Применение Следствия 1 завершает доказательство.
чтд.

Следствие 3:
Пусть $A = A^* > 0$. Тогда метод Зейделя сходится в среднеквадратичной норме при любом начальном приближении x_0.

Доказательство:
$B - 0.5\tau A > 0$
$B = D + R_1 - \frac{1}{2}(R_1 + D + R_2) > 0$
$\Rightarrow D + R_1 - R_2 > 0;$
$((D + R_1 - R_2)x, x) > 0;$
$(Dx, x) + (R_1 x, x) - (R_1^* x, x) > 0;$
т.к. $(R_1 x, x) = (R_1^* x, x)$, то $(Dx, x) > 0$
чтд.

Следствие 4:
$A = A^* > 0$ (в вещественном пространстве).
Метод простой итерации - $\frac{x^{n+1} - x^n}{\tau} + Ax^n = f$.
Если выбрать τ: $0 < \tau < \frac{2}{\gamma_2}$, $\gamma_2 = \max\limits_k \lambda_k^A$ (модуль не нужен, матрица положительно определена), тогда метод простой итерации сходится $\forall x_0$ в среднеквадратичной норме.[метод релаксации] [очень похоже на производную по времени] [простейший с точки зрения реализации] [два параметра для ускорения

сходимости, но встает вопрос выбора]

Примечание:

На втором курсе рассматривалась обусловленность. Трудно строить сильно разбросанные спектры, $\min << \max$, граница для выбора τ очень узкая.

Например, если $\frac{\min}{\max} = 10^{-6}$, нельзя $\tau = 1$. Чем компактнее спектр, тем нам удобнее и быстрее. Эти вопросы мы рассмотрим в жёстких системах.

Доказательство:

Для сходимости необходимо $B - 0.5\tau A = 0$. Должно быть $E - 0.5\tau A > 0$.

$1 - 0.5\lambda_k^A > 0$.

Через собственные значения: $1 - 0.5\tau\gamma_2 > 0$ - нас устраивает.

Отсюда $\tau < \frac{2}{\gamma_2}$. $\tau < 0$ - потому что τ должно быть положительным. Как было для уравнений теплопроводности: от $\pi\frac{k}{2}$ до ∞. В разностной схеме мы это почувствуем, когда будем строить границы.

§ 7. Оценка скорости сходимости итерационных методов

В этом параграфе мы будем получать оценку скорости сходимости итерационных методов — оценивать количество итераций для достижения заданной точности и находить оценки — при каких соотношениях сходится или сходилось.

Сначала получим общие оценки для двухслойного итерационного метода, потом будем применять к конкретным.

По-прежнему нам интересна система

$$Ax = f, \tag{1.36}$$

где $|A| \neq 0, A \in R^{m\times m}$.

$$B\frac{x^{n+1} - x^n}{\tau} + Ax^n = f, \tag{1.37}$$

где $\tau > 0,\ \exists B^{-1},\ n = 0, 1, ...\ x^0$—задан.

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

$$B\frac{v^{n+1} - v^n}{\tau} + Av^n = f, \tag{1.38}$$

$n = 0, 1, ...\ \ v^0 = x^0 - x$.

Ставим задачу получить в нормах оценки вида

$$||v^{n+1}|| \leq \rho||v^n||, \tag{1.39}$$

$n = 0, 1, ...\ \ 0 < \rho < 1$.

Пока не фиксируем норму где удастся. Это пока не важно.

Откуда сходимость? Применим как рекуррентное. $||v^n|| \le \rho^n ||v^0||$, $n \to \infty \Rightarrow \rho^n \to 0 \Rightarrow ||v^n|| \to 0$ сходится.

Это уже известно из теоремы Самарского, теперь необходимо найти скорость сходимости.

$0 < \rho < 1$ — сходимость, это понятно. Когда приближаем ρ к 0, сходиться будет быстрее. ρ называют *затуханием погрешности*, ρ-оценка.

Если получим 1.39, то получим нужное количество итераций. Это главный критерий, по которому отбираем качество итерационного метода.

(Решить задачу с точностью до ε — значит получить $||x^n - x|| \le \varepsilon ||x^0 - x||$.)

Рекуррентная формула даёт соотношение $||x^n - x|| \le \rho^n ||x^0 - x||$. Добьёмся точности, когда $\rho^n \le \varepsilon$.

Перевернём: $\frac{1}{\varepsilon} \le \frac{1}{\rho^n}$.

Логарифмируем: $n \ln \frac{1}{p} \ge \ln \frac{1}{\varepsilon}$.

$n_0(\varepsilon) = \left[\frac{\ln \frac{1}{\varepsilon}}{\ln \frac{1}{p}}\right]$ — тогда $n \ge n_0(\varepsilon)$, оценка выполнена. $\ln \frac{1}{p}$ — скорость сходимости итерационного метода: чем она больше, тем меньше $n_0(\varepsilon)$, тем меньше число итераций.

В дальнейшем мы научимся находить число. Важнейшая задача! ρ получается через спектр и иными методами.

Опять вводим вещественное пространство H, $\dim H = m$.

Скалярное произведение: $\forall x, y \in H : (x, y) = \sum\limits_{i=1}^{m} x_i y_i$. $||x|| = (x, x)^{1/2} = \sqrt{\sum\limits_{i=1}^{m} x_i y_i}$.

$B = B^* > 0$ (пока не связан с 1.37, но потом будет). Энергетическая норма $||x||_B = (Bx, x)^{1/2}$.

Теорема 1:

Пусть $A = A^* > 0$, $B = B^* > 0$. (Ужесточили, у Самарского могло быть $B \ne B^*$. Но с точки зрения класса задач всё в порядке, B мы выбираем сами.)
$\exists \rho,\ 0 < \rho < 1 \Rightarrow$

$$\frac{1-\rho}{\tau} B \le A \le \frac{1+\rho}{\tau} B \tag{1.40}$$

Тогда итерационный метод 1.37 решения 1.36 сходится, и верна оценка:

$$||v^{n+1}||_B \leq \rho ||v^n||_B, \tag{1.41}$$

$n = 0, 1, \ldots$ (априорная оценка).
Отсюда мы можем оценить число итераций.

Доказательство:

$$\rho < 1, \quad A \leq \frac{1+\rho}{\tau} B$$

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

Значит сходимость будет. Но оценка получается в энергетической норме.

Замечание:

Из 1.41 можно получить и в норме $||\cdot||_A$. Удобнее: B — наш выбор, A — «заказчик».

Идея: доказательство из двух этапов. Строим матрицу перехода от (n)-й к $(n+1)$-й итерации, строим её спектр. У неё все $|\lambda_i| < \rho < 1$. А так как $A = A^* \Rightarrow$ есть ортонормированный базис собственных векторов.

Получим оценку 1.41:
$B = B^* > 0$, тогда существуют $B^{1/2} = (B^{1/2})^* > 0$, $B^{-1/2} = (B^{-1/2})^* > 0$.

Домножаем 1.38 на $B^{-1/2}$: $B^{1/2}\frac{v^{n+1}-v^n}{\tau} + B^{-1/2}Av^n = 0$.

Вводим $z^n = B^{1/2}v^n$. Достаточно для 1.41 получить оценку

$$||z^{n+1}|| \leq \rho ||z^n|| \tag{1.42}$$

В самом деле: $||z^n||^2 = (z_n, z_n) = (B^{1/2}v^n, B^{1/2}v^n) = (Bv^n, v^n) = ||v^n||_B$.

Кстати, отсюда $v^n = B^{-1/2}z^n$.

Уравнение примет вид:

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

Выражаем z^{n+1}:

$$z^{n+1} = z^n - \tau B^{-1/2}AB^{-1/2}z^n = Sz^n$$

$$S = E - \tau B^{-1/2}AB^{-1/2} \tag{1.43}$$

Пометим $Se_k = s_k e_k$, $k = 1...m$, $l_k \neq 0$, $s_k = eigenvalues$(задача на собственные значения).

Самосопряжённость S:

$$S^* = (E - \tau B^{-1/2} A B^{-1/2})^* = E^* - \tau (B^{-frac12})^* A^* (B^{-\frac{1}{2}})^* =$$

$$E - \tau (B^*)^{-\frac{1}{7}2} A (B^*)^{-\frac{1}{2}} = E - \tau B^{-frac12} A B^{-\frac{1}{2}} = S$$

Самосопряжённость B - это очень удобно.

Надо доказать, что $|s_k| \leq \rho$. Задача на собственные значения:

$$(E - \tau B^{-\frac{1}{2}} A B^{-\frac{1}{2}}) e_k = s_k e_k,$$

$e_k \neq 0,\ k = 1...m.$

Подействуем на всё $(B^{\frac{1}{2}} - \tau A B^{-\frac{1}{2}}) e_k = s_k B^{1/2} e_k$.
Обозначим $y = B^{-\frac{1}{2}} e_k$. $e_k = B^{\frac{1}{2}} y,\ y \neq 0$.
Переписываем задачу:

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

Отсюда

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

или

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

Подстроились под 1.40. Это означает — если скалярно умножим на $y \neq 0$:

$$(Ay, y) = \frac{1 - s_k}{\tau} (By, y)$$

Через 1.40:

$$\frac{1 - \rho}{\tau} (By, y) \leq \frac{1 - s_k}{\tau} (By, y) \leq \frac{1 + \rho}{\tau} (By, y),$$

$(By, y) > 0$ из $y \neq 0$! Так и сократим: $|s_k| \leq \rho,\ k = 1...m$.

$D = D^* \Rightarrow \exists$ ортонормированный базис собственных векторов. У самосопряжённых нет замкнутой теории. Пропадают замечательные качества, вместо равества Парсеваля — неравенство Риса.

Формально: $De_k = d_k l_k,\ k = 1...m,\ (l_k, l_e) = \delta_{kl} = (k == l)$. $\forall x \in H : x = c_1 e_1 + ... + c_k e_k$. Тогда $||x||^2 = \sum i = 1mc_i^2$. Это равенство Парсеваля. Оно ещё встретится в этом курсе.

Теперь докажем оценку для z: $z^{n+1} = S z^n$. Помним: $S e_k = s_k e_k,\ k = 1...m$.
Раскладываем по базису $\Rightarrow$

$$z^n = \sum k = 1mc_k^{(n)} l_k = \sum k = 1mc_k^{(n)} S e_k = \sum k = 1mc_k^{(n)} s_k e_k$$

Найдем коэффициенты Фурье. По Парсевалю (дважды): $||z^{n+1}||^2 = \sum k = 1m \left(c_k^{(n)}\right)^2 s_k^2$.

$$||z^{n+1}||^2 \le \rho^2 \sum k = 1m \left(c_k^{(n)}\right)^2 = \rho^2 ||z^n||^2$$

То есть $||z^{n+1}|| \le \rho ||z^n||$. Условий оказалось достаточно для 1.41.
чтд.

Следствие 1:
$A = A^* > 0$, $B = B^* > 0$, $\exists \gamma_2 > \gamma_1 > 0$: $\gamma_1 B \le A \le \gamma_2 B$.
Тогда, если $\tau = \tau_0 = \frac{2}{\gamma_1 + \gamma_2}$, то выполняется оценка 1.41:

$$||v^{n+1}||_B \le \rho ||v^n||_B,$$

$\rho = \frac{1-\xi}{1+\xi}$, $\xi = \frac{\gamma_1}{\gamma_2}$.

Гаммы — это верхняя/нижняя граница по модулю задачи на собственные значения.

Попадаем в условие теоремы: $\frac{1-\rho}{\tau} = \gamma_1$, $\frac{1+\rho}{\tau} = \gamma_2$.
Складываем: $\frac{2}{\tau} = \gamma_1 + \gamma_2$; $\tau = \frac{2}{\gamma_1 + \gamma_2}$.
Вычитаем: $\gamma_2 - \gamma_1 = \frac{2\rho}{\tau}$.
Тогда $\rho = \frac{\gamma_2 - \gamma_1}{2} \cdot \frac{2}{\gamma_1 + \gamma_2} = \frac{1 - \frac{\gamma_1}{\gamma_2}}{1 + \frac{\gamma_1}{\gamma_2}} = \frac{1-\xi}{1+\xi}$.
Математики хотят решить максимально широкий класс задач, но надо давать достаточные условия для необразованных инженеров. ;)

Следствие 2:
Метод простой итерации:

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

Пусть $A = A^* > 0$,
$\gamma_1 = \min_{1 \le k \le m} \lambda_k^A$,
$\gamma_2 = \max_{1 \le k \le m} \lambda_k^A$ ($\gamma_2 > \gamma_1 > 0$ автоматически), $\tau = \frac{2}{\gamma_1 + \gamma_2}$.

Тогда

$$||v^{n+1}|| \le \rho ||v^n||, \ \rho = \frac{1-\xi}{1+\xi}, \ \xi = \frac{\gamma_1}{\gamma_2},$$

где ξ - число обусловленности!

Доказательство:

Через Следствие 1: $B = E$, $\gamma_1 E \le A \le \gamma_2 E$, $||\cdot||_B = ||\cdot||_E = ||\cdot||$, дальше рассуждения аналогичные.

чтд.

§ 8. Исследование сходимости попеременного треугольного итерационного метода(ПТИМ)

Решаем СЛАУ:

$$Ax = f \tag{1.44}$$

где $detA \neq 0,\ A(m,m)$

Решим эту систему ПТИМ. Для этого представим матрицу $A = R_1 + R_2$, где

$$R_1 = \begin{bmatrix} 0.5a_{11} & 0 & 0 & \dots & 0 \\ a_{21} & 0.5a_{22} & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & a_{m3} & \dots & 0.5a_{mm} \end{bmatrix},\ R_2 = \begin{bmatrix} 0.5a_{11} & a_{12} & a_{13} & \dots & a_{1m} \\ 0 & 0.5a_{22} & a_{23} & \dots & a_{2m} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0.5a_{mm} \end{bmatrix}$$

Тогда оператор

$$B = (E + wR_1)(E + wR_2) \Rightarrow B\frac{x^{n+1} - x^n}{\tau} + Ax^n = f, \tag{1.45}$$

$n = 1, 2, \dots,\ \tau > 0,\ w > 0,\ x$—задано.

Метод сугубо неявный из-за оператора B. Реализация метода проблем не вызывает.

τ, w - итерационные параметры.

Теорема1: (о достаточном условии сходимости)

Пусть матрица A - самосопряженная и положительно определена ($A = A^* > 0$). Пусть $w > \frac{\tau}{4} \Rightarrow$ итерационный метод 1.45, реализующий задачу 1.44, сходится в средней квадратичной форме при любом начальном приближении.

Доказательство:

Опираемся на теорему Самарского $\Rightarrow B - 0.5\tau A > 0$.

Рассмотрим оператор B. Т.к. $A = A^* \Rightarrow R_1 = R_2^* \Rightarrow$

$$B = (E + wR_2^*)(E + wR_2) = E + w(R_2^* + R_2) + w^2 R_2^* R_2 = E + wA + w^2 R_2^* R_2$$

Запишем B по-другому: $B = (E - wR_2^*)(E - wR_2) + 2wA$.

Пусть $E - wR_2 = C \Rightarrow C^* = E - wR_2^* \Rightarrow (C^*Cx, x) = (Cx, Cx) \geq 0 \Rightarrow (E - wR_2^*)(E - wR_2) \geq 0$. Таким образом $B \geq 2wA$.

Также из теоремы Самарского $\Rightarrow B \geq 0.5\tau A$.

Тогда $B \geq 2wA > 0.5\tau A$. По свойству транзитивности $2w > 0.5\tau,\ w > \frac{\tau}{4}$. В этом случае работает достаточное условие т. Самарского.

чтд.

Теперь покажем, что скорость сходимости ПТИМ гораздо выше остальных методов.

Теорема 2: Об оценке скорости сходимости ПТИМ (Достаточное условие)

Пусть A самосопряженный положительный оператор, т.е. $A = A^* > 0$. Пусть существуют константы $\delta > 0,\ \Delta > 0$. Пусть для них выполнены условия:

$$A \geq \delta E_1,\ R_2^* R_2 \leq \frac{\Delta}{4} A \tag{1.46}$$

Положим $w = \frac{2}{\sqrt{\Delta\delta}},\ \tau = \frac{2}{\gamma_1 + \gamma_2}$, где

$$\gamma_1 = \frac{\delta}{2} \frac{\sqrt{\Delta\delta}}{\sqrt{\delta} + \sqrt{\Delta}},\ \gamma_2 = \frac{\sqrt{\delta\Delta}}{4} \tag{1.47}$$

γ_1, γ_2 - константы.

Тогда итерацонный метод 1.45 сходится, и для него справедлива оценка

$$\|v^{n+1}\|_B \leq \rho \|v^n\|_B, \tag{1.48}$$

$$\rho = \frac{1 - \sqrt{\eta}}{1 + 3\sqrt{\eta}},\ \eta = \frac{\delta}{\Delta} \tag{1.49}$$

Здесь $B = (E + wR_2^*)(E + wR_2)$. Для сходимости необходимо выполнение 2-х условий: $0 < \rho < 1,\ \delta < \Delta$.

Доказательство:

Убедимся, что $\delta < \Delta$, на основании Следствия 1.

Следствие 1:

Пусть $A = A^*,\ B = B^*$, если $\gamma_1 > \gamma_2 > 0 \Rightarrow$

$$\|x^{n+1} - x\|_B \leq\ \rho(w),\ \ \|x^n - x\|_B, \rho(w) = \frac{1 - \xi(w)}{1 + \xi(w)},\ \xi(w) = \frac{\gamma_1(w)}{\gamma_2(w)}$$

.

Поскольку ρ - переменная величина, то для неё необходимо найти min. В этой точке будет самая быстрая сходимость.

Из $A \geq \delta E \Rightarrow (Ax, x) \geq \delta(x, x) = \delta\|x\|^2$

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

Т.к. мы рассматриваем задачу в вещественном пространстве $\Rightarrow R_2 = R_2^* \Rightarrow (Ax, x) = (2R_2 x, x) = 2(R_2 x, x)$

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

$$\Rightarrow \delta \leq \Delta \Rightarrow \rho \leq 1$$

Чем ξ ближе к 1, тем меньше $\rho \Rightarrow w$ должно минимизировать ξ.

Рассмотрим фунцию $f(\xi) = \frac{\gamma_2(w)}{\gamma_1(w)}$.
Мы уже получали $B \geq 2wA,\ A \leq \frac{1}{2w}R \Rightarrow \gamma_2(w) = \frac{1}{2w}$

$$B = E + wA + w^2 R_2^* R_2 \leq \frac{1}{\delta}A + wA + w^2\frac{\Delta}{4}A =$$

$$= (\frac{1}{\delta} + w + \frac{w^2\Delta}{4}) \Rightarrow \gamma_1(w) = (\frac{1}{\delta} + w + \frac{w^2\Delta}{4})^{-1}$$

Для минимизаци ρ необходимо найти минимум $f(\xi) = f(w)$.

$$f(w) = \frac{\frac{1}{\delta} + w + \frac{\Delta}{4}w^2}{2w} = \frac{1}{2}(1 + \frac{1}{\delta w} + \frac{\Delta}{4}w)$$

$$f^{(1)} = \frac{1}{2}(\frac{\Delta}{4} - \frac{1}{\delta w^2}),\ \frac{\Delta}{4} = \frac{1}{\delta w^2},\ w^2 = \frac{4}{\delta\Delta} \Rightarrow w = w_0 = \frac{2}{\sqrt{\delta\Delta}}.$$

$$f^{(2)} = \frac{1\ 2}{2\delta w^3} > 0 \Rightarrow f(w_0) - min.$$

Теперь пересчитаем все константы.

$$\gamma_2(w) = \frac{1}{2w} = \frac{\sqrt{\Delta\delta}}{4}$$

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

$$\tau = \frac{2}{\gamma_1 + \gamma_2},$$

$$\rho = \frac{1-\xi(w)}{1+\xi(w)},\ \xi(w) = \frac{\gamma_1}{\gamma_2} = \frac{4\frac{\sqrt{\delta}\sqrt{\Delta\delta}}{2(\sqrt{\delta}+\sqrt{\delta})}}{\sqrt{\delta}\Delta} = \frac{2\sqrt{\delta}}{\sqrt{\delta}+\sqrt{\Delta}} \Rightarrow$$

$$\rho = \frac{1-\frac{2\sqrt{\delta}}{\sqrt{\delta}+\sqrt{\Delta}}}{1+\frac{2\sqrt{\delta}}{\sqrt{\Delta}+\sqrt{\delta}}} = \frac{\sqrt{\Delta}-\sqrt{\delta}}{3\sqrt{\delta}+\sqrt{\Delta}} = \frac{1-\sqrt{\frac{\delta}{\Delta}}}{1+3\sqrt{\frac{\delta}{\Delta}}}$$

При $\eta = \frac{\delta}{\Delta} \Rightarrow \rho = \frac{1-\sqrt{\eta}}{1+3\sqrt{\eta}}$ следует утверждение теоремы.

чтд.

Покажем качество ПИТМ. Сравниваем число операций $\eta_0(\xi) = \frac{ln\frac{1}{\xi}}{ln\frac{1}{\rho}}$, $\eta = O(m^{-2})$-маленькая величина в большинстве задач.

Оценим:

$$\frac{1}{\rho} = \frac{1+3\sqrt{\eta}}{1-\sqrt{\eta}} = \frac{(1+3\sqrt{\eta})(1+\sqrt{\eta})}{1-\eta} \approx 1+4\sqrt{\eta} \Rightarrow ln\frac{1}{\rho} \approx \sqrt{\eta} \Rightarrow n_0(\xi) \approx \frac{1}{\sqrt{\eta}} = O(m)$$

Посмотрим метод простой итерации и сравним количество действий:

$$\frac{x^{n+1}-x^n}{\tau} + Ax^n = f + (other) \Rightarrow \|x^{n+1}-x\| \le \rho\|x^n - x\|,\ \rho = \frac{1-\xi}{1+\xi},\ \xi = \frac{\gamma_1}{\gamma_2}$$

$$\gamma_1 = min\lambda_k^A$$
$$\gamma_2 = max\lambda_k^A,\ \xi = \eta$$

Полагая $\eta = O(m^{-2})$ найдём число итераций:

$$\Rightarrow \frac{1}{\rho} = \frac{1+\eta}{1-\eta} = \frac{(1+\eta)^2}{1-\eta^2} \approx 1+2\eta$$

$$ln\frac{1}{\rho} \approx \eta \Rightarrow n_0(\xi) \approx \frac{1}{\eta} \approx O(m^2)$$

В некоторых случаях $m = 10^6 \dots 10^7$

§ 9. Методы решения задач на собственные значения

Решение проблемы нахождения собственных значений сводится к решению характеристического уравнения.

Пусть дана произвольная матрица $A(m, m)$. Задача нахождения собственных значений

$$Ax = \lambda x,\ x \neq 0\ (!), \tag{1.50}$$

где λ - собственное значение, характеристический корень. x - собственный вектор.

Если задача решается над полем комплексных чисел, то собственные значения всегда существуют и их m штук. Нормальный оператор имеет базис из собственных векторов (самый широкий класс).

Необходимо узнать количество собственных значений. Все методы итерационные.

$$\|x\| = 1$$

- собственный вектор всегда нормированный.

Пусть у нас с учетом кратности m собственных значений (в том числе и комплексных). Занумируем их следующим образом:

$$\mid \lambda_1 \mid \leq \mid \lambda_2 \mid \leq \ldots \leq \mid \lambda_m \mid$$

Если матрица вещественная, а собственные значения комплексные, тогда собственные вектора - комплексные.

Проблемы:

1)Частичная проблема собственных значений

2)Полная проблема собственных значений (задача нахождения всего спектра)

§ 10. Степенной метод

Аналитически проблема собственных значений практически не разрешима. Один из самых простых методов для спектра - степенной метод.

Предварительные замечания:

В данном параграфе мы рассматриваем произвольную квадратную матрицу A. Никаких ограничений не накладываем, соответственно её спектр произвольный.

(Несмотря на то, что большинство задач работают с вещественной матрицей, а для итерационных методов требуется искать границы спектра (min и max собственных значений))

Рассмотрим степенной метод.

Пусть далее: x_n - n-ая итерация собственного вектора;

x_0 - начальное приближение

$$x_n + 1 = Ax_n \tag{1.51}$$

— рекурентная формула. : $n = 0, 1, ..., x_0$ — задано.

Примечание:

Здесь мы поймём, что для теоретического обоснования нам понадобятся некоторые ограничения. Мы будем использовать те подходы, которые можем применять без специальных (сложных) математических подходов. (Т.е. возможно в некоторых случаях условия можно и смягчить, но мы этим не занимаемся)

Можем выразить n-ую итерацию через x_0:

$$x_n = A^n x_0 \tag{1.52}$$

Собственные вектора упорядочим в порядке невозрастания модулей собственных значений: $|\lambda_1| \leq |\lambda_2| \leq \ldots \leq |\lambda_m|$.

Начинаем делать нужные допущения:

A) Пусть матрица обладает полным набором из собственных векторов (т.е. существует базис из собственных векторов; пока неважно - ортонормированный он или произвольный):

$$\{e_i\}_{i=1}^m : \ Ae_i = \lambda_i e_i, \ i = 1, \ldots, m.$$

$$x_n = c_1\lambda_1^n e_1 + c_2\lambda_2^n e_2 + \ldots + c_m\lambda_m^n e_m, \tag{1.53}$$

Примечание:

Самый широкий класс, для которого базис из собственных векторов существует, - это нормальный оператор.

Самосопряженные операторы входят в этот класс, а именно они часто встречаются на практике. Поэтому ограничение не такое уж и сильное.

B) Пусть $c_m \neq 0$, где m отвечает максимальному собственному значению (см. формулу 1.53).

C) $\left|\frac{\lambda_{m-1}}{\lambda_m}\right| < 1$.

Утверждение:

Пусть для матрицы выполнены условия A,B,C. Тогда степенной метод сходится по направлению к собственному вектору, отвечающему максимальному

по модулю собственному значению.

Примечание:

Ещё раз напомним: собственный вектор определен с точностью до константы. Поэтому всегда при счете проводится нормировка вектора. И мы будем вести речь не о длине собственного вектора, а о его направлении.

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

В соответствии с ограничением B:

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

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

Этот метод легко позволит найти максимальное значение λ_m.

Точнее, докажем что:

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

Пусть x_n^i - i-ая координата n-ой итерации.

Расписываем x_n^i и x_{n+1}^i:

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

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

Поделив $x_{n+1}^{(i)}$ на $x_n^{(i)}$, получаем:

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

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

Примечание:

Никто не гарантирует, что i-ая координата n-ой итерации не ноль. Этих нулей может быть $n-1$, но никак не n. Соответственно, при необходимости мы можем поменять координаты.

Таким образом, для произвольной матрицы, при выполнении условий A,B,C, степенной метод позволяет найти максимальное собственное значение и собственный вектор, ему отвечающий.

Если мы возьмем матрицу симметричную (у которой есть ОНБ из собственных векторов), то мы найдем собственный вектор быстрее, и сходимость будет не линейной, а в степени $2n$.

