

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ  
РОССИЙСКОЙ ФЕДЕРАЦИИ  
УРАЛЬСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ  
ИМЕНИ ПЕРВОГО ПРЕЗИДЕНТА РОССИИ Б. Н. ЕЛЬЦИНА

В. Г. Пименов, А. Б. Ложников

# ЧИСЛЕННЫЕ МЕТОДЫ

В двух частях

Часть 2

Рекомендовано методическим советом УрФУ  
в качестве учебного пособия для студентов, обучающихся  
по программе бакалавриата по направлениям подготовки  
010100 «Математика», 010200 «Математика и компьютерные  
науки», 010800 «Механика и математическое моделирование»,  
010300 «Фундаментальная информатика и информационные  
технологии», 230700 «Прикладная информатика»

Екатеринбург  
Издательство Уральского университета  
2014

УДК 519.6 (075.8)  
П325

**Рецензенты:**

кафедра высшей математики Уральского государственного  
лесотехнического университета (заведующий кафедрой кандидат  
физико-математических наук, доцент А. Ю. Вдовин);  
А. Л. Агеев, доктор физико-математических наук (Институт  
математики и механики им. Н. Н. Красовского УрО РАН)

Научный редактор  
Ю. А. Меленцова,  
кандидат физико-математических наук, доцент

**Пименов, В. Г.**

П325 Численные методы : в 2 ч. Ч. 2 : [учеб. пособие] / В. Г. Пименов, А. Б. Ложников ; [науч. ред. Ю. А. Меленцова] ; М-во образования и науки Рос. Федерации, Урал. федер. ун-т. -- Екатеринбург : Изд-во Урал. ун-та, 2014. -- 106 с.

ISBN 978-5-7996-1015-9

ISBN 978-5-7996-1342-6 (часть 2)

Даются основные понятия, изучаемые во второй части курса «Численные методы»: методы решения дифференциальных уравнений (задачи Коши), методы решения краевых задач, интерполяция сплайнами, метод наименьших квадратов, численное решение интегральных уравнений, численное решение уравнений математической физики.

Для студентов 3-го курса Института математики и компьютерных наук УрФУ всех направлений подготовки.

УДК 519.6 (075.8)

ISBN 978-5-7996-1015-9

© Уральский федеральный  
университет, 2014

ISBN 978-5-7996-1342-6 (часть 2)

© Пименов В. Г., Ложников А. Б., 2014

## Предисловие

Во второй части пособия излагаются разделы дисциплины «Численные методы», которые обычно изучаются во втором семестре годового курса. Эти разделы опираются на материал, изучаемый в первом семестре и изложенный в первой части пособия. На разных направлениях подготовки бакалавров Института математики и компьютерных наук на этот курс отводится различное количество часов, поэтому ряд разделов курса без ущерба для изучения основного материала на некоторых направлениях подготовки можно опустить.

В пособие вошли разделы «Численные методы решения дифференциальных уравнений (задачи Коши)», «Численные методы решения краевых задач», «Интерполяция сплайнами», «Метод наименьших квадратов», «Численное решение интегральных уравнений», «Численное решение уравнений математической физики». Материал отобран и изложен в соответствии со сложившимися традициями преподавания на кафедре вычислительной математики. В связи с этим авторы признательны Г. Н. Мильштейну, Л. Б. Рышко, С. А. Нефедову, И. А. Башкирцевой, С. А. Пьянзину, А. Л. Агееву, И. Н. Кандобе, О. Э. Соловьевой, которые в разные годы читали лекции по этому курсу. Авторы благодарят также сотрудников кафедры (Е. А. Конончук, О. В. Корниенко, А. В. Лекомцева, Г. Е. Лузину, Ю. А. Меленцову, С. И. Солодушкина и др.), которые вели и ведут практические и лабораторные занятия, тесно связанные с курсом лекций. На эти традиции в значительной мере повлияла и научная направленность сотрудников кафедры.

Литература ко второй части во многом совпадает с литературой, указанной в первой части пособия, поэтому не приводится. Но ко второй части пособия добавлен ряд книг специального характера [1–8], в которых на фундаментальном уровне и с большой полнотой излагаются вопросы изучаемых разделов.

# 1. Численные методы решения дифференциальных уравнений.

## Задача Коши

### 1.1. Постановка задачи и предположения

Рассмотрим обыкновенное дифференциальное уравнение первого порядка

$$y' = f(x, y)$$

с начальными условиями

$$y(x_0) = y_0.$$

Эту задачу называют обычно *задачей Коши*. Будем предполагать, что функция  $f(x, y)$  определена по первому аргументу на отрезке  $[x_0, x_0 + X]$ , по второму аргументу — на всей числовой прямой; в области определения функция  $f(x, y)$  непрерывна, а также выполняется условие липшицевости по второму аргументу: существует константа  $L$ , такая, что для всяких  $y_1$  и  $y_2$  выполняется

$$|f(x, y_1) - f(x, y_2)| \leq L|y_1 - y_2|.$$

Эти условия гарантируют существование и единственность решения  $y(x)$  задачи Коши. Если, кроме того, потребовать выполнение условия продолжимости, например выполнение условия подлинейного роста

$$|f(x, y)| \leq \alpha(x)(1 + |y|),$$

то решение определено на всем отрезке  $[x_0, x_0 + X]$ .

Будем искать приближение этого решения в дискретной схеме. Разобьем отрезок  $[x_0, x_0 + X]$  на  $N$  равных частей с шагом  $h = \frac{X}{N}$ , обозначив узлы  $x_i = x_0 + ih, i = 0, \dots, N$ . Приближенные значения решения в узлах будем обозначать  $y_i$ , в отличие от точных значений решения в узлах  $y(x_i), i = 0, \dots, N$ .

Задача состоит в конструировании алгоритмов для  $y_i$  и исследования условий, при которых точные и приближенные решения близки.

**З а м е ч а н и е.** Все излагаемые ниже результаты справедливы для систем дифференциальных уравнений, т. е. случая, когда  $y \in R^n$ , что дает возможность использовать излагаемые ниже методы также для численного решения дифференциальных уравнений высших порядков.

## 1.2. Методы, основанные на тейлоровском разложении

В этом разделе будем предполагать, что точное решение  $y(x)$  и правая часть дифференциального уравнения (функция  $f(x, y)$ ) достаточное число раз непрерывно дифференцируемы.

Разложим  $y(x_1)$  по формуле Тейлора в окрестности точки  $x_0$ :

$$y(x_1) = y(x_0) + y'(x_0)h + y''(x_0)\frac{h^2}{2} + \dots$$

Взяв в разложении несколько слагаемых и отбросив остальные, получаем тот или иной метод, при этом производные точного решения можно выразить в силу исходного уравнения. Например, простейший метод на одном шаге можно записать так:

$$y_1 = y_0 + hf(x_0, y_0).$$

Аналогично можно записать алгоритм при переходе к очередному узлу:

$$y_{n+1} = y_n + hf(x_n, y_n), \quad n = 0, 1, \dots, N - 1.$$

Этот метод называется методом Тейлора первого порядка, чаще его называют *методом Эйлера*. Геометрическая интерпретация метода Эйлера представляет собой движение по касательной к интегральным кривым (рис. 1). Сходимость метода

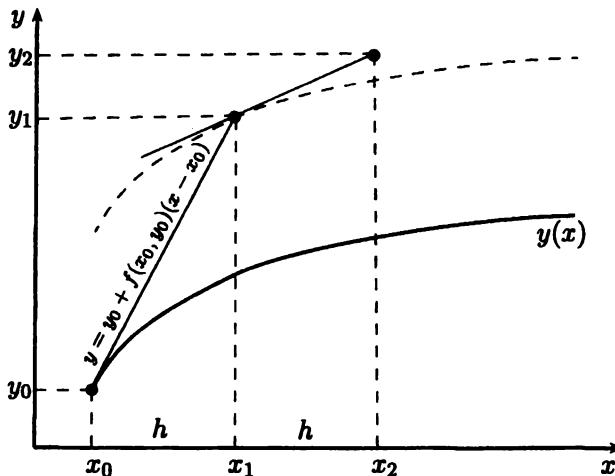


Рис. 1. Геометрическая интерпретация явного метода Эйлера

Эйлера рассмотрим чуть позже, а пока запишем метод Тейлора второго порядка.

Учитывая, что

$$\frac{df(x, y)}{dx} = \frac{\partial f(x, y)}{\partial x} + \frac{\partial f(x, y)}{\partial y} \frac{dy}{dx},$$

получаем

$$y_{n+1} = y_n + hf(x_n, y_n) + \frac{h^2}{2} \left[ \frac{\partial f(x_n, y_n)}{\partial x} + \frac{\partial f(x_n, y_n)}{\partial y} f(x_n, y_n) \right].$$

Этот метод, называемый *методом Тейлора второго порядка*, имеет существенный недостаток – необходимость считать частные производные, что предполагает аналитическую работу, которая ограничивает общность компьютерных программ. Аналогично можно выписать методы Тейлора третьего порядка, четвертого порядка и далее. И все эти методы, кроме метода Эйлера, обладают указанным недостатком.

### 1.3. Явный метод Эйлера

На примере метода Эйлера исследуем главный вопрос – сходимость. Назовем метод *сходящимся с порядком*  $p > 0$ , если найдется  $C$ , такое, что выполняется

$$|y_n - y(x_n)| \leq Ch^p$$

для всех  $n = 1, 2, \dots, N$ .

Цель этого раздела доказать, что метод Эйлера (его называют также явным методом Эйлера)

$$y_{n+1} = y_n + hf(x_n, y_n)$$

сходится с первым порядком.

На первом этапе исследуем локальную погрешность, называемую также погрешностью аппроксимации или невязкой метода. *Невязкой* метода назовем разность левой и правой части метода, где вместо приближенного решения подставлено точное решение. Так, невязка метода Эйлера определяется как

$$\psi_n = y(x_{n+1}) - y(x_n) - hf(x_n, y(x_n)).$$

**Лемма.** Если точное решение – дважды непрерывная дифференцируемая функция, то невязка метода Эйлера имеет второй порядок малости по  $h$ .

Доказательство. В определении невязки разложим  $y(x_{n+1})$  в окрестности точки  $x_n$ :

$$\psi_n = y(x_n) + y'(x_n)h + y''(\vartheta) \frac{h^2}{2} - y(x_n) - hf(x_n, y(x_n)),$$

где  $x_n \leq \vartheta \leq x_{n+1}$ . Так как вторая производная решения непрерывна на отрезке, то она ограничена и найдется константа  $C$ , такая, что

$$|\psi_n| \leq Ch^2.$$

**Теорема 1.** *Если точное решение – дважды непрерывная дифференцируемая функция, то метод Эйлера сходится с первым порядком.*

**Доказательство.** Обозначим величину погрешности через

$$\varepsilon_n = y(x_n) - y_n$$

и оценим  $|\varepsilon_{n+1}|$  через  $|\varepsilon_n|$ . Выразим  $y(x_{n+1})$  из определения невязки, а  $y_{n+1}$  из определения метода Эйлера:

$$\begin{aligned} |\varepsilon_{n+1}| &= |y(x_{n+1}) - y_{n+1}| = |y(x_n) + hf(x_n, y(x_n)) + \psi_n - \\ &- y_n - hf(x_n, y_n)| \leq |\varepsilon_n| + h|f(x_n, y(x_n)) - f(x_n, y_n)| + |\psi_n|. \end{aligned}$$

Используя для оценки второго слагаемого условие Липшица функции  $f(x, y)$  по второй переменной, а для оценки третьего слагаемого результат леммы, получаем

$$|\varepsilon_{n+1}| \leq (1 + Lh)|\varepsilon_n| + Ch^2.$$

Введем обозначения  $A = (1 + Lh)$ ,  $B = Ch^2$ , тогда

$$|\varepsilon_{n+1}| \leq A|\varepsilon_n| + B.$$

Последовательно получаем  $|\varepsilon_0| = 0$ ,  $|\varepsilon_1| \leq B$ ,  $|\varepsilon_2| \leq AB + B$ ,  $|\varepsilon_3| \leq (A^2 + A + 1)B$ , ...,  $|\varepsilon_n| \leq (A^{n-1} + \dots + A + 1)B$ . Используя формулу суммы геометрической прогрессии, получаем

$$|\varepsilon_n| \leq \frac{A^n - 1}{A - 1}B \leq \frac{A^N - 1}{A - 1}B.$$

Подставим в эту оценку выражения для  $A$  и  $B$ , воспользуемся также связью  $hN = X$ :

$$|\varepsilon_n| \leq \frac{(1 + Lh)^{\frac{X}{h}} - 1}{Lh}Ch^2.$$

Используя то, что во втором замечательном пределе  $\lim_{\alpha \rightarrow 0} (1 + \alpha)^{\frac{1}{\alpha}} = e$  выражение под знаком предела не превосходит предела, получаем оценку

$$|\varepsilon_n| \leq \frac{C}{L} (e^{LX} - 1)h,$$

равномерную по всем  $n = 1, 2, \dots, N$ . Эта оценка означает сходимость метода Эйлера с первым порядком.

#### 1.4. Метод Эйлера с пересчетом

Как известно из курса дифференциальных уравнений, задача Коши эквивалентна интегральному уравнению

$$y(x) = y_0 + \int_{x_0}^x f(s, y(s))ds.$$

Если применять на каждом шаге для подсчета интеграла метод левых прямоугольников, то получится метод Эйлера. Попробуем применить для подсчета интеграла более точную формулу трапеций. Получим на первом шаге

$$y_1 = y_0 + \frac{h}{2} (f(x_0, y_0) + f(x_1, y(x_1))).$$

Однако величина  $y(x_1)$  неизвестна, заменим ее величиной  $\hat{y} = y_0 + hf(x_0, y_0)$ , предварительно вычисленной по методу Эйлера. Получим метод

$$y_1 = y_0 + \frac{h}{2} (f(x_0, y_0) + f(x_1, y_0 + hf(x_0, y_0))),$$

или на  $n$ -м шаге

$$y_{n+1} = y_n + \frac{h}{2} (f(x_n, y_n) + f(x_n + h, y_n + hf(x_n, y_n))).$$

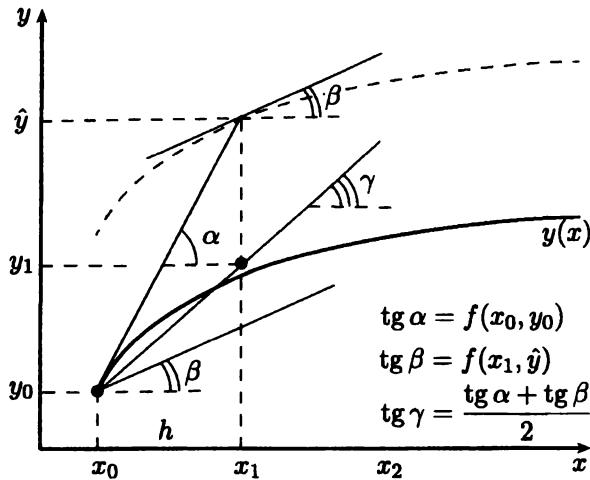


Рис. 2. Геометрическая интерпретация метода Эйлера с пересчетом

Этот метод называется *методом Эйлера с пересчетом* или *методом Хойна*. Геометрическая интерпретация метода Эйлера с пересчетом состоит в движении по прямой с углом наклона  $\gamma$  (рис. 2).

Покажем, что невязка метода имеет третий порядок. По определению невязка

$$\begin{aligned} \psi_n &= y(x_{n+1}) - y(x_n) - \\ &- \frac{h}{2}(f(x_n, y(x_n)) + f(x_n + h, y(x_n) + hf(x_n, y(x_n)))). \end{aligned}$$

Разложим решение  $y(x)$  в окрестности точки  $x_n$  и функцию  $f(x, y)$  в окрестности точки  $(x_n, y(x_n))$ . Для сокращения записи обозначим  $f = f(x_n, y(x_n))$ ,  $f'_x = \frac{\partial f(x_n, y(x_n))}{\partial x}$ ,  $f'_y = \frac{\partial f(x_n, y(x_n))}{\partial y}$ . Тогда

$$\psi_n = y(x_n) + hf + \frac{h^2}{2}(f'_x + f'_y f) + O(h^3) - y(x_n) -$$

$$-\frac{h}{2}(f + f + f'_x h + f'_y h f + O(h^2)) = O(h^3).$$

Позже будет доказано, что если невязка (локальная по-грешность) имеет порядок 3, то порядок сходимости этого метода равен 2.

### 1.5. Метод Коши

Применим для подсчета интеграла в интегральном уравнении формулу средних прямоугольников. Получим на первом шаге

$$y_1 = y_0 + h f(x_0 + \frac{h}{2}, y(x_0 + \frac{h}{2})).$$

Величина  $y(x_0 + \frac{h}{2})$  неизвестна, заменим ее величиной  $\hat{y} = y_0 + \frac{h}{2} f(x_0, y_0)$ , т. е. полшага по методу Эйлера. Получим метод

$$y_1 = y_0 + h f(x_0 + \frac{h}{2}, y_0 + \frac{h}{2} f(x_0, y_0)),$$

или на  $n$ -м шаге

$$y_{n+1} = y_n + h f(x_n + \frac{h}{2}, y_n + \frac{h}{2} f(x_n, y_n)).$$

Этот метод называется *методом Коши*. Геометрическая интерпретация метода Коши представлена на рис. 3. Подобно методу Эйлера с пересчетом проверяется, что невязка метода Коши имеет третий порядок малости по  $h$ , откуда следует, что метод сходится со вторым порядком. Метод Коши, так же как и метод Эйлера с пересчетом, требует на каждом шаге двух вычислений правой части дифференциального уравнения.

### 1.6. Семейство явных методов Рунге–Кутты второго порядка

Рассмотрим семейство методов, в которых на каждом шаге требуется последовательно вычислять два значения правой

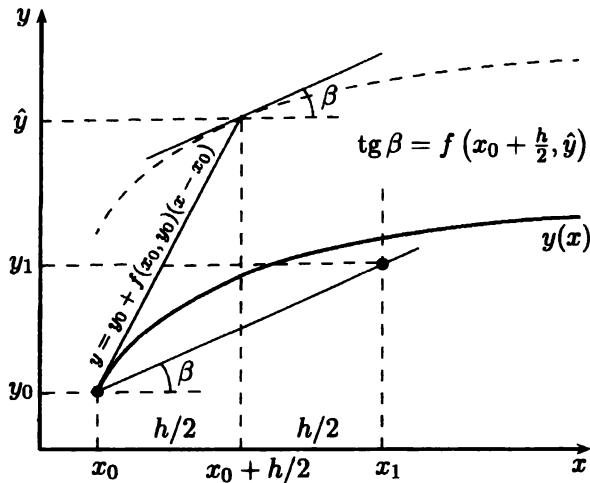


Рис. 3. Геометрическая интерпретация метода Коши

части дифференциального уравнения

$$y_{n+1} = y_n + \sigma_1 k_1 + \sigma_2 k_2,$$

где

$$k_1 = h f(x_n, y_n),$$

$$k_2 = h f(x_n + ah, y_n + bk_1).$$

Семейство содержит четыре параметра:  $\sigma_1, \sigma_2, a, b$ . В частности, для метода Эйлера  $\sigma_1 = 1, \sigma_2 = 0$ , для метода Эйлера с пересчетом  $\sigma_1 = \frac{1}{2}, \sigma_2 = \frac{1}{2}, a = 1, b = 1$ , для метода Коши  $\sigma_1 = 0, \sigma_2 = 1, a = \frac{1}{2}, b = \frac{1}{2}$ .

Потребуем, чтобы методы из семейства имели невязку третьего порядка, тогда они будут иметь второй порядок сходимости. По определению невязка будет

$$\begin{aligned} \psi_n = & y(x_{n+1}) - y(x_n) - \sigma_1 h f(x_n, y(x_n)) - \\ & - \sigma_2 h f(x_n + ah, y(x_n) + bh f(x_n, y(x_n))). \end{aligned}$$

Разложим по формуле Тейлора решение  $y(x)$  в окрестности точки  $x_n$  и функцию  $f(x, y)$  в окрестности точки  $(x_n, y(x_n))$  и, используя обозначения как в методе Эйлера с пересчетом, получаем

$$\begin{aligned}\psi_n = y(x_n) + hf + \frac{h^2}{2}(f'_x + f'_y f) + O(h^3) - y(x_n) - \sigma_1 hf - \\ - \sigma_2 h(f + f'_x ah + f'_y bhf + O(h^2)).\end{aligned}$$

Чтобы сократились слагаемые первого и второго порядков малости по  $h$ , нужно потребовать выполнения трех условий:

$$\begin{cases} \sigma_1 + \sigma_2 = 1 \\ a\sigma_2 = \frac{1}{2} \\ b\sigma_2 = \frac{1}{2}. \end{cases}$$

Таким образом, получаем однопараметрическое семейство методов, в которое входят, в частности, метод Эйлера с пересчетом и метод Коши. Семейство называется **явные методы Рунге–Кутты второго порядка**.

### 1.7. Явные методы Рунге–Кутты третьего и четвертого порядков

Среди семейства явных методов Рунге–Кутты третьего порядка отметим метод

$$y_{n+1} = y_n + \frac{1}{6}k_1 + \frac{1}{6}k_2 + \frac{4}{6}k_3,$$

где

$$k_1 = hf(x_n, y_n),$$

$$k_2 = hf(x_n + h, y_n + k_1),$$

$$k_3 = hf\left(x_n + \frac{h}{2}, y_n + \frac{1}{4}k_1 + \frac{1}{4}k_2\right).$$

Но наибольшее распространение получил один из методов Рунге–Кутты четвертого порядка (все семейство образует двухпараметрическое множество методов):

$$y_{n+1} = y_n + \frac{1}{6}k_1 + \frac{2}{6}k_2 + \frac{2}{6}k_3 + \frac{1}{6}k_4,$$

где

$$\begin{aligned} k_1 &= hf(x_n, y_n), \\ k_2 &= hf\left(x_n + \frac{h}{2}, y_n + \frac{1}{2}k_1\right), \\ k_3 &= hf\left(x_n + \frac{h}{2}, y_n + \frac{1}{2}k_2\right), \\ k_4 &= hf(x_n + h, y_n + k_3). \end{aligned}$$

Этот метод на каждом шаге требует четырех вычислений правой части дифференциального уравнения, но так как он четвертого порядка, то там, где метод Эйлера для достижения определенной точности требует 10 000 вычислений правой части дифференциального уравнения, метод Рунге–Кутты четвертого порядка — только 40, что ведет к уменьшению вычислительной погрешности.

З а м е ч а н и е. При пяти вычислениях правой части дифференциального уравнения не существует явного метода Рунге–Кутты пятого порядка, этот факт называется *барьером Бутчера*. Для построения метода пятого порядка нужно как минимум шесть вычислений.

## 1.8. Общий класс методов Рунге–Кутты

*Явными методами Рунге–Кутты* называются алгоритмы

$$y_{n+1} = y_n + \sum_{i=1}^m \sigma_i k_i,$$

где

$$k_1 = hf(x_n, y_n),$$

$$k_i = hf(x_n + a_i h, y_n + \sum_{j=1}^{i-1} b_{ij} k_j).$$

Число  $m$  называется числом этапов (количество вычислений правой части дифференциального уравнения на каждом шаге).

