

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ  
РОССИЙСКОЙ ФЕДЕРАЦИИ

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ  
МИРЭА

---

А.Б. САМОХИН, В.В. ЧЕРДЫНЦЕВ, А.А. ВОРОНЦОВ

# ЧИСЛЕННЫЕ МЕТОДЫ

УЧЕБНОЕ ПОСОБИЕ

для студентов,  
обучающихся по направлениям подготовки  
*«Информатика и вычислительная техника»*,  
*«Информационные системы и технологии»*,  
*«Программная инженерия»*, *«Прикладная математика»*

УДК 519.6(075)  
ББК 221.193я73  
С 17

*Утверждено редакционно-издательским советом МГТУ МИРЭА  
в качестве методического пособия для студентов*

Подготовлено на кафедре прикладной математики

Рецензенты: д.ф.-м.н., профессор ВИС НИВЦ МГУ А.В. Сетуха;  
д.ф.-м.н., профессор РОСНОУ А.С. Крюковский

**Самохин А.Б.**

С 17 Численные методы : учебное пособие / А.Б. Самохин,  
В.В. Чердынцев, А.А. Воронцов. — М.: МГТУ МИРЭА,  
2014. — 84 с.

ISBN 978-5-7339-1075-8

Рассмотрены численные методы решения прикладных математических задач. Учебное пособие написано для студентов факультета информационных технологий. Оно может быть полезным также при изучении дисциплин «Вычислительная математика», «Математическое моделирование», «Методы оптимизации» и др.

ISBN 978-5-7339-1075-8

© Самохин А.Б., Чердынцев В.В.,  
Воронцов А.А., 2014

## ВВЕДЕНИЕ

Пятидесятилетняя эволюция ЭВМ от первых ламповых до современных серийных с быстродействием порядка  $10^{12}$  операций в секунду привела к развитию математического моделирования и численного анализа практически во всех отраслях человеческого знания. Развитие технических возможностей, математического и программного обеспечения ЭВМ показали несовершенство некоторых классических методов решения инженерных и научно-технических задач, что обусловило развитие новых методов их численного решения. Проблема выбора оптимального численного метода решения с точки зрения как экономии ресурсов ЭВМ, так и снижения результирующей погрешности требует определенного опыта и вычислительной практики.

Настоящее пособие является введением в численные методы. В конце каждой темы приведены задания для практических занятий, выполнение которых позволяет глубже понять и усвоить вычислительные алгоритмы. При их решении допустимо использование инженерных калькуляторов и применение математических пакетов прикладных программ.

### 1. АБСОЛЮТНАЯ И ОТНОСИТЕЛЬНАЯ ПОГРЕШНОСТИ

Численные методы служат для нахождения приближенного решения математических задач. Любое приближенное решение связано с ошибкой (погрешностью). Виды ошибок:

Погрешность математической модели, связанная с неполными знаниями о процессе.

1. Погрешность упрощения модели.
2. Погрешность, связанная с приближенным характером начальных данных.
3. Погрешность округления при расчетах.

Первые две погрешности относятся к систематическим, а две последние — к статистическим ошибкам. Для их оценки вводится абсолютная и относительная погрешности.

Абсолютная (предельная) погрешность — определяет интервал, в котором лежит точное значение величины. Пусть  $A$  — точное значение величины (неизвестно), а  $a$  — приближенное значение величины (известно). За абсолютную погрешность  $\Delta a$  принимается минимальное число  $\Delta a$ , удовлетворяющее условию

$$\Delta a \geq |A - a|. \quad (1.1)$$

При статистических измерениях погрешность  $\Delta a$  задается с определенной достоверностью, т.е. вероятность события  $|A - a| \leq \Delta a$  больше определенной величины  $\gamma$ :  $P(|A - a| \leq \Delta a) > \gamma \leq 1$ . Перепишем определение:  $a - \Delta a \leq A \leq a + \Delta a$ , то есть точное значение  $A$  лежит в заданном интервале. Для оценки качества измерений вводится относительная погрешность:

$$\delta a = \frac{\Delta a}{|A|} \approx \frac{\Delta a}{|a|}. \quad (1.2)$$

Заданные величины  $\Delta a$  или  $\delta a$  позволяют записать точное значение  $A$  в символическом виде:  $A = a(1 \pm \delta a)$  или  $A = a \pm \Delta a$ .

### 1.1. Число верных знаков приближенного числа

Приближенное число можно представить в виде:

$$a = a_m \cdot 10^m + a_{m-1} \cdot 10^{m-1} + \dots + a_{m-n+1} \cdot 10^{m-n+1}, \quad (1.1.1)$$

где  $m$  — величина старшего разряда;  $n$  — текущий номер знака, отсчитываемый слева направо. Говорят, что  $k$  первых знаков приближенного числа верные, если абсолютная погрешность  $\Delta a$  удовлетворяют условию:  $|A - a| \leq \Delta a \leq 0,5 \cdot 10^{m-k+1}$ , то есть меньше половины соответствующего разряда. Подбирается минимальное число вида  $0,5 \cdot 10^l$  большее, чем  $\Delta a$ , и сравниваются разряды.

## 1.2. Погрешность функций

Пусть дана функция  $f(a_1, a_2, \dots, a_n)$  от  $n$  приближенных значений  $a_i$ , погрешности которых известны. Требуется определить погрешность функции ( $\Delta f = ?$ ):

$$\Delta f = \max |f(a_1 + \Delta_1, a_2 + \Delta_2, \dots, a_n + \Delta_n) - f(a_1, \dots, a_n)|,$$

где  $\Delta_i$  — абсолютная погрешность приближенной величины  $a_i$ . Если  $\Delta_i \ll |a_i|$ , то разность, стоящую в формуле можно оценить в линейном приближении:

$$\begin{aligned} f(a_1 + \Delta_1, a_2 + \Delta_2, \dots, a_n + \Delta_n) - f(a_1, a_2, \dots, a_n) &\approx \\ &\approx \frac{\partial f}{\partial a_1} \Delta_1 + \frac{\partial f}{\partial a_2} \Delta_2 + \dots + \frac{\partial f}{\partial a_n} \Delta_n. \end{aligned}$$

Отсюда следует выражение для погрешности функции:

$$\Delta f = \sum_{i=1}^n \left| \frac{\partial f}{\partial a_i} \right| \Delta a_i, \quad \delta f = \frac{1}{|f|} \sum_{i=1}^n \left| \frac{\partial f}{\partial a_i} \right| \Delta a_i = \sum_{i=1}^n \left| \frac{\partial}{\partial a_i} \cdot \ln f \right| \Delta a_i. \quad (1.2.1)$$

## 1.3. Погрешность простейших функций двух переменных

Погрешность суммы:

$$f = a_1 + a_2, \quad \frac{\partial f}{\partial a_1} = \frac{\partial f}{\partial a_2} = 1, \quad \Delta f = \Delta a_1 + \Delta a_2, \quad \delta f = \frac{\Delta a_1 + \Delta a_2}{|a_1 + a_2|}.$$

Погрешность разности:

$$f = a_1 - a_2, \quad \frac{\partial f}{\partial a_2} = -1, \quad \Delta f = \Delta a_1 + \Delta a_2, \quad \delta f = \frac{\Delta a_1 + \Delta a_2}{|a_1 - a_2|}.$$

При  $a_1 \approx a_2$  качество измерений разности ухудшается.

*Замечание:* Абсолютная погрешность суммы и разности  $n$  приближенных величин равна сумме их абсолютных погрешностей.

Погрешность произведения:

$$f = a_1 \cdot a_2; \quad \frac{\partial f}{\partial a_1} = a_2; \quad \frac{\partial f}{\partial a_2} = a_1; \quad \Delta_f = |a_2| \cdot \Delta a_1 + |a_1| \cdot \Delta a_2;$$

$$\delta_f = \frac{\Delta f}{|a_1 \cdot a_2|} = \frac{\Delta a_1}{|a_1|} + \frac{\Delta a_2}{|a_2|} = \delta a_1 + \delta a_2.$$

То есть предпочтительней сначала найти относительную погрешность, а затем искать абсолютную:  $\Delta f = |a_1 \cdot a_2| \cdot \delta_f$ .

*Замечания:*

- Относительная погрешность степени есть произведение модуля показателя на относительную погрешность основания степени:  $f = a^v$ ,  $\delta_f = |v| \cdot \delta_a$ .

- Относительная погрешность произведения  $n$  сомножителей приближенных величин равна сумме относительных погрешностей сомножителей:

$$f = a_1 \cdot a_2 \cdot \dots \cdot a_n, \quad \delta_f = \sum_{i=1}^n \delta a_i.$$

Погрешность частного:

$$f = \frac{a_1}{a_2}; \quad \frac{\delta f}{\delta a_1} = \frac{1}{a_2}; \quad \frac{\delta f}{\delta a_2} = -\frac{a_1}{a_2^2};$$

$$\Delta f = \frac{\Delta a_1}{|f|} = \frac{\Delta a_1}{|a_1|} + \left| \frac{a_1}{a_2^2} \right| \cdot \Delta a_2; \quad \delta_f = \frac{\Delta f}{|f|} = \frac{\Delta a_1}{|a_1|} + \frac{\Delta a_2}{|a_2|} = \delta_1 + \delta_2.$$

Все замечания, сделанные для произведения справедливы и в этом случае.

#### 1.4. Примеры и задания

*Пример.* Дано приближенное число 3457,0 погрешность — 0,6. Найти число верных знаков. Цифра 3 входит в число с весом  $10^3$ , (1.3), то есть  $m=3$ ;  $0,6 < 0,5 \cdot 10^k$ , минимальное  $k = 1$ ,  $(3 - n + 1 = 1)$ , то есть верны три знака  $\underbrace{345}_{\text{верны}} 7,0$ .

*Пример.* Дан куб, сторона которого  $a = 100$  см измерена с точностью  $\Delta a = 1$  см. Определить погрешности измерения поверхности и объема куба:

$$S = 6a^2; V = a^3.$$

$$\Delta V = \frac{da^3}{da} \Delta a = 3a^2 \Delta a = 30000 \text{ см}^3, \quad \delta V = \frac{3a^2 \Delta a}{a^3} = 3 \frac{\Delta a}{a} = 3\delta a = 0,03;$$

$$S = \frac{d6a^2}{da} \Delta a = 12a \Delta a = 1200 \text{ см}^2; \quad \delta S = \frac{12a \Delta a}{6a^2} = 2 \frac{\Delta a}{a} = 2\delta a = 0,02.$$

*Пример.* Расчет погрешности функции трех переменных (1.2.1):

$$f = \frac{(a_1 + a_3^\alpha)}{(a_1 + a_2^\beta) \cdot a_3}; \quad \Delta f = ?; \quad \delta f = ?.$$

$$\frac{\partial f}{\partial a_1} = \frac{a_2^\beta - a_3^\alpha}{a_3(a_1 + a_2^\beta)^2}, \quad \frac{\partial f}{\partial a_2} = \frac{a_1 + a_3^\alpha}{a_3(a_1 + a_2^\beta)^2} \beta a_2^{\beta-1}, \quad \frac{\partial f}{\partial a_3} = \frac{(\alpha - 1)a_3^{\alpha-1} - a_1}{a_3^2(a_1 + a_2^\beta)}.$$

$$\Delta f = \left| \frac{a_2^\beta - a_3^\alpha}{a_3(a_1 + a_2^\beta)^2} \right| \Delta_1 + \left| \frac{a_1 + a_3^\alpha}{a_3(a_1 + a_2^\beta)^2} \beta a_2^{\beta-1} \right| \Delta_2 + \left| \frac{(\alpha - 1)a_3^{\alpha-1} - a_1}{a_3^2(a_1 + a_2^\beta)} \right| \Delta_3.$$

*Пример.* Катеты прямоугольного треугольника  $a = 30$  см и  $b = 40$  см измерены с погрешностью  $\Delta = 1$  см. Определить погрешность измерения гипотенузы  $c$ .

$$c = \sqrt{a^2 + b^2} = 50 \text{ см}, \quad \frac{\partial c}{\partial a} = \frac{a}{c} = \frac{3}{5}, \quad \frac{\partial c}{\partial b} = \frac{b}{c} = \frac{4}{5},$$

$$\Delta_c = \frac{a}{c} \Delta + \frac{b}{c} \Delta = \frac{7}{5} = 1,4 \text{ см}.$$

В каждом варианте задания три задачи, ниже приведены последовательно первая, вторая и третья задачи вариантов.

**А.** Найти абсолютную и относительную ошибки выражения, где  $a_1$ ,  $a_2$  и  $a_3$  — приближенные величины данные с погрешностями  $\Delta_1$ ,  $\Delta_2$ ,  $\Delta_3$  — соответственно:

$$1) \frac{a_1 \times a_2^\alpha}{(a_1^\beta + a_2) \cdot a_3}, \alpha = 1/3, \beta = 3. \quad 2) \frac{(a_1^\alpha + a_2) \cdot a_3}{a_1 \cdot a_2^\beta}, \alpha = 1/5, \beta = 4.$$

$$3) (a_1^\alpha + a_2) \cdot a_3 + a_1 \cdot a_2^\beta, \alpha = 1/2, \beta = 3.$$

$$4) (a_1 + a_2^\beta) / a_3 + a_1 / a_2^\beta, \alpha = 1/5, \beta = 3/2.$$

$$5) \frac{a_1^\alpha + a_2 \cdot a_3}{a_1 + a_2^\beta}, \alpha = 1/2, \beta = 3. \quad 6) \frac{a_1 \times a_2^\alpha}{(a_1^\beta + a_2) \cdot a_3}, \alpha = 3, \beta = 2/7.$$

$$7) \frac{(a_1^\alpha + a_2) \cdot a_3}{a_1 \cdot a_2^\beta}, \alpha = 2/5, \beta = 3. \quad 8) (a_1^\alpha + a_3) \cdot a_2 + a_1 + a_2^\beta,$$

$$\alpha = 2/5, \beta = 3. \quad 9) \frac{a_1 + a_2^\alpha}{(a_1^\beta + a_2) \cdot a_3}, \alpha = 2, \beta = 2/3.$$

$$10) \frac{(a_1 + a_2^\alpha) \cdot a_3}{a_1 \cdot a_2^\beta}, \alpha = 3, \beta = 4/3.$$

$$11) (a_1^\alpha + a_2) / a_3 + a_1 + a_2^\beta, \alpha = 1/2, \beta = 3.$$

$$12) (a_1^\alpha + a_2) \cdot a_3 + a_1 / a_2^\beta, \alpha = 4, \beta = 1/3.$$

$$13) \frac{a_1^\alpha + a_2 / a_3}{a_1 + a_2^\beta}, \alpha = 1/2, \beta = 3. \quad 14) \frac{a_1 \times a_2^\alpha}{(a_1^\beta + a_2) \cdot a_3}, \alpha = 2/7, \beta = 5.$$

$$15) \frac{(a_1^\alpha + a_3) \cdot a_2}{a_1 \cdot a_2^\beta}, \alpha = 2/5, \beta = 3. \quad 16) (a_1^\alpha + a_2) \cdot a_3 + a_1 + a_3^\beta,$$

$$\alpha = 3, \beta = 4/3.$$

**Б.** Дано приближенное число и его погрешность. Найти количество верных знаков:

- 1) 23,587; 0,08 . 2) 13,58; 0,07. 3) 103,58; 0,03. 4) 1655; 6.
- 5) 323,07; 0,06. 6) 43,837; 0,008. 7) 16,402; 0,009. 8) 13,540; 0,006.
- 9) 31,541; 0,003. 10) 13,42; 0,03. 11) 137,5; 0,08. 12) 134; 20.
- 13) 3457,0; 0,6. 14) 4657; 8. 15) 16,47; 0,07. 16) 130,6; 0,06.

**В.** Дана геометрическая фигура. Определить в трехмерном случае объем и полную поверхность, а в плоском случае площадь и периметр. Погрешность определения размеров линейных элементов равна 1 см:

1) Равнобедренная трапеция со сторонами основания, равными 20 и 30 см, и высотой, равной 12 см.

2) Правильная четырехугольная пирамида со стороной основания, равной 10 см, и высотой, равной 12 см.

3) Конус с высотой, равной 30 см, и радиусом, равным 40 см.



4) Прямоугольный параллелепипед с высотой 30 см, стороной основания 60 см и диагональю основания 100 см.

5) Цилиндр с главной диагональю, равной 100 см, и радиусом, равным 40 см.

6) Равнобедренная трапеция со сторонами основания, равными 20 и 80 см, и высотой, равной 40 см.

7) Правильная четырехугольная пирамида со стороной основания, равной 60 см, и высотой, равной 40 см.

8) Прямоугольный параллелепипед с высотой 25 см, стороной основания 60 и диагональю основания 100 см.

9) Равнобедренная трапеция со сторонами основания, равными 34 и 58 см, и высотой, равной 5 см.

10) Правильная четырехугольная пирамида со стороной основания, равной 120 см, и высотой, равной 80 см.

11) Конус с высотой, равной 12 см, и радиусом основания, равным 5 см.

12) Прямоугольный параллелепипед с высотой 20 см стороной основания 50 и диагональю основания 130 см.

13) Цилиндр с образующей, равной 60 см, и главной диагональю, равной 100 см.

14) Равнобедренная трапеция со сторонами основания, равными 20 и 32 см, и высотой, равной 8 см.

15) Правильная четырехугольная пирамида со стороной основания, равной 24 см, и высотой, равной 5 см.

16) Прямоугольный параллелепипед со стороной основания 12 см, его диагональю 13 см и высотой 40 см.

## 2. ПРИБЛИЖЕНИЕ ФУНКЦИЙ

Во многих случаях функция задается таблично, т.е. известны её значения только в узловых точках (узлах):

Таблица 2.1

$N$	0	1	...	$n$
$X$	$X_0$	$X_1$	...	$X_n$
$Y$	$Y_0$	$Y_1$	...	$Y_n$

Существуют два основных критерия (условия) построения приближающих функций. Критерий интерполяции требует, чтобы приближающая функция проходила через узлы. Критерий аппроксимации требует минимизации некоторого функционала.

Полином степени  $n$  однозначно определяется своими значениями в  $(n+1)$ -й точке с попарно разными абсциссами:  $x_i \neq x_j$ , если  $i \neq j$ . Действительно, выпишем согласно критерию интерполяции систему уравнений  $P_n(x_i) = y_i$  или в развернутом виде:

Система  $(n+1)$ -го уравнения относительно  $a_i, 0 \leq i \leq n$  имеет единственное решение, если  $x_i \neq x_j$  так как в этом случае определитель не равен 0. Существуют методы, позволяющие избежать непосредственного решения системы уравнений для нахождения  $a_i$ .

Рассмотрим вначале  $n=1$  (2.1):

$$P_n(x) = a_0 + a_1x; \quad \begin{cases} a_0 + a_1x_0 = y_0 \\ a_0 + a_1x_1 = y_1 \end{cases};$$

$$a_1 = \frac{y_0 - y_1}{x_0 - x_1}; \quad a_0 = \frac{y_1 x_0 - y_0 x_1}{x_0 - x_1}.$$

Подставляя коэффициенты в  $P_1(x)$ , получим:

$$P_1(x) = y_0 \frac{x - x_1}{x_0 - x_1} + y_1 \frac{x - x_0}{x_1 - x_0}; \quad \text{т.е. полином представлен в виде}$$

суммы двух линейных функций, независимых от ординат, умноженных на ординаты и обладающих свойством

$$P_1(x_0) = y_0 1 + y_1 0, \quad P_1(x_1) = y_0 0 + y_1 1.$$

В этом состоит идея построения интерполяционного полинома Лагранжа. Для произвольного значения  $n$  запишем интерполяционный полином в виде

$$P_n(x) = y_0 \cdot L_0(x) + y_1 L_1(x) + \dots + y_n L_n(x),$$

где  $L_i$  полиномы степени не выше  $n$ , не зависящие от ординат, и обладающие следующим свойством:  $L_i(x_i) = 1$ ,  $L_i(x_j) = 0$ .

Из равенства,  $L_i(x_j) = 0$  следует, что  $L_i$  имеет  $n$  корней (рассматриваются однократные корни).

$$L_i(x) = \frac{(x - x_0)(x - x_1) \dots (x - x_{i-1})(x - x_{i+1}) \dots (x - x_n)}{N_i},$$

где  $N_i$  — коэффициент, который находится из условия  $L_i(x_i) = 1$ .

В результате интерполяционный полином Лагранжа имеет вид

$$P_n(x) = \sum_{i=0}^n y_i \frac{(x - x_0)(x - x_1) \dots (x - x_{i-1})(x - x_{i+1}) \dots (x - x_n)}{(x_i - x_0)(x_i - x_1) \dots (x_i - x_{i-1})(x_i - x_{i+1}) \dots (x_i - x_n)}. \quad (2.2.1)$$

Достоинства интерполяционного полинома Лагранжа является простота конструкции. При заданном наборе абсцисс узловых точек и выбранной расчетной точке  $x^*$  упрощается вычисления для различных ординат  $y_i$ . Недостаток — добавление  $(n+1)$ -ого узла  $(x_{n+1}, y_{n+1})$  требует перерасчета всех слагаемых.

Погрешность вычисления: пусть  $f(x)$  — функция, дифференцируемая  $n+1$  раз, и  $P_n(x)$  — приближающий её интерполяционный полином.

$$|f(x) - P_n(x)| \leq \frac{M_{n+1}}{(n+1)!} |(x-x_0)(x-x_1)\cdots(x-x_n)|,$$

где  $M_{n+1} = \max |f^{(n+1)}(x)|$ ,  $x \in [a, b]$ .

Интерполяционный полином Лагранжа при линейных преобразованиях  $x = at + b$  ( $t$  — новая переменная) сохраняет свой вид.

### 2.3. Интерполяционный полином Ньютона

Пусть  $n=0$ , тогда  $P_0(x) = y_0$ , если  $n = 1$ , то выражение для полинома можно записать в виде:  $P_1(x) = y_0 + \frac{y_1 - y_0}{x_1 - x_0}(x - x_0)$ , т.е. введение приближающей функции с добавлением узлов, уточняется вблизи точки  $x_0$ . Конструкция интерполяционного полинома Ньютона такова:

$$P_n(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1) + a_3(x - x_0)(x - x_1)(x - x_2) + \dots + a_n(x - x_0)(x - x_1)\dots(x - x_{n-1}).$$

Рассматривается равномерная сетка, т.е.  $x_i = x_0 + ih$ .