Покажем, что собственное значение также может быть вычислено по формуле:

$$\lambda_m^{(n)} = \frac{(x_{n+1}, x_n)}{(x_n, x_n)} = \frac{(Ax_n, x_n)}{(x_n, x_n)} \tag{1.54}$$

Рассмотрим 2 случая (и когда есть ОНБ из собственных векторов, и когда нет)

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

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

$$(e_l, e_j) = \delta_{lj}$$

- условие ортонормированности.

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

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

(первый знак равенства обеспечивает условие ортонормированности).

Таким образом, для симметричной матрицы мы можем найти собственный вектор быстрее, и сходимость будет не линейной, а степени $2n$.

Задача 1:
Доказать: $\lambda_1^{(n)} - \lambda_1 = o\left(\frac{\lambda_1}{\lambda_2}\right)^n$.
Задача 2:
Пусть $A = A^*$, $\exists A^{-1}$; $\lambda_1^{(n)} = \frac{(x_n, x_n)}{(x_{n+1}, x_n)}$. Доказать: $\lambda_1^{(n)} - \lambda_1 = O\left(\frac{\lambda_1}{\lambda_2}\right)^{2n}$
Решения будут позже

Таким образом, в предположениях A,B,C с помощью степенного метода мы можем найти собственный вектор, отвечающий максимальному собственному значению, и несколькими способами найти само это собственное значение.

Замечание о начальном приближении:
Матрица вещественная, но её собственные значения могут быть и комплексные (если матрица произвольная, а не самосопряженная). Тогда они сопряженные, и вектор, отвечающий этому собственному значению — комплексный, что в целом неправильно. Тогда мы не можем взять начальное приближение в виде вещественного вектора.

Метод обратных итераций.
Обобщаем метод для нахождения и других собственных значений, в частности - внутри спектра.

Сперва займемся минимальным по модулю собственным значением.
Нужно накладывать ограничения на матрицу A. Матрица вырожденная, если есть в спектре нулевое собственное значение. Поэтому рассматриваем только матрицы A, для которых существует обратная (т.е. невырожденные).

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

Метод неявный. Но умножив обе части слева на A^{-1}, получим:

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

Теперь метод превратился в степенной, но для обратной матрицы.
Таким образом, получится, что степенным методом мы найдём собственный вектор, соответствующий минимальному значению матрицы A.

Примечание:
Собственные значения обратной матрицы - обратные числа для :

$$\lambda_k^{A^{-1}} = \frac{1}{\lambda_k^A} \; \forall k, \; \lambda_k \neq 0$$

А собственные векторы матрицы A^{-1} — совпадают с собственными векторами A. Это важно, так как далее мы расписываем вектор через базис .
Вводим ограничения:

1. (A) Матрица A имеет базис из собственных векторов $\{e_i\}_{i=1}^{i=m}$
2. (B) $|\frac{\lambda_1}{\lambda_2}| < 1$
3. (C) $x_0 = c_1e_1 + c_2e_2 + \cdots + c_me_m$, где $c_1 \neq 0$

Тогда:

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

$$\lambda_1^n x_n = c_1e_1 + c_2\left(\frac{\lambda_1}{\lambda_2}\right)^n e_2 + \cdots + c_m\left(\frac{\lambda_1}{\lambda_m}\right)^n e_m$$

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

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

Метод обратных итераций со сдвигом.

Организуем метод следующим образом так:

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

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

При этом α - такое число, чтобы для матрицы $(A - \alpha E)$ существовала обратная: $(A - \alpha E)^{-1} = B$. Получим степенной метод для матрицы B:

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

Собственные значения матрицы B:

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

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

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

Если итерационный метод записать как

$$(A - \alpha E)x_{n+1} = x_n; \quad n = 0, 1, 2...$$

x_0 - задано, $\alpha \in \mathbb{R}$ такое, что $\exists (A - \alpha E)^{-1}$

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

Имеем $B = (A - \alpha E)^{-1}$ - для этой матрицы наш метод стал степенным.

Итерация будет сходится к собственному вектору $x_n \to \lambda_l$, для которого достигается max:

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

Здесь можно найти собственные значения:

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

§ 11. Приведение матрицы к верхней почти треугольной форме (ВПТФ)

A - произвольная квадратная матрица порядка m.

Удобно было бы привести матрицу A к C - диагональной или треугольной. Но нужно, чтобы процесс приведения матрицы сохранял её спектр: например через преобразования подобия:

$$C = Q^{-1}AQ \tag{1.55}$$

Под верхней почти треугольной формой матрицы подразумевают вид:

$$\begin{pmatrix} x & x & .... & x & x & x \\ x & x & .... & x & x & x \\ 0 & x & .... & x & x & x \\ 0 & 0 & .... & x & x & x \\ .... & .... & .... & .... & .... & .... \\ 0 & 0 & .... & 0 & x & x \end{pmatrix},$$

т.е. есть побочная диагональ в дополнение к треугольной матрице.

Мы рассмотрим метод элементарного отражения для приведения матрицы к ВПТФ.

Определение:

Матрица S - ортогональна, если $S^{-1} = S^T$ ($S^{-1} = S^*$ в унитарном пространстве).

Определение:

Пусть v - вектор-столбец

$$v = \begin{pmatrix} v_1 \\ v_2 \\ .... \\ v_m \end{pmatrix}$$

т.е. $v^T = (v_1, v_2, ..., v_m)$.

Элементарным отражением, соответствующим вектору-столбцу v называется преобразование, задаваемое матрицей:

$$H = E - \frac{2vv^T}{\|v\|^2} \tag{1.56}$$

У этого преобразования 3 важных свойства:

1. оно симметрично
2. оно ортогонально
3. действие на вектор - специфично

Вообще говоря: $\|v\|^2 = vv^T = v_1^2 + v_2^2 + ...v_m^2$

$vv^T = \begin{pmatrix} v_1^2 & v_1v_2 & .... & v_1v_m \\ v_2v_1 & v_2^2 & .... & v_2v_m \\ .... & .... & .... & .... \\ v_mv_1 & v_mv_2 & .... & v_m^2 \end{pmatrix}$ - симметричная матрица

Т.к. E - тоже симметричная $\Rightarrow H^T = H$, т.е. H - симметричная матрица.

Докажем, что H - ортогональна, т.е., что $H^T = H^{-1}$

$$H^TH = HH = H^2 =$$

$$= (E - \frac{2vv^T}{\|v\|^2})(E - \frac{2vv^T}{\|v\|^2}) = E - 4\frac{vv^T}{\|v\|^2} + 4\frac{vv^Tvv^T}{\|v\|^4} =$$

$$= E - 4\frac{vv^T}{\|v\|^2} + 4\frac{v\|v\|^2v^T}{\|v\|^4} = E - 4\frac{vv^T}{\|v\|^2} + 4\frac{vv^T}{\|v\|^2} = E$$

Следовательно H - ортогональна.

Поясним, что означает третье свойство.

Пусть задан произвольный вектор-столбец $x = (x_1, x_2, ..., x_m)^T$.

Тогда можно выбрать вектор v такой, что $Hx = \begin{pmatrix} -\sigma \\ 0 \\ .... \\ 0 \end{pmatrix}$, где $\sigma = \|x\|$.

Доказательство:

Положим $v = x + \sigma z$, где z - вектор специального вида $z = (1, 0, ..., 0)^T$ - длины m.

$$Hx = (E - \frac{2vv^T}{\|v\|^2})x = x - \frac{2(x + \sigma z)(x + \sigma z)^T}{(x + \sigma z)^T(x + \sigma z)}x =$$

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

числитель:

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

знаменатель:

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

$\sigma = \|x\|^2$;
Тогда числитель $= 2\sigma^2 + 2\sigma x_1$, знаменатель $= 2\sigma^2 + 2\sigma x_1$
Тогда

$$\frac{2(x + \sigma z)^T x}{(x + \sigma z)^T (x + \sigma z)} = 1$$

Тогда $Hx = x - x - \sigma z = \begin{pmatrix} -\sigma \\ 0 \\ .... \\ 0 \end{pmatrix}$

чтд.
Теперь докажем, что любую матрицу A можно свести к ВПТФ.

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

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

По третьему свойству специфичности для вектора x_{m-1} выберем вектор v:
$v = x_{m-1} - \sigma z_{m-1}$, где $\|x_{m-1}\| = \sigma, \quad z_{m-1} = (1, 0, ..., 0)$

Тогда $H_{m-1} x_{m-1} = -\sigma z_{m-1} = (-\sigma, 0, ..., 0)^T$
Введем U_1 :

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

Здесь $0_{12} = (0, ...0), \quad 0_{21} = (0, ..., 0)^T$ — вектора длины $m - 1$
Очевидно также, что $U_1 = U_1^T$
Проверим ортогональность матрицы U_1:

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

т.к. $H_{m-1} H_{m-1} = E$
Значит $U_1^{-1} = U_1^T = U_1$.
Тогда

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

$$= \begin{pmatrix} 1 & 0_{12} \\ 0_{21} & H_{m-1} \end{pmatrix} \begin{pmatrix} a_{11} & y_{m-1} \\ x_{m-1} & A_{m-1} \end{pmatrix} = \begin{pmatrix} a_{11} & y_{m-1} \\ H_{m-1}x_{m-1} & H_{m-1}A_{m-1} \end{pmatrix}$$

домножим справа на U_1, чтобы было преобразование подобия:

$$U_1^{-1}AU_1 = \begin{pmatrix} a_{11} & y_{m-1} \\ H_{m-1}x_{m-1} & H_{m-1}A_{m-1} \end{pmatrix} \begin{pmatrix} 1 & 0_{12} \\ 0_{21} & H_{m-1} \end{pmatrix}$$

$$= \begin{pmatrix} a_{11} & y_{m-1}H_{m-1} \\ -\sigma_1 z_{m-1} & H_{m-1}^{-1}A_{m-1}H_{m-1} \end{pmatrix},$$

здесь $\sigma_1 = \|x_{m-1}\|$.
Тогда матрица имеет вид:

$$U^{-1}AU = C_1 = \begin{pmatrix} x & x & & \\ x & x & & \\ 0 & x & & \\ 0 & x & & X \\ .... & .... & & \\ 0 & x & & \end{pmatrix}$$

Первый шаг завершен. На втором шаге берем $x_{m-2} = (c_{32}, c_{42}, ..., c_{m2})^T$, по нему опять выбираем вектор v и матрицу H_{m-2}: $\quad H_{m-2}x_{m-2} = -\sigma_2 z_{m-2}$

$$U_2 = \begin{pmatrix} \begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix} & 0_{12} \\ 0_{21} & H_{m-2} \end{pmatrix}$$

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

Т.е. через $(m-2)$ шага мы получим матрицу C:
$C = U_{m-2}^{-1}...U_2^{-1}U_1^{-1}AU_1U_2...U_{m-2}$ - матрицу ВПТФ.
Нужно убедиться, что преобразования сохраняли свойства подобия.
Обозначим $U = U_1U_2...U_{m-2}$
Найдем U^{-1}:

$$U^{-1} = (U_1U_2...U_{m-2})^{-1} = U_{m-2}^{-1}...U_2^{-1}U_1^{-1} = \text{ т.к. } U_i \text{ ортогональная } =$$

$$= U_{m-2}^T ... U_2^T U_1^T = U^T$$

Таким образом U - ортогональная.

Имеем $C = U^{-1}AU$, где C - имеет ВПТФ.

Мы показали, что любую A можно свести к ВПТФ преобразованиями подобия с помощью ортогональной матрицы.

Замечание 1:

Собственные значения матрицы C равны собственным значениями матрицы A: $\lambda_k^A = \lambda_k^C \quad k = 1, ..., m$

Доказательство:

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

Домножим слева на U^{-1}: $U^{-1}Ax = \lambda U^{-1}x;$

обозначим $y = U^{-1}x$ тогда $x = Uy$

тогда $U^{-1}AUy = \lambda y; \quad y \neq 0$

значит $Cy = \lambda y \quad \Rightarrow \lambda$ - собственное значение $C \quad (\forall \lambda_k)$ **чтд.**

Замечание 2:

Сохраняется симметрия, т.е. если $A = A^T$, то $C = C^T$

Доказательство:

$C = U^{-1}AU$

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

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

Имеется произвольная матрица - квадратная, порядка m.

$$A_{m \times m}$$

Поставим задачу - факторизовать матрицу A в вид: $A = QR$,
где Q - ортогональная, R - имеет верхнюю треугольную форму

Берем обозначения:

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

$v = x + \sigma z$, где $z = (1, 0, 0, ..., 0)$ — вектор размера m

строим $H_1^{m \times m}$:

$$H_1^{m \times m} = E - \frac{2vv^T}{\|v\|^2} \quad : H_1 = H_1^T = H_1^{-1}$$

подействуем матрицей H_1 на матрицу A:

$$H_1 A = \begin{pmatrix} x & x & \\ 0 & x & \\ 0 & x & X \\ 0 & x & \\ .... & .... & \\ 0 & x & \end{pmatrix}$$

Теперь по второму столбцу строим $H_2^{(m-1)\times(m-1)}$:

$$H_2 = \begin{pmatrix} 1 & 0_{12} \\ 0_{21} & H \end{pmatrix}$$

$H_2^{-1} = H_2^T = H_2$
Тогда

$$H_2^{-1} H_1^{-1} A = \begin{pmatrix} x & x & \\ 0 & x & \\ 0 & 0 & X \\ 0 & 0 & \\ .... & .... & \\ 0 & 0 & \end{pmatrix}$$

Через $m-1$ шаг получим верхнюю треугольную матрицу.
Получили матрицу $R = H_{m-1} H_{m-2} .... H_2 H_1 A$
Обозначим через $Q = H_1 H_2 ... H_{m-1}$
Найдем Q^{-1}:

$$Q^{-1} = (H_1 H_2 ... H_{m-1})^{-1} = H_{m-1}^{-1} .... H_1^{-1} = H_{m-1}^T .... H_2^T H_1^T = Q^T$$

Таким образом Q - ортогональная, значит:
$A = QR$. Следовательно любую матрицу можно QR разложить.

$A_{k+1} = Q_k^{-1} A_k Q_k$ — преобразование подобия с ортогональной матрицей
При $k \to \infty$ и вещественных λ_k:

$$A_k \to \begin{bmatrix} x & x & ... & x \\ 0 & x & ... & x \\ ... & ... & ... & ... \\ 0 & 0 & ... & x \end{bmatrix}$$

При комплексных λ_k:

$$A_k \to \begin{bmatrix} x & x & ... & x & x & x \\ 0 & x & ... & x & x & x \\ ... & ... & ... & ... & ... & ... \\ 0 & 0 & ... & \mathbf{x} & \mathbf{x} & x \\ 0 & 0 & ... & \mathbf{x} & \mathbf{x} & x \\ 0 & 0 & ... & 0 & 0 & x \end{bmatrix}$$

QR-алгоритм сходится к матрице (верхнетреугольной, если λ_k вещественные, и квазиверхнетреугольной, если λ_k комплексные). Таким образом, он применим к любой матрице.

§ 13. Предварительное преобразование матрицы к ВПТФ

$$A_k = Q_k R_k \tag{1.57}$$

$$A_{k+1} = R_k Q_k \; k = 0, 1, ... \tag{1.58}$$

$$A \to A_0$$

Лемма 1: пусть A, B — матрицы одного порядка, B имеет верхнюю треугольную форму, A — верхнюю почти треугольную форму, $C = BA$. Тогда C имеет верхнюю почти треугольную форму.

Доказательство: B в верхней треугольной — то есть $b_{il} = 0$ при $i > l$.

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

A в верхней почти треугольной — то есть $a_{lj} = 0$ при $l > j + 1$.

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

$c_{ij} = 0$ при $i > j + 1$ — то есть C в верхней почти треугольной форме.

Задача: A — верхняя почти треугольная, B — верхняя треугольная, $C = AB$. Доказать: C — верхняя почти треугольная.

Теперь мы знаем, что QR-алгоритм не портит ВПТФ. Перепишем (1.57) и (1.58):

$$Q_k = A_k R_k^{-1} = \{\text{ВПТФ}\} \times \{\text{ВТФ}\} = Q_k = \{\text{ВПТФ}\}$$

$$A_{k+1} = R_k Q_k - \text{ВПТФ}$$

Глава 2

Интерполирование и приближение функций

§ 1. Постановка задачи интерполирования

Интерполирование и приближение функций - это очень ёмкое понятие. Это можно решать бесчисленным количеством способов, в зависимости от поставленной цели.

Сейчас мы вспомним и полином Лагранжа, затем рассмотрим интерполяционный полином Ньютона. Далее начнём строить полином Эрмита и покажем, для чего нужны они. Ну и наилучшее квадратичное приближение функции.

Это наиболее древняя область, но она актуальна и поныне, поскольку без интерполирования математик обойтись не может. Это актуально для специальных функций, которые не выражаются через элементарные, но используются для решения дифуров. Обычно используются таблицы, но возникает вопрос, что делать, когда нужно получить значение для аргумента, которого нет в таблице. Первое решение — сгущать таблицы. Ещё одно — пользоваться интерполированием.

Где ещё требуется интерполирование — есть домна, за ней надо следить. Надо наставить датчиков, но много их поставить нельзя, иначе она разрушится, а информацию надо получать по всему профилю. Поэтому надо использовать интерполяцию.

Ешё одно применение — разностные схемы.

Ещё одно применение — экстраполирование.

Обычно будем использовать приближение полиномом.

$$f(x)\ x \in [a,b]\ a \le x_0 < x_1 < ... < x_n \le b$$

$\{x_i\}_0^n$ — узлы интерполяции.

$$f(x_i) = f_i, i = 0...n \tag{2.1}$$

$$P_n(x) = a_0 + a_1 x + ... + a_n x^n \tag{2.2}$$

$$P_n(x_i) = f_i, x = 0...n \tag{2.3}$$

$$\begin{cases} a_0 + a_1 x_0 + ... + a_n x_0^n = f_0 ... a_0 + a_1 x_n + ... + a_n x_n^n = f_n \end{cases} \tag{2.4}$$

Определитель системы — определитель Вандермонда:

$$\begin{vmatrix} 1 & x_0 & x_0^2 & ... & x_0^n \\ 1 & x_1 & x_1^2 & ... & x_1^n \\ ... & ... & ... & ... & ... \\ 1 & x_n & x_n^2 & ... & x_n^n \end{vmatrix} = \prod_{n \geq i > j \geq 0} (x_i - x_j) \neq 0$$

Так как он ненулевой, система имеет единственное решение.

§ 2. Интерполяционная формула Лагранжа

$$L_n(x_i) = f_i, i = 0...n \tag{2.5}$$

$c_k(x)$ — полином n-й степени.

$$L_n(x) = \sum_{k_0}^{n} c_k(x) f(x_k) \tag{2.6}$$

$$\omega(x) = \prod_{i=0}^{n} (x - x_i) \tag{2.7}$$

Продифференцируем $\omega(x)$:

$$\omega(x) = [...](x - x_k)$$

$$\omega'(x) = [...] + [...]'(x - x_k) = \prod_{k \neq j}^{n} (x_k - x_j)$$

Возьмём $c_k(x) = \frac{\omega(x)}{(x - x_k)\omega'(x_k)}$. Тогда

$$L_n(x) = \sum_{k=0}^{n} \frac{\omega(x)}{(x - x_k)\omega'(x_k)} f(x_k) \tag{2.8}$$

Погрешность для $L_n(x)$:

$$\Psi_{L_n(x)} = f(x) - L_n(x)$$

$$g(s) = f(s) - L_n(s) - Kw(s)$$

$g(x) = 0;\ K = \frac{f(x)-L_n(x)}{w(x)}$

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

$$|\Psi_{L_n(x)}| = \frac{M_{n+1}}{(n+1)!}|\omega(x)| \tag{2.9}$$

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

Замечание: если $f(x)$ — полином степени n или меньше, то погрешность будет равна нулю, и мы получим в точности его.

Ещё замечание: вообще говоря, $L_n(x) \not\to f(x)$.

§ 3. Разделенные разности

$$f(x)\ x \in [a,b]\ a \le x_0 < x_1 < ... < x_n \le b$$

Разделённая разность первого порядка определяется как

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

В частности,

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

Разделённая разность второго порядка:

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

Разделённая разность высших порядков определяется аналогично.

Выразим определяющую разность k-го порядка через кратность узла.

Утверждение:

Разделенная разность k-го порядка представим в виде

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

где

$$w(x) = (x - x_0)(x - x_1)\dots(x - x_n);$$
$$w_{\alpha,\beta}(x) = (x - x_\alpha)(x - x_{\alpha+1})\dots(x - x_\beta);$$
$$w^{'}(x_i) = w^{'}_{0,k}(x_i) = \Pi^{k}_{\substack{j=0 \\ j\neq i}}(x_i - x_j)$$

Доказательство:

По индукции:

1)По определению:

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

2)Пусть

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

3)

$$f(x_0, x_1, \dots, x_{l+1}) = \frac{f(x_1, \dots, x_{l+1}) - f(x_0, \dots, x_l)}{x_{l+1} - x_0}$$

В суммировании пределы разные. Выделим их:

$$\Rightarrow f(x_0, x_1, \dots, x_{l+1}) = \frac{f(x_0)}{(x_0 - x_{l+1})w^{'}_{0,l}(x_0)} + \frac{f(x_{l+1})}{(x_{l+1} - x_0)w^{'}_{1,l+1}(x_{l+1})} +$$

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

Также

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

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

Выразим значение функции в k-ой точке через значение в нулевой и разделенной разности k-го порядка.

Доказываем по по индукции:

$1)k = 1:$

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

$$(x_1 - x_0)f(x_0, x_1) = f(x_1) - f(x_0)f(x_1) = f(x_0) + (x - x_0)f(x_0, x_1)$$

$2)k = 2:$

$$f(x_0, x_1, x_2) = \frac{f(x_0)}{(x_0 - x_1)(x_0 - x_2)} + \frac{f(x_1)}{(x_1 - x_0)(x_1 - x_2)} + \frac{fx_2}{(x_2 - x_0)(x_2 - x_1)}$$

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

Заметим, что

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

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

$$f(x_2) = f(x_0) + (x_2 - x_0)f(x_0, x_1) + (x_2 - x_0)(x_2 - x_1)f(x_0, x_1, x_2) \Rightarrow$$

$$f(x_k) = f(x_0) + (x_k - x_0)f(x_0, x_1) + (x_k - x_1)(x_k - x_2)f(x_0, x_1, x_2) + \dots + (x_k - x_0)(x_k - x_1)\dots$$

§ 4. Интерполяционная формула Ньютона

Обозначим $n + 1$ узел как $\{x_i\}_0^n$, в этих узлах определена функция $f(x_i) = f_i$. Общий вид полинома Ньютона n-ой степени:

$$N_n(x) = f(x_0) + (x - x_0)f(x_0, x_1) + (x - x_0)(x - x_1)f(x_0, x_1, x_2) + \cdots$$

$$+(x - x_0)(x - x_1)\dots(x - x_{n-1})f(x_0, \dots, x_n)$$

Мы должны показать, что $N_n(x_i) = f(x_i), i = 0, \ldots, n$.

В этом случае полином будет интерполяционным.

$$N_n(x_i) = f(x_i) = f(x_0) + (x_i - x_0)f(x_0, x_1) + (x_i - x_0)(x_2 - x_1)f(x_0, x_1, x_2) + \cdots$$

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

Данное тождество совпадает с полиномом Лагранжа, только отличается внешним видом. В некоторых задачах вид полинома в форме Лагранжа не удобен.

Погрешность интерполяционного полинома Ньютона:

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

При $M_{n+1} = sup \mid f(x)^{n+1|} \Rightarrow \mid \Psi_{N_n(x)} \mid \leq \frac{M_{n+1}}{(n+1)!} \mid w(x) \mid$.

Там, где узлы увеличиваются (например при работе с таблицами Брадиса) обычно используют интерполяционный полином Ньютона.

§ 5. Интерполяцирование с кратными узлами. Полиномы Эрмита

Узлы обозначим как $\{x_i\}_0^m$. Пусть существуют

$$f(x_0), \ldots, f(x_m)$$

$$f^{'}(x_0), \ldots, f^{'}(x_m)\ldots f^{a_0-1}(x_0), \ldots, f^{a_m-1}(x_m),$$

где $a_i \in N$- кратность узла x_i:

$$H_n^{(i)}(x_k) = f^{(i)}(x_k) \tag{2.10}$$

Сумма кратностей $a_0 + \cdots + a_m = n + 1 \Rightarrow$

$$H_n(x) = \sum_{k=0}^{m}\sum_{i=0}^{a_k-1} c_{k,i}(x)f^{(i)}(x_k) \tag{2.11}$$

Построим полином Эрмита H_3 в заданных точках x_0, x_1, x_2. Точка x_1 является кратной. $\Rightarrow$

$$H_3(x_0) = f(x_0)$$

$$H_3(x_1) = f(x_1)$$

$$H_3(x_2) = f(x_2)$$

$$H_3^{'}(x_1) = f^{'}(x_1)$$

Общий вид полинома Эрмита третьей степени:

$$H_3(x) = c_0(x)f(x_0) + c_1(x)f(x_1) + c_2(x)f(x_2) + b_1(x)f^{'}(x_1) \tag{2.12}$$

При

$$c_0(x_0) = 1, c_1(x_0) = 0, c_2(x_0) = 0, b_1(x_0) = 0$$
$$c_0(x_1) = 0, c_1(x_1) = 1, c_2(x_1) = 0, b_1(x_1) = 0$$
$$c_0(x_2) = 0, c_1(x_2) = 0, c_2(x_2) = 1, b_1(x_2) = 0$$
$$c_0^{'}(x_1) = 0, c_1^{'}(x_1) = 1, c_2^{'}(1_2) = 0, b_1^{'}(x_1) = 1$$

Тогда H_3 удовлетворяет начальным условиям.
Пусть x_1—двукратный корень в c_0, c_2:

$$c_0 = k(x - x_2)(x - x_1)^2$$

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

Аналогично:

$$c_2(x) = \frac{(x - x_1)^2(x - x_0)}{(x_2 - x_1)^2(x_2 - x_0)}$$
$$(b_1(x) = k_1(x - x_0)(x - x_1)(x - x_2)$$
$$b_1(x) = (x - x_1)(k_1(x - x_0)(x - x_2))$$
$$b_1^{'}(x) = (k_1(x - x_0)(x - x_2)) + (x - x_1)(k_1(x - x_0)(x - x_2))^{'}$$
$$b_1^{'}(x_1) = 1 = k(x_1 - x_0)(x_1 - x_2)$$
$$b_1(x) = \frac{(x - x_0)(x - x_1)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)}$$
$$c_1(x) = (x - x_0)(x - x_2)(ax + b)$$
$$c_1(x_1) = 1 = (x_1 - x_0)(x_1 - x_2)(ax_1 + b) \Rightarrow$$
$$(ax_1 + b) = \frac{1}{(x_1 - x_0)(x_1 - x_2)}$$
$$c_1^{'}(x) = a(x - x_0)(x - x_2) + (ax + b)(2x - x_0 - x_2)$$
$$c_1^{'}(x) = 0 = a(x_1 - x_0)(x_1 - x_2) + (ax_1 + b)(ax_1 - x_0 - x_2)$$

Подставляя $(ax_1 + b)$:

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

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

Таким образом все коэффициенты однозначно находятся в явном виде.
Оценка погрешности $H_3(x)$, x_0, x_1, x_2 — так же, как делали для Лагранжа.

$$g(s) = f(s) - H_3(s) - K\omega(s)$$
$$\omega(s) = (x - x_0)(x - x_1)^2(x - x_2)$$

$$x_0 \le s \le x_2, x \in [x_0, x_2],$$

Константу K будем выбирать:

$$g(x) = 0 = f(x) - H_3(x) - K\omega(x)$$
$$K = \frac{f(x) - H_3(x)}{\omega(x)}$$

Дальнейший алгоритм полностью совпадает, но хитрость вокруг теоремы Ролля.