*Неявными методами Рунге–Кутты* называются алгоритмы

$$y_{n+1} = y_n + \sum_{i=1}^m \sigma_i k_i,$$

где

$$k_i = hf(x_n + a_i h, y_n + \sum_{j=1}^m b_{ij} k_j).$$

**З а м е ч а н и е.** Для того чтобы реализовать неявные методы, в отличие от явных, необходимо решать нелинейную систему относительно  $k_i$ , поэтому эти методы сложны в реализации. Однако эти методы обладают рядом преимуществ, в частности, ими можно решать жесткие системы, этот вопрос рассмотрим позже.

Коэффициенты метода удобно задавать в виде таблицы — матрицы Бутчера:

|       |            |            |     |            |
|-------|------------|------------|-----|------------|
| $a_1$ | $b_{11}$   | $b_{12}$   | ... | $b_{1m}$   |
| $a_2$ | $b_{21}$   | $b_{22}$   | ... | $b_{2m}$   |
| .     | .          | .          | ... | .          |
| $a_m$ | $b_{m1}$   | $b_{m2}$   | ... | $b_{mm}$   |
|       | $\sigma_1$ | $\sigma_2$ | ... | $\sigma_m$ |

Для явных методов Рунге–Кутты таблица Бутчера выглядит так:

|       |            |            |     |            |
|-------|------------|------------|-----|------------|
| 0     | 0          | 0          | ... | 0          |
| $a_2$ | $b_{21}$   | 0          | ... | 0          |
| .     | .          | .          | ... | .          |
| $a_m$ | $b_{m1}$   | $b_{m2}$   | ... | 0          |
|       | $\sigma_1$ | $\sigma_2$ | ... | $\sigma_m$ |

Например, эта таблица для метода Эйлера с пересчетом запишется в виде

|               |               |   |
|---------------|---------------|---|
| 0             | 0             | 0 |
| 1             | 1             | 0 |
| $\frac{1}{2}$ | $\frac{1}{2}$ |   |

Изучим условия, обеспечивающие сходимость явных методов Рунге–Кутты (определение порядка сходимости и невязки было дано выше, при изучении метода Эйлера). Невязка явного метода Рунге–Кутты определяется как

$$\psi_n = y(x_{n+1}) - y(x_n) - \sum_{i=1}^m \sigma_i \hat{k}_i,$$

где

$$\hat{k}_1 = hf(x_n, y(x_n)),$$

$$\hat{k}_i = hf(x_n + a_i h, y(x_n) + \sum_{j=1}^{i-1} b_{ij} \hat{k}_j).$$

**Теорема 2.** *Если явный метод Рунге–Кутты имеет невязку порядка  $p+1$ , то он сходится с порядком  $p$ .*

**Доказательство.** Обозначим величину погрешности через

$$\varepsilon_n = y(x_n) - y_n$$

и оценим  $|\varepsilon_{n+1}|$  через  $|\varepsilon_n|$ . Выразим  $y(x_{n+1})$  из определения невязки, а  $y_{n+1}$  из определения метода

$$\begin{aligned} |\varepsilon_{n+1}| &= |y(x_{n+1}) - y_{n+1}| = |y(x_n) + \sum_{i=1}^m \sigma_i \hat{k}_i + \psi_n - y_n - \sum_{i=1}^m \sigma_i k_i| \leq \\ &\leq |\varepsilon_n| + \sum_{i=1}^m |\sigma_i| |\hat{k}_i - k_i| + |\psi_n| \leq |\varepsilon_n| + \sigma \sum_{i=1}^m |\hat{k}_i - k_i| + |\psi_n|, \end{aligned}$$

где  $\sigma = \max_i |\sigma_i|$ .

Оценим величины  $r_i = |\hat{k}_i - k_i|$ , используя липшицевость функции  $f(x, y)$  по второму аргументу:

$$\begin{aligned} r_i &= h|f(x_n + a_i h, y(x_n) + \sum_{j=1}^{i-1} b_{ij} \hat{k}_j) - f(x_n + a_i h, y_n + \sum_{j=1}^{i-1} b_{ij} k_j)| \leq \\ &\leq hL(|\varepsilon_n| + \sum_{j=1}^{i-1} |b_{ij}| |\hat{k}_j - k_j|) \leq hL(|\varepsilon_n| + b \sum_{j=1}^{i-1} r_j), \end{aligned}$$

где  $b = \max_{i,j} |b_{ij}|$ .

Индукцией по  $i$  покажем справедливость оценки

$$r_i \leq (1 + Lbh)^{i-1} Lh |\varepsilon_n|.$$

База индукции. При  $i = 1$

$$r_1 = h|f(x_n, y(x_n)) - f(x_n, y_n)| \leq Lh |\varepsilon_n|.$$

Шаг индукции. Пусть доказываемая оценка справедлива при индексах  $\leq i - 1$ , докажем ее для индекса  $i$ :

$$r_i \leq hL(|\varepsilon_n| + b \sum_{j=1}^{i-1} r_j) \leq hL |\varepsilon_n| (1 + Lbh \sum_{j=1}^{i-1} (1 + Lbh)^{j-1}).$$

Суммируя геометрическую прогрессию, получаем

$$r_i \leq (1 + Lbh)^{i-1} Lh |\varepsilon_n|.$$

Доказанную мажоранту для  $r_i$  подставим в оценку для  $|\varepsilon_{n+1}|$ :

$$|\varepsilon_{n+1}| \leq |\varepsilon_n| + \sigma \sum_{i=1}^m r_i + |\psi_n| \leq |\varepsilon_n| + \sigma Lh |\varepsilon_n| \sum_{i=1}^m (1 + Lbh)^{i-1} + |\psi_n|.$$

Отметим, что величина

$$\alpha(h) = \sum_{i=1}^m (1 + Lbh)^{i-1}$$

при любом шаге  $h$ , не превосходящем длины отрезка  $X$ , ограничена некоторой величиной  $\bar{\alpha}$  (так как число этапов  $m$  фиксировано), поэтому

$$|\varepsilon_{n+1}| \leq (1 + \sigma\bar{\alpha}Lh)|\varepsilon_n| + |\psi_n|.$$

По условию теоремы невязка имеет порядок  $p+1$ , т. е. найдется константа  $C$ , такая, что  $|\psi_n| \leq Ch^{p+1}$ , поэтому

$$|\varepsilon_{n+1}| \leq (1 + \sigma\bar{\alpha}Lh)|\varepsilon_n| + Ch^{p+1}.$$

Введем обозначения  $A = (1 + \sigma\bar{\alpha}Lh)$ ,  $B = Ch^{p+1}$ , тогда

$$|\varepsilon_{n+1}| \leq A|\varepsilon_n| + B.$$

Аналогично оценке, приведенной в доказательстве сходимости метода Эйлера, получим

$$|\varepsilon_n| \leq \frac{A^n - 1}{A - 1}B \leq \frac{A^N - 1}{A - 1}B.$$

Подставим в эту оценку выражения для  $A$  и  $B$ , воспользуемся также связью  $hN = X$ :

$$|\varepsilon_n| \leq \frac{(1 + \sigma\bar{\alpha}Lh)^{\frac{X}{h}} - 1}{Lh} Ch^{p+1}.$$

Используя свойство второго замечательного предела, получаем оценку

$$|\varepsilon_n| \leq \frac{C}{L\sigma\bar{\alpha}}(e^{\sigma\bar{\alpha}LX} - 1)h^p,$$

равномерную по всем  $n = 1, 2 \dots, N$ . Эта оценка означает сходимость метода с  $p$ -м порядком.

## 1.9. Явные методы Адамса

Методы Рунге–Кутты основаны на многократном пересчитывании на каждом шаге значений правой части дифференциальных уравнений. Методы Адамса основаны на запоминании

и использовании на каждом шаге уже посчитанных ранее значений правой части дифференциальных уравнений. Воспользуемся интегральным представлением точного решения:

$$y(x_{n+1}) = y(x_n) + \int_{x_n}^{x_{n+1}} f(s, y(s)) ds.$$

Пусть вычислены и запомнены значения  $f_n = f(x_n, y_n)$ ,  $f_{n-1} = f(x_{n-1}, y_{n-1})$ , ...,  $f_{n-k+1} = f(x_{n-k+1}, y_{n-k+1})$ . По этим данным построим интерполяционный многочлен Лагранжа  $L_{k-1}(x)$  степени  $k - 1$ .

*Явным методом Адамса*  $k$ -го порядка назовем метод

$$y_{n+1} = y_n + \int_{x_n}^{x_{n+1}} L_{k-1}(s) ds.$$

Выведем некоторые частные случаи при разных  $k$ . Число  $k$  называется *количество шагов*, а методы Адамса, в отличие от методов Рунге–Кутты, относятся к многошаговым методам.

Пусть  $k = 1$ . Тогда  $L_{k-1}(x) = f_n$  и получается явный метод Эйлера:

$$y_{n+1} = y_n + h f_n.$$

Пусть  $k = 2$ . Запишем многочлен Лагранжа  $L_{k-1}(x) = L_1(x)$  в форме Ньютона:

$$L_1(x) = f_n + f(x_n, x_{n-1})(x - x_n) = f_n + \frac{f_n - f_{n-1}}{h}(x - x_n),$$

подставим его в интеграл, посчитаем этот интеграл и приведем подобные. Получаем

$$y_{n+1} = y_n + \frac{h}{2}(3f_n - f_{n-1}),$$

этот метод называется *явным методом Адамса второго порядка*.

При  $k = 3$  многочлен  $L_{k-1}(x)$  имеет вид

$$\begin{aligned}
 L_2(x) &= f_n + f(x_n, x_{n-1})(x - x_n) + \\
 &+ f(x_n, x_{n-1}, x_{n-2})(x - x_n)(x - x_{n-2}) = \\
 &= f_n + \frac{f_n - f_{n-1}}{h}(x - x_n) + \frac{f_n - 2f_{n-1} + f_{n-2}}{2h^2}(x - x_n)(x - x_{n-1}).
 \end{aligned}$$

Подставляя его в интеграл, производя вычисления (интеграл от последнего слагаемого проще вычислять по частям) и приводя подобные, получаем трехшаговый метод Адамса:

$$y_{n+1} = y_n + \frac{h}{12}(23f_n - 16f_{n-1} + 5f_{n-2}),$$

который называется *явным методом Адамса третьего порядка* или методом Адамса–Бэшфорта.

**З а м е ч а н и е.** Отметим важную особенность всех методов Адамса, начиная с двухшаговых: кроме начального значения  $y_0$  требуется знание стартовых значений  $y_1, \dots, y_{k-1}$ , соответственно,  $f_1, \dots, f_{k-1}$ . Процедура вычисления стартовых значений называется *разгоном*. Рекомендуется делать разгон методами (например, Рунге–Кутты) порядка не ниже чем  $k$ , а лучше  $k + 1$ .

## 1.10. Неявные методы Адамса

Если к данным  $f_n, f_{n-1}, \dots, f_{n-k+1}$  добавить значение  $f_{n+1} = f(x_{n+1}, y_{n+1})$ , то формально можно построить интерполяционный многочлен Лагранжа на единицу большей степени  $L_k(x)$ .

*Неявными методами Адамса* назовем методы

$$y_{n+1} = y_n + \int_{x_n}^{x_{n+1}} L_k(s)ds,$$

эти методы еще называются *экстраполяционными* методами Адамса, в отличие от явных, которые называются *интерполяционными*.

Выведем некоторые частные случаи при разных  $k$ .

При  $k = 0$  имеем  $L_k(x) = f_{n+1}$  и получается метод

$$y_{n+1} = y_n + h f(x_{n+1}, y_{n+1}),$$

который называется *неявным методом Эйлера*.

При  $k = 1$  интерполяционный многочлен имеет вид

$$L_1(x) = f_{n+1} + f(x_{n+1}, x_n)(x - x_{n+1}) = f_n + \frac{f_{n+1} - f_n}{h}(x - x_{n+1}),$$

подставим его в интеграл, посчитаем этот интеграл и приведем подобные. Получаем

$$y_{n+1} = y_n + \frac{h}{2}(f_{n+1} + f_n),$$

этот метод называется *неявным методом Адамса второго порядка* или *неявным методом трапеций*.

При  $k = 2$  интерполяционный многочлен имеет вид

$$\begin{aligned} L_2(x) &= f_{n+1} + f(x_{n+1}, x_n)(x - x_{n+1}) + \\ &+ f(x_{n+1}, x_n, x_{n-1})(x - x_{n+1})(x - x_n) = \\ &= f_n + \frac{f_{n+1} - f_n}{h}(x - x_{n+1}) + \frac{f_{n+1} - 2f_n + f_{n-1}}{2h^2}(x - x_{n+1})(x - x_n). \end{aligned}$$

Подставляя его в интеграл, производя вычисления и приводя подобные, получаем метод

$$y_{n+1} = y_n + \frac{h}{12}(5f_{n+1} + 8f_n - f_{n-1}),$$

этот метод называется *неявным методом Адамса третьего порядка* или *методом Адамса-Мултона*.

Сравнивая явные и неявные методы Адамса, можно отметить следующее:

1. Недостаток неявных методов состоит в необходимости на каждом шаге решать уравнение относительно неизвестной величины  $y_{n+1}$ .

2. Некоторое преимущество неявных методов состоит в точности: при одной и той же шаговости  $k$  неявные методы имеют порядок сходимости  $k + 1$ , в отличие от явных, у которых порядок сходимости  $k$ .

3. Главное преимущество неявных методов состоит в возможности решать жесткие системы.

## 1.11. Общий класс многошаговых методов

*Явные  $k$ -шаговые методы* имеют вид

$$y_n = \sum_{i=1}^k \alpha_i y_{n-i} + h \sum_{i=1}^k \beta_i f_{n-i},$$

где  $f_{n-i} = f(x_{n-i}, y_{n-i})$ .

*Неявные  $k$ -шаговые методы* имеют вид

$$y_n = \sum_{i=1}^k \alpha_i y_{n-i} + h \sum_{i=0}^k \beta_i f_{n-i}.$$

Класс явных  $k$ -шаговых методов имеет  $2k$  параметров, а класс неявных  $k$ -шаговых методов имеет  $2k + 1$  параметров, поэтому за счет выбора этих параметров можно получить локальную погрешность гораздо более высокого порядка малости, чем, например, у методов Адамса. Подбор параметров удобно проводить, используя понятие алгебраической степени точности. Напомним, что число  $N$  называется *алгебраической степенью точности* формулы, если:

- 1) она точна для всех многочленов степени  $N$  и ниже;
- 2) среди многочленов степени  $N + 1$  найдется хотя бы один, для которого формула неточна.

Без доказательства приведем

**Утверждение.** Если алгебраическая степень точности метода  $N$ , то его невязка имеет порядок  $N + 1$ .

В качестве примера найдем метод с наивысшей алгебраической степенью точности в классе явных двухшаговых методов:

$$y_n = \alpha_1 y_{n-1} + \alpha_2 y_{n-2} + h\beta_1 f_{n-1} + h\beta_2 f_{n-2}.$$

Подставляя в метод последовательно многочлены  $y \equiv 1$ ,  $y = x - x_{n-1}$ ,  $y = (x - x_{n-1})^2$ ,  $y = (x - x_{n-1})^3$ , получаем систему уравнений

$$\begin{cases} 1 = \alpha_1 + \alpha_2 \\ h = -\alpha_2 h + h(\beta_1 + \beta_2) \\ h^2 = \alpha_2 h^2 + h(-2h\beta_2) \\ h^3 = -\alpha_2 h^3 + h(3h^2\beta_2), \end{cases}$$

решая которую, получаем

$$\begin{cases} \alpha_1 = -4 \\ \beta_1 = 4 \\ \alpha_2 = 5 \\ \beta_2 = 2. \end{cases}$$

Таким образом, получаем метод

$$y_n = -4y_{n-1} + 5y_{n-2} + h(4f_{n-1} + 2f_{n-2}),$$

который имеет алгебраическую степень точности  $N = 3$  (как нетрудно проверить, для многочленов степени 4 метод неточен). Согласно приведенному утверждению, этот метод имеет невязку (локальную погрешность) порядка 4, что на порядок выше явного двухшагового метода Адамса из того же класса методов.

Однако этот метод расходится, как можно проверить на тестовом примере.

Рассмотрим уравнение

$$y' = 0$$

с начальными условиями  $y(0) = y_0 = 0$ , которое имеет тривиальное решение  $y \equiv 0$ .

Выведенный метод для данного уравнения имеет вид

$$y_n = -4y_{n-1} + 5y_{n-2}.$$

Чтобы считать по этому методу, нужно в качестве разгона получить  $y_1$ , допустим разгон сделали с малой ошибкой  $y_1 = \varepsilon$ .

Считая по методу, получаем  $\varepsilon_0 = 0$ ,  $\varepsilon_1 = \varepsilon$ ,  $\varepsilon_2 = -4\varepsilon$ ,  $\varepsilon_3 = 21\varepsilon$ ,  $\varepsilon_4 = -104\varepsilon, \dots$ . Ошибка возрастает по абсолютной величине, метод расходится. Такое явление называется неустойчивостью метода. Чтобы привести условия, гарантирующие устойчивость и в конечном итоге — сходимость многошаговых методов, рассмотрим элементы теории линейных разностных уравнений.

## 1.12. Линейные разностные уравнения и их устойчивость

Рассмотрим линейные однородные разностные уравнения порядка  $k$  с постоянными коэффициентами вида:

$$y_n + a_1 y_{n-1} + \dots + a_k y_{n-k} = 0.$$

Если заданы начальные условия — числа  $y_0, y_1, \dots, y_{k-1}$ , то последующие члены последовательности  $y_k, y_{k+1}, \dots$  определяются по явной формуле

$$y_n = -(a_1 y_{n-1} + \dots + a_k y_{n-k})$$

и решением является последовательность  $\{y_n, n = 0, 1, \dots\}$ .

Отметим некоторые свойства решений:

1. Среди решений есть тривиальное  $\{y_n = 0, n = 0, 1, \dots\}$ .

2. Любая линейная комбинация решений является решением.

3. Все решения представляют собой линейную комбинацию линейно независимых решений, называемых фундаментальной системой решений.

Для нахождения фундаментальной системы решений будем искать решение в виде  $y_n = \rho^n$ , получаем уравнение

$$\rho^n + a_1\rho^{n-1} + \cdots + a_k\rho^{n-k} = 0,$$

или, после сокращения на  $\rho^{n-k}$ , алгебраическое уравнение

$$\rho^k + a_1\rho^{k-1} + \cdots + a_k = 0.$$

Это уравнение называется *характеристическим*.

Если  $\rho$  — корень характеристического уравнения (возможно, комплексный), то ему соответствует решение — последовательность  $y = \{1, \rho, \rho^2, \dots, \rho^n, \dots\}$ , которая остается ограниченной тогда и только тогда, когда  $|\rho| \leq 1$ .

Если же корень  $\rho$  имеет кратность  $l > 1$ , то ему соответствует  $l$  линейно независимых решений вида  $y^{(1)} = \{1, \rho, \rho^2, \dots, \rho^n, \dots\}$ ,  $y^{(2)} = \{0, 1\rho, 2\rho^2, \dots, n\rho^n, \dots\}$ , ...,  $y^{(l)} = \{0, 1^{(l-1)}\rho, 2^{(l-1)}\rho^2, \dots, n^{(l-1)}\rho^n, \dots\}$ , которые остаются ограниченными тогда и только тогда, когда  $|\rho| < 1$ .

Тривиальное решение разностного уравнения называется *устойчивым*, если для любого  $\varepsilon > 0$  найдется  $\delta > 0$ , такое, что для всякого решения разностного уравнения  $y = \{y_0, y_1, \dots, y_n, \dots\}$  с начальными условиями  $|y_i| < \delta$ ,  $i = 0, 1, \dots, k-1$ , выполняется  $|y_i| < \varepsilon$ ,  $i = k, k+1, \dots$ . В этом случае говорят также, что *разностное уравнение устойчиво*.

Из представлений фундаментальной системы решений вытекает утверждение.

**Критерий устойчивости.** Для того чтобы разностное уравнение было устойчивым, необходимо и достаточно, чтобы для всех корней характеристического уравнения выполнялось условие  $|\rho| \leq 1$ , причем если выполняется  $|\rho| = 1$ , то корень не кратный.

### 1.13. 0-устойчивость и сходимость многошаговых методов

Вернемся к изучению многошаговых методов

$$y_n = \sum_{i=1}^k \alpha_i y_{n-i} + h \sum_{i=0}^k \beta_i f_{n-i}.$$

Назовем метод *0-устойчивым*, если разностное уравнение, получающееся при  $h = 0$ , является устойчивым.

В качестве примера исследуем на 0-устойчивость рассмотренный выше метод

$$y_n = -4y_{n-1} + 5y_{n-2} + h(4f_{n-1} + 2f_{n-2}).$$

Разностное уравнение, получающееся при  $h = 0$ , имеет вид

$$y_n = -4y_{n-1} + 5y_{n-2},$$

соответствующее характеристическое уравнение

$$\rho^2 + 4\rho - 5 = 0$$

имеет корни  $\rho_1 = 1$ ,  $\rho_2 = -5$ , таким образом, условие устойчивости  $|\rho| \leq 1$  не выполняется, метод не является 0-устойчивым.

В противоположность этому все методы Адамса являются устойчивыми, так как соответствующее разностное уравнение имеет вид  $y_{n+1} = y_n$ , характеристическое уравнение  $\rho = 1$  сразу дает корень, удовлетворяющий условию устойчивости.

Справедлива теорема, которую приводим без доказательства.

**Теорема 3.** *Если многошаговый метод имеет невязку порядка  $p+1$ , разгон порядка  $p$  и является 0-устойчивым, то он сходится с порядком  $p$ .*

**З а м е ч а н и е.** Для 0-устойчивых  $k$ -шаговых методов выполняются барьеры Далквиста: 1) порядок сходимости  $p \leq k+2$  для четных  $k$ ; 2)  $p \leq k+1$  для нечетных  $k$ ; 3)  $p \leq k$ , если  $\beta_0 \leq 0$  (в частности, для явных методов). Поэтому, по сравнению с методами Адамса, в общем классе многошаговых методов можно добиться большей точности только для неявных методов и только для четных  $k$ .

### 1.14. Жесткие системы. А-устойчивость

Наличие жесткости создает большие проблемы при численном решении дифференциальных уравнений. Рассмотрим простой пример.

Пусть дано уравнение  $y' = -\alpha y$  с начальным условием  $y(0) = 1$  и пусть  $\alpha \gg 1$ . Точное решение этой задачи  $y = e^{-\alpha x}$ .

Применим к решению этой задачи явный метод Эйлера, получим

$$y_{n+1} = y_n - h\alpha y_n = (1 - \alpha h)y_n.$$

Таким образом, если шаг большой:  $h > \frac{1}{\alpha}$ , то не сохраняется даже такой важный показатель решения, как положительность. Разностное решение, возникающее при применении явного метода Эйлера, устойчиво, только если  $h \leq \frac{2}{\alpha}$ . Но малый шаг ведет к накоплению вычислительной погрешности и этот метод как с большими, так и с малыми шагами может дать очень неточные результаты. Конечно, в данном примере вопрос можно решить с помощью переменного шага, малого на почти вертикальном участке и большого на почти горизонтальном участке, но если решается система уравнений, то переменный шаг не поможет.

Применим к решению этой задачи неявный метод Эйлера, получим

$$y_{n+1} = y_n - h\alpha y_{n+1}$$

или

$$y_{n+1} = \frac{1}{1 + \alpha h} y_n.$$

Это разностное уравнение устойчиво, так как корень  $\rho = \frac{1}{1+\alpha h} < 1$  характеристического уравнения удовлетворяет условию  $0 < \rho < 1$  и уравнение можно решать неявным методом Эйлера с любым шагом.