Для дальнейшего анализа вводится понятие конечной разности. Конечной разностью первого порядка называется величина

$$\Delta y(x) = y(x+h) - y(x), \quad x \in [x_0, x_n].$$

Конечная разность второго порядка определяется по первой

$$\Delta^2 y(x) = \Delta y(x+h) - \Delta y(x) = y(x+2h) - 2y(x+h) + y(x)$$

и т.д., конечная разность  $i$ -го порядка определяется через рекуррентное соотношение

$$\Delta^{(i+1)} y(x) = \Delta^{(i)} y(x+h) - \Delta^{(i)} y(x)$$

и зависит от значений  $y$  в  $(i+1)$ -й точке.

Выражение вида  $x^{[n]} = x(x-h)(x-2h)\dots(x-(n-1)h)$  называется обобщенным произведением. Его первая конечная разность равна

$$\Delta x^{[n]} = (x+h)^{[n]} - x^{[n]} = (x+h)x(x-h)\dots$$

$$\dots(x - (n - 2)h) - x(x - h)(x - 2h)\dots(x - (n - 1)h) = nhx^{[n-1]}. \quad (2.3.1)$$

Отсюда следует выражение для конечных разностей для обобщенного произведения высших порядков.

Подставляя  $x_0$  в  $P_n(x)$ , получим:  $a_0 = P_n(x_0) = y_0$ . Далее определим конечную разность в точке  $x_0$ . Из свойства (2.3.1) получим

$$\Delta P_n(x)|_{x=x_0} = \{a_1h + 2a_2h(x - x_0) + \dots \\ \dots + na_nh(x - x_0)(x - x_1)\dots(x - x_{n-2})\}|_{x=x_0} = a_1h = \Delta y_0.$$

Отсюда следует, что  $a_1 = \frac{\Delta y_0}{h}$ . Точно так же из (2.3.1) следует выражение для конечной разности второго порядка в точке  $x_0$ :

$$\Delta^2 P_n(x)|_{x=x_0} = \{2a_2h^2 + 3 \cdot 2a_3h^2(x - x_0) + \\ \dots + n(n-1)a_nh^2(x - x_0)(x - x_1)\dots(x - x_{n-3})\}|_{x=x_0} = 2a_2h^2 = \Delta^2 y_0.$$

Общая формула имеет следующий вид:  $a_i = \frac{\Delta^{(i)} y_0}{i!h^i}$ .

В результате получаем первый интерполяционный полином Ньютона:

$$P_n(x) = y_0 + \frac{\Delta y_0}{h^1}(x - x_0) + \frac{\Delta^{(2)} y_0}{2!h^2}(x - x_0)(x - x_1) + \dots \\ \dots + \frac{\Delta^{(n)} y_0}{n!h^n}(x - x_0)(x - x_1)\dots(x - x_{n-1}). \quad (2.3.2)$$

Построенный таким образом интерполяционный полином проходит через узловые точки.

Второй интерполяционный полином Ньютона позволяет начать интерполяцию с точки  $x_n$ , т.е. улучшить точность приближения на правой границе интервала интерполяции

$$P_n(x) = a_0 + a_1(x - x_n) + a_2(x - x_{n-1})(x - x_n) + \\ + a_3(x - x_{n-2})(x - x_{n-1})(x - x_n) + \dots + a_n(x - x_1)(x - x_{n-1})\dots(x - x_n).$$

Из структуры полинома следует, что  $a_0 = y_n$ .

$$\Delta P_n(x_{n-1})|_{x=x_n} = \Delta y_{n-1} = \{ha_1 + 2ha_2(x - x_n) + \dots\}|_{x=x_n};$$

$a_1 = \frac{\Delta y_{n-1}}{h \cdot 1!}$ ;  $a_2 = \frac{\Delta^2 y_{n-2}}{h^2 2!}$ ; и так далее. Окончательно получим:

$$P_n(x) = y_n + \frac{\Delta y_{n-1}}{h}(x - x_n) + \frac{\Delta^2 y_{n-2}}{h^2 2!}(x - x_n)(x - x_{n-1}) + \dots$$

$$\dots + \frac{\Delta^n y_0}{h^n n!}(x - x_n) \cdots (x - x_1). \quad (2.3.3)$$

При расчётах и алгоритмизации вычисления интерполяционного полинома применяется таблица конечных разностей:

Таблица 2.2

№	$x$	$y$	$\Delta y$	$\Delta^2 y$	$\Delta^3 y$	...
0	$x_0$	$y_0$	$\Delta y_0$	$\Delta^2 y_0$	$\Delta^3 y_0$	...
1	$x_1$	$y_1$	$\Delta y_1$	$\Delta^2 y_1$	...	
2	$x_2$	$y_2$	$\Delta y_2$	...		
3	$x_3$	$y_3$	...			
...	...	...				

Для построения 1-го интерполяционного полинома Ньютона необходима 1-я строка табл. 2.2. Для построения 2-го интерполяционного полинома Ньютона необходима побочная диагональ таблицы. Обычно при машинных расчётах массив ординат узловых точек последовательно преобразуется в массив коэффициентов  $a_i$  так, что они запоминаются в соответствующих элементах массива.

### 2.3. Примеры и задания для практических занятий

*Пример.* Дана таблица узлов. Построить интерполяционный полином Лагранжа и провести проверку табл. 2.3.

Таблица 2.3

$N$	0	1	2	3
$X$	0	0,5	1	1,5
$Y$	1	2	3	1

В выражение (2.2.1) для  $n=3$

$$P_3(x) = y_0 \frac{(x-x_1)(x-x_2)(x-x_3)}{(x_0-x_1)(x_0-x_2)(x_0-x_3)} + y_1 \frac{(x-x_0)(x-x_2)(x-x_3)}{(x_1-x_0)(x_1-x_2)(x_1-x_3)} +$$

$$+ y_2 \frac{(x-x_0)(x-x_1)(x-x_3)}{(x_2-x_0)(x_2-x_1)(x_2-x_3)} + y_3 \frac{(x-x_0)(x-x_1)(x-x_2)}{(x_3-x_0)(x_3-x_1)(x_3-x_2)}$$

необходимо подставить данные из табл. 2.3:

$$P_3(x) = 1 \frac{(x-0,5)(x-1)(x-1,5)}{(0-0,5)(0-1)(0-1,5)} + 2 \frac{(x-0)(x-1)(x-1,5)}{(0,5-0)(0,5-1)(0,5-1,5)} +$$

$$+ 3 \frac{(x-0)(x-0,5)(x-1,5)}{(1-0)(1-0,5)(1-1,5)} + 1 \frac{(x-0)(x-0,5)(x-1)}{(1,5-0)(1,5-0,5)(1,5-1)}.$$

После преобразований получим:  $P_3(x) = -4x^3 + 6x^2 + 1$ .

*Проверка:*

$$P_3(x_0) = P_3(0) = 1 \equiv y_0, \quad P_3(x_1) = P_3(0,5) = -4/8 + 6/4 + 1 = 2 \equiv y_1,$$

$$P_3(x_2) = P_3(1) = -4 + 6 + 1 = 3 \equiv y_2,$$

$$P_3(x_3) = P_3(1,5) = -4 \cdot 27/8 + 6 \cdot 9/4 + 1 = 1 \equiv y_3.$$

*Пример.* Построить интерполяционные полиномы Ньютона по предыдущей таблице узловых точек.

№	$x$	$y$	$\Delta y$	$\Delta^2 y$	$\Delta^3 y$
0	0	1	1	0	-3
1	0,5	2	1	-3	
2	1	3	-2		
3	1,5	1			

Первый интерполяционный полином Ньютона.

$$P_3(x) = y_0 + \frac{\Delta y}{h}(x-x_0) +$$

$$+ \frac{\Delta^2 y_0}{2h^2}(x-x_0)(x-x_1) +$$

$$+ \frac{\Delta^3 y_0}{6h^3}(x-x_0)(x-x_1)(x-x_2);$$

$$P_3(x) = 1 + 2x + 0 + \frac{-3}{6(0,5)^3} x(x-0,5)(x-1) + 1 + 2x -$$

$$-4(x^3 - 1,5x^2 + 0,5) = -4x^3 + 6x^2 + 1.$$

Второй интерполяционный полином Ньютона:

$$P_3(x) = y_3 + \frac{\Delta y_2}{h}(x - x_3) + \frac{\Delta^2 y_1}{2h^2}(x - x_3)(x - x_2) + \frac{\Delta^3 y_0}{6h^3}(x - x_3)(x - x_2)(x - x_1);$$

$$P_3(x) = 1 - 2\frac{x-1,5}{0,5} + \frac{-3}{2(0,5)^2}(x-1,5)(x-1) + \\ + \frac{-3}{6(0,5)^3}(x-1,5)(x-1)(x-0,5) = -4x^3 + 6x^2 + 1.$$

Варианты задаются по номерам столбцов табл.2.4 и 2.5 в виде дробей:  $\frac{N_y}{N_x}$ , например,  $\frac{9}{2}$  означает, что для узловых точек по  $x$  и  $y$  выбираются девятый и второй варианты соответственно. Каждый студент должен получить три таких дроби для расчета интерполяционного полинома Лагранжа, первого и второго интерполяционного полинома Ньютона. Результат необходимо представить в виде:  $P_3(x) = a_0 + a_1x + a_2x^2 + a_3x^3$ , где коэффициенты правильные или неправильные дроби, не десятичные. Проверка производится подстановкой узловых точек.

**Таблица 2.4**

	Варианты $N_x$		
$n$	1	2	3
0	0	-0,5	-1
1	0,5	0	-0,5
2	1	0,5	0
3	1,5	1	0,5

**Таблица 2.5**

$n$	Варианты $N_y$															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
0	-1	2	1	-1	2	1	-1	2	1	1	-1	0	2	2	1	3
1	2	-1	2	0	1	1	2	-2	2	-1	-1	-1	0	1	-2	0
2	1	1	-1	-1	-1	-1	-2	-1	-1	2	2	2	-1	0	-1	-1
3	1	0	1	2	2	2	1	1	-1	1	-2	1	2	-2	-1	2



### 3. ЧИСЛЕННЫЕ МЕТОДЫ РЕШЕНИЙ ТРАНСЦЕНДЕНТНЫХ И АЛГЕБРАИЧЕСКИХ УРАВНЕНИЙ

Общий вид уравнения  $f(x) = 0$ . Решить уравнение, т.е. найти его корень, означает определить  $x^*$  такое, что  $f(x^*) \equiv 0$ .

Во многих случаях точное значение  $x^*$  найти невозможно, поэтому используются приближенные методы, когда значение корня определяется с заданной точностью  $\varepsilon$ . Геометрически корень – это пересечение графиком функции  $f(x)$  оси  $x$ .

Задача делится на 2 этапа:

1. Локализация корня – т.е. нахождение интервала, на котором изолирован единственный нужный нам корень. Выбор интервала производится путем анализа знака  $f(x)$  в ряде пробных точек. Этот процесс в общем виде не алгоритмизируется.

2. Уточнение положения корня на интервале изоляции.

Свойства функции на интервале изоляции  $[a, b]$ :

- 2.1.  $f(x)$  непрерывна на  $[a, b]$ ;
- 2.2.  $f(x)$  монотонна на  $[a, b]$ , т.е.  $f'(x) > 0$  или  $f'(x) < 0$ , что обуславливает единственность корня;
- 2.3.  $f(x)$  меняет знак на  $[a, b]$ ,  $f(a)f(b) < 0$ , т.е. корень существует;
- 2.4.  $f(x)$  не имеет точек перегиба, т.е.  $f''(x) > 0$  или  $f''(x) < 0$ .

Последние условия не являются в общем случае обязательными, но для сходимости некоторых методов они необходимы. Так, если функция имеет корень в точке своего локального минимума, условие 2.3. не выполняется, однако оно необходимо для сходимости методов дихотомии, хорд и секущих. Для сходимости метода секущих также необходимо выполнение условия 2.4.

Нахождение приближенного значения корня – это итерационный процесс, когда по предыдущему (предыдущим) значениям корня находится следующее приближенное значение. Итерационный процесс прекращается, когда достигается заданная точность:

$$f(x_n - \varepsilon) \cdot f(x_n + \varepsilon) < 0. \quad (3.1)$$

Для этого необходимо, чтобы процесс итераций сходился. Рассмотрим несколько итерационных процедур.

### 3.1. Метод простой итерации для решения нелинейных и трансцендентных уравнений

Уравнение  $f(x) = 0$  преобразуется к виду

$$x = \phi(x), \quad (3.1.1)$$

и если выполняется условие

$$|\phi'(x)| < 1, \quad (3.1.2)$$

то итерационный процесс

$$x_{k+1} = \phi(x_k) \quad (3.1.3)$$

сходится к точному значению. Действительно,  $x_{k+1} - x_k = \phi(x_k) - \phi(x_{k-1})$ , из теоремы о среднем следует оценка:  $|x_{k+1} - x_k| \leq M_1 |x_k - x_{k-1}|$ , т.е., расстояние между точками последовательности уменьшается, если  $M_1 < 1$  — ( $M_1 = \max |\phi'| = q$  — знаменатель сходимости). По теореме о неподвижной точке в этом случае существует предел — решение уравнения. Начальная точка  $x_0$  — любая точка интервала локализации корня. Знаменатель сходимости зависит от вида  $\phi(x)$ . Уравнение  $f(x) = 0$  может быть преобразовано к итерационному виду (3.1.1) множеством различных способов — модификаций одношагового стационарного метода простой итерации (см. также 3.3), выбором которых можно добиться минимума знаменателя сходимости.

Например, исходное уравнение эквивалентно следующему:  $x = x - \lambda f(x)$ . Достаточное условие сходимости (3.1.2) выполняется, если  $\lambda = 1/M_1$ ,  $f'(x) > 0$ , где  $M_1 = \max |f'(x)|, x \in [a, b]$ .

### 3.2. Метод хорд и секущих

На интервале  $[a, b]$  заменим  $f(x)$  линейным интерполяционным полиномом, проходящем через точки  $(a, f(a))$  и  $(b, f(b))$ :

$$P_1(x) = f(a) + \frac{f(b) - f(a)}{b - a}(x - a).$$

В качестве первого приближенного значения корня выберем корень полинома  $P_1(x) = 0$ , тогда:

$$x_1 = a - \frac{f(a)(b-a)}{f(b)-f(a)}. \quad (3.2.1)$$

Далее, если поведение  $f''(x)$  неизвестно, то выбирают интервал, на котором  $f(x)$  меняет знак  $[a; x_1]$  или  $[x_1, a]$ , и на нем строят новую хорду (т.е. в формулу подставляем новые границы интервала), и т.д. до достижения заданной точности (3.1).

Если  $f(x)$  не имеет точки перегиба на  $[a; b]$ , то один из концов множества хорд неподвижен. Условие неподвижной точки:

$$c = \begin{cases} a, \text{ если } f(a)f''(a) > 0; \\ b, \text{ если } f(b)f''(b) > 0. \end{cases} \quad (3.2.2)$$

Анализ  $f''(x)$  позволяет определить неподвижную точку  $c$  и для нахождения  $x_{n+1}$  использовать итерационную формулу:

$$x_{n+1} = \phi(x_n) = x_n - \frac{f(x_n)}{f(c) - f(x_n)} \cdot (c - x_n), \quad (3.2.3)$$

причем  $x_0 = \bar{c}$ .

При отсутствии точки перегиба в области локализации корня более эффективным является двухшаговый метод секущих, в котором последующее приближенное значение корня находится по двум предыдущим. Через первые две точки проводится секущая, пересечение которой с осью абсцисс дает следующее приближенное значение. В результате приходим к итерационной формуле:

$$x_{n+1} = x_n - \frac{f(x_n)}{f(x_n) - f(x_{n-1})} \cdot (x_n - x_{n-1}). \quad (3.2.4)$$

Аналогичная формула получается, если в правой части формулы метода Ньютона вместо производной от функции подставить её конечноразностную аппроксимацию первого порядка в точке  $x_n$ .

### 3.3. Метод касательных (метод Ньютона)

В этом методе в качестве  $x_0$  выбирается одна из границ интервала  $[a, b]$  и из этой точки строится касательная. В качестве приближенного значения корня  $x_1$  принимается точка пересечения касательной с осью абсцисс. Из точки  $(x_1, f(x_1))$  проводится новая касательная и т. д., до достижения заданной точности (3.1).

Уравнение касательной в точке  $x_n$  имеет вид:

$$y_k(x) = f(x_n) + f'(x_n) \cdot (x - x_n), \quad y_k(x_{n+1}) = 0,$$

отсюда следует итерационный процесс:

$$x_{n+1} = \phi(x_n) = x_n - \frac{f(x_n)}{f'(x_n)}. \quad (3.3.1)$$

Выражение для начальной точки  $x_0$  совпадает с (3.2.2).

Метод Ньютона можно считать модификацией метода простой итерации (3.1.1) при  $\phi(x) = x - f(x)/f'(x)$ . Условия сходимости метода следуют из (3.1.2), а именно, для всех  $x$  из области локализации корня должно выполняться

$$q = \left| \frac{f(x)f''(x)}{(f'(x))^2} \right| < 1. \quad (3.3.2)$$

Из 3.3.2 следует, что чем меньше область локализации корня, тем меньше знаменатель  $q$  сходимости метода Ньютона и в пределе  $q \rightarrow 0$  при  $x \rightarrow x^*$ . Таким образом, при достаточно малой области локализации корня сходимость метода Ньютона безусловная.

### 3.4. Скорость сходимости итерационных методов

Введем обозначения:  $\Delta x_n^* = x_n - x^*$ ,  $\Delta x_n = x_n - x_{n-1}$ . Для оценки скорости сходимости необходимо определить зависимость между  $\Delta x_{n+1}^*$  и  $\Delta x_n^*$ .

Если в процессе итераций, начиная с некоторого  $n$ , выполняется

$$\frac{\Delta x_{n+1}^*}{(\Delta x_n^*)^r} \leq q, \quad \text{где } r, q = \text{const}, \quad (3.4.1)$$

то скорость сходимости итерационного процесса определяется показателем  $r$ . При  $r=1$  скорость сходимости линейная, при  $r=2$  – квадратичная, при  $1 < r < 2$  – сверхлинейная. Если (3.4.1) устанавливается при  $n \rightarrow \infty$ , то скорость сходимости называется асимптотической.

В случае метода простых итераций:

$$|x_{n+1} - x^*| = |\phi(x_n) - \phi(x^*)| \leq M_1 |x_n - x^*| \text{ или } \Delta x_{n+1}^* \leq q \Delta x_n^*,$$

то есть скорость сходимости со знаменателем  $q = M_1$ , по меньшей мере, линейная, однако, она может быть выше в конкретной реализации. Заметим, что контролируемые в процессе вычислений величины  $\Delta x_{n+1}$  и  $\Delta x_n$ , в общем случае простой итерации, также связаны между собой в первом приближении аналогичным неравенством

$$|\Delta x_{n+1}| = |\phi(x_n) - \phi(x_{n-1})| \leq q \Delta x_n. \quad (3.4.2)$$

Несмотря на схожесть выражений для метода хорд и секущих, скорость их сходимости различна. Так для метода хорд получим, разлагая выражение для  $\phi(x)$  в точке  $x^*$  в ряд Тейлора и ограничиваясь тремя слагаемыми:

$$x_{n+1} - x^* = x_n - x^* - \frac{f(x^*) + f'(x^*)(x_n - x^*) + f''(x^*)(x_n - x^*)/2}{f'(x^*)(x_n - c) + f''(x_n)((x_n - x^*)^2 - (c - x^*)^2)/2} \cdot (x_n - c).$$

Учитывая, что  $f(x^*) \equiv 0$ , сокращая в числителе и знаменателе  $f'(x^*)(x_n - c)$  и разлагая знаменатель в ряд, получим:

$$x_{n+1} - x^* = \frac{f''(x)}{2f'}(x_n - x^*)(c - x^*). \quad (3.4.3)$$

Оценка (3.8) с учетом того, что расстояние между точками  $x^*$  и  $c$  меньше длины интервала изоляции дает:

$$|x_{n+1} - x^*| \leq \frac{M_2}{2M_1}(b-a)|x_n - x^*|, \quad (3.4.4)$$

то есть скорость сходимости линейная.

В методе секущих в выражении (3.2.4) с необходимо заменить на  $x_{n-1}$ . Предположим, что соотношение для скорости сходимости имеет вид:

$$x_{n+1} - x^* = \left( \frac{f''(x)}{2f'} \right)^t (x_n - x^*)^r.$$

Подставляя полученное из него выражение для  $(x_{n-1} - x^*)$  в (3.4.3), получим для степеней  $r$  и  $t$ :

$$r = 1 + \frac{1}{r}, \text{ и } r \cdot t = 1, \quad r \approx 1,62, \quad t \approx 0,62.$$

Таким образом, сходимость метода секущих сверхлинейная.

Для метода касательных, вычитая из левой и правой части (3.3.1) значение корня и разлагая функцию в ряд, получим

$$x_{n+1} - x^* = x_n - x^* - \frac{f(x^*) + f'(x^*)(x_n - x^*) + f''(x^*)(x_n - x^*)^2 / 2}{f'(x_n)}.$$

Откуда

$$|x_{n+1} - x^*| \leq \frac{M_2}{2M_1} (x_n - x^*)^2, \quad (3.4.4)$$

то есть сходимость метода касательных квадратичная.

Метод хорд используется в тех случаях, когда анализ поведения второй производной затруднен. Метод является безусловно сходящимся, так же, как и известный метод дихотомии — деления отрезка локализации корня пополам. Оба метода обладают линейной скоростью сходимости и знаменателями сходимости, соответственно,  $q = \frac{M_2}{2M_1}(b-a)$  и  $q \in (0.5, 1)$ .

Если точки перегиба на интервале изоляции нет, то используется метод секущих. Если вычисление первой производной не требует значительного машинного времени и отсутствует точка перегиба, то целесообразно применять самый быстрый метод из рассмотренных — метод Ньютона (касательных).

Если вычисление первой производной не требует значительного машинного времени и отсутствует точка перегиба, то целесообразно применять самый быстрый метод из рассмотренных — метод Ньютона (касательных).