Рассмотрим $g(x)$. Она имеет по крайней мере 4 нуля на $[x_0, x_2]$: 3 во множителях и одно $g(x) = 0$.

Гладкости хватает. Применим теорему Ролля — у g' не менее 3 нулей, у g'' — не менее 2, у g''' — не менее 1, но этого мало, $\omega(s)$ — полином четвёртой степени.

Но из-за кратности узла x_1: есть ещё $g'(x_1) = 0$, так что у производной 4 нуля. У четвёртой ноль есть: $\exists \xi$, $g^{(4)}(\xi) = 0$.

Четырежды дифференцируем:

$$g^{(4)}(s) = f^{(4)}(s) - 4!K$$

Обозначим $M_4 = \sup\limits_{x_0 \le x \le x_2} |f^{(4)}(x)|$

Тогда

$$|f(x) - H_3(x)| = |\Psi_{H_3}(x)| \le \frac{M_4}{4!}|\omega(x)| = \frac{M_4}{4!}|(x - x_0)(x - x_1)^2(x - x_2)|$$

Если что-то обращается в ноль, приближение будет точным
Аналогично:

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

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

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

Придётся считать огромное количество этих интегралов, а аналитически это сложно! А численно лучше бы поточнее — здесь помогут формулы Гаусса.

Задача 1:

Необходимо из полинома Лагранжа предельным переходом получить полином Эрмита.

x_0, x_1, x_2 узлы, x_3 фиктивный, $x_3 \to x_1$. Доказать, что $\lim\limits_{x_3 \to x_1} L_3(x) = H_3(x)$.

Решение:

По заданным трем точкам можно однозначно построить полином $L_3(x)$:

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

Осуществим предельный переход от x_1 к x_3:

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

§ 6. Использование Полинома Эрмита для оценки погрешности квадратурной формулы Симпсона

Предварительные замечания:

Забегая вперёд: получится пятый порядок для частичного отрезка и четвёртый - для всего отрезка. А трапеция - только второй порядок.

Определенный интеграл $\int\limits_a^b f(x)\,dx$ вычисляем приближённо квадратурной формулой Симпсона. Для этого разбиваем отрезок на n частичных отрезков $a \le x_0 < x_1 < ... < x_N \le b$ так, что $x_i - x_{i-1} = h$.

$$\int\limits_{x_{i-1}}^{x_i} f(x)\,dx \ = \frac{h}{6}(f_{i-1} + 4f_{i-\frac{1}{2}} + f_i), \tag{2.13}$$

$$f_i = f(x_i),$$

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

— формула Симпсона на частичном отрезке.

Заменяем подынтегральное $f(x)$ полиномом, в нашем случае параболой. Погрешность — интеграл от погрешности! Из этих соображений выходит 4-й порядок, а мы получим 5-й.

Формула точна для полиномов до третьей степени:

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

— до x^2 формула Лагранжа точна, формула Симпсона точна по построению. Остаётся x^3.

Хотим показать:

$$\int\limits_{x_{i-1}}^{x_i} x^3\,dx = \frac{x_i^4 - x_{i-1}^4}{4} =$$

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

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

Теперь сводим формулу Симпсона к этому.

Записываем:

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

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

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

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

$$= \frac{h}{12}(x_i + x_{i-1})3(x_{i-1}^2 + x_i^2) =$$

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

Этот факт пригодится для получения точной оценки. Она будет точна для $H_3(x)$, так как это полином третьей степени.

Рассмотрим $H_3(x)$ в узлах $x_{i-1},\ x_{i-1/2},\ x_i$:

$$H_3(x_{i-1}) = f(x_{i-1});\ H_3(x_{i-1/2}) = f_{i-1/2};$$
$$H_3(x_i) = f_i;\ H_3'(x_{i-1/2}) = f_{i-1/2}'$$

Мы знаем о существовании и единственности, но в явном виде он нам не нужен. Нужна только погрешность и факт того, что он третьей степени.

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

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

Подставляем:

$$\int\limits_{x_{i-1}} x_i f(x)\,dx = \int\limits_{x_{i-1}} x_i H_3(x)\,dx + \int\limits_{x_{i-1}} x_i \Psi_{H_3}(x)\,dx =$$

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

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

Тогда погрешность:

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

$$M_4 = \sup_x |f^{(4)}(x)|$$

Перепишем функцию, чтобы была неотрицательная, убрав модуль:

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

Задача 1:

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

Решение:

Пусть $x = x_{i-1} + th,\ 0 \le x \le 1 \Rightarrow$

$$dx = hdt,\ x - x_{i-1} = th$$

$$x_i - x = h(1 - t),\ (x - x_{i-\frac{1}{2}})^2 = (t - \frac{1}{2})^2 \Rightarrow$$

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

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

Оценим погрешность на всем отрезке $[a, b]$. Надо оценивать полиномом Эрмита, а не Лагранжа. Второй дал бы грубую оценку. Но на всём отрезке — только 4:

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

$$|\Psi_h(f)| \le (\frac{h}{2})^4 \frac{b-a}{180},$$

$hN = b - a$.

§ 7. Наилучшее среднеквадратичное приближение функций

$H = L_2[a, b]$ (гильбертово пространство): $\forall f \in L_2[a, b] : \int\limits_a^b f^2(x)\,dx < \infty$

Введем норму:

$$\forall f, g \in L_2 : (f, g) = \int\limits_a^b f(x)g(x)\,dx; ||f||_{L_2} = \sqrt{\int\limits_a^b f^2(x)\,dx}$$

В этом пространстве и строится наилучшее среднеквадратичное приближение.
Теперь ставим задачу:

Пусть $\phi_0(x), \phi_1(x), ..., \phi_n(x) \in L_2$ - заданные $n+1$ функции, удовлетворяющие условию $\forall i : \int\limits_a^b \phi_i^2(x)\,dx < \infty$

Пусть также все функции линейно независимы.
Обобщённым многочленом назовем многочлен, имеющий вид:

$$\phi(x) = c_0\phi_0(x) + ... + c_n\phi_n(x) = \sum_{k=0}^{n} c_k\phi_k(x) \tag{2.14}$$

где $c_k, k = 0, 1, ..., n$ - вещественные числа.

Необходимо среди всех обобщённых многочленов вида 2.14 найти такой $\overline{\phi}(x) = \sum\limits_{k=0}^{n} \overline{c_k}\phi_k(x)$, который минимизирует норму

$$||f - \overline{\phi}||_{L_2} = \min_{\phi(x)} ||f - \phi||_{L_2}$$

$\overline{\phi}$ и есть наилучшее среднеквадратичное приближение f по системе $\{\phi_i(x)\}_0^n$.

Мы докажем, что оно всегда существует и единственно.

Чтобы понять, рассмотрим пример. Пусть $n = 0$, $\phi_0(x) \in L_2$. Тогда $\phi(x) = c_0\phi_0(x)$

$$F(c_0) = \int\limits_a^b (f(x) - c_0\phi_0(x))^2\,dx - \int\limits_a^b f^2(x)\,dx -$$

$$-2c_0 \int\limits_a^b f(x)\phi_0(x)\,dx + c_0^2 \int\limits_a^b \phi_0^2(x)\,dx$$

— квадратичная функция относительно x_0. Парабола вверх, поэтому ее минимум достигается в c_0, такой что: $F'(c_0) = 0$.

Следовательно:

$$2c_0 \int\limits_a^b f(x)\phi_0(x)\,dx = 2c_0 \int\limits_a^b \phi_0^2(x)\,dx \Rightarrow$$

$$\overline{c_0} = \frac{\int\limits_a^b f(x)\phi_0(x)\,dx}{\int\limits_a^b \phi_0^2(x)\,dx} = \frac{(f, \phi_0)}{(\phi_0, \phi_0)}$$

Тогда $\phi(x) = \bar{c}_0\phi_0(x)$. Она существует и единственна.

Если $\phi_0(x) = 1$:

$$\overline{c_0} = \frac{\int\limits_a^b f(x)\,dx}{b - a}$$

$$\overline{\phi}(x) = \overline{c_0} \times 1 = \text{среднее значение интеграла} = \frac{\int\limits_a^b f(x)\,dx}{b - a}$$

Просмотрели весь механизм. Дальше понимаем, что в случае большего количества переменных процесс будет аналогичен, породится система для нахождения

c_i:

$$\{\phi_i(x)\}_0^n \int\limits_a^b \phi_i^2(x)\,dx < \infty,$$

то есть $\phi_i(x) \in L_2[a,b]$.

$$F(c_0, c_1, ..., c_n) = \int\limits_a^b (f(x) - \phi(x))^2\,dx = \int\limits_a^b (f(x) - \sum_{k=0}^n c_k\phi_k(x))^2\,dx$$

— минимизируем этот интеграл.

Минимум достигается, где $\frac{\partial F}{\partial c_k} = 0,\ k = 0, ..., n$

$$F(c_0, c_1, ..., c_n) = \int\limits_a^b f^2(x)\,dx - 2\sum_{k=0}^n c_k \int\limits_a^b f(x)\phi_k(x)\,dx +$$

$$+ \sum_{k=0}^n c_k \sum_{l=0}^n c_l \int\limits_a^b \phi_k(x)\phi_l(x)\,dx =$$

$$= (f,f) - 2\sum_{k=0}^n c_k(f,\phi_k) + \sum_{k=0}^n c_k \sum_{l=0}^n c_l(\phi_k,\phi_l)$$

$$F(c_0, c_1, ..., c_n) =$$

$$= \int\limits_a^b f^2(x)\,dx - 2\sum_{k=0}^n c_k \int\limits_a^b f(x)\phi_k(x)\,dx + \sum_{k=0}^n c_k \sum_{l=0}^n c_l \int\limits_a^b \phi_k(x)\phi_l(x)\,dx =$$

$$= (f,f) - 2\sum_{k=0}^n c_k(f,\phi_k) + \sum_{k=0}^n c_k \sum_{l=0}^n c_l(\phi_k,\phi_l)$$

Рассмотрим теперь систему:

$$\frac{\delta F(c_0, ..., c_n)}{\delta c_k} = 0, \quad k = 0, 1, ..., n$$

$$\sum_{l=0}^n c_l(\phi_k,\phi_l) = (f,\phi_k), \quad k = 0, 1, ..., n$$

При $k = 0$:

$$\begin{cases} c_0(\phi_0,\phi_0)+c_1(\phi_0,\phi_1)+....+c_n(\phi_0,\phi_n)=(f,\phi_0) \\ c_0(\phi_1,\phi_0)+c_1(\phi_1,\phi_1)+....+c_n(\phi_1,\phi_n)=(f,\phi_1) \\ ..... \\ c_0(\phi_n,\phi_0)+c_1(\phi_n,\phi_1)+....+c_n(\phi_n,\phi_n)=(f,\phi_n) \end{cases} \tag{2.15}$$

$$\bar{c}_i=(f_i,\phi_i) \tag{2.16}$$

Выпишем матрицу системы:

$$\begin{pmatrix} (\phi_0,\phi_0) & (\phi_0,\phi_1) & .... & (\phi_0,\phi_n) \\ (\phi_1,\phi_0) & (\phi_1,\phi_1) & .... & (\phi_1,\phi_n) \\ .... & .... & .... & .... \\ (\phi_n,\phi_0) & (\phi_n,\phi_1) & .... & (\phi_n,\phi_n) \end{pmatrix} = G(\phi_0,....,\phi_n); \quad |G|>0$$

Получили матрицу Грама. Ее главное свойство - если система построена на матрице Грама, то система не вырождена. Таким образом, мы доказали, что применим критерий определенности системы (т.к. матрица вырожденная).

И мы находим $\bar{c}_0,\bar{c}_1,....,\bar{c}_n$.

И тогда $\bar{\phi}(x)=\sum\limits_{i=0}^{n}\bar{c}_i\phi_i(x)$ - будет наилучшим среднеквадратичным приближением.

Таким образом среднеквадратичное приближение существует и единственно.

чтд.

Если система $\{\phi_i(x)\}_0^n$ - ортонормированная, то матрица Грамма - единичная. И тогда вычисления будут осуществляться по формулам 2.16 и коэффициенты будут называться коэффициентами Фурье - они находятся гораздо проще.

По возможности для ЛНЗ систем проводят ортогонализацию и ортонормирование. Оказывается что при рассмотрении простого варианта системы:

$$1,x,x^2,....,x^n; \quad \phi_i(x)=x^i$$

Из этой системы скалярные произведения этих функций рассматриваются как интегралы:

$$\int\limits_{\alpha}^{\beta}\rho(x)\phi_k(x)\phi_l(x)\,dx=(\phi_k,\phi_l)$$

Здесь $\rho(x)>0$ - весовая функция.

Если брать различные веса, то можно построить систему ортогональных полиномов: Якоби, Лежандра, Чебышева и т.д.. По этим полиномам нам удобней

строить приближения. Если система ортонормированна, то тогда нетрудно вычислить отклонение наилучшего среднеквадратичного приближения:

$$F(c_0, c_1, ..., c_n) = \int_a^b [f(x) - \sum_{k=0}^{n} c_k \phi_k(x)]^2 \, dx = (f, f) - \sum_{k=0}^{n} c_k^2 \geq 0$$

Тогда получаем, что $(f, f) \geq \sum\limits_{k=0}^{n} c_k^2$.

Тогда выходит **неравенство Бесселя**: $\|f\|^2 \geq \sum\limits_{k=0}^{n} c_k^2$

В случае ОНБ - получаем **равенство Парсеваля**: $\|f^2\| = \sum\limits_{k=0}^{n} c_k^2$

Обобщение:
Пусть имеются базисные функции:
$\phi_0(x_i), \phi_1(x_i), ..., \phi_n(x_i)$ - заданные в узлах $\{x_i\}$

Тогда вводим пространство дискретных функций H. В нем вводим скалярное произведение для произвольных функций:

$$(f, g) = \sum_{i=0}^{n} f_i g_i$$

Тогда можно ввести норму, это будет аналог дискретной L2 нормы:

$$\|f\| = (f, f)^{\frac{1}{2}} = (\sum_{i=0}^{n} f_i^2)^{\frac{1}{2}})$$

Дальше строим $\phi(x_i) = \sum\limits_{k=0}^{n} c_k \phi_k(x_i)$, такой чтобы

$$\|f - \sum_{k=0}^{n} c_k \phi_k(x_i)\|^2 = F(c_0, c_1, ..., c_n) =$$

$$= (f, f) - 2 \sum_{k=0}^{n} c_k (f, \phi_k) + \sum_{k=0}^{n} c_k \sum_{l=0}^{n} c_l (\phi_k \phi_l)$$

Следовательно, в случае когда и базисные функции и приближаемые функции заданы дискретно, мы получим такую же систему уравнений для нахождения коэффициентов $\bar{c}_i$ обобщенного многочлена. Поэтому вопросы о существовании и единственности наилучшем среднеквадратичном приближении в дискретном случае решаются аналогично.

Глава 3

Численное решение нелинейных уравнений и систем нелинейных уравнений

§ 1. Введение

Изучая проблемы собственных значений, становится ясно, что сложно найти спектр операторов аналитически.

Проблема нахождения решения нелинейного уравнения $f(x) = 0$ - очень актуальна.

$f(x)$ - произвольная непрерывная функция.

В таких случаях начальное приближение нужно выбирать не случайным. Поэтому корень надо локализовать.

1 этап

Пусть x^* - корень, т.е. $f(x^*) = 0$;

Тогда надо указывать окрестность корня: $U_a(x^*) = \{x : |x^* - x| < a\}$

2 этап

Тут мы организоваваем итерационный сходящийся процесс $x_n \to x^*$

$$\begin{cases} f_1(x_1, x_2, ..., x_m) = 0 \\ f_2(x_1, x_2, ..., x_m) = 0 \\ ..... \\ f_m(x_1, x_2, ..., x_m) = 0 \end{cases} \tag{3.1}$$

$\bar{x} = (x_1, x_2, ..., x_m)$

$\bar{f} = (f_1, f_2, ..., f_m)$

$\bar{f}(\bar{x}) = 0$

$f : R_n \to R_n$

Стоит отметить, что нет канонического алгоритма для локализации корней. Указав окрестность - мы не всегда можем знать сколько корней в этой окрестности. Также могут быть кратные корни. В этой главе основное внимание будет уделено второму этапу.

1 способ локализации
Разобьем окрестность на узлы:

$$f(x): \quad \{x_i\}^n \quad f(x_i)$$

И если найдется пара узлов, такие, что $f(x_i)f(x_{i+1}) < 0 \Rightarrow$ между x_i и x_{i+1} есть хотя бы один корень.

Тогда уже этот отрезок можно размечать на узлы.

2 способ локализации (Бисекция)
Берется отрезок $[a;b]: \quad f(a) < 0, \quad f(b) > 0$
Рассматриваем середину $x_0 = \frac{a+b}{2}$, если $f(x_0) > 0 \Rightarrow x^* \in (a, x_0)$
Делим отрезок (a, x_0) пополам, и т.д.

Если выделить корень x^*, то для того, чтобы рассмотреть другие корни, можно разложить $f(x)$:

$$f(x) = (x - x^*)g(x)$$

и уже работать над $g(x)$.

§ 2. Метод простой итерации

$$f(x) = 0 \tag{3.2}$$

Все происходит в вещественном пространстве.

Теперь начинаем работает в локализованной окрестности - теперь считаем, что корень в окрестности есть и он один.

Записываем уравнение:

$$x = S(x) \tag{3.3}$$

$$x_{n+1} = S(x_n) \tag{3.4}$$

где $n = 0, 1, ... \quad x_0 \in U_a(x^*)$

Теперь надо выбрать $S(x)$ - именно выбором этой функции обуславливается метод:

$$S(x) = x + r(x)f(x) \tag{3.5}$$

Наложим ограничение:

$$sgn\, r(x) \neq 0 \quad x \in U_a(x^*) - \text{т.е. } r(x) \text{ не меняет знака в } U_a$$

Определение:
$S(x)$ удовлетворяет условию Липшица с константой Q на U_a, если $\forall x_1, x_2$ из U_a, выполнено:

$$|S(x_1) - S(x_2)| \leq Q|x_1 - x_2| \tag{3.6}$$

Утверждение:
Пусть $S(x)$ - непрерывна по Липшицу с константой $q < 1$. Пусть начальное приближение берется из окрестности U_a. Тогда метод простой итерации сходится.
Доказательство:
Покажем, что если $x_n \in U_a$, то и $x_{n+1} \in U_a$.
Оценим $|x_{n+1} - x^*|$:
Т.к. $x_{n+1} = S(x_n)$:

$$|x_{n+1} - x| = |S(x_n) - s(x^*)| = \{\text{по условию Липшица}\} \leq$$

$$\leq q|x_n - x^*| < |x_n - x^*| < a \quad \Rightarrow \quad x_{n+1} \in U_a$$

Значит из окрестности метод не выходит.
Применяя эту формулу рекуррентно, получим, что $|x_n - x^*| \leq q^n|x_0 - x^*|$,
а при $n \to \infty$: $\lim\limits_{n\to\infty} q^n = 0 \quad \Rightarrow \quad \lim\limits_{n\to\infty} x_n = x$.

Замечание 1:
Пусть $\max\limits_{x\in U_a(x^*)} |S'(x)| = q < 1$
Тогда метод простой итерации сходится.
Замечание 2:
Рассмотрим следующий итерационный процесс:

$$\frac{x_{n+1} - x_n}{\tau} + f(x_n) = 0; \quad n = 0, 1, ...; \quad x_0 \in U_a(x^*); \quad \tau > 0 \tag{3.7}$$

Пусть для определенности $f'(x) < 0$.

$S(x) = x - \tau f(x)$
Тогда в силу **замечания 1**, если $|S'(x)| < 1$, то сходимость обеспечена

$S'(x) = 1 - \tau f'(x)$, при предположении, что f - гладкая

Обозначим $M_1 = \max\limits_{x \in U_a(x^*)} |f'(x)|$
Тогда $|1 - \tau M_1| < 1 : \quad -1 < 1 - \tau M_1 < 1$

Значит, записанный в виде 3.7 итерационный метод с $\tau : \quad 0 < \tau < \frac{2}{M_1}$ - сходится

§ 3. Метод Эйткена ускорения сходимости итерационных методов

Пусть нам известно, что две соседние итерации отличаются:

$$x_n - x^* = Aq^n; \quad n = 0, 1, ...$$

$$\text{Тогда} \quad x_{n+1} - x^* \approx Aq^{n-1} \tag{3.8}$$

$$x_n - x^* \approx Aq^n \tag{3.9}$$

$$x_{n+1} - x^* \approx Aq^{n+1} \tag{3.10}$$

Из формулы 3.10:

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

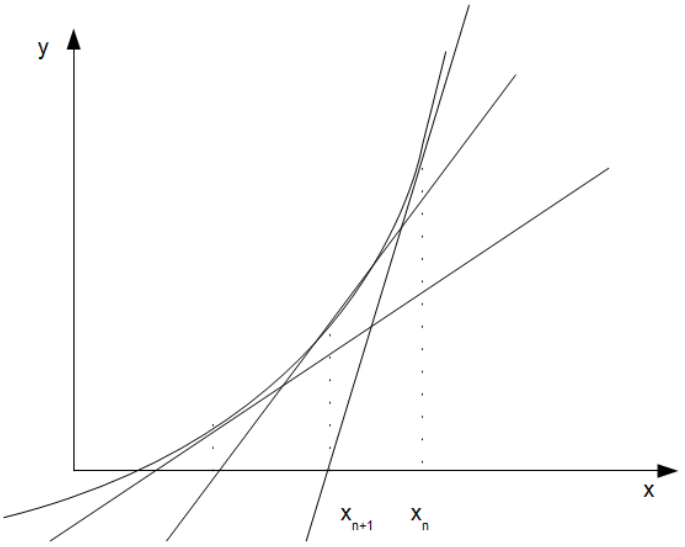

Рис. 3.1:

§ 4. Метод Ньютона и метод секущих

По-прежнему уравнение

$$f(x) = 0 \tag{3.11}$$

локализуем корень и окрестность $U_a(x_*)$. Обозначение: x^n — n-я итерация.

Метод Ньютона:

$$x^{n+1} = x^n - \frac{f(x^n)}{f'(x^n),} \tag{3.12}$$

где $n = 0, 1, 2...,\ \ x^0 \in U_a(x_*)$. В данном случае мы не заботимся о гладкости, чтобы было проще доказывать.

Также данный метод называют методом касательных.

Метод Ньютона сходится очень быстро, нужны обычно 3–4 итерации для хорошего результата.

Формула 3.12 получается разложением в окрестности корня по формуле Тейлора:

$$f(x_*) = 0 = f(x) + (x_* - x)f'(x) + \text{малое}$$

Если $x_* \to x^{n+1},\ x \to x^n \Rightarrow$

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

Производная не должна обращаться в ноль! Отсюда следует 3.12. Для системы будет всё аналогично.

Важно: применимость!

Неудачное применение метода Ньютона (зацикливание) изображено на рисунке 3.2:

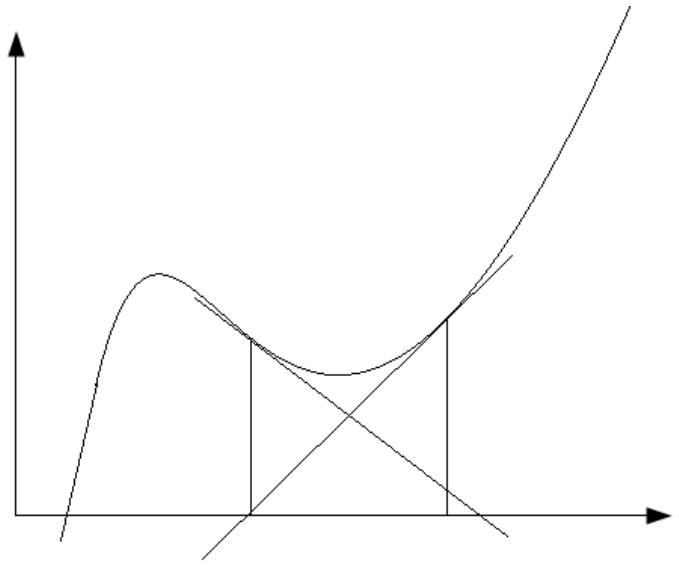

Рис. 3.2:

Модифицированный метод Ньютона:

$$x^{n+1} = x^n - \frac{f(x^n)}{f'(x^0)} \tag{3.13}$$

В этом случае мы всегда уверены в знаке, не надо лишний раз считать; серьёзная плата — очень плохая (медленная) сходимость.

Метод Ньютона очень быстр по сравнению с методом простой итерации.

Модифицированный метод применяется на практике: решаем серьёзную проблему, связанную с термоядерными реакциями. Сложные нелинейные дифференциальные уравнения. Решение аналитически затруднительно, возможно только итерационное решение. Будем применять метод Ньютона. Бывают очень быстротекущие процессы, бывают — не очень, но надо всё считать вместе, начальное приближение взять бывает тяжело.

Рассмотрим системы:

$$f_1(x_1, x_2) = 0; f_2(x_1, x_2) = 0$$

Решение — (x_1^*, x_2^*): $f_i(x_1^*, x_2^*) = 0;\ i = 1, 2.$

Гладкость по x_1, x_2 предполагаем какую угодно. (Непрерывности, очевидно, не хватит.)

Повторяем вывод:

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

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

Производим замену:

$$x_i \to x_i^n;\ x_i^* \to x_i^{n+1}$$

Отбрасывая малые, имеем систему:

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

$$f_2(x_1^n, x_2^n) + (x_1^{n+1} - x_1^n)\frac{\partial f_2(x_1^n, x_2^n)}{\partial x_1} + (x_2^{n+1} - x_2^n)\frac{\partial f_2(x_1^n, x_2^n)}{\partial x_2} = 0 \tag{3.14}$$

Чтобы удобно записывать, введём векторы $\bar{f} = (f_1, f_2)^T$, $\bar{x} = (x_1, x_2)^T$ и матрицу:

$$J(x) = \begin{bmatrix} \frac{\partial f_1(x)}{\partial x_1} & \frac{\partial f_1(x)}{\partial x_2} \\ \frac{\partial f_2(x)}{\partial x_1} & \frac{\partial f_2(x)}{\partial x_1 2} \end{bmatrix} \tag{3.15}$$

Замена, $x_i \to x_i^n;\ x_i^* \to x_i^{n+1}$ (опять такая же).