На рис. 4 жирной линией показано точное решение задачи Коши при  $\alpha = 10$ , ломаными линиями показаны приближенные решения, полученные явным (кружки) и неявным (квадратики) методом Эйлера при шаге  $h = 0.2$ .

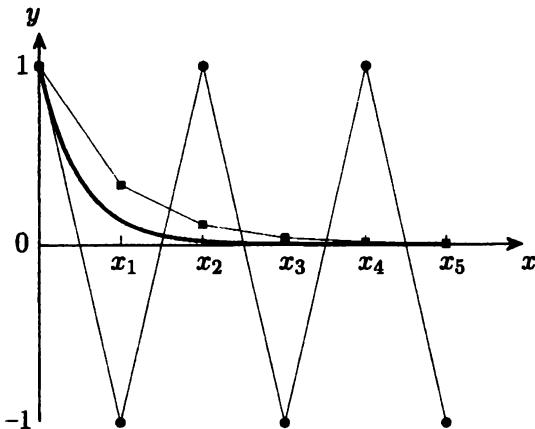


Рис. 4. Графики точного решения при  $\alpha = 10$ , а также приближенного, полученного явным и неявным методом Эйлера при  $h = 0.2$

Данный пример относится к числу жестких. Не существует единого определения жестких систем, приведем некоторые:

1. Жесткие системы — системы с наличием почти вертикальных и почти горизонтальных участков решения.
2. Жесткие системы — системы с большой константой Липшица.
3. Жесткие системы — системы, которые нельзя решать явными методами.

Приведем еще одно определение, которое является достаточно распространенным.

Рассмотрим систему дифференциальных уравнений в нормальной форме

$$y' = f(x, y), \quad y \in R^n$$

с начальными условиями  $y(x_0) = y_0$ . Пусть  $y(x)$  — точное решение этой задачи. Составим систему первого приближения вдоль этого решения

$$y' = A(x)y,$$

и пусть  $\lambda_i = \lambda_i(x)$  — собственные числа матрицы  $A(x)$ .

Если для некоторого  $x$  выполняются условия:

- 1)  $Re\lambda_i < 0$ ;
- 2)  $\frac{\max Re\lambda_i}{\min Re\lambda_i} \gg 1$ ,

то система называется *жесткой*.

Для решения жестких систем рекомендуется применять методы со специальным свойством *A-устойчивости*.

Рассмотрим некоторый метод, например многошаговый

$$y_n = \sum_{i=1}^k \alpha_i y_{n-i} + h \sum_{i=0}^k \beta_i f_{n-i},$$

и применим его к решению тестового уравнения

$$y' = \lambda y,$$

$\lambda$  — комплексное число. Получим разностное уравнение

$$y_n = \sum_{i=1}^k \alpha_i y_{n-i} + h\lambda \sum_{i=0}^k \beta_i y_{n-i}.$$

Областью устойчивости метода называется множество параметров  $h\lambda$  на комплексной плоскости, таких, что разностное уравнение устойчиво. Метод называется *A-устойчивым*, если область устойчивости содержит всю левую комплексную полу-плоскость  $Re(h\lambda) < 0$ .

В качестве примера исследуем на *A-устойчивость* явный метод Эйлера  $y_{n+1} = y_n + hf_n$ . Применяя его к тестовому уравнению, получаем  $y_{n+1} = y_n + h\lambda y_n$ , корень характеристического

уравнения  $\rho = 1 + h\lambda$ . Область устойчивости  $|1 + h\lambda| < 1$  представляет собой на комплексной плоскости внутренность единичного круга с центром в точке  $-1$  и не содержит всю левую комплексную полуплоскость (рис. 5а). Метод не является А-устойчивым.

Исследуем на А-устойчивость неявный метод Эйлера  $y_{n+1} = y_n + hf_{n+1}$ . Применяя его к тестовому уравнению, получаем  $y_{n+1} = y_n + h\lambda y_{n+1}$ , корень характеристического уравнения  $\rho = \frac{1}{1-h\lambda}$ . Область устойчивости  $|1 - h\lambda| > 1$  представляет собой на комплексной плоскости внешность единичного круга с центром в точке  $1$  и содержит всю левую комплексную полуплоскость (рис. 5б). Метод является А-устойчивым.

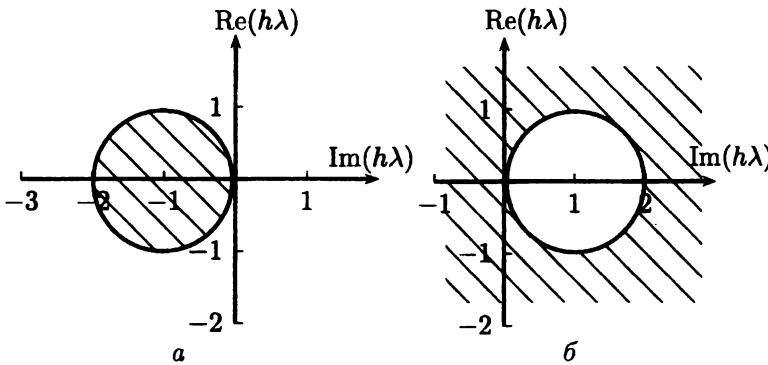


Рис. 5. Область устойчивости: а — явного метода Эйлера; б — неявного метода Эйлера

## 2. Численные методы решения краевых задач

### 2.1. Постановка краевых задач. Метод стрельбы

Краевые задачи для обыкновенных дифференциальных уравнений, в отличие от задачи Коши, имеют гораздо менее развитый аппарат аналитических методов решения, между тем в приложениях краевые задачи встречаются не реже. Поэтому на первый план выходят численные методы решения.

Рассмотрим дифференциальное уравнение второго порядка

$$y'' = f(x, y, y').$$

Простейшие краевые условия (первого рода) имеют вид

$$\begin{cases} y(a) = \alpha \\ y(b) = \beta. \end{cases}$$

Будем предполагать, что решение краевой задачи существует и единственно на отрезке  $[a, b]$  и обладает необходимой гладкостью.

Рассмотрим метод, который позволяет свести решение краевой задачи к решению задачи Коши и нелинейного уравнения. Введем параметр  $\mu$  и рассмотрим вспомогательную задачу Коши

$$\begin{cases} y'' = f(x, y, y') \\ y(a) = \alpha \\ y'(a) = \mu. \end{cases}$$

Эту задачу можно решать численно, например методами, рассмотренными в первой главе, следует только учитывать, что дифференциальное уравнение второго порядка нужно предварительно свести к системе в нормальной форме второго порядка. Предположим, что решение задачи Коши  $y(x, \mu)$  найдено.

Теперь нужно подобрать параметр  $\mu = \mu^*$ , чтобы выполнялось условие

$$y(b, \mu^*) = \beta,$$

для этого нужно решить нелинейное уравнение, что также можно сделать численными методами. Эта методика решения

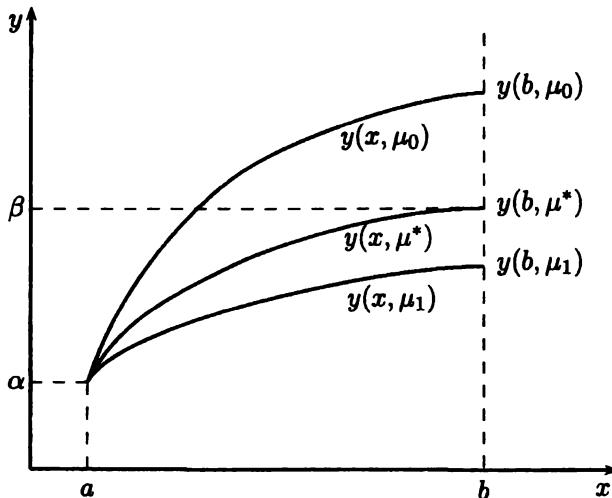


Рис. 6. Геометрическая интерпретация метода стрельбы

краевых задач называется **методом стрельбы** (рис. 6). Метод стрельбы представляет собой пару вложенных методов: внешний — для решения нелинейного уравнения, внутренний — для решения задачи Коши (например, метод деления отрезка пополам и явный метод Эйлера).

Особый интерес представляет алгоритм в случае, когда применяется метод Ньютона для решения нелинейного уравнения. Обозначим  $F(\mu) = y(b, \mu) - \beta$ , тогда метод Ньютона состоит в вычислениях по алгоритму  $\mu_{k+1} = \mu_k - \frac{F(\mu_k)}{F'(\mu_k)}$ , поэтому необходимо на каждом шаге получить  $F'(\mu_k) = y'_\mu(b, \mu_k)$ . Для того чтобы вычислить эту величину, воспользуемся урав-

нением в вариациях. В предположении достаточной гладкости продифференцируем уравнения задачи Коши по  $\mu$ :

$$\begin{cases} y'''_{xx\mu}(x, \mu) = f'_y y'_\mu + f'_{y'} y''_{x\mu} \\ y'_\mu(a, \mu) = 0 \\ y''_{x,\mu}(a, \mu) = 1. \end{cases}$$

Обозначим  $z(x, \mu) = y'_\mu(x, \mu)$ , тогда получаем задачу Коши для линейного уравнения второго порядка

$$\begin{cases} z''_{xx} = f'_y z + f'_{y'} z'_x \\ z(a, \mu) = 0 \\ z'_x(a, \mu) = 1. \end{cases}$$

Таким образом, метод стрельбы с применением метода Ньютона сводит решение краевой задачи к решению двух вложенных задач Коши.

## 2.2. Метод разностной прогонки

Рассмотрим линейное уравнение второго порядка

$$y'' = p(x)y + q(x),$$

с краевыми условиями первого рода

$$\begin{cases} y(a) = \alpha \\ y(b) = \beta. \end{cases}$$

Будем предполагать, что функции  $p(x) \geq \bar{p} > 0$  и  $q(x)$  непрерывны на  $[a, b]$ , решение краевой задачи существует и единственно на отрезке  $[a, b]$  и обладает необходимой гладкостью.

Разобьем отрезок  $[a, b]$  на  $N$  равных частей с шагом  $h = \frac{b-a}{N}$ , обозначив узлы  $x_i = a + ih$ ,  $i = 0, \dots, N$ . Приближенные значения решения в узлах будем обозначать  $y_i$ , в отличие от

точных значений решения в узлах  $y(x_i)$ ,  $i = 0, \dots, N$ . Задача состоит в конструировании алгоритма для вычисления  $y_i$ .

Обозначим  $p_i = p(x_i)$ ,  $q_i = q(x_i)$  — это известные величины при всех  $i = 0, \dots, N$ . Во внутренних узлах заменим вторую производную по формуле численного дифференцирования по трем узлам на середину

$$y''(x_i) \approx \frac{y_{i-1} - 2y_i + y_{i+1}}{h^2}, \quad i = 1, \dots, N-1.$$

В результате получим систему уравнений

$$\begin{cases} y_0 = \alpha, \\ \frac{y_{i-1} - 2y_i + y_{i+1}}{h^2} = p_i y_i + q_i, \quad i = 1, \dots, N-1, \\ y_N = \beta. \end{cases}$$

Данная линейная неоднородная система имеет основную матрицу трехдиагональной структуры и может быть эффективно решена методом прогонки. В самом деле,  $i$ -е уравнение имеет вид

$$y_{i-1} - (2 + p_i h^2) y_i + y_{i+1} = q_i h^2,$$

следовательно, при условии  $p(x) \geq \bar{p} > 0$  имеет диагональное преобладание  $\Delta = \bar{p}h^2$ .

Оценим погрешности  $\varepsilon_i = y(x_i) - y_i$  метода. Учитывая, что для внутренних узлов выполняется

$$y''(x_i) = \frac{y(x_{i-1}) - 2y(x_i) + y(x_{i+1})}{h^2} + R_i, \quad i = 1, \dots, N-1,$$

где погрешность формулы численного дифференцирования для второй производной определяется оценкой

$$|R_i| \leq \frac{1}{12} M_4 h^2, \quad M_4 = \max_{x \in [a, b]} |f'''(x)|,$$

получаем для погрешностей систему уравнений

$$\begin{cases} \varepsilon_0 = 0, \\ \varepsilon_{i-1} - (2 + p_i h^2) \varepsilon_i + \varepsilon_{i+1} = -R_i h^2, \quad i = 1, \dots, N-1, \\ \varepsilon_N = 0. \end{cases}$$

Используя результат об оценке решений системы с трехдиагональной матрицей, получаем

$$\max_{0 \leq i \leq N} |\varepsilon_i| \leq \max_{1 \leq i \leq N-1} \left\{ \frac{|R_i|}{p_i} \right\} \leq \frac{h^2}{12\bar{p}} M_4,$$

т. е. алгоритм сходится со вторым порядком.

### 2.3. Метод разностной прогонки.

#### Краевые условия третьего рода

Рассмотрим то же линейное уравнение второго порядка

$$y'' = p(x)y + q(x),$$

с краевыми условиями третьего рода на каждом из концов отрезка

$$\begin{cases} y'(a) = \alpha_0 y(a) + \alpha_1, \quad \alpha_0 > 0, \\ y'(b) = -\beta_0 y(b) + \beta_1, \quad \beta_0 > 0. \end{cases}$$

Будем предполагать, что функции  $p(x) \geq \bar{p} > 0$  и  $q(x)$  непрерывны на  $[a, b]$ , решение краевой задачи существует и единственno на отрезке  $[a, b]$  и обладает необходимой гладкостью.

Будем использовать те же обозначения, что и в методе разностной прогонки для краевой задачи первого рода, воспользуемся той же формулой численного дифференцирования для второй производной во внутренних узлах, а для производных, входящих в краевые условия, воспользуемся формулами численного дифференцирования на край, соответственно

$$y'(a) \approx \frac{y_1 - y_0}{h}, \quad y'(b) \approx \frac{y_N - y_{N-1}}{h}.$$

В результате получим систему уравнений

$$\begin{cases} (1 + \alpha_0 h)y_0 - y_1 = -\alpha_1 h, \\ y_{i-1} - (2 + p_i h^2)y_i + y_{i+1} = q_i h^2, \quad i = 1, \dots, N-1, \\ -y_{N-1} + (1 + \beta_0 h)y_N = \beta_1 h, \end{cases}$$

которая имеет трехдиагональную структуру и (при указанных выше условиях) диагональное преобладание, поэтому ее эффективно можно решать методом прогонки.

Оценим погрешности метода, учитывая не только погрешности формулы численного дифференцирования для второй производной во внутренних узлах, но и погрешности формул численного дифференцирования для первой производной на краях:

$$y'(a) = \frac{y(x_1) - y(x_0)}{h} + r_0, \quad y'(b) = \frac{y(x_N) - y(x_{N-1})}{h} + r_N,$$

где погрешность формулы численного дифференцирования на край для первой производной определяется оценкой

$$|r_0| \leq \frac{1}{2} M_2 h, \quad |r_N| \leq \frac{1}{2} M_2 h, \quad M_2 = \max_{x \in [a, b]} |f''(x)|.$$

Получаем для погрешностей систему уравнений

$$\begin{cases} (1 + \alpha_0 h) \varepsilon_0 - \varepsilon_1 = r_0, \\ \varepsilon_{i-1} - (2 + p_i h^2) \varepsilon_i + \varepsilon_{i+1} = -R_i h^2, \quad i = 1, \dots, N-1, \\ -\varepsilon_{N-1} + (1 + \beta_0 h) \varepsilon_N = -r_N. \end{cases}$$

Используя результат об оценке решений системы с трехдиагональной матрицей, получаем

$$\begin{aligned} \max_{0 \leq i \leq N} |\varepsilon_i| &\leq \max\left\{\frac{|r_0|}{\alpha_0}, \frac{|r_N|}{\beta_0}, \max_{1 \leq i \leq N-1} \left\{\frac{|R_i|}{p_i}\right\}\right\} \leq \\ &\leq \max\left\{\frac{h}{2\alpha_0} M_2, \frac{h}{2\beta_0} M_2, \frac{h^2}{12\bar{p}} M_4\right\}, \end{aligned}$$

т. е. алгоритм сходится только с первым порядком.

Для восстановления второго порядка сходимости используют разные способы: формулы дифференцирования для первой производной по трем узлам на край или *метод фиктивного узла*.

Рассмотрим идею метода фиктивного узла применительно к левой точке  $a = x_0$ . Введем фиктивный узел  $x_{-1} = a - h$  и заменим производную в точке  $a = x_0$  по формуле численного дифференцирования по двум узлам на середину

$$y'(a) \approx \frac{y_1 - y_{-1}}{2h}.$$

Кроме того, уравнение для внутренних узлов распространим и на узел  $x_0$ . Тогда первые два уравнения системы будут выглядеть так:

$$\begin{cases} -y_{-1} - 2\alpha_0 h y_0 + y_1 = 2\alpha_1 h, \\ y_{-1} - (2 + p_0 h^2) y_0 + y_1 = q_0 h^2, \\ \dots \dots \dots \end{cases}$$

Исключая из этой системы уравнений  $y_{-1}$ , получаем первое уравнение системы

$$-(2 + 2\alpha_0 h + p_0 h^2) y_0 + y_1 = 2\alpha_1 h + q_0 h^2,$$

таким образом, восстановили трехдиагональность и диагональное преобладание. Погрешность этого метода, как можно проверить, имеет второй порядок малости по  $h$ , так как погрешность формулы численного дифференцирования по двум узлам на середину имеет второй порядок малости по  $h$ .

## 2.4. Метод разностной прогонки для нелинейного уравнения

Рассмотрим нелинейное уравнение второго порядка

$$y'' = f(x, y),$$

с краевыми условиями первого рода

$$\begin{cases} y(a) = \alpha \\ y(b) = \beta. \end{cases}$$

После дискретизации и замены второй производной по формуле численного дифференцирования по трем узлам на середину возникает система нелинейных уравнений

$$\begin{cases} y_0 = \alpha, \\ \frac{y_{i-1} - 2y_i + y_{i+1}}{h^2} = f(x_i, y_i), \quad i = 1, \dots, N-1, \\ y_N = \beta. \end{cases}$$

Применим итерационный метод к решению этой системы:

$$\begin{cases} y_0^{k+1} = \alpha, \\ \frac{y_{i-1}^{k+1} - 2y_i^{k+1} + y_{i+1}^{k+1}}{h^2} = f(x_i, y_i^k), \quad i = 1, \dots, N-1, \\ y_N^{k+1} = \beta, \end{cases}$$

и на каждом шаге получаем систему линейных уравнений с трехдиагональной матрицей относительно  $y_i^{k+1}$ , которую можно решать методом прогонки. При определенных условиях метод сходится.

## 2.5. Самосопряженная краевая задача и ее физическая интерпретация

В следующем методе будем рассматривать линейную краевую задачу вида (это уравнение называется *самосопряженным*):

$$(k(x)y')' = p(x)y + q(x),$$

с краевыми условиями первого рода

$$\begin{cases} y(a) = \alpha \\ y(b) = \beta. \end{cases}$$

Несмотря на то, что самосопряженное уравнение не является самым общим в классе линейных дифференциальных уравнений второго порядка, оно широко применяется в моделировании. Рассмотрим в качестве примера процесс распространения

тепла в стержне длиной  $l = b - a$ . Пусть пространственная координата  $x \in [a, b]$ , временная независимая переменная  $t$ , искомая функция — температура  $y = y(x, t)$ . Пусть на концах стержня поддерживается постоянная температура, соответственно на левом конце  $\alpha$ , а на правом конце  $\beta$ , коэффициент теплопроводности точек стержня  $k(x)$ , коэффициент теплообмена с внешней средой  $p(x)$ , плотность (мощность) источников тепла или стоков тепла  $q(x)$  (рис. 7). Тогда процесс распространения

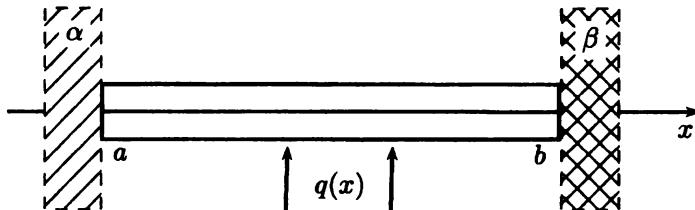


Рис. 7. Иллюстрация к физическому содержанию краевой задачи тепла в стержне описывается уравнением теплопроводности

$$-\frac{\partial y}{\partial t} + \frac{\partial}{\partial x} \left( k(x) \frac{\partial y}{\partial x} \right) = p(x)y + q(x),$$

однако если задача стационарная (коэффициенты не зависят от времени), то устанавливается стационарный режим, который описывается указанной самосопряженной задачей.

Отметим, что краевые условия можно свести к нулевым (однородным) путем замены  $y = z + \varphi(x)$ ,  $\varphi(x) = \alpha \frac{b-x}{b-a} + \beta \frac{x-a}{b-a}$ , тогда  $z(a) = z(b) = 0$ .

## 2.6. Вариационный метод Ритца

Рассмотрим однородную самосопряженную задачу

$$\begin{cases} (k(x)y')' = p(x)y + q(x) \\ y(a) = 0 \\ y(b) = 0. \end{cases}$$

Воспользуемся следующим результатом теории вариационного исчисления.

Простейшая задача вариационного исчисления состоит в том, чтобы минимизировать функционал

$$\int_a^b F(x, y(x), y'(x)) dx$$

на множестве дважды непрерывно дифференцируемых на отрезке  $[a, b]$  функций  $y(x)$  с нулевыми значениями на концах отрезка (это множество будем обозначать  $Y$ ). Основной результат теории вариационного исчисления заключается в необходимых условиях: если функция  $y(x)$  решает задачу вариационного исчисления, то она удовлетворяет уравнению Эйлера

$$\frac{d}{dx}(F'_{y'}) = F'_y.$$

Если в задаче вариационного исчисления определить функцию  $F(x, y(x), y'(x)) = k(x)[y'(x)]^2 + p(x)y^2 + 2q(x)y(x)$ , то решение задачи

$$J[y(\cdot)] = \int_a^b (k(x)[y'(x)]^2 + p(x)y^2 + 2q(x)y(x)) dx \rightarrow \min, \quad y(\cdot) \in Y,$$

является в силу уравнения Эйлера решением исходной краевой задачи.

В дальнейшем предположим, что непрерывно дифференцируемый коэффициент  $k(x)$  удовлетворяет условиям  $k(x) \geq \bar{k} > 0$ , коэффициенты  $p(x)$  и  $q(x)$  непрерывны, причем  $p(x) \geq 0$ .

Пусть  $Y_n$  конечномерное подпространство пространства  $Y$ , которое образовано линейными комбинациями известных функций:

$$Y_n = \{y \in Y, \quad y = \sum_{i=1}^n c_i \varphi_i(x)\},$$

тогда задачу минимизации функционала  $J[y(\cdot)]$  на множестве  $Y$  можно заменить приближенной задачей

$$J[y(\cdot)] \rightarrow \min, \quad y(\cdot) \in Y_n.$$

Будем предполагать, что система функций  $\varphi_i(x)$  линейно независима (тогда их называют базисными функциями).