Рассмотренные методы сходятся абсолютно при отсутствии точки перегиба.

### 3.5. Условие выхода из вычислительного процесса по заданной точности в методах простой итерации

Покажем практический способ выхода из процесса итераций гарантирующий достижение заданной точности вычислений в общем случае простой итерации со знаменателем  $q$ . Считается, что корень на  $n$ -й итерации вычислен с точностью  $\varepsilon$ , если  $\Delta x_n^* \leq \varepsilon$ . Контроль же в процессе вычислений поддается величина  $\Delta x_n$ . Установив связь между этими величинами, мы получим возможность проводить вычисления с заданной точностью. Заметим, что  $|x_{n+k} - x_n| \rightarrow \Delta x_n^*$  при  $k \rightarrow \infty$ . Далее, учитывая неравенство треугольника и (3.4.2):

$$\begin{aligned} |x_{n+k} - x_n| &\leq \Delta x_{n+k} + \Delta x_{n+k-1} + \dots + \Delta x_{n+1} \leq q^k \Delta x_n + q^{k-1} \Delta x_{n-1} + \dots + q \Delta x_n = \\ &= q(1 + q + \dots + q^{k-1}) \Delta x_n = \frac{q(1 - q^k)}{1 - q} \Delta x_n. \end{aligned}$$

При  $k \rightarrow \infty$  получаем

$$\Delta x_n^* = \frac{q}{1 - q} \Delta x_n.$$

Таким образом, требование

$$\Delta x_n \leq \frac{1 - q}{q} \varepsilon \quad (3.5.1)$$

обеспечивает заданную точность вычислений  $\varepsilon$ .

### 3.6. Пример и задание для практических занятий

*Пример.* Найти методом хорд, касательных и простой итерации корни уравнения:

$$x^3 - Kx + L = 0, K=20, L=10. \quad (3.6.1)$$

Каждый корень искать одним из предложенных методов. Для этого вначале необходимо отделить корни и выбрать метод решения. Рекомендуемый план решения приводится ниже.

1) Находятся первая и вторая производные:

$$f'(x) = 3x^2 - K, \quad f''(x) = 6x.$$

Очевидно, что корни (если они существуют) расположены левее, между и правее точек экстремума функции

$$x_{1,2} = \pm \sqrt{\frac{K}{3}} \approx \pm 2,582.$$

Выбираются три интервала  $[a, b]$  и проверяется условие (3.1) на каждом интервале.

2) Для метода простых итераций уравнение преобразуется к итерационному виду:  $x_{k+1} = \sqrt[3]{Kx_k - L} = \sqrt[3]{20x_k + 10}$  и выбирается интервал  $[a, b] = [3, 5]$ , на котором проверяется выполнение условия (3.1). В качестве начального значения выбирается  $x_0 = 3$ , тогда по (3.1.3) получается  $x_1 = 4,12$ ,  $x_2 = 4,52$ .

3) Для метода хорд выбирается интервал  $[a, b] = [-3, 3]$  и проверяется (3.1)  $f(3)f(-3) = -43 \cdot 23 < 0$ , неподвижной точки на этом интервале не существует, поэтому каждый раз находится новый интервал из условия (3.1), в результате, применяя (3.2.1) получим два последовательных приближенных значения корня:  $x_1 = -0,91$ ,  $x_2 = -0,33$ .

4) Для метода касательных выбирается интервал  $[a, b] = [-3, -5]$  и проверяется выполнение условия (3.1)  $f(-5)f(-3) = -35 \cdot 23 < 0$ , выбирается начальная точка из условия (3.2.2):  $f(-5)f''(-5) = (-35)(-30) > 0$ . По формуле (3.3.1) проводятся две итерации:  $x_1 = -4,36$ ,  $x_2 = -4,21$ .

Варианты для практических и лабораторных занятий приведены в табл. 4.1. Для лабораторных занятий следует графически локализовать корни, затем уточнить корни заданными методами с точностью  $\varepsilon = 10^{-15}$ , вычислить значение функции в каждом найденном корне.

**Таблица 4.1**

№	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$K$	2,3	5,3	3,2	4,2	4,5	3,8	2,8	2,5	3,4	3,9	4,3	4,7	4,9	5,1	2,5	4,8
$L$	4,9	2,9	1,5	2,1	1,9	1,6	1,0	0,8	1,7	2,3	2,1	2,3	2,6	1,8	0,8	1,6



#### 4. ЧИСЛЕННОЕ ИНТЕГРИРОВАНИЕ

Цель – приближенно вычислить определенный интеграл  $I = \int_a^b f(x)dx$  на  $[a, b]$ .

По теореме Ньютона – Лейбница он равен разности верхнего и нижнего пределов первообразной функции  $f(x)$  –  $F(x)$  ( $F'(x) = f(x)$ ). Но для табличных функций их первообразная не существует и даже для известных  $f(x)$  не всегда представима в виде комбинаций элементарных функций.

Интеграл геометрически равен площади криволинейной трапеции.

В численных методах интеграл ищется в виде квадратуры:  $\tilde{I} = \sum_{i=0}^n A_i f(x_i)$ . Необходимо найти оптимальным образом  $A_i$  и  $x_i$ .

Обычно коэффициенты подбираются так, чтобы квадратура давала точное значение для полинома максимально возможной степени.

##### 4.1. Метод Ньютона — Котеса

Предполагается, что значения аргументов известны и расположены равномерно. Требуется найти коэффициенты  $A$ .

Рассмотрим интервал:  $[\xi_0, \xi_n]$ ,  $\xi_i = \xi_0 + hi$ .

На интервале  $[\xi_0, \xi_n]$  заменим  $f(x)$  интерполяционным полиномом Лагранжа (2.1.1), подставляя в него переменную  $q$ , равную

$$q = \frac{\xi - \xi_0}{h} \quad \left( h = \frac{b - a}{n} \right), \quad \text{получим } P_n(q) = \sum_{i=0}^n y_i \frac{(-1)^{n-i}}{i!(n-i)!} \prod_{j=0, j \neq i}^n [q - j],$$

где штрих означает отсутствие в произведении сомножителя с  $j=i$

$$\int_{\xi_0}^{\xi_n} f(\xi) d\xi \approx \int_{\xi_0}^{\xi_n} P_n(\xi) d\xi = \sum_{i=0}^n y_i A_i,$$

коэффициенты  $A_i$  равны

$$A_i = \frac{b-a}{n} \frac{(-1)^{n-i}}{i!(n-i)!} \int_0^n \frac{q^{[n+1]}}{(q-i)} dq = (b-a)H_i, \quad (4.1.1)$$

где  $H_i$  — не зависящие от интервала  $[a, b]$  коэффициенты Котеса.

В дальнейшем рассматривается равномерная сетка узлов с шагом  $h$ .

## 4.2. Метод прямоугольников.

Степень полинома  $n = 0$   $P_0(\xi) = \text{const}$ . Коэффициент Котеса (4.1.1) при  $n = 0$  (вычисляется как предельный переход при  $n \rightarrow 0$ ) равен 1. Интервал  $[\xi_0, \xi_n]$  неопределен, так как есть только одна точка —  $\xi_0$ . Геометрически это обозначает, что  $f(x)$  заменяется на интервале каким-то значением ординаты. Если интервал  $[a, b]$  велик, то его разбивают точками  $x_i$  на  $n$  интервалов и на каждом применяют метод прямоугольников. Для первого интервала приближенное значение интеграла равно  $f(\tilde{x})(x_1 - x_0)$ , где  $\tilde{x} \in [x_0, x_1]$ . В качестве  $\tilde{x}$  обычно применяют:

$x_0$  — метод левых прямоугольников;

$x_1$  — метод правых прямоугольников.

На  $[x_1, x_2]$  повторяют ту же процедуру и результат суммируют:

$$\bar{I}_{\text{л.п}} = h \sum_{i=0}^{n-1} f(x_i), \quad \bar{I}_{\text{п.п}} = h \sum_{i=1}^n f(x_i). \quad (4.2.1)$$

Погрешность метода на интервале длиной  $h$  равна:

$R(h) = \int_x^{x+h} f(t)dt - f(x)h$ , разлагая подынтегральную функцию в ряд Тейлора, получим:

$$R(h) = \int_x^{x+h} (f(x) + f'(x)(t-x))dt - f(x)h = f'(x)\frac{h^2}{2}, \quad \theta \in [x, x+h].$$

Абсолютная погрешность на  $n$  интервалах суммируется. В результате, учитывая, что  $h = \frac{(b-a)}{n}$  получим:  $|I - I_{\text{пр}}| \leq \frac{(b-a)^2}{2n} M_1$ , где  $M_1 = \max |f'(x)|$ .

### 4.3. Метод трапеций

На частичном интервале функция заменяется линейной, т.е.  $n=1$ .  $H_0 = \int_0^1 (q-1)dq = \frac{1}{2}$ ,  $H_1 = \int_0^1 qdq = \frac{1}{2}$ . На интервале  $|x_i, x_{i+1}|$ , заменяя  $f(x)$  на  $P_1(x)$ , получим для равноотстоящих узлов:  $I = h(f_i + f_{i+1})/2$ . То есть, площадь криволинейной трапеции заменена площадью прямоугольной трапеции.

Суммируя по всем интервалам, приходим к выражению:  $I_{\text{тр}} = \frac{h}{2} \sum_{i=0}^{n-1} (f_i + f_{i+1})$ , в котором внутренние ординаты встречается дважды. Окончательно получим

$$I_{\text{тр}} = ((f(a) + f(b))/2 + \sum_{i=1}^{n-1} f_i)h. \quad (4.3.1)$$

Между методом трапеций и методом прямоугольников существует простая связь:

$$I_{\text{тр}} = \frac{I_{\text{л.п}} + I_{\text{п.п}}}{2}. \quad (4.3.2)$$

Погрешность метода

$$R = \int_x^{x+h} f dx - \frac{h}{2}(f(x) + f(x+h)),$$

разлагая функции в ряд Тейлора, получим

$$\begin{aligned} R(h) = \int_x^{x+h} (f(x) + f'(x)(t-x) + \frac{1}{2}f''(x)(t-x)^2)dt - \\ - \frac{h}{2}(f(x) + f(x) + f'(x)h + \frac{h^2}{2}f''(x)) = -f''(x)\frac{h^3}{12}. \end{aligned}$$

Погрешность на интервале интегрирования есть сумма погрешности на каждом частичном интервале, в результате получим

$|I - I_{\text{тр}}| \leq \frac{(b-a)^3}{12n^2} M_2$ ,  $M_2 = \max |f''|$ . Очевидно, что метод трапеций точен для линейной функции.

#### 4.4. Метод парабол (метод Симпсона)

Степень полинома  $n$  равна двум. Рассмотрим интервал длиной  $2h$ :  $[x_{i-1}, x_{i+1}]$ . Коэффициенты Котеса (4.1.1) равны:

$$H_0 = \frac{1}{4} \int_0^2 (q-1)(q-2) dq = \frac{1}{6}, \quad H_1 = -\frac{1}{2} \int_0^2 q(q-2) dq = \frac{2}{3},$$

$$H_2 = \frac{1}{4} \int_0^2 (q-1)q dq = \frac{1}{6}.$$

В результате квадратурная формула имеет вид

$$\int_{x-h}^{x+h} f(t) dt = \frac{h}{3} (f(x+h) + 4f(x) + f(x-h)).$$

Для применения метода парабол на  $[a, b]$ , его необходимо разбить на  $2n$  интервала, т.е. число интервалов должно быть четно. При суммировании по частичным интервалам внутренние четные точки удваиваются, В результате окончательная формула имеет вид:

$$I_{\text{пар}} = \frac{2}{3} h ((f_0 + f_{2n}) / 2 + 2 \sum_{i=1}^n f_{2i-1} + \sum_{i=1}^{n-1} f_{2i}), \quad (4.4.1)$$

где  $f_0 = f(a)$ ,  $f_{2n} = f(b)$ .

Оценка точности метода парабол:

$$R = \int_{x_{i-1}}^{x_{i+1}} y f(t) dt - \frac{h}{3} (f(x-h) + 4f(x) + f(x+h)),$$

и разлагая функции в ряд Тейлора до четвертой производной, получим

$$R = -\frac{h^5}{90} f^{IV}(x).$$

Погрешность  $R$  зависит не от третьей, а от четвертой производной, т.е. приближение имеет повышенную точность и формула парабол

верна для полиномов третьей степени. Окончательно, погрешность имеет вид

$$|R| = \frac{h^5 n}{90} |y^{IV}(\theta)| \leq \frac{(b-a)^5}{2880n^4} M_4, \quad M_4 = \max_{[a,b]} y^{IV}.$$

На практике, достижение заданной точности определяется путем сравнения значений интеграла, рассчитанных для текущего и удвоенного числа разбиений интервала.

#### 4.5. Квадратурные формулы Гаусса

Предварительно необходимо рассмотреть свойства полиномов Лежандра:  $P_n(x) = \frac{1}{2^n n!} \cdot \frac{d^n}{d^n x} (x^2 - 1)^n$  — полином степени  $n$ ,  $x \in [-1, 1]$ . Полиномы ортогональны, т.е.  $\int_{-1}^1 P_n(x) \cdot P_m(x) dx = \delta_{n,m}$ , где  $\delta_{n,m}$  — символ Кронекера.

Имеют  $n$  корней на  $[-1, 1]$ . Для любого полинома  $Q_k(x)$ :  $\int_{-1}^1 P_n(x) \cdot Q_k(x) \cdot dx = 0$ , если  $k < n$ , так как полином степени  $k$  представим в виде линейной комбинации полиномов Лежандра до степени  $k$  включительно.

Исходим из формулы общего вида:

$$\int_{-1}^1 f(t) \cdot dt = \sum_{i=1}^n A_i \cdot f(t_i).$$

Для произвольного отрезка  $[a, b]$  замена переменных  $x = \frac{a+b}{2} + \frac{b-a}{2} \cdot t$  переводит его в отрезок  $[-1, 1]$ , и квадратурная формула Гаусса имеет вид

$$I_g = \left[ \frac{b-a}{2} \sum_{i=1}^n A_i f \left( \frac{a+b}{2} + \frac{b-a}{2} t_i \right) \right]. \quad (4.5.1)$$

Потребуем, чтобы квадратурная формула была точна для полиномов максимальной степени  $2 \cdot n - 1$ , а следовательно, должна

быть точна для  $1, t, \dots, t^{2 \cdot n - 1}$ . Система уравнений  $\sum_{i=1}^n A_i \cdot t_i^k = \frac{1 + (-1)^k}{k + 1}$  нелинейная.

Используем свойство полинома Лежандра:

$$\int_{-1}^1 t^k \cdot P_n(t) \cdot dt = \sum_{i=1}^n A_i \cdot t_i^k \cdot P_n(t_i) = 0 \text{ при } k = 0, 1, \dots, n - 1.$$

Равенство интеграла нулю возможно, если  $t_i$  — корни полинома Лежандра, которые известны.

Полученные  $t_i$  подставляются в первые  $n$  уравнений системы для определения коэффициентов  $A_i$ :

$$A_1 \cdot t_1^i + A_2 \cdot t_2^i + \dots + A_n \cdot t_n^i = \frac{[1 + (-1)^i]}{i + 1}, \quad 0 \leq i \leq n - 1.$$

Определитель системы — определитель Вандермонда — не равен 0, и система имеет единственное решение.

Оценка точности квадратурной формулы Гаусса проводится по формуле

$$\left| I - \overline{I_g} \right| \leq \frac{(b - a)^{2n+1} \cdot (n!)^4 M_{2n}}{(2n!)^3 (2n + 1)}, \text{ где } M_{2n} = \max_{[a,b]} f^{(2n)}.$$

#### 4.6. Задание для практических занятий

В практической работе исследуется сходимость различных методов в зависимости от  $n$  — числа точек разбиения.

Рассматривается интеграл вида  $I = \int_a^b \frac{x + L}{x^2 + x + K} dx$ , где  $a = (K - L) / 2$ ,  $b = K + L$ , значения  $K$ ,  $L$  даны в табл. 4.3,  $n = 4, 6, 8$ .

Точное значение интеграла равно:

$$I = \left[ \frac{1}{2} \ln(x^2 + x + K) + \frac{L - \frac{1}{2}}{\sqrt{K - \frac{1}{4}}} \operatorname{arctg} \frac{x + \frac{1}{2}}{\sqrt{K - \frac{1}{4}}} \right]_a^b.$$

Сравнить его со значениями, полученными методом трапеций (4.3.1), методом парабол (4.5.1), методом Гаусса (4.7.1), коэффициенты этого метода приведены в табл. 4.1

Таблица 4.1

$n$	$i$	$t_i$	$A_i$
$n=4$	1,4	$\mp 0,861136$	0,347854
	2,3	$\mp 0,339981$	0,652145
$n=6$	1,6	$\mp 0,932464$	0,171324
	2,5	$\mp 0,661209$	0,360761
	3,4	$\mp 0,238619$	0,467913
$n=8$	1,8	$\mp 0,960289$	0,101228
	2,7	$\mp 0,796666$	0,222381
	3,6	$\mp 0,525532$	0,313706
	4,5	$\mp 0,183434$	0,362683

Результаты расчетов свести в табл. 4.2:

Таблица 4.2

$n$	4	6	8
$I_{\text{тр}}$	...	...	...
$I_{\text{пар}}$	...	...	...
$I_g$	...	...	...

Построить график зависимости величины интегралов от  $n$ , на который нанести результаты расчетов и точное значение интеграла. Оценить качественно скорость сходимости различных методов.

Таблица 4.3

№	1	2	3	4	5	6	7	8
$K$	3,2	3,4	3,6	3,8	4,0	2,2	2,4	2,6
$L$	1,6	1,8	2,0	2,2	2,4	1,2	1,4	1,6

Окончание табл. 4.3

№	9	10	11	12	13	14	15	16
$K$	2,8	3,0	1,2	1,4	1,6	1,8	4,2	4,4
$L$	1,8	2,2	0,8	1,0	1,2	1,4	3,2	3,4

## 5. ЧИСЛЕННЫЕ МЕТОДЫ ЛИНЕЙНОЙ АЛГЕБРЫ

Рассматриваются численные методы решения систем линейных алгебраических уравнений (СЛАУ), а также нахождения собственных значений и собственных векторов матриц.

### 5.1. Численное решение СЛАУ

СЛАУ используются во многих областях науки и техники и являются наиболее часто встречающимся типом задач вычислительной математики. В общем виде СЛАУ из  $n$  уравнений с  $n$  неизвестными записывается в виде:

$$A\vec{x} = \vec{b}. \quad (5.1)$$

Здесь  $\vec{x}$  — неизвестный вектор решения,  $\vec{b}$  — заданный вектор в  $n$ -мерном пространстве, а

$$A = \left( \begin{array}{cc|c} a_{11} & a_{12} & a_{1n} \\ a_{21} & a_{22} & a_{2n} \\ \hline a_{n1} & a_{n2} & a_{nn} \end{array} \right)$$

— линейный оператор в этом пространстве, заданная матрица размером  $n \times n$  или в другом виде  $A = \{a_{ij}\}$ ,  $i, j = 1, 2, \dots, n$ .

Доказывается, что если определитель матрицы не равен нулю, то СЛАУ имеет единственное решение. Ниже будем полагать, что это условие выполняется. Однако, отличие определителя  $A$  от нуля не могут служить гарантией того, что решение СЛАУ будет найдено численно с заданной точностью. Причиной этого может быть как плохая обусловленность самой системы, так и выбранного алгоритма. Заметим, что близость определителя к нулю и даже весьма малое его значение не свидетельствуют, вообще говоря, о плохой обусловленности системы. В качестве примера можно привести матрицу системы, у которой присутствует только главная диагональ с весьма малыми, но отличными от нуля коэффициентами. Определитель такой матрицы может быть машинный



нуль, в то же время свойства такой матрицы близки к единичной, а ошибка в решении порядка ошибки в задании исходных данных.

Для так называемых плохо обусловленных задач их решение принципиально нельзя получить совершенно точно. Для них малые изменения в исходных данных (коэффициентах матрицы и в векторе правой части), которые могут находиться в пределах точности их задания, приводят к несоразмерно большим изменениям в решении. В результате, в пределах точности задания исходных данных (например, в пределах ошибки округления из-за ограниченного формата числовых данных ЭВМ) может существовать множество различных решений, удовлетворяющих системе.

В качестве примера плохо обусловленной системы можно привести СЛАУ с почти линейно зависимыми строками (столбцами) в матрице. Плохо обусловленным алгоритмом для решения СЛАУ можно назвать метод Гаусса без выбора главного элемента.

Для характеристики обусловленности задачи вводят, так называемое, число обусловленности  $K$ . Для задачи решения СЛАУ в качестве числа обусловленности можно принять

$$K = \|A\| \cdot \|A^{-1}\|.$$

Здесь  $\|\cdot\|$  — какая-либо норма в пространстве  $n$ -мерных векторов, которая выражается через норму вектора следующим образом:

$$\|A\| = \max_{\|\vec{x}\| \neq 0} \frac{\|A \cdot \vec{x}\|}{\|\vec{x}\|} = \max_{\|\vec{x}\|=1} \|A \cdot \vec{x}\|.$$

Норма матрицы характеризует максимально возможное относительное увеличение по норме ненулевого вектора при воздействии на него матрицы.

Пусть решение  $\vec{x}$  СЛАУ получено с относительной ошибкой  $\delta\vec{x}$ . Тогда для нее справедлива оценка:

$$\|\delta\vec{x}\| < \approx K \varepsilon_{\text{маш}}.$$

Здесь  $\varepsilon_{\text{маш}}$  — машинная константа, т.е. наименьшее число, которое при прибавлении к единице ещё изменяет её значение в машинном представлении. Отметим, что оценка справедлива для малых ошибок в заданной матрице  $K \|\Delta A\| / \|A\| < 1$ .