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

Пусть у нас есть $J^{-1}(x^n)$. Тогда:

$$x^{n+1} = x^n - J^{-1}(x^n)f(x^n) = 0 \tag{3.16}$$

.

Если обращать, то важно, какой определитель, близок ли он к нулю.
Вводят вектор $\bar{v}^{n+1} = \bar{x}^{n+1} - \bar{x}^n$ и получают:

$$f(x^n) + J(x^n)v^n = 0$$

;

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

Для произвольного количества переменных — совершенно аналогично.

$$f_1(x_1, x_2, ..., x_m) = 0$$

$$f_2(x_1, x_2, ..., x_m) = 0$$

$$.....$$

$$f_m(x_1, x_2, ..., x_m) = 0$$

Гладкость, опять-таки, вся есть.
Матрица $J = (f_{ij})$, $f_{ij} = \frac{\partial f_i}{\partial x_j}$. Вектор $\bar{f} = (f_1, ..., f_m)^T$, $x = (x_1, ..., x_m)^T$
Опять тем же образом:

$$\bar{x}^{n+1} = \bar{x}^n - J^{-1}(\bar{x}^n)f(\bar{x}^n) \tag{3.17}$$

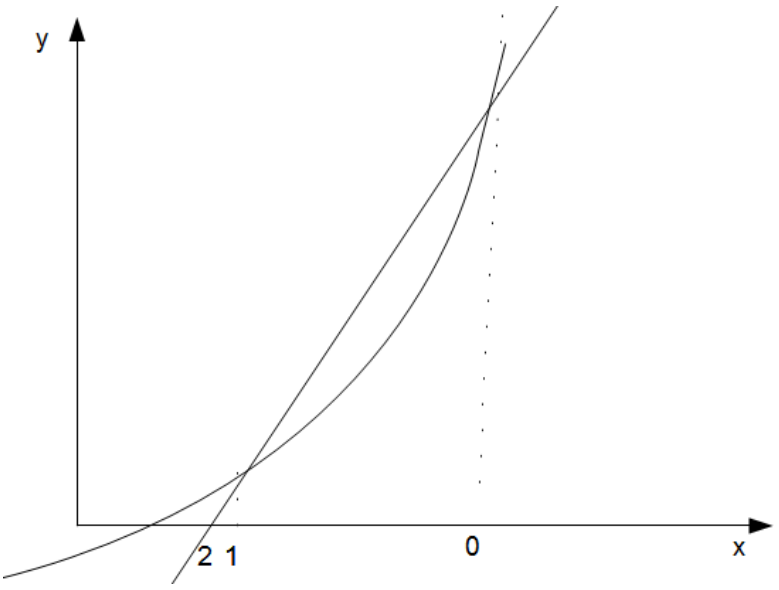

Рис. 3.3:

Замечание про $\bar{v}^{n+1}$ актуально.

Очевидно, случай одной переменной — частный случай этой формулы.

Несколько слов про метод секущих.

Производную считать очень тяжело.

Метод Ньютона:

$$f(x_*) = 0,$$

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

Многошаговые методы сходятся лучше, но им надо пространство для разгона.

$$f'(x^n) \approx \frac{f(x^n) - f(x^{n-1})}{x^n - x^{n-1}}$$

Заменяем:

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

$n = 1, 2, 3...$; x^0, x^1 где-то задаются.

Данный метод называется двушаговым. Иллюстрация данного метода изображена на рисунке 4.2.

Мы заменяем производную на полином первой степени, у которой всегда один ноль. Можно строить многошаговый метод, но там возникают сложности с нулями.

§ 5. Сходимость метода Ньютона. Оценка скорости сходимости

По прежнему рассматриваем:

$$f(x) = 0 \tag{3.18}$$

$$x^{n+1} = x^n - \frac{f(x^n)}{f'(x^n)}, \tag{3.19}$$

где $n = 0, 1, 2...$, $x^0 \in U_a(x_*)$.

Трактуем метод Ньютона как простую итерацию: $x^{n+1} = S(x^n)$. $S(x) = x - \frac{f(x)}{f'(x)}$. Гладкость: третья производная непрерывная— достаточное условие.
$|S'(x)| = q < 1,\ x \in U_a(x_*)$ — вот и сходимость.

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

Понятно, что $S'(x_*) = 0$. Это условие показывает, что погрешность на двух соседних итерациях связана квадратичным образом. В простой: связаны линейным соотношением.
$z_n = x^n - x_*$ — погрешность на итерации n.

Вновь Тейлор:

$$z_{n+1} = S(z^n + x_*) - S(x_*) = [S(x_*) + S'(x_*)]z_n + 0.5S''(\tilde{x})z_n^2 - S(x_*) =$$

$$= 0.5S''(\tilde{x})z_n^2,\ \tilde{x} = x_* + \theta x_n,\ |\theta| < 1$$

Придумаем константу M: $|0.5S''(\tilde{x})| \le M$. Тогда $|z_{n+1}| \le M|z_n^2|$.
Или $M|z_{n+1}| \le (M|z_n|)^2$.

$$v_n = M|z_n|,\ v_{n+1} \le v_n^2$$

Рекуррентно: $v_n \le v_0^{2^n}$; $M|z_n| \le (M|z_0|)^{2^n}$.

$$|z_n| \le \frac{1}{M}(M|z_0|)^2 \tag{3.20}$$

Обозначим $M|z_0| = q$. Если $q < 1$, то есть сходимость.
Как это связано с выбором начального приближения: $q < 1$, $|z_0| < \frac{1}{M}$

$$|x^0 - x_*| < \frac{1}{M}; \tag{3.21}$$

$|z_n| \to 0,\ n \to \infty$:

$$|x^n - x_*| \le \frac{1}{M}(M|x^0 - x_*|)^{2^n} \tag{3.22}$$

Пересчитываем :

$$0.5|(f(x)f''(x)(f'(x))^2)'| \le M \tag{3.23}$$

Утверждение:

Пусть $\exists M$ согласно 3.23, и пусть в окрестности $U_a(x_*),\ a = a(M)$ начальное приближение выбирается с условием 3.21. Тогда итерационный метод Ньютона сходится, и оценка 3.22 верна. Доказательство выше. Модифицированный метод Ньютона:

$$S'(x_*) = 1 - \frac{f'(x_*)}{f'(x^0)}$$

Возьмём x^0 возле x_*: если дробь близка к 1, S' — к 0, тогда повезло.

Нестационарное: будем решать задачу во времени по слоям; в качестве приближения — решение на предыдущем слое («не слишком сильно изменится»). В произвольной функции — метод пробуем угадать.

Глава 4

Разностные методы решения задач математической физики

§ 1. Введение

Задачи математической физики с помощью разностных схем.

Достаточно много времени потратим, но познакомимся всё равно поверхностно.

Важнейший раздел вычислительной математики. Будет на государственных экзаменахах. Теоретическая база построена Тихоновым и Самарским.

§ 2. Разностные схемы для первой краевой задачи уравнения теплопроводности

Есть куча других методов, вероятностные и всякие прочие, но разностные самые известные и наиболее подходящие.

Постановка:

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + f(x,t), \tag{4.1}$$

где $0 < x < 1$, $0 < t \leq T$.

Краевые:

$$u(0,t) = \mu_1(t); u(1,t) = \mu_2(t);\ 0 \leq t \leq T \tag{4.2}$$

— первого рода.

Начальное условие:

$$u(x,0) = u_0(x);\ 0 \leq x \leq 1 \tag{4.3}$$

Ищем уравнение, внутри удовлетворяющее 3.18, на соответствующих границах 4.2 и 4.3.

Знаем существование, единственность, устойчивость по правой части и по начальному условию — корректно поставленная задача.

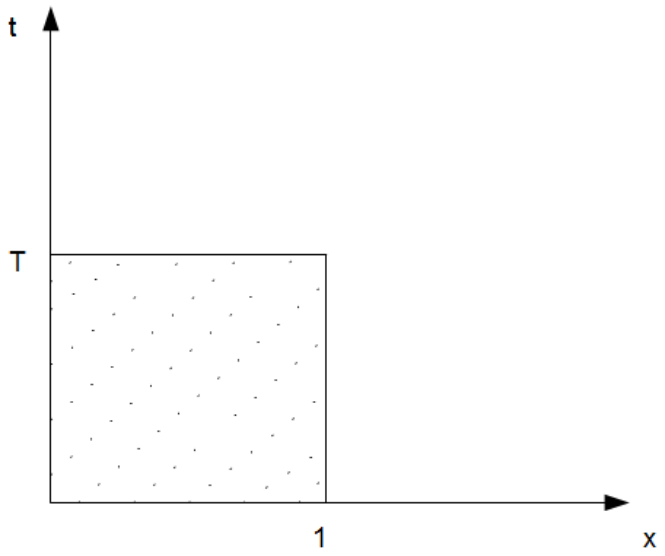

Рис. 4.1:

Ищем классические решения, обобщённые будут в курсе функционального анализа.

Что значит «решить разностным методом»?

Непрерывную область будем менять на дискретную (то есть сетку). Самостоятельная часть вычислительной математики.

Теория будет на сетках с постоянным шагом, там проще. Но не всегда это годится:

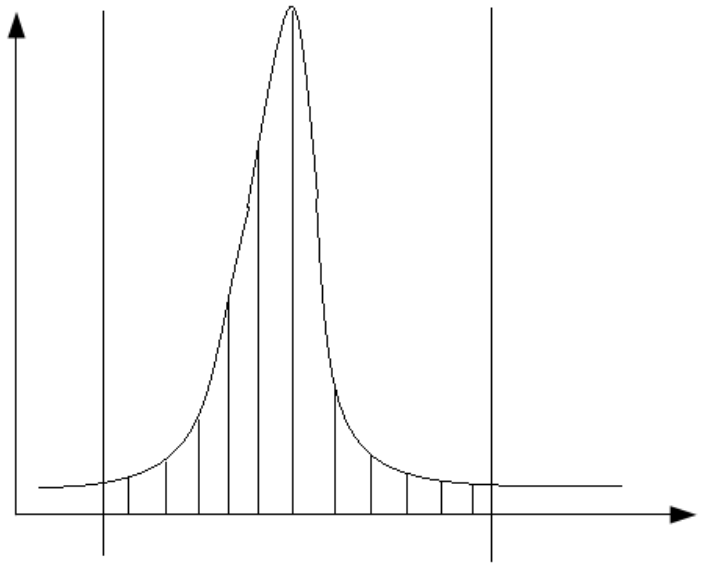

Рис. 4.2:

$$\omega_h = \{x_i = ih;\ i = 1, 2, ..., N-1;\ h = 1/N\},$$

где $h > 0$ — шаг по переменной x — внутренние узлы сетки

$\bar{\omega}_h = \{x_i = ih;\ i = 0, 1, ..., N;\ hN = 1\}$ - все узлы сетки (как бы замыкание)

$\omega_\tau = \{t_j = j\tau;\ j = 1, 2, ..., j_0;\ j_0\tau = T\}$
$\bar{\omega}_\tau = \{t_j = j\tau;\ j = 0, 1, ..., j_0;\ j_0\tau = T\}$ - шаг по времени $t - \tau > 0$.

Все внутренные узлы $\omega_{\tau h} = \omega_\tau \times \omega_h$. Также $\bar{\omega}_{\tau h} = \bar{\omega}_\tau \times \bar{\omega}_h$.

С замены области на сетку начинается любое решение разностным методом.

Рассмотрим область $D = \{(x, y) \in R^2 :\ 0 < x < 1,\ 0 < t \leq T\}$ (T – заданное положительное число).

Запишем первую краевую задачу для уравнения теплопроводности в этой области:

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + f(x, t), \quad (x, t) \in D, \tag{4.4}$$

краевые условия:

$$\begin{cases} u(0, t) = \mu_1(t), \\ u(1, t) = \mu_2(t), \end{cases} \tag{4.5}$$

начальное условие (начальная температура):

$$u(x, 0) = u_0(x). \tag{4.6}$$

Вводим следующие обозначения:

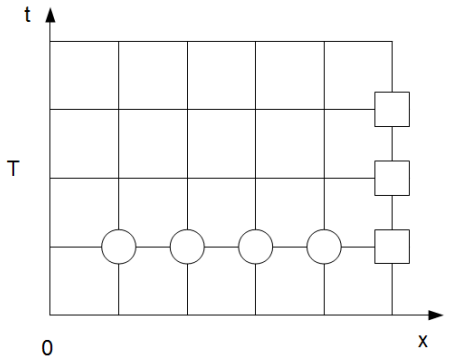

Рис. 4.3: Вводим сетку, по x - с шагом a, по t - с шагом τ

Возникают соответствующие узлы сетки. В реальности - внутренних узлов много больше, чем граничных. Совокупность всех узлов, соответсвующих моменту времени t_k (включая граничные) будем называть слоем.

Примечание.

T - искусственная граница, можно просто писать, что $t > 0$).

Пункт 1. Явная разностная схема

Решая любую дифференциальную задачу, мы движемся по следующей схеме: сначала вводим сетку, а затем рассматриваем сеточную функцию.

Будем обозначать численное решение поставленной задачи через $y(x, t)$. Пусть

$$y_i^n = y(x_i, t_n).$$

- ввели сеточную функцию (от двух переменных). f_i^n - значения в узлах. Покажем, что при стремлении шагов к нулю y будет сходиться к решению (с любой заданной точностью)

Запишем разностное уравнение:

$$\frac{y_i^{n+1} - y_i^n}{\tau} = \frac{y_{i-1}^n - 2y_i^n + y_{i+1}^n}{h^2} + f(x_i, t_n), \quad (x_i, t_n) \in \omega_{\tau h}, \tag{4.7}$$

т.е. точки (x_i, t_n) принадлежит внутренней части области.

Таким образом, поставили исходному уравнению 4.1 его дискретный аналог - уравнение 4.7.

Почему выбрали именно первое краевое? Просто проще делать нужной погрешность (в отличие от условия с производной). Задаём граничные и начальное условия:

$$\begin{cases} y_0^{n+1} = \mu_1(t_{n+1}), & t_{n+1} \in \overline{\omega}_\tau, \\ y_N^{n+1} = \mu_2(t_{n+1}), & t_{n+1} \in \overline{\omega}_\tau, \end{cases} \tag{4.8}$$

$$y_i^0 = u_0(x_i), \quad x_i \in \overline{\omega}_h. \tag{4.9}$$

Т.о., исходной задаче поставили дискретный аналог 4.7—4.9. Т.е. получили СЛАУ - она и называется разностной схемой.

Посмотрим, какие узлы использует эта разностная схема (это множество называется шаблоном разностной схемы)

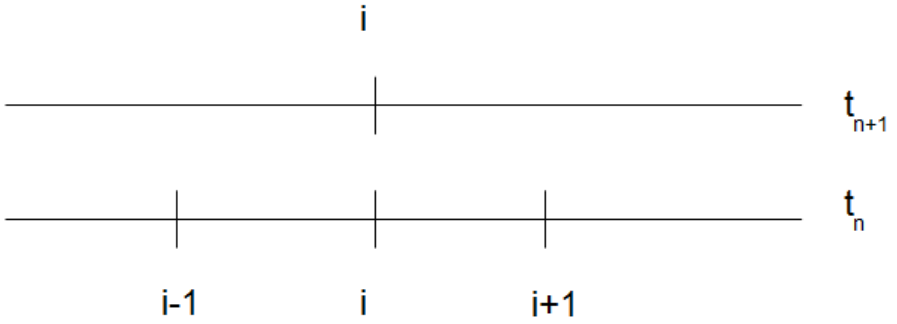

Рис. 4.4:

Используется 2 слоя: $t_n + 1$ и t_n Получается четырехточечный шаблон.

Примечание 1:

Мы стремимся к тому, чтобы схема апроксимировала более точно. Для этого можно задействовать больше узлов шаблона. Но если будет погрешность высокая, то из-за большого числа узлов мы можем потерять устойчивость.

Примечание 2:
90% студентов во время экзамена не понимают, что такое устойчивость разностной схемы. Нам вбили в головы, что это непрерывная зависимость от входных данных. Здесь это не так, у нас более жесткая зависимость!

Задача 4.7—4.9 решается на слоях. Нулевой слой задан. Проводим вычисления на первом слое; от него - переходим на второй, и т.д. Решение по явной формуле, на каждом слое не требуется никаких обращений матрицы и т.д.

$$y_i^{n+1} = y_i^n + \frac{\tau}{h^2}(y_{i-1}^n - 2y_i^n + y_{i+1}^n) + \tau f_i^n, \quad i = 1, \ldots, N-1. \tag{4.10}$$

Это записано по слоям, но внутри сетки; а снаружи — y_0^{n+1}, y_N^{n+1} — значения заданы по условиями.

Таким образом, преимущество данного метода в простоте реализации, все формулы явные.

Сформулируем вопросы, которые обычно решаются при изучении разностных схем.

1. Погрешность аппроксимации. (Очевидно, что если схема не приближает исходную задачу - нечего её исследовать.

   Далее: Про задачу 4.4—4.6 - задача корректно поставленная, но если неправильно выберем разностную задачу для 4.4—4.6 - может быть некорректной)

2. Существование решения и единственности решения разностной схемы (т.е. СЛАУ). (Мы должны доказать ещё, есть ли решение и единственное ли оно)

3. Алгоритм нахождения численного решения

   И ключевые вопросы:

4. Сходимость разностной схемы к решению исходной задачи (можно коротко: сходимость разностной схемы)

5. Исследование устойчивости решения разностной схемы по начальному условию и правой части (коротко говорят: по входным данным)

Примечание:
Каждый из этих пунктов для различных разностных схем решается с той или иной степенью точности. Дальше мы докажем теорему Филиппова, которая свяжет между собой вопросы 4—5(ключевые) и первый.

Для нашей схемы: существование, единственность, алгоритм - доказано наличием явных формул.

Займемся ключевыми вопросами.

Примечание:

Философский вопрос: будем сравнивать решение дискретное (функцию в узлах) и непрерывное решение. Как их сравнивать, они из разных пространств?

Во-первых, если получили сеточную функцию в узлах, можем восстановить её во всей области (сплайнами, квадратично, как угодно). Тогда будем сравнивать в метрике функционального пространства непрерывных функций (или более точно: в том пространстве, где была поставлена исходная дифференциальная задача). Или наоборот: будем проецировать функцию U внутри сетки. Мы будем идти по второму пути, считаем этот путь лучшим.

Наступил принципиальный момент: как раз здесь вопрос нормы становится принципиальным: мы не можем сказать, что если в одной норме схема сходится, то и в другой норме будет сходиться.

Схема у нас - условно сходящаяся и условно устойчивая: далее мы покажем, что если $\frac{\tau}{h^2} < 0.5$, то разностная схема сходится. Иначе - не будет сходиться.

Если схема сходится при любых шагах, то её называют абсолютно сходящейся и устойчивой.

Определим погрешность разностной схемы z_i^n так:

$$z_i^n = y_i^n - u_i^n. \tag{4.11}$$

при этом

$$z_0^{n+1} = 0,\ z_N^{n+1} = 0,\ z_i^0 = 0 \tag{4.12}$$

Примечание:

Погрешность разностной схемы 4.14 - $z_i^n = y_i^n - u$ - это не погрешность задачи.

Тогда (4.7) можно переписать следующим образом (в силу линейности задачи):

$$\frac{z_i^{n+1} - z_i^n}{\tau} = \frac{z_{i-1}^n - 2z_i^n + z_{i+1}^n}{h^2} + \psi_i^n, \quad (x_i, t_n) \in \overline{\omega}_{\tau h}. \tag{4.13}$$

Определим ещё функцию ψ_i^n:

$$\psi_i^n = \frac{u_{i-1}^n - 2u_i^n + u_{i+1}^n}{h^2} - \frac{u_i^{n+1} - u_i^n}{\tau} + f_i^n. \tag{4.14}$$

Определение. Функция ψ_i^n, определяемая равенством (4.14), называется погрешностью аппроксимации разностной схемы (4.7) — (4.9) на решение задачи исходной задачи 4.4 —4.6.

Примечание:

Надо в разностную схему вместо y подставить u, и разность левых и правой частей — и есть погрешность аппроксимации.

Задача: Доказать, что $\psi_i^n = O(\tau + h^2)$.

(Это будет рассказано в общем виде, но и сейчас всё видно; первая производная и вторая заменяются на какие-то О-большие)

Поэтому говорят, что для данной схемы погрешность аппроксимации первого порядка по τ и второго по h.

Решение:

Пусть $U(x,t)$ - достаточно гладкая функция.

Разложим $u(x_i, t_{n+1})$ в узле (x_i, t_n) по формуле Тейлора:

$$u(x_i, t_{n+1}) = u_i^{n+1} = u(x_i, t_n) + u_t(x_i, t_n)\tau + O(\tau^2).$$

Разложим $u(x_{i+1}, t_n)$ в узле (x_i, t_n) по формуле Тейлора:

$$u(x_{i+1}, t_n) = u_{i+1}^n = u(x_i, t_n) + u_x(x_i, t_n)h + \frac{1}{2}u_{xx}(x_i, t_n)h^2 + \frac{1}{6}u_{xxx}(x_i, t_n)h^3 + O(h^4).$$

Разложим $u(x_{i-1}, t_n)$ в узле (x_i, t_n) по формуле Тейлора:

$$u(x_{i-1}, t_n) = u_{i+1}^n = u(x_i, t_n) - u_x(x_i, t_n)h + \frac{1}{2}u_{xx}(x_i, t_n)h^2 - \frac{1}{6}u_{xxx}(x_i, t_n)h^3 + O(h^4).$$

Подставив выписанные разложения в (4.14), приведя подобные члены и воспользовавшись (4.4), получим

$$\psi_i^n = O(\tau + h^2).$$

Докажем, что условия (4.10) достаточно для сходимости в сильной норме, в норме С.

Введем норму на слое (и докажем сходимость в ней):

$$\|y^n\|_C = \max_{0 \le i \le N} |y_i^n|.$$

Введенная таким образом норма называется равномерной (сильной).

Докажем, что условие (4.16) является необходимым и достаточным для сходимости (и устойчивости) явной разностной схемы.

Докажем достаточность условия (4.16). Пусть это условие выполнено. Тогда

Выразим z_i^{n+1} в формуле (4.13): Рассматриваем только уравнение; начальные и краевые - там всё понятно.

Запишем разностную задачу для z (!!здесь ошибка 404 - смотрите у СПшников !!), оттуда:

$$z_i^{n+1} = (1 - 2\gamma)z_i^n + \gamma(z_{i-1}^n + z_{i+1}^n) + \tau\psi_i^n,$$

(коэффициент - неотрицательный при выполнении условия). Поэтому можем взять модуль и для коэффициента его не ставить.

$$|z_i^{n+1}| \le (1-2\gamma)|z_i^n| + \gamma(|z_{i-1}^n| + |z_{i+1}^n|) + \tau|\psi_i^n|,$$

Ясно, что неравенство только усилится, если вместо модулей поставим норму:

$$|z_i^{n+1}| \le (1-2\gamma)\|z^n\|_C + \gamma(\|z^n\|_c + \|z^n\|_C) + \tau\|\psi^n\|_C,$$

$$|z_i^{n+1}| \le \|z^n\|_C + \tau\|\psi^n\|_C,$$

поскольку это выполняется для всех i, то мы можем поставить норму:

$$\|z^{n+1}\|_C \le \|z^n\|_C + \tau\|\psi^n\|_C. \tag{4.15}$$

Это и есть ключевая оценка, которая позволит доказать сходимость, а потом и позволит сказать об устойчивости.

Применяя формулу (4.15) как рекуррентную, опустимся до нулевого слоя, получим:

$$\|z^{n+1}\|_C \le \|z^0\|_C + \tau\sum_{k=0}^{n}\|\psi^k\|_C, \tag{4.16}$$

поскольку $\|z^0\|_C = 0$, то

$$\|z^{n+1}\|_C \le \tau\sum_{k=0}^{n}\|\psi^k\|_C. \tag{4.17}$$

С легкостью получаем результат:

Т.к. $\psi_i^n = O(\tau + h^2)$, то $\exists\, M > 0 :\quad \|\psi^k\|_C \le M(\tau + h^2)$, M не зависит от τ и h.

Учитывая, что $\sum\limits_{k=0}^{n}\tau = t^{n+1} \le T$, имеем

$$\|z^{n+1}\|_C \le MT(\tau + h^2) = M_1(\tau + h^2).$$

При этом, M_1 не зависит от τ и h.

Мы получили априорную оценку

$$\|z^{n+1}\|_C \le M_1(\tau + h^2). \tag{4.18}$$

Из полученной оценки видно, что

$$\tau, h \to 0 \;\Rightarrow\; \|z^{n+1}\| \to 0, \text{ т.е. } \|y_i^{n+1} - u_i^{n+1}\| \to 0.$$

Таким образом, имеет место сходимость численного решения к решению исходной задачи с первым порядком по τ и вторым по h (говорят, что разностная схема имеет первый порядок точности по τ и второй по h).

Доказали сходимость, поговорим об устойчивости. Если мы рассмотрим нашу разностную схему при нулевых краевых условиях

$$y_0^{n+1} = y_n^{n+1} = 0,$$

то для сеточной функции y будет выполнено оценка вида 4.15.

$$\|y^{n+1}\|_C \leq \|y_0\|_C + \sum_{k=0}^{n} \tau \|f^k\|_C, \tag{4.19}$$

Наличие такой оценки (которую называют априорной) и называется устойчивостью. Говорят, что решение устойчиво по начальному условию и по правой части уравнения.

Примечание:
Оценка в доказательстве была получена в предположении $\frac{\tau}{h^2} < 0.5$, Так что, разностная схема - условно сходящаяся.

Докажем, что условие $\frac{\tau}{h^2} < 0.5$ является и необходимым.
Выпишем однородную систему:

$$\frac{y_i^{n+1} - y_i^n}{\tau} = \frac{y_{i-1}^n - 2y_i^n + y_{i+1}^n}{h^2}, \quad (x_i, t_n) \in \omega_{\tau h}. \tag{4.20}$$

Будем искать ее некоторые частные решения в виде $y_j^n = q^n e^{ijh\phi}$, где i - мнимая единица, $i^2 = -1$, $\phi \in R$, $q \in C$.
Подставим это в уравнение. Получим:

$$q = 1 + \gamma(e^{ih\phi} - 2 + e^{-ih\phi}) = 1 + \gamma(2\cos h\phi - 2) = 1 - 4\gamma \sin^2 \frac{h\phi}{2}.$$

Если взять ϕ такое, что $|q| > 1$, то получим $\gamma > \frac{1}{2}$,, - что будет означать неустойчивость.

Таким образом, условие (4.16) является необходимым и достаточным для сходимости и устойчивости явной разностной схемы.