Будем также предполагать, что при увеличении  $n$  последовательность пространств  $\{Y_n\}$  лежит плотно в  $Y$ , т. е. для любого  $\varepsilon > 0$  и для любого  $y \in Y$  найдутся такие  $n$  и  $y_n \in Y_n$ , что выполняется

$$\max_{x \in [a, b]} \|y(x) - y_n(x)\| < \varepsilon, \quad \max_{x \in [a, b]} \|y'(x) - y'_n(x)\| < \varepsilon.$$

Введем на  $Y$  и его подмножестве  $Y_n$  два скалярных произведения (при указанных условиях, как можно проверить, аксиомы скалярного произведения выполняются):

$$(\varphi, \psi)_1 = \int_a^b \varphi(x)\psi(x)dx,$$

$$(\varphi, \psi)_2 = \int_a^b (k(x)\varphi'(x)\psi'(x) + p(x)\varphi(x)\psi(x))dx,$$

тогда минимизируемый функционал можно представить в виде

$$J[y(\cdot)] = \int_a^b (k(x)[y'(x)]^2 + p(x)y^2 + 2q(x)y(x))dx = (y, y)_2 + 2(y, q)_1.$$

Заменив в этом выражении  $y = \sum_{i=1}^n c_i \varphi_i(x)$ , получаем, используя свойства скалярных произведений,

$$J[y] = (\sum_{i=1}^n c_i \varphi_i, \sum_{i=1}^n c_i \varphi_i)_2 + 2(\sum_{i=1}^n c_i \varphi_i, q)_1 =$$

$$= \sum_{i,k=1}^n c_i c_k (\varphi_i, \varphi_k)_2 + 2 \sum_{i=1}^n c_i (\varphi_i, q)_1.$$

Это выражение представляет собой линейно-квадратичную форму относительно неизвестных  $c_1, c_2, \dots, c_n$ .

Введем обозначения

$$G = \begin{pmatrix} g_{11} & \dots & g_{1n} \\ \vdots & \dots & \vdots \\ g_{n1} & \dots & g_{nn} \end{pmatrix}, \quad c = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}, \quad b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix},$$

где элементы матрицы  $G$  и вектора  $b$  определяются соотношениями  $g_{ik} = (\varphi_i, \varphi_k)_2$ ,  $b_i = (\varphi_i, q)_1$ , тогда  $F = (c, Gc) + 2(c, b)$ , где  $(c, b)$  обозначает скалярное произведение векторов в конечномерном евклидовом пространстве  $R^n$ .

Таким образом, получаем задачу

$$F(c) = (c, Gc) + 2(c, b) \rightarrow \min, \quad c \in R^n.$$

Отметим, что матрица  $G$  является, во-первых, симметричной, это следует из свойств скалярного произведения. Во-вторых, если система базисных функций  $\varphi_i(x)$  линейно независима, то матрица  $G$  является положительно определенной. В самом деле,

$$(c, Gc) = \left( \sum_{i=1}^n c_i \varphi_i, \sum_{i=1}^n c_i \varphi_i \right)_2 \geq 0.$$

При этом

$$(c, Gc) = \left( \sum_{i=1}^n c_i \varphi_i, \sum_{i=1}^n c_i \varphi_i \right)_2 = 0$$

тогда и только тогда, когда  $\sum_{i=1}^n c_i \varphi_i = 0$ , а это, в силу линейной независимости  $\varphi_i(x)$ , выполняется тогда и только тогда, когда  $c_1 = c_2 = \dots = c_n = 0$ , что по определению означает положительную определенность матрицы  $G$ .

Докажем вспомогательный результат из теории квадратичных форм.

**Лемма.** Если  $F(c) = (c, Gc) + 2(c, b)$ , где матрица  $G$  симметрична и положительно определена, то существует единственное решение задачи минимизации  $F(c)$ , равное  $\bar{c} = -G^{-1}b$ .

**Доказательство.** Представим минимизируемую функцию в виде

$$\begin{aligned} F(c) &= F(c - \bar{c} + \bar{c}) = (c - \bar{c} + \bar{c}, G(c - \bar{c} + \bar{c})) + 2(c - \bar{c} + \bar{c}, b) = \\ &= (\bar{c}, G\bar{c}) + 2(\bar{c}, b) + 2(c - \bar{c}, G\bar{c}) + 2(c - \bar{c}, b) + (c - \bar{c}, G(c - \bar{c})) = \\ &= F(\bar{c}) + (c - \bar{c}, G(c - \bar{c})). \end{aligned}$$

В силу положительной определенности  $G$  последнее слагаемое неотрицательно, причем оно равно нулю тогда и только тогда, когда  $c = \bar{c}$ , отсюда вытекает заключение леммы.

Используя эту лемму, получаем, что решение задачи минимизации  $F$  достигается на решении линейной системы

$$Gx = -b.$$

Таким образом, приближенное решение краевой задачи свелось к решению линейной системы, этот метод называется *вариационным методом Ритца*.

## 2.7. Метод Галеркина

Рассмотрим краевую задачу с однородными краевыми условиями

$$\begin{cases} Ly = q(x) \\ y(a) = 0 \\ y(b) = 0, \end{cases}$$

где  $Ly = a(x)y'' + b(x)y' + c(x)y$  — линейный дифференциальный оператор второго порядка, не обязательно самосопряженный. Требуется найти решения в пространстве  $Y$  (используем обозначения предыдущего раздела).

Пусть  $Y_n$  — конечномерное подпространство пространства  $Y$ , которое образовано линейными комбинациями базисных (линейно независимых) функций:

$$Y_n = \{y_n \in Y, y_n = \sum_{k=1}^n c_k \varphi_k(x)\},$$

будем находить приближенное решение  $y_n$  в этом пространстве.

*Невязкой* назовем величину

$$\mu(x) = Ly_n - q(x).$$

Так как невязка — функция, то можно предложить разные подходы к уменьшению невязки.

*Метод Галеркина* основан на том, чтобы сделать невязку ортогональной некоторой (вообще говоря, другой) системе базисных функций  $\{\psi_i, i = 1, 2, \dots, n\}$  из пространства  $Y$ : потребуем, чтобы

$$(Ly_n - q(x), \psi_i)_1 = 0, i = 1, 2, \dots, n.$$

Подставив для  $y_n$  разложение по системе базисных функций  $\{\varphi_k\}$  и воспользовавшись линейной независимостью функций  $\{\psi_i\}$ , получаем систему уравнений относительно коэффициентов разложения  $\{c_k\}$ :

$$\sum_{k=1}^n (L\varphi_k, \psi_i)_1 c_k = (q, \psi_i)_1, i = 1, 2, \dots, n.$$

**З а м е ч а н и е.** Метод Ритца является частным случаем метода Галеркина, если, во-первых, уравнение самосопряженное, а, во-вторых, системы базисных функций  $\{\psi_i\}$  и  $\{\varphi_i\}$  совпадают.

## 2.8. Метод наименьших квадратов для решения краевой задачи

Рассмотрим краевую задачу как в предыдущем разделе и будем минимизировать скалярное произведение  $F$  невязки самой на себя. Воспользовавшись свойствами скалярных произведений и линейностью оператора  $L$ , получаем

$$\begin{aligned} F &= (Ly_n - q, Ly_n - q)_1 = \left( \sum_{k=1}^n c_k L\varphi_k - q, \sum_{i=1}^n c_i L\varphi_i - q \right)_1 = \\ &= \sum_{k,i=1}^n (L\varphi_k, L\varphi_i)_1 c_k c_i - 2 \sum_{k=1}^n (L\varphi_k, q)_1 c_k + (q, q)_1. \end{aligned}$$

Это выражение представляет собой линейно-квадратичную форму относительно коэффициентов разложения. Введем обозначения

$$G = \begin{pmatrix} g_{11} & \dots & g_{1n} \\ \vdots & \dots & \vdots \\ g_{n1} & \dots & g_{nn} \end{pmatrix}, \quad c = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}, \quad b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix},$$

где элементы матрицы  $G$  и вектора  $b$  определяются соотношениями  $g_{ik} = (L\varphi_i, L\varphi_k)_1$ ,  $b_i = (L\varphi_i, q)_1$ , тогда  $F = (c, Gc) - 2(c, b) + (q, q)_1$ . При этом матрица  $G$  положительно определена тогда и только тогда, когда  $L\varphi_1, L\varphi_2, \dots, L\varphi_n$  линейно независимы. Если это условие выполняется, то воспользовавшись леммой о минимизации линейно-квадратичной формы (она доказана в методе Ритца), получаем, что метод **наименьших квадратов** сводится к решению линейной системы

$$Gx = b.$$

**З а м е ч а н и е.** Метод наименьших квадратов является частным случаем метода Галеркина, если в качестве системы базисных функций  $\{\psi_i\}$  выбрать  $\{L\varphi_i\}$ .

## 2.9. Метод коллокации

Рассмотрим краевую задачу как в методе Галеркина и за-нулим невязку в некоторых узлах, называемых *узлами коллокации*. Пусть узлы  $a \leq x_1 \leq x_2 \leq \dots \leq x_n = b$ , тогда

$$Ly_n(x)|_{x=x_i} = q(x_i), \quad i = 1, 2, \dots, n.$$

Подставляя для  $y_n$  линейную комбинацию базисных функций и пользуясь свойствами скалярных произведений и линейностью оператора  $L$ , получаем систему относительно неизвестных коэффициентов разложения

$$\sum_{k=1}^n c_k L\varphi_k(x)|_{x=x_i} = q(x_i), \quad i = 1, 2, \dots, n.$$

Главная проблема в этом *методе коллокации* состоит в таком выборе узлов, чтобы матрица полученной системы была невырожденной, а также в обосновании сходимости метода при данном способе увеличения количества узлов.

### 3. Интерполяция сплайнами

#### 3.1. Интерполяционный кубический сплайн

Рассмотрим задачу интерполяции: построить на отрезке  $[a, b]$  функцию  $f(x)$ , удовлетворяющую интерполяционным условиям  $f(x_i) = y_i$ ,  $i = 0, 1, \dots, n$ , где  $a = x_0 < x_1 < \dots < x_n = b$ . Эта задача была рассмотрена в пособии ранее в классе многочленов  $n$ -й степени, при этом решение было дано в форме Лагранжа и в форме Ньютона. Такая глобальная интерполяция обладает существенным недостатком: при больших  $n$  возникает большая погрешность, обусловленная неустойчивостью при решении систем большой размерности. Небольшая погрешность в исходных данных всего в одном узле может сильно возрасти при построении многочлена высокой степени, удовлетворяющего интерполяционным условиям. В противоположность этому локальная интерполяция, например кусочно-линейная, не учитывает погрешности в дальних узлах и поэтому устойчива к ней. Однако такая интерполяция дает малую точность, так как в узлах кусочно-линейной интерполяции результат интерполяции недостаточно гладкий. В приложениях хорошо зарекомендовал себя промежуточный вариант — интерполяция сплайнами, которая сочетает в себе достоинства глобальной и локальной интерполяции.

Сплайном назовем функцию, заданную кусочно многочленами на отрезках  $[x_{i-1}, x_i]$ ,  $i = 1, \dots, n$ . Среди всех сплайнов наибольшее распространение получили интерполяционные кубические сплайны. *Интерполяционным кубическим сплайнам* назовем дважды непрерывно дифференцируемую внутри отрезка  $[a, b]$  функцию, заданную на каждом отрезке  $[x_{i-1}, x_i]$  кубическим многочленом и удовлетворяющую интерполяционным условиям. В дальнейшем, так как будем рассматривать только интерполяционные кубические сплайны, коротко будем называть их просто сплайнами.

По определению сплайн  $S(x)$  задается формулами

$$S(x) = S_i(x), x \in [x_{i-1}, x_i], i = 1, \dots, n,$$

$$S_i(x) = A_i x^3 + B_i x^2 + C_i x + D_i, i = 1, \dots, n.$$

Всего формулы для сплайна содержат  $4n$  неизвестных коэффициентов. Выпишем для них уравнения. Из условий интерполяции

$$S_i(x_{i-1}) = y_{i-1}, i = 1, \dots, n,$$

$$S_i(x_i) = y_i, i = 1, \dots, n.$$

Из условий дважды непрерывной дифференцируемости выполняется

$$S'_i(x_i) = S'_{i+1}(x_i), i = 1, \dots, n-1,$$

$$S''_i(x_i) = S''_{i+1}(x_i), i = 1, \dots, n-1.$$

Итого  $4n - 2$  уравнения. Недостающие два уравнения задают разными способами.

*Дополнительные условия первого типа:*

$$\begin{cases} S'(a) = \alpha \\ S'(b) = \beta. \end{cases}$$

*Дополнительные условия второго типа:*

$$\begin{cases} S''(a) = \alpha \\ S''(b) = \beta. \end{cases}$$

*Дополнительные условия третьего типа* (используют для интерполяции периодических функций):

$$\begin{cases} S'(a) = S'(b) \\ S''(a) = S''(b). \end{cases}$$

*Дополнительные условия четвертого типа* (требование дополнительной гладкости в первом и предпоследнем узлах):

$$\begin{cases} S'''(x_1 - 0) = S'''(x_1 + 0) \\ S'''(x_{n-1} - 0) = S'''(x_{n-1} + 0). \end{cases}$$

В дальнейшем мы будем рассматривать дополнительные условия второго типа. Теперь имеем систему линейных уравнений для определения коэффициентов сплайна с  $4n$  неизвестными, в которой столько же уравнений. Главный эффект в построении сплайнов состоит в том, что за счет переобозначения неизвестных эта система приводится к трехдиагональному виду.

### 3.2. Эффективный способ построения сплайна

Так как сплайн на каждом отрезке  $[x_{i-1}, x_i]$  задается кубической функцией, то его вторая производная является линейной функцией. Обозначим  $M_i = S''(x_i)$  и пусть  $h_i = x_i - x_{i-1}$ . Тогда

$$S''(x) = S''_i(x) = M_{i-1} \frac{x_i - x}{h_i} + M_i \frac{x - x_{i-1}}{h_i}, \quad x \in [x_{i-1}, x_i].$$

Интегрируя, получаем

$$S'_i(x) = -M_{i-1} \frac{(x_i - x)^2}{2h_i} + M_i \frac{(x - x_{i-1})^2}{2h_i} + A_i,$$

$$S_i(x) = M_{i-1} \frac{(x_i - x)^3}{6h_i} + M_i \frac{(x - x_{i-1})^3}{6h_i} + A_i x + B_i,$$

где  $A_i$  и  $B_i$  – константы интегрирования. Чтобы их определить из интерполяционных условий, проще сделать невырожденную замену констант  $A_i x + B_i = a_i(x_i - x) + b_i(x - x_{i-1})$ , тогда

$$S_i(x) = M_{i-1} \frac{(x_i - x)^3}{6h_i} + M_i \frac{(x - x_{i-1})^3}{6h_i} + a_i(x_i - x) + b_i(x - x_{i-1}).$$

Используя условия

$$S_i(x_{i-1}) = y_{i-1}, \quad S_i(x_i) = y_i,$$

получаем систему уравнений

$$\begin{cases} M_{i-1} \frac{h_i^2}{6} + a_i h_i = y_{i-1} \\ M_i \frac{h_i^2}{6} + b_i h_i = y_i, \end{cases}$$

откуда

$$\begin{cases} a_i = \frac{y_{i-1}}{h_i} - M_{i-1} \frac{h_i}{6} \\ b_i = \frac{y_i}{h_i} - M_i \frac{h_i}{6}. \end{cases}$$

Подставляя эти константы в формулу для сплайна, получаем

$$S_i(x) = y_{i-1} \frac{(x_i - x)}{h_i} + y_i \frac{(x - x_{i-1})}{h_i} + \frac{h_i^2}{6} M_{i-1} \left[ \frac{(x_i - x)^3}{h_i^3} - \frac{(x_i - x)}{h_i} \right] + \frac{h_i^2}{6} M_i \left[ \frac{(x - x_{i-1})^3}{h_i^3} - \frac{(x - x_{i-1})}{h_i} \right].$$

Для того чтобы получить уравнения для  $M_i$ , воспользуемся условиями

$$S'_i(x_i) = S'_{i+1}(x_i), \quad i = 1, \dots, n-1.$$

Имеем

$$\begin{aligned} S'_i(x) &= \frac{y_i - y_{i-1}}{h_i} + \\ &+ \frac{h_i^2}{6} M_{i-1} \left[ -\frac{3(x_i - x)^2}{h_i^3} + \frac{1}{h_i} \right] + \frac{h_i^2}{6} M_i \left[ \frac{3(x - x_{i-1})^2}{h_i^3} - \frac{1}{h_i} \right], \\ S'_{i+1}(x) &= \frac{y_{i+1} - y_i}{h_{i+1}} + \\ &+ \frac{h_{i+1}^2}{6} M_i \left[ -\frac{3(x_{i+1} - x)^2}{h_{i+1}^3} + \frac{1}{h_{i+1}} \right] + \frac{h_{i+1}^2}{6} M_{i+1} \left[ \frac{3(x - x_i)^2}{h_{i+1}^3} - \frac{1}{h_{i+1}} \right], \end{aligned}$$

откуда

$$S'_i(x_i) = \frac{y_i - y_{i-1}}{h_i} + \frac{h_i}{6} M_{i-1} + \frac{h_i}{3} M_i,$$

$$S'_{i+1}(x_i) = \frac{y_{i+1} - y_i}{h_{i+1}} - \frac{h_{i+1}}{3} M_i - \frac{h_{i+1}}{6} M_{i+1}.$$

Приравнивая эти выражения, получаем систему уравнений

$$\frac{h_i}{6} M_{i-1} + \frac{h_i + h_{i+1}}{3} M_i + \frac{h_{i+1}}{6} M_{i+1} = \frac{y_{i+1} - y_i}{h_{i+1}} - \frac{y_i - y_{i-1}}{h_i},$$

где  $i = 1, \dots, n-1$ . Дополнив эту систему уравнениями из дополнительных условий второго рода

$$\begin{cases} M_0 = \alpha \\ M_n = \beta, \end{cases}$$

получаем систему с трехдиагональной матрицей, которую эффективно можно решать методом прогонки, так как имеется диагональное преобладание.

### 3.3. Экстремальное свойство сплайна

Обозначим через  $Y$  пространство дважды непрерывно дифференцируемых на отрезке  $[a, b]$  функций с закрепленными значениями в узлах  $f(x_i) = y_i$ ,  $i = 0, 1, \dots, n$ , где  $a = x_0 < x_1 < \dots < x_n = b$ .

Рассмотрим задачу: найти минимум на  $Y$  функционала

$$J[f] = \int_a^b (f''(x))^2 dx.$$

Механическая интерпретация данной задачи состоит в том, что нужно минимизировать потенциальную энергию упругого гибкого тела, например, металлической линейки, поставленной на ребро, с закрепленными точками. На рис. 8 изображена кривая  $S(x)$ , минимизирующая функционал  $J[f]$  среди всех функций  $f \in Y$ .

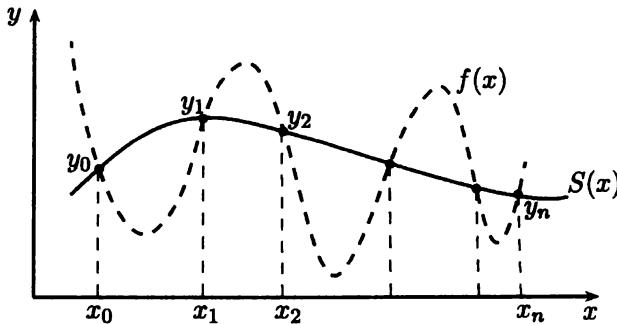


Рис. 8. Экстремальное свойство интерполяционного кубического сплайна  $S(x)$

**Теорема 4.** Единственным решением сформулированной задачи минимизации является интерполяционный кубический сплайн с дополнительными условиями  $S''(a) = S''(b) = 0$ .

**Доказательство.** Пусть  $f(x)$  — произвольная функция из  $Y$ , а  $S(x)$  — интерполяционный кубический сплайн с дополнительными условиями  $S''(a) = S''(b) = 0$ . Сначала докажем, что  $J[f] = J[S] + J[f - S]$ .

Обозначим для краткости через

$$\langle \varphi, \psi \rangle = \int_a^b \varphi''(x) \psi''(x) dx,$$

тогда

$$\begin{aligned} J[f] &= \langle f, f \rangle = \langle S + (f - S), S + (f - S) \rangle = \\ &= \langle S, S \rangle + 2 \langle S, f - S \rangle + \langle f - S, f - S \rangle = \\ &= J[S] + J[f - S] + 2 \langle S, f - S \rangle. \end{aligned}$$

Осталось доказать, что  $\langle S, f - S \rangle = 0$ . В самом деле, для того чтобы иметь возможность применять формулу интегрирования по частям для сплайнов, глобально не обладающих

достаточной гладкостью, разобьем интеграл на сумму интегралов

$$\begin{aligned} < S, f - S > &= \int_a^b S''(x)(f''(x) - S''(x))dx = \\ &= \sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} S''(x)(f''(x) - S''(x))dx. \end{aligned}$$

Вычислим отдельный интеграл из суммы, дважды применив формулу интегрирования по частям,

$$\begin{aligned} \int_{x_i}^{x_{i+1}} S''(x)(f''(x) - S''(x))dx &= \int_{x_i}^{x_{i+1}} S''(x)d(f'(x) - S'(x)) = \\ &= S''(x)(f'(x) - S'(x))|_{x_i}^{x_{i+1}} - \int_{x_i}^{x_{i+1}} S'''(x)(f'(x) - S'(x))dx = \\ &= S''(x)(f'(x) - S'(x))|_{x_i}^{x_{i+1}} - \int_{x_i}^{x_{i+1}} S'''(x)d(f(x) - S(x)) = \\ &= S''(x)(f'(x) - S'(x))|_{x_i}^{x_{i+1}} - S'''(x)(f(x) - S(x))|_{x_i}^{x_{i+1}} + \\ &+ \int_{x_i}^{x_{i+1}} S''''(x)(f(x) - S(x))dx = S''(x)(f'(x) - S'(x))|_{x_i}^{x_{i+1}}, \end{aligned}$$

так как  $f(x_i) = S(x_i)$  и  $S''''(x) \equiv 0$ .

Но тогда

$$\begin{aligned} \sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} S''(x)(f''(x) - S''(x))dx &= \sum_{i=0}^{n-1} S''(x)(f'(x) - S'(x))|_{x_i}^{x_{i+1}} = \\ &= S''(b)(f'(b) - S'(b)) - S''(a)(f'(a) - S'(a)) = 0 \end{aligned}$$

в силу дополнительных условий  $S''(a) = S''(b) = 0$ .

В силу доказанного задача минимизации функционала  $J[f]$  свелась к задаче минимизации функционала  $J[\bar{f} - S]$ .

Так как для любой функции  $f$  выполняется  $J[f - S] \geq 0$ , то при  $f \equiv S$  достигается минимум. Докажем его единственность от противного. Пусть  $J[\bar{f} - S] = 0$ , т. е.

$$\int_a^b (\bar{f}''(x) - S''(x))^2 dx = 0.$$

Так как под интегралом стоит непрерывная неотрицательная функция, то  $\bar{f}''(x) \equiv S''(x)$ , т. е.  $\bar{f}(x) = S(x) + Ax + B$ .

Подставляя в последнее равенство концы отрезка  $a$  и  $b$ , получаем систему

$$\begin{cases} \bar{f}(a) = S(a) + Aa + B \\ \bar{f}(b) = S(b) + Ab + B, \end{cases}$$

или, учитывая, что  $\bar{f}(a) = S(a)$ ,  $\bar{f}(b) = S(b)$ , систему

$$\begin{cases} Aa + B = 0 \\ Ab + B = 0, \end{cases}$$

которая имеет лишь тривиальное решение  $A = B = 0$ . Это доказывает единственность решения задачи минимизации.

### 3.4. Сплайн-квадратура

Среди многочисленных приложений сплайнов рассмотрим задачу численного интегрирования: найти

$$I[f] = \int_a^b f(x) dx,$$

используя значения подынтегральной функции  $f(x_i) = y_i$ ,  $i = 0, 1, \dots, n$ .