Введём понятие невязки  $\vec{h}$  решения:

$$\vec{h} = A \cdot \vec{x} - \vec{b}. \quad (5.2)$$

Заметим, что малость невязки  $\|\vec{h}\| < \varepsilon \|\vec{x}\|$ ,  $\varepsilon \ll 1$  не гарантирует малость ошибки  $\Delta\vec{x}$  в решении. Так, для невязки выполняется соотношение

$$\|\vec{h}\| \leq \|A\| \|\vec{x}\| \varepsilon_{\text{маш}},$$

в то время как для  $\Delta\vec{x}$  справедливо

$$\|\Delta\vec{x}\| \leq \|A^{-1}\| \|\vec{h}\|.$$

Норма обратной матрицы для плохо обусловленной СЛАУ велика, также как и число обусловленности  $K$ , характеризующее в этом случае близость матрицы к вырожденной (сингулярной), для которой  $\|A^{-1}\| \rightarrow \infty$ .

Существуют два основных класса методов для решения СЛАУ – прямые и итерационные. Прямые методы характеризуются тем, что при абсолютной точности вычислений (на гипотетической бесконечноразрядной ЭВМ) точное решение СЛАУ может быть получено с помощью конечного числа арифметических операций. Итерационные методы характеризуются тем, что даже при абсолютной точности вычислений за конечное число арифметических операций может быть получено лишь приближенное решение системы, хотя возможно и как угодно близкое к точному. Однако при реальных вычислениях на ЭВМ указанное различие теряет свой смысл, и для многих задач итерационные методы оказываются более предпочтительными, чем прямые в силу отсутствия накопления ошибок для сходящегося процесса и возможности приблизиться к решению с заданной точностью.

Рассмотрим сначала прямые методы. Наиболее известным является метод Гаусса, поскольку другие методы являются, как правило, его модификацией.

## 5.2. Прямые методы решения СЛАУ

Количество операций для решения системы  $\sim n^3$ . Матрица  $A$  либо неявно обращается, либо представляется в виде произведения матриц удобных для обращения.

В первом случае матрица  $A$  последовательно преобразуется с помощью элементарных (эквивалентных) преобразований:

1. Перестановка столбцов и строк.
2. Умножение столбцов и строк на число.
3. Прибавление к строке (столбцу) другой строки, умноженной на число.

Каждое элементарное преобразование можно представить в виде матрицы  $L_i$ , в результате последовательного умножения  $A$  на  $L_i$ , она преобразуется в единичную матрицу:

$$L_n \dots L_2 L_1 A \cdot \vec{x} = L_n \dots L_1 \vec{b}.$$

### 5.2.1. Метод Гаусса (метод исключений)

Формально метод Гаусса основан на последовательном применении матриц

$$L_i = \begin{pmatrix} 1 & 0 & 0 \\ 0 & l_{ii} & 0 \\ 0 & l_{ij} & 0 \\ 0 & l_{ni} & 0 \end{pmatrix}; \quad l_{ii} = \frac{1}{a_{ii}}; \quad l_{ij} = -\frac{a_{ij}}{a_{ii}}. \quad (5.2.1.1)$$

Пример для матрицы  $(3 \times 3)$ :

$$\begin{pmatrix} a_{11}^{-1} & 0 & 0 \\ -\frac{a_{21}}{a_{11}} & 1 & 0 \\ -\frac{a_{31}}{a_{11}} & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} =$$

$$= \begin{pmatrix} 1 & \frac{a_{12}}{a_{11}} & \frac{a_{13}}{a_{11}} \\ 0 & a_{22} - \frac{a_{12} \cdot a_{21}}{a_{11}} & a_{23} - \frac{a_{21} \cdot a_{13}}{a_{11}} \\ 0 & a_{32} - \frac{a_{12} \cdot a_{31}}{a_{11}} & a_{33} - \frac{a_{13} \cdot a_{31}}{a_{11}} \end{pmatrix}.$$

Действие матрицы  $L_i$  преобразуют элементы  $i$ -го столбца матрицы  $A$  ниже диагонали в нулевые (т.е. исключают их).

### 5.2.2. Вычислительная схема метода Гаусса

В каждом уравнении выделяется ведущий элемент, на который производится деление; пусть это будет  $a_{11}$ . Делим первое уравнение на  $a_{11}$ :

$$c_{1j} = \frac{a_{1j}}{a_{11}}; \quad g_1 = \frac{b_1}{a_{11}}.$$

Все остальные элементы преобразуются по схеме

$$a_{ij}^{(1)} = a_{ij} - a_{i1} \cdot \frac{a_{1j}}{a_{11}}; \quad b_i^{(1)} = b_i - a_{i1} \cdot \frac{b_1}{a_{11}}; \quad (5.2.2.1)$$

$$\left\{ \begin{array}{l} x_1 + c_{12}x_2 + \dots + c_{1n}x_n = g_1; \\ a_{22}^{(1)}x_2 + \dots + a_{2n}^{(1)}x_n = b_2^{(1)}; \\ \dots\dots\dots \\ a_{n2}^{(1)}x_2 + \dots + a_{nn}^{(1)}x_n = b_n^{(1)}. \end{array} \right.$$

На втором шаге ведущим элементом выбирается  $a_{22}^{(1)}$ , на него делится вторая строка, а все остальные элементы преобразуются по формулам

$$c_{2j} = \frac{c_{2j}^{(1)}}{a_{22}^{(1)}}; g_2 = \frac{b_2^{(1)}}{a_{22}^{(1)}}; \quad (5.2.2.2)$$

$$a_{ij}^{(2)} = a_{ij}^{(1)} - a_{i2}^{(1)} \cdot \frac{a_{2j}^{(1)}}{a_{22}^{(1)}}; b_i^{(2)} = b_i^{(1)} - a_{i2}^{(1)} \cdot \frac{b_2^{(1)}}{a_{22}^{(1)}}.$$

Элементы во втором столбце с  $i > 2$  становятся равны 0. В результате таких преобразований, мы приходим к верхней треугольной матрице с единичной диагональю:

$$\begin{pmatrix} 1 & c_{12} & \dots & c_{1n} \\ 0 & 1 & \dots & c_{2n} \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix} = \begin{pmatrix} g_1 \\ g_2 \\ \dots \\ g_n \end{pmatrix}.$$

Преобразование к верхней треугольной матрице называется прямым ходом.

Далее следует обратный ход: начиная с  $x_n$ , последовательно вычисляются компоненты вектора:

$$x_n = g_n; \quad x_{n-1} = g_{n-1} - c_{n-1,n} x_n; \\ x_k = g_k - \sum_{i=k+1}^n c_{ki} x_i, k = n, n-1, \dots, 1. \quad (5.2.2.3)$$

В машинных расчетах в качестве ведущего элемента обычно выбирается максимальный элемент  $i$ -го столбца с  $j > i$  или строки  $a_{ij}$  с  $i > j$ .

Эта строка (или столбец) переставляются на место  $i$ -й строки (столбца). Такой выбор уменьшает ошибки округления. При ручных расчетах элементы матрицы записываются вместе с элементами вектора  $\vec{b}$  в расширенную матрицу:

$$\left( \begin{array}{ccc|c} a_{11} & a_{12} & a_{13} & b_1 \\ a_{21} & a_{22} & a_{23} & b_2 \\ a_{31} & a_{32} & a_{33} & b_3 \end{array} \right),$$

далее из соображений удобства выбирают ведущий элемент, а преобразование остальных элементов на одном шаге прямого хода метода Гаусса проводят по *правилу прямоугольника*. В матрице выделяется прямоугольник, на главной диагонали которого расположены ведущий и преобразуемый элементы.

Пусть  $a_{ii}$  — ведущий элемент, тогда

$$a_{kt}^{(1)} = a_{kt} - \frac{a_{ki}a_{it}}{a_{ii}}. \quad (5.2.2.4)$$

Из преобразуемого элемента вычитается произведение элементов, стоящих на побочной диагонали, деленное на ведущий элемент.

### 5.2.3. Ортогонализация матриц

Матрица называется ортогональной, если  $A \cdot A^T = D$ , где  $D$  — диагональная матрица, т.е. в ней отличны от нуля только диагональные элементы, если  $A \cdot A^T = E$ , то  $A$  — ортонормированная матрица. Любая неособенная матрица  $A$  может быть представлена в виде:  $A = R \cdot T$ ,  $R$  — ортогональная, а  $T$  — верхняя треугольная матрица с единичной диагональю.

Рассмотрим матрицу  $A$ , как набор вектор-столбцов  $\vec{a}_i$ ,  $A = [\vec{a}_1 | \vec{a}_2 | \dots | \vec{a}_n]$  — вектора  $\vec{a}_i$  — линейно независимы, так как  $\det A \neq 0$ . Выберем первый столбец матрицы  $R = \vec{r}_1$ , равным  $\vec{a}_1$ ;  $\vec{r}_1 = \vec{a}_1$ . Запишем  $\vec{a}_2 = t_{12}\vec{r}_1 + \vec{r}_2$ , условие ортогональности  $R$   $(\vec{r}_1, \vec{r}_2) = 0$  позволяет получить  $t_{12}$ :

$$(\vec{r}_1, \vec{a}_2) = t_{12} \cdot (\vec{r}_1, \vec{r}_1), \quad t_{12} = \frac{(\vec{r}_1, \vec{a}_2)}{(\vec{r}_1, \vec{r}_1)},$$

следовательно, известен и вектор  $\vec{r}_2 = \vec{a}_2 - t_{12}\vec{r}_1$ . Аналогичным образом представляется и  $\vec{a}_3 = t_{13} \cdot \vec{r}_1 + t_{23} \cdot \vec{r}_2 + \vec{r}_3$ , где

$$t_{13} = \frac{(\vec{r}_1, \vec{a}_3)}{(\vec{r}_1, \vec{r}_1)}, t_{23} = \frac{(\vec{r}_2, \vec{a}_3)}{(\vec{r}_2, \vec{r}_2)}.$$

В общем случае получим выражения:

$$\vec{r}_k = a_k - \sum_{i=1}^{k-1} t_{ik} \cdot \vec{r}_i, t_{ik} = \frac{(\vec{r}_i, \vec{a}_k)}{(\vec{r}_i, \vec{r}_i)}. \quad (5.2.3.1)$$

Покажем, что  $t_{ik}$  – элементы матрицы  $T$ . Действительно:

$$\vec{a}_1 = \vec{r}_1;$$

$$\vec{a}_2 = t_{12} \cdot \vec{r}_1 + \vec{r}_2;$$

$$\vec{a}_3 = t_{13} \cdot \vec{r}_1 + t_{23} \cdot \vec{r}_2 + \vec{r}_3;$$

.....

или иначе

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots \\ a_{21} & a_{22} & a_{23} & \dots \\ a_{31} & a_{32} & a_{33} & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & \dots \\ r_{21} & r_{22} & r_{23} & \dots \\ r_{31} & r_{32} & r_{33} & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix} \cdot \begin{bmatrix} 1 & t_{12} & t_{13} & \dots \\ 0 & 1 & t_{23} & \dots \\ 0 & 0 & 1 & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix}$$

#### 5.2.4. Решение системы уравнений методом ортогонализации

Оптимальной является следующая схема, основанная на свойствах вектора  $\vec{r}$ . Запишем систему  $A \cdot \vec{x} = \vec{b}$  в виде:  $\vec{a}_1 \cdot x_1 + \vec{a}_2 \cdot x_2 + \dots + \vec{a}_n \cdot x_n = \vec{b}$ . Из структуры векторов  $\vec{r}$  следует, что  $(\vec{r}_i, \vec{a}_j) = 0, (i > j)$ . Умножаем систему слева на  $\vec{r}_n$ :

$$\vec{r}_n \cdot (\vec{a}_1 \cdot x_1 + \vec{a}_2 \cdot x_2 + \dots + \vec{a}_n \cdot x_n) = \vec{r}_n \cdot \vec{b},$$

в уравнении остается одно слагаемое:  $(\vec{r}_n \cdot \vec{a}_n) \cdot x_n = (\vec{r}_n \cdot \vec{b}),$

$$x_n = \frac{(\vec{r}_n, \vec{b})}{(\vec{r}_n, \vec{a}_n)};$$

$$\vec{a}_1 \cdot x_1 + \dots + \vec{a}_{n-1} \cdot x_{n-1} = \vec{b} - \vec{a}_n \frac{(\vec{r}_n \cdot \vec{b})}{(\vec{r}_n \cdot \vec{a}_n)} = \vec{b}^{(1)}.$$

Полученную систему умножим на  $\vec{r}_{n-1}$ , находим  $x_{n-1}$  и вычисляем  $\vec{b}^{(2)}$  и т.д.

$$x_i = \frac{(\vec{r}_i, \vec{b}^{(n-i)})}{(\vec{r}_i, \vec{a}_i)}; \quad \vec{b}^{(i)} = \vec{b}^{(i-1)} - x_{n-i+1} \cdot \vec{a}_{n-i+1}. \quad (5.2.4.1)$$

### 5.3. Итерационные методы решения СЛАУ

#### 5.3.1. Метод простой итерации

Многие итерационные методы могут быть сведены к процессу простой итерации. При этом исходное уравнение тем или иным способом должно быть сведено к уравнению

$$\vec{x} = B\vec{x} + \vec{b}. \quad (5.3.1.1)$$

Здесь  $\vec{x}$  – неизвестный вектор,  $\vec{b}$  – заданный вектор правой части,  $B$  – заданная матрица коэффициентов (оператор). Например, если задана СЛАУ (5.1), то непосредственно принимая

$$B = I - A, \quad (5.3.1.2)$$

где  $I$  – единичная матрица, приходим к (5.3.1.1).

Процесс простой итерации строится следующим образом:

$$\vec{x}^{(k+1)} = B\vec{x}^{(k)} + \vec{b}, \quad k = 0, 1, 2, \dots. \quad (5.3.1.3)$$

В качестве начального приближения  $\vec{x}^{(0)}$  можно принять  $\vec{x}^{(0)} = \vec{b}$ .

Заметим, что переход от (5.1) к (5.3.1.1) может быть выполнен не единственным способом, что приводит к различным модификациям метода простой итерации. Так, метод (5.3.1.3) с преобразованием (5.3.1.2) известен в литературе как метод Рундсона.



Другие методы простой итерации будут рассмотрены в разделе 5.3.2.

Процесс простой итерации может быть эквивалентно записан также в виде ряда по степеням оператора  $B$ , т.е. в виде так называемого ряда Неймана

$$\vec{x} = \vec{b} + B\vec{b} + B^2\vec{b} + \dots = \sum_{i=0}^{\infty} B^i \vec{b}. \quad (5.3.1.4)$$

Если матрица  $B$  постоянна (не зависит от номера итерации  $k$ ), то такой итерационный процесс называется стационарным.

Пусть  $\vec{x}^*$  – «гипотетическое» точное решение, строго удовлетворяющее  $\vec{x}^* = B\vec{x}^* + \vec{b}$ , а  $\Delta\vec{x}^{(k)} = \vec{x}^{(k)} - \vec{x}^*$  – ошибка на  $k$ -м шаге. Подставляя в формулу простой итерации получаем для соотношения ошибок на  $(k+1)$  и  $k$ -м шаге  $\Delta\vec{x}^{(k+1)} = B\Delta\vec{x}^{(k)}$ . Для нормы ошибки

$$\|\Delta\vec{x}^{(k+1)}\| \leq \|B\| \|\Delta\vec{x}^{(k)}\| \leq \|B\|^k \|\Delta\vec{x}^{(1)}\|.$$

Отсюда следует достаточное условие сходимости процесса простой итерации:  $\|B\| = \delta$ ,  $\delta < 1$ .

Действительно, тогда

$$\|\Delta\vec{x}^{(k+1)}\| \leq \|B\|^k \|\Delta\vec{x}^{(1)}\| = \delta^k \|\Delta\vec{x}^{(1)}\| \ll \|\Delta\vec{x}^{(1)}\|.$$

Оператор с  $\|B\| = \delta < 1$  называется сжимающим, а процесс (5.3.1.2), (5.3.1.3) для него сходящимся, т.к. ошибка убывает с каждым шагом независимо от её начальной величины.

Спектральным радиусом матрицы (конечномерного оператора)  $B$  называется  $\rho(B) = \max_i |\beta_i|$ , где  $\beta_i$  – собственные числа оператора  $B$  (см. 5.4).

Для любой нормы справедливо соотношение  $\rho(B) \leq \|B\|$ .

Доказывается, что необходимым и достаточным условием сходимости процесса простой итерации (5.3.1.3) является

$$\rho(B) < 1, \quad (5.3.1.5)$$

при этом итерации сходятся не хуже геометрической прогрессии со знаменателем  $q = \rho(B)$ .

Условие (5.3.1.5) является, как правило, сильным ограничением при непосредственном применении метода (5.3.1.2), (5.3.1.3) к заданной СЛАУ. Выбор нового оператора  $\tilde{B}$  с другим спектром при эквивалентности исходной системе (5.1) может значительно расширить область сходимости процесса простой итерации с его участием:

$$\vec{x}^{(k+1)} = \tilde{B}\vec{x}^{(k)} + \vec{\tilde{b}}, \quad k = 0, 1, 2, \dots \quad (5.3.1.6)$$

В качестве условия выхода из вычислительного процесса по достижении заданной точности решения  $\varepsilon$ , аналогично (3.5.1), можно принять:  $|\vec{x}^{(k+1)} - \vec{x}^{(k)}| \leq \frac{1-q}{q}\varepsilon$ , где  $q$  – спектральный радиус  $B$  или какая-либо оценка другой нормы  $B$ .

### 5.3.2. Метод Якоби и метод Зейделя

Исторически одними из самых ранних итерационных методов являются метод Якоби и метод Зейделя, которые могут быть представлены в виде модификации метода простой итерации. Перепишем (5.1) в следующем виде:

$$\left\{ \begin{array}{l} x_1 = \frac{1}{a_{11}}(b_1 - a_{12}x_2 - a_{13}x_3 - \dots - a_{1n}x_n), \\ \dots \dots \dots \\ x_i = \frac{1}{a_{ii}}(b_i - a_{i1}x_1 - \dots - a_{ii-1}x_{i-1} - a_{ii+1}x_{i+1} - \dots - a_{in}x_n), \\ \dots \dots \dots \\ x_n = \frac{1}{a_{nn}}(b_n - a_{n1}x_1 - a_{n2}x_2 - \dots - a_{nn-1}x_{n-1}). \end{array} \right. \quad (5.3.2.1)$$

Используем (5.3.2.1) для построения процесса итераций, начиная с  $\vec{x}^{(0)} = \vec{b}$  при  $k = 0, k = 0, 1, 2, \dots$ :

$$x_i^{(k+1)} = \frac{1}{a_{i,i}} (b_i - a_{i,1}x_1^{(k)} - \dots - a_{i,i-1}x_{i-1}^{(k)} - a_{i,i+1}x_{i+1}^{(k)} - \dots - a_{i,n}x_n^{(k)}),$$

$$i = 1, 2, \dots, n. \quad (5.3.2.2)$$

В матричных обозначениях метод Якоби можно записать следующим образом. Представим  $C = D - A$ , где  $D$  — диагональная матрица,  $\{D\}_{i,i} = \{A\}_{i,i}, i = 1, 2, \dots, n, \{D\}_{i,j} = 0, i \neq j$ .  $C$  — матрица с нулевой главной диагональю. Тогда справедлива запись уравнения аналогично (5.3.1.6), где

$$\tilde{B} = D^{-1}C, \quad \vec{\tilde{b}} = D^{-1}\vec{b}. \quad (5.3.2.3)$$

Матрица  $D^{-1}$  — диагональная и  $\{D^{-1}\}_{i,i} = 1/\{A\}_{i,i}, i = 1, 2, \dots, n$ .

Необходимые и достаточные условия сходимости метода Якоби:

$$\rho(D^{-1}(D - A)) < 1.$$

Другой известный метод простой итерации для случая, когда  $\tilde{B}$  строится на основе матрицы с нулевой главной диагональю — это метод Зейделя. Он отличается от метода Якоби тем, что при расчете координат вектора  $\vec{x}^{(k+1)}$  на текущей  $(k+1)$ -й итерации используются не только координаты вектора на предыдущей  $k$ -й итерации  $\vec{x}^{(k)}$ , но и уже ранее найденные на текущей итерации координаты вектора  $\vec{x}^{(k+1)}$ :

$$x_i^{(k+1)} = \frac{1}{a_{i,i}} (b_i - a_{i,1}x_1^{(k+1)} - \dots - a_{i,i-1}x_{i-1}^{(k+1)} - a_{i,i+1}x_{i+1}^{(k)} - \dots - a_{i,n}x_n^{(k)}),$$

$$i = 1, 2, \dots, n; \quad k = 0, 1, 2, \dots \quad (5.3.2.4)$$

В матричных обозначениях это соответствует представлению исходной матрицы  $A$  как  $A = L + D + U$ , где  $L$  — нижняя треугольная матрица,  $D$  — диагональная матрица,  $\{D\}_{i,i} = \{A\}_{i,i}, i = 1, 2, \dots, n$ , и  $U$  — верхняя треугольная матрица.

В отличие от метода Якоби действие оператора  $\tilde{B}$  на вектор предыдущей итерации разделяется здесь на две части:

$$\tilde{B}\vec{x}^{(k)} = -D^{-1}L\vec{x}^{(k+1)} - D^{-1}U\vec{x}^{(k)}, \quad (5.3.2.5)$$

и процесс его воздействия (но не результат!) нельзя свести к воздействию какой-либо матрицы на вектор предыдущей итерации.

Метод Зейделя хорошо алгоритмизируется. Если известна скорость сходимости метода, нет необходимости хранить оба вектора  $\vec{x}^{(k+1)}$  и  $\vec{x}^{(k)}$ .

Достаточными условиями сходимости методов Якоби и Зейделя является диагональное преобладание в матричных элементах:

$$q|a_{i,i}| \geq \sum_{j \neq i} |a_{i,j}|, \quad q < 1, \quad \text{для всех } i = 1, 2, \dots, n,$$

однако на практике область сходимости значительно шире и определяется условием (5.3.1.5) на спектральный радиус матрицы (5.3.2.3) для метода Якоби и оператора (5.3.2.5) для метода Зейделя. Для решения СЛАУ с ленточными матрицами метод Зейделя является превосходным инструментом. Так, для симметричных положительно определенных матриц он будет всегда сходящимся. Однако возможно улучшение сходимости как метода Зейделя, так и любого другого метода простой итерации с помощью изложенного ниже метода оптимального спектрального параметра.

### **5.3.3. Метод оптимального спектрального параметра (ОСП) для простой итерации**

Рассмотрим случай, когда спектр оператора  $B$  выходит за границы единичного круга на комплексной  $\beta$ -плоскости собственных чисел. В этом случае ряд простой итерации (5.3.1.3) расходится.

Определим выпуклую оболочку спектра оператора  $B$  как выпуклую замкнутую кривую наименьшей меры, полностью охватывающую спектр оператора на  $\beta$ -плоскости. Доказывается, что если точка  $l$  находится вне выпуклой оболочки спектра, то можно построить сходящийся ряд простой итерации с новым

$$\tilde{\lambda} = \frac{\lambda - k_0}{1 - k_0}$$

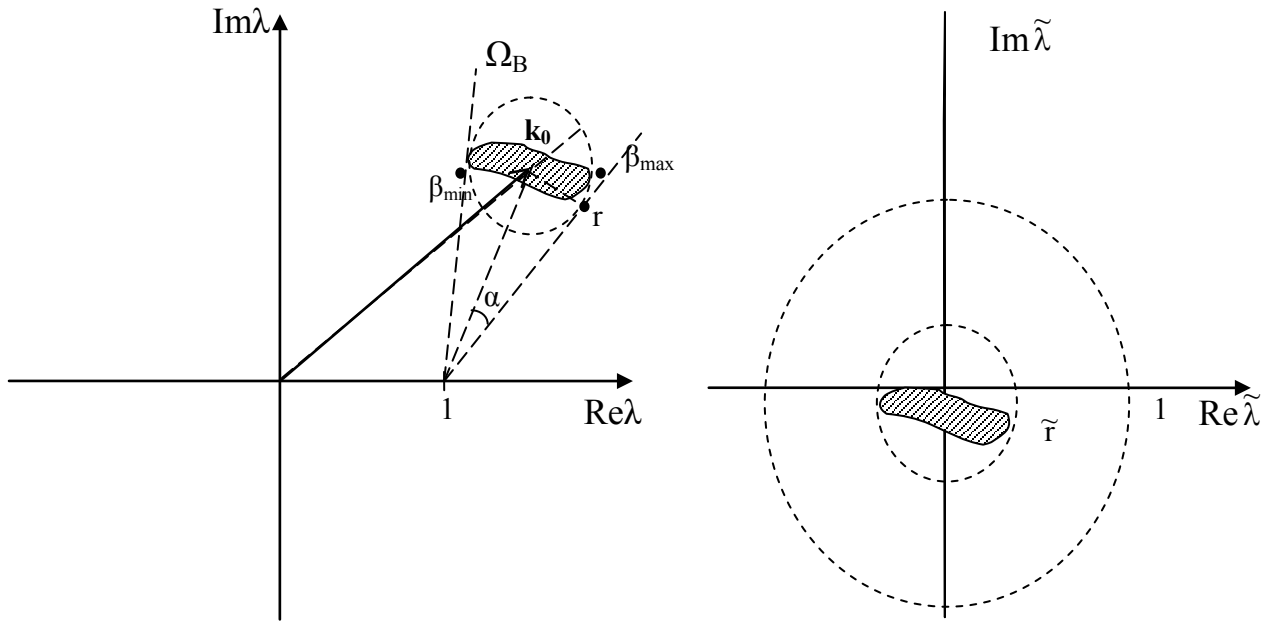


Рис. 5.1

оператором  $\tilde{B} = f(B)$ . Дадим конструктивный способ построения такого сходящегося ряда. Примем:

$$\tilde{B} = \frac{B - kI}{1 - k}, \quad \vec{\tilde{b}} = \frac{\vec{b}}{1 - k}, \quad (5.3.3.1)$$

где  $k$  – комплексный параметр. При  $k \neq 1$  исходные уравнения (5.3.1.1) с операторами  $B$  и  $\tilde{B}$  эквивалентны. Выбором  $k$  попробуем добиться сходимости ряда (5.3.1.6).

Пусть  $\Omega_B$  – один из множества кругов радиуса  $r$ , полностью охватывающих спектр оператора  $B$ , и пусть при этом точка  $1 \notin \Omega_B$  (Рис.1). Очевидно, что  $\Omega_B$  включает в себя выпуклую оболочку спектра. Вектор из начала  $\beta = 0$  в центр этого круга обозначим  $\vec{k}_0$ . При дробно-линейном преобразовании (5.3.3.1) с  $k = k_0$  круг  $\Omega_B$  переходит в круг  $\Omega_{\tilde{B}}$  с центром в точке  $\beta = 0$  и радиусом

с  $|\tilde{r}| = \left| \frac{r}{1 - k_0} \right|$ . Если  $|\tilde{r}| < 1$ , то ряд (5.3.1.6) сходится.

Найдем минимум значения  $|\tilde{r}|$ . Пусть круг  $\Omega_B$  «виден» из точки 1 под углом  $2\alpha$ . Пусть  $\vec{r}$  – вектор из центра круга  $\vec{k}_0$  в точку касания луча из т.1 и круга. Из рис. 1 очевидно, что  $|\tilde{r}| = \left| \frac{\vec{r}}{1 - \vec{k}_0} \right| = \sin(\alpha)$  и, следовательно,  $|\tilde{r}| < 1$ .

Таким образом, если  $\Omega_B$  такой круг, что точка  $l \notin \Omega_B$  и «видимый» из точки  $l$  под наименьшим углом  $2\alpha$ , то комплексное расстояние до центра этого круга есть оптимальный параметр для сходимости (5.3.1.6), а скорость сходимости ряда (5.3.1.6) не хуже, чем у геометрической прогрессии со знаменателем  $\sin(\alpha)$ .

Пусть для спектра  $\{\beta_v\}$  известны оценки для  $\beta_{\min}$ ,  $\beta_{\max}$  – минимального и максимального по модулю собственного числа (или нижней и верхней границы расстояния от т.  $\theta$  до области расположения спектра в случае непрерывного спектрального множества). Тогда, если весь спектр оператора размещается в круге  $\Omega_B$ , натянутом на точки  $\beta_{\min}$ ,  $\beta_{\max}$  как на концевые точки диаметра и точка  $l \notin \Omega_B$ , для оптимального параметра верна простая приближенная формула

$$k_0 = \frac{\beta_{\min} + \beta_{\max}}{2}. \quad (5.3.3.2)$$

Если граница круга принадлежит спектру, то формула (5.3.3.2) точная. Точная она также и в случае вещественного спектра. Формулу (5.3.3.2) можно улучшить, учитывая более точную конфигурацию спектральной области, например, если область расположения спектра – прямая линия. С помощью формулы (5.3.3.2) во многих случаях можно найти значение близкое к оптимальному параметру в условиях неполного знания свойств спектра, но при известных минимальных и максимальных по модулю собственных числах.

Сходимость каждого из рассмотренных методов простой итерации зависит от конкретного вида исходной матрицы, а точнее,

от свойств её спектра. Можно привести примеры матриц, для которых сходится только один из рассмотренных методов, однако комбинация метода простой итерации, Зейделя или Якоби с методом оптимального спектрального параметра (ОСП) позволяют добиться сходимости в случаях, когда каждый из этих методов по отдельности расходится.

Рассмотрим применение метода ОСП на примерах конкретных матричных задач.

Пусть элементы матрицы  $A$  при  $n=2$  следующие:  $a_{11}=2$ ,  $a_{22}=-2$ ,  $a_{12}=3$ ,  $a_{21}=7$ . Собственные числа матрицы  $B$  (5.3.1.2) равны  $\beta_1=-4$ ,  $\beta_2=6$  и располагаются по разные стороны от точки  $I$  на прямой, проходящей через неё. В этом случае точка  $I$  принадлежит выпуклой оболочке спектра и дробно-линейным преобразованием (5.3.3.1) нельзя добиться сходимости итерационного процесса. Собственные же числа матрицы Якоби (5.3.2.3) равны  $Y_1=-2.3i$ ,  $Y_2=2.3i$  (здесь  $i$  – мнимая единица) и точка  $I$  находится вне выпуклой оболочки спектра. То же самое можно утверждать и о спектре оператора Зейделя. Однако, непосредственное применение метода Якоби или Зейделя не приведёт к сходящемуся ряду, т.к.  $|Y_v|>1$  и не выполняется (5.3.1.5). Заклучая спектр  $Y_v$  в круг  $\Omega_Y$  с центром в т.  $-8$  приходим к сходящемуся методу Якоби – ОСП с параметром  $k=8$ . Для метода Зейделя – ОСП оптимальный параметр  $k=1$  приводит к быстро сходящемуся процессу. Решение СЛАУ (5.1) с правой частью  $b_v=1, i=1,2$  и точностью  $\varepsilon=10^{-5}$  достигается за  $m=20$  итераций ряда (5.3.1.6).

Наоборот, если матрица Якоби (оператор Зейделя) имеют спектр, выпуклая оболочка которого содержит т.  $I$ , то никакие модификации этих методов не приведут к сходящемуся процессу. Применение метода ОСП непосредственно к исходной матрице в виде (1.2) может привести в этом случае к сходимости. Такова матрица с элементами  $a_{11}=5$ ,  $a_{22}=-0.7$ ,  $a_{12}=-4$ ,  $a_{21}=2$ , для которой собственные числа матрицы (1.2)  $\beta_1=-1.5$ ,  $\beta_2=0.8$ , а соб-

ственные числа матрицы (2.3) -  $Y_1 = -1.5$ ,  $Y_2 = 1.5$ . Применение методов Якоби и Зейделя и их модификаций дают расходящийся процесс, т.к. точка  $I$  принадлежит выпуклой оболочке спектра. Применение же метода ОСП к простой итерации с матрицей (5.3.1.2) дает быстро сходящийся ряд. Решение СЛАУ (5.1) с точностью  $\varepsilon = 10^{-5}$  достигается за  $m = 9$  итераций ряда (5.3.1.6).

Применение метода ОСП наиболее успешно в том случае, когда спектр оператора  $B$  в (1.1) локализован в небольшой окрестности с центром в т.  $k_0$  вдали от точки  $I$ . Тогда применение этого метода с оптимальным параметром  $k = -k_0$  является самым удачным среди одношаговых стационарных методов и приводит к быстро сходящемуся ряду простой итерации. В качестве примера рассмотрим СЛАУ с матрицей  $a_{11} = 7$ ,  $a_{22} = 3$ ,  $a_{12} = 4$ ,  $a_{21} = -1$ . В этом примере для матриц (1.2) и (2.3) имеем следующие собственные числа  $\beta_1 = -4$ ,  $\beta_2 = -4$  и  $Y_1 = -0.436i$ ,  $Y_2 = 0.436i$ . Значение

оптимального параметра  $k_0 = -\frac{\beta_1 + \beta_2}{2} = 4$  переводит в данном

случае точку  $\beta = 4$ , в которой находится весь спектр матрицы  $B$ , в точку  $\tilde{\beta} = 0$ , в которой находится спектр матрицы  $\tilde{B}$ . Таким образом, скорость сходимости ряда (5.3.1.6) с матрицей (5.3.3.1), (5.3.1.2) в данном случае очень высокая, т.к.  $\rho(\tilde{B}) = 0$ . Решение СЛАУ (1) с точностью до машинной константы достигается за  $m = 2$  итерации. Решение той же задачи методами Якоби и Зейделя требует гораздо большего количества итераций -  $m = 48$  и  $m = 23$  соответственно. Для метода Якоби применение ОСП не даст улучшения сходимости, т.к. центр спектра и так находится в

точке  $0$  и оптимальный параметр  $k_0 = -\frac{Y_1 + Y_2}{2} = 0$ . Для метода же

Зейделя спектр оператора (5.3.2.5) отличается от спектра матрицы (5.3.2.3) и использование метода Зейделя-ОСП с оптимальным параметром  $k_0 = 0.076$ , т.е. ряда (1.6) с оператором (5.3.3.1),



(5.3.2.5), приводит к уменьшению требуемого количества итераций –  $m = 15$ .

Пусть рассмотренная матрица продолжена на большую трехдиагональную матрицу с  $n = 100$  и такими же элементами, т.е. на главной диагонали чередуются значения 7 и 3, а на двух соседних соответственно 4 и  $-1$ . Спектр исходной матрицы существенно трансформируется из точки в протяженную область на комплексной плоскости, но при этом значение оптимального параметра, полученного по формуле (5.3.3.2) с участием минимального и максимального по модулю собственного числа матрицы  $B$  (5.3.1.2), остается неизменным  $k_0 = 4$ . Это справедливо для любой трехдиагональной матрицы, полученной таким периодическим продолжением из малой матрицы. Однако это значение  $k_0$  все же приближенное в силу того, что матрица не является положительно определенной и другие комплексные собственные числа выходят за пределы круга, натянутого на  $[\beta_{\min}, \beta_{\max}]$  как на диаметр. Опытным путем для сравнительно малых матриц с  $n = 10$  значение оптимального параметра можно уточнить до  $k_0 = 5.9$  и это значение остается практически неизменным для всех больших матриц такого вида. Для параметров  $k_0 = 4$  и  $k_0 = 5.9$  и точности решения  $\varepsilon = 10^{-5}$  получаем соответственно число требуемых итераций  $m = 129$  и  $m = 46$ . Впечатляющий результат для данной задачи приносит метод Зейделя-ОСП. Если для обычного метода Зейделя число итераций  $m = 190$ , то с применением ОСП при  $k_0 = 0.17$  число требуемых итераций снижается до  $m = 6$ !

Конечно, задача определения спектра матрицы в общем случае ничем не проще задачи решения СЛАУ прямыми методами. Однако, для ряда матриц приближенное значение оптимального параметра  $k_0$  для метода ОСП в применении к простой итерации (5.3.1.2), (5.3.1.3) находится весьма просто через её коэффициенты. Например, для большой трехдиагональной матрицы с двумя постоянными диагоналями возле главной и с чередующимися зна-

чениями  $a$  и  $b$  коэффициентов на главной диагонали. Для такой матрицы  $A$  в (5.1) значение оптимального параметра в (5.3.1.6) с (5.3.3.1), (5.3.1.2) равно  $k_0 = (a+b)/2 - 1$  и, если  $A$  – положительно определенная матрица, то это значение точное. Это не значит, что для любой матрицы такого типа можно построить сходящийся итерационный процесс, но если можно добиться сходимости, то при таком  $k = k_0$  метод сходится.

Кроме того, для физических и технических задач область локализации спектра оператора часто известна, т.к. она соответствует физически нерегулярным и резонансным решениям.

Преобразование оператора (5.3.3.1) можно использовать в условиях неполной информации об его спектре. Так, например, если известна в точности только одна граница вещественного спектра. Более определенно, пусть известно, что собственные числа  $\beta_v$  находятся на интервале  $[-M, m]$  и значение  $M \geq 1$  известно точно, а для  $m$  известно лишь, что  $m \in (0, 1)$ . Т.к. для данного случая  $\rho(\beta_v) \geq 1$ , то ряд простой итерации расходится, но в силу того, что  $1 \notin \Omega_B$  можно построить сходящийся ряд. Действительно, принимая  $k = M$ , получаем сходящийся ряд простой итерации для оператора  $\tilde{B}$ , спектр которого лежит на интервале  $\left[0, \frac{m+M}{1+M}\right]$ , причем

$\frac{m+M}{1+M} < 1$ , т.е.  $\rho(\tilde{\beta}_v) < 1$ . Можно показать также, что в условиях неопределенности данной задачи  $0 < m < 1$  лучший результат даст  $k = M / 2$ .

Если даже приходится детально исследовать спектр задачи для построения быстро сходящегося итерационного процесса то, однажды его построив, можно затем многократно использовать для расчетов с различными источниками - правыми частями  $\vec{b}$ .

Преимущества же быстро сходящихся итерационных процессов перед прямыми методами известны. Это:

- количество арифметических операций  $\approx m * n^2$  (здесь  $m$  - число итераций), вместо  $n^3$ ;
- отсутствие накопления ошибок в процессе итераций со сжимающим оператором;
- пониженные требования к оперативной памяти ЭВМ.

Особенно эти преимущества заметны для задач с большими матрицами  $n \geq 1000$ . Решение СЛАУ с  $n = 1000$  стандартным методом Mathcad на ЭВМ Р-2 750 МГц занимает около 3 мин машинного времени, в то время как решение той же системы быстро сходящимся итерационным методом с  $m \approx 10...20$  требует всего около 1...2 с.

#### 5.4. Нахождение собственных векторов и собственных значений матриц

Собственными векторами и собственными значениями матрицы  $A$  называются вектора и числа, удовлетворяющие соотношению:  $A\vec{x} = \lambda\vec{x}$ , причем собственный вектор определен с точностью до постоянного множителя.

В дальнейшем рассматриваются невырожденные матрицы, имеющие различные собственные значения. Для нахождения собственных значений необходимо решить уравнение:  $\det(A - E\lambda) = 0$ . Нахождение коэффициентов характеристического полинома

$$\det(A - E\lambda) = (-1)^n D(\lambda) = (-1)^n (\lambda^n + d_1\lambda^{n-1} + \dots + d_n)$$

непосредственным раскрытием определителя достаточно громоздко. В методе Крылова используется то, что подстановка в характеристический полином вместо переменной матрицы  $A$  дает в результате нулевую матрицу:  $D(A) = O$ . Это тождество помножается слева на произвольный вектор  $\vec{y}_0$ :

$$\vec{y}_n + d_1\vec{y}_{n-1} + \dots + d_n\vec{y}_0 = 0, \text{ где } \vec{y}_k = A^k \vec{y}_0,$$

т. е. получается СЛАУ относительно коэффициентов характеристического полинома  $d_k$ . Для определения собственных векторов вводится система полиномов

$$R_i(\lambda) = \frac{D(\lambda)}{\lambda - \lambda_i} = \lambda^{n-1} + r_{1,i}\lambda^{n-2} + \dots + r_{n-1,i},$$

$$R_i(\lambda_j) = 0, \text{ если } i \neq j.$$

Учитывается, что собственные вектора  $\vec{x}_k$  линейно независимые, то есть любой вектор можно представить в виде их линейной комбинации:

$$\vec{y}_0 = c_1\vec{x}_1 + c_2\vec{x}_2 + \dots c_n\vec{x}_n.$$

Собственный вектор является линейной комбинацией векторов  $\vec{y}_k$  и коэффициентов полинома  $R_i(\lambda)$ . Действительно:

$$\begin{aligned} & \vec{y}_{n-1} + r_{1,i}\vec{y}_{n-2} + \dots + r_{n-1,i}\vec{y}_0 = \\ & = c_1\lambda_1^{n-1}\vec{x}_1 + c_2\lambda_2^{n-1}\vec{x}_2 + \dots + c_n\lambda_n^{n-1}\vec{x}_n + \\ & + r_{1,i}(c_1\lambda_1^{n-2}\vec{x}_1 + c_2\lambda_2^{n-2}\vec{x}_2 + \dots + c_n\lambda_n^{n-2}\vec{x}_n) + \\ & \dots \\ & + r_{n-1,i}(c_1\vec{x}_1 + c_2\vec{x}_2 + \dots + c_n\vec{x}_n). \end{aligned}$$

Коэффициенты при собственных векторах представляют собой  $c_j R_i(\lambda_j)$ , которые все равны нулю кроме коэффициента с  $i = j$ , стоящего перед  $\vec{x}_i$ . То есть данная линейная комбинация является собственным вектором.

## 5.5. Примеры и задания к теме

### 5.5.1. Прямые методы решения СЛАУ

*Пример, метод Гаусса.*

Последовательно выбираются ведущие элементы. Преобразованная с помощью правила прямоугольника матрица записывается в следующую расширенную матрицу, подчеркнуты ведущие элементы:

$$\begin{aligned} \left[ \begin{array}{ccc|c} 1 & 1 & -1 & 0 \\ 2 & 1 & 1 & 7 \\ 1 & -1 & 1 & 2 \end{array} \right] &\Rightarrow \left[ \begin{array}{ccc|c} 1 & 1 & -1 & 0 \\ 0 & -1 & 3 & 7 \\ 0 & -2 & 2 & 2 \end{array} \right] \Rightarrow \\ &\Rightarrow \left[ \begin{array}{ccc|c} 1 & 1 & -1 & 0 \\ 0 & 1 & -3 & -7 \\ 0 & 0 & -4 & -12 \end{array} \right] \Rightarrow \left[ \begin{array}{ccc|c} 1 & 1 & -1 & 0 \\ 0 & 1 & -3 & -7 \\ 0 & 0 & 1 & 3 \end{array} \right]. \end{aligned}$$

После чего с помощью обратного хода находятся компоненты вектора:  $x_3 = 3$ ,  $x_2 = -7 - (-3)3 = 2$ ,  $x_1 = 0 - 2 + 3 = 1$ .

*Метод ортогонализации:*

$$\left[ \begin{array}{ccc|c} 1 & 1 & -1 & 0 \\ 2 & 1 & 1 & 7 \\ 1 & -1 & 1 & 2 \end{array} \right]$$

СЛАУ записывается в векторном представлении и выбирается первый вектор ортогональной матрицы  $R$ :

$$\vec{a}_1 x_1 + \vec{a}_2 x_2 + \vec{a}_3 x_3 = \vec{b},$$

$$\vec{r}_1 = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} = \vec{a}_1.$$

Вектор  $\vec{a}_2$  записывается в виде линейной комбинации двух ортогональных векторов, умножается скалярно на  $\vec{r}_1$  и определяется коэффициент  $t_{12}$ :

$$\vec{r}_1 \cdot \vec{a}_2 = t_{12} \vec{r}_1 + \vec{r}_2,$$

$$t_{12} = \frac{(\vec{r}_1, \vec{a}_2)}{r_1^2} = \frac{1+2-1}{1+4+1} = \frac{1}{3}.$$

Далее вычисляется  $\vec{r}_2$ :

$$\vec{r}_2 = \vec{a}_2 - t_{12} \vec{r}_1 = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} - \frac{1}{3} \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{2}{3} \\ \frac{1}{3} \\ -\frac{4}{3} \end{pmatrix}.$$

Вычисляется вектор  $\vec{r}_3$ :

$$\vec{a}_3 = t_{13}\vec{r}_1 + t_{23}\vec{r}_2 + \vec{r}_3,$$

$$t_{13} = \frac{(\vec{r}_1, \vec{a}_3)}{r_1^2} = \frac{-1+2+1}{6} = \frac{1}{3}, \quad t_{23} = \frac{(\vec{r}_2, \vec{a}_3)}{r_2^2} = \frac{-\frac{2}{3} + \frac{1}{3} - \frac{4}{3}}{\frac{4}{9} + \frac{1}{9} + \frac{16}{9}} = -\frac{5}{7};$$

$$\vec{r}_3 = \vec{a}_3 - t_{13}\vec{r}_1 - t_{23}\vec{r}_2 = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix} - \frac{1}{3} \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} + \frac{5}{7} \begin{pmatrix} \frac{2}{3} \\ \frac{1}{3} \\ -\frac{4}{3} \end{pmatrix} = \begin{pmatrix} -\frac{6}{7} \\ \frac{4}{7} \\ -\frac{2}{7} \end{pmatrix};$$

$$A = RT = \begin{pmatrix} 1 & \frac{2}{3} & -\frac{6}{7} \\ 2 & \frac{1}{3} & \frac{4}{7} \\ 1 & -\frac{4}{3} & -\frac{2}{7} \end{pmatrix} \cdot \begin{pmatrix} 1 & \frac{1}{3} & \frac{1}{3} \\ 0 & 1 & -\frac{5}{7} \\ 0 & 0 & 1 \end{pmatrix}.$$

Вычисление компонент вектора  $\vec{x}$ :

$$\vec{a}_1 \cdot x_1 + \vec{a}_2 \cdot x_2 + \vec{a}_3 \cdot x_3 = \vec{b}, \quad x_3 = \frac{(\vec{r}_3, \vec{b})}{(\vec{r}_3, \vec{a}_3)} = \frac{0+4-\frac{4}{7}}{\frac{6}{7} + \frac{4}{7} - \frac{2}{7}} = 3;$$

$$\vec{a}_1 x_1 + \vec{a}_2 x_2 = \begin{pmatrix} 0 \\ 7 \\ 2 \end{pmatrix} - 3 \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \\ -1 \end{pmatrix} = \vec{b}^{(1)},$$

$$x_2 = \frac{(\vec{r}_2, \vec{b}^{(1)})}{(\vec{r}_2, \vec{a}_2)} = \frac{2 + \frac{4}{3} + \frac{4}{3}}{\frac{2}{3} + \frac{1}{3} + \frac{4}{3}} = 2;$$

$$\vec{a}_1 x_1 = \begin{pmatrix} 3 \\ 4 \\ -1 \end{pmatrix} - 2 \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} = \vec{b}^{(2)}, \quad x_1 = \frac{(\vec{r}_1, \vec{b}^{(2)})}{(\vec{r}_1, \vec{a}_1)} = 1.$$

*Варианты заданий.*

Найти решение СЛАУ методом Гаусса и методом ортогонализации:

$$1) \begin{pmatrix} 9 & 8 & 3 & | & 4 \\ 3 & 6 & 3 & | & -3 \\ 5 & 3 & 1 & | & 2 \end{pmatrix}; \quad 2) \begin{pmatrix} 1 & 2 & 1 & | & 8 \\ 5 & -4 & -7 & | & 0 \\ 2 & 1 & -1 & | & 7 \end{pmatrix}; \quad 3) \begin{pmatrix} 2 & 1 & 7 & | & 2 \\ 5 & 2 & 1 & | & -11 \\ 1 & 1 & 6 & | & 3 \end{pmatrix};$$

$$4) \begin{pmatrix} 1 & 2 & 8 & | & -5 \\ 7 & 1 & 6 & | & -11 \\ 1 & 7 & 2 & | & 11 \end{pmatrix}; \quad 5) \begin{pmatrix} 5 & 5 & 5 & | & 0 \\ 7 & 7 & 3 & | & -4 \\ 4 & 1 & 3 & | & -1 \end{pmatrix}; \quad 6) \begin{pmatrix} 4 & 1 & 1 & | & -1 \\ 2 & 1 & 4 & | & 4 \\ 1 & 1 & 5 & | & 6 \end{pmatrix};$$

$$7) \begin{pmatrix} 1 & 4 & -3 & | & 3 \\ 4 & 2 & 1 & | & 11 \\ 1 & -1 & 0 & | & 1 \end{pmatrix}; \quad 8) \begin{pmatrix} 8 & 8 & 1 & | & 2 \\ 0 & 2 & 4 & | & 6 \\ 1 & 1 & 0 & | & 0 \end{pmatrix}; \quad 9) \begin{pmatrix} 5 & 5 & -1 & | & -2 \\ 1 & 1 & 1 & | & 2 \\ 1 & 0 & -2 & | & -3 \end{pmatrix};$$

$$10) \begin{pmatrix} 1 & 1 & 6 & | & 5 \\ 3 & 5 & 2 & | & 3 \\ 5 & 0 & 1 & | & -14 \end{pmatrix}; \quad 11) \begin{pmatrix} 1 & 8 & 3 & | & 14 \\ 5 & 1 & 7 & | & 0 \\ 5 & 4 & 2 & | & 11 \end{pmatrix}; \quad 12) \begin{pmatrix} 1 & 9 & 3 & | & -7 \\ 4 & 3 & 1 & | & 5 \\ 0 & 3 & 0 & | & 0 \end{pmatrix};$$

$$13) \begin{pmatrix} 1 & 1 & 4 & | & 3 \\ 5 & 4 & 0 & | & -3 \\ 2 & 0 & 1 & | & 3 \end{pmatrix}; \quad 14) \begin{pmatrix} 8 & 8 & 4 & | & 4 \\ 0 & 2 & 0 & | & -4 \\ 1 & 3 & -1 & | & -5 \end{pmatrix}; \quad 15) \begin{pmatrix} 4 & 5 & 4 & | & 2 \\ 2 & 3 & 2 & | & 2 \\ 6 & 4 & 2 & | & 0 \end{pmatrix};$$

$$16) \begin{pmatrix} 1 & 1 & 1 & | & -1 \\ 5 & 4 & 3 & | & -2 \\ 1 & 2 & 0 & | & -4 \end{pmatrix}; \quad 17) \begin{pmatrix} 1 & 5 & 7 & | & 3 \\ 4 & 5 & 7 & | & 3 \\ 0 & 1 & 1 & | & 1 \end{pmatrix}; \quad 18) \begin{pmatrix} 6 & 3 & 1 & | & 4 \\ 4 & 5 & 3 & | & -4 \\ 0 & 3 & 4 & | & -5 \end{pmatrix};$$

$$19) \begin{pmatrix} 3 & 2 & 6 & | & 4 \\ 1 & 4 & 0 & | & 10 \\ 0 & 1 & 4 & | & -2 \end{pmatrix}; \quad 20) \begin{pmatrix} 1 & 1 & 1 & | & 1 \\ 3 & 3 & 4 & | & 4 \\ 3 & 0 & 1 & | & -2 \end{pmatrix}.$$

### 5.5.2. Итерационные методы решения СЛАУ

Найти решение СЛАУ с матрицей  $A = \{a_{i,j}\}$ ,  $i, j = 1..n$  и правой частью  $\vec{b}$  итерационными методами Якоби, Зейделя и ОСП. Решение получить с заданной относительной точностью  $\varepsilon$ . Указать количество итераций  $m_{iter}$  необходимых каждому методу для достижения заданной точности. На практических занятиях с использованием калькуляторов в случаях слабой сходимости ограничиться числом  $m_{iter} = 10..15$ , отразив это в результатах. На лабораторных работах с использованием пакета Mathcad это ограничение снимается. Отметить также случаи явной расходимости метода.

В качестве оптимального параметра  $k$  для сходимости метода ОСП в задаче с матрицей (5.3.1.2) размером  $n * n$  следует принимать:

- $k = \frac{a_{1,1} + a_{2,2}}{2} - 1$ , в случае  $n = 2$  (и в случае периодического продолжения на трехдиагональную матрицу с большим значением  $n$ )
- $k = d - 1$ , для задач с трехдиагональной матрицей с постоянной главной диагональю, где  $d = a_{i,i}, i = 1, 2, \dots, n$ . В частности, для матриц с  $n = 3$ .

В случаях расходимости всех трех используемых методов следует применить комбинированный метод Якоби-Зейделя и ОСП. Для этого потребуются найти собственные числа матрицы Якоби (5.3.2.3) и на этой основе сделать вывод о значении  $k$  для ряда оптимальной простой итерации с матрицей (5.3.3.1), в которой  $B$ -матрица (5.3.2.3). Этот же оптимальный параметр можно использовать для построения ряда простой итерации с оператором (5.3.3.1), где  $B$ -оператор Зейделя (5.3.2.5). Однако, в последнем случае оптимальный параметр, как правило, может быть значительно улучшен и в необходимых случаях он указан.



## Варианты заданий.

№ вар.	$A$	$\vec{b}$	$\varepsilon$	Оптимальный па- раметр для мето- да Зейделя-ОСП
1.	$\begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 1 \end{pmatrix}$	$10^{-10}$	
2.	$\begin{pmatrix} 3 & -1 \\ 1 & 5 \end{pmatrix}$	$\begin{pmatrix} 2 \\ 6 \end{pmatrix}$	$10^{-8}$	
3.	$\begin{pmatrix} 7 & -1 \\ 4 & 3 \end{pmatrix}$	$\begin{pmatrix} 6 \\ 7 \end{pmatrix}$	$10^{-8}$	$k=0.1$
4.	$\begin{pmatrix} 15 & -1 \\ 12 & 8 \end{pmatrix}$	$\begin{pmatrix} 16 \\ 4 \end{pmatrix}$	$10^{-8}$	$k=0.1$
5.	$\begin{pmatrix} 14 & -2 \\ 7 & 6 \end{pmatrix}$	$\begin{pmatrix} 10 \\ 19 \end{pmatrix}$	$10^{-5}$	$k=0.2$
6. *	$\begin{pmatrix} 5 & 2 \\ -4 & -0.7 \end{pmatrix}$	$\begin{pmatrix} 3 \\ -3.3 \end{pmatrix}$	$10^{-5}$	
7.	$\begin{pmatrix} 14 & -2 \\ 7 & 1 \end{pmatrix}$	$\begin{pmatrix} 12 \\ 8 \end{pmatrix}$	$10^{-5}$	$k=0.25$ ( $n=2$ ), $k=0.65$ ( $n=100$ )
8.	$\begin{pmatrix} 2 & 3 \\ 7 & -2 \end{pmatrix}$	$\begin{pmatrix} 5 \\ 5 \end{pmatrix}$	$10^{-3}$	$k=1$ ( $n=2$ ), $k=1.85$ ( $n=100$ )
9.	$\begin{pmatrix} 14 & -2 \\ 10 & 1 \end{pmatrix}$	$\begin{pmatrix} 12 \\ 11 \end{pmatrix}$	$10^{-5}$	$k=0.3$ ( $n=2$ ), $k=0.8$ ( $n=100$ )
10.	$\begin{pmatrix} 14 & 1 \\ -4 & -2 \end{pmatrix}$	$\begin{pmatrix} 15 \\ -6 \end{pmatrix}$	$10^{-8}$	$k=-0.17$
11.	$\begin{pmatrix} 15 & -1 \\ 3 & 8 \end{pmatrix}$	$\begin{pmatrix} 14 \\ 11 \end{pmatrix}$	$10^{-8}$	
12.	$\begin{pmatrix} 14 & -1 \\ 3 & 6 \end{pmatrix}$	$\begin{pmatrix} 13 \\ 9 \end{pmatrix}$	$10^{-5}$	

13	$\begin{pmatrix} 1 & 10 & 10 \\ 0 & 1 & 0 \\ 0 & 10 & 1 \end{pmatrix}$	$\begin{pmatrix} 21 \\ 1 \\ 11 \end{pmatrix}$	$10^{-15}$	
14. *	$\begin{pmatrix} 1 & 10 & 0 \\ 10 & 1 & -10 \\ 0 & 10 & 1 \end{pmatrix}$	$\begin{pmatrix} 11 \\ 1 \\ 11 \end{pmatrix}$	$10^{-10}$	
15 *	$\begin{pmatrix} 5 & -10 & 0 \\ 10 & 5 & 10 \\ 0 & 10 & 5 \end{pmatrix}$	$\begin{pmatrix} -5 \\ 25 \\ 15 \end{pmatrix}$	$10^{-10}$	

При выполнении лабораторных работ с помощью пакета Mathcad указанные варианты видоизменяются до больших трех-диагональных матриц с  $n=10$  и  $n=100$ . Для этого главная и две побочные диагонали периодически продолжаются на большую матрицу. Остальные коэффициенты матрицы нулевые. Так, матрица варианта № 2 выглядит следующим образом:

$$A = \left( \begin{array}{cccc|c} 3 & -1 & 0 & 0 & 0 \\ 1 & 5 & -1 & 0 & 0 \\ 0 & 1 & 3 & -1 & 0 \\ 0 & 0 & 1 & 5 & -1 \\ \hline 0 & 0 & 0 & 1 & 5 \end{array} \right).$$

Оптимальный параметр  $k$  для метода ОСП остаётся при этом неизменным (несмотря на трансформацию спектра матрицы – увеличение радиуса круга  $\Omega_B$  при неизменном положении его центра) и определяется так, как указано выше для малых матриц.

Решить задачу для различных векторов правой части:

1).  $b_i = 0, b_1 = 1$ ; 2).  $b_i = 0, b_n = 1$ ; 3).  $b_i = 1$  для всех  $i = 0, 1, 2, \dots$ .  
Относительная точность вычислений для всех вариантов  $\varepsilon = 10^{-15}$ .

Варианты помеченные \* используются только для практических работ с  $n = 2$  и для  $n = 100$  решений в виде сходящегося

итерационного процесса не имеют. В этом случае выбирается вариант с номером, равным 16 – № вар.

В результате работы представить для каждого метода:

- вектор решения (несколько первых компонент)
- число итераций
- невязку решения (по норме)
- спектр матрицы (с помощью стандартных функций Mathcad)

Сравнить значение оптимального параметра, полученного исходя из знания спектра оператора, и найденного по вышеизложенным правилам.

Сравнить решение СЛАУ, полученное стандартным методом Mathcad и Вашим итерационным методом.

Сделать выводы о причинах хорошей (плохой) сходимости итерационного метода и о её зависимости от начального приближения (правой части  $\vec{b}$ ).

Найти число обусловленности исходной матрицы (с помощью стандартных функций Mathcad).

Для вариантов задания с быстрой сходимостью ( $m < \approx 10..20$ ) сравнить при  $n = 1000$  время решения СЛАУ стандартным методом Mathcad и итерационным методом.

### 5.5.3. Нахождение собственных значений и векторов

*Пример:* Найти собственные вектора и значения матрицы:

$$A = \begin{bmatrix} 4 & 2 \\ 0 & 3 \end{bmatrix}.$$

Выберем  $\vec{y}_0 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ , тогда:

$$\vec{y}_1 = A\vec{y}_0 = \begin{bmatrix} 4 & 2 \\ 0 & 3 \end{bmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \quad \vec{y}_2 = A\vec{y}_1 = \begin{bmatrix} 4 & 2 \\ 0 & 3 \end{bmatrix} \begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 14 \\ 9 \end{pmatrix}.$$

СЛАУ имеет следующий вид:

$$d_1 \begin{pmatrix} 2 \\ 3 \end{pmatrix} + d_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = - \begin{pmatrix} 14 \\ 9 \end{pmatrix},$$

откуда  $d_1 = -7$ ,  $d_2 = 12$  в результате получаем характеристический полином

$$\lambda^2 - 7\lambda + 12 = 0,$$

откуда  $\lambda_1 = 3$ ,  $\lambda_2 = 4$ .

Построим полиномы для нахождения собственных векторов:

$$R_1(\lambda) = \lambda - 4, \quad R_2(\lambda) = \lambda - 3,$$

Поэтому:

$$\vec{x}_1 = \vec{y}_1 - 4\vec{y}_0 = \begin{pmatrix} 2 \\ -1 \end{pmatrix}, \quad \vec{x}_2 = \vec{y}_1 - 3\vec{y}_0 = \begin{pmatrix} 2 \\ 0 \end{pmatrix} \propto \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \text{ так как собствен-}$$

ный вектор определен с точностью до произвольного множителя.

Проверка:

$$A\vec{x}_1 = \begin{pmatrix} 4 & 2 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 2 \\ -1 \end{pmatrix} = \begin{pmatrix} 6 \\ -3 \end{pmatrix} = 3 \begin{pmatrix} 2 \\ -1 \end{pmatrix} = 3\vec{x}_1,$$

$$A\vec{x}_2 = \begin{pmatrix} 4 & 2 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 4 \\ 0 \end{pmatrix} = 4 \begin{pmatrix} 1 \\ 0 \end{pmatrix} = 4\vec{x}_2.$$

*Задание для практических занятий.*

Дана матрица  $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ . Найти ее собственные значения и вектора.

Варианты:

- 1)  $a = 2, b = 1, c = 1, d = 2$ , 2)  $a = 1, b = 0, c = 1, d = 2$ ,
- 3)  $a = -1, b = 2, c = 0, d = 1$ , 4)  $a = -1, b = 2, c = 0.5, d = -1$ ,
- 5)  $a = -1, b = 3, c = 0, d = 2$ , 6)  $a = -1, b = 2, c = 2, d = -1$ ,
- 7)  $a = 1, b = 4, c = 1, d = 1$ , 8)  $a = -2, b = 4, c = 0, d = 1$ ,
- 9)  $a = 3, b = 2, c = 0, d = -1$ , 10)  $a = -2, b = 2, c = 1, d = -2$ ,
- 11)  $a = 1, b = 2, c = 0, d = -2$ , 12)  $a = 1, b = 0, c = 3, d = -1$ ,
- 13)  $a = -2, b = 2, c = 0, d = -3$ , 14)  $a = -2, b = 2, c = 0.5, d = -2$ ,
- 15)  $a = 2, b = 3, c = 3, d = 2$ , 16)  $a = 1, b = 3, c = 3, d = 1$ .

## 6. ЧИСЛЕННЫЕ МЕТОДЫ РЕШЕНИЯ ОБЫКНОВЕННЫХ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ

Рассматриваются уравнения первого порядка, разрешимые относительно первой производной с начальными условиями  $(x_0, y_0)$ :

$$\frac{dy}{dx} = f(x, y) . \quad (6.1)$$

Существует теорема Коши о единственности решения дифференциального уравнения при заданных начальных условиях  $(x_0, y_0)$ . Геометрически  $f(x, y)$  определяет поле направлений на плоскости  $(x, y)$ , а решения обыкновенного дифференциального уравнения (ОДУ) - интегральные кривые.

Численные методы решения задачи Коши для ОДУ основаны на том, что решение можно представить в виде разложения в ряд Тейлора с любой степенью точности.

### 6.1. Метод разложения в ряд Тейлора

Решение ищется в виде

$$y(x) = y(x_0) + y'(x_0)(x - x_0) + \frac{y''(x_0)}{2!}(x - x_0)^2 + \dots . \quad (6.1.1)$$

Функциональные зависимости  $y^{(k)}(x)$  известны:

$$\begin{aligned} y'(x) &= f(x, y), \\ y''(x) &= \frac{\partial f}{\partial x} + \frac{\partial f \cdot \partial y}{\partial y \cdot \partial x} = f_x + f_y \cdot f, \\ y'''(x) &= (f_{xx} + 2ff_{xy} + f^2 f_{yy} + f_y(f_x + ff_y)) \text{ и т.д.} \end{aligned} \quad (6.1.2)$$

Этот метод приводит к громоздким выражениям для производных, и в основном используются для получения других численных методов.

## 6.2. Общая схема метода Рунге—Кутты

Одношаговые методы позволяют получить заданную точность используя только предыдущее значение  $y(x)$ . Изменение  $y(x)$  на шаге  $h$  представляется в виде квадратурной формулы (типа Гаусса):

$$\Delta y = y(x+h) - y(x) = \int_x^{x+h} f(x, y) dx \approx \sum_{i=0}^n A_i \phi_i,$$

где  $\phi_i = hf(x + \alpha_i h, y + \beta_{i,0}\phi_0 + \beta_{i,1}\phi_1 + \dots + \beta_{i,i-1}\phi_{i-1})$ .

Для получения коэффициентов  $A_i$ ,  $\alpha_i$  и  $\beta_i$  квадратурная сумма разлагается в ряд по степеням  $h$ . Полученное разложение сравнивается с рядом Тейлора:

$$\Delta y(x) = y'(x)h + \frac{1}{2}y''(x)h^2 + \dots \quad (6.2.1)$$

В общем виде выражения для коэффициентов получить трудно, поэтому рассмотрим наиболее употребительные формулы.

Введем обозначения:

$$\begin{aligned} \phi_0(x) &= hf(x, y), \\ \phi_1(x) &= hf(x + \alpha_1 h, y + \beta_{1,0}\phi_0), \\ \phi_2(x) &= hf(x + \alpha_2 h, y + \beta_{2,0}\phi_0 + \beta_{2,1}\phi_1), \\ &\dots\dots\dots \\ \phi_i(x) &= hf(x + \alpha_i h, y + \beta_{i,0}\phi_0 + \dots + \beta_{i,i-1}\phi_{i-1}). \end{aligned} \quad (6.2.2)$$

Квадратурную формулу разлагаем в ряд по  $h$ :

$$\begin{aligned} \Delta y &= h(A_0 + A_1)f + h^2 A_1(\alpha_1 f_x + \beta_{1,0} f_y f) + \\ &+ h^3 / 2 A_1(\alpha_1^2 f_{xx} + 2\alpha_1 \beta_{1,0} f_{xy} f + \beta_{1,0}^2 f_{yy} f) + O(h^4), \end{aligned} \quad (6.2.3)$$

$f_x, f_y, f_{xx}, f_{yy}$  — частные производные по  $x$  и  $y$   $f(x, y)$ .

Полученное разложение сравнивается с рядом Тейлора (6.1.1).

Рассмотрим несколько частных случаев.

### 6.3 Методы Рунге–Кутты низших порядков

#### 6.3.1. Метод Эйлера

Иначе его называют методом ломаных и в силу простоты часто используют.

В квадратурной формуле ограничиваемся одним слагаемым:

$$\Delta y = y(x+h) - y(x) = A_0 \phi_0 = hf(x, y). \quad (6.3.1.1)$$

Интегральная кривая заменяется ломаной линией, состоящей из прямолинейных отрезков. Выбирается шаг  $h$  и значение функции в точке  $x = x+h$  ищется по формуле  $y(x+h) = y + f(x, y)h$ , т.е. в интегральном уравнении  $f(x, y)$  заменяется на константу.

Ошибки метода  $\propto h^2$ , так как в ряде Тейлора отбрасываются вторые производные.

#### 6.3.2. Метод трапеций и прямоугольника

Это популярные методы, иначе их называют метод Коши–Эйлера и модифицированный метод Эйлера, их ошибка  $\approx h^3$

Представление  $\Delta y = A_0 \phi_0 + A_1 \phi_1$  позволяет сравнить два первых слагаемых в разложении с рядом Тейлора:

$$A_0 + A_1 = 1, \quad \alpha_1 A_1 = \frac{1}{2}, \quad \beta_{10} A_1 = \frac{1}{2}.$$

Получены три уравнения для четырех неизвестных, что является общим свойством метода Рунге–Кутты. То есть для каждого порядка точности существует множество вычислительных схем:

$$\alpha_1 = \beta_{1,0} = \frac{1}{2A_1}, \quad A_0 = 1 - A_1.$$

Положим  $A_1 = 1/2$  (метод трапеций), тогда

$$\Delta y = \frac{1}{2}(\phi_0 + \phi_1) = \frac{h}{2}(f(x, y) + f(x+h, y+hf(x, y))), \quad (6.3.2.1)$$

то есть значение производной «подправляется» значением в предварительно определенной точке.

В методе прямоугольников  $A_1 = 1$ , тогда  $A_0 = 0$ ,  $\alpha = \beta = \frac{1}{2}$ .  
В этом случае

$$\Delta y = hf(x + h/2, y + hf(x, y)/2). \quad (6.3.2.2)$$

#### 6.4. Методы Рунге–Кутты высших порядков

В методе Рунге–Кутты третьего порядка точности:

$$\Delta y = A_0\phi_0 + A_1\phi_1 + A_2\phi_2.$$

Разлагая в ряд по  $h$  до  $h^3$  и сравнивая с рядом Тейлора (6.1.1), получим следующую систему из шести уравнений для восьми неизвестных:

$$\left. \begin{aligned} A_0 + A_1 + A_2 &= 1 \\ A_1 * \alpha_1 + A_2 * \alpha_2 &= 1/2 \\ A_1 * \alpha_1^2 + A_2 * \alpha_2^2 &= 1/3 \\ A_2 * \alpha_1 * \beta_{21} &= 1/6 \\ \beta_{20} + \beta_{21} &= \alpha_2 \\ \beta_{10} &= \alpha_1 \end{aligned} \right\}$$

Наиболее употребительна в этом случае симметричная разностная схема (аналог метода парабол при численном интегрировании):  $A_0 = A_2 = 1/6$ , тогда:

$$A_1 = 4/6, \alpha_1 = 1/2, \alpha_2 = 1, \beta_{10} = 1/2, \beta_{20} = -1, \beta_{21} = 2.$$

$$\Delta y = \frac{1}{6}(\phi_0 + 4\phi_1 + \phi_2),$$

$$\phi_0 = hf(x, y),$$

$$\phi_1 = hf(x + h/2, y + \phi_0/2),$$

$$\phi_2 = hf(x + h, y - \phi_0 + 2\phi_1).$$

В методе Рунге–Кутты точности порядка  $h^4$  получается система из 11 уравнений для 13 неизвестных.

Наиболее употребительны две вычислительные схемы:

1. Аналог метода 3/8 в численном интегрировании:

$$\Delta y = \frac{1}{8}(\phi_0 + 3\phi_1 + 3\phi_2 + \phi_3),$$



где

$$\begin{aligned}\phi_0 &= hf(x, y), \\ \phi_1 &= hf(x + h/3, y + \phi_0/3), \\ \phi_2 &= hf(x + 2h/3, y - \phi_0/3 + \phi_1), \\ \phi_3 &= hf(x + h, y + \phi_0 - \phi_1 + \phi_2).\end{aligned}$$

2. Аналог метода парабол:

$$\Delta y = \frac{1}{6}(\phi_0 + 2\phi_1 + 2\phi_2 + \phi_3), \quad (6.4.1)$$

где  $\phi_0 = hf(x, y)$ ,

$$\begin{aligned}\phi_1 &= hf(x + h/2, y + \phi_0/2), \\ \phi_2 &= hf(x + h/2, y + \phi_1/2), \\ \phi_3 &= hf(x + h, y + \phi_2).\end{aligned}$$

Проблема выбора той или иной вычислительной схемы при заданной точности зависит от вида  $f(x, y)$ , так как от этого зависит величина остаточного члена.

### 6.5. Задание к теме и пример решения ОДУ

Найти решение задачи Коши для ОДУ:

$y' = x^2 + \frac{(K-1)}{2}y$ ,  $y(0) = L$  на интервале  $[0, 2]$ .  $K$  и  $L$  параметры из табл. 4.3

Решить пятью методами:

1. Метод вариации постоянных (точное решение).
2. Разложение в ряд Тейлора до четвертого порядка.
3. Метод Эйлера (6.3.1.1).
4. Метод трапеций (Коши–Эйлера) (6.3.2.1).
5. Метод Рунге–Кутты (6.4.1).

Построить графики и сравнить точность различных методов, шаг  $h = 0,5$ .

*Пример.*  $K=3$ ,  $L=2$ .  $y' = x^2 + y$ ,  $y(0) = 2$ .

1. В методе вариации постоянных решение ищется в виде  $y = C(x)y_{\text{одн}}$ . Однородное уравнение  $y' - y = 0$  имеет очевидное

решение  $y_{\text{одн}} = Ce^x$ . Подстановка в неоднородное уравнение дает уравнение для коэффициента:  $C'(x) = x^2 e^{-x}$ . После интегрирования и подстановки начального условия получим:  $y(x) = 4e^x - x^2 - 2x - 2$ .

2. Разложение в ряд Тейлора проводится в точке  $x = 0$ . Все производные в этой точке известны:  $y'(0) = y(0) = 2$ ;

$$y''(x) = 2x + y'(x), \quad y''(0) = y'(0) = 2;$$

$$y'''(x) = 2 + y''(x), \quad y'''(0) = 4;$$

$$y^{(4)}(x) = y'''(x), \quad y^{(4)}(0) = 4.$$

$$\begin{aligned} y_T(x) &= y(0) + y'(0)x + \frac{y''(0)}{2!}x^2 + \frac{y'''(0)}{3!}x^3 + \frac{y^{(4)}(0)}{4!}x^4 = \\ &= 2 + 2x + x^2 + \frac{2}{3}x^3 + \frac{1}{6}x^4. \end{aligned}$$

3. Метод Эйлера. Расчет ведется по формуле (6.3.1.1):

4. Метод Коши-Эйлера (метод трапеций). Вначале рассчитывается значение  $\tilde{y}_{k+1} = y_k + hf(x_k, y_k)$ , которое затем используется в окончательном выражении (6.3.2.1).

5. Метод Рунге-Кутты. Последовательно вычисляются значения производной в промежуточных точках и используются в окончательном выражении с заданными весами (6.4.1).

Результаты расчетов представлены в таблице:

$x$	0	0,5	1	1,5	2
Точное решение	2	3,34488	5,87313	10,6768	19,5562
Ряд Тейлора	2	3,3438	5,83333	10,3438	18
Метод Эйлера	2	3	4,625	7,4375	12,2812
Метод Коши-Эйлера	2	3,3125	5,72656	10,2432	18,4889
Метод Рунге-Кутты	2	3,34440	5,87111	10,6710	19,5423

## 7. ЧИСЛЕННОЕ РЕШЕНИЕ НАЧАЛЬНО-КРАЕВЫХ ЗАДАЧ ДЛЯ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ В ЧАСТНЫХ ПРОИЗВОДНЫХ

Рассматриваются простейшие уравнения математической физики гиперболического, параболического и эллиптического типов с начально-краевыми условиями для дифференциальных уравнений в частных производных (ДУЧП). Задачи такого типа возникают в физике, технике, других прикладных науках.

Для численного решения начально-краевых задач для ДУЧП используется метод конечных разностей, основанный на приближенных формулах для первой и второй производной функций. При этом начально-краевая задача заменяется на сеточные уравнения, связывающие значения искомой функции в узлах сетки.

### 7.1. Конечные разности

Область решения на плоскости двух переменных, например  $(t, x)$ , разбивается на дискретную сетку из узлов  $(t_i, x_j)$ ,  $i \in Z_1, j \in Z_2$ ,  $Z_1, Z_2$  – подмножества целых чисел. Например, в прямоугольнике  $0 \leq t \leq T, 0 \leq x \leq X$  узлы сетки

$$t_i = i \cdot \Delta t, \quad x_j = j \cdot \Delta x, \quad 0 \leq i \leq M, 0 \leq j \leq N, \quad (7.1.1)$$

где  $\Delta t = T/M, \Delta x = X/N$  – шаги сетки по координатам  $t$  и  $x$  соответственно;  $M, N$  – целые числа.

Неизвестная функция  $u(t, x)$ , участвующая в краевой задаче, заменяется искомой сеточной функцией  $u_{i,j} = u(t_i, x_j)$  на узлах сетки. Частные производные по координатам заменяются соответствующими конечными разностями, которые могут быть различного порядка точности по шагу сетки вдоль координаты. Пусть  $h$  – шаг сетки вдоль рассматриваемой координаты,  $f_0$  – значение функции в рассматриваемой точке  $(i, j)$ ,  $f_1, f_2, f_{-1}, f_{-2}$  – последующие и предыдущие значения сеточной функции по данной координате. Тогда первая производная по этой координате

может быть заменена правой или левой конечной разностью порядка  $O(h)$ :

$$f'_{0,+} = (f_1 - f_0) / h, \quad f'_{0,-} = (f_0 - f_{-1}) / h \quad (7.1.2)$$

или центральной конечной разностью порядка  $O(h^2)$ :

$$f'_0 = (f_1 - f_{-1}) / (2h). \quad (7.1.3)$$

Можно записать также для первой производной конечную разность порядка  $O(h^4)$ :

$$f'_0 = (-f_2 + 8f_1 - 8f_{-1} + f_{-2}) / (12h).$$

Центральные конечные разности для второй производной порядка  $O(h^2)$  и  $O(h^4)$  выглядят следующим образом:

$$f''_0 = (f_1 - 2f_0 + f_{-1}) / h^2, \quad (7.1.4)$$

$$f''_0 = (-f_2 + 16f_1 - 30f_0 + 16f_{-1} - f_{-2}) / (12h^2).$$

Например, для первой производной по времени можно принять правую конечную разность порядка  $O(h)$ :

$$\left. \frac{\partial u(t, x_j)}{\partial t} \right|_{t=t_i} \approx \frac{u_{i+1,j} - u_{i,j}}{\Delta t},$$

а для второй производной по координате  $x$  центральную конечную разность порядка  $O(h^2)$ :

$$\left. \frac{\partial^2 u(t_i, x)}{\partial x^2} \right|_{x=x_j} \approx \frac{u_{i,j+1} - 2u_{i,j} + u_{i,j-1}}{(\Delta x)^2}.$$

## 7.2. Гиперболические уравнения

В качестве примера гиперболического уравнения рассматривается волновое уравнение колебаний эластичной струны

$$u_{tt}(x,t) = c^2 u_{xx}(x,t), \quad x \in (0, X), t \in (0, T), \quad (7.2.1)$$

с граничными условиями (закрепление струны на концах)

$$u(0,t) = u_l(t) \equiv L, \quad u(X,t) = u_r(t) \equiv R, \quad t \in [0, T], \quad (7.2.2)$$

и начальными условиями первого рода для функции  $u(x,t)$  (отклонение от положения равновесия) и второго рода для её производной (скорости отклонения)

$$u(x,0) = f(x), \quad u_t(x,0) = g(x) \quad \text{для } x \in [0, X]. \quad (7.2.3)$$

На сетке (7.1.1) сеточная функция  $u_{i,j}$  удовлетворяет соотношениям, следующим из (7.2.1) и (7.1.4):

$$\frac{u_{i+1,j} - 2u_{i,j} + u_{i-1,j}}{(\Delta t)^2} = c^2 \frac{u_{i,j+1} - 2u_{i,j} + u_{i,j-1}}{(\Delta x)^2}.$$

Обозначим  $q = c \frac{\Delta t}{\Delta x}$ . После небольших преобразований получаем явное выражение для значения сеточной функции на  $(i+1)$ -м слое по времени через её значения на  $i$ -м и  $(i-1)$ -м слоях:

$$u_{i+1,j} = 2(1 - q^2)u_{i,j} + q^2(u_{i,j+1} + u_{i,j-1}) - u_{i-1,j} \quad (7.2.4)$$

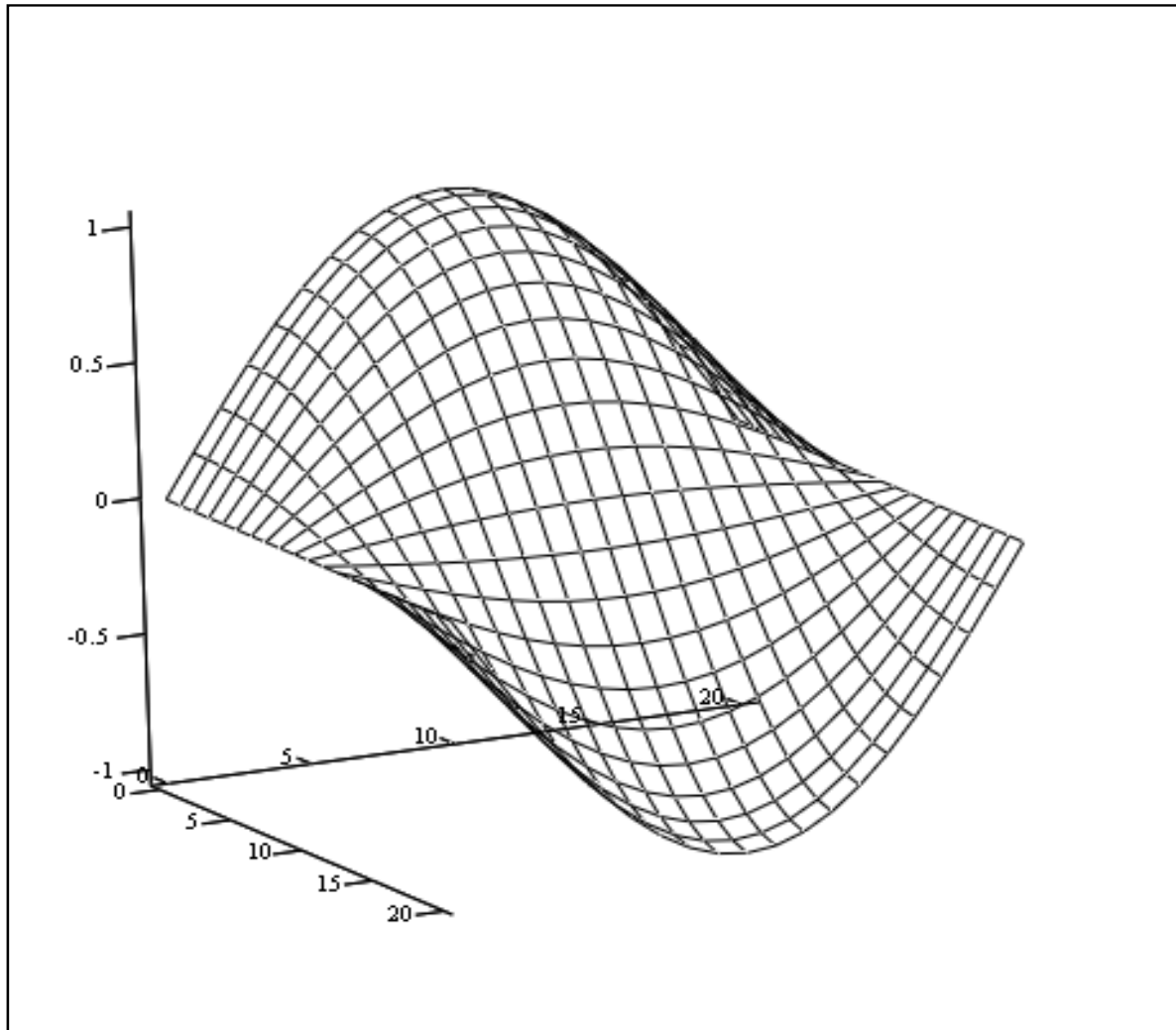
для  $j = 1, 2, \dots, N-1$ .

Сеточный шаблон вычислений по формуле (7.2.4) является пятиточечным, так как связывает между собой пять соседних узлов сетки вокруг точки  $(i, j)$ , включая её. Для того чтобы вычисления по формуле (7.2.4) были устойчивы, необходимо выполнение соотношения  $q \leq 1$ . Таким образом, условия устойчивости вычислений по явной схеме (7.2.4) накладывают ограничения на

шаг по времени при заданном шаге по пространственной координате.

Для определения значений сеточной функции на двух начальных слоях  $i = 0$  и  $i = 1$  используем заданные начальные условия (7.2.2), (7.2.3) и правую конечную разность порядка  $O(\Delta t)$  для аппроксимации первой производной в начальной точке. Использование конечной разности первого порядка точности вносит на начальном этапе ошибку большую, чем при аппроксимации уравнения во внутренних точках. В результате получаем для значений на слоях  $i = 0, 1$

$$u_{0,j} = f_j, \quad u_{1,j} = f_j + \Delta t \cdot g_j. \quad (7.2.5)$$



u

Рис. 7.1

На рис. 7.1 показан полупериод колебаний струны, рассчитанный по формулам (7.2.4), (7.2.5) при следующих исходных данных, граничных и начальных условиях:  $X = 1$ ,  $T = 1$ ,  $c = 1$ ,  $L = R = 0$ ,  $f(x) = \sin(\pi x)$ ,  $g(x) = 0$ . Шаги сетки по пространственной координате и по времени  $\Delta x = 0.05$ ,  $\Delta t = 0.05$ .

Число шагов по  $x$  и по  $t$  соответственно  $N = 20$ ,  $M = 20$ .

### 7.3. Параболические уравнения

Данный тип уравнений рассмотрим на примере одномерного нестационарного уравнения теплопроводности (7.3.1) с граничными (7.3.2) и начальными условиями (7.3.3), описывающего процесс установления температуры в изолированном стержне, имеющем на концах постоянную температуру  $L$  и  $R$  и заданное начальное распределение температуры вдоль стержня  $f(x)$ :

$$u_t(x, t) = c^2 u_{xx}(x, t), \quad 0 < x < X, \quad 0 \leq t < T; \quad (7.3.1)$$

$$u(0, t) = u_l(t) \equiv L, \quad u(X, t) = u_r(t) \equiv R, \quad t \in [0, T]; \quad (7.3.2)$$

$$u(x, 0) = f(x), \quad x \in [0, X]. \quad (7.3.3)$$

Для аппроксимации уравнения (7.3.1) используем конечные разности (7.1.2) и (7.1.4)

$$\frac{u_{i+1,j} - u_{i,j}}{\Delta t} = c^2 \frac{u_{i,j+1} - 2u_{i,j} + u_{i,j-1}}{(\Delta x)^2}.$$

Обозначим  $q = c^2 \frac{\Delta t}{(\Delta x)^2}$ . После преобразований получаем явную четырехточечную сеточную схему, в которой значение функции на  $(i+1)$ -м слое по времени выражается через три соседних значения на нижнем,  $i$ -м слое:

$$u_{i+1,j} = (1 - 2q)u_{i,j} + q(u_{i,j+1} + u_{i,j-1}). \quad (7.3.4)$$

Формула (7.3.4) позволяет последовательно найти все значения сеточной функции, начиная со слоя  $i = 0$ , на котором заданы начальные условия (7.3.3). Однако вычисления по этой формуле устойчивы только в том случае, если выполняется условие  $0 \leq q \leq 0.5$ . Это накладывает жесткие ограничения на шаг сетки по времени, обязывая выбирать этот шаг намного меньшим, чем шаг по пространственной координате, что существенно увеличивает время расчета и ограничивает применимость явной схемы.

Для аппроксимации уравнения (7.3.1) может быть использована левая конечная разность (7.1.2):

$$\frac{u_{i,j} - u_{i-1,j}}{\Delta t} = c^2 \frac{u_{i,j+1} - 2u_{i,j} + u_{i,j-1}}{(\Delta x)^2},$$

что приводит к неявной четырёхточечной разностной схеме :

$$u_{i,j+1} - \left(2 + \frac{1}{q}\right) u_{i,j} + u_{i,j-1} = -\frac{1}{q} u_{i-1,j}, \quad (7.3.5)$$

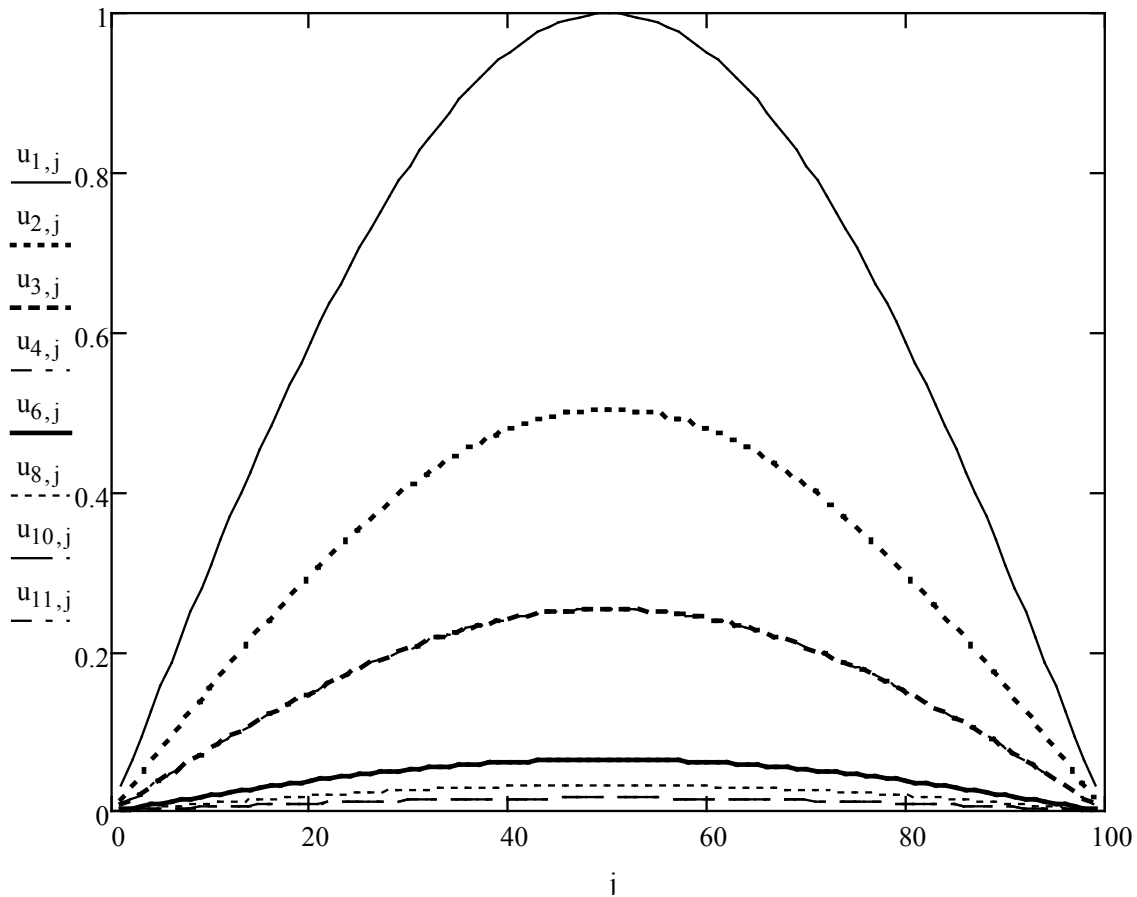
которая устойчива при любых соотношениях шагов сетки.

Из (7.3.5) следует, что для каждого слоя  $i$  по времени значения неизвестной сеточной функции  $u_{i,j}$ ,  $j = 1, 2, \dots, N-1$  связаны СЛАУ с трехдиагональной матрицей. В этой матрице на главной диагонали находится значение  $-(2 + \frac{1}{q})$ , а на двух соседних диагоналях -1. Значение на главной диагонали близко к  $-2$ , т.к. значение  $q$ , как правило,  $\gg 1$ . Вектор в правой части (7.3.5) (при постоянном значении  $i = \text{const}$ ) известен из вычислений на предыдущем шаге по времени и входит в правую часть СЛАУ.

Последовательно решая СЛАУ (7.3.5), начиная со слоя  $i = 1$ , можно вычислить сеточную функцию во всей области решения. Система (7.3.5) может быть решена как стандартным методом



(так как порядок системы не слишком велик:  $N-1$ ), так и специальными методами применяемыми для решения систем с трехдиагональными матрицами, например, методом прогонки [2].



**Рис. 7.2**

На рис. 7.2 представлен расчет установления температуры в стержне, проведенный по неявной схеме (7.3.5), при следующих начальных и граничных условиях:  $X=1$ ,  $T=1$ ;  $u(0,t)=0$ ,  $u(X,t)=0$ ,  $t \in [0,T]$ ;  $u(x,0)=f(x)=\sin(\pi x)$ ,  $x \in [0,X]$ . Шаги сетки по времени и по пространственной координате  $\Delta t=0.1$ ,  $\Delta x=0.01$ . При данном значении  $q=1000$  расчеты по явной схеме (7.3.4) были бы невозможны из-за большой неустойчивости. Число шагов по  $t$  и по  $x$  соответственно  $M=10$ ,  $N=100$ .

### 7.4. Уравнения эллиптического типа

Двумерные краевые задачи для уравнений данного типа рассмотрим на примере уравнений Лапласа, Пуассона и Гельмгольца.

Обозначим, как обычно, оператор Лапласа

$$\Delta u = u_{xx} + u_{yy}, \quad x, y \in S.$$

Тогда указанные уравнения имеют вид:

- 1) уравнение Лапласа:  $\Delta u = 0$ ,
- 2) уравнение Пуассона:  $\Delta u = f(x, y)$ .
- 3) уравнение Гельмгольца:  $\Delta u + g(x, y)u = f(x, y)$ .

Граничные условия задаются на границе области  $S$ :  $u_\Gamma = G(x, y)$ , в частности, на границе прямоугольника  $x \in [0, X], y \in [0, Y]$ :  $u(x, 0) = u_0(x)$ ,  $u(x, Y) = u_Y(x)$ ,  $u(0, y) = u_0(y)$ ,  $u(X, y) = u_X(y)$ .

#### 7.4.1. Разностная схема уравнений

Разностную схему рассмотрим на примере уравнения Пуассона в прямоугольнике, используя для аппроксимации второй производной конечные разности второго порядка точности (7.1.4). Вводя сетку  $y_i = i \cdot \Delta y$ ,  $x_j = j \cdot \Delta x$ ,  $0 \leq i \leq M$ ,  $0 \leq j \leq N$ , получаем

$$\frac{u_{i,j+1} - 2u_{i,j} + u_{i,j-1}}{(\Delta x)^2} + \frac{u_{i+1,j} - 2u_{i,j} + u_{i-1,j}}{(\Delta y)^2} = f_{i,j},$$

или, введя обозначение  $q = (\Delta x)^2 / (\Delta y)^2$ , получаем пятиточечную разностную схему для внутренних узлов прямоугольника:

$$u_{i,j+1} - 2(1+q)u_{i,j} + u_{i,j-1} + q(u_{i+1,j} + u_{i-1,j}) = (\Delta x)^2 f_{i,j}. \quad (7.4.1)$$

Данная неявная схема охватывает все внутренние точки области  $1 \leq i \leq M-1, 1 \leq j \leq N-1$ , их количество  $n = (M-1)(N-1)$ .

Таково же число уравнений и неизвестных в СЛАУ, построенной на основе (7.4.1).

Пусть для простоты  $f(x, y) = -1$  для всех  $x, y \in S$ , т.е.  $f_{i,j} \equiv -1$  для всех внутренних точек  $1 \leq i \leq M-1$ ,  $1 \leq j \leq N-1$ , а граничные условия таковы: внизу  $u(x, 0) = 0$ , слева и справа  $u(0, y) = 0$ ,  $u(X, y) = 0$  и только наверху задана отличная от нуля функция  $u(x, Y) = \sin(\pi x)$ . Зададим  $\Delta x = \Delta y = 0.2$ ,  $X = Y = 1$ . Тогда  $M = 5, N = 5$ , а число внутренних точек и уравнений  $n = 16$ . Матрица СЛАУ для данной задачи задается по следующему закону (на языке пакета Mathcad):

$$A := \text{matrix}(n, n, a)$$

$$a(i, j) := \begin{cases} -2 \cdot (1 + q) & \text{if } i = j \\ 1 & \text{if } j = i + 1 \wedge \text{mod}(j, N - 1) \neq 0 \\ 1 & \text{if } i = j + 1 \wedge \text{mod}(i, N - 1) \neq 0 \\ q & \text{if } j = i + N - 1 \\ q & \text{if } j = i - N + 1 \\ 0 & \text{otherwise} \end{cases}$$

Здесь  $0 \leq i \leq n-1$ ,  $0 \leq j \leq n-1$  – это индексы матрицы  $A$ , они связаны с другими, ранее введенными индексами  $i, j$  для узлов сетки. Сеточная функция  $u_{i,j}$ ,  $1 \leq i \leq M-1$ ,  $1 \leq j \leq N-1$  выражается через найденный в результате решения СЛАУ вектор решения  $v_l$ ,  $1 \leq l \leq n-1$  следующим образом:  $u_{i,j} = v_{j-1+(N-1)(i-1)}$ .

Пятидиагональная матрица  $A$  имеет следующее строение:

$$A =$$

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	-4	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
1	1	-4	1	0	0	1	0	0	0	0	0	0	0	0	0	0
2	0	1	-4	1	0	0	1	0	0	0	0	0	0	0	0	0
3	0	0	1	-4	0	0	0	1	0	0	0	0	0	0	0	0
4	1	0	0	0	-4	1	0	0	1	0	0	0	0	0	0	0
5	0	1	0	0	1	-4	1	0	0	1	0	0	0	0	0	0
6	0	0	1	0	0	1	-4	1	0	0	1	0	0	0	0	0
7	0	0	0	1	0	0	1	-4	0	0	0	1	0	0	0	0
8	0	0	0	0	1	0	0	0	-4	1	0	0	1	0	0	0
9	0	0	0	0	0	1	0	0	1	-4	1	0	0	1	0	0
10	0	0	0	0	0	0	1	0	0	1	-4	1	0	0	1	0
11	0	0	0	0	0	0	0	1	0	0	1	-4	0	0	0	1
12	0	0	0	0	0	0	0	0	1	0	0	0	-4	1	0	0
13	0	0	0	0	0	0	0	0	0	1	0	0	1	-4	1	0
14	0	0	0	0	0	0	0	0	0	0	1	0	0	1	-4	1
15	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	-4

Вектор правой части СЛАУ  $A\vec{v} = \vec{b}$  задается в данной задаче по закону:

$$b_i := \begin{cases} -(\Delta x)^2 & \text{if } i < IK \\ -(\Delta x)^2 - q \cdot \sin[\pi \cdot (i + 1 - IK) \cdot \Delta x] & \text{otherwise} \end{cases},$$

где  $IK = (M - 2)N - M + 2$  – число, определяющее в индексах вектора решения начало последнего слоя внутренних узлов по оси  $y$ , на которых учитывается заданное граничное условие. Заданная функция из уравнения, в данной задаче являющаяся константой,  $f(x, y) \equiv -1$ , входит в правую часть СЛАУ в виде слагаемого  $-1 \cdot (\Delta x)^2$ . Значения вектора правой части в данной задаче:

$b^T =$		0	1	2	3	4	5	6	7
	0	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04	-0.04
	8	9	10	11	12	13	14	15	
	-0.04	-0.04	-0.04	-0.04	0.628	0.991	0.991	0.628	

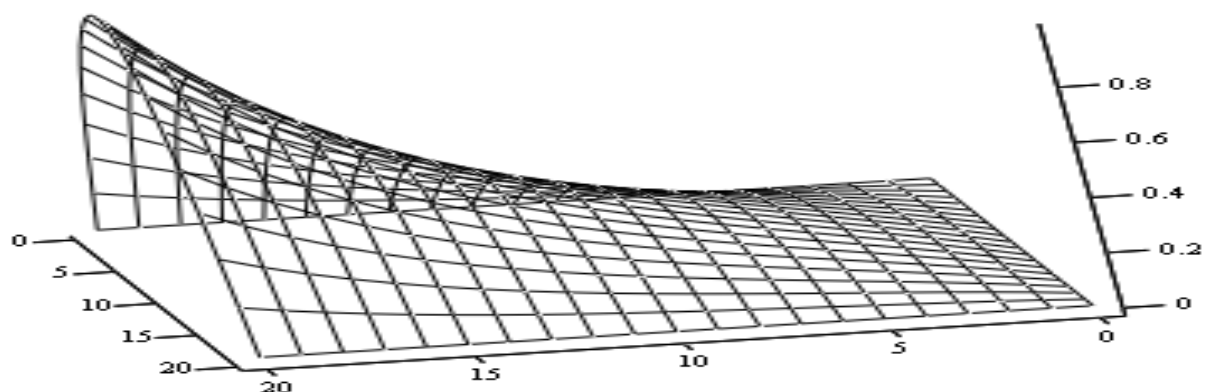


Рис. 7.3

На рис. 7.3 показано распределение функции решения аналогичной краевой задачи в двумерной области при порядке СЛАУ  $n = 361$ . Решение получено комбинированным методом Зейделя-ОСП при оптимальном параметре  $k_0 = -0.425$  за  $m = 41$  итераций с относительной точностью решения в  $\varepsilon = 10^{-5}$ . Обычный метод Зейделя сходится здесь лишь за  $m = 305$  итераций и сопоставим по времени решения с прямым методом. Еще большее число требуемых итераций показывают в данной задаче метод ОСП с матрицей (1.2) -  $m = 546$ .

Отметим, что матрица задачи при  $\Delta x = \Delta y$  и заданном  $n$  постоянна и не зависит от краевых условий и источников  $f(x, y)$ , которые входят в правую часть СЛАУ. Соответственно задачи с различными краевыми условиями и источниками могут решаться с тем же самым оптимальным параметром, найденным один раз для данной сетки.

### 7.5. Лабораторные задания к теме «Численное решение уравнений в частных производных»

Лабораторные работы по теме могут быть выполнены с помощью математических пакетов программ Mathcad или Matlab. В результате работы должна быть представлена искомая сеточная функция в виде матрицы значений в узлах сетки либо в виде по-

слоного по времени распределения значений сеточных векторов. Следует также привести графическое представление результатов.

### 7.5.1. Гиперболические уравнения

Варианты заданий для одномерного волнового уравнения с граничными и начальными условиями (см. 7.2).

№ п/п	$X$	$T$	$L$	$R$	$f(x)$	$g(x)$	$c$	$\Delta x$	Метод*
1.1	1	1	0	0	$\sin(\pi x)$	0	2	0.1, 0.05	1
1.2	1	2	0	0	$\sin(\pi x) + \sin(2\pi x)$	0	1	0.1, 0.05	1
1.3	$\pi$	1	0	0	$\begin{cases} x, x < 0.5X \\ -x + X, x \geq 0.5X \end{cases}$	0	2	0.1, 0.05	1
1.4	1	2	0	0	$\begin{cases} x, x < 0.75 \\ -3x + 3, x \geq 0.75 \end{cases}$	0	1	0.1, 0.05	1
1.5	$\pi$	1	0	0	$x(\pi - x)$	0	2	0.1, 0.05	1
1.6	1	2	0	0	$x - x^3$	0	1	0.1, 0.05	1
1.7	$\pi$	1	0	0	$x \sin(x)$	0	2	0.1, 0.05	1
1.8	1	2	0	0	$x^2(1 - x^2)$	0	1	0.1, 0.05	1
1.9	$\pi$	1	0	0	$x(\pi^2 - x^2)$	0	2	0.1, 0.05	1
1.10	1	2	0	0	$x(1 - x^3)$	0	1	0.1, 0.05	1

### 7.5.2. Параболические уравнения

Варианты заданий для одномерного уравнения теплопроводности с граничными и начальными условиями (см. 7.3).

№ п/п	$X$	$T$	$L$	$R$	$f(x)$	$g(x)$	$c$	$\Delta x$	Метод *
2.1	1	0.1	0	0	$\sin(\pi x)$	0	1	0.01	1
2.2	1	1	0	0	$\sin(\pi x)$	0	1	0.01	2
2.3	1	0.1	0	0	$\sin(\pi x) + \sin(2\pi x)$	0	1	0.01	1
2.4	1	2	0	0	$\sin(\pi x) + \sin(2\pi x)$	0	1	0.01	2
2.5	$\pi$	0.1	0	0	$\begin{cases} x, x < 0.5X \\ -x + X, x \geq 0.5X \end{cases}$	0	1	$0.01\pi$	1
2.6	$\pi$	1	0	0	$\begin{cases} x, x < 0.5X \\ -x + X, x \geq 0.5X \end{cases}$	0	1	$0.01\pi$	2
2.7	$\pi$	0.1	0	0	$x(\pi - x)$	0	1	$0.01\pi$	1
2.8	$\pi$	2	0	0	$x(\pi - x)$	0	1	$0.01\pi$	2
2.9	1	0.1	0	0	$x - x^3$	0	1	0.01	1
2.10	1	1	0	0	$x - x^3$	0	1	0.01	2
2.11	$\pi$	0.1	0	0	$x \sin(x)$	0	1	0.01..	1
2.12	$\pi$	2	0	0	$x \sin(x)$	0	1	$0.01\pi$	2

\* Методы: 1- явный; 2 - неявный, сведением к СЛАУ и последующим решением стандартным методом. Сравнить со строгим решением.

### 7.5.3. Эллиптические уравнения

Решить заданную краевую задачу методом сеток, сведением её к СЛАУ и последующим решением прямым (стандартным) и итерационным (Зейделя–ОСП) методами. Сравнить с существующим строгим решением.

Варианты заданий для краевой задачи с уравнениями эллиптического типа (см. 7.4).

№ п/п	Уравнение *	$X$	$Y$	$u_0(x)$	$u_Y(x)$	$u_0(y)$	$u_X(y)$	$\Delta x = \Delta y$	$f(x, y)$
3.1	1	1	1	$\sin(\pi y)$	0	0	0	0.2, 0.1	–
3.2	2	1	1	$x^2$	$(x-1)^2$	$y^2$	$(y-1)^2$	0.2, 0.1	–1
3.3	1	1	1	$u_\Gamma = x + y$				0.2, 0.1	–
3.4	3	1	1	$u_\Gamma = \cos(2x) + \sin(2y)$				0.2, 0.1	0, $g(x, y) = 4$
3.5	1	1	1	0	$x$	0	$y$	0.2, 0.1	–
3.6	2	1	1	$x^3$	$x^3$	0	1	0.2, 0.1	$y$
3.7	1	4	4	10	120	90	40	0.5, 0.25	-
3.8	2	1	1	$u_\Gamma = x + y$				0.2, 0.1	–2
3.9	1	2	2	10	20	30	40	0.25, 0.1	–
3.10	2	1	1	$x^3$	$(x-1)^3$	$y^3$	$(y-1)^3$	0.2, 0.1	2

\* Уравнения: 1 – Лапласа, 2 – Пуассона, 3 – Гельмгольца.



## ЛИТЕРАТУРА

1. **Бахвалов Н.С., Жидков Н.П., Кобельков Г.М.** Численные методы. – М.: Лаборатория Базовых Знаний, 2001. – 632 с.
2. **Мэтьюз Д.Г., Финк К.Д.** Численные методы. Использование MATLAB : пер. с англ. 3-е издание. – М.: Изд. «Вильямс», 2001. – 720 с.
3. **А.Б.Самохин, А.С. Самохина.** Численные методы и программирование на Фортране для персонального компьютера. – М.: Радио и связь, 1996. – 224 с.
4. **Демидович Б.П., Марон И.А.** Основы вычислительной математики. – М.: Наука, 1970. – 644 с.
5. **Вержбицкий В.М.** Численные методы (Математический анализ и обыкновенные дифференциальные уравнения). – М.: Высшая школа, 2001. – 382 с.
6. **Заварыкин В.М., Житомирский В.Г., Лапчин М.П.** Численные методы. – М.: Просвещение , 1991. – 176 с.

## Содержание

Введение.....	3
1. Абсолютная и относительная погрешности.....	3
1.1. Число верных знаков приближенного числа.....	4
1.2. Погрешность функций.....	5
1.3. Погрешность простейших функций двух переменных.....	5
1.4. Примеры и задания .....	6
2. Приближение функций.....	9
2.2. Интерполяционный полином Лагранжа .....	10
2.3. Интерполяционный полином Ньютона .....	12
2.3. Примеры и задания для практических занятий .....	14
3. Численные методы решений трансцендентных и алгебраических уравнений.....	17
3.1. Метод простой итерации для решения нелинейных и трансцендентных уравнений.....	18
3.2. Метод хорд и секущих.....	18
3.3. Метод касательных .....	20
3.4. Скорость сходимости итерационных методов.....	20
3.5. Пример и задание для практических занятий .....	20
4. Численное интегрирование .....	25
4.1. Метод Ньютона—Котеса.....	25
4.2. Метод прямоугольников.....	26
4.3. Метод трапеций.....	27
4.4. Метод парабол (метод Симпсона).....	28
4.5. Квадратурные формулы Гаусса .....	29
4.6. Задание для практических занятий .....	30
5. Численные методы линейной алгебры.....	32
5.1. Численное решение СЛАУ.....	32
5.2. Прямые методы решения СЛАУ .....	35
5.2.1. Метод Гаусса (метод исключений) .....	35
5.2.2. Вычислительная схема метода Гаусса .....	36
5.2.3. Ортогонализация матриц.....	38
5.2.4. Решение системы уравнений методом ортогонализации .....	39
5.3. Итерационные методы решения СЛАУ.....	40
5.3.1. Метод простой итерации.....	40
5.3.2. Метод Якоби и метод Зейделя.....	42
5.3.3. Метод оптимального спектрального параметра (ОСП) для простой итерации .....	44
5.4. Нахождение собственных векторов и собственных значений матриц.....	51

5.5. Примеры и задания к теме.....	52
5.5.1. Прямые методы решения СЛАУ .....	52
5.5.2. Итерационные методы решения СЛАУ .....	56
5.5.3. Нахождение собственных значений и векторов .....	59
6. Численные методы решения обыкновенных дифференциальных уравнений .....	61
6.1. Метод разложения в ряд Тейлора.....	61
6.2. Общая схема метода Рунге—Кутты.....	62
6.3 Методы Рунге—Кутты низших порядков .....	63
6.3.1. Метод Эйлера .....	63
6.3.2. Метод трапеций и прямоугольника .....	63
6.4. Методы Рунге—Кутты высших порядков.....	64
6.5. Задание к теме и пример решения ОДУ .....	65
7. Численное решение начально-краевых задач для дифференциальных уравнений в частных производных .....	67
7.1. Конечные разности. ....	67
7.2. Гиперболические уравнения .....	69
7.3. Параболические уравнения .....	71
7.4. Уравнения эллиптического типа .....	74
7.4.1. Разностная схема уравнений.....	74
7.5. Лабораторные задания к теме «Численное решение уравнений в частных производных» .....	77
7.5.1. Гиперболические уравнения .....	78
7.5.2. Параболические уравнения .....	78
7.5.3. Эллиптические уравнения.....	80
Литература .....	81

---

*Учебное издание*

**Самохин Александр Борисович**  
**Чердынцев Виктор Викторович**  
**Воронцов Александр Алексеевич**

**ЧИСЛЕННЫЕ МЕТОДЫ**  
**Учебное пособие**

Печатается в авторской редакции

---

Подписано в печать 20.11.2014. Формат 60×84 1/16.  
Физ. печ. л. 5,25. Тираж 100 экз. Изд. № 105. Заказ № 576

---

ФГБОУ ВПО «МГТУРЭА (МИРЭА)»  
119454, Москва, пр. Вернадского, д. 78