Подводя итог, отметим следующие моменты рассмотренной схемы:

1. чрезвычайно удобна для реализации
2. первый порядок по τ, второй по h
3. условно сходящаяся и условно устойчивая

Чисто неявная разностная схема (схема с опережением).

Ставим в соответствие исходной задаче 4.1—4.3 следующую разностную схему:

$$\frac{y_i^{n+1} - y_i^n}{\tau} = \frac{y_{i-1}^{n+1} - 2y_i^{n+1} + y_{i+1}^{n+1}}{h^2} + f(x_i, t_{n+1}), \quad (x_i, t_{n+1}) \in \omega_{\tau h}, \tag{4.21}$$

$$\begin{cases} y_0^{n+1} = \mu_1(t_{n+1}), & t_{n+1} \in \overline{\omega}_\tau, \\ y_N^{n+1} = \mu_2(t_{n+1}), & t_{n+1} \in \overline{\omega}_\tau, \end{cases} \tag{4.22}$$

$$y_i^0 = u_0(x_i), \quad x_i \in \overline{\omega}_h. \tag{4.23}$$

Разностная схема записана на четырехточечном шаблоне

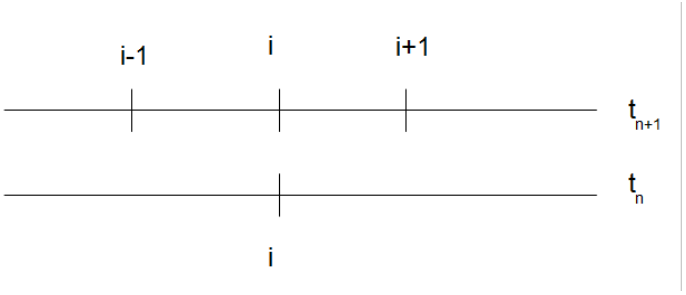

Рис. 4.5:

Эта разностная схема будет абсолютно сходящаяся и *абсолютно* устойчивая, в той же норме, с тем же порядком что и явная.

Но при этом неизвестно, имеет ли решение СЛАУ (схема неявная, нужно обращать матрицу). Эту разностную схему придется решать методом прогонки, это сложнее чем находить решение на явной схеме.

Мы записали так:

$$\gamma y_{i-1}^{n+1} - (1 + 2\gamma) y_i^{n+1} + \gamma y_{i+1}^{n+1} = -F_i(y^n), \quad i = 1, \ldots, N-1,$$

где $F_i = (y_i^n + f_i^{n+1})$

Выпишем матрицу данной системы:

$$S = -\begin{pmatrix} 1+2\gamma & -\gamma & 0 & 0 & \ldots & 0 \\ -\gamma & 1+2\gamma & -\gamma & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & 0 & -\gamma & 1+2\gamma & -\gamma \\ 0 & 0 & \ldots & 0 & -\gamma & 1+2\gamma \end{pmatrix}$$

Образуется матрица с диагональным преобладанием. Следовательно, решение существует, единственно, алгоритм решения - прогонка.

Рассмотрим неявную разностную схему, схему с опережением:

$$\frac{y_i^{n+1} - y_i^n}{\tau} = \frac{y_{i-1}^{n+1} - 2y_i + y_{i+1}^{n+1}}{n^2} + f(x, t_{n+1}) \tag{4.24}$$

$$y_0^{n+1} = \mu_1(t_{n+1});\ y_n^{n+1} = \mu_2(t_{n+1});\ t_{n+1} \in \bar{\omega_\tau} \tag{4.25}$$

$$y_i^0 = u_0(x_i);\ x_i \in \bar{\omega_\tau} \tag{4.26}$$

Обсудим вопрос сходимости и устойчивости разностной схемы. Необходимо определиться с тем, в какой области осуществляется сходимость. Покажем, что эти разностные схемы сходятся абсолютно.

Введем сетчатую функцию:

$$z_i^n = y_i^n - u(x_i, t_n) = y_i^n - u_i^n$$

Запишем уравнение

$$\frac{z_i^{n+1} - z_i^n}{\tau} = \frac{z_{i-1}^{n+1} - 2z_i^{n+1} + z_{i+1}^{n+1}}{h^2} + \psi_i^n \tag{4.27}$$

Нулевые начальные и краевые условия:

$$z_0^{n+1} = z_N^{n+1} = 0 \tag{4.28}$$

$$z_i^0 = 0 \tag{4.29}$$

Применяем принцип максимума (применяем для явной разностной схемы).
Погрешность аппроксимации на решении:

$$\psi_i^n = -\frac{u_i^{n+1} - u_i^n}{\tau} + \frac{u_{i-1}^{n+1} - 2u_i^{n+1} + u_{i+1}^{n+1}}{h^2} - f_i^{n+1} \tag{4.30}$$

Задача:
Показать, что $\psi_i^n = O(\tau + h^2)$
Решение:
Разложим $u(x_i, t_{n+1})$ в точке (x_n, 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 точке (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 точке (x_i, t_n) по формуле Тейлора:

$$u(x_{i-1}, t_n) = u_{i+1}^n = u(x_i, t_n) - u_x(x_i, t_n)h + \frac{1}{2}u_{xx}(x_i, t_n)h^2 - \frac{1}{6}u_{xxx}(x_i, t_n)h^3 + O(h^4)$$

Подставим данные разложения в 4.30, приведем подобные слагаемые и, принимая во внимание вид рассматриваемой задачи, получим

$$\psi_i^n = O(\tau + n^2)$$

Получаем оценку в сильной норме C. Мы знаем, что существуют нетривиальные решения $\Rightarrow \exists x_{i0} : max_{0\le i\le N} \mid z_i^{n+1} \mid = \mid z_{i_0}^{n+1} \mid = \|z^{n+1}\|_C$

В этом узле запишем уравнение 4.27:

$$(1+\frac{\tau}{h^2})z_{i_0}^{n+1} = \frac{\tau}{h^2}(z_{i_0-1}^{n+1} + z_{i_0+1}^{n+1}) + z_{i_0}^n + \tau\psi_{i_0}^n$$

Пусть $\gamma = \frac{\tau}{h^2} \Rightarrow$

$$(1+2\gamma) \mid z_{i_0}^{n+1} \mid \le \gamma(\mid z_{i_0-1}^{n+1} \mid + \mid z_{i_0+1}^{n+1} \mid) + \mid z_{i_0}^n \mid + \tau \mid \psi_{i_0}^n \mid$$

Усилим норму:

$$(1+2\gamma) \mid z_{i_0}^{n+1} \mid \le \gamma(\|z^{n+1}\|_C + \|z^{n+1}\|_C) + \|z^n\|_C + \tau\|\psi^n\|_C$$

$$(1+2\gamma)\|z^{n+1}\|_C \le 2\gamma\|z^{n+1}\|_C + \|z^n\|_C + \tau\|\psi^n\|_C$$

$$\|z^{n+1}\|_C \le \|z^n\|_C + \tau\|\psi^n\|_C$$

Мы получили ту же самую оценку, что и для явной разностной схемы, но без каких-либо ограничений на шаги сетки τ и h.

$$\|z^{n+1}\|_C \le \|z^0\|_C + \Sigma_{k=0}^n \tau\|\psi^k\|_C,$$

где $\|z^0\|_C = 0$

$$\|z^{n+1}\|_C \le \Sigma_{k=0}^n \tau\|\psi^n\|_C$$

$$\|\psi^n\| \le M(\tau + h^2),$$

где M не зависит от τ и h. Это важно при стремлении M к бесконечности.

$$\|z^{n+1}\|_C \le M_1(\tau + h^2),\ M_1 = TM$$

$$\|z^{n+1}\|_C \to 0,\ \tau, h \to 0 \Rightarrow$$

Значит имеет место сходимость разностной схемы к решению точной задачи.

Данная разностная схема имеет 1-й порядок точности по τ и второй по h.

Если в 4.24—4.26 при $y_0^{n+1} = y_N^{n+1} = 0 \Rightarrow$ { Если даже начальная функция не нулевая} $\Rightarrow$ получим ту же самую оценку:

$$\|y^{n+1}\|_C \le \|u_0\|_C + \tau\Sigma_{k=0}^n\|f^k\|_C,$$

где u_0 - начальное условие, а f_k-правая часть.

В этом случае говорят, что разностная схема устойчива по начальному условию и правой части. Если $\|y^{n+1}\| \le M_1\|u_0\|_C + M_2\|f\|_C$, где M_1, M_2 не зависят от шагов. Отсюда следует априорная оценка, означающая устойчивость по начальному условию и правой части.

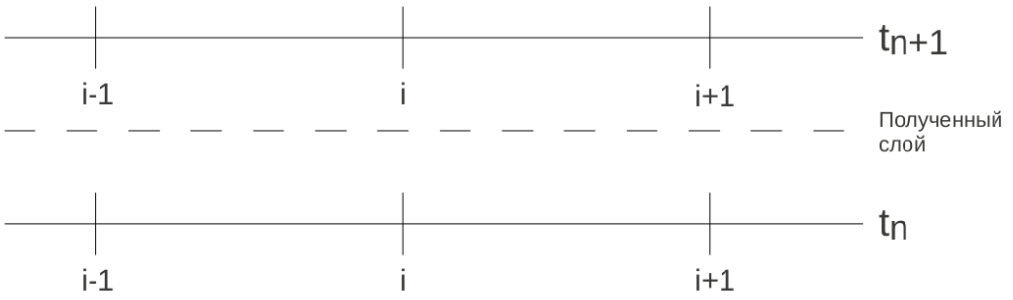

Рис. 4.6: Шаблон типа "ящик"

§ 3. Симметричная разностная схема (Схема Кранка-Никольсона)

На практике данная схема применяется чаще, поскольку в оценке второй порядок точности по τ и h.

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + f(x,t), 0 \le x \le 1, 0 \le t \le T \tag{4.31}$$

$$y(0,t) = \mu_1(t),\ y(1,t) = \mu_2(t), 0 \le t \le T \tag{4.32}$$

$$y(x,0) = u_0(x), 0 \le x \le 1 \tag{4.33}$$

$$y^n_{\bar{x}x,i} = \frac{y^n_{i+1} - 2y^n_i + y_{i-1}}{n^2}$$

Вид схемы:

$$\frac{y^{n+1}_i - y^n_i}{\tau} = 0.5(y^{n+1}_{\bar{x}x,i} + y^n_{\bar{x}x,i}) + f(x_i, t_{n+\frac{1}{2}}), (x_{i,t_{n+\frac{1}{2}}}) \in \omega_{\tau n}, \tag{4.34}$$

$$(x_i, t_{n+\frac{1}{2}}) = (x_i, t_n + 0.5\tau)$$

$$y^{n+1}_0 = \mu_1(t_{n+1}), y^{n+1}_N = \mu_2(t_{n+1}), t_{n+1} \in \bar{\omega}_\tau \tag{4.35}$$

$$y^0_i = u_0(x_i), x_i \in \bar{\omega}_h \tag{4.36}$$

Мы сопоставляем задаче 4.31—4.33 разностную схему 4.34—4.36.

Симметричный шаблон. Решение разностной схемы существует и единственно. Способ решения - прогонка. Если не говорить про сходимость по какой именно норме, то такая фраза не имеет смысла.

Покажем абсолютную сходимость. Введем погрешность $z_i^n = y_i^n - u_i^n$:

$$\frac{z_i^{n+1} - z_i^n}{\tau} = 0.5(z_{\bar{x}x,i}^{n+1} + z_{\bar{x}x,i}^n) + \psi_i^n$$

$$\psi_i^n = -\frac{u_i^{n+1} - u_i^n}{\tau} + 0.5(u_{\bar{x}x,i}^{n+1} + u_{\bar{x}x,i}^n) + f(x_i, t_n + \frac{1}{2}\tau) \tag{4.37}$$

Задача:
Показать, что $\psi_i^n = O(\tau^2 + n^2)$
Решение:
Разложим $u_{i\pm1}^{n+1}$ и u_i^n в ряд Тейлора в окрестности точки $(x_i, t_{n+0.5})$:

$$u_i^{n+1} = u_i^{n+0.5} + u_{t,i}^{n+0.5}\frac{\tau}{2} + \frac{1}{2}u_{tt,i}^{n+0.5}(\frac{\tau}{2})^2 + O(\tau^3)$$

$$u_i^n = u_i^{n+0.5} - u_{t,i}^{(}n + 0.5) - u_{t,i}^{n+0.5}\frac{\tau}{2} + \frac{1}{2}u_{tt,i}^{n+0.5}(\frac{\tau}{2})^2 + O(\tau^3)$$

Подставим полученные разложения в формулу 4.37:

$$\psi_i^n = -u_{t,i}^{n+0.5} + O(\tau^2) + 0.5(u_{\bar{x}x,i}^{n+1} + u_{\bar{x}x,i}^n) + f_i^{n+0.5}$$

В представлении фторой разностной производной разложим все вхождения функции в ряд Тейлора. Получим:

$$u_{\bar{x}x,i}^n = u_{xx,i}^n + u_{xxxx,i}^n\frac{h^2}{12} + O(h^4)$$

Применим полученное разложение к $u_{\bar{x}x,i}^{n+1}$, затем проведем ещё одно разложение в ряд Тейлора в точке $(x_i, t_{n+0.5})$:

$$u_{\bar{x}x,i}^{n+1} = u_{xx,i}^{n+1} + u_{xxxx,i}^{n+1}\frac{h^2}{12} + O(h^4) = u_{xx,i}^{n+0.5} + u_{xxt,i}^{n+0.5}\frac{\tau}{2} + u_{xxxx,i}^{n+0.5}\frac{h^2}{12} + u_{xxxxt,i}^{n+0.5}\frac{h^2}{12}\frac{\tau}{2} + O(\tau^2 + h^4)$$

Аналогично для $u_{\bar{x}x,i}^n$:
Подставим полученные разложения в выражение для ψ_i^n:

$$\psi_i^n = (-u_{t,i}^{n+0.5} + u_{xx,i}^{n+0.5} + f_i^{n+0.5}) + u_{xxxx,i}^{n+0.5}\frac{h^2}{12} + O(\tau^2 + h^4) = O(\tau^2 + h^2)$$

§ 4. Задача Штурма-Луивилля

$$\frac{d^2u}{dx^2} + \lambda u(x) = 0, \tag{4.38}$$

где $u(x) \neq 0,\ 0 \leq x \leq 1,\ u(0) = u(1) = 0$ - начальные условия. λ - собственное значение, $u(x)$- собственная функция.

Решение задачи $\lambda_k = \pi^2 k^2,\ k = 1, 2, \ldots$
$\lambda = 0$ не дает решение, поэтому в спектр не входит. $0 < \lambda_1 < \ldots < \lambda_n < \ldots$

Отвечающие им собственные значения функции $u_k(x) = c\, sin(\pi k x),\ c \neq 0,\ c = \sqrt{2} \to u_k(x) = \sqrt{2} sin(\pi k x)$

L_2- пространство функций, интегрируемых с квадратом. В нем скалярное произведение: $(u_k, u_l) = \sigma_{kl}$.

$\{u_k\}^\infty$-ортонормированный базис $\Rightarrow$

$$\forall f \in L_2, f(x) = \sum_{k=1}^{\infty} c_k u_k(x) \Rightarrow$$

$$\|f\|_{L_2}^2 = \sum_{k=1}^{\infty} c_k^2$$

- равенство Парсеваля.

$$y_{\bar{x}x_i} + \lambda y(x_i) = 0, x_i \in \omega_n, y(x_i) \neq 0 \tag{4.39}$$

$$y_0 = y_N = 0 \tag{4.40}$$

Проблема собственных значений аналитически решается редко.

$$y_{i+1} - 2y_i + y_{i+1} + h^2 \lambda y_i = 0$$

$$y(x_i) = sin(\alpha x_i),\ \alpha \in R,\ i = 1, \ldots, N-1$$

$$y_{i+1} + y_{i-1} = (2 - h^2\lambda) y_i$$

$$y_{i+1} + y_{i+1} = y(x_i + h) + y(x_i - h) = sin\alpha(x_i + h) + sin\alpha(x_i - h) = 2sin(\alpha x_i) cos(\alpha h)$$

$$2cos(\alpha x) sin(\alpha x_i) = (2 - h^2\lambda) sin(\alpha x_i)$$

$$2cos(\alpha h) = 2 - h^2\lambda$$

$$k^2\lambda = 2(1 - cos(\alpha h)) = 4sin^2 \frac{\alpha h}{2} \Rightarrow \lambda = \frac{4}{h^2} sin^2 \frac{\alpha h}{2}$$

$$y_N = 0 = sin\alpha \Rightarrow \alpha = \pi k \Rightarrow \lambda_k = \frac{4}{h^2} sin^2 \frac{\pi k h}{2},\ k = 1, \ldots, N-1 \tag{4.41}$$

$$y_k(x_i) = csin(\pi k x_i),\ x_i \in \bar{\omega}_n,\ k = 1, \ldots, N-1,\ c = \sqrt{2} \tag{4.42}$$

$$H_{N-1} : \forall f, g \in H_{n-1},\ dim H_{n-1} = N - 1,\ (f, g) = \sum_{i=1}^{N-1} f_i g_i h \Rightarrow$$

$$\|f\|_{L_2(\omega_n)} = \|f\| = \sqrt{\sum_{i=1}^{N-1} f_i^2 h}$$

$\{y_k\}_{k=1}^{N-1}$- Ортонормированный базис H_{N-1}; H_{N-1}— пространство сетчатых функций.

Обозначим $y_k(x_i) = \mu_k(x_i)$.
Если $c = \sqrt{2} \Rightarrow y_k(x_i) = \sqrt{2} sin\pi k x_i,\ i, k = 1, \ldots, N-1$

$$(y_k, y_l) = \sum_{k,l} \Rightarrow \forall f \in H_{N-1}\ f = \Sigma_{k=1}^{N-1} c_k \mu_k$$

Используем равенство Парсеваля:

$$\|f\|^2_{L_2(\omega_n)} = \sum_{k=1}^{N-1} c_k^2$$

$$z_i^k = \sum_{k=1}^{N} c_k(t_n)\mu_k(x_i), x_i \in \bar{\omega_h}$$

$$\psi_i^n = \sum_{k=1}^{N} \psi^{(k)}(t_n)\mu_k(x_i)$$

- погрешность аппроксимации.

$$\sum_{k=1}^{N} \frac{c_k(t_{n+1}) - c_k(t_n)}{\tau}\mu_k(x_i) = \sum_{k=1}^{N-1} 0.5(c_k(t_{n+1}) + c_k(t_n))\mu^{(k)}_{\bar{x}x,i} + \sum_{k=1}^{N-1} \psi^{(k)}\mu_k(x_i)$$

$$\mu^{(k)}_{\bar{x}x,i} = -\lambda_k \mu_k(x_i) \Rightarrow$$

рассматриваем только при k=1:

$$\Rightarrow \frac{c_k(t_{n+1}) - c_k(t_n))}{\tau} + 0.5\lambda_k(c_k(t_{n+1}) + c_k(t_n)) = \psi^{(k)}(t_n)$$

Получили задачу нахождения c_k.

$$(1 + 0.5\tau\lambda_k)c_k(t_{n+1}) = (1 - 0.5\tau\lambda_k)c_k(t_n) + \tau\psi^{(k)}(t_n)$$

$$c_k(t_{n+1}) = \frac{1 - 0.5\tau\lambda_k}{1 + 0.5\tau\lambda_k}c_k(t_n) + \frac{\tau}{1 + 0.5\tau\lambda_k}\psi^k(t_n) = q_k$$

$$\frac{c_k(t_{n+1}) - c_k(t_n)}{\tau} + 0.5\lambda_k(c_k(t_{n+1}) + c_k(t_n)) = \psi^{(k)}(t_n); \quad k = 1, 2, ..., n-1$$

$$(1 + 0,5\lambda_k\tau)c_k(t_{n+1}) = (1 - 0.5\tau\lambda_k)c_k(t_n) + \tau\psi^{(k)}(t_n)$$

$$c_k(t_{n+1}) = \frac{(1 - 0.5\lambda_k)}{1 + 0.5\tau\lambda_k}c_k(t_n) + \frac{\tau}{1 + 0.5\lambda_k}\psi^{(k)}(t_n)$$

$$|q_k| \leq 1$$

$$Z_i^{n+1} = \sum_{k=1}^{N-1} c_k(t_{n+1})\mu_k(x_i) = \sum_{k=1}^{N-1} q_k c_k(t_n)\mu_k(x_i) + \sum_{k=1}^{N-1} \frac{\tau}{1 + 0.5\tau\lambda_k}\psi^{(k)}(t_n)\mu_k(x_i);$$

последняя сумма $= W_i^{n+1}$

Тогда получим $\|Z^{n+1}\|_{L+2(w_n)} = \|Z^{n+1}\| \leq \|V^{n+1}\| + \|W^{n+1}\|$

$$\|V^{n+1}\|^2 = \ \{\text{согласно равенству Парсеваля}\} \ = \sum_{k=1}^{N-1} q_k^2 c_k^2(t_n) \leq \sum_{k=1}^{N-1} c_k^2(t_n) = \|Z^n\|^2$$

$$\|W^{n+1}\|^2 = \ \{\text{аналогично}\} \ = \sum \frac{\tau^2}{(1 + 0.5\tau\lambda_k)^2}(\psi^{(k)}(t_n))^2 = \tau^2\|\psi^n\|$$

$\|W^{n+1}\| \leq \tau\|\psi^n\|$
$\|Z^{n+1}\| \leq \|Z^n\| + \tau\|\psi^n\|$

Запишем решение на $n+1$ слое через нулевой слой:
$\|Z^{n+1}\| \leq \|Z^n\| + \sum_{j=0}^{n} \tau\|\psi(t_j)\|$
Тогда $\|z^0\| = 0$

$\psi^n = 0(\tau^n + h^2)$
M не зависит от τ и h
Это значит, что $\|\psi(t_j)\| \leq M(\tau^2 + h^2)$

$\|Z^{n+1}\|_{L_2(w_n)}$

$y_0^{n+1} = y_N^{n+1} = 0$

$\|y^{n+1}\|_{L_2(w_n)} \le \|U_o\| + \sum\limits_{j=0}^{n} \tau \|f\|$

Введем вещественный параметр, при определенных значениях которого будут получаться схемы, которые мы уже рассматривали. Но также может получаться бесконечное число других схем. Это семейство схем называется *схемы с весами.*

Выведем и покажем, как выводится погршеность аппроксимации при решении для любых разностных схем, которые мы рассматриваем.

Разностная схема с весами:

Погрешность аппроксимации на решении.

$$\frac{y_i^{n+1} - y_i^n}{\tau} = \sigma y_{\bar{x}x,i}^{n+1} + (1-\sigma) y_{\bar{x}x,i}^n + \phi_i^n \tag{4.43}$$

$+$ краевые условия (5) $+$ начальные условия (6). $\sigma \in \mathbb{R}$

1. Если мы возьмем $\sigma = 0; \phi_i^n = f_i^n$ то мы получим явную разностную схему.
2. $\sigma = 1; \phi_i^n = f_i^{n+1}$ - чисто неявная схема
3. $\sigma = 0.5; \quad \phi_i^n = f_i^{n+\frac{1}{2}}$ - симметричная схема Кранка-Никольсона

Для всех схем с весами легко получить все оценки, которые мы получали для среднеквадратичных норм - получить методом разделения переменных.