Если заменить подынтегральную функцию сплайном, который совпадает с исходной функцией в узлах, то полученная формула называется *сплайн-квадратурой*. Выведем эти формулы. Интерполяционный кубический сплайн задается следующим образом:

$$S(x) = S_i(x), x \in [x_{i-1}, x_i], i = 1, \dots, n,$$

$$S_i(x) = y_{i-1} \frac{(x_i - x)}{h_i} + y_i \frac{(x - x_{i-1})}{h_i} +$$

$$+ \frac{h_i^2}{6} M_{i-1} \left[ \frac{(x_i - x)^3}{h_i^3} - \frac{(x_i - x)}{h_i} \right] + \frac{h_i^2}{6} M_i \left[ \frac{(x - x_{i-1})^3}{h_i^3} - \frac{(x - x_{i-1})}{h_i} \right].$$

Обозначим линейную часть сплайна через  $S^1(x)$ , а кубическую через  $S^3(x)$ :

$$S_i^1(x) = y_{i-1} \frac{(x_i - x)}{h_i} + y_i \frac{(x - x_{i-1})}{h_i},$$

$$S_i^3(x) = \frac{h_i^2}{6} M_{i-1} \left[ \frac{(x_i - x)^3}{h_i^3} - \frac{(x_i - x)}{h_i} \right] +$$

$$+ \frac{h_i^2}{6} M_i \left[ \frac{(x - x_{i-1})^3}{h_i^3} - \frac{(x - x_{i-1})}{h_i} \right],$$

$$S^1(x) = S_i^1(x), x \in [x_{i-1}, x_i], i = 1, \dots, n,$$

$$S^3(x) = S_i^3(x), x \in [x_{i-1}, x_i], i = 1, \dots, n,$$

тогда

$$S(x) = S^1(x) + S^3(x), S_i(x) = S_i^1(x) + S_i^3(x).$$

Заменим сначала подынтегральную функцию  $f(x)$  линейным сплайном  $S^1(x)$ , получим после вычислений

$$I[S^1] = \sum_{i=1}^n \int_{x_{i-1}}^{x_i} S_i^1(x) dx =$$

$$\begin{aligned}
&= \sum_{i=1}^n \int_{x_{i-1}}^{x_i} \left( y_{i-1} \frac{(x_i - x)}{h_i} + y_i \frac{(x - x_{i-1})}{h_i} \right) dx = \\
&= \sum_{i=1}^n \frac{y_i + y_{i-1}}{2} h_i.
\end{aligned}$$

При подсчете интегралов рекомендуется делать замены  $s = \frac{(x_i - x)}{h_i}$  и  $s = \frac{(x - x_{i-1})}{h_i}$  соответственно.

Полученная формула является составной формулой трапеций для случая неравномерного шага.

Для подсчета интеграла от кубической части вычислим (используя те же замены):

$$\begin{aligned}
\int_{x_{i-1}}^{x_i} S_i^3(x) dx &= \int_{x_{i-1}}^{x_i} \left( \frac{h_i^2}{6} M_{i-1} \left[ \frac{(x_i - x)^3}{h_i^3} - \frac{(x_i - x)}{h_i} \right] + \right. \\
&\quad \left. + \frac{h_i^2}{6} M_i \left[ \frac{(x - x_{i-1})^3}{h_i^3} - \frac{(x - x_{i-1})}{h_i} \right] \right) dx = -\frac{h_i^3}{24} (M_{i-1} + M_i).
\end{aligned}$$

Можно заметить, что погрешность элементарной формулы трапеций  $R_i = -\frac{f''(\xi_i)h_i^3}{12}$ , т. е. в сплайн-квадратуре вместо  $f''(\xi_i)$ ,  $\xi_i \in [x_{i-1}, x_i]$ , берется аппроксимация  $\frac{M_{i-1} + M_i}{2}$ , где  $M_{i-1} = S''(x_{i-1})$ ,  $M_i = S''(x_i)$ .

Итоговая формула для сплайн-квадратуры

$$I[f] \approx \sum_{i=1}^n \left( \frac{y_i + y_{i-1}}{2} h_i - \frac{M_{i-1} + M_i}{24} h_i^3 \right).$$

Заметим, без доказательства, что она имеет четвертый порядок погрешности по максимальному шагу, как и составная формула Симпсона.

### 3.5. Метод конечных элементов

Метод конечных элементов является одним из самых распространенных численных методов решения уравнений математической физики. Здесь мы рассмотрим простой вариант этого метода, применительно к однородной самосопряженной краевой задаче

$$\begin{cases} (k(x)y')' = p(x)y + q(x) \\ y(a) = 0 \\ y(b) = 0. \end{cases}$$

В методе Ритца эта задача сводилась к задаче минимизации интегрального функционала  $J[y(\cdot)]$  на множестве  $Y_n = \{y_n = \sum_{i=1}^n c_i \varphi_i(x)\}$  линейных комбинаций базисных функций, которая, в свою очередь, сводилась к решению линейной системы  $Gx = -b$ .

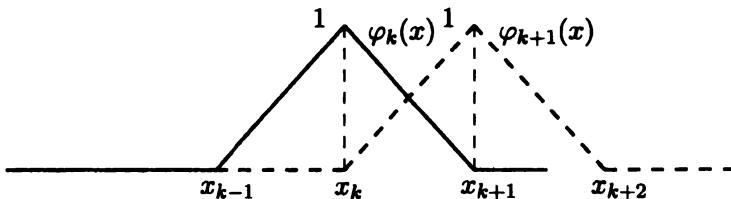


Рис. 9. Функции  $\varphi_k(x)$  и  $\varphi_{k+1}(x)$

Пусть  $a = x_0 < x_1 < \dots < x_{n+1} = b$  — равномерное разбиение отрезка  $[a, b]$ . Если в качестве  $Y_n$  выбрать множество кусочно-линейных функций (сплайнов первого порядка) с нулевыми значениями в точках  $a$  и  $b$ , то в качестве базисных функций можно взять кусочно-линейную функцию  $\varphi_i(x)$ , определенную во внутренних узлах  $x_k$ ,  $k = 1, 2, \dots, n$ , соотношениями

$$\varphi_i(x_k) = \begin{cases} 1, & k = i \\ 0, & k \neq i. \end{cases}$$

Функции данного вида представлены на рис. 9. Тогда матрица  $G$  линейной системы будет иметь трехдиагональную структуру, так как ее элементами являются скалярные произведения базисных функций  $g_{ik} = (\varphi_i, \varphi_k)_2$  и выполняется  $(\varphi_i, \varphi_k)_2 = 0$  при  $|k - i| \geq 2$ .

## 4. Метод наименьших квадратов

### 4.1. Метод наименьших квадратов в задаче решения линейных систем

Идеология метода наименьших квадратов широко распространена в приложениях математики. Мы уже рассматривали этот метод применительно к решению краевых задач. Теперь рассмотрим применительно к задаче решения линейных систем. Возьмем пример, который особенно часто встречается в статистике.

П р и м е р. Заданы точки  $(x_i, y_i)$ ,  $i = 1, 2, \dots, m$  на плоскости. Требуется провести через них прямую линию  $y = kx + b$ . С «классической» точки зрения система уравнений  $y_i = kx_i + b$ , вообще говоря, неразрешима относительно неизвестных  $k$  и  $b$ , если, например,  $m > 2$ . Однако в статистике отказываются от мысли провести прямую через все точки и проводят прямую так, чтобы сумма квадратов отклонений  $y_i - kx_i - b$  была наименьшей (уравнение такой прямой в статистике называется *уравнением регрессии*). На рис. 10 приведен график этой прямой.

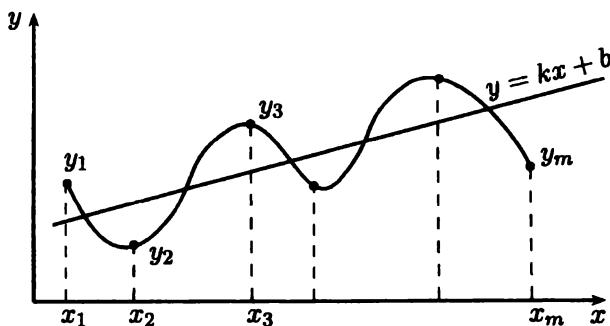


Рис. 10. Регрессионная прямая

Перейдем к общей задаче о решении линейной системы

$$Ax = y,$$

где  $x \in R^n$  — неизвестный вектор;  $y \in R^m$  — известный вектор;  $A$  — известная матрица размерности  $m \times n$ .

Рассмотрим задачу о минимизации квадрата невязки

$$\min_{x \in R^n} \|Ax - y\|^2.$$

Пользуясь свойствами скалярного произведения, распишем величину минимизируемой функции:

$$\begin{aligned} J(x) &= \|Ax - y\|^2 = (Ax - y, Ax - y) = (Ax, Ax) - 2(Ax, y) + (y, y) = \\ &= (x, A^T Ax) - 2(x, A^T y) + (y, y). \end{aligned}$$

Последнее слагаемое не зависит от аргумента. Обозначим  $G = A^T A$ ,  $b = -A^T y$ ,  $F(x) = (x, A^T Ax) - 2(x, A^T y) = (x, Gx) + 2(x, b)$ , тогда, используя Лемму из раздела 2.6, получаем, что если матрица  $G = A^T A$  положительно определенная, то существует единственное решение задачи минимизации, определяемое формулой

$$x = -G^{-1}b,$$

т. е. решение удовлетворят системе

$$A^T Ax = A^T y.$$

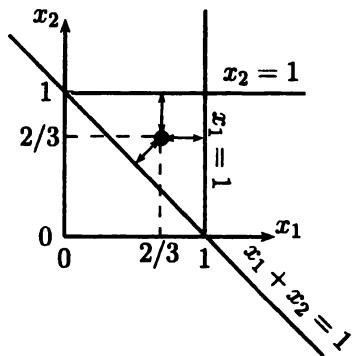
Это решение называется *MНK-решением* линейной системы.

Выясним условия, при которых матрица  $G = A^T A$  является положительно определенной, в данном случае, так как она всегда неотрицательно определена, — просто невырожденной.

Обозначим столбцы матрицы  $A$  через  $h_1, h_2, \dots, h_n$ , тогда  $Ax = \sum_{i=1}^n h_i x_i$ , и получаем  $(x, A^T Ax) = (Ax, Ax) = 0$  тогда и

только тогда, когда  $\sum_{i=1}^n h_i x_i = 0$ . А это, в свою очередь, означает, что матрица  $G = A^T A$  является положительно определенной тогда и только тогда, когда столбцы  $h_1, h_2, \dots, h_n$  матрицы  $A$  линейно независимы, т. е. ранг матрицы  $A$  равен  $n$ . Это условие называется *невырожденный МНК*.

В качестве примера решим несовместную систему (рис. 11):



$$\begin{cases} x_1 + x_2 = 1 \\ x_1 = 1 \\ x_2 = 1. \end{cases}$$

Здесь

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad y = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

Рис. 11. Иллюстрация к примеру решения системы

Тогда

$$A^T = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \quad A^T A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, \quad A^T y = \begin{pmatrix} 2 \\ 2 \end{pmatrix}.$$

МНК-система имеет вид

$$\begin{cases} 2x_1 + x_2 = 2 \\ x_1 + 2x_2 = 2, \end{cases}$$

система невырождена, а ее решение (МНК-решение):

$$\begin{cases} x_1 = \frac{2}{3} \\ x_2 = \frac{2}{3}. \end{cases}$$

На рис. 11 решение отмечено точкой. Сумма квадратов расстояний от этой точки до прямых — наименьшая среди всех возможных точек.

## 4.2. Вырожденный метод наименьших квадратов в задаче решения линейных систем

Пусть теперь наблюдается вырождение МНК, т. е.  $r(A) < n$ . В частности, вырожденный МНК имеет место, когда  $m < n$ , т. е. число уравнений в системе меньше числа неизвестных, что приводит к неединственности решения. Задача относится к числу некорректных и регуляризированный МНК позволяет отыскать решение с минимальной нормой.

Приведем алгоритм вырожденного МНК. Добавим к величине минимизируемой функции  $(Ax - y, Ax - y)$  (квадрата невязки) квадрат  $x$  с некоторым положительным коэффициентом  $\alpha$  (параметр регуляризации), таким образом, минимизируемая величина представляется в виде

$$\begin{aligned} J(x) &= \|Ax - y\|^2 + \alpha\|x\|^2 = (Ax - y, Ax - y) + \alpha(x, x) = \\ &= (Ax, Ax) - 2(Ax, y) + (y, y) + \alpha(x, x) = \\ &= (x, (A^T A + \alpha)x) - 2(x, A^T y) + (y, y). \end{aligned}$$

Обозначим  $G = A^T A + \alpha$ ,  $b = -A^T y$ , тогда, используя ту же лемму из раздела 2.6, получаем, что если матрица  $G = A^T A + \alpha$  положительно определенная (а она является такой), то существует единственное решение задачи минимизации, определяемое формулой

$$(A^T A + \alpha)x = A^T y.$$

Если существует предел решения этой системы при  $\alpha \rightarrow 0$ , то он называется *вырожденным МНК-решением* линейной системы. Это решение формально записывается в виде

$$x = \lim_{\alpha \rightarrow 0} (A^T A + \alpha)^{-1} A^T y.$$

По аналогии с обратной матрицей  $A^{-1}$ , с помощью которой можно записать решение системы  $Ax = y$  в случае невырож-

денной матрицы  $A$  в виде  $x = A^{-1}y$ , матрицу

$$A^+ = \lim_{\alpha \rightarrow 0} (A^T A + \alpha)^{-1} A^T$$

называют *псевдообратной* к  $A$ . Псевдообратная матрица обладает свойством: произведение  $A^+ A$  представляет собой диагональную матрицу, у которой на главной диагонали стоит несколько единиц, а все остальные элементы нули.

### 4.3. Метод наименьших квадратов в задаче приближения функций (дискретный вариант)

Пусть в узлах  $x_1, x_2, \dots, x_m$  заданы значения функции  $y_1, y_2, \dots, y_m$ . Требуется в классе функций  $y = f(x)$ , являющихся линейной комбинацией базисных функций  $\varphi_i(x)$ ,  $i = 1, 2, \dots, n$ , выбрать функцию, наилучшим образом приближающую значения в узлах. Если в этом классе существует функция, в точности удовлетворяющая интерполяционным условиям  $f(x_j) = y_j$ , для  $j = 1, 2, \dots, m$ , то решается задача *интерполяции*. В более общем случае, когда интерполяционные условия не выполняются, то рассматривают задачу *приближения*, в которой минимизируют отклонение от этих условий. В методе наименьших квадратов в качестве критерия выбрана сумма квадратов отклонений в узлах

$$J(C) = J(c_1, c_2, \dots, c_n) = \sum_{j=1}^m \left( \sum_{i=1}^n c_i \varphi_i(x_j) - y_j \right)^2,$$

эту величину нужно минимизировать за счет выбора коэффициентов  $c_1, c_2, \dots, c_n$ , такой вариант задачи приближения называется *дискретным методом наименьших квадратов*.

Введем скалярное произведение в пространстве функций, определенных на узлах

$$\langle f(x), g(x) \rangle = \sum_{j=1}^m f(x_j)g(x_j),$$

а также вектор  $Y = (y_1, y_2, \dots, y_m)$ , тогда минимизируемая величина записывается как

$$\begin{aligned} J(C) &= \left\langle \sum_{i=1}^n c_i \varphi_i(x) - Y, \sum_{i=1}^n c_i \varphi_i(x) - Y \right\rangle = \\ &= \sum_{i,k=1}^n c_i c_k \langle \varphi_i(x), \varphi_k(x) \rangle - 2 \sum_{i=1}^n c_i \langle \varphi_i(x), Y \rangle + \langle Y, Y \rangle. \end{aligned}$$

Получили задачу о минимизации линейно-квадратичной формы. Обозначим через  $G$  матрицу с элементами  $\langle \varphi_i(x), \varphi_k(x) \rangle$ , через  $b$  — вектор с координатами  $\langle \varphi_i(x), Y \rangle$ . Используя лемму из раздела 2.6, получаем, что если матрица  $G$  невырождена, то для нахождения вектора  $C$ , образованного искомыми неизвестными коэффициентами, нужно решить линейную систему

$$GC = b.$$

Удобство такого варианта состоит в том, что коэффициенты матрицы  $G$  и вектора  $b$  легко считать, так они представляют собой конечномерные скалярные произведения, т. е. конечные суммы. В то же время метод имеет существенный недостаток: при увеличении числа базисных функций  $n$  метод вырождается.

#### 4.4. Метод наименьших квадратов в задаче приближения функций (непрерывный вариант)

Задана некоторая функция  $y = f(x)$ , определенная на отрезке  $[a, b]$ . Задана также некоторая система базисных функций  $\varphi_i(x)$ ,  $i = 1, 2, \dots, n$ , также определенных на этом отрезке. Требуется в классе линейных комбинаций  $y = \sum_{i=1}^n c_i \varphi_i(x)$  найти

наилучшее приближение для функции  $f(x)$  в среднеквадратичном смысле, т. е. минимизировать

$$J(C) = J(c_1, c_2, \dots, c_n) = \int_a^b \left( \sum_{i=1}^n c_i \varphi_i(x) - f(x) \right)^2 dx.$$

Введем скалярное произведение в пространстве функций, определенных на  $[a, b]$ :

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

тогда минимизируемая величина запишется как

$$\begin{aligned} J(C) &= \left\langle \sum_{i=1}^n c_i \varphi_i - f, \sum_{i=1}^n c_i \varphi_i - f \right\rangle = \\ &= \sum_{i,k=1}^n c_i c_k \langle \varphi_i, \varphi_k \rangle - 2 \sum_{i=1}^n c_i \langle \varphi_i, f \rangle + \langle f, f \rangle. \end{aligned}$$

Получили задачу о минимизации линейно-квадратичной формы. Обозначим через  $G$  матрицу с элементами  $\langle \varphi_i, \varphi_k \rangle$ , через  $b$  вектор с координатами  $\langle \varphi_i, f \rangle$  (но скалярные произведения другие по сравнению с дискретным вариантом). Используя лемму из раздела 2.6, получаем, что если матрица  $G$  невырождена, то для нахождения вектора  $C$ , образованного искомыми неизвестными коэффициентами, нужно решить линейную систему

$$GC = b,$$

этот вариант называется *непрерывным МНК*.

В отличие от дискретного варианта, для подсчета коэффициентов этой системы нужно считать интегралы, что создает определенные трудности. Но при увеличении числа базисных

функций  $n$  непрерывный МНК не вырождается. Чтобы посмотреть это, предположим для простоты, что система базисных функций ортонормальна, тогда матрица  $G$  становится единичной и мы получаем, что

$$c_i = \langle \varphi_i, f \rangle.$$

Тогда значение функционала записывается в виде

$$J(C) = \sum_{i=1}^n c_i^2 - 2 \sum_{i=1}^n c_i^2 + \langle f, f \rangle = \langle f, f \rangle - \sum_{i=1}^n \langle \varphi_i, f \rangle.$$

Так как  $J(C) \geq 0$ , то получаем известное неравенство Бесселя

$$\langle f, f \rangle \geq \sum_{i=1}^n \langle \varphi_i, f \rangle.$$

Если же приближаемая функция  $f$  берется из линейной комбинации базисных функций, то неравенство Бесселя переходит в равенство Парсеваля

$$\langle f, f \rangle = \sum_{i=1}^n \langle \varphi_i, f \rangle$$

и ошибка приближения равна нулю. При увеличение числа базисных функций, как показывает неравенство Бесселя, ошибка приближения лишь уменьшается.

## 5. Численное решение интегральных уравнений

### 5.1. Классификация линейных интегральных уравнений

Типовым линейным интегральным уравнением является *уравнение Фредгольма второго рода*

$$y(x) - \lambda \int_a^b K(x, s)y(s)ds = f(x),$$

где  $y(x)$  — неизвестная функция;  $K(x, s)$  — известная функция двух переменных (ядро);  $f(x)$  — известная функция (неоднородность);  $\lambda$  — собственное число (как правило, известно). Если  $f(x) \equiv 0$ , то уравнение называется *однородным*.

*Уравнение Фредгольма первого рода* имеет вид

$$\int_a^b K(x, s)y(s)ds = f(x)$$

и, в отличие от уравнения Фредгольма второго рода, представляет собой некорректную задачу.

*Уравнение Вольтерра второго рода* имеет вид

$$y(x) - \lambda \int_a^x K(x, s)y(s)ds = f(x)$$

и формально представляет собой частный случай уравнения Фредгольма второго рода с треугольным ядром. Однако этот частный случай представляет значительный интерес. Именно к такому уравнению сводится, например, задача Коши для линейного дифференциального уравнения.

Уравнение Вольтерра первого рода имеет вид

$$\int_a^x K(x, s)y(s)ds = f(x).$$

Если  $f(x) \equiv 0$ , то уравнение любого рассмотренного вида называется *однородным*. Иногда для однородного уравнения в отдельный вид выделяют задачу для нахождения не только  $y(x)$ , но и числа  $\lambda$  (задача на собственные функции и собственные значения), например, для уравнения Фредгольма второго рода

$$y(x) = \lambda \int_a^b K(x, s)y(s)ds.$$

В дальнейшем будем рассматривать неоднородное уравнение Фредгольма второго рода, останавливаясь на других видах уравнений только для рассмотрения коренных отличий.

## 5.2. Метод разложения по собственным значениям

В уравнении Фредгольма второго рода обозначим через  $I$  интегральный оператор

$$Iy = \int_a^b K(x, s)y(s)ds,$$

тогда уравнение перепишется в виде

$$(E - \lambda I)y = f,$$

где  $E$  — единичный оператор.

Если оператор  $E - \lambda I$  обратим, то

$$y(x) = (E - \lambda I)^{-1}f(x),$$

причем справедлива формула

$$(E - \lambda I)^{-1} = \sum_{k=0}^{\infty} (\lambda I)^k,$$

которая доказывается подобно аналогичной формуле для матриц, доказанной в теореме о сходимости метода простой итерации для решения линейных систем. Таким образом, получаем формулу

$$y(x) = \sum_{k=0}^{\infty} (\lambda I)^k f(x).$$

Отметим, что на практике этот алгоритм применять затруднительно из-за медленной сходимости при суммировании ряда.

### 5.3. Метод замены интеграла конечной суммой

Разобьем отрезок интегрирования  $[a, b]$  на части с шагом  $h = \frac{b-a}{n}$  и введем узлы  $x_i = a + ih$ ,  $i = 0, 1, \dots, n$ . Точные значения решения интегрального уравнения  $y(x_i)$  в узлах будем приближать величинами  $y_i$ .

Для интеграла в интегральном уравнении применим какую-либо квадратурную формулу с узлами  $x_i$ :

$$\int_a^b F(x) ds \approx \sum_{i=0}^n A_i F(x_i).$$

Самый распространенный вариант – составные квадратурные формулы, у которых при увеличении  $n$  погрешность уменьшается, а коэффициенты  $A_i$  стремятся к нулю, так как они содержат в качестве множителя шаг  $h$ , именно это мы и будем предполагать.