Вводим $Z_i^n = y_i^n - U_i^n$

$$\frac{Z_i^{n+1} - Z_i^n}{\tau} = \sigma Z_{\bar{x}x,i}^{n+1} + (1-\sigma) y_{\bar{x},i}^n + \psi_i^n$$

$$\psi_i^n = \sigma U_{\bar{x}x,i}^{n+1} + (1-\sigma) U_{\bar{x}x,i}^n - \frac{U_i^{n+1} - U_i^n}{\tau} + \phi_i^n \tag{4.44}$$

В узле $(x_i, t_{n+\frac{1}{2}})$

$$U_{i+1} = U_i + \frac{h}{2} U$$

Договоримся об обозначениях для частичных производных:

$$\frac{\partial U}{\partial t}(x, t_n + \frac{1}{2}) = \dot{U}$$

$$\frac{\partial U}{\delta x} = U'$$

$$U_{i+1} = Ui + hU_i' + \frac{h^2}{2} + \frac{h^2}{2}U_i'' + \frac{h^3}{6}U_i''' + ...$$
$$U_{i-1} = U_i - hUi' + \frac{h^2}{2}U_i'' - \frac{h^3}{6}U_i''' + ...$$

А на полуцелом слое:

$$U_i^{n+1} = U_i(t_{n+\frac{1}{2}}) - \frac{\tau}{2}\dot{U}_i(t_{n+\frac{1}{2}}) + \frac{\tau^2}{8}\ddot{U}_i'(t_{n+\frac{1}{2}})$$

$$U_I^n = U_i(t_{n+\frac{1}{2}}) - \frac{\tau}{2}\dot{U}_i(t_{n+\frac{1}{2}}) + \frac{\tau}{8}\dot{U}_i(t_{n+\frac{1}{2}}) - ....$$

$$U_{\bar{x}x,i} = \frac{U_{i+1} - 2U_i + U_{i-1}}{h^2} = U_i'' + \frac{h^2}{12}u_i'''' + O(h^4)$$

Поэтому когда мы говорим $U_{\bar{x}x,i} - U_i'' = \frac{h^2}{12}U_i'''' = O(n^2)$

первая производная по времени:$\frac{U_I^{n+1} - u_i^n}{\tau} = U_i(.)(t_{n+\frac{1}{2}}) + O(\tau^2)$

Теперь можем рассматривать ψ_i^n:

$$\psi_i^n = \sigma(U'' + \frac{\tau}{2}U''(.) + \frac{h^2}{12}U'''' + O(\tau h^2)) + O(h^4))+$$
$$+(1-\sigma)(U'' - \frac{\tau}{2}U''(.) + \frac{h^2}{12}U'''') - U(.) + \phi^n + O(\tau^2 + h^4)$$

В новых обозначениях: $U'' = \dot{U} - f$ $U'''' = \dot{u}' - f''$
Тогда мы получаем

$$U'' - \dot{U} + \phi_i^n + (\sigma - 0.5)\tau U''\dot{+}\frac{h^2}{12} + O(\tau^2 + h^4) =$$
$$U'' - \dot{U} + f(x_i, t_{n+\frac{1}{2}}) = \phi_i'' - f(x_i, t_{n+\frac{1}{2}}) + \tau[(\sigma - 0.5)+$$
$$+\frac{h^2}{12}]\dot{U}'''' - \frac{h^2}{12}f(\text{непонятно}) + O(\tau^2 + h^4)$$

$\sigma_* = \sigma = \frac{1}{2} - \frac{h^2}{12\tau}$
$\phi_i^n = f(x_i, t_{n+\frac{1}{2}}) + \frac{h^2}{12}f''(x_i, t_{n+\frac{1}{2}})$

$\psi = O(\tau^2 + h^4)$ - схема повышенного порядка аппроксимации
$\sigma = 0; \quad \phi_i^n = f_i^n; \quad O(\tau + h^2)$
$\sigma = 1; \quad \phi_i^n = f_i^{n+1}; \quad O(\tau + h^2)$
$\sigma = \frac{1}{2}; \quad \phi_i^n = f_i(t_n + \frac{1}{2})$
Когда $\sigma \neq$ рассмотренным σ - будем получать $O(\tau + h^2)$
Переходим к построению и изучению нового класса разностных схем - для уравнения Пуассона

§ 5. Разностные схемы для уравнения Пуассона. (Задача Дирихле)

Рассматриваем уравнение:

$$\frac{\partial^2 U}{\partial x_1^2} + \frac{\partial^2 U}{\partial x_2^2} = f(x_1, x_2); \quad (x_1, x_2) \in (D) \tag{4.45}$$

$$U(x_1, x_2) \text{на границе } \Gamma = \mu(x_1, x_2) \tag{4.46}$$

$D = \{(x_1, x_2); \quad 0 < x_1 < l_1, 0 < x_2 < l_2\}$

Начинаем решение с построения сетки:

$$\omega_h = \{ \quad (x_1^{(i)}, x_i^{(j)}) = x_{ij}, \quad x_1^{(i)} = ih; \quad N_1 h_1 = l_1; \quad x_2^{(j)} = jh_2; \quad N_2 h_2 = l_2 \quad \}$$

Напомним:

$$\frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2} = f(x_1, x_2) \;\; (x_1, x_2) \in D \tag{4.47}$$

$$u(x_1, x_2)|_r = \mu(x_1, x_2) \tag{4.48}$$

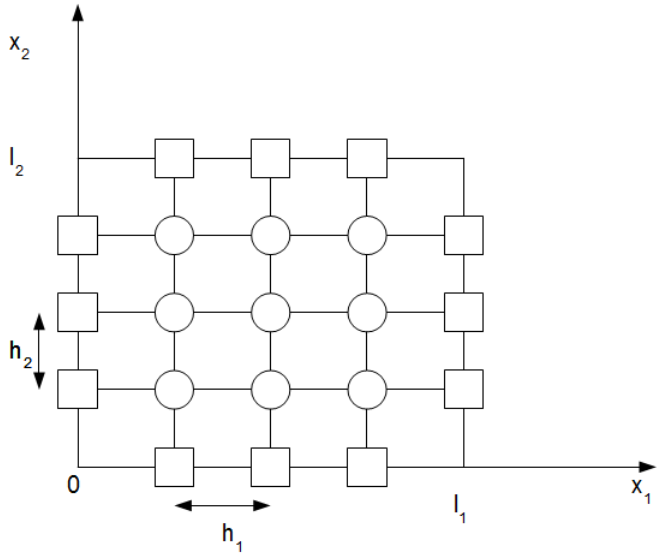

Рис. 4.7:

$$\Gamma_n = \{x_{0,j}\}_1^{N_2-1} \cup \{x_{N_1,j}\}_1^{N_2-1} \cup \{x_{i,0}\}_1^{N_1-1} \cup \{x_{i,N_2}\}_1^{N_1-1}$$

$\bar{\omega}_n = \omega_n$кружочки $\cup\, \Gamma_n$квадратики — все узлы
Вторая производная:

$$y_{\bar{x}_1 x_1; i,j} = \frac{y_{i+1,j} - 2y_{i,j} + y_{i-1,j}}{h_1^2}$$

$$f_{i,j} = f(\bar{x}_1^{(i)}, x_2^{(i)}) = f(x_{i,j})$$

$$y_{\bar{x}_1 x_1; i,j} + y_{\bar{x}_2 x_2; i,j} = f_{i,j} \;\; x_{i,j} \in \omega_h \tag{4.49}$$

— разностная аппроксимация

Граничное условие первого рода — поэтому аппроксимируем точно $y_{i,j}|_{\Gamma_n} = \mu_{i,j}$

$$\begin{cases} 1 \le i \le N_1 - 1 \\ 1 \le j \le N_2 - 1 \end{cases} \tag{4.50}$$

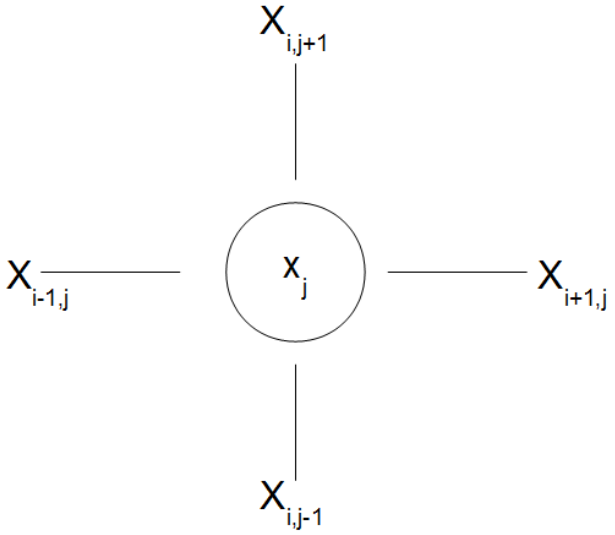

Рис. 4.8: шаблон разностной схемы типа «крест»

Изучим 5 вопросов: существование-единственность решения, погрешность аппроксимации (самый лёгкий)...

Первое суждение о погрешности:

$$z_{i,j} = y_{i,j} - u_{i,j}$$

$$z_{\bar{x}_1 x_1; i,j} + z_{\bar{x}_2 x_2; i,j} = -\Psi_{i,j} \tag{4.51}$$

$$z_{i,j}|\Gamma_n = 0 \tag{4.52}$$

$$\Psi_{i,j} = u_{\bar{x}_1 x_1; i,j} + u_{\bar{x}_2 x_2; i,j} - f_{i,j} \tag{4.53}$$

— невязка.

Задача:
Показать что $\Psi_{i,j} = O(h_1^2 + h_2^2)$б $\quad (u(x_1, x_2) \in C^4(\bar{D}))$
Подсказка: $\Psi_{i,j} = \frac{\partial^2 u}{\partial} x_1^2 + \frac{h^2}{12}(...)$

§ 6. Разрешимость разностной задачи. Сходимость разностной схемы

Распишем 4.49 относительно центрального узла, записав координатно:

$$(\frac{2}{h_1^2}+\frac{2}{h_2}^2)y_{i,j} = \frac{y_{i-1,j}+y_{i+1,j}}{h_1^2}+\frac{y_{i,j+1}+y_{i,j-1}}{h_2^2}-f_{i,j} \;\; 1 \le i \le N_1-1, 1 \le j \le N_2-1 \tag{4.54}$$

$$y_{i,j}|_{\Gamma_n} = \mu_{i,j}$$

Явно мы ничего вычислить не можем. Для доказательства существования будем опираться на аппарат алгебры. (Ранги расширенных матриц должны совпадать с числом неизвестных; другими словами, однородная система имеет только тривиальное решение — значит, определитель будет отличен от нуля.)

Записываем соответствующую однородную задачу.

$$(\frac{2}{h_1^2}+\frac{2}{h_2}^2)v_{i,j} = \frac{v_{i-1,j}+v_{i+1,j}}{h_1^2}+\frac{v_{i,j+1}+v_{i,j-1}}{h_2^2} \;\; 1 \le i \le N_1-1, 1 \le j \le N_2-1 \tag{4.55}$$

$$v_{i,j}|_{\Gamma_n} = 0$$

Матрица получилась бы чрезвычайно любопытного вида: матрица из трёхдиагональных блоков. Алгоритм решения — потом. А сейчас всё-таки существование и единственность.

Теорема 1:
Система 4.55 имеет только тривиальное решение:

$$v_{i,j} === 0 \;\; x_{i,j} \in \bar{\omega}_n$$

Доказательство:
Доказательство — на основе принципа максимума.

$\exists$ узел $x_{i_0,j_0} : |v_{i_0,j_0}| = \max\limits_{\text{от 0 до } N_n i,j} |v_{i,j}| = \|v\|_C$; $v_{i_0,j_0} \neq 0$. А это норма - не та! Это $C(\bar{D})$!

Среди этих узлов выберем один с двумя свойствами:
1. $|v_{i_0,j_0}| = \|v\|_C$;

2. Хотя бы в одном из четырёх соседних узлов: $|v_{i,j}| < |v_{i_0,j_0}|$.

Он существует, по крайней мере потому, что на границах ноль, а в узле — нет.

Наше уравнение в узле:

$$(\frac{2}{h_1^2} + \frac{2}{h_2}^2)v_{i_0,j_0} = \frac{v_{i_0-1,j_0} + v_{i_0+1,j_0}}{h_1^2} + \frac{v_{i_0,j_0+1} + v_{i_0,j_0-1}}{h_2^2}$$

Мажорируем:

$$(\frac{2}{h_1^2} + \frac{2}{h_2}^2)|v_{i_0,j_0}| \leq \frac{|v_{i_0-1,j_0}| + |v_{i_0+1,j_0}|}{h_1^2} + \frac{|v_{i_0,j_0+1}| + |v_{i_0,j_0-1}|}{h_2^2}$$

Из четырёх слагаемых есть хотя бы одно, в котором значение строго меньше, ведь мы ставили свойство узла 2.

$$(\frac{2}{h_1^2} + \frac{2}{h_2}^2)\|v\|_C < (\frac{2}{h_1^2} + \frac{2}{h_2}^2)\|v\|_C$$

— противоречие. Доказали, что однородная система имеет только тривиальное решение.

чтд.

Следствие:

Возвращаемся к исходной схеме 4.49 - 4.50. Эта схема имеет единственное решение при любых μ и $0f$.

Ключевой вопрос — сходимость и устойчивость.

Сразу будем доказывать сходимость, потому что она итак есть по теореме Кэли. Доказывать будем в норме C.

Даже аспиранты при сдаче кандминимума от этого воют.

Погрешность оценивается несложно, имеет второй порядок. Из всего, что было, знаем, что если получим оценку: $\|z\|_C \leq M(h_1^2 + h_2^2)$, где M не зависит от h_1, h_2, то при $h_1 \to 0, h_2 \to 0$: $\|z\|_C = \|y - u\|_C \to 0$

Может сразу разрешим относительно центрального узла?

$$\frac{y_{i-1,j} - 2y_{i,j} + y_{i+1,j}}{h_1^2} + \frac{y_{i,j-1} - 2y_{i,j} + y_{i,j+1}}{h_2^2} = f_{i,j}$$

Для большей общности введём оператор:

$$L_h v_{i,j} = (\frac{2}{h_1^2} + \frac{2}{h_2^2})v_{i,j} - \frac{v_{i-1,j} + v_{i+1,j}}{h_1^2} - \frac{v_{i,j-1} + v_{i,j+1}}{h_1^2}$$

$$x_{i,j} \in \omega_h$$

Лемма:
Пусть $v_{i,j} \geq 0$ на границе $x_{i,j} \in \Gamma_h$, и пусть $L_h v_{i,j} \geq 0$ внутри $x_{i,j} \in \omega_h$. Тогда $v_{i,j} \geq 0$ везде $x_{i,j} \in \bar{\omega}_h$. (принцип max)

Доказательство:
Как и положено, доказываем от противного.
$\exists$ узел x_{i_0,j_0}: $v_{i_0,j_0} < 0$.
Выберем, опять-таки, узел с двумя условиями:
1. $v_{i_0,j_0} < 0$;
2. $v_{i_0,j_0} < v_{\text{сосед}}$.

Перепишем в этом узле.

$$L_h v_{i_0,j_0} = \frac{v_{i_0,j_0} - v_{i_0-1,j_0}}{h_1^2} + \frac{v_{i_0,j_0} - v_{i_0+1,j_0}}{h_1^2} + \frac{v_{i_0,j_0} - v_{i_0,j_0-1}}{h_2^2} + \frac{v_{i_0,j_0} - v_{i_0,j_0+1}}{h_2^2}$$

Одно из слагаемых строго меньше нуля по второму условию. Тогда $L_h v_{i_0,j_0} < 0$ — противоречие.

Подойдём к теореме сравнения, там будет мажоранта.

Следствие 1 (дискретный аналог теоремы сравнения):

$$L_n y_{i,j} = \phi_{i,j} \tag{4.56}$$

$$L_n Y_{i,j} = \Phi_{i,j} \tag{4.57}$$

(оба — $x_{i,j} \in \omega_n$). $y_{i,j}|\Gamma_n$, $Y_{i,j}|\Gamma_n$ задано.

Утверждается: если $|y_{i,j}| \leq Y_{i,j}$ на границе $x_{i,j} \in \Gamma_n$. (Y — мажоранта, она всегда неотрицательна, можно так сравнивать с модулем), и если $|\phi_{i,j}| \leq \Phi_{i,j}$ на ω, то $|y_{i,j}| \leq Y_{i,j}$ — везде.

Что это нам дает: мажоранту выберем так, что оценку сделаем через шаги, и получим априорную оценку, которая доказывает сходимость и является оценкой устойчивости.

$v_{i,j} = Y_{i,j} - y_{i,j}$; $w_{i,j} = Y_{i,j} + y_{i,j}$. Подействуем L_h:

$L_h v_{i,j} = \Phi_{i,j} - \phi_{i,j} \geq 0$ (потому что $v_{i,j} \geq 0, x_{i,j} \in \Gamma_n$) — следовательно $v_{i,j} \geq 0, x_{i,j} \in \omega_h$.

$L_h w_{i,j} = \Phi_{i,j} + \phi_{i,j} \geq 0$ (потому что $w_{i,j} \geq 0, x_{i,j} \in \Gamma_n$) — следовательно $w_{i,j} \geq 0, x_{i,j} \in \omega_h$.

Это означает, что $|y_{i,j}| \leq Y_{i,j}$.

Займёмся построением мажорант.

Перепишем задачу для z через оператор. $L_h z_{i,j} = \Psi_{i,j}$.

Знак здесь действительно поменяется на противоположный, это не опечатка.

Наша задача:

$$\begin{cases} L_h z_{i,j} = \Psi_{i,j} \;\; x_{i,j} \in \omega_n \\ z_{i,j}|_{\Gamma_n} = 0 \end{cases} \tag{4.58}$$

$$Y_{i,j} = K(l_1^2 + l_2^2 - (x_1^{(i)})^2 - (x_2^{(j)})^2) \tag{4.59}$$

$K > 0$ — константа (выберем потом).
$Y_{i,j} \geq 0 \; x_{i,j} \in \bar{\omega}_h$.

Задача:
Показать, что $L_h Y_{i,j} = 4K$

Решение:
Положим $L_h Y_{ij} = K_1, \; K_1 > 0$.
Представим $Y_{ij} = (l_1^2 + l_2^2 - (x_1{}^{(i)})^2 - (x_2{}^{(j)})^2)K, \; K > 0$

$$Y_{ij} L_h \geq 0, \; x_{ij} \in \bar{\omega} \Rightarrow$$

Мы берем от каждого по двойке, а две константы обнуляем.

$$Y_{ij}L_h = 4K$$

Положим $|||\psi||_c = 4K$

$$\begin{cases} L_n Y_{i,j} = \|\Psi\|_C, x_{i,j} \in \omega_h; 4K = \|\Psi\|_C \\ Y_{i,j}|_{\Gamma_n} \geq 0, x_{i,j} \in \Gamma_n \end{cases} \tag{4.60}$$

Попадаем в условия следствия 1, поэтому $|z_{i,j}| \leq Y_{i,j}$ на $x_{i,j} \in \bar{\omega}_h$.

$0 \leq Y_{i,j} \leq \frac{l_1^2+l_2^2}{4}\|\Psi\|_C$

$M = \frac{l_1^2+l_2^2}{4}$ - Получается, что M зависит только от границ области.

$$\|z\|_C \leq \frac{l_1^2 + l_2^2}{4}\|\Psi\|_C \tag{4.61}$$

Это означает устойчивость для y: вместо Ψ станет ϕ:
4.49 $+y_{i,j}|_{\Gamma_n} = 0$ — в точности та задача, которую мы исследовали для z. Значит,

$$\|y\|_C \leq \frac{l_1^2 + l_2^2}{4}\|\phi\|_C \tag{4.62}$$

— она и означает устойчивость.

Теорема 2:

$$u(x_1, x_2) \in C^4(\bar{D})$$

Тогда разностная схема 4.49 - 4.50 сходится к решению задачи 4.47 - 4.48 со вторым порядком по h_1 и h_2 (или имеет второй порядок точности).

Доказательство:
$\|\Psi\|_C \leq M(h_1^2 + h_2^2)$, $M > 0$ не зависит от h_1, h_2.
Получаем: $\|z\|_C \leq M_1(h_1^2 + h_2^2)$. $M_1 = M\frac{l_1^2,l_2^2}{4}$ — не зависит от h_1, h_2.
А так как $\|z\|_C = \|y - u\|_C$ — вот и доказали.
чтд.

Единственный зависший вопрос — как решать-то? Покажем на следующей лекции: есть и прямые методы, но самые распространённые — итерационные. (Метод Самарского про переменный треугольник)

Методы решения разностной задачи Дирихле.
Развернутый вид:

$$\frac{y_{i-1}-2y_{ij}+y_{i+1,j}}{h_1^2}+\frac{y_{i,j-1}-2y_{i,j}+y_{i,j+1}}{h_2^2}=f_{i,j},\ i=1,\dots,N_1-1,\ j=1,\dots,N_2-1 \tag{4.63}$$

$$y_{ij}\mid_{\Gamma_n}=\mu_{ij},\ x_{ij}\in\Gamma_n \tag{4.64}$$

т.к необходима хорошая точность ($h_1\to 0,\ h_2\to 0$), то чем меньше мы возьмем шаг, тем выше будет точность.
Возможна ситуация, когда необходимо решение оператора Лапласса с 1000 уравнений. Встает вопрос: "Как решать"? Везде применять - это нерационально. Применение метода Гаусса - нерационально. (Сложность $\sim n^3$).
Самое широкое применение получил итерационый метод.

Уравнение для центрального узла:

$$(\frac{2}{h_1^2}+\frac{2}{h_2^2})y_{ij}=\frac{y_{i-1,j}+y_{i+1,j}}{h_1^2}+\frac{y_{i,j-1}+y_{i,j+1}}{h_2^2}-f_{ij},$$

$y_{ij}^{(S)}-S$-я итерация.

Метод попеременной итерации.

$$(\frac{2}{h_1^2}+\frac{2}{h_2^2})y_{ij}^{(s+1)}=\frac{y_{i-1}^{(s)}+y_{i+1,j}^{(s)}}{h_1^2}+\frac{y_{i,j-1}^{(s)}+y_{i,j+1}^{(s)}}{h_2^2}-f_{ij}, \tag{4.65}$$

y_{ij}^0- начальное приближение $S=0,1,\dots$
Примем $h=h_1=h_2\Rightarrow h_0(\varepsilon)\sim O(h^{-2})\Rightarrow$ сходимость чрезвычайно медленная. Это не экономичный метод.

Метод Зейделя (неявный)

$$\frac{2}{h_1^2}+\frac{2}{h_2^2})y_{ij}^{(s+1)}=\frac{y_{i-1,j}^{(s+1)}+y_{i+1,j}^{(s)}}{h_1^2}+\frac{y_{i,j-1}^{(s+1)}+y_{i,j+1}^{(s)}}{h_2^2}-f_{ij}, \tag{4.66}$$

$y_{i,j}^{(0)}$-задан, $s=0,1,2,\dots$

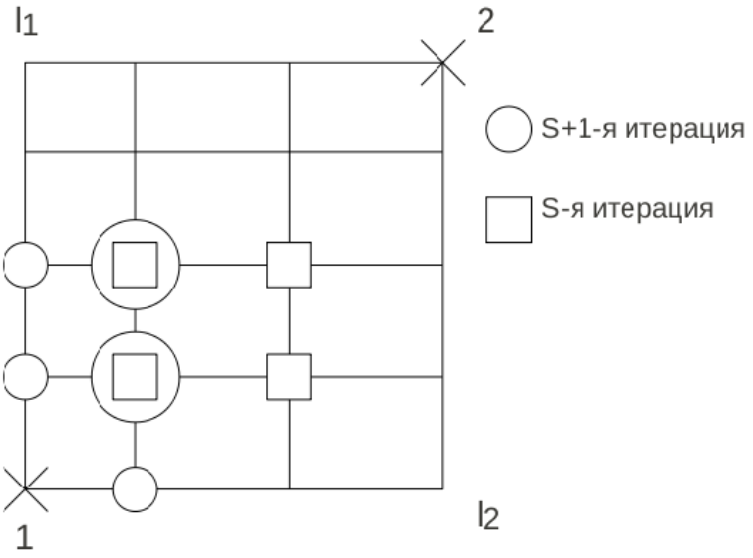

Рис. 4.9:

Мы движемся из 1 в 2.
$y_{1,j} : j = 1, \ldots, N_2 - 1$
$y_{2,j} : j = 1, \ldots, N_2 - 1$
СЛАУ: $Ay = \varphi$ - разностная задача. $A = A^* > 0$
Представим $A = R_1 + R_2$, где R_1-матрица, имеющая нижнюю треугольную форму с $0.5a_i$ на диагонали. K_2-матрица, имеющая верхнюю треугольную форму.

$$(E + wR_1)(E + wR_2)\frac{y^{(s+1)} - y^{(s)}}{\tau} + Ay^{(s)} = \varphi \tag{4.67}$$

где $\tau \geq 0,\ w > 0,\ w > \frac{\tau}{4}$ –необходимое условие сходимости.
Начальное условие y_0-задано, $s = 0, 1, \ldots$

Обозначим:

$$1)(e + wR_2)\frac{y^{(s+1)} - y^{(s)}}{\tau} = W^{(s+1)}$$

$$2)(E + wR_1)W^{(s+1)} = \varphi - Ay^{(s)}$$

$$3)V^{(s+1)} = \frac{y^{(s+1)} - y^{(s)}}{\tau}$$

Из 1-го $\rightarrow W^{(s+1)}$
Из 2-го $\rightarrow V^{(s+1)}$
Из 3-го $\rightarrow y^{(s+1)} = y^{(s)} + \tau v^{(s+1)} \Rightarrow n_0(\varepsilon) = O(h^{-1})$. Этот метод является ведущим.

§ 7. Основные понятия теории разностных схем: аппроксимация, устойчивость, сходимость

Следующие высказывания присуще всем линейным задачам математической физики.

$$LU(x) = f(x),\ x \in G \tag{4.68}$$

где L-линейный дифференциальный оператор.
$x = (x_1, \ldots, x_m)$-линейный вектор.

Начинаем с построения сетки.
$G \to G_h$, h-некоторая норма шагов (обобщенная характеристика). На практике выбор сетки - серьезный вопрос.

Вводим сеточные функции. y_h, L_H.

Разностная схема:

$$L_H y_h(x) = \varphi_n(x),\ x \in G_n,\ h =\mid h \mid \tag{4.69}$$

Аппроксимируем исходную дифференциальную задачу. Научимся измерять расстояние между функциями из разностных нормированных пространств.
$y(x) \in B_0,\ x \in G$
$y_h(x) \in B_h,\ x \in G_h$
$P_h : B_i \to B_h$- оператор проектирования.
$P_h U = U_{h(x)}$- на сетке $x \in G_h$

Вводим нормы, чтобы исследовать сходимость на норме.
$\|.\|_0$ -в B_0; $\|.\|_h$ - в B_h Нормы должны быть согласованы $\lim\limits_{h \to 0} \|u_h\| = \|u\|_0$

Пусть область $G : 0 \le x \le 1$. Построим сетку $G_h = \{x_i = ih, i = 0, 1, \ldots, N,\ hN = 1\}$.

Рассмотрим сильную норму C:
$\|u\|_0 = max|u(x)| = \|u\|_C, 0 \le x \le 1$
$\|y\|_h = max|y_i| = \|y\|_C, 0 \le x \le 1$

Нормы должны быть согласованы.

Пример согласованной нормы:

$$\|u\|_0 = (\int\limits_0^1 u^2(x)dx)^{\frac{1}{2}}$$

$$\|y\|_h = (\sum_{i=0}^{N} y_i^2 h)^{\frac{1}{2}}$$

Пример несогласованной нормы:

$$\|y\|_h = (\sum_{i=0}^{N} y_i^2)^{\frac{1}{2}}$$

- не согласуется ни с одной нормой в B_0

$$u(x) \equiv 1 \|u_h\|_h = (\sum_{i=0}^{N} 1)^{\frac{1}{2}} = (N+1)^{\frac{1}{2}}$$

Если норма не согласована, то возможно существование 2-х функций, к которым сходится искомая.

$$P_h(u)_i = u(x_i)$$

$$P_h(u)_i = \frac{1}{h} \int\limits_{x_i - 0.5h}^{x_i + 0.5h} u(x)dx, \ i = 1, \ldots, N-1$$

$$(P_h u)_0 = \frac{1}{0.5h} \int\limits_0^{0.5h} u(x)dx, (P_h u)_N = \frac{1}{0.5} \int\limits_{1-0.5h}^{1} u(x)dx$$

Можем выбирать оператор проектирования как нам удобно.

Пусть

$$z_h = y_h - u_h \tag{4.70}$$

Сетчатая функция 4.70 называется погрешностью разностной схемы
$\Rightarrow y_h = z_h + u_h$
$L_h z_h + L_h u_h = \varphi_n$
$L_n z_n = \psi_n$ -разность,

$$\psi_h - L_h u_h \tag{4.71}$$

- невязка

Сетчатая функция 4.71 называется погрешностью аппроксимации разностной схемы на решении исходной задачи.

Определение 2:
Говорят, что разностная схема имеет к-й порядок аппроксимации, если $\exists M_1 > 0, k > 0$, не зависящие от шагов, для которых справедлива оценка $\|\psi_n\|_h \leq M_1 h$
k-не обязательно натуральное число.

Задача 4.68 построена коорректно, т.к.
1) $\exists!\ u(x) : \forall f(x), x \in G$,
2) непрерывно зависит от правой части f(x)

Определение 3:
Говорят, что разностная схема называется корректно поставленной, если
1)$\exists!\ y_h \in B_h\ \forall \varphi_n$, 2)$\exists M_2 > 0$, не зависящее от шагов :

$$\|y_h\|_h \leq M_2 \|\psi_n\|_h \tag{4.72}$$

Определение 4:
Говорят, что решение разностной задачи 4.69 сходится к решению разностной задачи 4.68, если $\|z_h\|_h = \|y_h - u_h\|_h \to 0, h \to \infty$

Определение 5:
Говорят, что разностная схема имеет к-й порядок точности (сходится с к-м порядком), если $\exists M_3 > 0$, k- не зависит от h $\Rightarrow \|z_h\|_h \leq M_3 h^k$

Теорема 1:
Пусть исходная задача 4.68 корректно поставлена и пусть разностная схема 4.69, аппроксимирующая задачу 4.68 - корректна $\Rightarrow$ решение разностной задачи

сходится к решению исходной с порядком погрешности аппроксимации.

Доказательство:
$\|y_h\|_h \leq M_2\|\varphi_h\|_h,\ M_2$ не зависит от h.
$\|z_h\|_h \leq M_2\|\psi_h\|_h$

$\|\psi_h\|_h \leq M_1 h^k,\ M_1$ не зависит от h.

$\|z_h\|_h \leq M_3 h^k, M_3 = M_1 M_2$ не зависят от h.

$\lim\limits_{h\to 0} \|z_h\|_h = 0$
чтд.

Выводы: Изучение разностных схем для линейных задач проходит в 2 этапа:

1. вычисляется погрешность аппроксимации
2. находится априорная оценка (то есть исследуется устойчивость)

При доказательстве теоремы 1 из прошлой лекции мы не использовали согласованность норм. Покажем, что если услвоия согласованности не было, то мы не могли бы гарантировать, что есть к нужному вектору.

Мы показали, что при $h \to 0$ разность между сеточной функцией и её проекцией в узлы точечного решения стремится к нулю:

$$\|y_h - u_h\|_h \to 0, h \to 0,$$

т.е. существует $u(x) \in B_0$ - решение $Lu = f$.

Но доказанная теорема не утверждает, что не существует функции $v(x) \in B_0$ такой, что$\|y_h - v_h\|_h \to 0$, и при этом *не является решением*

А обеспечивается это как раз условием согласованности:

$$\|u_h - v_h\|_h = \|u_h - y_h + y_h - v_h\| \leq \text{(неравенство треугольника)}$$

$$\leq \| - y_h + u_h\|_h + \|y_h - v_h\|_h \to 0 \text{ при } h \to 0$$

Так как норма согласованная, то:

$$\lim_{h\to 0} \|u_h - v_h\|_h = \|u - v\|_0 = 0 \Rightarrow u(x) \equiv v(x)$$

Таким образом, задача сходится к единственному решению.

Глава 5

Методы решения обыкновенных дифференциальных уравнений (ОДУ) и систем ОДУ

§ 1. Постановка задачи Коши и примеры численных методов решения задачи Коши

Теперь переходим к заключительной главе курса.

Предмет изучения - задача Коши для системы ОДУ.

$$\begin{cases} \frac{du}{dt} = f(t, u(t)), \quad t > 0, \\ u(0) = u_0; \end{cases} \tag{5.1}$$

$$u(t) = (u_1(t), u_2(t), \ldots, u_m(t))^T$$

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

Первый факт - задача нелинейная (математики говорят, что это *слабая нелинейность*. Значит, разностные схемы будут требовать подхода, отличного от того, что мы применяли ранее.

Рассмотрим параллелепипед $R = \{(t,u), |t| \le a, |u - u_0| \le b\}$.

Из курса обыкновенных дифференциальных уравнений известно, что если решаем корректные задачи,то мы должны иметь представление о существовании и единственноти решения, их непрерывной зависимости от входных данных.

Если f - непрерывная функция, по второму аргументу удовлетворяющая условию Липшица:

$$|f(t,u) - f(t,v)| \le L|u - v|, \quad L = const,$$

то решение задачи Коши существует и единственно для некоторых $t > 0$ (говорят: *существует в малом*).
Доказывают это утверждение методом Пикара, переходя к интегральному соотношению

$$u(t) = u(0) + \int_0^t f(t, u(x))dx$$

и организуя последовательные приближения

$$u_{n+1}(t) = u(0) + \int_0^t f(t, u_n(x))dx, \quad n = 0, 1, \dots$$

На практике же интеграл в правой части может не вычисляться, поэтому существуют численные методы решения поставленной задачи.

Среди численных методов решения задачи Коши широкое распространение нашли 2 подхода:

1. методы Рунге-Кутта

2. многошаговые разностные методы

И та, и другая группа находит широкое применение на практике. Мы скажем, какой из них когда удобнее, где можно добиться большей точности и т. д.
Рассмотрим примеры.

Введем сетку:

$$\omega_\tau = \{t_n = n\tau, \tau > 0, n = 0, 1, 2 \dots\}$$

Пример 1:
Схема Эйлера. Явная и неявная схема Эйлера обладают одинаковой погрешностью, но с точки зрения устойчивости и сходимости разные. Явные схемы обычно условно устойчивые, в то время как неявные - абсолютно устойчивые.

Пусть на введенной сетке: $u_n = u(t_n), f(t_n, y(t_n)) = f_n; y_n -$. Явная схема Эйлера имеет вид:

$$\begin{cases} \frac{y_{n+1}-y_n}{\tau} = f_n, \quad t_n \in \omega_\tau, \\ y(0) = u_0; \end{cases} \tag{5.2}$$

y_{n+1} явно выражается из первого уравнения:

$$y_{n+1} = y_n + \tau f_n$$

Все компоненты в правой части известны, то есть y_{n+1} можно найти в явном виде.
Невязка 4.71:

$$\psi_n = -\frac{u_{n+1} - u_n}{\tau} + f(t_n, u_n) \tag{5.3}$$

Сеточная функция 5.3 называется погрешностью апроксимации разностной схемы 5.2 на решение исходной задачи 5.1.

Разложим u_{n+1} в ряд Тейлора:

$$\frac{u_{n+1} - u_n}{\tau} = u'_n + O(\tau)$$

$$\psi_n = -u'_n + f(t_n, u_n) + O(\tau)$$

Учитывая, что $-u'_n + f(t_n, u_n) = 0$, получаем:

$$\psi_n = \underline{O}(\tau)$$

Соответственно, мы можем ожидать сходимость не выше первого порядка.

Погрешность:

$$|y_n - u(t_0)| \leq M\tau, \quad M > 0 \text{не зависит от } \tau$$

Позднее мы покажем, что схема имеет первый порядок точности.

Пример 2:
Метод Рунге-Кутта, двухэтапный (или схема *предиктор-корректор*).

(Схема предиктор-корректор, схема второго порядка, более точная, широко используется на практике)

Переход от t_n к t_{n+1} осуществляют в 2 этапа:

$$t_n \to t_{n+\frac{1}{2}} \to t_{n+1}$$

Поставим в соответствие задаче 5.1 разностную схему, введя при этом полуцелый слой:

$$\begin{cases} \frac{y_{n+\frac{1}{2}}-y_n}{0.5\tau} = f(t_n, y_n) \\ \frac{y_{n+1}-y_n}{0.5\tau} = f(t_{n+\frac{1}{2}}, y_{n+\frac{1}{2}}) \\ y(0) = u_0; n = 0, 1, 2, ... \end{cases} \tag{5.4}$$

Решение реализуется следующим образом:

$$y_{n+\frac{1}{2}} = y_n + 0.5 f_n$$

$$y_{n+1} = y_n + \tau f(t_{n+\frac{1}{2}}, y_n + 0.5\tau f(t_n, y_n))$$

У нас 2 этапа, поэтому она и называется схемой предиктор-корректор.

Вывод погрешности аппроксимации выведем для общего случая. Позднее получим и точность.

Погрешность аппроксимации общего двухэтапного метода Рунге-Кутта

$$\begin{cases} \frac{y_{n+1}-y_n}{\tau} = \sigma_1 K_1 + \sigma_2 K_2, \sigma_1, \sigma_2 \in R \\ y_0 = u_0 \\ K_1 = f(t_n, y_n), \\ K_2 = f(t_n + a_2\tau, y_n + b_{21}\tau K_1); \end{cases} \tag{5.5}$$

Вещественые числа a и b будут управлять погрешностью аппроксимации (индексы выбраны для соответствия с общим подходом, рассмотренным далее).

Вопрос на экзамене: Мы выбрали сигма 3/2 и 4/3 и попросим посчитать. На самом деле сумма сигм должна быть равна 1, иначе нет никакой аппроксимации!

Перепишем уравнение в задаче 5.5 в виде

$$\frac{y_{n+1}-y_n}{\tau} = \sigma_1 f(t_n, y_n) + \sigma_2 f(t_n + a_2\tau, y_n + b_2\tau f(t_n, y_n))$$

Сеточную функцию:

$$\psi_n = -\frac{u_{n+1}-u_n}{\tau} + \sigma_1 f(t_n, u_n) + \sigma_2 f(t_n + a_2\tau, u_n + b_{21}\tau f(t_n, u_n)) \tag{5.6}$$

5.6 называется погрешностью аппроксимации. Получим оценку для ψ_n.

Расскладываем в окрестности (t_n, u_n):

$$\frac{u_{n+1}-u_n}{\tau} = u'_n + \frac{\tau}{2}u''_n + \underline{O}(\tau^2)$$

$$f(t_n + a_2\tau, y_n + b_{21}\tau f(t_n, u_n)) = f(t_n, u_n) + \frac{\partial f_n}{\partial t}a_2\tau + \frac{\partial f_n}{\partial u}b_{21}\tau f(t_n, u_n) + \underline{O}(\tau_2)$$

Перепишем теперь ψ_n учитывая эти представления, сгруппировав в итоге слагаемые, удобным для оценки погрешности образом:

$$\psi_n = -\left(u'_n + 0.5\tau\left(\frac{\partial f_n}{\partial t} + \frac{\partial f_n}{\partial u}f_n\right)\right) + \sigma_1 f(t_n, u_n) + \sigma_2 f(t_n, u_n) +$$

$$+\sigma_2\frac{\partial f_n}{\partial t}a_2\tau + \sigma_2\frac{\partial f_n}{\partial u}b_{21}\tau f(t_n, u_n) + \underline{O}(\tau^2) =$$

$$= (\text{ так как } u''_n = \frac{d}{dt}(f(t, u_n(t))) = \frac{\partial f_n}{\partial t} + \frac{\partial f_n}{\partial u}f_n) =$$

$$= -u'_n + (\sigma_1 + \sigma_2)f(t_n, u_n) +$$

$$+\tau\left((\sigma_2 a_2 - 0.5)\frac{\partial f_n}{\partial t} + ((\sigma_2 b_{21} - 0.5))\frac{\partial f_n}{\partial u}\tau f(t_n, u_n)\right) + \underline{O}(\tau^2)$$

Потребуем выполнений следующих условий:

1. $\sigma_1 + \sigma_2 = 1$ (обеспечивает наличие погрешности аппроксимации)

2. $\sigma_2 a_2 = \sigma_2 b_{21} = 0.5$ (для достижения второго порядка аппроксимации)

Полагают обычно $\sigma_2 = \sigma, \sigma_1 = 1 - \sigma$ Получается семейство параметрических схем:

$$\frac{y_{n+1} - y_n}{\tau} = (1 - \sigma)K_1 + \sigma K_2$$

В приведенном выше примере 2: $\sigma = 1, a_2 = 0.5, b_{21} = 0.5$, схема имеет второй порядок аппроксимации.

Если положить $\sigma = 0.5, a_2 = b_{21} = 1$, то получим так называемую симметричную разностную схему.

$$\frac{y_{n+1} - y_n}{\tau} = 0.5(f(t_n, y_n) + f(t_{n+1}, y_{n+1}))$$

Эта схема также имеет 2-й порядок погрешности аппроксимации (позднее докажем, что она имеет и 2-й порядок точности)

§ 2. Общая схема Рунге-Кутта

По своей идее многоэтапный метод означает, что переходя от t_n к t_{n+1}, мы используем несколько (m) промежуточных этапов (и, соответственно, множество параметров, влияющих на погрешность аппроксимации).

$$\frac{y_{n+1} - y_n}{\tau} = \sigma_1 K_1 + \sigma_2 K_2 + \cdots + \sigma_m K_m$$

$$y_0 = u_0, n = 0, 1, 2, \ldots$$

$$K_1 = f(t_n, y_n)$$

$$K_2 = f(t_n + a_2\tau, y_n + b_{21}\tau K_1)$$

$$K_3 = f(t_n + a_3\tau, y_n + b_{31}\tau K_1 + b_{32}\tau K_2)$$

$$\ldots$$

$$K_m = f(t_n + a_m\tau, y_n + b_{m1}\tau K_1 + b_{m2}\tau K_2 + \cdots + b_{mm-1}\tau K_{m-1})$$

При этом должно быть выполнено условие

$$\sum_{i=1}^{m} \sigma_i = 1 \text{ — условие аппроксимации}$$

При больших m схемы Рунге-Кутта на практике обычно не применяются, в силу громоздкости этих методов. Используют схемы третьего и четвертого порядка. Но есть общее утверждение о том, что ***метод имеет ту же точность, что и погрешность аппроксимации***.

Приведем примеры для третьего и четвертого порядка:

Пример. Схема Рунге-Кутта третьего порядка.

$$\frac{y_{n+1} - y_n}{\tau} = \frac{1}{6}(K_1 + 4K_2 + K_3)$$

$$K_1 = f(t_n, y_n)$$

$$K_2 = f(t_n + 0.5\tau, y_n + 0.5\tau K_1)$$

$$K_3 = f(t_n + \tau, y_n - \tau K_1 - 2\tau K_2)$$

Пример. Схема Рунге-Кутта четвертого порядка.

$$\frac{y_{n+1} - y_n}{\tau} = \frac{1}{6}(K_1 + 2K_2 + 2K_3 + K_4)$$

$$K_1 = f(t_n, y_n)$$

$$K_2 = f(t_n + 0.5\tau, y_n + 0.5\tau K_1)$$

$$K_3 = f(t_n + 0.5\tau, y_n + 0.5\tau K_2)$$

$$K_4 = f(t_n + \tau, y_n + \tau K_3)$$

Мы видим, что здесь получить схему высокого порядка погрешности аппроксимации требует больших вычислений в правых частях. Этот фактор ставит метод

Рунге-Кутта ниже многошаговых методов.

Оценка точности на примере двухэтапного метода Рунге-Кутты:

Вся сложность порождена нелинейностью задачи

$$\frac{dU}{dt} = f(t, U(t)), \quad t > 0 \tag{5.7}$$

$$U(0) = U_0$$

В прошлый раз мы показали, что если:

$\sigma a_2 = \sigma b_2 = \frac{1}{2}$, то тогда разностная схема имеет второй порядок погрешности аппроксимации.

$$\frac{y_{n+1} - y_n}{\tau} = (1 - \sigma) f_n + \sigma f(t_n + a\tau, y_n + a\tau f_n) \tag{5.8}$$

$$y_0 = U_0; \quad n = 0, 1, ...$$

Положим для определенности $0 \leq \sigma \leq 1, \quad a_2 = b_2 = a$

Как обычно вводим погрешность $z_n = y_n - U_n; \quad U_n = U(t_n)$

Для погрешности получаем следующую задачу:

$$z_n = y_n - U_n$$

$$\frac{z_{n+1} - z_n}{\tau} = -\frac{U_{n-1} - U_n}{\tau} + (1 - \sigma) f(t_n, y_n) + \sigma f(t_n + a\tau, y_n + a\tau f(t_n, y_n)) \tag{5.9}$$

Перепишем 5.9 в эквивалентном виде, но сформировав погрешность аппроксимации на решении:

$$\frac{z_{n+1} - z_n}{\tau} = -\frac{U_{n-1} - U_n}{\tau} + (1 - \sigma) f(t_n, U_n) +$$

$$+\sigma f(t_n + a\tau, U_n + a\tau f(t_n, U_n)) + (1-\sigma)(f(t_n, y_n) - f(t_n, y_n)) +$$

$$+\sigma(f(t_n + a\tau, y_n + a\tau f(t_n, y_n)) - f(t_n + a\tau, U_n + a\tau f(t_n, U_n)))$$

С такой правой частью работать сложно.

Обозначим $f(t_n, y_n) - f(t_n, U_n) = \phi_n^{(1)}$
Обозначим $f(t_n + a\tau, y_n + a\tau f(t_n, y_n) - f(t_n + a\tau, U_n + a\tau f(t_n, U_n)) = \phi_n^{(2)}$

Тогда

$$\frac{z_{n+1} - z_n}{\tau} = \psi_n + \phi_n^{(1)} + \phi_n^{(2)}$$

Эти три слагаемых проще оценить последовательно.

Делаем предположения, что выполнено усл. Липшица по второму аргументу.

$$|f(t,u) - f(t,v)| \le L|u - v|,$$

Тогда

$$|\phi_n^{(1)}| = |f(t_n, y_n) - f(t_n, U_n)| \le L|y_n - U_n| = L|z_n|$$

$$|\phi_n^{(2)}| = |f(t_n + a\tau, y_n + a\tau f(t_n, y_n)) - f(t_n + a\tau, U_n + a\tau f(t_n, U_n))| \le$$

$$\le L|y_n + a\tau f(t_n, y_n) - U_n - a\tau f(t_n, U_n)| \le$$

$$\le L(|y_n - U_n| + a\tau|f(t_n, y_n) - f(t_n, y_n)|)$$

$$\le L(|z_n| + a\tau L|z_n|)$$

Тогда:

$$z_{n+1} = z_n + \tau\psi_n + (1-\sigma)\tau\phi_n^{(1)} + \sigma\tau\phi_n^{(2)};$$

$$|z_{n+1}| \le \tau|\psi_n| + (1-\sigma)\tau L|z_n| + \sigma\tau L(|z_n| + a\tau L|z_n|) =$$

$$= \tau L|z_n| + \tau L(\sigma + \sigma a \tau L)|z_n| + \tau|\psi_n| + |z_n|$$

Пусть $a < \frac{1}{2}$, тогда имеем

$$|z_{n+1}| \le \tau|\psi_n| + (1 + \tau L + 0.5\tau^2 L^2)|z_n|$$

$$|z_{n+1}| \le e^{\tau L}|z_n| + \tau|\phi_n|; \quad e^{\tau L} = \rho > 0$$

$$|z_{n+1}| \le \rho|z_n| + \tau|\psi_n|$$

- применим это n раз реккурентно

$$|z_{n+1}| \le \rho^n|z_0| + \sum_{j=0}^{n} \tau|\psi_j|; \quad \text{т.к.}|z_0| = 0$$

$$|z_{n+1}| \le \sum_{j=0}^{n} \tau|\psi_j| = \max_{0\le j\le n}|\psi_j| \sum_{j=0}^{n} \tau\rho^{n-j} =$$

$$= t_{n+1} \max_{0\le j\le n}|\psi_j|\rho^n = t_{n+1}e^{\tau n L}\|\psi\|_c$$

Окончательно приходим к оценке $|z_{n+1}| \le M\|\psi\|_c$,
где M - положительная константа, не зависящая от τ

Ясно, что $\lim\limits_{n\to\infty}|z_n| \to 0$
Значит имеет место сходимость метода.

Напомним, что схема "предиктор-корректор":
1) $\sigma = 1; \qquad a = 0.5; \qquad \psi_n = O(\tau^2); \qquad |z_{n+1}| \le M\tau^2$
Симметричная разностная схема:
2) $\sigma = 0.5; \qquad a = 1; \qquad \psi_n = O(\tau^2); \qquad |z_{n+1}| \le M\tau^2$

Если $\sigma = 0, \quad a$ - любое: тогда $\psi_n = O(\tau); \qquad |z_{n+1}| \le M\tau$
Это схема Эйлера.

§ 3. Многошаговые разностные методы

В таких методах для вычислений на t_n-ом шаге используются данные с предыдущих m шагов.

Определение:
Линейным m-шаговым разностным методом решения задачи (1) называется метод, записанный уравнением:

$$\sum_{k=0}^{m} \frac{a_k}{\tau} y_{n-k} = \sum_{k=0}^{m} b_k f_{n-k} \tag{5.10}$$

$$y_{n-k} = y(t_n - k\tau)$$

$$f_{n-k} = f(t_n - k\tau, y_{n-k})$$

Здесь $a_k, b_k -$ действительные числа $\quad k = 0, 1, ..., m$, причем:
$a_0 \neq 0$
$b_m \neq 0$
$\tau > 0$

Если $b_0 = 0$, метод называется явным, если $b \neq 0$ - неявным.

$y_0, y_1, ..., y_{m-1}$ - т.н. "разгонный этап".
На этом этапе эти значения обычно получаются другим методом - например методом Рунге-Кутты. Будем считать, что они уже заданы.

Попробуем сравнить этот метод с методом Рунге-Кутты:

Недостатки:
- наличие разгонного этапа
- нужно помнить на y_n шаге - шаги $y_{n-1}, ..., y_{n-m}$
Достоинства:
- компактная, простая формула
- легко можно получить более высокий порядок аппроксимации
(Позже отметим, что высший порядок $p = 2m$)

Условие нормировки:

$$\sum_{k=0}^{m} b_k = 1 \tag{5.11}$$

Переходим к вычислению оценки погрешности аппроксимации.

Предполагаем, что исходная функция обладает нужной гладкостью.

$$\psi_n = -\sum_{k=0}^{m} \frac{a_k}{\tau} U_{n-k} + \sum_{k=0}^{m} b_k f(t_n - k\tau, U_{n-k}) \tag{5.12}$$

Оценим ψ_n.
Напомним, что $U_{n-k} = U(t_n - k\tau)$
Разложим в ряд Тейлора в окрестности t_n:

$$U_{n-k} = \sum_{l=0}^{p} \frac{(-k\tau)^l}{l!} U^{(l)}(t_n) + O(\tau^{p+1})$$

Теперь разложим правую часть:

$$f(t_n - k\tau, U_{n-k}) = U'_{n-k} = \sum_{l=0}^{p-1} \frac{(-k\tau)^l}{l!} U^{(l+1)}(t_n) + O(\tau^p)$$

Подставляем в формулу для ψ_n:

$$\psi_n = -\sum_{lk=0}^{m} \frac{a_k}{\tau} \sum_{l=0}^{p} \frac{(-k\tau)^l}{l!} U^{(l)}(t_n) + \sum_{k=0}^{m} b_k \sum_{l=0}^{p-1} \frac{(-k\tau)^l}{l!} U(t_n) + O(\tau^p)$$

Смещаем индексы: $l + 1 = l'$

$$\psi_n = -\sum_{l=0}^{p} \sum_{k=0}^{m} \frac{a_k}{\tau} \frac{(-k\tau)^l}{l!} U^{(l)}(t_n) +$$

$$+ \sum_{l=1}^{p} \sum_{k=0}^{m} b_k \frac{l(-k\tau)^{l-1}}{l(l-1)!} U^{(l)}(t_n) + O(\tau^p) =$$

$$= -\sum_{k=0}^{m} \frac{a_k}{\tau} U(t_n) + \sum_{l=1}^{p} (-\sum_{k=0}^{m} \frac{a_k}{\tau} \frac{(-k\tau)^l}{l!} U^{(l)} +$$

$$+ \sum_{k=0}^{m} l b_k \frac{(-k\tau)^{l-1}}{l!} U^{(l)}(t_n)) + O(\tau^p) =$$

$$= \{ \sum_{k=0}^{m} a_k = 0 \text{ - условие аппроксимации } \} = \tag{5.13}$$

$$= -\sum_{k=0}^{m} \frac{a_k}{\tau} U(t_n) + \sum_{l=1}^{p} (-\sum_{k=0}^{m} \frac{(-k\tau)^{l-1}}{l!} U^{(l)}(t_n) \cdot (ka_k + lb_k)) + O(\tau^p)$$

$$\sum_{k=0}^{m} k^{l-1}(ka_k + lb_k) = 0; \qquad l = 1, 2, ..., p \tag{5.14}$$

$$a_0, ..., a_m$$

$$b_0, ..., b_m$$

Значит у нас $2m + 2$ уравнения из этой системы и ещё $p + 2$ уравнений.
Чтобы система разрешалась, надо чтобы $p + 2 \leq 2m + 2$
То есть, надо чтобы выполнялось: $p \leq 2m$
Значит метод может иметь наивысший порядок - $2m$ для данного m.

Напомним, что

$$\frac{z_{n+1} - z_n}{\tau} = \psi_n + \phi_n^{(1)} + \phi_n^{(2)}$$

ϕ_n - пограничная аппроксимация на решении.
$\phi_n^{(1)} = (1 - \sigma)(f(t_n, y_n) - f(t_n, U_n))$
$\phi_n^{(2)} = \sigma[(f(t_n + a\tau, y_n + a\tau f(t_n, y_n)) - f(t_n + a\tau, U_n + a\tau f(t_n, U_n))]$

Полагая, что функция $f(t, u)$ удовлетворяет условию Липшица с константой L по 2-ому аргументу, получим:
$|\phi_n^{(1)}| \leq (1 - \sigma)L|y_n - U_n| = (1 - \sigma)L|z_n|$ - по Липшицу

Пусть $0 \leq \sigma \leq 1; \qquad a \geq 0$

$$|\phi_n^{(2)}| \leq \sigma L|y_n + a\tau f(t_n, y_n) - U_n - a\tau f(t_n, U_n)| \leq$$

$$\leq \sigma L(|y_n - U_n| + a\tau L|y_n - U_n|) = \sigma L(1 + a\tau L)|z_n|$$

$$\phi_n^{(1)}| + |\phi_n^{(2)}| \leq L|z_n| - \sigma L|z_n| + \sigma L|z_n| + \sigma a\tau L^2|z_n| = L|z_n| + \sigma a\tau L^2|z^n|$$

Здесь делаем допущение, что $\sigma a \leq 0.5$
Тогда

$$|\phi_n^{(1)}| + |\phi_n^{(2)}| \le (L + 0.5\tau L^2)|z_n|$$

А теперь, т.к. $z_{n+1} = z_n + \tau\psi_n + \tau(\phi_n^{(1)} + \phi_n^{(2)})$, - получаем оценку:

$$|z_{n+1}| \le (1 + \tau L + 0.5\tau^2 L^2)|z_n| + \tau|\psi_n|$$

Причем можно отметить, что $(1 + \tau L + 0.5\tau^2 L^2) \le e^{\tau L} = \rho$

Тогда получена система

$$\sum_{k=0}^{m} a_k = 0 \tag{5.15}$$

$$\sum_{k=0}^{m} k^{l-1}(ka_k + lb_k) = 0 \qquad l = 1, ..., p \tag{5.16}$$

$$\sum_{k=0}^{m} b_k = 1 \tag{5.17}$$

$$\sum_{k=0}^{m} ka_k = -\sum_{k=0}^{m} b_k = -1 \tag{5.18}$$

$$\sum_{k=0}^{m} k^{l-1}(ka_k + lb_k) = 0 \qquad l = 2, ..., p$$

$$b_0 = 1 - \sum_{k=1}^{m} b_k; \qquad a_0 = -\sum_{k=1}^{m} a_k$$

Всего тогда имеется p уравнений.
В качестве неизвестных: $a_1, ..., a_m$, $b_1, ..., b_m$
То есть имеем $2m$ неизвестных

Для того, чтобы система не была переопределенной, получаем $p \le 2m$

$$\sum_{k=0}^{m} \frac{a_k}{\tau} y_{n-k} = \sum_{k=0}^{m} b_k f_{n-k}$$

Если $b_0 = 0$, то можно при $k = 0$ выделить слагаемое:

$$\frac{a_k}{\tau}y_n = \sum_{k=1}^{m} b_k f_{n-k} - \frac{a_k}{\tau}y_{n-k})$$

Здесь нам известны компоненты правой части. Тогда явный разностный метод вычисляется просто через правую часть.