Введем обозначения  $K(x_i, s_j) = K_{ij}$ ,  $f(x_i) = f_i$ , тогда уравнение Фредгольма второго рода заменится системой линейных уравнений

$$y_j - \lambda \sum_{i=0}^n A_i K_{ij} y_i = f_j, \quad j = 0, 1, \dots, n.$$

Главная матрица этой системы имеет вид

$$\begin{pmatrix} 1 - \lambda K_{00} A_0 & -\lambda K_{01} A_1 & \dots & -\lambda K_{0n} A_n \\ -\lambda K_{10} A_0 & 1 - \lambda K_{11} A_1 & \dots & -\lambda K_{1n} A_n \\ \dots & \dots & \dots & \dots \\ -\lambda K_{n0} A_0 & -\lambda K_{n1} A_1 & \dots & 1 - \lambda K_{nn} A_n \end{pmatrix}.$$

Если предположить, что ядро  $K(x, s)$  ограничено, то в силу того, что коэффициенты  $A_i$  стремятся к нулю при увеличении  $n$ , эта матрица стремится к единичной, и система становится хорошо обусловленной.

Для сравнения рассмотрим уравнение Фредгольма первого рода. Приводя подобным образом все выкладки, получаем линейную систему с главной матрицей вида

$$\begin{pmatrix} \lambda K_{00} A_0 & \lambda K_{01} A_1 & \dots & \lambda K_{0n} A_n \\ \lambda K_{10} A_0 & \lambda K_{11} A_1 & \dots & \lambda K_{1n} A_n \\ \dots & \dots & \dots & \dots \\ \lambda K_{n0} A_0 & \lambda K_{n1} A_1 & \dots & \lambda K_{nn} A_n \end{pmatrix}.$$

Если ядро  $K(x, s)$  непрерывно, то при увеличении  $n$  соседние строки становятся близкими, что приводит к плохой обусловленности системы.

Для уравнений Вольтерра в этом методе наблюдаются те же эффекты, только добавляется важное свойство: главная матрица системы становится треугольной.

#### 5.4. Метод замены ядра на вырожденное

Уравнение называется с вырожденным ядром, если ядро представимо в виде

$$K(x, s) = \sum_{i=1}^n \varphi_i(x) \psi_i(s).$$

Покажем, что если уравнение Фредгольма второго рода имеет вырожденное ядро, то нахождение его решения сводится к решению системы линейных алгебраических уравнений.

Будем искать решение в виде

$$y(x) = \lambda \sum_{i=1}^n C_i \varphi_i(x) + f(x)$$

и подберем коэффициенты  $C_i$ . Подставим  $y(x)$  в уравнение и воспользуемся тем, что ядро вырожденное:

$$\lambda \sum_{i=1}^n C_i \varphi_i(x) - \lambda \int_a^b \sum_{i=1}^n \varphi_i(x) \psi_i(s) [\lambda \sum_{j=1}^n C_j \varphi_j(s) + f(s)] ds = 0.$$

После преобразований получаем

$$\sum_{i=1}^n \varphi_i(x) [C_i - \sum_{j=1}^n C_j \lambda \int_a^b \psi_i(s) \varphi_j(s) ds - \lambda \int_a^b \psi_i(s) f(s) ds] = 0.$$

Приравнивая к нулю выражения при  $\varphi_i(x)$ , получаем систему уравнений относительно  $C_i$ :

$$C_i - \sum_{j=1}^n C_j \lambda \int_a^b \psi_i(s) \varphi_j(s) ds - \lambda \int_a^b \psi_i(s) f(s) ds = 0, \quad i = 1, \dots, n.$$

**З а м е ч а н и е.** Если ядро уравнения невырожденное, то можно приблизить его уравнением с вырожденным ядром, используя, например, разложение Тейлора.

В качестве примера приближенно решим уравнение

$$y(x) - \int_0^1 x e^{xs} ds = x.$$

Ядро уравнения невырожденное. Однако можно, используя разложение

$$e^{xs} = 1 + xs + \frac{(xs)^2}{2} + \dots,$$

приблизить ядро

$$xe^{xs} \approx x + x^2 s.$$

Тогда  $\varphi_1(x) = x$ ,  $\psi_1(s) = 1$ ,  $\varphi_2(x) = x^2$ ,  $\psi_2(s) = s$  (выбор неоднозначен) и ищем решение в виде

$$y(x) = C_1 x + C_2 x^2 + x.$$

Посчитав интегралы, получаем систему

$$\begin{cases} C_1(1 - \frac{1}{2}) - C_2 \frac{1}{3} = \frac{1}{2} \\ -C_1 \frac{1}{3} + C_2(1 - \frac{1}{4}) = \frac{1}{3}, \end{cases}$$

решив которую, получаем приближенное решение интегрального уравнения

$$y = \frac{54}{19}x + \frac{24}{19}x^2.$$

## 6. Численное решение уравнений математической физики

Среди уравнений математической физики наиболее распространены уравнения в частных производных второго порядка эллиптического, параболического и гиперболического типов. Эти уравнения описывают реальные физические процессы и, в силу сложности таких задач, на первый план выходят численные алгоритмы их решения. Не претендуя на полноту, познакомимся с численными методами решения на примерах простейших задач.

### 6.1. Решение уравнения эллиптического типа методом сеток

Рассмотрим задачу Дирихле для уравнения Пуассона:

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = f(x, y),$$

где  $x \in (0, 1)$ ,  $y \in (0, 1)$  — независимые переменные;  $u(x, y)$  — искомая функция. На границах области (квадрата) заданы граничные условия

$$\begin{aligned} u(0, y) &= \varphi_0(y), \quad y \in [0, 1], \\ u(1, y) &= \varphi_1(y), \quad y \in [0, 1], \\ u(x, 0) &= \psi_0(x), \quad x \in [0, 1], \\ u(x, 1) &= \psi_1(x), \quad x \in [0, 1]. \end{aligned}$$

Предполагаем, что задача имеет единственное решение, определяемое в классическом смысле.

Проведем дискретизацию всех входящих в задачу величин. Введем шаг  $h = \frac{1}{n}$ ; точки разбиения  $x_i = ih$ ,  $i = 0, 1, \dots, n$ , и  $y_j = jh$ ,  $j = 0, 1, \dots, n$ ; пары  $(x_i, y_j)$  будем называть *узлами сетки*; точное решение в узлах  $u(x_i, y_j)$  будем заменять приближенным  $u_{i,j}$ . Обозначим также  $f(x_i, y_j) = f_{i,j}$ ,  $\varphi_0(y_j) = \varphi_{0j}$ ,  $\varphi_1(y_j) = \varphi_{1j}$ ,  $\psi_0(x_i) = \psi_{0i}$ ,  $\psi_1(x_i) = \psi_{1i}$ .

Для замены оператора Лапласа в уравнении Пуассона используем формулу численного дифференцирования на середину для вторых производных по каждой переменной (схема *крест*), получим сеточное уравнение

$$\frac{u_{i-1,j} - 2u_{i,j} + u_{i+1,j}}{h^2} + \frac{u_{i,j-1} - 2u_{i,j} + u_{i,j+1}}{h^2} = f_{i,j},$$

для  $i = 1, \dots, n - 1$ , и  $j = 1, \dots, n - 1$ . Эта система уравнений дополняется граничными условиями

$$\begin{aligned} u_{0,j} &= \varphi_{0j}, \quad j = 0, 1, \dots, n, \\ u_{n,j} &= \varphi_{1j}, \quad j = 0, 1, \dots, n, \\ u_{i,0} &= \psi_{0i}, \quad i = 1, \dots, n - 1, \\ u_{i,n} &= \psi_{1i}, \quad i = 1, \dots, n - 1. \end{aligned}$$

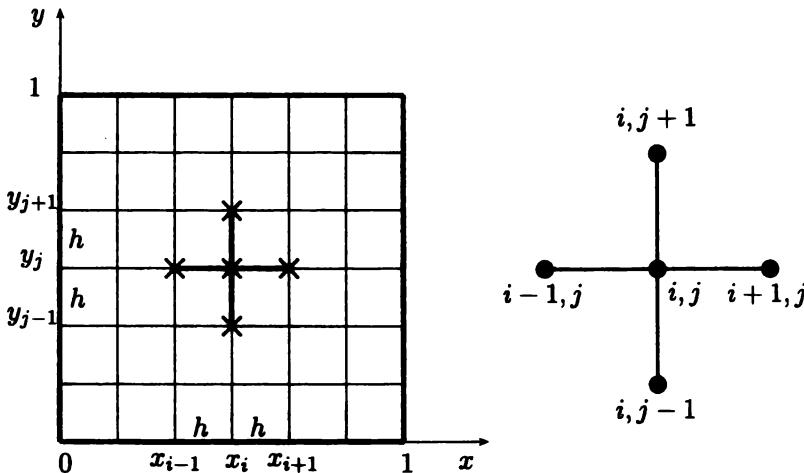


Рис. 12. Конечно-разностная сетка и шаблон схемы «крест» для уравнения эллиптического типа

Эта система размерности  $(n + 1)^2$  не обладает трехдиагональной структурой, поэтому ее решение — трудоемкий про-

цесс. Один из вариантов ее решения — вариант простой итерации в форме Якоби

$$u_{i,j}^{k+1} = \frac{1}{4}(u_{i-1,j}^k + u_{i+1,j}^k + u_{i,j-1}^k + u_{i,j+1}^k) - \frac{1}{4}h^2 f_{i,j},$$

который хотя и сходится, но достаточно медленно (нет диагонального преобладания).

Другой способ решения этой задачи — схема счета на установление, будет рассмотрен позже.

Конечно-разностная сетка и шаблон схемы «крест» описанного метода решения уравнения эллиптического типа показаны на рис. 12.

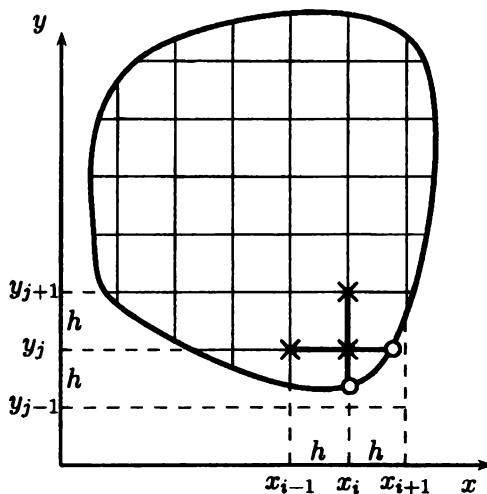


Рис. 13. Модификация схемы «крест» для непрямоугольной области

**З а м е ч а н и е.** Метод можно модифицировать для любой непрямоугольной области (рис. 13). Для этого в схеме нужно на узлах, примыкающих к границе области, применять формулы численного дифференцирования по неравноотстоящим узлам.

## 6.2. Решение одномерного уравнения теплопроводности (явная схема)

Рассмотрим уравнение параболического типа — одномерное уравнение теплопроводности

$$\frac{\partial u}{\partial t} = a^2 \frac{\partial^2 u}{\partial x^2} + f(x, t),$$

где  $x \in (0, l)$ ,  $t > 0$  — независимые переменные;  $u(x, t)$  — искаемая функция.

Заданы начальные условия

$$u(x, 0) = \varphi(x), \quad x \in [0, l],$$

и граничные условия

$$u(0, t) = g_0(t), \quad t > 0,$$

$$u(l, t) = g_1(t), \quad t > 0.$$

Эта задача интерпретируется как процесс распространения температуры  $u$  по однородному стержню длиной  $l$  ( $x$  — пространственная координата стержня) в зависимости от времени  $t$ . Здесь  $a^2$  — коэффициент теплопроводности,  $f(x, t)$  — мощность источников или стоков температуры. Задана начальная температура стержня  $\varphi(x)$ , на границах стержня поддерживается заданный температурный режим: на левой границе  $g_0(t)$ , на правой —  $g_1(t)$  (краевая задача первого рода). Предполагается, что задача имеет единственное решение, определяемое в классическом смысле.

Численные методы решения этой задачи можно разбить на два класса: сеточные методы и метод прямых. Рассмотрим **явный сеточный метод**.

Разобьем отрезок  $[0, l]$  точками  $x_i = ih$ ,  $i = 0, 1, \dots, n$ , где  $h = \frac{l}{n}$  — пространственный шаг. Пусть  $\tau > 0$  — временной шаг и  $t_j = j\tau$ , для  $j = 0, 1, \dots, J$ , и пусть  $\tau J = T$ . Пары  $(x_i, t_j)$  будем

называть узлами сетки; точное решение в узлах  $u(x_i, t_j)$  будем заменять приближенным  $u_i^j$ . Обозначим также  $f(x_i, t_j) = f_i^j$ ,  $\varphi(x_i) = \varphi_i$ ,  $g_0(y_j) = g_0^j$ ,  $g_1(y_j) = g_1^j$ . Индекс  $j$  в этих обозначениях обычно называют временным слоем.

Выберем для аппроксимации частной производной по времени  $t$  формулу численного дифференцирования по двум узлам на нижний край (слой), а для второй производной по пространственной переменной  $x$  выберем формулу численного дифференцирования по трем узлам на середину при фиксированном нижнем слое  $j$ , получим

$$\frac{u_i^{j+1} - u_i^j}{\tau} = a^2 \frac{u_{i-1}^j - 2u_i^j + u_{i+1}^j}{h^2} + f_i^j,$$

для  $i = 1, \dots, n-1$ , и  $j = 0, 1, \dots, J-1$ . Эта система уравнений дополняется граничными условиями

$$\begin{aligned} u_0^j &= g_0^j, & j &= 1, \dots, J, \\ u_n^j &= g_1^j, & j &= 1, \dots, J, \end{aligned}$$

и начальными условиями

$$u_i^0 = \varphi_i, \quad i = 0, 1, \dots, n.$$

Из этой системы уравнений неизвестное на  $j+1$  слое выражается явно:

$$u_i^{j+1} = \sigma u_{i-1}^j + (1 - 2\sigma) u_i^j + \sigma u_{i+1}^j + \tau f_i^j,$$

где

$$\sigma = \frac{a^2 \tau}{h^2} > 0.$$

Конечно-разностная сетка и шаблон явной схемы для уравнения параболического типа показаны на рис. 14.

Изучим вопрос о сходимости метода. Метод называется *сходящимся с порядком*  $\tau^p + h^q$ , если найдется такая константа  $C$ , что  $|u(x_i, t_j) - u_i^j| \leq C(\tau^p + h^q)$  для всех  $i = 0, \dots, n$  и

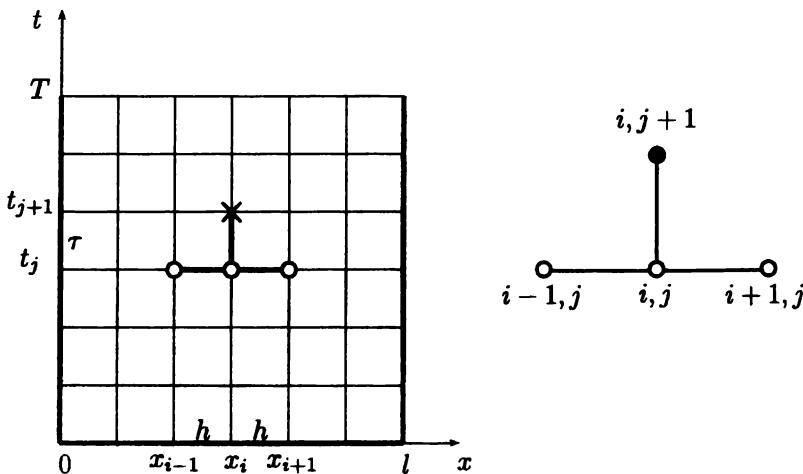


Рис. 14. Конечно-разностная сетка и шаблон явной схемы для одномерного уравнения параболического типа

$j = 0, \dots, J$ . Этот вопрос решается с помощью определения порядка невязки и характера устойчивости.

*Невязкой* метода называется разность левой и правой части метода, если вместо приближенного решения в метод подставить точное решение. В частности, для явного сеточного метода невязка определяется следующим образом:

$$\nu_i^j = \frac{u(x_i, t_{j+1}) - u(x_i, t_j)}{\tau} - a^2 \frac{u(x_{i-1}, t_j) - 2u(x_i, t_j) + u(x_{i+1}, t_j)}{h^2} - f_i^j.$$

Порядок невязки (порядок аппроксимации метода) определяется стандартным образом с помощью тейлоровских разложений при условии достаточной гладкости решений. Проделаем этот вывод для явной схемы. Во всех остальных методах, которые будут рассмотрены ниже, вывод порядка невязки проводится аналогично.

**Лемма.** Если точное решение  $u(x, t)$  четырежды непрерывно дифференцируемо по  $x$  и дважды непрерывно дифференцируемо по  $t$ , то невязка явной схемы имеет порядок  $\tau + h^2$ .

**Доказательство.** Разложим по формуле Тейлора все значения решения в окрестности точки  $(x_i, t_j)$ :

$$\begin{aligned} u(x_i, t_{j+1}) &= u(x_i, t_j) + \frac{\partial u}{\partial t}(x_i, t_j)\tau + \frac{\partial^2 u}{\partial t^2}(x_i, t_j)\frac{\tau^2}{2} + \dots \\ u(x_{i-1}, t_j) &= u(x_i, t_j) + \frac{\partial u}{\partial x}(x_i, t_j)(-h) + \frac{\partial^2 u}{\partial x^2}(x_i, t_j)\frac{h^2}{2} + \\ &\quad + \frac{\partial^3 u}{\partial x^3}(x_i, t_j)(-\frac{h^3}{6}) + \frac{\partial^4 u}{\partial x^4}(x_i, t_j)\frac{h^4}{24} + \dots \\ u(x_{i+1}, t_j) &= u(x_i, t_j) + \frac{\partial u}{\partial x}(x_i, t_j)h + \frac{\partial^2 u}{\partial x^2}(x_i, t_j)\frac{h^2}{2} + \frac{\partial^3 u}{\partial x^3}(x_i, t_j)\frac{h^3}{6} + \\ &\quad + \frac{\partial^4 u}{\partial x^4}(x_i, t_j)\frac{h^4}{24} + \dots \end{aligned}$$

и подставим в определение невязки. Воспользовавшись ограниченностью соответствующих производных и определением точного решения, получаем

$$\nu_i^j = \frac{\partial u}{\partial t}(x_i, t_j) + O(\tau) - a^2 \frac{\partial^2 u}{\partial x^2}(x_i, t_j) + O(h^2) - f(x_i, t_j) = O(\tau + h^2).$$

Вопрос с устойчивостью явной схемы решается неоднозначно: устойчивость зависит от соотношения шагов  $h$  и  $\tau$ .

**Пример.** Рассмотрим однородное уравнение  $f(x, t) = 0$  с нулевыми граничными условиями  $g_0(t) = g_1(t) = 0$ . Пусть начальные условия также нулевые, за исключением одной точки  $x_{i_0}$ , где задана начальная температура  $\varepsilon > 0$ :

$$\varphi_i = \begin{cases} 0, & i \neq i_0, \\ \varepsilon, & i = i_0. \end{cases}$$

Используя расчетную формулу

$$u_i^{j+1} = \sigma u_{i-1}^j + (1 - 2\sigma)u_i^j + \sigma u_{i+1}^j,$$

можем посчитать, если задано  $\sigma$ , значения искомой функции на разных временных слоях. Рассмотрим два случая.

1.  $\sigma = \frac{1}{4}$ , тогда  $u_i^{j+1} = \frac{1}{4}u_{i-1}^j + \frac{1}{2}u_i^j + \frac{1}{4}u_{i+1}^j$ , и последовательно получаем

на нулевом слое:  $0, 0, \dots, 0, 0, \varepsilon, 0, 0, \dots, 0, 0$ ,

на первом слое:  $0, 0, \dots, 0, \frac{\varepsilon}{4}, \frac{\varepsilon}{2}, \frac{\varepsilon}{4}, 0, \dots, 0, 0$ ,

на втором слое:  $0, 0, \dots, \frac{\varepsilon}{16}, \frac{\varepsilon}{4}, \frac{3\varepsilon}{8}, \frac{\varepsilon}{4}, \frac{\varepsilon}{16}, \dots, 0, 0$ ,

и так далее. Видно, что расчеты хорошо описывают физический процесс: температурный импульс «растекается» (рис. 15).

2.  $\sigma = 1$ , тогда  $u_i^{j+1} = u_{i-1}^j - u_i^j + u_{i+1}^j$ , и последовательно получаем

на нулевом слое:  $0, 0, \dots, 0, 0, \varepsilon, 0, 0, \dots, 0, 0$ ,

на первом слое:  $0, 0, \dots, 0, \varepsilon, -\varepsilon, \varepsilon, 0, \dots, 0, 0$ ,

на втором слое:  $0, 0, \dots, \varepsilon, -2\varepsilon, 3\varepsilon, -2\varepsilon, \varepsilon, \dots, 0, 0$ ,

и так далее. Налицо неустойчивость, значения не только растут по абсолютной величине, но даже становятся отрицательными, что противоречит физическому смыслу (рис. 16).

Разностная схема называется *условно устойчивой*, если она устойчива при определенных соотношениях параметров задачи и шагов дискретизации. Выведем условия устойчивости явной схемы с помощью приема, который называется *методом гармоник*. Рассмотрим явную схему для однородного уравнения с нулевыми граничными условиями и произвольными начальными условиями

$$u_k^{j+1} = \sigma u_{k-1}^j + (1 - 2\sigma)u_k^j + \sigma u_{k+1}^j$$

(индекс  $i$  заменен на  $k$ , чтобы не было путаницы с мнимой единицей). Будем искать решение в виде

$$u_k^j = q^j e^{ik\phi h},$$

где  $q$  — действительное число, которое подлежит определению;  $\phi$  — произвольное вещественное число;  $i$  — мнимая единица. Подставляя в схему, получаем после сокращения на  $q^j e^{ik\phi h}$

$$q = 1 + \sigma(e^{-i\phi h} - 2 + e^{i\phi h}).$$

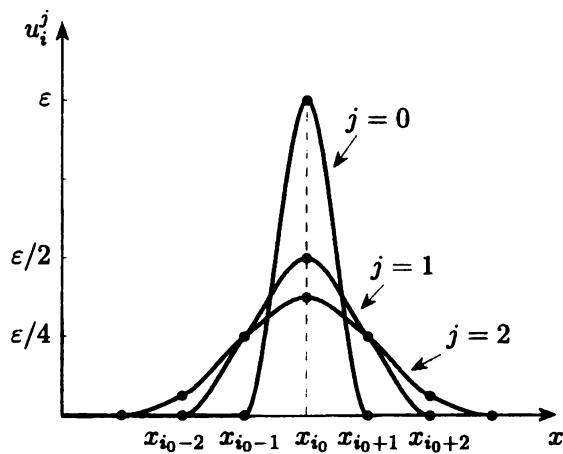


Рис. 15. Значения  $u_i^j$  на первых трех временных слоях при  $\sigma = \frac{1}{4}$

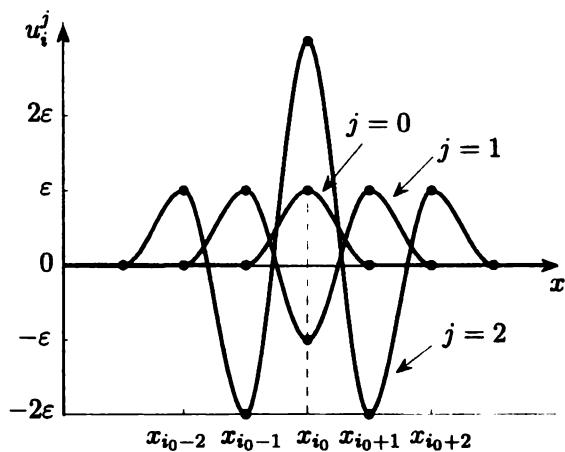


Рис. 16. Значения  $u_i^j$  на первых трех временных слоях при  $\sigma = 1$

Используя формулу Эйлера  $e^{i\phi h} = \cos \phi h + i \sin \phi h$ , получаем

$$q - 1 = 2\sigma(\cos \phi h - 1) = -2\sigma 2 \sin^2 \frac{\phi h}{2},$$

или

$$q = 1 - 4\sigma \sin^2 \frac{\phi h}{2}.$$

Решение будет устойчивым, если его модуль не будет расти, а это выполняется при условии  $|q| \leq 1$ . Но в нашем случае всегда  $q \leq 1$ , а условие  $q \geq -1$  выполняется, если  $4\sigma \leq 2$ , или

$$\sigma \leq \frac{1}{2},$$

это *условие устойчивости явной разностной схемы*.

В терминах шагов это условие выглядит как  $a\tau \leq \frac{1}{2}h^2$  и является весьма обременительным. Например, при  $a = 1$ ,  $l = 1$ ,  $T = 1$ , при шаге  $h = 0.1$  требуется выбирать временной шаг  $\tau \leq 0.005$ , а при  $h = 0.01$  требуется выбирать временной шаг  $\tau \leq 0.00005$ , т. е. 20 000 временных слоев. Это ведет к увеличению вычислительной погрешности, и поэтому явная схема может применяться только для расчетов с небольшой точностью.

### 6.3. Неявная и другие схемы для решения одномерного уравнения теплопроводности

Рассмотрим ту же задачу, что и в предыдущем разделе. При дискретизации будем использовать те же обозначения. Выберем для аппроксимации частной производной по времени  $t$  формулу численного дифференцирования по двум узлам на нижний слой, а для второй производной по пространственной переменной  $x$  выберем формулу численного дифференцирования по трем узлам на середину при фиксированном верхнем слое  $j + 1$ , получим

$$\frac{u_i^{j+1} - u_i^j}{\tau} = a^2 \frac{u_{i-1}^{j+1} - 2u_i^{j+1} + u_{i+1}^{j+1}}{h^2} + f_i^{j+1},$$

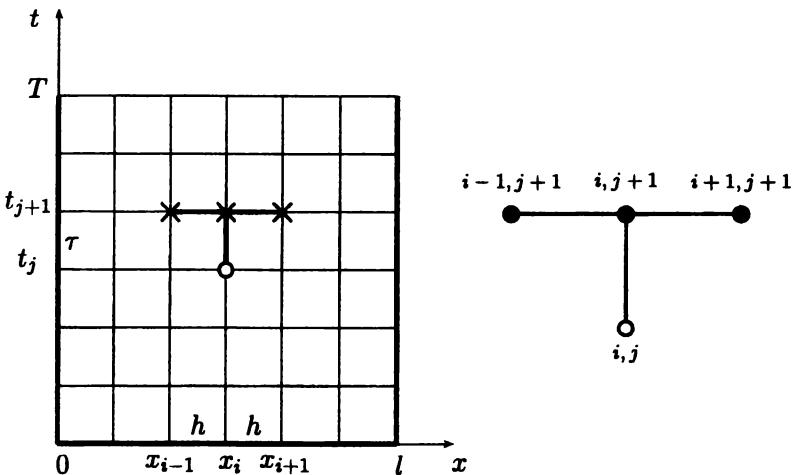


Рис. 17. Конечно-разностная сетка и шаблон неявной схемы для одномерного уравнения параболического типа.

для  $i = 1, \dots, n - 1$  и  $j = 0, 1, \dots, J - 1$ . Система дополняется теми же начальными и граничными условиями. Приводя подобные при неизвестных на  $j + 1$  слое, получим

$$\sigma u_{i-1}^{j+1} - (1 + 2\sigma)u_i^{j+1} + \sigma u_{i+1}^{j+1} = -\tau f_i^{j+1} - u_i^j,$$

т. е. трехдиагональную систему с диагональным преобладанием. Эта система может быть эффективно решена методом прогонки, причем число операций (умножения) на одном слое пропорционально  $n$ , а всего для реализации алгоритма требуется операций порядка  $nJ$ . Заметим, что для явной схемы порядок операций тот же. Это наименьший возможный порядок для данной задачи. Алгоритмы с таким порядком называются **экономичными**. Явная и неявная схемы относятся к числу экономичных. Конечно-разностная сетка и шаблон неявной схемы для уравнения параболического типа показаны на рис. 17.

Так же как в случае явной схемы, проверяется утверждение

о том, что при условии соответствующей гладкости решения порядок невязки неявной схемы  $O(\tau + h^2)$ .

Исследуем вопрос об устойчивости с помощью метода гармоник. Будем искать решение однородной задачи с нулевыми граничными условиями

$$\sigma u_{k-1}^{j+1} - (1 + 2\sigma)u_k^{j+1} + \sigma u_{k+1}^{j+1} = -u_k^j$$

в виде  $u_k^j = q^j e^{ik\phi h}$ . Подставляя в схему, получаем после сокращения на  $q^j e^{ik\phi h}$  и использования формулы Эйлера

$$\sigma q^2 \cos \phi h - (1 + 2\sigma)q = -1.$$

После тригонометрических преобразований, выражая  $q$ , получаем

$$q = \frac{1}{1 + 4\sigma \sin^2 \frac{\phi h}{2}} \leq 1.$$

Таким образом неявная схема устойчива при любом  $\sigma$ . Схемы, устойчивые при любом соотношении шагов, называются *абсолютно устойчивыми*. Неявная схема абсолютно устойчива и сходится с порядком  $O(\tau + h^2)$ .

Низкий порядок сходимости, особенно по  $\tau$ , вызывает вопрос о конструировании схем с более высоким порядком сходимости. Например, можно попробовать заменить формулу численного дифференцирования для первой производной по времени по двум узлам на формулу численного дифференцирования по трем узлам. Получим схему «явный крест», которая имеет порядок невязки  $O(\tau^2 + h^2)$ , но является абсолютно неустойчивой. Геометрическая интерпретация этой схемы представлена на рис. 18.

Можно рассмотреть комбинации явной и неявной схем. Среди них выделятся схема Кранка–Никольсон:

$$\frac{u_i^{j+1} - u_i^j}{\tau} = \frac{a^2}{2} \left( \frac{u_{i-1}^j - 2u_i^j + u_{i+1}^j}{h^2} + \frac{u_{i-1}^{j+1} - 2u_i^{j+1} + u_{i+1}^{j+1}}{h^2} \right) + f_i^{j+\frac{1}{2}}.$$

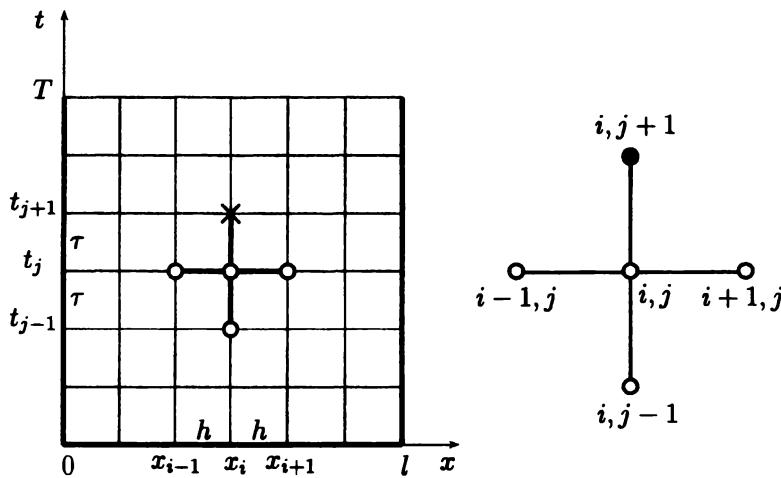


Рис. 18. Конечно-разностная сетка и шаблон схемы «явный крест» для одномерного уравнения параболического типа

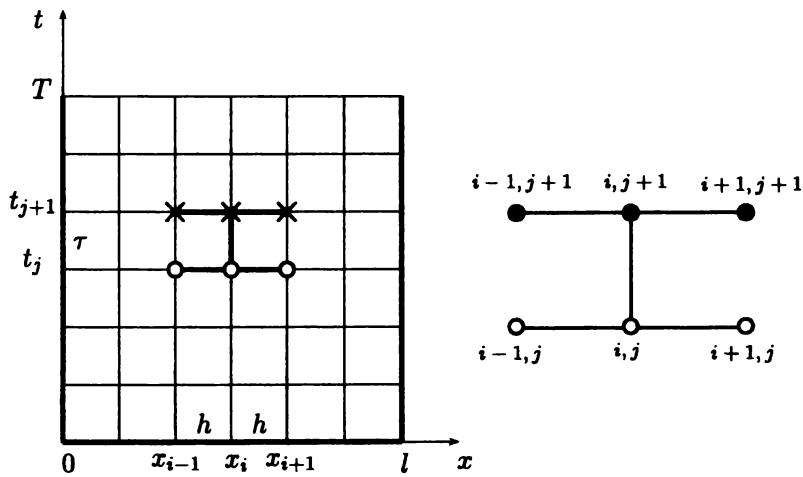


Рис. 19. Конечно-разностная сетка и шаблон неявной схемы Кранка-Никольсон для одномерного уравнения параболического типа

Эта схема, так же как и неявная схема, сводится к решению линейной системы методом прогонки, имеет порядок невязки  $O(\tau^2 + h^2)$  и является абсолютно устойчивой. Конечно-разностная сетка и шаблон схемы Кранка–Никольсон для уравнения параболического типа показаны на рис. 19.

#### 6.4. Метод прямых для решения одномерного уравнения теплопроводности

Проведем дискретизацию только по одной, пространственной переменной.

Разобьем отрезок  $[0, l]$  точками (узлами)  $x_i = ih$ ,  $i = 0, 1, \dots, n$ , где  $h = \frac{l}{n}$  — шаг. Точное решение в узлах  $u(x_i, t)$  будем заменять приближенным  $u_i(t)$ . Обозначим также  $f(x_i, t) = f_i(t)$ ,  $\varphi(x_i) = \varphi_i$ .

Заменим в уравнении теплопроводности вторую производную по пространственной переменной по формуле численного дифференцирования по трем узлам на середину, тогда уравнение теплопроводности заменится системой обыкновенных дифференциальных уравнений

$$\dot{u}_i(t) = a^2 \frac{u_{i-1}(t) - 2u_i(t) + u_{i+1}(t)}{h^2} + f_i(t),$$

для  $i = 1, \dots, n - 1$ . Система дополняется еще двумя уравнениями  $u_0(t) = g_0(t)$ ,  $u_n(t) = g_1(t)$  и начальными условиями

$$u_i(0) = \varphi_i, \quad i = 1, \dots, n - 1.$$

Решения  $u_i(t)$  ( $i = 1, \dots, n - 1$ ) находятся вдоль соответствующих прямых  $x = x_i$ . На рис. 20 точки и жирные линии показывают места, где значения функции  $u$  заданы.

Численные методы решения такой задачи Коши хорошо разработаны (см. соответствующий раздел). Однако имеется проблема: при увеличении размерности системы  $n$  увеличивается жесткость системы, так как в коэффициенты системы входит множитель  $\frac{1}{h^2}$ . Отметим, кроме того, что после выбора метода численного интегрирования этой системы по времени все

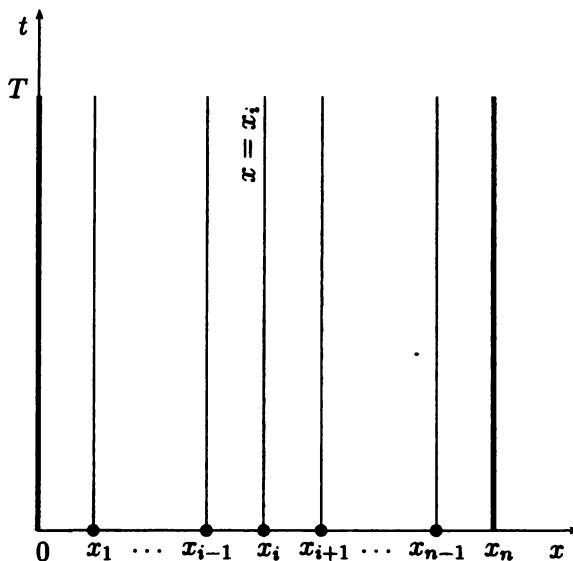


Рис. 20. Схема метода прямых с дискретизацией по  $x$

равно происходит дискретизация по двум переменным, и метод прямых оказывается просто эквивалентной формой двумерных разностных схем.

### 6.5. Численные методы решения уравнения гиперболического типа

Рассмотрим уравнение гиперболического типа с двумя независимыми переменными

$$\frac{\partial^2 u}{\partial t^2} = a^2 \frac{\partial^2 u}{\partial x^2} + f(x, t),$$

где  $x \in (0, l)$ ,  $t > 0$  — независимые переменные;  $u(x, t)$  — искомая функция.

Заданы начальные условия

$$u(x, 0) = \varphi(x), x \in [0, l],$$

$$\frac{\partial u(x, 0)}{\partial t} = \psi(x), x \in [0, l],$$

и граничные условия

$$u(0, t) = g_0(t), t > 0,$$

$$u(l, t) = g_1(t), t > 0.$$

Эта задача интерпретируется как процесс колебаний (отклонения  $u(x, t)$  от положения равновесия) однородного твердого тела (струны) длиной  $l$  ( $x$  – пространственная координата точки тела) в зависимости от времени  $t$ . Здесь  $a^2$  – коэффициент упругости,  $f(x, t)$  – действующая на тело сила. Заданы начальное положение  $\varphi(x)$  точек струны и их начальная скорость  $\psi(x)$ , на границах струны поддерживаются заданные отклонения: на левой границе  $g_0(t)$ , на правой –  $g_1(t)$  (краевая задача первого рода). Это уравнение еще называют *волновым уравнением*. Предполагается, что задача имеет единственное решение, определяемое в классическом смысле.

Численные методы решения этой задачи также можно разбить на два класса: сеточные методы и метод прямых. Рассмотрим *явный сеточный метод*.

Разобьем отрезок  $[0, l]$  точками  $x_i = ih$ ,  $i = 0, 1, \dots, n$ , где  $h = \frac{l}{n}$  – пространственный шаг. Пусть  $\tau > 0$  – временной шаг и  $t_j = j\tau$ , для  $j = 0, 1, \dots, J$ , и пусть  $\tau J = T$ . Пары  $(x_i, t_j)$  будем называть узлами сетки; точное решение в узлах  $u(x_i, t_j)$  будем заменять приближенным  $u_i^j$ . Обозначим также  $f(x_i, t_j) = f_i^j$ ,  $\varphi(x_i) = \varphi_i$ ,  $g_0(y_j) = g_0^j$ ,  $g_1(y_j) = g_1^j$ . Индекс  $j$  будем называть номером временного слоя.

Выберем для аппроксимации вторых частных производных по времени  $t$  и пространству  $x$  формулы дифференцирования

по трем узлам на середину, получим

$$\frac{u_i^{j+1} - 2u_i^j + u_i^{j-1}}{\tau^2} = a^2 \frac{u_{i-1}^j - 2u_i^j + u_{i+1}^j}{h^2} + f_i^j,$$

для  $i = 1, \dots, n-1$ , и  $j = 1, \dots, J-1$ . Эта система уравнений дополняется граничными условиями

$$\begin{aligned} u_0^j &= g_0^j, \quad j = 1, \dots, J, \\ u_n^j &= g_1^j, \quad j = 1, \dots, J. \end{aligned}$$

Из этой системы уравнений неизвестное на  $j+1$  слое выражается явно:

$$u_i^{j+1} = -u_i^{j-1} + \sigma u_{i-1}^j + 2(1-\sigma)u_i^j + \sigma u_{i+1}^j + \tau^2 f_i^j,$$

где

$$\sigma = \frac{a^2 \tau^2}{h^2} > 0.$$

Аналогично явной схеме для уравнения теплопроводности проверяется, что при условии соответствующей гладкости решения  $u(x, t)$  невязка метода имеет порядок  $O(\tau^2 + h^2)$ . Конечно-разностная сетка и шаблон данной схемы такие же, как и для схемы «явный крест», показанной на рис. 18.

Разностная схема относится к числу *трехслойных*, поэтому нужно задать значения на первых двух слоях.

На нулевом слое значения искомой функции можно взять из начальных условий

$$u_i^0 = \varphi_i, \quad i = 0, 1, \dots, n.$$

Простейший способ определить значения на первом слое — использовать аппроксимацию для первой производной в начальных условиях по двум узлам на нижний край:

$$\frac{u_i^1 - u_i^0}{\tau} = \psi_i,$$

откуда

$$u_i^1 = \varphi_i + \tau \psi_i, \quad i = 0, 1, \dots, n.$$

Этот способ аппроксимации имеет первый порядок малости по  $\tau$ , другой способ — *метод фиктивного узла* (см. раздел 2.3), который имеет второй порядок малости по  $\tau$ .

Исследуем методом гармоник вопрос об устойчивости для однородного уравнения явной схемы

$$u_k^{j+1} = -u_k^{j-1} + \sigma u_{k-1}^j + 2(1 - \sigma)u_k^j + \sigma u_{k+1}^j.$$

Будем искать решение в виде

$$u_k^j = q^j e^{ik\phi h}.$$

Подставляя его в уравнение, получаем после сокращения на  $q^j e^{ik\phi h}$  и использования формулы Эйлера квадратное уравнение

$$q^2 - 2q(\sigma \cos \phi h - \sigma + 1) + 1 = 0,$$

или, после тригонометрических преобразований,

$$q^2 - 2q(1 - 2\sigma \sin^2 \frac{\phi h}{2}) + 1 = 0.$$

Дискриминант этого уравнения вычисляется как

$$D = (1 - 2\sigma \sin^2 \frac{\phi h}{2})^2 - 1.$$

В случае положительности дискриминанта оба корня вещественные, различные, по теореме Виета один из них по модулю больше единицы, что дает неустойчивость. В случае, если  $D \leq 0$ , оба корня по теореме Виета имеют модули, равные единице, и имеет место устойчивость. Но в случае  $\sigma \leq 1$  условие  $D \leq 0$  выполняется при любых  $\phi h$ .

Таким образом, условие

$$\sigma \leq 1,$$

известное как *условие Куранта*, обеспечивает устойчивость явной разностной схемы. Отметим, что, в отличие от явной схемы для уравнения теплопроводности, это условие гораздо менее обременительное, так как в определении  $\sigma$  шаги  $\tau$  и  $h$  входят с одним порядком.

Рассмотрим *метод прямых* для волнового уравнения. Главное отличие от метода прямых для уравнения теплопроводности состоит в том, что возникает система обыкновенных дифференциальных уравнений, каждое уравнение не первого, а второго порядка (т. е. фактически порядок системы в два раза выше):

$$\ddot{u}_i(t) = a^2 \frac{u_{i-1}(t) - 2u_i(t) + u_{i+1}(t)}{h^2} + f_i(t),$$

для  $i = 1, \dots, n - 1$ . Система дополняется еще двумя уравнениями  $u_0(t) = g_0(t)$ ,  $u_n(t) = g_1(t)$  и начальными условиями

$$u_i(0) = \varphi_i, \quad \dot{u}_i(0) = \psi_i, \quad i = 1, \dots, n - 1.$$

Все замечания, высказанные при рассмотрении метода прямых для уравнения теплопроводности, справедливы и в этом случае.

## 6.6. Численные методы решения пространственного уравнения параболического типа

Рассмотрим уравнение параболического типа с двумя независимыми пространственными переменными  $x$  и  $y$  и временной независимой переменной  $t$ :

$$\frac{\partial u}{\partial t} = a^2 \left( \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right) + f(x, y, t).$$

Здесь  $0 < x < 1$ ,  $0 < y < 1$ ,  $t > 0$ . Предполагается, что заданы соответствующие начальные и граничные условия.

Проведем дискретизацию по трем переменным, введем  $h = \frac{1}{n}$ ,  $\tau > 0$ , точки  $x_i = ih$ ,  $i = 0, 1, \dots, n$ ,  $y_j = jh$ ,  $j = 0, 1, \dots, n$ ,  $t_k = k\tau$ ,  $k = 0, 1, \dots, m$ , узлы  $(x_i, y_j, t_k)$ . Точное решение в узлах  $u(x_i, y_j, t_k)$  будем заменять приближенным  $u_{i,j}^k$ , обозначим также  $f(x_i, y_j, t^k) = f_{i,j}^k$ .

Применяя для аппроксимации частной производной по времени  $t$  формулу численного дифференцирования по двум узлам на нижний слой, а для второй производной по пространственным переменным  $x$  и  $y$  — формулы численного дифференцирования по трем узлам на середину при фиксированном нижнем слое  $k$ , получим явную схему

$$\begin{aligned} \frac{u_{i,j}^{k+1} - u_{i,j}^k}{\tau} = a^2 \left( \frac{u_{i-1,j}^k - 2u_{i,j}^k + u_{i+1,j}^k}{h^2} + \right. \\ \left. + \frac{u_{i,j-1}^k - 2u_{i,j}^k + u_{i,j+1}^k}{h^2} \right) + f_{i,j}^k. \end{aligned}$$

Явная схема проста в реализации, но является условно устойчивой, причем условие устойчивости  $\sigma \leq \frac{1}{4}$ , где  $\sigma = \frac{\tau a^2}{h^2}$ , является очень жестким. Геометрическая интерпретация явной схемы показана на рис. 21.

Неявная схема

$$\begin{aligned} \frac{u_{i,j}^{k+1} - u_{i,j}^k}{\tau} = a^2 \left( \frac{u_{i-1,j}^{k+1} - 2u_{i,j}^{k+1} + u_{i+1,j}^{k+1}}{h^2} + \right. \\ \left. + \frac{u_{i,j-1}^{k+1} - 2u_{i,j}^{k+1} + u_{i,j+1}^{k+1}}{h^2} \right) + f_{i,j}^{k+1} \end{aligned}$$

абсолютно устойчива, но, в отличие от случая одной пространственной переменной, не приводит к системе с трехдиагональной структурой, поэтому неэффективна. Геометрическая интерпретация неявной схемы показана на рис. 22.

К настоящему времени создано несколько вариантов схем, которые сочетают устойчивость и эффективность счета и в

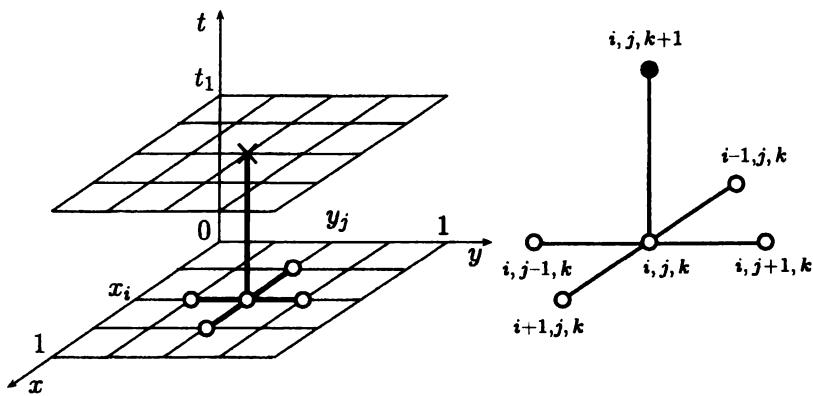


Рис. 21. Конечно-разностная сетка и шаблон явной схемы для пространственного уравнения параболического типа.

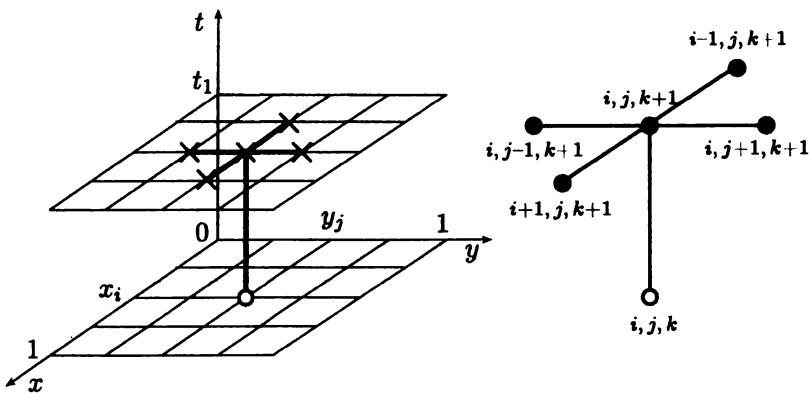


Рис. 22. Конечно-разностная сетка и шаблон неявной схемы для пространственного уравнения параболического типа.

многомерных случаях, эти схемы получили названия *схема продольно-поперечной прогонки, метод дробных шагов, схема переменных направлений*. Рассмотрим один из вариантов, который был предложен в 1956 г. Дугласом, Писменом и Ричфордом.

Для удобства записи введем разностные операторы:

$$\Lambda_x : u_{i,j}^k \rightarrow \frac{u_{i-1,j}^k - 2u_{i,j}^k + u_{i+1,j}^k}{h^2},$$

$$\Lambda_y : u_{i,j}^k \rightarrow \frac{u_{i,j-1}^k - 2u_{i,j}^k + u_{i,j+1}^k}{h^2},$$

тогда, например, рассмотренные выше схемы запишутся так:

$$\frac{u_{i,j}^{k+1} - u_{i,j}^k}{\tau} = a^2(\Lambda_x u_{i,j}^k + \Lambda_y u_{i,j}^k) + f_{i,j}^k$$

(явная схема),

$$\frac{u_{i,j}^{k+1} - u_{i,j}^k}{\tau} = a^2(\Lambda_x u_{i,j}^{k+1} + \Lambda_y u_{i,j}^{k+1}) + f_{i,j}^{k+1}$$

(неявная схема).

Разделим результат действия операторов на одном слое, а затем поменяем их местами. Для удобства введем промежуточный слой с индексом  $k + \frac{1}{2}$ , тогда метод запишется в виде

$$\frac{u_{i,j}^{k+\frac{1}{2}} - u_{i,j}^k}{\tau/2} = a^2(\Lambda_x u_{i,j}^k + \Lambda_y u_{i,j}^{k+\frac{1}{2}}) + f_{i,j}^k,$$

$$\frac{u_{i,j}^{k+1} - u_{i,j}^{k+\frac{1}{2}}}{\tau/2} = a^2(\Lambda_x u_{i,j}^{k+1} + \Lambda_y u_{i,j}^{k+\frac{1}{2}}) + f_{i,j}^{k+1}.$$

Геометрическая интерпретация этой схемы показана на рис. 23. На каждом дробном шаге нужно решать линейную систему с трехдиагональной структурой. Если применять прогонку, то

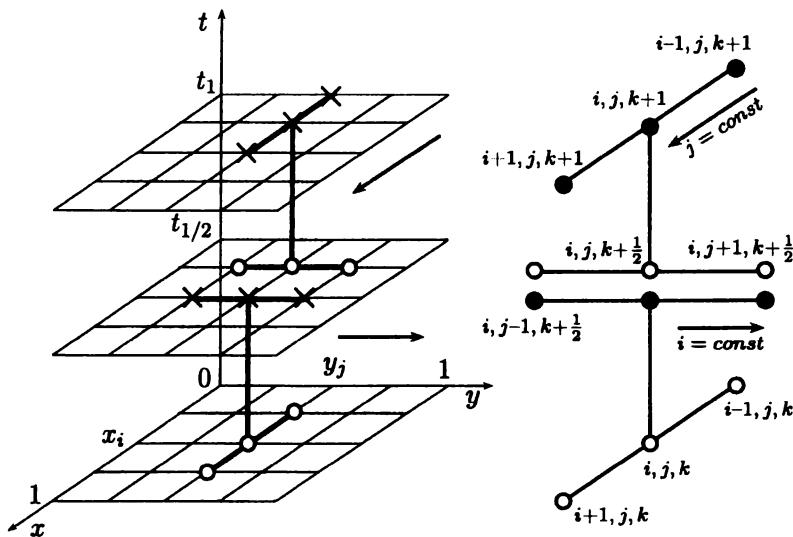


Рис. 23. Конечно-разностная сетка и шаблон схемы переменных направлений для пространственного уравнения параболического типа: стрелками показаны направления действия частей шаблона

общее число операций пропорционально  $n^2m$ , т. е. метод относится к числу экономичных. Кроме того, метод абсолютно устойчив. Этим же методом можно решать уравнения с числом независимых пространственных переменных, больших, чем две.

## 6.7. Схема счета на установление для решения эллиптического уравнения

Вернемся к рассмотрению задачи эллиптического типа

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = f(x, y),$$

где  $x \in (0, 1)$ ,  $y \in (0, 1)$ , на границах области  $\Gamma$  заданы условия

$$u(x, y)|_{(x, y) \in \Gamma} = g(x, y) = \begin{cases} \varphi_0(y), & x = 0, y \in [0, 1] \\ \varphi_1(y), & x = 1, y \in [0, 1] \\ \psi_0(x), & y = 0, x \in [0, 1] \\ \psi_1(x), & y = 1, x \in [0, 1]. \end{cases}$$

Введем вспомогательную задачу параболического типа с двумя пространственными независимыми переменными и временной независимой переменной

$$\frac{\partial u}{\partial t} = a^2 \left( \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right) + f(x, y, t),$$

$$0 < x < 1, \quad 0 < y < 1, \quad t > 0,$$

с граничными условиями  $u(x, y, t)|_{(x, y) \in \Gamma} = g(x, y)$ . Зададим также каким-то образом начальные условия  $u(x, y, 0) = \varphi(x, y)$  (начальное приближение в решении эллиптической задачи).

Если при  $t \rightarrow \infty$  решение этой вспомогательной задачи  $u(x, y, t)$  стремится к функции  $u(x, y)$ , не зависящей от времени  $t$ , то при определенных условиях эта функция будет являться решением исходного эллиптического уравнения. Вспомогательную задачу можно решать схемой переменных направлений, при этом число операций пропорционально  $n^2 m$ , где  $n$  — число точек разбиения по пространственной переменной, а  $m$  — число шагов по времени до установления. Это может оказаться эффективней, чем решать возникающую при прямом сеточном методе систему размерности  $n^2$ .

## 6.8. Элементы общей теории разностных схем

Во всех разностных схемах решения задач математической физики (и не только математической физики) сходимость получается как результат наличия двух факторов: локальной аппроксимации и устойчивости (тезис Лакса). Изложим этот тезис в очень общем виде.

Пусть имеется исходное абстрактное уравнение

$$Au = F,$$

где  $u(x)$  – неизвестная функция, определенная на множестве  $D$  из  $R^n$ . Отметим, что в это уравнение можно включить за счет повышения размерности  $n$  пространства не только основное уравнение, но и начальные и граничные условия. Будем считать, что  $u(\cdot)$  – элемент банахова пространства функций  $B$ , правая часть уравнения  $F(\cdot)$  – из банахова пространства функций  $F$ ,  $A$  – оператор, переводящий  $B$  в  $F$ .

Будем предполагать, что эта задача корректно поставлена, т. е. имеет решение, притом единственное и непрерывно зависящее от входящих в нее параметров. В случае линейности оператора  $A$  требование корректности эквивалентно обратимости и ограниченности этого оператора.

Пусть  $D_h$  – некоторое множество (*сетка*) из  $R^n$ , состоящее из конечного числа элементов  $x_h$ , называемых *узлами сетки*. Пусть сетка характеризуется некоторым векторным параметром  $h = (h_1, h_2, \dots, h_n)$ , положительные координаты  $h_i$  которого характеризуют расстояние по координатам между узлами сетки. Его единая характеристика – число  $|h|$ , называется *модулем сетки*.

Наряду с исходным уравнением, рассмотрим *сеточное уравнение*

$$A_h y_h = F_h,$$

где  $y_h(x_h)$  – неизвестная функция, определенная на множестве  $D_h$ . Будем считать, что  $y_h(\cdot)$  – элемент банахова пространства функций  $B_h$ , правая часть уравнения  $F_h(\cdot)$  – из банахова пространства функций  $F_h$ ,  $A_h$  – оператор, переводящий  $B_h$  в  $F_h$ .

Назовем сеточное уравнение *устойчивым*, если оно корректно поставлено, т. е. имеет решение, притом единственное и непрерывно зависящее от входящих в уравнение параметров. Назовем сеточное уравнение *условно устойчивым*, если

оно корректно поставлено при определенных соотношениях координат вектора  $h$ .

Пусть задан оператор проецирования  $\varrho$ , переводящий функции из  $\mathbf{B}$  в  $\mathbf{B}_h$ , в частности, если  $D_h \subset D$ , то оператор проецирования можно задать тривиальным образом  $u_h(x_h) = u(x_h)$ . С помощью оператора проецирования решение исходного уравнения индуцирует сеточную функцию  $u_h = \varrho u$ .

Будем говорить, что решение сеточного уравнения *сходится* к решению исходного уравнения, если

$$\|y_h(\cdot) - u_h(\cdot)\|_{\mathbf{B}_h} \rightarrow 0 \text{ при } |h| \rightarrow 0.$$

Будем говорить, что решение сеточного уравнения *сходится с порядком  $p$*  к решению исходного уравнения, если найдется такая константа  $C$ , что

$$\|y_h(\cdot) - u_h(\cdot)\|_{\mathbf{B}_h} \leq C|h|^p.$$

*Невязкой* назовем сеточную функцию

$$\psi = A_h u_h - F_h.$$

Будем говорить, что сеточное уравнение *аппроксимирует* исходное уравнение, если невязка стремится к нулю (в соответствующей норме) при стремлении  $|h|$  к нулю. *Порядком аппроксимации* назовем число  $p > 0$ , такое, что найдется константа  $C$ , для которой выполняется оценка

$$\|\psi(\cdot)\|_{\mathbf{F}_h} \leq C|h|^p.$$

**Теорема 5.** *Пусть сеточное уравнение корректно и аппроксимирует исходное уравнение, тогда его решение сходится к решению исходного уравнения. Если оператор  $A_h$  в сеточном уравнении линеен и порядок аппроксимации  $p > 0$ , то имеет место сходимость с тем же порядком  $p$ .*

**Доказательство.** Запишем два уравнения: сеточное уравнение

$$A_h y_h = F_h$$

и определение невязки

$$A_h u_h = F_h + \psi,$$

которое также можно рассматривать как сеточное уравнение с другой правой частью. В силу корректности (устойчивости по правой части) этого уравнения для любого  $\varepsilon > 0$  найдется  $\delta > 0$ , такое, что из неравенства

$$\|\psi\|_{\mathbf{F}_h} < \delta$$

следует

$$\|y_h(\cdot) - u_h(\cdot)\|_{\mathbf{B}_h} < \varepsilon.$$

Но, так как сеточное уравнение аппроксимирует исходное уравнение, то по числу  $\delta > 0$  найдется число  $\eta > 0$ , такое, что из  $|h| < \eta$  следует  $\|\psi\|_{\mathbf{F}_h} < \delta$ , что и доказывает первую часть утверждения теоремы.

Если  $A_h$  — линейный оператор, то имеем

$$A_h(y_h - u_h) = \psi,$$

а так как этот оператор обратимый и ограниченный, то

$$\|y_h - u_h\|_{\mathbf{B}_h} \leq \|A_h^{-1}\| \|\psi\| \leq MC|h|^p.$$

Это неравенство доказывает сходимость с порядком  $p$ .

В качестве примера рассмотрим вложение в эту общую схему доказательства сходимости неявной схемы для уравнения параболического типа.

Введем множество  $D = \{(x, t) : 0 \leq x \leq 1, 0 \leq t\}$  и для функций  $u(x, y)$ , определенных на  $D$ , введем оператор

$$Au = \begin{cases} \frac{\partial u}{\partial t} - a^2 \frac{\partial^2 u}{\partial x^2}, & 0 < x < 1, t > 0, \\ u(x, y), & t = 0, x \in [0, 1], \\ u(x, y), & x = 0, t > 0, \\ u(x, y), & x = 1, t > 0 \end{cases}$$

и функцию  $F(x, t)$ , определенную на  $D$ :

$$F = \begin{cases} f(x, t), & 0 < x < 1, t > 0, \\ \varphi(x), & t = 0, x \in [0, 1], \\ g_0(t), & x = 0, t > 0, \\ g_1(t), & x = 1, t > 0, \end{cases}$$

тогда уравнение теплопроводности, начальные и краевые условия можно записать в виде одного уравнения:

$$Au = F.$$

Разобьем отрезок  $[0, 1]$  точками  $x_i = ih_x$ ,  $i = 0, 1, \dots, n$ , где  $h_x = \frac{1}{n}$ . Пусть  $\tau > 0$  и  $t_j = j\tau$ , для  $j = 0, 1, \dots, J$ , где  $\tau J = T$ . Пары  $(x_i, t_j)$  будем называть узлами сетки, а их множество (сетку) обозначим  $D_h$ . Сетка характеризуется векторным параметром  $h = (h_x, \tau)$ , модуль можно выбрать, например, так:  $|h| = \max\{h_x^2, \tau\}$ .

Обозначим сеточную функцию  $y_h(x_i, t_j) = y_i^j$  и рассмотрим сеточное уравнение

$$A_h y_h = F_h,$$

где

$$A_h y_i^j = \begin{cases} \frac{y_i^{j+1} - y_i^j}{\tau} - a^2 \frac{y_{i-1}^{j+1} - 2y_i^{j+1} + y_{i+1}^{j+1}}{h^2}, & 1 \leq i \leq n-1, 0 \leq j \leq J-1, \\ y_i^j, & i = 0, 1, \dots, n, j = 0, \\ y_i^j, & i = 0, j = 1, \dots, J, \\ y_i^j, & i = n, j = 1, \dots, J, \end{cases}$$

$$F_h = \begin{cases} f(x_i, t_{j+1}), & i = 1, \dots, n-1, j = 0, 1, \dots, J-1, \\ \varphi(x_i), & i = 0, 1, \dots, n, j = 0, \\ g_0(t_j), & i = 0, j = 1, \dots, J, \\ g_1(t_j), & i = n, j = 1, \dots, J. \end{cases}$$

Так как  $D_h \subseteq D$ , то оператор проектирования — тождественный оператор, который индуцирует из точного решения

сеточную функцию  $u_h = u(x_i, t_j)$ . С помощью тейлоровского разложения (см. лемму в описании явной схемы для уравнения теплопроводности) устанавливается, что при определенной гладкости решения сеточное уравнение аппроксимирует исходное с порядком  $h^2 + \tau$ , или, согласно введенному модулю, с первым порядком по  $|h|$ .

Главный вопрос — проверить корректность сеточного уравнения. В нашем случае нужно доказать устойчивость по правой части  $f$ , по начальным и граничным функциям. Проведем доказательство в самом сложном моменте — по начальной функции  $\varphi$ . Предполагая, что  $f = 0$ ,  $g_0 = 0$ ,  $g_1 = 0$ , перепишем сеточное уравнение в виде

$$\sigma y_{i-1}^{j+1} - (1 + 2\sigma)y_i^{j+1} + \sigma y_{i+1}^{j+1} = -y_i^j,$$

где  $i = 1, \dots, n-1$ ,  $j = 0, 1, \dots, J-1$ , с начальными условиями

$$y_i^0 = \varphi(x_i), \quad x = 0, 1, \dots, n,$$

где  $\sigma = \frac{a^2 \tau}{h^2} > 0$ .

Введем послойный вектор  $y^j = (y_0^j, y_1^j, \dots, y_n^j)$  и для него норму

$$\|y^j\| = \max_{i=0,1,\dots,n} |y_i^j|.$$

Тогда, если

$$\|y^0\| = \max_{i=0,1,\dots,n} |\varphi(x_i)| = \varepsilon,$$

то в силу того, что сеточное уравнение представляет собой трехдиагональную линейную систему с диагональным преобладанием  $\Delta = 1$ , выполняется

$$\|y^{j+1}\| \leq \varepsilon,$$

для всех  $j = 0, 1, \dots, J-1$ , что и означает устойчивость.

## Список библиографических ссылок

1. Годунов С. К., Рябенский В. С. Введение в теорию разностных схем. М. : Физматгиз, 1962. 340 с.
2. Завьялов Ю. С., Квасов Б. И., Мирошниченко В. Л. Методы сплайн-функций. М. : Наука, 1980. 352 с.
3. Крылов В. И., Бобков В. В., Монастырный П. И. Начала теории вычислительных методов : в 5 т. Дифференциальные уравнения. Минск : Наука и техника, 1982. 286 с.
4. Крылов В. И., Бобков В. В., Монастырный П. И. Начала теории вычислительных методов : в 5 т. Интегральные уравнения, некорректные задачи и улучшение сходимости. Минск : Наука и техника, 1984. 263 с.
5. Крылов В. И., Бобков В. В., Монастырный П. И. Начала теории вычислительных методов : в 5 т. Уравнения в частных производных. Минск : Наука и техника, 1986. 311 с.
6. Самарский А. А. Теория разностных схем. М. : Наука, 1989. 616 с.
7. Хайрер Э., Норсетт С., Ваннер Г. Решение обыкновенных дифференциальных уравнений. Нежесткие задачи. М. : Мир, 1990. 512 с.
8. Хайрер Э., Ваннер Г. Решение обыкновенных дифференциальных уравнений. Жесткие и дифференциально-алгебраические задачи. М. : Мир, 1999. 685 с.

# Оглавление

|   |           |
|---|-----------|
| <b>Предисловие</b>  | <b>3</b>  |
| <b>1. Численные методы решения дифференциальных уравнений. Задача Коши</b>  | <b>4</b>  |
| 1.1. Постановка задачи и предположения . . . . .                            | 4         |
| 1.2. Методы, основанные на тейлоровском разложении . . . . .                | 5         |
| 1.3. Явный метод Эйлера . . . . .   | 7         |
| 1.4. Метод Эйлера с пересчетом . . . . .                                    | 9         |
| 1.5. Метод Коши . . . . .   | 11        |
| 1.6. Семейство явных методов Рунге–Кутты второго порядка . . . . .          | 11        |
| 1.7. Явные методы Рунге–Кутты третьего и четвертого порядков . . . . .      | 13        |
| 1.8. Общий класс методов Рунге–Кутты . . . . .                              | 14        |
| 1.9. Явные методы Адамса . . . . .  | 18        |
| 1.10. Неявные методы Адамса . . . . .                                       | 20        |
| 1.11. Общий класс многошаговых методов . . . . .                            | 22        |
| 1.12. Линейные разностные уравнения и их устойчивость . . . . .             | 24        |
| 1.13. 0-устойчивость и сходимость многошаговых методов . . . . .            | 26        |
| 1.14. Жесткие системы. A-устойчивость . . . . .                             | 27        |
| <b>2. Численные методы решения краевых задач</b>                            | <b>31</b> |
| 2.1. Постановка краевых задач. Метод стрельбы . . . . .                     | 31        |
| 2.2. Метод разностной прогонки . . . . .                                    | 33        |
| 2.3. Метод разностной прогонки. Краевые условия третьего рода . . . . .     | 35        |
| 2.4. Метод разностной прогонки для нелинейного уравнения . . . . .          | 37        |
| 2.5. Самосопряженная краевая задача и ее физическая интерпретация . . . . . | 38        |
| 2.6. Вариационный метод Ритца . . . . .                                     | 39        |
| 2.7. Метод Галеркина . . . . .  | 43        |

|  |           |
|--|-----------|
| 2.8. Метод наименьших квадратов для решения краевой задачи . . . . .                         | 45        |
| 2.9. Метод коллокации . . . . .  | 46        |
| <b>3. Интерполяция сплайнами</b>   | <b>47</b> |
| 3.1. Интерполяционный кубический сплайн . . . . .  | 47        |
| 3.2. Эффективный способ построения сплайна . . . . .   | 49        |
| 3.3. Экстремальное свойство сплайна . . . . .  | 51        |
| 3.4. Сплайн-квадратура . . . . .   | 54        |
| 3.5. Метод конечных элементов . . . . .  | 57        |
| <b>4. Метод наименьших квадратов</b>   | <b>59</b> |
| 4.1. Метод наименьших квадратов в задаче решения линейных систем . . . . .                   | 59        |
| 4.2. Вырожденный метод наименьших квадратов в задаче решения линейных систем . . . . .       | 62        |
| 4.3. Метод наименьших квадратов в задаче приближения функций (дискретный вариант) . . . . .  | 63        |
| 4.4. Метод наименьших квадратов в задаче приближения функций (непрерывный вариант) . . . . . | 64        |
| <b>5. Численное решение интегральных уравнений</b>   | <b>67</b> |
| 5.1. Классификация линейных интегральных уравнений . . . . .                                 | 67        |
| 5.2. Метод разложения по собственным значениям . . . . .                                     | 68        |
| 5.3. Метод замены интеграла конечной суммой . . . . .  | 69        |
| 5.4. Метод замены ядра на вырожденное . . . . .  | 70        |
| <b>6. Численное решение уравнений математической физики</b>                                  | <b>73</b> |
| 6.1. Решение уравнения эллиптического типа методом сеток . . . . .                           | 73        |
| 6.2. Решение одномерного уравнения теплопроводности (явная схема) . . . . .                  | 76        |

|  |            |
|--|------------|
| 6.3. Неявная и другие схемы для решения одномерного уравнения теплопроводности . . . . . | 82         |
| 6.4. Метод прямых для решения одномерного уравнения теплопроводности . . . . .           | 86         |
| 6.5. Численные методы решения уравнения гиперболического типа . . . . .                  | 87         |
| 6.6. Численные методы решения пространственного уравнения параболического типа . . . . . | 91         |
| 6.7. Схема счета на установление для решения эллиптического уравнения . . . . .          | 95         |
| 6.8. Элементы общей теории разностных схем . . . . .                                     | 96         |
| <b>Список библиографических ссылок</b>   | <b>102</b> |

*Учебное издание*

Пименов Владимир Германович  
Ложников Андрей Борисович

**ЧИСЛЕННЫЕ МЕТОДЫ**  
**В двух частях**  
**Часть 2**

Учебное пособие

Заведующий редакцией М. А. Овочкина  
Редактор Н. В. Чапаева  
Корректор Н. В. Чапаева  
Оригинал-макет В. Г. Пименов

План выпуска 2014 г. Подписано в печать 20.11.2014.  
Формат 60×84 1/16. Бумага офсетная. Гарнитура Times.  
Уч.-изд. л. 4,6. Усл. печ. л. 6,27. Тираж 160 экз. Заказ 1693.  
Издательство Уральского университета  
620000, г. Екатеринбург, ул. Тургенева, 4.  
Отпечатано в Издательско-полиграфическом центре УрФУ  
620000, Екатеринбург, ул. Тургенева, 4.  
Тел.: + (343) 350-56-64, 350-90-13  
Факс: + 7 (343) 358-93-06  
E-mail: [press-urfu@mail.ru](mailto:press-urfu@mail.ru)

Для заметок

Д л я з а м е т о к