Если $b_0 \neq 0$, тогда метод - неявный, т.к. справа есть y_n;
Тогда: вновь выделяем $\frac{a_k}{\tau}y_n - b_0 f_n$; здесь $f_n = f(t_n, y_n)$ - неизвестно.
Остальное - в правую часть:

$$\frac{a_k}{\tau}y_n - b_0 f_n = F(y_{n-1}, y_{n-2}, ..., y_{n-m}) = \sum_{k=1}^{m} (b_k f_{n-k} - \frac{a_k}{\tau}y_{n-k}) \tag{5.19}$$

Получили неявное нелинейное уравнение. Такое уравнение решается методом Ньютона, в качестве начального приближения выбирают значения в предыдущие момент времени.

Метод Адамса:

$$\frac{y_n - y_{n-1}}{\tau} = \sum_{k=0}^{m} b_k f_{n-k}$$

$$a_0 = 1; \qquad a_1 = -1;$$

$$a_k = 0; \qquad k \geq 2$$

§ 4. Понятие устойчивости разностного метода

Уравнение:

$$y_{n+1} = qy_n \qquad n = 0, 1, ... \quad y_0 - \text{задано} \tag{5.20}$$

Покажем, что если $|q| > 1$, то процесс - неустойчив.

Очевидно, что на каждом шаге находим приближенные значения. Например пусть здесь влияет округление значений в процессе вычислений - $|\delta_n|$

$$\tilde{y}_n = y_n + \delta_n$$

Тогда $\tilde{y}_n = qy_n + q\delta_n = y_{n+1} + \delta_{n+1}$

$\delta_{n+1} = q\delta_n$, значит, если $|q| > 1$, то погрешность вычислений δ_{n+1} - возрастает неограниченно. Т.е. нет устойчивости.

При $|q| \leq 1$, δ_{n+1} - не возрастает, и метод устойчив.
Фактически будет выполняться оценка $|y_{n+1} \leq |y_n|$

Рассмотрим все вышесказанное на модельной задаче:

$$U'(t) + \lambda U(t) = 0; \qquad t > 0; \tag{5.21}$$

$$U(0) = U_0; \qquad \lambda \in \mathbb{R} \quad \lambda > 0$$

Аналитически посчитаем:

$U(t) = U_0 e^{-\lambda t}$
При $\lambda > 0$ - быстро убывает экспонента, всё в порядке.

Решаем задачу численно:

Явная схема Эйлера:

$$\frac{dU}{dt} = f(t, U(t)) \qquad t > 0 \tag{5.22}$$

Схема Эйлера для (2) выглядит так:

$$\frac{y_n - y_{n-1}}{\tau} = f(t_n, y_n)$$

$$\frac{y_n - y_{n-1}}{\tau} + \lambda y_{n-1} = 0 \qquad y_0 = U_0$$

$$y_{n+1} = (1 - \tau\lambda) y_n; \qquad \text{появилось } q = 1 - \tau\lambda$$

Если $|q| \leq 1$ - тогда метод устойчив.

$$-1 \leq 1 - \tau\lambda \leq 1; \qquad \tau\lambda \leq 2; \qquad \Rightarrow \quad 0 \leq \tau \leq \frac{2}{\lambda} \tag{5.23}$$

Если τ выйдет из этого интервала - устойчивости нет. 5.23 - условие устойчивости явной схем Эйлера.

Если взять λ большим - экспонента очень быстро убывает, но парадокс - надо брать очень маленький шаг.

Неявная схема Эйлера:

$$\frac{y_{n+1}-y_n}{\tau}=f(t_{n+1},y_{n+1})$$

$$\frac{y_{n+1}-y_n}{\tau}+\lambda y_{n+1}=0$$

$$y_{n+1}+\tau\lambda y_{n+1}=y_n$$

$$y_n=(1+\tau\lambda)y_{n+1}$$

$$y_{n+1}=\frac{1}{1+\tau\lambda}y_n$$

$$q=\frac{1}{1+\tau\lambda}\qquad 0<q<1;\quad |q|<1$$

Этот метод уже устойчив независимо от выбора шага τ

Общий m-шаговый разностный метод:

$$\frac{dU}{dt}=f(t,U(t))\qquad t>0 \tag{5.24}$$

$$U(0)=U_0$$

$$\frac{dU}{dt}+\lambda U(t)=0;\qquad t>0$$

$$U(0)=U_0\qquad -\text{задача} \tag{5.25}$$

$$\sum_{k=0}^{m}\frac{a_k}{\tau}y_{n-k}=\sum_{k=0}^{m}b_k f_{n-k} \tag{5.26}$$

$y_0, y_1, ..., y_m$ - считаем заданными (разгонный этап);

Подчеркнем, что a_k, b_k - не зависят от τ

$$\sum_{k=0}^{m}(\frac{a_k}{\tau}+b_k\lambda)y_{n-k}=0 \tag{5.27}$$

Перепишем:

$$\sum_{k=0}^{m}(a_k + \tau b_k \lambda) y_{n-k} = 0 \tag{5.28}$$

Решение уравнения 5 ищется в виде:

$$y_j = q^j$$

Сокращаем на q^{n-m}

$$F(q,\tau) = \sum_{k=0}^{m}(a_k + \tau\lambda b_k) q^{m-k} = 0 \tag{5.29}$$

Уравнение 5.28 - характеристическое уравнение для разностных схем

Исследуем $|q| < 1$ и $|q| \geq 1$;
Аналитически - практически невозможно.
Но τ - мало, положим его формально равным нулю: $\tau = 0$
Получим упрощенное характеристическое уравнение:

$$\sum_{k=0}^{m} a_k q^{m-k} = 0 \tag{5.30}$$

Вводя и исследуя устойчивость - математики используют именно это уравнение. Здесь нет правой части, мы судим об устойчивости по производной.

Определение:
Говорят, что разностная схема удовлетворяет условнию (α), если все корни хар-кого уравнения лежат внутри либо на границе единичного круга комплексной плоскости, причем на границе нет кратных корней.

С точки зрения устойчивости это позволяет сформулировать следующую теорему:

Теорема (без док-ва):
Пусть разностная схема удовлетворяет условию (α);
Пусть $|f_n| \leq L \qquad 0 \leq t \leq T; \quad t$ - конечное.
Тогда, для $t_n = n\tau : \qquad 0 \leq t_n \leq T$
и всех достаточно малых τ - выполняется оценка:

$$|z_n| = |y(t_n) - U(t_n)| \leq M(\sum_{j=m}^{n} \tau|\psi_j| + \max_{0\leq i\leq m-1} |y_i - U(t_i)|)$$

M - не зависит от τ

M - функция от LT : $\quad M = M(LT)$

Замечание 1:
Методы Адамса удовлетворяют условию (α);

$$\frac{y_{n+1} - y_n}{\tau} = \sum_{k=0}^{m} b_k f_{n-k}; \qquad y_0 = U_0$$

Сведем к виду : получаем $q = 1$;

Замечание 2:
Это замечание касается понятий абсолютной и условной устойчивости - мы не делаем разницы между ними.

Замечание 3:
Пусть m - нечетное. Тогда наивысший порядок устойчивого разностного метода - $(m + 1)$;

Пусть m - четное. Тогда наивысший порядок устойчивого разностного метода - $(m + 2)$
Если метод явный, тогда $p = m$;

Пример:
Рассмотрим разностную схему:

$$\frac{y_n + 4y_{n+1} - 5y_{n-2}}{6\tau} = \frac{2f_{n-1} + f_{n-2}}{3} \tag{5.31}$$

- явная схема.

Задача 1:
Показать, что для 5.31 - погрешность аппроксимации - третьего порядка $O(\tau^3)$
Решение:
Обозначим

$$\psi_n = -\frac{u_n + 4u_{n-1} - 5u_{n-2}}{6\tau} + \frac{2f_{n-1} + f_{n-2}}{3}$$

Условия, которым должен удовлетворять многошаговый разностный метод, для того, чтобы погрешность аппроксимации имела порядок $O(n^3)$:

$$b_0 = 1 - \sum_{k=1}^{m} b_k$$

$$a_0 = -\sum_{k=1}^{m} a_k$$

$$\sum_{k=1}^{m} a_k k = -1$$

$$\sum_{k=1}^{m} k^{l-1}(a_k k b_k) = 0,\ l = 2, 3$$

При $m = 2,\ a_0 = \frac{1}{6},\ a_1 = \frac{2}{3},\ a_2 = -\frac{5}{6},\ b_0 = 0,\ b_1 = \frac{2}{3},\ b_2 = -\frac{1}{3}$ условия выполняются и справедлива оценка $O(n^3)$.

Покажем, что схема не удовлетворяет условию (α).
Получаем хар-кое уравнение:

$$q^2 + 4q - 5 = 0$$

$$q_1 = 1; \quad q_2 = -5$$

Т.к. $|q_2| = 5 > 1$, значит - неустойчивый метод.

§ 5. Жёсткие системы ОДУ

Круг задач, связанных с жёсткими системами, широкий, и в современной практике идет процесс изучения этих численных алгоритмов. Но есть и завершенные результаты, о которых и поговорим в этом параграфе.

Для того чтобы понять смысл жёстких систем, начнем рассмотрение модельной задачи, которая будеть казаться довольно искусственной, но даёт возможность понимания жёсткости систем ОДУ.

Рассмотрим систему из двух независимых функций (вообще говоря):

$$\begin{cases} \frac{du_1}{dt} + a_1 u_1(t) = 0, \quad t > 0, \\ u_1(0) = u_{10}, \quad a_1 > 0 \\ \frac{du_2}{dt} + a_2 u_2(t) = 0, \quad t > 0, \\ u_2(0) = u_{20}, \quad a_2 > 0. \end{cases} \tag{5.32}$$

(условия 1 и 2), при этом $a_1 >> a_2$ (a_1 много больше a_2, обычно — на несколько порядков).

Каждая из компонент убывает, устойчивость есть по каждому условию, а значит, существует и решение:

$$\overline{u}(t) = (u_1(t), u_2(t))^T$$
$$u_1(t) = u_{10}e^{-a_1 t},$$
$$u_2(t) = u_{20}e^{-a_2 t}.$$

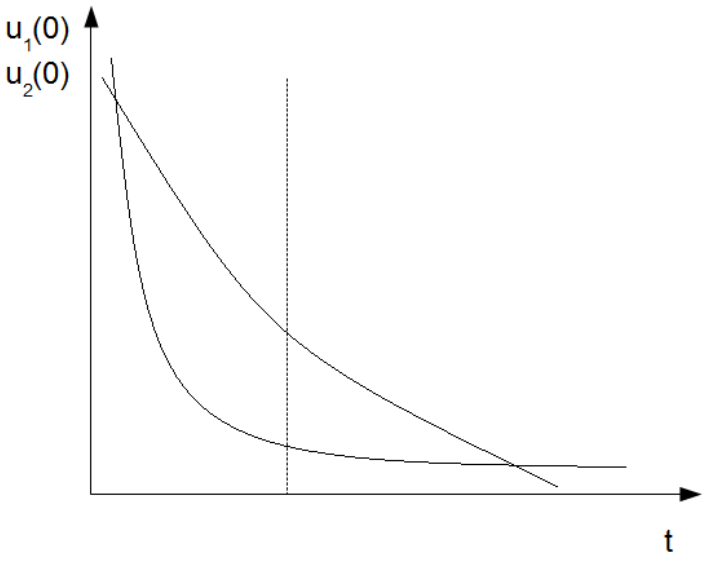

Рис. 5.1: $u_2(0)$ быстро убывает, и при некотором t^* станет практически равным нулю.

Если будем решать задачу численно, то результат будет зависеть от того, какую схему выберем.

- Явная схема Эйлера.

$$\begin{cases} \frac{y_1^{n+1}-y_1^n}{\tau} + a_1 y_1^n = 0, \\ \\ \frac{y_2^{n+1}-y_2^n}{\tau} + a_2 y_2^n = 0, \end{cases} \tag{5.33}$$

(условия 3 и 4)

На прошлой лекции показали, что устойчивость схемы 5.22 будет при условии $0 < \tau < \frac{2}{a_1}$. А для схемы 5.23 — $0 < \tau < \frac{2}{a_2}$

Чтобы обеспечить устойчивость системы, мы должны выбрать минимальный шаг, т.е. второй вариант условия.
Таким образом, спустя некоторый момент времени решение полностью определяется первой компонентой, но при этом шаг счета лимитируется второй компонентой.

Мы могли бы идти с большим шагом, но для устойчивости нам приходится идти с очень малым шагом.

Поэтому явные схемы оказываются непригодными в данном случае. Выход - в применении неявных схем.

- Неявная схема Эйлера.

$$\begin{cases} \frac{y_1^{n+1}-y_1^n}{\tau} + a_1 y_1^{n+1} = 0, \\ \\ \frac{y_2^{n+1}-y_2^n}{\tau} + a_2 y_2^{n+1} = 0, \end{cases} \tag{5.34}$$

Схемы эти абсолютно устойчивы, шаг регламентирован только условиями точности, но никак не устойчивости.

Несмотря на искусственность данного примера, в жестких системах такая же картина. В системе из n уравнений есть быстроубывающие компоненты и медленно убывающие. Это будет связано со спектром матрицы системы, действительные части собственных значений которой будут сильно отличаться.

Начнем с линейных систем:

$$\frac{d\overline{u}}{dt} + A\overline{u}(t) = 0, \quad t > 0 \tag{5.35}$$

$A(m \times m)$ с постоянными числами (не зависит от времени), $\overline{u}(0) = \overline{u}_0$. (система 5)

Определение:
Система линейных уравнений (5) называется жесткой, если:
1)

$$Re\,\lambda_k^A > 0, k = \overline{1,m}$$

2)

$$s = \frac{max_{1 \le k \le m} |Re\,\lambda_k^A|}{min_{1 \le k \le m} |Re\,\lambda_k^A|} >> 1$$

(s – число жесткости).

Примечание:

Если различие в условии 2 - на 2-3 порядка, то уже систему называют жесткой. Первое условие означает устойчивость по Ляпунову. А второе — чило жесткости — должно быть большим. Это как раз говорит о разбросе собственных значений (их действительных частей).

Рассмотрим теперь нелинейную задачу Коши.

$$\frac{d\overline{u}}{dt} = f(t, \overline{u}(t)), t > 0 \tag{5.36}$$

$$\overline{u}(0) = \overline{u}_0$$

$$\overline{u}(t) = (u_1(t), u_2(t), \ldots, u_m(t))^T$$

.

$$\overline{f}(t, \overline{u}(t)) = (f_1(t, \overline{u}(t)), f_2(t, \overline{u}(t)), \ldots, f_m(t, \overline{u}(t)))^T$$

(система 6)

Большинство методов, которые мы использовали при решении одного уравнения, сюда легко переносятся. Но возникают особенности, связанные с тем, что компоненты вектора u могут вести себя по-разному (быстро и медленно убывать), что усложняет численное интегрирование.

Введем понятие жесткости. При определении исходят из *линеаризованной системы.*

Проведем процесс линеаризации в окрестности некоторого известного решения.

Пусть $v(t)$ — некоторое известное решение задачи. Рассмотрим вектор

$$\overline{z}(t) = \overline{u}(t) - \overline{v}(t)$$

Правую часть (в предположении нужной гладкости) раскладываем по формуле Тейлора:

$$\frac{d\overline{z_k}}{dt} = f_k(t, \overline{v}(t) + \overline{z}(t)) - f_k(t, \overline{v}(t)), k = \overline{1, m}$$

$$\frac{d\overline{z_k}}{dt} = f_k(t, \overline{v}(t)) + \frac{\partial f_k}{\partial u_1}(t, \overline{v}(t))z_1(t) + \ldots + \frac{\partial f_k}{\partial u_m}(t, \overline{v}(t))z_m(t) + o(|z|) - f_k(t, \overline{v}(t))$$

Таким образом получаем:

$$\frac{\partial \overline{z}}{\partial t} = J(t, \overline{v}(t))\overline{z} \tag{5.37}$$

(система 7)

$$J(t, \overline{v}(t))\overline{z} = a_{ij} = \frac{\partial f_i(t, \overline{v}(t))}{\partial u_j}, i, j = \overline{1, n}.$$

Система (7) называется системой первого приближения.

Введем понятие жесткости:

$$s = \frac{\max Re\lambda_k^J}{\min Re\lambda_k^J}$$

Определение:

Система (6) называется жесткой на решении $v(t)$ и моменте времени $0 \leq t \leq T$, если выполнены 2 условия:

1. $Re\lambda_k^J < 0$
2. $s(t) >> 1$

§ 6. Дальнейшее определение устойчивости и примеры разностных схем интегрирования жестких систем ДУ

Конечно, исследуя устойчивость жестких систем, мы можем исходить из нашего старого определения устойчивости. Но при интегрировании жестких систем вводят более узкие определения.

Поставим исходную задачу:

$$\frac{du}{dt} = f(t, u(t)), \quad t > 0, u(0) = u_0 \tag{5.38}$$

Линеаризуя, получаем

$$\frac{du}{dt} = \Lambda u(t), \quad t > 0 \tag{5.39}$$

$$\overline{u}(0) = \overline{u}_0$$

Λ – собственные значение матрицы первого приближения J:

$$\Lambda = \Lambda^J$$

Проведем аппроксимацию — будет возникать комплексный параметр:

$$\tau\lambda = \mu, \mu = \mu_0 + i\mu_1$$

Определение:

Областью устойчивости разностного метода для задачи 5.38 называется множество точек комплексной плоскости $\mu = \tau\lambda$, для которых данный метод, примененный к уравнению 5.39, устойчив.

Явная схема Эйлера:

$$\frac{y_{n+1} - y_n}{\tau} = f(t_n, y_n)$$

$$\frac{y_{n+1} - y_n}{\tau} = \lambda y_n$$

(применительно к задаче 1 и 2 соответственно).

$$y_{n+1} = y_n + \tau\lambda y_0 = (1 + \mu)y_n$$
$$|1 + \mu| \leq 1$$
$$|1 + \mu_0 + i\mu_1| \leq 1$$
$$(1 + \mu_0)^2 + \mu_1^2 \leq 1$$

Область устойчивости — внутренность круга с центром $(-1, 0)$.

Неявная схема Эйлера:

$$\frac{y_{n+1} - y_n}{\tau} = f(t_{n+1}, y_{n+1})$$

$$\frac{y_{n+1} - y_n}{\tau} - \lambda y_{n+1} = 0$$

(применительно к двум задачам соответственно)

$$y_{n+1} = y_n + \tau\lambda y_{n+1}$$

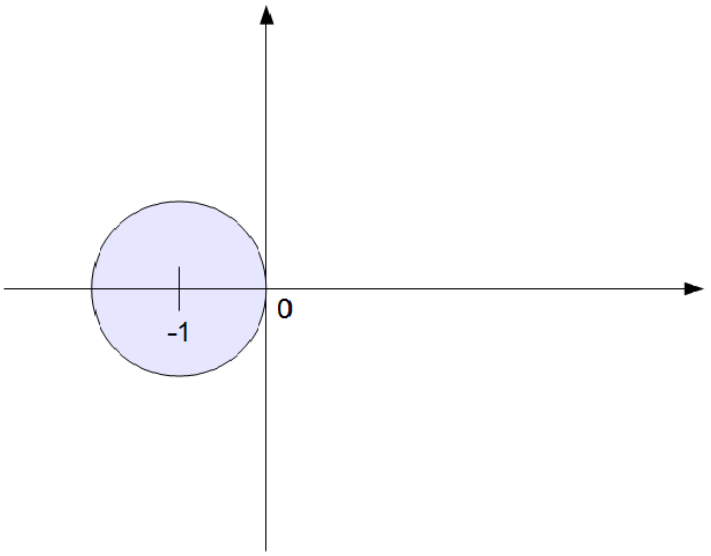

Рис. 5.2:

$$(1-\mu)y_{n+1} = y_n$$

$$y_{n+1} = \frac{1}{1-\mu}y_n$$

Для обеспечения устойчивости требуем:

$$|\frac{1}{1-\tau\lambda}| \leq 1$$

$$|1-\mu| \geq 1$$

$$(1-\mu_0)^2 + \mu_1^2 \geq 1$$

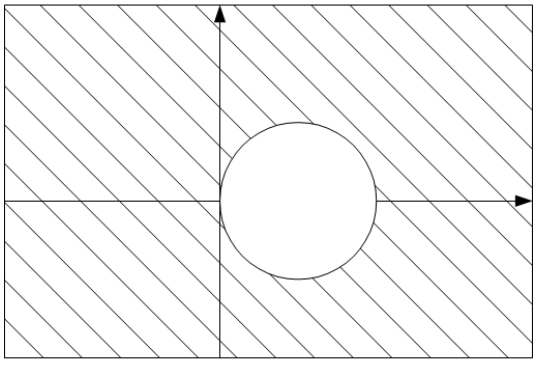

Рис. 5.3:

Определение:

Разностный метод A-устойчивый, если область его устойчивости содержит всю левую полуплоскость комплексной плоскости (т.е. $Re(\mu) < 0$). (рис. 5.4)

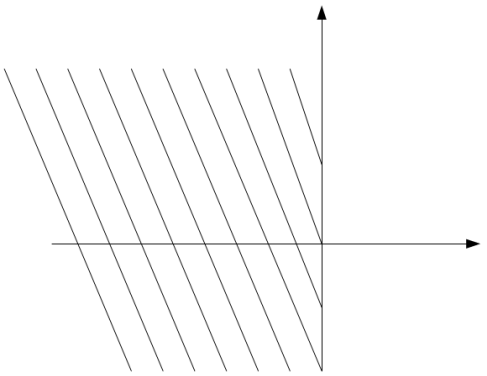

Рис. 5.4:

Таким образом, явная схема Эйлера не является -устойчивой, а неявная — A-устойчива.

Замечание: Если разностный метод A-устойчив, то он абсолютно устойчив, т.е. Устойчив при любых $\tau > 0$.

Оказывается, этим условием A-устойчивости обладает очень узкий набор схем.

Доказано, что явных A-устойчивых методов в природе не существует. Также доказано, что среди неявных существуют разностные методы не выше второго порядка.

В качестве примера, 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 задаче 5.39, разностная схема примет вид:

$$\frac{y_{n+1} - y_n}{\tau} = 0.5\lambda(y_n + y_{n+1})$$

$$(y_{n+1} - y_n) - 0.5\mu(y_n + y_{n+1}) = 0$$

$$(1 - 0.5\mu)y_{n+1} = (1 + 0.5\mu)y_n$$

$$y_{n+1} = qy_n, q = \frac{(1 + 0.5\mu)}{(1 - 0.5\mu)}$$

$$|q| \leq 1$$

$$|1 + 0.5\mu| \leq |1 - 0.5\mu|$$

$$(1 + 0.5\mu_0)^2 + \mu_1^2 \leq (1 - 0.5\mu_0) + \mu_1^2$$

$$1 + \mu_0 + 0.25\mu_0^2 \leq 1 - \mu_0 + 0.25\mu_0^2$$

$$\mu_0 \leq 0$$

.

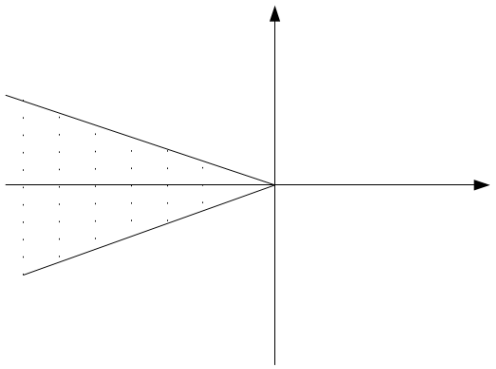

Рис. 5.5:

Таким образом, схема является A-устойчивой.

Коль круг оказался узкий, то было сделано смягчение в определении, и было введено понятие $A(\alpha)$-устойчивого метода:

Определение: Разностный метод называется $A(\alpha)$-устойчивым, если область его устойчивости содержит угол левой полуплоскости: $|\arg(-\mu)| < \alpha$

В частности, $(\frac{\pi}{2})$-устойчивый метод есть A-устойчивый.

Оказалось, что явных $A(\alpha)$-устойчивых методов не существует. А среди неявных были построены схемы третьего и даже четвертого порядка точности (чисто неявные, правая часть берется только на T_n, на последнем временном моменте).

В заключение, приведем пример схемы 4 порядка:

$$\frac{25y_{n+4} - 48y_{n+3} + 36y_{n+2} - 16y_{n+1} + 3y_n}{12\tau} = f(t_{n+4}, y_{n+4})$$

Такая схема имеет четвертый порядок и при некотором $\alpha > 0$ является $A(\alpha)$-устойчивой.

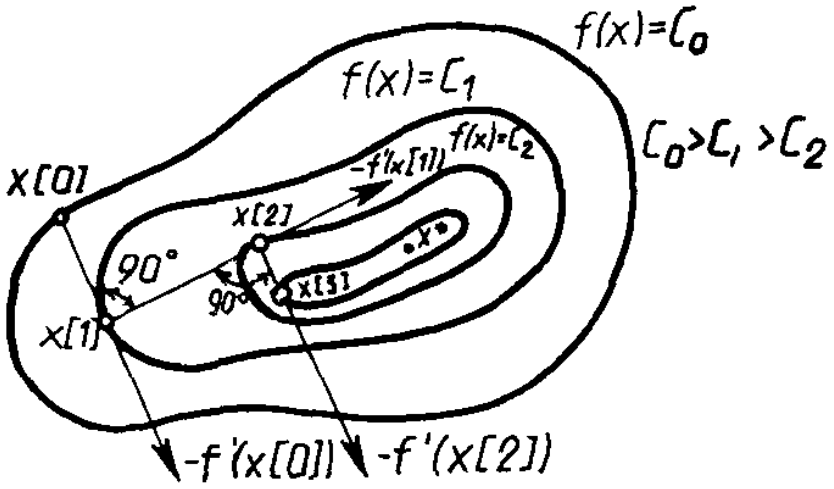

Рис. 5.6:

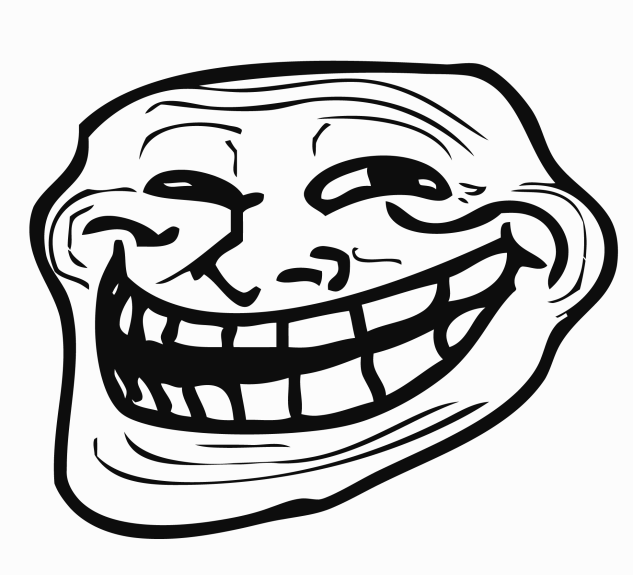

Рис. 5.7: