

В. И. КИРЕЕВ, А. В. ПАНТЕЛЕЕВ

ЧИСЛЕННЫЕ МЕТОДЫ

в примерах и задачах

Издание четвертое, исправленное

*ДОПУЩЕНО
УМО по образованию в области прикладной математики
и управления качеством в качестве учебного пособия
для студентов вузов, обучающихся по направлению
«Прикладная математика»*



САНКТ-ПЕТЕРБУРГ
МОСКВА
КРАСНОДАР
2022

ББК 22.193я73

К 43

Киреев В. И., Пантелеев А. В.

К 43 Численные методы в примерах и задачах: Учебное пособие. — 4-е изд., испр. — СПб.: Издательство «Лань», 2022. — 448 с.: ил. — (Учебники для вузов. Специальная литература).

ISBN 978-5-8114-1888-6

Пособие охватывает классические разделы численного анализа: методы алгебры, теории приближения функций одной переменной с их приложениями, разностные методы решения задач Коши и краевых задач для обыкновенных дифференциальных уравнений, численные методы решения уравнений математической физики с двумя и тремя независимыми переменными. Наряду с традиционными методами изложены новые экономичные, устойчивые и простые в реализации методы приближения функций, численного дифференцирования и интегрирования, решения задач Коши, основанные на применении интегрально-дифференциальных сплайнов.

В каждом разделе кратко изложены основные теоретические сведения, приведены решения типовых примеров и задачи для самостоятельного решения. Учебное пособие поддерживает компетентностную модель обучения: содержит модели требуемых знаний и умений решать типовые задачи предмета.

Для студентов, обучающихся по направлению «Прикладная математика» и для других математических, инженерно-технических и авиационных специальностей вузов, а также для аспирантов и научных работников.

ББК 22.193я73

Рецензенты:

В. А. КУДИНОВ — доктор физико-математических наук, профессор, зав. кафедрой «Теоретические основы теплотехники и гидродинамики» Самарского государственного технического университета;
Э. М. КАРТАШОВ — доктор физико-математических наук, профессор кафедры высшей и прикладной математики Московского государственного университета тонких химических технологий им. М. В. Ломоносова.

Обложка
E. A. ВЛАСОВА

*Охраняется Законом РФ об авторском праве.
Воспроизведение всей книги или любой ее части
запрещается без письменного разрешения издателя.
Любые попытки нарушения закона
будут преследоваться в судебном порядке.*

© Издательство «Лань», 2022
© В. И. Киреев, А. В. Пантелеев, 2022
© Издательство «Лань»,
художественное оформление, 2022

ВВЕДЕНИЕ

В.1. ПОНЯТИЕ О ЧИСЛЕННЫХ МЕТОДАХ

Проектирование и отработка современных летательных аппаратов, их отдельных узлов и блоков, а также других технических систем связаны с теоретическими расчетами и исследованиями, предваряющими выбор определяющих параметров конструкций. Эти расчеты проводятся с использованием вычислительных средств (компьютеров и их систем) и вычислительных методов. При этом обычно выполняются следующие этапы.

1. *Физическая постановка задачи.* Результатом этого этапа является общая формулировка задачи в содержательных терминах, т. е. что дано и что требуется определить. Например, рассчитать траекторию полета ракеты при заданных тяге двигателей, массе ракеты, ее аэродинамических характеристиках, при определенных метеоусловиях и др. Как правило, этот этап выполняется специалистом в конкретной предметной области.

2. *Поиск, выбор или модификация некоторой математической модели,* адекватной физической постановке задачи. На этом этапе осуществляются:

- выделение (запись) основных математических уравнений, соотношений, аппроксимационных формул, описывающих задачу;
- выделение (запись) дополнительных математических уравнений, связей, граничных или краевых условий;
- предварительное (априорное) обоснование математической модели.

Этот этап является очень важным, так как ошибочная или неудачная модель, неадекватная физической, сводит «на нет» все дальнейшие усилия по проектированию изделия. Заметим, что при решении многих задач выбираются, как правило, общепринятые математические модели.

3. *Разработка, выбор или модификация математического (аналитического, приближенно-аналитического или численного) метода,* наиболее целесообразного и экономичного. Этот этап осуществляется на основе имеющихся у исследователей знаний (субъективный подход), а также исходя из ресурсов компьютера — оперативной и внешней памяти, быстродействия, возможностей представления информации (объективный подход).

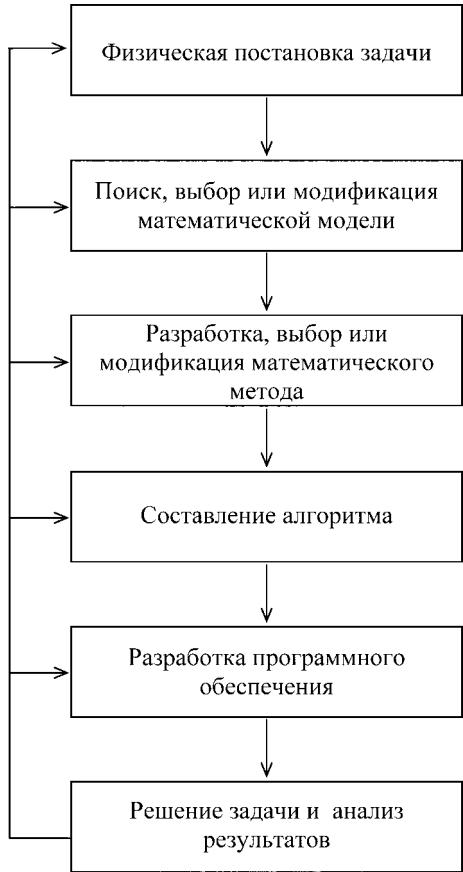


Рис. В.1

4. Составление алгоритма.
5. Разработка программного обеспечения.

6. Решение задачи: апостериорное обоснование модели и метода путем их методических и параметрических компьютерных исследований в привязке к реальному объекту. Этот этап включает выдачу рекомендаций и характеристик объекта в проектно-конструкторские подразделения.

В результате анализа полученного решения задачи может осуществляться переход к любому из описанных этапов для внесения соответствующих изменений. Изложенные этапы исследования прикладных задач схематически показаны на рисунке В.1.

Рассмотрим пример простейшей физической и математической моделей. При заданном количестве микробов или концентрации некоторого вещества y_0 в начальный момент времени $t = t_0$ получить зависимость, описывающую их дальнейшее изменение во времени. При $t > t_0$ данная физическая модель описывается обыкновенным дифференциальным уравнением и начальным условием:

$$\frac{dy}{dt} - y = 0; \quad (B.1)$$

$$y(t_0) = y_0. \quad (B.2)$$

Таким образом, уравнение (B.1) и дополнительное (начальное) условие (B.2) составляют математическую модель (дифференциальную, так как ей соответствует дифференциальное уравнение). Сопоставление решения задачи (B.1), (B.2) с результатами эксперимента подтверждает адекватность физической и математической моделей.

Другими классическими примерами простых математических моделей являются: определенный интеграл для нахождения площадей, объемов, длин кривых, центра тяжести тела и др.; система уравнений движения центра масс летательного аппарата с начальными условиями, заданными на старте; уравнение теплообмена, например для изучения процесса охлаждения или нагрева тела при подводе тепла.

Классическим средством изучения математических моделей и исследований на их основе свойств реальных объектов являются *аналитические*

методы, позволяющие получать точные решения в виде математических формул. Эти методы дают наиболее полную информацию о решении задачи, и они до настоящего времени не утратили своего значения. Однако, к сожалению, класс задач, для которого они могут использоваться, весьма ограничен. Поэтому решение широкого класса задач при отработке современных технических систем, как правило, осуществляется численными методами.

Численные методы — это методы приближенного решения задач прикладной математики, основанные на реализации алгоритмов, соответствующих математическим моделям. Наука, изучающая численные методы, называется также численным анализом, или вычислительной математикой.

Численные методы, в отличие от аналитических, дают *не общие, а частные решения*, которые определяются не в континуальных (Ω), а в дискретных областях изменения независимых переменных (Ω_h). При этом требуется выполнить достаточное количество арифметических и логических действий над числовыми и логическими массивами. В силу приближенного характера вычислений этот процесс в свою очередь связан с некоторыми основными требованиями или понятиями, относящимися к конкретным задачам и численным методам (схемам), — *устойчивостью*, зависящей от *хорошей обусловленности* задачи; *сходимостью*, *высокой точностью*, *экономичностью*, и параметрами методов — шагами дискретизации h_Ω или разбиения исходной области Ω , в которой решается задача, количеством итераций (для итерационных методов), соотношениями шагов для неравномерного разбиения и др.

Некоторые из перечисленных здесь требований являются противоречивыми, поэтому при выполнении исследований чем-то приходится жертвовать, например точностью или экономичностью метода. Часть из указанных понятий (сходимость, устойчивость, хорошая обусловленность) рассматривается и наполняется конкретным содержанием при рассмотрении задач, которые решаются в основных разделах книги, поэтому дадим только их краткие определения.

Численный метод называется *сходящимся*, если при стремлении параметров метода к определенным предельным значениям (например, шагов сетки h_Ω к нулю (при $h_\Omega \rightarrow 0$)) результаты расчета стремятся к точному решению, т. е. $\hat{y}|_{h_\Omega \rightarrow 0} \rightarrow y|_\Omega$ (\hat{y} и y — приближенное и точное решения соответственно).

Задача является *хорошо обусловленной*, если при небольших изменениях входных данных результаты ее решения изменяются незначительно (непрерывная зависимость решения от исходных данных) и при любых исходных данных из возможного диапазона их изменения задача однозначно разрешима.

Рассмотрим пример плохо обусловленной задачи.

Пример В.1. Пусть задана система двух линейных алгебраических уравнений с двумя неизвестными:

$$\begin{aligned} 300x_1 + 400x_2 &= 700, \\ 100x_1 + 133x_2 &= 233. \end{aligned} \tag{B.3}$$

Выяснить, является ли эта задача хорошо обусловленной.

□ Система (B.3) имеет точное решение $x_* = (x_{*1}, x_{*2})^T = (1; 1)^T$. Пусть одно из исходных данных — число $b = 233$ изменилось на доли процента, и вместо

него приняли число $\tilde{b} = 232$. Тогда получается решение $\tilde{x}_{*1} = -3$; $\tilde{x}_{*2} = 4$. Таким образом, при изменении b на $0,43\% \left(\frac{\Delta b}{233} \cdot 100\% = \frac{1}{233} \cdot 100\% = 0,43\% \right)$ компоненты решения x_{*1}, x_{*2} изменились по модулю соответственно в 3 и 4 раза. В соответствии с определением обусловленности задача (B.3) является плохо обусловленной, так как при небольшом изменении входных данных результаты ее решения изменились значительно. ■

Численный метод называется *устойчивым*, если результаты расчета непрерывно зависят от входных (исходных) данных задачи (т. е. выполняется условие хорошей обусловленности задачи) и погрешность округления, связанная с реализацией численного метода, при заданных пределах изменения параметров численного метода остается ограниченной.

Решения, получаемые численными методами, в силу их приближенности содержат некоторые погрешности. Рассмотрим их источники и типы.

В.2. ПОГРЕШНОСТИ ВЫЧИСЛЕНИЙ

Один из типов погрешностей обусловлен *неадекватностью* выбранной математической модели исходной физической. Эта неадекватность в большей или меньшей степени присуща всем приближенно решаемым задачам. Данная погрешность является неустранимой и определяется апостериорным путем (на шестом этапе решения задачи (см. рис. В.1)). Остальные три типа погрешностей являются сугубо вычислительными и обусловлены следующими причинами.

Неточность (неопределенность) задания исходных данных приводит также к *неустранимым погрешностям*, связанным с точностью измерений или вычислений или округлением данных.

Если мы устраним неопределенность в исходных данных, например путем их фиксирования, и найдем решение с помощью какого-либо численного метода, то получим результат, не в точности соответствующий исходным данным. Это есть *погрешность численного* или *какого-либо другого* приближенного *метода* (например, приближенно-аналитического); именно такие погрешности будут оцениваться при рассмотрении численных методов. Эти оценки могут получаться до выполнения вычислений (априорные оценки) и после них (апостериорные оценки).

В компьютере все числа представляются в конечном виде, и поэтому при использовании вычислительного алгоритма реализуются ошибки арифметических и других операций над числами, а также *ошибки округления*.

Дадим некоторые понятия из теории погрешностей вычислительных действий над приближенными величинами.

Пусть x — точное, но, как правило, неизвестное значение величины, а \hat{x} — ее известное приближенное значение.

Абсолютной погрешностью приближения \hat{x} называется разность $\Delta\hat{x} = |x - \hat{x}|$ (в общем случае $\Delta\hat{x}$ имеет размерность величины x).

Относительная погрешность приближения \hat{x} обозначается δ и выражается отношением $\delta = \frac{\Delta\hat{x}}{|\hat{x}|}$ (δ — безразмерная величина, $\hat{x} \neq 0$). Часто величина δ вычисляется в процентах, и тогда она умножается на 100.

Так как величина x , как правило, неизвестна, а погрешность необходимо определять, то в рассмотрение вводится *предельная абсолютная погрешность* $\Delta(\hat{x})$:

$$\Delta\hat{x} = |x - \hat{x}| \leq \Delta(\hat{x}).$$

Раскрывая в этом неравенстве модуль, получаем соотношение, задающее отрезок, которому принадлежит точное значение:

$$\hat{x} - \Delta(\hat{x}) \leq x \leq \hat{x} + \Delta(\hat{x}).$$

Таким образом, величина x находится в Δ -окрестности (дельта-окрестности), определяемой величинами \hat{x} и $\Delta(\hat{x})$ и составляющей отрезок $[a, b]$ (рис. В.2).

Предельная относительная погрешность приближения \hat{x} определяется отношением

$$\delta(\hat{x}) = \frac{\Delta(\hat{x})}{|\hat{x}|}.$$

Отсюда получается часто используемое соотношение

$$\Delta(\hat{x}) = \delta(\hat{x}) \cdot |\hat{x}|. \quad (\text{B.4})$$

В дальнейшем рассматриваются только предельные погрешности, и поэтому в тексте при указании на абсолютную и относительную погрешности для сокращения слова «предельная» будет опускаться.

Значащими цифрами приближенного числа называют все цифры в его записи, начиная с первой ненулевой слева.

Первые n значащих цифр приближенного числа называются *верными*, если абсолютная погрешность этого числа не превышает половины единицы разряда, соответствующего n -й значащей цифре, считая слева направо. Излишние сохраненные цифры, помимо верных, называются *сомнительными*.

Вычислить приближенное число с точностью $\varepsilon = 10^{-n}$ означает, что необходимо сохранить верной значащую цифру, стоящую в n -м разряде после запятой.

На практике возникает надобность в *округлении* приближенного числа, т. е. замене его числом с меньшим количеством значащих цифр.

Для округления числа до n значащих цифр следует отбросить все его цифры, стоящие справа от n -й значащей цифры. При этом:

а) если первая из отброшенных цифр меньше 5, то оставшиеся десятичные знаки сохраняются без изменения;

б) если первая из отброшенных цифр больше 5 либо равна 5 и среди остальных отброшенных цифр есть ненулевые, то к последней оставшейся цифре прибавляется единица;

в) если первая из отброшенных цифр равна 5 и остальные отброшенные цифры нулевые, то последняя оставшаяся цифра не изменяется, если она четная, и увеличивается на единицу, если она нечетная.

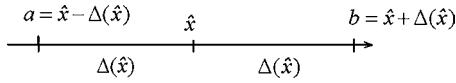


Рис. В.2

Пример В.2. Найти абсолютные и относительные погрешности числа $\pi = 3,141592654\dots$, заданного двумя и тремя цифрами после запятой.

□1. Пусть $\hat{x} = 3,14$. Тогда справедливо

$$|\pi - 3,14| \leq 0,002 = \Delta(3,14); \quad \delta(3,14) = (0,002/3,14) \cdot 100\% \approx 0,064\%.$$

2. Если $\hat{x} = 3,141$, то

$$|\pi - 3,141| \leq 0,0006 = \Delta(3,141); \quad \delta(3,141) = (0,0006/3,141) \cdot 100\% \approx 0,02\%. \blacksquare$$

Абсолютная и относительная погрешности записываются в виде чисел с одной или двумя значащими цифрами и округляются с избытком. В записи приближенных чисел они указываются так:

$$x = \hat{x} \pm \Delta; \quad x = \hat{x}(1 \pm \delta). \quad (\text{B.5})$$

Например:

$$\pi = 3,141 \pm 0,0006; \quad \pi = 3,141(1 \pm 0,02\%).$$

Если Δ в записи числа не указано, то подразумевается, что x имеет точность половины единицы ($1/2$ ед.) младшего разряда. Так, для $\hat{x} = 5,63$ абсолютная погрешность $\Delta = 0,005$.

Пример В.3. Округлить число $x = 3,928525750$ до семи, шести, пяти и т. д. десятичных знаков и до единиц.

□Пользуясь правилом округления (см. п. «а»–«в»), получаем $x = 3,9285258$ (согласно п. «в», так как первая из отброшенных цифр равна 5, остальные отброшенные цифры нулевые, последняя оставшаяся цифра нечетная); $x = 3,928526$ (согласно п. «б», так как первая из отброшенных цифр больше 5); $x = 3,92853$ (согласно п. «б», так как первая из отброшенных цифр равна 5 и среди отброшенных цифр есть ненулевые); $x = 3,9285$ (см. п. «а»); $x = 3,929$ (см. п. «б»); $x = 3,93$ (см. п. «б»); $x = 3,9$ (см. п. «а»); $x = 4$ (см. п. «б»). ■

Пример В.4. Выделить значащие цифры следующих чисел:

$$2,396037; \quad 0,00167; \quad 3250000; \quad 0,00005.$$

□Поскольку значащими являются все цифры в записи числа, начиная с первой ненулевой слева, то получаем:

$$\underline{2,396037}; \quad \underline{0,00167}; \quad \underline{3250000}; \quad \underline{0,00005}. \blacksquare$$

Пример В.5. Выделить верные значащие цифры чисел:

- а) $x = 0,004507; \quad \Delta = 0,00001;$
- б) $x = 0,004507; \quad \Delta = 0,00006;$
- в) $x = 0,0208700; \quad \Delta = 0,0000005;$
- г) $x = 12,396; \quad \Delta = 0,03;$
- д) $x = 0,037862; \quad \Delta = 0,007;$
- е) $x = 9,999785; \quad \Delta = 0,0004.$

□В записи рассматриваемых чисел подчеркнем верные значащие цифры:

а) $0,00\underline{4}507 \pm 0,00001$, так как $\Delta = 0,00001 < 0,5 \cdot 10^{-4}$;

б) $0,00\underline{4}507 \pm 0,00006$, так как $0,5 \cdot 10^{-4} < \Delta = 0,00006 < 0,5 \cdot 10^{-3}$;

в) $0,02\underline{0}8700 \pm 0,0000005$, поскольку $\Delta = 0,0000005 = 0,5 \cdot 10^{-6}$;

г) $12,396 \pm 0,03$, так как $\Delta = 0,03 < 0,5 \cdot 10^{-1}$;

д) $0,037862 \pm 0,007$, поскольку $0,5 \cdot 10^{-2} < \Delta = 0,007 < 0,5 \cdot 10^{-1}$ (все цифры сомнительные);

е) $9,999785 \pm 0,0004$, так как $\Delta = 0,0004 < 0,5 \cdot 10^{-3}$. ■

Погрешности округления в ЭВМ числа \hat{x} , обусловленные конечностью разрядной сетки, для различных компьютеров могут быть вычислены по формуле

$$\delta(\hat{x}) = \frac{1}{\alpha_m} \left(\frac{1}{s} \right)^{n-1}, \quad (\text{B.6})$$

где α_m — первая значащая (отличная от нуля) цифра; s — основание системы счисления, использующееся в компьютере; n — разрядность компьютера ($n = 32$ при стандартной точности и $n = 64$ при двойной).

Рассмотрим правило вычисления погрешностей арифметических операций и функций по погрешности аргументов (без учета ошибок округления).

При вычислении абсолютных погрешностей обычно используются формулы дифференцирования, в которых дифференциалы независимых переменных заменяются абсолютными погрешностями ($dx_i = \Delta \hat{x}_i$).

Пусть $x_1 > 0$, $x_2 > 0$ и заданы предельные абсолютные погрешности $\Delta(\hat{x}_1)$, $\Delta(\hat{x}_2)$, т. е. $x_1 = \hat{x}_1 \pm \Delta(\hat{x}_1)$, $x_2 = \hat{x}_2 \pm \Delta(\hat{x}_2)$. Требуется найти погрешность суммы $x = x_1 + x_2$.

Так как $dx = dx_1 + dx_2$, то $\Delta x = \Delta \hat{x}_1 + \Delta \hat{x}_2 \leq \Delta(\hat{x}_1) + \Delta(\hat{x}_2)$. Отсюда следует, что

$$\Delta(\hat{x}_1 + \hat{x}_2) = \Delta(\hat{x}_1) + \Delta(\hat{x}_2). \quad (\text{B.7})$$

Теперь предположим, что $x_1 > x_2$, и найдем погрешность разности $x = x_1 - x_2$. Тогда $dx = dx_1 - dx_2$ и $\Delta x = \Delta \hat{x}_1 - \Delta \hat{x}_2 \leq \Delta(\hat{x}_1) + \Delta(\hat{x}_2)$, поэтому

$$\Delta(\hat{x}_1 - \hat{x}_2) = \Delta(\hat{x}_1) + \Delta(\hat{x}_2). \quad (\text{B.8})$$

Таким образом, *абсолютная погрешность суммы и разности двух приближенных чисел равна сумме абсолютных погрешностей слагаемых*.

Это правило справедливо для произвольного числа слагаемых. Так, если \hat{x}_1 , \hat{x}_2 , ..., \hat{x}_n имеют одну и ту же погрешность $\Delta(\hat{x})$, то $\Delta(\hat{x}_1 + \hat{x}_2 + \dots + \hat{x}_n) = n \cdot \Delta(\hat{x})$. Но реально погрешности могут иметь разные знаки и потому взаимно компенсировать друг друга. По правилу Чеботарева при $n > 10$ можно принять

$$\Delta(\hat{x}_1 + \hat{x}_2 + \dots + \hat{x}_n) = \sqrt{3n} \cdot \Delta(\hat{x}).$$

Для относительной погрешности суммы и разности двух чисел $x_1 \pm x_2$ получаем

$$\begin{aligned} \delta(\hat{x}_1 \pm \hat{x}_2) &= \frac{\Delta(\hat{x}_1 \pm \hat{x}_2)}{(\hat{x}_1 \pm \hat{x}_2)} = \frac{\Delta(\hat{x}_1) + \Delta(\hat{x}_2)}{(\hat{x}_1 \pm \hat{x}_2)} = \frac{\Delta(\hat{x}_1) \cdot \hat{x}_1}{(\hat{x}_1 \pm \hat{x}_2) \cdot \hat{x}_1} + \frac{\Delta(\hat{x}_2) \cdot \hat{x}_2}{(\hat{x}_1 \pm \hat{x}_2) \cdot \hat{x}_2} = \\ &= \frac{\hat{x}_1}{(\hat{x}_1 \pm \hat{x}_2)} \delta(\hat{x}_1) + \frac{\hat{x}_2}{(\hat{x}_1 \pm \hat{x}_2)} \delta(\hat{x}_2). \end{aligned}$$

Для произвольного числа слагаемых

$$\delta(\hat{x}) = \frac{\hat{x}_1}{\hat{x}} \delta(\hat{x}_1) + \frac{\hat{x}_2}{\hat{x}} \delta(\hat{x}_2) + \dots + \frac{\hat{x}_n}{\hat{x}} \delta(\hat{x}_n), \quad (\text{B.9})$$

где $\hat{x} = \hat{x}_1 + \hat{x}_2 + \dots + \hat{x}_n$; $\hat{x}_i > 0$, $i = 1, \dots, n$.

Пусть $\max_i \delta(\hat{x}_i) = M$, $\min_i \delta(\hat{x}_i) = m$, тогда

$$\delta(\hat{x}) \leq \frac{\hat{x}_1 \cdot M + \dots + \hat{x}_n \cdot M}{\hat{x}} = M; \quad \delta(\hat{x}) \geq \frac{\hat{x}_1 \cdot m + \dots + \hat{x}_n \cdot m}{\hat{x}} = m.$$

Поэтому

$$m \leq \delta(\hat{x}) \leq M. \quad (\text{B.10})$$

Аналогично при выполнении умножения и деления получаем погрешности (при тех же предположениях):

$$\Delta(\hat{x}_1 \cdot \hat{x}_2) = \Delta\hat{x}_1 \cdot \hat{x}_2 + \Delta\hat{x}_2 \cdot \hat{x}_1 \leq \hat{x}_1 \cdot \Delta(\hat{x}_2) + \hat{x}_2 \cdot \Delta(\hat{x}_1); \quad (\text{B.11})$$

$$\delta(\hat{x}_1 \cdot \hat{x}_2) = \frac{\Delta(\hat{x}_1 \cdot \hat{x}_2)}{\hat{x}_1 \cdot \hat{x}_2} = \delta(\hat{x}_1) + \delta(\hat{x}_2); \quad (\text{B.12})$$

$$\Delta\left(\frac{\hat{x}_1}{\hat{x}_2}\right) = \frac{\Delta\hat{x}_1 \cdot \hat{x}_2 - \Delta\hat{x}_2 \cdot \hat{x}_1}{\hat{x}_2^2} \leq \frac{\hat{x}_1 \cdot \Delta(\hat{x}_2) + \hat{x}_2 \cdot \Delta(\hat{x}_1)}{\hat{x}_2^2}; \quad (\text{B.13})$$

$$\delta\left(\frac{\hat{x}_1}{\hat{x}_2}\right) = \delta(\hat{x}_1) + \delta(\hat{x}_2). \quad (\text{B.14})$$

Из последних соотношений следует, что *относительная погрешность произведения двух чисел равна сумме относительных погрешностей его сомножителей*. Аналогичное правило выполняется и для частного от деления двух чисел.

Замечание. При вычитании близких чисел на ЭВМ происходит потеря верных значащих цифр, поскольку разность $\hat{x}_1 - \hat{x}_2$ близка к нулю и, следовательно, относительная погрешность разности значительно больше относительной погрешности уменьшаемого и вычитаемого.

Пример B.6. Требуется вычислить разность $y = x_1 - x_2$ двух близких чисел $x_1 = 0,01\underline{234}$ и $x_2 = 0,01\underline{231}$, для которых $\Delta(\hat{x}) = 0,000002$ (подчеркнуты верные значащие цифры).

□ Величина $y = x_1 - x_2 = 0,00003$ имеет только одну значащую цифру, так как $\Delta(\hat{y}) = 0,000004 < 0,5 \cdot 10^{-5}$ ($\Delta(\hat{y}) = \Delta(\hat{x}_1) + \Delta(\hat{x}_2)$) (абсолютная погрешность результата удвоилась). Таким образом, произошла потеря трех верных значащих цифр. Это следует учитывать при составлении алгоритма, по возможности избегая потери точности с помощью алгебраических преобразований формулы и изменения в ней последовательности вычислений. ■

Аналогично получаются формулы для погрешностей арифметических действий при вычислении функций многих переменных. Так, для $z = f(x_1, x_2, \dots, x_n)$ имеем следующие формулы для погрешностей:

$$\begin{aligned} \Delta(\hat{z}) &= \left| \frac{\partial f}{\partial x_1} \right| \cdot \Delta(\hat{x}_1) + \left| \frac{\partial f}{\partial x_2} \right| \cdot \Delta(\hat{x}_2) + \dots + \left| \frac{\partial f}{\partial x_n} \right| \cdot \Delta(\hat{x}_n); \\ \delta(\hat{z}) &= \left| \frac{\hat{x}_1}{f} \cdot \frac{\partial f}{\partial x_1} \right| \cdot \delta(\hat{x}_1) + \left| \frac{\hat{x}_2}{f} \cdot \frac{\partial f}{\partial x_2} \right| \cdot \delta(\hat{x}_2) + \dots + \left| \frac{\hat{x}_n}{f} \cdot \frac{\partial f}{\partial x_n} \right| \cdot \delta(\hat{x}_n) = \\ &= \left| \hat{x}_1 \frac{\partial \ln f}{\partial x_1} \right| \cdot \delta(\hat{x}_1) + \left| \hat{x}_2 \frac{\partial \ln f}{\partial x_2} \right| \cdot \delta(\hat{x}_2) + \dots + \left| \hat{x}_n \frac{\partial \ln f}{\partial x_n} \right| \cdot \delta(\hat{x}_n). \end{aligned} \quad (\text{B.15})$$

Пример В.7. Стороны треугольника $\hat{a} = 17,3$ см, $\hat{b} = 23,6$ см, $\hat{c} = 14,2$ см замерены с абсолютной погрешностью $\Delta(\hat{a}) = \Delta(\hat{b}) = \Delta(\hat{c}) = 0,1$ см. Определить абсолютную погрешность периметра $p = a + b + c$ этого треугольника.

□ Получаем $\hat{p} = 17,3 + 23,6 + 14,2 = 55,1$ см; $\Delta(\hat{p}) = 3 \cdot 0,1 = 0,3$ см. Таким образом, $\hat{p} = 55,1 \pm 0,3$ см (последняя цифра в числе $\hat{p} = 55,1$ сомнительная, так как $\Delta = 0,3 < 0,5 \cdot 10^0$). ■

Пример В.8. Найти относительные погрешности измерения сторон треугольника в примере В.7 и пределы относительной погрешности для периметра.

□ Вычислим относительные погрешности:

$$\delta(\hat{a}) = \frac{0,1}{17,3} \cdot 100\% = 0,58\%;$$

$$\delta(\hat{b}) = \frac{0,1}{23,6} \cdot 100\% = 0,42\% = m;$$

$$\delta(\hat{c}) = \frac{0,1}{14,2} \cdot 100\% = 0,70\% = M.$$

Пределы изменения относительной погрешности найдем по формуле (В.10):

$$0,42\% \leq \delta(\hat{p}) \leq 0,70\%. ■$$

Пример В.9. Ребра прямоугольного параллелепипеда $\hat{a} = 4,3$ см, $\hat{b} = 1,6$ см, $\hat{c} = 2,8$ см замерены с абсолютной погрешностью $\Delta(\hat{a}) = \Delta(\hat{b}) = \Delta(\hat{c}) = 0,1$ см.

Определить абсолютную и относительную погрешности его объема $V = abc$.

□ Воспользуемся обобщением соотношения (В.12):

$$\hat{V} = \hat{a} \cdot \hat{b} \cdot \hat{c} = 19,264; \quad \delta(\hat{V}) = \delta(19,264) = \frac{0,1}{4,3} + \frac{0,1}{1,6} + \frac{0,1}{2,8} = 0,121.$$

По формуле (В.4) имеем $\Delta(\hat{V}) = \hat{V} \cdot \delta(\hat{V}) = 19,264 \times 0,121 \approx 2,330$. Так как в $\Delta(\hat{V}) = 2,330 < 0,5 \cdot 10^1$, то в $\hat{V} = 19,264$ верна лишь цифра десятков, а остальные цифры сомнительные. Округляя \hat{V} , оставим две сомнительные цифры: $\hat{V} \approx 19,3$ см³. ■

Замечания

1. В арифметике целых чисел на ЭВМ операции сложения, вычитания и умножения выполняются точно. Деление целых чисел осуществляется с округлением — заменой частного его целой частью. Если в вычислениях присутствует большое число операций деления, то в целочисленной арифметике это может привести к большой погрешности из-за ошибок округления.

2. В арифметике вещественных чисел на ЭВМ все четыре действия, как правило, выполняются с округлением. Исключение составляют действия над числами, которые не выходят за разрядность ЭВМ.

3. Оценка погрешностей округлений при решении многих задач на ЭВМ может быть найдена апостериорным путем с помощью двух повторных расчетов в арифметике одинарной и двойной точности и составления разности полученных результатов. Этот прием используется также и для оценки погрешностей численного метода (модели).

В.3. НЕКОТОРЫЕ ПОНЯТИЯ МАТЕМАТИЧЕСКОГО АНАЛИЗА, ИСПОЛЬЗУЮЩИЕСЯ В ЧИСЛЕННЫХ МЕТОДАХ

Для получения оценок погрешностей аппроксимационных формул в книге используются понятия « O большое от h^k » (h — шаг сетки), класс функций $C_k[a, b]$, разложение функции по формуле Тейлора, теорема о среднем значении. Рассмотрим эти понятия.

1. Пусть $R(h)$ — некоторая функция переменной h (как правило, $R(h)$ — остаточное слагаемое некоторой аппроксимационной формулы) с конечной областью определения D_R на полуоси $h > 0$, причем $h \in D_R$.

Тогда, если при некотором $h \leq h_0$ справедливо неравенство $|R(h)| \leq ch^k$, где $c = \text{const}$, не зависящая от h , k — целое число, $h_0 > 0$, то пишут $R(h) = O(h^k)$ и говорят, что $R(h)$ есть « O большое от h^k » при $h \rightarrow 0$.

В этом случае значение k соответствует порядку аппроксимации аппроксимационной формулы (оператора).

Согласно приведенному определению справедливы свойства:

а) если $R_1(h) = O(h^k)$, $R_2(h) = O(h^k)$, причем $D_{R_1} = D_{R_2}$, то $R_1(h) \pm R_2(h) = O(h^k)$, поскольку справедливо $|R_1(h) \pm R_2(h)| \leq |R_1(h)| + |R_2(h)|$, т. е. $O(h^k) \pm O(h^k) = O(h^k)$;

б) $\frac{R_1(h)}{h^p} = O(h^{k-p})$; $R_1(h) \cdot h^p = O(h^{k+p})$;

в) если α — произвольная константа, не зависящая от h , то $\alpha R_1(h) = O(h^k)$;

г) если $R_1(h) = O(h^{k_1})$, $R_2(h) = O(h^{k_2})$, то произведение $R_1(h) \cdot R_2(h) = O(h^{k_1+k_2})$, а сумма $R_1(h) + R_2(h) = O(h^{k_3})$, где $k_3 = \min(k_1, k_2)$.

2. Пусть функция $f(x)$ определена на отрезке $[a, b] \equiv G$ и имеет на нем непрерывные производные до порядка k включительно. Тогда считается, что $f(x)$ принадлежит классу $C_k[G]$, и для обозначения этого используется запись $f \in C_k[G]$. При $k = 0$ вместо $C_0[G]$ будет использоваться обозначение $C[G]$ (функция непрерывна на отрезке $[a, b]$).

Запись $f \in C_k[G]$ означает, что на некотором интервале $\tilde{G} \equiv (A, B)$, который включает в себя $[a, b]$, существует k раз непрерывно дифференцируемая функция, такая, что $\tilde{f}(x) \equiv f(x)$ на $G = [a, b]$. Тогда значения производных от $f(x)$ на концах $[a, b]$ будут соответствовать производным функции $\tilde{f}(x)$ в этих точках (данное обобщение приведено в связи с тем, что в математическом анализе производные определяются во внутренних точках области определения функции, от которой вычисляются $f^{(k)}(x)$).

3. Пусть задана функция $f(x) \in C_{k+1}[a, b]$. Тогда в точке $x_0 \in [a, b]$ справедлива формула Тейлора:

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + \dots + \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k + \\ + \frac{f^{(k+1)}(\xi)}{(k+1)!}(x - x_0)^{k+1} = \sum_{j=0}^k \frac{f^{(j)}(x_0)}{j!}(x - x_0)^j + R_{k+1}(x), \quad (\text{B.16})$$

где $x \in [a, b]$, $x \neq x_0$; ξ — некоторая точка, лежащая строго между точками x и x_0 ; $R_{k+1}(x) = \frac{f^{(k+1)}(\xi)}{(k+1)!}(x - x_0)^{k+1}$ — остаточный член в форме Лагранжа. Индексы, указанные в круглых скобках вверху, здесь и далее означают порядки производных.

Так как по предположению производная $f^{(k+1)}(x)$ непрерывна на отрезке $[a, b]$, то она ограничена на этом отрезке, т. е. величина $M_{k+1} = \max_{[a,b]} |f^{(k+1)}(x)| < +\infty$. Поэтому справедлива оценка остаточного члена

$$|R_{k+1}(x)| \leq \frac{M_{k+1}}{(k+1)!} |x - x_0|^{k+1}.$$

Это неравенство свидетельствует о том, что

$$|R_{k+1}(x)| = O(|x - x_0|^{k+1}) = O(h^{k+1}),$$

где $h = |x - x_0|$.

Тогда формула (B.16) может быть записана в виде

$$f(x) = \sum_{j=0}^k \frac{f^{(j)}(x_0)}{j!} (x - x_0)^j + O(h^{k+1}). \quad (\text{B.17})$$

При $k = 1$ формулу Тейлора можно записать в виде

$$\begin{aligned} f(x_0 + h) &= f(x_0) + f'(x_0)h + O(h^2); \\ f(x_0 - h) &= f(x_0) - f'(x_0)h + O(h^2), \end{aligned}$$

а при $k = 2$ — в форме

$$\begin{aligned} f(x_0 + h) &= f(x_0) + f'(x_0)h + \frac{h^2}{2} f''(x_0) + O(h^3); \\ f(x_0 - h) &= f(x_0) - f'(x_0)h + \frac{h^2}{2} f''(x_0) + O(h^3). \end{aligned} \quad (\text{B.18})$$

Рассмотрим частный случай формулы (B.17). Положим $x = x_{i+1}$, $x_0 = x_i$, $h_{i+1} = x_{i+1} - x_i$, где значения x_i , x_{i+1} задают отрезок $[x_i, x_{i+1}] \subset [a, b]$. Тогда получаем

$$f(x_{i+1}) = f(x_i) + f'(x_i)h_{i+1} + \frac{f''(x_i)}{2!} h_{i+1}^2 + \dots + \frac{f^{(k)}(x_i)}{k!} h_{i+1}^k + O(h_{i+1}^{k+1}). \quad (\text{B.19})$$

Обозначая $f_{i+1} = f(x_{i+1})$, $f_i = f(x_i)$, $f_i^{(p)} = f^{(p)}(x_i)$, $p = 1, 2, \dots$, и учитывая, что остаточный член в форме Лагранжа имеет вид $\frac{f^{(k)}(\xi)}{(k+1)!} h_{i+1}^{k+1}$, $\xi \in (x_i, x_{i+1})$, формулу (B.19) можно записать в форме

$$f_{i+1} = f_i + f'_i h_{i+1} + \frac{f''_i}{2!} h_{i+1}^2 + \dots + \frac{f_i^{(k)}}{k!} h_{i+1}^k + \frac{h_{i+1}^{k+1}}{(k+1)!} f^{(k+1)}(\xi), \quad \xi \in (x_i, x_{i+1}). \quad (\text{B.20})$$

4. При изложении будет использовано следующее утверждение [6].

Утверждение B.1. Пусть $f(x) \in C[a, b]$, $\xi_i \in [a, b]$, ($i = \overline{1, n}$) — произвольные точки, принадлежащие отрезку $[a, b]$. Тогда существует такая точка $\xi \in [a, b]$, что

$$\frac{f(\xi_1) + \dots + f(\xi_n)}{n} = f(\xi).$$

В.4. ДИСКРЕТИЗАЦИЯ И ПРИНЦИП СООТВЕТСТВИЯ ПОРЯДКОВ АППРОКСИМАЦИИ ДИСКРЕТНЫХ МОДЕЛЕЙ ДЛЯ ИНТЕГРАЛОВ, ФУНКЦИЙ И ПРОИЗВОДНЫХ

Предметом данной книги является изложение наиболее широко распространенных математических моделей, увязывающих функции, интегралы, производные, а также их применение в теории решения обыкновенных дифференциальных уравнений и уравнений в частных производных. Эта теория тесно связана с теорией приближения или с заменой непрерывных математических моделей их дискретными аналогами (дискретизацией задач). Например, для решения задачи Коши

$$\frac{dy}{dx} = f(x, y); \quad y(x_0) = y_0, \quad x \in [a, b] \equiv \Omega \quad (\text{B.21})$$

дифференциальная модель (B.21) (обозначим ее $M(Ly, y_0 \rightarrow y)$, где в скобках слева от стрелки указаны известные данные — дифференциальный оператор Ly и начальное значение y_0 , а справа от стрелки — искомая функция) заменяется дискретной моделью (ДМ) $(L_h y, y_0 \rightarrow \hat{y}_i)$, в которой вместо непрерывной области Ω изменения x рассматривается сетка $\Omega_n = \{x_0, x_1, \dots, x_n\}$ со значениями $x_{i+1} = x_i + h_{i+1}$, $i = \overline{0, n-1}$, $h_{i+1} = x_{i+1} - x_i$ и с разностным оператором $L_h y$, определенным в x_i (узлах сетки).

Данный процесс замены модели $M(Ly, y_0 \rightarrow y)$ моделью $D\mathcal{M}(L_h y, y_0 \rightarrow \hat{y}_i)$, соответствующей дискретной задаче, называется *процессом дискретизации* модели. При этом в книге используются сетки двух типов — *равномерные* (или регулярные), если $h_{i+1} = \text{const}$ при всех i , и *неравномерные* (или нерегулярные) ($h_{i+1} = \text{var}$) — в противном случае.

После получения решения дискретной задачи найденные значения \hat{y}_i могут обрабатываться с помощью других дискретных моделей типа дифференцирования $D\mathcal{M}_1(\hat{y}_i \rightarrow \hat{y}'_i)$ или интегрирования $D\mathcal{M}_2(\hat{y}_i \rightarrow \hat{I}_i^{i+1})$. Эти дискретные модели могут использоваться в составе или независимо от каких-либо более общих моделей дифференциального типа, на основе которых получаются решения \hat{y}_i . Подчеркнем, что указанные совокупности операций дифференцирования и интегрирования выполняются не с помощью классического математического анализа, приспособленного только к непрерывным функциям, а на основе аппарата теории приближений сеточных функций, т. е. на дискретных моделях. Класс этих дискретных моделей (по сравнению с традиционными) расширен за счет использования теории интегрально-дифференциальных сплайнов (ИД-сплайнов) в [15–24].

Действительно, здесь реализованы модели:

- численного дифференцирования двух типов: $D\mathcal{M}_1(\hat{f}_i, \hat{I}_i^{i+1} \rightarrow \hat{f}_i^{(p)})$ (\hat{f}_i — исходная сеточная функция, p — порядок производных, $\hat{f}_i^{(p)}$ — либо значение производной, либо ее аппроксимационная формула), $D\mathcal{M}_1(\hat{I}_i^{i+1} \rightarrow \hat{f}_i^{(p)})$;
- численного интегрирования двух типов: $D\mathcal{M}_2(\hat{f}_i, \hat{f}_i^{(p)} \rightarrow \hat{I}_i^{i+1})$ ($p = 1, 2, 3$); $D\mathcal{M}_2(\hat{f}_i \rightarrow \hat{I}_i^{i+1})$;
- восстановления функций по интегралам и по совокупности интегралов и производных: $D\mathcal{M}_3(\hat{I}_i^{i+1} \rightarrow \hat{f}_i)$, $D\mathcal{M}_3(\hat{I}_i^{i+1}, \hat{f}_i^{(p)} \rightarrow \hat{f}_i)$ и др.

Здесь и далее символом I_i^{i+1} обозначается определенный интеграл $I_i^{i+1} = \int_{x_1}^{x_{i+1}} f(x) dx$ ($y = f(x)$), а символом \hat{I}_i^{i+1} — его приближенное значение, вычисленное, например, по одной из квадратурных формул (см. главу 5).

Заметим, что часть из указанных моделей может быть использована при решении какой-либо другой более общей задачи. В этом случае все входящие в задачу величины определенным образом увязываются между собой в единых формулах. Кроме того, увязка между функциями, производными и интегралами существует в вычислительной практике, например при формулировке задач обработки данных, аппроксимации результатов экспериментов, анализе силовых и тепловых характеристик летательных аппаратов и других технических систем.

При решении таких задач необходимо уметь сохранять требуемый порядок аппроксимации непрерывной модели ее дискретным представлением (о порядке аппроксимации см. п. В.3). Для этого следует увязывать по порядку аппроксимации все математические объекты, входящие в одну модель, или разные модели, соединяемые в одну. Данный порядок определяется величинами $P_{0,0}, P_{0,1}, P_{0,2}, \dots$ — порядками аппроксимации (точности) исходных данных, указанных слева от стрелки в приведенных выше обозначениях математических моделей, и конкретной связью известных и искомых величин, входящих в эту модель. Для получения P_1 — априорно заданной точности результата — необходимо, чтобы $P_{0,0}, P_{0,1}, P_{0,2}$ по величине не были меньше P_1 . Без учета этого соответствия требуемый порядок результата может не достигаться, что приводит к понижению его точности.

Проблема увязки порядков $P_{0,0}, P_{0,1}, P_{0,2}, \dots$ и P_1 до настоящего времени решалась на субъективном уровне, так как отсутствовал четкий принцип соответствия между порядками аппроксимации входящих в различные дискретные модели значений функций, производных, интегралов и других математических объектов.

Данный принцип следует из представления значения $F(x_{i+1})$ первообразной функции $F(x)$ в точке x_{i+1} на некотором опорном отрезке $[x_i, x_{i+1}]$ относительно точки x_i .

Действительно, применим формулу Тейлора (В.20) для функции $F(x) \in C_{k+1}[a, b]$:

$$F(x_{i+1}) = F(x_i) + F'(x_i)h_{i+1} + \frac{F''(x_i)}{2!}h_{i+1}^2 + \dots + \frac{F^{(k)}(x_i)}{k!}h_{i+1}^k + O(h_{i+1}^{k+1}).$$

Так как $F(x)$ первообразная, то справедливы равенства $F'(x_i) = f(x_i)$, $F''(x_i) = f'(x_i)$, ..., $F^{(k)}(x_i) = f^{(k-1)}(x_i)$. Обозначая $F_{i+1} = F(x_{i+1})$, $F_i = F(x_i)$, $f_i = f(x_i)$, $f'_i = f'(x_i) = F''(x_i)$ и т. д., имеем

$$F_{i+1} = F_i + h_{i+1} \cdot f_i + \frac{h_{i+1}^2}{2} f'_i + \dots + \frac{h_{i+1}^k}{k!} f_i^{(k-1)} + O(h_{i+1}^{k+1}). \quad (\text{B.22})$$

Учитывая, что здесь остаточное слагаемое в форме Лагранжа имеет вид

$$\frac{F^{(k+1)}(\xi)}{(k+1)!} h_{i+1}^{k+1} = \frac{f^{(k)}(\xi)}{(k+1)!} h_{i+1}^{k+1}, \quad \xi \in (x_i, x_{i+1}),$$

формулу (B.22) можно переписать следующим образом:

$$F_{i+1} = F_i + h_{i+1} \cdot \hat{f}_i + \frac{h_{i+1}^2}{2} f'_i + \dots + \frac{h_{i+1}^k}{k!} f_i^{(k-1)} + \frac{f^{(k)}(\xi)}{(k+1)!} h_{i+1}^{k+1}. \quad (\text{B.23})$$

Из (B.22) с использованием формулы Ньютона — Лейбница

$$I_i^{i+1} = \int_{x_i}^{x_{i+1}} f(x) dx = F(x_{i+1}) - F(x_i) = F_{i+1} - F_i$$

получается следующее представление интеграла:

$$I_i^{i+1} = h_{i+1} \cdot \hat{f}_i + \frac{h_{i+1}^2}{2} f'_i + \dots + \frac{h_{i+1}^k}{k!} f_i^{(k-1)} + O(h_{i+1}^{k+1}). \quad (\text{B.24})$$

Формула (B.24) хотя и используется в дальнейшем для построения аппроксимационных соотношений для интегралов и производных (см. п. 5.2.2), здесь рассматривается только как абстрактная связь между интегралом, функцией и ее производными различных порядков.

Эта связь фактически фиксирует соответствие порядков аппроксимации объектов интегрального, функционального и дифференциального типов, составляющих некоторую, вообще говоря произвольную, модель (аппроксимацию).

Рассмотрим способ использования связи (B.24) для анализа некоторых из вышеуказанных моделей, например,

$$\text{ДМ}_2(\hat{f}_i \rightarrow \hat{I}_i^{i+1}). \quad (\text{B.25})$$

Величины, указываемые в скобках сразу за обозначением модели, будем называть ее *параметрами* (в данном случае параметрами аппроксимации). Включаемый в число параметров моделей интеграл является односторонним, т. е. он определяется на одном отрезке, однако его аппроксимация может строиться не на одном отрезке $[x_i, x_{i+1}]$, а на некотором *шаблоне* или *опорном отрезке*, который может включать большее число узлов сетки.

В модели (B.25) интегралы \hat{I}_i^{i+1} определяются по значениям функции \hat{f}_i , которые измерены или вычислены с некоторой погрешностью. Пусть эта погрешность составляет величину $O(h_{i+1}^3)$ и при вычислении интегралов на каждом из отрезков $[x_i, x_{i+1}]$ указанную точность необходимо сохранить. Иными словами, нам требуется найти порядок аппроксимации интеграла \hat{I}_i^{i+1} , который соответствовал бы порядку точности заданной функции, по которой вычисляется интеграл. С этой целью из (B.24) устанавливается связь между двумя включенными в модель (B.25) объектами \hat{I}_i^{i+1} и \hat{f}_i : $\hat{I}_i^{i+1} = h_{i+1} \hat{f}_i$. Учитывая в правой части этого равенства остаточное слагаемое для значений \hat{f}_i , получаем

$$\hat{I}_i^{i+1} = h_{i+1} \left(\hat{f}_i + O(h_{i+1}^3) \right) = h_{i+1} \hat{f}_i + O(h_{i+1}^4).$$

Таким образом, расчет интегралов \hat{I}_i^{i+1} на каждом из отрезков $[x_i, x_{i+1}]$ с точностью $O(h_{i+1}^4)$ соответствует точности $O(h_{i+1}^3)$ для \hat{f}_i (или наоборот).

Если же взять модель $\text{ДМ}_2(\hat{f}, \hat{f}'_i \rightarrow \hat{I}_i^{i+1})$, в которой интегралы определяются по \hat{f}_i и \hat{f}'_i , то для соблюдения соответствия порядков аппроксимации \hat{I}_i^{i+1} ,

\hat{f}_i , \hat{f}'_i производные должны быть рассчитаны (аппроксимированы) с точностью не ниже $O(h_{i+1}^2)$.

Для модели $\text{ДМ}_3(\hat{I}_i^{i+1} \rightarrow f_i)$, являющейся обратной по отношению к модели (В.25), вышеописанное соответствие также справедливо, т. е. при задании интегралов \hat{I}_i^{i+1} с точностью не ниже $O(h_{i+1}^4)$ модель обеспечивает вычисление функции с точностью не ниже $O(h_{i+1}^3)$.

Аналогичный анализ может быть выполнен и для других моделей, в частности для моделей, использующих другие сочетания заданных и определяемых математических объектов.

В общей форме рассмотренный *принцип соответствия порядков аппроксимации* интегралов, функций и производных можно сформулировать в виде утверждения.

Утверждение В.2. *Если некоторая аппроксимационная формула или дискретная математическая модель содержит в качестве исходных и искомых параметров сеточную функцию f_i , ее одну или несколько производных $\hat{f}_i^{(p)}$, и односторонние интегралы \hat{I}_i^{i+1} , имеющие порядки аппроксимации P_f , $P_{f^{(p)}}$, P_I соответственно, то эта модель обеспечивает соответствие указанных порядков в случае, если $P_I \geq P_f + 1$; $P_{f^{(p)}} \geq P_f + p$, где p — порядок производной.*

Замечания

1. В данном утверждении порядок P_f принят как бы в качестве базового, с которым сопоставляются порядки $P_{f^{(p)}}$ и P_I . В качестве базового порядка может быть принят также и любой другой.

2. Сформулированный принцип соответствия легко распространяется также на модели, включающие дифференциальные уравнения.

ЧАСТЬ ПЕРВАЯ

ЧИСЛЕННЫЕ МЕТОДЫ
АЛГЕБРЫ

ГЛАВА
1

МЕТОДЫ РЕШЕНИЯ СИСТЕМ ЛИНЕЙНЫХ АЛГЕБРАИЧЕСКИХ УРАВНЕНИЙ

1.1. ПОСТАНОВКА ЗАДАЧИ

Прикладные задачи, характерные для проектирования современных объектов новой техники, часто сводятся к многомерным в общем случае нелинейным уравнениям, которые решаются методом линеаризации, т. е. сведением нелинейных уравнений к линейным. В общем случае система n уравнений с n неизвестными записывается в виде:

$$\begin{aligned} f_1(x_1, x_2, \dots, x_n) &= 0, \\ f_2(x_1, x_2, \dots, x_n) &= 0, \\ \dots & \\ f_n(x_1, x_2, \dots, x_n) &= 0, \end{aligned} \tag{1.1}$$

где f_1, f_2, \dots, f_n — функции n переменных, нелинейные или линейные (x_i в функции f_j входят в первых или частично в нулевых степенях).

Здесь рассматривается частный случай задачи (1.1) — линейная однородная задача для систем линейных алгебраических уравнений (СЛАУ), которая сокращенно записывается в виде

$$Ax = b \text{ или } \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}, \tag{1.2}$$

где $A = (a_{ij}) \in R^{n \times n}$ — действительная матрица размеров $(n \times n)$, i, j — переменные, соответствующие номерам строк и столбцов (целые числа); $b = (b_1, \dots, b_n)^T \in R^n$ — вектор-столбец размеров $(n \times 1)$; $x = (x_1, \dots, x_n)^T \in R^n$ — вектор-столбец неизвестных; R^n — n -мерное евклидово пространство, верхний индекс « T » здесь и далее обозначает операцию транспонирования.

Требуется найти решение $x_* = (x_{*1}, \dots, x_{*n})^T \in R^n$ системы (1.2), подстановка которого в (1.2) приводит к верному равенству $Ax_* = b$.

Замечания

1. Из линейной алгебры известно, что решение задачи (1.2) существует и единствено, если детерминант матрицы A отличен от нуля, т. е. $\det A \equiv |A| \neq 0$ (A — невырожденная матрица, называемая также неособенной).

2. Поставленная задача часто именуется *первой задачей линейной алгебры*. Подчеркнем, что в ней входными (исходными) данными являются матрица A и вектор b , а выходными — вектор x .

3. Задача (1.2) имеет следующие особенности:

а) задача линейная (все переменные x_i , входящие в систему, имеют степени не выше первой) и неоднородная ($b \neq 0$);

б) количество уравнений равно количеству неизвестных (система замкнута);

в) количество уравнений для некоторых практических задач велико: $n > k \cdot 10^3$ ($k = 1 \div 10^3$);

г) при больших n использовать формулу $x = A^{-1}b$ не рекомендуется в силу трудностей нахождения обратной матрицы.

4. Важнейшим *признаком* любой математической задачи, который надо в первую очередь принимать во внимание при ее анализе и выборе метода решения, является ее *линейность* или *нелинейность*. Это связано с тем, что нелинейные задачи с вычислительной точки зрения являются наиболее трудными. Так, нелинейная задача (1.1) является достаточно сложной при числе уравнений n , пропорциональном 10^2 , а линейная задача — при n , пропорциональном 10^6 .

Характер задачи и точность получаемого решения в большой степени зависят от ее *обусловленности*, являющейся важнейшим математическим понятием, влияющим на выбор метода ее решения. Поясним это понятие на примере двумерной задачи (см. также пример В.1):

$$a_{11}x_1 + a_{12}x_2 = b_1,$$

$$a_{21}x_1 + a_{22}x_2 = b_2.$$

Точным решением этой задачи является вектор $x_* = (x_{*1}, x_{*2})^T$, компоненты которого определяются координатами точки пересечения двух прямых, соответствующих уравнениям $a_{11}x_1 + a_{12}x_2 = b_1$, $a_{21}x_1 + a_{22}x_2 = b_2$ (рис. 1.1а).

На рисунке 1.1б применительно к трем наборам входных данных, заданных с некоторыми погрешностями и соответствующих различным системам линейных уравнений, иллюстрируется характер обусловленности системы.

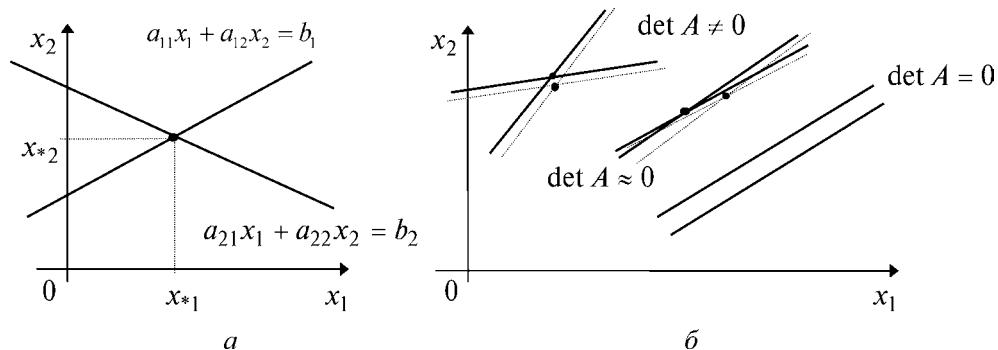


Рис. 1.1

Если $\det A$ существенно отличен от нуля, то точка пересечения пунктирных прямых, смещенных относительно сплошных прямых из-за погрешностей задания A и b , сдвигается незначительно. Это свидетельствует о хорошей обусловленности системы. При $\det A \approx 0$ небольшие погрешности в коэффициентах могут привести к большим погрешностям в решении (плохо обусловленная задача), поскольку прямые близки к параллельным. При $\det A = 0$ прямые параллельны или совпадают, и тогда решение задачи не существует или оно не единственное.

Более строго обусловленность задачи характеризуется числом обусловленности $v(A) = \|A\| \cdot \|A^{-1}\|$, где $\|A\|$ — норма матрицы A , а $\|A^{-1}\|$ — норма обратной матрицы. Чем больше это число, тем хуже обусловленность системы (при $v(A) \approx 10^3 \div 10^4$ система линейных алгебраических уравнений плохо обусловлена). В качестве нормы матрицы может быть принято число, являющееся максимальным из сумм (по модулю) элементов всех строк этой матрицы (подробнее см. в п. 1.3.1).

Подчеркнем, что реализация хорошей или плохой обусловленности в корректной и некорректной задачах напрямую связана с вытекающей отсюда численной устойчивостью или неустойчивостью. При этом для решения некорректных задач обычно применяются специальные методы или математические преобразования этих задач к корректным.

В численном анализе используются два класса численных методов решения систем линейных алгебраических уравнений.

1. *Прямые методы*, позволяющие найти решение за определенное число операций. К прямым методам относятся: метод Гаусса и его модификации (в том числе метод прогонки), метод LU — разложения и др.

2. *Итерационные методы*, основанные на использовании повторяющегося (циклического) процесса и позволяющие получить решение в результате последовательных приближений. Операции, входящие в повторяющийся процесс, составляют итерацию. К итерационным методам относятся: метод простых итераций, метод Зейделя и др.

1.2. ПРЯМЫЕ МЕТОДЫ

1.2.1. МЕТОД ГАУССА

Метод Гаусса состоит в исключении слагаемых системы путем ее равносильного преобразования. При этом в численном анализе с целью хорошей алгоритмизации задачи преобразованию подвергаются не сами уравнения, а исходные данные — матрица A и вектор b , соединенные в одну (расширенную) матрицу. В этом случае метод разбивается на две совокупности операций, которые условно названы *прямым ходом* и *обратным ходом*.

Методика решения задачи

1. *Прямой ход* состоит в исключении элементов, расположенных ниже элементов, соответствующих главной диагонали матрицы A . Схематически прямой ход представляется в виде

$$A_1 = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} & : & b_1 \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} & : & b_2 \\ \dots & \dots & \dots & \dots & \dots & : & \dots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nn} & : & b_n \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & c_{12} & c_{13} & \dots & c_{1n} & : & d_1 \\ 0 & 1 & c_{23} & \dots & c_{2n} & : & d_2 \\ \dots & \dots & \dots & \dots & \dots & : & \dots \\ 0 & 0 & 0 & \dots & 1 & : & d_n \end{pmatrix} = \hat{A}_1, \quad (1.3)$$

где \hat{A}_1 — матрица, которая содержит *опорные строки* (о. с.), получающиеся при выполнении прямого хода, а затем использующиеся в обратном ходе.

Особенности прямого хода:

а) матрица A преобразуется к верхнетреугольному виду с единицами на главной диагонали;

б) получение матрицы \hat{A}_1 в прямом ходе включает следующие действия:

- деление на *ведущие элементы* — на a_{11} в первой строке, на \tilde{a}_{22} — во второй, на \tilde{a}_{33} — в третьей и т. д. (при этом полагается, что они отличны от нуля, и в этом случае рассматриваемый метод Гаусса называется методом *единственного деления* в смысле отсутствия перебора ведущих элементов);
- исключение элементов a_{ij} столбцов, лежащих ниже опорных строк матрицы \hat{A}_1 . Для исключения элементов первого столбца используется первая о. с., для элементов второго столбца — вторая о. с. и т. д. Все остальные действия поясним на примере.

2. *Обратный ход* — от \hat{A}_1 переходим к системе, включающей x_1, \dots, x_n , и, начиная с последнего уравнения, последовательно определяем x_n, x_{n-1}, \dots, x_1 :

$$\begin{aligned} x_n &= d_n; \\ x_{n-1} &= d_{n-1} - c_{n-1,n} \cdot x_n, \\ &\dots \\ x_1 &= d_1 - c_{12}x_2 - \dots - c_{1n}x_n. \end{aligned}$$

Здесь и далее знак «*» у компонент решения системы опущен.

Пример 1.1. Методом Гаусса единственного деления найти решение системы линейных алгебраических уравнений ($n = 3$):

$$\begin{aligned} 5x_1 + 0x_2 + x_3 &= 11, \\ 2x_1 + 6x_2 - 2x_3 &= 8, \\ -3x_1 + 2x_2 + 10x_3 &= 6. \end{aligned}$$

□1. Прямой ход.

Пояснения действий по преобразованию матрицы A_1 .

5	0	1	11	Исходная расширенная матрица A_1 , в прямоугольник заключен ведущий элемент (в. э.) первой строки.
2	6	-2	8	
-3	2	10	6	
1	0	0,2	2,2	Поделили первую строку на первый в. э. — получили первую о. с.
0	6	-2,4	3,6	Умножили первую о. с. на $a_{21} = 2$ и результат вычли из второй строки.
0	2	10,6	12,6	Умножили первую о. с. на $a_{31} = -3$ и результат вычли из третьей строки.

Исключение элементов первого столбца по первой о. с. выполнено. После этого делением на $\tilde{a}_{22} = 6$ готовится вторая о. с., и далее исключаются элементы второго столбца в строках, расположенных ниже второй о. с., и т. д.

Весь прямой ход легко прослеживается по результирующей таблице преобразования матрицы A_1 в матрицу \hat{A}_1 .

Начинаем с первой строки.

5	0	1	11	
2	6	-2	8	
-3	2	10	6	
1	0	0,2	2,2	Первая о. с. (результат деления 1-й строки на $a_{11} = 5$).

Продолжаем со второй строки.

0	6	-2,4	3,6	
0	2	10,6	12,6	
0	1	-0,4	0,6	Вторая о. с. (результат деления на $\tilde{a}_{22} = 6$).
0	0	11,4	11,4	

Продолжаем с третьей строки.

0	0	1	1	Третья о. с. (результат деления на $\tilde{a}_{33} = 11,4$).
---	---	---	---	---

Таким образом, в результате прямого хода получили

$$\hat{A}_1 = \begin{pmatrix} 1 & 0 & 0,2 & 2,2 \\ 0 & 1 & -0,4 & 0,6 \\ 0 & 0 & 1 & 1 \end{pmatrix}.$$

2. Обратный ход. По матрице \hat{A}_1 составим систему уравнений:

$$\begin{pmatrix} 1 & 0 & 0,2 \\ 0 & 1 & -0,4 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2,2 \\ 0,6 \\ 1 \end{pmatrix}, \text{ или } \begin{aligned} x_1 + 0 \cdot x_2 + 0,2 \cdot x_3 &= 2,2; \\ x_2 - 0,4 \cdot x_3 &= 0,6; \\ x_3 &= 1. \end{aligned}$$

Решая эту систему, начиная с последнего уравнения, получаем: $x_3 = 1$; $x_2 = 1$; $x_1 = 2$. ■

Замечания

1. На основе выполненного прямого хода путем перемножения ведущих элементов a_{11} , \tilde{a}_{22} , \tilde{a}_{33} и т. д. вычисляется определитель матрицы A :

$$|A| = \det A = a_{11} \cdot \tilde{a}_{22} \cdot \tilde{a}_{33} \cdots \tilde{\tilde{a}}_{nn}.$$

Для примера 1.1: $\det A = 5 \cdot 6 \cdot 11,4 = 342$.

2. Изложенный метод имеет ограничение, связанное с тем, что ведущие элементы a_{11} , a_{22} и т. д. должны быть отличны от нуля и не должны быть малыми по модулю, поскольку тогда погрешности вычислений будут большими.

Поэтому в *методе Гаусса — Жордана* (с выбором главного элемента) в качестве ведущих элементов выбираются максимальные по модулю a_{ii} путем перебора этих элементов по столбцу, соответствующему этому в. э., или по всем столбцам.

3. При решении систем линейных алгебраических уравнений методом Гаусса единственного деления необходимо выполнить количество операций, пропорциональное n^3 . Оно может быть существенно снижено, если система имеет разреженную матрицу A (см. метод прогонки).

Процесс решения иногда удобно представлять в виде последовательности матриц, получаемых в результате элементарных преобразований.

Пример 1.2. Решить систему линейных алгебраических уравнений:

$$2x_1 + x_2 + 4x_3 = 16,$$

$$3x_1 + 2x_2 + x_3 = 10,$$

$$x_1 + 3x_2 + 3x_3 = 16$$

методом Гаусса единственного деления.

□1. Прямой ход. Запишем расширенную матрицу и реализуем прямой ход с помощью описанных преобразований:

$$\begin{aligned} A_1 &= \begin{pmatrix} 2 & 1 & 4 & 16 \\ 3 & 2 & 1 & 10 \\ 1 & 3 & 3 & 16 \end{pmatrix} \xrightarrow{k=1} \begin{pmatrix} 1 & \frac{1}{2} & 2 & 8 \\ 0 & \boxed{\frac{1}{2}} & -5 & -14 \\ 0 & \frac{5}{2} & 1 & 8 \end{pmatrix} \xrightarrow{k=2} \begin{pmatrix} 1 & \frac{1}{2} & 2 & 8 \\ 0 & 1 & -10 & -28 \\ 0 & 0 & \boxed{26} & 78 \end{pmatrix} \xrightarrow{k=3} \\ &\xrightarrow{k=3} \begin{pmatrix} 1 & \frac{1}{2} & 2 & 8 \\ 0 & 1 & -10 & -28 \\ 0 & 0 & 1 & 3 \end{pmatrix} = \hat{A}_1. \end{aligned}$$

При этом в соответствии с вышеизложенным алгоритмом можно использовать два правила:

- каждый элемент строки, в которой находится ведущий элемент, делится на него: $a_{kj}^{(k)} = \frac{a_{kj}^{(k-1)}}{a_{kk}^{(k-1)}}$, $j = 1, \dots, n+1$; k — номер шага (формирование о. с.);
 - элементы строк, находящихся ниже строки с ведущим элементом, подсчитываются по правилу прямоугольника, схематически показанного на рисунке 1.2 (исключение элементов, лежащих ниже в. э. текущей о. с.).
- Поясним алгоритм исключения (второе правило) на рисунке 1.2. Пусть рассчитывается значение $a_{ij}^{(k)}$ на k -м шаге. Следует соединить элемент $a_{ij}^{(k-1)}$

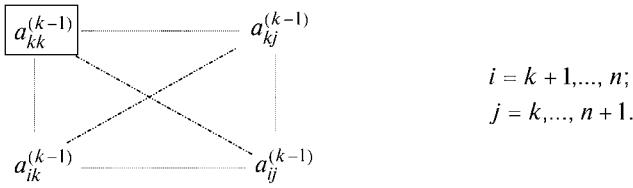


Рис. 1.2

с ведущим элементом $a_{kk}^{(k-1)}$. Получена одна из диагоналей прямоугольника. Вторую диагональ образует соединение элементов $a_{ik}^{(k-1)}$ и $a_{kj}^{(k-1)}$. Для нахождения значения $a_{ij}^{(k)}$ из его текущего значения $a_{ij}^{(k-1)}$ вычитается произведение элементов $a_{ik}^{(k-1)}$ и $a_{kj}^{(k-1)}$, деленное на ведущий элемент.

2. Обратный ход. По матрице \hat{A}_1 составим систему уравнений:

$$\begin{pmatrix} 1 & \frac{1}{2} & 2 \\ 0 & 1 & -10 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 8 \\ -28 \\ 3 \end{pmatrix} \text{ или } \begin{aligned} x_1 + \frac{1}{2}x_2 + 2x_3 &= 8, \\ x_2 - 10x_3 &= -28, \\ x_3 &= 3. \end{aligned}$$

Решая эту систему, начиная с последнего уравнения, находим: $x_3 = 3$, $x_2 = 2$, $x_1 = 1$. ■

Пример 1.3. Методом Гаусса с выбором ведущего элемента по столбцам решить систему:

$$\begin{aligned} -3x_1 + 2,099x_2 + 6x_3 &= 3,901; \\ 10x_1 - 7x_2 + 0x_3 &= 7; \\ 5x_1 - x_2 + 5x_3 &= 6. \end{aligned}$$

□1. Прямой ход. Реализуем поиск ведущего элемента по правилу: на k -м шаге переставляются $(n - k + 1)$ оставшихся уравнений так, чтобы наибольший по модулю коэффициент при x_k попал на главную диагональ:

$$\begin{aligned} A_1 &= \begin{pmatrix} -3 & 2,099 & 6 & 3,901 \\ 10 & -7 & 0 & 7 \\ 5 & -1 & 5 & 6 \end{pmatrix} \xrightarrow{k=1} \begin{pmatrix} \boxed{10} & -7 & 0 & 7 \\ -3 & 2,099 & 6 & 3,901 \\ 5 & -1 & 5 & 6 \end{pmatrix} \xrightarrow{k=2} \\ &\xrightarrow{k=1} \begin{pmatrix} 1 & -0,7 & 0 & 0,7 \\ 0 & -0,001 & 6 & 6,001 \\ 0 & 2,5 & 5 & 2,5 \end{pmatrix} \xrightarrow{k=2} \begin{pmatrix} 1 & -0,7 & 0 & 0,7 \\ 0 & \boxed{2,5} & 5 & 2,5 \\ 0 & -0,001 & 6 & 6,001 \end{pmatrix} \xrightarrow{k=2} \\ &\xrightarrow{k=2} \begin{pmatrix} 1 & -0,7 & 0 & 0,7 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & \boxed{6,002} & 6,002 \end{pmatrix} \xrightarrow{k=3} \begin{pmatrix} 1 & -0,7 & 0 & 0,7 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix} = \hat{A}_1. \end{aligned}$$

2. Обратный ход. По матрице \hat{A}_1 составим систему уравнений:

$$\begin{pmatrix} 1 & -0,7 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0,7 \\ 1 \\ 1 \end{pmatrix} \text{ или } \begin{aligned} x_1 - 0,7x_2 &= 0,7, \\ x_2 + 2x_3 &= 1, \\ x_3 &= 1. \end{aligned}$$

Решая ее, последовательно получаем: $x_3 = 1$, $x_2 = -1$, $x_1 = 0$. ■

Пример 1.4. Решить систему уравнений:

$$\begin{aligned}x_1 - x_2 + x_3 - 4x_4 &= -2, \\2x_1 + x_2 - 5x_3 + x_4 &= 2, \\8x_1 - x_2 - x_3 + 2x_4 &= 11, \\x_1 + 6x_2 - 2x_3 - 2x_4 &= -7\end{aligned}$$

методом Гаусса единственного деления.

□1. Прямой ход.

$$\left(\begin{array}{ccccc} 1 & -1 & 1 & -4 & -2 \\ 2 & 1 & -5 & 1 & 2 \\ 8 & -1 & -1 & 2 & 11 \\ 1 & 6 & -2 & -2 & -7 \end{array} \right) \xrightarrow{k=1} \left(\begin{array}{ccccc} 1 & -1 & 1 & -4 & -2 \\ 0 & 3 & -7 & 9 & 6 \\ 0 & 7 & -9 & 34 & 27 \\ 0 & 7 & -3 & 2 & -5 \end{array} \right) \xrightarrow{k=2} \left(\begin{array}{ccccc} 1 & -1 & 1 & -4 & -2 \\ 0 & 1 & -\frac{7}{3} & 3 & 2 \\ 0 & 0 & \frac{22}{3} & 13 & 13 \\ 0 & 0 & \frac{40}{3} & -\frac{57}{3} & -19 \end{array} \right) \xrightarrow{k=3} \right.$$

$$\left. \begin{array}{ccccc} 1 & -1 & 1 & -4 & -2 \\ 0 & 1 & -\frac{7}{3} & 3 & 2 \\ 0 & 0 & 1 & \frac{39}{22} & \frac{39}{22} \\ 0 & 0 & 0 & -\frac{2814}{66} & -\frac{2814}{66} \end{array} \right) \xrightarrow{k=4} \left(\begin{array}{ccccc} 1 & -1 & 1 & -4 & -2 \\ 0 & 1 & -\frac{7}{3} & 3 & 2 \\ 0 & 0 & 1 & \frac{39}{22} & \frac{39}{22} \\ 0 & 0 & 0 & 1 & 1 \end{array} \right).$$

2. Обратный ход.

$$\left(\begin{array}{cccc} 1 & -1 & 1 & -4 \\ 0 & 1 & -\frac{7}{3} & 3 \\ 0 & 0 & 1 & \frac{39}{22} \\ 0 & 0 & 0 & 1 \end{array} \right) \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} -2 \\ 2 \\ \frac{39}{22} \\ 1 \end{pmatrix} \quad \begin{array}{l} x_1 - x_2 + x_3 - 4x_4 = -2 \\ x_2 - \frac{7}{3}x_3 + 3x_4 = 2, \\ x_3 + \frac{39}{22}x_4 = \frac{39}{22}, \\ x_4 = 1. \end{array}$$

Отсюда $x_4 = 1$, $x_3 = \frac{39}{22} - \frac{39}{22}x_4 = 0$, $x_2 = 2 + \frac{7}{3}x_3 - 3x_4 = -1$, $x_1 = -2 + x_2 - x_3 + 4x_4 = 1$.

В результате получено решение: $x_* = (1; -1; 0; 1)^T$. ■

1.2.2. МЕТОД ПРОГОНКИ

Метод применим в случае, когда матрица A — трехдиагональная. Общая постановка задачи имеет следующий вид.

Дана система линейных алгебраических уравнений с трехдиагональной матрицей A . Развернутая запись этой системы имеет вид

$$\alpha_i x_{i-1} - \beta_i x_i + \gamma_i x_{i+1} = \delta_i, \quad \alpha_1 = \gamma_n = 0, \quad i = \overline{1, n}, \quad (1.4)$$

которому соответствует расширенная матрица

$$A_1 = \begin{pmatrix} -\beta_1 & \gamma_1 & 0 & 0 & \cdots & 0 & 0 & \delta_1 \\ \alpha_2 & -\beta_2 & \gamma_2 & 0 & \cdots & 0 & 0 & \delta_2 \\ 0 & \alpha_3 & -\beta_3 & \gamma_3 & \cdots & 0 & 0 & \delta_3 \\ \cdots & \cdots \\ 0 & 0 & \cdots & \cdots & \cdots & \alpha_n & -\beta_n & \delta_n \end{pmatrix}.$$

Здесь первое и последнее уравнения, содержащие по два слагаемых, могут рассматриваться как краевые условия. Знак «–» при коэффициенте β_i взят для более удобного представления расчетных формул метода.

Требуется найти решение $x_* = (x_{*1}, \dots, x_{*n})^T$ системы (1.4) методом исключения Гаусса.

Если к (1.4) применить алгоритм прямого хода метода Гаусса, то вместо A_1 получится \hat{A}_1 :

$$\hat{A}_1 = \begin{pmatrix} 1 & -P_1 & 0 & 0 & \cdots & Q_1 \\ 0 & 1 & -P_2 & 0 & \cdots & Q_2 \\ 0 & 0 & 1 & -P_3 & \cdots & Q_3 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 1 & \cdots & Q_n \end{pmatrix}.$$

Учитывая, что последний столбец в этой матрице соответствует правой части, и переходя к системе, включающей неизвестные, получаем рекуррентную формулу:

$$x_i = P_i x_{i+1} + Q_i, \quad i = \overline{1, n-1}. \quad (1.5)$$

Соотношение (1.5) есть формула для обратного хода, а формулы для коэффициентов P_i , Q_i , которые называются *прогоночными*, определяются из (1.4), (1.5). Запишем (1.5) для индекса $i-1$:

$$x_{i-1} = P_{i-1} x_i + Q_{i-1}$$

и подставим в (1.4). Получим

$$\alpha_i(P_{i-1} x_i + Q_{i-1}) - \beta_i x_i + \gamma_i x_{i+1} = \delta_i.$$

Приводя эту формулу к виду (1.5) и сравнивая полученное выражение с (1.5), получаем рекуррентные соотношения для P_i , Q_i :

$$P_i = \frac{\gamma_i}{\beta_i - \alpha_i P_{i-1}}, \quad Q_i = \frac{\alpha_i Q_{i-1} - \delta_i}{\beta_i - \alpha_i P_{i-1}}, \quad i = \overline{1, n-1}. \quad (1.6)$$

Определение прогоночных коэффициентов по формулам (1.6) соответствует *прямому ходу* метода прогонки.

Обратный ход метода прогонки начинается с вычисления x_n . Для этого используется последнее уравнение, коэффициенты которого определены в прямом ходе, и последнее уравнение исходной системы:

$$\begin{aligned} x_{n-1} &= P_{n-1} x_n + Q_{n-1}, \\ \alpha_n x_{n-1} - \beta_n x_n + 0 \cdot x_{n+1} &= \delta_n. \end{aligned}$$

Тогда определяется x_n :

$$x_n = \frac{\alpha_n Q_{n-1} - \delta_n}{\beta_n - \alpha_n P_{n-1}} = Q_n, \text{ т.е. } x_n = Q_n. \quad (1.7)$$

Остальные значения неизвестных находятся рекуррентно по формуле (1.5).

Все соотношения для выполнения вычислений получены. Тогда можно провести расчеты по методу Гаусса, используя прямой и обратный ход.

Методика решения задачи

Прямой ход.

1. Вычислить $P_1 = \frac{\gamma_1}{\beta_1}$; $Q_1 = -\frac{\delta_1}{\beta_1}$ (в (1.6) подставить $\alpha_1 = 0$).

2. Вычислить прогоночные коэффициенты $P_2, Q_2; P_3, Q_3; \dots; P_{n-1}, Q_{n-1}$ по формулам (1.6).

Обратный ход.

1. Найти $x_n = \frac{\alpha_n Q_{n-1} - \delta_n}{\beta_n - \alpha_n P_{n-1}}$.

2. Значения $x_{n-1}, x_{n-2}, \dots, x_1$ определить по формуле (1.5):

$$x_{n-1} = P_{n-1}x_n + Q_{n-1}, x_{n-2} = P_{n-2}x_{n-1} + Q_{n-2}, \dots, x_1 = P_1x_2 + Q_1.$$

Замечания

1. Данный метод называется методом скалярной прогонки, так как при решении задачи на каждом i -м шаге определяется скалярная величина x_i ($i = \overline{1, n}$).

2. Аналогичный подход используется для решения систем линейных алгебраических уравнений с пятидиагональными матрицами.

3. Алгоритм метода прогонки называется *корректным*, если для всех $i = \overline{1, n}$ $\beta_i - \alpha_i P_{i-1} \neq 0$, и *устойчивым*, если $|P_i| < 1$, $i = \overline{1, n-1}$.

4. Достаточным условием *корректности* и *устойчивости* прогонки является *условие преобладания диагональных элементов* в матрице A , в которой $\alpha_i \neq 0$ и $\gamma_i \neq 0$ ($i = \overline{2, n-1}$):

$$|\beta_i| \geq |\alpha_i| + |\gamma_i| \quad (1.8)$$

и в (1.8) имеет место строгое неравенство хотя бы при одном i .

5. Алгоритм метода прогонки является очень экономичным и требует для своей реализации количество операций, пропорциональное n .

Пример 1.5. Данна система линейных алгебраических уравнений с трехдиагональной матрицей A ($n = 4$):

$$\begin{aligned} 5x_1 + 3x_2 &= 8, \\ 3x_1 + 6x_2 + x_3 &= 10, \\ x_2 + 4x_3 - 2x_4 &= 3, \\ x_3 - 3x_4 &= -2 \end{aligned}$$

($\alpha_1 = 0, \gamma_4 = 0$). Решить эту систему методом прогонки.

□ Данная система удовлетворяет условию преобладания диагональных элементов (1.8): в первом уравнении $5 > 3$, во втором уравнении $6 > 3 + 1$, в третьем уравнении $4 > 1 + 2$, в четвертом уравнении $3 > 1$. Далее выполняем прямой и обратный ход, учитывая, что матрица

$$A_1 = \begin{pmatrix} 5 & 3 & 0 & 0 & 8 \\ 3 & 6 & 1 & 0 & 10 \\ 0 & 1 & 4 & -2 & 3 \\ 0 & 0 & 1 & -3 & -2 \end{pmatrix}.$$

Прямой ход. Вычислим прогоночные коэффициенты:

$$P_1 = \frac{\gamma_1}{\beta_1} = \frac{3}{-5} = -\frac{3}{5}; \quad Q_1 = -\frac{\delta_1}{\beta_1} = -\frac{8}{-5} = \frac{8}{5};$$

$$P_2 = \frac{\gamma_2}{\beta_2 - \alpha_2 P_1} = \frac{1}{-6 - 3 \cdot \left(-\frac{3}{5}\right)} = -\frac{5}{21}; \quad Q_2 = \frac{\alpha_2 Q_1 - \delta_2}{\beta_2 - \alpha_2 P_1} = \frac{3 \cdot \frac{8}{5} - 10}{-6 - 3 \cdot \left(-\frac{3}{5}\right)} = \frac{26}{21}.$$

Подчеркнем, что $\beta_1 = -5$; $\beta_2 = -6$; $\beta_3 = -4$; $\beta_4 = 3$, так как в (1.4) во втором слагаемом взят знак «минус»:

$$P_3 = \frac{\gamma_3}{\beta_3 - \alpha_3 P_2} = \frac{-2}{-4 - 1 \cdot \left(-\frac{5}{21}\right)} = \frac{42}{79}; \quad Q_3 = \frac{\alpha_3 Q_2 - \delta_3}{\beta_3 - \alpha_3 P_2} = \frac{1 \cdot \frac{26}{21} - 3}{-4 - 1 \cdot \left(-\frac{5}{21}\right)} = \frac{37}{79}.$$

Обратный ход:

$$x_4 = Q_4 = \frac{\alpha_4 Q_3 - \delta_4}{\beta_4 - \alpha_4 P_3} = \frac{1 \cdot \frac{37}{79} + 2}{3 - 1 \cdot \left(\frac{42}{79}\right)} = 1; \quad x_3 = P_3 x_4 + Q_3 = \frac{42}{79} \cdot 1 + \frac{37}{79} = 1;$$

$$x_2 = P_2 x_3 + Q_2 = -\frac{5}{21} \cdot 1 + \frac{26}{21} = 1; \quad x_1 = P_1 x_2 + Q_1 = -\frac{3}{5} \cdot 1 + \frac{8}{5} = 1.$$

Подстановкой решения $x_* = (1; 1; 1; 1)^T$ в исходную систему убеждаемся, что задача решена верно. Путем анализа всех полученных $\beta_i - \alpha_i P_{i-1}$ и P_i

Таблица 1.1

i	α_i	β_i	γ_i	δ_i	P_i	Q_i	x_i
1	0	-5	3	8	$-\frac{3}{5}$	$\frac{8}{5}$	1
2	3	-6	1	10	$-\frac{5}{21}$	$\frac{26}{21}$	1
3	1	-4	-2	3	$\frac{42}{79}$	$\frac{37}{79}$	1
4	1	3	0	-2	—	1	1

Прямой ход

Обратный ход

легко убедиться также в том, что для данного примера прогонка оказалась корректной и устойчивой, поскольку $\beta_i - \alpha_i P_{i-1} \neq 0$, $i = \overline{1, 4}$; $|P_i| < 1$, $i = \overline{1, 3}$.

Для наглядности представления информации исходные данные и результаты расчетов поместим в таблицу 1.1, где в первых четырех колонках содержатся исходные данные, а в последних трех — полученные результаты. ■

Пример 1.6. Данна система линейных алгебраических уравнений с трехдиагональной матрицей A :

$$\begin{aligned}x_1 + 2x_2 &= 5, \\2x_1 - x_2 + x_3 &= 3, \\x_2 - x_3 + x_4 &= 3, \\x_3 + x_4 &= 7.\end{aligned}$$

Решить систему методом прогонки.

□ Результаты расчетов в прямом и обратном ходе занесены в таблицу 1.2.

Таблица 1.2

i	α_i	β_i	γ_i	δ_i	P_i	Q_i	x_i
1	0	-1	2	5	-2	5	1
2	2	1	1	3	$\frac{1}{5}$	$\frac{7}{5}$	2
3	1	1	1	3	$\frac{5}{4}$	-2	3
4	1	-1	0	7	—	4	4

В результате получено решение: $x_* = (1; 2; 3; 4)^T$. Заметим, что условие преобладания диагональных элементов в данном примере не выполнено, но алгоритм метода прогонки позволил получить точное решение. При этом обратим внимание на небольшой порядок системы и отсутствие погрешностей вычислений. ■

Пример 1.7. Решить систему уравнений:

$$\begin{aligned}2x_1 + x_2 &= 4, \\2x_1 + 3x_2 - x_3 &= 9, \\x_2 - x_3 + 3x_4 &= 12, \\x_3 - x_4 &= -4\end{aligned}$$

или

$$\begin{pmatrix} 2 & 1 & 0 & 0 \\ 2 & 3 & -1 & 0 \\ 0 & 1 & -1 & 3 \\ 0 & 0 & 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 4 \\ 9 \\ 12 \\ -4 \end{pmatrix}$$

методом прогонки.

□ *Прямой ход.* Вычислим прогоночные коэффициенты:

$$P_1 = \frac{\gamma_1}{\beta_1} = -\frac{1}{2}; \quad Q_1 = -\frac{\delta_1}{\beta_1} = -\frac{4}{-2} = 2;$$

$$P_2 = \frac{\gamma_2}{\beta_2 - \alpha_2 P_1} = \frac{-1}{-3 - 2 \cdot \left(-\frac{1}{2}\right)} = \frac{1}{2}; \quad Q_2 = \frac{\alpha_2 Q_1 - \delta_2}{\beta_2 - \alpha_2 P_1} = \frac{2 \cdot 2 - 9}{-3 - 2 \cdot \left(-\frac{1}{2}\right)} = \frac{5}{2};$$

$$P_3 = \frac{\gamma_3}{\beta_3 - \alpha_3 P_2} = \frac{3}{1 - 1 \cdot \frac{1}{2}} = 6; \quad Q_3 = \frac{\alpha_3 Q_2 - \delta_3}{\beta_3 - \alpha_3 P_2} = \frac{1 \cdot \frac{5}{2} - 12}{1 - 1 \cdot \frac{1}{2}} = -19.$$

Обратный ход:

$$Q_4 = \frac{\alpha_4 Q_3 - \delta_4}{\beta_4 - \alpha_4 P_3} = \frac{1 \cdot (-19) - (-4)}{1 - 1 \cdot 6} = 3 = x_4; \quad x_3 = P_3 x_4 + Q_3 = 6 \cdot 3 - 19 = -1;$$

$$x_2 = P_2 x_3 + Q_2 = \frac{1}{2} \cdot (-1) + \frac{5}{2} = 2; \quad x_1 = P_1 x_2 + Q_1 = -\frac{1}{2} \cdot 2 + 2 = 1.$$

Таблица 1.3

i	α_i	β_i	γ_i	δ_i	P_i	Q_i	x_i
1	0	-2	1	4	$-\frac{1}{2}$	2	1
2	2	-3	-1	9	$\frac{1}{2}$	$\frac{5}{2}$	2
3	1	1	3	12	6	-19	-1
4	1	1	0	-4	—	3	3

Результаты расчетов приведены в таблице 1.3.

Получено решение системы: $x_* = (1; 2; -1; 3)^T$. ■

1.2.3. МЕТОД LU-РАЗЛОЖЕНИЯ

Рассмотрим еще один метод решения задачи (1.2). Метод опирается на возможность представления матрицы A системы в форме произведения двух треугольных матриц:

$$A = LU, \tag{1.9}$$

где L — нижняя, а U — верхняя треугольные матрицы, имеющие вид

$$L = \begin{pmatrix} l_{11} & 0 & \cdots & 0 \\ l_{21} & l_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ l_{n1} & l_{n2} & \cdots & l_{nn} \end{pmatrix}, \quad U = \begin{pmatrix} 1 & u_{12} & \cdots & u_{1n} \\ 0 & 1 & \cdots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}.$$

С учетом (1.9) система $Ax = b$ представляется в форме

$$LUx = b. \tag{1.10}$$

Решение системы (1.10) сводится к последовательному решению двух простых систем с треугольными матрицами. В итоге процедура решения состоит из двух этапов.

Прямой ход. Произведение Ux обозначим через y . В результате решения системы $Ly = b$ находится вектор y .

Обратный ход. В результате решения системы $Ux = y$ находится решение задачи — вектор x .

В силу треугольности матриц L и U решения обеих систем находятся рекуррентно (как в обратном ходе метода Гаусса).

Из общего вида элемента произведения $A = LU$, а также структуры матриц L и U следуют формулы для определения элементов этих матриц:

$$\begin{aligned} l_{ij} &= a_{ij} - \sum_{s=1}^{j-1} l_{is} u_{sj}, \quad i \geq j; \\ u_{ij} &= \frac{1}{l_{ii}} \left(a_{ij} - \sum_{s=1}^{i-1} l_{is} u_{sj} \right), \quad i < j. \end{aligned} \quad (1.11)$$

Результат представления матрицы A в виде произведения двух треугольных матриц (операции *факторизации*) удобно хранить в одной матрице следующей структуры:

$$\left(\begin{array}{c|ccccc} l_{11} & u_{12} & u_{13} & \cdots & u_{1n} \\ l_{21} & l_{22} & u_{23} & \cdots & u_{2n} \\ l_{31} & l_{32} & l_{33} & \cdots & u_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ l_{n1} & l_{n2} & l_{n3} & \cdots & l_{nn} \end{array} \right).$$

Вычисления на k -м шаге метода LU -разложения удобно производить, пользуясь двумя схемами, изображенными на рисунке 1.3.

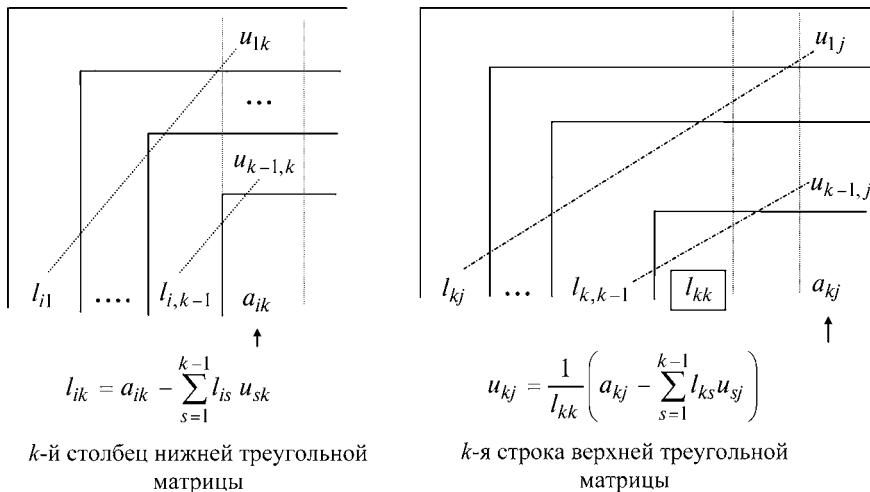


Рис. 1.3

Замечание. Всякую квадратную матрицу A , имеющую отличные от нуля угловые миноры $\Delta_1 = a_{11} \neq 0$, $\Delta_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \neq 0$, ..., $\Delta_n = |A| \neq 0$, можно представить в виде LU -разложения, причем это разложение будет единственным.

Методика решения задачи

1. Выполнить операцию факторизации исходной матрицы A , применяя схемы (рис. 1.3) или формулы (1.11), и получить матрицы L и U .

2. Решить систему $Ly = b$.

3. Решить систему $Ux = y$.

Пример 1.8. Решить систему линейных алгебраических уравнений:

$$2x_1 + x_2 + 4x_3 = 16,$$

$$3x_1 + 2x_2 + x_3 = 10,$$

$$x_1 + 3x_2 + 3x_3 = 16$$

методом LU -разложения.

□1. Выполним операцию факторизации:

$$\left(\begin{array}{ccc|c} 2 & 1 & 4 \\ 3 & 2 & 1 \\ 1 & 1 & 1 \end{array} \right) \xrightarrow{k=1} \left(\begin{array}{ccc|c} 2 & 0,5 & -2 \\ 3 & 2 & 1 \\ 1 & 3 & 1 \end{array} \right) \xrightarrow{k=2} \left(\begin{array}{ccc|c} 2 & 0,5 & 2 \\ 3 & 0,5 & -10 \\ 1 & 2,5 & 1 \end{array} \right) \xrightarrow{k=3} \left(\begin{array}{ccc|c} 2 & 0,5 & 2 \\ 3 & 0,5 & -10 \\ 1 & 2,5 & 26 \end{array} \right);$$

при $k = 1$ $l_{11} = a_{11} = 2$; $l_{21} = a_{21} = 3$; $l_{31} = a_{31} = 1$; $u_{12} = \frac{1}{2}a_{12} = 0,5$; $u_{13} = \frac{1}{2}a_{13} = \frac{4}{2} = 2$;

при $k = 2$ $l_{22} = a_{22} - l_{21}u_{12} = 2 - 3 \cdot 0,5 = 0,5$; $l_{32} = a_{32} - l_{31}u_{12} = 3 - 1 \cdot 0,5 = 2,5$;

$$u_{23} = \frac{1}{l_{22}}(a_{23} - l_{21}u_{13}) = \frac{1}{0,5}(1 - 3 \cdot 2) = -10;$$

$$\text{при } k = 3 \quad l_{33} = a_{33} - l_{31}u_{13} - l_{32}u_{23} = 3 - 1 \cdot 2 - 2,5 \cdot (-10) = 26.$$

В результате получены две треугольные матрицы:

$$L = \begin{pmatrix} 2 & 0 & 0 \\ 3 & 0,5 & 0 \\ 1 & 2,5 & 26 \end{pmatrix}, \quad U = \begin{pmatrix} 1 & 0,5 & 2 \\ 0 & 1 & -10 \\ 0 & 0 & 1 \end{pmatrix}.$$

2. Решим систему $Ly = b$:

$$\underbrace{\begin{pmatrix} 2 & 0 & 0 \\ 3 & 0,5 & 0 \\ 1 & 2,5 & 26 \end{pmatrix}}_L \cdot \underbrace{\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}}_y = \underbrace{\begin{pmatrix} 16 \\ 10 \\ 16 \end{pmatrix}}_b \quad \begin{aligned} 2y_1 &= 16, \\ 3y_1 + 0,5 \cdot y_2 &= 10, \\ y_1 + 2,5 \cdot y_2 + 26y_3 &= 16. \end{aligned}$$

Отсюда $y_1 = 8$, $y_2 = (10 - 3 \cdot 8) \cdot 2 = -28$, $y_3 = \frac{16 - 8 + 70}{26} = 3$.

3. Решим систему $Ux = y$:

$$\underbrace{\begin{pmatrix} 1 & 0,5 & 2 \\ 0 & 1 & -10 \\ 0 & 0 & 1 \end{pmatrix}}_U \cdot \underbrace{\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}}_x = \underbrace{\begin{pmatrix} 8 \\ -28 \\ 3 \end{pmatrix}}_y \quad \begin{aligned} x_1 + 0,5x_2 + 2x_3 &= 8, \\ x_2 - 10x_3 &= -28, \\ x_3 &= 3. \end{aligned}$$

Отсюда $x_3 = 3$, $x_2 = -28 + 10 \cdot 3 = 2$, $x_1 = 8 - 2 \cdot 3 - 0,5 \cdot 2 = 1$. Получен ответ: $x_* = (1; 2; 3)^T$. ■

Пример 1.9. Решить систему линейных алгебраических уравнений:

$$\begin{aligned}3x_1 - x_2 &= 5, \\-2x_1 + x_2 + x_3 &= 0, \\2x_1 - x_2 + 4x_3 &= 15\end{aligned}$$

методом LU-разложения.

□1. Выполним операцию факторизации:

$$\begin{array}{c}\left(\begin{array}{ccc|c}3 & -1 & 0 & \\-2 & 1 & 1 & \\2 & -1 & 4 & \end{array}\right) \xrightarrow{k=1} \left(\begin{array}{ccc|c}3 & -\frac{1}{3} & 0 & \\-2 & 1 & 1 & \\2 & -1 & \frac{4}{3} & \end{array}\right) \xrightarrow{k=2} \left(\begin{array}{ccc|c}3 & -\frac{1}{3} & 0 & \\-2 & \frac{1}{3} & 3 & \\2 & -\frac{1}{3} & \frac{4}{3} & \end{array}\right) \xrightarrow{k=3} \\ \xrightarrow{k=3} \left(\begin{array}{ccc|c}3 & -\frac{1}{3} & 0 & \\-2 & \frac{1}{3} & 3 & \\2 & -\frac{1}{3} & 5 & \end{array}\right);\end{array}$$

$$\text{при } k = 1 \quad l_{11} = a_{11} = 3; l_{21} = a_{21} = -2; l_{31} = a_{31} = 2; u_{12} = \frac{a_{12}}{l_{11}} = -\frac{1}{3}; u_{13} = \frac{a_{13}}{l_{11}} = 0;$$

$$\text{при } k = 2 \quad l_{22} = a_{22} - l_{21}u_{12} = 1 - 2 \cdot \frac{1}{3} = \frac{1}{3}; l_{32} = a_{32} - l_{31}u_{12} = -1 + 2 \cdot \frac{1}{3} = -\frac{1}{3};$$

$$u_{23} = \frac{1}{l_{22}}(a_{23} - l_{21}u_{13}) = \frac{1}{1/3} \cdot (1 - (-2) \cdot 0) = 3;$$

$$\text{при } k = 3 \quad l_{33} = a_{33} - l_{31}u_{13} - l_{32}u_{23} = 4 - 2 \cdot 0 - 3 \cdot \left(-\frac{1}{3}\right) = 5.$$

2. Решим систему $Ly = b$:

$$\underbrace{\left(\begin{array}{ccc|c}3 & 0 & 0 & \\-2 & \frac{1}{3} & 0 & \\2 & -\frac{1}{3} & 5 & \end{array}\right)}_L \cdot \underbrace{\begin{pmatrix}y_1 \\ y_2 \\ y_3\end{pmatrix}}_y = \underbrace{\begin{pmatrix}5 \\ 0 \\ 15\end{pmatrix}}_b \quad \text{или} \quad \begin{aligned}3y_1 &= 5, \\-2y_1 + \frac{1}{3}y_2 &= 0, \\2y_1 - \frac{1}{3}y_2 + 5y_3 &= 15.\end{aligned}$$

Отсюда $y_1 = \frac{5}{3}$, $y_2 = 10$, $y_3 = 3$.

3. Решим систему $Ux = y$:

$$\underbrace{\left(\begin{array}{ccc|c}1 & -\frac{1}{3} & 0 & \\0 & 1 & 3 & \\0 & 0 & 1 & \end{array}\right)}_U \cdot \underbrace{\begin{pmatrix}x_1 \\ x_2 \\ x_3\end{pmatrix}}_x = \underbrace{\begin{pmatrix}\frac{5}{3} \\ 10 \\ 3\end{pmatrix}}_y \quad \text{или} \quad \begin{aligned}x_1 - \frac{1}{3}x_2 &= \frac{5}{3}, \\x_2 + 3x_3 &= 10, \\x_3 &= 3.\end{aligned}$$

Отсюда $x_3 = 3$, $x_2 = 1$, $x_1 = 2$. ■

Пример 1.10. Решить систему линейных алгебраических уравнений

$$\begin{aligned} 2x_1 + x_2 &= 4, \\ 2x_1 + 3x_2 - x_3 &= 9, \\ x_2 - x_3 + 3x_4 &= 12, \\ x_3 - x_4 &= -4 \end{aligned}$$

методом LU-разложения.

□1. Выполним процедуру факторизации:

$$\begin{array}{c} \left(\begin{array}{cccc} 2 & 1 & 0 & 0 \\ 2 & 3 & -1 & 0 \\ 0 & 1 & -1 & 3 \\ 0 & 0 & 1 & -1 \end{array} \right) \xrightarrow{k=1} \left(\begin{array}{cccc} 2 & \frac{1}{2} & 0 & 0 \\ 2 & 3 & -1 & 0 \\ 0 & 1 & -1 & 3 \\ 0 & 0 & 1 & -1 \end{array} \right) \xrightarrow{k=2} \left(\begin{array}{cccc} 2 & \frac{1}{2} & 0 & 0 \\ 2 & \frac{1}{2} & -\frac{1}{2} & 0 \\ 0 & 1 & -1 & 3 \\ 0 & 0 & 1 & -1 \end{array} \right) \xrightarrow{k=3} \\ \xrightarrow{k=3} \left(\begin{array}{cccc} 2 & \frac{1}{2} & 0 & 0 \\ 2 & 2 & -\frac{1}{2} & 0 \\ 0 & 1 & \boxed{\frac{1}{2}} & -6 \\ 0 & 0 & 1 & -1 \end{array} \right) \xrightarrow{k=4} \left(\begin{array}{cccc} 2 & \frac{1}{2} & 0 & 0 \\ 2 & 2 & -\frac{1}{2} & 0 \\ 0 & 1 & -\frac{1}{2} & -6 \\ 0 & 0 & 1 & 5 \end{array} \right). \end{array}$$

В результате получаем

$$L = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 2 & 2 & 0 & 0 \\ 0 & 1 & -\frac{1}{2} & 0 \\ 0 & 0 & 1 & 5 \end{pmatrix}, \quad U = \begin{pmatrix} 1 & \frac{1}{2} & 0 & 0 \\ 0 & 1 & -\frac{1}{2} & 0 \\ 0 & 0 & 1 & -6 \\ 0 & 0 & 0 & 1 \end{pmatrix};$$

при $k = 1$ $l_{11} = 2, l_{21} = 2, l_{31} = 0, l_{41} = 0, u_{12} = \frac{1}{2} \cdot 1 = \frac{1}{2}, u_{13} = \frac{1}{2} \cdot 0 = 0, u_{14} = \frac{1}{2} \cdot 0 = 0;$

при $k = 2$ $l_{22} = 3 - 2 \cdot \frac{1}{2} = 2, l_{32} = 1 - 0 \cdot \frac{1}{2} = 1, l_{42} = 0 - 0 \cdot \frac{1}{2} = 0, u_{23} = \frac{1}{2}(-1 - 2 \cdot 0) = -\frac{1}{2}, u_{24} = \frac{1}{2}(0 - 2 \cdot 0) = 0;$

при $k = 3$ $l_{33} = -1 - 1 \cdot \left(-\frac{1}{2}\right) = -\frac{1}{2}, l_{43} = 1 - 0 \cdot 0 - 0 \cdot \left(-\frac{1}{2}\right) = 1, u_{34} = \frac{1}{-\frac{1}{2}}(3 - 0 \cdot 0 - 1 \cdot 0) = -6;$

при $k = 4$ $l_{44} = -1 - (0 \cdot 0 + 0 \cdot 0 + 1 \cdot (-6)) = 5.$

2. Решим систему $Ly = b$:

$$\begin{pmatrix} 2 & 0 & 0 & 0 \\ 2 & 2 & 0 & 0 \\ 0 & 1 & -\frac{1}{2} & 0 \\ 0 & 0 & 1 & 5 \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} 4 \\ 9 \\ 12 \\ -4 \end{pmatrix} \quad \begin{array}{l} 2y_1 = 4, \\ 2y_1 + 2y_2 = 9, \\ y_2 - \frac{1}{2}y_3 = 12, \\ y_3 + 5y_4 = -4. \end{array}$$

Отсюда $y_1 = 2, y_2 = \frac{5}{2}, y_3 = -19, y_4 = 3.$

3. Решим систему $Ux = y$:

$$\begin{pmatrix} 1 & \frac{1}{2} & 0 & 0 \\ 0 & 1 & -\frac{1}{2} & 0 \\ 0 & 0 & 1 & -6 \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 2 \\ \frac{5}{2} \\ -19 \\ 3 \end{pmatrix} \quad \begin{array}{l} x_1 + \frac{1}{2}x_2 = 2, \\ x_2 - \frac{1}{2}x_3 = \frac{5}{2}, \\ x_3 - 6x_4 = -19, \\ x_4 = 3. \end{array}$$

Отсюда получаем: $x_4 = 3$, $x_3 = -1$, $x_2 = 2$, $x_1 = 1$ или $x_* = (1; 2; -1; 3)^T$. ■

Замечания

1. При большом числе уравнений (больше 100) прямые методы решения систем линейных алгебраических уравнений становятся труднореализуемыми на ЭВМ прежде всего из-за сложности хранения промежуточных результатов и операций с матрицами большой размерности.

2. Существуют различные способы представления матрицы A в виде

$$A = A_1 \cdot A_2 \cdot \dots \cdot A_m,$$

где каждая матрица A_i имеет форму, удобную для решения системы линейных уравнений [11]. Как правило, $m \leq 5$, причем главным образом $m = 2$ и $m = 3$. Тогда в результате решения последовательности систем

$$A_1 b_1 = b, \quad A_2 b_2 = b_1, \dots, \quad A_m b_m = b_{m-1}$$

можно найти искомое решение $x_* = b_m$.

1.3. ИТЕРАЦИОННЫЕ МЕТОДЫ

1.3.1. МЕТОД ПРОСТЫХ ИТЕРАЦИЙ

Альтернативой прямым методам являются итерационные методы, основанные на многократном уточнении $x^{(0)}$ — приближенно заданного решения задачи $Ax = b$. Верхним индексом в скобках здесь и далее по тексту обозначается номер итерации (совокупности повторяющихся действий).

Суть простейшего итерационного метода — метода простых итераций, состоит в выполнении следующих процедур.

1. Исходная задача $Ax = b$ преобразуется к равносильному виду:

$$x = \alpha x + \beta, \tag{1.12}$$

где $\alpha = \{\alpha_{ij}\}$ — квадратная матрица, $\beta = \{\beta_i\}$ — вектор, $i, j = 1, \dots, n$. Это преобразование может быть выполнено различными путями, но для обеспечения сходимости итераций (см. процедуру 2) нужно добиться, чтобы $\|\alpha\| < 1$ (чтобы норма α была меньше единицы). Понятие нормы вводится ниже).

2. Вектор β принимается в качестве начального приближения $x^{(0)} = \beta$ и далее многократно выполняются действия по уточнению решения согласно рекуррентному соотношению

$$x^{(k+1)} = \alpha x^{(k)} + \beta, \quad k = 0, 1, \dots \tag{1.13}$$

или в развернутом виде

$$\begin{aligned}x_1^{(k+1)} &= \alpha_{11}x_1^{(k)} + \alpha_{12}x_2^{(k)} + \dots + \alpha_{1n}x_n^{(k)} + \beta_1, \\x_2^{(k+1)} &= \alpha_{21}x_1^{(k)} + \alpha_{22}x_2^{(k)} + \dots + \alpha_{2n}x_n^{(k)} + \beta_2, \\&\vdots \\x_n^{(k+1)} &= \alpha_{n1}x_1^{(k)} + \alpha_{n2}x_2^{(k)} + \dots + \alpha_{nn}x_n^{(k)} + \beta_n.\end{aligned}$$

3. Итерации прерываются при выполнении условия

$$\frac{\|\alpha\|}{1-\|\alpha\|} \|x^{(k+1)} - x^{(k)}\| \leq \varepsilon, \quad (1.14)$$

где $\varepsilon > 0$ — заданная точность, которую необходимо достигнуть при решении задачи, или более простого условия:

$$\|x^{(k+1)} - x^{(k)}\| < \varepsilon. \quad (1.15)$$

Оба условия носят апостериорный характер (*a posteriori* — после).

Замечания

1. Процесс (1.13) называется *параллельным итерированием*, так как для вычисления $(k+1)$ -го приближения всех неизвестных учитываются вычисленные ранее их k -е приближения.

2. Начальное приближение $x^{(0)}$ может выбираться произвольно или из некоторых соображений. При этом может использоваться априорная информация о решении или просто «грубая» прикидка.

При выполнении итераций (любых) возникают следующие вопросы:

а) сходится ли процесс (1.13), т. е. имеет ли место $x^{(k)} \rightarrow x_*$ при $k \rightarrow \infty$, где x_* — точное решение?

б) если сходимость есть, то какова ее скорость?

в) какова погрешность найденного решения $x^{(k+1)}$, т. е. чему равна норма разности $\|x^{(k)} - x_*\|$?

Для их разрешения необходимо сравнивать векторы $x^{(k)}$ и $x^{(k+1)}$. Это сравнение (а также математическое обоснование и других задач вычислительной математики) осуществляется на основе норм матриц и векторов. Приведем основные сведения о них.

Нормой матрицы $A = \{a_{ij}\}$ ($i, j = 1, \dots, n$) называется действительное число (обозначаемое $\|A\|$), удовлетворяющее условиям:

1) $\|A\| > 0$ при $A \neq 0$ и $\|A\| = 0$ тогда и только тогда, когда A нулевая матрица ($A = 0$);

2) $\|\alpha A\| = |\alpha| \cdot \|A\|$ для любого действительного α ; ($|\alpha|$ — модуль α);

3) $\|A_1 + A_2\| \leq \|A_1\| + \|A_2\|$, где A_1 и A_2 — некоторые матрицы (неравенство треугольника);

4) $\|A_1 \cdot A_2\| \leq \|A_1\| \cdot \|A_2\|$.

Здесь A_1 и A_2 — матрицы, для которых соответствующие операции имеют смысл.

Условия 1–4 — это условия нормы, а их непосредственное вычисление может производиться различными путями. Наиболее употребительными яв-

ляются следующие формулы для вычисления значений норм матриц и векторов, образованных действительными компонентами.

Нормы матрицы A

$$1) \|A\|_1 = \max_i \sum_{j=1}^n |a_{ij}|$$

$$2) \|A\|_2 = \max_j \sum_{i=1}^n |a_{ij}|$$

$$3) \|A\|_3 = \sqrt{\sum_{i=1}^n \sum_{j=1}^n a_{ij}^2}$$

Нормы вектора x

$$\|x\|_1 = \max_i |x_i|$$

$$\|x\|_2 = \sqrt{\sum_{i=1}^n |x_i|^2}$$

$$\|x\|_3 = \sqrt{\sum_{i=1}^n x_i^2}$$

Здесь приведены согласованные нормы матриц и векторов. Их согласование осуществляется связью:

$$\|A\| = \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|}.$$

Заметим, что в случае $\|x\|_3$ согласованная норма равна $\sqrt{\lambda_{\max}(A^T A)}$, где $\lambda_{\max}(A^T A)$ — максимальное собственное значение матрицы $A^T A$. Однако ее вычисление связано с весьма трудоемкими операциями. Можно доказать, что справедливо неравенство

$$\sqrt{\lambda_{\max}(A^T A)} \leq \|A\|_3 = \sqrt{\sum_{i=1}^n \sum_{j=1}^n a_{ij}^2}.$$

Так как величина $\|A\|_3$ вычисляется относительно просто, то она часто используется в оценках. Третья норма $\|\cdot\|_3$ называется *евклидовой*, для вектора она соответствует длине вектора, исходящего из начала координат. В множестве действительных чисел $\|x\| = |x|$.

Ответ на вопросы о сходимости дают следующие две теоремы, приводящиеся без доказательства.

Теорема 1.1 (о достаточном условии сходимости метода простых итераций).

Метод простых итераций, реализующийся в процессе последовательных приближений (1.13), сходится к единственному решению исходной системы $Ax = b$ при любом начальном приближении $x^{(0)}$ со скоростью не медленнее геометрической прогрессии, если какая-либо норма матрицы α меньше единицы, т. е. $\|\alpha\|_s < 1$ ($s \in \{1, 2, 3\}$).

Замечания

1. Условие теоремы 1.1, как достаточное, предъявляет завышенные требования к матрице α , и потому иногда сходимость будет, если даже $\|\alpha\| \geq 1$.

2. Сходящийся процесс обладает свойством *самоисправляемости*, т. е. отдельная ошибка в вычислениях не отразится на окончательном результате, так как ошибочное приближение можно рассматривать как новое начальное.

3. Условия сходимости выполняются, если в матрице A диагональные элементы преобладают, т. е.

$$|a_{ii}| \geq |a_{i1}| + \dots + |a_{i,i-1}| + |a_{i,i+1}| + \dots + |a_{in}|, \quad i = 1, \dots, n, \quad (1.16)$$

и хотя бы для одного i неравенство строгое. Иначе модули диагональных коэффициентов в каждом уравнении системы больше суммы модулей недиагональных коэффициентов (свободные члены не рассматриваются).

4. Чем меньше величина нормы $\|\alpha\|$, тем быстрее сходимость метода.

Теорема 1.2 (о необходимом и достаточном условии сходимости метода простых итераций).

Для сходимости последовательности (1.13) при любых $x^{(0)}$ и β необходимо и достаточно, чтобы собственные значения матрицы α были по модулю меньше единицы, т. е. $|\lambda_i(\alpha)| < 1$, $i = 1, \dots, n$.

Замечание. Хотя теорема 1.2 дает более общие условия сходимости метода простых итераций, чем теорема 1.1, однако ею воспользоваться сложнее, так как нужно предварительно вычислить границы собственных значений матрицы α или сами собственные значения.

Рассмотрим последовательность $\{x^{(k)}\}$, сходящуюся к x_* . Предположим, что все ее элементы различны и ни один из них не совпадает с x_* . Наиболее эффективный способ оценивания скорости сходимости состоит в сопоставлении расстояния между $x^{(k+1)}$ и x_* с расстоянием между $x^{(k)}$ и x_* .

Последовательность $\{x^{(k)}\}$ называется *сходящейся с порядком p* , если p — максимальное число, для которого

$$0 \leq \lim_{k \rightarrow \infty} \frac{\|x^{(k+1)} - x_*\|}{\|x^{(k)} - x_*\|^p} < \infty.$$

Поскольку величина p определяется предельными свойствами $\{x^{(k)}\}$, она называется *асимптотической скоростью сходимости*.

Если последовательность $\{x^{(k)}\}$ — сходящаяся с порядком p , то число

$$c = \lim_{k \rightarrow \infty} \frac{\|x^{(k+1)} - x_*\|}{\|x^{(k)} - x_*\|^p}$$

называется *асимптотическим параметром ошибки*.

Если $p = 1$, $c < 1$, то сходимость линейная, если $p = 2$ — квадратичная, если $p = 3$ — кубичная и т. д. Если $p > 1$ или $p = 1$, $c = 0$, то сходимость сверхлинейная. Линейная сходимость является синонимом сходимости со скоростью геометрической прогрессии. Сверхлинейная сходимость является более быстрой, чем определяемая любой геометрической прогрессией.

Теорема 1.3 (о погрешности приближений, вычисляемых методом простых итераций).

Если в итерационном процессе норма матрицы α , согласованная с нормой вектора x , меньше единицы ($\|\alpha\| < 1$), то справедлива следующая оценка погрешности:

$$\|x^{(k)} - x_*\| \leq \frac{\|\alpha\|}{1 - \|\alpha\|} \|x^{(k)} - x^{(k-1)}\|. \quad (1.17)$$

Это соотношение может быть переписано через начальное приближение $x^{(0)} = \beta$:

$$\|x^{(k)} - x_*\| \leq \frac{\|\alpha\|^{k+1}}{1-\|\alpha\|} \|\beta\|. \quad (1.18)$$

Неравенство (1.17) позволяет провести апостериорную оценку погрешности k -го приближения посредством вычисления нормы разности двух последовательных приближений $x^{(k)}$, $x^{(k-1)}$ и нормы матрицы α . На основе этой оценки осуществляется выход из итерационного процесса по результатам расчета (см. (1.14)):

$$\|x^{(k+1)} - x^{(k)}\| \leq \frac{(1-\|\alpha\|)}{\|\alpha\|} \varepsilon.$$

На основе неравенства (1.18) можно записать *aприорную оценку* погрешности (*a priori* — до)

$$\frac{\|\alpha\|^{k+1}}{1-\|\alpha\|} \|\beta\| \leq \varepsilon,$$

из которой еще до расчета можно получить число итераций k , требуемых для достижения заданной точности:

$$k+1 \geq \frac{\lg \varepsilon + \lg(1-\|\alpha\|) - \lg \|\beta\|}{\lg \|\alpha\|}. \quad (1.19)$$

Преобразование системы $Ax = b$ к виду $x = \alpha x + \beta$ с матрицей α , удовлетворяющей условиям сходимости, может быть выполнено несколькими способами. Приведем способы, используемые наиболее часто.

1. Уравнения, входящие в систему $Ax = b$, переставляются так, чтобы выполнялось условие (1.16) преобладания диагональных элементов (для той же цели можно использовать другие элементарные преобразования). Затем первое уравнение разрешается относительно x_1 , второе — относительно x_2 и т. д. При этом получается матрица α с нулевыми диагональными элементами.

Например, система

$$\begin{aligned} -2,8x_1 + x_2 + 4x_3 &= 60, \\ 10x_1 - x_2 + 8x_3 &= 10, \\ -x_1 + 2x_2 - 0,6x_3 &= 20 \end{aligned}$$

с помощью перестановки уравнений приводится к виду

$$\begin{aligned} 10x_1 - x_2 + 8x_3 &= 10, \\ -x_1 + 2x_2 - 0,6x_3 &= 20, \\ -2,8x_1 + x_2 + 4x_3 &= 60, \end{aligned}$$

где $|10| > |-1| + |8|$, $|2| > |-1| + |-0,6|$, $|4| > |-2,8| + |1|$, т. е. диагональные элементы преобладают.

Выражая x_1 из первого уравнения, x_2 — из второго, а x_3 — из третьего, получаем систему:

$$\begin{aligned}x_1 &= 0 \cdot x_1 + 0,1x_2 - 0,8x_3 + 1, \\x_2 &= 0,5x_1 + 0 \cdot x_2 + 0,3x_3 + 10, \\x_3 &= 0,7x_1 - 0,25x_2 + 0 \cdot x_3 + 15,\end{aligned}$$

где $\alpha = \begin{pmatrix} 0 & 0,1 & -0,8 \\ 0,5 & 0 & 0,3 \\ 0,7 & -0,25 & 0 \end{pmatrix}$, $\beta = \begin{pmatrix} 1 \\ 10 \\ 15 \end{pmatrix}$.

Заметим, что $\|\alpha\|_1 = \max\{0,9; 0,8; 0,95\} = 0,95 < 1$, т. е. условие теоремы 1.1 выполнено.

Проиллюстрируем применение других элементарных преобразований. Так, система

$$\begin{aligned}4x_1 + x_2 + 9x_3 &= -7, \\3x_1 + 8x_2 - 7x_3 &= -6, \\x_1 + x_2 - 8x_3 &= 7\end{aligned}$$

путем сложения первого и третьего уравнений и вычитания из второго уравнения третьего уравнения преобразуется к виду

$$\begin{aligned}5x_1 + 2x_2 + x_3 &= 0, \\2x_1 + 7x_2 + x_3 &= -13, \\x_1 + x_2 - 8x_3 &= 7\end{aligned}$$

с преобладанием диагональных элементов.

2. Уравнения преобразуются так, чтобы выполнялось условие преобладания диагональных элементов, но при этом коэффициенты α_{ii} не обязательно равнялись нулю.

Например, систему

$$\begin{aligned}1,02x_1 - 0,15x_2 &= 2,7, \\0,8x_1 + 1,05x_2 &= 4\end{aligned}$$

можно записать в форме

$$\begin{aligned}x_1 &= -0,02x_1 + 0,15x_2 + 2,7, \\x_2 &= -0,8x_1 - 0,05x_2 + 4,\end{aligned}$$

для которой $\|\alpha\|_1 = \max\{0,17; 0,85\} = 0,85 < 1$.

3. Если $\det A \neq 0$, систему $Ax = b$ следует умножить на матрицу $D = A^{-1} - \varepsilon$, где $\{\varepsilon_{ij}\}$ — матрица с малыми по модулю элементами. Тогда получается система $(A^{-1} - \varepsilon)Ax = Db$ или $A^{-1}Ax - \varepsilon Ax = Db$, которую можно записать в форме $x = \alpha x + \beta$, где $\alpha = \varepsilon A$, $\beta = Db$. Если $|\varepsilon_{ij}|$, $i, j = 1, \dots, n$, достаточно малы, условие сходимости выполняется.

Методика решения задачи

1. Преобразовать систему $Ax = b$ к виду $x = \alpha x + \beta$ одним из описанных способов.

2. Задать начальное приближение решения $x^{(0)}$ произвольно или положить $x^{(0)} = \beta$, а также малое положительное число ε (точность). Положить $k = 0$.

3. Вычислить следующее приближение $x^{(k+1)}$ по формуле $x^{(k+1)} = \alpha x^{(k)} + \beta$.

4. Если выполнено условие $\|x^{(k+1)} - x^{(k)}\| < \varepsilon$ или (1.14), процесс завершить и положить $x_* \approx x^{(k+1)}$. Иначе положить $k = k + 1$ и перейти к п. 3.

Пример 1.11. Методом простых итераций с точностью $\varepsilon = 0,01$ решить систему линейных алгебраических уравнений:

$$\begin{aligned} 2x_1 + 2x_2 + 10x_3 &= 14, \\ 10x_1 + x_2 + x_3 &= 12, \\ 2x_1 + 10x_2 + x_3 &= 13. \end{aligned}$$

Предварительно определить число итераций.

□1. Так как $|2| < |2| + |10|$, $|1| < |10| + |1|$, $|1| < |2| + |10|$, условие (1.16) не выполняется. Переставим уравнения местами так, чтобы выполнялось условие преобладания диагональных элементов:

$$\begin{aligned} 10x_1 + x_2 + x_3 &= 12, \\ 2x_1 + 10x_2 + x_3 &= 13, \\ 2x_1 + 2x_2 + 10x_3 &= 14. \end{aligned}$$

Получаем $|10| > |1| + |1|$, $|10| > |2| + |1|$, $|10| > |2| + |2|$. Выразим из первого уравнения x_1 , из второго — x_2 , из третьего — x_3 :

$$\begin{aligned} x_1 &= -0,1 \cdot x_2 - 0,1 \cdot x_3 + 1,2, \\ x_2 &= -0,2 \cdot x_1 - 0,1 \cdot x_3 + 1,3, \\ x_3 &= -0,2 \cdot x_1 - 0,2 \cdot x_2 + 1,4; \end{aligned} \quad \alpha = \begin{pmatrix} 0 & -0,1 & -0,1 \\ -0,2 & 0 & -0,1 \\ -0,2 & -0,2 & 0 \end{pmatrix}; \quad \beta = \begin{pmatrix} 1,2 \\ 1,3 \\ 1,4 \end{pmatrix}.$$

Заметим, что $\|\alpha\|_1 = \max\{0,2; 0,3; 0,4\} = 0,4 < 1$, следовательно, условие сходимости (теорема 1.1) выполнено.

По формуле (1.19) вычисляем число итераций, обеспечивающих заданную точность:

$$k + 1 \geq \frac{-2 + \lg 0,6 - \lg 1,4}{\lg 0,4} = 5,95; \quad k \geq 4,95.$$

Таким образом, для решения задачи необходимо выполнить не менее пяти итераций.

2. Зададим $x^{(0)} = \beta = \begin{pmatrix} 1,2 \\ 1,3 \\ 1,4 \end{pmatrix}$. В поставленной задаче $\varepsilon = 0,01$.

3. Выполним расчеты по формуле (1.13):

$$x^{(k+1)} = \begin{pmatrix} 0 & -0,1 & -0,1 \\ -0,2 & 0 & -0,1 \\ -0,2 & -0,2 & 0 \end{pmatrix} \cdot \begin{pmatrix} x_1^{(k)} \\ x_2^{(k)} \\ x_3^{(k)} \end{pmatrix} + \begin{pmatrix} 1,2 \\ 1,3 \\ 1,4 \end{pmatrix}, \quad k = 0, 1, \dots,$$

или

$$\begin{aligned} x_1^{(k+1)} &= -0,1x_2^{(k)} - 0,1x_3^{(k)} + 1,2; \\ x_2^{(k+1)} &= -0,2x_1^{(k)} - 0,1x_3^{(k)} + 1,3; \quad k = 0, 1, \dots, \\ x_3^{(k+1)} &= -0,2x_1^{(k)} - 0,2x_2^{(k)} + 1,4; \end{aligned}$$

до выполнения условия окончания и результаты занесем в таблицу 1.4.

4. Расчет закончен, поскольку условие окончания $\|x^{(k+1)} - x^{(k)}\| = 0,0027 < \varepsilon$ выполнено.

Приближенное решение задачи: $x_* \approx \begin{pmatrix} 0,9996 & 0,9995 & 0,9993 \end{pmatrix}^T$. Очевидно, точное решение: $x_* = (1; 1; 1)^T$.

На рисунке 1.4 показан характер сходимости $x_3^{(k)}$ ($k = 0, 1, 2, 3, 4, 5$) к $x_{*3} = 1$. Видно, что значения $x_3^{(k)}$ получаются то больше, то меньше единицы, но с увеличением k все меньше отличаются от нее.

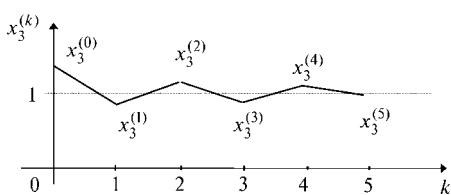


Рис. 1.4

Таблица 1.4

k	$x_1^{(k)}$	$x_2^{(k)}$	$x_3^{(k)}$	$\ x^{(k)} - x^{(k-1)}\ _1$
0	1,2000	1,3000	1,4000	—
1	0,9300	0,9200	0,900	0,5
2	1,0180	1,0240	1,0300	0,13
3	0,9946	0,9934	0,9916	0,0384
4	1,0015	1,0020	1,0024	0,0108
5	0,9996	0,9995	0,9993	0,0027 < ε

Таблица 1.5

k	$x_1^{(k)}$	$x_2^{(k)}$	$x_3^{(k)}$	$\ x^{(k)} - x^{(k-1)}\ _1$
0	1,2000	0	0	—
1	1,2000	1,0600	1,1600	1,1600
2	0,9780	0,9440	0,9480	0,2220
3	1,0108	1,0096	1,0156	0,0676
4	0,9975	0,9963	0,9959	0,0133
5	1,0008	0,0009	1,0012	0,0053
6	0,9998	0,9997	0,9997	0,0015
7	1,0001	1,0001	1,0001	0,0004 < ε

Это соответствует *двусторонней сходимости* метода простых итераций для заданной задачи. В других задачах может проявляться *односторонняя сходимость*, когда стремление к x_{*i} осуществляется с одной стороны от x_{*i} .

Приведем результаты расчетов для другого начального приближения $x^{(0)} = (1,2; 0; 0)^T$ и $\varepsilon = 0,001$ (табл. 1.5).

Приближенное решение задачи: $x_* \approx (1,0001; 1,0001; 1,0001)^T$. ■

1.3.2. МЕТОД ЗЕЙДЕЛЯ

Этот метод является модификацией метода простых итераций и в некоторых случаях приводит к более быстрой сходимости.

Итерации по методу Зейделя отличаются от простых итераций (1.13) тем, что при нахождении i -й компоненты $(k+1)$ -го приближения сразу используются уже найденные компоненты $(k+1)$ -го приближения с меньшими номерами $1, 2, \dots, i-1$. При рассмотрении развернутой формы системы итерационный процесс записывается в виде

$$\begin{aligned} x_1^{(k+1)} &= \alpha_{11}x_1^{(k)} + \alpha_{12}x_2^{(k)} + \alpha_{13}x_3^{(k)} + \dots + \alpha_{1n}x_n^{(k)} + \beta_1, \\ x_2^{(k+1)} &= \alpha_{21} \boxed{x_1^{(k+1)}} + \alpha_{22}x_2^{(k)} + \alpha_{23}x_3^{(k)} + \dots + \alpha_{2n}x_n^{(k)} + \beta_2, \\ x_3^{(k+1)} &= \alpha_{31} \boxed{x_1^{(k+1)}} + \alpha_{32} \boxed{x_2^{(k+1)}} + \alpha_{33}x_3^{(k)} + \dots + \alpha_{3n}x_n^{(k)} + \beta_3, \\ \vdots \\ x_n^{(k+1)} &= \alpha_{n1} \boxed{x_1^{(k+1)}} + \alpha_{n2} \boxed{x_2^{(k+1)}} + \alpha_{n3} \boxed{x_3^{(k+1)}} + \dots + \alpha_{nn-1} \boxed{x_{n-1}^{(k+1)}} + \alpha_{nn}x_n^{(k)} + \beta_n. \end{aligned} \quad (1.20)$$

В каждое последующее уравнение подставляются значения неизвестных, полученных из предыдущих уравнений, что показано в записи (1.20) стрелками.

Теорема 1.4 (о достаточном условии сходимости метода Зейделя).

Если для системы $x = \alpha x + \beta$ какая-либо норма матрицы α меньше единицы, т. е. $\|\alpha\|_s < 1$ ($s \in \{1, 2, 3\}$), то процесс последовательных приближений (1.20) сходится к единственному решению исходной системы $Ax = b$ при любом начальном приближении $x^{(0)}$.

Записывая (1.20) в матричной форме, получаем

$$x^{(k+1)} = Lx^{(k+1)} + Ux^{(k)} + \beta, \quad (1.21)$$

где L, U являются разложениями матрицы α :

$$L = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ \alpha_{21} & 0 & 0 & \cdots & 0 \\ \alpha_{31} & \alpha_{32} & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \alpha_{n1} & \alpha_{n2} & \alpha_{n3} & \cdots & 0 \end{pmatrix}, \quad U = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} & \cdots & \alpha_{1n} \\ 0 & \alpha_{22} & \alpha_{23} & \cdots & \alpha_{2n} \\ 0 & 0 & \alpha_{33} & \cdots & \alpha_{3n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \alpha_{nn} \end{pmatrix}.$$

Преобразуя (1.21) к виду $x = \alpha x + \beta$, получаем матричную форму итерационного процесса метода Зейделя:

$$x^{(k+1)} = (E - L)^{-1}Ux^{(k)} + (E - L)^{-1}\beta. \quad (1.22)$$

Тогда достаточное, а также необходимое и достаточное условия сходимости будут соответственно такими (см. теоремы 1.1 и 1.2):

$$\|\alpha\| = \|(E - L)^{-1}U\| < 1, \quad |\lambda_i(\alpha)| = |\lambda_i((E - L)^{-1}U)| < 1.$$

Замечания

1. Для обеспечения сходимости метода Зейделя требуется преобразовать систему $Ax = b$ к виду $x = \alpha x + \beta$ с преобладанием диагональных элементов в матрице α (см. метод простых итераций).

Например, в системе

$$\begin{aligned} 2x_1 + x_2 &= 2, \\ x_1 - 2x_2 &= -2 \end{aligned}$$

диагональные элементы преобладают, так как $|2| > 1, |-2| > 1$.

Соотношения метода Зейделя (1.20) принимают вид

$$\begin{aligned} x_1^{(k+1)} &= -\frac{x_2^{(k)}}{2} + 1, \\ x_2^{(k+1)} &= \frac{x_1^{(k+1)}}{2} + 1. \end{aligned}$$

Выберем в качестве начального приближения $x^{(0)} = (0; 0)^T$ (рис. 1.5a). Тогда $x_1^{(1)} = -\frac{x_2^{(0)}}{2} + 1 = 1$. Так как при этом $x_2^{(0)} = 0$, то вычислению $x_1^{(1)}$ соответствует движение по горизонтали до пересечения с прямой, описываемой первым уравнением. Далее $x_2^{(1)} = \frac{x_1^{(1)}}{2} + 1 = \frac{3}{2}$. Вычислению $x_2^{(1)}$ соответствует движение по вертикали до пересечения с прямой, описываемой вторым уравнением. Продолжая вычисления, получаем $x_1^{(2)} = -\frac{x_2^{(1)}}{2} + 1 = -\frac{3}{4} + 1 = \frac{1}{4}$, $x_2^{(2)} = \frac{x_1^{(2)}}{2} + 1 = \frac{1}{8} + 1 = \frac{9}{8}$ и т. д. В результате имеем процесс, *сходящийся* к точке $x_* = \left(\frac{2}{5}; \frac{6}{5}\right)^T$.

Переставим уравнения в системе:

$$\begin{aligned} x_1 - 2x_2 &= -2, \\ 2x_1 + x_2 &= 2. \end{aligned}$$

В полученной системе диагональные элементы не преобладают. Уравнения метода Зейделя имеют вид:

$$\begin{aligned} x_1^{(k+1)} &= 2x_2^{(k)} - 2, \\ x_2^{(k+1)} &= -2x_1^{(k+1)} + 2. \end{aligned}$$

При $x^{(0)} = (0; 0)^T$ получаем $x_1^{(1)} = -2$, $x_2^{(1)} = 6$ и т. д. В результате имеем *расходящийся процесс* (рис. 1.5б).

2. Условие преобладания диагональных элементов является достаточным для сходимости, но не является необходимым.

Например, в системе

$$x_1 + 2x_2 = 3,$$

$$x_1 - 4x_2 = -3$$

в первом уравнении диагональный элемент не является преобладающим, а процесс итераций по методу Зейделя сходится (рис. 1.6).

3. Процесс (1.20) называется *последовательным итерированием*, так как на каждой итерации полученные из предыдущих уравнений значения подставляются в последующие. Как правило, метод Зейделя обеспечивает лучшую

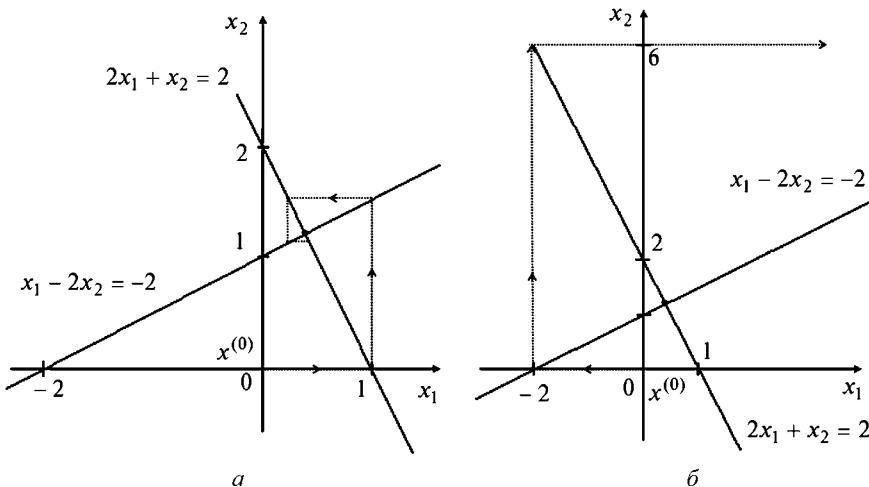


Рис. 1.5

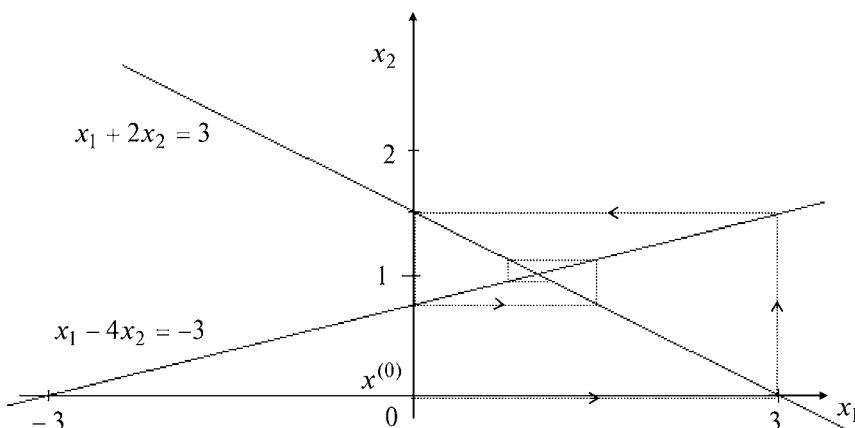


Рис. 1.6

сходимость, чем метод простых итераций (за счет накопления информации, полученной при решении предыдущих уравнений). Метод Зейделя может сходиться, если расходится метод простых итераций, и наоборот.

4. При расчетах на ЭВМ удобнее пользоваться формулой (1.22).

5. Преимуществом метода Зейделя, как и метода простых итераций, является его *самоисправляемость*.

6. Метод Зейделя имеет преимущества перед методом простых итераций, так как он всегда сходится для *нормальных* систем линейных алгебраических уравнений, т. е. таких систем, в которых матрица A является симметрической и положительно определенной. Систему линейных алгебраических уравнений с невырожденной матрицей A всегда можно преобразовать к нормальной, если ее умножить слева на матрицу A^T . Таким образом, система $A^T A x = A^T b$ является нормальной, а матрица $A^T A$ — симметрической.

Методика решения задачи

1. Преобразовать систему $Ax = b$ к виду $x = \alpha x + \beta$ одним из описанных способов.

2. Задать начальное приближение решения $x^{(0)}$ произвольно или положить $x^{(0)} = \beta$, а также малое положительное число ε (точность). Положить $k = 0$.

3. Произвести расчеты по формуле (1.20) или (1.21) и найти $x^{(k+1)}$.

4. Если выполнено условие окончания $\|x^{(k+1)} - x^{(k)}\| < \varepsilon$, процесс завершить и положить $x_* \cong x^{(k+1)}$. Иначе положить $k = k + 1$ и перейти к п. 3.

Пример 1.12. Методом Зейделя с точностью $\varepsilon = 0,001$ решить систему линейных алгебраических уравнений:

$$\begin{aligned} 2x_1 + 2x_2 + 10x_3 &= 14, \\ 10x_1 + x_2 + x_3 &= 12, \\ 2x_1 + 10x_2 + x_3 &= 13. \end{aligned}$$

□1. Приведем систему $Ax = b$ к виду $x = \alpha x + \beta$ так же, как в примере 1.11:

$$\begin{aligned} x_1 &= -0,1 \cdot x_2 - 0,1 \cdot x_3 + 1,2, \\ x_2 &= -0,2 \cdot x_1 - 0,1 \cdot x_3 + 1,3, \\ x_3 &= -0,2 \cdot x_1 - 0,2 \cdot x_2 + 1,4; \end{aligned} \quad \alpha = \begin{pmatrix} 0 & -0,1 & -0,1 \\ -0,2 & 0 & -0,1 \\ -0,2 & -0,2 & 0 \end{pmatrix}; \quad \beta = \begin{pmatrix} 1,2 \\ 1,3 \\ 1,4 \end{pmatrix}.$$

Так как $\|\alpha\|_1 = \max\{0,2; 0,3; 0,4\} = 0,4 < 1$, условие сходимости выполняется.

2. Зададим $x^{(0)} = (1,2; 0; 0)^T$. В поставленной задаче $\varepsilon = 0,001$.

Таблица 1.6

k	$x_1^{(k)}$	$x_2^{(k)}$	$x_3^{(k)}$	$\ x^{(k)} - x^{(k-1)}\ _1$
0	1,2000	0	0	—
1	1,2000	1,0600	0,9480	1,0600
2	0,9992	1,0054	0,9991	0,1008
3	0,9996	1,0002	1,0000	0,0052
4	1,0000	1,0000	1,0000	0,0004 < ε

3. Выполним расчеты по формуле (1.20):

$$\begin{aligned}x_1^{(k+1)} &= -0,1x_2^{(k)} - 0,1x_3^{(k)} + 1,2; \\x_2^{(k+1)} &= -0,2x_1^{(k+1)} - 0,1x_3^{(k)} + 1,3; \quad k = 0,1,\dots, \\x_3^{(k+1)} &= -0,2x_1^{(k+1)} - 0,2x_2^{(k+1)} + 1,4;\end{aligned}$$

и результаты занесем в таблицу 1.6.

Очевидно, найденное решение $x_* = (1; 1; 1)^T$ является точным.

4. Расчет завершен, поскольку условие окончания $\|x^{(k+1)} - x^{(k)}\| = 0,0004 < \varepsilon$ выполнено. ■

Пример 1.13. Методом Зейделя с точностью $\varepsilon = 0,005$ решить систему линейных алгебраических уравнений:

$$\begin{aligned}4x_1 - x_2 + x_3 &= 4, \\x_1 + 6x_2 + 2x_3 &= 9, \\-x_1 - 2x_2 + 5x_3 &= 2.\end{aligned}$$

□1. Так как $|4| > |-1| + |1|$, $|6| > |1| + |2|$, $|5| > |-1| + |-2|$, в данной системе диагональные элементы преобладают. Выразим из первого уравнения x_1 , из второго — x_2 , из третьего — x_3 :

$$\begin{aligned}x_1 &= \frac{1}{4}x_2 - \frac{1}{4}x_3 + 1, \\x_2 &= -\frac{1}{6}x_1 - \frac{1}{3}x_3 + \frac{3}{2}, \quad \alpha = \begin{pmatrix} 0 & \frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{6} & 0 & -\frac{1}{3} \\ \frac{1}{5} & \frac{2}{5} & 0 \end{pmatrix}; \quad \beta = \begin{pmatrix} 1 \\ \frac{3}{2} \\ \frac{2}{5} \end{pmatrix}.\end{aligned}$$

2. Зададим $x^{(0)} = (0; 0; 0)^T$. В поставленной задаче $\varepsilon = 0,005$.

3. Выполним расчеты по формулам (1.20):

$$\begin{aligned}x_1^{(k+1)} &= \frac{1}{4}x_2^{(k)} - \frac{1}{4}x_3^{(k)} + 1; \\x_2^{(k+1)} &= -\frac{1}{6}x_1^{(k+1)} - \frac{1}{3}x_3^{(k)} + \frac{3}{2}; \quad k = 0,1,\dots, \\x_3^{(k+1)} &= \frac{1}{5}x_1^{(k+1)} + \frac{2}{5}x_2^{(k+1)} + \frac{2}{5};\end{aligned}$$

и результаты занесем в таблицу 1.7.

Таблица 1.7

k	$x_1^{(k)}$	$x_2^{(k)}$	$x_3^{(k)}$	$\ x^{(k)} - x^{(k-1)}\ _1$
0	0	0	0	—
1	1,0000	1,3333	1,1333	1,3333
2	1,0500	0,9473	0,9889	0,3860
3	0,9896	1,0050	0,9999	0,0604
4	1,0010	0,9999	1,0000	0,0114
5	1,0000	1,0000	1,0000	0,0010 < ε

Очевидно, найденное решение $x_* = (1; 1; 1)^T$ является точным.

4. Расчет завершен, поскольку условие окончания $\|x^{(k+1)} - x^{(k)}\| = 0,001 < \varepsilon$ выполнено. ■

Пример 1.14. Для системы линейных алгебраических уравнений

$$x_1 + 2x_2 = 3,$$

$$3x_1 + 4x_2 = 7$$

построить сходящийся итерационный процесс.

□1. Матрица $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ этой системы не имеет диагонального преобразования. Достаточное условие сходимости метода простых итераций $\|\alpha\| < 1$ не выполняется. Поэтому сначала проведем нормализацию исходной СЛАУ с помощью умножения слева на матрицу A^T . Проделав умножение

$$\begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} 3 \\ 7 \end{pmatrix},$$

получим:

$$10x_1 + 14x_2 = 24,$$

$$14x_1 + 20x_2 = 34.$$

Полученная матрица $\hat{A} = \begin{pmatrix} 10 & 14 \\ 14 & 20 \end{pmatrix}$ является симметрической и положительно определенной, так как все ее угловые миноры строго положительны: $\Delta_1 = 10 > 0$, $\Delta_2 = |A| = 200 - 14^2 = 4 > 0$. Поэтому система является нормальной, а для нее метод Зейделя сходится.

2. Положим $x^{(0)} = (2,4; 1,02)^T$.

3. Расчеты произведем по формулам (1.20):

$$x_1^{(k+1)} = -1,4x_2^{(k)} + 2,4,$$

$$x_2^{(k+1)} = -0,7x_1^{(k+1)} + 1,7.$$

Таким образом,

$$L = \begin{pmatrix} 0 & 0 \\ -0,7 & 0 \end{pmatrix}, \quad U = \begin{pmatrix} 0 & -1,4 \\ 0 & 0 \end{pmatrix}, \quad \beta = \begin{pmatrix} 2,4 \\ 1,7 \end{pmatrix}.$$

Последовательные приближения, получаемые методом Зейделя, приведены в таблице 1.8.

Таблица 1.8

k	$x_1^{(k)}$	$x_2^{(k)}$	$\ x^{(k)} - x^{(k-1)}\ $
0	2,40000	1,02000	—
1	0,97200	1,01960	1,42800
2	0,97256	1,019208	0,00056
3	0,97311	1,018823	0,00055
4	0,97365	1,018445	0,00054
5	0,97418	1,018076	0,00053

Из таблицы 1.8 видно, что норма разности между двумя последовательными приближениями с увеличением номера итерации уменьшается, что характеризует сходящийся процесс (этот вывод относится ко всем приведенным примерам). Видно также, что, начиная со 2-й итерации, $x_1^{(k)}$ и $x_2^{(k)}$ с увеличением k увеличиваются и уменьшаются соответственно и стремятся к верному (точному) решению: $x_{1*} = 1; x_{2*} = 1$. ■

Пример 1.15. Решить систему

$$\begin{aligned}x_1 - x_2 + x_3 - 4x_4 &= -2, \\2x_1 + x_2 - 5x_3 + x_4 &= 2, \\8x_1 - x_2 - x_3 + 2x_4 &= 11, \\x_1 + 6x_2 - 2x_3 - 2x_4 &= -7\end{aligned}$$

методами простых итераций и Зейделя.

□1. Переставляя уравнения местами, преобразуем систему к виду с преобладанием диагональных элементов:

$$\begin{aligned}8x_1 - x_2 - x_3 + 2x_4 &= 11, \\x_1 + 6x_2 - 2x_3 - 2x_4 &= -7, \\2x_1 + x_2 - 5x_3 + x_4 &= 2, \\x_1 - x_2 + x_3 - 4x_4 &= -2.\end{aligned}$$

Так как $|8| > |-1| + |-1| + |2|, |6| > |1| + |-2| + |-2|, |-5| > |2| + |1| + |1|, |-4| > |1| + + |-1| + |1|$, то условие преобладания диагональных элементов выполняется.

2. Выразим из первого уравнения x_1 , из второго — x_2 , из третьего — x_3 , а из четвертого — x_4 :

$$\begin{aligned}x_1 &= \frac{1}{8}x_2 + \frac{1}{8}x_3 - \frac{1}{4}x_4 + \frac{11}{8}, \\x_2 &= -\frac{1}{6}x_1 + \frac{1}{3}x_3 + \frac{1}{3}x_4 - \frac{7}{6}, \\x_3 &= \frac{2}{5}x_1 + \frac{1}{5}x_2 + \frac{1}{5}x_4 - \frac{2}{5}, \\x_4 &= \frac{1}{4}x_1 - \frac{1}{4}x_2 + \frac{1}{4}x_3 + \frac{1}{2}.\end{aligned}$$

3. Зададим начальное приближение $x^{(0)} = (0; 0; 1; 2)^T$ и $\varepsilon = 0,001$.

Расчеты методом простых итераций выполним по следующим формулам (результаты приведены в таблице 1.9, где $\Delta = \|x^{(k)} - x^{k-1}\|_1$):

$$\begin{aligned}x_1^{(k+1)} &= \frac{1}{8}x_2^{(k)} + \frac{1}{8}x_3^{(k)} - \frac{1}{4}x_4^{(k)} + \frac{11}{8}, \\x_2^{(k+1)} &= -\frac{1}{6}x_1^{(k)} + \frac{1}{3}x_3^{(k)} + \frac{1}{3}x_4^{(k)} - \frac{7}{6}, \\x_3^{(k+1)} &= \frac{2}{5}x_1^{(k)} + \frac{1}{5}x_2^{(k)} + \frac{1}{5}x_4^{(k)} - \frac{2}{5}, \\x_4^{(k+1)} &= \frac{1}{4}x_1^{(k)} - \frac{1}{4}x_2^{(k)} + \frac{1}{4}x_3^{(k)} + \frac{1}{2}.\end{aligned}$$

k	0	1	2	3	4	5	6	7	8	9	10
x_1	0	1	1,1667	1,0562	0,9708	0,9989	1,0045	1,0018	0,9992	0,99994	1,0001
x_2	0	-0,1667	-1,0833	-1,0583	-0,9760	-0,9752	-1,0021	-1,0019	-0,9995	-0,9993	-1,0000
x_3	1	0	0,1166	0,0083	0,02916	-0,0007	0,003	0,00006	0,0008	-0,00004	0,00009
x_4	2	0,75	0,7916	1,0916	1,0307	0,9940	0,9933	1,0025	1,0009	0,99986	0,99979
Δ	—	1,25	0,9166	0,3	0,0854	0,0367	0,0269	0,009	0,0024	0,00104	0,0007

Расчеты методом Зейделя выполним по формулам (результаты приведены в табл. 1.10):

$$\begin{aligned}x_1^{(k+1)} &= \frac{1}{8}x_2^{(k)} + \frac{1}{8}x_3^{(k)} - \frac{1}{4}x_4^{(k)} + \frac{11}{8}, \\x_2^{(k+1)} &= -\frac{1}{6}x_1^{(k+1)} + \frac{1}{3}x_3^{(k)} + \frac{1}{3}x_4^{(k)} - \frac{7}{6}, \\x_3^{(k+1)} &= \frac{2}{5}x_1^{(k+1)} + \frac{1}{5}x_2^{(k+1)} + \frac{1}{5}x_4^{(k)} - \frac{2}{5}, \\x_4^{(k+1)} &= \frac{1}{4}x_1^{(k+1)} - \frac{1}{4}x_2^{(k+1)} + \frac{1}{4}x_3^{(k+1)} + \frac{1}{2}.\end{aligned}$$

Таблица 1.10

k	0	1	2	3	4	5	6	7
x_1	0	1	1,1458	1,0052	1,0059	1,0004	1,00025	1,0002
x_2	0	-0,3333	-0,9409	-0,9713	-0,9967	-0,9988	-0,9998	-0,9999
x_3	1	0,3333	0,0534	0,0148	0,0026	0,00067	0,000125	0,00003
x_4	2	0,9166	1,0351	0,9978	1,0013	0,99995	1,00005	0,999999
Δ	—	1,083	0,6075	0,1406	0,0254	0,0055	0,0011	0,00023

Методом простых итераций получено приближенное решение $x_* \approx (1,0001; -1,0000; 0,00009; 0,99979)^T$, а методом Зейделя решение: $x_* \approx (1,0002; -0,9999; 0,00003; 0,99999)^T$. Точное решение: $x_* = (1; -1; 0; 1)^T$. Очевидно, метод Зейделя сорвался за меньшее число итераций. ■

Задачи для самостоятельного решения

1. Найти численное решение системы, характеризующихся расширенными матрицами A_1 , методом Гаусса единственного деления:

$$\begin{array}{lll}1) \begin{pmatrix} 5 & 0 & 1 & 11 \\ 1 & 3 & -1 & 4 \\ -3 & 2 & 10 & 6 \end{pmatrix}; & 2) \begin{pmatrix} 2 & 0 & -1 & -3 \\ -1 & 3 & 1 & 2 \\ 1 & -1 & 4 & 3 \end{pmatrix}; & 3) \begin{pmatrix} 2 & 0 & -1 & 1 \\ 1 & -3 & 1 & 2 \\ 1 & 1 & 3 & 4 \end{pmatrix}; \\4) \begin{pmatrix} 5 & 1 & -1 & -5 \\ -1 & 3 & 1 & 5 \\ 1 & -2 & 4 & 1 \end{pmatrix}; & 5) \begin{pmatrix} 3 & 1 & -1 & 1 \\ -2 & 4 & 1 & 5 \\ 1 & 1 & 3 & -3 \end{pmatrix}; & 6) \begin{pmatrix} 3 & 1 & -1 & 6 \\ 2 & 4 & 1 & 9 \\ 1 & -1 & 3 & 4 \end{pmatrix};\end{array}$$

$$\begin{array}{lll}
7) \begin{pmatrix} 2 & -1 & 0 & -2 \\ 2 & 5 & -2 & -4 \\ 1 & -1 & 3 & 2 \end{pmatrix}; &
8) \begin{pmatrix} 3 & -1 & 1 & 1 \\ 0 & 2 & -1 & 3 \\ -1 & 1 & 5 & -5 \end{pmatrix}; &
9) \begin{pmatrix} 4 & 1 & -1 & 7 \\ 2 & 3 & 0 & 7 \\ 1 & -1 & 5 & 11 \end{pmatrix}; \\
10) \begin{pmatrix} 2 & 0 & -1 & 1 \\ 1 & -4 & 2 & -5 \\ 1 & 1 & 3 & 6 \end{pmatrix}; &
11) \begin{pmatrix} 2 & -1 & 0 & 3 \\ 0 & 5 & 2 & 7 \\ 1 & -1 & 3 & 4 \end{pmatrix}; &
12) \begin{pmatrix} 3 & 1 & -1 & -2 \\ 3 & 5 & -1 & 8 \\ 2 & 0 & 3 & 1 \end{pmatrix}; \\
13) \begin{pmatrix} 3 & 0 & -1 & -4 \\ 2 & -5 & 1 & 9 \\ 2 & -2 & 6 & 8 \end{pmatrix}; &
14) \begin{pmatrix} 4 & 1 & 1 & 9 \\ 1 & 5 & 1 & 14 \\ 2 & 2 & 5 & 21 \end{pmatrix}; &
15) \begin{pmatrix} 3 & 1 & 1 & 6 \\ 1 & 4 & 2 & 9 \\ 2 & 2 & 5 & 14 \end{pmatrix}; \\
16) \begin{pmatrix} 2 & 1 & 0 & 4 \\ 1 & 3 & 1 & 9 \\ 1 & 1 & 4 & 11 \end{pmatrix}; &
17) \begin{pmatrix} 3 & 1 & 0,5 & 7,5 \\ 1 & 3 & 1 & 6 \\ 1 & 1 & 4 & 7 \end{pmatrix}; &
18) \begin{pmatrix} 3 & 4 & 5 & 26 \\ 2 & 1 & 1 & 7 \\ 4 & 3 & 7 & 31 \end{pmatrix}; \\
19) \begin{pmatrix} 5 & 1 & 2 & 12 \\ 2 & 4 & 1 & 16 \\ 1 & 1 & 4 & 12 \end{pmatrix}; &
20) \begin{pmatrix} 3 & 1 & 5 & 21 \\ 2 & 1 & 1 & 7 \\ 2 & 3 & 7 & 25 \end{pmatrix}; &
21) \begin{pmatrix} 4 & 1 & 1 & 9 \\ 1 & 5 & 1 & 14 \\ 2 & 2 & 5 & 21 \end{pmatrix}.
\end{array}$$

2. Данна система с трехдиагональной матрицей A :

$$\begin{aligned}
2x_1 + x_2 &= \frac{5}{2} + \frac{1}{3} \cdot N, \\
x_1 + 3x_2 + x_3 &= \frac{17}{6} + \frac{1}{2} \cdot N, \\
-2x_2 + 3x_3 + x_4 &= 1, \\
x_3 - 2x_4 &= -\frac{5}{3} + \frac{1}{3} \cdot N,
\end{aligned}$$

где N — параметр, соответствующий номеру по списку группы.

Требуется:

- а) убедиться, что матрица A этой системы, если ее представить в виде $Ax = b$, трехдиагональная и удовлетворяет условию преобладания диагональных элементов;
- б) решить систему методом прогонки;
- в) подстановкой полученного решения в исходную систему убедиться в правильности результата.

3. Методом простых итераций решить систему линейных алгебраических уравнений $Ax = b$, где

$$A = \begin{pmatrix} 5 & 1 & 1 \\ 2 & 6 & -1 \\ 2 & 3 & 10 \end{pmatrix}, \quad b = \begin{pmatrix} 11 + 0,8 \cdot N \\ 13 + 0,9 \cdot N \\ 18 - 0,5 \cdot N \end{pmatrix}, \quad N — \text{параметр},$$

с точностью $\varepsilon = 0,001$, предварительно вычислив количество итераций по формуле (1.19).

4. Сравнить количество итераций, требующихся для получения решения системы $Ax = b$, где

$$A = \begin{pmatrix} 12 & -3 & 2 & -1 \\ -1 & 6 & -1 & 1 \\ 3 & 2 & -8 & 2 \\ 2 & -1 & -1 & 5 \end{pmatrix}, \quad b = \begin{pmatrix} 8 \\ 12 \\ -9 \\ 17 \end{pmatrix},$$

методом простых итераций и методом Зейделя с точностью $\varepsilon = 0,001$, прекратив итерационный процесс апостериорным путем.

5. Методом Зейделя с точностью $\varepsilon = 0,001$ найти решение двух систем линейных алгебраических уравнений, преобразуя их предварительно к виду, обеспечивающему сходимость:

$$A_1 = \begin{pmatrix} 2,7 & 9,3 & 1,3 \\ 3,5 & 1,7 & 2,8 \\ 4,1 & 5,8 & -1,7 \end{pmatrix}, \quad b_1 = \begin{pmatrix} 2,1 \\ 1,7 \\ 0,8 \end{pmatrix}; \quad A_2 = \begin{pmatrix} 1,7 & -2,8 & 1,9 \\ 2,1 & 3,4 & 1,8 \\ 4,2 & -1,7 & 1,3 \end{pmatrix}, \quad b_2 = \begin{pmatrix} 0,7 \\ 1,1 \\ 2,8 \end{pmatrix}.$$

ГЛАВА 2

МЕТОДЫ РЕШЕНИЯ ЗАДАЧ О СОБСТВЕННЫХ ЗНАЧЕНИЯХ И СОБСТВЕННЫХ ВЕКТОРАХ МАТРИЦ

2.1. ПОСТАНОВКА ЗАДАЧИ

Пусть A — действительная числовая квадратная матрица размеров $(n \times n)$. Ненулевой вектор $X = (x_1, \dots, x_n)^T$ размеров $(n \times 1)$, удовлетворяющий условию

$$AX = \lambda X, \quad (2.1)$$

называется *собственным вектором* матрицы A . Число λ в равенстве (2.1) называется *собственным значением*. Говорят, что собственный вектор X соответствует (принадлежит) собственному значению λ .

Равенство (2.1) равносильно однородной относительно X системе:

$$(A - \lambda E)X = 0, \quad (X \neq 0). \quad (2.2)$$

Система (2.2) имеет ненулевое решение для вектора X (при известном λ) при условии $|A - \lambda E| = 0$. Это равенство есть *характеристическое уравнение*:

$$|A - \lambda E| = P_n(\lambda) = 0, \quad (2.3)$$

где $P_n(\lambda)$ — *характеристический многочлен* n -й степени. Корни $\lambda_1, \lambda_2, \dots, \lambda_n$ *характеристического уравнения* (2.3) являются *собственными (характеристическими) значениями* матрицы A , а соответствующие каждому собственному значению $\lambda_i, i = 1, \dots, n$, ненулевые векторы X^i , удовлетворяющие системе

$$AX^i = \lambda_i X^i \text{ или } (A - \lambda_i E)X^i = 0, \quad i = 1, 2, \dots, n, \quad (2.4)$$

являются *собственными векторами*.

Требуется найти собственные значения и собственные векторы заданной матрицы. Поставленная задача часто именуется *второй задачей линейной алгебры*.

Проблема собственных значений (частот) возникает при анализе поведения мостов, зданий, летательных аппаратов и других конструкций, характеризующихся малыми смещениями от положения равновесия, а также при анализе устойчивости численных схем. Характеристическое уравнение вме-

сте с его собственными значениями и собственными векторами является основным в теории механических или электрических колебаний на макроскопическом или микроскопическом уровнях.

Различают полную и частичную проблему собственных значений, когда необходимо найти весь спектр (все собственные значения) и собственные векторы либо часть спектра, например: $\rho(A) = \max_i |\lambda_i(A)|$ и $\min_i |\lambda_i(A)|$. Величина $\rho(A)$ называется *спектральным радиусом*.

Замечания

- Если для собственного значения λ_i найден собственный вектор X^i , то вектор μX^i , где μ — произвольное число, также является собственным вектором, соответствующим этому же собственному значению λ_i .

- Попарно различным собственным значениям соответствуют линейно независимые собственные векторы; k -кратному корню характеристического уравнения соответствует не более k линейно независимых собственных векторов.

- Симметрическая матрица имеет полный спектр λ_i , $i = \overline{1, n}$, действительных собственных значений; k -кратному корню характеристического уравнения симметрической матрицы соответствует ровно k линейно независимых собственных векторов.

- Положительно определенная симметрическая матрица имеет полный спектр действительных положительных собственных значений.

2.2. МЕТОД НЕПОСРЕДСТВЕННОГО РАЗВЕРТЫВАНИЯ

Полную проблему собственных значений для матриц невысокого порядка ($n \leq 10$) можно решить методом непосредственного развертывания. В этом случае будем иметь

$$|A - \lambda E| = \begin{vmatrix} a_{11} - \lambda & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} - \lambda & a_{23} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nn} - \lambda \end{vmatrix} = P_n(\lambda) = 0. \quad (2.5)$$

Уравнение $P_n(\lambda) = 0$ является нелинейным (методы его решения изложены в главе 3). Его решение дает n , вообще говоря, комплексных собственных значений $\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_n$, при которых $P_n(\lambda_i) = 0$, $i = \overline{1, n}$. Для каждого λ_i может быть найдено решение однородной системы $(A - \lambda_i E)X^i = 0$, $i = \overline{1, n}$. Эти решения X^i , определенные с точностью до произвольной константы, образуют систему n , вообще говоря, различных векторов n -мерного пространства. В некоторых задачах несколько этих векторов (или все) могут совпадать.

Методика решения задачи

- Для заданной матрицы A составить характеристическое уравнение (2.5):

$$|A - \lambda E| = 0.$$

Для развертывания детерминанта $|A - \lambda E|$ можно использовать различные методы, например метод Крылова, метод Данилевского или другие [2], [9], [14].

2. Решить характеристическое уравнение и найти собственные значения $\lambda_1, \lambda_2, \dots, \lambda_n$. Для этого можно применить методы, изложенные в п. 3.1.

3. Для каждого собственного значения составить систему (2.4):

$$(A - \lambda_i E)X^i = 0, \quad i = 1, 2, \dots, n,$$

и найти собственные векторы X^i .

Замечание. Каждому собственному значению соответствует один или несколько векторов. Поскольку определитель $|A - \lambda_i E|$ системы равен нулю, то ранг матрицы системы меньше числа неизвестных: $\text{rang}(A - \lambda_i E) = r < n$ и в системе имеется ровно r независимых уравнений, а $(n - r)$ уравнений являются зависимыми. Для нахождения решения системы следует выбрать r уравнений с r неизвестными так, чтобы определитель составленной системы был отличен от нуля. Остальные $(n - r)$ неизвестных следует перенести в правую часть и считать параметрами. Придавая параметрам различные значения, можно получить различные решения системы. Для простоты, как правило, попаременно полагают значение одного параметра равным 1, а остальные равными 0.

Пример 2.1. Найти собственные значения и собственные векторы матрицы $A \in R^{2 \times 2}$, где $A = \begin{pmatrix} 3 & -2 \\ -4 & 1 \end{pmatrix}$.

□ Воспользуемся методикой.

1. Запишем уравнение (2.5):

$$|A - \lambda E| = \begin{vmatrix} 3 - \lambda & -2 \\ -4 & 1 - \lambda \end{vmatrix} = \lambda^2 - 4\lambda - 5 = 0,$$

отсюда получаем характеристическое уравнение

$$P_2(\lambda) \equiv \lambda^2 - 4\lambda - 5 = 0.$$

2. Находим его корни (собственные значения): $\lambda_1 = 5, \lambda_2 = -1$.

3. Составим систему $(A - \lambda_i E)X^i = 0, i = 1, 2$, для каждого собственного значения и найдем собственные векторы:

$$\begin{pmatrix} 3 - \lambda_1 & -2 \\ -4 & 1 - \lambda_1 \end{pmatrix} \cdot \begin{pmatrix} x_1^1 \\ x_2^1 \end{pmatrix} = 0 \quad \text{или} \quad \begin{cases} -2x_1^1 - 2x_2^1 = 0, \\ -4x_1^1 - 4x_2^1 = 0. \end{cases}$$

Отсюда $x_1^1 = -x_2^1$. Если $x_2^1 = \mu$, то $x_1^1 = -\mu$. В результате получаем

$$X^1 = \{x_1^1, x_2^1\}^T = \{\mu(-1; 1)\}^T.$$

Для $\lambda_2 = -1$ имеем

$$\begin{pmatrix} 3 - \lambda_2 & -2 \\ -4 & 1 - \lambda_2 \end{pmatrix} \cdot \begin{pmatrix} x_1^2 \\ x_2^2 \end{pmatrix} = 0 \quad \text{или} \quad \begin{cases} 4x_1^2 - 2x_2^2 = 0, \\ -4x_1^2 + 2x_2^2 = 0. \end{cases}$$

Отсюда $x_2^2 = 2x_1^2$. Если $x_1^2 = \mu$, то $x_2^2 = 2\mu$. В результате получаем

$$X^2 = \{x_1^2, x_2^2\}^T = \{\mu(1; 2)\}^T,$$

где μ — произвольное действительное число. ■

Пример 2.2. Найти собственные значения и собственные векторы матрицы

$$A = \begin{pmatrix} 2 & -1 & 1 \\ -1 & 2 & -1 \\ 0 & 0 & 1 \end{pmatrix}.$$

□ Воспользуемся методикой.

1. Запишем характеристическое уравнение (2.5):

$$\begin{vmatrix} 2-\lambda & -1 & 1 \\ -1 & 2-\lambda & -1 \\ 0 & 0 & 1-\lambda \end{vmatrix} = 0 \text{ или } (1-\lambda)[(2-\lambda)^2 - 1] = 0.$$

2. Корни характеристического уравнения: $\lambda_{1,2} = 1$ (кратный корень), $\lambda_3 = 3$ — собственные значения матрицы.

3. Найдем собственные векторы.

Для $\lambda_{1,2} = 1$ запишем систему $(A - \lambda_{1,2}E) \cdot X^{1,2} = 0$:

$$\begin{pmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} x_1^{1,2} \\ x_2^{1,2} \\ x_3^{1,2} \end{pmatrix} = 0.$$

Поскольку $\text{rang}(A - \lambda_{1,2}E) = 1$, в системе имеется одно независимое уравнение $x_1^{1,2} - x_2^{1,2} + x_3^{1,2} = 0$ или $x_1^{1,2} = x_2^{1,2} - x_3^{1,2}$.

Полагая $x_2^{1,2} = 1$, $x_3^{1,2} = 0$, получаем $x_1^{1,2} = 1$ и собственный вектор $X^1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$.

Полагая $x_2^{1,2} = 0$, $x_3^{1,2} = 1$, получаем $x_1^{1,2} = -1$ и другой собственный вектор $X^2 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$. Заметим, что оба собственных вектора линейно независимы.

Для собственного значения $\lambda_3 = 3$ запишем систему $(A - \lambda_3 E) \cdot X^3 = 0$:

$$\begin{pmatrix} -1 & -1 & 1 \\ -1 & -1 & -1 \\ 0 & 0 & -2 \end{pmatrix} \cdot \begin{pmatrix} x_1^3 \\ x_2^3 \\ x_3^3 \end{pmatrix} = 0.$$

Поскольку $\text{rang}(A - \lambda_3 E) = 2$, то выбираем два уравнения:

$$-x_1^3 - x_2^3 + x_3^3 = 0, \quad -2x_3^3 = 0.$$

Отсюда $x_3^3 = 0$, $x_1^3 = -x_2^3$. Полагая $x_2^3 = 1$, получаем $x_1^3 = -1$ и собственный вектор $X^3 = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$. ■

2.3. МЕТОД ИТЕРАЦИЙ

Для решения частичной проблемы собственных значений и собственных векторов в практических расчетах часто используется метод итераций (*степенной метод*). На его основе можно определить приближенно собственные значения матрицы A и *спектральный радиус* $\rho(A) = \max|\lambda_i(A)|$.

Пусть матрица A имеет n линейно независимых собственных векторов X^i , $i = 1, n$, и собственные значения матрицы A таковы, что:

$$\rho(A) = |\lambda_1(A)| > |\lambda_2(A)| \geq \dots \geq |\lambda_n(A)|.$$

Методика решения задачи

1. Выбрать произвольное начальное (нулевое) приближение собственного вектора $X^{1(0)}$ (второй индекс в скобках здесь и ниже указывает номер приближения, а первый индекс без скобок соответствует номеру собственного значения). Положить $k = 0$.

2. Найти $X^{1(1)} = AX^{1(0)}$, $\lambda_1^{(1)} = \frac{x_i^{1(1)}}{x_i^{1(0)}}$, где i — любой номер $1 \leq i \leq n$, и положить $k = 1$.

3. Вычислить $X^{1(k+1)} = AX^{1(k)}$.

4. Найти $\lambda_1^{(k+1)} = \frac{x_i^{1(k+1)}}{x_i^{1(k)}}$, где $x_i^{1(k+1)}$, $x_i^{1(k)}$ — соответствующие координаты векторов $X^{1(k+1)}$ и $X^{1(k)}$. При этом может быть использована любая координата с номером i , $i = 1, n$.

5. Если $\Delta = |\lambda_1^{(k+1)} - \lambda_1^{(k)}| \leq \varepsilon$, процесс завершить и положить $\lambda_1 \cong \lambda_1^{(k+1)}$. Если $\Delta > \varepsilon$, положить $k = k + 1$ и перейти к п. 3.

Замечания

1. Процесс последовательных приближений

$$X^{1(1)} = AX^{1(0)}, X^{1(2)} = AX^{1(1)} = A^2X^{1(0)}, \dots, \\ X^{1(k)} = AX^{1(k-1)} = A \cdot A^{k-1}X^{1(0)} = A^kX^{1(0)}, \dots$$

сходится, т. е. при $k \rightarrow \infty$ вектор $X^{1(k)}$ стремится к собственному вектору X^1 . Действительно, разложим $X^{1(0)}$ по всем собственным векторам:

$$X^{1(0)} = \sum_{i=1}^n c_i X^i.$$

Так как, согласно (2.4), $AX^i = \lambda_i X^i$, то

$$AX^{1(0)} = X^{1(1)} = \sum_{i=1}^n c_i \lambda_i X^i; \quad AX^{1(1)} = A^2X^{1(0)} = X^{1(2)} = \sum_{i=1}^n c_i \lambda_i^2 X^i, \dots, \\ A^k X^{1(0)} = X^{1(k)} = \sum_{i=1}^n c_i \lambda_i^k X^i = \lambda_1^k \left[c_1 X^1 + c_2 \left(\frac{\lambda_2}{\lambda_1} \right)^k X^2 + \dots + c_n \left(\frac{\lambda_n}{\lambda_1} \right)^k X^n \right].$$

При большом k дроби $\left(\frac{\lambda_2}{\lambda_1}\right)^k, \dots, \left(\frac{\lambda_n}{\lambda_1}\right)^k$ малы, и поэтому $A^k X^{1(0)} = c_1 \lambda_1^k X^1$, т. е. $X^{1(k)} \rightarrow X^1$ при $k \rightarrow \infty$. Одновременно $\lambda_1 = \lim_{k \rightarrow \infty} \frac{x_i^{1(k+1)}}{x_i^{1(k)}}$.

2. Вместо применяемой в п. 4 методики формулы для $\lambda_1^{(k+1)}$ можно взять среднее арифметическое соответствующих отношений для разных координат.

3. Метод может использоваться и в случае, если наибольшее по модулю собственное значение матрицы A является кратным, т. е. $\lambda_1 = \lambda_2 = \dots = \lambda_s$, и $|\lambda_1| > |\lambda_k|$ при $k > s$.

4. При неудачном выборе начального приближения $X^{1(0)}$ предел отношения $\frac{x_i^{1(k+1)}}{x_i^{1(k)}}$ может не существовать. В этом случае следует задать другое начальное приближение.

5. Рассмотренный итерационный процесс для λ_1 сходится линейно с параметром $c = \frac{\lambda_2}{\lambda_1}$ и может быть очень медленным. Для его ускорения используется алгоритм Эйткена.

6. Если $A = A^T$ (матрица A симметрическая), то сходимость процесса при определении $\rho(A)$ может быть ускорена.

7. Используя λ_1 , можно определить следующее значение λ_2 по формуле

$$\lambda_2 = \frac{x_i^{1(k+1)} - \lambda_1 x_i^{1(k)}}{x_i^{1(k)} - \lambda_1 x_i^{1(k-1)}} \quad (i=1, 2, \dots, n).$$

Эта формула дает грубые значения для λ_2 , так как значение λ_1 является приближенным. Если модули всех собственных значений различны, то на основе последней формулы можно вычислять и остальные λ_j ($j = 3, 4, \dots, n$).

8. После проведения некоторого числа итераций рекомендуется «гасить» растущие компоненты получающегося собственного вектора. Это осуществляется нормировкой вектора, например по формуле $\frac{X^{1(k)}}{\|X^{1(k)}\|}$.

Пример 2.3. Для матрицы $A = \begin{pmatrix} 5 & 1 & 2 \\ 1 & 4 & 1 \\ 2 & 1 & 3 \end{pmatrix}$ найти спектральный радиус степенным методом с точностью $\varepsilon = 0,1$.

□1. Выбирается начальное приближение собственного вектора $X^{(0)} = (1; 1; 1)^T$. Положим $k = 0$.

2. Найдем

$$X^{1(1)} = AX^{1(0)} = \begin{pmatrix} 5 & 1 & 2 \\ 1 & 4 & 1 \\ 2 & 1 & 3 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 8 \\ 6 \\ 6 \end{pmatrix}; \quad \lambda_1^{(1)} = \frac{x_1^{1(1)}}{x_1^{1(0)}} = \frac{8}{1} = 8,$$

положим $k = 1$.

k	$x_1^{1(k)}$	$x_2^{1(k)}$	$x_3^{1(k)}$	$\lambda_1^{(k)}$	$ \lambda_1^{(k)} - \lambda_1^{(k-1)} $
0	1	1	1	—	—
1	8	6	6	8	—
2	58	38	40	7,25	0,75
3	408	250	274	7,034	0,116
4	2838	1682	1888	6,9559	0,078 < ε

3. Вычислим

$$X^{1(2)} = AX^{1(1)} = \begin{pmatrix} 5 & 1 & 2 \\ 1 & 4 & 1 \\ 2 & 1 & 3 \end{pmatrix} \cdot \begin{pmatrix} 8 \\ 6 \\ 6 \end{pmatrix} = \begin{pmatrix} 58 \\ 38 \\ 40 \end{pmatrix}.$$

4. Найдем

$$\lambda_1^{(2)} = \frac{x_1^{1(2)}}{x_1^{1(1)}} = \frac{58}{8} = 7,25.$$

5. Так как $|\lambda_1^{(2)} - \lambda_1^{(1)}| = 0,75 > \varepsilon$, то процесс необходимо продолжить.

Результаты вычислений удобно представить в виде таблицы 2.1.

Точность по $\lambda_1^{(k)}$ достигнута на четвертой итерации. Таким образом, в качестве приближенного значения λ_1 берется 6,9559, а в качестве собственного вектора принимается $X^1 = (2838, 1682, 1888)^T$. Так как собственный вектор определяется с точностью до постоянного множителя, то X^1 лучше пронормировать, т. е. поделить все его компоненты на величину нормы. Для рассматриваемого примера получим

$$X^1 = \frac{1}{2838} \begin{pmatrix} 2838 \\ 1682 \\ 1888 \end{pmatrix} = \begin{pmatrix} 1,000 \\ 0,5927 \\ 0,6652 \end{pmatrix}.$$

Согласно замечаниям, в качестве собственного значения λ_1 можно взять не только отношение $\frac{x_1^{1(4)}}{x_1^{1(3)}} = \frac{2838}{408} = 6,9559$, но и $\frac{x_2^{1(4)}}{x_2^{1(3)}} = \frac{1682}{250} = 6,7280$; $\frac{x_3^{1(4)}}{x_3^{1(3)}} = \frac{1888}{274} = 6,8905$, а также их среднее арифметическое $\frac{6,9559 + 6,7280 + 6,8905}{3} = 6,8581$. ■

Пример 2.4. Найти максимальное по модулю собственное значение мат-

рицы $A = \begin{pmatrix} 2 & -1 & 1 \\ -1 & 2 & -1 \\ 0 & 0 & 1 \end{pmatrix}$ и соответствующий собственный вектор.

□1. Зададим начальное приближение $X^{1(0)} = (1; -1; 1)^T$ и $\varepsilon = 0,0001$.

Выполним расчеты согласно методике (см. табл. 2.2).

k	$x_1^{1(k)}$	$x_2^{1(k)}$	$x_3^{1(k)}$	$\lambda_1^{(k)}$	$ \lambda_1^{(k)} - \lambda_1^{(k-1)} $
0	1	-1	1	—	—
1	4	-4	1	4	—
2	13	-13	1	3,25	0,75
3	40	-40	1	3,0769	0,17307
4	121	-121	1	3,025	0,0519
5	364	-364	1	3,00826	0,01673
6	1093	-1093	1	3,002747	0,005512
7	3280	-3280	1	3,000914	0,00183
8	9841	-9841	1	3,000304	0,000609
9	29524	-29524	1	3,000101	0,000202
10	88573	-88573	1	3,000034	0,000067

В результате получено собственное значение $\lambda_1 \approx 3,00003$ и собственный вектор $X^1 = (88573; -88573; 1)^T$ или после выполнения нормировки $X^1 = \frac{1}{88573}(88573; -88573; 1)^T = (1; -1; 0,0000118)^T$ (сравните с примером 2.2). ■

2.4. МЕТОД ВРАЩЕНИЙ

Метод используется для решения полной проблемы собственных значений симметрической матрицы и основан на преобразовании подобия исходной матрицы $A \in R^{n \times n}$ с помощью ортогональной матрицы H .

Две матрицы A и $A^{(i)}$ называются *подобными* ($A \sim A^{(i)}$ или $A^{(i)} \sim A$), если $A^{(i)} = H^{-1}AH$ или $A = HA^{(i)}H^{-1}$, где H — невырожденная матрица.

В методе вращений в качестве H берется *ортогональная матрица*, такая, что $HH^T = H^TH = E$, т. е. $H^T = H^{-1}$. В силу свойства ортогонального преобразования евклидова норма исходной матрицы A не меняется. Для преобразованной матрицы $A^{(i)}$ сохраняется ее след и собственные значения λ_i :

$$\text{tr}A = \sum_{i=1}^n a_{ii} = \sum_{i=1}^n \lambda_i(A) = \text{tr}A^{(i)}.$$

При реализации метода вращений преобразование подобия применяется к исходной матрице A многократно:

$$A^{(k+1)} = (H^{(k)})^{-1}A^{(k)}H^{(k)} = (H^{(k)})^TA^{(k)}H^{(k)}, \quad k = 0, 1, \dots \quad (2.6)$$

Формула (2.6) определяет итерационный процесс, где начальное приближение $A^{(0)} = A$. На каждой k -й итерации для некоторого выбранного при решении задачи недиагонального элемента $a_{ij}^{(k)}$, $i \neq j$, определяется ортогональная матрица $H^{(k)}$, приводящая этот элемент $a_{ij}^{(k+1)}$ (а также и $a_{ji}^{(k+1)}$) к нулю. При этом на каждой итерации в качестве $a_{ij}^{(k)}$ выбирается наибольший по модулю. Матрица $H^{(k)}$, называемая *матрицей вращения Якоби*, зависит от угла $\varphi^{(k)}$ и имеет вид

$$H^{(k)} = \begin{pmatrix} 1 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \cos \varphi^{(k)} & 0 & \cdots & 0 & -\sin \varphi^{(k)} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 1 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \sin \varphi^{(k)} & 0 & \cdots & 0 & \cos \varphi^{(k)} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 1 \end{pmatrix} \begin{array}{l} \leftarrow i\text{-я строка} \\ \leftarrow j\text{-я строка} \end{array}$$

В данной ортогональной матрице элементы на главной диагонали единичные, кроме $h_{ii}^{(k)} = \cos\varphi^{(k)}$ и $h_{jj}^{(k)} = \cos\varphi^{(k)}$, а $h_{ij}^{(k)} = -\sin\varphi^{(k)}$, $h_{ji}^{(k)} = \sin\varphi^{(k)}$ (h_{ij} — элементы матрицы H).

Угол поворота $\phi^{(k)}$ определяется по формуле

$$\operatorname{tg} 2\varphi^{(k)} = \frac{2a_{ij}^{(k)}}{a_{ii}^{(k)} - a_{jj}^{(k)}} = \bar{P}_k; \quad \varphi^{(k)} = \frac{1}{2} \operatorname{arctg} \bar{P}_k, \quad (2.7)$$

где $|2\varphi^{(k)}| \leq \frac{\pi}{2}$, $i < j$ (a_{ij} выбирается в верхней треугольной наддиагональной части матрицы A). Заметим, что при $a_{ii}^{(k)} = a_{jj}^{(k)}$ получается $\varphi^{(k)} = \frac{\pi}{4}$.

В процессе итераций сумма квадратов всех недиагональных элементов $\sigma(A^{(k)})$ при возрастании k уменьшается, так что $\sigma(A^{(k+1)}) < \sigma(A^{(k)})$. Элементы $a_{ij}^{(k)}$, приведенные к нулю на k -й итерации, на последующей итерации немного возрастают. При $k \rightarrow \infty$ получается монотонно убывающая ограниченная снизу нулем последовательность $\sigma(A^{(1)}) > \sigma(A^{(2)}) > \dots > \sigma(A^{(k)})$. Поэтому $\sigma(A^{(k)}) \rightarrow 0$ при $k \rightarrow \infty$. Это и означает сходимость метода. При этом

$$A^{(k)} \rightarrow \Lambda = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}.$$

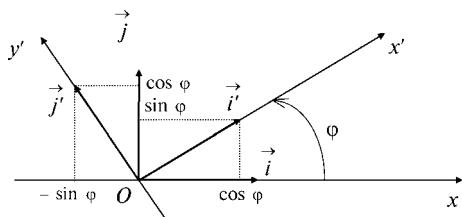


Рис. 2.1

Замечание. В двумерном пространстве с введенной в нем системой координат Oxy с ортонормированным базисом $\{\vec{i}, \vec{j}\}$ матрица вращения легко получается из рисунка 2.1, где система координат $Ox'y'$ повернута на угол Φ :

$$\vec{i}' = \vec{i} \cos\varphi + \vec{j} \sin\varphi.$$

$$\vec{i}' = -\vec{i} \sin \omega + \vec{i} \cos \omega$$

Таким образом, для компонент \vec{i}', \vec{j}' будем иметь

$$\left(\begin{array}{c} \vec{i}' \\ \vec{j}' \end{array} \right) = \left(\begin{array}{c} \vec{i} \\ \vec{j} \end{array} \right) \cdot \begin{pmatrix} \cos \varphi & -\sin \varphi \\ \sin \varphi & \cos \varphi \end{pmatrix}.$$

Отсюда следует, что в двумерном пространстве матрица вращения имеет вид

$$H = \begin{pmatrix} \cos \varphi & -\sin \varphi \\ \sin \varphi & \cos \varphi \end{pmatrix}.$$

Отметим, что при $n = 2$ для решения задачи требуется одна итерация.

Методика решения задачи

1. Положить $k = 0$, $A^{(0)} = A$ и задать $\varepsilon > 0$.
2. Выделить в верхней треугольной наддиагональной части матрицы $A^{(k)}$ максимальный по модулю элемент $a_{ij}^{(k)}$, $i < j$.
Если $|a_{ij}^{(k)}| \leq \varepsilon$ для всех $i \neq j$, процесс завершить. Собственные значения определяются по формуле

$$\lambda_i(A^{(k)}) = a_{ii}^{(k)}, \quad i = \overline{1, n}.$$

Собственные векторы X^i находятся как i -е столбцы матрицы, получающейся в результате перемножения:

$$v_k = H^{(0)}H^{(1)}H^{(2)} \dots H^{(k-1)} = (X^1, X^2, X^3, \dots, X^n).$$

Если $|a_{ij}^{(k)}| > \varepsilon$, процесс продолжается.

3. Найти угол поворота по формуле

$$\varphi^{(k)} = \frac{1}{2} \operatorname{arctg} \frac{2a_{ij}^{(k)}}{a_{ii}^{(k)} - a_{jj}^{(k)}} \quad \left(\text{при } a_{ii}^{(k)} = a_{jj}^{(k)} \text{ получается } \varphi^{(k)} = \frac{\pi}{4} \right).$$

4. Составить матрицу вращения $H^{(k)}$.

5. Вычислить очередное приближение

$$A^{(k+1)} = (H^{(k)})^T A^{(k)} H^{(k)}.$$

Положить $k = k + 1$ и перейти к п. 2.

Замечания

1. Используя обозначение $\bar{P}_k = \frac{2a_{ij}^{(k)}}{a_{ii}^{(k)} - a_{jj}^{(k)}}$, можно в п. 3 методики вычислять элементы матрицы вращения по формулам:

$$\sin \varphi^{(k)} = \operatorname{sign} \bar{P}_k \cdot \sqrt{\frac{1}{2} \left(1 - \frac{1}{\sqrt{1 + \bar{P}_k^2}} \right)}; \quad \cos \varphi^{(k)} = \sqrt{\frac{1}{2} \left(1 + \frac{1}{\sqrt{1 + \bar{P}_k^2}} \right)}.$$

2. Контроль правильности выполнения действий по каждому повороту осуществляется путем проверки сохранения следа преобразуемой матрицы.

Пример 2.5. Для матрицы $A = \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}$ методом вращений найти собственные значения и собственные векторы.

□1. Положим $k = 0$, $A^{(0)} = A = \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}$, $\varepsilon = 10^{-10}$.

2⁰. Выше главной диагонали имеется только один элемент $a_{ij} = a_{12} = 1$.

3⁰. Находим угол поворота матрицы по формуле (2.7), используя в расчетах 11 цифр после запятой в соответствии с заданной точностью:

$$\operatorname{tg} 2\varphi^{(0)} = \frac{2a_{ij}}{a_{ii} - a_{jj}} = \frac{2}{2-3} = -2;$$

$$\sin \varphi^{(0)} = -0,52573111212; \cos \varphi^{(0)} = 0,85065080835.$$

4⁰. Сформируем матрицу вращения:

$$H^{(0)} = \begin{pmatrix} \cos \varphi^{(0)} & -\sin \varphi^{(0)} \\ \sin \varphi^{(0)} & \cos \varphi^{(0)} \end{pmatrix} = \begin{pmatrix} 0,85065080835 & 0,52573111212 \\ -0,52573111212 & 0,85065080835 \end{pmatrix}.$$

5⁰. Выполним первую итерацию:

$$\begin{aligned} A^{(1)} &= (H^{(0)})^T A^{(0)} H^{(0)} = \\ &= \begin{pmatrix} 0,85065080835 & -0,52573111212 \\ 0,52573111212 & 0,85065080835 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} 0,85065080835 & 0,52573111212 \\ -0,52573111212 & 0,85065080835 \end{pmatrix} = \\ &= \begin{pmatrix} 1,38196601125 & -4,04620781325 \cdot 10^{-12} \\ -4,04587474634 \cdot 10^{-12} & 3,61803398874 \end{pmatrix}. \end{aligned}$$

Очевидно, след матрицы с заданной точностью сохраняется, т. е.

$$\sum_{i=1}^2 a_{ii}^{(1)} = \sum_{i=1}^2 a_{ii}^{(0)} = 5.$$

Положим $k = 1$ и перейдем к п. 2.

2¹. Максимальный по модулю наддиагональный элемент

$$|a_{12}| = 4,04620781325 \cdot 10^{-12} < \varepsilon = 10^{-10}.$$

Для решения задачи (подчеркнем, что $n = 2$) с принятой точностью потребовалась одна итерация, полученную матрицу можно считать диагональной. Найдены следующие собственные значения и собственные векторы:

$$\lambda_1 = 1,38196601125; \quad \lambda_2 = 3,61803398874;$$

$$X^1 = \begin{pmatrix} 0,85065080835 \\ -0,52573111212 \end{pmatrix}; \quad X^2 = \begin{pmatrix} 0,52573111212 \\ 0,85065080835 \end{pmatrix}. \blacksquare$$

Пример 2.6. Найти собственные значения и собственные векторы матрицы

$$A = \begin{pmatrix} 5 & 1 & 2 \\ 1 & 4 & 1 \\ 2 & 1 & 3 \end{pmatrix}.$$

□1. Положим $k = 0$, $A^{(0)} = A$, $\varepsilon = 0,001$.

2⁰. Выделим максимальный по модулю элемент в наддиагональной части: $a_{13}^{(0)} = 2$. Так как $a_{13} = 2 > \varepsilon = 0,001$, то процесс продолжается.

3⁰. Находим угол поворота:

$$\begin{aligned}\varphi^{(0)} &= \frac{1}{2} \operatorname{arctg} \frac{2 \cdot a_{13}^{(0)}}{a_{11}^{(0)} - a_{33}^{(0)}} = \frac{1}{2} \operatorname{arctg} \frac{4}{5-3} = \frac{1}{2} \operatorname{arctg} 2 = 0,553574; \\ \sin \varphi^{(0)} &= 0,52573; \quad \cos \varphi^{(0)} = 0,85065.\end{aligned}$$

4⁰. Сформируем матрицу вращения:

$$H^{(0)} = \begin{pmatrix} 0,85065 & 0 & -0,52573 \\ 0 & 1 & 0 \\ 0,52573 & 0 & 0,85065 \end{pmatrix}.$$

5⁰. Выполним первую итерацию:

$$A^{(1)} = (H^{(0)})^T A^{(0)} H^{(0)} = \begin{pmatrix} 6,236 & 1,376 & 2,33 \cdot 10^{-6} \\ 1,376 & 4 & 0,325 \\ 2,33 \cdot 10^{-6} & 0,325 & 1,764 \end{pmatrix}.$$

Положим $k = 1$ и перейдем к п. 2.

2¹. Максимальный по модулю наддиагональный элемент $a_{12}^{(1)} = 1,376$. Так как $a_{12}^{(1)} > \varepsilon$, процесс продолжается.

3¹. Найдем угол поворота:

$$\begin{aligned}\varphi^{(1)} &= \frac{1}{2} \operatorname{arctg} \frac{2 \cdot a_{12}^{(1)}}{a_{11}^{(1)} - a_{22}^{(1)}} = \frac{1}{2} \operatorname{arctg} \frac{2 \cdot 1,376}{6,236 - 4} = \frac{1}{2} \operatorname{arctg} 1,230769 = 0,444239; \\ \sin \varphi^{(1)} &= 0,429770; \quad \cos \varphi^{(1)} = 0,902937.\end{aligned}$$

4¹. Сформируем матрицу вращения:

$$H^{(1)} = \begin{pmatrix} 0,902937 & -0,429770 & 0 \\ 0,429770 & 0,902937 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

5¹. Выполним вторую итерацию:

$$A^{(2)} = (H^{(1)})^T A^{(1)} H^{(1)} = \begin{pmatrix} 6,891 & 2,238 \cdot 10^{-4} & 0,14 \\ 2,238 \cdot 10^{-4} & 3,345 & 0,293 \\ 0,14 & 0,293 & 1,764 \end{pmatrix}.$$

Положим $k = 2$ и перейдем к п. 2.

2². Максимальный по модулю наддиагональный элемент $a_{23}^{(2)} = 0,293 > \varepsilon$.

3². Найдем угол поворота:

$$\begin{aligned}\varphi^{(2)} &= \frac{1}{2} \operatorname{arctg} \frac{2 \cdot a_{23}^{(2)}}{a_{22}^{(2)} - a_{33}^{(2)}} = \frac{1}{2} \operatorname{arctg} \frac{2 \cdot 0,293}{3,345 - 1,764} = \frac{1}{2} \operatorname{arctg} 0,370651 = 0,177476; \\ \sin \varphi^{(2)} &= 0,1765460; \quad \cos \varphi^{(2)} = 0,9842924.\end{aligned}$$

4². Сформируем матрицу вращения:

$$H^{(2)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0,9842924 & -0,1765460 \\ 0 & 0,1765460 & 0,9842924 \end{pmatrix}.$$

5². Выполним третью итерацию:

$$A^{(3)} = (H^{(2)})^T A^{(2)} H^{(2)} = \begin{pmatrix} 6,891 & 0,025 & 0,138 \\ 0,025 & 3,398 & 3,375 \cdot 10^{-7} \\ 0,138 & 3,375 \cdot 10^{-7} & 1,711 \end{pmatrix}.$$

Положим $k = 3$ и перейдем к п. 2.

2³. Максимальный по модулю наддиагональный элемент $a_{13}^{(3)} = 0,138 > \varepsilon$.

3³. Найдем угол поворота:

$$\varphi^{(3)} = \frac{1}{2} \operatorname{arctg} \frac{2 \cdot a_{13}^{(3)}}{a_{11}^{(3)} - a_{33}^{(3)}} = \frac{1}{2} \operatorname{arctg} \frac{2 \cdot 0,138}{6,891 - 1,711} = \frac{1}{2} \operatorname{arctg} 0,05328 = 0,026615;$$

$$\sin \varphi^{(3)} = 0,026611; \cos \varphi^{(3)} = 0,999646.$$

4³. Сформируем матрицу вращения:

$$H^{(3)} = \begin{pmatrix} 0,999646 & 0 & -0,026611 \\ 0 & 1 & 0 \\ 0,026611 & 0 & 0,999646 \end{pmatrix}.$$

5³. Выполним четвертую итерацию:

$$A^{(4)} = (H^{(3)})^T A^{(3)} H^{(3)} = \begin{pmatrix} 6,895 & 0,025 & 8,406 \cdot 10^{-6} \\ 0,025 & 3,398 & -6,649 \cdot 10^{-4} \\ 8,406 \cdot 10^{-6} & -6,649 \cdot 10^{-4} & 1,707 \end{pmatrix}.$$

Положим $k = 4$ и перейдем к п. 2.

2⁴. Так как $a_{12}^{(4)} = 0,025 > \varepsilon$, процесс повторяется.

3⁴. Найдем угол поворота:

$$\varphi^{(4)} = \frac{1}{2} \operatorname{arctg} \frac{2 \cdot a_{12}^{(4)}}{a_{11}^{(4)} - a_{22}^{(4)}} = \frac{1}{2} \operatorname{arctg} \frac{2 \cdot 0,025}{6,895 - 3,398} = 0,0071484;$$

$$\sin \varphi^{(4)} = 0,0071483; \cos \varphi^{(4)} = 0,9999744.$$

4⁴. Сформируем матрицу вращения:

$$H^{(4)} = \begin{pmatrix} 0,9999744 & -0,0071483 & 0 \\ 0,0071483 & 0,9999744 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

5⁴. Выполним пятую итерацию:

$$A^{(5)} = (H^{(4)})^T A^{(4)} H^{(4)} = \begin{pmatrix} 6,895 & 4,774 \cdot 10^{-7} & 3,653 \cdot 10^{-6} \\ 4,774 \cdot 10^{-7} & 3,398 & -6,649 \cdot 10^{-4} \\ 3,653 \cdot 10^{-6} & -6,649 \cdot 10^{-4} & 1,707 \end{pmatrix}.$$

Положим $k = 5$ и перейдем к п. 2.

2⁵. Так как наибольший по модулю наддиагональный элемент удовлетворяет условию $|-6,649 \cdot 10^{-4}| < \varepsilon = 0,001$, процесс завершается.

Собственные значения:

$$\lambda_1 \cong a_{11}^{(5)} = 6,895; \quad \lambda_2 \cong a_{22}^{(5)} = 3,398; \quad \lambda_3 \cong a_{33}^{(5)} = 1,707.$$

Для нахождения собственных векторов вычислим

$$v_5 = H^{(0)} \cdot H^{(1)} \cdot H^{(2)} \cdot H^{(3)} \cdot H^{(4)} = \begin{pmatrix} 0,753 & -0,458 & -0,473 \\ 0,432 & 0,886 & -0,171 \\ 0,497 & -0,076 & 0,864 \end{pmatrix},$$

отсюда

$$X^1 = \begin{pmatrix} 0,753 \\ 0,432 \\ 0,497 \end{pmatrix}, \quad X^2 = \begin{pmatrix} -0,458 \\ 0,886 \\ -0,076 \end{pmatrix}, \quad X^3 = \begin{pmatrix} -0,473 \\ -0,171 \\ 0,864 \end{pmatrix},$$

или после нормировки

$$X^1 = \begin{pmatrix} 1 \\ 0,5737 \\ 0,660 \end{pmatrix}, \quad X^2 = \begin{pmatrix} -0,517 \\ 1 \\ -0,0858 \end{pmatrix}, \quad X^3 = \begin{pmatrix} -0,5474 \\ -0,1979 \\ 1 \end{pmatrix}. \blacksquare$$

Задачи для самостоятельного решения

1. Для следующих матриц с точностью $\varepsilon = 0,01$ определить методами вращения и непосредственного развертывания собственные значения и собственные векторы. Методом итераций найти спектральный радиус.

$$1) \begin{pmatrix} 2,1 & 1 & 1,1 \\ 1 & 2,6 & 1,1 \\ 1,1 & 1,1 & 3,1 \end{pmatrix}; \quad 2) \begin{pmatrix} 2,4 & 1 & 1,4 \\ 1 & 2,3 & 1,4 \\ 1,4 & 1,4 & 3,4 \end{pmatrix}; \quad 3) \begin{pmatrix} 1,3 & 0,4 & 0,5 \\ 0,4 & 1,3 & 0,3 \\ 0,5 & 0,3 & 1,3 \end{pmatrix};$$

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

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

2. Методом итераций найти спектральный радиус для матриц:

$$1) \begin{pmatrix} 4 & 2 & 2 \\ 2 & 5 & 1 \\ 2 & 1 & 6 \end{pmatrix};$$

$$2) \begin{pmatrix} 4 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{pmatrix};$$

$$3) \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 2 & 3 \\ 3 & 2 & 1 & 2 \\ 4 & 3 & 2 & 1 \end{pmatrix};$$

$$4) \begin{pmatrix} 2 & 3 & 5 \\ 3 & 1 & 6 \\ 5 & 6 & 2 \end{pmatrix};$$

$$5) \begin{pmatrix} 1,6 & 1,2 & 0,4 \\ 1,2 & 0,5 & 1,0 \\ 0,4 & 1,0 & 0,8 \end{pmatrix};$$

$$6) \begin{pmatrix} 1,2 & 2,0 & 0,5 \\ 2,0 & 1,4 & 1,7 \\ 0,5 & 1,7 & 0,3 \end{pmatrix}.$$

ГЛАВА 3

МЕТОДЫ РЕШЕНИЯ НЕЛИНЕЙНЫХ УРАВНЕНИЙ И СИСТЕМ

3.1. МЕТОДЫ РЕШЕНИЯ НЕЛИНЕЙНЫХ УРАВНЕНИЙ

3.1.1. ПОСТАНОВКА ЗАДАЧИ

Пусть дано нелинейное уравнение

$$f(x) = 0, \quad (3.1)$$

где $f(x)$ — функция, определенная и непрерывная на некотором промежутке. В некоторых случаях на функцию $f(x)$ могут быть наложены дополнительные ограничения, например непрерывность первой и второй производных, что специально оговаривается. Функция $f(x)$ может быть задана в виде алгебраического многочлена или трансцендентной функции (тогда ей соответствует алгебраическое или трансцендентное уравнение).

Требуется найти *корни* уравнения (3.1), т. е. числа x_{*1}, x_{*2}, \dots , которые путем подстановки их в (3.1) превращают уравнение в верное числовое равенство. Числа x_{*1}, x_{*2}, \dots называются также *нулями* функции $f(x)$.

На практике часто бывает выгодно уравнение (3.1) заменить равносильным ему уравнением (уравнения равносильны, если имеют одинаковые корни):

$$f_1(x) - f_2(x) = 0, \quad (3.2)$$

где функции $f_1(x), f_2(x)$ — более простые, чем функция $f(x)$. Тогда при задании уравнения в виде (3.1) нулями функции $f(x)$ являются точки пересечения $f(x)$ с осью Ox (рис. 3.1а), а при задании в виде (3.2) — абсциссы точек пересечения функций $f_1(x)$ и $f_2(x)$ (рис. 3.1б).

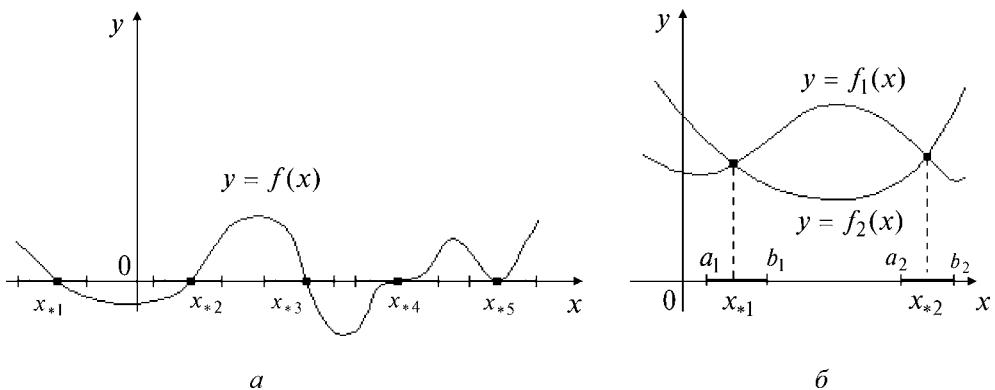


Рис. 3.1

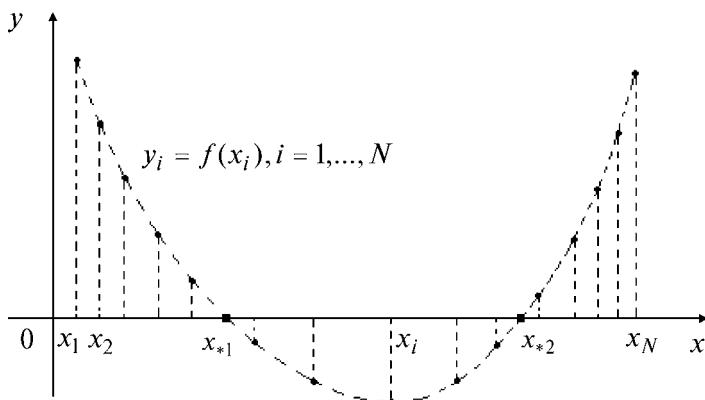


Рис. 3.2

Замечания

1. Если $f(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_0 = P_n(x)$ — алгебраический многочлен, то уравнение (3.1) называется также *алгебраическим n-й степени*:

$$P_n(x) \equiv a_n x^n + a_{n-1} x^{n-1} + \dots + a_0 = 0, \quad (3.3)$$

где a_n, \dots, a_0 — действительные числа, коэффициенты уравнения.

2. На практике встречаются задачи нахождения корней уравнения $f(x_i) = 0$, левая часть которого задана сеточной функцией $y_i = f(x_i)$, $i = 1, \dots, N$ (рис. 3.2).

Число x_* есть *корень уравнения (3.1) кратности k*, если при $x = x_*$ вместе с функцией $f(x)$ обращаются в нуль ее производные до $(k-1)$ -го порядка включительно, т. е. $f(x_*) = f'(x_*) = \dots = f^{(k-1)}(x_*) = 0$, а $f^{(k)}(x_*) \neq 0$. Корень кратности $k=1$ называется *простым*. На рисунке 3.1а простыми корнями являются x_{*1}, x_{*2}, x_{*3} , а корни x_{*4}, x_{*5} — кратные.

В соответствии с классическим результатом Галуа алгебраическое уравнение (3.1) при $n \geq 5$ не имеет решения в замкнутом (формульном) виде. Сеточные уравнения вообще не имеют формульных решений. Поэтому корни

алгебраических ($n > 2$), трансцендентных и сеточных уравнений, как правило, определяются приближенно с заданной точностью.

Решение осуществляется в два этапа:

Первый этап. Находятся отрезки $[a_i, b_i]$, внутри каждого из которых содержится один простой или кратный корень ($x_{*i} \in [a_i, b_i]$) (см. рис. 3.1). Этот этап называется процедурой *отделения корней*. По сути на нем осуществляется грубое нахождение корней x_{*i} .

Второй этап. Грубое значение каждого корня x_{*i} уточняется до заданной точности одним из численных методов, в которых реализуются последовательные приближения. Порядок (скорость) сходимости метода определяется так же, как в п. 1.3.1.

3.1.2. ОТДЕЛЕНИЕ КОРНЕЙ

Для отделения действительных корней полезно определять заранее число корней, а также верхнюю и нижнюю границы их расположения. Для этого используется ряд теорем.

Теорема 3.1 (о числе корней алгебраического уравнения (3.3)).

Алгебраическое уравнение (3.3) n -й степени имеет ровно n корней, действительных или комплексных, при условии, что каждый корень считается столько раз, какова его кратность.

Теорема 3.2 (о свойстве парной сопряженности комплексных корней уравнения (3.3)).

Если $x_{*i} = \alpha + \beta i$ — корень алгебраического уравнения (3.3) кратности k , то число $\bar{x}_{*i} = \alpha - \beta i$ также является корнем той же кратности.

Следствие. Алгебраическое уравнение нечетной степени имеет по крайней мере один действительный корень.

Теорема 3.3 (об оценке модулей корней уравнения (3.3)).

Пусть $A = \max\{|a_{n-1}|, \dots, |a_0|\}$, $B = \max\{|a_n|, |a_{n-1}|, \dots, |a_1|\}$, где a_k , $k = \overline{0, n}$ — коэффициенты уравнения $a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0 = 0$.

Тогда модули всех корней x_{*i} ($i = 1, \dots, n$) уравнения удовлетворяют неравенству

$$\frac{1}{1 + \frac{B}{|a_0|}} < |x_{*i}| \leq 1 + \frac{A}{|a_n|}, \quad i = 1, \dots, n, \quad (3.4)$$

т. е. корни уравнения расположены в кольце.

Следствие. Числа $r = \frac{1}{1 + \frac{B}{|a_0|}}$ и $R = 1 + \frac{A}{|a_n|}$ являются соответственно нижней и верхней границами положительных корней алгебраического уравнения $r < x_{*i}^+ < R$. Аналогично числа $-R$ и $-r$ служат нижней и верхней границами отрицательных корней уравнения $-R < x_{*i}^- < -r$.

Приведем полезные теоремы, используемые для более точного установления границ действительных корней алгебраических уравнений.

Теорема 3.4 (теорема Лагранжа о верхней границе положительных корней уравнения (3.3)).

Пусть $a_n > 0$ и a_i — первый отрицательный коэффициент в последовательности $a_n, a_{n-1}, a_{n-2}, \dots, a_1, a_0$; C — наибольшая из абсолютных величин отрицательных коэффициентов. Тогда за верхнюю границу положительных корней уравнения (3.3) может быть принято число

$$R = 1 + \sqrt[n-i]{\frac{C}{a_n}}. \quad (3.5)$$

Теорема 3.5 (о нижних и верхних границах положительных и отрицательных корней алгебраического уравнения).

Пусть R — верхняя граница положительных корней уравнения $P_n(x) = 0$, R_1 — верхняя граница положительных корней уравнения $P^1(x) = x^n P_n\left(\frac{1}{x}\right) = 0$, R_2 — верхняя граница положительных корней уравнения $P^2(x) = P_n(-x) = 0$, R_3 — верхняя граница положительных корней уравнения $P^3(x) = x^n P_n\left(-\frac{1}{x}\right) = 0$.

Тогда положительные корни x_{*i}^+ и отрицательные корни x_{*i}^- уравнения (3.3) удовлетворяют неравенствам

$$\frac{1}{R_1} \leq x_{*i}^+ \leq R; \quad -R_2 \leq x_{*i}^- \leq -\frac{1}{R_3}. \quad (3.6)$$

Теорема 3.6 (теорема Декарта о количестве действительных корней алгебраических уравнений).

Число S_1 положительных корней (с учетом их кратностей) алгебраического уравнения $P_n(x) = 0$ равно числу перемен знаков в последовательности коэффициентов a_n, a_{n-1}, \dots, a_0 (коэффициенты, равные нулю, не учитываются) многочлена $P_n(x)$ или меньше этого числа на четное число. Число S_2 отрицательных корней (с учетом их кратностей) алгебраического уравнения $P_n(x) = 0$ равно числу перемен знаков в последовательности a_n, a_{n-1}, \dots, a_0 многочлена $P_n(-x)$ или меньше этого числа на четное число.

Теорема 3.7 (теорема Гюа о необходимом условии действительности всех корней алгебраического уравнения).

Если алгебраическое уравнение (3.3) имеет все действительные корни, то квадрат каждого некрайнего коэффициента больше произведения двух его соседних коэффициентов.

Следствие. Если при каком-нибудь k выполнено неравенство $a_k^2 \leq a_{k-1} a_{k+1}$, то уравнение (3.3) имеет по крайней мере одну пару комплексных корней.

Для отделения корней применяется следующая теорема.

Теорема 3.8. Если функция $f(x)$, определяющая уравнение $f(x) = 0$, на концах отрезка $[a_i, b_i]$ принимает значения разных знаков, т. е. $f(a_i) \cdot f(b_i) < 0$, то на этом отрезке содержится по крайней мере один корень уравнения. Если же $f(x)$ непрерывна и дифференцируема и ее первая производная сохраняет знак внутри отрезка $[a_i, b_i]$ ($\underset{[a_i, b_i]}{\text{sign } f'(x)} = \text{const}$), то на $[a_i, b_i]$ находится только один корень x_{*i} уравнения.

В вычислительной практике обычно используются следующие способы отделения корней:

1) средствами компьютерной графики: функция $f(x)$ представляется на мониторе, и приближенно определяются отрезки, которым принадлежат точки x_{*i} ;

2) средствами математического анализа с помощью исследования функций и построения графиков (см. рис. 3.1а);

3) формированием простых функций $f_1(x)$ и $f_2(x)$ таких, что получается равносильное уравнение в виде (3.2), и дальнейшим построением графиков этих функций (см. рис. 3.1б).

Пример 3.1. Определить число положительных и отрицательных корней, а также их границы для уравнения

$$P_5(x) = x^5 + 2x^4 - 5x^3 + 8x^2 - 7x - 3 = 0.$$

□ В данной задаче $n = 5$, $a_5 = 1$, $a_4 = 2$, $a_3 = -5$, $a_2 = 8$, $a_1 = -7$, $a_0 = -3$.

Согласно теореме 3.1 уравнение имеет пять корней. Поскольку $n = 5$, то по следствию из теоремы 3.2 уравнение имеет по крайней мере один действительный корень.

Определим модули корней по теореме 3.3. Так как $A = \max\{|2|, |-5|, |8|, |-7|, |-3|\} = 8$, $B = \max\{|1|, |2|, |-5|, |8|, |-7|\} = 8$, то

$$\frac{1}{1 + \frac{8}{|-3|}} < |x_{*i}| < 1 + \frac{8}{|1|} \text{ или } \frac{3}{11} < |x_{*i}| < 9,$$

т. е. все корни лежат внутри данного кольца. По следствию из теоремы 3.3 это означает, что положительные корни удовлетворяют неравенству $\frac{3}{11} < x_{*i}^+ < 9$, а отрицательные — неравенству $-9 < x_{*i}^- < -\frac{3}{11}$.

Применим теоремы 3.4 и 3.5 для уточнения приведенных результатов. Найдем верхнюю границу положительных корней. Так как $a_3 = -5$ — первый отрицательный коэффициент в последовательности $1, 2, -5, 8, -7, -3$, то $i = 3$, а $C = \max\{|-5|, |-7|, |-3|\} = 7$ — наибольшая из абсолютных величин отрицательных коэффициентов. Следовательно,

$$R = 1 + \sqrt[5-3]{\frac{7}{1}} = 1 + \sqrt{7} \approx 3,646.$$

Найдем нижнюю границу положительных корней. Составим уравнение:

$$\begin{aligned} P^1(x) &= x^n P_n\left(\frac{1}{x}\right) = x^5 P_5\left(\frac{1}{x}\right) = x^5 \left(\frac{1}{x^5} + 2 \frac{1}{x^4} - 5 \frac{1}{x^3} + 8 \frac{1}{x^2} - 7 \frac{1}{x} - 3 \right) = \\ &= -3x^5 - 7x^4 + 8x^3 - 5x^2 + 2x + 1 = 0 \end{aligned}$$

или

$$3x^5 + 7x^4 - 8x^3 + 5x^2 - 2x - 1 = 0$$

(старший коэффициент должен быть положительным). Для этого уравнения $i = 3$, $C = \max\{|-8|, |-2|, |-1|\} = 8$, поэтому

$$R_1 = 1 + \sqrt[5-3]{\frac{8}{3}} = 1 + \sqrt{\frac{8}{3}} \approx 2,632.$$

Отсюда

$$\frac{1}{R_1} = 0,38 \leq x_{*i}^+ \leq R = 3,646.$$

Уточним границы отрицательных корней. Составим уравнение:

$$P^2(x) = P_5(-x) = -x^5 + 2x^4 + 5x^3 + 8x^2 + 7x - 3 = 0$$

или

$$x^5 - 2x^4 - 5x^3 - 8x^2 - 7x + 3 = 0.$$

Для этого уравнения $i = 4$, $C = \max\{|-2|, |-5|, |-6|, |-7|\} = 8$, поэтому

$$R_2 = 1 + \sqrt[5-4]{\frac{8}{1}} = 9.$$

Составим уравнение

$$\begin{aligned} P^3(x) &= x^n P_n\left(-\frac{1}{x}\right) = x^5 P_5\left(-\frac{1}{x}\right) = x^5 \left(-\frac{1}{x^5} + 2 \frac{1}{x^4} + 5 \frac{1}{x^3} + 8 \frac{1}{x^2} + 7 \frac{1}{x} - 3\right) = \\ &= -3x^5 + 7x^4 + 8x^3 + 5x^2 + 2x - 1 = 0 \end{aligned}$$

или

$$3x^5 - 7x^4 - 8x^3 - 5x^2 - 2x + 1 = 0.$$

Для этого уравнения $i = 4$, $C = \max\{|-7|, |-8|, |-5|, |-2|\} = 8$, поэтому

$$R_3 = 1 + \sqrt[5-4]{\frac{8}{3}} = 1 + \frac{8}{3} \approx 3,666.$$

Отсюда находим:

$$-R_2 = -9 \leq x_{*i}^- \leq -\frac{1}{R_3} = -0,272.$$

Заметим, что данный результат совпадает с полученным ранее.

Исследуем структуру корней уравнения. Так как квадрат каждого не-крайнего коэффициента больше произведения двух его соседних коэффициентов, то по теореме 3.7 необходимое условие действительности всех корней уравнения выполняется.

На основе теоремы 3.6 определим число положительных и отрицательных корней. Выписываем коэффициенты многочлена $P_n(x) = P_5(x)$: 1, 2, -5, 8, -7, -3. Так как число перемен знака $S_1 = 3$, то число положительных корней равно трем или меньше на четное число, т. е. равно 1. Далее выписываем коэффициенты многочлена $P_5(-x)$: -1, 2, 5, 8, 7, -3. Так как число перемен знаков $S_2 = 2$, то число отрицательных корней равно двум или меньше на четное число, т. е. их вообще нет. ■

Пример 3.2. Отделить корни уравнения $x^3 - x + 1 = 0$.

□ Согласно теореме 3.1 уравнение имеет три корня, среди которых по крайней мере один действительный (следствие из теоремы 3.2, поскольку это уравнение нечетной степени).

Оценим модули корней уравнения по теореме 3.3. Так как $A = \max\{|0|, |-1|, |1|\} = 1$, $B = \max\{|1|, |0|, |-1|\} = 1$, то

$$\frac{1}{1 + \frac{1}{|1|}} < |x_{*i}| < 1 + \frac{1}{|1|}, \text{ или } \frac{1}{2} < |x_{*i}| < 2.$$

Отсюда

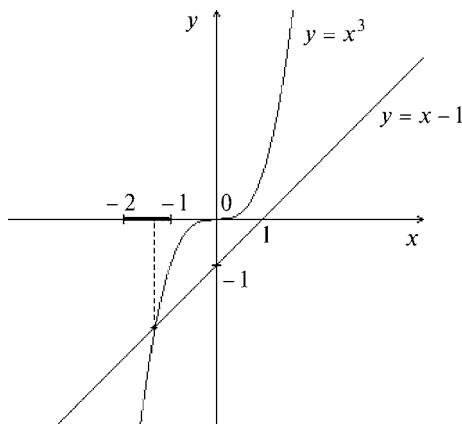


Рис. 3.3

$$\frac{1}{2} < x_{*i}^+ < 2 \text{ и } -2 < x_{*i}^- < -\frac{1}{2}.$$

Определим число положительных и отрицательных корней. Выписываем коэффициенты многочлена $P_3(x)$: 1, 0, -1, 1. Так как число перемен знака $S_1 = 2$ (нулевой коэффициент не учитывается), то число положительных корней равно двум или меньше на четное число, т. е. они отсутствуют. Далее выписываем коэффициенты многочлена $P_3(-x) = -x^3 + x + 1$: -1, 0, 1, 1. Так как число перемен знака $S_2 = 1$ (нулевой коэффициент не учитывается), то число отрицательных корней равно единице.

Отделим корни третьим способом. Для этого преобразуем уравнение к равносильному виду (3.2): $x^3 = x - 1$ и найдем точки пересечения графиков $y = x^3$ и $y = x - 1$ (рис. 3.3).

Очевидно, корень уравнения $x_* \in [-2; -1]$. ■

Пример 3.3. Отделить корни уравнения $x^3 - x^2 - 9x + 9 = 0$.

□ Согласно теореме 3.1, уравнение имеет три корня, среди которых по крайней мере один действительный (по следствию из теоремы 3.2). Оценим модули корней уравнения по теореме 3.3.

Так как $A = \max\{|-1|, |-9|, |9|\} = 9$, $B = \max\{|1|, |-1|, |-9|\} = 9$, то

$$\frac{1}{1 + \frac{9}{|9|}} < |x_{*i}| < 1 + \frac{9}{|9|} \text{ или } \frac{1}{2} < |x_{*i}| < 10.$$

Отсюда

$$\frac{1}{2} < x_{*i}^+ < 10 \text{ и } -10 < x_{*i}^- < -\frac{1}{2}.$$

Определим число положительных и отрицательных корней. Выписываем коэффициенты многочлена $P_3(x)$: 1, -1, -9, 9. Так как число перемен знака $S_1 = 2$, то уравнение имеет два положительных корня или ни одного.

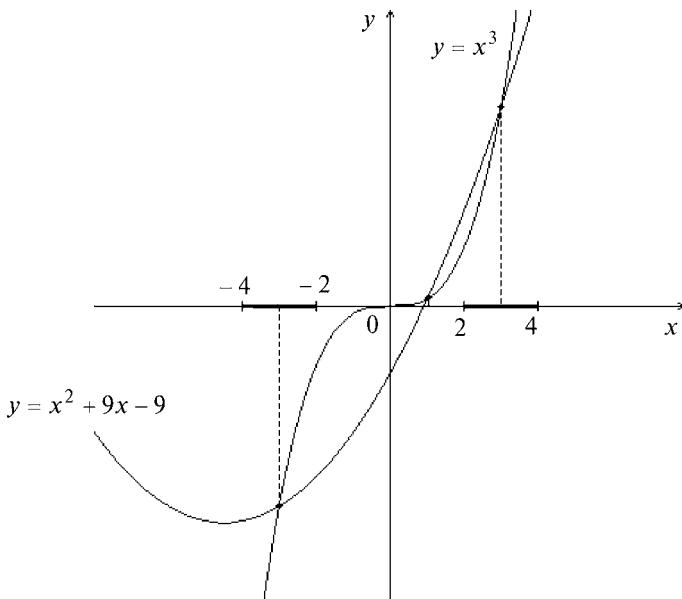


Рис. 3.4

Далее выписываем коэффициенты многочлена $P_3(-x) = -x^3 - x^2 + 9x + 9: -1, -1, 9, 9$. Так как число перемен знака $S_2 = 1$, то имеется один отрицательный корень.

Отделим корни третьим способом. С этой целью преобразуем уравнение к равносильному виду: $x^3 = x^2 + 9x - 9$. Найдем абсциссы точек пересечения графиков $y = x^3$ и $y = x^2 + 9x - 9$ (на рис. 3.4 указаны два из трех полученных промежутков).

Результат отделения корней — три промежутка $[0,5; 2]$, $[2; 4]$, $[-4; -2]$. Заметим, что отрезки могут быть сужены, например вместо отрезка $[2; 4]$ можно принять $[2,5; 4]$. ■

3.1.3. МЕТОД ПОЛОВИННОГО ДЕЛЕНИЯ

Пусть дано уравнение $f(x) = 0$ и отделен простой корень x_* , т. е. найден такой отрезок $[a_0, b_0]$, что $x_* \in [a_0, b_0]$, и на концах отрезка функция имеет значения, противоположные по знаку ($f(a_0) \cdot f(b_0) < 0$). Отрезок $[a_0, b_0]$ называется *начальным интервалом неопределенности*, потому что известно, что корень ему принадлежит, но его местоположение с требуемой точностью не определено.

Процедура уточнения положения корня заключается в построении последовательности вложенных друг в друга отрезков, каждый из которых содержит корень уравнения. Для этого находится середина текущего интервала неопределенности $c_k = \frac{a_k + b_k}{2}$, $k = 0, 1, \dots$ и в качестве следующего интервала неопределенности из двух возможных выбирается тот, на концах которого функция $f(x)$ имеет разные знаки (рис. 3.5).

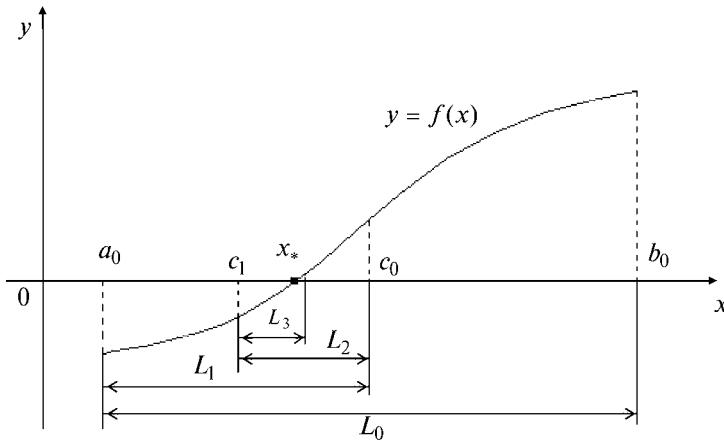


Рис. 3.5

Процесс завершается, когда длина текущего интервала неопределенности становится меньше заданной величины ε , задающей точность нахождения корня. В качестве приближенного значения корня берется середина последнего интервала неопределенности.

Методика решения задачи

1. Найти начальный интервал неопределенности $L_0 = [a_0, b_0]$ одним из методов отделения корней, задать малое положительное число ε . Положить $k = 0$.
2. Найти середину текущего интервала неопределенности:

$$c_k = \frac{a_k + b_k}{2}.$$

3. Если $f(a_k) \cdot f(c_k) < 0$, то положить $a_{k+1} = a_k$, $b_{k+1} = c_k$, а если $f(c_k) \cdot f(b_k) < 0$, то принять $a_{k+1} = c_k$, $b_{k+1} = b_k$. В результате находится текущий интервал неопределенности $L_{k+1} = [a_{k+1}, b_{k+1}]$.

4. Если $b_{k+1} - a_{k+1} \leq \varepsilon$, то процесс завершить: $x_* \in L_{k+1} = [a_{k+1}, b_{k+1}]$. Приближенное значение корня можно найти по формуле

$$x_* \cong \frac{a_{k+1} + b_{k+1}}{2}.$$

Если $b_{k+1} - a_{k+1} > \varepsilon$, положить $k = k + 1$ и перейти к п. 2.

Замечания

1. Метод имеет линейную, но безусловную сходимость, и его погрешность за каждую итерацию уменьшается в два раза:

$$|b_1 - a_1| = \frac{|b_0 - a_0|}{2}; |b_2 - a_2| = \frac{|b_1 - a_1|}{2} = \frac{|b_0 - a_0|}{2^2}, \dots; |b_k - a_k| = |b_0 - a_0| \cdot 2^{-k}.$$

Из последнего соотношения можно оценить число итераций k для достижения заданной точности ε :

$$|b_0 - a_0| \cdot 2^{-k} \leq \varepsilon; \quad k \geq \log_2 \frac{(b_0 - a_0)}{\varepsilon}.$$

Отсюда видно, что, например, для достижения точности $\varepsilon \approx 10^{-3}$ при $(b_0 - a_0) \cong 1$ необходимо выполнить примерно десять итераций.

2. К достоинствам метода следует отнести то, что он позволяет найти простой корень уравнения $x_* \in [a_0, b_0]$ для любых непрерывных функций $f(x)$ при любых значениях a_0, b_0 , таких, что $f(a_0) \cdot f(b_0) < 0$. Недостатки метода — он не обобщается на системы нелинейных уравнений и не может использоваться для нахождения корней четной кратности.

Пример 3.4. Найти корень уравнения $x^3 - x + 1 = 0$ методом половинного деления с точностью $\varepsilon_1 = 0,01$ и $\varepsilon_2 = 0,0005$.

□ В примере 3.2 корень уравнения был отделен: $x_* \in [-2; -1]$, поэтому $a_0 = -2, b_0 = -1$.

Очевидно, функция непрерывна на отрезке $[-2; -1]$, имеет единственный простой корень (см. рис. 3.3). На концах отрезка функция имеет значения $f(-2) = -5, f(-1) = 1$, противоположные по знаку. Результаты расчетов приведены в таблице 3.1.

Таблица 3.1

k	$f(a_k)$	a_k	b_k	$f(b_k)$	$c_k = \frac{a_k + b_k}{2}$	$f(c_k)$	$a_k - b_k$
0	-5	-2	-1	1	-1,5	-0,875	1
1	-0,875	-1,5	-1	1	-1,25	0,2965	0,5
2	-0,875	-1,5	-1,25	0,2965	-1,375	-0,224	0,25
3	-0,224	-1,375	-1,25	0,2965	-1,3125	0,05	0,125
4	-0,224	-1,375	-1,3125	0,05	-1,34375	-0,08	0,0625
5	-0,08	-1,34375	-1,3125	0,05	-1,3282	-0,015	0,03125
6	-0,015	-1,3282	-1,3125	0,05	-1,3204	0,018	0,0156
7	-0,015	-1,3282	-1,3204	0,018	-1,3243	0,0018	0,00781
8	-0,015	-1,3282	-1,3243	0,0018	-1,3263	-0,007	0,0039
9	-0,007	-1,3263	-1,3243	0,0018	-1,3253	-0,0025	0,002
10	-0,0025	-1,3253	-1,3243	0,0018	-1,3248	-0,0003	0,001
11	-0,0003	-1,3248	-1,3243	0,0018	—	—	0,0005

Если $\varepsilon_1 = 0,01$, корень $x_* \in [-1,3282; -1,3204]$, а если $\varepsilon_2 = 0,0005$ — корень $x_* \in [-1,3248; -1,3243]$ или $x_* \cong \frac{-1,3282 - 1,3204}{2} = -1,3243$ при $\varepsilon_1 = 0,01$; $x_* \cong \frac{-1,3248 - 1,3243}{2} = -1,3245$ при $\varepsilon_2 = 0,0005$. ■

Пример 3.5. Найти корень уравнения $x^3 - x^2 - 9x + 9 = 0$ методом половинного деления с точностью $\varepsilon = 0,01$.

□ В примере 3.3 были отделены корни уравнения. Уточним корень, лежащий на отрезке $[2,5; 4]$. Результаты расчетов поместим в таблицу 3.2.

k	$f(a_k)$	a_k	b_k	$f(b_k)$	$c_k = \frac{a_k + b_k}{2}$	$f(c_k)$	$a_k - b_k$
0	-4,125	2,5	4	21	3,25	3,5156	1,5
1	-4,125	2,5	3,25	3,5156	2,875	-1,3769	0,75
2	-1,3769	2,875	3,25	3,5156	3,0625	0,78149	0,375
3	-1,3769	2,875	3,0625	0,7815	2,96875	-0,3672	0,1875
4	-0,3672	2,9687	3,0625	0,7815	3,01562	0,18945	0,09375
5	-0,3672	2,9687	3,0156	0,1894	2,99218	-0,0933	0,04687
6	-0,0933	2,9921	3,0156	0,1894	3,0039	0,0469	0,02344
7	-0,0933	2,9921	3,0039	0,0469	2,99804	-0,02349	0,01172
8	-0,02349	2,99804	3,0039	0,0469	3,00097	0,011647	0,00586

В результате найден интервал $[2,99804; 3,0039]$ и приближенное значение корня $x_* \approx 3,00097$. ■

Пример 3.6. Отделить корни трансцендентного уравнения $x^2 - e^{-x} = 0$ и найти один из корней с точностью $\varepsilon = 0,01$.

1. Можно отделить корни уравнения, преобразовав его к равносильному виду $x^2 = e^{-x}$ и определив абсциссу точек пересечения графиков функций $y = f_1(x) = x^2$ и $y = f_2(x) = e^{-x}$: $x_* \in [0,5; 1]$. При этом $f(0,5) \cdot f(1) < 0$.

2. Уточним корень. Результаты расчетов приведем в таблице 3.3.

k	$f(a_k)$	a_k	b_k	$f(b_k)$	$c_k = \frac{a_k + b_k}{2}$	$f(c_k)$	$a_k - b_k$
0	-0,3565	0,5	1	0,6321	0,75	0,09013	0,5
1	-0,3565	0,5	0,75	0,09013	0,625	-0,1446	0,25
2	-0,1446	0,625	0,75	0,09013	0,6875	-0,0301	0,125
3	-0,0301	0,6875	0,75	0,09013	0,7187	0,0292	0,0625
4	-0,0301	0,6875	0,7187	0,0292	0,7031	-0,00069	0,0312
5	-0,00069	0,7031	0,7187	0,0292	0,7109	0,0142	0,0156
6	-0,00069	0,7031	0,7109	0,0142	0,7070	0,0067	0,0078

Вычисления показывают, что $b_6 - a_6 = 0,7109 - 0,7031 = 0,0078 < \varepsilon = 0,01$.

Поэтому $x_* \approx \frac{0,7109 + 0,7031}{2} = c_6 = 0,707$. ■

3.1.4. МЕТОД ХОРД

Этот метод при тех же предположениях обеспечивает более быстрое нахождение корня, чем метод половинного деления. Для этого отрезок $[a, b]$ делится не пополам, а в отношении $|f(a)| : |f(b)|$.

Геометрически метод хорд эквивалентен замене кривой $y = f(x)$ хордой, проходящей через точки $(a, f(a))$ и $(b, f(b))$ (рис. 3.6).

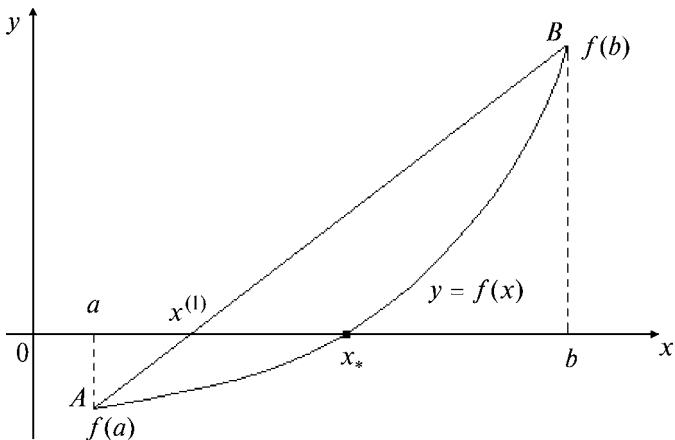


Рис. 3.6

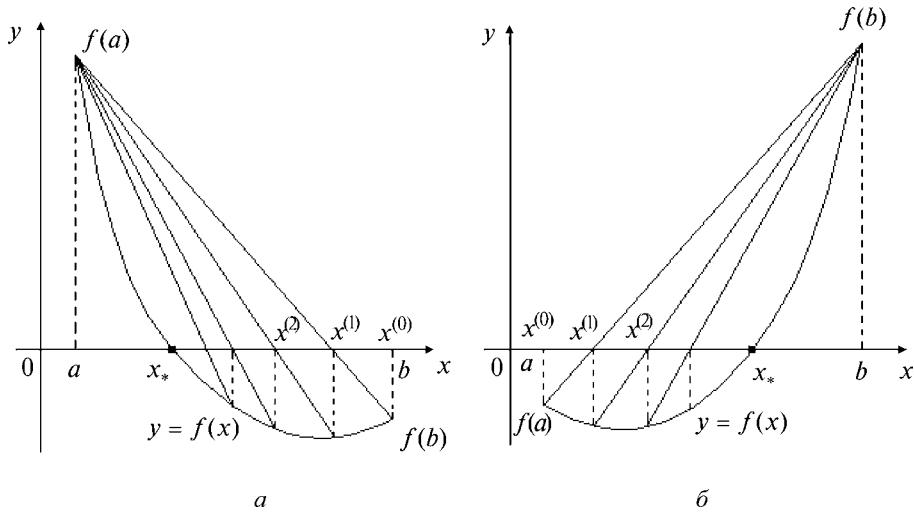


Рис. 3.7

Уравнение хорды AB имеет вид

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

Полагая $x = x^{(1)}$ и $y = 0$, получаем

$$x^{(1)} = a - \frac{f(a)}{f(b)-f(a)}(b-a).$$

Предположим, что вторая производная $f''(x)$ сохраняет постоянный знак, и рассмотрим два случая: $f''(x) > 0$, $f''(x) > 0$ (рис. 3.7а) и $f''(x) < 0$, $f''(x) > 0$ (рис. 3.7б). Случай $f''(x) < 0$ сводится к рассматриваемому, если уравнение записать в форме:

$$-f(x) = 0.$$

Первому случаю (рис. 3.7а) соответствует формула

$$\begin{aligned}x^{(0)} &= b, \\x^{(k+1)} &= x^{(k)} - \frac{f(x^{(k)})}{f(x^{(k)}) - f(a)}(x^{(k)} - a), \quad k = 0, 1, \dots,\end{aligned}\tag{3.7}$$

а второму случаю (рис. 3.7б):

$$\begin{aligned}x^{(0)} &= a, \\x^{(k+1)} &= x^{(k)} - \frac{f(x^{(k)})}{f(b) - f(x^{(k)})}(b - x^{(k)}), \quad k = 0, 1, \dots\end{aligned}\tag{3.8}$$

В первом случае остается неподвижным конец a , а во втором случае — конец b .

Замечание. Для выявления неподвижного конца используется условие $f''(x) \cdot f(t) > 0$, где $t = a$ или $t = b$. Если неподвижен конец a , применяется формула (3.7), а если конец b — формула (3.8).

Пример 3.7. Найти корень уравнения $x^3 - x + 1 = 0$ методом хорд с точностью $\varepsilon = 0,001$.

□ Рассмотрим задачу нахождения корня на отрезке $[-2; -1]$ (см. пример 3.2). Так как $f(a) = f(-2) = -5$, $f(b) = f(-1) = 1$, а $f''(x) = 6x < 0$ на отрезке $[-2; -1]$, то умножим уравнение на (-1) . Получим $f(x) = -x^3 + x - 1 = 0$. Тогда $f''(x) = -6x > 0$ на отрезке $[-2; -1]$, $f(a) = f(-2) = 5 > 0$, а $f''(x) \cdot f(a) > 0$ и, следовательно, имеем первый случай (рис. 3.7а).

Положим $x^{(0)} = b = -1$, $k = 0$. Тогда по формуле (3.7) получаем

$$x^{(1)} = x^{(0)} - \frac{f(x^{(0)})}{f(x^{(0)}) - f(a)} \cdot (x^{(0)} - a) = -1 - \frac{-1}{-1 - 5} \cdot (-1 - (-2)) = -1,1667.$$

Так как $\Delta_1 = |x^{(1)} - x^{(0)}| = 0,1667 > \varepsilon$, то положим $k = 1$ и продолжим процесс:

$$\begin{aligned}x^{(2)} &= x^{(1)} - \frac{f(x^{(1)})}{f(x^{(1)}) - f(a)} \cdot (x^{(1)} - a) = \\&= -1,1667 - \frac{-0,5786}{-0,5786 - 5} \cdot (-1,1667 - (-2)) = -1,2531.\end{aligned}$$

Так как $\Delta_2 = |x^{(2)} - x^{(1)}| = 0,0864 > \varepsilon$, то положим $k = 2$ и продолжим процесс:

$$\begin{aligned}x^{(3)} &= x^{(2)} - \frac{f(x^{(2)})}{f(x^{(2)}) - f(a)} \cdot (x^{(2)} - a) = \\&= -1,2531 - \frac{-0,2854}{-0,2854 - 5} \cdot (-1,2531 - (-2)) = -1,2934.\end{aligned}$$

Поскольку $\Delta_3 = |x^{(3)} - x^{(2)}| = 0,0403 > \varepsilon$, положим $k = 3$:

$$\begin{aligned}x^{(4)} &= x^{(3)} - \frac{f(x^{(3)})}{f(x^{(3)}) - f(a)} \cdot (x^{(3)} - a) = \\&= -1,2934 - \frac{-0,1297}{-0,1297 - 5} \cdot (-1,2934 - (-2)) = -1,3113.\end{aligned}$$

Так как $\Delta_4 = |x^{(4)} - x^{(3)}| = 0,0179 > \varepsilon$, положим $k = 4$:

$$x^{(5)} = -1,3113 - \frac{-0,0565}{-0,0565 - 5} \cdot (-1,3113 - (-2)) = -1,319.$$

Поскольку $\Delta_5 = |x^{(5)} - x^{(4)}| = 0,00769 > \varepsilon$, положим $k = 5$:

$$x^{(6)} = -1,319 - \frac{-0,0243}{-0,0243 - 5} \cdot (-1,319 - (-2)) = -1,3223.$$

Так как $\Delta_6 = |x^{(6)} - x^{(5)}| = 0,00328 > \varepsilon$, положим $k = 6$:

$$x^{(7)} = -1,3223 - \frac{-0,0103}{-0,0103 - 5} \cdot (-1,3223 - (-2)) = -1,3237.$$

Поскольку $\Delta_7 = |x^{(7)} - x^{(6)}| = 0,00139 > \varepsilon$, положим $k = 7$:

$$x^{(8)} = -1,3237 - \frac{-0,00434}{-0,00434 - 5} \cdot (-1,3237 - (-2)) = -1,3243.$$

Так как $\Delta_8 = |x^{(8)} - x^{(7)}| = 0,000586 < \varepsilon = 0,001$, то корень уравнения $x_* \approx -1,3243$. Из анализа поведения Δ_k следует, что сходимость метода хорд линейная, однако более быстрая, чем сходимость метода половинного деления. ■

3.1.5. МЕТОД ПРОСТЫХ ИТЕРАЦИЙ

Пусть известно, что корень x_* уравнения $f(x) = 0$ лежит на отрезке $G = \{a \leq x \leq b\}$.

Методика решения задачи

1. Уравнение $f(x) = 0$ равносильным преобразованием привести к виду $x = \varphi(x)$. Это преобразование может быть осуществлено различными путями, но для сходимости нужно обеспечить выполнение условия $|\varphi'(x)| \leq \chi < 1$ (χ — некоторая константа). При этом геометрически задача сводится к нахождению абсциссы точки пересечения прямой $y = x$ и кривой $y = \varphi(x)$ (рис. 3.8).

2. Задать начальное приближение $x^{(0)} \in [a, b]$ и малое положительное число ε . Положить $k = 0$.

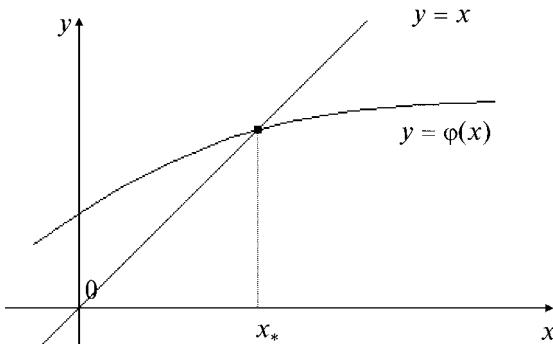


Рис. 3.8

3. Вычислить следующее приближение:

$$x^{(k+1)} = \varphi(x^{(k)}). \quad (3.9)$$

4. Если $|x^{(k+1)} - x^{(k)}| \leq \varepsilon$, итерации завершаются и $x_* \cong x^{(k+1)}$. Если $|x^{(k+1)} - x^{(k)}| > \varepsilon$, положить $k = k + 1$ и перейти к п. 3.

Замечание. В качестве условия завершения итераций при известном значении χ может быть использовано неравенство

$$|x^{(k+1)} - x^{(k)}| \leq \frac{1-\chi}{\chi} \varepsilon.$$

Проблемы сходимости и единственности численного решения, являющиеся главными при использовании этого метода, решаются и исследуются с помощью понятия о сжимающем отображении и теоремы о достаточном условии сходимости метода.

Отображение (функция) $\varphi(x)$ называется *сжимающим* в области G с коэффициентом χ ($0 \leq \chi < 1$), если для любых двух x' , x'' из G выполнено неравенство

$$|\varphi(x'') - \varphi(x')| \leq \chi |x'' - x'|. \quad (3.10)$$

Теорема 3.9 (о сходимости метода простых итераций и единственности получаемого численного решения).

Пусть выполнены условия:

1. Нелинейное уравнение $x = \varphi(x)$ имеет решение $x_* \in G$.
2. Отображение $\varphi(x)$ является сжимающим в области G с некоторым коэффициентом χ ($0 \leq \chi < 1$).

Тогда:

- a) решение x_* является единственным решением в области G ;
- б) последовательность $x^{(0)}, x^{(1)}, \dots, x^{(k+1)}, \dots$, определяемая по отображению на основе итерационного процесса, сходится к решению x_* со скоростью геометрической прогрессии, т. е. при выборе $x^{(0)}$ из условия $|x_* - x^{(0)}| < R$, где $R > 0$ — некоторое малое число, справедливо неравенство $|x_* - x^{(k)}| \leq \chi^k |x_* - x^{(0)}|$, $k = 0, 1, \dots$

Теорема 3.9 утверждает, что при выполнении условий 1, 2 существует окрестность $G(x_*, R)$, такая, что если взять $x^{(0)}$ в этой окрестности и вычислять $x^{(1)}, x^{(2)}, \dots$ по формуле (3.9), то в результате с любой наперед заданной точностью можно вычислить $x^{(k+1)} \approx x_*$, соответствующее искомому (единственному) корню. Но так как эта окрестность неизвестна, то можно взять произвольное $x^{(0)} \in G$. Если при этом вычисляется последовательность $x^{(0)}, x^{(1)}, x^{(2)}, \dots, x^{(k)}, \dots$, сходящаяся к некоторому значению \hat{x} , то в силу теоремы $\hat{x} = x_*$. Если сходимость отсутствует, то надо взять другое $x^{(0)} \in G$ и повторить расчет.

Теорема 3.10 (о достаточном условии сходимости метода простых итераций).

Пусть выполнены условия:

1. Функция $\varphi(x)$ имеет производные для всех $x \in G$.
2. Существует число χ ($0 \leq \chi < 1$, $\chi = \text{const}$), такое, что $|\varphi'(x)| \leq \chi$ для всех $x \in G$.

Тогда отображение $\varphi(x)$ является сжимающим в G с коэффициентом сжатия χ и последовательность $x^{(0)}, x^{(1)}, \dots, x^{(k+1)}, \dots$, определяемая на основе итерационного процесса, сходится к решению x_* , т. е. $x^{(k)} \rightarrow x_*$ при $k \rightarrow \infty$.

Доказательство. В силу условия 1 справедлива теорема Лагранжа о конечных приращениях:

$$\varphi(x'') - \varphi(x') \leq (x'' - x')\varphi'(\delta) \text{ для всех } x', x'' \in G,$$

где δ — точка, лежащая между x' и x'' . Но тогда с учетом условия 2 и замены $\varphi'(\delta)$ ее мажорантой χ получим

$$|\varphi(x'') - \varphi(x')| \leq \chi|x'' - x'|,$$

т. е. отображение $\varphi(x)$ — сжимающее, и, таким образом, последовательность $x^{(0)}, x^{(1)}, \dots, x^{(k+1)}, \dots$ сходится к некоторому \hat{x} . Но тогда в силу непрерывности $\varphi(x)$ в G имеет место равенство $\hat{x} = x_*$.

Геометрическая интерпретация процесса сходимости и расходимости в зависимости от выполнения или невыполнения достаточного условия сходимости представлена на рисунке 3.9.

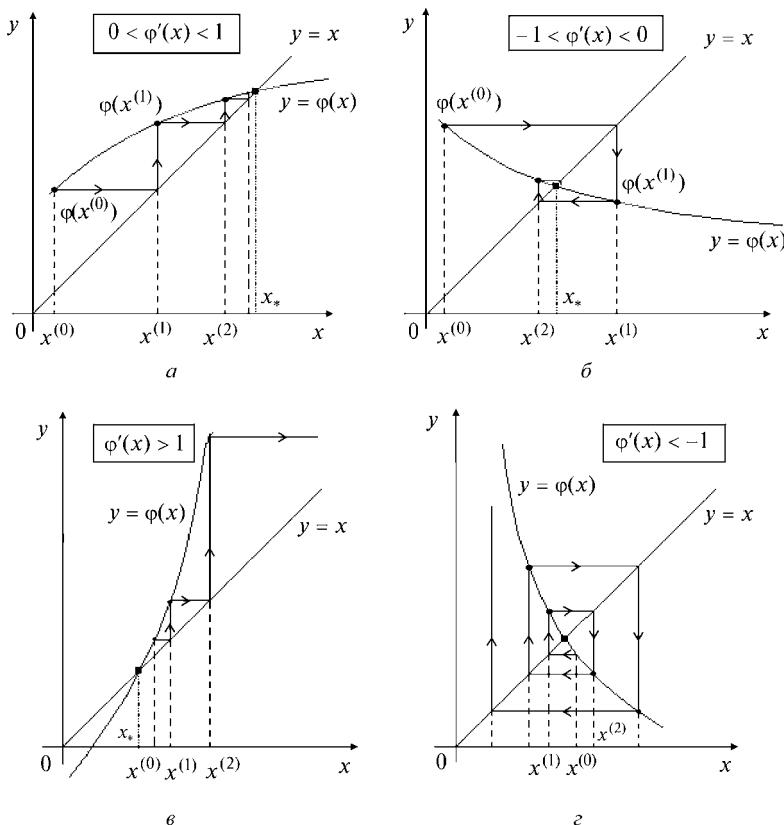


Рис. 3.9

Из рисунка 3.9 видно, что при $0 < \varphi'(x) < 1$ и при $-1 < \varphi'(x) < 0$ (рис. 3.9а, б) итерационные последовательности $x^{(0)}, x^{(1)}, x^{(2)}, \dots$ сходятся к x_* . Причем в первом случае реализуется односторонняя (монотонная) сходимость, а во втором — двусторонняя (немонотонная). При $|\varphi'(x)| > 1$ (рис. 3.9в, г) процесс расходится, несмотря на то, что точка $x^{(0)}$ очень близка к x_* . Процессы на рисунке 3.9а, б соответствуют сжимающему отображению.

Преобразование уравнения $f(x) = 0$ к равносильному виду $x = \varphi(x)$ может быть выполнено неоднозначно. Рассмотрим универсальные практические приемы равносильного преобразования $f(x) = 0$, дающие сжимающие отображения.

1. Можно заменить уравнение $f(x) = 0$ на равносильное $x = x + cf(x)$, где $c = \text{const} \neq 0$. Тогда, принимая правую часть этого уравнения за $\varphi(x)$ и раскрывая $|\varphi'(x)| = |1 + cf'(x)| < 1$, получаем условие

$$-2 < cf'(x) < 0. \quad (3.11)$$

Таким образом, можно найти константу c на отрезке $G = [a, b]$ так, чтобы удовлетворялись неравенства (3.11). При этом надо стремиться получить такую постоянную c , которая бы больше отличалась от нуля, и тогда будет реализовываться более быстрая сходимость.

Можно показать, что абсолютная погрешность метода оценивается по формуле (эта оценка выражается через два последовательных приближения)

$$|x^{(k+1)} - x_*| \leq \frac{\chi}{1-\chi} \cdot |x^{(k+1)} - x^{(k)}| \quad (3.12)$$

или по формуле (эта оценка выражается через два приближения $x^{(0)}$ и $x^{(1)}$)

$$|x^{(k+1)} - x_*| \leq \frac{\chi^{k+1}}{1-\chi} \cdot |x^{(1)} - x^{(0)}|. \quad (3.13)$$

Из оценки следует условие выхода из итерационного процесса. Если значение χ неизвестно, то выход осуществляется по условию

$$|x^{(k+1)} - x^{(k)}| \leq \varepsilon.$$

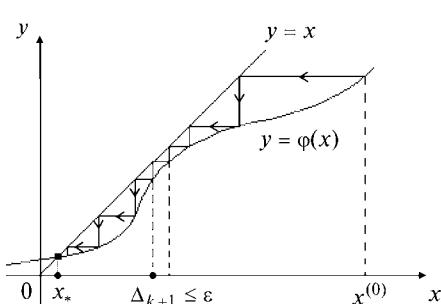


Рис. 3.10

2. Уравнение $f(x) = 0$ заменяется равносильным:

$$x = x \mp \frac{f(x)}{\max |f'(x)|} \equiv \varphi(x) \text{ при } x \in G,$$

где знак в правой части выбирается из условия $|\varphi'(x)| < 1$.

3. Можно выразить x из уравнения $f(x) = 0$ так, чтобы для полученного уравнения $x = \varphi(x)$ выполнялось условие сходимости $|\varphi'(x)| < 1$ в окрестности искомого корня.

Замечание. Выполнение условия $\Delta_{k+1} = |x^{(k+1)} - x^{(k)}| \leq \varepsilon$ не гарантирует близости к точному решению (на рисунке 3.10 достигается область, где усло-

вие окончания выполняется, но корень x_* расположен достаточно далеко от этой области).

Пример 3.8. Методом простых итераций с точностью $\varepsilon = 0,001$ уточнить корень трансцендентного уравнения $x^2 - e^{-x} = 0$, причем искомый корень $x_* \in [0,5; 1,0] \equiv G$.

□1. Преобразуем исходное уравнение к виду $x = \varphi(x)$: $x = e^{-\frac{x}{2}} = \varphi(x)$. Приверкой можно убедиться, что $\varphi'(x) < 1$ при $x \in G$.

2. В качестве начального приближения выберем $x^{(0)} = \frac{0,5+1,0}{2} = 0,75$.

3, 4. Выполняем последовательные действия по формуле (3.9): $x^{(k+1)} = e^{-\frac{x^{(k)}}{2}}$.

Результаты расчетов приведены в таблице 3.4.

Таблица 3.4

k	0	1	2	3	4	5
$x^{(k)}$	0,7500	0,6873	0,7091	0,7015	0,7042	0,7032
$ x^{(k)} - x^{(k-1)} $	—	0,0627	0,0218	0,0076	0,0027	0,0010

На пятой итерации выполнилось условие $|x^{(5)} - x^{(4)}| \leq \varepsilon = 0,001$, поэтому процесс завершен. В качестве приближенного решения берется $x_* \approx x^{(5)} = 0,703$.

Подстановка $x^{(5)}$ в исходное уравнение дает $(x^{(5)})^2 - e^{-x^{(5)}} = 0,00088$, т. е. равенство в уравнении также выполняется с заданной точностью. Видно, что сходимость двусторонняя (это следует также из неравенства $-1 < \varphi'(x) < 0$) и линейная, так как отношения $\left| \frac{x^{(k+1)} - x^{(k)}}{x^{(k)} - x^{(k-1)}} \right|$ при $k = 1, 2, 3, 4, 5$ примерно одинаковые и равны $\approx 0,35$ (знаменатель геометрической прогрессии). ■

Пример 3.9. Найти решение уравнения $x^3 - x + 1 = 0$ методом простых итераций с точностью $\varepsilon_1 = 0,01$ и $\varepsilon_2 = 0,001$.

□1. В примере 3.2 был отделен корень уравнения $x_* \in [-2; -1]$. Преобразуем уравнение к виду $x = \varphi(x)$. Для этого запишем его сначала в форме $x = x^3 + 1$. Легко показать, что функция $\varphi(x) = x^3 + 1$ не удовлетворяет условию сходимости, поскольку $\varphi'(x) = 3x^2$, $\varphi'(-2) = 12 > 1$, $\varphi'(-1) = 3 > 1$. Поэтому воспользуемся другим преобразованием.

В результате получим $x = \sqrt[3]{x-1}$. Можно проверить, что $|\varphi'(x)| < 1$ на отрезке $[-2; -1]$, т. е. достаточные условия сходимости выполняются.

2. Зададим начальное приближение $x^{(0)} = -1$. Решим задачу с различной точностью ε_1 и ε_2 .

3, 4. Выполним расчеты по формуле (3.9):

$$x^{(k+1)} = \sqrt[3]{x^{(k)} - 1}, \quad k = 0, 1, 2, \dots$$

Результаты расчетов приведены в таблице 3.5.

Если $\varepsilon_1 = 0,01$, то $x_* \approx -1,3223$, а если $\varepsilon_2 = 0,001$, то $x_* \approx -1,3246$. ■

Таблица 3.5

k	0	1	2	3	4	5
$x^{(k)}$	-1,000	-1,2599	-1,3123	-1,3223	-1,3243	-1,3246
$ x^{(k)} - x^{(k-1)} $	—	0,2599	0,0524	0,0100	0,0020	0,0003

Пример 3.10. Найти корни уравнения $x^3 - x^2 - 9x + 9 = 0$ методом простых итераций с точностью $\varepsilon = 0,001$.

□1. Преобразуем уравнение к виду $x = \varphi(x)$: $x = \sqrt[3]{x^2 + 9x - 9}$.

В примере 3.3 была выполнена операция отделения корней и получены отрезки $[-4; -2]$, $[0,5; 2]$, $[2; 4]$.

Можно показать, что на отрезках $[2; 4]$, $[-4; -2]$ функция $\varphi(x) = \sqrt[3]{x^2 + 9x - 9}$ удовлетворяет условию $|\varphi'(x)| < 1$.

На отрезке $[0,5; 2]$ используем другой вид уравнения: $x = \frac{x^3}{9} - \frac{x^2}{9} + 1$. Также легко проверить, что функция $\varphi(x) = \frac{x^3}{9} - \frac{x^2}{9} + 1$ удовлетворяет достаточному условию сходимости на отрезке $[0,5; 2]$.

2. В качестве начальных приближений выберем:

- точку $x^{(0)} = -2$ на отрезке $[-4; -2]$;
- точку $x^{(0)} = 0,5$ на отрезке $[0,5; 2]$;
- точку $x^{(0)} = 2$ на отрезке $[2; 4]$.

В поставленной задаче $\varepsilon = 0,001$.

3, 4. Выполним расчеты по формуле

$$x^{(k+1)} = \sqrt[3]{(x^{(k)})^2 + 9x^{(k)} - 9}, \quad k = 0, 1, 2, \dots$$

с начальными значениями $x^{(0)} = -2$ и $x^{(0)} = 2$ и по формуле

$$x^{(k+1)} = \frac{(x^{(k)})^3}{9} - \frac{(x^{(k)})^2}{9} + 1, \quad k = 0, 1, 2, \dots$$

с начальным значением $x^{(0)} = 0,5$.

Результаты расчетов занесены в таблицы 3.6–3.8.

Таблица 3.6

k	$x^{(k)}$	$ x^{(k)} - x^{(k-1)} $
0	2,0000	—
1	2,3513	0,3513
2	2,6056	0,2543
3	2,7694	0,1638
4	2,8682	0,0988
5	2,9255	0,0573
6	2,9582	0,0327
7	2,9767	0,0185
8	2,9870	0,0102
9	2,9927	0,0057
10	2,9959	0,0032
11	2,9977	0,0018
12	2,9987	0,0010

Таблица 3.7

k	$x^{(k)}$	$ x^{(k)} - x^{(k-1)} $
0	-2,0000	—
1	-2,8438	0,8438
2	-2,9816	0,1378
3	-2,9979	0,0163
4	-2,9997	0,0018
5	-2,99997	0,00027

Таблица 3.8

k	$x^{(k)}$	$ x^{(k)} - x^{(k-1)} $
0	0,50000	—
1	0,98611	0,4861
2	0,99849	0,01238
3	0,99983	0,00134
4	0,99998	0,00015

В результате получены приближенные значения корней: $x_{1*} \approx -2,99997$, $x_{2*} \approx 0,99998$, $x_{3*} \approx 2,9987$.

Обратим внимание на сильное различие в числе итераций, потребовавшихся для нахождения корней $x_* = 3$ (табл. 3.6) и $x_* = -3$ (табл. 3.7) с помощью одной и той же формулы. Заметим, что в окрестности корня $x_* = 3$ значения модуля производной функции $\varphi(x) = \sqrt[3]{x^2 + 9x - 9}$ равны: $|\varphi'(2)| = 0,784$; $|\varphi'(2,3513)| = 0,673$; $|\varphi'(2,6056)| = 0,618$; $|\varphi'(2,9977)| = 0,556$. С другой стороны, в окрестности корня $x_* = -3$ имеем: $|\varphi'(-2)| = 0,206$; $|\varphi'(-2,8438)| = 0,124$; $|\varphi'(-2,9977)| = 0,111$. Анализ результатов показывает, что чем меньше значения модуля производной $|\varphi'(x)|$, тем быстрее сходимость. ■

3.1.6. МЕТОД НЬЮТОНА

Метод Ньютона (*метод касательных*, или *метод линеаризации*) является одним из наиболее популярных численных методов. Он быстро сходится (имеет квадратичную сходимость) и допускает различные модификации, приспособленные для решения векторных задач и сеточных уравнений. Однако этот метод эффективен при весьма жестких ограничениях на характер функции $f(x)$:

- 1) существование второй производной функции $f(x)$ на множестве $G = \{a \leq x \leq b\}$;
- 2) удовлетворение первой производной условию $f'(x) \neq 0$ для всех $x \in G$;
- 3) знакопостоянство $f'(x), f''(x)$ для всех $x \in G$.

Поэтому его желательно использовать совместно с другими методами, например методом половинного деления, чтобы достигнуть диапазона $\tilde{a} \leq x \leq \tilde{b}$, где указанные условия начинают выполняться.

Геометрическая интерпретация метода Ньютона состоит в следующем. Задается начальное приближение $x^{(0)}$. Далее проводится касательная к кривой $y = f(x)$ в точке $x^{(0)}$ (рис. 3.11), т. е. кривая заменяется прямой линией. В качестве следующего приближения выбирается точка пересечения этой касательной с осью абсцисс. Процесс построения касательных и нахождения точек пересечения с осью абсцисс повторяется до тех пор, пока приращение не станет меньше заданной величины ε .

Получим расчетную формулу метода Ньютона. Вместо участка кривой BC (точка C соответствует x_*) возьмем участок AB — касательную, проведенную в точке $(x^{(0)}, f(x^{(0)}))$. Для этого отрезка

справедливо конечное соотношение:

$$\frac{f(x^{(0)}) - 0}{x^{(0)} - x^{(1)}} = f'(x^{(0)}) \equiv \operatorname{tg} \alpha,$$

где α — угол наклона касательной в точке $(x^{(0)}, f(x^{(0)}))$ к оси абсцисс. Разрешая это соотношение относительно $x^{(1)}$, получаем

$$x^{(1)} = x^{(0)} - \frac{f(x^{(0)})}{f'(x^{(0)})}.$$

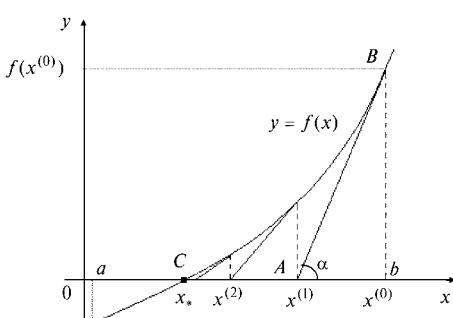


Рис. 3.11

Повторяя процесс, находим общую формулу

$$x^{(k+1)} = x^{(k)} - \frac{f(x^{(k)})}{f'(x^{(k)})}, \quad k=0,1,2,\dots \quad (3.14)$$

Подчеркнем, что если отбросить итерационный индекс, то (3.14) записывается в виде нелинейного уравнения

$$x = x - \frac{f(x)}{f'(x)} \equiv \varphi(x), \quad (3.15)$$

которое, однако, на $[a, b]$ не равносильно исходному, а является таковым только в одной точке при $x = x_*$. Поэтому данный метод не служит разновидностью метода простых итераций.

Применим теперь для вывода формулы (3.14) метод линеаризации. Положим, что итерационный процесс имеет вид

$$x^{(k+1)} = x^{(k)} + \delta^{(k)}, \quad k=0,1,2,\dots, \quad (3.16)$$

где $\delta^{(k)}$ — поправка к k -му приближению, которую необходимо найти. Предполагая, что $f(x)$ имеет непрерывную вторую производную, разложим $f(x^{(k)} + \delta^{(k)})$ по формуле Тейлора относительно точки $x^{(k)}$:

$$f(x^{(k)} + \delta^{(k)}) = f(x^{(k)}) + \delta^{(k)} f'(x^{(k)}) + \frac{(\delta^{(k)})^2}{2} f''(\xi),$$

где $\xi \in (x^{(k)}, x^{(k+1)})$. Учитывая, что $f(x^{(k)} + \delta^{(k)}) = 0$ (это соответствует нахождению точки пересечения с осью абсцисс), и оставляя только линейную (относительно $\delta^{(k)}$) часть разложения (отсюда и название — *метод линеаризации*), записываем линейное относительно $\delta^{(k)}$ уравнение

$$f(x^{(k)}) + \delta^{(k)} f'(x^{(k)}) = 0.$$

Отсюда выражается поправка $\delta^{(k)} = -\frac{f(x^{(k)})}{f'(x^{(k)})}$. Подставляя $\delta^{(k)}$ в (3.16), получаем (3.14).

Теорема 3.11 (о достаточных условиях сходимости метода Ньютона).

Пусть выполняются следующие условия.

1. Функция $f(x)$ определена и дважды дифференцируема на $[a, b]$.

2. Отрезку $[a, b]$ принадлежит только один простой корень x_* , так что $f(a) \cdot f(b) < 0$.

3. Производные $f'(x)$, $f''(x)$ на $[a, b]$ сохраняют знак, и $f'(x) \neq 0$.

4. Начальное приближение $x^{(0)}$ удовлетворяет неравенству $f(x^{(0)}) \times f''(x^{(0)}) > 0$ (знаки функций $f(x)$ и $f''(x)$ в точке $x^{(0)}$ совпадают).

Тогда с помощью метода Ньютона (3.14) можно вычислить корень уравнения $f(x) = 0$ с любой точностью.

Замечания

1. Метод Ньютона характеризуется вторым порядком сходимости вблизи корня и первым порядком — вдали от него. Данную оценку проверяем для двух случаев, когда $x^{(k)}$ находится далеко от корня x_* (это возможно на первых итерациях) и когда $x^{(k)}$ располагается вблизи x_* .

Первый случай. Пусть отрезок $\omega = [x^{(k)}, x_*]$ не мал. Тогда на основе теоремы Лагранжа получим

$$|f(x^{(k)})| = |f(x_*) - f(x^{(k)})| = |f'(\xi)| \cdot |x_* - x^{(k)}| \geq m_1 |x_* - x^{(k)}|, \quad (3.17)$$

где $m_1 = \min_{\omega} |f'(x)|$.

Преобразуем с использованием (3.17) итерационную формулу (3.14):

$$x^{(k+1)} = x^{(k)} - \frac{f(x^{(k)})}{f'(x^{(k)})}; \quad |x^{(k+1)} - x^{(k)}| = \frac{|f(x^{(k)})|}{|f'(x^{(k)})|} \geq \frac{m_1 |x_* - x^{(k)}|}{M_1},$$

где $M_1 = \max_{[a,b]} |f'(x)|$.

В силу монотонности последовательности $x^{(0)}, x^{(1)}, \dots$

$$\begin{aligned} x_* - x^{(k+1)} &= x_* - x^{(k)} - (x^{(k+1)} - x^{(k)}), \\ |x_* - x^{(k+1)}| &= |x_* - x^{(k)}| - |x^{(k+1)} - x^{(k)}| \geq |x_* - x^{(k)}| - \frac{m_1}{M_1} |x_* - x^{(k)}| = \\ &= \left(1 - \frac{m_1}{M_1}\right) |x_* - x^{(k)}|. \end{aligned}$$

Таким образом, вдали от корня получаем линейную сходимость

$$|x_* - x^{(k+1)}| \leq c \cdot |x_* - x^{(k)}|,$$

где $0 \leq c = 1 - \frac{m_1}{M_1} < 1$.

Второй случай. Пусть теперь отрезок $\omega = [x^{(k)}, x_*]$ мал, т. е. итерации выполняются вблизи корня. Тогда, полагая, что указанные выше предположения 1–4 теоремы 3.10 выполнены, разложим функцию $f(x)$ в окрестности корня x_* относительно точки $x^{(k)}$, учитывая члены до второго порядка. Получим

$$f(x)|_{x=x^{(k)}} = f(x^{(k)}) + (x_* - x^{(k)})f'(x^{(k)}) + (x_* - x^{(k)})^2 \frac{f''(\xi)}{2} = 0,$$

где $\xi \in (x^{(k)} - \delta, x^{(k)} + \delta)$ ($\delta > 0$ — малая величина). Из данного соотношения выражаем x_* :

$$x_* = x^{(k)} - \frac{f(x^{(k)})}{f'(x^{(k)})} - \frac{(x_* - x^{(k)})^2}{2f'(x^{(k)})} f''(\xi).$$

Но первые два слагаемых в правой части в соответствии с (3.14) равны $x^{(k+1)}$.

Тогда будем иметь

$$x_* - x^{(k+1)} = -\frac{(x_* - x^{(k)})^2}{2f'(x^{(k)})} f''(\xi).$$

Из последнего соотношения следует оценка погрешности $(k+1)$ -го приближения через погрешность k -го приближения:

$$|x_* - x^{(k+1)}| \leq \frac{M_2}{2m_1} |x_* - x^{(k)}|^2, \quad (3.18)$$

где принято $M_2 = \max_{\omega} |f''(x)|$, $m_1 = \min_{\omega} |f'(x)|$. Оценка (3.18) свидетельствует о квадратичной сходимости метода касательных вблизи корня. С вычислительной точки зрения это означает, что на каждом приближении количество верных цифр результата удваивается.

2. Для нахождения комплексных корней уравнения $f(z) = 0$ можно также использовать (3.14) в форме

$$z^{(k+1)} = z^{(k)} - \frac{f(z^{(k)})}{f'(z^{(k)})}, \quad k = 0, 1, \dots,$$

где $z = x + iy$ — комплексная переменная.

Метод Ньютона может применяться не только для нахождения простых корней, но для определения кратных корней, т. е. когда на отрезке $[a, b]$, содержащем корень, не выполняется условие $f(a) \cdot f(b) < 0$ (условие 2 теоремы 3.11). Наряду с теоремой 3.11 удобно также пользоваться следующей теоремой.

Теорема 3.12 (достаточные условия сходимости метода Ньютона).

Пусть:

- а) дана функция $f(x): D \rightarrow R$, где D — открытый интервал, и $f'(x) \in \text{Lip}_\gamma(D)$;
- б) для некоторого $\rho > 0$ выполняется условие $|f'(x)| \geq \rho$ при всех x из D .

Если уравнение $f(x) = 0$ имеет решение $x_* \in D$, то существует некоторое $\eta > 0$, такое, что если $|x^{(0)} - x_*| < \eta$, то последовательность $\{x^{(0)}, x^{(1)}, x^{(2)}, \dots\}$, задаваемая формулой (3.14), существует и сходится к x_* . Более того, справедлива оценка

$$|x^{(k+1)} - x_*| \leq \frac{\gamma}{2\rho} |x^{(k)} - x_*|^2, \quad k = 0, 1, 2, \dots$$

Здесь обозначение $f'(x) \in \text{Lip}_\gamma(D)$ означает, что функция $f'(x)$ непрерывна по Липшичу с константой γ на множестве D , т. е. $|f'(x) - f'(y)| \leq \gamma \cdot |x - y|$ для любых x, y из D .

Замечания

1. Требование теоремы 3.12, чтобы производная $f'(x)$ имела ненулевую нижнюю границу в D , определяет, что значение $f'(x_*)$ должно быть ненулевым для квадратичной сходимости метода Ньютона. Если же $f'(x_*) = 0$, то x_* является кратным корнем, а метод Ньютона сходится лишь линейно.

2. Выполнение условий теоремы 3.12 гарантирует сходимость метода Ньютона только из «хорошего» начального приближения $x^{(0)}$.

3. Метод Ньютона является локально сходящимся, так как он сходится с определенной скоростью к истинному решению при условии, что стартует в достаточной близости от этого решения.

Методика решения задачи

1. Задать начальное приближение $x^{(0)}$ так, чтобы выполнялось неравенство $f(x^{(0)}) \cdot f''(x^{(0)}) > 0$, а также малое положительное число ε . Положить $k = 0$.

2. Вычислить $x^{(k+1)}$ по формуле

$$x^{(k+1)} = x^{(k)} - \frac{f(x^{(k)})}{f'(x^{(k)})}.$$

3. Если $|x^{(k+1)} - x^{(k)}| \leq \varepsilon$, процесс завершить и положить $x_* \cong x^{(k+1)}$.

Если $|x^{(k+1)} - x^{(k)}| > \varepsilon$, положить $k = k + 1$ и перейти к п. 2.

Пример 3.11. Методом Ньютона с точностью $\varepsilon = 0,001$ уточнить корень трансцендентного уравнения $x^2 - e^{-x} = 0$, причем искомый корень $x_* \in [0,5; 1,0] \equiv G$.

□ Можно проверить, что на множестве G удовлетворяются условия 1, 2, 3 теоремы 3.11, обеспечивающие сходимость метода касательных.

1. Зададим начальное приближение из условия 4. Так как для $f(x) = x^2 - e^{-x}$ справедливо $f''(x) = 2 - e^{-x} > 0$ на множестве G , $f(a) = f(0,5) < 0$, $f(b) = f(1) > 0$, то $f(1) \cdot f''(1) > 0$, поэтому $x^{(0)} = 1$.

2, 3. Результаты расчетов по формуле (3.14)

$$x^{(k+1)} = x^{(k)} - \frac{(x^{(k)})^2 - e^{-x^{(k)}}}{2x^{(k)} + e^{-x^{(k)}}}, \quad k = 0, 1, \dots$$

помещены в таблицу 3.9.

Из таблицы 3.9 можно сделать следующие выводы:

1. Для достижения заданной точности, проверяемой по модулю разности $|x^{(k)} - x^{(k-1)}|$, потребовалось выполнить три приближения. Если же выход организовать по условию $|f(x^{(k)})| \leq \varepsilon$, то достаточно двух приближений.

2. Вблизи корня x_* количество верных цифр в результатах $x^{(k)}$ удваивается, так что все цифры в $x^{(3)}$ являются верными, тогда как в $x^{(2)}$ верными являются первые три цифры после запятой.

3. Скорость сходимости метода Ньютона выше скорости сходимости метода простых итераций (также точность была достигнута за пять итераций). ■

Пример 3.12. Методом Ньютона найти корень уравнения $x^3 - x + 1 = 0$.

□ В примере 3.2 корень был отделен: $x_* \in [-2; -1]$.

1. Зададим начальное приближение $x^{(0)}$. Так как $f'(x) = 3x^2 - 1$, $f''(x) = 6x$, то $f(-2) = -5$, $f''(x) < 0$ при $x \in [-2; -1]$. Поэтому $f(-2)f''(-2) > 0$ и $x^{(0)} = -2$. Положим $\varepsilon = 0,001$.

2, 3. Результаты расчетов по формуле (3.14)

$$x^{(k+1)} = x^{(k)} - \frac{(x^{(k)})^3 - x^{(k)} + 1}{3(x^{(k)})^2 - 1}, \quad k = 0, 1, \dots$$

приведены в таблице 3.10.

Найденное приближенное решение $x_* \cong -1,32472$. ■

Таблица 3.9

k	0	1	2	3
$x^{(k)}$	1,000	0,73304	0,70381	0,703467
$ x^{(k)} - x^{(k-1)} $	—	0,27	0,029	0,00034
$f(x^{(k)})$	0,63212	0,05690	0,00065	$8,25 \cdot 10^{-7}$

Таблица 3.10

k	0	1	2	3	4	5
$x^{(k)}$	-2,0000	-1,545455	-1,359615	-1,325801	-1,324719	-1,324718
$ x^{(k)} - x^{(k-1)} $	—	0,454545	0,185840	0,033814	0,001082	0,000001

Пример 3.13. Найти корни уравнения $x^3 - x^2 - 9x + 9 = 0$ методом Ньютона с точностью $\varepsilon = 0,001$.

□Процедура отделения корней была выполнена в примере 3.3. В качестве отрезков $[a_i, b_i]$, которым принадлежат корни уравнения, выберем $[-4; -2], [2,5; 4], [0,5; 2]$.

Так как $f(-4) = -3,5; f(-2) = 15$, т. е. $f(-4)f(-2) < 0$, производные $f''(x) = 6x - 2 < 0, f'(x) = 3x^2 - 2x - 9 > 0$ сохраняют знак при $x \in [-4; -2]$, то условия сходимости выполняются.

Так как $f(2,5) = -4,125; f(4) = 21$, т. е. $f(2,5)f(4) < 0$, и производные $f'(x) > 0, f''(x) > 0$ сохраняют знак при $x \in [2,5; 4]$, то условия сходимости на этом отрезке тоже выполняются.

Так как $f(0,5) = 4,375; f(2) = -5$, т. е. $f(0,5)f(2) < 0$, и производные $f'(x) < 0, f''(x) > 0$ сохраняют знак при $x \in [0,5; 2]$, то условия сходимости выполняются.

1. Зададим начальные приближения: на отрезке $[0,5; 2]$ выберем $x^{(0)} = 0,5$, так как $f(0,5) \cdot f''(0,5) > 0$; на отрезке $[-4, -2]$ выберем $x^{(0)} = -4$, так как $f(4) \cdot f''(4) > 0$; аналогично на отрезке $[2,5; 4]$ выберем $x^{(0)} = 4$. В поставленной задаче $\varepsilon = 0,001$.

2, 3. Результаты расчетов по формуле (3.14)

$$x^{(k+1)} = x^{(k)} - \frac{(x^{(k)})^3 - (x^{(k)})^2 - 9x^{(k)} + 9}{3(x^{(k)})^2 - 2x^{(k)} - 9}, \quad k = 0, 1, \dots$$

приведены в таблицах 3.11–3.13.

В результате получены приближенные значения корней: $x_{*1} \approx -3,0000; x_{*2} \approx 1,0000; x_{*3} \approx 3,0000$. ■

Пример 3.14. Найти корень уравнения $x^2 - 2x + 1 = 0$ методом Ньютона с точностью $\varepsilon = 0,01$.

Таблица 3.11

k	0	1	2	3	4
$x^{(k)}$	-4,00000	-3,255319	-3,023383	-3,000225	-3,000000
$ x^{(k)} - x^{(k-1)} $	—	0,744681	0,231936	0,023158	0,000225

Таблица 3.12

k	0	1	2	3
$x^{(k)}$	0,5000	0,972973	0,9998246	1,0000000
$ x^{(k)} - x^{(k-1)} $	—	0,472973	0,0268516	0,0001754

Таблица 3.13

k	0	1	2	3	4	5
$x^{(k)}$	4,0000	3,322581	3,051484	3,001674	3,0000002	3,0000
$ x^{(k)} - x^{(k-1)} $	—	0,6774194	0,2710969	0,049809	0,001671	$2 \cdot 10^{-6}$

Таблица 3.14

k	0	1	2	3	4	5	6	7
$x^{(k)}$	2	1,5	1,25	1,125	1,0625	1,03125	1,015625	1,007813
$ x^{(k)} - x^{(k-1)} $	—	0,5	0,25	0,125	0,0625	0,03125	0,015625	0,0078125

□ Очевидно, уравнение имеет один кратный корень $x_* = 1$. Согласно замечанию к теореме 3.12 применение метода Ньютона для решения поставленной задачи сопровождается линейной сходимостью (табл. 3.14).

Полученное приближенное решение $x_* \approx 1,007813$. ■

3.1.7. МОДИФИКАЦИИ МЕТОДА НЬЮТОНА

Метод касательных, являясь весьма эффективным средством численного анализа, к сожалению, имеет достаточно жесткие ограничения. Действительно, он не может применяться для сеточных уравнений (рис. 3.2); при нарушении знакопостоянства производных (рис. 3.12); при существовании неограниченных вторых производных и др. Так, если даже условие знакопостоянства нарушено вдали от корня, где выбрано $x^{(0)}$, а вблизи корня выпол-

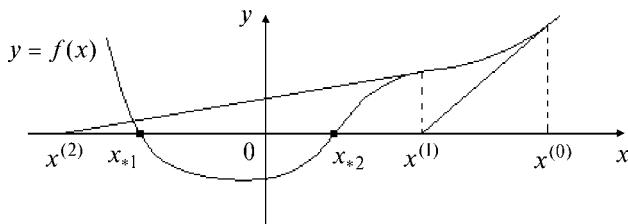


Рис. 3.12

няется, то все равно метод касательных неприменим (рис. 3.12), если не произвести сужения начального отрезка.

Кроме того, если функция $f(x)$ очень сложная, то будет сложной и ее производная, и поэтому на каждой итерации приходится рассчитывать две функции, что снижает эффективность метода касательных.

В силу этого в ряде случаев могут оказаться более предпочтительными модификации метода касательных. Рассмотрим три из них.

А. Упрощенный метод Ньютона. Методика его применения совпадает с изложенной в п. 3.1.6, но вместо формулы (3.14) используется

$$x^{(k+1)} = x^{(k)} - \frac{f(x^{(k)})}{f'(x^{(0)})}, \quad k = 0, 1, \dots \quad (3.19)$$

Отличие от метода Ньютона заключается в том, что производная функции $f(x)$ подсчитывается только в точке начального приближения, а на последующих итерациях не уточняется. Процесс последовательных приближений отражен на рисунке 3.13. Первая итерация совпадает с первой итерацией метода Ньютона. На последующих итерациях соответствующие отрезки параллельны касательной, проведенной в начальной точке.

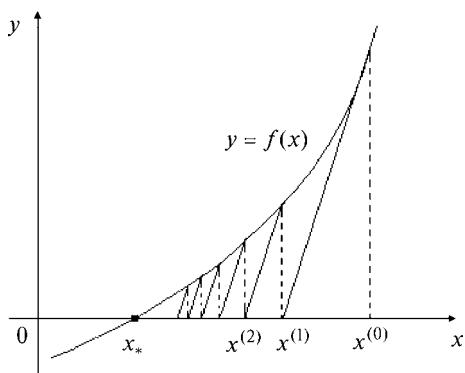


Рис. 3.13

Для этой модификации снимаются некоторые ограничения метода касательных, например условие знакопостоянства производных. Сходимость упрощенного метода Ньютона линейная.

Пример 3.15. Найти корень уравнения $x^3 - x + 1 = 0$ упрощенным методом Ньютона.

□ Корень уравнения отделен в примере 3.2: $x_* \in [-2; -1]$.

1. Выберем начальное приближение $x^{(0)} = -2$ и зададим $\varepsilon_1 = 0,01$ и $\varepsilon_2 = 0,001$.

2, 3. Выполним расчеты по формуле (3.19):

$$x^{(k+1)} = x^{(k)} - \frac{(x^{(k)})^3 - x^{(k)} + 1}{3(x^{(k)})^2 - 1} = x^{(k)} - \frac{(x^{(k)})^3 - x^{(k)} + 1}{11}, \quad k = 0, 1, \dots$$

Результаты расчетов приведены в таблице 3.15.

При $\varepsilon = 0,01$ получено решение $x_* \approx -1,3388$, а при $\varepsilon = 0,001$ — решение $x_* \approx -1,3259$. Очевидно, по сравнению с методом Ньютона сходимость замедляется (см. пример 3.12). ■

Б. Метод Ньютона — Бройдена. Этот метод позволяет увеличить скорость сходимости последовательных приближений благодаря использованию формулы

$$x^{(k+1)} = x^{(k)} - c_k \frac{f(x^{(k)})}{f'(x^{(k)})}, \quad k = 0, 1, \dots, \quad (3.20)$$

где c_k — число, которое выбирается на каждой итерации так, чтобы уменьшить значение $|f(x^{(k+1)})|$ по сравнению с $|f(x^{(k)})|$. При $c_k = 1$ метод Ньютона — Бройдена совпадает с методом Ньютона.

Как правило, при плохой сходимости или ее отсутствии полагают $0 < c_k < 1$ (рис. 3.14a), а при хорошей сходимости для $c_k = 1$ полагают $c_k > 1$ (это ускоряет сходимость (рис. 3.14б)).

На рисунке 3.14 прямоугольниками отмечены точки $x^{(1)}$, полученные при $c_0 = 1$, $\delta^{(k)} = \frac{f(x^{(k)})}{f'(x^{(k)})}$, $k = 0, 1, \dots$ — поправка, соответствующая методу Ньютона, а точки $x^{(1)} = x^{(0)} - c_0 \delta^{(0)}$ получены по методу Ньютона — Бройдена.

В. Метод секущих. В этом методе производная функции $f(x)$ подсчитывается с помощью конечно-разностных соотношений:

- в точке $x^{(0)}$ используется формула

$$f'(x^{(0)}) \approx \frac{f(x^{(0)}) - f(x^{(0)} - \delta)}{\delta},$$

где δ — малая положительная величина;

- в точках $x^{(k)}$, $k = 1, 2, \dots$ используется формула

$$f'(x^{(k)}) \approx \frac{f(x^{(k)}) - f(x^{(k-1)})}{x^{(k)} - x^{k-1}}.$$

Вычисленное значение $f'(x^{(k)})$ определяет тангенс угла наклона секущей (рис. 3.15).

k	$x^{(k)}$	$ x^{(k)} - x^{(k-1)} $	k	$x^{(k)}$	$ x^{(k)} - x^{(k-1)} $
0	-2,0000	—	6	-1,3388	0,0092
1	-1,5455	0,4545	7	-1,3333	0,0005
2	-1,4413	0,1042	8	-1,3299	0,00034
3	-1,3911	0,0502	9	-1,3279	0,0020
4	-1,3637	0,0274	10	-1,3267	0,0012
5	-1,3480	0,0157	11	-1,3259	0,0008

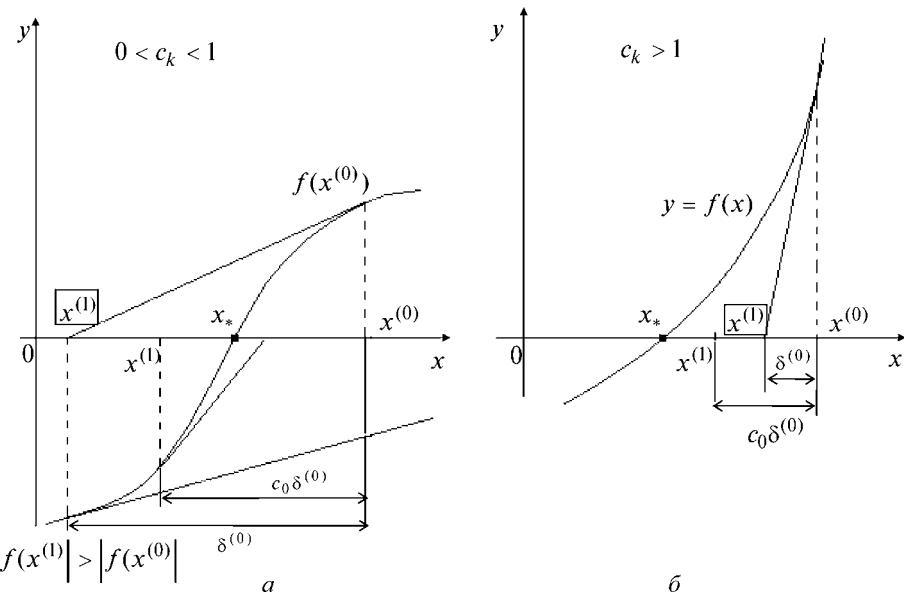


Рис. 3.14

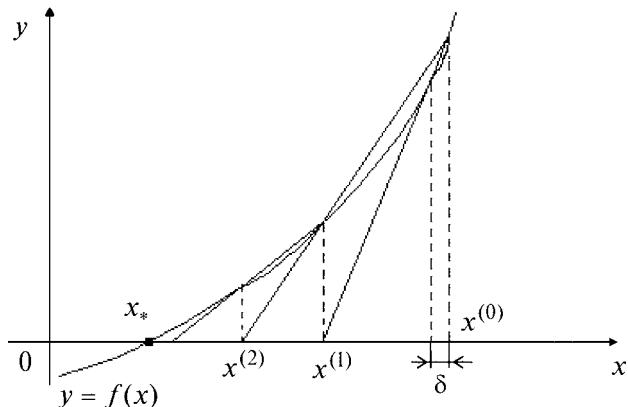


Рис. 3.15

Методика применения метода секущих совпадает с описанной в п. 3.1.6, но вместо (3.14) используется формула

$$x^{(k+1)} = x^{(k)} - \frac{f(x^{(k)})}{f(x^{(k)}) - f(x^{(k-1)})} \cdot (x^{(k)} - x^{(k-1)}), \quad k=1,2,\dots \quad (3.21)$$

Замечания

1. Метод секущих является более экономичным по сравнению с методом Ньютона по количеству функций, подлежащих расчету: на каждой итерации в методе секущих необходимо рассчитать только значение $f(x^{(k)})$, так как величина $f(x^{(k-1)})$ уже подсчитана на предыдущей итерации.

2. Метод секущих может применяться и для решения сеточных уравнений. Для сеточной функции производные в методе секущих определяются так же, но для определения $f(x^{(k)})$ в промежуточных точках осуществляется аппроксимация (как правило, интерполяция) функции $f(x_i)$ (рис. 3.16). На этом рисунке штриховой линией показана аппроксимационная кривая.

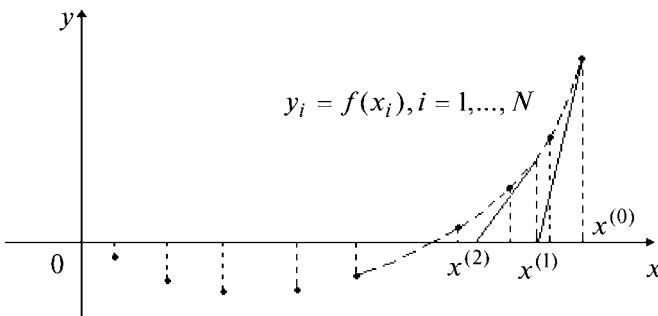


Рис. 3.16

3. Для всех описанных модификаций скорость сходимости p по сравнению с методом касательных снижается: $p < 2$. Однако для некоторых из них (метод секущих) значение $p > 1$ и может достигать $p = 1,5$.

Пример 3.16. Методом секущих найти корень уравнения $x^3 - x + 1 = 0$ с точностью $\varepsilon = 0,001$.

□1. Зададим начальное приближение $x^{(0)} = -2$ (см. пример 3.12).

2. Для вычисления $f'(x^{(0)})$ зададим $\delta = 0,1$. Тогда

$$f'(x^{(0)}) \approx \frac{f(-2) - f(-2,1)}{0,1} = \frac{-5 - (-6,161)}{0,1} = 11,61.$$

Отсюда

$$x^{(1)} = x^{(0)} - \frac{f(x^{(0)})}{f'(x^{(0)})} = -2 - \frac{-5}{11,61} = -1,56934.$$

Дальнейшие расчеты выполняются по формуле (3.21):

$$x^{(2)} = x^{(1)} - \frac{f(x^{(1)})}{f(x^{(1)}) - f(x^{(0)})} \cdot (x^{(1)} - x^{(0)}) = -1,56934 - \frac{-1,29567}{3,704325} \cdot 0,43066 = -1,41871;$$

$$\begin{aligned}
x^{(3)} &= x^{(2)} - \frac{f(x^{(2)})}{f(x^{(2)}) - f(x^{(1)})} \cdot (x^{(2)} - x^{(1)}) = -1,41871 - \frac{-0,43678}{0,858893} \cdot 0,15063 = -1,34211; \\
x^{(4)} &= x^{(3)} - \frac{f(x^{(3)})}{f(x^{(3)}) - f(x^{(2)})} \cdot (x^{(3)} - x^{(2)}) = -1,34211 - \frac{-0,07538}{0,361404} \cdot 0,0766 = -1,32613; \\
x^{(5)} &= x^{(4)} - \frac{f(x^{(4)})}{f(x^{(4)}) - f(x^{(3)})} \cdot (x^{(4)} - x^{(3)}) = -1,32613 - \frac{-0,00603}{0,069348} \cdot 0,01598 = -1,32474; \\
x^{(6)} &= x^{(5)} - \frac{f(x^{(5)})}{f(x^{(5)}) - f(x^{(4)})} \cdot (x^{(5)} - x^{(4)}) = -1,32474 - \frac{-0,000139}{0,005936} \cdot 0,00139 = \\
&= -1,32472 \approx x_*.
\end{aligned}$$

Результаты расчетов приведены в таблице 3.16.

Таблица 3.16

k	0	1	2	3	4	5	6
$x^{(k)}$	-2,00000	-1,56934	-1,41871	-1,34211	-1,32613	-1,32474	-1,32472
$ x^{(k)} - x^{(k-1)} $	—	0,43066	0,15063	0,07660	0,01598	0,00139	0,00002

Очевидно, метод сходится чуть хуже метода Ньютона (см. пример 3.12), однако скорость сходимости выше линейной. ■

3.2. МЕТОДЫ РЕШЕНИЯ СИСТЕМ НЕЛИНЕЙНЫХ УРАВНЕНИЙ

3.2.1. ПОСТАНОВКА ЗАДАЧИ

Дана система n нелинейных уравнений с n неизвестными:

$$\begin{aligned}
f_1(x_1, \dots, x_n) &= 0, \\
f_2(x_1, \dots, x_n) &= 0, \\
&\vdots \\
f_n(x_1, \dots, x_n) &= 0,
\end{aligned} \tag{3.22}$$

где $f_i(x_1, \dots, x_n): R^n \rightarrow R$, $i = 1, \dots, n$ — нелинейные функции, определенные и непрерывные в некоторой области $G \subset R^n$, или в векторном виде

$$F(x) = 0, \tag{3.23}$$

где $x = (x_1, \dots, x_n)^T$, $F(x) = [f_1(x), \dots, f_n(x)]^T$.

Требуется найти такой вектор $x_* = (x_{*1}, \dots, x_{*n})^T$, который при подстановке в систему (3.22) превращает каждое уравнение в верное числовое равенство.

Замечания

1. Проблема решения системы (3.22) возникает при решении многих прикладных задач, например поиска безусловного экстремума функций многих переменных с помощью необходимых условий [31], при применении неявных методов интегрирования обыкновенных дифференциальных уравнений [33] и т. д.

2. Задача нахождения комплексных корней $f(z) = 0$ может быть сведена к проблеме решения двух уравнений с двумя неизвестными. Для этого следует положить $z = x + iy$ и выделить действительную и мнимую части функции [32]:

$$f(z) = u(x, y) + iv(x, y) = 0.$$

Отсюда получаем систему:

$$\begin{aligned} u(x, y) &= 0, \\ v(x, y) &= 0, \end{aligned}$$

которая может быть решена одним из рассматриваемых в данном разделе методов.

3. Для всех рассматриваемых далее методов требуется находить начальное приближение $x^{(0)}$. В случае $n = 2$ это можно сделать графически, определив координаты точки пересечения кривых, описываемых уравнениями $f_1(x_1, x_2) = 0$ и $f_2(x_1, x_2) = 0$.

4. Задача решения системы (3.22) может быть сведена к задаче поиска минимума функции $\Psi(x) = \sum_{i=1}^n f_i^2(x_1, \dots, x_n)$. Так как функция $\Psi(x)$ неотрицательная, ее минимальное значение, равное нулю, достигается в точке x_* , являющейся решением системы (3.22).

Для поиска минимума функции $\Psi(x)$ можно применить различные методы поиска безусловного экстремума функций многих переменных (первого, второго, нулевого порядков) [31].

Далее рассмотрим основные методы решения задачи (3.22).

3.2.2. МЕТОД ПРОСТЫХ ИТЕРАЦИЙ

Для применения метода требуется привести систему (3.22) к равносильному виду:

$$\begin{aligned} x_1 &= \varphi_1(x_1, \dots, x_n), \\ x_2 &= \varphi_2(x_1, \dots, x_n), \\ &\vdots \\ x_n &= \varphi_n(x_1, \dots, x_n), \end{aligned} \tag{3.24}$$

или в векторной форме

$$x = \Phi(x), \tag{3.25}$$

где $x = (x_1, \dots, x_n)^T$, $\Phi(x) = [\varphi_1(x), \dots, \varphi_n(x)]^T$, функции $\varphi_i(x)$ определены и непрерывны в окрестности изолированного решения x_* системы (3.24).

Методика решения задачи

1. Задать начальное приближение $x^{(0)} = (x_{10}, x_{20}, \dots, x_{n0})^T$ и малое положительное число ε (точность). Положить $k = 0$.

2. Вычислить $x^{(k+1)}$ по формуле

$$x^{(k+1)} = \Phi(x^{(k)}) \tag{3.26}$$

или

$$\begin{aligned} x_1^{(k+1)} &= \varphi_1(x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)}), \\ x_2^{(k+1)} &= \varphi_2(x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)}), \\ &\vdots \\ x_n^{(k+1)} &= \varphi_n(x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)}). \end{aligned} \quad (3.27)$$

3. Если $\Delta^{(k+1)} = \max_i |x_i^{(k+1)} - x_i^{(k)}| \leq \varepsilon$, процесс завершен и $x_* \cong x^{(k+1)}$. Если $\Delta^{(k+1)} > \varepsilon$, то положить $k = k + 1$ и перейти к п. 2.

Замечания

1. Итерационный процесс, реализуемый согласно (3.27), соответствует *параллельному итерированию*, так как для вычисления $(k+1)$ -го приближения всех неизвестных учитываются вычисленные ранее их k -е приближения.

2. Система (3.22) может быть преобразована к виду (3.24) различными способами, например с помощью выражения переменных x_i , $i = 1, \dots, n$, таким образом, чтобы выполнялось условие сходимости.

Другой способ заключается в замене системы $F(x) = 0$ системой

$$x = \underbrace{x + \Lambda F(x)}_{\Phi(x)},$$

где Λ — неособенная матрица. В качестве матрицы Λ можно использовать, например, $\Lambda = -W^{-1}(x^{(0)})$, если $\det W(x^{(0)}) \neq 0$, где

$$W(x) = \begin{pmatrix} \frac{\partial f_1(x)}{\partial x_1} & \dots & \frac{\partial f_1(x)}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_n(x)}{\partial x_1} & \dots & \frac{\partial f_n(x)}{\partial x_n} \end{pmatrix} \text{ — матрица Якоби.}$$

В случае, если $\det W(x^{(0)}) = 0$, следует выбрать другое начальное приближение. Заметим, что при таком выборе Λ метод простых итераций совпадает с упрощенным методом Ньютона (см. п. 3.2.5).

3. В качестве $\Delta^{(k+1)}$ можно использовать различные нормы векторов (см. п. 1.3.1).

Теорема 3.13 (о достаточном условии сходимости метода простых итераций).

Пусть функции $\varphi_i(x)$ и $\varphi'_i(x)$, $i = 1, \dots, n$ непрерывны в области G , причем выполнено неравенство

$$\max_{x \in G} \max_i \sum_{j=1}^n \left| \frac{\partial \varphi_i(x)}{\partial x_j} \right| \leq q < 1, \quad (3.28)$$

где q — некоторая постоянная.

Если последовательные приближения $x^{(k+1)} = \Phi(x^{(k)})$, $k = 0, 1, \dots$ не выходят из области G , то процесс последовательных приближений сходится:

$x_* = \lim_{k \rightarrow \infty} x^{(k)}$ и вектор x_* является в области G единственным решением системы (3.25).

Замечания

1. Вместо условия (3.28) можно также использовать

$$\max_{x \in G} \max_j \sum_{i=1}^n \left| \frac{\partial \varphi_i(x)}{\partial x_j} \right| \leq q < 1. \quad (3.29)$$

2. Условия (3.28), (3.29) выполняются, если для любого $x \in G$ справедливы неравенства $\left\| \frac{\partial \Phi}{\partial x} \right\|_1 < 1$, $\left\| \frac{\partial \Phi}{\partial x} \right\|_2 < 1$ соответственно, где

$$\frac{\partial \Phi}{\partial x} = \begin{pmatrix} \frac{\partial \varphi_1(x)}{\partial x_1} & \dots & \frac{\partial \varphi_1(x)}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial \varphi_n(x)}{\partial x_1} & \dots & \frac{\partial \varphi_n(x)}{\partial x_n} \end{pmatrix}.$$

Пример 3.17. Найти корни системы

$$\begin{aligned} 2x_1^2 - x_1 x_2 - 5x_1 + 1 &= 0, \\ x_1 + 3 \cdot \lg x_1 - x_2^2 &= 0, \end{aligned}$$

расположенные в первом квадранте, методом простых итераций с точностью $\varepsilon = 0,001$.

□ Преобразуем систему к виду (3.24) так, чтобы выполнялось условие сходимости:

$$\begin{aligned} x_1 &= \sqrt{\frac{x_1(x_2 + 5) - 1}{2}} = \varphi_1(x), \\ x_2 &= \sqrt{x_1 + 3 \cdot \lg x_1} = \varphi_2(x). \end{aligned}$$

Найдем частные производные:

$$\begin{aligned} \frac{\partial \varphi_1}{\partial x_1} &= \frac{x_2 + 5}{4\sqrt{\frac{x_1(x_2 + 5) - 1}{2}}}, \quad \frac{\partial \varphi_1}{\partial x_2} = \frac{x_1}{4\sqrt{\frac{x_1(x_2 + 5) - 1}{2}}}; \\ \frac{\partial \varphi_2}{\partial x_1} &= \frac{1 + \frac{3 \cdot 0,43429}{x_1}}{2\sqrt{x_1 + 3 \cdot \lg x_1}}, \quad \frac{\partial \varphi_2}{\partial x_2} = 0. \end{aligned}$$

Здесь принято $\lg e \approx 0,43429$. Далее воспользуемся методикой решения задачи.

1. Для выбора начального приближения найдем координаты точек пересечения кривых, соответствующих первому и второму уравнениям (рис. 3.17).

Находим приближенные значения координат решения (по условию задачи нас интересуют только корни с положительными координатами): $x^{(0)} = (3,5; 2,2)^T$.

Проверим выполнение условий сходимости. Будем рассматривать окрестность найденной точки $x^{(0)}$: $G = \{|x_1 - 3,5| \leq 0,1; |x_2 - 2,2| \leq 0,1\}$. Тогда

$$\left| \frac{\partial \varphi_1}{\partial x_1} \right| \leq \frac{2,3+5}{4\sqrt{\frac{3,4 \cdot (2,1+5)-1}{2}}} = 0,536 < 0,54; \quad \left| \frac{\partial \varphi_1}{\partial x_2} \right| \leq \frac{3,6}{4\sqrt{\frac{3,4 \cdot (2,1+5)-1}{2}}} = 0,265 < 0,27;$$

$$\left| \frac{\partial \varphi_2}{\partial x_1} \right| \leq \frac{1 + \frac{3 \cdot 0,43429}{3,4}}{2\sqrt{3,4 + 3 \cdot \lg 3,4}} = 0,309 < 0,31; \quad \left| \frac{\partial \varphi_2}{\partial x_2} \right| = 0.$$

Следовательно, можно получить оценки:

$$\begin{aligned} \left| \frac{\partial \varphi_1(x)}{\partial x_1} \right| + \left| \frac{\partial \varphi_2(x)}{\partial x_1} \right| &< 0,54 + 0,31 = 0,85 < 1, \\ \left| \frac{\partial \varphi_1(x)}{\partial x_2} \right| + \left| \frac{\partial \varphi_2(x)}{\partial x_2} \right| &< 0,27 + 0 = 0,27 < 1. \end{aligned}$$

Очевидно, условие (3.28) выполняется. Если последовательные приближения не будут выходить из области G , то согласно теореме 3.13 итерационный процесс будет сходящимся. В поставленной задаче $\varepsilon = 0,001$.

2, 3. Выполним расчеты по формулам (3.27):

$$\begin{aligned} x_1^{(k+1)} &= \sqrt{\frac{x_1^{(k)} \cdot [x_2^{(k)} + 5] - 1}{2}}, \quad k = 0, 1, \dots, \\ x_2^{(k+1)} &= \sqrt{x_1^{(k)} + 3 \cdot \lg x_1^{(k)}}, \end{aligned}$$

а результаты поместим в таблицу 3.17.

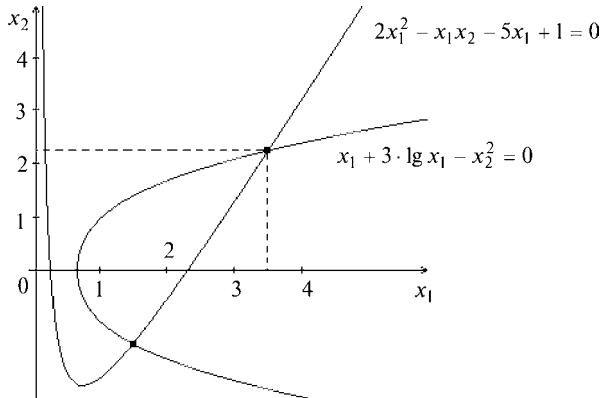


Рис. 3.17

Таблица 3.17

k	0	1	2	3	4
$x_1^{(k)}$	3,5000	3,4785	3,4837	3,4848	3,4957
$x_2^{(k)}$	2,2000	2,2654	2,2589	2,26049	2,26082
$\Delta^{(k+1)}$	—	0,0654	0,0065	0,00159	0,0009

Заметим, что величина $\Delta^{(k+1)}$ при увеличении номера итерации уменьшается, что характерно для любого сходящегося процесса. Найдено приближенное решение: $x_* \cong (3,4857; 2,2608)^T$. При этом $f_1(x_*) = -0,0083$, $f_2(x_*) = 0,00126$. ■

3.2.3. МЕТОД ЗЕЙДЕЛЯ

Метод Зейделя предназначен для решения систем, записанных в форме (3.24). Этот метод является модификацией метода простых итераций, где после задания начального приближения $x^{(0)}$ вместо параллельного итерирования производится *последовательное итерирование*, причем на каждой итерации в каждое последующее уравнение подставляются значения неизвестных, полученных из предыдущих уравнений.

Методика решения задачи

1. Задать начальное приближение $x^{(0)}$ и малое положительное число ε (точность). Положить $k = 0$.

2. Вычислить $x^{(k+1)}$ по формулам:

$$\begin{aligned} x_1^{(k+1)} &= \varphi_1\left(x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)}\right), \\ &\quad \downarrow \\ x_2^{(k+1)} &= \varphi_2\left(\boxed{x_1^{(k+1)}}, x_2^{(k)}, \dots, x_n^{(k)}\right), \\ &\quad \downarrow \\ \vdots & \\ x_n^{(k+1)} &= \varphi_n\left(\boxed{x_1^{(k+1)}}, \boxed{x_2^{(k+1)}}, \dots, \boxed{x_{n-1}^{(k+1)}}, x_n^{(k)}\right), \end{aligned} \quad (3.30)$$

где прямоугольниками отмечены значения, которые берутся из предшествующих уравнений на текущей итерации.

3. Если $\Delta^{(k+1)} = \max_i |x_i^{(k+1)} - x_i^{(k)}| \leq \varepsilon$, процесс завершить и положить $x_* \cong x^{(k+1)}$.

Если $\Delta^{(k+1)} > \varepsilon$, то положить $k = k + 1$ и перейти к п. 2.

Пример 3.18. Найти корни системы

$$\begin{aligned} f_1(x_1, x_2) &= 2x_1^2 - x_1x_2 - 5x_1 + 1 = 0, \\ f_2(x_1, x_2) &= x_1 + 3 \cdot \lg x_1 - x_2^2 = 0, \end{aligned}$$

расположенные в первом квадранте, методом Зейделя с точностью $\varepsilon = 0,001$.

□ Преобразование системы к виду (3.24) и поиск начального приближения описаны в примере 3.17.

1. Зададим начальное приближение $x^{(0)} = (3,5; 2,2)^T$. В поставленной задаче $\varepsilon = 0,001$.

2, 3. Выполним расчеты по формулам (3.30):

$$\begin{aligned} x_1^{(k+1)} &= \sqrt{\frac{x_1^{(k)} \cdot [x_2^{(k)} + 5] - 1}{2}}, \quad k = 0, 1, \dots, \\ x_2^{(k+1)} &= \sqrt{x_1^{(k+1)} + 3 \cdot \lg x_1^{(k+1)}}, \end{aligned}$$

а результаты поместим в таблицу 3.18.

k	0	1	2	3	4	5
$x_1^{(k)}$	3,5000	3,4785	3,4821	3,484250	3,485537	3,486305
$x_2^{(k)}$	2,2000	2,2588	2,2600	2,260658	2,261049	2,26128
Δ	—	0,0588	0,0036	0,00215	0,00128	0,0007

Найдено приближенное решение $x_* \cong (3,4863; 2,2613)^T$. При этом $f_1(x_*) = -0,006425$, $f_2(x_*) = 0,000007$.

3.2.4. МЕТОД НЬЮТОНА

Метод используется для решения систем вида (3.22) или (3.23). Расчетная формула для нахождения решения является естественным обобщением формулы (3.14):

$$x^{(k+1)} = x^{(k)} - W^{-1}(x^{(k)}) \cdot F(x^{(k)}), \quad k = 0, 1, \dots, \quad (3.31)$$

где

$$W(x) = \begin{pmatrix} \frac{\partial f_1(x)}{\partial x_1} & \dots & \frac{\partial f_1(x)}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n(x)}{\partial x_1} & \dots & \frac{\partial f_n(x)}{\partial x_n} \end{pmatrix}$$

является матрицей Якоби.

Так как процесс вычисления обратной матрицы является трудоемким, преобразуем (3.31) следующим образом:

$$\Delta x^{(k)} = -W^{-1}(x^{(k)}) \cdot F(x^{(k)}), \quad k = 0, 1, \dots,$$

где $\Delta x^{(k)} = x^{(k+1)} - x^{(k)}$ — поправка к текущему приближению $x^{(k)}$.

Умножим последнее выражение слева на матрицу Якоби $W(x^{(k)})$:

$$W(x^{(k)}) \Delta x^{(k)} = -W(x^{(k)}) W^{-1}(x^{(k)}) F(x^{(k)}) = -F(x^{(k)}), \quad k = 0, 1, \dots$$

В результате получена система линейных алгебраических уравнений относительно поправки $\Delta x^{(k)}$. После ее определения вычисляется следующее приближение $x^{(k+1)} = x^{(k)} + \Delta x^{(k)}$.

Методика решения задачи

1. Задать начальное приближение $x^{(0)}$ и малое положительное число ε (точность). Положить $k = 0$.

2. Решить систему линейных алгебраических уравнений относительно поправки $\Delta x^{(k)}$:

$$W(x^{(k)}) \cdot \Delta x^{(k)} = -F(x^{(k)}). \quad (3.32)$$

3. Вычислить следующее приближение:

$$x^{(k+1)} = x^{(k)} + \Delta x^{(k)}.$$

4. Если $\Delta^{(k+1)} = \max_i |x_i^{(k+1)} - x_i^{(k)}| \leq \varepsilon$, процесс закончить и положить $x_* \cong x^{(k+1)}$.

Если $\Delta^{(k+1)} > \varepsilon$, то положить $k = k + 1$ и перейти к п. 2.

Теорема 3.14 (о достаточных условиях сходимости метода Ньютона).

Пусть функция $F(x)$ непрерывно дифференцируема в открытом выпуклом множестве $G \subset R^n$. Предположим, что существуют $x_* \in R^n$ и $r, \beta > 0$, такие, что $N(x_*, r) \subset G$, $F(x_*) = 0$, и существует $W^{-1}(x_*)$, причем $\|W^{-1}(x_*)\| \leq \beta$ и $W(x) \in Lip_\gamma(N(x_*, r))$. Тогда существует $\varepsilon > 0$ такое, что для всех $x^{(0)} \in N(x_*, \varepsilon)$ последовательность $x^{(1)}, x^{(2)}, \dots$, порожденная соотношением (3.31), сходится к x_* и удовлетворяет неравенству

$$\|x^{(k+1)} - x_*\| \leq \beta \gamma \|x^{(k)} - x_*\|^2, \quad k = 0, 1, \dots$$

Здесь использованы следующие обозначения: $N(x, r)$ — открытая окрестность радиуса r с центром в точке x : $N(x, r) = \{\bar{x} \in R^n : |\bar{x} - x| < r\}$; запись $W(x) \in Lip_\gamma(N(x_*, r))$ означает, что $W(x)$ непрерывна по Липшицу, где γ — константа Липшица, т. е. $\|W(y) - W(x)\| \leq \gamma \|y - x\|$, $x, y \in N(x_*, r)$.

Замечания

1. Теорема 3.14 свидетельствует о локальной квадратичной сходимости метода Ньютона.

2. К недостаткам метода Ньютона следует отнести:

- необходимость задавать достаточно хорошее начальное приближение;
- отсутствие глобальной сходимости для многих задач;
- необходимость вычисления матрицы Якоби на каждой итерации;
- необходимость решения на каждой итерации системы линейных уравнений, которая может быть плохо обусловленной.

Достоинством метода является квадратичная сходимость из хорошего начального приближения при условии невырожденности матрицы Якоби.

Пример 3.19. Решить систему:

$$\begin{aligned} x_1 + x_2 - 3 &= 0, \\ x_1^2 + x_2^2 - 9 &= 0 \end{aligned}$$

методом Ньютона с точностью $\varepsilon = 0,001$.

□ Очевидно, корнями системы являются $x_{*1} = (3; 0)^T$, $x_{*2} = (0; 3)^T$. Найдем приближенно второй корень x_{*2} .

1. Зададим начальное приближение $x^{(0)} = (1; 5)^T$. В поставленной задаче $\varepsilon = 0,001$. Положим $k = 0$.

2⁰. Составим систему (3.32). Так как матрица

$$W(x) = \begin{pmatrix} 1 & 1 \\ 2x_1 & 2x_2 \end{pmatrix},$$

то система $W(x^{(0)}) \cdot \Delta x^{(0)} = -F(x^{(0)})$ имеет вид

$$\begin{pmatrix} 1 & 1 \\ 2 & 10 \end{pmatrix} \cdot \begin{pmatrix} \Delta x_1^{(0)} \\ \Delta x_2^{(0)} \end{pmatrix} = -\begin{pmatrix} 3 \\ 17 \end{pmatrix}.$$

Отсюда

$$\Delta x^{(0)} = \begin{pmatrix} \Delta x_1^{(0)} \\ \Delta x_2^{(0)} \end{pmatrix} = \begin{pmatrix} -\frac{13}{8} \\ -\frac{11}{8} \end{pmatrix}.$$

Для вычисления $\Delta x^{(k)}$, $k = 0, 1, \dots$, здесь и далее используется метод Гаусса единственного деления (см. п. 1.2.1).

3⁰. Вычислим

$$x^{(1)} = x^{(0)} + \Delta x^{(0)} = \begin{pmatrix} 1 \\ 5 \end{pmatrix} + \begin{pmatrix} -\frac{13}{8} \\ -\frac{11}{8} \end{pmatrix} = \begin{pmatrix} -\frac{5}{8} \\ \frac{29}{8} \end{pmatrix} = \begin{pmatrix} -0,625 \\ 3,625 \end{pmatrix}.$$

4⁰. Так как $\Delta^{(1)} = \max \left\{ \left| -\frac{13}{8} \right|, \left| -\frac{11}{8} \right| \right\} = \frac{13}{8} > \varepsilon$, то положим $k = 1$ и перейдем к п. 2.

2¹. Составим систему $W(x^{(1)}) \cdot \Delta x^{(1)} = -F(x^{(1)})$:

$$\begin{pmatrix} 1 & 1 \\ -\frac{5}{4} & \frac{29}{4} \end{pmatrix} \cdot \Delta x^{(1)} = -\begin{pmatrix} 0 \\ \frac{145}{32} \end{pmatrix}.$$

Отсюда

$$\Delta x^{(1)} = \begin{pmatrix} \frac{145}{272} \\ -\frac{145}{272} \end{pmatrix} = \begin{pmatrix} 0,5333 \\ -0,5333 \end{pmatrix}.$$

3¹. Вычислим

$$x^{(2)} = x^{(1)} + \Delta x^{(1)} = \begin{pmatrix} -0,625 \\ 3,625 \end{pmatrix} + \begin{pmatrix} 0,5333 \\ -0,5333 \end{pmatrix} = \begin{pmatrix} -0,0919 \\ 3,0919 \end{pmatrix}.$$

4¹. Так как $\Delta^{(2)} = 0,5333 > \varepsilon$, то положим $k = 2$ и перейдем к п. 2. Результаты дальнейших вычислений приведены в таблице 3.19.

Таблица 3.19

k	0	1	2	3	4	5
$x_1^{(k)}$	1	-0,625	-0,091911	-0,002653	-0,0000023	0,0000
$x_2^{(k)}$	5	3,625	3,091911	3,002653	3,0000023	3,0000
$\Delta^{(k+1)}$	—	1,625	0,5333	0,089258	0,0026507	0,0000023

Найденное приближенное решение $x_* \approx (0; 3,0000)^T$. Из анализа решения, приведенного в таблице 3.19, следует, что количество верных знаков на каждой итерации удваивается, что соответствует квадратичной сходимости. ■

Пример 3.20. Найти решение системы

$$\begin{aligned} f_1(x) &= x_1 + 3 \cdot \lg x_1 - x_2^2 = 0, \\ f_2(x) &= 2x_1^2 - x_1 x_2 - 5x_1 + 1 = 0, \end{aligned}$$

расположенное в первом квадранте, с точностью $\varepsilon = 10^{-5}$.

□1. Выберем начальное приближение $x^{(0)} = (3,5; 2,2)^T$. В поставленной задаче $\varepsilon = 10^{-5}$. Положим $k = 0$.

2⁰. Составим систему (3.32).

Так как

$$W(x) = \begin{pmatrix} \frac{\partial f_1(x)}{\partial x_1} & \frac{\partial f_1(x)}{\partial x_2} \\ \frac{\partial f_2(x)}{\partial x_1} & \frac{\partial f_2(x)}{\partial x_2} \end{pmatrix} = \begin{pmatrix} 1 + \frac{3 \cdot 0,43429}{x_1} & -2x_2 \\ 4x_1 - x_2 - 5 & -x_1 \end{pmatrix},$$

то

$$\begin{aligned} W(x^{(0)}) &= \begin{pmatrix} 1 + \frac{3 \cdot 0,43429}{3,5} & -2 \cdot 2,2 \\ 4 \cdot 3,5 - 2,2 - 5 & -3,5 \end{pmatrix} = \begin{pmatrix} 1,372248571 & -4,4 \\ 6,8 & -3,5 \end{pmatrix}, \\ &\quad \begin{pmatrix} 1,372248571 & -4,4 \\ 6,8 & -3,5 \end{pmatrix} \Delta x^{(0)} = -F(x^{(0)}) = \\ &= -\begin{pmatrix} 3,5 + 3 \cdot \lg 3,5 - 2,2^2 \\ 2 \cdot 3,5^2 - 3,5 \cdot 2,2 - 5 \cdot 3,5 + 1 \end{pmatrix} = \begin{pmatrix} -0,29220413 \\ -0,3 \end{pmatrix}. \end{aligned}$$

Отсюда с помощью метода Гаусса единственного деления находим

$$\Delta x^{(0)} = \begin{pmatrix} -0,011835967 \\ 0,062718692 \end{pmatrix}.$$

3⁰. Вычислим

$$x^{(1)} = x^{(0)} + \Delta x^{(0)} = \begin{pmatrix} 3,5 \\ 2,2 \end{pmatrix} + \begin{pmatrix} -0,011835967 \\ 0,062718692 \end{pmatrix} = \begin{pmatrix} 3,488164032 \\ 2,262718691 \end{pmatrix}.$$

Этот же результат может быть непосредственно получен по формуле (3.31):

$$x^{(1)} = \begin{pmatrix} 3,5 \\ 2,2 \end{pmatrix} - \frac{1}{25,11713} \begin{pmatrix} -3,5 & 4,4 \\ -6,8 & 1,372248571 \end{pmatrix} \begin{pmatrix} -0,29220413 \\ -0,3 \end{pmatrix} = \begin{pmatrix} 3,488164032 \\ 2,262718691 \end{pmatrix}.$$

4⁰. Так как $\Delta^{(1)} = \max\{0,011835967; 0,062718692\} = 0,062718692 > \varepsilon$, то положим $k = 1$ и перейдем к п. 2.

2¹. Составим систему (3.32):

$$\begin{pmatrix} 1,373511678 & -4,525437382 \\ 6,689937437 & -3,488164032 \end{pmatrix} \Delta x^{(1)} = \begin{pmatrix} 0,00394114 \\ -0,00102252 \end{pmatrix}.$$

Применяя метод Гаусса единственного деления, получаем

$$\Delta x^{(1)} = \begin{pmatrix} -7,21032383 \cdot 10^{-4} \\ -0,001089727 \end{pmatrix}.$$

3¹. Вычислим

$$x^{(2)} = x^{(1)} + \Delta x^{(1)} = \begin{pmatrix} 3,487443 \\ 2,261628964 \end{pmatrix}.$$

k	0	1	2	3
$x_1^{(k)}$	3,500000	3,488164032	3,487443000	3,487442788
$x_2^{(k)}$	2,200000	2,262718691	2,261628964	2,261628631
Δ	—	0,062718692	0,001089727	$3,334461599 \cdot 10^{-7}$

4¹. Так как $\Delta^{(2)} = \max\{|-7,21032383 \cdot 10^{-4}|; |-0,001089727|\} = 0,001089727 > \varepsilon$, то положим $k = 2$ и продолжим решение по алгоритму. Результаты дальнейших расчетов приведены в таблице 3.20.

Из сопоставления полученных результатов с таблицей 3.17 следует, что по методу простых итераций точность $\varepsilon = 0,001$ достигается за четыре итерации, а методом Ньютона точность $\varepsilon = 10^{-5}$ достигнута всего за три приближения. ■

Пример 3.21. Найти решение системы:

$$\begin{aligned}x_1^2 + x_2^2 + x_3^2 &= 1, \\2x_1^2 + x_2^2 - 4x_3 &= 0, \\3x_1^2 - 4x_2 + x_3^2 &= 0\end{aligned}$$

методом Ньютона с точностью $\varepsilon = 0,005$.

□1. Выберем в качестве начального приближения $x^{(0)} = (0,5; 0,5; 0,5)^T$. В поставленной задаче $\varepsilon = 0,005$. Положим $k = 0$.

2⁰, 3⁰. Воспользуемся формулой (3.31). Матрица Якоби имеет вид

$$W(x) = \begin{pmatrix} 2x_1 & 2x_2 & 2x_3 \\ 4x_1 & 4x_2 & -4 \\ 6x_1 & -4 & 2x_3 \end{pmatrix}.$$

В точке $x^{(0)}$ справедливо

$$\begin{aligned}W(x^{(0)}) &= \begin{pmatrix} 1 & 1 & 1 \\ 2 & 1 & -4 \\ 3 & -4 & 1 \end{pmatrix}, \quad F(x^{(0)}) = \begin{pmatrix} 0,25 + 0,25 + 0,25 - 1 \\ 0,5 + 0,25 - 2 \\ 0,75 - 2 + 0,25 \end{pmatrix} = \begin{pmatrix} -0,25 \\ -1,25 \\ -1 \end{pmatrix}, \\ \det W(x^{(0)}) &= -40, \quad W^{-1}(x^{(0)}) = -\frac{1}{40} \begin{pmatrix} -15 & -5 & -5 \\ -14 & -2 & 6 \\ -11 & 7 & -1 \end{pmatrix} = \begin{pmatrix} \frac{3}{8} & \frac{1}{8} & \frac{1}{8} \\ \frac{7}{20} & \frac{1}{20} & -\frac{3}{20} \\ \frac{11}{40} & -\frac{7}{40} & \frac{1}{40} \end{pmatrix}.\end{aligned}$$

В результате

$$\begin{aligned}x^{(1)} &= x^{(0)} - W^{-1}(x^{(0)}) \cdot F(x^{(0)}) = \\&= \begin{pmatrix} 0,5 \\ 0,5 \\ 0,5 \end{pmatrix} - \begin{pmatrix} \frac{3}{8} & \frac{1}{8} & \frac{1}{8} \\ \frac{7}{20} & \frac{1}{20} & -\frac{3}{20} \\ \frac{11}{40} & -\frac{7}{40} & \frac{1}{40} \end{pmatrix} \begin{pmatrix} -0,25 \\ -1,25 \\ -1 \end{pmatrix} = \begin{pmatrix} 0,5 \\ 0,5 \\ 0,5 \end{pmatrix} + \begin{pmatrix} 0,375 \\ 0 \\ -0,125 \end{pmatrix} = \begin{pmatrix} 0,875 \\ 0,5 \\ 0,375 \end{pmatrix}.\end{aligned}$$

4⁰. Так как $\Delta^{(1)} = \max\{0,375; 0; |-0,125|\} = 0,375 > \varepsilon$, то положим $k = 1$ и перейдем к п. 2.

2¹. В точке $x^{(1)}$ справедливо

$$W(x^{(1)}) = \begin{pmatrix} 2 \cdot 0,875 & 2 \cdot 0,5 & 2 \cdot 0,375 \\ 4 \cdot 0,875 & 2 \cdot 0,5 & -4 \\ 6 \cdot 0,875 & -4 & 2 \cdot 0,375 \end{pmatrix} = \begin{pmatrix} 1,75 & 1 & 0,75 \\ 3,5 & 1 & -4 \\ 5,25 & -4 & 0,75 \end{pmatrix},$$

$$F(x^{(1)}) = \begin{pmatrix} 0,875^2 + 0,5^2 + 0,375^2 - 1 \\ 2 \cdot 0,875^2 + 0,5^2 - 4 \cdot 0,375 \\ 3 \cdot 0,875^2 - 4 \cdot 0,5 + 0,375^2 \end{pmatrix} = \begin{pmatrix} 0,15625 \\ 0,28125 \\ 0,4375 \end{pmatrix}.$$

Составим и решим систему

$$\begin{pmatrix} 1,75 & 1 & 0,75 \\ 3,5 & 1 & -4 \\ 5,25 & -4 & 0,75 \end{pmatrix} \cdot \Delta x^{(1)} = -\begin{pmatrix} 0,15625 \\ 0,28125 \\ 0,4375 \end{pmatrix}.$$

С помощью метода Гаусса единственного деления получаем

$$\Delta x^{(1)} = \begin{pmatrix} -0,084995 \\ -0,0032467 \\ -0,0056818 \end{pmatrix}.$$

3¹. Вычислим

$$x^{(2)} = x^{(1)} + \Delta x^{(1)} = \begin{pmatrix} 0,790005 \\ 0,496753 \\ 0,369318 \end{pmatrix}.$$

4¹. Так как $\Delta^{(2)} = \max_i |\Delta x_i^{(1)}| = 0,084995 > \varepsilon = 0,005$, то положим $k = 2$ и перейдем к п. 2.

2². Составим и решим систему

$$\begin{pmatrix} 1,58001 & 0,993506 & 0,738636 \\ 3,16002 & 1,987012 & -4 \\ 4,74003 & -4 & 0,738636 \end{pmatrix} \Delta x^{(2)} = -\begin{pmatrix} 0,007267 \\ 0,017707 \\ 0,021707 \end{pmatrix}.$$

Применяя метод Гаусса, находим

$$\Delta x^{(2)} = \begin{pmatrix} -0,004785 \\ -0,000136 \\ 0,000579 \end{pmatrix}.$$

3². Вычислим

$$x^{(3)} = x^{(2)} + \Delta x^{(2)} = \begin{pmatrix} 0,785220 \\ 0,496617 \\ 0,369897 \end{pmatrix}.$$

4². Так как $\Delta^{(3)} = \max_i |\Delta x_i^{(2)}| = 0,004785 < 0,005$, то процесс завершен и $x_* \cong x^{(3)}$. ■

3.2.5. МОДИФИКАЦИИ МЕТОДА НЬЮТОНА

А. Упрощенный метод Ньютона. В этом методе в отличие от метода Ньютона (3.31) обратная матрица ищется только один раз в начальной точке $x^{(0)}$:

$$x^{(k+1)} = x^{(k)} - W^{-1}(x^{(0)}) \cdot F(x^{(k)}), \quad k = 0, 1, \dots \quad (3.33)$$

Заметим, что при решении одного уравнения $f(x) = 0$ упрощенным методом Ньютона производная функции вычисляется также один раз в начальной точке (см. п. 3.1.7).

Методика решения задачи аналогична изложенной в п. 3.2.4, где вместо (3.32) используется система

$$W(x^{(0)}) \cdot \Delta x^{(k)} = -F(x^{(k)}), \quad k = 0, 1, \dots,$$

матрица которой $W(x^{(0)})$ не изменяется от итерации к итерации.

Очевидно, сходимость упрощенного метода Ньютона в общем случае хуже по сравнению со сходимостью процесса (3.31).

Пример 3.22. Найти положительное решение системы

$$f_1(x) = x_1^2 + x_2^2 - 1 = 0,$$

$$f_2(x) = x_1^3 - x_2 = 0$$

упрощенным методом Ньютона с точностью $\varepsilon = 10^{-4}$.

□1. Выберем начальное приближение. Из рисунка 3.18 следует, что для нахождения положительного решения системы можно принять $x^{(0)} = (0,9; 0,5)^T$.

В поставленной задаче $\varepsilon = 10^{-4}$. Положим $k = 0$.

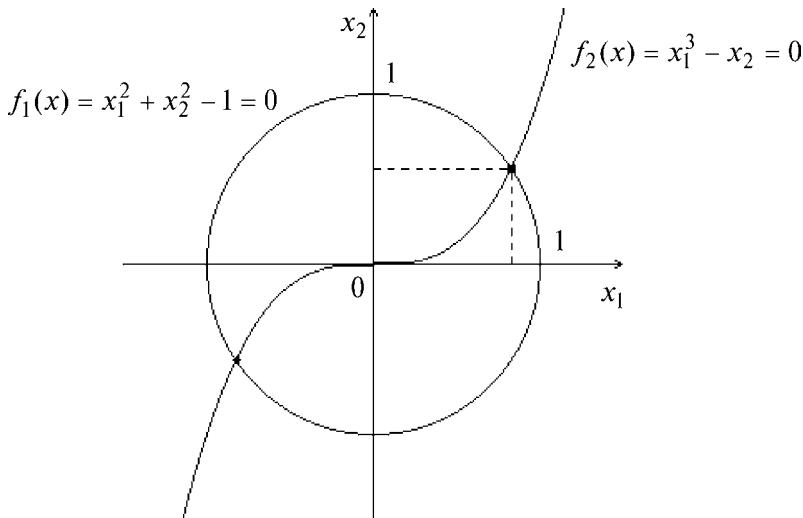


Рис. 3.18

$2^0, 3^0$. Так как

$$W(x) = \begin{pmatrix} 2x_1 & 2x_2 \\ 3x_1^2 & -1 \end{pmatrix}, \quad F(x) = \begin{pmatrix} x_1^2 + x_2^2 - 1 \\ x_1^3 - x_2 \end{pmatrix},$$

то для выполнения вычислений по формуле (3.33) найдем обратную матрицу $W^{-1}(x^{(0)})$:

$$\begin{aligned} W(x^{(0)}) &= \begin{pmatrix} 1,8 & 1 \\ 2,43 & -1 \end{pmatrix}, \quad \det W(x^{(0)}) = -1,8 - 2,43 = -4,23; \\ W^{-1}(x^{(0)}) &= -\frac{1}{-4,23} \begin{pmatrix} -1 & -1 \\ -2,43 & 1,8 \end{pmatrix} = \begin{pmatrix} 0,2364 & 0,2364 \\ 0,5745 & -0,4255 \end{pmatrix}. \end{aligned}$$

Поэтому

$$\begin{aligned} x^{(1)} &= x^{(0)} - W^{-1}(x^{(0)}) \cdot F(x^{(0)}) = \\ &= \begin{pmatrix} 0,9 \\ 0,5 \end{pmatrix} - \begin{pmatrix} 0,2364 & 0,2364 \\ 0,5745 & -0,4255 \end{pmatrix} \begin{pmatrix} 0,060 \\ 0,229 \end{pmatrix} = \begin{pmatrix} 0,9 \\ 0,5 \end{pmatrix} - \begin{pmatrix} 0,06832 \\ -0,06297 \end{pmatrix} = \begin{pmatrix} 0,83167 \\ 0,56298 \end{pmatrix}. \end{aligned}$$

4^0 . Так как $\Delta^{(1)} = \max\{|-0,06832|, |0,06297|\} = 0,06832 > \varepsilon$, то положим $k = 1$ и перейдем к п. 2.

$2^1, 3^1$. По формуле (3.33) получаем

$$\begin{aligned} x^{(2)} &= x^{(1)} - W^{-1}(x^{(0)}) \cdot F(x^{(1)}) = \\ &= \begin{pmatrix} 0,83167 \\ 0,56298 \end{pmatrix} - \begin{pmatrix} 0,2364 & 0,2364 \\ 0,5745 & -0,4255 \end{pmatrix} \begin{pmatrix} 0,008619 \\ 0,012267 \end{pmatrix} = \\ &= \begin{pmatrix} 0,83167 \\ 0,56298 \end{pmatrix} - \begin{pmatrix} 0,004937 \\ -0,000268 \end{pmatrix} = \begin{pmatrix} 0,826732 \\ 0,563246 \end{pmatrix}. \end{aligned}$$

4^1 . Так как $\Delta^{(2)} = 0,004937 > \varepsilon$, то положим $k = 2$ и перейдем к п. 2.

$2^2, 3^2$. По формуле (3.33) получаем

$$\begin{aligned} x^{(3)} &= x^{(2)} - W^{-1}(x^{(0)}) \cdot F(x^{(2)}) = \\ &= \begin{pmatrix} 0,826732 \\ 0,563246 \end{pmatrix} - \begin{pmatrix} 0,2364 & 0,2364 \\ 0,5745 & -0,4255 \end{pmatrix} \begin{pmatrix} 0,000732 \\ 0,001814 \end{pmatrix} = \\ &= \begin{pmatrix} 0,826732 \\ 0,563246 \end{pmatrix} - \begin{pmatrix} 0,000602 \\ -0,000351 \end{pmatrix} = \begin{pmatrix} 0,82613 \\ 0,56359 \end{pmatrix}. \end{aligned}$$

4^2 . Поскольку $\Delta^{(3)} = 0,000602 > \varepsilon$, то положим $k = 3$ и перейдем к п. 2.

$2^3, 3^3$. По формуле (3.33) имеем

$$\begin{aligned} x^{(4)} &= x^{(3)} - W^{-1}(x^{(0)}) \cdot F(x^{(3)}) = \\ &= \begin{pmatrix} 0,82613 \\ 0,56359 \end{pmatrix} - \begin{pmatrix} 0,2364 & 0,2364 \\ 0,5745 & -0,4255 \end{pmatrix} \begin{pmatrix} 0,000124 \\ 0,000236 \end{pmatrix} = \\ &= \begin{pmatrix} 0,82613 \\ 0,56359 \end{pmatrix} - \begin{pmatrix} 8,524 \cdot 10^{-5} \\ -2,89692 \cdot 10^{-5} \end{pmatrix} = \begin{pmatrix} 0,8260447 \\ 0,5636189 \end{pmatrix}. \end{aligned}$$

4^3 . Поскольку $\Delta^{(4)} = 8,524 \cdot 10^{-5} < \varepsilon = 0,0001$, то процесс завершен и $x_* \cong x^{(4)}$. ■

Б. Метод секущих. Идея метода секущих (*метода Бройдена*) заключается в аппроксимации матрицы Якоби с использованием уже вычисленных значений функций, образующих систему (сравните с методом секущих для уравнения $f(x) = 0$, изложенным в п. 3.1.7).

Методика решения задачи

1. Задать начальное приближение $x^{(0)}$ и малое положительное число ε .
2. Положить $k = 0$ и $A_0 = W(x^{(0)})$, где $W(x)$ — матрица Якоби.
3. Решить систему линейных алгебраических уравнений

$$A_k s_k = -F(x^{(k)})$$

относительно s_k — поправки к текущему приближению.

4. Вычислить $x^{(k+1)} = x^{(k)} + s_k$.

5. Если $\|s_k\| \leq \varepsilon$, процесс завершить и положить $x_* = x^{(k+1)}$. Если $\|s_k\| > \varepsilon$, вычислить

$$y_k = F(x^{(k+1)}) - F(x^{(k)}), \quad A_{k+1} = A_k + \frac{(y_k - A_k s_k) \cdot s_k^T}{s_k^T s_k},$$

положить $k = k + 1$ и перейти к п. 3.

Замечания

1. Можно доказать [11], что если $x^{(0)}$ достаточно близко к корню x_* , где матрица Якоби $W(x_*)$ не вырождена, и если A_0 достаточно близка к $W(x^{(0)})$, то последовательность итераций $\{x^{(k)}\}$ сходится к x_* сверхлинейно.

2. Когда метод секущих сходится к x_* сверхлинейно, нельзя предполагать, что последняя из аппроксимаций якобиана A_k будет приблизительно равняться $W(x_*)$, хотя часто это именно так.

Пример 3.23. Решить систему

$$\begin{aligned} f_1(x) &= x_1 + x_2 - 3 = 0, \\ f_2(x) &= x_1^2 + x_2^2 - 9 = 0 \end{aligned}$$

методом секущих.

□1. Как следует из примера 3.18, для нахождения корня $x_* = (0; 3)^T$ можно выбрать начальное приближение $x^{(0)} = (1; 5)^T$ и $\varepsilon = 0,001$.

2. Положим $k = 0$ и

$$A_0 = W(x^{(0)}) = \begin{pmatrix} 1 & 1 \\ 2 & 10 \end{pmatrix}, \text{ так как } W(x) = \begin{pmatrix} 1 & 1 \\ 2x_2 & 2x_2 \end{pmatrix}.$$

3⁰. Решим систему $A_0 \cdot s_0 = -F(x^{(0)})$. Так как $F(x^{(0)}) = \begin{pmatrix} 3 \\ 17 \end{pmatrix}$, то

$$s_0 = -A_0^{-1} F(x^{(0)}) = \begin{pmatrix} -1,625 \\ -1,375 \end{pmatrix}.$$

4⁰. Вычислим

$$x^{(1)} = x^{(0)} + s_0 = \begin{pmatrix} 1 \\ 5 \end{pmatrix} + \begin{pmatrix} -1,625 \\ -1,375 \end{pmatrix} = \begin{pmatrix} -0,625 \\ 3,625 \end{pmatrix}.$$

5⁰. Поскольку $\|s_0\|_1 = 1,625 > \varepsilon$, то вычислим

$$y_0 = F(x^{(1)}) - F(x^{(0)}) = \begin{pmatrix} 0 \\ 4,53125 \end{pmatrix} - \begin{pmatrix} 3 \\ 17 \end{pmatrix} = \begin{pmatrix} -3 \\ -12,46875 \end{pmatrix},$$

$$s_0^T s_0 = (-1,625 \quad -1,375) \begin{pmatrix} -1,625 \\ -1,375 \end{pmatrix} = 4,53125,$$

$$y_0 - A_0 s_0 = \begin{pmatrix} -3 \\ -12,46875 \end{pmatrix} - \begin{pmatrix} 1 & 1 \\ 2 & 10 \end{pmatrix} \begin{pmatrix} -1,625 \\ -1,375 \end{pmatrix} = \begin{pmatrix} -3 \\ -12,46875 \end{pmatrix} - \begin{pmatrix} -3 \\ -17 \end{pmatrix} = \begin{pmatrix} 0 \\ 4,53125 \end{pmatrix},$$

$$(y_0 - A_0 s_0) \cdot s_0^T = \begin{pmatrix} 0 \\ 4,53125 \end{pmatrix} (-1,625 \quad -1,375);$$

$$A_1 = A_0 + \frac{(y_0 - A_0 s_0) \cdot s_0^T}{s_0^T s_0} = \begin{pmatrix} 1 & 1 \\ 2 & 10 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ -1,625 & -1,375 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0,375 & 8,625 \end{pmatrix}.$$

Положим $k = 1$ и перейдем к п. 3.

3¹. Решим систему $A_1 \cdot s_1 = -F(x^{(1)})$, т. е.

$$\begin{pmatrix} 1 & 1 \\ 0,375 & 8,625 \end{pmatrix} s_1 = -\begin{pmatrix} 0 \\ 4,53125 \end{pmatrix}.$$

Применяя метод Гаусса, получаем

$$s_1 = -A_1^{-1} \cdot F(x^{(1)}) = \begin{pmatrix} 0,549 \\ -0,549 \end{pmatrix}.$$

4¹. Вычислим

$$x^{(2)} = x^{(1)} + s_1 = \begin{pmatrix} -0,625 \\ 3,625 \end{pmatrix} + \begin{pmatrix} 0,549 \\ -0,549 \end{pmatrix} = \begin{pmatrix} -0,076 \\ 3,076 \end{pmatrix}.$$

5¹. Поскольку $\|s_1\|_1 = 0,549 > \varepsilon$, то вычислим

$$A_2 = \begin{pmatrix} 1 & 1 \\ -0,799 & 8,201 \end{pmatrix},$$

положим $k = 2$ и перейдем к п. 3. Результаты вычислений приведены в таблице 3.21.

Полученное приближенное решение $x_* \approx (0,0000013; 3,0000013)^T$. ■

Таблица 3.21

k	0	1	2	3	4	5
$x_1^{(k)}$	1	-0,625	-0,0757575	-0,0127942	-0,0003138	0,0000013
$x_2^{(k)}$	5	3,625	3,0757575	3,0127942	3,0003138	3,0000013
Δ	—	1,625	0,549	0,0629	0,01248	0,00031

Пример 3.24. Решить систему

$$x_1^2 + x_2^2 - 2 = 0,$$

$$e^{x_1-1} + x_2^3 - 2 = 0$$

методом секущих с точностью $\varepsilon = 0,01$.

k	0	1	2	3	4	5	6
$x_1^{(k)}$	1,5	0,8060692	0,7410741	0,8022786	0,9294701	1,004003	1,003084
$x_2^{(k)}$	2	1,457948	1,277067	1,159900	1,070406	1,003084	0,9992213
Δ	—	0,69393	0,18088	0,11717	0,12719	0,07453	0,0038

□ Очевидно, точное решение системы $x_* = (1; 1)^T$. Выберем начальное приближение $x^{(0)} = (1,5; 2)^T$. В поставленной задаче $\varepsilon = 0,01$.

Так как $W(x) = \begin{pmatrix} 2x_1 & 2x_2 \\ e^{x_1-1} & 3x_2^2 \end{pmatrix}$, то положим

$$A_0 = W(x^{(0)}) = \begin{pmatrix} 3 & 4 \\ 1,6487 & 12 \end{pmatrix}.$$

Результаты расчетов приведены в таблице 3.22.

Найдено приближенное решение: $x_* \approx (1,003084; 0,9992213)^T$. ■

Задачи для самостоятельного решения

1. Методами половинного деления, хорд, простых итераций, Ньютона, упрощенным методом Ньютона найти корни уравнений ($\varepsilon = 10^{-3}$):

- 1) $x^4 - 2x - 4 = 0$ (положительный корень);
- 2) $2 - \lg x - x = 0$;
- 3) $\operatorname{tg}(1,9x) - 2,8x = 0$;
- 4) $\sin(2,2x) - x = 0$;
- 5) $\ln(8x) = 9x - 3$;
- 6) $0,7e^{-0,59x} - x = 0$;
- 7) $5,6\sin(4,8x) - 4,5x = 0$;
- 8) $x^4 - 4x - 1 = 0$;
- 9) $\operatorname{tg}x = x$ (наименьший положительный корень);
- 10) $x^4 - 3x^2 + 75x - 10\,000 = 0$ (отрицательный корень);
- 11) $x^4 + 2x^3 - x - 1 = 0$;
- 12) $x^3 - 0,2x^2 - 0,2x - 1,2 = 0$ (положительный корень);
- 13) $x^5 - x - 0,2 = 0$ (положительный корень);
- 14) $x^3 - 2x - 5 = 0$;
- 15) $x^3 - 2x^2 - 4x - 7 = 0$;
- 16) $x^3 + 2x - 7 = 0$;
- 17) $x^4 + 3x - 20 = 0$.

2. Методами простых итераций, Зейделя, Ньютона, секущих, упрощенным методом Ньютона решить системы ($\varepsilon = 10^{-3}$):

$$\begin{array}{lll} 1) \begin{aligned} 2x_1^3 - x_2^2 - 1 &= 0, \\ x_1x_2^3 - x_2 - 4 &= 0; \end{aligned} & 2) \begin{aligned} \sin(x_1 + 1) - x_2 &= 1, \\ 2x_1 + \cos x_2 &= 2; \end{aligned} & 3) \begin{aligned} 2x_1 + \operatorname{tg}(x_1 x_2) &= 0, \\ (x_2^2 - 6)^2 + \ln x_1 &= 0; \end{aligned} \\ 4) \begin{aligned} \operatorname{tg}(x_1 x_2 + 0,2) &= x_1^2, \\ x_1^2 + 2x_2^2 &= 1; \end{aligned} & 5) \begin{aligned} x_1 - 2x_2^2 + 1 &= 0, \\ -x_1^2 + 2x_2 - 1 &= 0; \end{aligned} & 6) \begin{aligned} \cos(x_1 + 0,5) - x_2 &= 2, \\ \sin x_2 - 2x_1 &= 1. \end{aligned} \end{array}$$

ЧАСТЬ ВТОРАЯ

ЧИСЛЕННЫЕ МЕТОДЫ
ТЕОРИИ ПРИБЛИЖЕНИЙ

ГЛАВА
4

МЕТОДЫ ПРИБЛИЖЕНИЯ СЕТОЧНЫХ ФУНКЦИЙ

4.1. ОБЩАЯ ПОСТАНОВКА ЗАДАЧИ И КЛАССИФИКАЦИЯ МЕТОДОВ

Получаемые при компьютерных вычислениях, в экспериментальных исследованиях или задаваемые при проектировании элементов конструкций летательных аппаратов сеточные (табличные) функции

$$y_i = f(x_i), \quad x_i \in [a, b], \quad i = \overline{0, n},$$

малоинформативны. Они определены только в узлах x_i ($i = \overline{0, n}$) сетки Ω_n , а их значения в промежуточных точках, а также значения производных в узлах не известны, интегралы от них нельзя вычислить классическими методами. Каждая сетка характеризуется шагами $h_{i+1} = x_{i+1} - x_i$ неравномерного или ($h_{i+1} = \text{const}$) равномерного разбиения.

Однако значения функции должны быть известны при любом значении аргумента $x \neq x_i$, а в самих узлах x_i требуется знать также первые и вторые производные, поэтому сеточные функции $y_i = f(x_i)$ необходимо восполнять. Данная проблема решается с помощью методов теории приближений (теории аппроксимации) — одного из важнейших разделов численного анализа.

Для восполнения исходных (аппроксимируемых) функций $y_i = f(x_i)$, $i = \overline{0, n}$, искомыми (аппроксимирующими) функциями $\tilde{y} = \tilde{f}_m(x, \bar{a})$, как правило, используются алгебраические многочлены

$$\tilde{f}_m(x, \bar{a}) = \sum_{j=0}^m a_j x^j = a_0 + a_1 x + \dots + a_m x^m, \quad (4.1)$$

где $\bar{a} = (a_0, a_1, \dots, a_m)^T$ — вектор неизвестных параметров, m — степень многочлена, подлежащие выбору или определению (обычно $m \leq n$).

Замечания

1. Аппроксимация тригонометрическими многочленами с помощью рядов Фурье изучается в классических курсах математического анализа. Алгебраические многочлены являются частным случаем обобщенных многочленов, рассматриваемых в п. 4.4. Все понятия, вводимые в данном разделе, справедливы и для обобщенных многочленов.

2. Иногда в качестве аппроксимируемых функций берутся сложные формульные функции $y = f(x)$, которые желательно заменить более простыми.

Методы приближения различаются выбором различных по характеру условий согласования функций $\tilde{f}_m(x, \bar{a})$ и $y_i = f(x_i)$. В данной книге рассматриваются два типа условий согласования — точечные (дискретные) и интегральные.

Применительно к некоторой точке x_i сетки $\Omega_n = \{x_i, i = \overline{0, n}\}$ дискретные условия согласования записываются в виде нулевых невязок искомой функции и ее производных:

$$\delta\tilde{f}_m^{(p)}(x_i, \bar{a}) = \tilde{f}_m^{(p)}(x_i, \bar{a}) - f^{(p)}(x_i) = 0, \quad (4.2)$$

где p — порядок производной. При $p = 0$ условие (4.2) является функциональным и оно традиционно называется условием интерполяции:

$$\delta\tilde{f}_m(x_i, \bar{a}) = \tilde{f}_m(x_i, \bar{a}) - f(x_i) = 0. \quad (4.3)$$

Возможны два типа интегральных условий:

$$1) \sqrt{\frac{1}{n+1} \sum_{i=0}^n [\tilde{f}_m(x_i, \bar{a}) - f(x_i)]^2} \rightarrow \min_{\bar{a}} \quad (m \leq n); \quad (4.4)$$

$$2) \delta\tilde{f}\left(I_{x_i}^{x_{i+1}}\right) = \int_{x_i}^{x_{i+1}} \tilde{f}_m(x, \bar{a}) dx - I_i^{i+1} = 0$$

или (4.5)

$$\delta\tilde{f}\left(I_{x_0}^{x_i}\right) = \int_{x_0}^{x_i} \tilde{f}_m(x, \bar{a}) dx - I_0^i = 0, \quad (4.5)$$

где $I_i^{i+1} = \int_{x_i}^{x_{i+1}} f(x) dx$, $I_0^i = \int_{x_0}^{x_i} f(x) dx$. Условие (4.4) выражает минимум среднеквадратичной погрешности (или отклонения) представления заданной сеточной функции $f(x_i)$, $i = 0, n$, с помощью аппроксимирующей функции $\tilde{f}_m(x, \bar{a})$. Оно относится, как правило, ко всей области определения функции $f(x_i)$, т. е. к отрезку $[a, b]$. Сомножитель $\frac{1}{n+1}$ иногда опускают, так как его наличие или отсутствие влияет только на величину погрешности, но не влияет на вектор \bar{a} , обеспечивающий ее минимум. Условие (4.5) может относиться как к одному частичному отрезку $[x_i, x_{i+1}]$, так и к отрезку $[x_0, x_i]$, а также ко всему отрезку $[a, b]$. Условие (4.5) в одномерном случае, когда функция $f(x)$ зависит только от x , выражает равенство площадей под кривыми $f(x)$ и $\tilde{f}_m(x, \bar{a})$. В двумерном случае, когда функция зависит от двух переменных, условие (4.5) обобщается и выражает равенство объемов (аппроксимация таких функций здесь не рассматривается). Таким образом, применение условия (4.5) приводит к сохранению интегральных свойств аппроксимируемой функции, а также предоставляет исследователям новые ранее неиспользованные возможности при построении новых нетрадиционных методов решения задач аппроксимации [15], [16], [19–21].

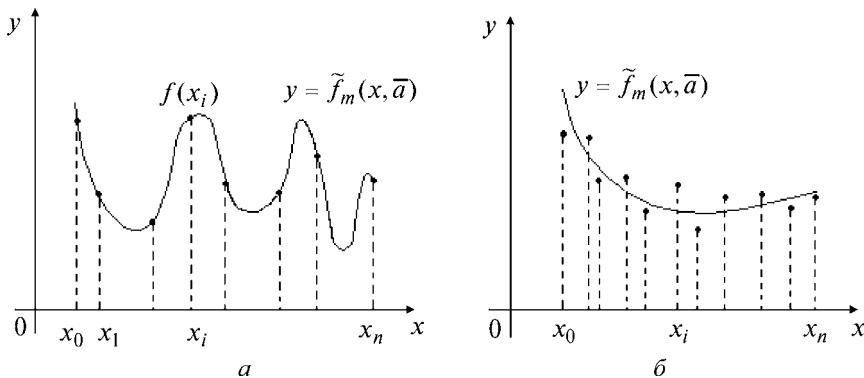


Рис. 4.1

Метод приближения, построенный на базе условия (4.3), будем называть *методом интерполяции*, или *функциональным методом* (рис. 4.1а).

Метод, использующий условие (4.2) при $p > 0$, называется *дифференциальным*.

Если совместно с условием (4.2) применяется интегральное условие, метод называется *интегрально-дифференциальным*.

Метод, построенный на базе условия (4.3) и интегрального условия, называется *интегрально-функциональным*. Методы приближения, сконструированные только на базе условия (4.4) или (4.5), выполняют сглаживание исходной функции (рис. 4.1б) и поэтому называются методами *сглаживания (интегрального сглаживания)*.

Если условия (4.5) используются отдельно, метод сглаживания называется *восстанавливющим*, а если совместно с условиями (4.3), то — *интерполяционно-сглаживающим*.

Названия многочленов, получаемых в результате применения методов, соответствуют названиям этих методов.

Можно выделить четыре способа применения методов аппроксимации, отличающихся областями их «действия».

1. *Глобальный* способ, в котором для всей области $\Omega = [a, b]$ определяется одна функция $\tilde{f}_m(x, \bar{a})$.

2. *Локальный* способ, когда функция восполняется только в окрестности некоторой точки x_i . Это восполнение обычно осуществляется на основе формулы Тейлора. Данный способ рассматривается в курсе математического анализа.

3. *Кусочный* способ, когда ищется одна или несколько функций $\tilde{f}_{ki}(x, \bar{a})$, $i = 0, 1, \dots$, каждая из которых является многочленом степени k и имеет область определения в виде частичного отрезка $\Omega_{ik} = [x_i, x_{i+k}]$ ($1 \leq k < n$, $k = 1, 2, \dots$), называемого «окном» аппроксимации, которое составляет *шаблон* $(x_i, x_{i+1}, \dots, x_{i+k})$. Саму функцию $\tilde{f}_{ki}(x, \bar{a})$, построенную на одном шаблоне, будем называть *звеном*.

4. *Кусочно-глобальный* способ, в котором область Ω представляется совокупностью N частичных отрезков Ω_{ik} , таких, что $\Omega = \bigcup_{i=1}^N \Omega_{ik}$. На первом этапе

на каждом из отрезков ищется функция $\tilde{f}_{ki}(x, \bar{a})$ — i -е звено с применением кусочного способа аппроксимации. На следующем этапе производится объединение всех звеньев в одну многозвенную функцию, т. е. $\tilde{f}_k(x, \bar{a}) = \bigcup_{i=1}^N \tilde{f}_{ki}(x, \bar{a})$.
Данный способ применяется, например, в п. 4.5 при построении сплайнов.

4.2. МЕТОДЫ ФУНКЦИОНАЛЬНОЙ ИНТЕРПОЛЯЦИИ

4.2.1. ПОСТАНОВКА ЗАДАЧИ

Пусть на множестве $\Omega = [a, b]$ задана сетка $\Omega_n = \{x_i, i = \overline{0, n}\}$, определяемая $n + 1$ точкой x_0, x_1, \dots, x_n , а на сетке задана сеточная функция $y_i = f(x_i)$, $i = \overline{0, n}$:

$$y_0 = f(x_0), \quad y_1 = f(x_1), \dots, \quad y_n = f(x_n). \quad (4.6)$$

В некоторых случаях $y_i = f(x_i)$, $i = \overline{0, n}$, является сеточным представлением заданной формульной функции $y = f(x)$. Сеточная функция может задаваться совокупностью пар: $(x_0, y_0), (x_1, y_1), \dots, (x_n, y_n)$.

Требуется найти функцию $y = F(x)$, принимающую в точках x_0, x_1, \dots, x_n те же значения, что и функция $y_i = f(x_i)$, $i = \overline{0, n}$, т. е. $F(x_i) = y_i$, $i = \overline{0, n}$.

Точки x_0, x_1, \dots, x_n называются *узлами интерполяции*, а искомая функция $y = F(x)$ — *интерполирующей*.

Геометрически это означает, что нужно найти кривую, проходящую через заданное множество точек (x_i, y_i) , $i = \overline{0, n}$ (рис. 4.2).

Одной из целей задачи интерполяции является вычисление значения функции в произвольной точке x_* . При этом различаются собственно *интерполирование*, когда точка $x_* \in [x_0, x_n]$, и *экстраполирование*, когда $x_* \notin [x_0, x_n]$.

Заметим, что можно провести бесчисленное множество «плавных» кривых, проходящих через заданное множество точек. Поэтому задача интерполяции в общей постановке не имеет единственного решения.

Если в качестве интерполирующей функции выбрать алгебраический многочлен, степень которого связана с числом заданных узлов интерполяции (на единицу меньше), решение задачи является единственным. Покажем это.

Воспользуемся сначала *кусочным способом*. Выделим из отрезка $[x_0, x_n]$ частичный отрезок $[x_i, x_{i+k}]$ и рассмотрим сеточную функцию $y_i = f(x_i)$, заданную в $(k+1)$ -м узлах $x_i, x_{i+1}, \dots, x_{i+k}$ (узлы не совпадают):

$$y_i = f(x_i), \quad y_{i+1} = f(x_{i+1}), \dots, \quad y_{i+k} = f(x_{i+k}). \quad (4.7)$$

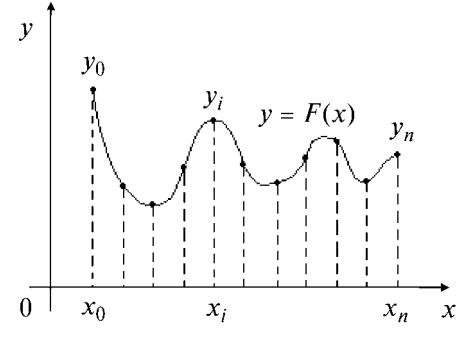


Рис. 4.2

В качестве интерполирующей функции выберем алгебраический многочлен k -й степени (степень многочлена на единицу меньше количества узлов):

$$F(x) = \tilde{f}_k(x, \bar{a}) = \sum_{j=0}^k a_j x^j = a_0 + a_1 x + \dots + a_k x^k. \quad (4.8)$$

Будем искать неизвестные коэффициенты a_0, a_1, \dots, a_k из условия интерполяции (4.3), т. е. $\delta \tilde{f}_k(x_i, \bar{a}) = 0$:

$$\tilde{f}_k(x_i, \bar{a}) = y_i, \quad \tilde{f}_k(x_{i+1}, \bar{a}) = y_{i+1}, \dots, \quad \tilde{f}_k(x_{i+k}, \bar{a}) = y_{i+k}. \quad (4.9)$$

Теорема 4.1 (о единственности решения задачи интерполяции).

Задача о нахождении интерполяционного многочлена $\tilde{f}_k(x, \bar{a}) = \sum_{j=0}^k a_j x^j$, удовлетворяющего условиям (4.9), на частичном отрезке $[x_i, x_{i+k}]$ по заданной сеточной функции (4.7) имеет единственное решение.

Доказательство. Запишем условия интерполяции (4.9) с учетом (4.8) и обозначения $y_i = f(x_i) = f_i$:

$$\begin{aligned} a_0 + a_1 x_i + a_2 x_i^2 + \dots + a_k x_i^k &= f_i, \\ a_0 + a_1 x_{i+1} + a_2 x_{i+1}^2 + \dots + a_k x_{i+1}^k &= f_{i+1}, \\ \dots & \\ a_0 + a_1 x_{i+k} + a_2 x_{i+k}^2 + \dots + a_k x_{i+k}^k &= f_{i+k}. \end{aligned} \quad (4.10)$$

Эта система линейных алгебраических уравнений относительно коэффициентов a_0, a_1, \dots, a_k имеет единственное решение, так как определитель матрицы системы

$$C_k(x_i, x_{i+1}, \dots, x_{i+k}) = \begin{pmatrix} 1 & x_i & x_i^2 & \dots & x_i^k \\ 1 & x_{i+1} & x_{i+1}^2 & \dots & x_{i+1}^k \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{i+k} & x_{i+k}^2 & \dots & x_{i+k}^k \end{pmatrix} \quad (4.11)$$

не равен нулю (доказательство последнего факта содержится в курсе линейной алгебры, где этот определитель называется определителем Вандермонда). Следовательно, задача интерполяции также имеет единственное решение.

Полагая $i = 0, k = n$, приходим к *глобальному способу* решения поставленной задачи. А именно, если задана сеточная функция в $(n + 1)$ -м узлах x_0, x_1, \dots, x_n и требуется найти алгебраический многочлен с использованием условий интерполяции, то единственным решением задачи интерполяции является интерполяционный многочлен n -й степени:

$$F(x) = \tilde{f}_n(x, \bar{a}) = \sum_{j=0}^n a_j x^j = a_0 + a_1 x + \dots + a_n x^n, \quad (4.12)$$

коэффициенты которого находятся из системы (4.10) при $i = 0, k = n$.

Замечания

1. Матрица (4.11) имеет дискретный (точечный) характер, так как ее элементы вычисляются по дискретным значениям $x_i, x_{i+1}, \dots, x_{i+k}$.

2. При решении поставленной задачи предполагается, что исходная сеточная функция задана своими точными значениями, хотя класс задач, для которых используются такие функции, ограничен.

3. Имеются и другие формы записи интерполяционных многочленов. По теореме 4.1 все эти многочлены степени n , удовлетворяющие функциональным условиям интерполяции и построенные по одним и тем же точкам, являются одним многочленом, записанным в разных формах.

4. При большом числе узлов решение системы (4.10) затруднительно. Искомый интерполяционный многочлен можно построить, не решая этой системы. Многочлены могут быть построены так, чтобы в самой структуре формулы многочлена условие интерполяции учитывалось (см. п. 4.2.2 и 4.2.3).

5. При решении задачи функциональной интерполяции и в ее приложениях требуется:

а) выбрать наиболее удобную форму и степень интерполяционного многочлена. При этом можно использовать многочлены Лагранжа (см. п. 4.2.2) или Ньютона (см. п. 4.2.3), а также формулу (4.8);

б) оценить погрешность интерполяции;

в) определить значения функции в точках, не совпадающих с узлами;

г) вычислить значения производных или определенных интегралов с использованием полученных интерполяционных многочленов.

Методика решения задачи интерполяции

1. По заданной сеточной функции составить интерполяционный многочлен определенной степени. При выборе степени многочлена следует руководствоваться желаемой точностью интерполяции.

2. Вычислить значения интерполяционного многочлена в заданных точках x_{*j} , $j = 1, \dots, p$, путем их подстановки в формулу многочлена.

4.2.2. МНОГОЧЛЕН ЛАГРАНЖА

ПОСТАНОВКА ЗАДАЧИ И ПОСТРОЕНИЕ МНОГОЧЛЕНА

Пусть исходная сеточная функция задана в $(n + 1)$ -й точках сетки Ω_n : $y_i = f(x_i)$ $i = \overline{0, n}$, где $x_i \in [a, b] = [x_0, x_n]$ — в общем случае неравноотстоящие узлы, определяемые шагами $h_{i+1} = x_{i+1} - x_i$ ($h_{i+1} = \text{var}$).

Воспользуемся сначала *кусочным способом*. Здесь и далее будем использовать обозначение $f_i = f(x_i)$.

Выделим «окно» или частичный отрезок $[x_i, x_{i+1}]$, содержащий только две точки (шаблон (x_i, x_{i+1})). Тогда многочлен Лагранжа, интерполирующий исходную функцию на данном шаблоне, имеет вид

$$L_1(x) = \frac{(x - x_{i+1})}{(x_i - x_{i+1})} f_i + \frac{(x - x_i)}{(x_{i+1} - x_i)} f_{i+1} = P_{1i}(x) \cdot f_i + P_{1i+1}(x) \cdot f_{i+1}, \quad (4.13)$$

где $P_{1i}(x)$, $P_{1i+1}(x)$ — коэффициенты.

Действительно, путем подстановки легко убедиться в том, что $L_1(x)$ — алгебраический многочлен первой степени, который удовлетворяет условиям функциональной интерполяции (4.3), т. е. $L_1(x_i) = f_i$, $L_1(x_{i+1}) = f_{i+1}$.

Выделим «окно» в виде двойного частичного отрезка $[x_i, x_{i+2}]$ с шаблоном (x_i, x_{i+1}, x_{i+2}) . Тогда многочлен Лагранжа записывается в виде

$$\begin{aligned} L_2(x) = & \frac{(x - x_{i+1}) \cdot (x - x_{i+2})}{(x_i - x_{i+1}) \cdot (x_i - x_{i+2})} f_i + \frac{(x - x_i) \cdot (x - x_{i+2})}{(x_{i+1} - x_i) \cdot (x_{i+1} - x_{i+2})} f_{i+1} + \\ & + \frac{(x - x_i) \cdot (x - x_{i+1})}{(x_{i+2} - x_i) \cdot (x_{i+2} - x_{i+1})} f_{i+2} = P_{2i}(x) \cdot f_i + P_{2i+1}(x) \cdot f_{i+1} + P_{2i+2}(x) \cdot f_{i+2}, \end{aligned} \quad (4.14)$$

где $P_{2i}(x), P_{2i+1}(x), P_{2i+2}(x)$ — коэффициенты.

Легко проверить, что (4.14) — многочлен второй степени и также удовлетворяет условиям функциональной интерполяции:

$$L_2(x_i) = f_i; \quad L_2(x_{i+1}) = f_{i+1}; \quad L_2(x_{i+2}) = f_{i+2}.$$

Обобщив запись многочлена на «окно» для k -кратного частичного отрезка $[x_i, x_{i+k}]$ с шаблоном $(x_i, x_{i+1}, \dots, x_{i+k})$, можно записать многочлен Лагранжа в виде

$$\begin{aligned} L_k(x) = & \sum_{m=i}^{i+k} \frac{(x - x_i) \cdot (x - x_{i+1}) \dots (x - x_{m-1}) \cdot (x - x_{m+1}) \dots (x - x_k)}{(x_m - x_i) \cdot (x_m - x_{i+1}) \dots (x_m - x_{m-1}) \cdot (x_m - x_{m+1}) \dots (x_m - x_k)} f_m = \\ & = \sum_{m=i}^{i+k} P_{km}(x) \cdot f_m, \end{aligned} \quad (4.15)$$

где $P_{km}(x)$ — коэффициенты Лагранжа, которые для внутренних точек шаблона записываются следующим образом:

$$P_{km}(x) = \frac{(x - x_i) \cdot (x - x_{i+1}) \dots (x - x_{m-1}) \cdot (x - x_{m+1}) \dots (x - x_k)}{(x_m - x_i) \cdot (x_m - x_{i+1}) \dots (x_m - x_{m-1}) \cdot (x_m - x_{m+1}) \dots (x_m - x_k)}.$$

Легко проверить, что $P_{km}(x)$ удовлетворяют условию

$$P_{km}(x_j) = \begin{cases} 1, & j = m, \\ 0, & j \neq m, \end{cases} \quad i \leq j \leq i+k. \quad (4.16)$$

Если положить $i = 0, k = n$, то приходим к *глобальному способу решения задачи*. Тогда *интерполяционный многочлен Лагранжа n -й степени* имеет вид

$$L_n(x) = \sum_{i=0}^n \frac{(x - x_0) \cdot (x - x_1) \dots (x - x_{i-1}) \cdot (x - x_{i+1}) \dots (x - x_n)}{(x_i - x_0) \cdot (x_i - x_1) \dots (x_i - x_{i-1}) \cdot (x_i - x_{i+1}) \dots (x_i - x_n)} f_i = \sum_{i=0}^n P_{ni}(x) \cdot f_i, \quad (4.17)$$

где *коэффициенты Лагранжа* $P_{ni}(x)$ во внутренних точках отрезка записываются в форме

$$P_{ni}(x) = \frac{(x - x_0) \cdot (x - x_1) \dots (x - x_{i-1}) \cdot (x - x_{i+1}) \dots (x - x_n)}{(x_i - x_0) \cdot (x_i - x_1) \dots (x_i - x_{i-1}) \cdot (x_i - x_{i+1}) \dots (x_i - x_n)}.$$

Очевидно, многочлен $L_n(x)$, заданный равенством (4.17), является многочленом степени n и удовлетворяет функциональным условиям интерполяции (4.3): $L_n(x_i) = f_i, i = \overline{0, n}$.

Для записи интерполяционного многочлена Лагранжа удобно пользоваться таблицей 4.1.

Тогда многочлен Лагранжа может быть записан в форме

$$L_n(x) = \Pi_{n+1}(x) \sum_{i=0}^n \frac{f_i}{D_i}. \quad (4.18)$$

Замечания

1. Если заданная сеточная функция такая, что $f_i = 1$ ($i = \overline{0, n}$), то из (4.17) следует, что $L_n(x) \equiv 1$, и поэтому справедливо равенство $\sum_{i=0}^n P_{ni}(x) = 1$ (сумма всех коэффициентов Лагранжа в точке x равна единице), которое можно использовать для контроля правильности расчетов.

2. Коэффициенты Лагранжа $P_{ni}(x)$ для некоторой функции $y_i = f(x_i)$ определяются лишь узлами сетки и точкой x , в которой необходимо вычислить значение многочлена. Если в некоторой точке x_* требуется определить значения нескольких интерполируемых функций $f_1(x_i), f_2(x_i), \dots$ ($i = \overline{1, n}$), то коэффициенты Лагранжа для всех исходных функций подсчитываются только один раз.

3. При введении дополнительных узлов интерполяции все коэффициенты многочлена Лагранжа необходимо пересчитывать заново, что неудобно на практике. От этого недостатка свободны многочлены Ньютона (см. п. 4.2.3).

Перейдем к рассмотрению примеров решения задачи интерполяции на основе вышеизложенной методики.

Пример 4.1. Построить многочлен Лагранжа третьей степени для сеточной функции, заданной таблицей 4.2. Вычислить значение функции в точке $x_* = 2,5$. Для записи многочлена использовать формулу (4.18).

□1. Построим многочлен Лагранжа. Для этого составим таблицу 4.3, соответствующую таблице 4.1.

Таблица 4.1

$x - x_0$	$x_0 - x_1$	$x_0 - x_2$	\dots	$x_0 - x_n$	D_0	f_0
$x_1 - x_0$	$x - x_1$	$x_1 - x_2$	\dots	$x_1 - x_n$	D_1	f_1
$x_2 - x_0$	$x_2 - x_1$	$x - x_2$	\dots	$x_2 - x_n$	D_2	f_2
\vdots	\vdots	\vdots		\vdots	\vdots	\vdots
$x_n - x_0$	$x_n - x_1$	$x_n - x_2$	\dots	$x - x_n$	D_n	f_n
$\Pi_{n+1}(x) = (x - x_0)(x - x_1)(x - x_2) \cdot \dots \cdot (x - x_n)$					D_i	f_i

Примечания. D_i — произведение элементов i -й строки, $\Pi_{n+1}(x)$ — произведение элементов главной диагонали, $y_i = f(x_i) = f_i$, $i = 0, 1, \dots, n$.

Таблица 4.2

i	0	1	2	3
x_i	2	3	4	5
$f(x_i) = f_i$	7	5	8	7

Таблица 4.3

$x - 2$	-1	-2	-3	$-6 \cdot (x - 2)$	7
1	$x - 3$	-1	-2	$2 \cdot (x - 3)$	5
2	1	$x - 4$	-1	$-2 \cdot (x - 4)$	8
3	2	1	$x - 5$	$6 \cdot (x - 5)$	7
$\Pi_4(x) = (x - 2)(x - 3)(x - 4)(x - 5)$			D_i	f_i	

По формуле (4.18) получаем:

$$\begin{aligned}
 L_3(x) = & \Pi_4(x) \sum_{i=0}^3 \frac{f_i}{D_i} = \frac{(x-2)(x-3)(x-4)(x-5) \cdot 7}{-6 \cdot (x-2)} + \frac{(x-2)(x-3)(x-4)(x-5) \cdot 5}{2 \cdot (x-3)} + \\
 & + \frac{(x-2)(x-3)(x-4)(x-5) \cdot 8}{-2 \cdot (x-4)} + \frac{(x-2)(x-3)(x-4)(x-5) \cdot 7}{6 \cdot (x-5)} = \\
 = & -\frac{7}{6} \cdot (x-3)(x-4)(x-5) + \frac{5}{2} \cdot (x-2)(x-4)(x-5) - 4 \cdot (x-2)(x-3)(x-5) + \\
 & + \frac{7}{6} \cdot (x-2)(x-3)(x-4) = -\frac{3}{2}x^3 + 16x^2 - \frac{107}{2}x + 62.
 \end{aligned}$$

2. Вычислим значение функции в заданной точке: $L_3(2,5) = 4,8125$. ■

ПОГРЕШНОСТЬ ИНТЕРПОЛЯЦИИ МНОГОЧЛЕНАМИ ЛАГРАНЖА

При определении значения $f(x)$, $x \neq x_i$, для функции $y_i = f(x_i)$ ($i = \overline{0, n}$) с помощью многочлена Лагранжа возникает *погрешность* или *остаточное слагаемое* $R_n(x)$:

$$f(x) = L_n(x) + R_n(x). \quad (4.19)$$

Здесь предполагается, что используется глобальный способ интерполяции и что $f(x) \in C_{n+1}[a, b]$. Последнее предположение требуется для применения соответствующих теорем математического анализа, однако приведенное ниже соотношение для $R_n(x)$ может использоваться и для сеточных функций.

На основе указанных предположений доказано [2], [6], [14], что при интерполяции функции $y_i = f(x_i)$, заданной в общем случае на неравномерной сетке Ω_n , интерполяционным многочленом Лагранжа $L_n(x) = \sum_{i=0}^n P_{ni}(x)f_i$ для произвольного значения $x \in (x_0, x_n)$ возникает погрешность

$$R_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \omega_n(x), \quad (4.20)$$

где $\omega_n(x) = (x - x_0)(x - x_1) \dots (x - x_n)$ — многочлен $(n+1)$ -й степени, а $\xi \in (a, b)$.

Поскольку точно найти $R_n(x)$ нельзя (из-за неопределенности точки ξ), то при проведении вычислений обычно находятся только приближенные оценки погрешностей интерполяции, которые являются априорными.

Оценка погрешности интерполяции в некоторой произвольной фиксированной точке $x_ \in [a, b]$* имеет вид

$$|f(x_*) - L_n(x_*)| \leq \frac{M_{n+1}}{(n+1)!} |\omega_n(x_*)|, \quad (4.21)$$

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

Оценка максимальной погрешности интерполяции в любой точке $x \in [a, b]$, т. е. на всем отрезке $[a, b]$:

$$|f(x) - L_n(x)| \leq \frac{M_{n+1}}{(n+1)!} \max_{[a,b]} |\omega_n(x)|. \quad (4.22)$$

Замечание. Для сеточных функций с фиксированными узлами сетки (узлами интерполяции) также можно проводить оценки погрешности по формулам (4.21), (4.22), однако для этого необходимо численно определять M_{n+1} с помощью аппарата численного дифференцирования. При этом следует учитывать, что при вычислении производных высокого порядка возникают большие погрешности (см. главу 5).

Пример 4.2. С какой точностью можно вычислить значение $f(x) = \sqrt{x}|_{x_*=85}$, если вычисления производить на основе интерполяционного многочлена Лагранжа первой и второй степени, а в качестве сеточной функции принять $(x_i, y_i) = \{(16; 4), (36; 6), (100; 10)\}, (i = 0, 1, 2)$?

□ Так как требуется вычислить погрешность только в одной точке $x_* = 85$, то необходимо использовать формулу (4.21).

Найдем оценку погрешности в точке $x_* = 85$ для $L_1(x)$. В качестве «окна» линейной интерполяции ($n = 1$) выбираем отрезок $[x_1, x_2] = [36; 100]$. Определим величину M_2 :

$$f'(x) = \frac{1}{2}x^{-\frac{1}{2}}, \quad f''(x) = -\frac{1}{4}x^{-\frac{3}{2}}; \quad M_2 = \max_{[36;100]} |f''(x)| = \frac{1}{4\sqrt{36^3}} = \frac{1}{864}.$$

Тогда

$$\omega_1(x_*) = (85 - 36)(85 - 100) = -735; \quad |R_1(x)| \leq \frac{M_2}{(n+1)!} \cdot |\omega_1(x_*)| = \frac{735}{2 \cdot 864} = 0,425.$$

Найдем оценку погрешности в точке x_* для $L_2(x)$. В качестве «окна» квадратичной интерполяции ($n = 2$) выбираем отрезок $[x_0, x_2] = [16; 100]$. Определим величину M_3 :

$$f'''(x) = \frac{3}{8}x^{-\frac{5}{2}}; \quad M_3 = \max_{[16;100]} |f'''(x)| = \frac{3}{8\sqrt{16^5}} = \frac{1}{2730,7}.$$

Тогда

$$\omega_2(x_*) = (85 - 16)(85 - 36)(85 - 100) = -50715; \quad |R_2(x)| \leq \frac{M_3}{(n+1)!} |\omega_2(x_*)| = 3,1.$$

Погрешность для многочлена $L_2(x)$ получилась больше погрешности для многочлена $L_1(x)$, что обусловлено величиной $\omega_2(x_*)$. ■

О СХОДИМОСТИ ФУНКЦИОНАЛЬНОГО ИНТЕРПОЛЯЦИОННОГО ПРОЦЕССА ДЛЯ НЕПРЕРЫВНЫХ ФУНКЦИЙ

Как отмечалось выше (см. п. 4.2.1), выбор сетки и соответствующей степени интерполяционного многочлена при интерполяции сеточных функций является одной из важных задач, решить которую можно, рассмотрев проблему сходимости интерполяционных процессов [6], [40] для непрерывных функций $y = f(x) \in C_{n+1}[a, b]$.

Будем считать, что интерполяция проводится на последовательности сеток $\Omega_1 = \{x_0^{(1)}, x_1^{(1)}\}$, $\Omega_2 = \{x_0^{(2)}, x_1^{(2)}, x_2^{(2)}\}$, ... $\Omega_n = \{x_0^{(n)}, x_1^{(n)}, \dots, x_n^{(n)}\}$ с возрастающим разбиением k отрезка $[a, b]$: $k_1 = 1; k_2 = 2; \dots; k_n = n$ и т. д.

Если при данных разбиениях при возрастающем $k = 1, 2, \dots, n, \dots$ определяются значения $L_k(x_*)$ в некоторой промежуточной точке x_* , то реализуется интерполяционный процесс, характеризующийся последовательностью значений многочленов: $L_1(x_*), L_2(x_*), \dots, L_n(x_*), \dots$

Интерполяционный процесс для функции $f(x)$ сходится в точке $x_* \in [a, b]$, если существует $\lim_{k \rightarrow \infty} L_k(x_*) = f(x_*)$ (поточечная сходимость).

Для отрезка $[a, b]$ существует понятие равномерной сходимости в некоторой норме, например $\max_{\substack{x \in [a,b] \\ n \rightarrow \infty}} |f(x) - L_n(x)| \rightarrow 0$.

Характер сходимости или расходимости интерполяционного процесса зависит как от гладкости и поведения функции $f(x)$, так и от выбора последовательности сеток Ω_n ($n = 1, 2, \dots$). Так, показано, что если $f(x)$ непрерывна на $[a, b]$, то найдется такая последовательность Ω_n ($n = 1, 2, \dots$), для которой интерполяционный процесс сходится равномерно на $[a, b]$ (теорема Марцинкевича [40]). Однако для дискретных функций, рассматриваемых в данном разделе, эта теорема не применима. Отметим также, что построены расходящиеся интерполяционные процессы и для формульных функций, например $f(x) = |x|$, $x \in [-1; 1]$ и $f(x) = (1 + 25x^2)^{-1}$, $x \in [-1; 1]$. Кроме того, применение многочленов высоких степеней приводит к так называемым «провалам» между узлами интерполяции, часто называемым осцилляциями. Указанные свойства интерполяционных процессов обусловливают нецелесообразность применения интерполяционных многочленов высоких степеней. В связи с этим в вычислительной практике для сеточных функций степень n не берут выше $5 \div 8$ ($n \leq 5 \div 8$) и задание частичного отрезка согласуют с выбранной степенью многочлена.

Рассмотрим часто использующиеся на практике линейную и параболическую интерполяцию.

ЛИНЕЙНАЯ И ПАРАБОЛИЧЕСКАЯ ИНТЕРПОЛЯЦИЯ С ПОМОЩЬЮ МНОГОЧЛЕНА ЛАГРАНЖА

В прикладных расчетах часто применяется простейшая кусочная интерполяция, основанная на многочленах первой степени $L_1(x)$ или второй степени $L_2(x)$. В этом случае функциональная интерполяция называется линейной или параболической (квадратичной) соответственно. Рассмотрим возможные способы их реализации и найдем оценки их погрешностей.

Пусть для сеточной функции $y_i = f(x_i)$, заданной на сетке $\Omega_n = \{x_0, x_1, \dots, x_n\}$, требуется выполнить интерполяционный процесс для определения значения $f(x_*)$, где $x_* \neq x_i$ ($i = \overline{0, n}$), и оценить погрешности. Для обоснованного выбора степени интерполяционного многочлена необходимо указать, какую погрешность имеют значения исходной функции $f(x_i)$ в узлах. Если эта погрешность составляет величину $O(h_{i+1}^2)$ или $O(h_{i+1}^3)$, а в широком классе вычислительных задач обеспечиваются именно такие погрешности, следует использовать линейную или параболическую интерполяцию.

Методика решения задачи линейной интерполяции

1. По расположению заданной точки x_* на оси Ox выбрать из всех частичных отрезков $[x_0, x_1], [x_1, x_2], \dots, [x_i, x_{i+1}], \dots, [x_{n-1}, x_n]$, заданных своими крайними значениями и в совокупности образующих сетку Ω_n , «окно» интерполяции $O \equiv [x_i, x_{i+1}]$, такое, что $x_i < x_* < x_{i+1}$.
2. Для отрезка $[x_i, x_{i+1}]$ вычислить значения коэффициентов Лагранжа $P_{1i}(x_*) = \frac{x_* - x_{i+1}}{x_i - x_{i+1}}$ и $P_{1i+1}(x_*) = \frac{x_* - x_i}{x_{i+1} - x_i}$, входящих в формулу (4.13) для многочлена $L_1(x)$. Правильность полученных значений $P_{1i}(x_*)$, $P_{1i+1}(x_*)$ проверить по условию $P_{1i}(x_*) + P_{1i+1}(x_*) = 1$, которое должно выполняться.
3. Вычислить искомое значение $f(x_*)$ согласно (4.13):

$$f(x_*) \approx L_1(x_*) = P_{1i}(x_*)f_i + P_{1i+1}(x_*)f_{i+1}.$$

Как показано ниже, порядок этой аппроксимации равен двум, т. е. $O(h_{i+1}^2)$.

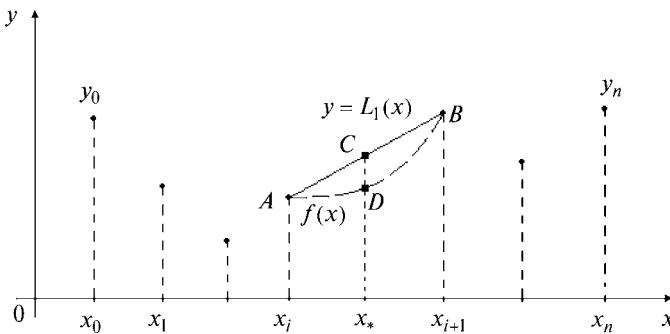


Рис. 4.3

Геометрическая интерпретация линейной интерполяции при известной формульной функции $y = f(x)$ (штриховая линия) изображена на рисунке 4.3. Здесь прямая AB соответствует графику функции $y = L_1(x)$ на отрезке $[x_i, x_{i+1}]$. Приближенное значение функции равно $L_1(x_*)$ (точка C), и оно отстоит от точного значения $f(x_*)$ на величину $CD \approx O(h_{i+1}^2)$ вдоль оси Oy .

Методика решения задачи параболической интерполяции

1. Из всей совокупности спаренных частичных отрезков $[x_0, x_1], [x_1, x_2], \dots, [x_{i-1}, x_{i+1}], [x_i, x_{i+2}], \dots, [x_{n-1}, x_n]$, образующих сетку Ω_n , по заданной величине x_* выбрать два пересекающихся «окна» $O_1 \equiv [x_{i-1}, x_{i+1}], O_2 \equiv [x_i, x_{i+2}]$ (предполагается, что x_* принадлежит внутреннему отрезку $[x_i, x_{i+1}]$).

2. Для «окон» O_1 и O_2 вычислить значения коэффициентов Лагранжа: $P_{2,i-1}^{(1)}(x_*)$, $P_{2,i}^{(1)}(x_*)$, $P_{2,i+1}^{(1)}(x_*)$ — для «окна» O_1 и $P_{2,i}^{(2)}(x_*)$, $P_{2,i+1}^{(2)}(x_*)$, $P_{2,i+2}^{(2)}(x_*)$ — для «окна» O_2 . Путем суммирования проверить правильность полученных значений коэффициентов:

$$\left(\sum_{k=i-1}^{i+1} P_{2,k}(x_*) = 1 \text{ и } \sum_{k=i}^{i+2} P_{2,k}(x_*) = 1 \right).$$

3. Используя значения коэффициентов Лагранжа, вычислить значения $L_2^{(1)}(x_*)$, $L_2^{(2)}(x_*)$ по формуле (4.14).

Если в расчетах не требуется высокая точность интерполяции, то можно ограничиться выбором одного «окна», например O_2 , и тогда $f(x_*) \approx L_2^{(2)}(x_*)$. Для достижения повышенной точности интерполяцию провести для двух «окон» и результаты $L_2^{(1)}(x_*)$, $L_2^{(2)}(x_*)$ осреднить:

$$f(x_*) \approx [L_2^{(1)}(x_*) + L_2^{(2)}(x_*)]/2 \equiv L_{2\text{cp}}(x_*).$$

При этом порядок интерполяции повышается на единицу, т. е. $L_2(x_*)$ аппроксимирует точное значение с четвертым порядком, поскольку погрешность составляет величину $O(H^4)$, где $H = \max\{h_i, h_{i+1}, h_{i+2}\}$.

Замечание. Если $x_* \in [x_0, x_2]$ или $x_* \in [x_{n-1}, x_n]$, то выбирается одно «окно» $[x_0, x_2]$ или $[x_{n-1}, x_n]$ соответственно.

Геометрическая интерпретация параболической интерполяции изображена на рисунке 4.4. Параболе $y = L_2^{(1)}(x)$ соответствует кривая $A_1 A_2 B_1$, параболе $y = L_2^{(2)}(x)$ — кривая $A_2 B_1 B_2$. Точка C соответствует значению $L_2^{(1)}(x_*)$, точка D — значению $L_2^{(2)}(x_*)$, точка E — значению $L_2(x_*)$.

Приведем оценки погрешностей линейной и параболической интерполяции. Вначале предположим, что сетка Ω_n равномерная (это имеет значение только для параболической интерполяции). Формулы (4.13), (4.14) и оценки (4.22), записанные для «окон» интерполяции, упрощаются, если ввести в рассмотрение новую переменную — *фазу интерполяции* $u = \frac{x - x_i}{h}$:

$$L_1(u) = (1-u)f_i + uf_{i+1}; \quad L_2(u) = \frac{(u-1)(u-2)}{2}f_i - u(u-2)f_{i+1} + \frac{u(u-1)}{2}f_{i+2}.$$

Здесь учтено, что $x - x_{i+1} = x - (x_i + h) = uh - h = h(u-1)$, $x - x_{i+2} = h(u-2)$. Очевидно, что для $L_1(u)$ величина u изменяется в диапазоне $0 \leq u \leq 1$, а для $L_2(u)$ — в диапазоне $0 \leq u \leq 2$. Для получения мажорант в оценке (4.22) необходимо найти $\max|\omega_1(x)|$ и $\max|\omega_2(x)|$. Преобразуя зависимости $\omega_1(x)$ и $\omega_2(x)$

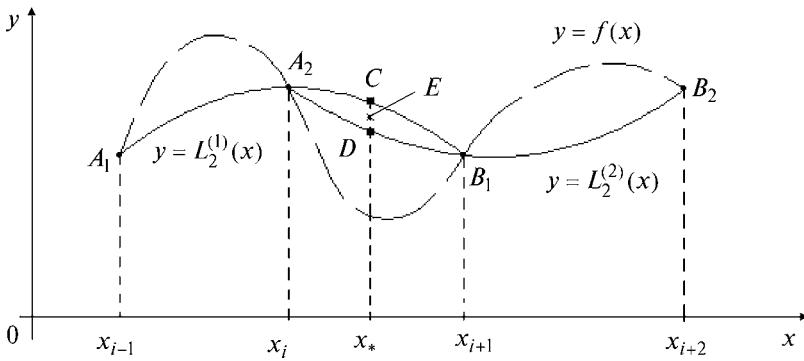


Рис. 4.4

к новой переменной u , получаем $\omega_1(u) = h^2 u \cdot (1 - u)$, $\omega_2(u) = h^3 u \cdot (1 - u) \cdot (2 - u)$ и находим максимумы:

$$\Omega^1 = \max_{0 \leq u \leq 1} |\omega_1(u)| = \frac{h^2}{4}; \quad \Omega^2 = \max_{0 \leq u \leq 2} |\omega_2(u)| = \frac{2\sqrt{3}h^2}{9}.$$

Таким образом, реализуются следующие оценки погрешностей линейной и параболической интерполяции, справедливых для соответствующих «окон»:

$$|f(u) - L_1(u)| \leq \frac{h^2}{8} M_{2,i}; \quad (4.23)$$

$$|f(u) - L_2(u)| \leq \frac{\sqrt{3}h^3}{27} M_{3,i}, \quad (4.24)$$

где $M_{2,i} = \max_{[x_i, x_{i+1}]} |f''(x)|$, $M_{3,i} = \max_{[x_i, x_{i+2}]} |f'''(x)|$. При этом предполагается, что $f(x)$ принадлежит классам функций $f(x) \in C_2[a, b]$, $f(x) \in C_3[a, b]$ соответственно для линейной и параболической интерполяции. Таким образом, из оценок (4.23), (4.24) следует, что линейная интерполяция обеспечивает на частичном отрезке $[x_i, x_{i+1}]$ второй порядок аппроксимации или погрешности по h , а параболическая (без осреднения) на двойном отрезке $[x_i, x_{i+2}]$ — третий порядок. Данные погрешности, как отмечалось во введении и предыдущих разделах, сокращенно записываются как $O(h^2)$ и $O(h^3)$. При реализации алгоритма с осреднением порядок параболической интерполяции становится равным $O(h^4)$.

Замечания

1. Оценка (4.23) для линейной интерполяции инвариантна по отношению к виду сетки Ω_n (равномерной или неравномерной). Параболическая интерполяция также сохраняет указанную погрешность при выполнении интерполяции на неравномерной сетке Ω_n . В [13] для $L_2(x)$ при выполнении условия $f(x) \in C_3[a, b]$ приведена оценка

$$\max_{[x_i, x_{i+2}]} |R_2(x)| \leq \frac{\sqrt{3}}{27} H^3 M_{3,i}, \quad (4.25)$$

где $H = \max(h_{i+1}, h_{i+2})$, $h_{i+2} = x_{i+2} - x_{i+1}$, $M_{3,i} = \max_{[x_i, x_{i+2}]} |f'''(x)|$.

2. Если гладкость функции $f(x)$ не достигает вышеуказанной и класс гладкости понижен на единицу ($f(x) \in C_2[a, b]$), то порядок параболической интерполяции также понижается на единицу:

$$\max_{[x_i, x_{i+2}]} |R_2(x)| \leq 0,1546 \cdot H^2 M_{2,i}. \quad (4.26)$$

3. Повышение класса гладкости функции $f(x)$ выше $C_3[a, b]$ не приводит к увеличению порядка параболической интерполяции относительно H , т. е. происходит как бы его «замораживание» или «насыщение». Указанные свойства, вытекающие из оценок (4.25), (4.26), носят общий характер и имеют место в других оценках аппроксимации многочленов, сплайнов, производных и интегралов.

4. Для произвольной степени интерполяционного многочлена при $h = \text{const}$, $f(x) \in C_{n+1}[a, b]$ погрешность функциональной интерполяции на отрезке $[x_0, x_n]$ выражается следующим образом [6]:

$$\|f(x) - L_n(x)\|_{[a, b]} = \max_{[a, b]} |f(x) - L_n(x)| \leq h^{n+1} \frac{M_{n+1}}{(n+1)!} \Omega^n, \quad (4.27)$$

где $\Omega^n = \max_{0 \leq u \leq n} |\omega_n(u)|$, $M_{n+1} = \max_{[x_0, x_n]} |f^{(n+1)}(x)|$, $\omega_n(u) = u(u-1) \cdot \dots \cdot (u-n)$. Для многочленов с $n \leq 5$ величины Ω^n или их оценки являются такими: $\Omega^1 = \frac{1}{4}$; $\Omega^2 = \frac{2\sqrt{3}}{9}$; $\Omega^3 = 1$; $\Omega^4 < 3,7$; $\Omega^5 < 17$. Из (4.27) следует, что на отрезке $[x_0, x_n]$ величина $\|R_n(x)\|$ есть $O(h^{n+1})$ и при уменьшении шага h в два раза мажоранта уменьшается по крайней мере в 2^{n+1} раз. Отсюда вытекает, что для непрерывных функций можно по заданной точности интерполяции выбирать шаг h . При этом можно путем использования кусочной интерполяции в некоторых пределах изменять степень интерполяционного многочлена. Мажоранту в оценке (4.27) при кусочной интерполяции можно также снизить путем выбора «окна» интерполяции $[x_i, x_{i+k}]$ так, чтобы точка x_* располагалась как можно ближе к его середине. Это обусловлено тем, что колебания функции $\omega_n(x)$ вблизи середины $[x_i, x_{i+k}]$ меньше, чем у его концов [6].

5. В широком классе задач математической физики применяются расчетные схемы в основном второго (и иногда третьего и выше) порядка точности. При их реализации, как правило, используются встроенные интерполяционные алгоритмы, основанные на многочленах и сплайн-функциях. Степени интерполяционных многочленов при этом должны выбираться из условия соответствия порядков их аппроксимации порядкам точности схем. Если эти порядки одинаковы, то порядок точности схем сохраняется, хотя константа в оценке погрешности схемы изменяется. Если же порядок встроенных интерполяционных алгоритмов хотя бы на единицу выше порядка точности схемы, то вместе с порядком точности схемы сохраняется и указанная константа. Отсюда следует, что необходимо выбирать такую степень интерполяционного многочлена, которая либо обеспечивает равенство порядка аппроксимации порядку точности схемы, либо на единицу превышает последний. Таким образом, использование параболической интерполяции в качестве встроенных алгоритмов или для восполнения численных решений, полученных по схемам второго порядка, позволяет сохранить требуемую точность расчета, а также не дает избыточный порядок и, следовательно, не усложняет алгоритм. Это замечание носит общий характер и относится к любым аппроксимационным алгоритмам, выполняющим функцию восполнения или интерполирования.

Перейдем к рассмотрению примеров решения задач линейной и параболической интерполяции.

Пример 4.3. Данна сеточная функция, являющаяся сеточным представлением формулы функции $f(x) = x^3$ (табл. 4.4).

Найти значение $f(x_*)$ при $x_* = 2$ с помощью линейной и параболической интерполяции.

i	0	1	2	3	4
x_i	-1	0	1	3	4
f_i	-1	0	1	27	64

□ Применение линейной интерполяции. Воспользуемся методикой.

1. Выбираем «окно» интерполяции $[x_i, x_{i+1}] = [1; 3]$ ($i = 2, i + 1 = 3$).
2. Вычислим коэффициенты Лагранжа:

$$P_{12} = \frac{x_* - x_3}{x_2 - x_3} = 0,5; \quad P_{13} = \frac{x_* - x_2}{x_3 - x_2} = 0,5.$$

Вычисления выполнены верно, так как $P_{12} + P_{13} = 1$.

3. Определим искомое значение $f(2)$ по формуле (4.13):

$$f(2) \approx L_1(2) = P_{12}f_2 + P_{13}f_3 = 14.$$

Применение параболической интерполяции с осреднением. Также воспользуемся соответствующей методикой.

1. Выбираем «окна» интерполяции $O_1 \equiv [x_1; x_3] = [0; 3]; O_2 \equiv [x_2; x_4] = [1; 4]$.

2. Вычислим коэффициенты Лагранжа $P_{21}^{(1)}, P_{22}^{(1)}, P_{23}^{(1)}$ для «окна» O_1 и $P_{22}^{(2)}, P_{23}^{(2)}, P_{24}^{(2)}$ для «окна» O_2 . Получим $P_{21}^{(1)} = -\frac{1}{3}, P_{22}^{(1)} = 1, P_{23}^{(1)} = \frac{1}{3}, P_{22}^{(2)} = \frac{1}{3}, P_{23}^{(2)} = 1, P_{24}^{(2)} = -\frac{1}{3}$. Вычисления выполнены правильно, так как $\sum_{k=1}^3 P_{2k}^{(1)} = 1; \sum_{k=2}^4 P_{2k}^{(2)} = 1$.

3. Определим значения $L_2^{(1)}(x_*)$, $L_2^{(2)}(x_*)$ по формуле (4.14):

$$L_2^{(1)}(2) = -\frac{1}{3} \cdot 0 + 1 \cdot 1 + \frac{1}{3} \cdot 27 = 10; \quad L_2^{(2)}(2) = \frac{1}{3} \cdot 1 + 1 \cdot 27 - \frac{1}{3} \cdot 64 = 6.$$

Искомое значение получим в результате осреднения:

$$f(2) \approx \frac{L_2^{(1)}(2) + L_2^{(2)}(2)}{2} = 8.$$

Найдено точное значение $f(2) = 8$, что объясняется тем, что оператор осреднения повышает порядок интерполяции. В этом случае в остаточное слагаемое входит $f^{(4)}(\xi) \equiv 0$ и оно равно нулю. ■

4.2.3. МНОГОЧЛЕНЫ НЬЮТОНА

РАЗДЕЛЕННЫЕ И КОНЕЧНЫЕ РАЗНОСТИ

В практике функционального интерполирования иногда удобнее использовать многочлены Ньютона, степень которых можно последовательно повышать путем добавления очередных слагаемых, имеющих более высокую степень. Такие несимметричные многочлены, альтернативные симметричным

многочленам Лагранжа, основаны на разделенных и конечных разностях, вычисляемых по интерполируемой сеточной функции.

Разделенные разности вводятся для функции $y_i = f(x_i) = f_i$, $i = \overline{0, n}$, заданной на неравномерной сетке ($h_{i+1} = \text{var}$), а *конечные разности* — для функции $y_i = f(x_i) = f_i$, $i = 0, n$, определенной на равномерной сетке ($h_{i+1} = \text{const}$).

Выбрав внутри неравномерной или равномерной сетки соответствующие шаблоны интерполяции $(x_i, x_{i+1}), (x_i, x_{i+1}, x_{i+2}), \dots, (x_i, x_{i+1}, \dots, x_{i+k})$, введем следующие определения разделенных и конечных разностей:

- *разделенная разность первого порядка*:

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

- *разделенная разность второго порядка*:

$$f(x_i, x_{i+1}, x_{i+2}) = \frac{f(x_{i+1}, x_{i+2}) - f(x_i, x_{i+1})}{x_{i+2} - x_i};$$

.....

- *разделенная разность k-го порядка*:

$$f(x_i, x_{i+1}, \dots, x_{i+k}) = \frac{f(x_{i+1}, x_{i+2}, \dots, x_{i+k}) - f(x_i, x_{i+1}, \dots, x_{i+k-1})}{x_{i+k} - x_i};$$

- *конечная разность первого порядка*:

$$\Delta f_i = f_{i+1} - f_i;$$

- *конечная разность второго порядка*:

$$\Delta^2 f_i = \Delta(\Delta f_i) = \Delta f_{i+1} - \Delta f_i = f_{i+2} - 2f_{i+1} + f_i;$$

.....

- *конечная разность k-го порядка*:

$$\Delta^k f_i = \Delta(\Delta^{k-1} f) = \sum_{j=0}^k (-1)^j C_k^j f_{i+j},$$

где $C_k^j = \frac{k!}{(k-j)! j!}$.

Последовательность получения разделенных и конечных разностей при $k = 3$ для произвольной функции наглядно представляют таблицы 4.5 и 4.6.

Для гладких функций числовые значения $f(x_j, x_{j+1}, \dots, x_{j+k})$ и $\Delta^k f_i$ при возрастании k уменьшаются и стремятся к нулю, т. е. $f(x_j, x_{j+1}, \dots, x_{j+k}) \rightarrow 0$ и $\Delta^k f_i \rightarrow 0$ при $k \rightarrow \infty$.

Связь между разделенными и конечными разностями k -го порядка при $h = \text{const}$ устанавливается следующим соотношением:

$$f(x_i, x_{i+1}, \dots, x_{i+k}) = \frac{\Delta^k f_i}{k! h^k}. \quad (4.28)$$

Действительно, при $k = 1$ для разделенной разности $f(x_i, x_{i+1})$ получаем

$$f(x_i, x_{i+1}) = \frac{f_{i+1} - f_i}{h} = \frac{\Delta f_i}{1! h^1},$$

т. е. (4.28) при $k = 1$ справедливо.

x_j	f_j	$f(x_j, x_{j+1})$	$f(x_j, x_{j+1}, x_{j+2})$	$f(x_j, x_{j+1}, x_{j+2}, x_{j+3})$
x_i	f_i	$f(x_i, x_{i+1})$	$f(x_i, x_{i+1}, x_{i+2}) = -f(x_i, x_{i+1})$	$f(x_i, x_{i+1}, x_{i+2}, x_{i+3}) = -f(x_i, x_{i+1}, x_{i+2})$
x_{i+1}	f_{i+1}	$f(x_{i+1}, x_i) = -f(x_i, x_{i+1})$	$f(x_i, x_{i+1}, x_{i+2}) = -f(x_i, x_{i+1}, x_{i+2})$	$f(x_i, x_{i+1}, x_{i+2}, x_{i+3}) = -f(x_i, x_{i+1}, x_{i+2}, x_{i+3})$
x_{i+2}	f_{i+2}	$f(x_{i+2}, x_{i+1}) = -f(x_{i+1}, x_i)$	$f(x_{i+1}, x_i) = -f(x_{i+2}, x_{i+1})$	$f(x_i, x_{i+1}, x_{i+2}, x_{i+3}) = -f(x_{i+1}, x_{i+2}, x_{i+3})$
x_{i+3}	f_{i+3}	$f(x_{i+3}, x_{i+2}) = -f(x_{i+2}, x_{i+1})$	$f(x_{i+2}, x_{i+1}) = -f(x_{i+3}, x_{i+2})$	$f(x_i, x_{i+1}, x_{i+2}, x_{i+3}) = -f(x_{i+3}, x_{i+2}, x_{i+1})$

x_j	f_j	Δf_j	$\Delta^2 f_j$	$\Delta^3 f_j$
x_i	f_i	Δf_i	$\Delta^2 f_i$	$\Delta^3 f_i$
x_{i+1}	f_{i+1}	$\Delta f_{i+1} = f_{i+1} - f_i$	$\Delta^2 f_{i+1} = \Delta f_{i+1} - \Delta f_i$	$\Delta^3 f_{i+1} = \Delta^2 f_{i+1} - \Delta^2 f_i$
x_{i+2}	f_{i+2}	$\Delta f_{i+2} = f_{i+2} - f_{i+1}$	$\Delta^2 f_{i+2} = \Delta f_{i+2} - \Delta f_{i+1}$	$\Delta^3 f_{i+2} = \Delta^2 f_{i+2} - \Delta^2 f_{i+1}$
x_{i+3}	f_{i+3}	$\Delta f_{i+3} = f_{i+3} - f_{i+2}$		

Пусть $k = 2$. Тогда получаем

$$\begin{aligned} f(x_i, x_{i+1}, x_{i+2}) &= \frac{f(x_{i+1}, x_{i+2}) - f(x_i, x_{i+1})}{x_{i+2} - x_i} = \\ &= \frac{1}{2h} \left(\frac{f_{i+2} - f_{i+1}}{h} - \frac{f_{i+1} - f_i}{h} \right) = \frac{\Delta f_{i+1} - \Delta f_i}{2h^2} = \frac{\Delta^2 f_i}{2!h^2}. \end{aligned}$$

Таким образом, связь (4.28) выполняется и при $k = 2$. Справедливость (4.28) при произвольном k можно доказать методом математической индукции.

ИНТЕРПОЛЯЦИОННЫЙ МНОГОЧЛЕН НЬЮТОНА ДЛЯ НЕРАВНОМЕРНОЙ СЕТКИ

Пусть исходная (интерполируемая) сеточная функция $y_i = f(x_i)$, $i = \overline{0, n}$, задана на неравномерной сетке $\Omega_n \equiv \{x_0, x_1, x_2, \dots, x_n\}$, характеризующейся шагами $h_{i+1} = x_{i+1} - x_i = \text{var}$.

Воспользуемся сначала *кусочным способом*. Из всей совокупности узлов выбираем шаблон $(x_i, x_{i+1}, \dots, x_{i+k})$, соответствующий некоторому «окну» интерполяции $[x_i, x_{i+k}]$.

Тогда для функциональной интерполяции может быть использован *многочлен Ньютона*, основанный на разделенных разностях:

$$N_k(x) = f_i + f(x_i, x_{i+1}) \cdot (x - x_i) + f(x_i, x_{i+1}, x_{i+2}) \cdot (x - x_i) \cdot (x - x_{i+1}) + \dots + f(x_i, x_{i+1}, \dots, x_{i+k}) \cdot (x - x_i) \cdot (x - x_{i+1}) \cdot \dots \cdot (x - x_{i+k-1}). \quad (4.29)$$

Действительно, $N_k(x)$ — многочлен k -й степени, что определяется сомножителями последнего слагаемого (разделенные разности, входящие в качестве одного из сомножителей в эти произведения, есть числа). Кроме того, для многочлена $N_k(x)$ удовлетворяются функциональные условия интерполяции: $N_k(x_j) = f_j$, $j = i, \dots, i+k$. Проверим их справедливость при $k = 1$ (шаблон (x_i, x_{i+1})) и $k = 2$ (шаблон (x_i, x_{i+1}, x_{i+2})).

Пусть $k = 1$. Тогда

$$N_1(x) = f_i + f(x_i, x_{i+1})(x - x_i), \quad (4.30)$$

и поэтому

$$N_1(x_i) = f_i; \quad N_1(x_{i+1}) = f_i + \frac{f_{i+1} - f_i}{x_{i+1} - x_i} (x_{i+1} - x_i) = f_{i+1}.$$

Таким образом, условия интерполяции для $N_1(x)$ выполнены, следовательно, многочлен (4.30) может быть использован для линейной интерполяции кусочным способом.

Пусть $k = 2$. Тогда

$$N_2(x) = f_i + f(x_i, x_{i+1})(x - x_i) + f(x_i, x_{i+1}, x_{i+2})(x - x_i)(x - x_{i+1}), \quad (4.31)$$

и поэтому

$$\begin{aligned} N_2(x_i) &= f_i; \quad N_2(x_{i+1}) = f_i + \frac{f_{i+1} - f_i}{x_{i+1} - x_i} (x_{i+1} - x_i) + 0 = f_{i+1}; \\ N_2(x_{i+2}) &= f_i + \frac{f_{i+1} - f_i}{x_{i+1} - x_i} (x_{i+2} - x_i) + \frac{1}{x_{i+2} - x_i} \left(\frac{f_{i+2} - f_{i+1}}{x_{i+2} - x_{i+1}} - \frac{f_{i+1} - f_i}{x_{i+1} - x_i} \right) \times \\ &\quad \times (x_{i+2} - x_i)(x_{i+2} - x_{i+1}) = f_{i+2}. \end{aligned}$$

Таким образом, условия интерполяции для многочлена $N_2(x)$ также выполнены, и он может использоваться для параболической интерполяции кусочным способом.

Для произвольного k справедливость равенств $N_k(x_i) = f_i$, $j = \overline{i, i+k}$, проверяется методом математической индукции.

Полагая $i = 0$, $k = n$, приходим к *глобальному способу*. Тогда *интерполяционный многочлен Ньютона n -й степени* имеет вид

$$\begin{aligned} N_n(x) &= f_0 + f(x_0, x_1)(x - x_0) + f(x_0, x_1, x_2)(x - x_0)(x - x_1) + \dots + \\ &\quad + f(x_0, x_1, \dots, x_n)(x - x_0)(x - x_1) \cdot \dots \cdot (x - x_{n-1}). \end{aligned} \quad (4.32)$$

Замечания

1. Согласно теореме 4.1 многочлен Ньютона (4.32) является тождественным многочленом $\tilde{f}_n(x, \bar{a}) = \sum_{j=0}^n a_j x^j$ с коэффициентами, получаемыми из системы (4.10), либо многочлену Лагранжа, т. е. $L_n(x) = N_n(x) = \tilde{f}_n(x, \bar{a})$, если узлы интерполяции и интерполируемая функция одинаковы.

2. Интерполяционный многочлен Ньютона (4.29) или (4.32) (так же, как и многочлен Ньютона, выражаемый ниже через конечные разности) записан не через значения функции, как это имеет место для многочлена Лагранжа, а через разделенные разности. Поэтому при изменении степени k в процессе интерполяирования у многочлена Ньютона $N_k(x)$ требуется только добавить или отбросить соответствующее число слагаемых. Это иногда упрощает алгоритм интерполяирования.

3. При интерполяции на основе (4.29) или (4.32) узлы интерполяции $x_i, x_{i+1}, \dots, x_{i+k}$ или x_0, x_1, \dots, x_n , определяющие шаблоны интерполяции, целе-

сообразно выбирать так, чтобы точка x_* была расположена возможно ближе к середине отрезка $[x_i, x_{i+k}]$ или $[x_0, x_n]$.

4. Остаточное слагаемое многочлена (4.32) совпадает с остаточным слагаемым многочлена Лагранжа, и оценки (4.21), (4.22), справедливые для точки x_* и всего отрезка $[x_0, x_n]$, сохраняются.

ИНТЕРПОЛЯЦИОННЫЕ МНОГОЧЛЕНЫ НЬЮТОНА ДЛЯ РАВНОМЕРНОЙ СЕТКИ

Сначала рассмотрим решение задачи *кусочной интерполяции* (применение кусочного способа). Если функция $y_i = f(x_i)$ ($i=0, n$) задана на равномерной сетке Ω_n , характеризующейся $h_{i+1} = \text{const}$ для всех i , то многочлен (4.29), соответствующий шаблону $(x_i, x_{i+1}, \dots, x_{i+k})$, путем подстановки в него вместо разделенных разностей их выражений через конечные разности, согласно (4.28), преобразуется к виду

$$N_k^{(I)}(q) = f_i + \frac{\Delta f_i}{1!} q + \frac{\Delta^2 f_i}{2!} q(q-1) + \dots + \frac{\Delta^k f_i}{k!} q(q-1)\dots(q-k+1), \quad (4.33)$$

где $q = \frac{x - x_i}{h}$ — фаза интерполяции, определенная относительно точки x_i ; $\Delta^j f_i$ ($j = 1, 2, \dots, k$) — конечные разности.

В соответствии с формулой для фазы интерполяции q точка начала ее отсчета расположена в узле x_i , и входящие в (4.33) конечные разности относятся к этой же точке x_i . В связи с этим (4.33) удобно применять в начале выделенного шаблона $(x_i, x_{i+1}, \dots, x_{i+k})$, когда $q > 0$. Если $x_i = x_0$, а $x_* < x_0$, то этот же многочлен используется и для экстраполяции левее точки x_0 ($q < 0$). Поэтому многочлен $N_k^{(I)}(q)$ называется *интерполяционным многочленом для интерполяции вперед (в начале таблицы)* или *для экстраполяции назад*. В (4.33) этот многочлен обозначен цифрой I, указанной в скобках вверху. Для определенности назовем его *первым интерполяционным многочленом Ньютона*.

Полагая $i = 0, k = n$, получаем решение задачи *глобальной интерполяции* на всем отрезке $[x_0, x_n]$:

$$N_n^{(I)}(q) = f_0 + \frac{\Delta f_0}{1!} q + \frac{\Delta^2 f_0}{2!} q(q-1) + \dots + \frac{\Delta^n f_0}{n!} q(q-1)\dots(q-n+1), \quad (4.34)$$

где $q = \frac{x - x_0}{h}$. Остаточное слагаемое этого многочлена имеет вид

$$R_n(q) = h^{n+1} \frac{q(q-1)\dots(q-n)}{(n+1)!} f^{(n+1)}(\xi),$$

где $\xi \in (a, b)$ — некоторое промежуточное значение между узлами x_0, x_1, \dots, x_n и точкой x .

Если фазу интерполяции определить относительно x_{i+k} некоторой конечной точки шаблона $(x_i, x_{i+1}, \dots, x_{i+k})$, т. е. $\hat{q} = \frac{x - x_{i+k}}{h}$, то вместо (4.33) получается *второй интерполяционный многочлен Ньютона*:

$$N_k^{(II)}(\hat{q}) = f_{i+k} + \frac{\Delta f_{i+k-1}}{1!} \hat{q} + \frac{\Delta^2 f_{i+k-2}}{2!} \hat{q}(\hat{q}+1) + \dots + \frac{\Delta^k f_i}{k!} \hat{q}(\hat{q}+1)\dots(\hat{q}+k-1). \quad (4.35)$$

Данный многочлен удобно применять в конце выделенного шаблона или всей таблицы $y_i = f(x_i)$ ($i = 0, n$).

Если $i + k = n$, а $x_* > x_n$, то (4.35) используется для экстраполяции правее точки $x_n (\hat{q} > 0)$. Поэтому многочлен $N_k^{(II)}(\hat{q})$ называется *многочленом для интерполяции назад (в конце таблицы)* или *экстраполяции вперед*.

Полагая $i = 0, i + k = n$, получаем решение задачи *глобальной интерполяции — второй интерполяционный многочлен Ньютона n -й степени*:

$$N_n^{(II)}(\hat{q}) = f_n + \frac{\Delta f_{n-1}}{1!} \hat{q} + \frac{\Delta^2 f_{n-2}}{2!} \hat{q}(\hat{q}+1) + \dots + \frac{\Delta^n f_0}{n!} \hat{q}(\hat{q}+1)\dots(\hat{q}+n-1), \quad (4.36)$$

где $\hat{q} = \frac{x - x_n}{h}$.

Остаточное слагаемое многочлена (4.36) имеет вид

$$R_n(\hat{q}) = h^{n+1} \frac{\hat{q}(\hat{q}+1)\dots(\hat{q}+n)}{(n+1)!} f^{(n+1)}(\xi),$$

где $\xi \in (a, b)$.

Схема выбора узлов интерполяции при изменении степеней интерполяционных многочленов $N_k^{(I)}(q)$, $N_k^{(II)}(\hat{q})$ ($k = 1, 2, 3$) в одном алгоритме показана на рисунке 4.5. Если точка x_* находится в начале отрезка $[x_0, x_n]$, например на частичном отрезке $[x_0, x_1]$, то применяется первый интерполяционный многочлен Ньютона, а если в конце, например на частичном отрезке $[x_{n-1}, x_n]$, то — второй (рис. 4.5a). Если точка x_* находится вдали от концов отрезка $[x_0, x_n]$, то может применяться как первый интерполяционный многочлен, так и второй (рис. 4.5б).

Замечания

1. Из формул интерполяционных многочленов $N_k^{(I)}(q)$ и $N_k^{(II)}(\hat{q})$ видно, что повышение их степеней в процессе реализации алгоритма не требует пе-

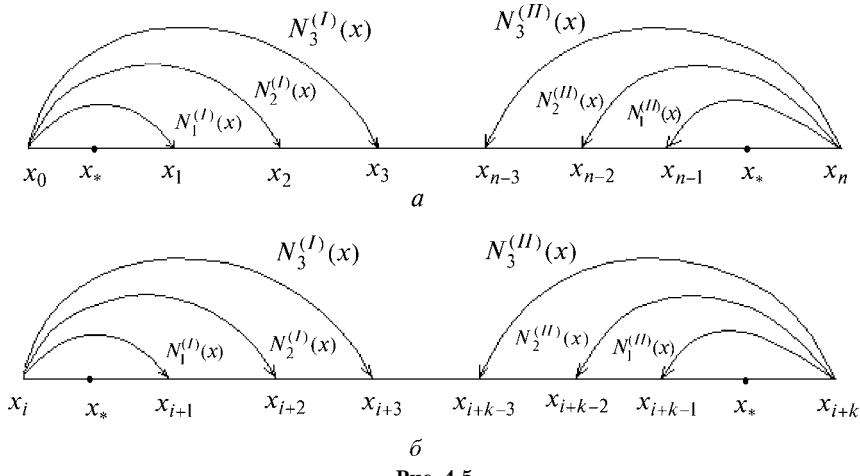


Рис. 4.5

рассчета предыдущих слагаемых, входящих в многочлены с меньшими степенями.

2. Для гладких функций при повышении порядка конечных разностей справедливо свойство: $\Delta^k f_i \rightarrow 0$ при $k \rightarrow \infty$, и поэтому, как только очередное слагаемое в рассматриваемых многочленах становится меньше требуемой точности интерполяции, увеличение степени $N_k^{(I)}(q)$, $N_k^{(II)}(\hat{q})$ следует прекратить. Это замечание справедливо также и для $N_k(x)$.

Пример 4.4. Построить многочлен Ньютона третьей степени для сеточной функции, заданной таблицей 4.7. Вычислить значение функции в точке $x_* = 2,5$. Решить задачу интерполяции при включении одного дополнительного значения сеточной функции: $f(1) = 5$.

Таблица 4.7

i	0	1	2	3
x_i	2	3	4	5
$f(x_i) = f_i$	7	5	8	7

□Построим многочлен Ньютона (4.32), справедливый для произвольного расположения узлов. Для этого составим таблицу 4.8, аналогичную таблице 4.5:

$$f(x_0, x_1) = \frac{5-7}{3-2} = -2; \quad f(x_1, x_2) = \frac{8-5}{4-3} = 3; \quad f(x_2, x_3) = \frac{7-8}{5-4} = -1;$$

$$f(x_0, x_1, x_2) = \frac{3-(-2)}{4-2} = \frac{5}{2}; \quad f(x_1, x_2, x_3) = \frac{-1-3}{5-3} = -2; \quad f(x_0, x_1, x_2, x_3) = \frac{-2-\frac{5}{2}}{5-2} = -\frac{3}{2}.$$

Таблица 4.8

x_i	f_i	$f(x, x_{j+1})$	$f(x, x_{j+1}, x_{j+2})$	$f(x, x_{j+1}, x_{j+2}, x_{j+3})$
2	7	-2	$\frac{5}{2}$	$-\frac{3}{2}$
3	5	3	-2	$-\frac{3}{2}$
4	8	-1	-2	$-\frac{3}{2}$
5	7			

По формуле (4.32) для $n = 3$ имеем

$$\begin{aligned} N_3(x) &= f(x_0) + (x - x_0)f(x_0, x_1) + (x - x_0)(x - x_1)f(x_0, x_1, x_2) + \\ &+ (x - x_0)(x - x_1)(x - x_2)f(x_0, x_1, x_2, x_3) = 7 + (x-2)\cdot(-2) + (x-2)\cdot(x-3)\cdot\frac{5}{2} + \\ &+ (x-2)\cdot(x-3)\cdot(x-4)\cdot\left(-\frac{3}{2}\right) = -\frac{3}{2}x^3 + 16x^2 - \frac{107}{2}x + 62. \end{aligned}$$

Поскольку в данной задаче заданы равноточащие узлы, воспользуемся также формулой (4.34) для первого интерполяционного многочлена Ньютона:

$$N_3^{(I)}(q) = f(x_0) + \frac{\Delta f_0}{1!} q + \frac{\Delta^2 f_0}{2!} q \cdot (q-1) + \frac{\Delta^3 f_0}{3!} q \cdot (q-1) \cdot (q-2),$$

где $q = \frac{x - x_0}{h} = \frac{x - 2}{1} = x - 2$. Составим таблицу 4.9, аналогичную таблице 4.6.

Таблица 4.9

x_i	$f_i = f(x_i)$	Δf_i	$\Delta^2 f_i$	$\Delta^3 f_i$
2	7	-2		
3	5	2	5	-9
4	8	-1	-4	
5	7			

Имеем: $\Delta f_0 = 5 - 7 = -2$; $\Delta f_1 = 8 - 5 = 3$; $\Delta f_2 = 7 - 8 = -1$; $\Delta^2 f_0 = 3 - (-2) = 5$; $\Delta^2 f_1 = -1 - 3 = -4$; $\Delta^3 f_0 = -4 - 5 = -9$. Поэтому

$$\begin{aligned} N_3^{(I)}(x) &= 7 + \frac{(-2)}{1!} q + \frac{5}{2!} q \cdot (q-1) + \frac{(-9)}{3!} q \cdot (q-1) \cdot (q-2) = -\frac{3}{2} q^3 + 7q^2 - \frac{15}{2} q + 7 \Big|_{q=\frac{x-2}{1}} = \\ &= -\frac{3}{2} \cdot (x^3 - 6x^2 + 12x - 8) + 7 \cdot (x^2 - 4x + 4) - \frac{15}{2} \cdot (x - 2) + 7 = -\frac{3}{2} x^3 + 16x^2 - \frac{107}{2} x + 62. \end{aligned}$$

Запишем также второй интерполяционный многочлен Ньютона (4.36):

$$\begin{aligned} N_3^{(II)}(x) &= f_3 + \frac{\Delta f_2}{1!} \hat{q} + \frac{\Delta^2 f_1}{2!} \hat{q} \cdot (\hat{q}+1) + \frac{\Delta^3 f_0}{3!} \hat{q} \cdot (\hat{q}+1) \cdot (\hat{q}+2) \Big|_{\hat{q}=\frac{x-x_3}{1}} = \\ &= 7 + \frac{(-1)}{1!} \hat{q} + \frac{-4}{2!} \hat{q} \cdot (\hat{q}+1) + \frac{(-9)}{3!} \hat{q} \cdot (\hat{q}+1) \cdot (\hat{q}+2) = -\frac{3}{2} \hat{q}^3 - \frac{13}{2} \hat{q}^2 - 6\hat{q} + 7 \Big|_{\hat{q}=x-5} = \\ &= -\frac{3}{2} x^3 + 16x^2 - \frac{107}{2} x + 62. \end{aligned}$$

Сравнивая с результатом примера 4.1, можно заключить, что $L_3(x) = N_3^{(I)}(x) = N_3^{(II)}(x)$. Это еще раз подтверждает единственность решения задачи интерполяции в классе многочленов, удовлетворяющих условиям теоремы 4.1.

Предположим, что в ходе некоторого эксперимента получен новый результат $f(1) = 5$, дополняющий заданную сеточную функцию. Тогда для ре-

Таблица 4.10

x_i	y_i	$f(x_j, x_{j+1})$	$f(x_j, x_{j+1}, x_{j+2})$	$f(x_j, x_{j+1}, x_{j+2}, x_{j+3})$	$f(x_0, x_1, x_2, x_3, x_4)$
2	7	-2			
3	5	2	5		
4	8	-1	-2	3	
5	7	1	-1/2	3/4	-3/4
1	5	1/2			

шения задачи интерполяции с помощью многочлена Ньютона можно использовать уже полученный результат. Для этого дополним таблицу 4.8. Новый узел и соответствующее значение функции поместим в конце таблицы 4.10.

По формуле (4.34) имеем

$$\begin{aligned}
 N_4(x) &= N_3(x) + (x - x_0)(x - x_1)(x - x_2)(x - x_3)f(x_0, x_1, x_2, x_3, x_4) = \\
 &= -\frac{3}{2}x^3 + 16x^2 - \frac{107}{2}x + 62 + (x - 2)(x - 3)(x - 4)(x - 5)(-\frac{3}{4}) = \\
 &= -\frac{3}{2}x^3 + 16x^2 - \frac{107}{2}x + 62 - \frac{3}{4} \cdot (x^4 - 14x^3 + 71x^2 - 154x + 120) = \\
 &= -\frac{3}{4}x^4 + 9x^3 - \frac{149}{4}x^2 + 62x - 28.
 \end{aligned}$$

Легко проверить, что новый узел можно добавить и в начале таблицы, т. е. будет найден тот же многочлен $N_4(x)$. На рисунке 4.6 изображены полученные интерполяционные многочлены.

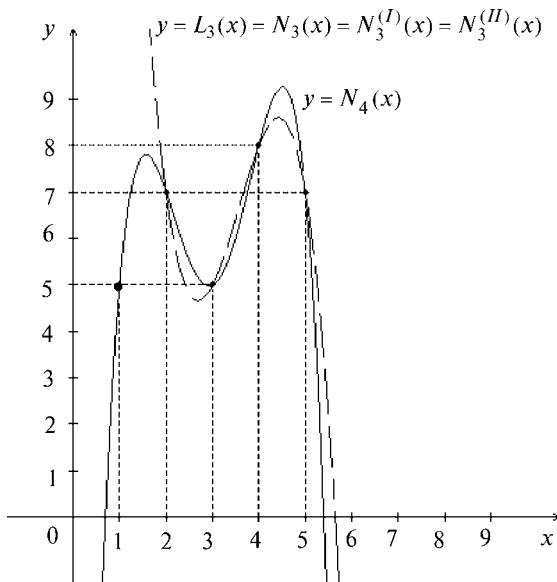


Рис. 4.6

Вычислим значение функции в точке $x_* = 2,5$:

$$N_3^{(I)}(x_*) = N_3^{(III)}(x_*) = 4,8125; \quad N_4^{(I)}(x_*) = 5,5156. \blacksquare$$

Пример 4.5. Для сеточной функции из примера 4.3 найти линейный и параболический многочлены Ньютона и на их основе подсчитать значение функции в точке $x_* = 2,5$. Оценить погрешность интерполяции.

□ Так как точка $x_* = 2,5$ находится вблизи узла $x_i = x_2$ и шаблон (x_2, x_3, x_4) имеет неравномерность по шагу, то для выполнения интерполяции выбираем многочлен Ньютона (4.29).

i	x_i	f_i	$f(x_i, x_{i+1})$	$f(x_i, x_{i+1}, x_{i+2})$
2	1	1	13	
3	3	27	37	-8
4	4	64		

1. Определим разделенные разности $f(x_i, x_{i+1})$, $f(x_i, x_{i+1}, x_{i+2})$ (табл. 4.11).

$$\text{Здесь } f(x_2, x_3) = \frac{27 - 1}{3 - 1} = 13, \quad f(x_3, x_4) = \frac{64 - 27}{4 - 3} = 37, \quad f(x_2, x_3, x_4) = \frac{37 - 13}{4 - 1} = 8.$$

Определим $N_1(x)$, $N_2(x)$:

$$\begin{aligned} N_1(x) &= f_2 + f(x_2, x_3) \cdot (x - x_2) = 1 + 13(x - 1) = 13x - 12; \\ N_2(x) &= f_2 + f(x_2, x_3) \cdot (x - x_2) + f(x_2, x_3, x_4)(x - x_2)(x - x_3) = \\ &= 13x - 12 + 8(x - 1)(x - 3) = 8x^2 - 19x + 12. \end{aligned}$$

Сравнивая $N_2(x)$ с соответствующим многочленом Лагранжа $L_2(x)$ (см. пример 4.3), видим, что они совпадают. Это подтверждает единственность решения задачи о построении интерполяционного многочлена.

2. Вычислим значения $N_1(x)|_{x=x_*}$, $N_2(x)|_{x=x_*}$:

$$\begin{aligned} N_1(2,5) &= 13 \cdot 2,5 - 12 = 20; \\ N_2(2,5) &= 8 \cdot 2,5^2 - 19 \cdot 2,5 + 12 = 14,5. \end{aligned}$$

Итак, линейная интерполяция приводит к отлинию от точного значения на $\left| \frac{20 - 15,625}{15,625} \right| \cdot 100\% = 28\%$, а квадратичная — на $\left| \frac{14,5 - 15,625}{15,625} \right| \cdot 100\% = 7\%$.

Таким образом, найдена фактическая погрешность, которая для квадратичной интерполяции получилась в 4 раза меньше.

3. Найдем априорную оценку погрешности линейной интерполяции на отрезке $[x_2, x_3]$ по формуле (4.23). Для этого необходимо сначала оценить производную $M_{2,i}$. При этом, как и в реальных задачах, будем определять $f''(x)$ численным дифференцированием. С этой целью используем многочлен $N_2(x)$: $N_2'(x) = 16x - 19$, $N_2''(x) = 16$. Приближенно можно принять, что $M_{2,i} = 16$. Тогда получаем

$$|R_1(x)| \leq \frac{2^2}{8} \cdot 16 = 8 \text{ или } \frac{8}{15,625} \cdot 100\% = 51,2\%.$$

Заметим, что фактическая погрешность получилась примерно в два раза меньше априорной. ■

Пример 4.6. Для сеточной функции, заданной в примере 4.4, построить интерполяционные многочлены первой и второй степени для нахождения значений в точках $x_{*1} = 2,5$; $x_{*2} = 4,5$; $x_{*3} = 3,5$; $x_{*4} = 5,5$.

Для подсчета значения функции в точке $x_{*1} = 2,5$ (согласно рис. 4.5a) выбираем шаблон $(x_0, x_1, x_2) = (2; 3; 4)$ для параболической интерполяции

$(i = 0, k = 2)$ и шаблон $(x_0, x_1) = (2; 3)$ для линейной интерполяции ($i = 0, k = 1$). Тогда по формуле (4.34)

$$N_1^{(I)}(q) = f_0 + \frac{\Delta f_0}{1!} q = 7 + \frac{-2}{1!} q = 7 - 2q;$$

$$N_2^{(I)}(q) = N_1^{(I)}(q) + \frac{\Delta^2 f_0}{2!} q(q-1) = 7 - 2q + \frac{5}{2} q(q-1) = \frac{5}{2} q^2 - \frac{9}{2} q + 7,$$

где $\bar{a} = \{a_0, a_1, \dots, a_m\}^T$ (см. табл. 4.9). В результате подстановки получаем

$$N_1^{(I)}(x) = -2x + 11; \quad N_2^{(I)}(x) = \frac{5}{2}x^2 - \frac{29}{2}x + 26.$$

В заданной точке

$$N_1^{(I)}(2,5) = 6; \quad N_2^{(I)}(2,5) = 5,375.$$

Для подсчета значения в точке $x_{*2} = 4,5$ выбираем шаблон $(3; 4; 5) = (x_1, x_2, x_3)$ для параболической интерполяции ($i = 1, k = 2$) и шаблон $(4; 5) = (x_2, x_3)$ для линейной интерполяции ($i = 2, k = 1$). Тогда по формуле (4.35)

$$N_1^{(II)}(\hat{q}) = f_3 + \frac{\Delta f_2}{1!} \hat{q} = 7 + \frac{-1}{1!} \hat{q} = 7 - \hat{q};$$

$$N_2^{(II)}(\hat{q}) = N_1^{(II)}(\hat{q}) + \frac{\Delta^2 f_1}{2!} \hat{q} \cdot (\hat{q}+1) = 7 - \hat{q} + \frac{-4}{2} \hat{q}(\hat{q}+1) = -2\hat{q}^2 - 3\hat{q} + 7,$$

где $\hat{q} = \frac{x - x_3}{h} = x - 5$. В результате подстановки получаем

$$N_1^{(II)}(x) = -x + 12; \quad N_2^{(II)}(x) = -2x^2 + 17x - 28.$$

В заданной точке

$$N_1^{(II)}(4,5) = 7,5; \quad N_2^{(II)}(4,5) = 8.$$

Многочлен $N_2^{(II)}(x)$ можно использовать для экстраполяции, т. е. для подсчета функции в точке $x_{*4} = 5,5$: $N_2^{(II)}(5,5) = 5$.

Чтобы подсчитать значения в точке $x_{*3} = 3,5$, сначала выберем шаблон $(3; 4; 5) = (x_1, x_2, x_3)$ для параболической интерполяции ($i = 1, k = 2$) и шаблон $(3; 4) = (x_1, x_2)$ для линейной интерполяции ($i = 1, k = 1$). Согласно рисунку 4.5б применим формулу (4.33):

$$N_1^{(I)}(q) = f_1 + \frac{\Delta f_1}{1!} q = 5 + \frac{3}{1} q = 5 + 3q;$$

$$N_2^{(I)}(q) = N_1^{(I)}(q) + \frac{\Delta^2 f_1}{2!} q(q-1) = 5 + 3q + \frac{-4}{2} q(q-1) = -2q^2 + 5q + 5,$$

где $q = \frac{x - x_1}{h} = x - 3$. После подстановки получаем

$$N_1^{(I)}(x) = 3x - 4; \quad N_2^{(I)}(x) = -2x^2 + 17x - 28 \text{ и } N_1^{(I)}(3,5) = 6,5; \quad N_2^{(I)}(3,5) = 7.$$

Подсчитать значения в точке $x_{3*} = 3,5$ (согласно рис. 4.5б) можно, использовав шаблон $(2; 3; 4) = (x_0, x_1, x_2)$ для параболической интерполяции

$(i = 0, k = 2)$ и шаблон $(3; 4) = (x_1, x_2)$ для линейной интерполяции $(i = 1, k = 1)$. По формуле (4.35) получаем

$$N_1^{(II)}(\hat{q}) = f_2 + \frac{\Delta f_1}{1!} \hat{q} = 8 + \frac{3}{1} \hat{q};$$

$$N_2^{(II)}(\hat{q}) = N_1^{(II)}(\hat{q}) + \frac{\Delta^2 f_0}{2!} \hat{q} = 8 + \frac{3}{1} \hat{q} \cdot (\hat{q} + 1) = \frac{5}{2} \hat{q}^2 + \frac{11}{2} \hat{q} + 8,$$

где $\hat{q} = \frac{x - x_2}{h} = x - 4$. После подстановки

$$N_1^{(II)}(x) = 3x - 4; \quad N_2^{(II)}(x) = \frac{5}{2}x^2 - \frac{29}{2}x + 26 \text{ и } N_1^{(II)}(3,5) = 6,5; \quad N_2^{(II)}(3,5) = 5,875. \blacksquare$$

4.3. МЕТОДЫ ИНТЕГРАЛЬНО-ДИФФЕРЕНЦИАЛЬНОЙ ИНТЕРПОЛЯЦИИ

4.3.1. ПОСТАНОВКА ЗАДАЧИ

Использование различных сочетаний функциональных, дифференциальных и интегральных условий согласования (4.2), (4.3), (4.5) позволяет конструировать различные типы интерполяционных или интерполяционно-сглаживающих многочленов. В данном разделе рассматриваются задачи построения двух интегрально-дифференциальных интерполяционных параболических многочленов $S_2^{(I)}(x)$ и $S_2^{(II)}(x)$.

Первый определяется по сеточной функции, заданной не только своими значениями f_i ($i = 0, n - 1$), но и значениями интегралов I_i^{i+1} ($i = 0, n - 1$). Так как для нахождения многочлена $S_2^{(I)}(x)$ используются совместно функциональное условие (4.3) и интегральное условие (4.5), то многочлен называется *интегрально-функциональным*. Поскольку функциональное условие (4.3) является частным случаем дифференциального условия (4.2), соответствующий метод относится к *интегрально-дифференциальному*.

Второй многочлен определяется по сеточной функции, заданной не только значениями производных $f_i^{(p)}$ ($i = 0, n$), но и значениями интегралов I_i^{i+1} ($i = 0, n - 1$). Так как для нахождения многочлена $S_2^{(II)}(x)$ совместно используются дифференциальное условие (4.2) при $p = 1$ и интегральное условие (4.5), то многочлен (и соответствующий метод) называется *интегрально-дифференциальным*.

4.3.2. ИНТЕРПОЛЯЦИОННЫЙ ПАРАБОЛИЧЕСКИЙ ИНТЕГРАЛЬНО-ФУНКЦИОНАЛЬНЫЙ МНОГОЧЛЕН

Пусть некоторая сеточная функция, определенная в общем случае на неравномерной сетке $\Omega_n = \{x_0, x_1, \dots, x_n\}$, задана своими значениями f_0, f_1, \dots, f_n и интегралами $I_0^1, I_1^2, \dots, I_{n-1}^n$ ($I_i^{i+1} = \int_{x_i}^{x_{i+1}} f(x) dx$), которые могут быть предварительно рассчитаны по одной из квадратурных формул (см. главу 5). Будем считать, что точность заданных значений f_i и вычисленных интегралов

I_i^{i+1} для всех узлов x_i ($i = \overline{0, n}$) и отрезков $[x_i, x_{i+1}]$ ($i = \overline{0, n-1}$) не ниже $O(h_{i+1}^3)$ и $O(h_{i+1}^4)$ соответственно. Это требование следует из утверждения В.2, устанавливающего принцип соответствия порядков аппроксимации (интерполяции).

Применим *кусочный способ* интерполяции. Для некоторого отрезка $[x_i, x_{i+1}]$ ($0 \leq i \leq n-1$) на двухточечном шаблоне требуется определить многочлен

$$S_{2,i}^{(I)}(x) = a_{0,i} + a_{1,i}(x - x_i) + a_{2,i}(x - x_i)^2,$$

где $a_{0,i}, a_{1,i}, a_{2,i}$ — неизвестные коэффициенты, удовлетворяющие функциональным (4.3) и интегральным условиям согласования (4.5):

$$\begin{aligned} \delta S_{2,i}^{(I)}(x_i) &= S_{2,i}^{(I)}(x_i) - f_i = 0, \quad \delta S_{2,i}^{(I)}(x_{i+1}) = S_{2,i}^{(I)}(x_{i+1}) - f_{i+1} = 0, \\ \delta S_{2,i}^{(I)}(I_i^{i+1}) &= \int_{x_i}^{x_{i+1}} S_{2,i}^{(I)}(x) dx - I_i^{i+1} = 0. \end{aligned}$$

Для определения трех коэффициентов параболического интегрально-функционального многочлена можно получить три линейных уравнения, являющиеся следствием трех условий согласования:

$$\begin{aligned} S_{2,i}^{(I)}(x_i) &= a_{0,i} = f_i \quad \text{или} \quad a_{0,i} = f_i, \\ S_{2,i}^{(I)}(x_{i+1}) &= f_i + a_{1,i}h_{i+1} + a_{2,i}h_{i+1}^2 = f_{i+1} \quad \text{или} \quad h_{i+1}a_{1,i} + h_{i+1}^2a_{2,i} = \Delta f_i, \\ \int_{x_i}^{x_{i+1}} [f_i + a_{1,i}(x - x_i) + a_{2,i}(x - x_i)^2] dx &= I_i^{i+1}. \end{aligned}$$

Вводя фазу интерполяции $u = \frac{x - x_i}{h_{i+1}}$, последнее условие после интегрирования можно записать в виде $h_{i+1}f_i + \frac{h_{i+1}^2}{2}a_{1,i} + \frac{h_{i+1}^3}{3}a_{2,i} = I_i^{i+1}$.

Итак, для оставшихся двух коэффициентов $a_{1,i}, a_{2,i}$ имеется система линейных алгебраических уравнений:

$$\begin{aligned} 3h_{i+1}^2a_{1,i} + 2h_{i+1}^3a_{2,i} &= 6\nabla I_i^{i+1}, \\ h_{i+1}a_{1,i} + h_{i+1}^2a_{2,i} &= \Delta f_i, \end{aligned}$$

где $\nabla I_i^{i+1} = I_i^{i+1} - h_{i+1}f_i$, $\Delta f_i = f_{i+1} - f_i$. Очевидно, определитель этой системы

$$\begin{vmatrix} 3h_{i+1}^2 & 2h_{i+1}^3 \\ h_{i+1} & h_{i+1}^2 \end{vmatrix} \neq 0,$$

что свидетельствует о существовании и единственности интегрально-функционального многочлена $S_{2,i}^{(I)}(x)$ на отрезке $[x_i, x_{i+1}]$. В силу произвольности отрезка $[x_i, x_{i+1}]$ данный многочлен $S_{2,i}^{(I)}(x)$ существует и является единственным на каждом частичном отрезке $[x_i, x_{i+1}]$, $i = \overline{0, n-1}$. Поэтому многочлен $S_{2,i}^{(I)}(x)$ может быть положен в основу кусочной многозвенной интерполяции, в том числе и сплайн-интерполяции (см. п. 4.5).

Определяя $a_{1,i}$, $a_{2,i}$ из приведенной системы, записываем параболический интегрально-функциональный многочлен $S_{2,i}^{(I)}(x)$ в полиномиальной форме:

$$S_{2,i}^{(I)}(x) = f_i + \left(\frac{6}{h_{i+1}^2} \nabla I_i^{i+1} - \frac{2}{h_{i+1}} \Delta f_i \right) (x - x_i) + \left(-\frac{6}{h_{i+1}^3} \nabla I_i^{i+1} + \frac{3}{h_{i+1}^2} \Delta f_i \right) (x - x_i)^2. \quad (4.37)$$

Если в (4.37) подставить вместо $(x - x_i)$ произведение $i h_{i+1}$ и выделить слагаемые при I_i^{i+1} , f_i , f_{i+1} , то получим параболический интегрально-функциональный многочлен в лагранжевой форме:

$$S_{2,i}^{(I)}(u) = \frac{1}{h_{i+1}} P_{2,I}(u) I_i^{i+1} + P_{2,i}(u) f_i + P_{2,i+1}(u) f_{i+1}, \quad (4.38)$$

где $P_{2,I}(u) = 6u(1-u)$; $P_{2,i}(u) = (1-u)(1-3u)$; $P_{2,i+1}(u) = u(3u-2)$ — коэффициенты Лагранжа, удовлетворяющие функциональным и интегральному условиям:

$$\begin{aligned} P_{2,I}(0) &= 0, \quad P_{2,I}(1) = 0, \quad \int_0^1 P_{2,I}(u) du = 1, \\ P_{2,i}(0) &= 1, \quad P_{2,i}(1) = 0, \quad \int_0^1 P_{2,i}(u) du = 0, \\ P_{2,i+1}(0) &= 0, \quad P_{2,i+1}(1) = 1, \quad \int_0^1 P_{2,i+1}(u) du = 0. \end{aligned} \quad (4.39)$$

Так же как и классические (функциональные) многочлены Лагранжа, интегрально-функциональный многочлен $S_{2,i}^{(I)}(x)$ линейно зависит от параметров I_i^{i+1} , f_i , f_{i+1} . При использовании кусочной интерполяции для определения значения функции $f(x_*)$ в точке x_* с помощью многочлена $S_{2,i}^{(I)}(x)$ выполняются действия, аналогичные применяемым при функциональной интерполяции.

Методика решения задачи интегрально-функциональной интерpolationи

1. Выбрать «окно» интерполяции $[x_i, x_{i+1}]$ так, чтобы $x_i \leq x_* \leq x_{i+1}$.
2. Определить фазу интерполяции $u = \frac{(x_* - x_i)}{h_{i+1}}$ и коэффициенты $P_{2,k}(u)$ ($k = i, i+1, I$).
3. По формуле (4.38) вычислить значение $f(x_*) \approx S_{2,i}^{(I)}(u)$.

Замечания

1. Правильность вычисления коэффициентов Лагранжа проверяется по условию $P_{2,I} + P_{2,i} + P_{2,i+1} = 1$, получающемуся путем подстановки в (4.38) функции $f(x) \equiv 1$.

2. При аппроксимации значения $f(x_*)$ значением многочлена $S_{2,i}^{(I)}(x_*)$ возникает погрешность $R_2^{(I)}(x_*)$: $f(x_*) = S_{2,i}^{(I)}(x_*) + R_2^{(I)}(x_*)$. При условии, что $f(x) \in C_3[a, b]$, можно получить оценку погрешности на отрезке $[x_i, x_{i+1}]$ ([18]):

$$\max_{[x_i, x_{i+1}]} |f(x) - S_{2,i}^{(I)}(x)| \leq \frac{h_{i+1}^3}{72\sqrt{3}} \tilde{M}_{3,i}, \quad (4.40)$$

где $\tilde{M}_{3,i} = \max_{[x_i, x_{i+1}]} |f'''(x)|$.

Если класс гладкости функции $f(x)$ понижен и $f(x) \in C_2[a, b]$, то порядок аппроксимации интегрально-функциональным многочленом $S_{2,i}^{(I)}(x)$ также понижается на единицу:

$$\max_{[x_i, x_{i+1}]} |f(x) - S_{2,i}^{(I)}(x)| \leq \frac{h_{i+1}^2}{25\sqrt{2}} M_{2,i}.$$

3. Сравним условия применения функционального (классического) многочлена Лагранжа $L_2(x)$, определенного на трехточечном шаблоне (x_i, x_{i+1}, x_{i+2}) , и интегрально-функционального многочлена $S_{2,i}^{(I)}(x)$:

а) аппроксимируемая функция для применения многочлена Лагранжа задается только своими значениями f_i ($i = 0, n$), т. е. функциональными параметрами, а для использования $S_2^{(I)}(x)$ функция определяется еще и интегралами I_i^{i+1} ($i = 0, n-1$). Если в постановке задачи интегралы неизвестны, то, как отмечено выше, они вычисляются предварительно по значениям f_i ($i = 0, n$) по тем или иным квадратурным формулам (см. главу 5). Интегрально-функциональная интерполяция предоставляет дополнительные возможности для более полного учета локальных свойств интерполируемых функций, которые в отдельных точках и областях могут иметь особенности: разрывы производных, локальные экстремумы и др. При этом локальные свойства учитываются путем соответствующего выбора трехточечного шаблона, на котором аппроксимируется интеграл I_i^{i+1} ;

б) многочлен $L_2(x)$ записывается на трехточечном шаблоне (x_i, x_{i+1}, x_{i+2}) , а интегрально-функциональный многочлен $S_{2,i}^{(I)}(x)$ — на двухточечном (x_i, x_{i+1}) ;

в) функциональный многочлен $L_2(x)$ при его записи на одном шаблоне определяется только функциональными параметрами f_i, f_{i+1}, f_{i+2} , а интегрально-функциональный многочлен $S_{2,i}^{(I)}(x)$ — двумя функциональными и интегральными I_i^{i+1} , поэтому при его построении можно учесть интегральные свойства аппроксимируемой функции;

г) по порядку аппроксимации оба многочлена обеспечивают одинаковый третий порядок аппроксимации, который при $f(x) \in C_3[a, b]$ является максимальным, однако константа в мажоранте для $S_{2,i}^{(I)}(x)$ в восемь раз меньше соответствующей константы в мажоранте для $L_2(x)$:

$$\max_{[x_i, x_{i+1}]} |f(x) - L_2(x)| \leq \frac{H^3\sqrt{3}}{27} M_{3,i}; \quad \max_{[x_i, x_{i+1}]} |f(x) - S_{2,i}^{(I)}(x)| \leq \frac{h_{i+1}^3\sqrt{3}}{216} \tilde{M}_{3,i},$$

где $H = \max_{k=i+1, i+2} h_k$, $M_{3,i} = \max_{[x_i, x_{i+2}]} |f'''(x)|$, $\tilde{M}_{3,i} = \max_{[x_i, x_{i+1}]} |f'''(x)|$.

Пример 4.7. Пусть сеточная функция на отрезке $[-1; 4]$ задана функциональными f_i ($i = 0, 4$) и интегральными параметрами I_i^{i+1} , $i = 0, 3$, приведенными в таблице 4.12. Данные параметры получены точно путем «обработки» сеточного представления функции $f(x) = x^3$. Интегралы вычислены точно по

i	0	1	2	3	4
x_i	-1	0	1	3	4
f_i	-1	0	1	27	64
I_i^{l+1}	$-\frac{1}{4}$	$\frac{1}{4}$	20	$\frac{175}{4}$	

формуле Ньютона — Лейбница $I_i^{l+1} = F_{i+1} - F_i$, где F_i — первообразная. Значения интегралов в таблице 4.12 помечены посередине между значениями функции.

Требуется найти значение $f(x_*)$ при $x_* = 2$ на основе интегрально-функционального многочлена $S_{2,i}^{(I)}(x)$.

□ 1. По значению $x_* = 2$ выбираем «окно» интерполяции $[x_2, x_3] = [1; 3]$, такое, что $x_2 < x_* < x_3$.

2. Вычислим

$$u = \frac{x_* - x_2}{h_3} = \frac{2 - 1}{2} = \frac{1}{2} \quad (h_3 = x_3 - x_2 = 2)$$

и коэффициенты Лагранжа

$$P_{2,1} = 6u \cdot (1-u) = \frac{3}{2}; \quad P_{2,2} = (1-u) \cdot (1-3u) = -\frac{1}{4}; \quad P_{2,3} = u \cdot (3u-2) = -\frac{1}{4}.$$

Так как $\sum_k P_{2,k} = 1$, коэффициенты вычислены правильно.

3. Найдем значение $f(2)$ по формуле (4.38) при $i = 2$:

$$f(2) \approx S_{2,2}^{(I)}(2) = \frac{P_{2,1}}{h_3} \cdot I_2^3 + P_{2,2} \cdot f_2 + P_{2,3} \cdot f_3 = \frac{3}{2} \cdot \frac{20}{2} - \frac{1}{4} \cdot 1 - \frac{1}{4} \cdot 27 = 8.$$

Получилось значение $S_{2,2}^{(I)}(2) = 8$, совпадающее с точным значением исходной формульной функции. ■

Пример 4.8. Данна сеточная функция, характеризующаяся функциональными и интегральными параметрами, заданными в таблице 4.12. Найти полиномиальный вид интегрально-функционального многочлена на отрезке $[x_2, x_3]$, внутри которого находится точка $x_* = 2$, и вычислить $S_2^{(I)}(2)$. Результаты сравнить с решением, полученным с помощью многочлена Лагранжа в примере 4.3.

□ Для решения задачи в соответствии с исходными данными используем интегрально-функциональный многочлен $S_2^{(I)}(x)$, записанный в форме (4.37).

1. В поставленной задаче «окно» интерполяции, как в примере 4.7, задано отрезком $[x_2, x_3] = [1; 3]$, поэтому $i = 2$.

2. Вычислим значения ∇I_2^3 , Δf_2 , h_3 и коэффициенты многочлена $S_{2,2}^{(I)}(x)$:

$$h_3 = x_3 - x_2 = 3 - 1 = 2; \quad \Delta f_2 = f_3 - f_2 = 27 - 1 = 26;$$

$$\nabla I_2^3 = I_2^3 - h_3 f_2 = 20 - 2 \cdot 1 = 18;$$

$$a_0 = f_2 = 1; \quad a_1 = \frac{6}{h_3^2} \nabla I_2^3 - \frac{2}{h_3} \Delta f_2 = 1; \quad a_2 = \left(-\frac{6}{h_3^3} \nabla I_2^3 + \frac{3}{h_3^2} \Delta f_2 \right) = 6.$$

Полученные коэффициенты подставим в формулу (4.37):

$$S_{2,2}^{(I)}(x) = 1 + (x - 1) + 6(x - 1)^2 = 6x^2 - 11x + 6.$$

3. Определим значение $S_{2,2}^{(I)}(x_*)$:

$$S_{2,2}^{(I)}(2) = (6x^2 - 11x + 6)|_{x=2} = 8.$$

Из сопоставления полученных результатов с результатами решения примера 4.3 следует, что значение $S_{2,2}^{(I)}(2) = 8$, полученное интегрально-функциональным методом, лучше соответствует точному значению, чем $L_2(2) = 6$. Это обусловлено тем, что кроме значений функции $f_i = f(x_i)$ в узлах задан еще и интеграл, значение которого является точным. В данном случае значение $S_{2,2}^{(I)}(2)$ совпало с точным значением, так как $x_* = 2$ находится в середине отрезка $[1; 3]$. ■

4.3.3. ИНТЕРПОЛЯЦИОННЫЙ ПАРАБОЛИЧЕСКИЙ ИНТЕГРАЛЬНО-ДИФФЕРЕНЦИАЛЬНЫЙ МНОГОЧЛЕН

Пусть некоторая сеточная функция на сетке Ω_n задана значениями производных $f'_i = f'(x_i)$ ($i = 0, n$) и интегралами I_i^{i+1} ($i = 0, n$), которые либо известны, либо могут быть предварительно рассчитаны по одной из квадратурных формул. Так же, как и для многочлена $S_{2,i}^{(I)}(x)$, рассмотренного в п. 4.3.2, будем считать, что f'_i и I_i^{i+1} заданы или вычислены с точностью не ниже $O(h_{i+1}^2)$ и $O(h_{i+1}^4)$ соответственно (см. утверждение В.2).

Применим *кусочный способ* интерполяции. Для некоторого отрезка $[x_i, x_{i+1}]$ ($0 \leq i \leq n - 1$) на двухточечном шаблоне требуется определить многочлен

$$S_{2,i}^{(\Pi)}(x) = a_{0,i} + a_{1,i}(x - x_i) + a_{2,i}(x - x_i)^2,$$

где $a_{0,i}$, $a_{1,i}$, $a_{2,i}$ — неизвестные коэффициенты, удовлетворяющие дифференциальным ((4.2) при $p = 1$) и интегральным (4.5) условиям согласования:

$$\delta S_{2,i}^{(\Pi)}(x_i) = S_{2,i}^{(\Pi)}(x_i) - f'_i = 0, \quad \delta S_{2,i}^{(\Pi)}(x_{i+1}) = S_{2,i}^{(\Pi)}(x_{i+1}) - f'_{i+1} = 0,$$

$$\delta S_{2,i}^{(\Pi)}(I_i^{i+1}) = \int_{x_i}^{x_{i+1}} S_{2,i}^{(\Pi)}(x) dx - I_i^{i+1} = 0.$$

Для нахождения трех коэффициентов многочлена можно получить три линейных уравнения, являющиеся следствием трех условий согласования, так же, как это сделано в п. 4.3.2. В результате имеем выражение для *парabolического интегрально-дифференциального многочлена* $S_{2,i}^{(\Pi)}(x)$ в полиномиальной форме:

$$S_{2,i}^{(\Pi)}(x) = \left(\frac{1}{h_{i+1}} I_i^{i+1} - \frac{h_{i+1}}{2} f'_i - \frac{h_{i+1}}{6} \Delta f'_i \right) + f'_i(x - x_i) + \frac{\Delta f'_i}{2h_{i+1}} (x - x_i)^2, \quad (4.41)$$

где $\Delta f'_i = f'_{i+1} - f'_i$.

Если подставить вместо $(x - x_i)$ произведение uh_{i+1} (u — фаза интерполяции) и выделить слагаемые при I_i^{i+1} , f'_i , f'_{i+1} , то получим *парabolический интегрально-дифференциальный многочлен в лагранжевой форме*:

$$S_{2,i}^{(\Pi)}(u) = \frac{1}{h_{i+1}} I_i^{i+1} + \frac{1}{6} (-3u^2 + 6u - 2) h_{i+1} f'_i + \frac{1}{6} (3u^2 - 1) h_{i+1} f'_{i+1}, \quad (4.42)$$

где $u = \frac{x - x_i}{h_{i+1}}$, $0 \leq u \leq 1$.

Замечания

1. При аппроксимации значения $f(x_*)$ значением многочлена $S_{2,i}^{(\Pi)}(x_*)$ возникает погрешность $R_2^{(\Pi)}(x_*)$:

$$f(x_*) = S_{2,i}^{(\Pi)}(x_*) + R_2^{(\Pi)}(x_*).$$

При условии, что $f(x) \in C_3[a, b]$, можно получить оценку погрешности на отрезке $[x_i, x_{i+1}]$ ([21]):

$$\max_{[x_i, x_{i+1}]} |f(x) - S_{2,i}^{(\Pi)}(x)| \leq \frac{h_{i+1}^3}{24} M_{3,i}.$$

2. Алгоритм аппроксимации сеточных функций на основе (4.41), (4.42) аналогичен изложенному в п. 4.3.2 для аппроксимации многочленом $S_{2,i}^{(\Pi)}(x)$.

4.4. МЕТОДЫ ИНТЕГРАЛЬНОГО СГЛАЖИВАНИЯ

4.4.1. ПОСТАНОВКА ЗАДАЧИ

Определение вида функциональных зависимостей, получаемых в физическом эксперименте, имеет очень важное значение. Обычно экспериментальные результаты представляются в виде таблиц или сеточных функций $y_i = f(x_i) \pm \varepsilon_i$, где ε_i — погрешности измерений. Если их построить на графике, соединяя экспериментальные точки отрезками прямых, то получится ломаная, не имеющая ничего общего с той или иной функциональной зависимостью, которая затем могла бы быть исследована средствами математического анализа. Форма этой ломаной из-за ошибок измерений при повторном эксперименте не воспроизводится. Поэтому экспериментатор на основе практического опыта предполагает, что полученная таблица является реализацией эмпирической гладкой зависимости $\tilde{f}_m(x, \bar{a})$ с неизвестным параметром \bar{a} , являющимся в общем случае вектором размерности $m + 1$ (см. рис. 4.1б). Для его определения используются различные формы записи зависимости $\tilde{f}_m(x, \bar{a})$ и различные условия согласования $\tilde{f}_m(x, \bar{a})$ с исходной функцией $y_i = f(x_i) \pm \varepsilon_i$.

На практике удобно представить искомую зависимость в виде *обобщенного многочлена*

$$\tilde{f}_m(x, \bar{a}) = \sum_{j=0}^m a_j \varphi_j(x) = a_0 \varphi_0(x) + a_1 \varphi_1(x) + \dots + a_m \varphi_m(x), \quad (4.43)$$

где $\bar{a} = \{a_0, a_1, \dots, a_m\}^T$ — вектор неизвестных коэффициентов, $\{\varphi_j\} = \{\varphi_0, \varphi_1, \dots, \varphi_m\}$ — заданная система *базисных функций*. В качестве базисных функций могут выбираться, например, степенные функции $\{\varphi_j\} = \{x^j\}$, многочлены Чебышева, тригонометрические функции $\{\varphi_j\} = \{\cos jx\}$. В первых двух случаях

ищется алгебраический полином, а в третьем — тригонометрический полином. Далее для решения задачи используются базисные функции только первых двух видов.

В зависимости от используемого условия согласования можно построить два основных альтернативных метода сглаживания — *метод наименьших квадратов*, являющийся традиционным, и *метод наилучшего интегрального приближения* или просто *метод интегрального приближения*, предложенный в работах [15–17], [19–23]. Соответствующий многочлен называется *многочленом интегрального приближения*.

Метод наименьших квадратов, являющийся методом наилучшего среднеквадратичного приближения, основан на применении интегрального условия (4.4), а метод интегрального приближения — на использовании интегрального условия (4.5).

Оба метода могут быть реализованы глобальным способом, т. е. для всего отрезка $[a, b]$ может быть получена одна функция $\tilde{f}_m(x, \bar{a})$. Как правило, принимают $m \leq 5-8$ ($m \ll n$) в связи с отмеченными в п. 4.2.2 свойствами многочленов высокой степени. Последнее обстоятельство обуславливает необходимость использования для сглаживания кусочно-глобального подхода, аналогичного сплайн-методу (см. п. 4.5), с шаблонами, количество точек в которых соответствует степени многочлена $\tilde{f}_m(x, \bar{a})$.

**ПОСТАНОВКА ЗАДАЧИ О НАХОЖДЕНИИ
СГЛАЖИВАЮЩИХ МНОГОЧЛЕНОВ
МЕТОДОМ НАИМЕНЬШИХ КВАДРАТОВ**

Пусть на множестве $\Omega = [a, b]$ задана сетка $\Omega_n = \{x_i, i = \overline{0, n}\}$, определяемая $n+1$ точкой x_0, x_1, \dots, x_n , а на сетке задана сеточная функция $y_i = f(x_i)$, $i = \overline{0, n}$:

$$y_0 = f(x_0), y_1 = f(x_1), \dots, y_n = f(x_n).$$

Предполагается, что сеточная функция получена, например, в физическом эксперименте с достаточно большой погрешностью относительно шага, т. е. $y_i = f(x_i) \pm \varepsilon_i$, причем значения ε_i превышают $O(h_{i+1})$, а узлы сетки могут быть заданы также с погрешностью. Как и ранее, будем использовать обозначение $f_i = f(x_i)$.

В качестве сглаживающей функции будем использовать обобщенный многочлен (4.43), в котором базисные функции $\varphi_i(x)$ являются степенными функциями или многочленами Чебышева, причем степень многочлена удовлетворяет условию $0 \leq m \leq n$.

Требуется найти такие коэффициенты многочлена a_0, a_1, \dots, a_m , чтобы выполнялось интегральное условие согласования (4.4):

$$\delta_m(\bar{a}) = \sqrt{\frac{1}{n+1} \sum_{i=0}^n [\tilde{f}_m(x_i, \bar{a}) - f_i]^2} \rightarrow \min_{\bar{a}}, \quad (4.44)$$

т. е. такой вектор $\bar{a} = \{a_0, a_1, \dots, a_m\}^T$, который обеспечивает минимум среднеквадратичной погрешности $\delta_m(\bar{a})$.

Метод решения поставленной задачи называется *методом наименьших квадратов* или *методом наилучшего среднеквадратичного приближения*. Согласно классификации метод является *сглаживающим* (см. рис. 4.7).

Отметим отличительные особенности решения задачи сглаживания методом наименьших квадратов по сравнению с решением задачи интерполяции.

1. Метод интерполяции — точечный метод, так как требует выполнения точечных условий интерполяции. Рассматриваемые здесь интегральные методы являются альтернативными точечным и не требуют точного удовлетворения функциональных условий, а требуют выполнения соответствия $\tilde{f}_m(x, \bar{a})$ и $y_i = y(x_i)$ в среднем (по интегралу или по интегральной сумме).

2. Исходная функция $y_i = f(x_i)$ задана не точно, а с погрешностью, существенно большей, чем в методе интерполяции. Эта функция, как правило, получена в результате некоторого физического эксперимента.

3. Количество точек x_i ($i = \overline{0, n}$), в которых задана исходная функция, как правило, значительно больше степени m многочлена ($n \gg m$). Поэтому между n и m нет строгого соответствия, как это имеет место в методе интерполяции.

ПОСТАНОВКА ЗАДАЧИ О НАХОЖДЕНИИ СГЛАЖИВАЮЩИХ МНОГОЧЛЕНОВ МЕТОДОМ ИНТЕГРАЛЬНОГО ПРИБЛИЖЕНИЯ

Пусть на множестве $\Omega = [a, b]$ задана сетка $\Omega_{m+1} = \{x_i, i = \overline{0, m+1}\}$, причем $x_0 = a < x_1 < x_2 < \dots < x_{m+1} = b$. Эта сетка, определяемая $(m + 2)$ несовпадающими узлами, порождает систему $(m + 1)$ частичных отрезков $[x_i, x_{i+1}]$ $i = \overline{0, m}$.

На рассматриваемой сетке задана сеточная функция одним из следующих способов.

Первый способ. На каждом частичном отрезке $[x_i, x_{i+1}]$, $i = \overline{0, m}$, сетки Ω_{m+1} задается величина интеграла I_i^{i+1} .

Второй способ. На частичных отрезках $[x_0, x_1], [x_0, x_2], \dots, [x_0, x_{m+1}]$ с фиксированным левым узлом задаются интегралы $I_0^1, I_0^2, \dots, I_0^{m+1}$.

Третий способ. Сеточная функция задается своими значениями $y_i = f(x_i)$, $i = \overline{0, m+1}$. Предполагается, что они известны с погрешностью ε_i , т. е. $y_i = f(x_i) \pm \varepsilon_i$. Соответствующие интегралы I_i^{i+1} , $i = \overline{0, m}$, и I_0^i , $i = \overline{1, m+1}$, вычисляются по квадратурным формулам (см. главу 5).

В качестве сглаживающей функции, так же как в методе наименьших квадратов, будем использовать обобщенный многочлен (4.43) степени m , в котором базисные функции $\varphi_i(x)$ являются степенными.

Требуется найти такие коэффициенты многочлена a_0, a_1, \dots, a_m , чтобы выполнялось интегральное условие согласования (4.5). Для первого способа задания функции оно имеет вид

$$\int_{x_i}^{x_{i+1}} \tilde{f}_m(x, \bar{a}) dx = I_i^{i+1}, \quad i = 0, 1, \dots, m, \quad (4.45)$$

а для второго способа —

$$\int_{x_0}^{x_i} \tilde{f}_m(x, \bar{a}) dx = I_0^i, \quad i = 1, \dots, m+1. \quad (4.46)$$

Замечание. Если вместе с интегральными условиями согласования использовать функциональные, то получаемые в результате многочлены называются интерполяционно-сглаживающими (см. п. 4.3).

4.4.2. МЕТОД НАИМЕНЬШИХ КВАДРАТОВ

В соответствии с постановкой задачи найдем коэффициенты a_0, a_1, \dots, a_m многочлена (4.43), обеспечивающие минимум среднеквадратичной погрешности (4.44):

$$\delta_m(\bar{a}) = \sqrt{\frac{1}{n+1} \sum_{i=0}^n [\tilde{f}_m(x_i, \bar{a}) - f_i]^2} \rightarrow \min_{a_0, a_1, \dots, a_m}. \quad (4.47)$$

Очевидно, минимум в (4.47) с учетом (4.43) достигается, если

$$\Delta = \sum_{i=0}^n [\varphi_0(x_i)a_0 + \varphi_1(x_i)a_1 + \dots + \varphi_m(x_i)a_m - f_i]^2 \rightarrow \min_{a_0, a_1, \dots, a_m}. \quad (4.48)$$

Так как на коэффициенты не наложено никаких ограничений, применим необходимые условия безусловного экстремума:

$$\frac{\partial \Delta}{\partial a_j} = 0, \quad j = 0, 1, \dots, m.$$

В результате получаем систему

$$\begin{aligned} \frac{\partial \Delta}{\partial a_0} &= 2 \sum_{i=0}^n [\varphi_0(x_i)a_0 + \varphi_1(x_i)a_1 + \dots + \varphi_m(x_i)a_m - f_i] \cdot \varphi_0(x_i) = 0, \\ \frac{\partial \Delta}{\partial a_1} &= 2 \sum_{i=0}^n [\varphi_0(x_i)a_0 + \varphi_1(x_i)a_1 + \dots + \varphi_m(x_i)a_m - f_i] \cdot \varphi_1(x_i) = 0, \\ &\dots \\ \frac{\partial \Delta}{\partial a_m} &= 2 \sum_{i=0}^n [\varphi_0(x_i)a_0 + \varphi_1(x_i)a_1 + \dots + \varphi_m(x_i)a_m - f_i] \cdot \varphi_m(x_i) = 0. \end{aligned} \quad (4.49)$$

Для компактной записи полученного результата удобно использовать скалярное произведение.

Скалярным произведением функций $\varphi_k(x)$ и $\varphi_l(x)$ на множестве точек $\{x_i, i = \overline{0, n}\}$ называется сумма произведений значений функций, вычисленных во всех точках, т. е.

$$(\varphi_k, \varphi_l) = \sum_{i=0}^n \varphi_k(x_i) \varphi_l(x_i).$$

Число $\|\varphi_k\| = \sqrt{(\varphi_k, \varphi_k)}$ называется *нормой* функции $\varphi_k(x)$ на множестве точек $\{x_i, i = \overline{0, n}\}$.

Тогда полученную систему (4.49) можно переписать в нормальной форме:

$$\begin{aligned} (\varphi_0, \varphi_0)a_0 + (\varphi_0, \varphi_1)a_1 + \dots + (\varphi_0, \varphi_m)a_m &= (f, \varphi_0), \\ (\varphi_1, \varphi_0)a_0 + (\varphi_1, \varphi_1)a_1 + \dots + (\varphi_1, \varphi_m)a_m &= (f, \varphi_1), \\ \dots \\ (\varphi_m, \varphi_0)a_0 + (\varphi_m, \varphi_1)a_1 + \dots + (\varphi_m, \varphi_m)a_m &= (f, \varphi_m), \end{aligned} \quad (4.50)$$

где $(f, \varphi_k) = \sum_{i=0}^n f_i \varphi_k(x_i)$. Таким образом, получено $(m+1)$ линейное уравнение с $(m+1)$ неизвестными a_0, a_1, \dots, a_m . В силу равенства $(\varphi_k, \varphi_l) = (\varphi_l, \varphi_k)$ матрица

$$A = \begin{pmatrix} (\varphi_0, \varphi_0) & (\varphi_0, \varphi_1) & \dots & (\varphi_0, \varphi_m) \\ (\varphi_1, \varphi_0) & (\varphi_1, \varphi_1) & \dots & (\varphi_1, \varphi_m) \\ \dots & \dots & \dots & \dots \\ (\varphi_m, \varphi_0) & (\varphi_m, \varphi_1) & \dots & (\varphi_m, \varphi_m) \end{pmatrix}$$

системы (4.50) является симметрической. Если базисные функции $\varphi_0, \varphi_1, \dots, \varphi_m$ линейно независимы, определитель матрицы A не равен нулю (он называется определителем Грама). Тогда решение системы (4.50) существует и единственно. Аналогичный вывод можно сделать и о задаче определения многочлена наилучшего среднеквадратичного приближения.

ПРИМЕНЕНИЕ СТЕПЕННЫХ БАЗИСНЫХ ФУНКЦИЙ

В качестве базисных функций используем степенные: $\varphi_j(x) = x^j, j = \overline{0, m}$. В этом случае обобщенный многочлен (4.43) имеет вид

$$\tilde{f}_m(x, \bar{a}) = \sum_{j=0}^m a_j x^j = a_0 + a_1 x + \dots + a_m x^m. \quad (4.51)$$

Тогда $(f, \varphi_j) = \sum_{i=0}^n f_i x_i^j, (\varphi_k, \varphi_l) = \sum_{i=0}^n x_i^{k+l}, (\varphi_k, \varphi_k) = \sum_{i=0}^n x_i^{2k}$ и система (4.50) имеет вид

$$\begin{aligned} \left(\sum_{i=0}^n 1 \right) a_0 + \left(\sum_{i=0}^n x_i \right) a_1 + \left(\sum_{i=0}^n x_i^2 \right) a_2 + \dots + \left(\sum_{i=0}^n x_i^m \right) a_m &= \sum_{i=0}^n f_i, \\ \left(\sum_{i=0}^n x_i \right) a_0 + \left(\sum_{i=0}^n x_i^2 \right) a_1 + \left(\sum_{i=0}^n x_i^3 \right) a_2 + \dots + \left(\sum_{i=0}^n x_i^{m+1} \right) a_m &= \sum_{i=0}^n x_i f_i, \\ \dots \\ \left(\sum_{i=0}^n x_i^m \right) a_0 + \left(\sum_{i=0}^n x_i^{m+1} \right) a_1 + \left(\sum_{i=0}^n x_i^{m+2} \right) a_2 + \dots + \left(\sum_{i=0}^n x_i^{2m} \right) a_m &= \sum_{i=0}^n x_i^m f_i. \end{aligned} \quad (4.52)$$

Обозначим

$$\begin{aligned} s_0 &= n+1, \quad t_0 = f_0 + f_1 + \dots + f_n, \\ s_k &= x_0^k + x_1^k + \dots + x_n^k, \quad k = 1, \dots, 2m; \\ t_k &= x_0^k f_0 + x_1^k f_1 + \dots + x_n^k f_n, \quad k = 1, \dots, m. \end{aligned}$$

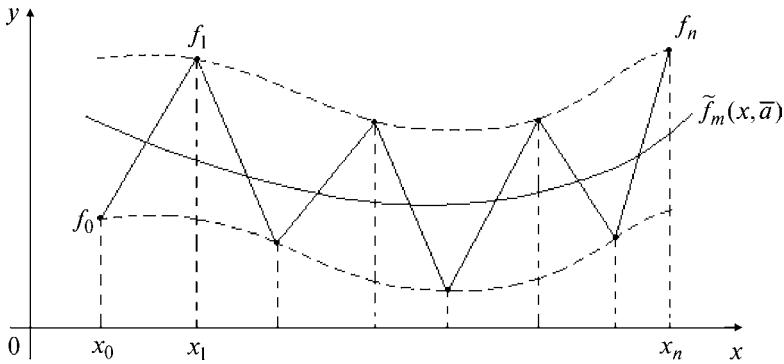


Рис. 4.7

Тогда система (4.52) преобразуется к виду

$$\begin{aligned} s_0 a_0 + s_1 a_1 + \dots + s_m a_m &= t_0, \\ s_1 a_0 + s_2 a_1 + \dots + s_{m+1} a_m &= t_1, \\ s_m a_0 + s_{m+1} a_1 + \dots + s_{2m} a_m &= t_m. \end{aligned} \quad (4.53)$$

Решая систему линейных алгебраических уравнений, находим неизвестные коэффициенты a_0, a_1, \dots, a_m . Подставляя решение в $\tilde{f}_m(x, \bar{a})$, получаем искомую формулу, которая сглаживает экспериментальные данные. Качественный характер поведения сглаживающей функции $\tilde{f}_m(x, \bar{a})$ относительно исходной функции показан на рисунке 4.7. Подчеркнем, что функция $\tilde{f}_m(x, \bar{a})$ расположена внутри полосы «разброса» функции $y_i = f(x_i)$ и является более плавной по сравнению с ломаной, проведенной через заданные экспериментальные точки. В этом состоит сглаживание исходной «пилообразной» функции или функции, заданной своими значениями с «разбросом».

Методика решения задачи сглаживания

1. Вычислить коэффициенты s_k , $k = \overline{0, 2m}$; t_k , $k = \overline{0, m}$, по заданной сеточной функции и записать систему (4.53).
2. Решить полученную систему одним из методов, изложенных в главе 1, и найти коэффициенты a_0, a_1, \dots, a_m .
3. Записать искомую сглаживающую функцию

$$\tilde{f}_m(x, \bar{a}) = a_0 + a_1 x + \dots + a_m x^m.$$

Замечания

1. Если для сеточной функции, заданной в $(n+1)$ -й точке x_0, x_1, \dots, x_n , определять многочлен степени $m = n$ методом наименьших квадратов, то тогда $\tilde{f}_m(x, \bar{a})$ совпадает с интерполяционным многочленом и метод становится эквивалентным методу интерполяции. При этом $\Delta = 0$ и $\delta_m(\bar{a}) = 0$.
2. В каждом конкретном случае может существовать «оптимальная» степень m , зависящая от конкретного поведения функции, числа n и вида базисных функций $\varphi_j(x)$. Предположим, что точность задания исходной функции характеризуется только одним значением ε , т. е. $\varepsilon_i = \varepsilon$ (ε_i все одинаковы).

Задавшись некоторым числом m и определив a_j , $j = \overline{0, m}$, требуется вычислить среднеквадратичное отклонение $\delta_m(\bar{a})$ и сравнить его с величиной ε .

Возможны три случая:

- если $\delta_m(\bar{a}) \gg \varepsilon$, то аппроксимация слишком грубая, степень m мала и необходимо ее увеличить;
- если $\delta_m(\bar{a}) \ll \varepsilon$, то аппроксимация физически недостоверна, найденная функция слишком «сплющена», старшие коэффициенты лишние и степень m надо уменьшить;
- если $\delta_m(\bar{a}) \approx \varepsilon$, степень многочлена m «оптимальна».

Расчет обычно начинают с $m = 1$, когда чаще всего $\delta_1(\bar{a}) \gg \varepsilon$, и увеличивают m , доходя до $\delta_m(\bar{a}) \approx \varepsilon$, если это возможно. Если $\delta_m(\bar{a})$ не становится близкой ε , то следует искать более подходящий вид базисных функций $\phi_j(x)$ или использовать, например, двухпараметрические зависимости.

3. Система (4.53) при $m \gg 1$ становится плохо обусловленной и определить полином $\tilde{f}_m(x, \bar{a})$ практически невозможно. Это связано с тем, что строчки матрицы A при больших m оказываются почти линейно зависимыми и, следовательно, $|A| \approx 0$. Чтобы этого избежать, необходимо вместо степенных базисных функций использовать полиномы Чебышева, способ реализации которых изложен ниже, либо строить аппроксимирующие полиномы методом интегрального приближения.

4. Метод наименьших квадратов реализует наилучшее в среднем приближение на всей области определения сеточной функции $y_i = f(x_i)$ и в некоторых случаях не учитывает локальных свойств аппроксимируемой функции (например, одиночный резкий «всплеск») [6].

5. Реализация метода наименьших квадратов с использованием степенных функций связана с решением системы линейных алгебраических уравнений относительно неопределенных коэффициентов, причем при изменении степени m все коэффициенты необходимо рассчитывать заново. Этот недостаток устраняется выбором ортогональных базисных функций $\phi_j(x)$.

Перейдем к рассмотрению примеров.

Пример 4.9. Решить задачу аппроксимации сеточной функции, заданной таблицей 4.13, при $m = 1$ и $m = 2$.

Таблица 4.13

i	0	1	2	3
x_i	2	3	4	5
$f(x_i) = f_i$	7	5	8	7

□ Пусть степень многочлена $m = 1$, тогда решение ищется в виде

$$\tilde{f}_1(x, \bar{a}) = a_0 + a_1 x.$$

1. Для составления системы (4.53):

$$\begin{aligned} s_0 a_0 + s_1 a_1 &= t_0, \\ s_1 a_0 + s_2 a_1 &= t_1 \end{aligned}$$

Таблица 4.14

x_i	f_i	1	x_i^2	$x_i f_i$
2	7	1	4	14
3	5	1	9	15
4	8	1	16	32
5	7	1	25	35
14	27	4	54	96
s_1	t_0	s_0	s_2	t_1

найдем ее коэффициенты s_0, s_1, s_2 . Расчеты поместим в таблицу 4.14, где в последней числовой строке находятся коэффициенты системы.

В результате получаем

$$4a_0 + 14a_1 = 27,$$

$$14a_0 + 54a_1 = 96.$$

2. Решение системы: $a_0 = 5,7; a_1 = 0,3$.

3. Искомая сглаживающая функция имеет вид $\tilde{f}_1(x, \bar{a}) = 5,7 + 0,3x$, а среднеквадратичная погрешность $\delta_1(\bar{a}) = 1,0368$.

Пусть $m = 2$, тогда решение ищется в виде $\tilde{f}_2(x, \bar{a}) = a_0 + a_1x + a_2x^2$.

1. Составим систему (4.53):

$$s_0a_0 + s_1a_1 + s_2a_2 = t_0,$$

$$s_1a_0 + s_2a_1 + s_3a_2 = t_1,$$

$$s_2a_0 + s_3a_1 + s_4a_2 = t_2.$$

Расчеты коэффициентов системы приведены в таблице 4.15.

Таблица 4.15

x_i	f_i	1	x_i^2	x_i^3	x_i^4	$x_i f_i$	$x_i^2 f_i$
2	7	1	4	8	16	14	28
3	5	1	9	27	81	15	45
4	8	1	16	64	256	32	128
5	7	1	25	125	625	35	175
14	27	4	54	224	978	96	376
s_1	t_0	s_0	s_2	s_3	s_4	t_1	t_2

В результате получаем систему

$$4a_0 + 14a_1 + 54a_2 = 27,$$

$$14a_0 + 54a_1 + 224a_2 = 96,$$

$$54a_0 + 224a_1 + 978a_2 = 376.$$

2. Решаем полученную систему методом Гаусса (см. главу 1).
 Прямой ход:

$$\left(\begin{array}{cccc} 4 & 14 & 54 & 27 \\ 14 & 54 & 224 & 96 \\ 54 & 224 & 978 & 376 \end{array} \right) \rightarrow \left(\begin{array}{cccc} 1 & \frac{7}{2} & \frac{27}{2} & \frac{27}{4} \\ 0 & 5 & 35 & 1,5 \\ 0 & 35 & 249 & 11,5 \end{array} \right) \rightarrow \left(\begin{array}{cccc} 1 & \frac{7}{2} & \frac{27}{2} & \frac{27}{4} \\ 0 & 1 & 7 & 0,3 \\ 0 & 0 & 4 & 1 \end{array} \right) \rightarrow \left(\begin{array}{cccc} 1 & \frac{7}{2} & \frac{27}{2} & \frac{27}{4} \\ 0 & 1 & 7 & 0,3 \\ 0 & 0 & 1 & \frac{1}{4} \end{array} \right).$$

Обратный ход:

$$\left(\begin{array}{ccc} 1 & \frac{7}{2} & \frac{27}{2} \\ 0 & 1 & 7 \\ 0 & 0 & 1 \end{array} \right) \cdot \left(\begin{array}{c} a_0 \\ a_1 \\ a_2 \end{array} \right) = \left(\begin{array}{c} \frac{27}{4} \\ 0,3 \\ \frac{1}{4} \end{array} \right) \text{ или } \begin{aligned} a_0 + \frac{7}{2}a_1 + \frac{27}{2}a_2 &= \frac{27}{4}, \\ a_1 + 7a_2 &= 0,3, \\ a_2 &= \frac{1}{4}. \end{aligned}$$

Отсюда

$$a_2 = \frac{1}{4}, \quad a_1 = \frac{3}{10} - 7 \cdot \frac{1}{4} = -\frac{29}{20},$$

$$a_0 = \frac{27}{4} - \frac{27}{4}a_2 - \frac{7}{2}a_1 = \frac{27}{4} - \frac{27}{4} \cdot \frac{1}{4} - \frac{7}{2} \cdot \left(-\frac{29}{20}\right) = \frac{169}{20}.$$

3. Искомая сглаживающая функция $\tilde{f}_2(x, \bar{a}) = \frac{169}{20} - \frac{29}{20}x + \frac{1}{4}x^2$, а среднеквадратичная погрешность $\delta_2(\bar{a}) = 1,0062$.

На рисунке 4.8 изображены заданная сеточная функция, сглаживающие многочлены при $m = 1$ и $m = 2$, а также интерполяционный многочлен, полученный в примерах 4.1 и 4.4 (при этом $m = 3$). ■

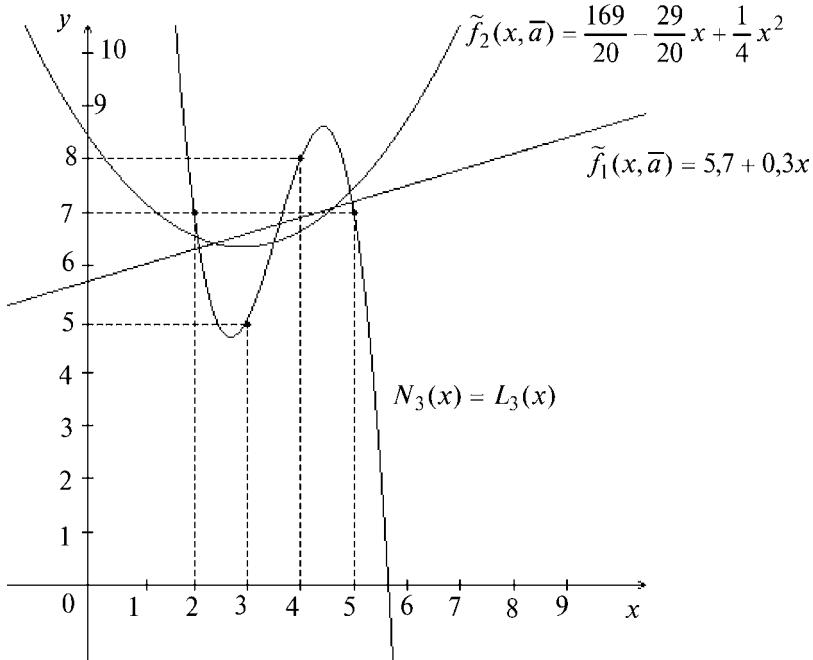


Рис. 4.8

i	0	1	2	3	4
x_i	1	2	3	4	5
f_i	-0,070	0,760	1,000	1,526	1,449

Пример 4.10. Пусть сеточное представление формульной функции $y = \ln x$ задано на отрезке $[1; 5]$ с погрешностью $\varepsilon = 0,1$ (табл. 4.16). По этим данным найти сглаживающий многочлен $\tilde{f}_m(x, \bar{a})$ методом наименьших квадратов с использованием последовательного выбора степени $m = 1, 2, \dots$

□ Примем сначала $m = 1$.

1. Искомая функция $\tilde{f}_1(x, \bar{a}) = a_0 + a_1 x$ определяется из системы (4.53):

$$\left. \begin{aligned} \left(\sum_{i=0}^4 1 \right) a_0 + \left(\sum_{i=0}^4 x_i \right) a_1 &= \sum_{i=0}^4 f_i, \\ \left(\sum_{i=0}^4 x_i \right) a_0 + \left(\sum_{i=0}^4 x_i^2 \right) a_1 &= \sum_{i=0}^4 x_i f_i \end{aligned} \right\} \Rightarrow \begin{aligned} 5a_0 + 15a_1 &= 4,665, \\ 15a_0 + 55a_1 &= 17,799. \end{aligned}$$

2. Решая эту систему, например методом Гаусса, получаем $a_0 = -2,2082$; $a_1 = 0,3804$.

3. Сглаживающий многочлен имеет вид $\tilde{f}_1(x, \bar{a}) = -0,2082 + 0,3804x$. Вычислим среднеквадратичную погрешность $\delta_1 = 0,205 > \varepsilon = 0,1$.

Так как условие $\delta_1 \approx \varepsilon$ не выполнено, возьмем $m = 2$:

$$\tilde{f}_2(x, \bar{a}) = a_0 + a_1 x + a_2 x^2.$$

Определяя a_0, a_1, a_2 из системы (4.53) при $m = 2$, получаем

$$\tilde{f}_2(x, \bar{a}) = -0,9726 + 1,0356x - 0,1092x^2.$$

Поскольку среднеквадратичная погрешность $\delta_2 = 0,094 \leq \varepsilon = 0,1$, задача решена. ■

ПРИМЕНЕНИЕ ЭМПИРИЧЕСКИХ ФОРМУЛ

Исходя из приблизительного сходства с экспериментальными данными выбирается одна из двухпараметрических формул:

$$\begin{aligned} \tilde{f}(x, a_0, a_1) &= a_0 + \frac{a_1}{x}; \quad \tilde{f}(x, a_0, a_1) = \frac{1}{a_0 + a_1 x}; \quad \tilde{f}(x, a_0, a_1) = \frac{x}{a_0 + a_1 x}; \\ \tilde{f}(x, a_0, a_1) &= a_0 a_1^x; \quad \tilde{f}(x, a_0, a_1) = a_0 e^{a_1 x}; \quad \tilde{f}(x, a_0, a_1) = \frac{1}{a_0 + a_1 e^{-x}}; \\ \tilde{f}(x, a_0, a_1) &= a_0 + a_1 \ln x; \quad \tilde{f}(x, a_0, a_1) = \frac{a_0}{a_1 + x}; \quad \tilde{f}(x, a_0, a_1) = \frac{a_0 x}{a_1 + x}, \dots, \end{aligned}$$

где a_0, a_1 — неизвестные коэффициенты.

Требуется найти коэффициенты a_0, a_1 , минимизирующие погрешность сглаживания на основе метода наименьших квадратов:

$$\Delta = \sum_{i=0}^n [\tilde{f}(x_i, a_0, a_1) - f_i]^2 \rightarrow \min_{a_0, a_1}$$

Для нахождения коэффициентов a_0, a_1 применяются необходимые условия экстремума:

$$\frac{\partial \Delta}{\partial a_0} = 0, \quad \frac{\partial \Delta}{\partial a_1} = 0.$$

В отличие от применения степенных функций, нахождение неизвестных коэффициентов сводится к решению двух нелинейных алгебраических уравнений с помощью одного из методов, изложенных в п. 3.2.

Замечание. В качестве эмпирических формул могут быть использованы также и трехпараметрические зависимости, например $\tilde{f}(x, a_0, a_1, a_2) = a_0 x^{a_1} + a_2$, $\tilde{f}(x, a_0, a_1, a_2) = a_0 e^{a_1 x} + a_2$ и др.

ПРИМЕНЕНИЕ ОРТОГОНАЛЬНЫХ БАЗИСНЫХ ФУНКЦИЙ

При нахождении коэффициентов обобщенного многочлена (4.43) с помощью ортогональных базисных функций нет необходимости решать систему (4.53).

Функции $\varphi_k(x)$ и $\varphi_l(x)$ называются *ортогональными* на множестве точек $\{x_i\}$, $i = 0, 1, \dots, n$, если их скалярное произведение на этом множестве равно нулю:

$$(\varphi_k, \varphi_l) = \sum_{i=0}^n \varphi_k(x_i) \cdot \varphi_l(x_i) = 0, \quad k \neq l.$$

Система функций $\{\varphi_j(x)\}$, $j = 0, 1, \dots, m$, называется *ортогональной* на множестве точек $\{x_i\}$, $i = 0, 1, \dots, n$, если все функции этой системы попарно ортогональны на этом множестве.

Так, если в обобщенном многочлене

$$Q_m(x) = a_0 \varphi_0(x) + a_1 \varphi_1(x) + \dots + a_m \varphi_m(x)$$

система базисных функций ортогональная, то $(\varphi_k, \varphi_l) = 0$, $k \neq l$, а (4.50) перепишется в виде

$$(\varphi_0, \varphi_0) a_0 = (f, \varphi_0),$$

$$(\varphi_1, \varphi_1) a_1 = (f, \varphi_1),$$

.....

$$(\varphi_m, \varphi_m) a_m = (f, \varphi_m),$$

т. е. все недиагональные элементы в матрице системы (4.50) становятся равными нулю. Следовательно, коэффициенты обобщенного многочлена находятся по формуле

$$a_j = \frac{(f, \varphi_j)}{(\varphi_j, \varphi_j)} = \frac{\sum_{i=0}^n f_i \cdot \varphi_j(x_i)}{\sum_{i=0}^n \varphi_j^2(x_i)}, \quad j = 0, 1, \dots, m. \quad (4.54)$$

Коэффициенты (4.54) обобщенного многочлена называются *коэффициентами Фурье* функции $y_i = f(x_i)$ относительно ортогональной системы функций.

Теорема 4.2 (о точечном среднеквадратичном приближении функций ортогональными многочленами [37]).

Для функции $y_i = f(x_i)$, определенной на множестве точек $\{x_i\}$, $i = 0, 1, \dots, n$, обобщенный многочлен m -й степени $Q_m(x)$ с коэффициентами Фурье относительно ортогональной на этом множестве точек системы функций $\{\varphi_j(x)\}$ является многочленом наилучшего среднеквадратичного приближения этой функции, причем квадрат наименьшего среднеквадратичного отклонения определяется соотношением

$$\rho^2(f, Q_m) = \|f\|^2 - \sum_{j=0}^m a_j^2 \|\varphi_j\|^2, \quad (4.55)$$

где a_j — коэффициенты Фурье, определяемые по формуле (4.54); $\|f\|^2 = \sum_{i=0}^n f_i^2$, $\|\varphi_j\|^2 = \sum_{i=0}^n \varphi_j^2(x_i)$, $\rho(f, Q_m) = \sqrt{\sum_{i=0}^n [Q_m(x_i) - f_i]^2}$.

Погрешность данного приближения вычисляется по формуле

$$\bar{\rho}(f, Q_m) = \frac{\rho(f, Q_m)}{\|f\|}. \quad (4.56)$$

Рассмотрим применение многочленов Чебышева в качестве системы базисных функций, которая используется для получения многочленов $Q_m(x)$.

Многочленами Чебышева $g_0(x), g_1(x), \dots, g_m(x)$ на множестве точек $\{x_i\}$, $i = 0, 1, \dots, n$, называются алгебраические многочлены, которые ортогональны на этом множестве, с нормой $\|g_k\|$, отличной от нуля, и определяются следующими рекуррентными формулами:

$$\begin{aligned} g_0(x) &\equiv 1; \quad g_1(x) = x - a; \\ g_k(x) &= (x - a_k)g_{k-1}(x) - b_k g_{k-2}(x), \quad k = 2, 3, \dots, m, \end{aligned} \quad (4.57)$$

где

$$a = \frac{1}{n+1} \sum_{i=0}^n x_i; \quad a_k = \frac{\sum_{i=0}^n x_i \cdot g_{k-1}^2(x_i)}{\sum_{i=0}^n g_{k-1}^2(x_i)}; \quad b_k = \frac{\sum_{i=0}^n x_i \cdot g_{k-2}(x_i) \cdot g_{k-1}(x_i)}{\sum_{i=0}^n g_{k-2}^2(x_i)}.$$

Соотношения для коэффициентов a , a_k , b_k получены из условия ортогональности.

i	0	1	2	3
x_i	0	1	3	4
f_i	4	0	1	2

Применение ортогональных базисных функций поясним на примере [37].

Пример 4.11. Пусть дана сеточная функция (табл. 4.17).

Требуется сгладить эту функцию алгебраическими многочленами наилучшего среднеквадратичного приближения $Q_0(x), Q_1(x), Q_2(x), Q_3(x)$, где

$$Q_0(x) = a_0 g_0(x),$$

$$Q_1(x) = a_0 g_0(x) + a_1 g_1(x),$$

$$Q_2(x) = a_0 g_0(x) + a_1 g_1(x) + a_2 g_2(x),$$

$$Q_3(x) = a_0 g_0(x) + a_1 g_1(x) + a_2 g_2(x) + a_3 g_3(x),$$

и найти погрешность каждого из указанных приближений, т. е. $\bar{\rho}(f, Q_0)$, $\bar{\rho}(f, Q_1)$, $\bar{\rho}(f, Q_2)$, $\bar{\rho}(f, Q_3)$.

□1. Определим ортогональные многочлены Чебышева $g_0(x), g_1(x), g_2(x), g_3(x)$ на множестве точек $\{0; 1; 3; 4\}$.

Сначала найдем $g_0(x), g_1(x)$. В соответствии с первой формулой из (4.57) функция $g_0(x) = 1$, а многочлен $g_1(x) = x - a$, $a = \frac{1}{n+1} \sum_{i=0}^n x_i$. При $n = 3$ получаем

$$a = \frac{1}{4} \sum_{i=0}^3 x_i = \frac{1}{4}(0+1+3+4) = 2.$$

Таким образом, $g_0(x) = 1; g_1(x) = x - 2$.

Теперь определим многочлен $g_2(x)$, для которого справедлива формула $g_2(x) = (x - a_2)g_1(x) - b_2g_0(x)$. Функции $g_0(x)$ и $g_1(x)$ получены. Требуется только найти a_2 и b_2 :

$$a_2 = \frac{\sum_{i=0}^3 x_i g_1^2(x_i)}{\|g_1\|^2}; \quad b_2 = \frac{\sum_{i=0}^3 x_i g_0(x_i) g_1(x_i)}{\|g_0\|^2}.$$

Вычислим сначала квадраты норм $\|g_0\|^2$ и $\|g_1\|^2$:

$$\begin{aligned} \|g_0\|^2 &= \sum_{i=0}^3 g_0^2(x_i) = 4; \quad \|g_1\|^2 = \sum_{i=0}^3 g_1^2(x_i) = g_1^2(0) + g_1^2(1) + g_1^2(3) + g_1^2(4) = \\ &= (0-2)^2 + (1-2)^2 + (3-2)^2 + (4-2)^2 = 10. \end{aligned}$$

Далее найдем a_2 и b_2 :

$$a_2 = \frac{1}{10} \sum_{i=0}^3 x_i \cdot g_1^2(x_i) = \frac{1}{10}[0 \cdot g_1^2(0) + 1 \cdot g_1^2(1) + 3 \cdot g_1^2(3) + 4 \cdot g_1^2(4)] = \frac{1}{10}(0+1+3+4) = 2;$$

$$b_2 = \frac{1}{4} \sum_{i=0}^3 x_i g_0(x_i) g_1(x_i) = \frac{1}{4}[0 \cdot 1 \cdot (-2) + 1 \cdot 1 \cdot (-1) + 3 \cdot 1 \cdot 1 + 4 \cdot 1 \cdot 2] = \frac{5}{2}.$$

Итак, многочлен

$$g_2(x) = (x-2)g_1(x) - \frac{5}{2} g_0(x) = (x-2)^2 - \frac{5}{2}.$$

Определим многочлен $g_3(x)$, для которого справедлива формула $g_3(x) = (x - a_3)g_2(x) - b_3g_1(x)$. Функции $g_1(x)$ и $g_2(x)$ получены. Требуется найти a_3, b_3 :

$$a_3 = \frac{\sum_{i=0}^3 x_i \cdot g_2^2(x_i)}{\|g_2\|^2}; \quad b_3 = \frac{\sum_{i=0}^3 x_i \cdot g_1(x_i) \cdot g_2(x_i)}{\|g_1\|^2}.$$

Вычислим сначала квадрат нормы $\|g_2\|^2$:

$$\begin{aligned} \|g_2\|^2 &= \sum_{i=0}^3 g_2^2(x_i) = \underbrace{\left[(x_0-2)^2 - \frac{5}{2}\right]^2}_{g_2^2(x_0)} + \underbrace{\left[(x_1-2)^2 - \frac{5}{2}\right]^2}_{g_2^2(x_1)} + \underbrace{\left[(x_2-2)^2 - \frac{5}{2}\right]^2}_{g_2^2(x_2)} + \\ &+ \underbrace{\left[(x_3-2)^2 - \frac{5}{2}\right]^2}_{g_2^2(x_3)} = \underbrace{\left(4 - \frac{5}{2}\right)^2}_{\frac{9}{4}} + \underbrace{\left(1 - \frac{5}{2}\right)^2}_{\frac{9}{4}} + \underbrace{\left(1 - \frac{5}{2}\right)^2}_{\frac{9}{4}} + \underbrace{\left(4 - \frac{5}{2}\right)^2}_{\frac{9}{4}} = 9. \end{aligned}$$

Далее найдем a_3, b_3 :

$$\begin{aligned} a_3 &= \frac{\sum_{i=0}^3 x_i \cdot g_2^2(x_i)}{\|g_2\|^2} = \frac{1}{9} \left(0 \cdot \frac{9}{4} + 1 \cdot \frac{9}{4} + 3 \cdot \frac{9}{4} + 4 \cdot \frac{9}{4} \right) = 2; \\ b_3 &= \frac{\sum_{i=0}^3 x_i \cdot g_1(x_i) \cdot g_2(x_i)}{\|g_1\|^2} = \frac{1}{10} \left[0 \cdot (-2) \cdot \frac{3}{2} + 1 \cdot (-1) \cdot \left(-\frac{3}{2}\right) + 3 \cdot 1 \cdot \left(-\frac{3}{2}\right) + 4 \cdot 2 \cdot \left(\frac{1}{2}\right) \right] = \frac{9}{10}. \end{aligned}$$

Итак, $g_3(x) = (x-2) \left[(x-2)^2 - \frac{5}{2} \right] - \frac{9}{10}(x-2) = (x-2)^3 - \frac{17}{5}(x-2)$.

Аналогичные вычисления выполним для многочлена $g_3(x)$:

$$\begin{aligned} \|g_3\|^2 &= \sum_{i=0}^3 g_3^2(x_i) = g_3^2(x_0) + g_3^2(x_1) + g_3^2(x_2) + g_3^2(x_3); \\ g_3(x_0)|_{x_0=0} &= -8 + \frac{34}{5} = -\frac{6}{5}; \quad g_3(x_1)|_{x_1=1} = -1 - \frac{17}{5} \cdot (-1) = \frac{12}{5}; \\ g_3(x_2)|_{x_2=3} &= 1 - \frac{17}{5} = -\frac{12}{5}; \quad g_3^2(x_3)|_{x_3=4} = 8 - \frac{17}{5} \cdot 2 = \frac{6}{5}; \\ \|g_3\|^2 &= \frac{36}{25} + \frac{144}{25} + \frac{144}{25} + \frac{36}{25} = \frac{72}{5}. \end{aligned}$$

Таблица 4.18

i	x_i	$G_0(x_i)$	$G_1(x_i)$	$G_2(x_i)$	$G_3(x_i)$
0	0	1	-2	1,5	-1,2
1	1	1	-1	-1,5	2,4
2	3	1	1	-1,5	-2,4
3	4	1	2	1,5	1,2

Значения многочленов $g_0(x), g_1(x), g_2(x), g_3(x)$ на множестве точек $\{0; 1; 3; 4\}$ представлены в таблице 4.18.

2. Определим многочлены наилучшего среднеквадратичного приближения $Q_0(x), Q_1(x), Q_2(x), Q_3(x)$:

- многочлен $Q_0(x)$:

$$Q_0(x) = a_0 g_0(x); \quad a_0 = \frac{(f, g_0)}{\|g_0\|^2} = \frac{\sum_{i=0}^3 f_i \cdot g_0(x_i)}{\|g_0\|^2} = \frac{1}{4}(4 \cdot 1 + 0 \cdot 1 + 1 \cdot 1 + 2 \cdot 1) = \frac{7}{4}.$$

Таким образом, сглаживающий многочлен $Q_0(x) = \frac{7}{4}$;

- многочлен $Q_1(x)$:

$$\begin{aligned} Q_1(x) &= a_0 g_0(x) + a_1 g_1(x); \quad a_1 = \frac{(f, g_1)}{\|g_1\|^2} = \frac{\sum_{i=0}^3 f_i \cdot g_1(x_i)}{\|g_1\|^2} = \\ &= \frac{4 \cdot (-2) + 0 \cdot (-1) + 1 \cdot 1 + 2 \cdot 2}{10} = -\frac{3}{10}. \end{aligned}$$

Таким образом, $Q_1(x) = \frac{7}{4} - \frac{3}{10}(x-2) = -\frac{3}{10}x + \frac{47}{20}$;

- многочлен $Q_2(x)$:

$$\begin{aligned} Q_2(x) &= a_0 g_0(x) + a_1 g_1(x) + a_2 g_2(x); \\ a_2 &= \frac{(f, g_2)}{\|g_2\|^2} = \frac{\sum_{i=0}^3 f_i \cdot g_2(x_i)}{\|g_2\|^2} = \frac{1}{9}\left(4 \cdot \frac{3}{2} - 0 \cdot \frac{3}{2} - 1 \cdot \frac{3}{2} + 2 \cdot \frac{3}{2}\right) = \frac{5}{6}; \\ Q_2(x) &= -\frac{3}{10}x + \frac{47}{20} + \frac{5}{6}\left[(x-2)^2 - \frac{5}{2}\right] = \frac{5}{6}x^2 - \frac{109}{30}x + \frac{18}{5}; \end{aligned}$$

- многочлен $Q_3(x)$:

$$\begin{aligned} Q_3(x) &= a_0 g_0(x) + a_1 g_1(x) + a_2 g_2(x) + a_3 g_3(x); \\ a_3 &= \frac{(f, g_3)}{\|g_3\|^2} = \frac{\sum_{i=0}^3 f_i \cdot g_3(x_i)}{\|g_3\|^2} = \frac{5}{72}\left[4 \cdot \left(-\frac{6}{5}\right) + 0 \cdot \frac{12}{5} + 1 \cdot \left(-\frac{12}{5}\right) + 2 \cdot \frac{6}{5}\right] = -\frac{1}{3}; \\ Q_3(x) &= \frac{5}{6}x^2 - \frac{109}{30}x + \frac{18}{5} - \frac{1}{3}\left[(x-2)^2 + \frac{17}{5}(x-2)\right] = -\frac{1}{3}x^3 + \frac{17}{6}x^2 - \frac{13}{2}x + 4. \end{aligned}$$

Значения алгебраических многочленов наилучшего среднеквадратичного приближения в точках x_i представлены в таблице 4.19.

Таблица 4.19

i	x_i	$Q_0(x_i)$	$Q_1(x_i)$	$Q_2(x_i)$	$Q_3(x_i)$	f_i
0	0	1,75	2,35	3,6	4,00	4
1	1	1,75	2,05	0,8	0,00	0
2	3	1,75	1,45	0,2	1,00	1
3	4	1,75	1,15	2,4	2,00	2

3. Найдем погрешности приближений функции $y_i = f(x_i)$ многочленами $Q_0(x), Q_1(x), Q_2(x), Q_3(x)$.

Для вычисления погрешности многочлена наилучшего среднеквадратичного приближения используются формулы (4.55), (4.56).

Определим $\|f\|$:

$$\|f\| = (f, f)^{\frac{1}{2}} = \sqrt{\sum_{i=0}^3 f_i} = \sqrt{(16 + 0 + 1 + 4)} = \sqrt{21}.$$

Тогда можно получить численное значение погрешности многочлена $Q_0(x)$:

$$\begin{aligned}\bar{\rho}(f, Q_0) &= \frac{1}{\sqrt{21}} \sqrt{\rho^2(f, Q_0)} = \frac{1}{\sqrt{21}} \sqrt{\|f\|^2 - a_0 \cdot \|g_0\|^2} = \\ &= \frac{1}{\sqrt{21}} \sqrt{21 - \frac{49}{16} \cdot 4} = \sqrt{\frac{50}{12}} \approx 0,645.\end{aligned}$$

Аналогично находятся

$$\bar{\rho}(f, Q_1) \approx 0,611; \quad \bar{\rho}(f, Q_2) \approx 0,276,$$

$$\bar{\rho}(f, Q_3) = \frac{1}{\|f\|} \sqrt{\|f\|^2 - \sum_{j=0}^3 a_j^2 \|g_j\|^2} = \frac{1}{\sqrt{21}} \sqrt{21 - \left(\frac{49}{16} \cdot 4 + \frac{9}{100} \cdot 10 + \frac{25}{36} \cdot 9 + \frac{1}{9} \cdot \frac{72}{5} \right)} = 0.$$

Очевидно, при изменении степени m нет необходимости рассчитывать все коэффициенты заново. Это является одним из преимуществ применения ортогональных многочленов. ■

4.4.3. МЕТОД НАИЛУЧШЕГО ИНТЕГРАЛЬНОГО ПРИБЛИЖЕНИЯ

Сначала рассмотрим применение *глобального способа*. В соответствии с постановкой задачи, изложенной в п. 4.4.1, будем искать коэффициенты сглаживающего многочлена

$$\tilde{f}_m(x, \bar{a}) \equiv S_m(x) = a_0 + a_1 x + a_2 x^2 + \dots + a_m x^m$$

из условия (4.45), если функция задана первым способом (т. е. интегралами $I_0^1, I_1^2, \dots, I_m^{m+1}$), или из условия (4.46), если функция задана вторым способом (интегралами $I_0^1, I_0^2, \dots, I_0^{m+1}$).

Для вычисления коэффициентов многочлена используются интегральные условия согласования (4.45) или (4.46).

Условие (4.45) имеет вид

$$\begin{aligned}
 & \int_{x_0}^{x_1} (a_0 + a_1 x + a_2 x^2 + \dots + a_m x^m) dx = I_0^1, \\
 & \int_{x_1}^{x_2} (a_0 + a_1 x + a_2 x^2 + \dots + a_m x^m) dx = I_1^2, \\
 & \dots \dots \dots \\
 & \int_{x_m}^{x_{m+1}} (a_0 + a_1 x + a_2 x^2 + \dots + a_m x^m) dx = I_m^{m+1}.
 \end{aligned} \tag{4.58}$$

Условие (4.46) приводит к системе

$$\begin{aligned} \int_{x_0}^{x_1} (a_0 + a_1 x + a_2 x^2 + \dots + a_m x^m) dx &= I_0^1, \\ \int_{x_0}^{x_2} (a_0 + a_1 x + a_2 x^2 + \dots + a_m x^m) dx &= I_0^2, \\ \dots \\ \int_{x_0}^{x_{m+1}} (a_0 + a_1 x + a_2 x^2 + \dots + a_m x^m) dx &= I_0^{m+1}. \end{aligned} \quad (4.59)$$

Интегрируя левые части каждой из систем (4.58), (4.59) и решая соответствующие системы $(m+1)$ линейных алгебраических уравнений с $(m+1)$ неизвестными, можно найти коэффициенты $a_0, a_1, a_2, \dots, a_m$ и искомые сглаживающие многочлены $\tilde{f}_m(x, \bar{a})$.

Замечания

- Системы (4.58), (4.59) равносильны и получаются одна из другой с помощью линейных комбинаций (попарного сложения и вычитания уравнений).
- Системы (4.58), (4.59) имеют единственное решение, так как можно показать, что определители матриц этих линейных систем не равны нулю, т. е.

$$\left| \begin{array}{cccc} \int_{x_0}^{x_1} dx & \int_{x_0}^{x_1} x dx & \dots & \int_{x_0}^{x_1} x^m dx \\ x_0 & x_0 & & x_0 \\ \int_{x_1}^{x_2} dx & \int_{x_1}^{x_2} x dx & \dots & \int_{x_1}^{x_2} x^m dx \\ x_1 & x_1 & & x_1 \\ \dots & \dots & \dots & \dots \\ \int_{x_m}^{x_{m+1}} dx & \int_{x_m}^{x_{m+1}} x dx & \dots & \int_{x_m}^{x_{m+1}} x^m dx \end{array} \right| \neq 0 \text{ и } \left| \begin{array}{cccc} \int_{x_0}^{x_1} dx & \int_{x_0}^{x_1} x dx & \dots & \int_{x_0}^{x_1} x^m dx \\ x_0 & x_0 & & x_0 \\ \int_{x_0}^{x_2} dx & \int_{x_0}^{x_2} x dx & \dots & \int_{x_0}^{x_2} x^m dx \\ x_0 & x_0 & & x_0 \\ \dots & \dots & \dots & \dots \\ \int_{x_0}^{x_{m+1}} dx & \int_{x_0}^{x_{m+1}} x dx & \dots & \int_{x_0}^{x_{m+1}} x^m dx \end{array} \right| \neq 0.$$

Таким образом, задача нахождения многочлена наилучшего интегрального приближения имеет единственное решение [16].

- Как показано в [16], при небольших m ($m \leq 7$) решать систему (4.58) необязательно, так как ее решение можно записать сразу, если ввести в рассмотрение конечно-интегральные разности.

- При больших m глобальный способ построения многочленов наилучшего интегрального приближения, как и интерполяционных многочленов, использовать не рекомендуется.

Применим теперь *кусочный способ*. Рассмотрим сеточную функцию, заданную на равномерной сетке ($h = \text{const}$): $y_i = f(x_i)$, $i = \overline{0, n}$. Пусть степень многочлена равна m , и этот многочлен строится на некотором внутреннем отрезке $[x_k, x_{k+m+1}]$, $k = 0, 1, \dots, n-m-1$. Этот отрезок, в свою очередь, заключает в себе $(m+1)$ частичных отрезков, на которых заданы интегралы $I_k^{k+1}, I_{k+1}^{k+2}, \dots, I_{k+m}^{k+m+1}$. Для большей унификации обозначений индексов будем счи-

тать, что для $m = 1$ на частичном отрезке $[x_k, x_{k+2}]$ определено дополнительное разбиение этого отрезка на два подотрезка ($p = 2$), для $m = 2$ на частичном отрезке $[x_k, x_{k+3}]$ определено разбиение на три подотрезка ($p = 3$) и т. д.

Конечно-интегральные разности определяются по следующим формулам:

- *конечно-интегральные разности первого порядка ($m = 1, p = 2$):*

$$\Delta I_k^{\frac{k+1}{p}} = I_{\frac{k+1}{p}}^{\frac{k+2}{p}} - I_k^{\frac{k+1}{p}};$$

- *конечно-интегральные разности второго порядка ($m = 2, p = 3$):*

$$\Delta^2 I_k^{\frac{k+1}{p}} = \Delta I_{\frac{k+1}{p}}^{\frac{k+2}{p}} - \Delta I_k^{\frac{k+1}{p}} = I_k^{\frac{k+1}{p}} - 2I_{\frac{k+1}{p}}^{\frac{k+2}{p}} + I_{\frac{k+2}{p}}^{\frac{k+3}{p}};$$

.....

- *конечно-интегральные разности i -го порядка ($m = i, p = i + 1$):*

$$\Delta^i I_k^{\frac{k+1}{p}} = \Delta^{i-1} I_{\frac{k+1}{p}}^{\frac{k+2}{p}} - \Delta^{i-1} I_k^{\frac{k+1}{p}} = \sum_{j=0}^i (-1)^j C_i^j I_{\frac{k+j}{p}}^{\frac{k+(j+1)}{p}}.$$

Здесь для многочленов степени $m = 1, m = 2, m = 3$ принимается $p = 2, p = 3, p = 4$ соответственно и т. д.

Тогда, например, для $S_1(x), S_2(x), S_3(x)$ на отрезках $[x_k, x_{k+2}], [x_k, x_{k+3}], [x_k, x_{k+4}]$ получаем [16]:

$$S_1(x) = \frac{2}{h_k} \left(I_k^{\frac{k+1}{2}} - \frac{1}{2} \Delta I_k^{\frac{k+1}{2}} \right) + \frac{2^2}{h_k^2} \Delta I_k^{\frac{k+1}{2}} (x - x_k); \quad (4.60)$$

$$S_2(x) = \frac{3}{h_k} \left(I_k^{\frac{k+1}{3}} - \frac{1}{2} \Delta I_k^{\frac{k+1}{3}} + \frac{1}{3} \Delta^2 I_k^{\frac{k+1}{3}} \right) + \frac{3^2}{h_k^2} \left(\Delta I_k^{\frac{k+1}{3}} - \Delta^2 I_k^{\frac{k+1}{3}} \right) \cdot (x - x_k) + \\ + \frac{3^3}{2h_k^3} \Delta^2 I_k^{\frac{k+1}{3}} (x - x_k)^2; \quad (4.61)$$

$$S_3(x) = \frac{4}{h_k} \left(I_k^{\frac{k+1}{4}} - \frac{1}{2} \Delta I_k^{\frac{k+1}{4}} + \frac{1}{3} \Delta^2 I_k^{\frac{k+1}{4}} - \frac{1}{4} \Delta^3 I_k^{\frac{k+1}{4}} \right) + \\ + \frac{4^2}{h_k^2} \left(\Delta I_k^{\frac{k+1}{4}} - \Delta^2 I_k^{\frac{k+1}{4}} + \frac{11}{12} \Delta^3 I_k^{\frac{k+1}{4}} \right) \times \\ \times (x - x_k) + \frac{4^3}{h_k^3} \left(\frac{1}{2} \Delta^2 I_k^{\frac{k+1}{4}} - \frac{3}{4} \Delta^3 I_k^{\frac{k+1}{4}} \right) \cdot (x - x_k)^2 + \frac{4^4}{6h_k^4} \Delta^3 I_k^{\frac{k+1}{4}} (x - x_k)^3. \quad (4.62)$$

Здесь $h_k = x_{k+m+1} - x_k$ — шаг, относящийся к внутреннему отрезку $[x_k, x_{k+m+1}]$. Для $S_1(x)$ используются интегралы $I_k^{\frac{k+1}{2}}, I_{\frac{k+1}{2}}^{\frac{k+1}{2}}$; для $S_2(x)$ — интегралы $I_k^{\frac{k+1}{3}}, I_{\frac{k+1}{3}}^{\frac{k+2}{3}}, I_{\frac{k+2}{3}}^{\frac{k+1}{2}}$; для $S_3(x)$ — интегралы $I_k^{\frac{k+1}{4}}, I_{\frac{k+1}{4}}^{\frac{k+1}{2}}, I_{\frac{k+1}{2}}^{\frac{k+3}{4}}, I_{\frac{k+3}{4}}^{\frac{k+1}{4}}$ (см. рис. 4.9).

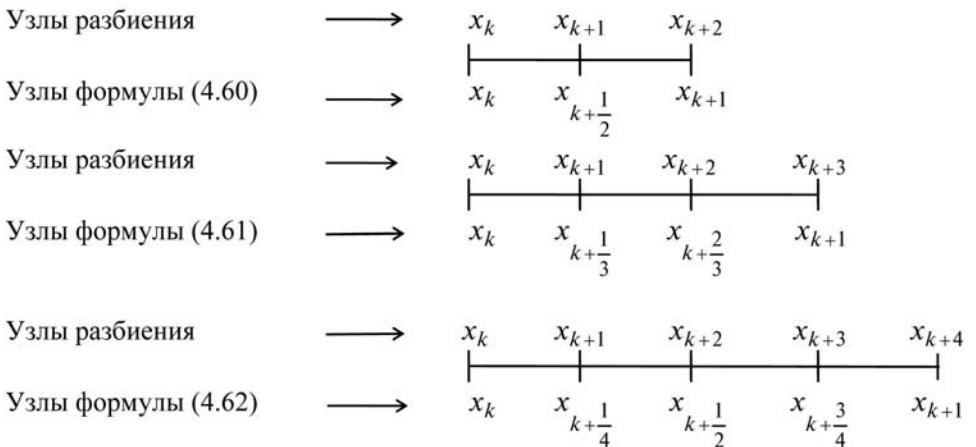


Рис. 4.9

Коэффициенты многочленов более высоких степеней могут быть получены рекуррентным способом.

Так, многочлены $S_m(x)$ до степени $m = 7$ имеют вид

$$\begin{aligned}
 S_m(x) = & \frac{p}{h_k} \left(I_k^{k+\frac{1}{p}} - \frac{1}{2} \Delta I_k^{k+\frac{1}{p}} + \frac{1}{3} \Delta^2 I_k^{k+\frac{1}{p}} - \frac{1}{4} \Delta^3 I_k^{k+\frac{1}{p}} + \frac{1}{5} \Delta^4 I_k^{k+\frac{1}{p}} - \right. \\
 & \quad \left. - \frac{1}{6} \Delta^5 I_k^{k+\frac{1}{p}} + \frac{1}{7} \Delta^6 I_k^{k+\frac{1}{p}} - \frac{1}{8} \Delta^7 I_k^{k+\frac{1}{p}} \right) + \\
 & + \frac{p^2}{h_k^2} \left(\Delta I_k^{k+\frac{1}{p}} - \Delta^2 I_k^{k+\frac{1}{p}} + \frac{11}{12} \Delta^3 I_k^{k+\frac{1}{p}} - \frac{10}{12} \Delta^4 I_k^{k+\frac{1}{p}} + \right. \\
 & \quad \left. + \frac{137}{180} \Delta^5 I_k^{k+\frac{1}{p}} - \frac{7}{10} \Delta^6 I_k^{k+\frac{1}{p}} + \frac{363}{560} \Delta^7 I_k^{k+\frac{1}{p}} \right) \cdot (x - x_k) + \\
 & + \frac{p^3}{h_k^3} \left(\frac{1}{2} \Delta^2 I_k^{k+\frac{1}{p}} - \frac{3}{4} \Delta^3 I_k^{k+\frac{1}{p}} + \frac{7}{8} \Delta^4 I_k^{k+\frac{1}{p}} - \right. \\
 & \quad \left. - \frac{15}{16} \Delta^5 I_k^{k+\frac{1}{p}} + \frac{29}{30} \Delta^6 I_k^{k+\frac{1}{p}} - \frac{469}{480} \Delta^7 I_k^{k+\frac{1}{p}} \right) \cdot (x - x_k)^2 + \\
 & + \frac{p^4}{h_k^4} \left(\frac{1}{6} \Delta^3 I_k^{k+\frac{1}{p}} - \frac{1}{3} \Delta^4 I_k^{k+\frac{1}{p}} + \frac{17}{36} \Delta^5 I_k^{k+\frac{1}{p}} - \frac{7}{12} \Delta^6 I_k^{k+\frac{1}{p}} + \frac{967}{1440} \Delta^7 I_k^{k+\frac{1}{p}} \right) (x - x_k)^3 + \\
 & + \frac{p^5}{h_k^5} \left(\frac{1}{24} \Delta^4 I_k^{k+\frac{1}{p}} - \frac{5}{48} \Delta^5 I_k^{k+\frac{1}{p}} + \frac{25}{144} \Delta^6 I_k^{k+\frac{1}{p}} - \frac{245}{1008} \Delta^7 I_k^{k+\frac{1}{p}} \right) (x - x_k)^4 + \\
 & + \frac{p^6}{h_k^6} \left(\frac{1}{120} \Delta^5 I_k^{k+\frac{1}{p}} - \frac{1}{40} \Delta^6 I_k^{k+\frac{1}{p}} + \frac{23}{480} \Delta^7 I_k^{k+\frac{1}{p}} \right) (x - x_k)^5 + \\
 & + \frac{p^7}{h_k^7} \left(\frac{1}{720} \Delta^6 I_k^{k+\frac{1}{p}} - \frac{7}{1440} \Delta^7 I_k^{k+\frac{1}{p}} \right) (x - x_k)^6 + \frac{p^8}{h_k^8} \frac{1}{5040} \Delta^7 I_k^{k+\frac{1}{p}} (x - x_k)^7,
 \end{aligned} \tag{4.63}$$

где $p = m + 1$.

Для того чтобы воспользоваться формулой (4.63) при $m < 7$, необходимо выбрать его конкретное численное значение, найти интегральные разности порядка m включительно и подставить их в формулу (4.63), в которой учитываются только слагаемые, имеющие конечно-интегральные разности до порядка m . Если $S_m(x)$ строится на всем отрезке $[a, b]$, то принимается $h_k = b - a$. Правило усечения формулы (4.63) легко устанавливается путем ее сопоставления с формулами (4.60)–(4.62), справедливыми при $m = 1, m = 2, m = 3$.

Подчеркнем, что в отличие от метода наименьших квадратов, метод наилучшего интегрального приближения при $m \leq 7$ может быть реализован по готовым формулам, и поэтому не требуется решать системы алгебраических уравнений относительно коэффициентов.

Методика решения задачи наилучшего интегрального приближения

1. Выбрать степень m (как правило, $m \leq 7$) сглаживающего многочлена и отрезок $[x_k, x_{k+m+1}]$, на котором он строится. Если многочлен строится на всем отрезке $[a, b]$, то выбрать $m = 7$.

2. Подсчитать конечно-интегральные разности до m -го порядка включительно.

3. Выписать сглаживающий многочлен по формуле (4.63).

Пример 4.12. Для задачи, решенной в примере 4.10 методом наименьших квадратов, построить сглаживающий многочлен наилучшего интегрального приближения.

□1. Сеточная функция, приведенная в таблице 4.16, задана с шагом $h = 1$, и погрешность ε_i составляет величину 0,1. Выберем степень многочлена $m = 3$. Поскольку $[x_0, x_n] = [x_0, x_4]$, то $n = 4$. Так как $n - m - 1 = 4 - 3 - 1 = 0$, то $k = 0$, а частичный отрезок $[x_0, x_{0+3+1}] = [x_0, x_4]$ совпадает с отрезком $[a, b] = [1; 5]$.

2. Рассчитаем интегралы $I_i^{i+\frac{1}{4}}$ ($i = 0, \dots, 4$) и конечно-интегральные разности до третьего порядка (см. рис. 4.9). Для нахождения интегралов используем простейшую квадратурную формулу трапеций

$$I_i^{i+\frac{1}{4}} = \frac{h}{2} \left(f_i + f_{i+\frac{1}{4}} \right).$$

Результаты расчетов $I_i^{i+\frac{1}{4}}$ и $\Delta^k I_i^{i+\frac{1}{4}}$ ($k = 1, 2, 3$) приведены в таблице 4.20.

Таблица 4.20

x_i	f_i	$I_i^{i+\frac{1}{4}}$	$\Delta I_i^{i+\frac{1}{4}}$	$\Delta^2 I_i^{i+\frac{1}{4}}$	$\Delta^3 I_i^{i+\frac{1}{4}}$
1	-0,070	0,3448			
2	0,760	0,8800	0,5352	-0,1522	-0,0063
3	1,000	1,2630	0,3830	-0,1585	
4	1,526		0,2245		
5	1,449	1,4875			

3. Подставляя значения конечно-интегральных разностей, представленных верхними значениями в колонках таблицы 4.19 и $h_k = 5 - 1 = 4$ в формулу (4.62), получаем сглаживающий полином

$$S_3(x) = 0,2804 + 0,0682 \cdot (x - 1) - 0,07137 \cdot (x - 1)^2 - 0,00105 \cdot (x - 1)^3.$$

Рассчитывая среднеквадратичную погрешность для сглаживающего многочлена, получаем значение $\delta = 0,11$, которое приблизительно равно ε . Это свидетельствует об обоснованности применения метода сглаживания многочленом наилучшего интегрального приближения. Отличие многочлена $S_3(x)$ от самой функции $f(x) = \ln x$ на отрезке $[1; 5]$ не превышает 0,064. ■

4.5. МЕТОДЫ ИНТЕРПОЛЯЦИИ И СГЛАЖИВАНИЯ НА ОСНОВЕ СПЛАЙНОВ

4.5.1. ПОСТАНОВКА ЗАДАЧИ И ОСНОВНЫЕ ПОЛОЖЕНИЯ

Как отмечалось в п. 4.1, в вычислительной практике при глобальном способе аппроксимации высокие степени интерполяционных многочленов использовать нецелесообразно. Поэтому часто применяется кусочно-глобальный способ на основе, например, линейной и квадратичной (параболической) интерполяции. Однако производные таких интерполяционных многочленов, получающихся на частичных отрезках, в местах их стыка терпят разрывы. На рисунке 4.10 a показано поведение совокупности трех таких параболических многочленов (звеньев). Легко видеть, что производные в точках стыка частичных отрезков (шаблонов (x_i, x_{i+1}, x_{i+2}) , $(x_{i+2}, x_{i+3}, x_{i+4})$, $(x_{i+4}, x_{i+5}, x_{i+6})$) терпят разрывы. Это является крайне нежелательным, если вся совокупность $\bigcup_i S_{2,i}(x)$ далее используется, например, для аппроксимации обводов конструктивных элементов сложных поверхностей. Для аппроксимации таких поверхностей требуется задание в узлах не только значений

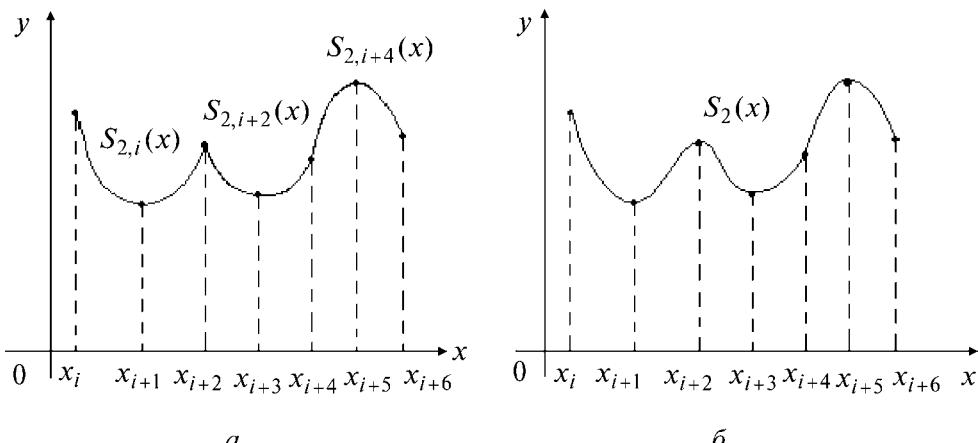


Рис. 4.10

$S_{k,i}(x_i)$, но и первой производной $S'_{k,i}(x_i)$ или даже второй производной $S''_{k,i}(x_i)$. Таким образом, вместе с функцией $y_i = f(x_i)$, $i=0, n$, требуется достаточно хорошо аппроксимировать и ее производные. Если восполняемая функция достаточно гладкая, то и ее производные тоже гладкие, и поэтому они должны быть непрерывными во всех внутренних узлах. В связи с этим к аппроксимирующему функциям $S_m(x)$ (m — степень многочлена) предъявляются требования непрерывности $S_m(x)$, $S'_m(x)$ и по возможности $S''_m(x)$ или $S'''_m(x)$ во всех точках $x \in [a, b]$ (рис. 4.10б). Это и обуславливает необходимость построения сплайн-функций, обладающих указанными свойствами и имеющих интерполяционный или сглаживающий характер.

Пусть на отрезке $[a, b]$ задана сеточная функция $y_i = f(x_i)$ на сетке Ω_n . Требуется восполнить ее функцией $S_m(x)$, где m — степень многочлена, кусочно-глобальным способом.

Дадим сначала упрощенное определение сплайна, которое затем уточним.

Сплайн-функцией или *сплайном* называется совокупность $S_{m,i}(x)$ алгебраических многочленов степени m (звеньев), определенных на частичных отрезках $[x_i, x_{i+1}]$, $i = 0, n-1$, и соединенных вместе по всем частичным отрезкам так, чтобы можно было составить многозвенную функцию $S_m(x) = \bigcup_{i=0}^{n-1} S_{m,i}(x)$, определенную и непрерывную на всем отрезке $[a, b]$ вместе со всеми своими производными $S_m^{(p)}(x)$ до некоторого их порядка $p = 1, 2, \dots$

Разность между m и наибольшим порядком производной, непрерывной на отрезке $[a, b]$, определяет *дефект сплайна* q .

Условиями согласования звеньев $S_{m,i}(x)$ сплайна с исходной функцией $y_i = f(x_i)$ на соответствующем частичном отрезке $[x_i, x_{i+1}]$ называются условия, накладываемые на невязки дифференциального и интегрального типов $\delta S_{m,i}^{(p)}(x_i)$, $\delta S_{m,i}(I_i^{i+1})$ и использующиеся для вывода формулы одного звена сплайна на указанном частичном отрезке.

Сплайн, удовлетворяющий условиям (4.3) нулевых функциональных невязок, называется *интерполяционным*, а удовлетворяющий только интегральным условиям (4.5), — *сглаживающим* или *интегрально-сглаживающим*.

Сплайн, построение которого основано только на дифференциальных условиях согласования (4.2) при $p = 0, 1$ или $p = 0, 2$, называется *дифференциальным* (*D-сплайном*). Если же наряду с дифференциальными условиями согласования используется интегральное условие вида (4.5), то сплайн называется *интегрально-дифференциальным* (*ID-сплайном*). Сплайны такого типа впервые предложены в [16–23].

Количество условий согласования, необходимых для получения формулы одного звена сплайна, должно соответствовать степени сплайна (число условий на единицу больше m). После определения формулы $S_{m,i}(x)$ ее правая часть выражается через известные и неизвестные величины (параметры): $f^{(p)}(x_i)$ — для дифференциального сплайна или через совокупность $f^{(p)}(x_i)$

и $I_i^{i+1} = \int_{x_i}^{x_{i+1}} f(x) dx$ — для интегрально-дифференциального. Эти величины

будут называться *параметрами* сплайна. В зависимости от того, заданы те или иные параметры сплайна в постановке задачи или нет, они именуются *определенными* или *неопределенными*. Последние вычисляются на основе условий непрерывности (гладкости) сплайна $S_m(x)$, которые здесь называются *условиями стыковки*. Для некоторого узла x_i , общего для звена $S_{m,i-1}(x)$, относящегося к частичному отрезку $[x_{i-1}, x_i]$, и звена $S_{m,i}(x)$, определенного на частичном отрезке $[x_i, x_{i+1}]$, это условие имеет вид

$$S_{m,i-1}^{(p)}(x)|_{x=x_i} = S_{m,i}^{(p)}(x)|_{x=x_i}.$$

При решении задачи аппроксимации с помощью сплайн-функций условия стыковки преобразуются к соотношениям, связывающим определенные и неопределенные параметры, которые называются *параметрическими соотношениями* (уравнениями).

Параметрические соотношения, записанные в виде уравнений, могут использоваться для нахождения неопределенных параметров применительно к сплайн-аппроксимации либо для выражения одних параметров через другие. Последний способ в данной книге применяется для вывода новых типов формул численного дифференцирования и интегрирования (см. главу 5). Кроме того, параметрические соотношения на основе изложенного во введении принципа согласования порядков численных величин устанавливают соответствие этих порядков для различных параметров, входящих в то или иное соотношение.

При решении задачи сплайн-аппроксимации необходимо предварительно вычислить неопределенные параметры сплайна для всех его звеньев. Этот этап соответствует решению *параметрической задачи*. Для пояснения ее характера введем понятие *порядка параметров* сплайн-функции, который обозначим через l . Этот порядок зависит от порядка производных исходной функции $y = f(x)$: для параметра I_i^{i+1} примем $l = -1$, а для $f^{(p)}(x_i)$ положим $l = p$ ($p = 0, 1, 2, \dots$).

Параметрическая задача называется *прямой (обратной)*, если порядки всех неопределенных параметров больше (меньше) порядков определенных параметров. Если в задаче аппроксимации порядки одних неопределенных параметров больше, а порядки других меньше порядков определенных параметров, то параметрическая задача называется *смешанной*. Данная классификация используется далее при описании конкретных методов построения параболических интегрально-дифференциальных сплайн-функций.

Две совокупности параметров, относящиеся к параметрическим уравнениям и сплайнам разных степеней, называются *подобными*, если они могут быть получены друг из друга путем соответствующего изменения порядков параметров l на одинаковую величину.

Параметрические уравнения или их системы, соответствующие двум сплайнам разных степеней, называются *подобными*, если они включают подобные параметры и имеют одинаковые коэффициенты, зависящие от шагов, определяющих сетку.

Два или несколько сплайнов различных степеней называются *подобными*, если формулы звеньев можно получить друг из друга путем дифференци-

рования или интегрирования и замены всех параметров одного сплайна соответствующими подобными параметрами другого сплайна.

Сплайн называется *локальным*, если все неопределенные параметры, относящиеся к каждому его звену $S_{m,t}(x)$, $t = i$, при $x \in [x_i, x_{i+1}]$, $i = 0, n-1$, находятся локально, т. е. независимо от параметров, характеризующих все остальные (или почти все остальные) звенья $S_{m,t}(x)$ ($t = 0, 1, \dots, i-1, i+1, \dots, n-1$).

Сплайн-аппроксимация (как правило, это интерполяция) локальными сплайнами сводится, таким образом, к получению конкретных значений коэффициентов многочленов $S_{m,i}(x)$ для каждого звена результирующего сплайна $S_m(x)$ и неопределенных параметров путем их вычисления по аппроксимационным формулам. Недостаток локальной интерполяции состоит в том, что таким способом не удается обеспечить минимальный дефект сплайна, т. е. максимально возможную его гладкость в смысле непрерывности производных как можно большего порядка ($p = 0, 1, 2, \dots$).

Альтернативой локальному сплайну является *глобальный сплайн*, для которого неопределенные параметры, относящиеся к каждому его звену $S_{m,t}(x)$, $t = i$, при $x \in [x_i, x_{i+1}]$, $i = 0, n-1$, находятся совместно с параметрами, характеризующими все остальные звенья $S_{m,t}(x)$, $t \neq i$. Неопределенные параметры в глобальных сплайнах для всех звеньев вычисляются, как правило, путем решения системы линейных алгебраических уравнений трехдиагонального вида методом прогонки, причем это обеспечивает непрерывность одной из производных, а именно той, которая не включена в условия согласования. Если в условие согласования включена производная другого порядка, то ее непрерывность также гарантируется. Глобальный способ аппроксимации по сравнению с локальным способом обеспечивает минимально возможный дефект сплайна. Поэтому глобальные сплайны широко используются в вычислительной практике.

В заключение данного раздела дадим развернутое определение глобального интегрально-дифференциального сплайна.

Функция $S_m(x) = \bigcup_{i=0}^{n-1} S_{m,i}(x)$, которая определена на отрезке $[a, b]$, принадлежит классу гладкости $C_r[a, b]$ и составлена из совокупности звеньев $S_{m,i}(x)$, $i = 0, n-1$, определенных на каждом частичном отрезке $[x_i, x_{i+1}]$ сетки Ω_n , называется алгебраическим интегрально-дифференциальным сплайном степени m и дефекта q ($0 \leq r \leq m$, $q = m - r$) с узлами на сетке Ω_n , если каждое его звено $S_{m,i}(x)$ $x \in [x_i, x_{i+1}]$, $i = 0, n-1$, представляется в виде алгебраического многочлена степени m :

$$S_{m,i}(x) = \sum_{k=0}^m a_{k,i}(x - x_i)^k, \quad (4.64)$$

с коэффициентами $a_{k,i}$, выражаемыми из совокупности $(m+1)$ интегральных и (или) дифференциальных условий согласования:

$$\delta S_{m,i}^{-1}(x_i, x_{i+1}) = \int_{x_i}^{x_{i+1}} [S_{m,i}(x) - f(x)] dx = 0; \quad (4.65)$$

$$\delta S_{m,i}^{(p_1)}(x_j) = S_{m,i}^{(p_1)}(x)|_{x=x_j} - f^{(p_1)}(x)|_{x=x_j} = 0, \quad j = i, i+1, \quad (4.66)$$

и из условий стыковки (непрерывности) звеньев $S_{m,i}(x)$ по производным $S_{m,i}^{(p_2)}$:

$$S_{m,i-1}^{(p_2)}(x)|_{x=x_i} = S_{m,i}^{(p_2)}(x)|_{x=x_i}, \quad i = \overline{1, n-1}. \quad (4.67)$$

Здесь p_1 ($0 \leq p_1 \leq r$) принимает значения из совокупности нескольких целых чисел, соответствующих порядкам производных, а p_2 — целое число или в общем случае несколько целых чисел, таких, что $0 \leq p_2 \leq r$, причем ни одно из этих чисел не равно ни одному значению чисел p_1 .

Замечания

1. Данное определение является обобщенным и справедливо как для интегрально-дифференциального сплайна, так и для классического дифференциального. В последнем случае условия (4.66) не используются.
2. Условия стыковки (4.67) вместе с дифференциальными условиями согласования (4.66) обеспечивают непрерывность $S_m^{(p)}(x)$, $p = 0, 1, \dots, r$, во всех внутренних узлах x_i , определяющих точки стыковки звеньев, т. е. $(x_i, S_m(x_i))$, $(i = 1, n-1)$. Это гарантирует выполнение условия $S_m(x) \in C_r[a, b]$.

3. Алгоритм конструирования сплайна сочетает в себе два способа аппроксимации — кусочный и глобальный, и поэтому является кусочно-глобальным. На кусочном способе основан процесс отыскания звеньев сплайна $S_{m,i}(x)$, т. е. коэффициентов $a_{k,i}$ многочлена (4.64), а на глобальном — соединение всех звеньев в единую (многозвенную) функцию $S_m(x)$ ($x \in [a, b]$) со стыковкой звеньев так, чтобы обеспечивалась непрерывность $S_m^{(p)}(x)$ при $p = 0, 1, 2, \dots$, т. е. непрерывность как самой функции $S_m(x)$, так и ее производных.

4. Порядки производных p_1 , выбранных для условий согласования (4.66), не должны совпадать с порядками производных, используемых в условиях стыковки (4.67), но вся совокупность порядков производных из p_1 и p_2 должна обеспечивать выполнение условия $S_m(x) \in C_r[a, b]$.

5. Интегральное условие согласования (4.65) может быть переписано в виде $\int_{x_i}^{x_{i+1}} S_{m,i}(x) dx = I_i^{i+1}$, и тогда I_i^{i+1} вместе со значениями $f^{(p_1)}(x_i)$, входящими в дифференциальное условие согласования, причисляется к параметрам ИД-сплайна. Данные параметры могут быть заданы при формулировке задачи аппроксимации либо определены при ее решении.

Рассмотрим сначала кубические дифференциальные сплайны, являющиеся традиционными и широко используемыми при интерполяционном восполнении сеточных функций [1], [2], [13], [14], [25], [44].

4.5.2. ИНТЕРПОЛЯЦИОННЫЕ ДИФФЕРЕНЦИАЛЬНЫЕ КУБИЧЕСКИЕ СПЛАЙНЫ

Рассмотрим задачу восполнения заданной сеточной функции $y_i = f(x_i)$, $i = 0, n$, $x_i \in [a, b]$ на базе интерполяционных глобальных кубических дифференциальных сплайнов дефекта $q = 1$, т. е. $S_3(x) \in C_2[a, b]$, которая изучена в [1], [2], [13], [25] и др. работах. При этом предположим, что восполняемая функция достаточно гладкая.

Решить эту задачу можно с помощью двух алгоритмов восполнения исходной сеточной функции, различающихся выбором порядков производных, на основе которых записываются условия согласования (4.66). Первый способ, наиболее широко распространенный, соответствует выбору вторых производных $m_i = f_i''$ ($i = \overline{0, n}$), а второй — выбору первых производных $\bar{m}_i = \bar{f}_i'$.

Оба способа основаны на решении параметрически прямой задачи и весьма близки по характеру вычислительных процедур.

Основные этапы реализации первого способа.

Первый этап. Для алгебраического многочлена (полинома)

$$S_{3,i}(x) = a_{0,i} + a_{1,i}(x - x_i) + a_{2,i}(x - x_i)^2 + a_{3,i}(x - x_i)^3, \quad (4.68)$$

относящегося к одному i -му звену сплайна, для которого $x \in [x_i, x_{i+1}]$, выбираются два дифференциальных условия согласования (4.66) с порядком $p_1 = \{0; 2\}$, каждое из которых относится к концам отрезка $[x_i, x_{i+1}]$:

$$\begin{aligned} \delta S_{3,i}(x)|_{x=x_i} &= S_{3,i}(x_i) - f(x_i) = 0, \\ \delta S_{3,i}(x)|_{x=x_{i+1}} &= S_{3,i}(x_{i+1}) - f(x_{i+1}) = 0; \end{aligned} \quad (4.69)$$

$$\begin{aligned} \delta S_{3,i}''(x)|_{x=x_i} &= S_{3,i}''(x_i) - f''(x_i) = 0, \\ \delta S_{3,i}''(x)|_{x=x_{i+1}} &= S_{3,i}''(x_{i+1}) - f''(x_{i+1}) = 0. \end{aligned} \quad (4.70)$$

Выбор значения $p = 0$ обусловлен тем, что отыскивается $S_3(x)$ — интерполяционный многочлен, а выбор производных порядка $p = 2$ зависит от выбора способа. Условия (4.69), (4.70) задают параметры сплайна f_i , $m_i = f''(x_i)$, $i = \overline{0, n}$, первые из которых (функциональные) являются определенными, а вторые (дифференциальные) — неопределенными.

Подстановка в соотношения (4.69), (4.70) полинома (4.68) приводит к системе четырех линейных алгебраических уравнений относительно неизвестных коэффициентов $a_{k,i}$ ($k = 0, 1, 2, 3$). Разрешая их, подставляя полученные формулы для $a_{k,i}$ в (4.68) и распространяя этот результат на все частичные отрезки, имеем:

$$\begin{aligned} S_{3,i}(x) &= f_i + \left(\frac{1}{h_{i+1}} \Delta f_i - \frac{h_{i+1}}{2} m_i - \frac{h_{i+1}}{6} \Delta m_i \right) \cdot (x - x_i) + \\ &+ \frac{m_i}{2} \cdot (x - x_i)^2 + \frac{1}{6h_{i+1}} \Delta m_i \cdot (x - x_i)^3, \quad i = \overline{1, n-1}, \end{aligned} \quad (4.71)$$

где $h_{i+1} = x_{i+1} - x_i$, $\Delta f_i = f_{i+1} - f_i$, $\Delta m_i = m_{i+1} - m_i$.

Таким образом, найдена общая формула одного звена сплайна, которая выражается через определенные (f_i) и неопределенные (m_i) параметры ($i = \overline{0, n}$). Эти параметры обеспечивают непрерывность сплайна: $S_{3,i-1}(x)|_{x=x_i} = S_{3,i}(x)|_{x=x_i}$, $i = \overline{1, n-1}$, и его вторых производных $S_{3,i-1}''(x)|_{x=x_i} = S_{3,i}''(x)|_{x=x_i}$, $i = \overline{1, n-1}$, для всех внутренних узлов.

Второй этап. При вычислении неопределенных параметров m_i ($i = \overline{0, n}$), входящих в (4.71), с учетом условия стыковки (4.67) для всех внутренних узлов устанавливается связь между f_i и m_i . Эта связь получается путем записи правой части (4.71) для звена $S_{3,i-1}(x)$ при $x \in [x_{i-1}, x_i]$, ее дифференцирования

и приравнивания производной правой части (4.71) для звена $S_{3,i}(x)$ при $x \in [x_i, x_{i+1}]$ в соответствии с условием стыковки (4.67) ($p_2 = 1$).

Проведя несложные выкладки и распространив найденное соотношение на все внутренние узлы, получим параметрическую систему — систему линейных алгебраических уравнений трехдиагонального вида, устанавливающую линейную связь комбинаций определенных f_i и неопределенных параметров m_i во внутренних узлах:

$$\frac{h_i}{6}m_{i-1} + \frac{h_i + h_{i+1}}{3}m_i + \frac{h_{i+1}}{6}m_{i+1} = \frac{\Delta f_i}{h_{i+1}} - \frac{\Delta f_{i-1}}{h_i}, \quad i = \overline{1, n-1}. \quad (4.72)$$

Данная система относительно m_i ($i = \overline{0, n}$) является незамкнутой, так как не хватает двух уравнений. Для ее замыкания используют различные пути выбора аппроксимаций производных на концах (граничных условий):

1. В простейшем случае вторые производные на концах (краевые условия) принимаются нулевыми, т. е.

$$m_0 = 0; m_n = 0.$$

Такие краевые условия называются условиями *натурального сплайна*.

Из рисунка 4.11 видно, что задание условий $m_0 = 0$ и $m_n = 0$ приводит к разрыву вторых производных $f''(x)|_{x=x_0}$ и $f''(x)|_{x=x_n}$ на концах отрезка $[a, b]$, что может вызвать возрастание погрешности интерполяции.

2. Для первых двух и последних двух отрезков можно применить условие равенства третьей производной. Тогда для этих отрезков получается соотношение $\Delta m_{k-1}/h_k = \Delta m_k/h_{k+1}$ ($k = 1, n-1$). Отсюда следуют два трехточечных уравнения:

$$\begin{aligned} \frac{m_0}{h_1} - \left(\frac{1}{h_1} + \frac{1}{h_2} \right) m_1 + \frac{m_2}{h_2} &= 0; \\ \frac{m_{n-2}}{h_{n-1}} - \left(\frac{1}{h_{n-1}} + \frac{1}{h_n} \right) m_{n-1} + \frac{m_n}{h_n} &= 0. \end{aligned} \quad (4.73)$$

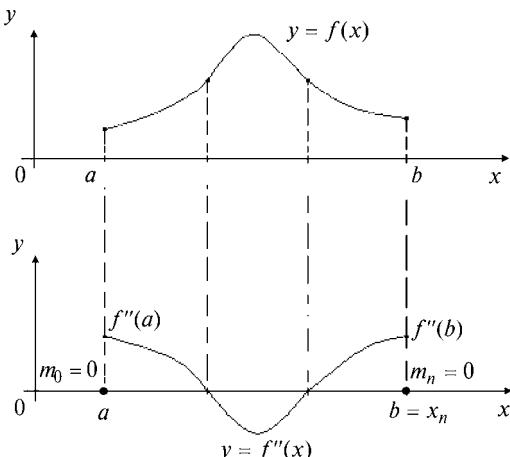


Рис. 4.11

Соотношения (4.73) позволяют замкнуть систему (4.72) при $h = \text{var}$ (неравномерная сетка). При $h = \text{const}$ вторые производные на концах могут быть вычислены по явным формулам второго порядка аппроксимации [6]:

$$\begin{aligned} m_0 &= \frac{1}{h^2} (2f_0 - 5f_1 + 4f_2 - f_3); \\ m_n &= \frac{1}{h^2} (-f_{n-3} + 4f_{n-2} - 5f_{n-1} + 2f_n). \end{aligned} \quad (4.74)$$

Отметим, что как (4.73), так и формулы (4.74) для m_0 и m_n имеют порядки аппроксимации, соответствующие порядку сходимости $S_3(x)$ к $f(x)$ (см. п. 2 замечаний ниже).

Третий этап. Определяются значения m_i ($i = \overline{0, n}$) путем решения систем (4.72), (4.73) (параметрически прямая задача). Для этого используется метод прогонки, причем необходимо предварительно из первых двух и последних двух уравнений исключить слагаемые с m_2 и m_{n-2} соответственно. В случае равномерной сетки, как отмечено выше, система (4.72) замыкается аппроксимацией (4.74).

Четвертый этап. Формируются массивы коэффициентов $a_{0,i}$, $a_{1,i}$, $a_{2,i}$, $a_{3,i}$ для всех звеньев сплайна ($i = \overline{0, n-1}$), и определяется глобальный интерполяционный сплайн $S_3(x) = \bigcup_{i=0}^{n-1} S_{3,i}(x)$ путем составления многозвенной функции.

Пятый этап. При необходимости полученная сплайн-функция $S_3(x)$ используется для вычисления значений функции и производных порядка p : $f^{(p)}(x) = S_3^{(p)}(x) + O(h_{i+1}^{4-p})$ ($p = 0, 1, 2, \dots$) в произвольных точках $x \in [x_i, x_{i+1}]$, $i = \overline{0, n-1}$, или определенных интегралов $I_a^b = \int_a^b S_3(x) dx$ (в приложениях теории приближений).

Основные этапы реализации второго способа.

Первый этап. Для алгебраического многочлена (4.68), обозначаемого $\bar{S}_{3,i}(x)$, вместо вторых производных в (4.70) используются первые производные. Тогда эти условия определяют параметры сплайна f_i , $\bar{m}_i = f'(x_i)$, $i = \overline{0, n}$, первые из которых (функциональные) являются определенными, а вторые (дифференциальные) — неопределенными ($p = 1$). Проделав выкладки (аналогично первому этапу первого способа), получим общую формулу i -го звена искомого сплайна, подобную формуле (4.37), справедливой для звена $\bar{S}_{2,i}(x)$:

$$\begin{aligned} \bar{S}_{3,i}(x) &= f_i + \bar{m}_i \cdot (x - x_i) + \left(\frac{3}{h_{i+1}^2} \Delta f_i - \frac{3}{h_{i+1}} \bar{m}_i - \frac{1}{h_{i+1}} \Delta \bar{m}_i \right) \cdot (x - x_i)^2 + \\ &+ \frac{1}{h_{i+1}^2} \left(-\frac{2}{h_{i+1}} \Delta f_i + 2\bar{m}_i + \Delta \bar{m}_i \right) (x - x_i)^3, \quad \Delta \bar{m}_i = \bar{m}_{i+1} - \bar{m}_i, \end{aligned} \quad (4.75)$$

которая зависит от определенных (f_i) и неопределенных ($\bar{m}_i = f'$) параметров сплайна. Данные параметры обеспечивают непрерывность всего сплайна и его производных $\bar{S}_{3,i-1}(x)|_{x=x_i} = \bar{S}_{3,i}(x)|_{x=x_i}$, $\bar{S}'_{3,i-1}(x)|_{x=x_i} = \bar{S}'_{3,i}(x)|_{x=x_i}$ во всех внутренних узлах ($i = 1, n-1$).

Второй этап. С использованием условийстыковки (4.67) по вторым производным ($p_2 = 2$) во внутренних узлах получается система линейных алгебраических уравнений трехдиагонального типа, выражающая параметрическую связь между \bar{m}_i и f_i :

$$\frac{1}{h_i} \bar{m}_{i-1} + 2\left(\frac{1}{h_i} + \frac{1}{h_{i+1}}\right) \bar{m}_i + \frac{1}{h_{i+1}} \bar{m}_{i+1} = 3\left(\frac{\Delta f_{i-1}}{h_i^2} + \frac{\Delta f_i}{h_{i+1}^2}\right), \quad i = \overline{1, n-1}. \quad (4.76)$$

Замыкание системы (4.76) осуществляется аналогично пункту 2 на втором этапе реализации первого способа. При $h = \text{var}$ получаются два соотношения (граничные условия), не нарушающие трехдиагональность системы:

$$\begin{aligned} \bar{m}_0 + \frac{H_1^2}{h_2} \bar{m}_1 &= \frac{h_1^2}{H_1^2} \left(\frac{2H_1^2 + h_1}{h_1^3} \Delta f_0 + \frac{1}{h_2^2} \Delta f_1 \right); \\ \frac{H_{n-1}^n}{h_{n-1}} \bar{m}_{n-1} + \bar{m}_n &= \frac{h_n^2}{H_{n-1}^n} \left(\frac{1}{h_{n-1}^2} \Delta f_{n-2} + \frac{2H_{n-1}^n + h_n}{h_n^3} \Delta f_{n-1} \right), \end{aligned} \quad (4.77)$$

где $H_{k-1}^k = h_{k-1} + h_k$ ($k = 2, n$).

Соотношения (4.77) аппроксимируют производные на концах с третьим порядком точности, что соответствует четвертому порядку сходимости $\bar{S}_3(x)$ к $f(x)$ (см. п. 2 замечаний ниже). При $h = \text{const}$ для определения значений производных на концах можно использовать соответствующие аппроксимационные формулы (см. главу 5).

Третий этап. Трехдиагональная система линейных алгебраических уравнений (4.76), (4.77) при конструировании глобального сплайна решается методом прогонки (параметрически прямая задача).

Четвертый и пятый этапы аналогичны изложенным для первого способа.

Замечания

1. Существование и единственность глобальных сплайнов $S_3(x)$ и $\bar{S}_3(x)$ следуют из рассмотренного выше построения общих формул их звеньев $S_{3,i}(x)$, $\bar{S}_{3,i}(x)$ и единственности решения трехдиагональных систем (4.72), (4.73) или (4.74) и (4.76), (4.77) в силу выполнения условия диагонального преобладания.

2. Сходимость процесса построения глобальных сплайнов $S_{3,i}(x)$, $\bar{S}_{3,i}(x)$ доказывается для непрерывных (формульных) функций, таких, что $f(x) \in C_4[a, b]$. Процесс является *сходящимся*, если при неограниченном увеличении числа n узлов сетки соответствующая последовательность сплайн-функций, построенных на этих сетках, сходится к исходной функции $f(x)$. Доказано, что на равномерной сетке сплайн-функции $S_{3,i}(x)$, $\bar{S}_{3,i}(x)$ сходятся к $f(x) \in C_4[a, b]$ с четвертым порядком, причем справедливы оценки

$$\|f^{(p)}(x) - S_3^{(p)}(x)\|_{C[a,b]} = \max_{[a,b]} |f^{(p)}(x) - S_3^{(p)}(x)| \leq M_4 h^{4-p}, \quad p = 0, 1, 2. \quad (4.78)$$

Таким образом, при использовании сплайн-функций для вычисления производных также реализуется сходимость, но ее порядок понижается на величину p .

3. В соответствии с принципом соответствия порядков аппроксимации (см. п. В.4) точность вычисления интегралов на частичном отрезке $[x_i, x_{i+1}]$, т. е. $I_i^{i+1} = \int_{x_i}^{x_{i+1}} f(x)dx \approx \hat{I}_i^{i+1} = \int_{x_i}^{x_{i+1}} S_3(x)dx$, на основе сплайнов $S_{3,i}(x)$, $\bar{S}_{3,i}(x)$ составляет $O(h_{i+1}^5)$, а интеграла $I_a^b = \int_a^b f(x)dx$ на отрезке $[a, b]$ составляет $O(H^4)$, $H = \max_i h_{i+1}$ (потеря порядка происходит из-за суммирования значений \hat{I}_i^{i+1} и остаточного слагаемого при переходе от \hat{I}_i^{i+1} , $i = \overline{0, n-1}$, к $\hat{I}_a^b = \sum_{i=0}^{n-1} \hat{I}_i^{i+1}$).

4. Выражая из системы (4.72) комбинацию значений функций f_{i-1}, f_i, f_{i+1} через комбинацию m_{i-1}, m_i, m_{i+1} , получаем трехдиагональную систему линейных алгебраических уравнений относительно f_i :

$$\begin{aligned} & \frac{1}{h_i} f_{i-1} - \left(\frac{1}{h_i} + \frac{1}{h_{i+1}} \right) f_i + \frac{1}{h_{i+1}} f_{i+1} = \\ & = \frac{h_i}{6} m_{i-1} + \frac{h_i + h_{i+1}}{3} m_i + \frac{h_{i+1}}{6} m_{i+1}, \quad i = \overline{1, n-1}. \end{aligned} \quad (4.79)$$

Если считать, что вторые производные для некоторой функции $y_i = f(x_i)$ заданы или каким-либо способом определены, а f_0 и f_n на концах $[a, b]$ известны, то из (4.79) можно восстановить значения \tilde{f}_i во внутренних узлах x_i ($i = \overline{1, n-1}$) путем решения параметрически обратной задачи. В этом случае сплайн $S_3(x)$, построенный по \tilde{f}_i, m_i ($i = \overline{0, n}$) и (4.71), является в строгом понимании не интерполяционным, а восстанавливающим функцию $y_i = f(x_i)$. Единственность решения задачи обеспечивается преобладанием диагональных коэффициентов в системе (4.79) относительно значений f_i ($i = \overline{0, n}$).

5. На основе соотношения (4.76) при заданных двух значениях функций f_0, f_1 или f_{n-1}, f_n и всех значениях m_i ($i = \overline{0, n}$) можно восстановить функцию $\tilde{f}(x_i)$ во всех точках сетки x_i ($i = \overline{2, n}$). С этой целью система (4.76) разрешается относительно приращений функций Δf_i , которые затем вычисляются рекуррентно. При задании неравномерной сетки устойчивость процесса достигается при регулярном сгущении сетки в направлении слева направо $\left(\delta_{i+1} = \frac{h_{i+1}}{h_i} \leq 1 \right)$ или при ее регулярном разрежении ($\delta_{i+1} > 1$). В первом случае восстановление осуществляется слева направо при заданных f_0, f_1 :

$$\Delta f_i = -\delta_{i+1}^2 \Delta f_{i-1} + L_h(\bar{m}_{i-1}, \bar{m}_i, \bar{m}_{i+1}),$$

где $L_h = \frac{1}{3} \left[\frac{1}{h_i} \bar{m}_{i-1} + 2 \cdot \left(\frac{1}{h_i} + \frac{1}{h_{i+1}} \right) \cdot \bar{m}_i + \frac{1}{h_{i+1}} \bar{m}_{i+1} \right]$, а во втором — справа налево при заданных f_n, f_{n-1} . Такой алгоритм обеспечивает устойчивость процесса обратной прогонки.

6. Локальный способ построения интерполяционных кубических сплайнов для заданной функции $y_i = f(x_i)$ основан на многочлене (4.75) и состоит в определении \bar{m}_i , $i = \overline{0, n}$, по аппроксимационным формулам третьего порядка

ка (см. главу 5 и [6]), в дальнейшей их подстановке в (4.75) и в формировании $\bar{S}_3(x) = \bigcup_{i=0}^{n-1} \bar{S}_{3,i}(x)$. Необходимость аппроксимации третьего порядка следует из принципа соответствия порядков, изложенного в п. В.4, а также из параметрической связи (4.76).

Методика построения интерполяционного кубического сплайна (первый способ)

1. Используя систему (4.72), условия $m_0 = 0$, $m_n = 0$ или (4.73) (при $h = \text{const}$ можно использовать условия (4.74)), сформировать замкнутую систему относительно неопределенных параметров.

2. Вычислить коэффициенты системы (4.72), зависящие от шагов сетки Ω_n , и решить ее методом прогонки. В результате найти неопределенные параметры m_0, m_1, \dots, m_n .

3. Записать выражения для звеньев сплайна по формуле (4.71):

$$S_{3,i}(x) = f_i + \left(\frac{1}{h_{i+1}} \Delta f_i - \frac{h_{i+1}}{2} m_i - \frac{h_{i+1}}{6} \Delta m_i \right) \cdot (x - x_i) + \\ + \frac{m_i}{2} \cdot (x - x_i)^2 + \frac{1}{6h_{i+1}} \Delta m_i \cdot (x - x_i)^3, \quad i = \overline{0, n-1},$$

вычислить коэффициенты звеньев и сформировать многозвенную сплайн-функцию $S_3(x) = \bigcup_{i=0}^{n-1} S_{3,i}(x)$, являющуюся сплайном дефекта $q = 1$.

Пример 4.13. Для сеточной функции $y_i = \cos x_i$, $x_i = 0, \frac{\pi}{6}, \frac{\pi}{3}, \frac{\pi}{2}$ ($i = \overline{0, 3}$) найти интерполяционную кубическую сплайн-функцию $S_3(x)$ первым способом.

В поставленной задаче $x_0 = 0$; $x_1 = \frac{\pi}{6}$; $x_2 = \frac{\pi}{3}$; $x_3 = \frac{\pi}{2}$, $n = 3$, $h = \pi/6 = \text{const}$; $f_0 = f(x_0) = 1$; $f_1 = f(x_1) = 0,8660254$; $f_2 = f(x_2) = 0,500000$; $f_3 = f(x_3) = 0$.

1. Запишем систему (4.72) для внутренних узлов:

$$\frac{h_1}{6} m_0 + \frac{h_1 + h_2}{3} m_1 + \frac{h_2}{6} m_2 = \frac{f_2 - f_1}{h_2} - \frac{f_1 - f_0}{h_1}, \\ \frac{h_2}{6} m_1 + \frac{h_2 + h_3}{3} m_2 + \frac{h_3}{6} m_3 = \frac{f_3 - f_2}{h_3} - \frac{f_2 - f_1}{h_2}.$$

В силу того, что $h = h_1 = h_2 = h_3 = \pi/6 = \text{const}$, для нахождения значений m_0, m_3 выбираются формулы (4.74):

$$m_0 = \frac{1}{h^2} (2f_0 - 5f_1 + 4f_2 - f_3), \\ m_3 = \frac{1}{h^2} (-f_0 + 4f_1 - 5f_2 + 2f_3).$$

2. Вычислим коэффициенты системы и значения m_0 , m_3 :

$$m_0 = \frac{36}{\pi^2} (2f_0 - 5f_1 + 4f_2 - f_3) = \frac{36}{\pi^2} (2 \cdot 1 - 5 \cdot 0,8660254 + 4 \cdot 0,5 + 0) = -1,2041589;$$

$$\frac{\pi}{36} m_0 + \frac{2\pi}{18} m_1 + \frac{\pi}{36} m_2 = \frac{6}{\pi} (f_2 - 2f_1 + f_0);$$

$$\frac{\pi}{36} m_1 + \frac{2\pi}{18} m_2 + \frac{\pi}{36} m_3 = \frac{6}{\pi} (f_3 - 2f_2 + f_1);$$

$$m_3 = \frac{36}{\pi^2} (-f_0 + 4f_1 - 5f_2 + 2f_3) = \frac{36}{\pi^2} = (-1 + 4 \cdot 0,8660254 - 5 \cdot 0,5 + 0) = -0,1309416.$$

Решение этой системы методом прогонки (см. п. 1.2.2): $m_1 = -0,84642$; $m_2 = -0,4886807$.

3. Подставляя параметры m_i , $i = 0, 1, 2, 3$, в формулу звена сплайна $S_{3,i}(x)$, получаем следующую трехзвенную сплайн-функцию:

$$S_3(x) = \begin{cases} 1 + \left(-\frac{6}{\pi} \cdot 0,13397 + \frac{\pi}{12} \cdot 1,20416 - \frac{\pi}{36} \cdot 0,35774 \right) \cdot (x - 0) - \\ - \frac{1,20416}{2} \cdot (x - 0)^2 + \frac{1}{\pi} \cdot 0,35774 \cdot (x - 0)^3 & \text{при } x \in \left[0; \frac{\pi}{6} \right]; \\ 0,86602 + \left(-\frac{6}{\pi} \cdot 0,36660 + \frac{\pi}{12} \cdot 0,84642 - \frac{\pi}{36} \cdot 0,35774 \right) \cdot \left(x - \frac{\pi}{6} \right) - \\ - \frac{0,84642}{2} \cdot \left(x - \frac{\pi}{6} \right)^2 + \frac{1}{\pi} \cdot 0,35774 \cdot \left(x - \frac{\pi}{6} \right)^3 & \text{при } x \in \left[\frac{\pi}{6}; \frac{\pi}{3} \right]; \\ 0,5 + \left(-\frac{6}{\pi} \cdot 0,5 + \frac{\pi}{12} \cdot 0,48868 - \frac{\pi}{36} \cdot 0,35774 \right) \cdot \left(x - \frac{\pi}{3} \right) - \\ - \frac{0,48868}{2} \cdot \left(x - \frac{\pi}{3} \right)^2 + \frac{1}{\pi} \cdot 0,35774 \cdot \left(x - \frac{\pi}{3} \right)^3 & \text{при } x \in \left[\frac{\pi}{3}; \frac{\pi}{2} \right] \end{cases}$$

или

$$S_3(x) = \begin{cases} 1 + 0,02816x - 0,60208x^2 + 0,11387x^3 & \text{при } x \in \left[0; \frac{\pi}{6} \right]; \\ 0,86602 - 0,50864 \left(x - \frac{\pi}{6} \right) - 0,42321 \left(x - \frac{\pi}{6} \right)^2 + 0,11387 \left(x - \frac{\pi}{6} \right)^3 & \text{при } x \in \left[\frac{\pi}{6}; \frac{\pi}{3} \right]; \\ 0,5 - 0,8582 \left(x - \frac{\pi}{3} \right) - 0,24434 \left(x - \frac{\pi}{3} \right)^2 + 0,11387 \left(x - \frac{\pi}{3} \right)^3 & \text{при } x \in \left[\frac{\pi}{3}; \frac{\pi}{2} \right]. \end{cases}$$

Полученная сплайн-функция является непрерывной и обеспечивает непрерывность первых и вторых производных. ■

4.5.3. ИНТЕРПОЛЯЦИОННЫЕ ДИФФЕРЕНЦИАЛЬНЫЕ ПАРАБОЛИЧЕСКИЕ СПЛАЙНЫ

Выше отмечено, что применение параболических сплайнов более оправданно при аппроксимационной обработке широкого класса сеточных функций, например, рассчитанных методами сеток второго и третьего порядка. Сплайн-методы развиваются и как самостоятельный класс методов решения задач с обыкновенными дифференциальными уравнениями (см. главу 6).

Следует отметить, что широко распространенные на практике кубические дифференциальные сплайны теоретически обеспечивают четвертый порядок сходимости. Однако для их корректного применения необходимо, чтобы исходные данные (аппроксимируемая функция $y_i = f(x_i)$) были заданы с порядком точности, не меньшим третьего. Это условие при обработке рассчитанных на основе методов сеток функций $y_i = f(x_i)$, как правило, не выполняется, и поэтому применять кубические сплайны не всегда целесообразно. Это обуславливает необходимость использования параболических сплайн-функций. Однако полученные традиционным способом сплайны $S_2(x)$ неустойчивы в смысле отсутствия их сходимости к исходной функции, что связано с несимметричным характером задания условий согласования и, как следствие, с аналогичным характером граничных условий. Поэтому для регуляризации сплайнов $S_2(x)$ Шенберг в [59] построил такой алгоритм аппроксимации с помощью Д-сплайнов, в котором узлы сплайна сдвинуты относительно узлов сеточной функции. Для нахождения параметров этого сплайна глобальным способом получается трехдиагональная система линейных алгебраических уравнений с преобладанием диагональных элементов. Однако такая модификация существенно усложняет расчетные формулы алгоритма, и в ней используется дополнительное, обычно в задачах не требующееся условие непрерывности производных в серединах отрезков между узлами интерполяции. В ряде случаев это дополнительное требование может привести к непредсказуемому снижению точности аппроксимации и к отсутствию сходимости.

Рассмотрим принцип построения параболических дифференциальных сплайнов со сдвигом узлов сплайна, следуя [44].

Пусть исходная функция $y_i = f(x_i)$, $i = 0, n$, задана в узлах $x_i \in [a, b]$, называемых *узлами интерполяции* и принадлежащими некоторой сетке Δ : $a = x_0 < x_1 < \dots < x_n < b$ ($n \geq 2$). Введем также в рассмотрение другую сетку $\bar{\Delta}$: $a = \bar{x}_0 < \bar{x}_1 < \dots < \bar{x}_n < b = \bar{x}_{n+1}$, узлы которой сдвинуты по отношению к узлам сетки Δ влево, так что $x_{i-1} < \bar{x}_i < x_i$ ($i = 1, n$).

Функция $S_2(x) = \bigcup_{i=0}^{n-1} S_{2,i}(x)$ называется *интерполяционной параболической сплайн-функцией*, интерполирующей $f(x_i)$ на сетке Δ и принадлежащей классу гладкости $C_1[a, b]$, если она составлена из совокупности звеньев $S_{2,i}(x)$ ($i = 0, n+1$), определенных на каждом частичном отрезке $[\bar{x}_i, \bar{x}_{i+1}]$ сетки $\bar{\Delta}$ и представленных в виде алгебраических многочленов второй степени:

$$S_{2,i}(x) = \sum_{k=0}^2 a_{k,i} (x - \bar{x}_i)^k = a_{0,i} + a_{1,i} \cdot (x - \bar{x}_i) + a_{2,i} \cdot (x - \bar{x}_i)^2, \quad (4.80)$$

где $a_{k,i}$ – коэффициенты, которые находятся из дифференциальных условий согласования

$$\delta S_{2,i}^{(p_1)}(x_k) = S_{2,i}^{(p_1)}(x)|_{x=\bar{x}_k} - f^{(p_1)}(x)|_{x=\bar{x}_k} = 0 \quad (k = i, i+1), \quad (4.81)$$

где $p_1 = 0$ или $p_1 = 1$. Звенья затем стыкуются по условию непрерывности (стыковки):

$$S_{2,i-1}^{(p_2)}(x)|_{x=\bar{x}_i} = S_{2,i}^{(p_2)}(x)|_{x=\bar{x}_i} \quad (4.82)$$

с порядками p_2 , отличными от p_1 , но такими, что $0 \leq p_2 \leq 2$.

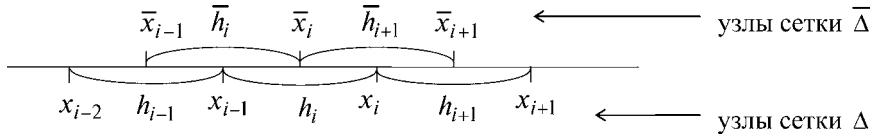


Рис. 4.12

В соответствии с этим определением приведем два сплайна $S_2(x)$ и $\bar{S}_2(x)$ и рассмотрим алгоритм построения одного из них, а именно $S_2(x)$.

Для некоторого отрезка $[\bar{x}_i, \bar{x}_{i+1}]$ формула $S_{2,i}(x)$ имеет вид (4.80)

$$S_{2,i}(x) = a_{0,i} + a_{1,i} \cdot (x - \bar{x}_i) + a_{2,i} \cdot (x - \bar{x}_i)^2.$$

На отрезке $[\bar{x}_i, \bar{x}_{i+1}]$ для определения $a_{0,i}$, $a_{1,i}$, $a_{2,i}$ используются условия согласования (4.81):

$$\delta S_{2,i}(x_i) = 0, \quad \delta S'_{2,i}(\bar{x}_i) = 0, \quad \delta S'_{2,i}(\bar{x}_{i+1}) = 0.$$

Для простоты предположим, что узлы сплайна расположены в серединах отрезков $[x_i, x_{i+1}]$, $i = 0, n-1$, составленных по узлам интерполяции.

На рисунке 4.12 изображена схема расположения узлов сетки Δ по отношению к узлам сетки Δ . Между шагами сеток Δ и $\bar{\Delta}$ в этом случае существует простая связь

$$\bar{h}_{i+1} = \frac{h_i}{2} + \frac{h_{i+1}}{2} = \frac{H_i^{i+1}}{2}.$$

Из условия $\delta S'_{2,i}(\bar{x}_i) = 0$ следует равенство $a_{1,i} = \bar{m}_i$ ($\bar{m}_i = f'(\bar{x}_i)$), а из условия $\delta S'_{2,i}(\bar{x}_{i+1}) = 0$ вытекает, что

$$\bar{m}_i + 2a_{2,i} \frac{H_i^{i+1}}{2} = \bar{m}_{i+1}; \quad a_{2,i} = \frac{\Delta \bar{m}_i}{H_i^{i+1}}.$$

Из условия интерполяции $\delta S_{2,i}(x_i) = 0$ определяется коэффициент

$$a_{0,i} = f_i - \frac{h_i}{2} \bar{m}_i - \frac{h_i^2 \Delta \bar{m}_i}{4 H_i^{i+1}}.$$

Итак, звено сплайна $S_{2,i}(x)$ на произвольном отрезке $[\bar{x}_i, \bar{x}_{i+1}]$ выражается формулой

$$S_{2,i}(x) = f_i - \frac{h_i}{2} \bar{m}_i - \frac{h_i \Delta \bar{m}_i}{4(1 + \delta_{i+1})} + \bar{m}_i \cdot (x - \bar{x}_i) + \frac{\Delta \bar{m}_i}{h_i(1 + \delta_{i+1})} \cdot (x - \bar{x}_i)^2, \quad (4.83)$$

где $\delta_k = h_{k+1}/h_k$ ($k = 1, 2, \dots, i, i+1, \dots$).

Подставив (4.83) в (4.82), получим параметрическую систему относительно \bar{m}_i :

$$\begin{aligned} & \left(\frac{h_i}{2} - \frac{\delta_i}{4(\delta_i + 1)} H_{2(i-1)}^i \right) \bar{m}_{i-1} + \frac{h_i}{4} \left(\frac{2\delta_{i+1} + 1}{\delta_{i+1} + 1} + \frac{H_{2(i-1)}^i}{h_{i-1}(\delta_i + 1)} \right) \bar{m}_i + \\ & + \frac{h_i}{4(\delta_{i+1} + 1)} \bar{m}_{i+1} = \Delta f_{i-1}, \quad i = \overline{1, n}, \end{aligned} \quad (4.84)$$

где $H_{2(i-1)}^i = 2h_{i-1} + h_i$. Система (4.84) обладает свойством преобладания диагональных элементов, но она незамкнута, так как число неопределенных параметров равно $n + 2$. В качестве двух дополнительных уравнений (граничных условий) могут быть приняты значения производных, аппроксимированных на концах со вторым порядком аппроксимации (см. главу 5):

$$\begin{aligned}\bar{m}_0 \equiv \bar{f}'_0 &= \frac{1}{\bar{H}_1^2} \cdot \left(-\frac{\bar{h}_1}{\bar{h}_2} \bar{f}_2 + \frac{(\bar{H}_1^2)^2}{\bar{h}_1 \bar{h}_2} \bar{f}_1 - \frac{\bar{H}_1^2 + \bar{h}_1}{\bar{h}_1} \bar{f}_0 \right), \\ \bar{m}_{n+1} \equiv \bar{f}'_{n+1} &= \frac{1}{\bar{H}_n^{n+1}} \cdot \left(\frac{\bar{H}_n^{n+1} + \bar{h}_{n+1}}{\bar{h}_{n+1}} \bar{f}_{n+1} - \frac{(\bar{H}_n^{n+1})^2}{\bar{h}_n \bar{h}_{n+1}} \bar{f}_n + \frac{\bar{h}_{n+1}}{\bar{h}_n} \bar{f}_{n-1} \right),\end{aligned}\quad (4.85)$$

где $\bar{H}_n^{n+1} = \bar{h}_n + \bar{h}_{n+1}$, $\bar{f}_i = f(\bar{x}_i)$. Здесь шаги $\bar{h}_1, \bar{h}_2, \bar{h}_{n-1}, \bar{h}_n, \bar{h}_{n+1}$ и их комбинации взяты на сетке $\bar{\Delta}$. После определения всех параметров \bar{m}_i их подставляют во все звенья $S_{2,i}(x)$, $i = \overline{0, n}$, из которых затем формируется результирующая многозвенная сплайн-функция $S_2(x)$, имеющая минимальный дефект $q = 1$.

При $h = \text{const}$ формулы (4.83), (4.84), (4.85) упрощаются:

$$\begin{aligned}S_{2,i}(x) &= f_i - \frac{h}{2} \bar{m}_i - \frac{h}{8} \Delta \bar{m}_i + \bar{m}_i(x - \bar{x}_i) + \frac{\Delta \bar{m}_i}{2h} (x - \bar{x}_i)^2 \quad (i = \overline{0, n+1}); \\ \bar{m}_{i-1} + 6\bar{m}_i + \bar{m}_{i+1} &= \frac{8\Delta \bar{f}_{i-1}}{h} \quad (i = \overline{1, n}), \\ \bar{m}_0 &= \frac{1}{2h} (-3\bar{f}_{0\Delta} + 4\bar{f}_1 - \bar{f}); \quad \bar{m}_{n+1} = \frac{1}{2h} (\bar{f}_{n-1} - 4\bar{f}_n + 3\bar{f}_{n+1\Delta}).\end{aligned}\quad (4.86)$$

Здесь вместо сетки $\bar{\Delta}$ используется равномерная сетка $\bar{\Delta}_\Delta = (\bar{x}_{0\Delta}, \bar{x}_1, \dots, \bar{x}_n, \bar{x}_{n+1\Delta})$, где $\bar{x}_{0\Delta} = a - \frac{h}{2}$, $\bar{x}_{n+1\Delta} = b + \frac{h}{2}$ — дополнительные узлы; $\bar{f}_{0\Delta} = f(\bar{x}_{0\Delta})$, $\bar{f}_{n+1\Delta} = f(\bar{x}_{n+1\Delta})$.

Другой параболический сплайн $\bar{S}_{2,i}(x)$ интерполяционного типа приведен в [44], и он построен для более общего случая, когда узлы \bar{x}_i не совпадают с серединами отрезков $[x_i, x_{i+1}]$. При этом $\bar{S}_{2,i}(x)$ записывается для отрезка $[x_i, x_{i+1}]$ с узлами интерполяции x_i, x_{i+1} :

$$\bar{S}_{2,i}(x) = f_i + \bar{m}_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - \bar{x}_{i+1})_+^2, \quad (4.87)$$

где, как и раньше, $x_{i-1} < \bar{x}_i < x_i$, $(i = \overline{1, n})$, а

$$(x - \bar{x}_{i+1})_+ = \begin{cases} x - \bar{x}_{i+1} & \text{при } x > \bar{x}_{i+1}, \\ 0 & \text{при } x \leq \bar{x}_{i+1}. \end{cases} \quad (4.88)$$

Здесь коэффициенты при $(x - x_0)^0$ и $(x - x_i)^1$ найдены с учетом условий согласования $\delta \bar{S}_{2,i}(x_i) = 0$ и $\delta \bar{S}'_{2,i}(x_i) = 0$. Коэффициенты c_i и d_i определяются из аналогичных условий согласования, но заданных в точке x_{i+1} отрезка $[x_i, x_{i+1}]$:

$$\begin{aligned} d_i &= \frac{\Delta f_i}{\bar{h}_{i+1}(\bar{h}_{i+1} - h_{i+1})} - \frac{\bar{m}_i + \bar{m}_{i+1}}{2} \cdot \frac{h_{i+1}}{\bar{h}_{i+1}(\bar{h}_{i+1} - h_{i+1})}; \\ c_i &= \frac{\Delta \bar{m}_i}{2h_{i+1}} - \frac{\Delta f_i}{h_{i+1}(\bar{h}_{i+1} - h_{i+1})} + \frac{\bar{m}_i + \bar{m}_{i+1}}{2(\bar{h}_{i+1} - h_{i+1})}. \end{aligned} \quad (4.89)$$

Тогда для получения связи определенных и неопределенных параметров используется условие стыковки (4.82) при $p_2 = 2$, т. е. условие непрерывности второй производной в точке x_i : $\bar{S}_{2,i-1}''(x)|_{x=x_i} = \bar{S}_{2,i}''(x)|_{x=x_i}$. Учтем, что $\bar{S}_{2,i-1}(x)|_{x=x_i} = c_{i-1} + d_{i-1}$, а $\bar{S}_{2,i}''(x)|_{x=x_i} = c_i$ в силу (4.88).

Подставляя в выражение $c_{i-1} + d_{i-1} = c_i$ соотношения (4.89), получаем систему линейных алгебраических уравнений относительно параметров \bar{m}_i ($i = 1, n-1$):

$$\begin{aligned} \frac{h_i - \bar{h}_i}{h_i \bar{h}_i} \bar{m}_{i-1} + \left(\frac{h_i + \bar{h}_i}{h_i \bar{h}_i} + \frac{2h_{i+1} - \bar{h}_{i+1}}{h_{i+1}(h_{i+1} - \bar{h}_{i+1})} \right) \cdot \bar{m}_i + \frac{\bar{h}_{i+1}}{h_{i+1}(h_{i+1} - \bar{h}_{i+1})} \bar{m}_{i+1} = \\ = 2 \cdot \left(\frac{\Delta f_{i-1}}{h_i \bar{h}_i} + \frac{\Delta f_i}{h_{i+1}(h_{i+1} - \bar{h}_{i+1})} \right), \quad i = \overline{1, n-1}. \end{aligned} \quad (4.90)$$

В качестве двух недостающих условий могут быть взяты аппроксимационные выражения (4.85). Случаи задания других типов краевых условий приведены в [44] и здесь не рассматриваются.

Отметим, что алгоритм построения рассмотренного сплайна является достаточно сложным, так как $\bar{S}_{2,i}(x)$ имеет дополнительное слагаемое $d_i(x - \bar{x}_{i+1})^2$, а коэффициенты сплайна и системы (4.90) зависят от шагов h и \bar{h} , характеризующих две сетки Δ и $\bar{\Delta}$, которые также достаточно сложны. Кроме того, требование непрерывности второй производной, присущее сплайну второй степени, может привести не к улучшению качества аппроксимации, а к ее ухудшению, так как эта производная для $\bar{S}_2(x)$ является константой.

Методика построения параболического сплайна аналогична методике, изложенной в п. 4.5.2 для кубических сплайнов.

Замечание. Использование интегрального подхода (ИД-сплайна), предложенного в [15–21] и изложенного в п. 4.5.4, 4.5.5, позволяет строить устойчивые алгоритмы аппроксимации функций параболическими сплайнами без смещения узлов сплайна относительно узлов интерполяции. Кроме того, эти сплайны являются консервативными, в том смысле, что интегралы от $S_2(x)$ и $y_i = f(x_i)$ равны как на $[a, b]$, так и на всех частичных отрезках $[x_i, x_{i+1}]$ ($i = 0, n-1$). С помощью интегрального подхода могут быть построены замкнутые алгоритмы аппроксимации сплайнами произвольной четной степени [21].

Пример 4.14. Для сеточного представления функции $y = x^3$, т. е. $y_i = x_i^3$ ($x_i = -2; -1; 0; 1; 2$, ($i = \overline{0, 4}$), найти интерполяционную параболическую сплайн-функцию $S_2(x)$ со сдвигом узлов на основе формул (4.86).

Для построения равномерной сетки $\bar{\Delta}_d$ введем в рассмотрение фиктивные узлы $\bar{x}_{0,d} = -\frac{5}{2}$; и $\bar{x}_{5,d} = \frac{5}{2}$ со значениями функции $\bar{f}_{0,d} = -15,625$ и $\bar{f}_{5,d} = 15,625$.

Таким образом, будем определять сплайн $S_2(x)$ на отрезке $[-2,5; 2,5]$ с сетками $\bar{\Delta}_d = \{-2,5; -1,5; -0,5; 0,5; 1,5; 2,5\}$, $\Delta = \{-2; -1; 0; 1; 2\}$.

1. Запишем систему (4.86):

$$\begin{aligned}\bar{m}_0 &= \frac{1}{2h}(-3\bar{f}_{0,\Delta} + 4\bar{f}_1 - \bar{f}_2), \\ \bar{m}_{i-1} + 6\bar{m}_i + \bar{m}_{i+1} &= \frac{8\Delta f_{i-1}}{h}, \quad i = \overline{1, 4}; \\ \bar{m}_5 &= \frac{1}{2h}(\bar{f}_3 - 4\bar{f}_4 + 3\bar{f}_{5,\Delta}).\end{aligned}$$

2. Вычислим значения производных на концах отрезка $[-2,5; 2,5]$:

$$\begin{aligned}\bar{m}_0 &= \frac{1}{2}[-3 \cdot (-15,625) + 4 \cdot (-3,375) + 0,125] = 16,75; \\ \bar{m}_5 &= \frac{1}{2}(0,125 - 4 \cdot 3,375 + 3 \cdot 15,625) = 16,75.\end{aligned}$$

Составим трехдиагональную систему относительно \bar{m}_i :

$$\begin{aligned}\bar{m}_0 &= 16,75; \\ \bar{m}_0 + 6\bar{m}_1 + \bar{m}_2 &= 8 \cdot 7 = 56 \quad (\Delta f_0 = 7); \\ \bar{m}_1 + 6\bar{m}_2 + \bar{m}_3 &= 8 \cdot 1 = 8 \quad (\Delta f_1 = 1); \\ \bar{m}_2 + 6\bar{m}_3 + \bar{m}_4 &= 8 \cdot 1 = 8 \quad (\Delta f_2 = 1); \\ \bar{m}_3 + 6\bar{m}_4 + \bar{m}_5 &= 8 \cdot 7 = 56 \quad (\Delta f_3 = 7); \\ \bar{m}_5 &= 16,75.\end{aligned}$$

Решим систему методом прогонки. В результате получаем значения производных \bar{m}_i (табл. 4.21).

Таблица 4.21

\bar{x}_i	$\bar{x}_{0,\Delta} = -2,5$	-1,5	-0,5	0,5	1,5	$\bar{x}_{5,\Delta} = 2,5$
\bar{m}_i	16,75	6,5061	0,2134	0,2134	6,5061	16,75
i	0	1	2	3	4	5

3. Построим многозвенную (в данном случае пятизвенную) сплайн-функцию $S_2(x)$. Для простоты ограничимся построением многочлена только для одного звена $S_{2,0}(x)$:

$$\begin{aligned}S_{2,0}(x) &= -15,625 - 0,5 \cdot 16,75 - \frac{1}{8} \cdot (6,5061 - 16,75) + 16,75 \cdot (x + 2,5) + \\ &+ \frac{(6,5061 - 16,75)}{2} \cdot (x + 2,5)^2 = -22,719 + 16,75 \cdot (x + 2,5) - 5,1220 \cdot (x + 2,5)^2.\end{aligned}$$

остальные звенья строятся аналогично. ■

4.5.4. ВОССТАНАВЛИВАЮЩИЕ, ИНТЕРПОЛЯЦИОННЫЕ И СГЛАЖИВАЮЩИЕ ИНТЕГРАЛЬНО-ДИФФЕРЕНЦИАЛЬНЫЕ ПАРАБОЛИЧЕСКИЕ СПЛАЙНЫ

В п. 4.3 приведены два интегрально-дифференциальных многочлена (4.37) (4.41), которые могут использоваться для интерполяции и сглаживания некоторых функций, заданных своими значениями и интегралами, либо интегралами и производными.

Здесь рассмотрим основанные на этих многочленах глобальные способы конструирования различных параболических ИД-сплайнов дефекта $q = 1$, позволяющие восстанавливать некоторую функцию либо ее сглаживать. Рассмотрим применение многочлена (4.37).

ПОСТРОЕНИЕ ВОССТАНАВЛИВАЮЩИХ СПЛАЙНОВ (ЗАДАЧА 1)

Пусть некоторая сеточная функция определена не значениями f_i в узлах x_i в общем случае неравномерной сетки Ω_n , а значениями интегралов I_i^{i+1} ($i = 0, n-1$). Требуется восстановить эту функцию путем построения сплайн-функции $\tilde{S}_{2,\text{id}1}(x)$ дефекта $q = 1$.

Для решения этой задачи воспользуемся параболическим многочленом (4.37). Для получения $\tilde{S}_{2,\text{id}1}(x)$ при известных параметрах I_i^{i+1} ($i = 0, n-1$) необходимо сначала определить функциональные параметры f_i ($i = 0, n$). Сплайн, составленный из звеньев (4.37), при любых значениях f_i обеспечивает непрерывность $\tilde{S}_{2,\text{id}1}(x)$. Поэтому для получения параметрической связи между f_i и I_i^{i+1} используются условия стыковки $\tilde{S}_{2,\text{id}1}(x)$ по первым производным:

$$\tilde{S}'_{2,i-1,\text{id}1}(x)|_{x=x_i} = \tilde{S}'_{2,i,\text{id}1}(x)|_{x=x_i} \quad (i = \overline{1, n-1}). \quad (4.91)$$

Преобразование (4.91) после подстановки многочлена (4.37) приводит к параметрической системе линейных алгебраических уравнений относительно f_i :

$$\frac{1}{h_i} f_{i-1} + 2 \cdot \left(\frac{1}{h_i} + \frac{1}{h_{i+1}} \right) f_i + \frac{1}{h_{i+1}} f_{i+1} = 3 \cdot \left(\frac{I_{i-1}^i}{h_i^2} + \frac{I_i^{i+1}}{h_{i+1}^2} \right) \quad (i = \overline{1, n-1}). \quad (4.92)$$

Из сопоставления систем (4.92) и (4.76), являющихся следствием условий стыковки параболических ИД-сплайнов и кубических Д-сплайнов, видно, что они являются подобными. Действительно, обе системы имеют одинаковые коэффициенты, а параметры в них, входящие в левую и правую части, подобны друг другу, так как различаются только порядком p . Если их представить в виде $f_k^{(p)}(k = i - 1, i, i + 1)$ и считать $p = -1$ для $I_i^{i+1} = \Delta F_i = F_{i+1} - F_i$ ($F(x)$ — первообразная, $F_i = F(x_i)$), то все порядки параметров, входящих в систему (4.76), будут на единицу больше соответствующих порядков параметров системы (4.92). Рассмотренные соотношения соответствуют также подобным звеньям (4.75) и (4.37), относящимся к кубическим и параболическим сплайнам. Действительно, звено параболического сплайна может быть получено путем дифференцирования звена кубического сплайна (4.75) и замены параметров \bar{m}_i , Δf_i , $\Delta \bar{m}_i$ на f_i , $I_i^{i+1} = \Delta F_i$, Δf_i .

Вернемся к задаче построения параболического сплайна. В качестве граничных условий, замыкающих систему (4.92), используем интегрально-функциональные связи:

$$\begin{aligned} f_0 + \frac{H_1^2}{h_2} f_1 &= \frac{h_1^2}{H_1^2} \left(\frac{2H_1^2 + h_1}{h_1^3} I_0^1 + \frac{1}{h_2^2} I_1^2 \right) \quad (i = 0); \\ \frac{H_{n-1}^n}{h_{n-1}} f_{n-1} + f_n &= \frac{h_n^2}{H_{n-1}^n} \left(\frac{1}{h_{n-1}^2} I_{n-2}^{n-1} + \frac{2H_{n-1}^n + h_n}{h_n^3} I_{n-1}^n \right) \quad (i = n), \end{aligned} \quad (4.93)$$

являющиеся подобными (4.77) и полученные разрешением системы (4.92) и соотношений $a_{2,0} = a_{2,1}, a_{2,n} = a_{2,n-1}$, где $a_{2,k}, k = 0, 1$ и $k = n, n - 1$ — коэффициенты при $(x - x_i)^2$ (последние равенства для коэффициентов справедливы в силу постоянства второй производной на двух соседних отрезках). Подчеркнем, что соотношения (4.93) не нарушают трехдиагональности системы (4.92) и имеют порядок аппроксимации, равный трем, что соответствует порядку сходимости выстраиваемых ИД-сплайнов.

Разрешение системы (4.92), (4.93) методом прогонки, достаточное условие устойчивости которого для данной системы выполняется безусловно, т. е. при любых разбиениях, позволяет вычислить f_i ($i = 0, n$) (параметрически прямая задача). Подстановка f_i, I_i^{i+1} в формулу (4.37) для звеньев $i = 0, n - 1$ определяет интегрально-дифференциальную сплайн-функцию $\tilde{S}_{2,\text{id1}}(x)$ восстанавливающего типа.

Из подобия параболического ИД-сплайна $\tilde{S}_{2,\text{id1}}(x)$ Д-сплайну $\bar{S}_3(x)$ и возможности его получения путем дифференцирования кубического Д-сплайна, а также на основе соотношения (4.78), устанавливающего порядок сходимости Д-сплайна $\bar{S}_3(x)$, можно получить оценку порядка сходимости параболического ИД-сплайна $\tilde{S}_{2,\text{id1}}(x)$:

$$f^{(p)}(x) - \tilde{S}_{2,\text{id1}}^{(p)}(x) |_{C[a,b]} \leq M_3 h^{3-p}, \quad p = 0, 1, 2, \quad (4.94)$$

свидетельствующую о сходимости $\tilde{S}_{2,\text{id1}}(x)$ с третьим порядком ($p = 0$), а производных $\tilde{S}_{2,\text{id1}}^{(p)}(x)$ с порядком ($3 - p$) ($p = 1, 2$).

Методика получения восстанавливающих сплайнов аналогична изложенной выше для кубического сплайна. Приведем пример построения глобального интегрально-дифференциального сплайна для функции, заданной интегралами на частичных отрезках.

Пример 4.15. Пусть исходная функция задана значениями интегралов на частичных отрезках $[x_i; x_{i+1}]$, $i = \overline{0, 3}$ (табл. 4.22).

Требуется найти восстанавливающий сплайн $\tilde{S}_{2,\text{id1}}(x)$ дефекта $q = 1$.

Таблица 4.22

i	0	1	2	3	4
x_i	0	1	2	3	4
I_i^{i+1}	$-0, (6)$	$7, (3)$	$44, (3)$	$159, 363$	

1. Запишем систему (4.92) для внутренних узлов x_i ($i = \overline{1, 3}$). При этом учтем, что шаг сетки постоянный:

$$\begin{aligned} \frac{1}{h} f_0 + \frac{4}{h} f_1 + \frac{1}{h} f_2 &= 3 \frac{I_0^1 + I_1^2}{h^2} \quad (i=1); \\ \frac{1}{h} f_1 + \frac{4}{h} f_2 + \frac{1}{h} f_3 &= 3 \frac{I_1^2 + I_2^3}{h^2} \quad (i=2); \\ \frac{1}{h} f_2 + \frac{4}{h} f_3 + \frac{1}{h} f_4 &= 3 \frac{I_2^3 + I_3^4}{h^2} \quad (i=3). \end{aligned}$$

Данная система замыкается условиями (4.93), записанными также при $h = \text{const}$:

$$f_0 + 2f_1 = \frac{h}{2} \left(\frac{5}{h^2} I_0^1 + \frac{1}{h^2} I_1^2 \right) \quad (i=0);$$

$$2f_3 + f_4 = \frac{h}{2} \left(\frac{1}{h^2} I_2^3 + \frac{5}{h^2} I_3^4 \right) \quad (i=4).$$

2. Подставляя $h = 1$ и значения интегралов в правые части, получаем систему линейных алгебраических уравнений, которую в соответствии с методом прогонки можно записать в виде

$$\alpha_i f_{i-1} - \beta_i f_i + \gamma_i f_{i+1} = \delta_i \quad (i=\overline{0,4}), \quad \alpha_0 = \gamma_4 = 0.$$

Коэффициенты системы, прогоночные коэффициенты P_i , Q_i и результат решения системы методом прогонки приведены в таблице 4.23.

Таблица 4.23

i	α_i	β_i	γ_i	δ_i	P_i	Q_i	f_i
0	0	-1	2	2,000	-2	2,000	0,94418
1	1	-4	1	20,000	-1/2	9,000	0,52791
2	1	-4	1	155,000	-2/7	41,714	16,94418
3	1	-4	1	611,089	-7/26	153,293	86,69534
4	2	-1	0	420,574	—	246,974	247,36348

3. Подставляя значения f_i и $\nabla I_i^{i+1} = I_i^{i+1} - f_i h$ в формулу (4.37) на каждом отрезке $[x_i, x_{i+1}]$, $i=\overline{0,3}$, находим искомый параболический восстанавливающий сплайн.

Для первого и второго отрезков получаем следующие функции:

$$\tilde{S}_{2,0}(x) = 0,94418 - 8,832536 \cdot x + 8,416264 \cdot x^2;$$

$$\tilde{S}_{2,1}(x) = 0,527908 + 8 \cdot (x-1) + 8,416264 \cdot (x-1)^2.$$

Для остальных двух отрезков выполняются аналогичные действия. Легко убедиться в том, что полученная сплайн-функция имеет дефект $q = 1$, т. е. первая производная во внутренних узлах x_1, x_2, x_3 непрерывна. Таким образом, для x_1 имеем

$$\tilde{S}'_{2,0}(x)|_{x=1} = -8,832536 + 2 \cdot 8,416264 \cdot x|_{x=1} = 8;$$

$$\tilde{S}'_{2,1}(x)|_{x=1} = 8 + 2 \cdot 8,416264 \cdot (x-1)|_{x=1} = 8.$$

Аналогичное условие выполняется и для других оставшихся узлов. ■

ПОСТРОЕНИЕ ИНТЕРПОЛЯЦИОННЫХ СПЛАЙНОВ (ЗАДАЧА 2)

Пусть на равномерно сгущающейся или разрежающейся сетке Ω_n заданы значения сеточной функции $f_i = f(x_i)$, $i = \overline{0, n}$.

Требуется восполнить эту функцию глобальным интерполяционным интегрально-функциональным сплайном $S_{2,\text{ннд}}^{(\text{инт})}(x)$ дефекта $q = 1$.

Решение этой задачи осуществляется путем объединения многочленов-звеньев (4.37) после определения интегралов I_i^{l+1} , выраженных из (4.92) по рекуррентным право- и левосторонним соотношениям (параметрически обратная задача):

$$I_i^{l+1} = -\delta_{i+1}^2 I_{i-1}^l + \frac{h_{i+1}^3}{3} L_i(h, f), \quad i = \overline{1, n-1} \quad (\delta_{i+1} \leq 1); \quad (4.95)$$

$$I_{i-1}^l = -(\delta_{i+1}^{-1})^2 I_i^{l+1} + \frac{h_i^3}{3} L_i(h, f), \quad i = \overline{n-1, 1} \quad (\delta_{i+1}^{-1} \leq 1), \quad (4.96)$$

где $L_i(h, f) = \frac{1}{h_i} f_{i-1} + 2 \cdot \left(\frac{1}{h_i} + \frac{1}{h_{i+1}} \right) \cdot f_i + \frac{1}{h_{i+1}} f_{i+1}$ — функция, равная сумме «удельных интегралов» по каждой паре смежных отрезков; $\delta_{i+1} = \frac{h_{i+1}}{h_i}$ — параметр, характеризующий неравномерность сетки, который называется *параметром нерегулярности*. Соотношения (4.95), (4.96) записаны в виде, учитывающем характер задания сеточной функции на неравномерной сетке, которая, по предположению, либо равномерно сгущается вправо, либо равномерно разрежается. Тогда (4.95) соответствует правостороннему обратному ходу метода прогонки, а (4.96) — левостороннему обратному ходу. Необходимость такого порядка действий обусловлена достаточным условием устойчивости прогонки $\delta_{i+1} \leq 1$ для (4.95) и $\delta_{i+1}^{-1} \leq 1$ для (4.96). Заметим, что в данном случае выполняется только обратный ход метода прогонки. Соотношения (4.95), (4.96) при $f(x) \in C_m[a, b]$ ($m \geq 3$) обеспечивают четвертый порядок точности вычисления интегралов.

В качестве краевых (граничных) условий для право- и левостороннего соотношений значения интегралов I_0^l и I_{n-1}^l вычисляются по одноинтервальным, трехточечным, лево- и правосторонним квадратурным формулам [17, 24] (см. также главу 5):

$$I_{i-1}^l = \frac{h_i^3}{6H_i^{l+1}} \left[\frac{H_{2i}^{3(i+1)}}{h_i^2} f_{i-1} + \frac{H_i^{i+1} H_{i+1}^{3(i+1)}}{h_i^2 h_{i+1}} f_i - \frac{1}{h_{i+1}} f_{i+1} \right] \quad (i=1); \quad (4.97)$$

$$I_i^{l+1} = \frac{h_{i+1}^3}{6H_i^{l+1}} \left[-\frac{1}{h_i^2} f_{i-1} + \frac{H_i^{i+1} H_{3i}^{i+1}}{h_i h_{i+1}^2} f_i + \frac{H_{3i}^{2(i+1)}}{h_{i+1}} f_{i+1} \right] \quad (i=n-1), \quad (4.98)$$

где $H_{ki}^{m(i+1)} = kh_i + mh_{i+1}$.

При $h = \text{const}$ (сетка Ω_n равномерная) данные формулы упрощаются:

$$I_{i-1}^l = \frac{h}{12} (5f_{i-1} + 8f_i - f_{i+1}); \quad I_i^{l+1} = \frac{h}{12} (-f_{i-1} + 8f_i + 5f_{i+1}). \quad (4.99)$$

В работах [17, 24] показано, что квадратурные формулы (4.97), (4.98) обеспечивают четвертый порядок аппроксимации интегралов на трехточечном шаблоне (x_{i-1}, x_i, x_{i+1}) , и поэтому порядок аппроксимации I_i^{l+1} соотношением (4.95) или (4.96) при фиксированном i за счет краевых условий не понижается. Показано также, что в данном способе построения интерполяционного сплайна достигается третий порядок сходимости. Однако способ

имеет недостаток, состоящий в том, что при аппроксимации негладких функций ошибки, возникающие при расчете интегралов по формулам (4.95) или (4.96), переносятся на последующие значения. Данный недостаток устранен в задаче построения слабо сглаживающего параболического сплайна, изложенного в п. 4.5.5.

ПОСТРОЕНИЕ СГЛАЖИВАЮЩИХ СПЛАЙНОВ (ЗАДАЧА 3)

Пусть в общем случае на неравномерной сетке Ω_n задана сеточная функция, определенная, например, в физическом эксперименте с достаточно большими погрешностями, определяющими разброс значений $f(x_i)$, т. е. $f(x_i) \pm \varepsilon_i$ ($i = 0, n$), где ε_i намного превышают величину $O(h_{i+1}^3)$.

Требуется восполнить эту функцию параболическим сглаживающим сплайном $\tilde{S}_{2,\text{ид1}}(x)$ дефекта $q = 1$.

Поставленная задача решается глобальным способом путем сведения ее к задаче 1. Рассмотрим его особенности.

Сначала путем визуального изучения исходной функции $y_i = f(x_i) \pm \varepsilon_i$ исследователь-вычислитель формирует «полосу разброса» данных, ограниченную двумя функциями — «верхней» $y_i^B = \hat{f}^{(B)}(x_i)$ и «нижней» $y_i^H = \hat{f}^{(H)}(x_i)$ (на рис. 4.13 квадратиками условно показаны значения, полученные, например, в результате физического эксперимента). Если на некоторых участках разброс данных не заметен, функции $y_i^B = \hat{f}^{(B)}(x_i)$, $y_i^H = \hat{f}^{(H)}(x_i)$ могут быть совмещены друг с другом. В случае отсутствия значений $f(x_i)$ на концах отрезка $[a, b]$ они определяются приближенно путем линейной или параболической экстраполяции.

Далее на каждом отрезке $[x_i, x_{i+1}]$, $i = \overline{0, n-1}$, по формулам (4.97) или (4.98) вычисляются интегралы $I_i^{i+1(B)}$ и $I_i^{i+1(H)}$ для функций $y_i^B = \hat{f}^{(B)}(x_i)$ и $y_i^H = \hat{f}^{(H)}(x_i)$ соответственно. Недостающие значения функций на линиях, ограничивающих «полосу разброса», определяются путем дополнительной интерполяции. Следует также учесть, что интеграл I_0^1 на отрезке $[x_0, x_1]$ вычисляется по левосторонней формуле (4.97), интеграл I_{n-1}^n на отрезке $[x_{n-1}, x_n]$ —

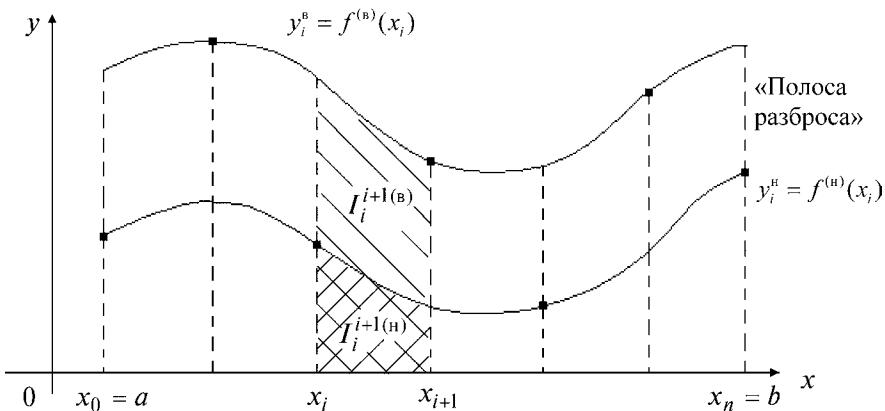


Рис. 4.13

по правосторонней, а интегралы по всем внутренним отрезкам могут рассчитываться по любой из формул (4.97), (4.98).

После вычисления интегралов $I_i^{i+1(\text{в})}$ и $I_i^{i+1(\text{н})}$ ($i = \overline{0, n-1}$) по всем частичным отрезкам находятся их средние значения на каждом частичном отрезке:

$$I_i^{i+1} = \frac{I_i^{i+1(\text{в})} + I_i^{i+1(\text{н})}}{2}, \quad i = \overline{0, n-1}.$$

Полученные значения интегралов I_i^{i+1} используются далее в качестве исходных данных для решения задачи 1.

Таким образом, существенной особенностью сглаживающих сплайнов данного типа является формирование «полосы разброса», включающей все экспериментальные данные, которые затем двукратно усредняются. При этом в отличие от метода наименьших квадратов с помощью данного метода реализуется более полный учет локальных свойств аппроксимируемой функции.

4.5.5. СЛАБО СГЛАЖИВАЮЩИЕ ИНТЕРПОЛЯЦИОННЫЕ ИНТЕГРАЛЬНО-ДИФФЕРЕНЦИАЛЬНЫЕ ПАРАБОЛИЧЕСКИЕ СПЛАЙНЫ

Пусть в общем случае на неравномерной сетке Ω_n задана сеточная функция $y_i = f(x_i)$, $i = \overline{0, n}$, определенная либо в численном эксперименте, либо путем измерений с небольшими погрешностями ε_i , не превышающими $O(h_{i+1}^3)$.

Применительно к данной задаче вводится понятие *слабого сглаживания*, в котором учитывается характер задания исходной функции с некоторой погрешностью, не превышающей $O(h_{i+1}^3)$. При этом допускается возможность отклонения получаемой в алгоритме сплайн-функции в узлах $\tilde{S}_{2,\text{идк}}^{(\text{слг})}(x_i)$ от $f(x_i)$ на величину не более $O(h_{i+1}^3)$, т. е. реализуется отклонение $\max |\tilde{S}_{2,\text{идк}}^{(\text{слг})}(x) - f(x)| = O(H^3)$ по всему отрезку $[a, b]$, включая и узлы интерполяции. Здесь индекс « k » ($k = 1$ или $k = 2$) соответствует способу построения сплайна. Подчеркнем, что для традиционных интерполяционных сплайнов данное отклонение в узлах отсутствует.

Требуется восполнить данную функцию параболическим интерполяционным слабо сглаживающим сплайном $\tilde{S}_{1,\text{идк}}^{(\text{слг})}(x)$ ($k = 1, 2$), компонующимся на основе ИД-многочленов (4.37) или (4.41).

Обоснованность такого расширения понятия интерполяции (дополнение его понятием слабого сглаживания) обусловлена следующими причинами.

1. Интерполяционные сплайн-функции, как и классические интерполяционные многочлены, часто применяются для восполнения и обработки сеточных функций, рассчитанных методами сеток второго и третьего порядка, и потому данные функции (исходные для задачи аппроксимации) неточные, и их погрешность обычно составляет $O(h_{i+1}^3)$. Поэтому применение условий интерполяции, обеспечивающих нулевые невязки в узлах, вообще говоря, не является оправданным.

2. Если даже значения исходной функции заданы точно (хотя класс задач, в которых значения $f(x_i)$ являются точными, указать трудно), то точность сплайн-интерполяции из-за этого не повышается. Для параболичес-

ких сплайн-функций порядок сходимости оценивается как третий, а для кубических — как четвертый.

Построение слабо сглаживающего параболического ИД-сплайна может быть осуществлено так же, как и построение кубического дифференциального сплайна (изложенного в п. 4.5.2) двумя способами, различающимися порядками производных, заданных в условиях согласования и условиях стыковки. Это соответствует отмеченному выше (см. п. 4.5.4) принципу подобия дифференциальных и интегрально-дифференциальных сплайнов, описывающихся формулами (4.75) и (4.37). Аналогично легко показать подобие Д-сплайнов (4.71) и (4.41).

Поэтому, так же как и для Д-сплайнов, здесь рассматриваются два сглаживающих интегрально-дифференциальных сплайна $\tilde{S}_{2,\text{id}1}^{(\text{сл})}(x)$ и $\tilde{S}_{2,\text{id}2}^{(\text{сл})}(x)$.

Первый способ решения задачи слабого сглаживания (построение сплайна $\tilde{S}_{2,\text{id}1}^{(\text{сл})}(x)$). Алгоритм восполнения функции $y_i = f(x_i)$ методом слабого сглаживания аналогичен алгоритму восстановления сеточной функции в задаче 1 (см. п. 4.5.4) и основан на формуле ИД-многочлена (4.37), принимаемой в качестве общей формулы звена искомого сплайна и параметрической системы (4.92), которые имеют вид

$$\begin{aligned}\tilde{S}_{2,i,\text{id}1}^{(\text{сл})}(x) &= \tilde{f}_i + \left(\frac{6}{h_{i+1}^2} \nabla I_i^{i+1} - \frac{2}{h_{i+1}} \Delta \tilde{f}_i \right) \cdot (x - x_i) + \\ &+ \left(-\frac{6}{h_{i+1}^3} \nabla I_i^{i+1} + \frac{3}{h_{i+1}^2} \Delta \tilde{f}_i \right) \cdot (x - x_i)^2 \quad (i = \overline{0, n-1});\end{aligned}\quad (4.100)$$

$$\frac{1}{h_i} \tilde{f}_{i-1} + 2 \cdot \left(\frac{1}{h_i} + \frac{1}{h_{i+1}} \right) \tilde{f}_i + \frac{1}{h_{i+1}} \tilde{f}_{i+1} = 3 \cdot \left(\frac{I_{i-1}^i}{h_i^2} + \frac{I_i^{i+1}}{h_{i+1}^2} \right) \quad (i = \overline{1, n-1}). \quad (4.101)$$

Волнистая черта указывает на отличие используемых значений \tilde{f}_i от значений f_i исходной функции. Здесь $h_{i+1} = x_{i+1} - x_i$, $\Delta \tilde{f}_i = \tilde{f}_{i+1} - \tilde{f}_i$, $\nabla I_i^{i+1} = I_i^{i+1} - h_{i+1} \tilde{f}_i$.

В правой части равенства (4.100) обе группы параметров \tilde{f}_i ($i = \overline{0, n}$) и I_i^{i+1} ($i = \overline{0, n-1}$) считаются неопределенными. При этом сначала вычисляются интегралы I_i^{i+1} на всех частичных отрезках $[x_i, x_{i+1}]$ ($i = \overline{0, n-1}$) по одной из явных квадратурных формул (4.97) или (4.98). После этого проводится пересчет функциональных параметров из системы (4.101), которая замыкается связями между \tilde{f}_0 , \tilde{f}_1 , I_0^1 , I_1^2 и между \tilde{f}_{n-1} , \tilde{f}_n , I_{n-2}^{n-1} , I_{n-1}^n (граничными условиями) соответственно на левом и правом концах отрезка $[a, b]$:

$$\begin{aligned}\tilde{f}_0 + \frac{H_1^2}{h_2} \tilde{f}_1 &= \frac{h_1^2}{H_1^2} \left(\frac{2H_1^2 + h_1}{h_1^3} I_0^1 + \frac{1}{h_2^2} I_1^2 \right) \quad (i = 0); \\ \frac{H_{n-1}^n}{h_n} \tilde{f}_{n-1} + \tilde{f}_n &= \frac{h_n^2}{H_{n-1}^n} \left(\frac{1}{h_{n-1}^2} I_{n-2}^{n-1} + \frac{2H_{n-1}^n + h_n}{h_n^3} I_{n-1}^n \right) \quad (i = n),\end{aligned}\quad (4.102)$$

совпадающими с граничными условиями, принятymi в задаче 1 (см. п. 4.5.4).

Отметим, что оба типа неопределенных параметров находятся путем решения смешанной параметрической задачи, так как сначала определялись интегралы I_i^{i+1} ($p = -1$), а затем значения функции \tilde{f}_i ($p = 0$).

Определение \tilde{f}_i ($i = \overline{0, n}$) из (4.101), (4.102), подстановка \tilde{f}_i ($i = \overline{0, n}$) и I_i^{i+1} ($i = 0, n-1$) в (4.100) для всех звеньев и их объединение в общую формулу $\tilde{S}_{2,\text{id}1}^{(\text{сгл})}(x)$ завершают решение задачи слабого сглаживания исходной приближенно заданной сеточной функции, погрешность определения которой в узлах не превышает $O(h_{i+1}^3)$. Если же погрешность задания исходной сеточной функции больше $O(h_{i+1}^3)$, то погрешность вычисления интегралов превысит величину $O(h_{i+1}^3)$ и третий порядок точности аппроксимации $\tilde{S}_{2,\text{id}1}^{(\text{сгл})}(x)$ не будет достигнут.

Первый способ решения задачи аппроксимации слабо сглаживающим сплайном является сходным с решением задачи 2 о восполнении сеточной функции интерполяционным интегрально-функциональным сплайном $S_{2,\text{id}1}^{(\text{инт})}(x)$. Различие состоит в том, что для $S_{2,\text{id}1}^{(\text{инт})}(x)$ неопределенными параметрами являются только I_i^{i+1} ($i = 0, n-1$), а определенными — значения f_i ($i = 0, n$), и они не изменяются при решении задачи. Таким образом, задача 2 решается проще, однако изложенный способ построения сглаживающего сплайна характеризуется большей устойчивостью, так как он основан на решении трехдиагональной системы линейных алгебраических уравнений, которая удовлетворяет условию преобладания диагональных элементов.

В [21] доказана следующая теорема.

Теорема 4.3 (о сходимости глобального параболического ИД-сплайна $\tilde{S}_{2,\text{id}1}^{(\text{сгл})}(x)$).

Пусть функцию $f(x)$ ($x \in [a, b]$), определенную в общем случае на неравномерной сетке Ω_n ($h_i = \text{var}, H = \max_{i=1,n} h_i$) с параметром неравномерности $\Delta = \max_{i=1,n} h_i / \min_{i=1,n} h_i$ аппроксимирует глобальный параболический слабосглаживающий ИД-сплайн $\tilde{S}_{2,\text{id}1}^{(\text{сгл})}(x)$. Тогда, если $f(x) \in C_3[a, b]$ и интегралы I_i^{i+1} ($i = 0, n-1$) находятся по квадратурным формулам (4.97) или (4.98), а \tilde{f}_i ($i = \overline{0, n}$) — из трехдиагональной системы (4.101), (4.102), то справедливы оценки

$$\max_{[a, b]} |\tilde{S}_{2,\text{id}1}^{(\text{сгл})}(x) - f^{(p)}(x)| \leq H^{3-p} (T_{3,p}^{(\text{id}1)} + K_p \cdot \Delta^{1+p}) \cdot M_3, \quad (4.103)$$

где p — порядок производной, $M_3 = \max_{[a, b]} |f'''(x)|$, а константы имеют значения $T_{3,0}^{(\text{id}1)} = \frac{1}{72\sqrt{3}}$; $T_{3,1}^{(\text{id}1)} = \frac{1}{12}$; $K_0 = \frac{11}{48}$, $K_1 = \frac{25}{24}$.

Таким образом, при $f(x) \in C_3[a, b]$ слабо сглаживающие сплайны $\tilde{S}_{2,\text{id}1}^{(\text{сгл})}(x)$ равномерно сходятся к функции $f(x)$ на последовательности сеток $\Omega_1 = \{x_0^{(1)}, x_1^{(1)}\}$, $\Omega_2 = \{x_0^{(2)}, x_1^{(2)}, x_2^{(2)}\}$, ..., $\Omega_n = \{x_0^{(n)}, x_1^{(n)}, \dots, x_n^{(n)}\}$ с возрастающим разбиением k отрезка $[a, b]$: $k_1 = 1$; $k_2 = 2$; ...; $k_n = n$ и т. д. по крайней мере с кубичной скоростью, а их производные $\tilde{S}_{2,\text{id}1}^{(\text{сгл})}(x)$ с ростом n равномерно сходятся к $f'(x)$ по крайней мере с квадратичной скоростью.

Второй способ решения задачи слабого сглаживания (построение сплайна $\tilde{S}_{2,\text{id}2}^{(\text{сгл})}(x)$). Алгоритм восполнения функции $y_i = f(x_i)$ в данной задаче основан на формуле ИД-многочлена (4.41), принимаемой в качестве общей фор-

мулы звена искомого сплайна, и на параметрической системе трехдиагонального вида, вытекающей из условия непрерывности сплайна $\tilde{S}_{2,\text{ид}2,i}^{(\text{спл})}(x)$:

$$\begin{aligned}\tilde{S}_{2,\text{ид}2,i}^{(\text{спл})}(x) = & \left(\frac{1}{h_{i+1}} I_i^{i+1} - \frac{h_{i+1}}{2} \bar{m}_i - \frac{h_{i+1}}{6} \Delta \bar{m}_i \right) + \bar{m} \cdot (x - x_i) + \\ & + \frac{1}{2h_{i+1}} \Delta \bar{m}_i \cdot (x - x_i)^2 \quad (i = \overline{0, n-1}),\end{aligned}\quad (4.104)$$

$$\frac{1}{6} h_{i+1} \bar{m}_{i-1} + \frac{1}{3} (h_i + h_{i+1}) \cdot \bar{m}_i + \frac{1}{6} h_{i+1} \bar{m}_{i+1} = \frac{I_i^{i+1}}{h_{i+1}} - \frac{I_{i-1}^i}{h_i} \quad (i = \overline{1, n-1}), \quad (4.105)$$

где $\bar{m}_i = f'(x)$. Подчеркнем, что формула (4.104) для $\tilde{S}_{2,\text{ид}2,i}^{(\text{спл})}(x)$ и параметрическая система являются подобными соответствующим соотношениям кубического дифференциального сплайна $S_3(x)$ (4.71) и (4.72), в которых порядки всех параметров повышенны на единицу.

Аналогично первому способу будем считать, что все параметры, входящие в (4.104), являются неопределенными. Их значения находятся из смещенной параметрической задачи, в которой вначале по формулам (4.97) или (4.98) определяются интегралы I_i^{i+1} ($i = \overline{0, n-1}$), а затем с использованием системы (4.105) трехдиагонального типа методом прогонки — производные \bar{m}_i ($i = \overline{0, n}$). Для замыкания (4.105) можно использовать два различных подхода.

В рамках первого подхода производные на концах отрезка $[a, b]$ определяются по аппроксимационным формулам второго порядка аппроксимации (см. [17], [23] и главу 5):

$$\begin{aligned}\bar{m}_{i-1} &= \left(-\frac{H_i^{i+1} + h_i}{h_i} f_{i-1} + \frac{(H_i^{i+1})^2}{h_i h_{i+1}} f_i - \frac{h_i}{h_{i+1}} f_{i+1} \right) \quad (i = 1); \\ \bar{m}_{i+1} &= \frac{1}{H_i^{i+1}} \left(\frac{h_{i+1}}{h_i} f_{i-1} - \frac{(H_i^{i+1})^2}{h_i h_{i+1}} f_i + \frac{H_i^{i+1} + h_{i+1}}{h_{i+1}} f_{i+1} \right) \quad (i = n-1).\end{aligned}$$

Во втором подходе при $h_{i+1} = \text{var}$ система (4.105) замыкается двумя трехточечными соотношениями:

$$\begin{aligned}\frac{\bar{m}_i}{h_{i+1}} - \left(\frac{1}{h_{i+1}} + \frac{1}{h_{i+2}} \right) \cdot \bar{m}_{i+1} + \frac{\bar{m}_{i+2}}{h_{i+2}} &= 0 \quad (i = 0); \\ \frac{\bar{m}_{i-2}}{h_{i-1}} - \left(\frac{1}{h_{i-1}} + \frac{1}{h_i} \right) \cdot \bar{m}_{i-1} + \frac{\bar{m}_i}{h_i} &= 0 \quad (i = n),\end{aligned}$$

вытекающими из условия равенства второй производной на двух первых и последних частичных отрезках.

В случае, когда $h = \text{const}$ (сетка равномерная), производные на концах могут быть определены также и путем интегральных аппроксимаций, приведенных в главе 5:

$$\begin{aligned}\bar{m}_i &= \frac{1}{h^2} (-2I_i^{i+1} + 3I_{i+1}^{i+2} - I_{i+2}^{i+3}) \quad (i = 0); \\ \bar{m}_i &= \frac{1}{h_2} (I_{i-3}^{i-2} - 3I_{i-2}^{i-1} + 2I_{i-1}^i) \quad (i = n).\end{aligned}$$

Выбор конкретных типов граничных условий обуславливается особенностями постановки задачи и свойствами аппроксимируемой функции.

Построение слабо сглаживающего сплайна завершается так же, как и в первом способе. Отличительной особенностью второго способа является то, что при решении смешанной параметрической задачи (ее прямой части) вычисляются не сами функции f_i , а их производные. Иногда это более удобно, например, когда в задаче восполнения требуется найти значения производных в узлах.

Выбор конкретного алгоритма сглаживания зависит от априорной информации, от аппроксимируемой функции и от требований к выходным данным.

В заключение отметим, что для сглаживающего сплайна $\tilde{S}_{2,\text{id}2}^{(\text{сгл})}(x)$, как и для сплайна $\tilde{S}_{2,\text{id}1}^{(\text{сгл})}(x)$, имеет место третий порядок сходимости [20], [21].

Задачи для самостоятельного решения

1. Для сеточной функции $x_0 = -1; y_0 = -3; x_1 = 2; y_1 = 4; x_2 = 3; y_2 = 5$ найти значение $y(1)$ с помощью линейной и параболической интерполяций.

2. Функция $f(x)$ аппроксимируется интерполяционным многочленом Лагранжа второй степени для системы равномерно расположенных на отрезке узлов. Оценить погрешность интерполяции в точке x_* :

а) $f(x) = \cos x^2, x_* = \frac{\pi}{12} \quad (0 \leq x \leq \frac{\pi}{4});$

б) $f(x) = x \cos x, x_* = \frac{\pi}{12} \quad (0 \leq x \leq \frac{\pi}{4});$

в) $f(x) = e^{-x^2}, x_* = 0,3 \quad (0 \leq x \leq 0,4).$

3. Сеточное представление функции $f(x) = e^x$ приведено в таблице 4.24.

Найти значения многочлена Лагранжа, интерполирующего функцию $f(x)$ в точках $x_* = 1,03; x_* = 1,07; x_* = 1,13; x_* = 1,18$. Полученные значения сравнить с точными значениями и априорным путем оценить погрешность в указанных точках. Дать анализ погрешностей.

Таблица 4.24

x_i	1,00	1,05	1,10	1,15	1,20
f_i	2,7183	2,8577	3,0042	3,1582	3,3201

4. Пусть сеточные представления функций:

1) $y = e^x$; 2) $y = shx$; 3) $y = chx$; 4) $y = sinx$;

5) $y = cosx$; 6) $y = lnx$; 7) $y = e^{-x}$

заданы на отрезке $[1,00; 1,20]$ для равномерного разбиения с шагом $h = 0,04$.

Построить кубический интерполяционный сплайн способом 1. Задать граничные условия в следующем виде:

а) $f''_0 = 0, f''_n = 0$; б) аппроксимации (4.73).

Найти значения обоих полученных сплайнов в точках $x_* = 1,05; x_* = 1,09; x_* = 1,13; x_* = 1,15; x_* = 1,17$. Проанализировать результаты для обеих сплайн-функций.

Получить оценку точности априорным путем по формуле (4.78).

5. Для условия задачи 4 построить кубический интерполяционный сплайн способом 2. Границные условия аппроксимировать по соотношениям (4.77).

6. Для условия задачи 4 построить параболический слабо сглаживающий сплайн по формулам (4.100)–(4.102). Априорным путем оценить погрешность. Сопоставить полученные результаты с точными значениями функций в промежуточных точках.

7. Для сеточных функций, заданных в таблицах 4.25 и 4.26, требуется:

а) решить задачу интерполяции методами Лагранжа и Ньютона;

б) решить задачу сглаживания методом наименьших квадратов при $m = 1$ и $m = 2$;

в) на одном графике построить исходную сеточную функцию, а также решения задач интерполяции и сглаживания.

Таблица 4.25

x_i	2	4	6	8
f_i	5	2	3	1

Таблица 4.26

x_i	1	2	3	4	5
f_i	4	2	5	8	2

ГЛАВА 5

МЕТОДЫ ЧИСЛЕННОГО ДИФФЕРЕНЦИРОВАНИЯ И ИНТЕГРИРОВАНИЯ

5.1. ПОСТАНОВКА ЗАДАЧИ И ПРИНЦИПЫ КОНСТРУИРОВАНИЯ АППРОКСИМАЦИОННЫХ ФОРМУЛ

Если функция $f(x)$ непрерывна на отрезке $[a, b]$ и известна ее первообразная $F(x)$, то определенный интеграл от этой функции может быть вычислен по формуле Ньютона — Лейбница

$$\int_a^b f(x)dx = F(b) - F(a),$$

где $F'(x) = f(x)$. Однако во многих случаях возникают большие трудности, связанные с нахождением первообразной, или эта задача не может быть решена элементарными способами. Например, в элементарных функциях не выражается интеграл $\int_1^2 \frac{dx}{\ln x}$.

Кроме того, в вычислительной практике часто требуется находить значения производных и определенных интегралов от сеточных функций, заданных в общем случае на неравномерной сетке $\Omega_n = \{x_0, x_1, \dots, x_n\}$, $x_{i+1} = x_i + h_{i+1}$, $i = \overline{0, n-1}$, $h_{i+1} = x_{i+1} - x_i$.

В связи с этим в численном анализе имеется специальный математический аппарат численного дифференцирования и интегрирования, отличный от соответствующего аппарата математического анализа.

Традиционно интегралы вычисляются с помощью квадратурных формул (явно), выраждающихся линейными комбинациями значений функции $y_i = f(x_i)$ в узлах сетки Ω_n , функционально-дифференциальной квадратурной формулы Эйлера — Маклорена, формул Гаусса — Кристоффеля, Маркова и различных нестандартных формул [2], [4], [6], [9], [14], [26], [28], [29], [35]. В данной главе кратко описаны как классические, так и новые методы численного дифференцирования и интегрирования, разработанные на основе аппарата интегрально-дифференциальных сплайнов [15]–[24]. Приведен ряд обобщенных явных квадратурных формул функционально-дифференциального типа, а также описан новый способ вычисления интегралов на основе

решения систем линейных алгебраических уравнений. Последний подход в отличие от способа вычисления интегралов по квадратурным формулам носит неявный характер и связан с построением *неявных алгоритмов*.

Особенность аппарата численного дифференцирования и интегрирования, изложенного в данной главе, состоит в том, что в некоторых аппроксимационных операторах дифференцирования могут использоваться как значения сеточной функции в узлах, так и значения определенных интегралов на частичных отрезках $[x_i, x_{i+1}]$, а в аппроксимационных операторах интегрирования — значения функции и ее производных в узлах. В соответствии с этим далее выделяются классические постановки задач численного дифференцирования и интегрирования, формулируемые обычно на равномерной сетке ($h_{i+1} = \text{const}$), и обобщенные постановки, учитывающие связи производных с интегралами в задаче численного дифференцирования и интегралов с производными в задаче численного интегрирования, формулируемые на неравномерной сетке ($h_{i+1} = \text{var}$). Классические постановки ниже нумеруются цифрой 1, а обобщенные — цифрой 2. Данные постановки относятся только к явным аппроксимационным формулам численного дифференцирования и интегрирования. Соответствующие неявные алгоритмы рассматриваются отдельно.

ПОСТАНОВКА ЗАДАЧИ 1 ЧИСЛЕННОГО ДИФФЕРЕНЦИРОВАНИЯ

Пусть на отрезке $[a, b]$ на равномерной сетке $\Omega_n (h_{i+1} = h = \text{const})$ заданы:

- сеточная функция $y_i = f(x_i)$, $i = 0, n$, своими значениями $f_i = f(x_i)$;
- точки x_j сетки Ω_n , в которых требуется найти значения производных;
- желаемый порядок t точности (аппроксимации) относительно h .

Требуется с заданным порядком точности (аппроксимации) вычислить значения производных $\hat{f}^{(p)}(x)|_{x=x_j}$ в точках x_j сетки, где p — порядок производной.

Иначе, требуется получить аппроксимационный оператор $\hat{f}^{(p)}(x_j)$, удовлетворяющий условию $|f^{(p)}(x_j) - \hat{f}^{(p)}(x_j)| \leq Ch^t$, где $C = \text{const}$, не зависящая от величины шага h .

ПОСТАНОВКА ЗАДАЧИ 2 ЧИСЛЕННОГО ДИФФЕРЕНЦИРОВАНИЯ

Пусть на отрезке $[a, b]$ в общем случае на неравномерной сетке $\Omega_n = \{x_0, x_1, \dots, x_n\}$, $x_{i+1} = x_i + h_{i+1}$, $i = 0, n-1$, $h_{i+1} = x_{i+1} - x_i$ ($h_{i+1} = \text{var}$) заданы:

- сеточная функция $y_i = f(x_i)$, $i = 0, n$, своими значениями $f_i = f(x_i)$ и/или значениями интегралов $I_i^{i+1} = \int_{x_i}^{x_{i+1}} f(x) dx$ на частичных отрезках $[x_i, x_{i+1}]$;
- точки x_j сетки Ω_n , в которых требуется найти значения производных;
- желаемый порядок t точности (аппроксимации) относительно величины шага.

Требуется с заданным порядком точности (аппроксимации) вычислить значения производных $\hat{f}^{(p)}(x)|_{x=x_j}$ в точках x_j сетки (получить аппроксимационный оператор), где p — порядок производной.

ПОСТАНОВКА ЗАДАЧИ 1 ЧИСЛЕННОГО ИНТЕГРИРОВАНИЯ

Пусть на отрезке $[a, b]$ на равномерной сетке $\Omega_n(h_{i+1} = h = \text{const})$ заданы:

- а) сеточная функция $y_i = f(x_i)$, $i = \overline{0, n}$, своими значениями $f_i = f(x_i)$ или сеточное представление формульной функции $y = f(x)$;
- б) желаемый порядок t точности (аппроксимации) относительно величины шага h .

Требуется с заданным порядком точности вычислить значение интеграла

$$\hat{I}_a^b \cong I_a^b = \int_a^b f(x) dx.$$

Иначе, требуется получить аппроксимационный оператор интегрирования \hat{I}_a^b , удовлетворяющий условию $|\hat{I}_a^b - I_a^b| \leq Ch^t$, где $C = \text{const}$, не зависящая от h .

ПОСТАНОВКА ЗАДАЧИ 2 ЧИСЛЕННОГО ИНТЕГРИРОВАНИЯ

Пусть на отрезке $[a, b]$ в общем случае на неравномерной сетке $\Omega_n = \{x_0, x_1, \dots, x_n\}$ ($x_{i+1} = x_i + h_{i+1}$, $i = \overline{0, n-1}$, $h_{i+1} = x_{i+1} - x_i = \text{var}$) заданы:

- а) сеточная функция $y_i = f(x_i)$, $i = \overline{0, n}$, своими значениями $f_i = f(x_i)$ и возможно значениями производных $f^{(p)}(x_i)$;
- б) желаемый порядок t точности (аппроксимации) относительно шага.

Требуется с заданным порядком точности вычислить значение интеграла (получить аппроксимационный оператор)

$$\hat{I}_a^b \cong I_a^b = \int_a^b f(x) dx.$$

Отметим, что символом « \wedge » здесь и далее обозначаются операторы дифференцирования и интегрирования.

Построение аппроксимационных формул для производных и интегралов в данной книге выполняется различными путями на основе:

- 1) разложения функции $f(x)$ по формуле Тейлора;
- 2) разложения первообразной функции $F(x)$ по формуле Тейлора;
- 3) замены заданной функции интерполяционными многочленами и последующим дифференцированием или интегрированием этих многочленов. Дополнительные возможности построения аппроксимационных операторов предоставляют интегрально-дифференциальные сплайны (ИД-сплайны).

Общая интегрально-функционально-дифференциальная формула для аппроксимационного оператора дифференцирования $\hat{f}^{(p)}(x_j)$ может быть записана в форме

$$\hat{f}^{(p)}(x_j) = \sum_{k_1} \left(\sum_{l_1} a_{k_1, l_1}(h) \right) I_{k_1}^{l_1+1} + \sum_{k_2} \left(\sum_{l_2} b_{k_2, l_2}(h) \right) f_{k_2} + \sum_{k_3} \left(\sum_{l_3} c_{k_3, l_3}(h) \right) f'_{k_3} + \dots, \quad (5.1)$$

где $a_{k_1, l_1}(h)$, $b_{k_2, l_2}(h)$, $c_{k_3, l_3}(h), \dots$ — некоторые коэффициенты, индексы при h для упрощения записи не указаны.

Общая функционально-дифференциальная формула для аппроксимационного оператора интегрирования \hat{I}_i^{i+1} может быть записана в виде

$$\hat{I}_i^{i+1} = h \sum_{k_1} a_{k_1}(h) f_{k_1} + h^2 \sum_{k_2} b_{k_2}(h) f'_{k_2} + h^3 \sum_{k_3} c_{k_3}(h) f''_{k_3} + \dots, \quad (5.2)$$

где $a_{k_1}(h)$, $b_{k_2}(h)$, $c_{k_3}(h)$, ... — некоторые коэффициенты.

Возможность такого комбинированного представления следует из разложения первообразной $F(x)$ по формуле Тейлора относительно точки x_i и последующего выражения из него производной некоторого порядка или определенного интеграла.

В формулы (5.1), (5.2), очевидно, входят суммы нескольких групп линейных комбинаций. Так, вторая сумма в (5.1) и первая сумма в (5.2) соответствуют функциональным комбинациям, первая сумма в (5.1) соответствует интегральной части суммы. Дифференциальным комбинациям соответствуют третья и последующие части в (5.1) и вторая, третья и т. д. суммы в (5.2).

Если в (5.1) присутствуют только интегральные комбинации, формула численного дифференцирования называется *интегральной*, если только функциональные комбинации — *функциональной* (или *точечной*), а если только дифференциальные комбинации — *дифференциальной*. Если же в (5.1) отсутствуют интегральные комбинации, формула называется *функционально-дифференциальной*, если отсутствуют функциональные комбинации — *интегрально-дифференциальной*, если отсутствуют дифференциальные комбинации — *интегрально-функциональной*.

Аналогичная классификация справедлива и для формулы (5.2) численного интегрирования.

5.2. МЕТОДЫ ЧИСЛЕННОГО ДИФФЕРЕНЦИРОВАНИЯ

5.2.1. ФОРМУЛЫ, ПОЛУЧЕННЫЕ НА ОСНОВЕ РАЗЛОЖЕНИЯ ФУНКЦИЙ ПО ФОРМУЛЕ ТЕЙЛORA

Рассмотрим решение задач 1 и 2 численного дифференцирования на различных регулярных и нерегулярных шаблонах.

А. Двухточечный шаблон. Выберем шаблон $\mathcal{W}_{2,i} = (x_i, x_{i+1})$ на неравномерной сетке Ω_n . Предполагая, что $f(x) \in C_2[a, b]$, разложим функцию $f(x)$ по формуле Тейлора (B.20) при $k = 1$ относительно точки x_i с остаточным слагаемым в форме Лагранжа и найдем выражение для $f_{i+1} = f(x_{i+1})$:

$$f_{i+1} = f_i + h_{i+1} f'_i + \frac{h_{i+1}^2}{2} f''(\xi), \quad (5.3)$$

где $\xi \in (x_i, x_{i+1})$, $h_{i+1} = x_{i+1} - x_i$. Отсюда получаем

$$f'_i = \frac{f_{i+1} - f_i}{h_{i+1}} - \frac{h_{i+1}}{2} f''(\xi).$$

Очевидно, справедлива оценка

$$\left| -\frac{h_{i+1}}{2} f''(\xi) \right| \leq \frac{h_{i+1}}{2} \max_{[x_i, x_{i+1}]} |f''(x)| = \frac{h_{i+1} \cdot M_{2,i}}{2},$$

где $M_{2,i} = \max_{[x_i, x_{i+1}]} |f''(x)|$.

Отсюда следует *функциональная формула* (функциональный оператор) для первой производной:

$$\hat{f}'_{i,c} = \frac{\hat{f}_{i+1} - \hat{f}_i}{h_{i+1}} \quad \left(\frac{h_{i+1}}{2} M_{2,i} \right). \quad (5.4)$$

В скобках справа от аппроксимационных операторов здесь и далее указываются правые части оценок их погрешностей.

Отметим, что формула (5.4) является несимметричной, односторонней (левосторонней). Если функцию $f(x)$ разложить по формуле Тейлора относительно точки x_{i+1} , то получим правостороннюю формулу

$$\hat{f}'_{i+1,c} = \frac{\hat{f}_{i+1} - \hat{f}_i}{h} \quad \left(\frac{h}{2} M_{2,i} \right).$$

Б. Трехточечный шаблон. На неравномерной сетке Ω_n выбираем трехточечный (двуухшаговый) шаблон $\Pi_{3,i} = (x_{i-1}, x_i, x_{i+1})$, характеризующийся шагами $h_{i+1} = x_{i+1} - x_i$, $h_i = x_i - x_{i-1}$ и параметром его нерегулярности $\delta_{i+1} = \frac{h_{i+1}}{h_i}$, который в общем случае не равен единице.

Аппроксимационные функциональные (точечные) формулы второго порядка в левой крайней, центральной и правой крайней точках шаблона можно получить на основе разложения функции $f(x)$ по формуле Тейлора с остаточным слагаемым в форме Лагранжа. При этом предполагается, что $f(x) \in C_3[a, b]$. Это позволяет получить разностные дифференциальные операторы $\hat{f}'(x_t)$ ($t = i-1, i, i+1$) и провести оценки их погрешностей.

Разложим функцию $f(x)$ при $x = x_i$ и $x = x_{i+1}$ по формуле Тейлора при $k = 2$ относительно точки x_{i-1} с остаточным слагаемым в форме Лагранжа. В результате находим соотношения, определяющие $f_i = f(x_i)$ и $f_{i+1} = f(x_{i+1})$:

$$f_i = f_{i-1} + h_i f'_{i-1} + \frac{h_i^2}{2} f''_{i-1} + \frac{h_i^3}{6} f'''(\xi_-); \quad (5.5)$$

$$f_{i+1} = f_{i-1} + H_i^{i+1} f'_{i-1} + \frac{(H_i^{i+1})^2}{2} f''_{i-1} + \frac{(H_i^{i+1})^3}{6} f'''(\xi_+), \quad (5.6)$$

где $H_i^{i+1} = h_i + h_{i+1}$, $\xi_- \in (x_{i-1}, x_i)$, $\xi_+ \in (x_{i-1}, x_{i+1})$, $f_{i-1}^{(p)} = f^{(p)}(x_{i-1})$, $p = 0, 1, 2$.

Исключая из (5.5), (5.6) слагаемое, содержащее вторую производную, и выражая из полученного соотношения f'_{i-1} , получаем следующую аппроксимацию *первой производной в левой крайней точке* (левостороннюю формулу или оператор)

$$\hat{f}'_{i-1,v} = \frac{1}{H_i^{i+1}} \left(-(2 + \delta_{i+1}) f_{i-1} + \frac{(1 + \delta_{i+1})^2}{\delta_{i+1}} f_i - \delta_{i+1}^{-1} f_{i+1} \right). \quad (5.7)$$

При $h = \text{const}(\delta_{i+1} = 1)$ формула (5.7) упрощается и приводится к известному виду [6]:

$$\hat{f}'_{i-1,c} = \frac{1}{2h}(-3f_{i-1} + 4f_i - f_{i+1}) \quad \left(\frac{h^2}{3}M_{3,i}\right). \quad (5.8)$$

Данную формулу можно записать через конечные разности:

$$\hat{f}'_{i-1,c} = \frac{1}{2h}(3\Delta f_{i-1} - \Delta f_i).$$

Здесь нижние индексы v и c , относящиеся к аппроксимационным операторам (5.7) и (5.8), указывают на тип шаблона — *нерегулярный* ($h_{i+1} = \text{var}$) и *регулярный* ($h_{i+1} = \text{const}$). Остаточное слагаемое для (5.7) получается равным $\frac{h_i^2(1+\delta_{i+1})}{6}f'''(\xi)$, $\xi \in (x_{i-1}, x_{i+1})$, и поэтому для этой аппроксимации справедлива следующая оценка погрешности: $|f'_{i-1} - \hat{f}'_{i-1,v}| \leq \frac{h_i^2(1+\delta_{i+1})}{6}M_{3,i}$, где $M_{3,i} = \max_{x \in \mathbb{W}_{3,i}} (f'''(x))$.

Приводимые здесь и ниже остаточные слагаемые для дифференциальных операторов также получаются путем дифференцирования остаточных слагаемых $R(x)$ интерполяционных многочленов соответствующей степени. Из (4.20) для трехточечных формул при $n = 2$ следует соотношение

$$R(x) = \frac{1}{3!}f'''(\xi) \cdot \omega(x), \quad \omega(x) = (x - x_{i-1})(x - x_i)(x - x_{i+1}).$$

Аналогично, разложив функцию $f(x)$ относительно точки x_{i+1} и получив соотношения для f_{i-1} , f_i , найдем \hat{f}'_{i+1} — разностный оператор, аппроксимирующий *первую производную в правой крайней точке* (правосторонняя формула):

$$\begin{aligned} \hat{f}'_{i+1,v} &= \frac{1}{H_i^{i+1}} \left(\delta_{i+1}f_{i-1} - \frac{(1+\delta_{i+1})^2}{\delta_{i+1}}f_i + \frac{2\delta_{i+1}+1}{\delta_{i+1}}f_{i+1} \right); \\ \hat{f}'_{i+1,c} &= \frac{1}{2h}(f_{i-1} - 4f_i + 3f_{i+1}) \text{ или } \hat{f}'_{i+1,c} = \frac{1}{2h}(3\Delta f_i - \Delta f_{i-1}) \quad \left(\frac{h^2}{3}M_{3,i}\right). \end{aligned} \quad (5.9)$$

Оператор $\hat{f}'_{i+1,v}$ имеет остаточное слагаемое $\frac{h_i^2}{6}\delta_{i+1}(1+\delta_{i+1})f'''(\xi)$.

Разложение функции $f(x)$ относительно центральной точки x_i шаблона, получение выражений для f_{i-1} , f_{i+1} и исключение из них слагаемого со второй производной приводят к следующим разностным операторам функционального типа, аппроксимирующими *первую производную в центральной точке* (формула центрального вида):

$$\begin{aligned} \hat{f}'_{i,v} &= \frac{1}{H_i^{i+1}} \left(\delta_{i+1}\Delta f_{i-1} + \frac{1}{\delta_{i+1}}\Delta f_i \right) = \frac{1}{H_i^{i+1}} \left(-\delta_{i+1}f_{i-1} + \frac{(\delta_{i+1}^2 - 1)}{\delta_{i+1}}f_i + \frac{1}{\delta_{i+1}}f_{i+1} \right); \\ \hat{f}'_{i,c} &= \frac{1}{2h}(f_{i+1} - f_{i-1}) \text{ или } \hat{f}'_{i,c} = \frac{1}{2h}(\Delta f_i + \Delta f_{i-1}) \quad \left(\frac{h^2}{6}M_{3,i}\right). \end{aligned} \quad (5.10)$$

Оператор $\hat{f}'_{i,v}$ имеет остаточное слагаемое $\frac{h_i^2}{6} \delta_{i+1} f'''(\xi)$.

Приведенные остаточные слагаемые разностных операторов обусловливают следующие оценки их погрешностей:

$$\begin{aligned} |f'_{i-1} - \hat{f}'_{i-1,v}| &\leq \frac{h_i^2}{6} (1 + \delta_{i+1}) M_{3,i}, \quad |f'_i - \hat{f}'_{i,v}| \leq \frac{h_i^2}{6} \delta_{i+1} M_{3,i}; \\ |f'_{i+1} - \hat{f}'_{i+1,v}| &\leq \frac{h_i^2}{6} (1 + \delta_{i+1}) \delta_{i+1} M_{3,i}, \quad M_{3,i} = \max_{x \in \mathbb{W}_{3,i}} |f'''(x)|. \end{aligned} \quad (5.11)$$

Замечания

1. Далее в тексте оценочная константа $M_{p,i}$ для краткости будет использоваться без дополнительного ее описания. В нижнем индексе этой константы всюду указывается p — порядок производной.

2. Из оценок (5.11) вытекает, что разностные операторы \hat{f}_{i-1} , \hat{f}_i , \hat{f}_{i+1} аппроксимируют при $h_{i+1} = \text{var}$ соответствующие производные f'_{i-1} , f'_i , f'_{i+1} со вторым порядком, если шаблон произвольный (*безусловная аппроксимация*). Если же на шаблоне с $\delta_{i+1} \ll 1$ наложить условие, например $\delta_{i+1} \leq h_i$, т. е. $h_{i+1} \leq h_i^2$, то порядок аппроксимации \hat{f}'_i , \hat{f}'_{i+1} может быть повышен до третьего (*условная аппроксимация*). Такая возможность повышения порядка аппроксимации относительно h_i без увеличения количества точек шаблона обеспечивается введением в мажоранты, соответствующие аппроксимациям, параметра δ_{i+1} , на который в случае необходимости можно наложить условие $\delta_{i+1} \leq h_i$. При этом следует иметь в виду, что данный параметр входит в знаменатель некоторых слагаемых аппроксимационных формул (5.9), (5.10) и при его уменьшении увеличиваются погрешности арифметических операций.

3. Аппроксимации (5.8), (5.9) являются условными, так как для них справедливо условие $\delta_{i+1} = 1$.

Предположим, что $f(x) \in C_4[a, b]$, и разложим функцию $f(x)$ в точках x_i , x_{i+1} на трехточечном нерегулярном шаблоне $\mathbb{W}_{3,i} = (x_{i-1}, x_i, x_{i+1})$ до слагаемого четвертого порядка относительно шага. Складывая эти разложения и выражая из суммы вторую производную, получаем *функционально-дифференциальную формулу для второй производной*:

$$f''_i = \frac{2(f_{i-1} - 2f_i + f_{i+1})}{h_{i+1}^2 + h_i^2} - \frac{2(h_{i+1} - h_i)}{h_{i+1}^2 + h_i^2} f'_i - \frac{2(h_{i+1}^3 - h_i^3)}{6(h_{i+1}^2 + h_i^2)} f'''_i - \frac{2(h_{i+1}^4 + h_i^4)}{4!(h_{i+1}^2 + h_i^2)} f^{(4)}(\xi). \quad (5.12)$$

Подставляя в (5.12) аппроксимационную формулу (5.10) для первой производной, находим разностный аппроксимационный оператор \hat{f}''_i , выраженный через параметры δ_{i+1} , h_i^2 и *аппроксимирующий* (безусловно) *вторую производную* f''_i на нерегулярном шаблоне с первым порядком:

$$\hat{f}''_{i,v} = \frac{2}{h_i^2} \left(\frac{1}{1 + \delta_{i+1}} f_{i-1} - \frac{1}{\delta_{i+1}} f_i + \frac{1}{(1 + \delta_{i+1}) \delta_{i+1}} f_{i+1} \right). \quad (5.13)$$

Выражение (5.13) можно преобразовать к виду

$$\hat{f}''_{i,v} = \frac{2}{H_i^{i+1}} \left(\frac{\Delta f_i}{h_{i+1}} - \frac{\Delta f_{i-1}}{h_i} \right) = \frac{2}{H_i^{i+1}} \left[\frac{f_{i-1}}{h_i} - \left(\frac{1}{h_i} + \frac{1}{h_{i+1}} \right) f_i + \frac{f_{i+1}}{h_{i+1}} \right].$$

Если сетка равномерная ($\delta_{i+1} = 1$), то указанный порядок условной аппроксимации возрастает на единицу, так как третье слагаемое в (5.12) становится равным нулю. В этом случае из (5.13) получаем широко распространенный оператор, аппроксимирующий вторую производную на регулярном шаблоне [6]:

$$\hat{f}_{i,c}'' = \frac{1}{h^2}(f_{i-1} - 2f_i + f_{i+1}) \text{ или } \hat{f}_{i,c}'' = \frac{1}{h^2}(\Delta f_i - \Delta f_{i-1}) \quad \left(\frac{h^2}{12} M_{4,i} \right). \quad (5.14)$$

Замечание. Из (5.12) следует, что порядок условной аппроксимации (5.13) можно повысить на единицу и на нерегулярном шаблоне, если принять $|\delta_{i+1} - 1| < h_i$, т. е. $h_i - h_i^2 \leq h_{i+1} \leq h_i + h_i^2$ (квазиравномерная сетка).

Как следует из приведенных выше постановок задач, в вычислительной практике аппроксимационные формулы (операторы) для производных используются для вычисления значений производных либо для замены ими соответствующих дифференциальных операторов. В связи с этим приведем методику вычисления $f^{(p)}(x)|_{x=x_j}$.

Методика вычисления значений производных

1. Выбрать конкретную аппроксимационную формулу (или несколько разных формул), в которой порядок аппроксимации должен соответствовать заданному в задаче порядку точности t .

2. Выбрать наборы точек (шаблоны $\Pi_{k,i}$), которым принадлежат точки x_j ($x_j \in \Pi_{k,i}$), причем для каждого из наборов расположение точки x_j должно быть зафиксировано относительно точек шаблона (k — количество точек, определяющих шаблон). Эта фиксация определяется структурой формулы. Например, если формула имеет центральный тип (см. формулу (5.10)), то точка x_j должна совпадать со средней точкой шаблона, а если формула левосторонняя (см. формулы (5.7) и (5.8)), то точка x_j должна совпадать с левой крайней точкой шаблона и т. д.

3. В правую часть выбранной формулы (или формул) подставить значения функций и (или) интегралов, которые соответствуют выбранным точкам шаблона (шаблонов).

4. Произвести требуемые вычисления с учетом того, что количество сохраняемых цифр должно приблизительно соответствовать величине остаточного слагаемого аппроксимационной формулы и порядку точности t .

Пример 5.1. Данна сеточная функция (табл. 5.1), являющаяся сеточным представлением формульной функции $y(x) = \frac{1}{x}$.

Заданы также порядок $t = 2$ относительно шага h , который необходимо обеспечить при решении задачи, и точка $x_j = 1,4$.

Таблица 5.1

i	0	1	2	3	4	5
x_i	1	1,2	1,4	1,6	1,8	2,0
f_i	1,000000	0,83333333	0,7142857	0,6250000	0,5555555	0,500000

Требуется вычислить значение первой производной $f'(1,4)$ и второй производной $f''(1,4)$ с помощью различных шаблонов и соответствующих формул.

□ Воспользуемся вышеприведенной методикой.

1. Так как шаг задания сеточной функции постоянный $h = x_{i+1} - x_i = 0,2$, точка $x_j = 1,4$ находится внутри сетки Ω_n , то для вычисления производной в этой точке выбирается вторая формула из (5.10), имеющая второй порядок аппроксимации относительно шага h . При этом центральная точка шаблона совпадает с точкой $x_j = 1,4$.

2. Выберем трехточечный шаблон $\Pi_{3,i} = (x_{i-1}, x_i, x_{i+1}) = (1,2; 1,4; 1,6)$, в котором $x_i = 1,4$ ($i = 2$); $x_{i-1} = 1,2$ ($i - 1 = 1$); $x_{i+1} = 1,6$ ($i + 1 = 3$). В данном шаблоне центральная точка $x_i = 1,4$, что соответствует центральному типу аппроксимационной формулы.

3. Подсчитаем искомое значение производной по формуле (5.10):

$$\hat{f}'_{i,c} = \frac{f_{i+1} - f_{i-1}}{2h} = \frac{0,6250000 - 0,8333333}{2 \cdot 0,2}.$$

4. Прежде чем выполнить вычисление, необходимо определить количество знаков, которое сохраняется при этом. Остаточное слагаемое выбранной формулы равно $\frac{h^2}{6} M_{3,i}$. Для его вычисления необходимо сначала определить $M_{3,i} = \max_{[x_{i-1}, x_{i+1}]} |f'''(x)|$. Поэтому воспользуемся интерполяционным многочленом Ньютона с конечными разностями:

$$f'''_i \approx N_3'''(x_i) = \frac{\Delta^3 f}{h^3},$$

где $\Delta^3 f$ — конечная разность третьего порядка. Эта разность может быть вычислена по значениям функции f_i в четырех точках. Возьмем точки x_2, x_3, x_4, x_5 . При этом будем считать, что $M_{3,i} \approx f'''(x_i)$ (порядок вычислений конечных разностей описан в главе 4 и отображен в таблице 4.2).

Вычисление дает

$$f'''(x_i) \approx -\frac{0,005935}{0,008} = -0,741875.$$

Тогда остаточное слагаемое по модулю будет равно

$$\frac{0,04 \cdot 0,74405}{6} \approx 0,0049 < 0,01.$$

На основе полученного приближенного значения остаточного слагаемого можно заключить, что в вычислениях ожидается одна верная цифра после запятой. Обычно в расчетах оставляют еще одну или две дополнительные цифры (в нашем примере это составляет всего 3 цифры). Оставляя три цифры после запятой, получаем результат:

$$\hat{f}'(x)|_{x=1,4} = -0,521.$$

Фактическая абсолютная погрешность составляет

$$\left| -0,521 + \frac{1}{1,4^2} \right| = \left| -0,521 + 0,5102 \right| = 0,0108,$$

т. е. относительная погрешность равна

$$\frac{0,0108}{0,5102} \cdot 100\% = 2,1\%.$$

Если эта погрешность не устраивает вычислителя, необходимо повышать порядок точности относительно h , например, до $t = 3$. В дальнейшем приводятся соответствующие примеры с порядком $t = 3$.

Для вычисления первой производной можно было использовать и другие формулы. При выборе шаблона $\Pi_{3,i} = (x_{i-1}, x_i, x_{i+1}) = (1,4; 1,6; 1,8)$ по формуле (5.8) имеем

$$\begin{aligned} f'(1,4) &\approx \hat{f}'_{i-1,c} = \frac{1}{2h}(-3f_{i-1} + 4f_i - f_{i+1}) = \frac{1}{2h}[-3f(1,4) + 4f(1,6) - f(1,8)] = \\ &= \frac{1}{2 \cdot 0,2}[-3 \cdot 0,7142857 + 4 \cdot 0,625 - 0,5555] = -0,496017. \end{aligned}$$

Фактическая абсолютная погрешность составляет $| -0,496017 - 0,510204 | \approx 0,0142$, относительная погрешность равна 2,78%.

Если выбрать шаблон $\Pi_{3,i} = (x_{i-1}, x_i, x_{i+1}) = (1; 1,2; 1,4)$, то по формуле (5.9) получаем

$$\begin{aligned} f'(1,4) &\approx \hat{f}'_{i+1,c} = \frac{1}{2h}(f_{i-1} - 4f_i + 3f_{i+1}) = \frac{1}{2 \cdot 0,2}[f(1) - 4f(1,2) + 3f(1,4)] = \\ &= \frac{1}{0,4}[1,0 - 4 \cdot 0,83333 + 3 \cdot 0,7142857] = -0,476187. \end{aligned}$$

Фактическая абсолютная погрешность равна $| -0,476187 - 0,510204 | \approx 0,03401$, относительная погрешность составляет 6,66%.

Для вычисления второй производной можно взять формулу (5.14) на шаблоне $\Pi_{3,i} = (x_{i-1}, x_i, x_{i+1}) = (1,2; 1,4; 1,6)$:

$$\begin{aligned} \hat{f}''_{i,c} &= \frac{1}{h^2}(f_{i-1} - 2f_i + f_{i+1}) = \frac{1}{0,2^2}[f(1,2) - 2 \cdot f(1,4) + f(1,6)] = \\ &= \frac{1}{0,04}[0,8333 - 2 \cdot 0,7142857 + 0,625] = 0,743965. \end{aligned}$$

Точное значение $f''(1,4) = \frac{2}{1,4^3} = 0,7288629$. Фактическая абсолютная погрешность равна 0,0151, относительная погрешность 2,07%. ■

В вычислительной практике иногда применяют аппроксимационные формулы, порядок аппроксимации которых не известен или не приведен в используемом источнике, и в этом случае, прежде чем использовать эту формулу, нужно получить оценку ее погрешности.

Данную процедуру можно выполнить с помощью различных подходов, один из которых основан на разложении функций, входящих в правую часть оператора, по формуле Тейлора относительно той точки x_j , для которой записан этот оператор. Другой подход использует анализ остаточного слагаемого, полученного для интерполяционного многочлена Лагранжа. При этом рассматривается соотношение (4.19), которое дифференцируется нужное количество раз:

$$f^{(p)}(x) = L_n^{(p)} + R_n^{(p)}(x).$$

Тогда если нужно найти погрешность численного дифференцирования в точке $x = x_j$, то осуществляется подстановка $x = x_j$. В результате находится оценка погрешности в точке $x = x_j$:

$$|R_n^{(p)}(x)|_{x=x_j} \leq \frac{M_{n+1}}{(n+1)!} \omega_n^{(p)}(x_j),$$

где n — степень алгебраического многочлена, по которому получен аппроксимационный оператор. Если априори степень n неизвестна, то она может быть определена путем подстановки в правую часть оператора произвольных узлов x_i и значений $f(x_i)$ для многочленов $N_1(x) = ax + b$; $N_2(x) = ax^2 + bx + c$ и т. д. Максимальная степень многочлена $N_n(x)$, для которого остаточное слагаемое равно нулю, является искомой.

Если необходимо оценить погрешность аппроксимационного оператора $\hat{f}^{(p)}(x)$ не в точке, а на всем отрезке $[a, b]$, то для этого используется неравенство (4.22), которое дифференцируется p раз:

$$\max_{[a, b]} |R_n^{(p)}(x)| \leq \frac{M_{n+1}}{(n+1)!} \max_{[a, b]} |\omega_n^{(p)}(x)|.$$

Изложим простейшую методику оценки погрешности в точке $x = x_j$, когда степень многочлена $L_{k-1}(x)$ и шаблон $\mathbb{W}_{k,i}$, на котором этот многочлен получен, известны.

Методика оценки погрешности аппроксимационного оператора

1. На заданном шаблоне, который в общем случае имеет структуру $\mathbb{W}_{k,i} = (x_{i-r}, \dots, x_i, \dots, x_{i+s})$, $k = r + s + 1$, записать выражение для оценки остаточного слагаемого в точке $x = x_j \in \mathbb{W}_{k,i}$:

$$|R_{k-1}^{(p)}(x)|_{x=x_j} \leq \frac{M_k}{k!} |\omega_{k-1}^{(p)}(x)|_{x=x_j}.$$

Здесь r и s — количество точек, расположенных соответственно левее и правее точки x_j , которая фиксируется на этом шаблоне.

2. Найти производную $\omega_{k-1}^{(p)}(x) = [(x - x_{i-r})(x - x_{i-r+1}) \dots (x - x_{i+s})]^{(p)}$.

3. В полученную производную подставить значение x_j . Далее преобразовать ее, выразив через h (при $h = \text{const}$) или через параметр нерегулярности $\delta_{i+1} = h_{i+1}/h_i$, и записать окончательно оценку погрешности.

Пример 5.2. Для заданного дифференциального оператора (5.8), записанного для производной в левой крайней точке шаблона $\Pi_{3,i} = (x_{i-1}, x_i, x_{i+1})$, требуется найти остаточное слагаемое и оценить погрешность в точке $x_j = x_{i-1}$.

□1. На шаблоне $\Pi_{3,i} = (x_{i-1}, x_i, x_{i+1})$ записываем выражение ($k = 3$):

$$\begin{aligned}|R'_2(x)|_{x=x_{i-1}} &\leq \frac{M_3}{3!} |\omega'_2(x)|_{x=x_{i-1}} \text{ или} \\|R'_2(x)|_{x=x_{i-1}} &\leq \frac{M_3}{3!} |[(x-x_{i-1})(x-x_i)(x-x_{i+1})']|_{x=x_{i-1}}.\end{aligned}$$

2. Находим производную многочлена $\omega_2(x)$:

$$\omega'_2(x) = (x - x_i)(x - x_{i+1}) + (x - x_{i-1})(x - x_{i+1}) + (x - x_{i-1})(x - x_i).$$

3. Так как $h = \text{const}$, то

$$\begin{aligned}\omega'_2(x)|_{x=x_{i-1}} &= (x_{i-1} - x_i)(x_{i-1} - x_{i+1}) + (x_{i-1} - x_{i-1})(x_{i-1} - x_i) + \\&+ (x_{i-1} - x_{i-1})(x_{i-1} - x_i) = (-h)(-2h) = 2h^2.\end{aligned}$$

В результате получаем

$$|R'_2(x)|_{x=x_{i-1}} \leq \frac{M_3}{3} h^2.$$

Именно эта оценка и приведена в скобках справа от формулы (5.8). ■

В. Четырехточечный шаблон. Формулы третьего порядка для первых производных на регулярном шаблоне $\Pi_{4,i} = (x_{i-2}, x_{i-1}, x_i, x_{i+1})$ имеют вид [6]:

$$\hat{f}'_{i-2,c} = \frac{1}{6h} (-11f_{i-2} + 18f_{i-1} - 9f_i + 2f_{i+1}) \quad \left(\frac{h^3}{4} M_{4,i} \right),$$

или

$$\begin{aligned}\hat{f}'_{i-2,c} &= \frac{1}{6h} (11\Delta f_{i-2} - 7\Delta f_{i-1} + 2\Delta f_i); \\ \hat{f}'_{i-1,c} &= \frac{1}{6h} (-2f_{i-2} - 3f_{i-1} + 6f_i - f_{i+1}) \quad \left(\frac{h^3}{12} M_{4,i} \right),\end{aligned}$$

или

$$\begin{aligned}\hat{f}'_{i-1,c} &= \frac{1}{6h} (2\Delta f_{i-2} + 5\Delta f_{i-1} - \Delta f_i); \\ \hat{f}'_{i,c} &= \frac{1}{6h} (f_{i-2} - 6f_{i-1} + 3f_i + 2f_{i+1}) \quad \left(\frac{h^3}{12} M_{4,i} \right),\end{aligned}\tag{5.15}$$

или

$$\begin{aligned}\hat{f}'_{i,c} &= \frac{1}{6h} (-\Delta f_{i-2} + 5\Delta f_{i-1} + 2\Delta f_i); \\ \hat{f}'_{i+1,c} &= \frac{1}{6h} (-2f_{i-2} + 9f_{i-1} - 18f_i + 11f_{i+1}) \quad \left(\frac{h^3}{4} M_{4,i} \right),\end{aligned}$$

или

$$\hat{f}'_{i+1,c} = \frac{1}{6h} (2\Delta f_{i-2} - 7\Delta f_{i-1} + 11\Delta f_i).$$

Замечание. Вариант записи производных через конечные разности здесь и выше приведен для того, чтобы в дальнейшем можно было преобразовать эти формулы на основе теории подобия для аппроксимации (восстановления) функций по интегралам (см. замечание в конце параграфа).

Формулы второго порядка на регулярном шаблоне для вторых производных имеют вид

$$\begin{aligned}\hat{f}_{i-2,c}'' &= \frac{1}{h^2}(2f_{i-2} - 5f_{i-1} + 4f_i - f_{i+1}) \quad \left(\frac{11h^2}{12}M_{4,i}\right); \\ \hat{f}_{i-1,c}'' &= \frac{1}{h^2}(f_{i-2} - 2f_{i-1} + f_i) \quad \left(\frac{h^2}{12}M_{4,i}\right); \\ \hat{f}_{i,c}'' &= \frac{1}{h^2}(f_{i-1} - 2f_i + f_{i+1}) \quad \left(\frac{h^2}{12}M_{4,i}\right); \\ \hat{f}_{i+1,c}'' &= \frac{1}{h^2}(-f_{i-2} + 4f_{i-1} - 5f_i + 2f_{i+1}) \quad \left(\frac{11h^2}{12}M_{4,i}\right).\end{aligned}\tag{5.16}$$

Г. Пятиточечный шаблон. Формулы четвертого порядка для первых производных на регулярном шаблоне $\Pi_{5,i} = (x_{i-2}, x_{i-1}, x_i, x_{i+1}, x_{i+2})$ имеют вид

$$\begin{aligned}\hat{f}_{i-2,c}' &= \frac{1}{12h}(-25f_{i-2} + 48f_{i-1} - 36f_i + 16f_{i+1} - 3f_{i+2}) \quad \left(\frac{h^4}{5}M_{5,i}\right); \\ \hat{f}_{i-1,c}' &= \frac{1}{12h}(-3f_{i-2} - 10f_{i-1} + 18f_i - 6f_{i+1} + f_{i+2}) \quad \left(\frac{h^4}{20}M_{5,i}\right); \\ \hat{f}_{i,c}' &= \frac{1}{12h}(f_{i-2} - 8f_{i-1} + 8f_{i+1} - f_{i+2}) \quad \left(\frac{h^4}{30}M_{5,i}\right); \\ \hat{f}_{i+1,c}' &= \frac{1}{12h}(-f_{i-2} + 6f_{i-1} - 18f_i + 10f_{i+1} + 3f_{i+2}) \quad \left(\frac{h^4}{20}M_{5,i}\right); \\ \hat{f}_{i+2,c}' &= \frac{1}{12h}(3f_{i-2} - 16f_{i-1} + 36f_i - 48f_{i+1} + 25f_{i+2}) \quad \left(\frac{h^4}{5}M_{5,i}\right).\end{aligned}\tag{5.17}$$

Формулы третьего порядка для вторых производных на указанном шаблоне имеют вид

$$\begin{aligned}\hat{f}_{i-2,c}'' &= \frac{1}{12h^2}(35f_{i-2} - 104f_{i-1} + 114f_i - 56f_{i+1} + 11f_{i+2}); \\ \hat{f}_{i-1,c}'' &= \frac{1}{12h^2}(11f_{i-2} - 20f_{i-1} + 6f_i + 4f_{i+1} - f_{i+2}); \\ \hat{f}_{i,c}'' &= \frac{1}{12h^2}(-f_{i-2} + 16f_{i-1} - 30f_i + 16f_{i+1} - f_{i+2}); \\ \hat{f}_{i+1,c}'' &= \frac{1}{12h^2}(-f_{i-2} + 4f_{i-1} + 6f_i - 20f_{i+1} + 11f_{i+2}); \\ \hat{f}_{i+2,c}'' &= \frac{1}{12h^2}(11f_{i-2} - 56f_{i-1} + 114f_i - 104f_{i+1} + 35f_{i+2}).\end{aligned}\tag{5.18}$$

Пример 5.3. Для сеточной функции из примера 5.1 вычислить значение первой производной $f'(x)|_{x_j=1,4}$ и второй производной $f''(x)|_{x_j=1,4}$, используя каждую из приведенных выше формул для четырехточечного и пятиточечного шаблонов.

Для вычисления производных воспользуемся соответствующей методикой.

Используем сначала четырехточечные шаблоны.

Для шаблона $\Pi_{4,i} = (x_{i-2}, x_{i-1}, x_i, x_{i+1}) = (1,4; 1,6; 1,8; 2)$ по первой из формул (5.15) при $x_{i-2} = x_j$ имеем

$$\begin{aligned}\hat{f}'_{i-2,c} &= \frac{1}{6h}(-11f_{i-2} + 18f_{i-1} - 9f_i + 2f_{i+1}) = \frac{1}{6 \cdot 0,2}[-11f(1,4) + 18f(1,6) - 9f(1,8) + 2f(2)] = \\ &= \frac{1}{1,2}[-11 \cdot 0,7142857 + 18 \cdot 0,625 - 9 \cdot 0,5555 + 2 \cdot 0,5] = -0,505535.\end{aligned}$$

Фактическая абсолютная ошибка равна 0,00467, относительная погрешность 0,915%.

Значение второй производной найдем по первой из формул (5.16) при $x_{i-2} = x_j$:

$$\begin{aligned}\hat{f}''_{i-2,c} &= \frac{1}{h^2}(2f_{i-2} - 5f_{i-1} + 4f_i - f_{i+1}) = \frac{1}{0,2^2}[2f(1,4) - 5f(1,6) + 4f(1,8) - f(2)] = \\ &= \frac{1}{0,04}[2 \cdot 0,7142857 - 5 \cdot 0,625 + 4 \cdot 0,5555 - 0,5] = 0,639285.\end{aligned}$$

Фактическая абсолютная ошибка равна 0,10468, относительная погрешность 14,07%.

Заметим, что величина мажоранты в оценке остаточного слагаемого для использованной формулы в 11 раз больше, чем для двух других формул в (5.16), аппроксимирующих производные в точках, расположенных внутри шаблона.

Для шаблона $\Pi_{4,i} = (x_{i-2}, x_{i-1}, x_i, x_{i+1}) = (1,2; 1,4; 1,6; 1,8)$ по второй из формул (5.15) при $x_{i-1} = x_j$ получаем

$$\begin{aligned}\hat{f}'_{i-1,c} &= \frac{1}{6h}(-2f_{i-2} - 3f_{i-1} + 6f_i - f_{i+1}) = \frac{1}{6 \cdot 0,2}[-2f(1,2) - 3f(1,4) + 6f(1,6) - f(1,8)] = \\ &= \frac{1}{1,2}[-2 \cdot 0,8333 - 3 \cdot 0,7142857 + 6 \cdot 0,625 - 0,5555] = -0,51256.\end{aligned}$$

Фактическая абсолютная погрешность равна 0,002357, относительная погрешность 0,462%.

Значение второй производной найдем по формуле

$$\hat{f}''_{i-1,c} = \frac{1}{h^2}(f_{i-2} - 2f_{i-1} + f_i) = \frac{1}{0,2^2}[f(1,2) - 2f(1,4) + f(1,6)] = 0,743965.$$

Фактическая абсолютная погрешность равна 0,0151, относительная погрешность 2,07% (см. пример 5.1).

Для шаблона $\Pi_{4,i} = (x_{i-2}, x_{i-1}, x_i, x_{i+1}) = (1; 1,2; 1,4; 1,6)$ при $x_i = x_j$ имеем

$$\begin{aligned}\hat{f}'_{i,c} &= \frac{1}{6h}(f_{i-2} - 6f_{i-1} + 3f_i + 2f_{i+1}) = \frac{1}{6 \cdot 0,2}[f(1) - 6f(1,2) + 3f(1,4) + 2f(1,6)] = \\ &= \frac{1}{1,2}[1,0 - 6 \cdot 0,8333 + 3 \cdot 0,7142857 + 2 \cdot 0,625] = -0,505785.\end{aligned}$$

Фактическая абсолютная ошибка равна 0,00442, относительная погрешность 0,866%.

Значение второй производной найдем по формуле

$$\hat{f}''_{i,c} = \frac{1}{h^2}(f_{i-1} - 2f_i + f_{i+1}) = \frac{1}{0,2^2}[f(1,2) - 2f(1,4) + f(1,6)] = 0,743965.$$

Фактическая абсолютная погрешность равна 0,0151, относительная погрешность 2,07% (см. пример 5.1).

Используем формулы (5.17), (5.18) для пятиточечного шаблона.

Для шаблона $\Pi_{5,i} = (x_{i-2}, x_{i-1}, x_i, x_{i+1}, x_{i+2}) = (1,2; 1,4; 1,6; 1,8; 2)$ имеем

$$\begin{aligned}\hat{f}'_{i-1,c} &= \frac{1}{12h}(-3f_{i-2} - 10f_{i-1} + 18f_i - 6f_{i+1} + f_{i+2}) = \\ &= \frac{1}{12 \cdot 0,2}[-3f(1,2) - 10f(1,4) + 18f(1,6) - 6f(1,8) + f(2)] = \\ &= \frac{1}{2,4}[-3 \cdot 0,83333 - 10 \cdot 0,7142857 + 18 \cdot 0,625 - 6 \cdot 0,555 + 0,5] = -0,5107696.\end{aligned}$$

Фактическая абсолютная погрешность равна 0,000566, относительная погрешность 0,11%.

Значение второй производной вычислим по формуле

$$\begin{aligned}\hat{f}''_{i-1,c} &= \frac{1}{12h^2}(11f_{i-2} - 20f_{i-1} + 6f_i + 4f_{i+1} - f_{i+2}) = \\ &= \frac{1}{12 \cdot 0,2^2}[11f(1,2) - 20f(1,4) + 6f(1,6) + 4f(1,8) - f(2)] = \\ &= \frac{1}{0,48}[11 \cdot 0,83333 - 20 \cdot 0,7142857 + 6 \cdot 0,625 + 4 \cdot 0,555 - 0,5] = 0,735241.\end{aligned}$$

Фактическая абсолютная погрешность равна 0,00637, относительная погрешность 0,875%.

Для шаблона $\Pi_{5,i} = (x_{i-2}, x_{i-1}, x_i, x_{i+1}, x_{i+2}) = (1; 1,2; 1,4; 1,6; 1,8)$ получаем

$$\begin{aligned}\hat{f}'_{i,c} &= \frac{1}{12h}(f_{i-2} - 8f_{i-1} + 8f_{i+1} - f_{i+2}) = \frac{1}{12 \cdot 0,2}[f(1) - 8f(1,2) + 8f(1,6) - f(1,8)] = \\ &= \frac{1}{2,4}[1 - 8 \cdot 0,8333 + 8 \cdot 0,625 - 0,555] = -0,5092.\end{aligned}$$

Фактическая абсолютная погрешность равна 0,00098, относительная погрешность 0,192%.

Вычислим вторую производную по формуле

$$\begin{aligned}\hat{f}_{i,c}'' &= \frac{1}{12h^2}(-f_{i-2} + 16f_{i-1} - 30f_i + 16f_{i+1} - f_{i+2}) = \\ &= \frac{1}{12 \cdot 0,2^2}[-f(1) + 16f(1,2) - 30f(1,4) + 16f(1,6) - f(1,8)] = \\ &= \frac{1}{0,48}[-1,0 + 16 \cdot 0,8333 - 30 \cdot 0,7142857 + 16 \cdot 0,625 - 0,5555] = 0,72751.\end{aligned}$$

Фактическая абсолютная погрешность равна 0,00134, относительная погрешность 0,18%.

Анализ точности аппроксимации показывает, что значения производных в точках, расположенных вблизи центра шаблона, вычисляются с большей точностью, чем значения производных в точках, расположенных вблизи концов шаблона. При повышении порядка аппроксимации точность вычисления производных возрастает. ■

Замечание. Формулы, записанные выше для аппроксимации производных через приращения функций, могут быть переписаны для восстановления функции $y = f(x)$ по значениям интегралов. С этой целью можно использовать изложенный в главе 4 метод подобия, состоящий в соответствующем изменении порядка производной в левой и правой частях аппроксимационного выражения. Так, из оператора \hat{f}'_k (k — номер точки шаблона) можно получить оператор \hat{f}_k путем замены $\Delta f_{k-1} = f_k - f_{k-1}$ на интеграл $I_{k-1}^k = F_k - F_{k-1}$. Проделав это, вместо формулы, следующей за (5.8), вторых из формул (5.9), (5.10) и формул (5.15) получим операторы, восстанавливающие функцию $y = f(x)$ по значениям интегралов:

$$\begin{aligned}\hat{f}_{i-1,c} &= \frac{1}{2h}(3I_{i-1}^i - I_i^{i+1}); \quad \hat{f}_{i,c} = \frac{1}{2h}(I_i^{i+1} + I_{i-1}^i); \\ \hat{f}_{i-2,c} &= \frac{1}{6h}(11I_{i-2}^{i-1} - 7I_{i-1}^i + 2I_i^{i+1}); \quad \hat{f}_{i-1,c} = \frac{1}{6h}(2I_{i-2}^{i-1} + 5I_{i-1}^i - I_i^{i+1}); \\ \hat{f}_{i,c} &= \frac{1}{6h}(-I_{i-2}^{i-1} + 5I_{i-1}^i + 2I_i^{i+1}); \quad \hat{f}_{i+1,c} = \frac{1}{6h}(2I_{i-2}^{i-1} - 7I_{i-1}^i + 11I_i^{i+1}).\end{aligned}$$

Первые две формулы имеют второй порядок аппроксимации по h (они записаны на трехточечном шаблоне), а последние четыре — третий порядок (они записаны на четырехточечном шаблоне).

Вместо формулы (5.14) имеем следующую:

$$\hat{f}_{i,c}' = \frac{1}{h^2}(I_i^{i+1} - I_{i-1}^i).$$

В п. 5.2.2 эта формула выведена другим способом.

Аналогичным путем можно найти и другие интегральные аппроксимации первой производной.

5.2.2. ФОРМУЛЫ, ПОЛУЧЕННЫЕ НА ОСНОВЕ РАЗЛОЖЕНИЯ ПЕРВООБРАЗНЫХ ПО ФОРМУЛЕ ТЕЙЛОРА

А. Двухточечный шаблон. На неравномерной сетке Ω_n рассмотрим двухточечный шаблон $\Pi_{2,i} = (x_i, x_{i+1})$ с шагом $h_{i+1} = x_{i+1} - x_i$. Предположив, что $f(x) \in C_2[a, b]$, разложим первообразную $F(x)$ относительно точки x_i по формуле Тейлора при $k = 2$ и найдем выражение для $F(x)|_{x=x_{i+1}} = F_{i+1}$:

$$F_{i+1} = F_i + h_{i+1} f_i + \frac{h_{i+1}^2}{2} f'_i + \frac{h_{i+1}^3}{6} f''(\xi),$$

где $\xi \in (x_i, x_{i+1})$. Выражая из этого разложения сначала первую производную, а затем интеграл $I_i^{i+1} = F_{i+1} - F_i$, получаем

$$f'_i = \frac{2I_i^{i+1}}{h_{i+1}^2} - \frac{2f_i}{h_{i+1}} - \frac{h_{i+1}}{3} f''(\xi); \quad I_i^{i+1} = h_{i+1} f_i + \frac{h_{i+1}^2}{2} f'_i + \frac{h_{i+1}^3}{6} f''(\xi).$$

Первые два слагаемых в правых частях последних двух соотношений представляют собой *интегрально-функциональную* (интегрально-точечную) аппроксимацию производной f'_i и *функционально-дифференциальную* аппроксимацию интеграла I_i^{i+1} соответственно:

$$\hat{f}'_i = \frac{2I_i^{i+1}}{h_{i+1}^2} - \frac{2f_i}{h_{i+1}} \quad \left(\frac{h_{i+1}}{3} M_2 \right); \quad (5.19)$$

$$\hat{I}_i^{i+1} = h_{i+1} f_i + \frac{h_{i+1}^2}{2} f'_i \quad \left(\frac{h_{i+1}^3}{6} M_2 \right). \quad (5.20)$$

Порядки этих аппроксимаций устанавливают остаточные слагаемые в выражениях для f'_i и I_i^{i+1} , из которых следуют оценки, правые части которых указаны в скобках рядом с формулами.

Данные оценки свидетельствуют о том, что порядок аппроксимации, обеспечиваемый (5.19), равен единице, а обеспечиваемый (5.20) равен трем при условии, что I_i^{i+1} и f'_i для (5.19) известны с точностью не ниже $O(h_{i+1}^3)$ и $O(h_{i+1}^2)$, а f_i и f'_i для (5.20) с точностью не ниже $O(h_{i+1}^2)$ и $O(h_{i+1})$. Этот же результат следует из рассмотренного п. В.4 правила соответствия порядков аппроксимации математических моделей различного типа.

Замечания

1. Выразив из разложения для F_{i+1} непосредственно функцию f_i , получим еще одно аппроксимационное выражение:

$$\hat{f}_i = \frac{1}{h_{i+1}} I_i^{i+1} - \frac{h_{i+1}}{2} f'_i,$$

позволяющее восстанавливать со вторым порядком аппроксимации значение функции f_i в точке x_i по значениям I_i^{i+1} и f'_i , известным с точностью не ниже $O(h_{i+1}^3)$ и $O(h_{i+1})$ соответственно.

2. Если в точке x_i известны производные более высоких порядков, то могут быть построены аппроксимации \hat{f}'_i , \hat{I}_i^{i+1} , \hat{f}_i более высокого порядка, которые здесь не рассматриваются.

Б. Трехточечный шаблон. Возьмем на неравномерной сетке Ω_n трехточечный шаблон $\Pi_{3,i} = (x_{i-1}, x_i, x_{i+1})$, характеризующийся шагами $h_{i+1} = \frac{h_{i+1}}{h_i} = x_{i+1} - x_i$, $h_i = x_i - x_{i-1}$ и параметром нерегулярности $\delta_{i+1} = \frac{h_{i+1}}{h_i}$. Предположив, что $f(x) \in C_3[a, b]$, найдем разложения для первообразной $F(x)$ при $x = x_{i+1}$ и $x = x_{i-1}$ относительно точки x_i и выпишем выражения для $F(x_{i+1})$, $F(x_{i-1})$ по формуле Тейлора (B.23) при $k = 3$:

$$F_{i+1} = F_i + h_{i+1} f'_i + \frac{h_{i+1}^2}{2} f''_i + \frac{h_{i+1}^3}{6} f'''(\xi_1), \quad \xi_1 \in (x_i, x_{i+1});$$

$$F_{i-1} = F_i - h_i f'_i + \frac{h_i^2}{2} f''_i - \frac{h_i^3}{6} f'''(\xi_2), \quad \xi_2 \in (x_{i-1}, x_i).$$

Умножая первое соотношение на h_i , а второе на h_{i+1} , складывая их с учетом равенства $I_i^{i+1} = F_{i+1} - F_i$ и разрешая относительно f'_i , получаем

$$f'_i = \frac{2}{h_i(1+\delta_{i+1})} \left(\frac{I_i^{i+1}}{h_{i+1}} - \frac{I_{i-1}^i}{h_i} \right) - \frac{h_i(\delta_{i+1}-1)}{3} f''_i - \frac{h_i^2}{12(1+\delta_{i+1})} (\delta_{i+1}^3 f'''(\xi_1) + f'''(\xi_2)).$$

Отсюда следует *интегрально-дифференциальная аппроксимационная формула для первой производной f'_i на нерегулярном шаблоне:*

$$\hat{f}'_{i,v} = \frac{2}{h_i(1+\delta_{i+1})} \left(\frac{I_i^{i+1}}{h_{i+1}} - \frac{I_{i-1}^i}{h_i} \right) - \frac{h_i(\delta_{i+1}-1)}{3} f''_i. \quad (5.21)$$

Формула (5.21) при заданных интегралах I_i^{i+1} , I_{i-1}^i (известных с точностью не ниже $O(h_{i+1}^3)$ и $O(h_i^3)$) и производной f''_i (известной с точностью не ниже первого порядка) аппроксимирует первую производную f'_i на нерегулярном шаблоне со вторым порядком (безусловная аппроксимация).

При условной аппроксимации, когда $\delta_{i+1} = 1$ (равномерная сетка), из (5.21) следует *интегральная аппроксимационная формула для первой производной f'_i на регулярном шаблоне:*

$$\hat{f}'_{i,c} = \frac{1}{h^2} (I_i^{i+1} - I_{i-1}^i) - \left(\frac{h^2}{12} M_3 \right). \quad (5.22)$$

Эта формула аппроксимирует первую производную f'_i со вторым порядком.

Из сопоставления мажорант оценок аппроксимационных операторов $\hat{f}'_{i,c}$ (см. (5.10)) и $\hat{f}'_{i,v}$ (см. (5.22)) вытекает, что они имеют одинаковый (второй) порядок аппроксимации, однако мажоранта или остаточное слагаемое оператора интегрального типа содержит константу $\left(\frac{1}{12}\right)$, в два раза меньшую соответствующей константы в мажоранте оператора точечного (функционального) типа.

Замечание. При условии $f(x) \in C_2[a, b]$ на регулярном шаблоне из разложения первообразных получается аппроксимационная *интегрально-функциональная формула для второй производной*:

$$\hat{f}_i'' = \frac{3}{h^2} I_{i-1}^{i+1} - \frac{6}{h^2} f_i \quad (O(h)), \quad (5.23)$$

из которой можно выразить интеграл I_{i-1}^{i+1} через значения функции f_i и производной f_i'' :

$$\hat{I}_{i-1}^{i+1} = 2hf_i + \frac{h^3}{3} f_i''.$$

Пример 5.4. Пусть некоторая функция (взята функция $f(x) = x^3$) на трехточечном шаблоне $x_1 = 1; x_2 = 1,5; x_3 = 2$ ($h = 0,5 = \text{const}$) задана двумя независимыми способами:

- 1) значениями функции $f_1 = 1; f_2 = 3,375; f_3 = 8$;
- 2) значениями интегралов на двух соседних отрезках: на отрезке $[x_1, x_2]$: $I_1^2 = 1,015625$, а на отрезке $[x_2, x_3]$: $I_2^3 = 2,734375$ (значения функций $f(x_i)$ и интегралов вычислены по $f(x) = x^3$ точно).

Требуется найти значение производной $f'(x)|_{x=1,5}$ с помощью функциональной и интегральной аппроксимаций.

□ Для определения $f'(x)|_{x=1,5}$ сначала воспользуемся функциональной аппроксимационной формулой (5.10):

$$\hat{f}'_{2,c} = \frac{f_3 - f_1}{2h} = \frac{8 - 1}{1} = 7 \text{ (3,7%).}$$

Применим интегральную формулу (5.22):

$$\hat{f}'_{2,c} = \frac{I_2^3 - I_1^2}{h^2} = \frac{2,734375 - 1,015625}{0,25} = 6,875 \text{ (1,8%).}$$

В скобках, сразу за численными результатами значений производных указанна относительная погрешность в процентах. Из сопоставления полученных приближенных значений с точными следует, что интегральная аппроксимация дает лучший результат. ■

Пример 5.5. Пусть сеточная функция (табл. 5.2), являющаяся представлением $f(x) = x^4$, задана интегралами на отрезках $[x_i, x_{i+1}]$ ($i = 1, 2, 3$), $h = 1$.

Требуется вычислить производную $f'(x)|_{x=2}$ с использованием интегральной формулы (5.22), имеющей второй порядок аппроксимации.

□ Вычисление производной основывается на методике, изложенной выше.

1. Так как в задаче указан конкретный аппроксимационный оператор, то выбор формулы осуществлять не нужно.

2. Формула (5.22) записана через значение двух интегралов I_{i-1}^i и I_i^{i+1} , что соответствует двухшаговому (трехточечному) шаблону $\Pi_{3,i} = (x_{i-1}, x_i, x_{i+1})$,

Таблица 5.2

i	0	1	2	3	4
x_i	0	1	2	3	4
I_i^{i+1}	$\frac{1}{5}$	$\frac{31}{5}$	$\frac{211}{5}$	$\frac{781}{5}$	

в котором $x_i = 2$ ($i = 2$), $x_{i-1} = 1$ ($i - 1 = 1$), $x_{i+1} = 3$ ($i + 1 = 2$). В данном шаблоне центральная точка $x_i = 2$ совпадает с точкой $x_j = 2$.

3. Получим искомое значение

$$\hat{f}' = \frac{I_i^{i+1} - I_{i-1}^i}{h^2} = \frac{\frac{211}{5} - \frac{31}{5}}{1} = \frac{180}{5} = 36.$$

Здесь вычисления выполнены точно, хотя результат является приближенным.

4. Используя точное значение производной $f'(2) = 32$, можно получить фактическую погрешность результата

$$\frac{|36 - 32|}{32} \cdot 100\% = 12,5\%. \blacksquare$$

Изложенная в п. 5.2.1 методика определения остаточного слагаемого основана на соответствующей теореме для интерполяционного многочлена Лагранжа. Полезно рассмотреть и другой способ получения оценки погрешности, использующий разложение функций по формуле Тейлора.

*Методика оценки погрешности
дифференциальных операторов на основе
разложения функции по формуле Тейлора*

1. Составить разность $f_i^{(p)} - \hat{f}_i^{(p)}$, где $\hat{f}_i^{(p)}$ — обозначение оператора, в качестве которого может быть принят оператор как функционального типа, так и интегрального или интегрально-функционального, а $f_i^{(p)}$ — точное значение производной.

2. Все функции, входящие в $\hat{f}_i^{(p)}$, разложить по формуле Тейлора с остаточным слагаемым в форме Лагранжа относительно точки x_i , в которой записан $\hat{f}_i^{(p)}$. Если оператор $\hat{f}_i^{(p)}$ выражается через интегралы, то они записываются через разности первообразных ($F(x)$). При этом предполагается, что функции $f(x)$, $F(x)$ являются непрерывными, и для них существуют непрерывные производные соответствующего порядка.

3. Подставить разложения функции в разность (см. п. 1) и выполнить ее преобразование, в процессе которого получается остаточное слагаемое.

4. На основе остаточного слагаемого оценить погрешность аппроксимации.

Пример 5.6. Для заданного дифференциального оператора (5.22) $\hat{f}_i' = \frac{I_i^{i+1} - I_{i-1}^i}{h^2}$, записанного на шаблоне $\mathbb{W}_{3,i} = (x_{i-1}, x_i, x_{i+1})$, найти остаточное слагаемое и оценить погрешность в точке $x_j = x_i$.

□1. Составим разность $f_i' - \hat{f}_i' = f_i' - \frac{1}{h^2}(I_i^{i+1} - I_{i-1}^i)$, где f_i' — точное значение производной, а \hat{f}_i' — аппроксимирующий ее оператор.

2. Интегралы заменим разностями первообразных:

$$I_i^{i+1} = F_{i+1} - F_i \quad (F_{i+1} = F(x_{i+1}), F_i = F(x_i)), \quad I_{i-1}^i = F_i - F_{i-1}.$$

Функцию $F(x)$ при $x = x_i$, $x = x_{i+1}$ разложим по формуле Тейлора с остаточным слагаемым в форме Лагранжа относительно точки x_i и выпишем выражения для $F(x_{i+1})$, $F(x_{i-1})$:

$$\begin{aligned} F(x_{i+1}) &= F_i + h f'_i + \frac{h^2}{2} f''_i + \frac{h^3}{6} f'''_i (\xi_+) \\ F(x_{i-1}) &= F_i - h f'_i + \frac{h^2}{2} f''_i - \frac{h^3}{6} f'''_i (\xi_-), \end{aligned}$$

где $\xi_+ \in (x_i, x_{i+1})$, $\xi_- \in (x_{i-1}, x_i)$.

3. Подставим данные разложения в разность, составленную в п. 1, и осуществим преобразования:

$$\begin{aligned} f'_i - \frac{1}{h^2} \left(F_i + h \cdot f'_i + \frac{h^2}{2} f''_i + \frac{h^3}{6} f'''_i (\xi_+) - F_i - F_i + F_i - h f'_i + \frac{h^2}{2} f'_i - \frac{h^3}{6} f''_i + \frac{h^4}{24} f'''_i (\xi_-) \right) = \\ = \frac{h^2}{24} (f'''(\xi_+) + f'''(\xi_-)). \end{aligned}$$

Применяя теорему о среднем (см. утверждение В.1), находим остаточное слагаемое

$$f'_i - \hat{f}'_i = \frac{h^2}{12} f'''(\xi_i), \quad \xi_i \in (x_{i-1}, x_{i+1}).$$

4. С помощью последнего соотношения получим оценку погрешности аппроксимации f'_i интегральным оператором (5.22):

$$\left| f'_i - \frac{1}{h^2} (I_i^{i+1} - I_{i-1}^i) \right| \leq \frac{M_{3,i}}{12} h^2,$$

где $M_{3,i} = \max_{[x_{i-1}, x_i]} |f'''(x)|$. ■

5.2.3. ФОРМУЛЫ, ПОЛУЧЕННЫЕ НА ОСНОВЕ СПЛАЙНОВ

ЯВНЫЕ ФОРМУЛЫ

A. Двухточечный шаблон. Из рассмотрения интегрально-дифференциальных сплайн-многочленов второй степени в работе [17] получены *интегрально-функциональные* аппроксимационные формулы для первых производных на двухточечном шаблоне $\Pi_{2,i} = (x_i, x_{i+1})$:

$$\begin{aligned} \hat{f}'_{i-1} &= \frac{6}{h_i^2} I_{i-1}^i - \frac{2}{h_i} (2f_{i-1} + f_i); \\ \hat{f}'_i &= \frac{2}{h_i} (f_{i-1} + 2f_i) - \frac{6}{h_i^2} I_{i-1}^i. \end{aligned} \tag{5.24}$$

Формулы (5.24) в силу их одноинтервального характера могут использоваться при $h_i = \text{var}$, если значение интеграла I_{i-1}^i либо известно, либо заранее

вычислено с точностью не ниже $O(h_i^4)$. Для операторов \hat{f}'_{i-1} , \hat{f}'_i справедливы оценки

$$|\hat{f}'_k - f'_k| \leq \frac{h_i^2}{12} M_{3,i}, \quad k = i-1, i.$$

Замечание. Из (5.24) можно выразить величину I_{i-1}^i , и тогда получаются следующие функционально-дифференциальные квадратурные формулы:

$$\begin{aligned}\hat{I}_{i-1}^i &= \frac{h_i}{3}(2f_{i-1} + f_i) + \frac{h_i^2}{6}f'_{i-1}; \\ \hat{I}_{i-1}^i &= \frac{h_i}{3}(f_{i-1} + 2f_i) - \frac{h_i^2}{6}f'.\end{aligned}$$

Б. Трехточечный шаблон. Из рассмотрения интегрально-дифференциальных сплайн-многочленов третьей степени на трехточечном шаблоне $\mathbb{W}_{3,i} = (x_{i-1}, x_i, x_{i+1})$ получены следующие интегрально-функциональные формулы:

- для левой, центральной и правой точек нерегулярного шаблона аппроксимационные формулы для первой производной имеют вид

$$\begin{aligned}\hat{f}'_{i-1,v} &= \frac{1}{H_i^{i+1}} \left[\begin{array}{l} 4 \left(\frac{h_i}{h_{i+1}^2} I_i^{i+1} + \frac{(2H_i^{i+1} + h_i)}{h_i^2} I_{i-1}^i \right) - \\ - \left(\frac{h_i}{h_{i+1}} f_{i+1} + \frac{3(H_i^{i+1})^2}{h_i h_{i+1}} f_i + \frac{5H_i^{i+1} + h_i}{h_i} f_{i-1} \right) \end{array} \right], \\ \hat{f}'_{i,v} &= \frac{1}{H_i^{i+1}} \left[\begin{array}{l} 4 \left(\frac{h_i}{h_{i+1}^2} I_i^{i+1} - \frac{h_{i+1}}{h_i^2} I_{i-1}^i \right) - \\ - \left(\frac{h_i}{h_{i+1}} f_{i+1} + \frac{3(h_i^2 - h_{i+1}^2)}{h_i h_{i+1}} f_i - \frac{h_{i+1}}{h_i} f_{i-1} \right) \end{array} \right], \\ \hat{f}'_{i+1,v} &= \frac{1}{H_i^{i+1}} \left[\begin{array}{l} -4 \left(\frac{(2H_i^{i+1} + h_{i+1})}{h_{i+1}^2} I_i^{i+1} + \frac{h_{i+1}}{h_i^2} I_{i-1}^i \right) + \\ + \left(\frac{5H_i^{i+1} + h_{i+1}}{h_{i+1}} f_{i+1} + \frac{3(H_i^{i+1})^2}{h_i h_{i+1}} f_i + \frac{h_{i+1}}{h_i} f_{i-1} \right) \end{array} \right];\end{aligned}$$

- для левой, центральной и правой точек регулярного шаблона аппроксимационные формулы для первой производной принимают форму

$$\begin{aligned}\hat{f}'_{i-1,c} &= \frac{1}{2h} \left[\frac{4}{h} (I_i^{i+1} + 5I_{i-1}^i) - (f_{i+1} + 12f_i + 11f_{i-1}) \right] \left(\frac{h^3}{60} M_{4,i} \right), \\ \hat{f}'_{i,c} &= \frac{1}{2h} \left[\frac{4}{h} (I_i^{i+1} - I_{i-1}^i) - (f_{i+1} - f_{i-1}) \right] = \frac{1}{2h} \left[\frac{4}{h} \Delta I_i^{i+1} - (\Delta f_i + \Delta f_{i-1}) \right] \left(\frac{h^4}{360} M_{5,i} \right), \\ \hat{f}'_{i+1,c} &= \frac{1}{2h} \left[-\frac{4}{h} (5I_i^{i+1} + I_{i-1}^i) + (11f_{i+1} + 12f_i + f_{i-1}) \right] \left(\frac{h^3}{60} M_{4,i} \right);\end{aligned}$$

- для левой, центральной и правой точек *нерегулярного шаблона* аппроксимационные формулы для второй производной имеют вид

$$\begin{aligned}\hat{f}_{i-1,v}'' &= \frac{6}{H_i^{i+1}} \left[-4 \left(\frac{1}{h_{i+1}^2} I_i^{i+1} + \frac{H_{2i}^{i+1}}{h_i^3} I_{i-1}^i \right) + \frac{1}{h_{i+1}} f_{i+1} + \frac{H_i^{i+1}(2H_i^{i+1} + h_i)}{h_i^2 h_{i+1}} f_i + \frac{2H_i^{i+1} + h_i}{h_i^2} f_{i-1} \right], \\ \hat{f}_{i,v}'' &= \frac{6}{H_i^{i+1}} \left[4 \left(\frac{1}{h_{i+1}^2} I_i^{i+1} + \frac{1}{h_i^2} I_{i-1}^i \right) - \left(\frac{f_{i+1}}{h_{i+1}} + \frac{3H_i^{i+1}}{h_{i+1} h_i} f_i + \frac{f_{i-1}}{h_i} \right) \right], \\ \hat{f}_{i+1,v}'' &= \frac{6}{H_i^{i+1}} \left[-4 \left(\frac{H_i^{2(i+1)}}{h_{i+1}^3} I_i^{i+1} + \frac{1}{h_i^2} I_{i-1}^i \right) + \frac{2H_i^{i+1} + h_{i+1}}{h_{i+1}^2} f_{i+1} + \frac{H_i^{i+1}(2H_i^{i+1} + h_{i+1})}{h_i h_{i+1}^2} f_i + \frac{f_{i-1}}{h_i} \right];\end{aligned}$$

- для левой, центральной и правой точек *регулярного шаблона* аппроксимационные формулы для второй производной принимают форму

$$\begin{aligned}\hat{f}_{i-1,c}'' &= \frac{3}{h^2} \left[-\frac{4}{h} (I_i^{i+1} + 3I_{i-1}^i) + (f_{i+1} + 10f_i + 5f_{i-1}) \right] \quad \left(\frac{7}{20} h^2 M_{4,i} \right), \\ \hat{f}_{i,c}'' &= \frac{3}{h^2} \left[\frac{4}{h} (I_i^{i+1} + I_{i-1}^i) - (f_{i+1} + 6f_i + f_{i-1}) \right] \quad \left(\frac{1}{20} h^2 M_{4,i} \right), \\ \hat{f}_{i+1,c}'' &= \frac{3}{h^2} \left[-\frac{4}{h} (3I_i^{i+1} + I_{i-1}^i) + (5f_{i+1} + 10f_i + f_{i-1}) \right] \quad \left(\frac{7}{20} h^2 M_{4,i} \right).\end{aligned}$$

Подчеркнем, что правые части данных аппроксимационных соотношений записаны через интегралы I_{i-1}^i , I_i^{i+1} на двух смежных отрезках, составляющих шаблон, и через значения функций в точках этого шаблона. Если известны интегралы и значения самой функции, то по этим формулам можно вычислить первые производные с третьим порядком, а вторые производные — со вторым (остаточные слагаемые некоторых аппроксимационных формул указаны в скобках, расположенных рядом с этими формулами). При этом формула для $\hat{f}_{i,c}'$, имеющая симметричный вид относительно центральной точки x_i и содержащая интегральные и функциональные разности, обеспечивает повышенный (четвертый) порядок аппроксимации. Если интегралы для исследуемой функции неизвестны, они должны быть предварительно рассчитаны с порядком, по крайней мере на два превышающим порядок аппроксимации дифференциальных операторов.

Замечание. В вычислительной практике могут оказаться полезными еще два аппроксимационных оператора $\hat{f}_{\Pi(\Pi)v}''$, $\hat{f}_{i,c}''$:

$$\hat{f}_{\Pi(\Pi)v}'' = \frac{6}{h_{i+1}^2} (f_i + f_{i+1}) - \frac{12}{h_{i+1}^3} I_i^{i+1} \quad \left(\frac{h_{i+1}}{2} M_{3,i} \right); \quad (5.25)$$

$$\hat{f}_{i,c}'' = \frac{3}{2h^2} (f_{i-1} + f_{i+1}) - \frac{3}{2h_i} I_{i-1}^{i+1} \quad \left(\frac{h^2}{10} M_{4,i} \right), \quad (5.26)$$

где $\hat{f}_{\Pi(\Pi)v}''$ — лево- или правосторонний оператор, а $\hat{f}_{i,c}''$ — центральный оператор, записанный на трехточечном шаблоне при $h = \text{const}$.

В. Четырехточечный шаблон. Выше приведены формулы численного дифференцирования на трехточечном (или для некоторых формул на двухточеч-

ном) шаблоне, имеющие порядок аппроксимации $h^{3-(p-1)}$, где p — порядок производных, для которых записаны эти формулы. В дополнение к изложенному материалу приведем формулы, аппроксимирующие производную $f^{(p)}(p=1, 2)$ с порядками $h^{3-(p-1)}$ на четырехточечном шаблоне. Данные формулы получены в работах [17], [23] путем анализа кубических дифференциальных и интегрально-дифференциальных сплайнов.

1. Формулы для первых производных третьего порядка аппроксимации на шаблоне $\Pi_{4,i} = (x_{i-2}, x_{i-1}, x_i, x_{i+1})$:

- для лево- и правосторонних внутренних точек x_{i-1} , x_i нерегулярного шаблона справедливы функциональные формулы:

$$\begin{aligned}\hat{f}'_{i-1,v} &= \frac{1}{a} \left\{ -h_i^2 h_{i+1}^2 (H_i^{i+1})^2 f_{i-2} + h_{i+1} \left(h_i^2 (H_i^{i+1})^2 - h_{i-1}^2 K_{2i-1}^2 \right) f_{i-1} + \right. \\ &\quad \left. + h_{i-1}^2 \left[h_i^2 H_{i-1}^i + K_{2i-1}^2 h_{i+1} \right] f_i - h_i^2 h_{i-1}^2 H_{i-1}^i f_{i+1} \right\};\end{aligned}\quad (5.27)$$

$$\begin{aligned}\hat{f}'_{i,v} &= \frac{1}{a} \left\{ h_i^2 h_{i+1}^2 H_i^{i+1} f_{i-2} - h_{i+1}^2 (h_i^2 H_i^{i+1} + h_{i-1} K_{2i}^2) f_{i+1} + \right. \\ &\quad \left. + h_{i-1} \left[h_{i+1}^2 K_{2i}^2 - h_i^2 (H_{i-1}^i)^2 \right] f_i + h_i^2 h_{i-1} (H_{i-1}^i)^2 f_{i+1} \right\},\end{aligned}\quad (5.28)$$

где $a = H_{i-1}^{i+1} H_{i-1}^i H_i^{i+1} \Pi_{i-1}^{i+1}$; $K_{2i}^2 = h_{i-1} (H_{i-1}^{i+1} + 2h_i) + h_i (2H_i^{i+1} + h_i)$; $K_{2i-1}^2 = h_{i+1} (H_{i-1}^{i+1} + 2h_i) + h_i (2H_{i-1}^i + h_i)$; $\Pi_{i-1}^{i+1} = h_{i-1} h_i h_{i+1}$; $H_{i-1}^{i+1} = h_{i-1} + h_i + h_{i+1}$;

- для левой и правой крайних точек x_{i-2} , x_{i+1} нерегулярного шаблона справедливы функционально-дифференциальные формулы рекуррентного типа:

$$\hat{f}'_{i-2,v} = \frac{h_{i-1}^2}{H_{i-1}^i} \left(\frac{2H_{i-1}^i + h_{i-1}}{h_{i-1}^3} \Delta f_{i-2} + \frac{\Delta f_{i-1}}{h_i^2} \right) - \frac{H_{i-1}^i}{h_i} \hat{f}'_{i-1}; \quad (5.29)$$

$$\hat{f}'_{i+1,v} = \frac{h_{i+1}^2}{H_i^{i+1}} \left(\frac{2H_i^{i+1} + h_{i+1}}{h_{i+1}^3} \Delta f_i + \frac{\Delta f_{i-1}}{h_i^2} \right) - \frac{H_i^{i+1}}{h_i} \hat{f}'_i. \quad (5.30)$$

На регулярном четырехточечном шаблоне при $h = \text{const}$ формулы (5.27)–(5.30) упрощаются и сводятся к формулам (5.15). Можно показать, что (5.27)–(5.30), так же как и формулы (5.15), имеют третий порядок аппроксимации.

Парabolicкие и кубические интегрально-дифференциальные сплайны позволяют сконструировать интегральные формулы численного дифференцирования [15]–[24].

На четырехточечном нерегулярном шаблоне $\Pi_{4,i+1} = (x_{i-1}, x_i, x_{i+1}, x_{i+2})$ получаются следующие интегральные и рекуррентные формулы для первых производных:

$$\hat{f}' = \frac{2}{A} \left\{ \frac{h_i^2 - h_{i+2}^2}{h_{i+2}} I_{i+1}^{i+2} + \left[3h_{i+1} H_{i+1}^{i+2} + (h_{i+2}^2 - h_i^2) \right] \frac{1}{h_{i+1}} I_i^{i+1} - \frac{H_{i+1}^{i+2} H_{2(i+1)}^{i+2}}{h_i} I_{i-1}^i \right\}; \quad (5.31)$$

$$\hat{f}'_{i+1} = \frac{2}{A} \left\{ \frac{H_i^{i+1} H_i^{2(i+1)}}{h_{i+2}} I_{i+1}^{i+2} - \frac{3h_{i+1} H_i^{i+1} + (h_i^2 - h_{i+2}^2)}{h_{i+1}} I_i^{i+1} + \frac{h_{i+1}^2 - h_{i+2}^2}{h_i} I_{i-1}^i \right\}; \quad (5.32)$$

$$\hat{f}'_{i-1} = \frac{H_{i-1}^i}{h_i} \hat{f}'_i - \frac{h_{i-1}}{h_i} \hat{f}'_{i+1}; \quad \hat{f}'_{i+2} = \frac{H_{i+1}^{i+2}}{h_i} \hat{f}'_{i+1} - \frac{h_{i+2}}{h_{i+1}} \hat{f}'_i, \quad (5.33)$$

где $A = h_{i+1}^2(2h_i + h_{i+1} + 2h_{i+2}) + h_{i+1}(h_i^2 + h_{i+2}^2) + h_i h_{i+2}(h_i + 3h_{i+1} + h_{i+2})$.

Формулы (5.31), (5.32) относятся к внутренним точкам шаблона и при $h = \text{const}$ переходят в (5.22), а формулы (5.33) (рекуррентные) — к крайним точкам шаблона и являются подобными формулам (5.36), справедливыми для вторых производных.

На регулярном шаблоне из (5.33) при $h = \text{const}$ легко получаются явные трехинтервальные аппроксимации интегрального типа:

$$\begin{aligned}\hat{f}'_{i-1,c} &= \frac{1}{h^2}(-2I_{i-1}^i + 3I_i^{i+1} - I_{i+1}^{i+2}) \quad \left(\frac{11}{12}h^2 M_{3,i}\right); \\ \hat{f}'_{i+2,c} &= \frac{1}{h^2}(I_{i-1}^i - 3I_i^{i+1} + 2I_{i+1}^{i+2}) \quad \left(\frac{11}{12}h^2 M_{3,i}\right).\end{aligned}$$

Последние две формулы могут быть получены также методом подобия из первой и последней формул (5.16). Их можно записать через интегральные разности:

$$\begin{aligned}\hat{f}'_{i-1,c} &= \frac{1}{h^2}(2\Delta I_i^{i+1} - \Delta I_{i+1}^{i+2}); \\ \hat{f}'_{i+2,c} &= \frac{1}{h^2}(2\Delta I_{i+1}^{i+2} - \Delta I_i^{i+1}),\end{aligned}$$

где $\Delta I_k^{k+1} = I_k^{k+1} - I_{k-1}^k$.

2. Формулы для *вторых производных* второго порядка аппроксимации на шаблоне $\mathbb{W}_{4,i} = (x_{i-2}, x_{i-1}, x_i, x_{i+1})$:

- для *лево- и правосторонних внутренних точек* x_{i-1}, x_i *нерегулярного шаблона* справедливы следующие функциональные формулы:

$$\begin{aligned}\hat{f}''_{i,v} &= \frac{2}{a} \left[K_i^{i+1} \Delta h_{i+1} f_{i-2} + H_{i-1}^i h_{i+1} (H_{i-1}^{2i} h_{i-1} - H_i^{i+1} \Delta h_{i+1}) f_i - \right. \\ &\quad \left. - H_i^{i+1} h_{i-1} (H_{i-1}^{2i} H_{i-1}^i - h_{i+1} \Delta h_i) f_i + K_{i-1}^i H_{i-1}^{2i} h_{i+1} \right], \quad (5.34)\end{aligned}$$

$$\begin{aligned}\hat{f}''_{i-1,v} &= \frac{2}{a} \left[K_i^{i+1} H_{2i}^{i+1} f_{i-2} - h_{i+1} H_i^{i+1} (H_{2i}^{i+1} H_i^{i+1} + h_{i-1} \Delta h_i) f_{i-1} + \right. \\ &\quad \left. + H_i^{i+1} h_{i-1} (h_{i+1} H_{2i}^{i+1} + H_{i-1}^i \Delta h_i) f_i - K_{i-1}^i \Delta h_i f_{i+1} \right], \quad (5.35)\end{aligned}$$

где $K_t^{t+1} = \Pi_t^{t+1} H_t^{t+1}$; $\Pi_t^{t+1} = h_t h_{t+1}$; $t = i-1, i$; $H_{i-1}^{2i} = h_{i-1} + 2h_i$; $H_{2i}^{i+1} = 2h_i + h_{i+1}$; $\Delta h_{i+1} = h_{i+1} - h_i$; $a = H_{i-1}^{i+1} H_{i-1}^i H_i^{i+1} \Pi_{i-1}^{i+1}$;

- для *левой и правой крайних точек нерегулярного шаблона* справедливы дифференциальные формулы рекуррентного типа:

$$\begin{aligned}\hat{f}''_{i-2,v} &= \frac{H_{i-1}^i}{h_i} \hat{f}''_{i-1,v} - \frac{h_{i-1}}{h_i} \hat{f}''_{i,v}; \\ \hat{f}''_{i+1,v} &= \frac{H_i^{i+1}}{h_i} \hat{f}''_{i,v} - \frac{h_{i+1}}{h_i} \hat{f}''_{i-1,v}.\end{aligned} \quad (5.36)$$

Формулы (5.27), (5.28) и (5.34), (5.35) могут использоваться для расчета значений производных во внутренних точках сетки $\{x_i\}$, $i = \overline{1, n-1}$, а (5.29), (5.30) и (5.36) — для расчета производных в крайних точках x_0 , x_n сетки Ω_n .

На регулярном шаблоне при $h = \text{const}$ последняя группа формул для аппроксимации вторых производных во внутренних точках шаблона преобразуется к традиционным:

$$\begin{aligned}\hat{f}_{i,c}'' &= \frac{1}{h^2}(f_{i-1} - 2f_i + f_{i+1}) \quad \left(\frac{h^2}{12}M_{4,i}\right); \\ \hat{f}_{i-1,c}'' &= \frac{1}{h^2}(f_{i-2} - f_{i-1} + f_i),\end{aligned}$$

а (5.36) — к рекуррентным формулам:

$$\hat{f}_{i-2,c}'' = 2\hat{f}_{i-1,c}'' - \hat{f}_{i,c}''; \quad \hat{f}_{i+1,c}'' = 2\hat{f}_{i,c}'' - \hat{f}_{i-1,c}'',$$

а также к первой и последней формулам из (5.16), имеющим второй порядок аппроксимации.

НЕЯВНЫЕ АЛГОРИТМЫ ЧИСЛЕННОГО ДИФФЕРЕНЦИРОВАНИЯ

Путем несложного анализа параболических и кубических дифференциальных сплайнов получаются нижеследующие *неявные алгоритмы* вычисления первых и вторых производных сеточных функций. Значения производных могут быть получены не по явным формулам, а в результате решения трехдиагональных систем линейных алгебраических уравнений методом прогонки.

Первые производные по заданной сеточной функции $y_i = f(x_i)$, $i = \overline{0, n}$, можно вычислить несколькими способами:

а) следствием параметрических соотношений параболических сплайнов является система линейных алгебраических уравнений относительно первых производных:

$$\frac{h_i}{2}\hat{f}'_{i-1} + \frac{1}{2}(h_i + h_{i+1})\hat{f}'_i + \frac{h_{i+1}}{2}\hat{f}'_{i+1} = \Delta f_{i-1} + \Delta f_i, \quad i = \overline{1, n-1}, \quad (5.37)$$

где $\Delta f_i = f_{i+1} - f_i$; $h_{i+1} = x_{i+1} - x_i$.

При заданных значениях производных на концах отрезка $[x_0, x_n]$ эта система позволяет *со вторым порядком аппроксимации* вычислить значения первых производных \hat{f}'_i во всех внутренних точках.

На регулярном шаблоне при $h = \text{const}$ эта система упрощается:

$$\hat{f}'_{i-1} + 2\hat{f}'_i + \hat{f}'_{i+1} = \frac{2(f_{i+1} - f_{i-1})}{h}, \quad i = \overline{1, n-1};$$

б) на регулярном шаблоне значения первых производных \hat{f}'_i со вторым порядком аппроксимации могут быть вычислены также из системы, удовлетворяющей свойству преобладания диагональных элементов [44]:

$$\hat{f}'_{i-\frac{1}{2}} + 6\hat{f}'_{i+\frac{1}{2}} + \hat{f}'_{i+\frac{3}{2}} = \frac{8\Delta f_i}{h_i}, \quad i = \overline{0, n-1}. \quad (5.38)$$

Решением этой системы будут производные в узлах, сдвинутых влево на полшага. Если производные на концах неизвестны, то их необходимо предварительно вычислить с порядком не ниже второго;

в) первые производные с третьим порядком аппроксимации могут быть определены из системы, являющейся следствием кубических дифференциальных сплайнов [1] (см. (4.76)):

$$\frac{\hat{f}'_{i-1}}{h_i} + 2\left(\frac{1}{h_{i+1}} + \frac{1}{h_i}\right)\hat{f}'_i + \frac{\hat{f}'_{i+1}}{h_{i+1}} = 3\left(\frac{\Delta f_i}{h_{i+1}^2} + \frac{\Delta f_{i-1}}{h_i^2}\right), \quad i = \overline{1, n-1}. \quad (5.39)$$

Данная система, удовлетворяющая свойству преобладания диагональных элементов, может быть замкнута либо значениями \hat{f}'_0 , \hat{f}'_n , либо двумя функционально-дифференциальными граничными соотношениями (4.77);

г) первые производные со вторым порядком аппроксимации могут вычисляться также по значениям интегралов с использованием системы, подобной системе (4.72), в которой порядок производных понижен на единицу (в этом случае вместо Δf_i и Δf_{i-1} следует подставить $\Delta F_i = I_i^{i+1}$ и $\Delta F_{i-1} = I_{i-1}^i$):

$$h_i \hat{f}'_{i-1} + 2(h_i + h_{i+1})\hat{f}'_i + h_{i+1} \hat{f}'_{i+1} = 6\left(\frac{I_i^{i+1}}{h_{i+1}} - \frac{I_{i-1}^i}{h_i}\right), \quad i = \overline{1, n-1}. \quad (5.40)$$

Замыкающие соотношения формируются аналогично предыдущему случаю.

На регулярном шаблоне система (5.40) для \hat{f}'_i записывается через интегральные приращения:

$$\hat{f}'_{i-1} + 4\hat{f}'_i + \hat{f}'_{i+1} = \frac{6}{h}\left(\Delta I_i^{i+1}\right), \quad i = \overline{1, n-1}.$$

Вторые производные со вторым порядком аппроксимации могут быть определены из системы, являющейся следствием применения кубических дифференциальных сплайнов (см. систему (4.72)):

$$h_i \hat{f}''_{i-1} + 2(h_i + h_{i+1})\hat{f}''_i + h_{i+1} \hat{f}''_{i+1} = 6\left(\frac{\Delta f_i}{h_{i+1}} - \frac{\Delta f_{i-1}}{h_i}\right), \quad i = \overline{0, n-1}. \quad (5.41)$$

Данная система, удовлетворяющая условию преобладания диагональных элементов, может быть замкнута либо известными значениями \hat{f}''_0 , \hat{f}''_n , либо двумя граничными соотношениями, следующими из равенства

$$\frac{\Delta \hat{f}''_i}{h_{i+1}} = \frac{\Delta \hat{f}''_{i-1}}{h_i} \left(\frac{\hat{f}''_{i+1} - \hat{f}''_i}{h_{i+1}} = \frac{\hat{f}''_i - \hat{f}''_{i-1}}{h_i} \right),$$

которое записывается для трех крайних точек сетки Ω_n , т. е. для (x_0, x_1, x_2) и (x_{n-2}, x_{n-1}, x_n) .

Подчеркнем, что неявные алгоритмы вычисления производных предпочтительно использовать в случае, когда для заданной сеточной функции необходимо определять производные во всех узлах.

x_k	$f(x_k)$	Δf_k	$f'(x_k)$
$x_{i-1} = 2$	4	12	4
$x_i = 4$	16	33	8
$x_{i+1} = 7$	49		14

Пример 5.7. Проверить соотношение (5.37) при фиксированном i путем подстановки в него степенной функции.

□ Соотношение (5.37) является следствием применения параболических сплайнов, поэтому оно должно выполняться для сеточного представления степенной функции (параболы) $f(x) = x^2$. Возьмем трехточечный нерегулярный шаблон $\Pi_{3,i} = (x_{i-1}, x_i, x_{i+1})$ с шагами $h_{i+1} = 3; h_i = 2; x_{i-1} = 2; x_i = 4; x_{i+1} = 7$. Значения функции $f(x_k)$ ($k = i-1, i, i+1$) и ее производной $f'(x)|_{x=x_k} = 2x_k$ вместе с функциональными приращениями приведены в таблице 5.3.

Тогда соотношению (5.37) при фиксированном i и выбранном шаблоне будет соответствовать численное равенство

$$\frac{2}{2} \cdot 4 + \frac{1}{2} (2+3) \cdot 8 + \frac{3}{2} \cdot 14 = 12 + 33.$$

В левой и правой частях получаются одинаковые значения, что свидетельствует о правильности системы (5.37). ■

Методика вычисления производных по неявным алгоритмам

1. С учетом характера задания сеточной функции (заданы значения функции $y_i = f(x_i)$, $i = 0, n$, или значения интегралов I_i^{i+1} ($i = 0, n-1$)) и порядка аппроксимации t , который необходимо обеспечить в алгоритме, выбрать систему алгебраических уравнений относительно значений производных во всех внутренних узлах сетки. При этом, если данная система является следствием параболических сплайнов, то для первых производных $t = 2$, для вторых производных $t = 1$. В случае, когда система получена из кубических сплайнов, порядок t возрастает на единицу. Данные системы являются незамкнутыми (число неизвестных превышает на два число уравнений).

2. Замкнуть выбранную систему двумя граничными условиями на концах сетки Ω_n . Эти условия могут выбираться либо в виде явных аппроксимационных формул, либо в виде двух дополнительных алгебраических соотношений, включающих по два слагаемых с $f_0^{(p)}, f_1^{(p)}$ и $f_{n-1}^{(p)}, f_n^{(p)}$, $p = 1, 2$. При этом необходимо соблюсти соответствие порядков аппроксимации последних соотношений (или формул) и исходного порядка t .

3. Решить полученную замкнутую систему линейных алгебраических уравнений трехдиагонального вида методом прогонки.

Приведенная методика является составной частью построения кубического дифференциального сплайна. Она была применена при решении примера 4.13.

5.3. МЕТОДЫ ЧИСЛЕННОГО ИНТЕГРИРОВАНИЯ

5.3.1. ФОРМУЛЫ, ПОЛУЧЕННЫЕ НА ОСНОВЕ ИНТЕРПОЛЯЦИОННЫХ МНОГОЧЛЕНОВ

Одним из классических методов вычисления определенных интегралов является применение *функциональных квадратурных формул*

$$I_a^b = \int_a^b f(x)dx \cong \sum_{j=1}^N q_j f(x_j) \equiv \hat{I}_a^b, \quad (5.42)$$

где q_j — весовые коэффициенты; $x_j, j = \overline{1, N}$, — некоторые точки отрезка $[a, b]$; N — число точек (узлов квадратурной формулы).

Квадратурная формула называется *точной* для многочленов степени m , если при замене функции $f(x)$ на произвольный алгебраический многочлен степени не выше m приближенное равенство (5.42) становится точным. В этом случае говорят, что квадратурная формула обладает *m-свойством*.

Замечания

1. При приближенном вычислении интеграла, как правило, отрезок $[a, b]$ представляется в виде объединения l непересекающихся частичных отрезков вида $[x_{i-r}, x_{i+s}]$, которым соответствует шаблон $\Pi_{k,i} = (x_{i-r}, \dots, x_i, \dots, x_{i+s})$, где i — номер базового узла сетки; r и s — количество узлов левее и правее узла с номером i ; $k = r + s + 1$ — общее число узлов (точек) в шаблоне (рис. 5.1). На каждом частичном отрезке с номером $j = 1, \dots, l$ вычисляется интеграл по соответствующей квадратурной формуле

$$I_{i-r}^{i+s,j} = \int_{x_{i-r}}^{x_{i+s}} f(x)dx \cong \hat{I}_{i-r}^{i+s,j} \equiv \hat{I}^j, \quad j = 1, \dots, l, \quad (5.43)$$

а затем полученные значения суммируются по всем частичным отрезкам, т. е.

$$\hat{I}_a^b = \sum_{j=0}^l \hat{I}^j = \sum_{j=0}^l \hat{I}_{i-r}^{i+s,j}. \quad (5.44)$$

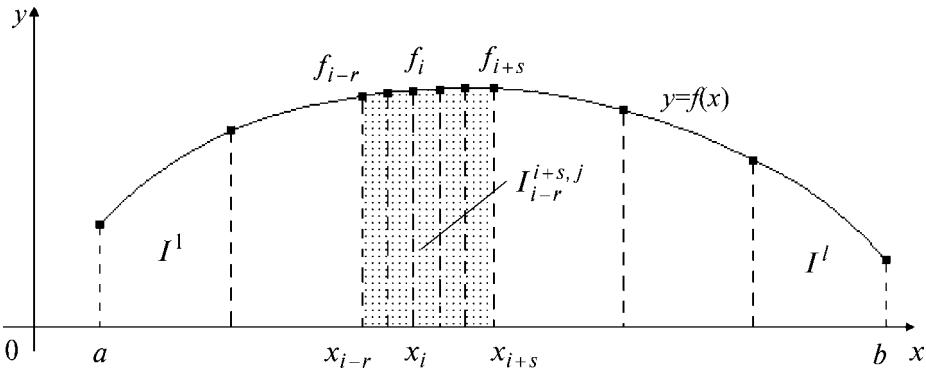


Рис. 5.1

2. Далее в силу использования представления (5.44) проблеме вычисления интеграла (5.43) на частичном отрезке уделяется основное внимание. По заданной сеточной функции или сеточному представлению формульной функции на частичном отрезке строится интерполяционный многочлен некоторой степени. Значение \hat{I}_{i-r}^{i+s} определяется величиной интеграла от этого многочлена.

Как следует из замечаний, для вычисления интеграла могут использоваться различные частичные отрезки и соответствующие им шаблоны. Рассмотрим процедуру получения простейших формул численного интегрирования на двухточечном и трехточечном шаблонах, наиболее часто описываемых в учебной литературе и используемых на практике.

А. Двухточечный шаблон. На частичном отрезке $[x_i, x_{i+1}]$, которому соответствует двухточечный шаблон $\Pi_{2,i} = (x_i, x_{i+1})$, где $r = 0, s = 1$, заменим функцию $f(x)$ двумя способами:

а) интерполяционным многочленом нулевой степени:

$$L_0(x) = f\left(\frac{x_i + x_{i+1}}{2}\right),$$

построенным по значению функции $f_{\frac{i+1}{2}} = f\left(\frac{x_{i+\frac{1}{2}}}{2}\right)$ в середине частичного отрезка $x_{i+\frac{1}{2}} = \frac{x_i + x_{i+1}}{2}$ (рис. 5.2а);

б) интерполяционным многочленом первой степени $L_1(x)$ с узловыми значениями x_i, x_{i+1} :

$$L_1(x) = \frac{x - x_{i+1}}{x_i - x_{i+1}} f_i + \frac{x - x_i}{x_{i+1} - x_i} f_{i+1} = \left(1 - \frac{x - x_i}{x_{i+1} - x_i}\right) f_i + \frac{x - x_i}{x_{i+1} - x_i} f_{i+1} = (1 - q) f_i + q f_{i+1},$$

где q — фаза интерполяции, $q = \frac{x - x_i}{h_{i+1}}$, $h_{i+1} = x_{i+1} - x_i$ (рис. 5.2б).

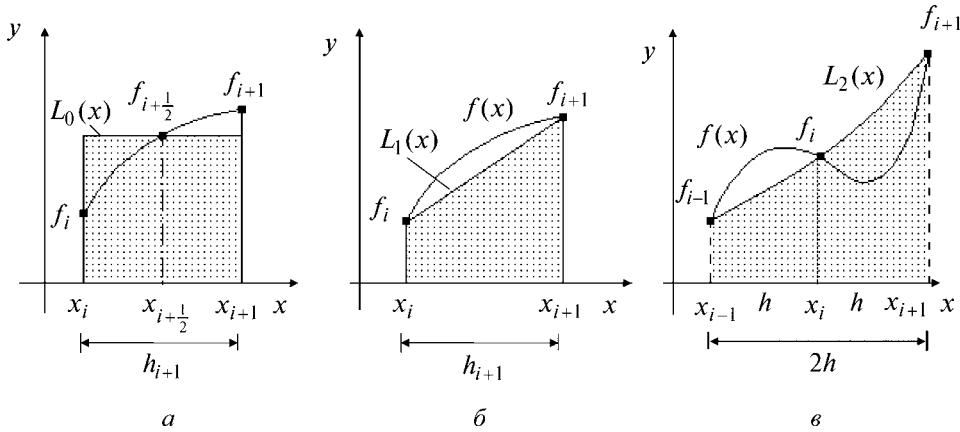


Рис. 5.2

В результате получим простейшие одиноинтервальные квадратурные формулы прямоугольников и трапеций:

$$\hat{I}_{i,\text{пр}}^{i+1} = \int_{x_i}^{x_{i+1}} L_0(x) dx = \int_{x_i}^{x_{i+1}} f\left(\frac{x_i + x_{i+1}}{2}\right) dx = h_{i+1} f_{\frac{i+1}{2}}; \quad (5.45)$$

$$\hat{I}_{i,\text{тр}}^{i+1} = \int_{x_i}^{x_{i+1}} L_1(x) dx = \int_0^1 [(1-q)f_i + qf_{i+1}] h_{i+1} dq = h_{i+1} \frac{f_i + f_{i+1}}{2}, \quad (5.46)$$

где при интегрировании учитывалось, что $dx = h_{i+1} \cdot dq$; $q = 0$ при $x = x_i$ и $q = 1$ при $x = x_{i+1}$.

Отметим, что понятие одноинтервальной формулы относится не к шаблону, а к отрезку, на котором вычисляется интеграл. Нижние индексы соответствуют названию квадратурной формулы (пр — прямоугольников, тр — трапеций).

Подчеркнем, что данные формулы справедливы и для нерегулярного шаблона, хотя последующее их суммирование по всем частичным отрезкам $[x_i, x_{i+1}]$ традиционно выполняется при $h = \text{const}$.

Б. Трехточечный шаблон. Пусть отрезок $[a, b]$ разбит на четное количество одинаковых частичных отрезков, т. е. $n = 2k$, где k — число пар. Проделав аналогичные выкладки для многочлена $L_2(x)$ второй степени, записываемого при $h = \text{const}$ на шаблоне $\Pi_{3,i} = (x_{i-1}, x_i, x_{i+1})$ (по одной паре отрезков при $r = 1, s = 1$), получим двухинтервальную квадратурную формулу парабол, или формулу Симпсона:

$$\hat{I}_{i-1,\text{пар}}^{i+1} = \int_{x_{i-1}}^{x_{i+1}} L_2(x) dx = \frac{h}{3} (f_{i-1} + 4f_i + f_{i+1}). \quad (5.47)$$

Эта же формула может быть получена из (5.23) путем разрешения ее относительно I_{i-1}^{i+1} и подстановки в нее конечно-разностной аппроксимации второй производной $\hat{f}_{i+1}'' = \frac{1}{h^2} (f_{i-1} - 2f_i + f_{i+1})$, имеющей второй порядок, или из формулы (4.92) с учетом $I_{i-1}^{i+1} = I_{i-1}^i + I_i^{i+1}$ и $h = \text{const}$.

Дадим геометрические интерпретации квадратурных формул прямоугольников (рис. 5.2а), трапеций (рис. 5.2б) и парабол (рис. 5.2в).

Интегралы обычно определяются не на частичных отрезках, а на всем отрезке $[a, b]$, и поэтому путем суммирования левых и правых частей (5.45)–(5.47) получаются так называемые составные квадратурные формулы.

Для сеточной функции $f(x_i)$, $i = 0, n$, заданной на регулярном шаблоне при $h = \text{const}$, эти формулы имеют вид

$$\hat{I}_{a,\text{пр}}^b = h \left(f_{1/2} + f_{3/2} + \dots + f_{\frac{n-1}{2}} \right) = h \sum_{i=0}^{n-1} f_{\frac{i+1}{2}}; \quad (5.48)$$

$$\begin{aligned} \hat{I}_{a,\text{тр}}^b &= h \left(\frac{f_0 + f_1}{2} + \frac{f_1 + f_2}{2} + \dots + \frac{f_{n-1} + f_n}{2} \right) = \frac{h}{2} [f_0 + 2(f_1 + f_2 + \dots + f_{n-1}) + f_n] = \\ &= \frac{h}{2} \left(f_0 + 2 \sum_{i=1}^{n-1} f_i + f_n \right); \end{aligned} \quad (5.49)$$

$$\begin{aligned}\hat{I}_{a,\text{пар}}^b &= \frac{h}{3} [f_0 + 4(f_1 + f_3 + \dots + f_{2k-1}) + 2(f_2 + f_4 + \dots + f_{2k-2}) + f_n] = \\ &= \frac{h}{3} \left[f_0 + 4 \sum_{i=1}^k f_{2i-1} + 2 \sum_{i=1}^{k-1} f_{2i} + f_{2k} \right].\end{aligned}\quad (5.50)$$

Подчеркнем, что в составной квадратурной формуле парабол индекс « k » указывает на число пар отрезков разбиения, которое предполагается четным ($n = 2k$). Если это условие не выполняется, то интеграл вычисляется для четного количества отрезков и к полученному значению добавляется величина I_{n-1}^n , рассчитанная с порядком $O(h^5)$ по формулам, приведенным далее.

ОЦЕНКА ПОГРЕШНОСТЕЙ КВАДРАТУРНЫХ ФОРМУЛ

Проведем оценку погрешностей одноинтервальных квадратурных формул прямоугольников и трапеций. С этой целью составим разность $I_i^{i+1} - \hat{I}_i^{i+1}$, которую затем преобразуем путем разложения первообразных и функций по формуле Тейлора [17], [23], [24]. Для этого I_i^{i+1} заменим разностью $F_{i+1} - F_i$, а $F(x)$ и функции, входящие в правые части квадратурной формулы, разложим по формуле Тейлора относительно точки x_i ($F(x)$ — первообразная, $F_i = F(x_i)$). Так, для формулы (5.46) при условии, что $f(x) \in C_2[a, b]$, получим

$$\begin{aligned}I_i^{i+1} - \hat{I}_{i,\text{тр}}^{i+1} &= F_{i+1} - F_i - \frac{h_{i+1}}{2}(f_i + f_{i+1}) = F_i + h_{i+1}f_i + \frac{h_{i+1}^2}{2}f'_i + \frac{h_{i+1}^3}{6}f''(\xi_i) - F_i - \\ &- \frac{h_{i+1}}{2} \left(f_i + f_i + h_{i+1}f'_i + \frac{h_{i+1}^2}{2}f''(\xi_i) \right) = \frac{h_{i+1}^3}{6}f''(\xi_i) - \frac{h_{i+1}^3}{4}f''(\xi_i) = -\frac{1}{12}h_{i+1}^3f''(\xi_i),\end{aligned}$$

где $\xi_i \in (x_i, x_{i+1})$ — одна и та же точка для функций $F(x)$ и $f(x)$. Знак полученной разности указывает на то, что если вторая производная $f''(x)$ на частичном отрезке $[x_i, x_{i+1}]$ положительна, то формула (5.46) аппроксимирует I_i^{i+1} с избытком. Из полученного соотношения для модуля разности $|I_i^{i+1} - \hat{I}_{i,\text{тр}}^{i+1}|$ вытекает оценка

$$|I_i^{i+1} - \hat{I}_{i,\text{тр}}^{i+1}| \leq \frac{M_{2,i}}{12}h_{i+1}^3,$$

где $M_{2,i} = \max_{[x_i, x_{i+1}]} |f''(x)|$, которая означает, что одноинтервальная формула трапеций аппроксимирует I_i^{i+1} с третьим порядком по h .

Чтобы перейти к оценке погрешности аппроксимации интеграла I_a^b по составной формуле трапеций на всем отрезке, соотношения для разности $I_i^{i+1} - \hat{I}_{i,\text{тр}}^{i+1} = -\frac{1}{12}h_{i+1}^3f''(\xi_i)$ необходимо просуммировать по всем частичным отрезкам (при этом полагается $h = \text{const}$):

$$I_a^b - \hat{I}_{a,\text{тр}}^b = -\frac{1}{12}h^3(f''(\xi_0) + f''(\xi_1) + f''(\xi_2) + \dots + f''(\xi_{n-1})).$$

Применяя к сумме в правой части утверждение В.1, получаем

$$I_a^b - \hat{I}_{a,\text{тр}}^b = -\frac{h^3}{12} n f''(\xi), \quad \xi \in [a, b].$$

Учитывая, что $n = \frac{b-a}{h}$, $M_2 = \max_{[a,b]} |f''(x)|$, и переходя к неравенству, находим оценку

$$\left| I_a^b - \hat{I}_{a,\text{тр}}^b \right| \leq \frac{M_2}{12} (b-a) h^2. \quad (5.51)$$

Отметим, что для оценки погрешностей квадратурных формул имеются и другие способы, изложенные, например в [29], [40].

Оценочная формула (5.51) позволяет сделать следующие выводы.

1. Порядок аппроксимации составной квадратурной формулы трапеций на единицу меньше порядка аппроксимации одноинтервальной формулы и равен двум.

2. Величина остаточного слагаемого зависит от длины отрезка интегрирования $[a, b]$ и M_2 . Константа в правой части равна $\frac{M_2(b-a)}{12}$.

Проделав аналогичные выкладки для одноинтервальной и составной формул прямоугольников и парабол, получим оценки [40]:

$$\left| I_a^b - \hat{I}_{a,\text{пр}}^b \right| \leq \frac{M_2}{24} (b-a) h^2; \quad (5.52)$$

$$\left| I_a^b - \hat{I}_{a,\text{пар}}^b \right| \leq \frac{M_4}{180} (b-a) h^4, \quad (5.53)$$

где $M_4 = \max_{[a,b]} |f^{(4)}(x)|$ при $f(x) \in C_4[a, b]$. Отсюда вытекает, что составная формула прямоугольников имеет на отрезке $[a, b]$ второй порядок аппроксимации и константу $\frac{M_2(b-a)}{24}$, которая в два раза меньше соответствующей константы в оценке погрешности квадратурной формулы трапеций. При $f''(x) > 0$ на частичном отрезке $[x_i, x_{i+1}]$ одноинтервальная формула (5.45) аппроксирует I_l^{i+1} с недостатком.

Двухинтервальная и составная формулы парабол аппроксируют I_{i-1}^{i+1} и I_a^b соответственно с пятым и четвертым порядком по h . Причем повышенный на единицу порядок аппроксимации формулы (5.47) по сравнению с порядком, устанавливаемым по правилу соответствия порядков (последний для двухинтервальной формулы парабол равен четырем), получился в силу симметричности этой формулы при $h = \text{const}$.

Замечание. Между максимальной степенью алгебраических многочленов, для которых квадратурная формула является точной, и порядком аппроксимации по отношению к шагу h имеется связь. Формулы прямоугольников и трапеций точны для многочленов первой степени и обладают вторым порядком аппроксимации. Формула Симпсона является точной для многочленов третьей степени и имеет четвертый порядок аппроксимации.

Вышеприведенные оценки погрешностей аппроксимационных квадратурных формул позволяют для непрерывных функций априорным путем

вычислять шаг интегрирования, который обеспечивает заданную точность. Для сеточных функций априорные оценки погрешностей выполняются путем предварительной аппроксимации функции $y_i = f(x_i)$ и определения соответствующей величины M_k ($k = 2, 3, 4$) по формулам численного дифференцирования. Далее опишем методику вычисления интеграла с априорным нахождением шага h .

Методика вычисления определенного интеграла с заданной точностью и априорным нахождением шага интегрирования

1. Для правой части формулы оценки погрешностей вычислить константу $M_p = \max_{[a,b]} |f^{(p)}(x)|$. С этой целью необходимо продифференцировать функцию p раз и вычислить ее максимальное значение на отрезке $[a, b]$, где p — порядок аппроксимации квадратурной формулы.

2. Из условия

$$\frac{M_p}{A} (b-a) \cdot h^p \leq \varepsilon,$$

где $\frac{M_p}{A}$ — константа, входящая в правую часть оценки погрешностей, определяется величина h :

$$h \leq \sqrt[p]{\frac{A \cdot \varepsilon}{M_p (b-a)}}.$$

3. По значению h вычислить n — количество разбиений отрезка $[a, b]$ и сформировать сеточное представление функции $y = f(x)$, т. е. $y_i = f(x_i)$, $x_0 = a$; $x_1 = a + h$; $x_2 = a + 2h$; ...; $x_n = a + n \cdot h$ ($i = 0, 1, \dots, n$).

4. Полученную сеточную функцию подставить в правую часть соответствующей квадратурной формулы и вычислить искомое значение \hat{I}_a^b . При этом значение интеграла в силу справедливости оценки удовлетворяет заданной точности ε .

Замечание. Данная методика может быть использована в основном для сеточных представлений формульных функций, для которых можно выбирать величину шага.

Пример 5.8. Вычислить на основе квадратурных формул трапеций интеграл $I_1^2 = \int_1^2 \frac{dx}{x}$ с точностью $\varepsilon = 0,01$. Оценить фактическую погрешность вычисленного значения.

□ Воспользуемся методикой.

1. Поскольку $p = 2$, вычислим $M_2 = \max_{[1;2]} |f''(x)|$. Дифференцируя $f(x) = \frac{1}{x}$ дважды, получаем $f''(x) = \frac{2}{x^3}$. Данная функция монотонно убывающая, поэтому она достигает максимума в точке $x = 1$. Таким образом, $M_2 = 2$.

2. Найдем шаг интегрирования, который обеспечивает заданную точность, из условия $\frac{M_2}{12}(b-a)h^2 \leq \varepsilon$. Таким образом,

$$h \leq \sqrt{\frac{12 \cdot \varepsilon}{(b-a)M_2}} = \sqrt{\frac{12 \cdot 0,1}{1 \cdot 2}} = 0,245.$$

Примем $h = 0,2$.

3. По значению $h = 0,2$ вычислим $n = \frac{b-a}{h} = \frac{1}{0,2} = 5$ и сформируем сеточное представление функции, т. е.

$$f_0 = \frac{1}{x_0} = \frac{1}{1} = 1, \quad f_1 = \frac{1}{x_0 + h} = \frac{1}{1,2} = 0,83(3), \quad f_2 = \frac{1}{x_0 + 2h} = \frac{1}{1,4} = 0,7142,$$

$$f_3 = \frac{1}{x_0 + 3h} = \frac{1}{1,6} = 0,6250, \quad f_4 = \frac{1}{x_0 + 4h} = \frac{1}{1,8} = 0,5555, \quad f_5 = \frac{1}{x_0 + 5h} = \frac{1}{2} = 0,5000.$$

4. Вычислим интеграл по формуле (5.49) при $n = 5$:

$$\begin{aligned} \hat{I}_{1,\text{tp}}^2 &= \frac{h}{2} \cdot [f_0 + 2(f_1 + f_2 + f_3 + f_4) + f_5] = \\ &= 0,1 \cdot [1 + 2 \cdot [0,83(3) + 0,7142 + 0,625 + 0,5(5)] + 0,5] = 0,695. \end{aligned}$$

Точное значение интеграла $I_1^2 = \ln 2 = 0,693147$. Оценим фактическую абсолютную погрешность полученного (приближенного) значения:

$$\delta_{\text{факт}} = |0,69563 - 0,693147| = 0,002483.$$

Из оценки следует, что фактическая погрешность не превышает 0,01, т. е. $\delta_{\text{факт}} < 0,01 = \varepsilon$. ■

Пример 5.9. Вычислить интеграл $I_1^2 = \int_1^2 \frac{dx}{x}$ на основе квадратурной формулы парабол с точностью $\varepsilon = 0,01$. Результат сопоставить с полученным в примере 5.8.

□1, 2. Аналогично примеру 5.8 найдем шаг интегрирования из неравенства

$$\frac{h^4(b-a)}{180} M_4 < \varepsilon, \quad M_4 = \max_{[1;2]} \left| \frac{24}{x^5} \right| = 24, \quad \text{т. е. } h \leq \sqrt[4]{\frac{180 \cdot \varepsilon}{(b-a)M_4}} = 0,52.$$

Примем $h = 0,5$.

3. По значению $h = 0,5$ вычислим $n = \frac{b-a}{h} = \frac{2-1}{0,5} = 2 = 2k$. Следовательно, $k = 1$. Сформируем сеточное представление функции:

$$f_0 = f(1) = 1, \quad f_1 = f(1,5) = \frac{2}{3}, \quad f_2 = f(2) = \frac{1}{2}.$$

Вычислим интеграл по формуле (5.50):

$$\hat{I}_{1,\text{пар}}^2 = \frac{h}{3} [f_0 + 4f_1 + f_2] = 0,6944.$$

Фактическая абсолютная погрешность $\delta_{\text{факт}} = |0,69444 - 0,693147| = 0,001293$. Таким образом, фактическая погрешность не превышает $\varepsilon = 0,01$.

Шаг интегрирования по формуле парабол получился более чем в два раза крупнее шага h , соответствующего квадратурной формуле трапеций. Это является следствием того, что (5.50) имеет четвертый порядок аппроксимации, а (5.49) — второй. В связи с повышенной точностью квадратурной формулы парабол она нашла широкое применение в вычислительной практике. ■

Замечания

1. Рассмотренный способ вычисления интегралов, когда с использованием оценок и точности ε предварительно вычисляется шаг интегрирования h , является способом *с априорным определением шага h* . В вычислительной практике большое распространение получил и другой *апостериорный способ вычисления h* , основанный на многократном дроблении отрезка $[a, b]$. В этом случае интеграл вычисляется с помощью итерационного алгоритма методом Рунге (см. п. 5.4).

2. Существуют и другие широко используемые квадратурные формулы.

Одноинтервальная формула прямоугольников (немодифицированная) на двухточечном шаблоне $\Pi_{2,i} = (x_i, x_{i+1})$, где $r = 0, s = 1$ (рис. 5.3а):

$$\hat{I}_{i,c}^{i+1} = h \cdot f_i \quad (O(h^2)).$$

Четырехинтервальная формула Боде на пятиточечном шаблоне $\Pi_{5,i} = (x_{i-2}, x_{i-1}, x_i, x_{i+1}, x_{i+2})$, где $r = 2, s = 2$ (рис. 5.3б):

$$\hat{I}_{i-2,c}^{i+2} = \frac{2h}{45}(7f_{i-2} + 32f_{i-1} + 12f_i + 32f_{i+1} + 7f_{i+2}) \quad (O(h^7)).$$

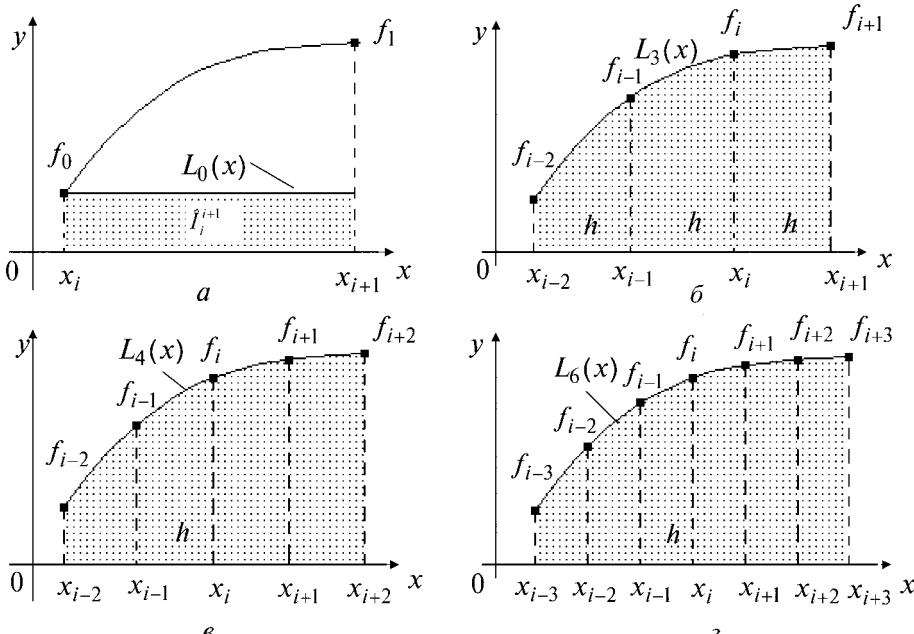


Рис. 5.3

Шестиинтервальная формула Уэдделля на семиточечном шаблоне $\Pi_{7,i} = (x_{i-3}, x_{i-2}, x_{i-1}, x_i, x_{i+1}, x_{i+2}, x_{i+3})$, где $r = 3, s = 3$ (рис. 5.3 σ):

$$\hat{I}_{i-3,c}^{i+3} = \frac{3h}{10}(f_{i-3} + 5f_{i-2} + f_{i-1} + 6f_i + f_{i+1} + 5f_{i+2} + f_{i+3}) \quad (O(h^7)).$$

Формулы Ньютона — Комеса (приведем два частных случая):

- трехинтервальная формула на четырехточечном шаблоне $\Pi_{4,i} = (x_{i-2}, x_{i-1}, x_i, x_{i+1})$, где $r = 2, s = 1$ (формула «трех восьмых», рис. 5.3 δ):

$$\hat{I}_{i-2,c}^{i+1} = \frac{3h}{8}(f_{i-2} + 3f_{i-1} + 3f_i + f_{i+1}) \quad (O(h^5));$$

- шестиинтервальная формула на семиточечном шаблоне $\Pi_{7,i} = (x_{i-3}, x_{i-2}, x_{i-1}, x_i, x_{i+1}, x_{i+2}, x_{i+3})$, где $r = 3, s = 3$ (рис. 5.3 σ):

$$\hat{I}_{i-3,c}^{i+3} = \frac{h}{140}(41f_{i-3} + 216f_{i-2} + 27f_{i-1} + 272f_i + 27f_{i+1} + 216f_{i+2} + 41f_{i+3}) \quad (O(h^9)).$$

3. Если по условию задачи имеется право выбора узлов и функция обладает высокой степенью гладкости, то часто применяется квадратурная формула Гаусса, в которой весовые коэффициенты и узлы подбираются так, чтобы приближенное равенство было точным для всех многочленов наивысшей степени.

Пример 5.10. Вычислить интегралы

$$I_1 = \int_0^2 x dx; \quad I_2 = \int_0^2 x^2 dx; \quad I_3 = \int_0^2 x^3 dx; \quad I_4 = \int_0^2 x^4 dx$$

по формулам прямоугольников, трапеций, парабол с шагом $h = 1$. Найти оценки погрешностей.

□ Точные значения интегралов:

$$I_1 = \int_0^2 x dx = \frac{x^2}{2} \Big|_0^2 = 2; \quad I_2 = \int_0^2 x^2 dx = \frac{x^3}{3} \Big|_0^2 = \frac{8}{3};$$

$$I_3 = \int_0^2 x^3 dx = \frac{x^4}{4} \Big|_0^2 = 4; \quad I_4 = \int_0^2 x^4 dx = \frac{x^5}{5} \Big|_0^2 = \frac{32}{5} = 6,4.$$

Для формул прямоугольников и трапеций порядок аппроксимации $p = 2$, а для формулы парабол $p = 4$. В поставленной задаче $a = 0, b = 2$. Сначала получим оценки погрешностей априорным способом.

Найдем $M_2 = \max |f''(x)|$:

- $M_2 = 0$ для функции $f(x) = x$;
- $M_2 = 2$ для функции $f(x) = x^2$;
- $M_2 = 12$ для функции $f(x) = x^3$;
- $M_2 = 48$ для функции $f(x) = x^4$.

Найдем $M_4 = \max_{[0,2]} |f^{(4)}(x)|$:

- $M_4 = 0$ для функций $f(x) = x; f(x) = x^2; f(x) = x^3$;
- $M_4 = 24$ для функции $f(x) = x^4$.

Согласно (5.51)–(5.53), справедливы оценки:

$$|\varepsilon_{\text{пп}}| \leq \frac{M_2}{24}(b-a)h^2; \quad |\varepsilon_{\text{тр}}| \leq \frac{M_2}{12}(b-a)h^2; \quad |\varepsilon_{\text{папр}}| \leq \frac{M_4}{180}(b-a)h^4.$$

Оценки погрешностей формулы прямоугольников:

$$|\varepsilon_{\text{пп}}| = 0 \text{ для } f(x) = x;$$

$$|\varepsilon_{\text{пп}}| \leq \frac{2}{24} \cdot 2 \cdot 1^2 = 0,16(6) \text{ для } f(x) = x^2;$$

$$|\varepsilon_{\text{пп}}| \leq \frac{12}{24} \cdot 2 \cdot 1^2 = 1 \text{ для } f(x) = x^3;$$

$$|\varepsilon_{\text{пп}}| \leq \frac{48}{24} \cdot 2 \cdot 1^2 = 8 \text{ для } f(x) = x^4.$$

Оценки погрешностей формулы трапеций:

$$|\varepsilon_{\text{тр}}| = 0 \text{ для } f(x) = x;$$

$$|\varepsilon_{\text{тр}}| \leq \frac{2}{12} \cdot 2 \cdot 1^2 = 0,3(3) \text{ для } f(x) = x^2;$$

$$|\varepsilon_{\text{тр}}| \leq \frac{12}{12} \cdot 2 \cdot 1^2 = 2 \text{ для } f(x) = x^3;$$

$$|\varepsilon_{\text{тр}}| \leq \frac{48}{12} \cdot 2 \cdot 1^2 = 8 \text{ для } f(x) = x^4.$$

Оценки погрешностей формулы парабол:

$$|\varepsilon_{\text{папр}}| = 0 \text{ для } f(x) = x; f(x) = x^2; f(x) = x^3;$$

$$|\varepsilon_{\text{папр}}| \leq \frac{24}{180} \cdot 2 \cdot 1^2 = 0,26(6) \text{ для } f(x) = x^4.$$

Таким образом, подтверждается факт, что формулы прямоугольников и трапеций должны быть точными для многочленов первой степени, а формула парабол — для многочленов не выше третьей степени.

Теперь рассчитаем значения интегралов по соответствующим квадратурным формулам.

При $h = 1$ сеточное представление функций имеет вид

$$f_0 = f(0), \quad f_{\frac{1}{2}} = f\left(\frac{1}{2}\right), \quad f_1 = f(1), \quad f_{\frac{3}{2}} = f\left(\frac{3}{2}\right), \quad f_2 = f(2).$$

По формуле прямоугольников (5.48) получаем $\hat{I}_{\text{пп}} = h \cdot \left[f_{\frac{1}{2}} + f_{\frac{3}{2}} \right]$, в частности:

$$\hat{I}_1 = 1 \cdot \left[\frac{1}{2} + \frac{3}{2} \right] = 2 \text{ (0);} \quad \hat{I}_2 = 1 \cdot \left[\frac{1}{4} + \frac{9}{4} \right] = 2,5 \text{ (0,16(6));}$$

$$\hat{I}_3 = 1 \cdot \left[\frac{1}{8} + \frac{27}{8} \right] = \frac{7}{2} = 3,5 \text{ (0,5);} \quad \hat{I}_4 = 1 \cdot \left[\frac{1}{16} + \frac{81}{16} \right] = \frac{82}{16} = 5,125 \text{ (1,275).}$$

Здесь в скобках указана величина фактической ошибки.

По формуле трапеций (5.49) находим $\hat{I}_{\text{тр}} = \frac{h}{2} \cdot [f_0 + 2f_1 + f_2]$, в частности:

$$\hat{I}_1 = \frac{1}{2} \cdot [0 + 2 + 2] = 2 \text{ (0); } \hat{I}_2 = \frac{1}{2} \cdot [0 + 2 + 4] = 3 \text{ (0,3(3));}$$

$$\hat{I}_3 = \frac{1}{2} \cdot [0 + 2 + 8] = 5 \text{ (1); } \hat{I}_4 = \frac{1}{2} \cdot [0 + 2 + 16] = 9 \text{ (2,6).}$$

По формуле парабол (5.50), учитывая, что $n = 2k = 2$ и, следовательно, $k = 1$, получаем $\hat{I}_{\text{пар}} = \frac{h}{3} \cdot [f_0 + 4f_1 + f_2]$, в частности:

$$\hat{I}_1 = \frac{1}{3} \cdot [0 + 4 + 2] = 2 \text{ (0); } \hat{I}_2 = \frac{1}{3} \cdot [0 + 4 + 4] = \frac{8}{3} \text{ (0);}$$

$$\hat{I}_3 = \frac{1}{3} \cdot [0 + 4 + 8] = 4 \text{ (0); } \hat{I}_4 = \frac{1}{3} \cdot [0 + 4 + 16] = \frac{20}{3} = 6,6(6) \text{ (0,26(6)).}$$

Очевидно, полученные фактические погрешности соответствуют вычисленным ранее оценкам. ■

Пример 5.11. Вычислить интеграл $I = \int_0^{\frac{\pi}{2}} \sin x dx$ методами прямоугольников, трапеций, парабол, Боде, Уэддля, Ньютона — Котеса на семиточечном шаблоне, применяя разбиение на 1, 2, 4, 8 частичных отрезка.

□ Точное значение интеграла:

$$I = \int_0^{\frac{\pi}{2}} \sin x dx = 1.$$

Разобьем отрезок $\left[0; \frac{\pi}{2}\right]$ на $l = 1, 2, 4, 8$ частичных отрезка. Применим методы прямоугольников, трапеций, парабол, используя составные формулы (5.48)–(5.50). Остальные значения интеграла подсчитаем по формуле (5.44) при $l = 1, 2, 4, 8$, используя на каждом частичном отрезке соответствующую квадратурную формулу Боде, Уэддля, Ньютона — Котеса на семиточечном шаблоне. Результаты расчетов приведем в таблице 5.4.

Из данного примера видно, что для повышения точности возможны два альтернативных подхода:

а) увеличивать количество частичных отрезков, и на каждом из них использовать более простую квадратурную формулу;

б) применять более точные формулы на пяти- и семиточечном шаблонах при малом количестве частичных отрезков. ■

Таблица 5.4

Метод	$l = 1$	$l = 2$	$l = 4$	$l = 8$
Прямоугольников	1,110720338230	1,026171820190	1,006454224265	1,001607874019
Трапеций	0,785398006439	0,948059172335	0,987115496263	0,996784860265
Парабол	1,002279560960	1,000134270907	1,000007981598	1,000000202767
Боде	0,999991251569	0,999999562310	0,999999684178	0,999999686054
Уэддля	0,999999293425	0,999999680058	0,999999685980	0,999999686082
Ньютона — Котеса по 7 точкам	0,999999711921	0,999999686177	0,999999686084	0,999999686083

5.3.2. ФОРМУЛЫ, ПОЛУЧЕННЫЕ НА ОСНОВЕ СПЛАЙНОВ

ПРИМЕНЕНИЕ ПАРАБОЛИЧЕСКИХ ИНТЕГРАЛЬНО-ДИФФЕРЕНЦИАЛЬНЫХ СПЛАЙНОВ

Явные формулы

Пусть в общем случае на неравномерной сетке Ω_n с некоторой точностью задана сеточная функция $y_i = f(x_i)$, $i = \overline{0, n}$. Предположим, что точность этого задания в каждом из узлов не ниже $O(h_{i+1}^3)$. В частном случае $f(x_i)$ является сеточным представлением формульной функции, и тогда сетка может быть выбрана исходя из характера изменения функции $f(x)$. Например, в зонах больших изменений $f(x)$ или ее производных, которые могут быть разрывными, шаг сетки желательно уменьшать. Закон мельчения сетки можно выбирать, например, по формулам арифметической или геометрической прогрессий. В последнем случае в рассмотрение вводится параметр нерегулярности сетки $\delta_{i+1} = \frac{h_{i+1}}{h_i}$, совпадающий со знаменателем прогрессии. Поэтому данный параметр будем включать здесь в правые части некоторых квадратурных формул и с его использованием проводить оценку погрешности аппроксимации.

Из параболических интегрально-дифференциальных сплайнов путем анализа их параметрических соотношений на трехточечном шаблоне $\mathbb{W}_{3,i} = (x_{i-1}, x_i, x_{i+1})$ получаются следующие лево- и правосторонние безусловные аппроксимации интегралов I_{i-1}^i , I_i^{i+1} четвертого порядка (одноинтервальные функциональные аппроксимации):

$$\hat{I}_{i-1,v}^i = \frac{h_i^3}{6H_i^{i+1}} \left(-\frac{1}{h_{i+1}} f_{i+1} + \frac{H_i^{i+1} H_i^{3(i+1)}}{h_i^2 h_{i+1}} f_i + \frac{H_{2i}^{3(i+1)}}{h_i^2} f_{i-1} \right); \quad (5.54)$$

$$\hat{I}_{i,v}^{i+1} = \frac{h_{i+1}^3}{6H_i^{i+1}} \left(\frac{H_{3i}^{2(i+1)}}{h_{i+1}^2} f_{i+1} + \frac{H_i^{i+1} H_{3i}^{(i+1)}}{h_{i+1}^2 h_i} f_i - \frac{1}{h_i} f_{i-1} \right), \quad (5.55)$$

где $H_{ki}^{m(i+1)} = kh_i + mh_{i+1}$.

Предполагая, что $f(x) \in C_3[a, b]$, и используя методику оценки погрешности [40], для данных аппроксимаций нетрудно получить оценки:

$$|\hat{I}_{i,v}^{i+1} - I_i^{i+1}| \leq \frac{h_i^2 h_{i+1}^2}{72} \delta_{i+1} (\delta_{i+1} + 2) \cdot M_{3,i}; \quad |\hat{I}_{i-1,v}^i - I_{i-1}^i| \leq \frac{h_i^4}{72} (2\delta_{i+1} + 1) \cdot M_{3,i},$$

где $M_{3,i} = \max_{[x_{i-1}, x_{i+1}]} |f''(x)|$.

При $h = \text{const}$ квадратурные формулы (5.54), (5.55) упрощаются (условная аппроксимация):

$$\hat{I}_{i-1,c}^i = \frac{h}{12} (5f_{i-1} + 8f_i - f_{i+1}); \quad \hat{I}_{i,c}^{i+1} = \frac{h}{12} (-f_{i-1} + 8f_i + 5f_{i+1}).$$

Их оценки при $\delta_{i+1} = 1$ становятся одинаковыми (с мажорантой $\frac{h^4}{24} M_{3,i}$).

Таким образом, из оценок для $\hat{I}_{i,v}^{i+1}$, $\hat{I}_{i-1,v}^i$ следует, что при безусловной аппроксимации обе квадратурные формулы имеют четвертый порядок по h . Однако для квадратурной формулы (5.55) этот порядок можно повысить, если применить условную аппроксимацию, в которой положить $\delta_{i+1} \leq h_i$. В этом случае из формулы $\delta_{i+1} = \frac{h_{i+1}}{h_i} < h_i$ получаем неравенство $h_{i+1} < h_i^2$, устанавливающее очень малый шаг h_{i+1} , такой, что x_{i+1} отстоит от x_i на величину второго порядка по h_i . Этот тип шаблона можно назвать «сильно сжатым», так как он как бы устремляет трехточечный шаблон к двухточечному, и в связи с этим сама величина интеграла $\hat{I}_{i,v}^{i+1}$ является малой.

Суммируя левые и правые части квадратурных формул (5.54), (5.55), после преобразований получаем компактную, двухинтервальную, трехточечную обобщенную формулу парабол функционального типа:

$$\hat{I}_{i-1}^{i+1} = \frac{h_i(\delta_{i+1} + 1)}{6} \left[(2 - \delta_{i+1})f_{i-1} + \frac{(\delta_{i+1} + 1)^2}{\delta_{i+1}} f_i + \left(2 - \frac{1}{\delta_{i+1}}\right) f_{i+1} \right]. \quad (5.56)$$

Предполагая, что разбиение отрезка $[a, b]$ четное, т. е. $n = 2k$, где k — количество пар разбиения, просуммируем (5.56) по k . Получаем обобщенную на нерегулярный шаблон составную квадратурную формулу парабол

$$\begin{aligned} \hat{I}_a^b = & \frac{h_1(1 + \delta_2)(2 - \delta_2)}{6} f_0 + \sum_{i=1}^k \frac{(\delta_{2i} + 1)^3 h_{2i-1}}{6\delta_{2i}} f_{2i-1} + \\ & + \frac{1}{6} \sum_{i=1}^{k-1} f_{2i} \left[(1 + \delta_{2i})h_{2i-1} \left(2 - \frac{1}{\delta_{2i}}\right) + (1 + \delta_{2i+2})h_{2i+1}(2 - \delta_{2i+2}) \right] + \\ & + \frac{(1 + \delta_{2n})}{6} h_{2k-1} \left(2 - \frac{1}{\delta_{2n}}\right) f_{2k}. \end{aligned} \quad (5.57)$$

Известно, что для квадратурных формул важно, чтобы их коэффициенты были положительны. Это свойство, называемое *свойством положительности коэффициентов*, обеспечивает минимум погрешности вычисления интеграла по данной квадратурной формуле, а также сходимость вычислительного процесса на последовательно сгущающихся сетках [28]. В квадратурной формуле (5.56) свойство положительности будет выполнено, если $2 - \delta_{i+1} > 0$, $2 - \frac{1}{\delta_{i+1}} > 0$. Отсюда вытекают пределы изменения параметра нерегулярности δ_{i+1} : $0,5 < \delta_{i+1} < 2$.

Предполагая, что $f(x) \in C_3[a, b]$, для (5.56) получаем остаточное слагаемое и оценку погрешности:

$$|I_{i-1}^{i+1} - \hat{I}_{i-1}^{i+1}| \leq \frac{h_i^4 |\delta_{i+1}^2 - 1|}{72} (\delta_{i+1} + 1)^2 M_{3,i}. \quad (5.58)$$

Переходя в (5.58) к мажоранте для оценки интеграла на всем отрезке $[a, b]$, имеем

$$|I_a^b - \hat{I}_a^b| \leq \frac{H_m^4 \tilde{\Delta}_m n}{72} (\Delta_m + 1)^2 M_3 \leq \frac{(b-a) H_m^3 \tilde{\Delta}_m}{144} (\Delta_m + 1)^2 M_3,$$

где $H_m = \max_i |h_i|$, $i = \overline{1, n}$, $\tilde{\Delta}_m = \max_i |\delta_{i+1}^2 - 1|$, $\Delta_m = \max_i \delta_{i+1}$, $M_3 = \max_{[a,b]} |f''(x)|$.

Приведем анализ соотношения (5.58), устанавливающего порядок аппроксимации обобщенной квадратурной двухинтервальной формулы парабол.

В общем случае квадратурная формула (5.56) имеет четвертый порядок аппроксимации (безусловная аппроксимация). Возможны также условные аппроксимации двух типов:

а) при $\delta_{i+1} = 1$ реализуется регулярный шаблон и тогда порядок аппроксимации повышается на единицу, а (5.56) переходит в классическую формулу парабол (Симпсона);

б) при $|\delta_{i+1} - 1| \leq h_i$ реализуется нерегулярный шаблон, и тогда порядок аппроксимации может быть повышен на единицу за счет небольшого отклонения δ_{i+1} от единицы: $h_i - h_i^2 \leq h_{i+1} \leq h_i + h_i^2$. В этом случае шаблон является *квазирегулярным*. Такие разбиения с локальными сгущениями узлов сетки могут формироваться при численном интегрировании с помощью квадратурной формулы (5.56) сильно изменяющихся функций или функций, имеющих в некоторых окрестностях разрывы производных.

Неславные методы

Из анализа параметрических соотношений (4.101) и (4.105) двух типов параболических интегрально-дифференциальных сплайнов, изложенных в п. 4.5.5, для внутренних узлов получаются системы линейных алгебраических уравнений относительно интегралов:

$$\begin{aligned} \frac{h_i^2}{h_{i+1}^2} I_{i-1}^i + \frac{2}{h_{i+1}^2} I_i^{i+1} + \frac{1}{h_{i+2}^2} I_{i+2}^{i+2} &= \frac{1}{3} \left[\frac{1}{h_i} f_{i-1} + \left(\frac{2}{h_i} + \frac{3}{h_{i+1}} \right) f_i + \left(\frac{3}{h_{i+1}} + \frac{2}{h_{i+2}} \right) f_{i+1} + \frac{1}{h_{i+2}} f_{i+2} \right], \\ i &= \overline{1, n-2}; \\ \frac{1}{h_i} I_{i-1}^i - \frac{2}{h_{i+1}} I_i^{i+1} + \frac{1}{h_{i+2}} I_{i+2}^{i+2} &= \frac{1}{6} \left[-h_i f'_{i-1} - H_{2i}^{i+1} f'_i + H_{i+1}^{2(i+2)} f'_{i+1} + h_{i+2} f'_{i+2} \right], \quad i = \overline{1, n-2}, \end{aligned}$$

которые при $h = \text{const}$ преобразуются к виду

$$\begin{aligned} I_{i-1}^i + 2I_i^{i+1} + I_{i+1}^{i+2} &= \frac{h}{3} [f_{i-1} + 5 \cdot (f_i + f_{i+1}) + f_{i+2}], \quad i = \overline{1, n-2}; \\ I_{i-1}^i - 2I_i^{i+1} + I_{i+1}^{i+2} &= \frac{h^2}{6} [-f'_{i-1} + 3 \cdot (f'_{i+1} - f'_i) + f'_{i+2}], \quad i = \overline{1, n-2}. \end{aligned}$$

Для вычисления интегралов на всех отрезках эти системы необходимо дополнить двумя граничными условиями. Это можно сделать двумя способами.

1. Системы дополняют значениями интегралов I_0^1 и I_{n-1}^n на конечных отрезках, которые вычисляются по односторонним формулам (5.54), (5.55) соответственно.

2. Каждая из систем может быть дополнена двумя связями при $i = 1$ и $i = n - 1$, которые выражаются из соотношения

$$\frac{1}{h_i^2} I_{i-1}^i + \frac{1}{h_{i+1}^2} I_i^{i+1} = \frac{1}{3} \left[\frac{1}{h_i} f_{i-1} + 2 \left(\frac{1}{h_i} + \frac{1}{h_{i+1}} \right) f_i + \frac{1}{h_{i+1}} f_{i+1} \right].$$

Замечания

1. Первая из приведенных здесь систем связывает искомые интегралы со значениями интегрируемой функции, а вторая — со значениями производных.

2. Описанные методы применимы для сеточных функций с $n \geq 3$.

3. Полученные системы линейных алгебраических уравнений являются трехдиагональными и решаются методом прогонки (см. главу 1).

ПРИМЕНЕНИЕ КУБИЧЕСКИХ ИНТЕГРАЛЬНО-ДИФФЕРЕНЦИАЛЬНЫХ СПЛАЙНОВ

Явные формулы

Если кубические многочлены, построенные на $[x_i, x_{i+1}]$ с помощью функциональных, дифференциальных и интегральных параметров f_i, f'_i, I_i^{i+1} , использовать для вывода квадратурных формул, то получаются различные типы одноинтервальных функционально-дифференциальных квадратурных формул Эйлера пятого порядка.

Квадратурная формула Эйлера — Маклорена имеет вид [2]

$$\hat{I}_i^{i+1} = \frac{h_{i+1}}{2} (f_i + f_{i+1}) - \frac{h_{i+1}^2}{12} (f'_{i+1} - f'_i) \quad \left(\frac{h_{i+1}^5}{720} M_{4,i} \right). \quad (5.59)$$

Суммируя (5.59) по всем отрезкам, записываем *составную квадратурную формулу Эйлера* четвертого порядка аппроксимации на нерегулярной сетке:

$$\hat{I}_{a,v}^b = \frac{h_1}{2} f_0 + \sum_{i=1}^{n-1} H_i^{i+1} f_i + \frac{h_n}{2} f_n - \frac{1}{12} \sum_{i=0}^{n-1} h_{i+1}^2 \Delta f'_{i+1}. \quad (5.60)$$

Подчеркнем, что производные f'_i для подстановки в (5.60) предварительно должны быть вычислены по вышеупомянутым аппроксимационным формулам с порядком не ниже третьего. Сокращать производные в последнем слагаемом составной квадратурной формулы, как это обычно делается в учебниках по численным методам, не следует, поскольку при этом в сущности предполагается, что все производные во внутренних узлах известны точно. Это условие для сеточных функций, от которых рассчитываются интегралы, не должно использоваться, так как производные этих функций вычисляются с некоторой погрешностью, порядок которой при суммировании по всем частичным отрезкам понижается на единицу. Справедливость этого утверждения подтверждается следующим примером. После суммирования и сокращения в последней сумме квадратурной формулы (5.60) производных во всех внутренних узлах при $h = \text{const}$ формула (5.60) приобретает вид

$$\hat{I}_a^b = \frac{h}{2} \left(f_0 + 2 \sum_{i=1}^{n-1} f_i + f_n \right) - \frac{h^2}{12} (f'_n - f'_0). \quad (5.61)$$

Если квадратурную формулу (5.61) использовать для вычисления интеграла от функции с нулевыми производными на концах или с их одинаковыми

ми значениями $f'_n = f'_0$, то квадратурная формула четвертого порядка переходит в квадратурную формулу трапеций, имеющую второй порядок. Таким образом, реализуется потеря двух порядков, что свидетельствует о необоснованности проведенного сокращения.

Квадратурную формулу (5.59) целесообразно применить для получения функциональных квадратурных формул путем подстановки в них конкретных аппроксимаций \hat{f}'_{i+1} , \hat{f}'_i . Подставив, например, оператор $\hat{f}'_i = \frac{\hat{f}_{i+1} - \hat{f}_{i-1}}{2h}$, а также аппроксимации производных в левой и правой крайних точках четырехточечного шаблона, получим *одноинтервальные функциональные квадратурные формулы* — центральную и лево- и правостороннюю:

$$\hat{I}_{i,\text{центр}}^{i+1} = \frac{h}{24}(-\hat{f}_{i-1} + 13(f_i + \hat{f}_{i+1}) - \hat{f}_{i+2}) \quad \left(\frac{11}{720} h_{i+1}^5 M_{4,i} \right); \quad (5.62)$$

$$\hat{I}_{i,\text{лев}}^{i+1} = \frac{h}{24}(9f_i + 19f_{i+1} - 5f_{i+2} + f_{i+3}) \quad \left(\frac{19}{720} h_{i+1}^5 M_{4,i} \right); \quad (5.63)$$

$$\hat{I}_{i,\text{прав}}^{i+1} = \frac{h}{24}(f_{i-2} - 5f_{i-1} + 19f_i + 9f_{i+1}) \quad \left(\frac{19}{720} h_{i+1}^5 M_{4,i} \right). \quad (5.64)$$

Последние три квадратурные формулы (5.62)–(5.64) могут быть объединены в единую *составную квадратурную формулу*:

$$\hat{I}_a^b = \hat{I}_{0,\text{лев}}^1 + \frac{h}{24} \left(-f_0 + 12f_1 + 25f_2 + 24 \sum_{i=3}^{n-3} f_i + 25f_{n-2} + 12f_{n-1} - f_n \right) + \hat{I}_{n-1,\text{прав}}^n.$$

Данная формула для всех внутренних узлов удовлетворяет условию положительности коэффициентов. Отметим также, что хотя порядок аппроксимации формулы (5.62) равен порядку аппроксимации квадратурной формулы парабол, однако их аппроксимационные свойства различаются. Это связано с тем, что квадратурная формула (5.62) вытекает из полинома третьей степени, а формула парабол — из полинома второй степени. Поэтому квадратурная формула (5.62) учитывает, например, наличие точки перегиба, если она имеется в интегрируемых функциях, в то время как формула парабол этого не позволяет, так как получается из многочлена второй степени. Это обуславливает преимущество формулы (5.62) по сравнению с формулой парабол при интегрировании функций, имеющих указанную, а также другие особенности, присущие полиномам нечетной степени.

Неявный метод

Неявный метод численного интегрирования получается путем анализа параметрических соотношений кубических интегрально-дифференциальных сплайнов. После преобразования соотношения (5.59) путем подстановки в него производных из п. 5.2.3 на нерегулярной сетке для внутренних узлов получается система линейных алгебраических уравнений относительно интегралов

$$\begin{aligned}
& \frac{h_{i+1}^3}{3h_i^2 H_i^{i+1}} I_{i-1}^i + \left(1 - \frac{h_i}{3H_i^{i+1}} - \frac{h_{i+2}}{3H_{i+1}^{i+2}}\right) I_{i+1}^{i+1} + \frac{h_{i+1}^3}{3h_{i+2}^2 H_{i+1}^{i+2}} I_{i+2}^{i+2} = \\
& = \frac{h_{i+1}^3}{12h_i H_i^{i+1}} f_{i-1} + \frac{h_{i+1}}{2} \left(1 - \frac{h_{i+2}}{6H_{i+1}^{i+2}} + \frac{h_{i+1}^2 - h_i^2}{2h_i H_i^{i+1}}\right) f_i + \\
& + \frac{h_{i+1}}{2} \left(1 + \frac{h_{i+1}^2 - h_{i+2}^2}{2h_{i+2} H_{i+1}^{i+2}} - \frac{h_i}{6H_i^{i+1}}\right) f_{i+1} + \frac{h_{i+1}^3}{12h_{i+2} H_{i+1}^{i+2}} f_{i+2}, \quad i=1, 2, \dots, n-2.
\end{aligned} \tag{5.65}$$

На регулярной сетке эта система преобразуется к более простому виду:

$$I_{i-1}^i + 4I_i^{i+1} + I_{i+1}^{i+2} = \frac{h}{4} [f_{i-1} + 11 \cdot (f_i + f_{i+1}) + f_{i+2}], \quad i=1, 2, \dots, n-2. \tag{5.66}$$

Для вычисления интегралов на всех отрезках эти системы необходимо дополнить значениями I_0^1 и I_{n-1}^n , которые могут быть вычислены по квадратурной формуле (5.59) при $h_i = \text{var}$ или по формулам (5.63), (5.64) при $h = \text{const}$. Полученная в результате система линейных алгебраических уравнений решается методом прогонки (см. главу 1).

Система (5.65) удовлетворяет условию преобладания диагональных элементов в следующем диапазоне изменения параметра нерегулярности:

$$\delta_*^{-1} \leq \delta_{i+1} \leq \delta_*,$$

где $\delta_* \approx 1,722$.

Для системы (5.66) в [24] дана оценка погрешности вычисления интеграла \hat{I}_a^b на всем отрезке $[a, b]$, которая имеет вид

$$|I_a^b - \hat{I}_{a,\text{неяв}}^b| \leq \frac{h^4(a-b)}{360} M_4,$$

где $M_4 = \max_{[a,b]} |f^{(4)}(x)|$.

ПРИМЕНЕНИЕ ИНТЕГРАЛЬНО-ДИФФЕРЕНЦИАЛЬНЫХ СПЛАЙНОВ ЧЕТВЕРТОЙ СТЕПЕНИ

Явные формулы

В данном пункте приводятся функционально-дифференциальные квадратурные формулы, полученные из интегрально-дифференциальных многочленов четвертой степени [16]. В соответствии с порядком аппроксимации многочленами четвертой степени, который равен пяти, порядок аппроксимации интегралов равен шести.

Одноинтервальные явные квадратурные формулы, представляющие функционально-дифференциальные аппроксимации на трехточечном шаблоне $\mathbb{W}_{3,i} = (x_{i-1}, x_i, x_{i+1})$, имеют вид

$$\begin{aligned}
\hat{I}_{i-1,v}^i &= P_0 \left[\frac{2}{h_i^2} \left(\frac{7}{h_i} + \frac{6}{h_{i+1}} \right) f_{i-1} + \left(\frac{16}{h_i^3} + \frac{18}{h_i^2 h_{i+1}} - \frac{2}{h_{i+1}^3} \right) f_i + \frac{2}{h_{i+1}^3} f_{i+1} + \right. \\
& \left. + \frac{1}{h_i} \left(\frac{2}{h_i} + \frac{3}{2h_{i+1}} \right) f'_{i-1} - \left(\frac{3}{h_i^2} + \frac{9}{2h_i h_{i+1}} + \frac{3}{2h_{i+1}^2} \right) f'_i - \frac{1}{2h_{i+1}^2} f'_{i+1} \right];
\end{aligned} \tag{5.67}$$

$$\hat{I}_{i,v}^{i+1} = P_0 \left[\frac{2}{h_i^3} f_{i-1} + \left(-\frac{2}{h_i^3} + \frac{18}{h_i h_{i+1}^2} + \frac{16}{h_{i+1}^3} \right) f_i + \frac{2}{h_{i+1}^2} \left(\frac{6}{h_i} + \frac{7}{h_{i+1}} \right) f_{i+1} + \frac{1}{2h_i^2} f'_{i-1} + \left(\frac{3}{2h_i^2} + \frac{9}{2h_i h_{i+1}} + \frac{3}{h_{i+1}^2} \right) f'_i - \frac{1}{h_{i+1}} \left(\frac{3}{2h_i} + \frac{2}{h_{i+1}} \right) f'_{i+1} \right], \quad (5.68)$$

где $P_0 = \frac{30}{h_k^2} \left(\frac{1}{h_i} + \frac{1}{h_{i+1}} \right)$, $k = i, i+1$, соответственно для (5.67), (5.68).

На регулярном шаблоне эти формулы упрощаются:

$$\hat{I}_{i-1,c}^i = \frac{h}{30} (13f_{i-1} + 16f_i + f_{i+1}) - \frac{h^2}{120} (f'_{i+1} + 18f'_i - 7f'_{i-1}) - \left(\frac{h^6}{7200} M_{5,i} \right); \quad (5.69)$$

$$\hat{I}_{i,c}^{i+1} = \frac{h}{30} (f_{i-1} + 16f_i + 13f_{i+1}) - \frac{h^2}{120} (7f'_{i+1} - 18f'_i - f'_{i-1}) - \left(\frac{h^6}{7200} M_{5,i} \right). \quad (5.70)$$

Суммирование формул (5.69), (5.70) приводит к *двухинтервальной квадратурной формуле* (модифицированной формуле Эйлера):

$$\hat{I}_{i-1,c}^{i+1} = \frac{h}{15} (7f_{i-1} + 16f_i + 7f_{i+1}) - \frac{h^2}{15} (f'_{i+1} - f'_{i-1}) - \left(\frac{h^7}{4725} M_{5,i} \right). \quad (5.71)$$

Порядок аппроксимации в этой формуле повышается на единицу из-за симметричного вида формулы (5.71).

Подставляя в функционально-дифференциальные квадратурные формулы аппроксимации производных на пятиточечном шаблоне с порядком $O(h^4)$, получаем *функциональные квадратурные формулы шестого порядка*:

$$\hat{I}_{i-1}^i = \frac{h}{720} (-19f_{i-2} + 346f_{i-1} + 456f_i - 74f_{i+1} + 11f_{i+2}) - \left(\frac{7}{7200} h^6 M_{5,i} \right); \quad (5.72)$$

$$\hat{I}_i^{i+1} = \frac{h}{720} (11f_{i-2} - 74f_{i-1} + 456f_i + 346f_{i+1} - 19f_{i+2}); \quad (5.73)$$

$$\hat{I}_{i-1}^{i+1} = \frac{h}{90} (-f_{i-2} + 34f_{i-1} + 114f_i + 34f_{i+1} - f_{i+2}) - \left(\frac{h^7}{756} M_{6,i} \right). \quad (5.74)$$

Подстановка в формулы (5.69), (5.70) лево- и правосторонних аппроксимаций производных в крайних точках шаблона приводит к *лево- и правосторонним квадратурным формулам шестого порядка*:

$$\hat{I}_{i-1}^i = \frac{h}{720} (251f_{i-1} + 646f_i - 264f_{i+1} + 106f_{i+2} - 19f_{i+3});$$

$$\hat{I}_i^{i+1} = \frac{h}{720} (-19f_{i-3} + 106f_{i-2} - 264f_{i-1} + 646f_i + 251f_{i+1}).$$

Для расчета интегралов $\hat{I}_{i-1,v}^i$, $\hat{I}_{i,v}^{i+1}$ на нерегулярной сетке по формулам (5.67), (5.68) предварительно необходимо рассчитать производные в соответствующих точках с порядком не ниже четвертого. С этой целью необходимо воспользоваться интерполяционным многочленом четвертой степени.

Неявный метод

В заключение данного пункта приведем неявный метод вычислений интегралов на всех частичных отрезках путем решения системы линейных алгебраических уравнений трехдиагонального вида ($h = \text{const}$):

$$7I_{i-1}^i + 19I_i^{i+1} + 4I_{i+1}^{i+2} = h \cdot (2f_{i-1} + 15f_i + 12f_{i+1} + f_{i+2}) \quad (i = \overline{1, n-2}).$$

Данная система должна быть дополнена значениями интегралов I_0^1, I_{n-1}^n , которые рассчитываются по вышеприведенным квадратурным формулам.

Можно показать, что неявный алгоритм вычисления интеграла на отрезке $[a, b]$ имеет оценку [24]:

$$|I_a^b - \hat{I}_a^b| \leq \frac{h^5(b-a)}{180} M_5.$$

5.3.3. ФОРМУЛЫ, ПОЛУЧЕННЫЕ НА ОСНОВЕ РАЗЛОЖЕНИЯ ПЕРВООБРАЗНЫХ ПО ФОРМУЛЕ ТЕЙЛОРА

Пусть $f(x) \in C_{m+1}[a, b]$ — формульная функция. Тогда из разложений первообразной $F(x)$ по формуле Тейлора относительно точек x_i и x_{i+1} и выражений для $F_{i+1} = F(x_{i+1})$ и $F_i = F(x_i)$ после преобразований получаются одновременные обобщенные квадратурные формулы Эйлера, которые имеют функционально-дифференциальный тип (см. формулу (B.24) и п. 5.2.2):

$$\hat{I}_{i,1}^{i+1} = h_{i+1} f_i + \frac{h_{i+1}^2}{2} f'_i + \frac{h_{i+1}^3}{6} f''_i + \dots = h_{i+1} \sum_{k=0}^m \frac{h_{i+1}^k}{(k+1)!} f^{(k)}_i \left(\frac{h_{i+1}^{m+2} M_{m+1}}{(m+2)!} \right); \quad (5.75)$$

$$\hat{I}_{i,2}^{i+1} = h_{i+1} f_{i+1} - \frac{h_{i+1}^2}{2} f'_{i+1} + \frac{h_{i+1}^3}{6} f''_{i+1} + \dots = h_{i+1} \sum_{k=0}^m (-1)^k \frac{h_{i+1}^k}{(k+1)!} f^{(k)}_{i+1} \left(\frac{h_{i+1}^{m+2} M_{m+1}}{(m+2)!} \right). \quad (5.76)$$

В скобках справа от квадратурных формул (5.75), (5.76) указаны мажоранты, характеризующие оценки их погрешностей. На рисунке 5.4 показан

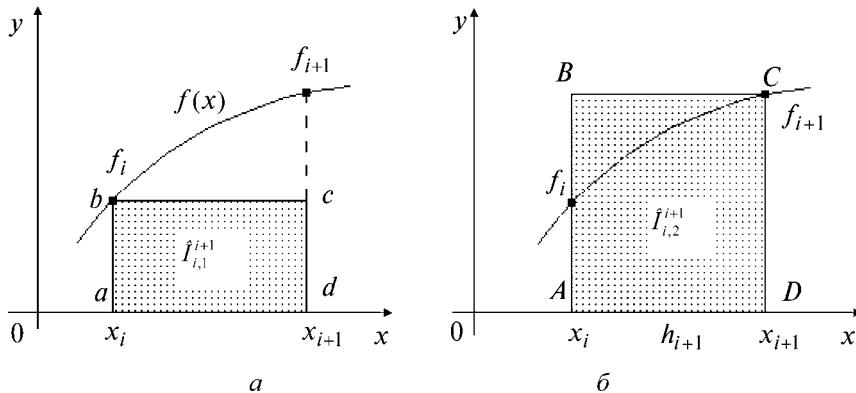


Рис. 5.4

характер аппроксимации интеграла \hat{I}_i^{i+1} на отрезке $[x_i, x_{i+1}]$ первыми (функциональными) слагаемыми комбинаций (5.75), (5.76). Эти аппроксимации представляют собой площади прямоугольников $abcd$ (рис. 5.4а) и $ABCD$ (рис. 5.4б), построенных по значениям функций $f(x)$ в левой точке x_i и в правой точке x_{i+1} соответственно.

Последующие слагаемые учитывают более высокие по порядку h_{i+1} величины, и они вычисляются по значению h_{i+1} и производным от первого до m -го порядка в точках x_i и x_{i+1} . В соответствии с этим (5.75), (5.76) можно назвать *обобщенными квадратурными формулами прямоугольников*.

Составляя линейную комбинацию (5.75), (5.76) (среднеарифметическое $\hat{I}_{i,1}^{i+1}$ и $\hat{I}_{i,2}^{i+1}$), получаем *обобщенную квадратурную формулу трапеций*:

$$\begin{aligned}\hat{I}_{i,3}^{i+1} &= h_{i+1} \frac{f_i + f_{i+1}}{2} - \frac{h_{i+1}^2}{4} (f'_{i+1} - f'_i) + \frac{h_{i+1}^3}{12} (f''_{i+1} + f''_i) - \frac{h_{i+1}^4}{48} (f'''_{i+1} - f'''_i) + \dots = \\ &= \frac{h_{i+1}}{2} \sum_{k=0}^m \frac{h_{i+1}^k}{(k+1)!} \cdot \left[(-1)^k f_{i+1}^{(k)} + f_i^{(k)} \right] \left(\frac{h^t}{t!} M_{t,i} \right),\end{aligned}\quad (5.77)$$

где $t = m + 3$ при четном m и $m = 0$; $t = m + 2$ при нечетном m . Первое слагаемое данной квадратурной формулы соответствует площади трапеции, основания которой равны f_i и f_{i+1} , а высотой является h_{i+1} (см. рис. 5.2б). Из выражения для $\hat{I}_{i,3}^{i+1}$ видно, что порядок аппроксимации интеграла этим слагаемым квадратурной формулы равен трем, в то время как для (5.75), (5.76) этот порядок равен двум. Это обусловлено знаками производных в комбинациях слагаемых квадратурной формулы (5.77) (при $m = 1$ во втором слагаемом производные f'_i и f'_{i+1} вычитаются, и если это слагаемое остаточное, то оно будет равно нулю, так как в этом случае обе производные записываются в одной и той же точке $\xi \in (x_i, x_{i+1})$).

Аналогичным путем на трехточечном шаблоне $\Pi_{3,i} = (x_{i-1}, x_i, x_{i+1})$ можно получить *двухинтервальную квадратурную формулу Эйлера*:

$$\hat{I}_{i-1}^{i+1} = \sum_{k=0}^m \frac{1}{(k+1)!} \cdot \left[h_{i+1}^{k+1} + (-1)^k h_i^{k+1} \right] \cdot f_i^{(k)} \quad \left(\frac{2H^{m+2}}{(m+2)!} \cdot M_{m+1,i} \right), \quad (5.78)$$

где $H = \max(h_i, h_{i+1})$.

Замечание. На основе рассмотренных здесь квадратурных формул дифференциального типа можно строить вложенные алгоритмы численного интегрирования с заданной точностью, состоящие в том, что на фиксированном шаге очередное слагаемое суммы вычисляется и добавляется к ранее вычисленному результату до тех пор, пока абсолютная величина этого слагаемого не станет меньше заданной величины $\varepsilon > 0$.

Пример 5.12. Применяя вложенный алгоритм, вычислить интеграл $I = \int_0^{\frac{\pi}{2}} \cos x dx$ с точностью $\varepsilon = 0,01$ на основе обобщенной квадратурной формулы трапеций (5.77). Результат сравнить с его точным значением.

□ Для функции $f(x)$ с ограниченными производными квадратурная формула (5.77) содержит слагаемые, которые при $h_{i+1} < 1$ уменьшаются по абсолютной величине при переходе от текущего слагаемого к последующему. Поэтому достигнутая точность может контролироваться в процессе вычислений по величине очередного слагаемого (по модулю). Как только это слагаемое по модулю становится меньше заданной точности ε , то вычисления можно закончить. Проведем вычисления слагаемых: 1-, 2-, 3-го и т. д. для заданной в примере функции $f(x) = \cos x$. При этом принимается $x_i = 0$, $x_{i+1} = \frac{\pi}{2}$. В результате имеем

$$h_{i+1} \frac{f_i + f_{i+1}}{2} = \frac{\pi}{2} \frac{\left(\cos 0 + \cos \frac{\pi}{2}\right)}{2} = \frac{\pi}{4} = 0,78537;$$

$$\frac{h_{i+1}^2}{4} (f'_{i+1} - f'_i) = -\frac{\pi^2}{4 \cdot 4} \left(\sin \frac{\pi}{2} - \sin 0\right) = -\frac{\pi^2}{16} = -0,61685;$$

$$\frac{h_{i+1}^3}{12} (f''_{i+1} + f''_i) = -\frac{\pi^3}{8 \cdot 12} \left(\cos \frac{\pi}{2} + \cos 0\right) = -\frac{\pi^3}{96} = -0,32298;$$

$$\frac{h_{i+1}^2}{48} (f'''_{i+1} - f'''_i) = \frac{\pi^4}{16 \cdot 48} \left(\sin \frac{\pi}{2} - \sin 0\right) = \frac{\pi^4}{16 \cdot 48} = 0,126834;$$

$$\frac{h_{i+1}^5}{240} (f^{(4)}_{i+1} + f^{(4)}_i) = \frac{\pi^5}{32 \cdot 240} \left(\cos \frac{\pi}{2} + \cos 0\right) = \frac{\pi^5}{32 \cdot 240} = 0,039846;$$

$$\frac{h_{i+1}^6}{1440} (f^{(5)}_{i+1} - f^{(5)}_i) = \frac{\pi^6}{64 \cdot 1440} \left(-\sin \frac{\pi}{2} + \sin 0\right) = -\frac{\pi^6}{64 \cdot 1440} = -0,01043.$$

Следующее седьмое слагаемое будет по модулю меньше заданной точности $\varepsilon = 0,01$, поэтому его и последующие слагаемые можно не учитывать. Сумма полученных слагаемых с учетом их знаков и знаков линейной комбинации (5.77) дает результат:

$$I_0^{\frac{\pi}{2}} = 0,78537 + 0,61685 - 0,32298 - 0,126834 + \\ + 0,039846 + 0,01043 = 1,002682.$$

Отличие полученного значения от точного не превышает 0,3%. Отметим, что для данной функции в качестве отрезка $[x_i, x_{i+1}]$ выбран целиком весь отрезок $[a, b] = \left[0; \frac{\pi}{2}\right]$. Для достижения заданной точности в примере 5.12 потребовалось вычислить шесть слагаемых. Если отрезок $[a, b]$ разбить на несколько частичных отрезков, например на два или три, то на каждом шаге потребовалось бы меньшее количество слагаемых и в результате общее их количество могло бы уменьшиться. В связи с этим в процессе реализации вычислительного алгоритма возникает задача выбора оптимальных шагов, которые для достижения той же точности обеспечивают минимальное общее количество слагаемых и в силу этого минимальное количество операций. ■

5.4. МЕТОД РУНГЕ УТОЧНЕНИЯ РЕЗУЛЬТАТОВ ЧИСЛЕННОГО ДИФФЕРЕНЦИРОВАНИЯ И ИНТЕГРИРОВАНИЯ

Этот метод является апостериорным, и он основан на получении поправки, выражаемой через два приближенных результата z_{hk} и z_h , получающихся на двух равномерных сетках с шагами h и kh , где k — коэффициент, характеризующий уменьшение шага ($k < 1$) или его увеличение ($k > 1$) при переходе с одной сетки на другую. Наиболее распространен способ генерации поправки в итерационном процессе с $k = \frac{1}{2}$ при необходимости снижения погрешности и $k = 2$ при возможном ее увеличении.

Общая формула уточненного результата, получающегося на двух сетках с h и kh , записывается в виде

$$z = z_{hk} + \frac{z_{hk} - z_h}{\left(\frac{1}{k}\right)^p - 1}, \quad (5.79)$$

где p — порядок точности метода, по которому рассчитаны z_{hk} и z_h . Второе слагаемое в правой части (5.79) представляет собой *поправку*, или *апостериорную оценку точности результата*, получаемого на сетке с параметром kh . Если заданная точность не достигается, то при $k < 1$ процесс продолжается путем дробления шага и перехода на новую, более мелкую сетку. Если же указанное слагаемое намного меньше заданной точности, то в алгоритме необходимо перейти на укрупненную сетку, полагая, например, $k = 2$.

Данную методику рассмотрим применительно к расчету производных на основе сгущающихся сеток ($k = \frac{1}{2}$), когда для первоначального вычисления производных принимается одна из формул (5.10). Схема уточнения приведена в таблице 5.5, где численные значения производных обозначены символом Z с двумя нижними индексами, первый из которых ($m \geq 1$) указывает на ступени уточнения ($m = 1$ — первая, $m = 2$ — вторая и т. д.), а второй — k — на порядковые номера значений производных, полученных при различных шагах на каждой ступени уточнения. Начиная с третьей колонки таблицы помещаются значения $Z_{0,0}, Z_{0,1}, \dots$, полученные по некоторой исходной аппроксимационной формуле при различных шагах разбиения сетки: $h_0, h_1 = \frac{h_0}{2}, h_2 = \frac{h_1}{2}, \dots, h_{k+1} = \frac{h_k}{2}, \dots$ В последующих колонках таблицы приводятся значения производных, рассчитанных по формуле (5.79), которая может быть записана в виде

$$Z_{m,k} = Z_{m-1,k+1} + \frac{Z_{m-1,k+1} - Z_{m-1,k}}{2^{2m} - 1}, \quad (5.80)$$

где значение $2m$ соответствует порядку аппроксимации p .

Порядок уточнения численных результатов.

Пусть по заданной сеточной функции $y_i = f(x_i)$ $i = \overline{0, n}$, в некоторой фиксированной точке $x_j = x_i$ производная вначале рассчитывается по формуле (5.10) второго порядка аппроксимации

$$Z_{0,k} = \frac{f(x_i + h_k) - f(x_i - h_k)}{2h_k} \quad (O(h^2)). \quad (5.81)$$

Тогда уточнение до порядка $O(h^6)$ производится следующим образом.

1. По формуле (5.81) рассчитываются значения $Z_{0,0}, Z_{0,1}, Z_{0,2}$ производных в заданной точке с использованием таблицы значений исходной сеточной функции.

При этом для вычисления $Z_{0,k}$, $k = 0, 1, 2$, берутся значения $f(x_i)$ в тех точках сетки, которые соответствуют шагам h_0, h_1 и h_2 . Все величины, входящие в третью колонку таблицы, имеют порядок аппроксимации $O(h^2)$, что указано в нижней строке таблицы 5.5.

Таблица 5.5

k	h_k	$Z_{0,k}$	$Z_{1,k}$	$Z_{2,k}$	$Z_{3,k}$
0	h_0	$Z_{0,0}$	$Z_{1,0}$	$Z_{2,0}$	$Z_{3,0}$
1	h_1	$Z_{0,1}$	$Z_{1,1}$	$Z_{2,1}$	
2	h_2	$Z_{0,2}$	$Z_{1,2}$		
3	h_3	$Z_{0,3}$			
		$O(h^2)$	$O(h^4)$	$O(h^6)$	$O(h^8)$

2. Реализуется первая ступень процесса уточнения по формуле (5.80), в которой следует принять $m = 1$ (знаменатель второго слагаемого формулы (5.80) равен 3), а $p = 2m = 2$. При этом вычисляются $Z_{1,0}$ и $Z_{1,1}$. Порядок аппроксимации результатов, получаемых в первой ступени уточнения, равен четырем.

3. Реализуется вторая ступень процесса уточнения по формуле (5.80), в которой следует принять $m = 2$ (знаменатель второго слагаемого формулы (5.80) равен 15), а $p = 2m = 4$. При этом вычисляется окончательное значение $Z_{2,0}$, порядок аппроксимации которого равен шести. Процесс уточнения до заданного порядка $p = 6$ закончен.

Замечание. В случае, когда необходимо обеспечить порядок аппроксимации, превышающий полученный, например восьмой, производятся дополнительные расчеты значений $Z_{0,k}$ при более мелких шагах и расчеты $Z_{1,2}, Z_{2,1}$ по первой и второй ступени уточнения. После этого рассчитывается $Z_{3,0}$ с порядком $O(h^8)$ (табл. 5.5). Данный процесс может быть продолжен и далее.

Пример 5.13. Для сеточного представления функции $y = \sin x$, $x \in [0; 2\pi/5]$, заданной на сетке с шагом $h = \pi/10$, на основе применения формулы (5.10) второго порядка и метода Рунге найти первую производную $f'(x)|_{x=\frac{\pi}{5}}$ с порядком $O(h^4)$.

□1. По формуле (5.81) рассчитываются значения $Z_{0,0}$ и $Z_{0,1}$ производных в точке $x_2 = \frac{\pi}{5}$. При этом $Z_{0,0}$ получается при $h_0 = \frac{2\pi}{10}$, а значение $Z_{0,1}$ при $h_1 = \frac{\pi}{10}$. Записывая формулу (5.81) для указанных шагов, получаем

$$f'_{2,h} = Z_{0,0} = \frac{f_4 - f_0}{2h_0} = \frac{0,951057 - 0}{2 \cdot \frac{\pi}{5}} = 0,756827;$$

$$f'_{\frac{2,h}{2}} = Z_{0,1} = \frac{f_3 - f_1}{2h_1} = \frac{0,809017 - 0,309017}{2 \cdot \frac{\pi}{10}} = 0,795775.$$

Исходная сеточная функция и результаты расчетов производной в точке x_2 помещены в таблице 5.6.

Таблица 5.6

i	0	1	2	3	4
x_i	0	$\frac{\pi}{10}$	$\frac{2\pi}{10}$	$\frac{3\pi}{10}$	$\frac{4\pi}{10}$
f_i	0	0,309017	0,587785	0,809017	0,951057
$f'_{2,h} = Z_{0,0}$			0,756827		
$f'_{\frac{2,h}{2}} = Z_{0,1}$			0,795775		

Стрелочками от значений производных показаны значения функции, использованные при расчете.

2. Реализуется первая (в данном случае и последняя) ступень процесса уточнения по формуле (5.80), в которой $m = 1$. При этом принимается $f'_{\frac{2,h}{2}} = Z_{0,1}$, $f'_{2,h} = Z_{0,0}$. Тогда имеем

$$Z_{1,0} \equiv f'_{2,\text{уточн}} = Z_{0,1} + \frac{Z_{0,1} - Z_{0,0}}{2^2 - 1} = 0,795775 + \frac{0,795775 - 0,756827}{3} = 0,808758.$$

Полученный результат соответствует порядку аппроксимации $O(h^4)$. ■

Замечание. Для уточнения результатов численного интегрирования метод Рунге реализуется аналогично.

Задачи для самостоятельного решения

1. Вычислить значения интеграла $I_0^2 = \int_0^2 f(x)dx$ по формулам трапеций и парабол при $h = 1$, сравнить полученные результаты с точным значением. В качестве подынтегральных функций взять:

- а) $f(x) = 1 + x$;
- б) $f(x) = 1 + x^2$;
- в) $f(x) = 1 + x^3$;
- г) $f(x) = 1 + x^4$.

2. Вычислить значение интеграла $I_{-1}^1 = \int_{-1}^1 (1 + x + x^2 + x^3 + x^4) dx$ методами прямоугольников, трапеций, парабол при $h = 1$.

3. Найти значения интегралов в задачах 1 и 2 с точностью $\varepsilon = 0,01$, используя методику априорного выбора шага. Вычислить фактические погрешности.

4. Найти приближенное значение интеграла $I_0^1 = \int_0^1 \sin x dx = 0,45970$ методом парабол при $h = 0,5$. Пояснить, почему получается значение, близкое к точному.

5. Найти приближенное значение интеграла $I_{0,1}^1 = \int_{0,1}^1 \frac{dx}{x} = 2,30259$ методом парабол при $h = 0,45; h = 0,225; h = 0,1125$. Пояснить, почему при первом значении шага получается значительная ошибка. Уточнить результат по методу Рунге до порядка $O(h^8)$.

6. Найти приближенное значение интеграла $I_0^{10} = \int_0^{10} e^{-x} dx = 0,999955$ методами прямоугольников, трапеций, парабол при $h = 1$. Сделать анализ полученных результатов.

7. Найти приближенное значение интеграла $I_{-1}^3 = \int_{-1}^3 \frac{dx}{2+x}$ методами прямоугольников, трапеций, парабол при $h = 1$. Оценить погрешности методов.

8. Найти значения первой и второй производной функции, заданной в таблице 5.7, в точках $x = 0; x = 0,1; x = 0,25$. Использовать формулы, соответствующие двух-, трех-, четырех- и пятиточечным шаблонам.

Таблица 5.7

x_i	0,00	0,05	0,10	0,15	0,20	0,25
y_i	0,00000	0,10017	0,20134	0,30452	0,41075	0,52110

ЧАСТЬ ТРЕТЬЯ

ЧИСЛЕННЫЕ
МЕТОДЫ РЕШЕНИЯ
ОБЫКНОВЕННЫХ
ДИФФЕРЕНЦИАЛЬНЫХ
УРАВНЕНИЙ

ГЛАВА 6

МЕТОДЫ РЕШЕНИЯ ЗАДАЧИ КОШИ

6.1. ПОСТАНОВКА ЗАДАЧИ И ОСНОВНЫЕ ПОЛОЖЕНИЯ

Обыкновенным дифференциальным уравнением называется уравнение, связывающее независимую переменную x , неизвестную функцию $y(x)$ этой независимой переменной и ее производные $y'(x), y''(x), \dots, y^{(n)}(x)$:

$$F(x, y(x), y'(x), \dots, y^{(n)}(x)) = 0, \quad (6.1)$$

где $F(x, y, y', \dots, y^{(n)})$ — функция указанных аргументов, заданная в некоторой области их изменения.

Порядком дифференциального уравнения называется порядок старшей производной, входящей в уравнение.

Если в соотношении (6.1) функция F такова, что его можно представить в виде

$$y^{(n)} = f(x, y(x), \dots, y^{(n-1)}(x)), \quad (6.2)$$

то уравнение называется обыкновенным дифференциальным уравнением n -го порядка, разрешенным относительно старшей производной.

Уравнение называется линейным, если функция F линейна относительно искомой функции и ее производных, т. е. если уравнение может быть записано в виде

$$a_n(x)y^{(n)} + a_{n-1}(x)y^{(n-1)} + \dots + a_0(x)y = f(x), \quad (6.3)$$

где $a_n(x), a_{n-1}(x), \dots, a_0(x), f(x)$ — известные в общем случае нелинейные функции x .

Решением дифференциального уравнения n -го порядка называется функция $y(x)$, непрерывная на некотором интервале (a, b) вместе со своими производными до $(n - 1)$ -го порядка включительно, имеющая производную $y^{(n)}(x)$ и такая, что подстановка $y(x)$ в уравнение обращает его в тождество.

График решения дифференциального уравнения называется интегральной кривой.

Одной из важнейших задач в теории и приложениях дифференциальных уравнений является задача Коши (начальная задача), в которой требуется

найти решение дифференциального уравнения, удовлетворяющее заданным начальными условиям. Для уравнения (6.2) она записывается следующим образом:

$$\begin{aligned} y^{(n)} &= f(x, y(x), \dots, y^{(n-1)}(x)); \\ y(x_0) &= y_0, y'(x_0) = y'_0, \dots, y^{(n-1)}(x_0) = y_0^{(n-1)}, \end{aligned} \quad (6.4)$$

где $x_0 \in (a, b)$, $y_0, y'_0, \dots, y_0^{(n-1)}$ — заданные числа.

Теорема 6.1 (о существовании и единственности решения задачи Коши (6.4)).

Пусть выполнены следующие условия:

- а) функция $f(x, y, \dots, y^{(n-1)})$ определена и непрерывна в некоторой замкнутой области \bar{D} , а также имеет в \bar{D} ограниченные частные производные по переменным $y, y', \dots, y^{(n-1)}$;
- б) точка $(x_0, y_0, y'_0, \dots, y_0^{(n-1)})$ лежит внутри области \bar{D} .

Тогда решение задачи Коши (6.4) существует и единствено.

Общим решением дифференциального уравнения n -го порядка в области $G \subset \bar{D}$ (\bar{D} — область, в которой выполнены условия теоремы 6.1) называется функция $y = y(x, C_1, \dots, C_n)$, зависящая от n произвольных постоянных, и такая, что при подстановке в уравнение она обращает его в тождество при любых значениях C_1, \dots, C_n . Геометрически общее решение в области G представляет собой семейство непересекающихся интегральных кривых, полностью покрывающих всю область.

Общим интегралом дифференциального уравнения называется соотношение вида

$$\varphi(x, y, C_1, \dots, C_n) = 0,$$

неявно определяющее общее решение.

При конкретных значениях C_1, \dots, C_n , включая $\pm\infty$, из общего решения выделяется *частное решение*, а общий интеграл становится *частным интегралом*. В каждой точке (x, y) частного решения или частного интеграла выполняются условия теоремы 6.1.

Наряду с проблемой решения дифференциальных уравнений n -го порядка на практике возникает проблема решения *систем обыкновенных дифференциальных уравнений первого порядка*, связывающих независимую переменную x , неизвестные функции $y_1(x), \dots, y_n(x)$ и их производные $y'_1(x), \dots, y'_n(x)$.

В случае, если уравнения разрешимы относительно производных, систему можно записать в *нормальной форме Коши*:

$$\begin{aligned} \frac{dy_1}{dx} &= f_1(x, y_1, \dots, y_n), \\ \frac{dy_2}{dx} &= f_2(x, y_1, \dots, y_n), \\ &\dots \\ \frac{dy_n}{dx} &= f_n(x, y_1, \dots, y_n), \end{aligned} \quad (6.5)$$

где $f_i(x, y_1, \dots, y_n)$, $i = \overline{1, n}$, — известные функции.

Решением системы (6.5) называется совокупность n функций $y_1(x), \dots, y_n(x)$, непрерывных на некотором интервале (a, b) и имеющих производные $y'_1(x), \dots, y'_n(x)$, такая, что подстановка этих функций в (6.5) обращает все уравнения в тождества.

Задача Коши для системы (6.5) состоит в нахождении решения системы, удовлетворяющего начальным условиям:

$$y_1(x_0) = y_{10}, \quad y_2(x_0) = y_{20}, \dots, \quad y_n(x_0) = y_{n0}, \quad (6.6)$$

где $y_{10}, y_{20}, \dots, y_{n0}$ — известные числа.

В векторной форме задача Коши (6.5), (6.6) имеет вид

$$Y' = F(x, Y), \quad Y(x_0) = Y_0, \quad (6.7)$$

где $Y = (y_1, \dots, y_n)^T$, $F(x, Y) = (f_1(x, Y), \dots, f_n(x, Y))^T$, $Y_0 = (y_{10}, \dots, y_{n0})^T$.

Теорема 6.2 (о существовании и единственности решения задачи Коши (6.5), (6.6)).

Пусть выполнены следующие условия:

a) функции $f_i(x, y_1, \dots, y_n)$, $i = \overline{1, n}$, определены и непрерывны в некоторой замкнутой области \bar{D} , а также имеют в \bar{D} ограниченные частные производные по переменным y_1, \dots, y_n ;

б) точка $(x_0, y_{10}, y_{20}, \dots, y_{n0})$ лежит внутри области \bar{D} .

Тогда решение задачи Коши (6.5), (6.6) существует и единственно.

Замечания

1. Во многих практических приложениях независимая переменная обозначается через t и имеет смысл времени [33], поэтому задача Коши называется начальной задачей.

2. Понятия общего и частного решений, общего и частного интегралов для уравнения первого порядка и систем совпадают по форме, если заменить функцию $y(x)$ на вектор-функцию $Y(x)$, $f(x, y)$ на $F(x, Y)$, а y_0 — на Y_0 .

Численные методы, рассматриваемые в данном пункте, пригодны для решения задач Коши, записанных в форме (6.5), (6.6). Чтобы решить задачу Коши (6.4) этими методами, ее необходимо привести к системе n уравнений первого порядка, т. е. к виду (6.5), (6.6).

Обозначая $y_1(x) = y(x)$, $y_2(x) = y'(x)$, ..., $y_n(x) = y^{(n-1)}(x)$, получаем

$$\begin{aligned} \frac{dy_1}{dx} &= y_2, & y_1(x_0) &= y_0, \\ \frac{dy_2}{dx} &= y_3, & y_2(x_0) &= y'_0, \\ \dots & & & \\ \frac{dy_n}{dx} &= f(x, y_1, \dots, y_n), & y_n(x_0) &= y_0^{(n-1)}. \end{aligned} \tag{6.8}$$

Получим точное решение модельного примера, используемого далее для демонстрации применения различных численных методов.

Пример 6.1. Найти аналитическое решение задачи Коши

$$Ty' + y = 1, \quad y(0) = 0,$$

где $T > 0$ — известное число, называемое *постоянной времени*.

□Решение задачи Коши найдем с помощью известной методики [33].

1. Определим общее решение однородного уравнения $Ty' + y = 0$. Поскольку корень $\lambda = -\frac{1}{T}$ соответствующего характеристического уравнения $T\lambda + 1 = 0$ действительный, то $y_0(x) = Ce^{\lambda x} = Ce^{-\frac{1}{T}x}$ — общее решение однородного уравнения.

2. Частное решение неоднородного уравнения ищется в форме $y_q(x) = A$, где $A = \text{const}$. После подстановки в решаемое уравнение получаем $y_q(x) = 1$.

3. Общее решение неоднородного уравнения получается как сумма общего решения однородного уравнения и частного решения неоднородного уравнения:

$$y(x) = y_0(x) + y_q(x) = Ce^{-\frac{1}{T}x} + 1.$$

Ему соответствует семейство интегральных кривых, характеризующееся параметром C , который может принимать произвольные значения.

4. Частное решение неоднородного уравнения находим из начального условия: $y(0) = C + 1 = 0$. Отсюда $C = -1$ и $y(x) = 1 - e^{-\frac{1}{T}x}$. ■

Пример 6.2. Записать дифференциальное уравнение второго порядка

$$\begin{aligned}2y'' + y' + 4y &= 6\sin x, \\y(0) = 1, \quad y'(0) &= 2\end{aligned}$$

в виде системы дифференциальных уравнений в нормальной форме Коши.

□Введем обозначения: $y_1(x) = y(x)$, $y_2(x) = y'_1(x)$ и запишем уравнение в форме $y'' = -\frac{1}{2}y' - 2y + 3\sin x$, разрешенной относительно старшей производной. Тогда получим

$$\begin{aligned}y'_1 &= y_2, \quad y_1(0) = 1, \\y'_2 &= -\frac{1}{2}y_2 - 2y_1 + 3\sin x, \quad y_2(0) = 2.\end{aligned}\blacksquare$$

Численные и приближенно-аналитические методы решения задачи Коши в отличие от аналитических методов позволяют найти искомую функцию $y(x)$ лишь приближенно. Но при этом численные методы являются более универсальными, так как с их помощью можно приближенно решать многие из задач, точные решения которых аналитическими методами не могут быть найдены. Аналитическими методами в основном решаются только линейные задачи и некоторые типы нелинейных задач, в то время как для численных методов эти ограничения отсутствуют.

Чтобы упростить изложение и в силу того, что численные методы легко обобщаются на системы уравнений, в дальнейшем будем рассматривать решение задачи Коши для уравнения первого порядка

$$y' = f(x, y), \quad y(x_0) = y_0, \quad x \in (a, b). \quad (6.9)$$

Численное решение задачи ищется в узлах сетки $\Omega_n = \{x_0, x_1, \dots, x_n\}$, где $h_{i+1} = x_{i+1} - x_i$, $i = \overline{0, n-1}$, — расстояние между соседними узлами, называемое шагом интегрирования (параметром сетки). Если $h_{i+1} = h = \text{const}$, сетка

называется *равномерной* (*регулярной*), а если $h_{i+1} = \text{var}$ — *неравномерной* (*нерегулярной*). В случае равномерной сетки узлы находятся по формуле

$$x_i = x_0 + ih, \quad i = \overline{0, n},$$

а в случае неравномерной:

$$\begin{aligned} x_1 &= x_0 + h_1, \quad x_2 = x_1 + h_2 = x_1 + \delta_2 h_1, \dots, x_{i+1} = x_i + \delta_{i+1} h_i, \dots, \\ x_n &= x_{n-1} + \delta_n h_{n-1}, \end{aligned}$$

где $\delta_{i+1} = \frac{h_{i+1}}{h_i}$ — *параметр нерегулярности*.

Решение находится в виде последовательности значений $\hat{y}_0, \hat{y}_1, \hat{y}_2, \dots, \hat{y}_n$, являющихся приближением значений $y_0, y(x_1), y(x_2), \dots, y(x_n)$ точного решения $y(x)$ в узлах сетки Ω_n (рис. 6.1).

Сеточное представление $y(x_i)$, $i = \overline{0, n}$, известной функции $y(x)$ (точного решения задачи Коши) называется *проекцией* $y(x)$ на сетку Ω_n .

Численные методы решения обыкновенных дифференциальных уравнений делятся на две группы:

1) *дискретные методы*, позволяющие найти решение только в узлах сетки. Эти методы называются еще *разностными методами*, или *методами сеток*;

2) *непрерывно-дискретные методы*, основанные на использовании дискретных методов и сплайн-функций для восполнения численных результатов. Они позволяют найти непрерывные решения дифференциальных уравнений.

Дискретные методы (методы сеток) подразделяются на *явные* и *неявные*.

Значение \hat{y}_{i+1} на $(i+1)$ -м шаге может определяться *явно*:

$$\hat{y}_{i+1} = \Phi(x_{i-k+1}, \dots, x_{i-1}, x_i, \hat{y}_{i-k+1}, \dots, \hat{y}_{i-1}, \hat{y}_i), \quad (6.10)$$

где $\Phi(\cdot)$ — некоторая функция, зависящая от конкретного метода (кроме последней рассчитанной точки (x_i, \hat{y}_i) могут использоваться еще $(k-1)$ предыдущих точек), или *неявно*:

$$\hat{y}_{i+1} = \Phi(x_{i-k+1}, \dots, x_{i-1}, x_i, x_{i+1}, \hat{y}_{i-k+1}, \dots, \hat{y}_{i-1}, \hat{y}_i, \hat{y}_{i+1}), \quad (6.11)$$

где искомая величина \hat{y}_{i+1} входит одновременно в левую, и в правую часть.

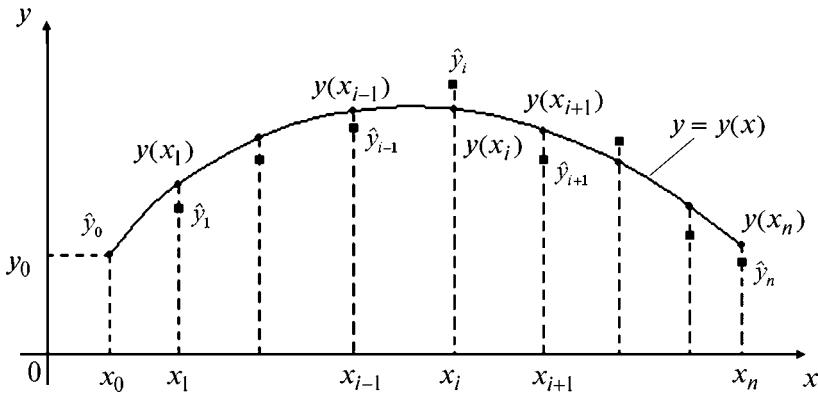


Рис. 6.1

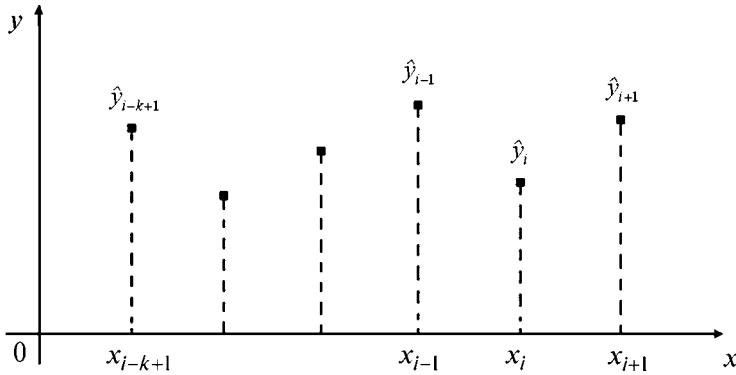


Рис. 6.2

Явные и неявные методы делятся также на *одношаговые* и *многошаговые* (k -шаговые). В *одношаговых* методах для расчета очередной точки (x_{i+1}, \hat{y}_{i+1}) требуется информация только о последней рассчитанной точке (x_i, \hat{y}_i) . В *k -шаговых* методах для нахождения точки (x_{i+1}, \hat{y}_{i+1}) требуется информация о k предыдущих точках (рис. 6.2).

Формулы (6.10), (6.11) в общем случае представляют собой нелинейные уравнения относительно \hat{y}_{i+1} и называются *разностными схемами*.

Численный алгоритм (метод) называется *устойчивым*, если численные результаты непрерывно зависят от входных данных и если погрешность остается ограниченной при заданных пределах изменения параметров численного алгоритма (шагов сетки, числа итераций и др.).

Сходимость приближенных методов является основной проблемой, от успешного преодоления которой зависит точность решения всей задачи. Численный алгоритм называется *сходящимся*, если при стремлении его параметров к определенным предельным значениям, например, при $h \rightarrow 0$ (или при $s \rightarrow \infty$, где s — число итераций), результаты стремятся к точному решению.

При известном точном решении некоторой модельной задачи сходимость может быть проверена следующим образом. Фиксируется некоторая точка $x > x_0$ и строится последовательность сеток Ω_n , таких что $h \rightarrow 0$, $x = x_n = x_0 + nh$. Здесь для простоты считаем, что все сетки, образующие указанную последовательность, являются равномерными. Тогда, если $|\hat{y}_n - y(x_n)| \rightarrow 0$ при $h \rightarrow 0$ ($n \rightarrow \infty$), то метод является *сходящимся* в точке x . Если метод сходится в каждой точке $x \in [c, d] \subset (a, b)$, то он сходящийся на отрезке $[c, d]$.

Локальной ошибкой численного метода на $(i+1)$ -м шаге называется величина

$$\varepsilon_{i+1}(h) = \hat{y}_{i+1} - y(x_{i+1}),$$

где $y(x_{i+1})$ — значение точного решения при $x = x_{i+1}$, а \hat{y}_{i+1} — приближенное решение, получаемое по формулам (6.10) или (6.11) при условии, что вместо приближенных значений $\hat{y}_i, \hat{y}_{i-1}, \dots, \hat{y}_{i-k+1}$ используются значения, соответствующие точному решению, т. е. $y(x_i), y(x_{i-1}), \dots, y(x_{i-k+1})$.

Глобальной ошибкой называется величина $e_n(h) = \hat{y}_n - y(x_n)$, где \hat{y}_n — значение, получаемое по формулам (6.10) или (6.11) при $i = n - 1$.

Глобальная ошибка определяется:

- а) ошибками округления и ошибками арифметических действий, обусловленными числом разрядов компьютера и характером выполняемых операций для расчета значения искомой функции в очередной точке x_{i+1} ;
- б) методическими ошибками, определяемыми выбранным алгоритмом;
- в) переходными ошибками, обусловленными тем, что при расчете значения \hat{y}_{i+1} вместо точных значений $y(x_i), y(x_{i-1}), \dots, y(x_{i-k+1})$ берутся приближенные значения $\hat{y}_i, \hat{y}_{i-1}, \dots, \hat{y}_{i-k+1}$, полученные на предыдущих шагах.

Локальные ошибки «переносятся» в точку x_n и формируют глобальную ошибку.

Число p называется *порядком (точностью) численного метода*, если его глобальная ошибка есть O большое от h^p , т. е. $e_n(h) = O(h^p)$.

На практике в качестве характеристики точности метода часто используется величина $\varepsilon(h) = \max_{i=0,1,\dots,n} |\hat{y}_i - y(x_i)|$.

Рассмотрим введенные понятия более подробно на примере явных одноступенчатых методов, построенных для задачи (6.9). При этом формулу (6.10) представим в виде

$$\hat{y}_{i+1} = \hat{y}_i + h \cdot \Psi(\hat{y}_i, x_i, h),$$

где $\Psi(\hat{y}_i, x_i, h)$ — некоторая функция, определяемая конструкцией того или иного метода.

Обозначим $y(x, x_i, \hat{y}_i)$ — решение задачи Коши

$$u' = f(x, u), \quad u(x_i) = \hat{y}_i.$$

Тогда локальная ошибка определяется выражением

$$\varepsilon_{i+1}(h) = \hat{y}_{i+1} + h \cdot \Psi(\hat{y}_i, x_i, h) - y(x_{i+1}, x_i, \hat{y}_i).$$

Геометрическая интерпретация возникновения локальных и глобальной ошибок изображена на рисунке 6.3.

Можно показать [53]–[55], что если локальная ошибка имеет порядок $(p+1)$, т. е. $\varepsilon_{i+1}(h) = O(h^{p+1})$, то глобальная погрешность имеет на единицу меньший порядок, т. е. $e_n(h) = O(h^p)$.

Перейдем теперь к рассмотрению устойчивости численных методов. Она проверяется на «тестовом примере»

$$y' = \mu y, \quad y(0) = 1, \tag{6.12}$$

где μ — в общем случае комплексная константа. Дифференциальное уравнение в (6.12) является простейшим линейным уравнением, и для него можно получить значимые критерии устойчивости в явной форме.

Метод называется *устойчивым (ограниченно устойчивым)*, если существует такое число $h_{kp} > 0$, что при использовании метода для решения задачи (6.12), где $\operatorname{Re}\mu < 0$ с шагом $0 < h < h_{kp}$ при $i \rightarrow \infty$ глобальная ошибка ограничена. Величина h_{kp} называется *критическим шагом*. Если $h > h_{kp}$, глобальная ошибка может неограниченно возрастать.

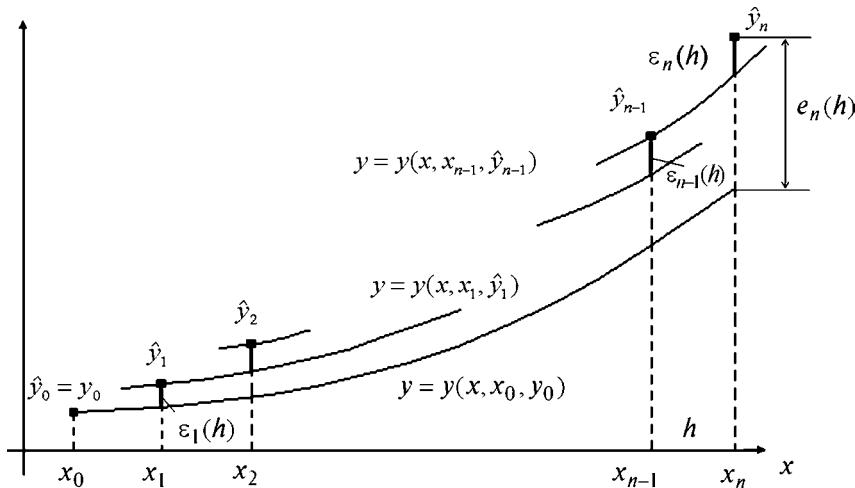


Рис. 6.3

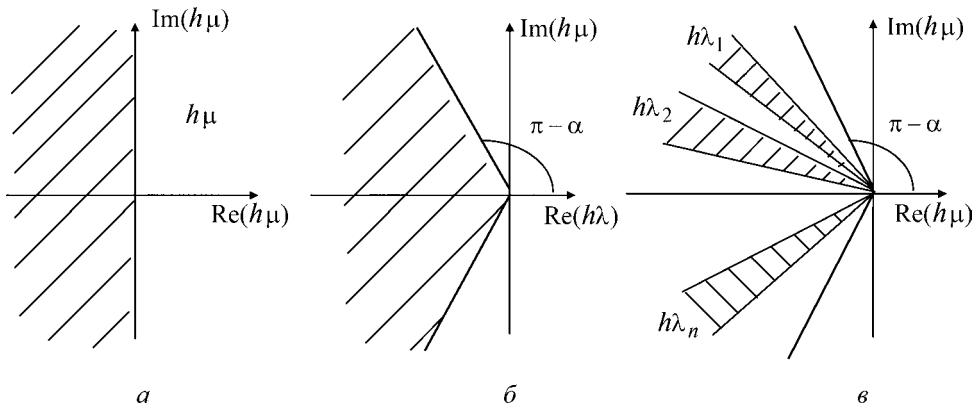


Рис. 6.4

В ограниченно устойчивых методах при задании величины шага h необходимо учитывать значение критического шага h_{kp} . Для сложных дифференциальных уравнений и систем нахождение h_{kp} является самостоятельной задачей, а свойство ограниченной устойчивости предупреждает вычислителя о возможных проблемах. Поэтому на практике становится актуальной задача конструирования таких методов, которые были бы устойчивы при любом значении шага, а его величина выбиралась бы только исходя из желаемой точности расчетов (при этом класс решаемых задач может быть ограничен).

Метод называется *A-устойчивым*, если при его применении с любым фиксированным положительным шагом h все численные решения задачи (6.12) с комплексной константой μ ($\operatorname{Re}\mu < 0$) стремятся к нулю при $i \rightarrow \infty$.

Область *A-устойчивости* — совокупность значений h и μ , удовлетворяющих условию $\operatorname{Re}(h\mu) < 0$. Она изображена на рисунке 6.4а. Выполнение свойства *A-устойчивости* является желательным, поскольку если решение задачи

(6.12) асимптотически устойчиво (в силу условия $\operatorname{Re}\mu < 0$ корень характеристического уравнения находится в левой полуплоскости), то погрешность численного решения стремится к нулю при любой величине шага $h > 0$.

При исследовании устойчивости численного метода необходимо использовать соответствующую ему разностную схему для решения задачи (6.12) и привести ее к линейному разностному уравнению с постоянными коэффициентами:

$$a_n y(i+n) + a_{n-1} y(i+n-1) + \dots + a_0 y(i) = g(i), \quad i = 0, 1, \dots$$

Известно, что критерием устойчивости решения линейного разностного уравнения является требование расположения корней λ_i соответствующего характеристического уравнения $a_n \lambda^n + a_{n-1} \lambda^{n-1} + \dots + a_0 = 0$ внутри круга единичного радиуса с центром в начале координат, т. е.

$$|\lambda_i| < 1, \quad i = \overline{1, n}. \quad (6.13)$$

Замечания

1. Численные методы, которые можно представить в виде

$$\sum_{j=0}^k (\alpha_j \cdot \hat{y}_{i+1-j} + h \cdot \beta_j \cdot f_{i+1-j}) = 0, \quad (6.14)$$

где $|\alpha_0| + |\beta_0| \neq 0$, $f_{i+1-j} = f(x_{i+1-j}, \hat{y}_{i+1-j})$, называются линейными k -шаговыми методами [54].

$$\text{Обозначим } \rho(\xi) = \sum_{j=0}^k \alpha_j \cdot \xi^{k-j}, \quad \sigma(\xi) = \sum_{j=0}^k \beta_j \cdot \xi^{k-j}.$$

Линейный многошаговый метод является устойчивым, если для фиксированного значения $h\mu$ корни уравнения

$$\rho(\xi) + h \cdot \mu \cdot \sigma(\xi) = 0 \quad (6.15)$$

лежат внутри круга единичного радиуса с центром в начале координат.

2. Для ограниченно устойчивых методов важной задачей является нахождение величины критического шага h_{kp} . Если константа μ в уравнении (6.12) действительная, то можно найти *интервал устойчивости*.

3. Существуют определения, смягчающие свойство A -устойчивости [53]–[55]. Приведем одно из них.

Метод называется $A(\alpha)$ -устойчивым, $\alpha \in (0, \frac{\pi}{2})$, если при его применении все численные решения уравнения (6.12) с фиксированным положительным шагом h стремятся к нулю при $i \rightarrow \infty$ для всех μ , удовлетворяющих условию

$$|\arg(-\mu)| < \alpha, \quad |\mu| \neq 0,$$

где $\arg(-\mu)$ — аргумент комплексного числа $(-\mu)$ [32]. Область $A(\alpha)$ -устойчивости показана на рисунке 6.4б.

Это условие применимо и для линейных систем с постоянными коэффициентами $y' = Ay$, где A — матрица коэффициентов, имеющая собственные значения λ_i , $i = \overline{1, n}$. Геометрическая интерпретация изображена на рисунке 6.4в.

4. Можно показать, что явные линейные многошаговые методы не могут быть A -устойчивыми.

Рассмотрим основные принципы конструирования разностных схем.

1. Производная $\left(\frac{dy}{dx}\right)_{x=x_i}$ аппроксимируется на заранее выбранном шаблоне $(x_{i-k+1}, x_{i-k+2}, \dots, x_{i+1})$, а правая часть $f(x, y)$ обыкновенного дифференциального уравнения заменяется сеточным представлением $f(x_i, \hat{y}_i)$ в точке (x_i, \hat{y}_i) . Условно этот принцип называется *аппроксимационным*. С его помощью строятся одношаговые и многошаговые схемы в зависимости от шаблона, по которому аппроксимируется производная.

2. Дифференциальное уравнение заменяется разностным аналогом на одиношаговом (двуточечном) шаблоне (x_i, x_{i+1}) (возможно использование дополнительных точек внутри шаблона) путем согласования с разложением функции $y = y(x)$ при $x = x_{i+1}$ по формуле Тейлора относительно точки $x = x_i$. По этому принципу строятся одношаговые схемы, в частности семейство схем Рунге — Кутты.

3. Осуществляется интегрирование обеих частей дифференциального уравнения вдоль решения на выбранном шаблоне. При этом правая часть уравнения предварительно заменяется интерполяционным многочленом, степень которого должна соответствовать порядку точности схемы. Данный принцип называется *интегрально-интерполяционным* и используется для построения одношаговых и многошаговых явных и неявных схем, в том числе схем Адамса.

4. Разностные схемы строятся на основе подобия соотношений теории приближения конструкциям схем. С этой целью используются следствия из интегрально-дифференциальных сплайнов — параметрические соотношения, формулы аппроксимации производных, квадратурные формулы, записанные в общем случае на нерегулярном шаблоне. Данный принцип, который можно назвать принципом *аналогий*, предложен в [18]. Далее в п. 6.2.4 он применен для конструирования двух- и трехшаговых явных и неявных схем.

6.2. ПРИНЦИПЫ ПОСТРОЕНИЯ РАЗНОСТНЫХ СХЕМ

6.2.1. ПРИНЦИП АППРОКСИМАЦИЙ

Рассмотрим проблему нахождения численного решения задачи Коши (6.9):

$$\frac{dy}{dx} = f(x, y), \quad y(x_0) = y_0.$$

Для аппроксимации производной $\left(\frac{dy}{dx}\right)_{x=x_i}$ используем формулу (5.4), записанную на двухточечном шаблоне $\Pi_{2,i} = (x_i, x_{i+1})$:

$$\left(\frac{dy}{dx}\right)_{x=x_i} = \frac{y_{i+1} - y_i}{h_{i+1}} + O(h_{i+1}) \left(\frac{h_{i+1}}{2} M_{2,i} \right), \quad (6.16)$$

и формулу (5.10), записанную на трехточечном нерегулярном шаблоне $\Pi_{3,i} = (x_{i-1}, x_i, x_{i+1})$:

$$\begin{aligned} \left. \left(\frac{dy}{dx} \right) \right|_{x=x_i} &= \frac{\Pi_i^{i+1}}{H_i^{i+1}} \left[\frac{\Delta y_{i-1}}{h_i^2} + \frac{\Delta y_i}{h_{i+1}^2} \right] = \\ &= \frac{1}{H_i^{i+1}} \left[-\delta_{i+1} \cdot y_{i-1} + \frac{(\delta_{i+1}^2 - 1)}{\delta_{i+1}} y_i + \frac{1}{\delta_{i+1}} y_{i+1} \right] + O(h_i h_{i+1}) \left(\frac{h_i^2}{6} \delta_{i+1} M_{3,i} \right), \end{aligned} \quad (6.17)$$

где $H_i^{i+1} = h_i + h_{i+1}$, $\Pi_i^{i+1} = h_i h_{i+1}$, $h_{i+1} = x_{i+1} - x_i$, $\delta_{i+1} = \frac{h_{i+1}}{h_i}$ — параметр нерегулярности.

При $\delta_{i+1} = 1$ шаблон становится регулярным, а среднее слагаемое с входящим в него y_i — равным нулю. При этом формула (6.17) сводится к традиционной:

$$\left. \left(\frac{dy}{dx} \right) \right|_{x=x_i} = \frac{y_{i+1} - y_{i-1}}{2h} + O(h^2) \left(\frac{h^2}{6} M_{3,i} \right). \quad (6.18)$$

Аппроксимация формульных функций, входящих в правую часть (6.9), осуществляется путем перехода к их сеточному представлению.

Методика построения разностных схем с помощью аппроксимаций производной

1. Вводится в общем случае неравномерная сетка $\Omega_n = (x_0, x_1, \dots, x_i, x_{i+1}, \dots, x_n)$. Величина шага $h_{i+1} = x_{i+1} - x_i$ выражается через узловые точки.

2. Для получения рассматриваемых здесь двух схем (одношаговой и двухшаговой) выбирается соответственно двухточечный (одношаговый) и трехточечный (двуихшаговый) шаблоны $\Pi_{2,i} = (x_i, x_{i+1})$ и $\Pi_{3,i} = (x_{i-1}, x_i, x_{i+1})$, на которых в точке x_i производится аппроксимация производной $\left. \left(\frac{dy}{dx} \right) \right|_{x=x_i}$ по формулам (6.16), (6.17) соответственно в левой и центральной точках x_i .

Далее заменяется правая часть уравнения (6.9) ее сеточным представлением, т. е. $f(x, y) \rightarrow f(x_i, y_i)$, а вместо $y(x)$ рассматривается сеточная функция $\hat{y}_i \approx y(x_i)$, которая определяется только в точках сетки.

3. Выполняется подстановка аппроксимаций (6.16), (6.17) и $f(x_i, y_i)$ в дифференциальное уравнение:

$$\begin{aligned} \frac{y_{i+1} - y_i}{h_{i+1}} + O(h_{i+1}) &= f(x_i, y_i); \\ \frac{1}{H_i^{i+1}} \left[-\delta_{i+1} \cdot y_{i-1} + \frac{(\delta_{i+1}^2 - 1)}{\delta_{i+1}} y_i + \frac{1}{\delta_{i+1}} y_{i+1} \right] + O(h_i h_{i+1}) &= f(x_i, y_i). \end{aligned}$$

После отбрасывания остаточных слагаемых получаются разностные схемы:

- явная схема Эйлера первого порядка (явный метод Эйлера):

$$\hat{y}_{i+1} = \hat{y}_i + h_{i+1} f(x_i, \hat{y}_i), \quad i = \overline{0, n-1}, \quad \hat{y}_0 = y_0; \quad (6.19)$$

- обобщенная на нерегулярный шаблон двухшаговая явная схема Эйлера второго порядка:

$$\hat{y}_{i+1} = \hat{y}_i - \delta_{i+1}^2 \Delta \hat{y}_i + H_i^{i+1} \delta_{i+1} f(x_i, \hat{y}_i), \quad i = \overline{1, n-1}, \quad (6.20)$$

где $\Delta \hat{y}_i = \hat{y}_{i+1} - \hat{y}_i$.

Обозначим схему (6.20) символами 2Я2Б (суть этого обозначения поясняется в п. 6.2.2). Анализ остаточного слагаемого $\left(\frac{h_i^2}{6} \delta_{i+1} M_{3,i}\right)$ обобщенной схемы Эйлера указывает на то, что при условном выборе шаблона, когда достигается соотношение $\delta_{i+1} \leq h_i$ (т. е. $h_{i+1} \leq h_i^2$), порядок аппроксимации этой схемы повышается на единицу без увеличения количества точек шаблона. Этот прием повышения точности может быть использован при конструировании вложенных алгоритмов путем измельчения шага.

При $\delta_{i+1} = 1$ (регулярный шаблон) схема (6.20) соответствует *методу Эйлера — Коши*:

$$\hat{y}_{i+1} = \hat{y}_i + 2h \cdot f(x_i, \hat{y}_i), \quad i = \overline{1, n-1}. \quad (6.21)$$

Для начала расчетов по формулам (6.20), (6.21) требуется иметь две «разгонные» точки \hat{y}_0 , \hat{y}_1 . Первая определяется известным начальным условием $\hat{y}_0 = y_0$, а вторая может быть найдена с помощью другого метода, например по формуле (6.19):

$$\hat{y}_1 = y_0 + h_1 f(x_0, y_0).$$

Приведем геометрическую интерпретацию явного метода Эйлера (6.19), которая показана на рисунке 6.5.

Пусть известна точка (x_i, \hat{y}_i) на искомой интегральной кривой $y = y(x)$. Через эту точку проведем к кривой $y = y(x)$ касательную L , тангенс угла наклона которой равен $y'(x_i) = f(x_i, \hat{y}_i)$. Ее уравнение имеет вид

$$y = \hat{y}_i + y'(x_i)(x - x_i) = \hat{y}_i + f(x_i, \hat{y}_i)(x - x_i).$$

Следующая точка \hat{y}_{i+1} приближенного решения получается при $x = x_{i+1}$ (в результате нахождения точки пересечения прямых L и $x_{i+1} = \text{const}$):

$$\hat{y}_{i+1} = \hat{y}_i + h_{i+1} f(x_i, \hat{y}_i).$$

Это соотношение совпадает с формулой (6.19).

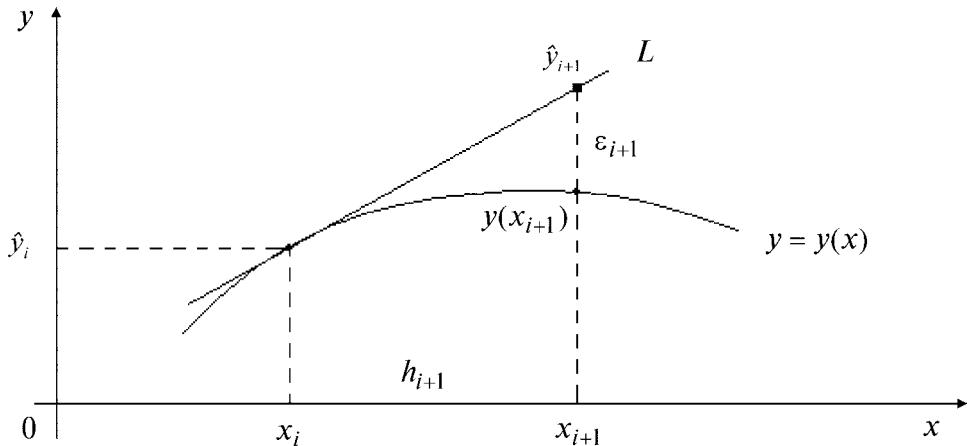


Рис. 6.5

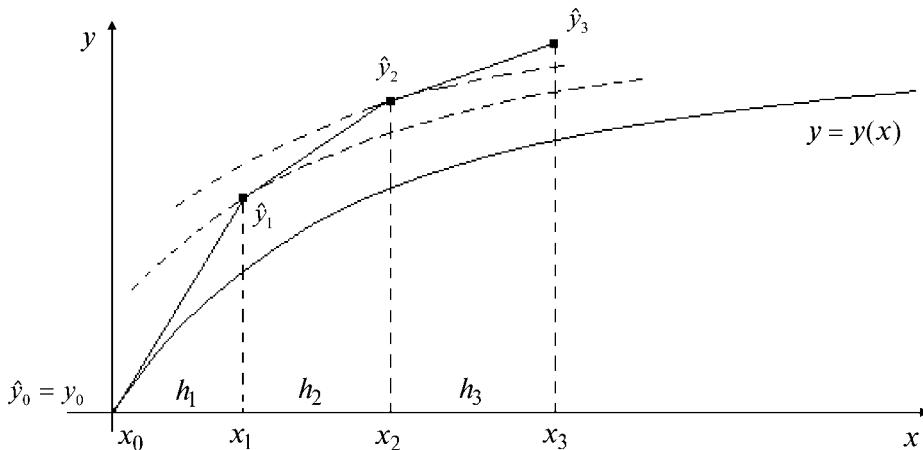


Рис. 6.6

Схема (6.19) иногда называется *схемой ломаных* (рис. 6.6), так как очередное значение \hat{y}_{i+1} получается в результате проведения касательной к интегральной кривой в точке (x_i, \hat{y}_i) . После выполнения каждого шага метода Эйлера приближенное решение переходит с одной интегральной кривой на другую, что обусловлено погрешностью численного решения. Большая погрешность схемы связана с тем, что порядок этой схемы равен единице, и поэтому ломаная, определяемая по (6.19), довольно значительно отклоняется от точного решения $y = y(x)$. В связи с этим данная схема используется, как правило, в качестве вспомогательной при конструировании схем повышенного порядка.

Для анализа устойчивости явного метода Эйлера применим формулу (6.19) при $h_{i+1} = h$ к решению задачи (6.12). В результате разностное уравнение будет иметь вид

$$\hat{y}_{i+1} = \hat{y}_i + h \cdot \mu \hat{y}_i \text{ или } \hat{y}_{i+1} - (1 + h \cdot \mu) \cdot \hat{y}_i = 0.$$

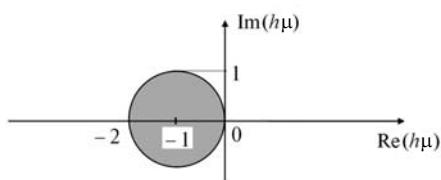


Рис. 6.7

Корень соответствующего характеристического уравнения $\lambda_1 = 1 + h \cdot \mu$. Условие (6.13) принимает форму $|1 + h \cdot \mu| < 1$. Отсюда следует, что область устойчивости представляется внутренней частью круга единичного радиуса с центром в точке $(-1; 0)$ (рис. 6.7). Если μ действительная константа, то можно указать интервал устойчивости $-2 < h\mu < 0$, т. е.

$h\mu \in (-2, 0)$, а при $\mu < 0$ имеем $0 < h < -\frac{2}{\mu}$. Поэтому явный метод Эйлера является ограниченно устойчивым с критическим шагом $h_{kp} = -\frac{2}{\mu}$.

Дадим геометрическую интерпретацию метода Эйлера — Коши (6.21) применительно к решению задачи (6.9) (рис. 6.8). Пусть известны две точки (x_{i-1}, \hat{y}_{i-1}) , (x_i, \hat{y}_i) на искомой интегральной кривой $y = y(x)$. Проведем

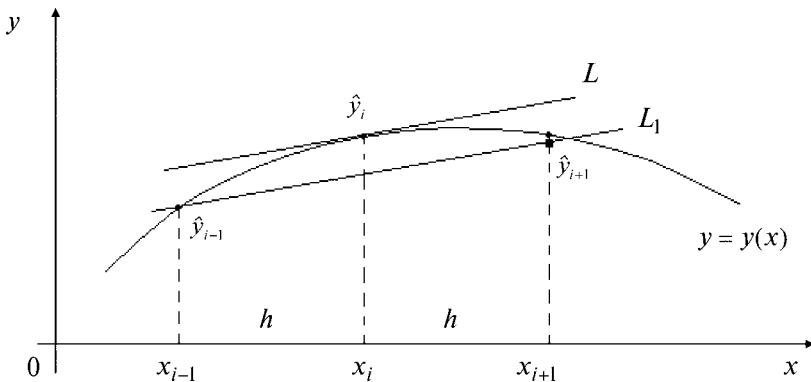


Рис. 6.8

через точку (x_i, \hat{y}_i) касательную L , тангенс угла наклона которой равен $y'(x_i) = f(x_i, \hat{y}_i)$. После этого через точку (x_{i-1}, \hat{y}_{i-1}) проведем прямую L_1 , параллельную L . Ее уравнение имеет вид

$$y = \hat{y}_{i-1} + f(x_i, \hat{y}_i)(x - x_{i-1}).$$

Следующая точка \hat{y}_{i+1} приближенного решения получается при $x = x_{i+1}$:

$$\hat{y}_{i+1} = \hat{y}_{i-1} + 2hf(x_i, \hat{y}_i).$$

Это соотношение совпадает с формулой (6.21).

Замечание. Все описанные здесь и далее методы легко переносятся на системы обыкновенных дифференциальных уравнений (6.7). Под y и $f(x, y)$ следует подразумевать векторы $Y = (y_1, \dots, y_n)^T$ и $F(x, Y) = (f_1(x, Y), \dots, f_n(x, Y))^T$. Например, для системы двух дифференциальных уравнений

$$\begin{aligned} y' &= f(x, y, z), \quad y(x_0) = y_0, \\ z' &= g(x, y, z), \quad z(x_0) = z_0 \end{aligned}$$

явный метод Эйлера (6.19) записывается в форме

$$\begin{aligned} \hat{y}_{i+1} &= \hat{y}_i + h_{i+1}f(x, \hat{y}_i, \hat{z}_i), \quad \hat{y}_0 = y_0, \\ \hat{z}_{i+1} &= \hat{z}_i + h_{i+1}g(x, \hat{y}_i, \hat{z}_i), \quad \hat{z}_0 = z_0. \end{aligned}$$

Пример 6.3. Исследовать устойчивость явного метода Эйлера на примере решения уравнения $Ty' + y = 1$, $y(0) = y_0$, где T — действительное положительное число. Найти приближенное решение задачи Коши

$$0,1y' + y = 1, \quad y(0) = 0$$

на отрезке $[0, 1]$ явным методом Эйлера.

□Точное решение задачи получено в примере 6.1: $y(x) = 1 - e^{-10x}$, поскольку $T = 0,1$. Уравнение $Ty' + y = 1$ перепишем в виде, аналогичном (6.12):

$$y' = -\frac{1}{T}y + \frac{1}{T},$$

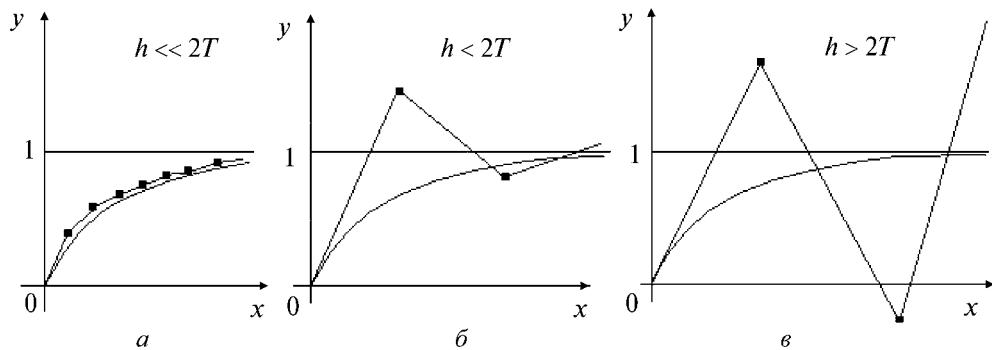


Рис. 6.9

Таблица 6.1

x_i	$\hat{y}_i (h = 0,05)$	$\hat{y}_i (h = 0,2)$	$\hat{y}_i (h = 0,5)$	$y(x_i)$
0,00	0,000000	0,000	0,000	0,000000
0,05	0,500000			0,393469
0,10	0,750000			0,632121
0,15	0,875000			0,776870
0,20	0,937500	2,000		0,864665
0,25	0,968750			0,917915
0,30	0,984375			0,950213
0,35	0,992187			0,969803
0,40	0,996094	0,000		0,981684
0,45	0,998047			0,988891
0,50	0,999023		5,000	0,993262
0,55	0,999511			0,995913
0,60	0,999755	2,000		0,997521
0,65	0,999877			0,998497
0,70	0,999938			0,999088
0,75	0,999969			0,999446
0,80	0,999984	0,000		0,999665
0,85	0,999992			0,999797
0,90	0,999996			0,999877
0,95	0,999998			0,999925
1,00	0,999999	2,000	-15,000	0,999955
$\max \varepsilon_i$	0,117879	1,135335	15,99995	

поэтому $\mu = -\frac{1}{T}$, а критический шаг равен $h_{\text{кр}} = -\frac{2}{\mu} = 2T$. Следовательно, при $h < 2T$ метод устойчив, а при $h > 2T$ неустойчив. На рисунке 6.9 схематически изображены приближенные решения уравнения $Ty' + y = 1$, $y(0) = 0$, соответствующие устойчивому процессу и хорошей точности численного решения (рис. 6.9a), устойчивому процессу и плохой точности (рис. 6.9б), а также неустойчивому процессу и плохой точности (рис. 6.9в).

Для уравнения $0,1y' + y = 1$ постоянная времени $T = 0,1$, поэтому $h_{\text{кр}} = 0,2$.

Решим задачу для значений шага $h = 0,05; 0,2; 0,5$. При $h = 0,05$ метод является устойчивым, при $h = 0,2$ находится на границе устойчивости, а при $h = 0,5$ является неустойчивым.

Переписав уравнение в виде $y' = 10 - 10y$, получим из (6.19) формулу для нахождения приближенного решения:

$$\hat{y}_{i+1} = \hat{y}_i + h \cdot (10 - 10\hat{y}_i) = 10h + (1 - 10h)\hat{y}_i.$$

Результаты расчетов, выполненных для различных значений h , приведены в таблице 6.1. При $h = 0,05$ приближенное решение достаточно близко к точному, при $h = 0,2$ «пила» не сходится и не расходится, а при $h = 0,5$ «пила», аппроксимирующая точное решение, расходится (рис. 6.9в). ■

Замечание. Порядок точности метода, как правило, определяется порядком аппроксимации схем. Например, для схемы (6.19) первого порядка, записанной в виде, соответствующем исходному уравнению (6.9), получаем (первое слагаемое аппроксимирует производную):

$$\frac{y(x_{i+1}) - y(x_i)}{h_{i+1}} - f(x_i, y_i) = \frac{y_i + h_{i+1}y'_i + O(h_{i+1}^2) - y_i}{h_{i+1}} - f_i = y'_i - f_i + O(h_{i+1}) = 0,$$

т. е. $y'_i = f_i + O(h_{i+1})$. Здесь были использованы формула Тейлора и свойства, приведенные в п. В.3.

Для двухшаговой схемы (6.21) аналогично получается второй порядок аппроксимации. Действительно, воспользуемся частным случаем формулы Тейлора (см. (B.18)) при $x_0 = x_i$, $x_{i+1} = x_i + h$, $x_{i-1} = x_i - h$, $f(x) = y(x)$:

$$y_{i+1} = y_i + hy'_i + \frac{h^2}{2}y''_i + O(h^3), \quad y_{i-1} = y_i - hy'_i + \frac{h^2}{2}y''_i + O(h^3).$$

Подставляя полученные соотношения в схему (6.21), записанную в виде $\frac{y_{i+1} - y_{i-1}}{2h} - f(x_i, y_i) = 0$, соответствующем дифференциальному уравнению (6.9), с учетом свойств, приведенных в п. В.3, имеем

$$\frac{y_i + hy'_i + \frac{h^2}{2}y''_i + O(h^3) - y_i + hy'_i - \frac{h^2}{2}y''_i + O(h^3)}{2h} - f_i = y'_i + O(h^2) - f_i = 0,$$

т. е. $y'_i = f_i + O(h^2)$.

Также легко проверить, что и схема (6.20) на нерегулярном шаблоне имеет второй порядок при безусловной аппроксимации. Если же $h_{i+1} \leq h_i^2$ (условная аппроксимация), то этот порядок равен трем.

6.2.2. ИНТЕГРАЛЬНО-ИНТЕРПОЛЯЦИОННЫЙ ПРИНЦИП

В данном подходе явные или неявные методы соответствующего порядка получаются путем интегрирования уравнения (6.9) вдоль решения $y(x)$. Так, на двухточечном шаблоне

$$\int_{y_i}^{y_{i+1}} dy = \int_{x_i}^{x_{i+1}} f(x, y(x)) dx. \quad (6.22)$$

Функция $f(x, y(x))$ в (6.22) заменяется интерполяционным многочленом соответствующей степени p ($p = 0$ для метода первого порядка, $p = 1$ для метода второго порядка, $p = 2$ для метода третьего порядка и т. д.). Подчеркнем, что замена $f(x, y)$ интерполяционным многочленом в (6.22) становится возможной благодаря тому, что $f(x, y)$ рассматривается не на всей плоскости (x, y) , а только вдоль интегральной кривой $y(x)$. Для вычисления интеграла в правой части могут быть использованы известные квадратурные формулы прямоугольников, трапеций, парабол (Симпсона) и др.

Примем x_{i+1} в качестве узловой точки многочлена $L_0(x)$, т. е. $L_0(x) = f(x_{i+1}, \hat{y}_{i+1})$. Тогда из (6.22) следует формула *неявного метода Эйлера*:

$$\hat{y}_{i+1} = \hat{y}_i + h_{i+1} f(x_{i+1}, \hat{y}_{i+1}) \equiv \Phi(x_i, x_{i+1}, \hat{y}_{i+1}), \quad i = \overline{0, n-1}. \quad (6.23)$$

Принимая для интерполяционного многочлена $L_1(x)$ в качестве узловых точек x_i и x_{i+1} и используя квадратурную формулу трапеций (5.46), получаем неявную одношаговую схему второго порядка (*метод трапеций*):

$$\hat{y}_{i+1} = \hat{y}_i + \frac{h_{i+1}}{2} [f_i + f(x_{i+1}, \hat{y}_{i+1})] \equiv \Phi(x_i, x_{i+1}, \hat{y}_{i+1}), \quad i = \overline{0, n-1}, \quad (6.24)$$

где $f_i = f(x_i, \hat{y}_i)$. Формулы (6.23), (6.24) имеют вид (6.11) при $k = 1$. Подчеркнем, что свойство неявности схем (6.23), (6.24) обусловлено наличием искомой величины \hat{y}_{i+1} в левой и правой частях в общем случае нелинейных уравнений.

При реализации алгоритма решения задачи Коши неизвестное значение \hat{y}_{i+1} вычисляется одним из методов решения нелинейных уравнений (см. главу 3). Применение метода Ньютона связано с записью уравнения в форме

$$\hat{y}_{i+1} - \Phi(x_i, x_{i+1}, \hat{y}_{i+1}) \equiv F(\hat{y}_{i+1}) = 0 \quad (6.25)$$

и с дифференцированием функции $F(\hat{y}_{i+1})$, что увеличивает время расчетов из-за возможной сложности вычисления производных или увеличивает погрешность в силу неточного задания правых частей.

Как правило, используется метод простых итераций:

$$\hat{y}_{i+1}^{(k+1)} = \Phi(x_i, x_{i+1}, \hat{y}_{i+1}^{(k)}), \quad k = 0, 1, \dots \quad (6.26)$$

Таким образом, уравнение (6.24) решается с помощью формулы (рекуррентного соотношения):

$$\hat{y}_{i+1}^{(k+1)} = \hat{y}_i + \frac{h_{i+1}}{2} \cdot [f_i + f(x_{i+1}, \hat{y}_{i+1}^{(k)})], \quad k = 0, 1, \dots \quad (6.27)$$

При применении методов Ньютона и простых итераций вначале задается или находится нулевое приближение решения по формуле $y_{i+1}^{(0)} = \hat{y}_i$ (так называемый «постоянный» прогноз) или явным методом Эйлера (6.19):

$$\hat{y}_{i+1}^{(0)} = \hat{y}_i + h_{i+1} f(x_i, \hat{y}_i).$$

Итерации завершаются при выполнении условия окончания

$$|\hat{y}_{i+1}^{(k+1)} - \hat{y}_{i+1}^{(k)}| \leq \varepsilon,$$

где ε — малое положительное число.

Для сокращения описания и удобства ссылок получаемые здесь схемы обозначаются буквенно-цифровыми символами. Первая цифра в нем обозначает «шаговость» схемы, следующая за ней буква «Я» или буквы «НЯ» — явный или неявный характер, затем цифра указывает порядок точности схемы и за ней буква — возможную модификацию. Таким образом, схема (6.24) обозначается 1НЯ2, т. е. она является одношаговой, неявной и имеет второй порядок точности. Последняя буква указывается только для тех схем, которые имеют модификации.

Приведем геометрическую интерпретацию неявного метода Эйлера (6.23) (рис. 6.10). Пусть известна точка (x_i, \hat{y}_i) на искомой интегральной кривой $y = y(x)$. Через эту точку проведем прямую, тангенс угла наклона которой равен $y'(x_{i+1}) = f(x_{i+1}, \hat{y}_{i+1})$, т. е. следующая точка \hat{y}_{i+1} приближенного решения берется на той кривой из семейства интегральных кривых, касательная к которой в точке (x_{i+1}, \hat{y}_{i+1}) проходит через точку (x_i, \hat{y}_i) . Поэтому $\frac{\hat{y}_{i+1} - \hat{y}_i}{h_{i+1}} = f(x_{i+1}, \hat{y}_{i+1})$.

Полученное соотношение совпадает с формулой (6.23).

Исследуем устойчивость неявного метода Эйлера и метода трапеций. Применим формулу (6.23) при $h_{i+1} = h$ к решению задачи (6.12). В результате получим разностное уравнение

$$\hat{y}_{i+1} = \hat{y}_i + h \cdot \mu \cdot \hat{y}_{i+1} \text{ или } \hat{y}_{i+1}(1 - h \cdot \mu) - \hat{y}_i = 0.$$

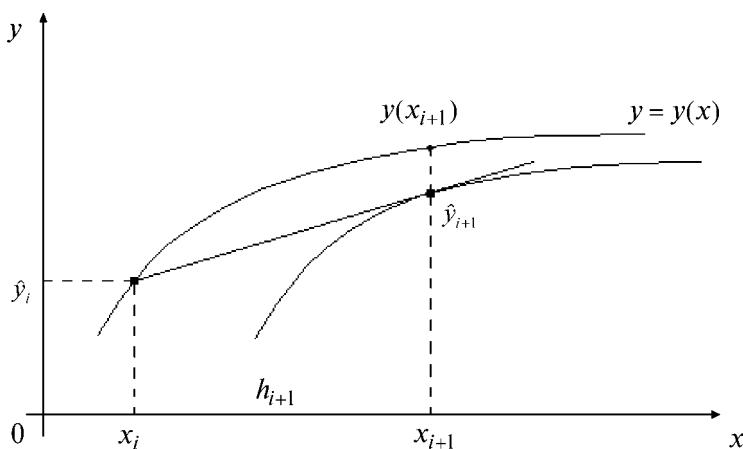


Рис. 6.10

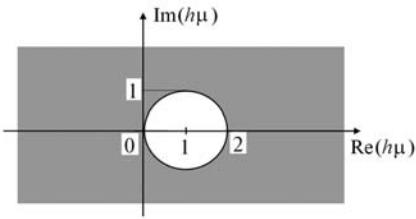


Рис. 6.11

Корень характеристического уравнения $\lambda_1 = \frac{1}{1-h \cdot \mu}$. Условие устойчивости (6.13) имеет вид $\left| \frac{1}{1-h \cdot \mu} \right| < 1$ или $|1 - h \cdot \mu| > 1$.

Область устойчивости представляется внешней частью круга единичного радиуса с центром в точке (1; 0). Из геометрической интерпретации (рис. 6.11) следует,

что неявный метод Эйлера обладает свойством A -устойчивости (см. рис. 6.4a), поскольку выделенная на рисунке 6.11 область включает левую полуплоскость.

Применим формулу (6.24) метода трапеций при $h_{i+1} = h$ к решению задачи (6.12). Получим разностное уравнение

$$\hat{y}_{i+1} = \hat{y}_i + \frac{h}{2}(\mu\hat{y}_i + \mu\hat{y}_{i+1}) \text{ или } \hat{y}_{i+1} \left(1 - \frac{h\mu}{2}\right) - \left(1 + \frac{h\mu}{2}\right)\hat{y}_i = 0.$$

Корень характеристического уравнения

$$\lambda_1 = \frac{1 + \frac{h\mu}{2}}{1 - \frac{h\mu}{2}}.$$

Условие устойчивости (6.13) принимает вид

$$\left| \frac{1 + \frac{h\mu}{2}}{1 - \frac{h\mu}{2}} \right| < 1 \text{ или } |2 + h\mu| < |2 - h\mu|.$$

Полученное неравенство выполняется в области, изображенной на рисунке 6.4a, т. е. метод трапеций является A -устойчивым.

Продолжим рассмотрение интегрально-интерполяционного метода. Использование квадратурной формулы прямоугольников дает *модифицированный метод Эйлера второго порядка*:

$$\begin{aligned} \hat{y}_{i+\frac{1}{2}} &= \hat{y}_i + \frac{h_{i+1}}{2} f(x_i, \hat{y}_i), \quad i = \overline{0, n-1}; \\ \hat{y}_{i+1} &= \hat{y}_i + h_{i+1} f\left(x_i + \frac{h_{i+1}}{2}, \hat{y}_{i+\frac{1}{2}}\right), \quad i = \overline{0, n-1}. \end{aligned} \tag{6.28}$$

Дадим геометрическую интерпретацию модифицированного метода Эйлера второго порядка (рис. 6.12).

Пусть известна точка (x_i, \hat{y}_i) на искомой интегральной кривой $y = y(x)$. Через эту точку проведем касательную L , тангенс угла наклона которой равен $y'(x_i) = f(x_i, \hat{y}_i)$. При $x = x_i + \frac{h_{i+1}}{2}$ получаем точку $\hat{y}_{i+\frac{1}{2}}$. Далее проведем прямую L_1 с тангенсом угла наклона $f\left(x_i + \frac{h_{i+1}}{2}, \hat{y}_{i+\frac{1}{2}}\right)$, а затем прямую L_2 , параллельную L_1 и проходящую через точку (x_i, \hat{y}_i) . Ее уравнение имеет вид

$$y = \hat{y}_i + f\left(x_i + \frac{h_{i+1}}{2}, \hat{y}_{i+\frac{1}{2}}\right) \cdot (x - x_i).$$

Следующая точка \hat{y}_{i+1} приближенного решения находится при $x = x_{i+1}$:

$$\hat{y}_{i+1} = \hat{y}_i + h_{i+1} f\left(x_i + \frac{h_{i+1}}{2}, \hat{y}_{i+\frac{1}{2}}\right).$$

Это соотношение совпадает с формулой (6.28).

Исследуем устойчивость модифицированного метода Эйлера второго порядка. Применим формулы (6.28) при $h_{i+1} = h$ к решению задачи (6.12). В результате получим разностное уравнение:

$$\hat{y}_{i+1} = \hat{y}_i + h\mu \left(\hat{y}_i + \frac{h}{2}\mu \hat{y}_i \right) \text{ или } \hat{y}_{i+1} - \hat{y}_i \left(1 + h\mu + \frac{(h\mu)^2}{2} \right) = 0.$$

Корень соответствующего характеристического уравнения

$$\lambda_1 = 1 + h\mu + \frac{(h\mu)^2}{2}.$$

Условие устойчивости (6.13) имеет вид

$$\left| 1 + h\mu + \frac{(h\mu)^2}{2} \right| < 1.$$

Если μ — действительное число, то, раскрывая модуль, получаем

$$-1 < 1 + h\mu + \frac{(h\mu)^2}{2} < 1.$$

Неравенство $2 + h\mu + \frac{(h\mu)^2}{2} > 0$ выполняется для любых $h\mu$, а неравенство $h\mu + \frac{(h\mu)^2}{2} < 0$, т. е. $h\mu \cdot (2 + h\mu) < 0$, справедливо при $-2 < h\mu < 0$. Таким образом, интервал устойчивости $h\mu \in (-2, 0)$ модифицированного метода Эйлера совпадает с интервалом устойчивости явного метода Эйлера (6.19).

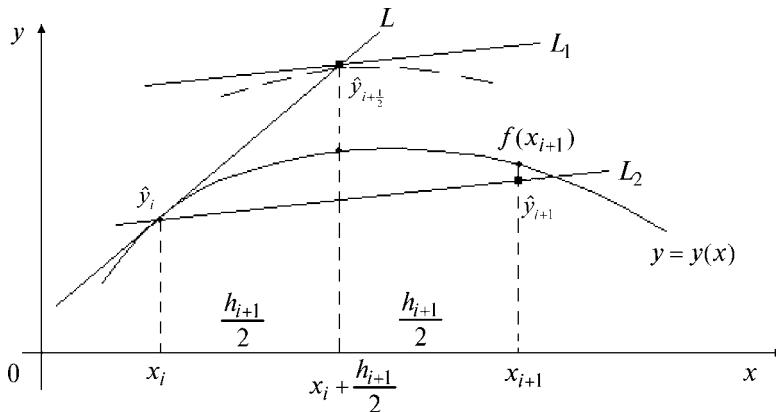


Рис. 6.12

Для использования регулярного трехточечного шаблона (x_{i-1} , x_i , x_{i+1}) заменим (6.22) аналогичным выражением:

$$\int_{y_{i-1}}^{y_{i+1}} dy = \int_{x_{i-1}}^{x_{i+1}} f(x, y(x)) dx.$$

Если функцию $f(x, y(x))$, входящую в правую часть, заменить многочленом Лагранжа второй степени и применить формулу (5.47), то получается *неявная схема парабол (Симпсона) четвертого порядка (2НЯ4)*:

$$\hat{y}_{i+1} = \hat{y}_{i-1} + \frac{h}{3} [f_{i-1} + 4f_i + f(x_{i+1}, \hat{y}_{i+1})]. \quad (6.29)$$

Отметим, что схема парабол имеет четвертый, а не третий порядок точности из-за регулярности шаблона.

В соответствии с классификацией, приведенной в [35], схема (6.29) относится к классу неявных схем Адамса. Для получения других *схем Адамса* следует переписать (6.22) в форме

$$\int_{y_i}^{y_{i+1}} dy = \int_{x_i}^{x_{i+1}} f(x, y(x)) dx = \int_{x_i}^{x_{i+1}} y'(x) dx \quad (6.30)$$

и заменить $y'(x)$ соответствующим интерполяционным полиномом, построенным на регулярном шаблоне с $h = \text{const}$.

В силу формулы (4.36) второго интерполяционного многочлена Ньютона получаем

$$y'(\hat{q}) = y'_i + \hat{q}\Delta y'_{i-1} + \frac{\Delta^2 y'_{i-2}}{2!} \hat{q} \cdot (\hat{q}+1) + \frac{\Delta^3 y'_{i-3}}{3!} \hat{q} \cdot (\hat{q}+1) \cdot (\hat{q}+2) + \dots$$

или

$$y'(\hat{q}) = y'_i + \hat{q}\Delta y'_{i-1} + \frac{(\hat{q}^2 + \hat{q})}{2} \Delta^2 y'_{i-2} + \frac{(\hat{q}^3 + 3\hat{q}^2 + 2\hat{q})}{6} \Delta^3 y'_{i-3} + \dots,$$

где $\hat{q} = \frac{x - x_i}{h}$ — фаза интерполяции.

Подставляя полученное выражение в правую часть (6.30) и учитывая, что $dx = h \cdot d\hat{q}$, вычисляем $\int_0^1 y'(\hat{q}) \cdot h \cdot d\hat{q}$, поскольку при $x = x_i$ и $x = x_{i+1}$ справедливо $\hat{q} = 0$ и $\hat{q} = 1$ соответственно. В результате имеем

$$\hat{y}_{i+1} - \hat{y}_i = h \cdot y'_i + \frac{1}{2} h \cdot \Delta y'_{i-1} + \frac{5}{12} h \cdot \Delta^2 y'_{i-2} + \frac{3}{8} h \cdot \Delta^3 y'_{i-3} + \dots$$

или с учетом (6.9) общую формулу *семейства явных схем Адамса*:

$$\hat{y}_{i+1} = \hat{y}_i + \eta_i + \frac{1}{2} \Delta \eta_{i-1} + \frac{5}{12} \Delta^2 \eta_{i-2} + \frac{3}{8} \Delta^3 \eta_{i-3} + \frac{251}{720} \Delta^4 \eta_{i-4} + \dots, \quad (6.31)$$

где $\eta_i = h \cdot f(x_i, \hat{y}_i) = h \cdot f_i$, а конечные разности вычисляются с помощью таблицы 6.2. Подчеркнем, что здесь интерполяционный многочлен записан на шаблоне, не включающем точку x_{i+1} .

x_k	\hat{y}_k	η_k	$\Delta\eta_k$	$\Delta^2\eta_k$	$\Delta^3\eta_k$
\dots	\dots	η_{i-3}	$\Delta\eta_{i-3}$	$\Delta^2\eta_{i-3}$	$\Delta^3\eta_{i-3}$
x_{i-2}	\hat{y}_{i-2}	η_{i-2}	$\Delta\eta_{i-2}$	$\Delta^2\eta_{i-2}$	$\Delta^3\eta_{i-2}$
x_{i-1}	\hat{y}_{i-1}	η_{i-1}	$\Delta\eta_{i-1}$	$\Delta^2\eta_{i-1}$	$\Delta^3\eta_{i-1}$
x_i	\hat{y}_i	η_i			

Если взять в правой части (6.31) два слагаемых, то получится явный метод Эйлера (6.19), если три слагаемых, то явный метод, описываемый формулой

$$\hat{y}_{i+1} = \hat{y}_i + h \cdot f_i + \frac{1}{2} [h \cdot f_i - h \cdot f_{i-1}] = \hat{y}_i + \frac{h}{2} [3f_i - f_{i-1}],$$

а если четыре слагаемых, то явный метод вида

$$\begin{aligned} \hat{y}_{i+1} &= \hat{y}_i + \frac{h}{2} [3f_i - f_{i-1}] + \frac{5}{12} [(h \cdot f_i - h \cdot f_{i-1}) - (h \cdot f_{i-1} - h \cdot f_{i-2})] = \\ &= \hat{y}_i + \frac{h}{12} [23f_i - 16f_{i-1} + 5f_{i-2}] \end{aligned}$$

и т. д.

В результате имеем схемы Адамса — Бэшфорта:

- второго порядка (2Я2В)

$$\hat{y}_{i+1} = \hat{y}_i + \frac{h}{2} [3f_i - f_{i-1}], \quad i = \overline{1, n-1}; \quad (6.32)$$

- третьего порядка (3Я3А)

$$\hat{y}_{i+1} = \hat{y}_i + \frac{h}{12} [23f_i - 16f_{i-1} + 5f_{i-2}], \quad i = \overline{2, n-1}; \quad (6.33)$$

- четвертого порядка (4Я4)

$$\hat{y}_{i+1} = \hat{y}_i + \frac{h}{24} [55f_i - 59f_{i-1} + 37f_{i-2} - 9f_{i-3}], \quad i = \overline{3, n-1}. \quad (6.34)$$

Для начала расчетов по формуле (6.32) требуются две «разгонные» точки: \hat{y}_0 , \hat{y}_1 , по формуле (6.33) — три «разгонные» точки: \hat{y}_0 , \hat{y}_1 , \hat{y}_2 , по формуле (6.34) — четыре «разгонные» точки: \hat{y}_0 , \hat{y}_1 , \hat{y}_2 , \hat{y}_3 . Их необходимо вычислить с порядком точности не меньше порядка точности схемы. Далее схема (6.32) обобщается на нерегулярный шаблон.

Если для построения интерполяционного многочлена использовать шаблоны, включающие точку x_{i+1} , то получится следующая формула семейства неявных схем Адамса [2], [42]:

$$\hat{y}_{i+1} = \hat{y}_i + \eta_{i+1} - \frac{1}{2}\Delta\eta_i - \frac{1}{12}\Delta^2\eta_{i-1} - \frac{1}{24}\Delta^3\eta_{i-2} - \frac{19}{720}\Delta^4\eta_{i-3} + \dots, \quad (6.35)$$

где $\eta_{i+1} = h \cdot f(x_{i+1}, \hat{y}_{i+1}) = h \cdot f_{i+1}$, а конечные разности вычисляются с помощью таблицы 6.3.

Таблица 6.3						
x_k	\hat{y}_k	η_k	$\Delta\eta_k$	$\Delta^2\eta_k$	$\Delta^3\eta_k$	$\Delta^4\eta_k$
...				
x_{i-3}	\hat{y}_{i-3}	η_{i-3}	$\Delta\eta_{i-3}$	$\Delta^2\eta_{i-3}$	$\Delta^3\eta_{i-3}$	$\Delta^4\eta_{i-3}$
x_{i-2}	\hat{y}_{i-2}	η_{i-2}	$\Delta\eta_{i-2}$	$\Delta^2\eta_{i-2}$	$\Delta^3\eta_{i-2}$	$\Delta^4\eta_{i-2}$
x_{i-1}	\hat{y}_{i-1}	η_{i-1}	$\Delta\eta_{i-1}$	$\Delta^2\eta_{i-1}$	$\Delta^3\eta_{i-1}$	$\Delta^4\eta_{i-1}$
x_i	\hat{y}_i	η_i	$\Delta\eta_i$	$\Delta^2\eta_i$	$\Delta^3\eta_i$	$\Delta^4\eta_i$
x_{i+1}	\hat{y}_{i+1}	η_{i+1}				

Если взять в правой части (6.35) два слагаемых, то получится неявный метод Эйлера (6.23), если три слагаемых — метод трапеций (6.24):

$$\hat{y}_{i+1} = \hat{y}_i + h \cdot f_{i+1} - \frac{1}{2}[h \cdot f_{i+1} - h \cdot f_i] = \hat{y}_i + \frac{h}{2}[f_i + f(x_{i+1}, \hat{y}_{i+1})],$$

если четыре слагаемых — неявный метод, описываемый соотношением

$$\begin{aligned} \hat{y}_{i+1} &= \hat{y}_i + \frac{h}{2}(f_i + f_{i+1}) - \frac{1}{12}[(h \cdot f_{i+1} - h \cdot f_i) - (h \cdot f_i - h \cdot f_{i-1})] = \\ &= \hat{y}_i + \frac{h}{12}[-f_{i-1} + 8f_i + 5f(x_{i+1}, \hat{y}_{i+1})], \end{aligned}$$

и т. д.

В результате получаются *неявные схемы Адамса — Мултона*:

- первого порядка (6.23);
- второго порядка (6.24);
- третьего порядка (ЗНЯ3)

$$\hat{y}_{i+1} = \hat{y}_i + \frac{h}{12}[-f_{i-1} + 8f_i + 5f(x_{i+1}, \hat{y}_{i+1})], \quad i = \overline{1, n-1}; \quad (6.36)$$

- четвертого порядка (ЗНЯ4)

$$\hat{y}_{i+1} = \hat{y}_i + \frac{h}{24}[f_{i-2} - 5f_{i-1} + 19f_i + 9f(x_{i+1}, \hat{y}_{i+1})], \quad i = \overline{2, n-1}. \quad (6.37)$$

Для расчетов по формуле (6.36) требуются две «разгонные» точки: \hat{y}_0 , \hat{y}_1 , по формуле (6.37) — три «разгонные» точки: \hat{y}_0 , \hat{y}_1 , \hat{y}_2 . Их необходимо вычислить с порядком точности не меньше порядка точности схемы.

Чтобы найти искомое значение \hat{y}_{i+1} с помощью формул (6.36), (6.37), так же как в неявном методе Эйлера и методе трапеций, требуется решить в общем случае нелинейное уравнение.

Замечания

1. Аналогичные методы могут быть получены из интегрального уравнения вида

$$\int_{y_{i-k}}^{y_{i+1}} dy = \int_{x_{i-k}}^{x_{i+1}} f(x, y(x)) dx, \quad k \geq 2.$$

2. Области устойчивости методов Адамса — Бэшфорта и Адамса — Мултона приведены в [54].

Пример 6.4. Найти приближенное решение задачи Коши

$$\begin{aligned} y' &= z - 1, \quad y(0) = 1, \\ z' &= -y - 2z, \quad z(0) = -1 \end{aligned}$$

на отрезке $[0; 1]$ с шагом $h = 0,1$ методами Эйлера (неявным и модифицированным), Адамса — Бэшфорта третьего порядка и трапеций.

□Путем прямой подстановки в систему легко убедиться в том, что точное решение задачи имеет вид

$$y(x) = -2 + 3e^{-x} + xe^{-x}; \quad z(x) = 1 - 2e^{-x} - xe^{-x}.$$

Используя соотношения (6.23), (6.24), (6.28), (6.33), выписываем формулы для нахождения приближенного решения указанными методами (при этом применяется векторная форма записи).

Для неявного метода Эйлера из (6.23) имеем

$$\begin{pmatrix} \hat{y}_{i+1} \\ \hat{z}_{i+1} \end{pmatrix} = \begin{pmatrix} \hat{y}_i \\ z_i \end{pmatrix} + 0,1 \cdot \begin{pmatrix} \hat{z}_{i+1} - 1 \\ -\hat{y}_{i+1} - 2\hat{z}_{i+1} \end{pmatrix} = \begin{pmatrix} \hat{y}_i + 0,1 \cdot \hat{z}_{i+1} - 0,1 \\ \hat{z}_i - 0,1 \cdot \hat{y}_{i+1} - 0,2 \cdot \hat{z}_{i+1} \end{pmatrix}, \quad \hat{y}_0 = y(0) = 1, \quad \hat{z}_0 = z(0) = -1.$$

Разрешая эту систему относительно \hat{y}_{i+1} , \hat{z}_{i+1} , окончательно получаем

$$\hat{y}_{i+1} = \frac{0,1 \cdot (\hat{z}_i - 0,1 \cdot \hat{y}_i + 0,001)}{1,21} - 0,1; \quad \hat{z}_{i+1} = \frac{\hat{z}_i - 0,1 \cdot \hat{y}_i + 0,01}{1,21}.$$

Соотношения (6.28) для модифицированного метода Эйлера принимают вид

$$\begin{aligned} \hat{y}_{i+\frac{1}{2}} &= \hat{y}_i + \frac{0,1}{2}(\hat{z}_i - 1); \quad \hat{z}_{i+\frac{1}{2}} = \hat{z}_i + \frac{0,1}{2}(-\hat{y}_i - 2\hat{z}_i); \\ \hat{y}_{i+1} &= \hat{y}_i + 0,1 \cdot \left(\hat{z}_{i+\frac{1}{2}} - 1 \right); \quad \hat{z}_{i+1} = \hat{z}_i + 0,1 \cdot \left(-\hat{y}_{i+\frac{1}{2}} - 2\hat{z}_{i+\frac{1}{2}} \right). \end{aligned}$$

Для метода Адамса — Бэшфорта третьего порядка из (6.33) находим

$$\begin{aligned} \hat{y}_{i+1} &= \hat{y}_i + \frac{0,1}{12} [23 \cdot (\hat{z}_i - 1) - 16 \cdot (\hat{z}_{i-1} - 1) + 5 \cdot (\hat{z}_{i-2} - 1)]; \\ \hat{z}_{i+1} &= \hat{z}_i + \frac{0,1}{12} [23 \cdot (-\hat{y}_i - 2\hat{z}_i) - 16 \cdot (-\hat{y}_{i-1} - 2\hat{z}_{i-1}) + 5 \cdot (-\hat{y}_{i-2} - 2\hat{z}_{i-2})]. \end{aligned}$$

Для определения «разгонных» точек $(x_0, \hat{y}_0, \hat{z}_0)$, $(x_1, \hat{y}_1, \hat{z}_1)$, $(x_2, \hat{y}_2, \hat{z}_2)$ воспользуемся модифицированным методом Эйлера. Точка $(x_0, \hat{y}_0, \hat{z}_0)$ определяется начальными условиями $(0; 1; -1)$.

Выпишем соотношения для метода трапеций (6.24) и разрешим их относительно неизвестных:

$$\begin{aligned} \begin{pmatrix} \hat{y}_{i+1} \\ \hat{z}_{i+1} \end{pmatrix} &= \begin{pmatrix} \hat{y}_i \\ \hat{z}_i \end{pmatrix} + \frac{0,1}{2} \left[\begin{pmatrix} \hat{z}_i - 1 \\ -\hat{y}_i - 2\hat{z}_i \end{pmatrix} + \begin{pmatrix} \hat{z}_{i+1} - 1 \\ -\hat{y}_{i+1} - 2\hat{z}_{i+1} \end{pmatrix} \right] = \\ &= \begin{pmatrix} \hat{y}_i + 0,05\hat{z}_i + 0,05\hat{z}_{i+1} - 0,1 \\ -0,05\hat{y}_i - 0,05\hat{y}_{i+1} + 0,9\hat{z}_i - 0,1\hat{z}_{i+1} \end{pmatrix}, \end{aligned}$$

откуда путем разрешения системы относительно \hat{y}_{i+1} , \hat{z}_{i+1} получаем

$$\begin{aligned} \hat{y}_{i+1} &= \hat{y}_i + 0,05\hat{z}_i + 0,05 \frac{-0,1\hat{y}_i + 0,8975\hat{z}_i + 0,05}{1,1025} - 0,1; \\ \hat{z}_{i+1} &= \frac{-0,1\hat{y}_i + 0,8975\hat{z}_i + 0,05}{1,1025}. \end{aligned}$$

Результаты проведенных расчетов даны в таблице 6.4 для \hat{y}_i , $y(x)$ и таблице 6.5 для \hat{z}_i , $z(x)$ соответственно. Из их анализа вытекает, что наиболее точно величину \hat{y}_i можно рассчитать методом Адамса — Бэшфорта, а величину \hat{z}_i — методом трапеций. ■

Пример 6.5. Исследовать устойчивость метода Адамса — Бэшфорта третьего порядка на примере решения задачи Коши [54]

$$y' = \mu \cdot (y - x^3) + 3x^2, \quad y\left(-\frac{1}{4}\right) = -0,015625$$

при $h = \frac{1}{8}$, $\mu = -1$ и $\mu = -100$.

□Точное решение задачи: $y(x) = x^3$. При $\mu = -1$ ($h\mu = -\frac{1}{8}$) дифференциальное уравнение имеет вид

$$y' = x^3 - y + 3x^2,$$

а при $\mu = -100$ ($h\mu = -\frac{100}{8}$):

$$y' = 100x^3 - 100y + 3x^2.$$

Формула (6.33) при $h = \frac{1}{8}$ принимает форму

$$\hat{y}_{i+1} = \hat{y}_i + \frac{1}{96}(23f_i - 16f_{i-1} + 5f_{i-2}), \quad i = 2, 3, \dots$$

Положим $f_0 = y'\left(-\frac{1}{4}\right) = 0,1875$; $\hat{y}_0 = y_0 = -0,015625$; $f_1 = y'\left(-\frac{1}{8}\right) = 0,046875$; $\hat{y}_1 = y_0 = -0,001953121$; $f_2 = y'(0) = 0$; $\hat{y}_2 = 0$, т. е. используем в качестве «разгонных» точек значения, соответствующие точному решению. Результаты нахождения приближенного решения задачи Коши приведены в таблице 6.6.

Таблица 6.4

x_i	Неявный метод Эйлера	Модифицированный метод Эйлера	Метод Адамса — Бэнфорта	Метод трапеций	Точное решение
0,0	1,00000	1,00000	1,00000	1,00000	1,00000
0,1	0,80992	0,80500	0,80499	0,80499	0,80500
0,2	0,62960	0,61997	0,61991	0,61991	0,61994
0,3	0,45885	0,44479	0,44469	0,44464	0,44470
0,4	0,29740	0,27924	0,27908	0,279000	0,27909
0,5	0,14500	0,12309	0,12285	0,12273	0,12286
0,6	0,00131	-0,02397	-0,02428	-0,02444	-0,02428
0,7	-0,13397	-0,16224	-0,16265	-0,16284	-0,16263
0,8	-0,26119	-0,29208	-0,29257	-0,29279	-0,29255
0,9	-0,38072	-0,41384	-0,41441	-0,41465	-0,41438
1,0	-0,49287	-0,52787	-0,52853	-0,52879	-0,52848
$\max \varepsilon_i$	0,03720415	0,00067291	0,00005942	0,00033550	—

Таблица 6.5

x_i	Неявный метод Эйлера	Модифицированный метод Эйлера	Метод Адамса — Бэнфорта	Метод трапеций	Точное решение
0,0	-1,00000	-1,00000	-1,00000	-1,00000	-1,00000
0,1	-0,90083	-0,90000	-0,90023	-0,90023	-0,90016
0,2	-0,80316	-0,80095	-0,80132	-0,80132	-0,80121
0,3	-0,70753	-0,70357	-0,70402	-0,70401	-0,70388
0,4	-0,61440	-0,60844	-0,60893	-0,60890	-0,60877
0,5	-0,52408	-0,51601	-0,51650	-0,51645	-0,51632
0,6	-0,43684	-0,42663	-0,42709	-0,42702	-0,42691
0,7	-0,35287	-0,34054	-0,34095	-0,34087	-0,34078
0,8	-0,27229	-0,25794	-0,25828	-0,25818	-0,25812
0,9	-0,19518	-0,17894	-0,17920	-0,17908	-0,17905
1,0	-0,12158	-0,10357	-0,10378	-0,10364	-0,10364
$\max \varepsilon_i$	0,01958133	0,00032586	0,00012145	0,00003005	—

Таблица 6.6

i	$\hat{y}_i \left(h\mu = -\frac{1}{8} \right)$	Ошибка	$\hat{y}_i \left(h\mu = -\frac{100}{8} \right)$	Ошибка
0	-0,01562500	0,00000000	-0,01562500	0,00000000
1	-0,001953125	0,00000000	-0,001953125	0,00000000
2	0,00000000	0,00000000	0,00000000	0,00000000
3	0,00195317	0,00000004	0,00195317	0,00000004
4	0,01562501	0,00000001	0,01562409	0,00000091
5	0,05273429	0,00000009	0,05275568	0,00002130
6	0,12499996	0,00000004	0,12449656	0,00050437
7	0,24414058	0,00000005	0,25408001	0,00993938
8	0,42187492	0,00000008	0,18516620	0,23670880
9	0,66992193	0,00000005	6,27264466	5,60272278

Из анализа решения, приведенного в таблице 6.6, видно, что в случае $h\mu = -\frac{100}{8}$ процесс не является устойчивым. Дадим обоснование этому факту. Запишем формулу (6.33) в виде (6.14):

$$\hat{y}_{i+1} - \hat{y}_i + h \cdot \left(-\frac{23}{12} f_i + \frac{16}{12} f_{i-1} - \frac{5}{12} f_{i-2} \right) = 0,$$

т. е. $\alpha_0 = 1$, $\alpha_1 = -1$, $\alpha_2 = \alpha_3 = 0$, $\beta_0 = 0$, $\beta_1 = -\frac{23}{12}$, $\beta_2 = \frac{16}{12}$, $\beta_3 = -\frac{5}{12}$, $k = 3$. Поэтому

$$\begin{aligned}\rho(\xi) &= \alpha_0 \xi^3 + \alpha_1 \xi^2 + \alpha_2 \xi + \alpha_3 = \xi^3 - \xi^2; \\ \sigma(\xi) &= \beta_0 \xi^3 + \beta_1 \xi^2 + \beta_2 \xi + \beta_3 = -\frac{1}{12}(23\xi^2 - 16\xi + 5).\end{aligned}$$

Составим уравнение (6.15) при каждом значении $h\mu$.

При $h\mu = -\frac{1}{8}$ все корни уравнения $\xi^3 - \frac{73}{96}\xi^2 - \frac{16}{96}\xi + \frac{5}{96} = 0$ лежат внутри круга единичного радиуса с центром в начале координат, поэтому процесс устойчив.

При $h\mu = -\frac{100}{8}$ один корень уравнения $\xi^3 + \frac{2204}{96}\xi^2 - \frac{1600}{96}\xi + \frac{500}{96} = 0$, равный $\frac{1}{3} \cdot \left(\frac{2204}{96}\right)$, очевидно не лежит внутри указанного круга. Поэтому процесс неустойчив. Эти выводы подтверждают результаты численного эксперимента. Ошибка вычисления значений \hat{y}_i при $h\mu = -\frac{100}{8}$ возрастает (см. правый столбец табл. 6.6). ■

6.2.3. ПРИНЦИП СОГЛАСОВАНИЯ С РАЗЛОЖЕНИЕМ ПО ФОРМУЛЕ ТЕЙЛORA

Рассмотрим задачу Коши (6.9) в общем случае на неравномерной сетке Ω_n . Для получения формулы нахождения последующего значения \hat{y}_{i+1} на шаблоне (x_i, x_{i+1}) разложим решение $y(x)$ при $x = x_{i+1}$ относительно точки x_i по формуле Тейлора на промежутке $x_i \leq x \leq x_{i+1}$:

$$y_{i+1} = y_i + h_{i+1} y'_i + \frac{1}{2} h_{i+1}^2 y''_i + \dots + \frac{1}{p!} h_{i+1}^{p+1} y_i^{(p)} + O(h_{i+1}^{p+1}). \quad (6.38)$$

Входящие в правую часть производные можно определить, дифференцируя уравнение (6.9) требуемое число раз:

$$y' = f(x, y), \quad y'' = \frac{d}{dx} f(x, y) = f_x + f_y \cdot y' = f_x + f \cdot f_y, \dots$$

Если функция $f(x, y)$ имеет p -е непрерывные производные по обоим аргументам, то в разложении (6.38) можно учесть слагаемые до $O(h_{i+1}^{p+1})$. Чем больше членов разложения используется для вычисления, тем точнее будет приближение и тем выше порядок схемы. Однако рассчитывать производные путем дифференцирования правой части уравнения (6.9) невыгодно.

Поэтому Рунге и Кутта независимо друг от друга предложили более удобную и рациональную форму представления численного решения, при использовании которой дифференцировать функцию $f(x, y)$ не требуется.

В общем случае *методы Рунге — Кутты* имеют следующую структуру:

$$\hat{y}_{i+1} = \hat{y}_i + h_{i+1} \cdot [b_1 K_{1,i} + b_2 K_{2,i} + \dots + b_s K_{s,i}], \quad \hat{y}_0 = y_0, \quad i = \overline{0, n-1},$$

где

$$\begin{aligned} K_{1,i} &= f(x_i, \hat{y}_i); \\ K_{2,i} &= f(x_i + c_2 h_{i+1}, \hat{y}_i + h_{i+1} a_{2,1} K_{1,i}); \\ K_{3,i} &= f(x_i + c_3 h_{i+1}, \hat{y}_i + h_{i+1} (a_{3,1} K_{1,i} + a_{3,2} K_{2,i})); \\ &\vdots \\ K_{s,i} &= f(x_i + c_s h_{i+1}, \hat{y}_i + h_{i+1} (a_{s,1} K_{1,i} + a_{s,2} K_{2,i} + \dots + a_{s,s-1} K_{s-1,i})), \end{aligned} \quad (6.39)$$

где s — число *стадий* (этапов); $K_{s,i}$ — значения коэффициентов схемы Рунге — Кутты, вычисленные на основе правой части уравнения (6.9), $c_j, j = 2, s$; $a_{l,m}, l = 2, s; m = 1, s-1$; $b_k, k = 1, s$. Первый индекс в обозначениях коэффициентов является порядковым номером, а второй соответствует индексу точки x_i — началу отрезка $[x_i, x_{i+1}]$, на котором производится расчет.

В некоторых методах кроме вычисления приближенного решения \hat{y}_{i+1} определяется еще дополнительное значение \tilde{y}_{i+1} по формуле

$$\tilde{y}_{i+1} = \hat{y}_i + h_{i+1} \cdot [\tilde{b}_1 K_{1,i} + \tilde{b}_2 K_{2,i} + \dots + \tilde{b}_s K_{s,i}], \quad (6.40)$$

порядок которого, как правило, на единицу больше или меньше обеспечивающего выражением для \hat{y}_{i+1} . Величина $|\hat{y}_{i+1} - \tilde{y}_{i+1}|$ служит для учета погрешности и управления величиной шага.

Коэффициенты схемы подбираются так, чтобы выражение для \hat{y}_{i+1} согласовывалось с формулой Тейлора (6.38) вплоть до членов степени h_{i+1}^p , где степень p меняется в зависимости от метода и соответствует порядку схемы.

Наибольшее распространение в вычислительной практике нашел *метод Рунге — Кутты четвертого порядка*:

$$\hat{y}_{i+1} = \hat{y}_i + \frac{h_{i+1}}{6} (K_{1,i} + 2K_{2,i} + 2K_{3,i} + K_{4,i}), \quad \hat{y}_0 = y_0, \quad i = \overline{0, n-1}, \quad (6.41)$$

где

$$\begin{aligned} K_{1,i} &= f_i = f(x_i, \hat{y}_i); \quad K_{2,i} = f\left(x_i + \frac{h_{i+1}}{2}, \hat{y}_i + \frac{h_{i+1}}{2} K_{1,i}\right); \\ K_{3,i} &= f\left(x_i + \frac{h_{i+1}}{2}, \hat{y}_i + \frac{h_{i+1}}{2} K_{2,i}\right); \quad K_{4,i} = f(x_i + h_{i+1}, \hat{y}_i + h_{i+1} \cdot K_{3,i}). \end{aligned}$$

Схема (6.41) является четырехчленной, первый коэффициент $K_{1,i}$ относится к точке x_i , второй и третий — к средней точке $x_i + \frac{h_{i+1}}{2}$, четвертый — к точке x_{i+1} . Для этой схемы выбираются следующие параметры, входящие в (6.39): $s = 4$, $c_2 = c_3 = \frac{1}{2}$; $c_4 = 1$; $a_{2,1} = \frac{1}{2}$; $a_{3,1} = a_{4,1} = a_{4,2} = 0$; $a_{3,2} = \frac{1}{2}$; $a_{4,3} = 1$.

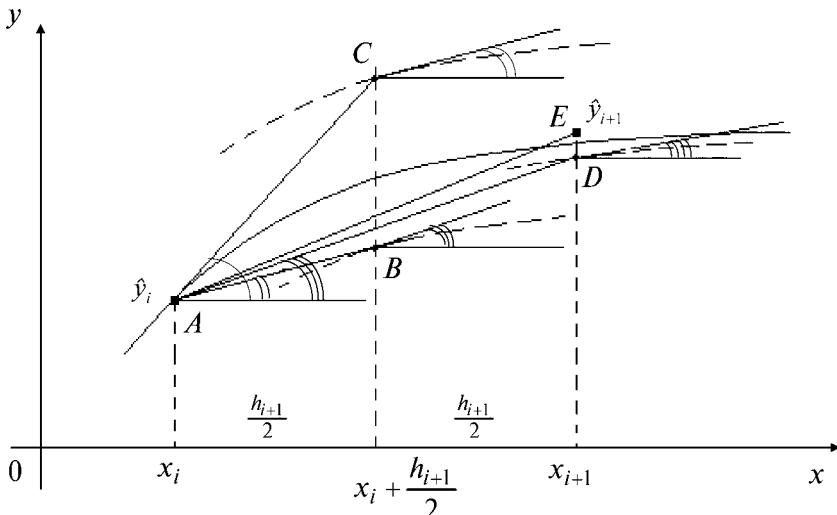


Рис. 6.13

Приведем геометрическую интерпретацию метода Рунге — Кутты четвертого порядка (рис. 6.13). Пусть известна точка (x_i, \hat{y}_i) на искомой интегральной кривой $y = y(x)$. Прямая AE , проходящая через точку (x_i, \hat{y}_i) и определяющая величину \hat{y}_{i+1} , имеет угловой коэффициент. Этот коэффициент рассчитывается по угловым коэффициентам касательных в точках A, B, C, D , проведенным к проходящим через эти точки интегральным кривым (штриховые линии). При этом коэффициент $K_{1,i}$ рассчитывается в точке A , коэффициент $K_{2,i}$ — в точке C , коэффициент $K_{3,i}$ — в точке B , а коэффициент $K_{4,i}$ — в точке D .

Замечания

1. Схемы Рунге — Кутты легко обобщаются на системы дифференциальных уравнений. Так, применительно к системе двух уравнений

$$\begin{aligned} y'(x) &= f(x, y, z), \quad y(x_0) = y_0, \\ z'(x) &= g(x, y, z), \quad z(x_0) = z_0 \end{aligned}$$

схема Рунге — Кутты четвертого порядка имеет вид

$$\begin{aligned} \hat{y}_{i+1} &= \hat{y}_i + \frac{h_{i+1}}{6}(K_{1,i} + 2K_{2,i} + 2K_{3,i} + K_{4,i}), \quad \hat{y}_0 = y_0, \quad i = \overline{0, n-1}, \\ \hat{z}_{i+1} &= \hat{z}_i + \frac{h_{i+1}}{6}(q_{1,i} + 2q_{2,i} + 2q_{3,i} + q_{4,i}), \quad \hat{z}_0 = z_0, \end{aligned} \tag{6.42}$$

где

$$\begin{aligned} K_{1,i} &= f(x_i, \hat{y}_i, \hat{z}_i); \\ q_{1,i} &= g(x_i, \hat{y}_i, \hat{z}_i); \\ K_{2,i} &= f\left(x_i + \frac{h_{i+1}}{2}, \hat{y}_i + \frac{1}{2}h_{i+1}K_{1,i}, \hat{z}_i + \frac{1}{2}h_{i+1}q_{1,i}\right); \\ q_{2,i} &= g\left(x_i + \frac{h_{i+1}}{2}, \hat{y}_i + \frac{1}{2}h_{i+1}K_{1,i}, \hat{z}_i + \frac{1}{2}h_{i+1}q_{1,i}\right); \end{aligned}$$

$$K_{3,i} = f\left(x_i + \frac{h_{i+1}}{2}, \hat{y}_i + \frac{1}{2}h_{i+1}K_{2,i}, \hat{z}_i + \frac{1}{2}h_{i+1}q_{2,i}\right);$$

$$q_{3,i} = g\left(x_i + \frac{h_{i+1}}{2}, \hat{y}_i + \frac{1}{2}h_{i+1}K_{2,i}, \hat{z}_i + \frac{1}{2}h_{i+1}q_{2,i}\right);$$

$$K_{4,i} = f(x_i + h_{i+1}, \hat{y}_i + h_{i+1}K_{3,i}, \hat{z}_i + h_{i+1}q_{3,i});$$

$$q_{4,i} = g(x_i + h_{i+1}, \hat{y}_i + h_{i+1}K_{3,i}, \hat{z}_i + h_{i+1}q_{3,i}).$$

2. Соотношения модифицированного метода Эйлера (6.28) можно записать в форме схемы Рунге — Кутты:

$$\hat{y}_{i+1} = \hat{y}_i + h_{i+1}K_{2,i}, \quad i = \overline{0, n-1}, \quad \hat{y}_0 = y_0;$$

$$K_{1,i} = f(x_i, \hat{y}_i); \quad K_{2,i} = f\left(x_i + \frac{h_{i+1}}{2}, \hat{y}_i + \frac{h_{i+1}}{2}K_{1,i}\right). \quad (6.43)$$

Покажем, что эта схема второго порядка. Разложение (6.38) с учетом равенства $y'' = f_x + f_i \cdot f_y$, где $f_i = f(x_i, y_i)$, частные производные f_x, f_y и вторая производная y'' вычисляются в точке (x_i, y_i) , принимает вид

$$y_{i+1} = y_i + h_{i+1}f_i + \frac{1}{2}h_{i+1}^2(f_x + f_i \cdot f_y) + O(h_{i+1}^3). \quad (6.44)$$

Разложение функции $f(x, y)$ относительно точки (x_i, y_i) можно записать следующим образом:

$$f(x, y) = f(x_i, y_i) + (x - x_i)f_x + (y - y_i)f_y + \dots, \quad (6.45)$$

где частные производные f_x, f_y вычисляются в точке (x_i, y_i) .

Тогда при $x = x_i + \frac{h_{i+1}}{2}$, $y = y_i + \frac{h_{i+1}}{2}f_i$ получаем

$$f\left(x_i + \frac{h_{i+1}}{2}, y_i + \frac{h_{i+1}}{2}f_i\right) = f_i + \frac{h_{i+1}}{2}f_x + \frac{h_{i+1}}{2}f_i \cdot f_y + O(h_{i+1}^2).$$

Следовательно, формулу (6.43) можно переписать в виде

$$\begin{aligned} \hat{y}_{i+1} &= \hat{y}_i + h_{i+1}f\left(x_i + \frac{h_{i+1}}{2}, \hat{y}_i + \frac{h_{i+1}}{2}f(x_i, \hat{y}_i)\right) = \\ &= \hat{y}_i + h_{i+1}f_i + \frac{1}{2}h_{i+1}^2(f_x + f_i \cdot f_y) + O(h_{i+1}^3). \end{aligned}$$

Сравнивая последнее соотношение с (6.44), можно сделать вывод о том, что схема модифицированного метода Эйлера согласуется с разложением по формуле Тейлора вплоть до членов степени h_{i+1}^2 и поэтому является методом Рунге — Кутты второго порядка.

3. Коэффициенты семейства методов Рунге — Кутты (6.39), (6.40) удобно записывать в виде таблицы 6.7. Таблица 6.8 соответствует классическому методу Рунге — Кутты четвертого порядка (6.41), а в таблице 6.9 приведены коэффициенты другого метода четвертого порядка (правило $\frac{3}{8}$).

Таблица 6.10 соответствует методу Рунге — Кутты (6.43) второго порядка, таблицы 6.11 и 6.12 — методам третьего порядка.

В таблице 6.13 даны коэффициенты *метода Фельберга четвертого порядка*, где формула для \hat{y}_{i+1} имеет пятый порядок, а в таблице 6.14 — семистадийного *метода Батчера шестого порядка*. В таблице 6.15 приведены коэффициенты *метода Рунге — Кутты — Мерсона четвертого порядка*, где формула для \hat{y}_{i+1} имеет пятый порядок для систем линейных дифференциальных уравнений с постоянными коэффициентами и третий порядок для нелинейных систем.

Параметры *метода Рунге — Кутты — Дормана — Принса пятого порядка* приведены в таблице 6.16. *Метод Фельберга седьмого порядка* и *метод Дормана — Принса восьмого порядка* изложены, например, в [55].

Таблица 6.7				Таблица 6.8				Таблица 6.9			
0				0				0			
c_2	$a_{2,1}$			$\frac{1}{2}$	$\frac{1}{2}$			$\frac{1}{3}$	$\frac{1}{3}$		
c_3	$a_{3,1}$	$a_{3,2}$		$\frac{1}{2}$	0	$\frac{1}{2}$		$\frac{2}{3}$	$-\frac{1}{3}$	1	
...	1	0	0	1	1	1	-1	1
c_s	$a_{s,1}$	$a_{s,2}$...					\hat{y}_{i+1}	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$
\hat{y}_{i+1}	B_1	b_2	...								
	\tilde{b}_1	\tilde{b}_2	...								
\tilde{y}_{i+1}				\hat{y}_{i+1}	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{2}{6}$	$\frac{1}{6}$			

Таблица 6.10				Таблица 6.11				Таблица 6.12			
0				0				0			
$\frac{1}{2}$		$\frac{1}{2}$		$\frac{1}{2}$	$\frac{1}{2}$			$\frac{1}{3}$	$\frac{1}{3}$		
				1	-1	2		$\frac{2}{3}$	0	$\frac{2}{3}$	
				\hat{y}_{i+1}	$\frac{1}{6}$	$\frac{2}{3}$	$\frac{1}{6}$				
\hat{y}_{i+1}		0	1					\hat{y}_{i+1}	$\frac{1}{4}$	0	$\frac{3}{4}$

Таблица 6.13						
0						
$\frac{1}{4}$		$\frac{1}{4}$				
$\frac{3}{8}$		$\frac{3}{32}$	$\frac{9}{32}$			
$\frac{12}{13}$	$\frac{1932}{2197}$		$\frac{7200}{2197}$	$\frac{7296}{2197}$		
1	$\frac{439}{216}$		-8	$\frac{3680}{513}$	$-\frac{845}{4104}$	
$\frac{1}{2}$	$-\frac{8}{27}$	2		$\frac{3544}{2565}$	$\frac{1859}{4104}$	$-\frac{11}{40}$
\hat{y}_{i+1}	$\frac{25}{216}$	0		$\frac{1408}{2565}$	$\frac{2197}{4104}$	$-\frac{1}{5}$
\tilde{y}_{i+1}	$\frac{16}{135}$	0		$\frac{6656}{12825}$	$\frac{28561}{56430}$	$-\frac{9}{50}$

0							
$\frac{1}{2}$	$\frac{1}{2}$						
$\frac{2}{3}$	$\frac{2}{9}$	$\frac{4}{9}$					
$\frac{1}{3}$	$\frac{7}{36}$	$\frac{2}{9}$	$-\frac{1}{12}$				
$\frac{5}{6}$	$-\frac{35}{144}$	$-\frac{55}{36}$	$\frac{35}{48}$	$\frac{15}{8}$			
$\frac{1}{6}$	$-\frac{1}{360}$	$-\frac{11}{36}$	$-\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{10}$		
1	$-\frac{41}{260}$	$\frac{22}{13}$	$\frac{43}{156}$	$-\frac{118}{39}$	$\frac{32}{195}$	$\frac{80}{39}$	
\hat{y}_{i+1}	$\frac{13}{200}$	0	$\frac{11}{40}$	$\frac{11}{40}$	$\frac{4}{25}$	$\frac{4}{25}$	$\frac{13}{200}$

0							
$\frac{1}{3}$		$\frac{1}{3}$					
$\frac{1}{3}$		$\frac{1}{6}$	$\frac{1}{6}$				
$\frac{1}{2}$		$\frac{1}{8}$	0	$\frac{3}{8}$			
1		$\frac{1}{2}$	0	$-\frac{3}{2}$	2		
\hat{y}_{i+1}		$\frac{1}{6}$	0	0	$\frac{2}{3}$	$\frac{1}{6}$	
\tilde{y}_{i+1}		$\frac{1}{10}$	0	$\frac{3}{10}$	$\frac{2}{5}$	$\frac{1}{5}$	

0							
$\frac{1}{5}$	$\frac{1}{5}$						
$\frac{3}{10}$	$\frac{3}{40}$	$\frac{9}{40}$					
$\frac{4}{5}$	$\frac{44}{45}$	$-\frac{56}{15}$	$\frac{32}{9}$				
$\frac{8}{9}$	$\frac{19372}{6561}$	$-\frac{25360}{2187}$	$\frac{64448}{6561}$	$-\frac{212}{729}$			
1	$\frac{9017}{3168}$	$-\frac{355}{33}$	$\frac{46732}{5247}$	$\frac{49}{176}$	$-\frac{5103}{18656}$		
1	$\frac{35}{384}$	0	$\frac{500}{1113}$	$\frac{125}{192}$	$-\frac{2187}{6784}$	$\frac{11}{84}$	
\hat{y}_{i+1}	$\frac{35}{384}$	0	$\frac{500}{1113}$	$\frac{125}{192}$	$-\frac{2187}{6784}$	$\frac{11}{84}$	0
\tilde{y}_{i+1}	$\frac{5179}{57600}$	0	$\frac{7571}{16695}$	$\frac{393}{640}$	$-\frac{92097}{339200}$	$\frac{187}{2100}$	$\frac{1}{40}$

4. Для s -стадийных методов Рунге — Кутты справедливо следующее условие устойчивости, определяемое для задачи (6.12):

$$\left| 1 + h\mu + \frac{(h\mu)^2}{2} + \frac{(h\mu)^3}{6} + \frac{(h\mu)^4}{24} + \dots + \frac{(h\mu)^s}{s!} \right| < 1.$$

Если μ — действительное число, можно указать интервалы устойчивости: $h\mu \in (-2; 0)$ при $s = 1, 2$; $h\mu \in (-2, 51; 0)$ при $s = 3$; $h\mu \in (-2, 78; 0)$ при $s = 4$ и т. д.

Пример 6.6. Для задачи Коши

$$y' = x + y, \quad y(0) = 1, \quad x \in [0; 0,2]$$

выполнить два шага методом Рунге — Кутты второго порядка (6.43) и один (первый шаг) методом четвертого порядка (6.41). Шаг интегрирования принять постоянным и равным 0,1. Сравнить оба численных решения с точным $y(x) = 2e^x - x - 1$.

□ В данной задаче $x_0 = 0, y_0 = 1$. Применим схему (6.43).

Первый шаг ($i = 0$). В соответствии с постановкой задачи принимается $h_1 = 0,1$. Вычислим

$$\begin{aligned} K_{1,0} &= f(x_0, \hat{y}_0) = f(x_0, y_0) = x_0 + y_0 = 0 + 1 = 1; \\ K_{2,0} &= f\left(x_0 + \frac{h_1}{2}, y_0 + \frac{h_1}{2} K_{1,0}\right) = 0,05 + 1,05 = 1,1. \end{aligned}$$

Тогда $\hat{y}_1 = \hat{y}_0 + h_1 K_{2,0} = 1 + 0,1 \cdot 1,1 = 1,11$.

Второй шаг ($i = 1$). Так как шаг постоянный, то $h_2 = 0,1$, а $x_1 = 0,1$ и, следовательно,

$$\begin{aligned} K_{1,1} &= f(x_1, \hat{y}_1) = x_1 + \hat{y}_1 = 0,1 + 1,11 = 1,12; \\ K_{2,1} &= f\left(x_1 + \frac{h_2}{2}, \hat{y}_1 + \frac{h_2}{2} K_{1,1}\right) = 0,1 + \frac{0,1}{2} + 1,11 + 0,05 \cdot 1,12 = 1,316. \end{aligned}$$

Тогда $\hat{y}_2 = \hat{y}_1 + h_2 K_{2,1} = 1,11 + 0,1 \cdot 1,316 = 1,2416$. Точное решение: $y(x_2) = 1,2428055$. Отличие приближенного решения от точного составляет 0,096%.

Выполним теперь первый шаг методом четвертого порядка по формуле (6.41), записанной при $i = 0$:

$$\hat{y}_1 = \hat{y}_0 + \frac{h_1}{6} (K_{1,0} + 2K_{2,0} + 2K_{3,0} + K_{4,0}).$$

При этом вычислим

$$K_{1,0} = x_0 + y_0 = 0 + 1 = 1;$$

$$K_{2,0} = \left(x_0 + \frac{h_1}{2}\right) + \left(\hat{y}_0 + \frac{h_1}{2} \cdot K_{1,0}\right) = 0,05 + (1 + 0,05 \cdot 1) = 1,1;$$

$$K_{3,0} = \left(x_0 + \frac{h_1}{2}\right) + \left(\hat{y}_0 + \frac{h_1}{2} \cdot K_{2,0}\right) = 0,05 + (1 + 0,05 \cdot 1) = 1,105;$$

$$K_{4,0} = (x_0 + h_1) + (\hat{y}_0 + h_1 \cdot K_{3,0}) = 0,1 + (1 + 0,1 \cdot 1,105) = 1,2105;$$

$$\hat{y}_1 = 1 + \frac{0,1}{6}(1 + 2 \cdot 1,1 + 2 \cdot 1,105 + 1,2105) = 1,1103416.$$

Точное решение: $y(0,1) = 1,1103418$.

Сопоставляя точное решения $y(x_1)$ в точке x_1 с результатами \hat{y}_1 , полученными по схемам второго и четвертого порядка, можно увидеть, что для схемы четвертого порядка результат \hat{y}_1 совпадает с точным решением в семи цифрах, а для схемы второго порядка — в трех знаках. ■

В заключение данного пункта отметим преимущества и недостатки одношаговых и многошаговых схем.

1. Одношаговые схемы позволяют достаточно легко строить алгоритмы с переменным шагом и заданной точностью, что необходимо при расчете функций в областях их больших изменений. Однако они имеют недостаток — приближенные значения искомой функции, рассчитанные на предыдущих шагах, в этих схемах не используются.

2. В многошаговых схемах, напротив, учитываются значения искомой функции, рассчитанные на предыдущих шагах интегрирования, что повышает порядок точности этих схем без использования дополнительных точек на отрезке $[x_i, x_{i+1}]$. Однако эти схемы требуют усложнения расчетных алгоритмов при реализации переменного шага.

Замечание. В явных одношаговых методах обеспечить заданную точность на каждом очередном шаге из текущей точки (x_i, \hat{y}_i) можно по следующей методике.

1. Задать начальную величину шага $h^{(0)} = x_{i+1}^{(0)} - x_i$ и рассчитать значение $\hat{y}_{i+1}^{(0)}$ в точке $x_{i+1}^{(0)}$.

2. Вычислить величину $\hat{y}_{i+1}^{(1)}$ в точке $x_{i+1}^{(0)}$ с половинным значением шага $h^{(1)} = \frac{h^{(0)}}{2}$ (до достижения точки $x_{i+1}^{(0)}$ выполняются два шага).

3. Если $|\hat{y}_{i+1}^{(1)} - \hat{y}_{i+1}^{(0)}| < \varepsilon$, то перейти к нахождению приближенного решения в точке $x_{i+2}^{(0)} = x_{i+1}^{(0)} + h^{(0)}$. Иначе осуществить еще одно дробление шага, т. е. вычислить $h^{(2)} = \frac{h^{(1)}}{2}$ и найти $\hat{y}_{i+1}^{(2)}$ в точке $x_{i+1}^{(1)} = x_i + h^{(1)}$.

4. Сравнить решения, полученные с шагом $h^{(1)}$ и $h^{(2)}$ в точке $x_{i+1}^{(1)}$. Если желаемая точность достигнута, расчеты продолжить с шагом $h^{(1)}$, а если нет, то процесс дробления шага продолжить до достижения заданной точности.

6.2.4. ПРИНЦИП АНАЛОГИЙ

Данный метод следует из обобщения интегрально-интерполяционного принципа. Он основан на применении интерполяционных интегрально-дифференциальных сплайн-функций, преобразования которых — параметрические соотношения, аппроксимационные формулы для производных и квадратурные формулы — используются для получения разностных схем решения задачи Коши [15]–[24]. Приведем некоторые из этих соотношений, соответствующих задаче аппроксимации некоторой произвольной функции $\Phi(x)$ многочленами второй степени:

а) параметрическое соотношение, связывающее интерполируемую функцию $\Phi(x)$ и ее производную $\bar{m}_i = \Phi'(x)|_{x=x_i}$:

$$\frac{\bar{m}_i}{h_{i+1}} - \frac{\bar{m}_{i-1}}{h_i} = \frac{\Delta\Phi_i}{h_{i+1}^2} - \frac{\Delta\Phi_{i-1}}{h_i^2}; \quad (6.46)$$

б) формула аппроксимации первой производной в точке x_{i+1} на трехточечном нерегулярном шаблоне (x_{i-1}, x_i, x_{i+1}) (см. первую формулу в (5.9)):

$$\bar{m}_{i+1} = \frac{1}{H_i^{i+1}} \left[\delta_{i+1} \Phi_{i-1} - \frac{(H_i^{i+1})^2}{\Pi_i^{i+1}} \Phi_i + \frac{H_i^{2(i+1)}}{h_{i+1}} \Phi_{i+1} \right] \left(\frac{h_i^2}{6} \delta_{i+1} (1 + \delta_{i+1}) M_{3,i} \right), \quad (6.47)$$

где $H_{si}^{p(i+1)} = s \cdot h_i + p \cdot h_{i+1}$; s, p — целые числа ($H_{si}^{p(i+1)} = h_i(s + p \cdot \delta_{i+1})$, $\delta_{i+1} = \frac{h_{i+1}}{h_i}$);

в) квадратурная формула экстраполяционного типа, имеющая четвертый порядок:

$$I_i^{i+1} = \frac{h_{i+1}^2}{6H_{i-1}^i} \left[\frac{H_{3i}^{2(i+1)}}{h_{i-1}} \Phi_{i-2} - \frac{3(H_{i-1}^i)^2 + 2h_{i+1}H_{i-1}^i}{\Pi_{i-1}^i} \Phi_{i-1} + \frac{6h_i H_{i-1}^i + h_{i+1}(3h_{i-1} + 6h_i + 2h_{i+1})}{\Pi_i^{i+1}} \Phi_i \right]. \quad (6.48)$$

Эта формула при $h_{i+1} = h = \text{const}$ упрощается:

$$I_i^{i+1} = \frac{h}{12} (5\Phi_{i-2} - 16\Phi_{i-1} + 23\Phi_i) \left(\frac{3}{8} h^4 M_{3,i} \right);$$

г) двухинтервальная (трехточечная) обобщенная на нерегулярный шаблон квадратурная формула парабол четвертого порядка на отрезке $[x_{i-1}, x_{i+1}]$, выражающая сумму «удельных» значений интегралов через линейную комбинацию значений $\Phi_{i-1}, \Phi_i, \Phi_{i+1}$:

$$3 \left(\frac{1}{h_i^2} I_{i-1}^i + \frac{1}{h_{i+1}^2} I_i^{i+1} \right) = \frac{1}{h_i} \Phi_{i-1} + 2 \left(\frac{1}{h_i} + \frac{1}{h_{i+1}} \right) \Phi_i + \frac{1}{h_{i+1}} \Phi_{i+1}; \quad (6.49)$$

д) одноинтервальная (трехточечная) функциональная квадратурная формула четвертого порядка на отрезке $[x_{i-1}, x_{i+1}]$

$$I_i^{i+1} = \frac{h_{i+1}^3}{6H_i^{i+1}} \left[-\frac{1}{h_i} \Phi_{i-1} + \frac{H_i^{i+1} H_{3i}^{i+1}}{\Pi_i^{i+1} h_{i+1}} \Phi_i + \frac{H_{3i}^{2(i+1)}}{h_{i+1}^2} \Phi_{i+1} \right] \left(\frac{h_i^2 h_{i+1}^2}{72} \delta_{n+1} (\delta_{n+1} + 2) \cdot M_{3,i} \right), \quad (6.50)$$

которая при $h_{i+1} = h = \text{const}$ преобразуется к более простому виду:

$$I_i^{i+1} = \frac{h}{12} (-\Phi_{i-1} + 8\Phi_i + 5\Phi_{i+1}).$$

Анализ остаточного слагаемого, получающегося при оценке погрешности квадратурной формулы (6.50) (см. мажоранту в скобках справа от дан-

ной формулы), указывает на то, что при сгущающейся вправо сетке (условная аппроксимация интеграла) порядок точности формулы может быть повышен на единицу без увеличения количества точек шаблона. Действительно, при $\delta_{i+1} < h_i$ легко получить приближенное условие повышения порядка [18]: $h_{i+1} \leq h_i^2$. Таким образом, в данном случае порядок повышается за счет значительного уменьшения шага, такого что h_{i+1} не превышает квадрата шага h_i . Этот факт может быть использован при решении задачи Коши на нерегулярном шаблоне;

е) две односторонние (двуточечные) квадратурные формулы функционально-дифференциального типа:

$$I_i^{i+1} = \frac{h_{i+1}}{3} (\Phi_i + 2\Phi_{i+1}) - \frac{h_{i+1}^2}{6} \Phi'_{i+1} \quad \left(\frac{5}{24} h_i^4 M_{3,i} \right); \quad (6.51)$$

$$I_i^{i+1} = h_{i+1} \Phi_i + \frac{h_{i+1}^2}{6} (2\Phi'_i + \Phi'_{i+1}) \quad \left(\frac{1}{24} h_i^4 M_{3,i} \right); \quad (6.52)$$

ж) интегрально-дифференциальное параметрическое соотношение, осуществляющее связь интегралов и производных:

$$\frac{I_i^{i+1} - I_{i-1}^i}{h_{i+1}} = \frac{1}{6} (h_{i+1} \bar{m}_{i+1} + 2H_i^{i+1} \cdot \bar{m}_i + h_i \cdot \bar{m}_{i-1}). \quad (6.53)$$

Дискретные явные или неявные схемы для решения задачи (6.9) получаются из приведенных соотношений либо посредством аппроксимации производной (см. п. 6.2.1), либо путем соответствующей замены одних параметров, относящихся к параметрическому соотношению или к квадратурной формуле, на другие параметры. Этот способ построения схем называется *способом аналогий*.

Так, заменяя в (6.46) $\Delta\Phi_k$ на $\Delta\hat{y}_k$, а \bar{m}_k на f_k ($k = i, i-1$) и разрешая относительно \hat{y}_{i+1} , получаем *двуточечную явную схему второго порядка (2Я2А)*:

$$\hat{y}_{i+1} = \hat{y}_i + \delta_{i+1}^2 \Delta\hat{y}_{i-1} + h_{i+1}^2 \quad \left(\frac{f_i}{h_{i+1}} - \frac{f_{i-1}}{h_i} \right), \quad (6.54)$$

которая при $h_{i+1} = h = \text{const}$ преобразуется к виду

$$\hat{y}_{i+1} = 2\hat{y}_i - \hat{y}_{i-1} + h(f_i - f_{i-1}).$$

Комбинация схем (6.54) и (6.24) дает *обобщенную на нерегулярный шаблон схему Адамса — Бэшфорта (2Я2В)*:

$$\hat{y}_{i+1} = \hat{y}_i + \frac{h_{i+1}^2}{2} \left(\frac{H_{2i}^{i+1}}{\Pi_{i+1}^{i+1}} f_i - \frac{1}{h_i} f_{i-1} \right), \quad (6.55)$$

которая при $h_{i+1} = h = \text{const}$ преобразуется к виду (6.32):

$$\hat{y}_{i+1} = \hat{y}_i + \frac{h}{2} (3f_i - f_{i-1}).$$

Аппроксимационная формула (6.47) определяет *неявную двухшаговую схему второго порядка* (2НЯ2):

$$\hat{y}_{i+1} = \frac{h_{i+1}}{H_i^{2(i+1)}} \left[\frac{(H_i^{i+1})^2}{\Pi_i^{i+1}} \hat{y}_i - \delta_{i+1} \hat{y}_{i-1} + H_i^{i+1} f(x_{i+1}, \hat{y}_{i+1}) \right], \quad (6.56)$$

которая при $h_{i+1} = h = \text{const}$ принимает вид

$$\hat{y}_{i+1} = -\frac{1}{3} \hat{y}_{i-1} + \frac{4}{3} \hat{y}_i + \frac{2}{3} h f(x_{i+1}, \hat{y}_{i+1}). \quad (6.57)$$

Замечание. Схема (6.57) является одним из методов Гира, получаемых на основе формул дифференцирования назад, связывающих значение производной в точке x_{i+1} со значениями функции в предыдущих точках x_i, \dots, x_{i-k+1} (см. рис. 6.2). Схема (6.57) является *схемой второго порядка*, а неявный метод Эйлера — схемой первого порядка. Приведем более точные многошаговые схемы, записанные на регулярных шаблонах:

- *схема третьего порядка*

$$\hat{y}_{i+1} = \frac{18}{11} \hat{y}_i - \frac{9}{11} \hat{y}_{i-1} + \frac{2}{11} \hat{y}_{i-2} + \frac{6}{11} h f(x_{i+1}, \hat{y}_{i+1}); \quad (6.58)$$

- *схема четвертого порядка*

$$\hat{y}_{i+1} = \frac{48}{25} \hat{y}_i - \frac{36}{25} \hat{y}_{i-1} + \frac{16}{25} \hat{y}_{i-2} - \frac{3}{25} \hat{y}_{i-3} + \frac{12}{25} h f(x_{i+1}, \hat{y}_{i+1});$$

- *схема пятого порядка*

$$\hat{y}_{i+1} = \frac{300}{137} \hat{y}_i - \frac{300}{137} \hat{y}_{i-1} + \frac{200}{137} \hat{y}_{i-2} - \frac{75}{137} \hat{y}_{i-3} + \frac{12}{137} \hat{y}_{i-4} + \frac{60}{137} h f(x_{i+1}, \hat{y}_{i+1});$$

- *схема шестого порядка*

$$\hat{y}_{i+1} = \frac{360}{147} \hat{y}_i - \frac{450}{147} \hat{y}_{i-1} + \frac{400}{147} \hat{y}_{i-2} - \frac{225}{147} \hat{y}_{i-3} + \frac{17}{147} \hat{y}_{i-4} - \frac{10}{147} \hat{y}_{i-5} + \frac{60}{147} h f(x_{i+1}, \hat{y}_{i+1}).$$

Известно, что схема (6.58) является $A(\alpha)$ -устойчивой с $\alpha \approx 68^\circ$.

Все рассматриваемые ниже схемы получаются из квадратурных формул (6.48)–(6.52), в которых вместо разности первообразных $F_{i+1} - F_i = I_i^{i+1}$ подставляется $\Delta \hat{y}_i$, а вместо Φ_k ($k = i-2, i-1, i$) и Φ'_k подставляются f_k и f'_k . При этом порядки точности схем на отрезке $[a, b]$ соответствуют порядкам точности вычисления интегралов на всем отрезке интегрирования.

Преобразуя таким образом квадратурную формулу (6.48), получаем *явную трехшаговую схему третьего порядка* (3Я3):

$$\begin{aligned} \hat{y}_{i+1} = & \hat{y}_i + \frac{h_{i+1}^2}{6H_{i-1}^i} \left[\frac{H_{3i}^{2(i+1)}}{h_{i-1}} f_{i-2} - \frac{3(H_{i-1}^i)^2 + 2h_{i+1}H_{i-1}^i}{\Pi_{i-1}^i} f_{i-1} + \right. \\ & \left. + \frac{6h_i H_{i-1}^i + h_{i+1}(H_{3(i-1)}^{6i} + 2h_{i+1})}{\Pi_i^{i+1}} f_i \right]. \end{aligned} \quad (6.59)$$

При $h_{i+1} = h = \text{const}$ она является традиционной, получающейся интегрально-интерполяционным методом (6.33):

$$\hat{y}_{i+1} = \hat{y}_i + \frac{h}{12}(5f_{i-2} - 16f_{i-1} + 23f_i).$$

Из обобщенной формулы (6.49) следует *первая двухшаговая неявная схема третьего порядка* (2НЯЗА)

$$\hat{y}_{i+1} = \hat{y}_i - \delta_{i+1}^2 \Delta \hat{y}_{i-1} + \frac{h_{i+1}^2}{3} \left[\frac{1}{h_i} f_{i-1} + 2 \cdot \left(\frac{1}{h_i} + \frac{1}{h_{i+1}} \right) f_i + \frac{1}{h_{i+1}} f(x_{i+1}, \hat{y}_{i+1}) \right], \quad (6.60)$$

которая при $h_{i+1} = h = \text{const}$ преобразуется к виду (6.29).

Одноинтервальная (трехточечная) квадратурная формула (6.50) определяет *вторую двухшаговую неявную схему третьего порядка* (2НЯЗБ)

$$\hat{y}_{i+1} = \hat{y}_i + \frac{h_{i+1}^3}{6H_i^{i+1}} \left(-\frac{1}{h_i} f_{i-1} + \frac{H_i^{i+1} H_{3i}^{i+1}}{\Pi_i^{i+1} h_{i+1}} f_i + \frac{H_{3i}^{2(i+1)}}{h_{i+1}^2} f(x_{i+1}, \hat{y}_{i+1}) \right). \quad (6.61)$$

Эта схема в соответствии с вышеприведенной оценкой точности интеграла при $\delta_{n+1} \leq h_i$ имеет не третий, а четвертый порядок точности. При $h = \text{const}$ эта схема преобразуется к виду (6.36):

$$\hat{y}_{i+1} = \hat{y}_i + \frac{h}{12}(-f_{i-1} + 8f_i + 5f(x_{i+1}, \hat{y}_{i+1})).$$

И наконец, последние приводимые здесь схемы следуют из квадратурных формул (6.51), (6.52) (*неявные одношаговые схемы*) и параметрического соотношения (6.53) (*неявная двухшаговая схема*):

$$\begin{aligned} \hat{y}_{i+1} &= \hat{y}_i + \frac{h_{i+1}}{3}[f_i + 2f(x_{i+1}, \hat{y}_{i+1})] - \frac{h_{i+1}^2}{6}f'(x_{i+1}, \hat{y}_{i+1}); \\ \hat{y}_{i+1} &= \hat{y}_i + h_{i+1}f_i + \frac{h_{i+1}^2}{6}[2f'_i + f'(x_{i+1}, \hat{y}_{i+1})]; \\ \hat{y}_{i+1} &= \hat{y}_i + \delta_{i+1}\Delta \hat{y}_{i-1} + \frac{h_{i+1}}{6}[h_i f'_{i-1} + 2H_i^{i+1} f'_i + h_{i+1} f'(x_{i+1}, \hat{y}_{i+1})], \end{aligned}$$

где $f' = f'_x + f'_y \cdot f$. Эти схемы обозначаются 1НЯЗА, 1НЯЗБ, 2НЯЗД соответственно.

При $h_{i+1} = h = \text{const}$ последняя схема преобразуется к виду

$$\hat{y}_{i+1} = -\hat{y}_{i-1} + 2\hat{y}_i + \frac{h^2}{6}[f'_{i-1} + 4f'_i + f'(x_{i+1}, \hat{y}_{i+1})].$$

Для применения разностных схем, полученных с помощью четырех различных подходов (см. п. 6.2.1–6.2.4), можно использовать общую методику.

Методика решения задачи Коши методами сеток

1. В соответствии с требуемой точностью решения задачи Коши выбрать численный метод (схему) определенного порядка аппроксимации и типа (явный или неявный, одношаговый или многошаговый).
2. Задать сетку Ω_n или выбрать ее параметр (h , если сетка равномерная; h_{i+1} , $i = \overline{0, n-1}$, если сетка неравномерная). Если выбранный метод является ограниченно устойчивым, то для задания параметра сетки оценить величину критического шага.
3. Если выбранный метод многошаговый, найти требуемое число «разгонных» точек.
4. Найти приближенное решение задачи Коши во всех точках сетки Ω_n . Если выбран явный метод, каждая следующая точка \hat{y}_{i+1} рассчитывается по явной формуле, а если неявный, то с помощью решения в общем случае нелинейного алгебраического уравнения (6.11) одним из численных методов.
5. Оценить достигнутую точность. В случае невыполнения требований повторить расчеты, меняя значение шага в соответствии в замечанием в конце п. 6.2.3 или используя более точный метод.

6.3. СОСТАВНЫЕ СХЕМЫ

Рассматриваемые здесь методы (схемы), называемые *составными*, известны под общим названием *методов прогноза и коррекции*. Из названия следует, что сначала «предсказывается» значение \hat{y}_{i+1} , а затем используется тот или иной метод для «корректировки» этого значения.

Таким образом, составные схемы включают в себя два шага (этапа) расчета очередного значения \hat{y}_{i+1} .

1. *Шаг «предиктор»* (предсказание), на котором рассчитывается предсказанное (предварительное) значение $\hat{y}_{i+1}^{(I)}$.

2. *Шаг «корректор»* (коррекция), на котором предсказанное значение уточняется. В результате находится значение $\hat{y}_{i+1}^{(K)}$, которое принимается за \hat{y}_{i+1} . Если промежуток интегрирования не исчерпан, оно далее используется при реализации очередного шага «предиктор» для нахождения следующего предсказанного значения $\hat{y}_{i+2}^{(I)}$.

Первый шаг реализуется с помощью явных методов, а второй шаг основан на применении формул неявных методов, в правую часть которых вместо неизвестного значения \hat{y}_{i+1} подставляется результат предсказания. Схемы такого типа называются также схемами «предиктор — корректор» и в итоге относятся к явным методам.

Приведем наиболее часто встречающиеся составные схемы.

А. Предсказание с помощью явного метода Эйлера (6.19) или метода Эйлера — Коши (6.21), коррекция по методу трапеций (6.24).

Шаг «предиктор»:

$$\hat{y}_{i+1}^{(I)} = \hat{y}_i + h_{i+1} f(x_i, \hat{y}_i), \quad (6.62)$$

или при условии $h_{i+1} = h = \text{const}$

$$\hat{y}_{i+1}^{(\Pi)} = \hat{y}_{i-1} + 2h \cdot f(x_i, \hat{y}_i), \quad (6.63)$$

где \hat{y}_i и \hat{y}_{i-1} рассчитаны на предыдущих шагах.

Шаг «корректор»:

$$\hat{y}_{i+1} \equiv \hat{y}_{i+1}^{(K)} = \hat{y}_i + \frac{h_{i+1}}{2} [f(x_i, \hat{y}_i) + f(x_i + h_{i+1}, \hat{y}_{i+1}^{(\Pi)})]. \quad (6.64)$$

Б. Предсказание по методу Адамса — Бэшфорта третьего (6.33) или четвертого (6.34) порядка, коррекция по методу Адамса — Мултона четвертого порядка (6.37) (при $h_{i+1} = h = \text{const}$).

Шаг «предиктор»:

$$\hat{y}_{i+1}^{(\Pi)} = \hat{y}_i + \frac{h}{12} [23f_i - 16f_{i-1} + 5f_{i-2}] \quad (6.65)$$

или

$$\hat{y}_{i+1}^{(\Pi)} = \hat{y}_i + \frac{h}{24} [55f_i - 59f_{i-1} + 37f_{i-2} - 9f_{i-3}]. \quad (6.66)$$

Шаг «корректор»:

$$\hat{y}_{i+1} \equiv \hat{y}_{i+1}^{(K)} = \hat{y}_i + \frac{h}{24} [f_{i-2} - 5f_{i-1} + 19f_i + 9f(x_{i+1}, \hat{y}_{i+1}^{(\Pi)})]. \quad (6.67)$$

В. Метод Хемминга четвертого порядка (при $h_{i+1} = h = \text{const}$).

Шаг «предиктор»:

$$\hat{y}_{i+1}^{(\Pi)} = \hat{y}_{i-3} + \frac{4h}{3} [2f_i - f_{i-1} + 2f_{i-2}]. \quad (6.68)$$

Шаг «корректор»:

$$\hat{y}_{i+1} \equiv \hat{y}_{i+1}^{(K)} = \frac{1}{8} (9\hat{y}_i - \hat{y}_{i-2}) + \frac{3h}{8} [-f_{i-1} + 2f_i + f(x_{i+1}, \hat{y}_{i+1}^{(\Pi)})]. \quad (6.69)$$

Замечания

1. Разработчикам прикладных программ рекомендуется при решении конкретных задач выбрать наиболее подходящий составной метод путем комбинирования схем на шагах «предиктор» и «корректор», соответствующих явным и неявным схемам. К числу этих схем относятся: П1К2 (1Я1, 1НЯ2), П2К2 (2Я2Б, 1НЯ2), П2К3 (2Я2Б, 2НЯЗА), П3К3 (3Я3, 2НЯЗА) и др. (табл. 6.17). Здесь буквы П и К указывают на шаги «предиктор» и «корректор», рядом

Таблица 6.17

Предикторы	Корректоры	Составная схема
1Я1 (6.19)	1НЯ2 (6.24)	П1К2 (6.62), (6.64)
2Я2Б (6.20)	1НЯ2 (6.24)	П2К2 (6.63), (6.64)
2Я2Б (6.20)	2НЯЗА (6.60)	П2К3
3Я3 (6.59)	2НЯЗА (6.60)	П3К3
3Я3А (6.33)	3НЯ4 (6.37)	П3К4 (6.65), (6.67)
4Я4 (6.34)	3НЯ4 (6.37)	П4К4 (6.66), (6.67)

с ними приведены порядки схем, а в скобках — ранее введенные обозначения формул для этих схем.

2. При использовании многошаговых схем (6.63), (6.65)–(6.69) на шаге «предиктор» и шаге «корректор» необходимо предварительно рассчитать требуемое число «разгонных» точек (см. п. 6.2.2). При этом желательно использовать схемы, порядок которых соответствует порядку основной схемы.

3. Составная схема (6.62), (6.64) является схемой второго порядка и одновременно относится к семейству схем Рунге — Кутты второго порядка (см. п. 6.2.3). Она может быть записана в форме

$$\begin{aligned} K_{1,i} &= f(x_i, \hat{y}_i); \\ K_{2,i} &= f(x_i + h_{i+1}, \hat{y}_i + h_{i+1}K_{1,i}); \\ \hat{y}_{i+1} &= \hat{y}_i + \frac{h_{i+1}}{2}(K_{1,i} + K_{2,i}). \end{aligned}$$

Кроме того, данная схема часто называется *исправленным методом Эйлера*, или *схемой Эйлера с пересчетом* [35]. Покажем, что эта схема второго порядка, хотя в ней используется предиктор (6.62) первого порядка. Как и при рассмотрении модифицированного метода Эйлера (6.43), получим разложение (6.45) функции $f(x, y)$ при $x = x_i + h_{i+1}$, $y = y_i + h_{i+1}f_i$:

$$f(x_i + h_{i+1}, y_i + h_{i+1}f_i) = f_i + h_{i+1}f_x + h_{i+1}f_i \cdot f_y + O(h_{i+1}^2),$$

где частные производные f_x , f_y вычисляются в точке (x_i, y_i) .

Следовательно,

$$\begin{aligned} \hat{y}_{i+1} &= \hat{y}_i + \frac{h_{i+1}}{2}[f_i + f(x_i + h_{i+1}, \hat{y}_i + h_{i+1}f_i(x_i, \hat{y}_i))] = \\ &= \hat{y}_i + \frac{h_{i+1}}{2}\left[f_i + f_i + h_{i+1}f_x + h_{i+1}f_i \cdot f_y + O(h_{i+1}^2)\right] = \\ &= \hat{y}_i + h_{i+1}f_i + \frac{1}{2}h_{i+1}^2(f_x + f_i \cdot f_y) + O(h_{i+1}^3). \end{aligned}$$

Сравнивая последнее соотношение с (6.38), можно сделать вывод о том, что схема исправленного метода Эйлера согласуется с разложением по формуле Тейлора вплоть до членов степени h_{i+1}^2 и поэтому является методом Рунге — Кутты второго порядка.

Приведем геометрическую интерпретацию схемы «предиктор — корректор» (6.62), (6.64) (рис. 6.14). Пусть известна точка (x_i, \hat{y}_i) на искомой интегральной кривой $y = y(x)$. Через эту точку проведем касательную L , тангенс угла наклона которой равен $y'(x_i) = f(x_i, \hat{y}_i)$. Ее уравнение имеет вид

$$y = \hat{y}_i + f(x_i, \hat{y}_i)(x - x_i).$$

В качестве результата предсказания берется значение $\hat{y}_{i+1}^{(\Pi)}$, получающееся при $x = x_{i+1}$: $\hat{y}_{i+1}^{(\Pi)} = \hat{y}_i + h_{i+1}f(x_i, \hat{y}_i)$. Далее через точку $(x_{i+1}, \hat{y}_{i+1}^{(\Pi)})$ к интегральной кривой проводится прямая L_1 с тангенсом угла наклона, равным

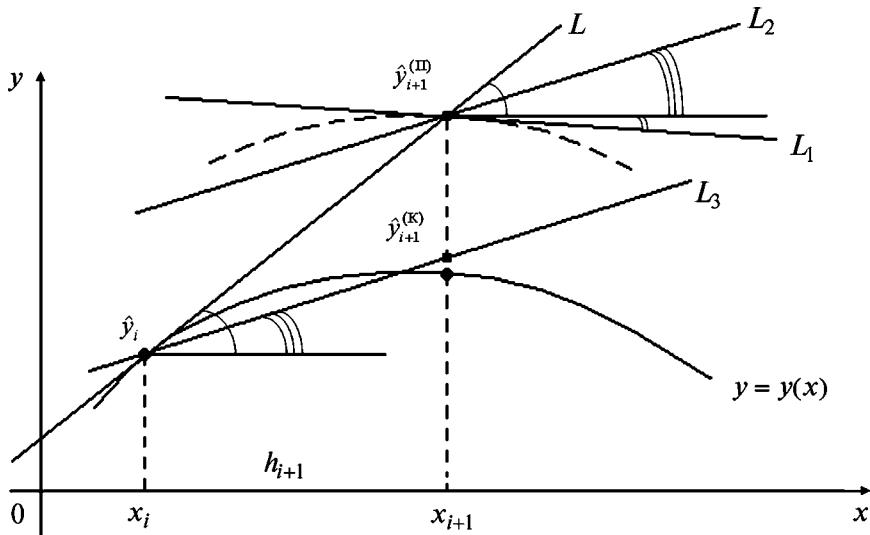


Рис. 6.14

$f(x_{i+1}, \hat{y}_{i+1}^{(II)})$. Средняя величина $\frac{f(x_i, \hat{y}_i) + f(x_{i+1}, \hat{y}_{i+1}^{(II)})}{2}$ порождает прямую L_2 , проходящую через точку $(x_{i+1}, \hat{y}_{i+1}^{(II)})$. Через точку (x_i, \hat{y}_i) проведем прямую L_3 , параллельную L_2 . Ее уравнение:

$$y = \hat{y}_i + \frac{f(x_i, \hat{y}_i) + f(x_{i+1}, \hat{y}_{i+1}^{(II)})}{2}(x - x_i).$$

Следующая точка $\hat{y}_{i+1} = \hat{y}_{i+1}^{(K)}$ приближенного решения (результат коррекции) получается при $x = x_{i+1}$:

$$\hat{y}_{i+1} = \hat{y}_{i+1}^{(K)} = \hat{y}_i + \frac{h_{i+1}}{2} [f(x_i, \hat{y}_i) + f(x_{i+1}, \hat{y}_{i+1}^{(II)})].$$

Данное соотношение соответствует (6.64). Если применяется предиктор (6.63) второго порядка, все приведенные построения сохраняются, только значение $\hat{y}_{i+1}^{(II)}$ находится так же, как на рисунке 6.8.

4. В составных методах может быть использована *модификация*. Например, в схеме (6.66), (6.67) после шага «предиктор» вычисляется модифицированное значение

$$\hat{y}_{i+1}^{(\text{ПМ})} = \hat{y}_{i+1}^{(\text{II})} + \frac{251}{270} (\hat{y}_i^{(\text{K})} - \hat{y}_i^{(\text{II})}),$$

где $\hat{y}_i^{(\text{K})}$, $\hat{y}_i^{(\text{II})}$ — значения, полученные на шагах «предиктор» и «корректор» в предыдущей точке.

Затем выполняется шаг «корректор»:

$$\hat{y}_{i+1}^{(\text{K})} = \hat{y}_i + \frac{h}{24} [f_{i-2} - 5f_{i-1} + 19f_i + 9f(x_{i+1}, \hat{y}_{i+1}^{(\text{ПМ})})].$$

Следующее значение \hat{y}_{i+1} вычисляется по формуле

$$\hat{y}_{i+1} = \hat{y}_{i+1}^{(K)} - \frac{19}{270}(\hat{y}_{i+1}^{(K)} - \hat{y}_{i+1}^{(\Pi)}).$$

Таким образом, здесь фактически выполняются три приближения.

5. При осуществлении шага «корректор» может быть выполнено несколько итераций. Например, для корректора (6.64) итерационная формула имеет вид

$$\hat{y}_{i+1}^{(k+1)} = \hat{y}_i + \frac{h_{i+1}}{2} [f(x_i, \hat{y}_i) + f(x_i + h_{i+1}, \hat{y}_{i+1}^{(k)})], \quad k = 0, 1, \dots, \quad \hat{y}_{i+1}^{(0)} = \hat{y}_{i+1}^{(\Pi)}.$$

Итерации прекращаются либо при выполнении условия $|\hat{y}_{i+1}^{(k+1)} - \hat{y}_{i+1}^{(k)}| \leq \varepsilon$ для некоторого положительного ε , либо при выполнении ровно m итераций (как правило, задается $m \leq 2$). Тогда полагают $\hat{y}_{i+1} = \hat{y}_{i+1}^{(k+1)}$.

Пример 6.7. Найти приближенное решение задачи Коши

$$y' = x + y, \quad y(0) = 1, \quad x \in [0; 0,2]$$

методом «предиктор — корректор» П1К2 (6.62), (6.64) второго порядка. Шаг интегрирования принять равным 0,1. Сравнить полученное решение с точным $y(x) = 2e^x - x - 1$.

□ Вычислим значения \hat{y}_1 и \hat{y}_2 в точках $x_1 = 0,1$ и $x_2 = 0,2$, полагая $\hat{y}_0 = 1$.

Шаг «предиктор»: по формуле (6.62) при $x_1 = 0,1$ получаем

$$\hat{y}_1^{(\Pi)} = \hat{y}_0 + h \cdot f(x_0, \hat{y}_0) = 1 + 0,1 \cdot (0 + 1) = 1,1.$$

Шаг «корректор»: по формуле (6.64) имеем

$$\hat{y}_1 \equiv \hat{y}_1^{(K)} = 1 + \frac{0,1}{2} [(0 + 1) + (0,1 + 1,1)] = 1,11.$$

Первый шаг по выбранной схеме выполнен. Реализуем следующий.

Шаг «предиктор»: по формуле (6.62) при $x_2 = 0,2$ получим

$$\hat{y}_2^{(\Pi)} = \hat{y}_1 + h \cdot f(x_1, \hat{y}_1) = 1,11 + 0,1 \cdot (0,1 + 1,11) = 1,231.$$

Шаг «корректор»: по формуле (6.64) имеем

$$\hat{y}_2 \equiv \hat{y}_2^{(K)} = 1,11 + \frac{0,1}{2} [(0,1 + 1,11) + (0,2 + 1,231)] = 1,24205.$$

Точное решение: $y(x_2) = 1,2428055$. Отличие приближенного решения от точного составляет 0,06 %. Так как отрезок интегрирования исчерпан, процесс закончен, иначе его следовало бы продолжить. ■

Пример 6.8. Приближенно решить задачу Коши

$$y' = -2y - 3x + 2, \quad y(0) = 0$$

на отрезке $[0; 1]$ с шагом $h = 0,1$ различными методами: Эйлера (явным и неявным), предсказания и коррекции (6.62), (6.64) второго порядка, Рунге — Кутты четвертого порядка и методом трапеций.

□ Используя исходный вид уравнения (подчеркнем, что оно линейное), из (6.19) получаем формулу явного метода Эйлера:

$$\hat{y}_{i+1} = \hat{y}_i + 0,1 \cdot (-2\hat{y}_i - 3x_i + 2), \quad \hat{y}_0 = 0, \quad i = \overline{0, 9};$$

из (6.62), (6.64) — формулы метода предсказания и коррекции:

$$\begin{aligned}\hat{y}_{i+1}^{(\Pi)} &= \hat{y}_i + 0,1 \cdot (-2\hat{y}_i - 3x_i + 2); \\ \hat{y}_{i+1} &= \hat{y}_i + \frac{0,1}{2}(-2\hat{y}_i - 3x_i + 2 - 2\hat{y}_{i+1}^{(\Pi)} - 3x_{i+1} + 2);\end{aligned}$$

из формулы (6.41) — формулы метода Рунге — Кутты четвертого порядка:

$$\begin{aligned}\hat{y}_{i+1} &= \hat{y}_i + \frac{0,1}{6}(K_{1,i} + 2K_{2,i} + 2K_{3,i} + K_{4,i}); \\ K_{1,i} &= -2\hat{y}_i - 3x_i + 2; \quad K_{2,i} = -2 \cdot (\hat{y}_i + 0,05 \cdot K_{1,i}) - 3 \cdot (x_i + 0,05) + 2; \\ K_{3,i} &= -2 \cdot (\hat{y}_i + 0,05 \cdot K_{2,i}) - 3 \cdot (x_i + 0,05) + 2; \\ K_{4,i} &= -2 \cdot (x_i + 0,1 \cdot K_{3,i}) - 3 \cdot (x_i + 0,1) + 2;\end{aligned}$$

из (6.23) — формулу неявного метода Эйлера:

$$\hat{y}_{i+1} = \hat{y}_i + 0,1 \cdot (-2\hat{y}_{i+1} - 3x_{i+1} + 2),$$

откуда

$$\hat{y}_{i+1} = \frac{\hat{y}_i - 0,3x_{i+1} + 0,2}{1,2};$$

из (6.24) — формулу метода трапеций:

$$\hat{y}_{i+1} = \hat{y}_i + \frac{0,1}{2} \cdot (-2\hat{y}_i - 3x_i + 2 - 2\hat{y}_{i+1} - 3x_{i+1} + 2),$$

откуда

$$\hat{y}_{i+1} = \frac{0,9 \cdot \hat{y}_i - \frac{3}{2} \cdot 0,1 \cdot x_i + 2 \cdot 0,1 - \frac{3}{2} \cdot 0,1 \cdot x_{i+1}}{1,1}.$$

Очевидно, в данном примере удалось получить явные формулы для нахождения \hat{y}_{i+1} неявным методом Эйлера и методом трапеций лишь в силу линейности решаемого уравнения. В общем случае применяются методы простых итераций (6.26) или Ньютона.

Точное решение рассматриваемой задачи Коши: $y(x) = 1,75 - 1,5x - 1,75e^{-2x}$. Результаты расчетов приведены в таблице 6.18, в последней строке которой указаны фактические погрешности.

Анализ результатов показывает, что при решении данной задачи метод предсказания и коррекции точнее явного и неявного методов Эйлера, но уступает методу Рунге — Кутты и совсем немного методу трапеций (порядок погрешности одинаков). ■

Таблица 6.18

x_i	Явный метод Эйлера	Неявный метод Эйлера	Метод Рунге — Кутты	Метод предсказания и коррекции	Метод трапеций	$y(x)$
0,0	0,000	0,000	0,000	0,000	0,000	0,000
0,1	0,200	0,142	0,167	0,165	0,168	0,167
0,2	0,330	0,238	0,277	0,273	0,279	0,277
0,3	0,404	0,287	0,340	0,335	0,342	0,340
0,4	0,433	0,306	0,364	0,359	0,366	0,364
0,5	0,427	0,297	0,356	0,351	0,358	0,356
0,6	0,391	0,264	0,323	0,318	0,325	0,323
0,7	0,333	0,212	0,268	0,264	0,270	0,268
0,8	0,256	0,143	0,197	0,192	0,199	0,197
0,9	0,165	0,061	0,111	0,107	0,112	0,111
1,0	0,062	-0,033	0,013	0,009	0,015	0,013
max ε_i	0,068	0,059	0,0000104	0,005	0,002	

Таблица 6.19

Тип схемы	П1	К2	П2	К2	П2	К3	П3	К3	Точное решение
x_i	1Я1	1НЯ2	2Я2Б	1НЯ2	2Я2Б	2НЯЗА	3Я3	2НЯЗА	$y(x_i)$
0,150	1,172711 0,08		1,174097 0,036		1,173700 0,0027		1,1736998 0,0027		1,1736684
0,285	1,372442 0,15		1,375370 0,061		1,374445 0,0057		1,374467 0,00415		1,374524
0,4065	1,593437 0,20		1,597867 0,079		1,596537 0,0043		1,596576 0,0019		1,596576

Пример 6.9. Найти приближенное решение задачи Коши

$$y' = x + y, \quad y(0) = 1, \quad x \in [0; 0,4065]$$

на нерегулярной сетке $\Omega_3 = (0,000; 0,150; 0,285; 0,4065)$ с $h_1 = 0,15; \delta_{i+1} = 0,9; i = 1,2$ с помощью методов предсказания и коррекции П1К2, П2К2, П2К3, П3К3.

□ Численные результаты помещены в таблице 6.19.

Для обеспечения расчета на первых шагах всех двухшаговых схем и одной трехшаговой предварительно рассчитывалось значение y в точке $x = 0,01$ по схеме Эйлера П1К2. В последнем столбце таблицы 6.19 указано точное значение, полученное аналитически. Относительные погрешности численных результатов указаны в строках, расположенных ниже строк, в которых помещены численные значения \hat{y}_i искомой функции. Анализ результатов свидетельствует об увеличении точности расчетов при повышении порядков схем как на шаге «предиктор», так и на шаге «корректор». Эти результаты оправдывают применение схем с многошаговым предиктором. Минимальная погрешность обеспечивается схемой П3К3, однако схема П2К3 также приводит к хорошим результатам. ■

6.4. ЭКСТРАПОЛЯЦИОННЫЕ МЕТОДЫ

Методы предсказания (см. п. 6.3) можно рассматривать как методы экстраполяции, где \hat{y}_{i+1} экстраполируется по известным предыдущим значениям. Здесь представлены другие экстраполяционные методы [53]–[55].

Процедура нахождения очередной рассчитываемой точки (x_{i+1}, \hat{y}_{i+1}) состоит из двух этапов.

Первый этап. Получение последовательности решений в точке $x = x_{i+1}$ с использованием некоторой стратегии уменьшения шага.

Второй этап. Уточнение численного решения на основе процедур полиномиальной или рациональной экстраполяции.

Способы реализации первого этапа. При решении конкретной задачи Коши следует задать число M , определяемое максимальным допустимым порядком экстраполяции, осуществляющей на втором этапе.

Рассмотрим процедуру интегрирования дифференциального уравнения (6.9) на двухточечном шаблоне (x_i, x_{i+1}) с длиной шага h , называемого *базовым*.

Определим последовательность чисел N_0, N_1, \dots, N_M , задающую количество частичных отрезков, располагающихся на промежутке $[x_i, x_{i+1}]$:

$$N_M > N_{M-1} > \dots > N_r > \dots > N_1 > N_0 \quad (0 \leq r \leq M), \quad (6.70)$$

и соответствующую последовательность значений шага $h_0, h_1, \dots, h_r, \dots, h_M$:

$$h_r = \frac{h}{N_r}, \quad r = 0, 1, \dots, M. \quad (6.71)$$

Разными авторами рекомендуются такие последовательности N_0, N_1, \dots, N_M :

- Дойфлхардом: $\{2, 4, 6, 8, 10, 12, 14, 16, 18, 20, \dots\}$ или $\{1, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, \dots\}$; «гармоническая» последовательность $\{1, 2, 3, 4, 5, 6, 7, 8, \dots\}$;
- Ромбергом: $\{2, 4, 8, 16, 32, 64, \dots\}$ или $\{1, 2, 4, 8, 16, 32, 64, \dots\}$;
- Булиршем и Штером: $\{2, 3, 4, 6, 8, 12, 16, 24, 32, \dots\}$ или $\{1, 2, 3, 4, 6, 8, 12, 16, 24, 32, \dots\}$.

Предположим, что точка (x_i, \hat{y}_i) известна. При каждом значении $r = \overline{0, M}$ генерируем сетку

$$t_s = x_i + s \cdot h_r, \quad s = 0, 1, \dots, N_r, \quad t_0 = x_i, \quad t_{N_r} = x_{i+1}, \quad (6.72)$$

и, применяя какой-либо разностный метод, интегрируем решаемое уравнение на отрезке $[x_i, x_{i+1}]$ с каждым значением шага h_r , начиная от $r = 0$ и кончая $r = M$. В результате получится $(M + 1)$ значений:

$$T_{r,0} \equiv y(x_{i+1}, h_r), \quad r = 0, 1, \dots, M, \quad (6.73)$$

где $y(x_{i+1}, h_r)$ — решение задачи Коши в точке x_{i+1} , найденное при интегрировании с шагом h_r .

Интегрирование рекомендуется проводить следующими методами [53]–[55]: а) явным методом Эйлера:

$$z_0 = \hat{y}_i, \quad z_{s+1} = z_s + h_r f(t_s, z_s), \quad s = 0, 1, \dots, N_r - 1, \quad T_{r,0} = z_{N_r}; \quad (6.74)$$

б) методом трапеций:

$$z_0 = \hat{y}_i, \quad z_{s+1} = z_s + \frac{h_r}{2} [f(t_s, z_s) + f(t_{s+1}, z_{s+1})], \quad s = 0, 1, \dots, N_r - 1, \quad T_{r,0} = z_{N_r};$$

в) методом Эйлера — Коши, модифицированным Грэггом:

$$\begin{aligned} z_0 &= \hat{y}_i, \quad z_1 = z_0 + h_0 f(t_0, z_0); \\ z_{s+1} &= z_{s-1} + 2h_r f(t_s, z_s) \quad s = 1, \dots, N_r - 1; \\ T_{r,0} &= \frac{1}{2} [z_{N_r-1} + z_{N_r} + h \cdot f(t_{N_r}, z_{N_r})]; \end{aligned} \quad (6.75)$$

г) обратным методом Эйлера:

$$z_0 = \hat{y}_i, \quad z_{s+1} = \frac{z_s^2}{z_s - h_r f(t_s, z_s)}, \quad s = 0, 1, \dots, N_r - 1, \quad T_{r,0} = z_{N_r}; \quad (6.76)$$

д) по правилу средней точки (значение z_{s+1} находится так же, как и в остальных неявных методах, изложенных в п. 6.2 и 6.3):

$$z_0 = \hat{y}_i, \quad z_{s+1} = z_s + h_r f\left(t_s + \frac{h_r}{2}, \frac{z_s + z_{s+1}}{2}\right), \quad s = 0, 1, \dots, N_r - 1, \quad T_{r,0} = z_{N_r};$$

е) методом Бадера и Дойфлхарда [53], который предложен для решения жестких автономных систем вида $Y' = F(Y)$. Заметим, что неавтономные системы (6.7) типа $Y' = F(x, Y)$ всегда можно свести к автономным путем введения новой переменной, удовлетворяющей уравнению

$$y'_{n+1} = 1, \quad y_{n+1}(x_0) = x_0,$$

и посредством рассмотрения расширенного вектора $\tilde{Y} = (Y, y_{n+1})$. С учетом обозначений

$$J = F_y(Y), \quad \bar{F}(Y) = F(Y) - J \cdot Y, \quad h_r = \frac{h}{2r},$$

метод записывается в форме

$$\begin{aligned} \eta_0 &= \hat{Y}_i, \quad \eta_1 = (E - h_r J)^{-1} [\hat{Y}_i + h_r \bar{F}(\hat{Y}_i)]; \\ \eta_{k+1} &= (E - h_r J)^{-1} [(E + h_r J)\eta_{k-1} + 2h_r \bar{F}(\eta_k)], \quad k = 1, 2, \dots, 2r; \\ T_{r,0} &= \frac{1}{2} (\eta_{2r-1} + \eta_{2r+1}). \end{aligned}$$

В результате применения каждого из методов получается столбец:

$$\begin{matrix} T_{0,0} \\ T_{1,0} \\ \vdots \\ T_{M-1,0} \\ T_{m,0} \end{matrix} \quad (6.77)$$

Способы реализации второго этапа. По полученным на первом этапе $(M+1)$ точкам $(h_r, y(x_{i+1}, h_r))$ могут быть построены либо интерполяционный полином M -й степени, либо рациональная функция.

Таким образом, для уточнения найденных на первом этапе решений используются:

a) *полиномиальная экстраполяция Эйткена — Невилла:*

$$T_{r,0} = y(x_{i+1}, h_r), \quad r = 0, 1, \dots, M; \\ T_{r,s} = T_{r+1,s-1} + \frac{T_{r+1,s-1} - T_{r,s-1}}{\left(\frac{h_r}{h_{r+s}}\right)^{\gamma} - 1}, \quad s = 1, 2, \dots, M; \quad r = 0, 1, \dots, M-s, \quad (6.78)$$

где, как правило, полагают $\gamma = 1$ или $\gamma = 2$.

По формуле (6.78) заполняется таблица 6.20.

В качестве уточненного результата \hat{y}_{i+1} принимается значение $T_{0,M}$;

б) *рациональная экстраполяция*, предложенная Булиршем и Штером:

$$T_{r,-1} = 0, \quad r = 1, \dots, M, \quad T_{r,0} = y(x_{i+1}, h_r), \quad r = 0, 1, \dots, M; \\ T_{r,s} = T_{r+1,s-1} + \frac{T_{r+1,s-1} - T_{r,s-1}}{\left(\frac{h_r}{h_{r+s}}\right)^{\gamma} \left(1 - \frac{T_{r+1,s-1} - T_{r,s-1}}{T_{r+1,s-1} - T_{r+1,s-2}}\right) - 1}, \quad s = 1, 2, \dots, M, \quad r = 0, 1, \dots, M-s. \quad (6.79)$$

Обычно полагают $\gamma = 1$ или $\gamma = 2$. В результате формируется таблица 6.21. В качестве уточненного результата \hat{y}_{i+1} принимается значение $T_{0,M}$.

Таблица 6.20

$T_{0,0}$	$T_{0,1}$	$T_{0,2}$	$T_{0,M-1}$	$T_{0,M}$
$T_{1,0}$	$T_{1,1}$			
$T_{2,0}$			$T_{1,M-1}$	
\vdots				
$T_{M-1,0}$	$T_{M-2,1}$	$T_{M-2,2}$		
$T_{M,0}$	$T_{M-1,1}$			

Таблица 6.21

$T_{0,0}$	$T_{0,1}$	$T_{0,2}$	$T_{0,M-1}$	$T_{0,M}$
$T_{1,0}$				
$T_{2,0}$				
\vdots				
$T_{M-1,0}$	$T_{M-2,1}$	$T_{M-2,2}$		
$T_{M,0}$	$T_{M-1,1}$			

Замечания

1. Если на первом этапе применять формулу (6.75), а на втором этапе — рациональную экстраполяцию, то получаем *метод Грэгга — Булирша — Штера*.
2. Рекомендуемый максимальный порядок экстраполяции $M = 6$.
3. Существуют различные модификации экстраполяционных методов, в том числе с динамическим изменением порядка экстраполяции, контролем точности, рекуррентной реализацией экстраполяции [53]–[55].

*Алгоритм применения
экстраполяционных методов*

1. Задать:

- а) порядок аппроксимации M ;
- б) базовый шаг h ;
- в) последовательность N_0, N_1, \dots, N_M .

2. Выбрать разностный метод для нахождения решений с величинами шага, равными h_0, h_1, \dots, h_M , и получить значения, составляющие столбец (6.77).

3. Выбрать метод экстраполяции, задать параметр γ и получить уточненное решение $T_{0,M}$ по формуле (6.78) при полиномиальной экстраполяции или (6.79) при рациональной экстраполяции. Перейти к следующему промежутку интегрирования.

Пример 6.10. Решить задачу Коши [53]

$$y' = 1 + y^2, \quad y(0) = 1$$

на отрезке $[0; 0,25]$ с базовым шагом $h = 0,25$ и порядком экстраполяции, равным $M = 3$.

□ Точное решение задачи: $y(x) = \operatorname{tg}\left(x + \frac{\pi}{4}\right)$, $y(0,25) = 1,685796417$. Будем использовать последовательность Ромберга. Так как $M = 3$, выберем из нее четыре первых числа: 2, 4, 8, 16, т. е. $N_0 = 2, N_1 = 4, N_2 = 8, N_3 = 16$. Тогда по формуле (6.71) получаем: $h_0 = \frac{0,25}{2}, h_1 = \frac{0,25}{4}, h_2 = \frac{0,25}{8}, h_3 = \frac{0,25}{16}$.

Для нахождения приближенного решения задачи на первом этапе применим явный метод Эйлера (6.74), а на втором этапе — рациональную экстраполяцию (6.79) при $\gamma = 1$ (табл. 6.22).

Таблица 6.22

0	1,5703125000	1,6720018467	1,6854629545	1,6857868712
0	1,6195625218	1,6814888520	1,6857435231	
0	1,6499448310	1,6845757643		
	1,6670804665			

Элементы второго столбца таблицы 6.22 получены по формулам (6.74), т. е.

$$T_{0,0} = y(0,25; h_0) = 1,5703125000;$$

$$T_{1,0} = y(0,25; h_1) = 1,6195625218;$$

$$T_{2,0} = y(0,25; h_2) = 1,6499448310;$$

$$T_{3,0} = y(0,25; h_3) = 1,6670804665.$$

Приведем примеры расчетов по формуле (6.79):

$$\begin{aligned} T_{0,1} &= T_{1,0} + \frac{T_{1,0} - T_{0,0}}{\left(\frac{h_0}{h_1}\right)^1 \left(1 - \frac{T_{1,0} - T_{0,0}}{T_{1,0} - T_{1,-1}}\right) - 1} = 1,6195625218 + \\ &+ \frac{1,6195625218 - 1,5703125000}{2^1 \cdot \left(1 - \frac{1,6195625218 - 1,5703125000}{1,6195625218 - 0}\right) - 1} = 1,6720018467; \end{aligned}$$

$$\begin{aligned} T_{1,1} &= T_{2,0} + \frac{T_{2,0} - T_{1,0}}{\left(\frac{h_1}{h_2}\right)^1 \left(1 - \frac{T_{2,0} - T_{1,0}}{T_{2,0} - T_{2,-1}}\right) - 1} = 1,6499448310 + \\ &+ \frac{1,6499448310 - 1,6195625218}{2^1 \cdot \left(1 - \frac{1,6499448310 - 1,6195625218}{1,6499448310 - 0}\right) - 1} = 1,6814888520; \end{aligned}$$

$$\begin{aligned} T_{0,2} &= T_{1,1} + \frac{T_{1,1} - T_{0,1}}{\left(\frac{h_0}{h_2}\right)^1 \left(1 - \frac{T_{1,1} - T_{0,1}}{T_{1,1} - T_{1,0}}\right) - 1} = 1,6814888520 + \\ &+ \frac{1,6814888520 - 1,6720018467}{4^1 \cdot \left(1 - \frac{1,6814888520 - 1,6720018467}{1,6814888520 - 1,6195625218}\right) - 1} = 1,6854629545; \end{aligned}$$

$$\begin{aligned} T_{0,3} &= T_{1,2} + \frac{T_{1,2} - T_{0,2}}{\left(\frac{h_0}{h_3}\right)^1 \left(1 - \frac{T_{1,2} - T_{0,2}}{T_{1,2} - T_{1,1}}\right) - 1} = 1,6857435231 + \\ &+ \frac{1,6857435231 - 1,6854629545}{8^1 \cdot \left(1 - \frac{1,6857435231 - 1,6854629545}{1,6857435231 - 1,6814888520}\right) - 1} = 1,6857868712. \end{aligned}$$

Остальные вычисления аналогичны. Погрешность полученного решения $\hat{y}(0,25) = 1,6857868712$ составляет

$$|1,6857868712 - 1,685796417| = 5,2894 \cdot 10^{-5}.$$

Если воспользоваться обратным методом Эйлера (6.76) на первом этапе, а полиномиальной экстраполяцией с параметром $\gamma = 1$ на втором этапе, то получим результаты, приведенные в таблице 6.23.

Здесь результаты экстраполяции находятся по формуле (6.78), например,

$$T_{0,3} = T_{1,2} + \frac{T_{1,2} - T_{0,2}}{\left(\frac{h_0}{h_3}\right)^1 - 1} = 1,6858697276 + \\ + \frac{1,6858697276 - 1,6865256650}{8^1 - 1} = 1,6857760212.$$

Погрешность полученного решения составляет

$$|1,6857760212 - 1,685796417| = 2,039 \cdot 10^{-5}.$$

Воспользуемся обратным методом Эйлера (6.76) и рациональной экстраполяцией при $\gamma = 1$ (табл. 6.24).

Фактическая погрешность полученного решения равна

$$|1,6857961912 - 1,685796417| = 7,2598 \cdot 10^{-6}.$$

Применим модифицированный метод Эйлера — Коши и рациональную экстраполяцию при $\gamma = 2$ (метод Грэгга — Булирша — Штера) (табл. 6.25).

Погрешность полученного решения:

$$|1,6857963393 - 1,685796417| = 3,4425 \cdot 10^{-7}.$$

Очевидно, найденное в таблице 6.25 решение точнее остальных. ■

Таблица 6.23

1,8028169014	1,6729289242	1,6865256650	1,6857760212
1,7378729128	1,6831264798	1,6858697276	
1,7104996963	1,6851839150		
1,6978418056			

Таблица 6.24

0	1,8028169014	1,6774452643	1,6858494038	1,6857961912
0	1,7378729128	1,6839754186	1,6858026769	
0	1,7104996963	1,6853698785		
	1,6978418056			

Таблица 6.25

0	1,6406250000	1,6842556897	1,6857813466	1,6857963393
0	1,6731318956	1,6856738366	1,6857960729	
0	1,6825207564	1,6857881562		
	1,6849701171			

6.5. НЕПРЕРЫВНО-ДИСКРЕТНЫЕ МЕТОДЫ

6.5.1. КОНСТРУИРОВАНИЕ ПОСЛЕДОВАТЕЛЬНЫХ СПЛАЙН-МЕТОДОВ

Результаты, получаемые численными методами сеток, имеют дискретный характер и, как отмечено ранее, требуют восполнения для определения значений функций в промежуточных точках и, возможно, для последующей обработки найденного решения. Этого недостатка лишены сплайн-методы решения задачи Коши. Однако классические сплайн-методы, основанные на разложении искомой функции в ряд Тейлора, неудобны в практической реализации, так как для нелинейных задач требуется решать нелинейные алгебраические уравнения [41]. В данном пункте излагаются непрерывно-дискретные методы, называемые *последовательными сплайн-методами* решения задачи Коши второго и третьего порядков, которые позволяют избежать решения нелинейных уравнений [18].

Процедура решения задачи Коши при переходе от известного значения \hat{y}_i в точке x_i к очередному рассчитываемому значению \hat{y}_{i+1} в точке x_{i+1} состоит из двух этапов.

Первый этап. Находится дискретное опорное решение задачи Коши с помощью явных или неявных методов второго и третьего порядков, изложенных выше.

Второй этап. Полученное на первом этапе опорное решение подставляется в соответствующую по порядку формулу многочлена (сплайн-функцию). В качестве этих сплайн-функций могут быть приняты многочлены второй и третьей степени, использующиеся в теории приближений:

$$S_{2,i}(x) = \hat{y}_i + \bar{m}_i \cdot (x - x_i) + \frac{1}{h_{i+1}} \left(\frac{\Delta \hat{y}_i}{h_{i+1}} - \bar{m}_i \right) \cdot (x - x_i)^2, \quad (i = 0, 1, \dots, n-1); \quad (6.80)$$

$$\begin{aligned} S_{3,i}(x) = & \hat{y}_i + \bar{m}_i \cdot (x - x_i) + \left(\frac{3\Delta \hat{y}_i}{h_{i+1}^2} - \frac{3\bar{m}_i}{h_{i+1}} - \frac{\Delta \bar{m}_i}{h_{i+1}} \right) \cdot (x - x_i)^2 + \\ & + \frac{1}{h_{i+1}^2} \left(-\frac{2}{h_{i+1}} \Delta \hat{y}_i + 2\bar{m}_i + \Delta \bar{m}_i \right) \cdot (x - x_i)^3, \quad (i = 0, 1, \dots, n-1), \end{aligned} \quad (6.81)$$

где $\bar{m}_i = \hat{y}'(x_i)$. Формула для $S_{2,i}(x)$ соответствует многочлену, получаемому по условиям согласования $\delta S_{2,i}(x_k) = 0$, ($k = i, i+1$); $\delta S'_{2,i}(x_i) = 0$, а формула для $S_{3,i}$ — формуле (4.75).

Таким образом, алгоритм получения непрерывного сплайн-решения задачи Коши на каждом частичном отрезке $[x_i, x_{i+1}]$ ($i = 0, 1, \dots, n-1$) разбиения отрезка $[a, b]$ на n промежутков (данное разбиение может осуществляться как заранее, так и в процессе решения задачи) содержит две совокупности вычислительных процедур, которые выполняются последовательно и независимо друг от друга. Последовательный характер данных процедур определяет название данного метода решения задачи Коши.

Рассмотрим совокупности этих процедур отдельно для схем второго и третьего порядков, так как они имеют некоторые особенности. Здесь, как и выше, решается задача Коши для уравнения первого порядка

$$y' = f(x, y), \quad y(x_0) = y_0.$$

6.5.2. СХЕМА ВТОРОГО ПОРЯДКА

Первый этап. Рассчитывается дискретное решение $\hat{y}_0, \hat{y}_1, \dots, \hat{y}_n$ по одной наиболее приемлемой явной или неявной несоставной схеме (см. п. 6.2) или по составной схеме, скомпонованной из совокупности явной и неявной схем (п. 6.3). Решение, полученное на данном этапе, называется *опорным*. При выборе конкретной схемы необходимо, чтобы ее порядок точности равнялся $p - 1 = 2$, где $p = 3$ — порядок сходимости сплайн-функции на отрезке $[x_i, x_{i+1}]$.

Второй этап. На каждом очередном отрезке $[x_i, x_{i+1}]$ дискретное решение преобразуется в непрерывное путем определения одного из звеньев сплайн-функции (6.80). Для этого сначала необходимо рассчитать коэффициенты $a_{0,i}, a_{1,i}, a_{2,i}$ многочлена

$$S_{2,i}(x) = a_{0,i} + a_{1,i}(x - x_i) + a_{2,i}(x - x_i)^2. \quad (6.82)$$

Из сопоставления формул (6.82) и (6.80) видно, что коэффициент $a_{0,i}$ определяется значением \hat{y}_i , которое известно, а коэффициент $a_{1,i}$, равный производной $y'(x_i) = \bar{m}_i$, находится путем дифференцирования в точке $x = x_i$ сплайн-функции $S_{2,i-1}(x)$, рассчитанной на предыдущем шаге интегрирования. Таким образом, для $a_{0,i}$ и $a_{1,i}$ имеем:

$$a_{0,i} = \hat{y}_i, \quad a_{1,i} = S'_{2,i-1} \Big|_{x=x_i} \quad (a_{1,i} = \bar{m}_i).$$

Последний коэффициент $a_{2,i}$ вычисляется с использованием полученного на первом этапе дискретного решения по формуле

$$a_{2,i} = \frac{1}{h_{i+1}} \left(\frac{\Delta \hat{y}_i}{h_{i+1}} - a_{1,i} \right),$$

следующей из (6.80). Здесь $\Delta \hat{y}_i = \hat{y}_{i+1} - \hat{y}_i$ (значения \hat{y}_i , $i = \overline{1, n}$, вычислены на первом этапе), $h_{i+1} = x_{i+1} - x_i$.

Замечания

1. При построении функции $S_{2,0}(x)$ на первом частичном отрезке следует положить $a_{0,0} = \hat{y}_0 = y_0$ (значение y_0 задано в задаче Коши (6.9)), а $a_{1,0} = f(x_0, \hat{y}_0)$.

2. Данный алгоритм построения сплайн-функции $S_2(x)$ обеспечивает непрерывность самого решения $\hat{y}(x) \equiv \bigcup_{i=0}^{n-1} S_{2,i}(x)$ на всем отрезке $[a, b]$, а также

его производной, т. е. $S'_{2,i-1}(x)|_{x=x_i} = S'_{2,i}(x)|_{x=x_i}$ ($i = \overline{1, n-1}$). Поэтому получаемое по данной схеме решение соответствует параболическому сплайну дефекта $q = 1$.

Пример 6.11. Для задачи Коши

$$y' = x + y; \quad y(0) = 1$$

найти численное решение на отрезке $[0; 0,4065]$ последовательным сплайн-методом второго порядка, задав число частичных отрезков $n = 3$; параметр нерегулярности сетки $\delta_{i+1} = 0,9$ ($i = 1; 2$), шаг $h_1 = 0,15$.

□1. Формирование узлов сетки. С учетом заданных параметров сетки находим:

$$\begin{aligned}x_0 &= 0; \quad x_1 = x_0 + h_1 = 0,15; \quad h_2/h_1 = 0,9; \quad h_2 = 0,15 \cdot 0,9 = 0,135; \\x_2 &= x_1 + h_2 = 0,15 + 0,135 = 0,285; \quad h_3/h_2 = 0,9; \quad h_3 = 0,135 \cdot 0,9 = 0,1215; \\x_3 &= 0,285 + 0,125 = 0,4065.\end{aligned}$$

2. Получение опорного решения (первый этап). В качестве опорного возьмем готовое решение, полученное выше (в п. 6.3) по схеме П2К2, имеющей второй порядок точности. Данное решение приведено в таблице 6.19 в третьей от левого края колонке. Нумерацию узлов для данных, помещенных в эту таблицу, следует начать с $i = 1$, так как узел $x_0 = 0$ в ней не содержится.

Построение формул сплайн-функций на трех частичных отрезках промежутка $[0; 0,4065]$ (второй этап).

Рассчитаем коэффициенты сплайн-функции на первом частичном отрезке $[0; 0,15]$:

$$\begin{aligned}a_{0,0} &= \hat{y}_0 = y_0 = 1; \quad a_{1,0} = f(x_0, \hat{y}_0) = x_0 + \hat{y}_0 = 1; \\&\Delta\hat{y}_0 = \hat{y}_1 - \hat{y}_0 = \hat{y}_1 - 1 = 1,1740967 - 1 = 0,1740967; \\a_{2,0} &= \frac{1}{h_1} \left(\frac{\Delta\hat{y}_0}{h_1} - a_{1,0} \right) = \frac{1}{0,15} \left(\frac{0,1740967}{0,15} - 1 \right) = 1,0709644.\end{aligned}$$

С учетом этих коэффициентов записываем формулу первого звена сплайн-функции:

$$S_{2,0}(x) = 1 + x + 1,0709644x^2.$$

Рассчитаем коэффициенты сплайн-функции на втором частичном отрезке $[0,15; 0,285]$:

$$\begin{aligned}a_{0,1} &= \hat{y}_1 = 1,174096; \quad a_{1,1} = S'_{2,0}(x)|_{x=0,15} = 1 + 2 \cdot 1,0709644 \cdot x|_{x=0,15} = 1,3212893; \\a_{2,1} &= \frac{1}{h_2} \left(\frac{\Delta\hat{y}_1}{h_2} - a_{1,1} \right) = \frac{1}{0,135} \left(\frac{1,3753698 - 1,1740967}{0,135} - 1,3212893 \right) = 1,2564633.\end{aligned}$$

С помощью коэффициентов записываем формулу второго звена сплайн-функции:

$$S_{2,1}(x) = 1,1740967 + 1,3212893 \cdot (x - 0,15) + 1,2564633 \cdot (x - 0,15)^2.$$

Аналогично рассчитываются коэффициенты сплайн-функции и записывается формула третьего звена:

$$S_{2,2}(x) = 1,3753698 + 1,6605343 \cdot (x - 0,285) + 1,4050819 \cdot (x - 0,285)^2.$$

В результате получено непрерывное решение задачи Коши, соответствующее параболическому сплайну дефекта $q = 1$. ■

Замечание. В примере 6.11, так же как и в последующих двух, связанных с применением кубических сплайн-функций $S_3(x)$, процессы получения дискретного и непрерывного решений (на первом и втором этапах) для упрощения полностью разделены. При решении реальных задач Коши более целесообразно их не разделять, а сразу после определения дискретного решения на очередном частичном отрезке $[x_i, x_{i+1}]$ строить соответствующее непрерывное решение, т. е. формировать очередное звено сплайн-функции.

6.5.3. СХЕМА ТРЕТЬЕГО ПОРЯДКА

Схема третьего порядка может быть реализована двумя различными способами, первый из которых основан на использовании формулы (6.71), а второй на иной конструкции формулы звена сплайна, получаемой в данном пункте из условий непрерывности первой и второй производной.

Первый способ реализации схемы третьего порядка.

Первый этап. Рассчитывается дискретное (опорное) решение \hat{y}_{i+1} в точке x_{i+1} по одной наиболее приемлемой явной или неявной несоставной схеме или по некоторой составной схеме, состоящей из явной и неявной схем. Порядок точности этой схемы должен быть не ниже трех, что соответствует $p = 1$, где $p = 4$ — порядок сходимости сплайн-функции $S_3(x)$.

Второй этап. При выполнении этого этапа дискретное решение \hat{y}_i , \hat{y}_{i+1} , соответствующее концам отрезка $[x_i, x_{i+1}]$, преобразуется в непрерывное путем построения одного звена сплайн-функции (6.81). Для этого, так же как и для схемы второго порядка, вначале рассчитываются коэффициенты $a_{0,i}, a_{1,i}, a_{2,i}, a_{3,i}$ многочлена $S_{3,i}(x)$:

$$S_{3,i}(x) = a_{0,i} + a_{1,i}(x - x_i) + a_{2,i}(x - x_i)^2 + a_{3,i}(x - x_i)^3, \quad i = \overline{0, n-1}. \quad (6.83)$$

Из сопоставления формул (6.81) и (6.83) видно, что $a_{0,i}$ определяется значением \hat{y}_i , которое рассчитано на предыдущем шаге. Коэффициент $a_{1,i}$ вычисляется по формуле $a_{1,i} = f(x_i, \hat{y}_i)$, а коэффициенты $a_{2,i}$ и $a_{3,i}$ — с использованием дискретного решения по формулам:

$$a_{2,i} = \frac{3\Delta\hat{y}_i}{h_{i+1}^2} - \frac{3\bar{m}_i}{h_{i+1}} - \frac{\Delta\bar{m}_i}{h_{i+1}}; \quad a_{3,i} = \frac{3}{h_{i+1}^2} \left(-\frac{2\Delta\hat{y}_i}{h_{i+1}} + 2\bar{m}_i + \Delta\bar{m}_i \right),$$

следующим из (6.81).

Здесь, как и выше, принято обозначение $\bar{m}_i = \hat{y}'(x_i)$. Величина \bar{m}_{i+1} вычисляется с использованием правой части исходного дифференциального уравнения, т. е. $\bar{m}_{i+1} = f(x_{i+1}, \hat{y}_{i+1})$. При построении функции $S_{3,0}(x)$ на первом отрезке следует положить $a_{0,0} = \hat{y}_0 = y_0$, а $a_{1,0} = f(x_0, \hat{y}_0)$.

Пример 6.12. Для задачи Коши

$$y' = x + y; \quad y(0) = 1$$

найти численное решение на отрезке $[0; 0,4065]$ последовательным сплайн-методом третьего порядка ($n = 3, \delta_{i+1} = 0,9, h_1 = 0,15$) в соответствии с изложенным первым способом.

□1. Формирование узлов сетки осуществляется аналогично примеру 6.11.

2. Получение опорного решения (первый этап). В качестве опорного возьмем решение, рассчитанное в предыдущем пункте по схеме П2К3, имеющей третий порядок точности. Данное решение приведено в таблице 6.19 в четвертой от левого края колонке. Нумерацию узлов для данных, помещенных в эту таблицу, следует начать с $i = 1$, так как узел $x_0 = 0$ в ней не содержится.

3. Построение формул сплайн-функции на трех частичных отрезках разбиения отрезка $[0; 0,4065]$ (второй этап).

Рассчитаем коэффициенты сплайн-функции на первом отрезке $[0; 0,15]$ (первое звено сплайна):

$$a_{0,0} = \hat{y}_0 = y_0 = 1; \quad a_{1,0} = \hat{y}'_0 = f(x_0, \hat{y}_0) = 1 = \bar{m}_0;$$

$$\bar{m}_1 = f(x_1, \hat{y}_1) = x_1 + \hat{y}_1 = 0,15 + 1,1736998 = 1,3236998;$$

$$a_{2,0} = \frac{3\Delta\hat{y}_0}{h_1^2} - \frac{3\bar{m}_0}{h_1} - \frac{\Delta\bar{m}_0}{h_1} = \frac{3(1,1736998 - 1)}{0,15^2} - \frac{3 \cdot 1}{0,15} - \frac{1,3236998 - 1}{0,15} = 1,0019746;$$

$$a_{3,0} = \frac{3}{h_1^2} \left(-\frac{2\Delta\hat{y}_0}{h_1} + 2\bar{m}_0 + \Delta\bar{m}_0 \right) = \frac{3}{0,15^2} \left(-2 \cdot \frac{0,1736998}{0,15} + 2 \cdot 1 + 0,3236998 \right) = 0,342332.$$

По значениям рассчитанных коэффициентов записывается формула первого звена сплайн-функции

$$S_{3,0}(x) = 1 + 1,0019746 \cdot x^2 + 0,342332 \cdot x^3.$$

Аналогично находятся и два оставшихся звена кубической сплайн-функции. Коэффициенты всех трех звеньев приведены в таблице 6.26. ■

Таблица 6.26

Номер звена i	$a_{0,i}$	$a_{1,i}$	$a_{2,0}$	$a_{3,0}$
0	1,0000	1,0000	1,0020	0,3423
1	1,1737	1,3237	1,1419	0,5016
2	1,3744	1,6594	1,3318	0,4507

Второй способ реализации схемы третьего порядка.

При необходимости строгого выполнения условия непрерывности первой и второй производной в узлах расчетной сетки коэффициенты искомого кубического сплайна определяются из двух условий интерполяции (согласования) на концах отрезка $[x_i, x_{i+1}]$ и двух условий стыковки (непрерывности) первой и второй производной в узлах, соединяющих два соседних отрезка $[x_{i-1}, x_i]$ и $[x_i, x_{i+1}]$, т. е.

$$S'_{3,i-1}(x)|_{x=x_i} = S'_{3,i}(x)|_{x=x_i}, \quad i = \overline{1, n-1};$$

$$S''_{3,i-1}(x)|_{x=x_i} = S''_{3,i}(x)|_{x=x_i}, \quad i = \overline{1, n-1}.$$

Производные, указанные в левых частях этих условий стыковки, при каждом значении i вычисляются с помощью многочлена, полученного на предыдущем шаге интегрирования, т. е. на отрезке $[x_{i-1}, x_i]$.

С учетом отмеченных выше двух условий согласования и двух условий стыковки нетрудно получить следующую формулу i -го звена сплайна:

$$S_{3,i}(x) = \hat{y}_i + \bar{m}_i(x - x_i) + \frac{m_i}{2} \cdot (x - x_i)^2 + \frac{1}{h_{i+1}^3} \left(\Delta \hat{y}_i - \bar{m}_i h_{i+1} - \frac{m_i}{2} h_{i+1}^2 \right) \cdot (x - x_i)^3, \quad (6.84)$$

где $\bar{m}_i = \hat{y}'(x_i) = S'_{3,i-1}(x_i)$; $m_i = \hat{y}''(x_i) = S''_{3,i-1}(x_i)$. Таким образом, коэффициенты звена сплайна (6.84) находятся по формулам

$$\begin{aligned} a_{0,i} &= \hat{y}_i; \quad a_{1,i} = \bar{m}_i = S'_{3,i-1}(x_i); \quad a_{2,i} = \frac{m_i}{2} = \frac{S''_{3,i-1}(x_i)}{2}; \\ a_{3,i} &= \frac{1}{h_{i+1}^3} \left(\Delta \hat{y}_i - a_{1,i} h_{i+1} - a_{2,i} h_{i+1}^2 \right), \quad i = \overline{1, n-1}. \end{aligned}$$

На первом отрезке для первого звена сплайна коэффициенты $a_{0,0}$, $a_{1,0}$, $a_{2,0}$ вычисляются по формулам

$$a_{0,0} = \hat{y}_0 = y_0; \quad a_{1,0} = \bar{m}_0 = f(x_0, \hat{y}_0); \quad a_{2,0} = \frac{m_0}{2} = \frac{y''(x_0)}{2} = \frac{f'(x_0, \hat{y}_0)}{2}.$$

Пример 6.13. Решить задачу, сформулированную в примерах 6.11 и 6.12, с помощью второго способа реализации схемы третьего порядка.

□1. Формирование узлов сетки аналогично примерам 6.11 и 6.12.

2. Получение опорного решения. Осуществляется аналогично изложенному в примере 6.12.

3. Построение звеньев сплайн-функции на трех частичных отрезках $[0; 0,15]$; $[0,15; 0,285]$; $[0,285; 0,4065]$.

Рассчитаем коэффициенты сплайн-функции на первом частичном отрезке $[0; 0,15]$ (первое звено сплайна):

$$\begin{aligned} a_{0,0} &= \hat{y}_0 = y_0 = 1; \quad a_{1,0} = \hat{y}'_0 = f(x_0, \hat{y}_0) = 1 = \bar{m}_0; \\ a_{2,0} &= \frac{\hat{y}''_0}{2} = \frac{f'(x, y)|_{x=x_0}}{2} = \frac{(x+y)'|_{x=0}}{2} = \frac{1+\hat{y}'_x|_{x=0}}{2} = \frac{1+1}{2} = 1; \\ a_{3,0} &= \frac{1}{h_1^3} \left(\Delta \hat{y}_0 - \bar{m}_0 \cdot h_1 - \frac{m_0}{2} h_1^2 \right) = \frac{1}{0,15^3} (0,1736998 - 1 \cdot 0,15 - 1 \cdot 0,15^2) = 0,3554963. \end{aligned}$$

С помощью этих коэффициентов записываем формулу первого звена сплайн-функции:

$$S_{3,0}(x) = 1 + x + x^2 + 0,3554963x^3.$$

Рассчитаем коэффициенты второго звена сплайн-функции на втором частичном отрезке $[0,15; 0,285] = [x_1, x_2]$ ($h_2 = 0,135$). По формуле первого звена $S_{3,0}(x)$ вычисляются первые и вторые производные $S'_{3,0}(x)|_{x=x_1}$, $S''_{3,0}(x)|_{x=x_1}$ соответственно:

$$\begin{aligned} S'_{3,0}(x)|_{x=0,15} &= (1 + 2x + 3 \cdot 0,3554963x^2)|_{x=0,15} = 1,323996 = \bar{m}_1; \\ S''_{3,0}(x)|_{x=0,15} &= (2 + 6 \cdot 0,3554963x)|_{x=0,15} = 2,3199466 = m_1. \end{aligned}$$

После этого легко найти все коэффициенты многочлена $S_{3,1}(x)$ второго звена сплайн-функции:

$$a_{0,1} = \hat{y}_1 = 1,1736998; \quad a_{1,1} = \bar{m}_1 = 1,323996;$$

$$a_{2,1} = \frac{m_1}{2} = \frac{2,3199466}{2} = 1,1599733; \quad a_{3,1} = \frac{1}{h_2^3} \left(\Delta \hat{y}_1 - \bar{m}_1 h_2 - \frac{m_1}{2} h_2^2 \right) =$$

$$= \frac{1}{0,135^3} \left[(1,374445 - 1,1736998) - 1,323996 \cdot 0,135 - \frac{2,3199466}{2} \cdot 0,135^2 \right] = 0,3528836.$$

С использованием этих коэффициентов записывается формула второго звена сплайн-функции:

$$S_{3,1}(x) = 1,1736998 + 1,323996 \cdot (x - 0,15) +$$

$$+ 1,1599733 \cdot (x - 0,15)^2 + 0,3528836 \cdot (x - 0,15)^3.$$

Процесс построения последующих звеньев сплайн-функции выполняется аналогично до конца отрезка $[a, b]$. В частности, для рассматриваемого примера получается такая формула последующего третьего звена сплайн-функции:

$$S_{3,2}(x) = 1,374445 + 1,6564827 \cdot (x - 0,285) +$$

$$+ 1,302912 \cdot (x - 0,285)^2 + 0,8894584 \cdot (x - 0,285)^3.$$

С помощью найденных звеньев формируется многозвенная кубическая сплайн-функция дефекта $q = 1$, являющаяся приближенным решением задачи Коши. ■

Задачи для самостоятельного решения

1. Для решения задачи Коши

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

записать формулы:

- а) явного и неявного методов Эйлера;
- б) метода Эйлера — Коши;
- в) метода трапеций;
- г) методов Адамса — Бэшфорта и Адамса — Мултона третьего и четвертого порядков.

2. Для решения задачи Коши

$$y' = y + 2z - 9x, \quad y(0) = 1;$$

$$z' = 2y + z - 4e^x, \quad z(0) = 2$$

записать формулы:

- а) явного и неявного методов Эйлера;
- б) метода Эйлера — Коши;
- в) метода трапеций;
- г) методов Адамса — Бэшфорта и Адамса — Мултона четвертого порядка.

3. Используя явный метод Эйлера с шагами $h = 1; 0,5; 0,25$, найти приближенное решение уравнения $y' = 0,04y; y(0) = 1000$, сравнить его с точным решением $y(x) = 1000e^{0,04x}$, вычислить значение $y(1)$, проанализировать сходимость.

4. Найти приближенное решение задачи Коши

$$y' = -0,01y; \quad y(0) = 100, \quad [a, b] = [0; 100]$$

и вычислить $y(100)$ (точный ответ равен 36,788) с помощью:

- а) явного метода Эйлера при $h = 25; 10; 5; 1;$
- б) модифицированного метода Эйлера при $h = 20; 10;$
- в) исправленного метода Эйлера при $h = 20; 10;$
- г) метода Рунге — Кутты четвертого порядка при $h = 100; 50;$
- д) метода прогноза и коррекции второго порядка при $h = 20$, если известно значение $y(20) = 81,8731$.

5. Найти приближенное решение задачи Коши

$$y' = 4 - 2x, \quad y(0) = 2, \quad [a, b] = [0; 1]$$

с помощью:

- а) явного метода Эйлера;
- б) модифицированного метода Эйлера;
- в) исправленного метода Эйлера.

Сравнить его с точным решением $y(x) = -x^2 + 4x + 2$. Исследовать поведение приближенного решения при изменении величины шага.

6. Найти приближенное решение задачи Коши

$$y' = -\frac{x}{y}, \quad y(0) = 20, \quad [a, b] = [0; 12]$$

с помощью:

- а) модифицированного метода Эйлера при $h = 2;$
- б) метода Рунге — Кутты четвертого порядка при $h = 4.$

7. Найти приближенное решение задачи Коши

$$y' = \frac{(6y + 2x + 1)}{x}, \quad y(1) = -\frac{17}{30}, \quad [a, b] = [1; 4]$$

при $h = 1$, вычислить $y(4)$ методом прогноза и коррекции второго порядка. Сравнить найденное решение с полученным при другом начальном условии $y(1) = -\frac{16}{30}$.

7.1. ПОСТАНОВКА ЗАДАЧИ И ОСНОВНЫЕ ПОЛОЖЕНИЯ

Рассмотрим *двуточечные краевые задачи*, часто встречающиеся в приложениях, например, при решении задач вариационного исчисления, оптимального управления, механики жидкости и газа и др. Пусть дано обыкновенное дифференциальное уравнение

$$F(x, y(x), y'(x), \dots, y^{(n)}(x)) = 0, \quad a \leq x \leq b \quad (7.1)$$

и краевые условия

$$\begin{aligned} \varphi_i(y(a), y'(a), \dots, y^{(n-1)}(a)) &= 0, \quad i = 1, 2, \dots, L; \\ \psi_j(y(b), y'(b), \dots, y^{(n-1)}(b)) &= 0, \quad j = L + 1, \dots, n, \end{aligned} \quad (7.2)$$

где $F(x, y, y', \dots, y^{(n)})$; $\varphi_i(y, y', \dots, y^{(n-1)})$, $i = \overline{1, L}$; $\psi_j(y, y', \dots, y^{(n-1)})$, $j = \overline{L+1, n}$ — функции указанных аргументов, заданные в некоторой области их изменения; L и $(n - L)$ — число условий на левом и правом концах отрезка $[a, b]$ соответственно. Общее количество условий равно порядку дифференциального уравнения.

Требуется найти функцию $y = y(x)$, которая на отрезке $[a, b]$ удовлетворяет уравнению (7.1), а на концах отрезка — краевым условиям (7.2).

Если уравнения (7.1), (7.2) линейны относительно искомой функции и ее производных, то краевая задача называется *линейной*.

Для простоты ограничимся частным случаем линейной краевой задачи для дифференциального уравнения второго порядка ($n = 2$), которая наиболее часто ставится в вычислительной практике и записывается в виде

$$y'' + p(x)y' - q(x)y = f(x), \quad a \leq x \leq b, \quad (\Omega \equiv [a, b]); \quad (7.3)$$

$$\begin{aligned} \alpha_0 y(a) + \beta_0 y'(a) &= A; \\ \alpha_1 y(b) + \beta_1 y'(b) &= B, \end{aligned} \quad (7.4)$$

где $p(x)$, $q(x)$, $f(x) \in C_2[a, b]$ — заданные функции, а α_0 , α_1 , β_0 , β_1 , A , B — заданные числа, $\alpha_j^2 + \beta_j^2 > 0$, $j = 0, 1$.

Требуется найти функцию $y(x)$, удовлетворяющую уравнению (7.3) и краевым условиям (7.4).

Краевые условия при $\alpha_j \neq 0, \beta_j \neq 0, j = 0, 1$, задают линейную связь между значениями искомого решения и его производной на концах отрезка $[a, b]$.

В простейшем случае, когда $\beta_0 = 0, \beta_1 = 0$, краевые условия задают на концах отрезка $[a, b]$ только значения функции $y(a), y(b)$. Такие *функциональные условия* называют *краевыми условиями первого рода*. В этом случае краевая задача называется *первой краевой задачей*.

В случае, когда $\alpha_0 = 0, \alpha_1 = 0$, т. е. на концах отрезка заданы только значения производных, краевые условия являются *дифференциальными*. Такие краевые условия называют *условиями второго рода*, или «*мягкими*». Последнее название обусловлено тем, что они определяют на концах отрезка $[a, b]$ всего лишь наклоны интегральных кривых, а не значения функции $y(x)$. В этом случае задача (7.3), (7.4) называется *второй краевой задачей*.

В общем случае, когда α_0 и/или $\alpha_1; \beta_0$ и/или β_1 не равны нулю, краевые условия носят *функционально-дифференциальный* характер и называются *условиями третьего рода*. Тогда задача (7.3), (7.4) называется *третьей краевой задачей*.

Например, условия

$$y(a) = A, \quad y(b) = B$$

являются условиями первого рода. Геометрически это означает, что при решении первой краевой задачи требуется найти интегральную кривую уравнения (7.3), проходящую через данные точки $(a, A), (b, B)$ (рис. 7.1a).

Условия

$$y'(a) = A, \quad y'(b) = B$$

являются условиями второго рода. Геометрически вторая краевая задача сводится к отысканию интегральной кривой уравнения, пересекающей прямые $x = a, x = b$ под заданными углами α, β , где $\operatorname{tg} \alpha = A, \operatorname{tg} \beta = B$ (рис. 7.1б).

Условия

$$y'(a) = A, \quad y(b) = B$$

являются частным случаем краевых условий третьего рода, так как $\alpha_0 = 0, \beta_0 = 1, \alpha_1 = 1, \beta_1 = 0$. Геометрически данная краевая задача сводится к отысканию интегральной кривой уравнения, пересекающей прямые $x = a, x = b$ под заданными углами α, β , где $\operatorname{tg} \alpha = A, \operatorname{tg} \beta = B$ (рис. 7.1в).

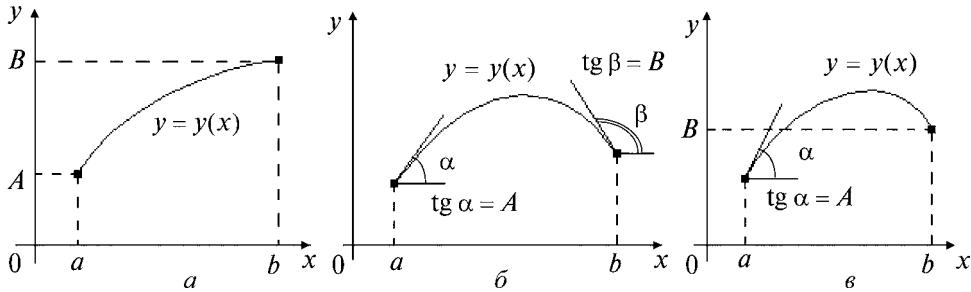


Рис. 7.1

канию интегральной кривой уравнения, проходящей через точку (b, B) и пересекающей прямую $x = a$ под данным углом α , где $\operatorname{tg}\alpha = A$ (рис. 7.1в).

В общем случае краевая задача может:

- а) иметь единственное решение;
- б) не иметь решений;
- в) иметь несколько или бесконечно много решений.

Утверждение 7.1 (о существовании и единственности решения краевой задачи (7.3), (7.4)) [6].

Для того чтобы существовало единственное решение краевой задачи (7.3), (7.4), необходимо и достаточно, чтобы однородная краевая задача

$$\begin{aligned}y'' + p(x)y' - q(x)y &= 0, \quad a \leq x \leq b; \\ \alpha_0y(a) + \beta_0y'(a) &= 0; \\ \alpha_1y(b) + \beta_1y'(b) &= 0\end{aligned}$$

имела только тривиальное решение $y(x) \equiv 0$.

Пример 7.1. Найти аналитическое решение следующих краевых задач:

- а) $y'' + y = 1$, $0 \leq x \leq \frac{\pi}{2}$, $y'(0) = 0$, $y\left(\frac{\pi}{2}\right) - y'\left(\frac{\pi}{2}\right) = 2$ (третья краевая задача);
- б) $y'' + y = -x$, $0 \leq x \leq 1$, $y(0) = 0$, $y(1) = 0$ (первая краевая задача).

□ Воспользуемся известной методикой отыскания общих решений дифференциальных уравнений [33]. Подставив в них заданные краевые условия, получим аналитические решения данных краевых задач.

1. Найдем общее решение однородного уравнения $y'' + y = 0$, одинакового для обеих рассматриваемых задач. Так как характеристическое уравнение $\lambda^2 + 1 = 0$ имеет комплексные сопряженные корни $\lambda_{1,2} = \pm i = \alpha \pm \beta i$ ($\alpha = 0$, $\beta = 1$), то общее решение будет $y_0(x) = e^{\alpha x}(C_1 \cos \beta x + C_2 \sin \beta x) = C_1 \cos x + C_2 \sin x$.

2. Частные решения неоднородных уравнений находятся методом подбора.

Подставляя $y_q(x) = C$ в уравнение $y'' + y = 1$, а $y_q(x) = Dx$ в уравнение $y'' + y = -x$, получаем $C = 1$, $D = -1$. Поэтому $y_q(x) = 1$ в случае «а», $y_q(x) = -x$ в случае «б».

3. Найдем общее решение неоднородного уравнения как сумму общего решения однородного уравнения и частного решения неоднородного уравнения:

- а) $y(x) = C_1 \cos x + C_2 \sin x + 1$;
- б) $y(x) = C_1 \cos x + C_2 \sin x - x$.

4. Определим значения произвольных постоянных из краевых условий третьего рода (случай «а») и первого рода (случай «б»):

- а) найдем $y'(x) = -C_1 \sin x + C_2 \cos x$. Тогда

$$y'(0) = C_2 = 0, \quad y\left(\frac{\pi}{2}\right) - y'\left(\frac{\pi}{2}\right) = C_2 + 1 + C_1 = 2.$$

Отсюда $C_1 = 1$ и $y(x) = 1 + \cos x$ — решение краевой задачи «а»;

б) $y(0) = C_1 = 0$, $y(1) = C_1 \cos 1 + C_2 \sin 1 - 1 = 0$. Отсюда $C_2 = \frac{1}{\sin 1}$ и $y(x) = \frac{\sin x}{\sin 1} - 1$ — решение краевой задачи «б».

Таким образом, решение краевой задачи представляет собой такое частное решение, которое удовлетворяет краевым условиям. ■

Рассмотренный метод нахождения аналитического решения краевых задач применим для ограниченного класса задач. Поэтому в вычислительной практике используются численные и приближенно-аналитические методы, позволяющие найти приближенное решение краевых задач, точные аналитические решения которых не могут быть найдены.

7.2. МЕТОД СЕТОК

Рассмотрим линейную краевую задачу с краевыми условиями первого рода (первую краевую задачу):

$$\begin{aligned} y'' + p(x)y' - q(x)y &= f(x), \quad a \leq x \leq b; \\ y(a) &= A, \quad y(b) = B, \end{aligned} \tag{7.5}$$

где $p(x), q(x), f(x) \in C_2[a, b]$ — заданные функции; A, B — заданные числа.

Очевидно, любой отрезок $[a, b]$, на котором ищется решение краевой задачи, можно привести к отрезку $[0; 1]$ с помощью линейного преобразования $\tilde{x} = \frac{x-a}{b-a}$. Действительно, тогда новая переменная $\tilde{x} \in [0; 1]$. В результате без ограничения общности краевая задача (7.5) может быть решена сначала на отрезке $[0; 1]$, а затем это решение с помощью преобразования $x = a + (b-a) \cdot \tilde{x}$ может быть записано на отрезке $[a, b]$. То же относится и к исследованию свойств полученного решения.

Утверждение 7.2 (о единственности решения краевой задачи (7.5)) [6].

Если функции $p(x), q(x), f(x)$ принадлежат классу $C_2[0; 1]$, $q(x) \geq 0$ на $[0; 1]$, то краевая задача (7.5) имеет единственное решение $y(x) \in C_4[0; 1]$.

Для решения задачи (7.5) применим метод сеток, получаемый путем аппроксимации первой и второй производных (см. также п. 6.2.1).

Введем равномерную сетку

$$\Omega_n = \{x_0, x_1, \dots, x_n\}, \quad x_i = x_0 + ih, \quad i = \overline{0, n}; \quad h = \frac{b-a}{n},$$

где n — число отрезков разбиения.

Функции $p(x), q(x), f(x)$ заменяются их проекциями на сетку Ω_n , т. е. $p(x) \rightarrow p(x_i) = p_i, q(x) \rightarrow q(x_i) = q_i, f(x) \rightarrow f(x_i) = f_i$. Вместо точного решения $y(x)$ отыскивается некоторое приближение $\hat{y}_i = \hat{y}(x_i) \approx y(x_i)$, $i = \overline{0, n}$.

Первая и вторая производные аппроксимируются на трехточечном шаблоне (x_{i-1}, x_i, x_{i+1}) по формулам второго порядка (5.10), (5.14):

$$y'(x_i) \cong \hat{y}'_i = \frac{\hat{y}_{i+1} - \hat{y}_{i-1}}{2h}; \quad y''(x_i) \cong \hat{y}''_i = \frac{\hat{y}_{i-1} - 2\hat{y}_i + \hat{y}_{i+1}}{h^2}.$$

Краевые условия для этой задачи аппроксимируются точно, т. е. $y(a)$ и $y(b)$ заменяются на \hat{y}_0 и \hat{y}_n . После замены от дифференциальной задачи (7.5) переходим к разностной схеме:

$$\begin{aligned} \frac{\hat{y}_{i+1} - 2\hat{y}_i + \hat{y}_{i-1}}{h^2} + p_i \frac{\hat{y}_{i+1} - \hat{y}_{i-1}}{2h} - q_i \hat{y}_i &= f_i, \quad i = \overline{1, n-1}; \\ \hat{y}_0 &= A, \quad \hat{y}_n = B, \end{aligned}$$

представляющей собой систему алгебраических уравнений трехдиагонального вида:

$$\begin{aligned} \alpha_i \hat{y}_{i-1} - \beta_i \hat{y}_i + \gamma_i \hat{y}_{i+1} &= \delta_i, \quad i = \overline{1, n-1}; \\ \hat{y}_0 &= A, \quad \hat{y}_n = B, \end{aligned} \quad (7.6)$$

где $\alpha_i = \frac{1}{h^2} - \frac{p_i}{2h}$, $\gamma_i = \frac{1}{h^2} + \frac{p_i}{2h}$, $\beta_i = \frac{2}{h^2} + q_i$, $\delta_i = f_i$. Здесь система (7.6) записана для внутренних узлов сетки Ω_n . Она является трехдиагональной системой линейных алгебраических уравнений и решается методом прогонки (см. п. 1.2.2).

Замечания

1. Изложенный метод сеток допускает обобщение. Например, его можно применять для решения нелинейной краевой задачи:

$$y'' = F(x, y), \quad x \in [a, b], \quad y(a) = A, \quad y(b) = B, \quad (7.7)$$

где $F(x, y)$ — нелинейная по y функция (в общем случае, который здесь не рассматривается, функция F зависит также и от y').

Рассуждая аналогично рассмотренному выше способу, перейдем к разностной задаче:

$$\begin{aligned} \hat{y}_{i-1} - 2\hat{y}_i + \hat{y}_{i+1} &= h^2 F(x_i, \hat{y}_i), \quad i = \overline{1, n-1}; \\ \hat{y}_0 &= A, \quad \hat{y}_n = B. \end{aligned} \quad (7.8)$$

В силу нелинейности правой части полученная алгебраическая система является нелинейной и для ее решения нельзя использовать метод прогонки в том виде, в каком он изложен для линейной задачи. Поэтому для ее решения используем метод простых итераций (см. п. 3.2.2), с помощью которого при фиксированном k (k — номер итерации) система алгебраических уравнений (7.8) превращается в линейную, так как величины, входящие в правую часть системы, известны из предыдущей итерации. Действительно, для k -й итерации получается система

$$\begin{aligned} \hat{y}_{i-1}^{(k)} - 2\hat{y}_i^{(k)} + \hat{y}_{i+1}^{(k)} &= h^2 F(x_i, \hat{y}_i^{(k-1)}), \quad i = \overline{1, n-1}; \\ \hat{y}_0^{(k)} &= A, \quad \hat{y}_n^{(k)} = B, \end{aligned}$$

которая решается на каждой итерации методом прогонки.

Можно показать [14], что итерации сходятся при выполнении условия $q = \frac{1}{8}(x_n - x_0)^2 M_1 < 1$, $M_1 = \max_{[a,b]} \left| \frac{\partial F}{\partial y} \right|$ с линейной скоростью.

2. Краевые условия второго и третьего рода в задаче, аналогичной (7.5), могут быть аппроксимированы несколькими способами.

Первый способ. Использование аппроксимационных формул (5.4) первого порядка

$$\hat{y}'_0 = \frac{\hat{y}_1 - \hat{y}_0}{h}, \quad \hat{y}'_n = \frac{\hat{y}_n - \hat{y}_{n-1}}{h}. \quad (7.9)$$

В силу первого порядка этих аппроксимаций метод сеток в этом случае также будет иметь первый порядок аппроксимации.

Второй способ. Применение формулы Тейлора (B.18) и ее преобразование с использованием дифференциального уравнения. Таким способом может быть достигнут второй порядок аппроксимации.

Третий способ. Применение левосторонней (5.8) и правосторонней (5.9) формул, аппроксимирующих производные со вторым порядком:

$$\hat{y}'_0 = \frac{1}{2h}(-3\hat{y}_0 + 4\hat{y}_1 - \hat{y}_2), \quad \hat{y}'_n = \frac{1}{2h}(\hat{y}_{n-2} - 4\hat{y}_{n-1} + 3\hat{y}_n). \quad (7.10)$$

3. Порядок аппроксимации схемы определяется минимальным порядком аппроксимации дифференциального уравнения и краевых условий.

Алгоритм применения метода сеток

1. Задать сетку Ω_n на отрезке $[a, b]$ или сформировать ее из условий достижения требуемой точности.

2. Используя аппроксимационные формулы (5.10), (5.14) и один из трех способов аппроксимации краевых условий (в случае, если они второго или третьего рода), перейти от исходной дифференциальной задачи к системе алгебраических уравнений (разностной схеме), неизвестными в которой являются величины, «близкие» к решению краевой задачи в узлах сетки.

3. Найти решение разностной задачи путем решения трехдиагональной системы уравнений и таким образом определить приближенное решение краевой задачи.

Пример 7.2. Найти приближенное решение краевой задачи

$$y'' + y = 1, \quad 0 \leq x \leq \frac{\pi}{2}; \\ y'(0) = 0, \quad y\left(\frac{\pi}{2}\right) - y'\left(\frac{\pi}{2}\right) = 2$$

при $n = 3$, используя первый способ аппроксимации краевых условий. Записать разностные схемы для второго и третьего способов при произвольном n .

□ В поставленной задаче $a = 0$, $b = \frac{\pi}{2}$, $p(x) = 0$, $q(x) = -1$, $f(x) = 1$, $\alpha_0 = 0$, $\beta_0 = 1$, $A = 0$, $\alpha_1 = 1$, $\beta_1 = -1$, $B = 2$. Для решения задачи воспользуемся методикой.

1. Так как $n = 3$, то сетка имеет вид $\Omega_3 = \{x_0, x_1, x_2, x_3\}$, где $x_i = ih$, $i = \overline{0, 3}$, $h = \frac{\frac{\pi}{2} - 0}{3} = \frac{\pi}{6}$. Тогда $y(0) = y_0$, $y\left(\frac{\pi}{6}\right) = y_1$, $y\left(\frac{\pi}{3}\right) = y_2$, $y\left(\frac{\pi}{2}\right) = y_3$. Будем искать приближенные значения \hat{y}_0 , \hat{y}_1 , \hat{y}_2 , \hat{y}_3 . Проекции функций $p(x)$, $q(x)$, $f(x)$ на сетку имеют вид $p_i = 0$, $q_i = -1$, $f_i = 1$.

2. Составим разностную схему. Согласно (7.6), для внутренних узлов сетки получаем

$$\frac{\hat{y}_{i+1} - 2\hat{y}_i + \hat{y}_{i-1}}{h^2} + \hat{y}_i = 1, \quad i = 1, 2, \text{ или } \hat{y}_{i-1} - (2 - h^2)\hat{y}_i + \hat{y}_{i+1} = h^2, \quad i = 1, 2.$$

Применим *первый способ* аппроксимации краевых условий. По формуле (5.4) с учетом условия $y'(0) = 0$ на левом конце имеем

$$\hat{y}'_0 = \frac{\hat{y}_1 - \hat{y}_0}{h} = 0 \text{ или } \hat{y}_1 - \hat{y}_0 = 0.$$

На правом конце $y\left(\frac{\pi}{2}\right) = y_3$, $y'\left(\frac{\pi}{2}\right) = y'_3$, и по второй из формул (7.9)

$$\hat{y}'_3 = \frac{\hat{y}_3 - \hat{y}_2}{h}.$$

Тогда краевое условие $y\left(\frac{\pi}{2}\right) - y'\left(\frac{\pi}{2}\right) = 2$ аппроксимируется выражением

$$\hat{y}_3 - \frac{\hat{y}_3 - \hat{y}_2}{h} = 2 \text{ или } \hat{y}_2 - (1-h)\hat{y}_3 = 2h.$$

В результате получаем разностную схему первого порядка аппроксимации (трехдиагональную систему линейных алгебраических уравнений)

$$\begin{aligned} \hat{y}_1 - \hat{y}_0 &= 0; \\ \hat{y}_{i-1} - (2-h^2)\hat{y}_i + \hat{y}_{i+1} &= h^2, \quad i = 1, 2; \\ \hat{y}_2 - (1-h)\hat{y}_3 &= 2h. \end{aligned}$$

Сравнивая первое уравнение этой системы с рекуррентным соотношением $\hat{y}_i = P_i \cdot \hat{y}_{i+1} + Q_i$ метода прогонки (см. п. 1.2.2), характеризующим обратный ход, получаем $P_0 = 1$, $Q_0 = 0$.

После этого вычисляются все последующие прогоночные коэффициенты по формулам (см. п. 1.2.2):

$$P_i = \frac{\gamma_i}{\beta_i - \alpha_i P_{i-1}}, \quad Q_i = \frac{\alpha_i Q_{i-1} - \delta_i}{\beta_i - \alpha_i P_{i-1}}, \quad i = 1, 2, 3.$$

Здесь α_i , β_i , γ_i соответствуют коэффициентам левой части полученной алгебраической системы, а δ_i — правой части.

Далее выполняется обратный ход:

$$\hat{y}_3 = Q_3, \quad \hat{y}_2 = P_2 \hat{y}_3 + Q_2, \quad \hat{y}_1 = P_1 \hat{y}_2 + Q_1.$$

Результаты решения краевой задачи приведены в таблице 7.1, в которой последний столбец соответствует точному решению $y(x) = 1 + \cos x$, найденному в примере 7.1.

В силу того, что краевые условия аппроксимированы с первым порядком относительно h , в данном случае получена разностная схема первого порядка,

Таблица 7.1

i	α_i	β_i	γ_i	δ_i	P_i	Q_i	\hat{y}_i	$y(x)$
0	0	-1,0000	-1	0,00000	1,00000	0	1,8648	2,0000
1	1	1,72584	1	0,27415	1,37771	-0,37770	1,8648	1,8666
2	1	1,72584	1	0,27415	2,87240	-1,87242	1,6277	1,5000
3	1	0,47640	—	1,04200	—	1,21853	1,21853	1,0000

так как порядок аппроксимации схемы определяется минимальным порядком аппроксимации дифференциального уравнения и краевых условий.

Воспользуемся *вторым способом* аппроксимации краевых условий для построения разностной схемы второго порядка аппроксимации. Разложим $y(x)$ в точке $x = x_1$ относительно точки x_0 по формуле Тейлора (B.18):

$$y_1 = y_0 + hy'(x_0) + \frac{h^2}{2}y''(x_0) + O(h^3).$$

Выразим из этого соотношения $y'(x_0)$ и подставим в него вместо $y''(x_0)$ выражение $y''(x_0) = 1 - y(x_0) = 1 - y_0$, определяемое исходным дифференциальным уравнением:

$$y'(x_0) = \frac{y_1 - y_0}{h} - \frac{h}{2}(1 - y_0) + O(h^2).$$

Как показывает это соотношение, дифференциальное условие на левой границе аппроксимируется на двухточечном шаблоне (x_0, x_1) со вторым порядком аппроксимации двухточечным алгебраическим уравнением:

$$\frac{\hat{y}_1 - \hat{y}_0}{h} - \frac{h}{2}(1 - \hat{y}_0) = 0 \text{ или } \frac{\hat{y}_1}{h} - \frac{1}{2h}(2 - h^2)\hat{y}_0 = \frac{h}{2}.$$

Аналогично получается двухточечное алгебраическое уравнение при $i = n - 1$ и $i = n$. Разложение $y(x)$ в точке $x = x_{n-1}$ относительно точки x_n по формуле Тейлора (B.18) имеет вид

$$y_{n-1} = y_n - h \cdot y'(x_n) + \frac{h^2}{2}y''(x_n) + O(h^3).$$

Выражая отсюда $y'(x_n)$ с учетом связи $y''(x_n) = 1 - y(x_n) = 1 - y_n$, следующей из исходного дифференциального уравнения, получаем

$$y'(x_n) = -\frac{y_{n-1}}{h} + \frac{y_n}{h} + \frac{h}{2}y''(x_n) = -\frac{y_{n-1}}{h} + \frac{y_n}{h} + \frac{h}{2}(1 - y_n).$$

Подставим это выражение в граничное условие:

$$y_n + \frac{y_{n-1}}{h} - \frac{y_n}{h} - \frac{h}{2}(1 - y_n) = 2 \text{ или } \frac{\hat{y}_{n-1}}{h} + \frac{1}{2h}(h^2 + 2h - 2)\hat{y}_n = 2 + \frac{h}{2}.$$

Таким образом, система линейных алгебраических уравнений в окончательном виде записывается следующим образом:

$$\begin{aligned} \frac{1}{2h}(h^2 - 2)\hat{y}_0 + \frac{1}{h}\hat{y}_1 &= \frac{h}{2}; \\ \hat{y}_{i-1} - (2 - h^2)\hat{y}_i + \hat{y}_{i+1} &= h^2, \quad i = \overline{1, n-1}; \\ \frac{\hat{y}_{n-1}}{h} + \frac{1}{2h}(h^2 + 2h - 2)\hat{y}_n &= 2 + \frac{h}{2}. \end{aligned}$$

Эта трехдиагональная система, отличающаяся от полученной первым способом только первым и последним уравнениями, решается численно методом прогонки.

Применим *третий способ аппроксимации краевых условий* для построения разностной схемы второго порядка. Так, для крайней левой точки используется левосторонняя формула (5.8):

$$\hat{y}'_0 = \frac{1}{2h}(-3\hat{y}_0 + 4\hat{y}_1 - \hat{y}_2).$$

Тогда получается трехточечное алгебраическое уравнение:

$$\frac{1}{2h}(-3\hat{y}_0 + 4\hat{y}_1 - \hat{y}_2) = 0 \text{ или } 3\hat{y}_0 - 4\hat{y}_1 + \hat{y}_2 = 0.$$

Аппроксимация производной $y'(\frac{\pi}{2})$ в крайней правой точке по правосторонней формуле $\hat{y}'_n = \frac{1}{2h}(\hat{y}_{n-2} - 4\hat{y}_{n-1} + 3\hat{y}_n)$ приводит к трехточечному алгебраическому уравнению:

$$\hat{y}_n - \frac{1}{2h}(\hat{y}_{n-2} - 4\hat{y}_{n-1} + 3\hat{y}_n) = 2 \text{ или } \hat{y}_{n-2} - 4\hat{y}_{n-1} + (3 - 2h)\hat{y}_n = -4h.$$

Тогда в этом случае получается следующая система линейных алгебраических уравнений:

$$\begin{aligned} 3\hat{y}_0 - 4\hat{y}_1 + \hat{y}_2 &= 0; \\ \hat{y}_{i-1} - (2 - h^2)\hat{y}_i + \hat{y}_{i+1} &= h^2, \quad i = \overline{1, n-1}; \\ \hat{y}_{n-2} - 4\hat{y}_{n-1} + (3 - 2h)\hat{y}_n &= -4h. \end{aligned}$$

Здесь \hat{y}_2 в первом уравнении и \hat{y}_{n-2} в последнем нарушают ее трехдиагональный характер. В этом случае система приводится к трехдиагональному виду путем исключения \hat{y}_2 и \hat{y}_{n-2} из первых двух и последних двух уравнений системы и после этого решается методом прогонки. ■

7.3. МЕТОДЫ МИНИМИЗАЦИИ НЕВЯЗКИ

Описываемые здесь методы относятся к приближенно-аналитическим и могут применяться при решении достаточно широкого класса задач. На основе одного из приближенно-аналитических методов (метода Галеркина) строится метод конечных элементов, излагаемый в п. 7.5.

Рассмотрим линейную краевую задачу (7.3), (7.4). Ее решение будем искать в виде

$$\hat{y}_m(x) = \varphi_0(x) + a_1 \cdot \varphi_1(x) + \dots + a_m \cdot \varphi_m(x), \quad (7.11)$$

где $\varphi_0(x), \varphi_1(x), \dots, \varphi_m(x)$ — элементы заданной системы функций; a_1, \dots, a_m — неопределенные коэффициенты. Заданная система функций называется *базисной*, и ее элементы должны удовлетворять условиям:

- а) $\varphi_j(x) \in C_2[a, b]$, $j = \overline{0, m}$;
- б) при любом конечном m функции $\varphi_1(x), \dots, \varphi_m(x)$ линейно независимы на отрезке $[a, b]$;

в) $\phi_0(x)$ удовлетворяет краевым условиям (7.4)

$$\begin{aligned}\alpha_0 \cdot \phi_0(a) + \beta_0 \cdot \phi'_0(a) &= A; \\ \alpha_1 \cdot \phi_0(b) + \beta_1 \cdot \phi'_0(b) &= B.\end{aligned}\quad (7.12)$$

г) $\phi_1(x), \dots, \phi_m(x)$ удовлетворяют условиям

$$\begin{aligned}\alpha_0 \cdot \phi_j(a) + \beta_0 \cdot \phi'_j(a) &= 0, \quad j = \overline{1, m}; \\ \alpha_1 \cdot \phi_j(b) + \beta_1 \cdot \phi'_j(b) &= 0, \quad j = \overline{1, m}.\end{aligned}\quad (7.13)$$

Функция

$$\varepsilon(x; a_1, \dots, a_m) = \hat{y}_m''(x) + p(x)\hat{y}'_m(x) - q(x)\hat{y}_m(x) - f(x) \quad (7.14)$$

называется *невязкой*. Она равна разности левой и правой частей уравнения (7.3), образующейся при подстановке $\hat{y}_m(x)$ вместо $y(x)$ в дифференциальное уравнение, и характеризует степень отклонения функции $\hat{y}_m(x)$ от точного решения краевой задачи. Если при некоторых значениях коэффициентов a_1, \dots, a_m невязка тождественно равна нулю на отрезке $[a, b]$, а именно

$$\varepsilon(x; a_1, \dots, a_m) \equiv 0, \quad a \leq x \leq b, \quad (7.15)$$

то функция $\hat{y}_m(x)$ совпадает с точным решением краевой задачи (7.3), (7.4), так как удовлетворяются и уравнение, и краевые условия.

Однако при решении краевых задач, как правило, не удается получить невязку тождественно равной нулю. Поэтому ставится задача: вычислить коэффициенты a_1, \dots, a_m таким образом, чтобы невязка в каком-либо смысле стала меньшей. Полученные в результате коэффициенты определяют приближенное решение (7.11).

Выражение для невязки $\varepsilon(x; a_1, \dots, a_m)$ с учетом (7.11) удобно записывать в следующей эквивалентной форме:

$$\varepsilon(x; a_1, \dots, a_m) = L\hat{y}_m - f(x) = L\phi_0(x) + \sum_{j=1}^m a_j \cdot L\phi_j(x) - f(x), \quad (7.16)$$

где $L\hat{y}_m \equiv \hat{y}_m''(x) + p(x)\hat{y}'_m(x) - q(x)\hat{y}_m(x)$, L — линейный оператор задачи (7.3), (7.4) (выполняются равенства $L(y + z) = Ly + Lz$, $L(Cy) = C \cdot Ly$ для любых y, z и постоянной C).

Рассмотрим различные методы, минимизирующие невязку.

А. Метод коллокации. На интервале (a, b) задаются m точек x_1, \dots, x_m (*точки коллокации*) и требуется, чтобы в каждой из них невязка (7.14) обращалась в нуль:

$$\begin{aligned}\varepsilon(x_1; a_1, \dots, a_m) &= 0; \\ &\vdots \\ \varepsilon(x_m; a_1, \dots, a_m) &= 0.\end{aligned}\quad (7.17)$$

С учетом (7.16) эта система принимает вид

$$\begin{aligned} a_1 L\varphi_1(x_1) + \dots + a_m L\varphi_m(x_1) &= f(x_1) - L\varphi_0(x_1); \\ a_1 L\varphi_1(x_2) + \dots + a_m L\varphi_m(x_2) &= f(x_2) - L\varphi_0(x_2); \\ &\vdots \\ a_1 L\varphi_1(x_m) + \dots + a_m L\varphi_m(x_m) &= f(x_m) - L\varphi_0(x_m). \end{aligned} \quad (7.18)$$

Если полученная система m линейных уравнений совместна, то из нее определяются коэффициенты a_1, \dots, a_m , которые затем подставляются в (7.11).

Б. Метод наименьших квадратов (непрерывный вариант). Неизвестные коэффициенты a_1, \dots, a_m должны обеспечивать минимум интеграла от квадрата невязки:

$$I = \int_a^b \varepsilon^2(x; a_1, \dots, a_m) dx \rightarrow \min_{a_1, \dots, a_m}.$$

Для решения задачи применяются необходимые условия безусловного экстремума [31]:

$$\frac{\partial I}{\partial a_j} = 2 \cdot \int_a^b \varepsilon(x; a_1, \dots, a_m) \frac{\partial \varepsilon(x; a_1, \dots, a_m)}{\partial a_j} dx = 0, \quad j = \overline{1, m}. \quad (7.19)$$

Подставляя (7.16) в (7.19), получаем систему m линейных алгебраических уравнений для нахождения коэффициентов a_1, \dots, a_m :

$$\begin{aligned} a_1(L\varphi_1, L\varphi_1) + \dots + a_m(L\varphi_m, L\varphi_1) &= (f - L\varphi_0, L\varphi_1); \\ a_1(L\varphi_1, L\varphi_2) + \dots + a_m(L\varphi_m, L\varphi_2) &= (f - L\varphi_0, L\varphi_2); \\ &\vdots \\ a_1(L\varphi_1, L\varphi_m) + \dots + a_m(L\varphi_m, L\varphi_m) &= (f - L\varphi_0, L\varphi_m), \end{aligned} \quad (7.20)$$

где $(f, g) = \int_a^b f(x)g(x)dx$ — скалярное произведение. В системе (7.20) все скалярные произведения предварительно вычисляются.

В. Метод наименьших квадратов (дискретный вариант). Неизвестные коэффициенты a_1, \dots, a_m должны обеспечивать минимум суммы квадратов значений невязки в заданном наборе точек x_1, \dots, x_n ; $n \geq m$; т. е. $x_i \in (a, b)$, $i = \overline{1, n}$:

$$S = \sum_{i=1}^n \varepsilon^2(x_i; a_1, \dots, a_m) \rightarrow \min_{a_1, \dots, a_m}.$$

Для решения задачи применяются необходимые условия безусловного экстремума [31]:

$$\frac{\partial S}{\partial a_j} = 2 \cdot \sum_{i=1}^n \varepsilon(x_i; a_1, \dots, a_m) \frac{\partial \varepsilon(x_i; a_1, \dots, a_m)}{\partial a_j} = 0, \quad j = \overline{1, m}. \quad (7.21)$$

Отсюда следует система m линейных уравнений для нахождения коэффициентов a_1, \dots, a_m , которая по форме записи совпадает с (7.20), но скалярное произведение определяется по формуле

$$(f, g) = \sum_{i=1}^n f(x_i)g(x_i).$$

Замечание. При $n = m$ результаты, полученные точечным методом наименьших квадратов и методом коллокации, совпадают. В этом случае точки x_1, \dots, x_n являются точками коллокации.

Г. Метод моментов (взвешенных невязок). Неизвестные коэффициенты a_1, \dots, a_m находятся из условия равенства нулю m моментов невязки:

$$I = \int_a^b \varepsilon(x; a_1, \dots, a_m) \cdot \psi_j(x) dx = 0, \quad j = \overline{1, m}, \quad (7.22)$$

где $\psi_j(x)$, $j = \overline{1, m}$ — функции, удовлетворяющие условиям:

- а) $\psi_j(x) \in C[a, b]$, $j = \overline{1, m}$;
- б) функции $\psi_j(x)$ являются элементами системы степеней x или системы тригонометрических функций.

Функции $\psi_j(x)$, $j = \overline{1, m}$, называются *весовыми*, а условие (7.22) является условием ортогональности невязки к весовым функциям.

Д. Метод Галеркина. Он является частным случаем метода моментов, когда в качестве весовых функций используются базисные. Коэффициенты a_1, \dots, a_m находятся из условия ортогональности функций базисной системы $\varphi_1(x), \dots, \varphi_m(x)$ к невязке:

$$\int_a^b \varepsilon(x; a_1, \dots, a_m) \varphi_j(x) dx = 0, \quad j = \overline{1, m}. \quad (7.23)$$

Отсюда следует система m линейных уравнений для нахождения коэффициентов:

$$\begin{aligned} a_1(L\varphi_1, \varphi_1) + \dots + a_m(L\varphi_m, \varphi_1) &= (f - L\varphi_0, \varphi_1); \\ a_1(L\varphi_1, \varphi_2) + \dots + a_m(L\varphi_m, \varphi_2) &= (f - L\varphi_0, \varphi_2); \\ &\vdots \\ a_1(L\varphi_1, \varphi_m) + \dots + a_m(L\varphi_m, \varphi_m) &= (f - L\varphi_0, \varphi_m), \end{aligned} \quad (7.24)$$

где $(f, g) = \int_a^b f(x)g(x)dx$ — скалярное произведение. В системе (7.24) все скалярные произведения предварительно вычисляются.

Известно, что при достаточно большом m условие (7.23) обеспечивает малость невязки в среднем [10].

Алгоритм применения методов минимизации невязки

1. В выражении (7.11) выбрать систему базисных функций, задать число m в зависимости от требуемой точности.
 2. Найти коэффициенты a_1, \dots, a_m путем решения одной из систем алгебраических уравнений (7.18), (7.20), (7.24) в зависимости от выбранного метода.
 3. Выписать приближенное решение краевой задачи по формуле (7.11).
- Пример 7.3.** Найти приближенное решение краевой задачи

$$\begin{aligned}y'' + y &= -x, \quad 0 \leq x \leq 1; \\y(0) &= 0, \quad y(1) = 0\end{aligned}$$

методом коллокации, интегральным методом наименьших квадратов, методом Галеркина.

В поставленной задаче $a = 0, b = 1, p(x) = 0, q(x) = -1, f(x) = -x, \alpha_0 = 1, \beta_0 = 0, A = 0, \alpha_1 = 1, \beta_1 = 0, B = 0, Ly = y'' + y$. Точное решение найдено в примере 7.1.

Воспользуемся сначала *методом коллокации*.

1. Зададим $m = 2$ и будем искать решение в виде

$$\hat{y}_2(x) = \varphi_0(x) + a_1 \cdot \varphi_1(x) + a_2 \cdot \varphi_2(x),$$

где $\varphi_0(x) \equiv 0$ (эта функция удовлетворяет каждому из краевых условий, т. е. $\varphi_0(0) = 0, \varphi_0(1) = 0$), функции $\varphi_1(x) = x(1-x)$, $\varphi_2(x) = x^2(1-x)$. Функции $\varphi_1(x), \varphi_2(x)$ линейно независимые, дважды непрерывно дифференцируемые и удовлетворяют условию (7.13). Действительно,

$$\begin{aligned}1 \cdot \varphi_j(0) + 0 \cdot \varphi'_j(0) &= 0, \quad j = 1, 2; \\1 \cdot \varphi_j(1) + 0 \cdot \varphi'_j(1) &= 0, \quad j = 1, 2.\end{aligned}$$

Таким образом, решение краевой задачи ищется в форме

$$\hat{y}_2(x) = a_1 \cdot x(1-x) + a_2 \cdot x^2(1-x).$$

2. Так как $m = 2$ и $\varphi_0(x) \equiv 0$, то система (7.18) имеет вид

$$\begin{aligned}a_1 L\varphi_1(x_1) + a_2 L\varphi_2(x_1) &= f(x_1); \\a_1 L\varphi_1(x_2) + a_2 L\varphi_2(x_2) &= f(x_2).\end{aligned}$$

Выберем узлы коллокации: $x_1 = \frac{1}{4}, x_2 = \frac{1}{2}$. Тогда

$$\begin{aligned}L\varphi_1(x) &= \varphi'_1(x) + \varphi_1(x) = -2 + x(1-x) = -2 + x - x^2; \\L\varphi_1(x_1) &= L\varphi_1\left(\frac{1}{4}\right) = -\frac{29}{16}; \quad L\varphi_1(x_2) = L\varphi_1\left(\frac{1}{2}\right) = -\frac{7}{4}; \\L\varphi_2(x) &= \varphi''_2(x) + \varphi_2(x) = 2 - 6x + x^2 - x^3; \\L\varphi_2(x_1) &= L\varphi_2\left(\frac{1}{4}\right) = \frac{35}{64}; \quad L\varphi_2(x_2) = L\varphi_2\left(\frac{1}{2}\right) = -\frac{7}{8}.\end{aligned}$$

Таким образом, имеем линейную систему относительно a_1 и a_2 :

$$\begin{aligned}\frac{29}{16}a_1 - \frac{35}{64}a_2 &= \frac{1}{4}; \\ \frac{7}{4}a_1 + \frac{7}{8}a_2 &= \frac{1}{2},\end{aligned}$$

решение которой дает

$$a_1 = \frac{6}{31}, \quad a_2 = \frac{40}{217}.$$

3. Приближенное решение задачи:

$$\hat{y}_2(x) = \frac{x(1-x)}{217}(42 + 40x).$$

Решим теперь задачу *методом наименьших квадратов* (см. непрерывный вариант).

1. Решение краевой задачи ищется в форме ($m = 2$):

$$\hat{y}_2(x) = a_1 \cdot x(1-x) + a_2 \cdot x^2(1-x).$$

2. Так как $f(x) = -x$, $\varphi_0(x) \equiv 0$, то система (7.20) имеет вид

$$\begin{aligned}a_1(L\varphi_1, L\varphi_1) + a_2(L\varphi_2, L\varphi_1) &= (-x, L\varphi_1); \\ a_1(L\varphi_1, L\varphi_2) + a_2(L\varphi_2, L\varphi_2) &= (-x, L\varphi_2),\end{aligned}$$

где

$$\begin{aligned}(L\varphi_1, L\varphi_1) &= \int_0^1 (-2 + x - x^2)^2 dx = \frac{101}{30}; \quad (L\varphi_2, L\varphi_2) = \int_0^1 (2 - 6x + x^2 - x^3)^2 dx = \frac{131}{35}; \\ (L\varphi_1, L\varphi_2) &= \int_0^1 (-2 + x - x^2)(2 - 6x + x^2 - x^3) dx = \frac{101}{60}; \\ (-x, L\varphi_1) &= \int_0^1 (-x)(-2 + x - x^2) dx = \frac{11}{12}; \quad (-x, L\varphi_2) = \int_0^1 (-x)(2 - 6x + x^2 - x^3) dx = \frac{19}{20}.\end{aligned}$$

Итак, имеем линейную систему относительно a_1 и a_2 :

$$\begin{aligned}\frac{101}{30}a_1 + \frac{101}{60}a_2 &= \frac{11}{12}; \\ \frac{101}{60}a_1 + \frac{131}{35}a_2 &= \frac{19}{20},\end{aligned}$$

решение которой дает $a_1 = 0,1875419$, $a_2 = 0,1694707$.

3. Приближенное решение задачи:

$$\hat{y}_2(x) = 0,1875419 \cdot x(1-x) + 0,1694707 \cdot x^2(1-x).$$

Решим задачу *методом Галеркина*.

1. Пусть сначала $m = 1$. Решение ищется в форме $\hat{y}_1(x) = a_1 \cdot x(1-x)$.

2. Тогда система (7.24) преобразуется к виду

$$a_1(L\varphi_1, \varphi_1) = (-x, \varphi_1) \text{ или } a_1(\varphi_1'' + \varphi_1) = (-x, \varphi_1).$$

Так как $\varphi_1(x) = x(1-x)$, $L\varphi_1(x) = \varphi_1''(x) + \varphi_1(x) = -2 + x(1-x)$, получаем

$$a_1 \int_0^1 (-2 + x(1-x)) \cdot x(1-x) dx = - \int_0^1 x^2(1-x) dx.$$

После вычисления интегралов имеем уравнение $-\frac{3}{10}a_1 = -\frac{1}{12}$, откуда $a_1 = \frac{5}{18}$.

3. Приближенное решение краевой задачи:

$$\hat{y}_1(x) = \frac{5}{18}x \cdot (1-x).$$

Пусть теперь $m = 2$.

1. Решение краевой задачи ищется в форме

$$\hat{y}_2(x) = a_1 \cdot x(1-x) + a_2 \cdot x^2(1-x).$$

2. Тогда система (7.24) имеет вид

$$\begin{aligned} a_1(L\varphi_1, \varphi_1) + a_2(L\varphi_2, \varphi_1) &= (f, \varphi_1); \\ a_1(L\varphi_1, \varphi_2) + a_2(L\varphi_2, \varphi_2) &= (f, \varphi_2). \end{aligned}$$

Так как $\varphi_1(x) = x(1-x)$, $\varphi_2(x) = x^2(1-x)$, $L\varphi_1(x) = \varphi_1''(x) + \varphi_1(x) = -2 + x(1-x)$, $L\varphi_2(x) = \varphi_2''(x) + \varphi_2(x) = 2 - 6x + x^2 - x^3$, получаем

$$\begin{aligned} a_1 \int_0^1 [-2 + x(1-x)] \cdot x(1-x) dx + a_2 \int_0^1 [2 - 6x + x^2(1-x)] \cdot x(1-x) dx &= - \int_0^1 x^2(1-x) dx; \\ a_1 \int_0^1 [-2 + x(1-x)] \cdot x^2(1-x) dx + a_2 \int_0^1 [2 - 6x + x^2(1-x)] \cdot x^2(1-x) dx &= - \int_0^1 x^3(1-x) dx. \end{aligned}$$

Вычисляя интегралы, находим

$$\begin{aligned} \frac{3}{10}a_1 + \frac{3}{20}a_2 &= \frac{1}{12}; \\ \frac{3}{20}a_1 + \frac{13}{105}a_2 &= \frac{1}{20}. \end{aligned}$$

Отсюда

$$a_1 = \frac{71}{369}, \quad a_2 = \frac{7}{41}.$$

3. Приближенное решение краевой задачи:

$$\hat{y}_2(x) = x(1-x) \cdot \left(\frac{71}{369} + \frac{7}{41}x \right).$$

x	$y_{\text{коллок}}$	$y_{\text{мет. наим. кв}}$	$y_{\text{Галеркина}}$	Точное решение
0,25	0,045	0,04311	0,0440	0,044014
0,50	0,071	0,06807	0,0698	0,069747
0,75	0,062	0,05899	0,0600	0,060050

Сопоставим полученные решения с точным (табл. 7.2).

Очевидно, метод Галеркина дал более точный результат. ■

Пример 7.4. Найти приближенное решение краевой задачи [42]

$$y'' + 2xy' - 2y = 2x^2, \quad 0 \leq x \leq 1;$$

$$y'(0) = -2, \quad y(1) + y'(1) = 0$$

методом Галеркина.

В поставленной задаче $a = 0, b = 1, p(x) = 2x, q(x) = 2, f(x) = 2x^2, \alpha_0 = 0, \beta_0 = 1, A = -2, \alpha_1 = 1, \beta_1 = 1, B = 0, Ly = y'' + 2xy' - 2y$.

1. Зададим $m = 2$ и подберем функции $\varphi_0(x), \varphi_1(x), \varphi_2(x)$, используя систему 1, x, x^2, \dots . Функция $\varphi_0(x)$ должна удовлетворять условиям (7.12):

$$\varphi'_0(0) = -2, \quad \varphi_0(1) + \varphi'_0(1) = 0.$$

Пусть $\varphi_0(x) = b + cx$, где b, c — неопределенные коэффициенты. Тогда

$$\varphi'_0(0) = c = -2, \quad \varphi_0(1) + \varphi'_0(1) = b + c + c = b + 2c = 0.$$

Отсюда $b = 4$ и $\varphi_0(x) = 4 - 2x$.

Функции $\varphi_1(x), \varphi_2(x)$ должны удовлетворять условиям (7.13):

$$\varphi'_j(0) = 0, \quad \varphi_j(1) + \varphi'_j(1) = 0, \quad j = 1, 2.$$

Первое условие выполняется для функций вида $\varphi_j(x) = x^{j+1} + b_j$. Значения b_j находятся из второго условия $1 + b_j + j + 1 = 0$, откуда $b_j = -j - 2$. Тогда получаем $\varphi_1(x) = x^2 - 3, \varphi_2(x) = x^3 - 4$.

Таким образом, решение краевой задачи ищется в форме

$$\hat{y}_2(x) = 4 - 2x + a_1 \cdot (x^2 - 3) + a_2 \cdot (x^3 - 4).$$

2. Тогда система (7.24) имеет вид

$$a_1(L\varphi_1, \varphi_1) + a_2(L\varphi_2, \varphi_1) = (f - L\varphi_0, \varphi_1),$$

$$a_1(L\varphi_1, \varphi_2) + a_2(L\varphi_2, \varphi_2) = (f - L\varphi_0, \varphi_2),$$

где

$$L\varphi_0(x) = \varphi''_0(x) + 2x\varphi'_0(x) - 2\varphi_0(x) = 0 + 2x(-2) - 8 + 4x = -8;$$

$$L\varphi_1(x) = \varphi''_1(x) + 2x\varphi'_1(x) - 2\varphi_1(x) = 2 + 2x2x - 2x^2 + 6 = 2x^2 + 8;$$

$$L\varphi_2(x) = \varphi''_2(x) + 2x\varphi'_2(x) - 2\varphi_2(x) = 6x + 2x3x^2 - 2x^3 + 8 = 4x^3 + 6x + 8.$$

$$(L\varphi_1, \varphi_1) = \int_0^1 (2x^2 + 8)(x^2 - 3) dx = -22,93333;$$

$$(L\varphi_2, \varphi_1) = \int_0^1 (4x^3 + 6x + 8)(x^2 - 3)dx = -31,16667;$$

$$(L\varphi_1, \varphi_2) = \int_0^1 (2x^2 + 8)(x^3 - 4)dx = -32,33333;$$

$$(L\varphi_2, \varphi_2) = \int_0^1 (4x^3 + 6x + 8)(x^3 - 4)dx = -44,22857;$$

$$(f - L\varphi_0, \varphi_1) = \int_0^1 (2x^2 + 8)(x^2 - 3)dx = -22,93333;$$

$$(f - L\varphi_0, \varphi_2) = \int_0^1 (2x^2 + 8)(x^3 - 4)dx = -32,33333.$$

Окончательно получаем

$$22,93333 \cdot a_1 + 31,16667 \cdot a_2 = 22,93333;$$

$$32,33333 \cdot a_1 + 44,22857 \cdot a_2 = 32,33333.$$

Отсюда $a_1 = 1$, $a_2 = 0$.

3. Приближенное решение краевой задачи $\hat{y}_2(x) = x^2 - 2x + 1$. ■

7.4. МЕТОДЫ СВЕДЕНИЯ КРАЕВОЙ ЗАДАЧИ К ЗАДАЧЕ КОШИ

7.4.1. МЕТОД СТРЕЛЬБЫ

Суть этого метода заключается в сведении решения краевой задачи к многократному решению задачи Коши. Принцип построения метода стрельбы рассмотрим на примере нелинейной краевой задачи:

$$y'' = f(x, y, y'), \quad a \leq x \leq b; \quad (7.25)$$

$$y(a) = A, \quad y(b) = B, \quad (7.26)$$

где $f(x, y, y')$ — нелинейная функция, обусловливающая нелинейность дифференциального уравнения (7.25).

При введении новой переменной $z = y'$ уравнение (7.25) записывается в нормальной форме Коши, а краевые условия видоизменяются:

$$\begin{aligned} y' &= z, \quad y(a) = A; \\ z' &= f(x, y, z), \quad z(a) = \eta, \end{aligned} \quad (7.27)$$

где $\eta = y'(a) = \operatorname{tg}\alpha$ — параметр, равный тангенсу угла наклона интегральной кривой в точке $x = a$. Угол α (параметр η) в процессе многократного решения краевой задачи должен принять такое значение, чтобы интегральная кривая «попала в цель», т. е. в точку (b, B) (рис. 7.2a). В общем случае полученное

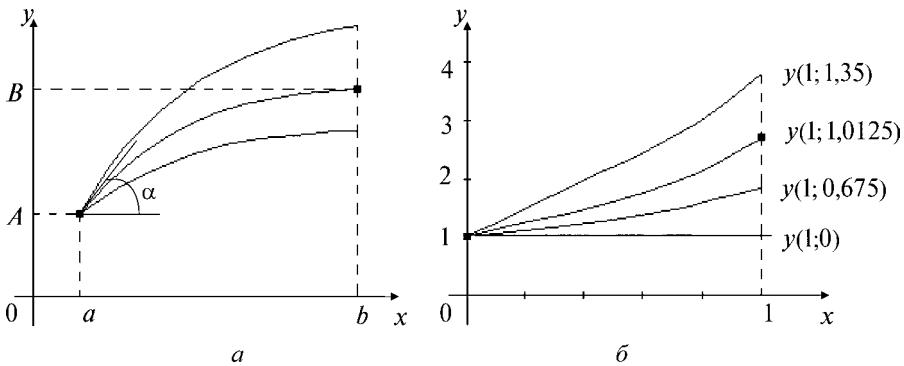


Рис. 7.2

при некотором значении η решение $y(x; \eta)$ не будет удовлетворять условию $y(b; \eta) = B$ на правом конце отрезка.

Следовательно, требуется найти такое значение параметра η , чтобы оно было корнем нелинейного уравнения

$$\Phi(\eta) = y(b; \eta) - B = 0.$$

Для решения этого уравнения, как правило, используются методы половинного деления (см. п. 3.1.3) или секущих (см. п. 3.1.7).

В случае использования метода половинного деления сначала делают «пробные» выстрелы при выбранных наугад или в соответствии с некоторым алгоритмом значениях η до тех пор, пока среди значений $\Phi(\eta)$ не окажется двух противоположных по знаку. Им соответствует начальный интервал неопределенности, который далее последовательно сокращается путем деления пополам.

При применении метода секущих используется формула

$$\eta^{(k+1)} = \eta^{(k)} - \frac{\Phi(\eta^{(k)}) \cdot [\eta^{(k)} - \eta^{(k-1)}]}{\Phi(\eta^{(k)}) - \Phi(\eta^{(k-1)})}, \quad k = 1, 2, \dots,$$

где $\eta^{(0)}, \eta^{(1)}$ — начальные значения параметра; k — номер итерации. Итерации прекращаются при выполнении условия окончания $|\Phi(\eta^{(k)})| \leq \varepsilon$ или $|\eta^{(k+1)} - \eta^{(k)}| \leq \varepsilon$ с некоторым положительным ε , характеризующим точность решения задачи.

Замечание. Точность решения краевой задачи зависит не только от точности определения параметра η , но также и от точности решения соответствующей задачи Коши. Поэтому одновременно с уточнением параметра η рекомендуется уменьшать шаг при решении задачи Коши, либо выбирать более точный метод.

Рассмотрим применение метода стрельбы для решения линейной краевой задачи (7.3), (7.4):

$$y'' + p(x)y' - q(x)y = f(x), \quad a \leq x \leq b;$$

$$\alpha_0 y(a) + \beta_0 y'(a) = A;$$

$$\alpha_1 y(b) + \beta_1 y'(b) = B.$$

Методика решения линейной краевой задачи

1. Решить две задачи Коши

$$y'' + p(x)y' - q(x)y = f(x), \quad a \leq x \leq b \quad (7.28)$$

при условии

$$\alpha_0 y(a) + \beta_0 y'(a) = A, \quad (7.29)$$

а именно, при $\beta_0 \neq 0$ задается $y(a) = \eta$, тогда $y'(a) = \frac{A - \alpha_0 \eta}{\beta_0}$, а при $\beta_0 = 0$ задается $y'(a) = \eta$ и тогда $y(a) = \frac{A}{\alpha_0}$ с разными произвольными значениями η , т. е. $\eta = \eta^{(0)}$ и $\eta = \eta^{(1)}$.

2. Найти значения $\Phi(\eta^{(0)})$, $\Phi(\eta^{(1)})$ функции $\Phi(\eta) = \alpha_1 \cdot y(b; \eta) + \beta_1 \cdot y'(b; \eta) - B$ и новое значение параметра:

$$\eta^{(2)} = \eta^{(1)} - \frac{\Phi(\eta^{(1)}) \cdot [\eta^{(1)} - \eta^{(0)}]}{\Phi(\eta^{(1)}) - \Phi(\eta^{(0)})}.$$

3. Решить краевую задачу (7.28), (7.29) при $\eta = \eta^{(2)}$. В результате получится приближенное решение исходной линейной краевой задачи, точность которого определяется точностью решения задачи Коши при значении $\eta = \eta^{(2)}$.

Таким образом, в силу линейности поставленной краевой задачи соответствующая задача Коши решается только три раза.

Пример 7.5. Методом стрельбы найти приближенное решение нелинейной краевой задачи

$$yy'' - y'^2 = 0, \quad y(0) = 1, \quad y(1) = 2,718.$$

Соответствующую задачу Коши решить явным методом Эйлера с шагом $h = 0,01$, а параметр η вычислить методом половинного деления.

□ Зададим начальные значения параметра η : $\eta^{(0)} = 0$, $\eta^{(1)} = 1,35$, которые обеспечивают разные по знаку значения функции $\Phi(\eta)$. Они являются соответственно левым и правым концами начального интервала неопределенности. Согласно методу половинного деления будем делить текущий интервал пополам и в качестве нового интервала выбирать тот, на концах которого функция $\Phi(\eta) = y(1, \eta) - 2,718$ имеет разные знаки. Процедуру поиска завершим при выполнении условия $|\eta^{(k+1)} - \eta^{(k)}| \leq \varepsilon$, где положим $\varepsilon = 0,01$.

Для применения явного метода Эйлера (6.19) уравнение следует переписать в форме $y'' = \frac{y'^2}{y}$, а затем с помощью введения новой переменной $z = y'$ — в виде (7.27):

$$\begin{aligned} y' &= z, \quad y(0) = 1; \\ z' &= \frac{z^2}{y}, \quad z(0) = \eta. \end{aligned}$$

Результаты интегрирования явным методом Эйлера с шагом $h = 0,01$ с параметром η , определяемым методом половинного деления, приведены в таблице 7.3. Четыре первых решения $y(x, \eta)$ изображены на рисунке 7.2б.

Последний интервал неопределенности удовлетворяет условию окончания процесса: $|1,0125 - 1,002734375| = 0,009765625 < \varepsilon = 0,01$. Полученное приближенное решение приведено в таблице 7.4.■

Пример 7.6. Найти приближенное решение краевой задачи

$$\begin{aligned} y'' + y &= -x, \quad 0 \leq x \leq 1; \\ y(0) &= 0, \quad y(1) = 0 \end{aligned}$$

методом стрельбы.

□ В поставленной задаче $a = 0, b = 1, p(x) = 0, q(x) = -1, f(x) = -x, \alpha_0 = 1, \beta_0 = 0, A = 0, \alpha_1 = 1, \beta_1 = 0, B = 0$. Точное решение найдено в примере 7.1.

Таблица 7.3

k	$\eta^{(k)}$	$y(1, \eta^{(k)})$	$\Phi(\eta^{(k)})$
0	0	1,000000	-1,718000
1	1,35	3,822746	1,104700
2	0,675	1,959583	-0,758417
3	1,0125	2,738495	0,020495
4	0,84375	2,316000	-0,402000
5	0,934375	2,534000	-0,184000
6	0,9734375	2,634597	-0,083403
7	0,99296875	2,686050	-0,031950
8	1,002734375	2,712146	-0,005850

Таблица 7.4

x_i	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
\hat{y}_i	1	1,105	1,2208	1,3489	1,4905	1,6468	1,8196	2,0106	2,2215	2,4546	2,7121

Таблица 7.5

η	$\eta^{(0)} = 1$		$\eta^{(1)} = -1$		$\eta^{(2)} = 0,18249$	
	$\hat{y}(x_i)$	$\hat{z}(x_i)$	$\hat{y}(x_i)$	$\hat{z}(x_i)$	$\hat{y}(x_i)$	$\hat{z}(x_i)$
0	0	1	0	-1	0	0,18249
0,1	$9,976 \cdot 10^{-2}$	0,9910	$9,9999 \cdot 10^{-2}$	-1	$1,81074 \cdot 10^{-2}$	0,17774
0,2	0,1977231	0,9620968	-0,2	-1	$3,51523 \cdot 10^{-2}$	0,16008
0,3	0,2919084	0,9135469	-0,3	-1	$4,99637 \cdot 10^{-2}$	0,13138
0,4	0,3803712	0,8458201	-0,4	-1	$6,13916 \cdot 10^{-2}$	$9,13343 \cdot 10^{-2}$
0,5	0,4612217	0,7595744	-0,5	-1	$6,83188 \cdot 10^{-2}$	$4,03417 \cdot 10^{-2}$
0,6	0,5326446	0,6556531	-0,6	-1	$0,06967 \cdot 10^{-2}$	$-2,11013 \cdot 10^{-2}$
0,7	0,5929167	0,5350776	-0,7	-1	$6,44322 \cdot 10^{-2}$	$-9,23912 \cdot 10^{-2}$
0,8	0,6404248	0,399037	-0,8	-1	$5,16459 \cdot 10^{-2}$	-0,17285
0,9	0,6736817	0,2488741	-0,9	-1	$3,04335 \cdot 10^{-2}$	-0,261608
1,0	0,691341	$8,6077 \cdot 10^{-2}$	-1	-1	$-8,22125 \cdot 10^{-7}$	-0,35786

Так как задача линейная, то воспользуемся приведенной выше методикой.

1. Решим задачу Коши с начальными условиями, соответствующими случаю $\beta_0 = 0$:

$$y'' + y = -x, \quad 0 \leq x \leq 1;$$

$$y(0) = \frac{A}{\alpha_0} = 0, \quad y'(0) = \eta$$

при двух значениях параметра η : $\eta^{(0)} = 1$ и $\eta^{(1)} = -1$. Для этого применим явный метод Эйлера с шагом $h = 0,01$. Исходное уравнение с учетом обозначений $z = y'$ предварительно запишем в форме системы

$$y' = z, \quad y(0) = 0;$$

$$z' = -y - x, \quad z(0) = \eta.$$

Численные результаты приведены в таблице 7.5.

2. Запишем функцию

$$\Phi(\eta) = \alpha_1 \cdot y(b; \eta) + \beta_1 \cdot y'(b; \eta) - B = 1 \cdot y(1; \eta) + 0 - 0 = y(1; \eta).$$

Найдем значения $\Phi(\eta^{(0)}) = y(1; 1) = 0,691341$ и $\Phi(\eta^{(1)}) = y(1; -1) = -1$, а затем

$$\eta^{(2)} = \eta^{(1)} - \frac{\Phi(\eta^{(1)}) \cdot [\eta^{(1)} - \eta^{(0)}]}{\Phi(\eta^{(1)}) - \Phi(\eta^{(0)})} = -1 - \frac{-1 \cdot (-2)}{-1,69134} = 0,18249.$$

3. Приближенное решение исходной задачи определяется решением задачи Коши с $\eta^{(2)} = 0,18249$ (две правые колонки в табл. 7.5). Сравнивая его с решением примера 7.1, можно сделать вывод о близости данного решения к точному. ■

7.4.2. МЕТОД ДИФФЕРЕНЦИАЛЬНОЙ ПРОГОНКИ

Отличие метода дифференциальной прогонки [42] от вышеизложенного метода стрельбы заключается в том, что решаются вспомогательные задачи Коши не для исходного дифференциального уравнения, а для других уравнений меньшего порядка. Рассмотрим его применение для решения линейной краевой задачи (7.3), (7.4).

*Методика решения линейной
краевой задачи*

Первый случай: $\beta_0 \neq 0$.

Прямая прогонка. Решить две задачи Коши:

$$Z'_1 = -Z_1^2 - p(x)Z_1 + q(x), \quad Z_1(a) = -\frac{\alpha_0}{\beta_0};$$

$$Z'_2 = -Z_2 \cdot [Z_1 + p(x)] + f(x), \quad Z_2(a) = \frac{A}{\beta_0}$$

относительно $Z_1(x)$ и $Z_2(x)$ на отрезке $[a, b]$.

Обратная прогонка. Решить задачу Коши

$$y' = Z_1(x)y + Z_2(x), \quad a \leq x \leq b; \quad y(b) = \frac{B - \beta_1 Z_2(b)}{\alpha_1 + \beta_1 Z_1(b)}.$$

Второй случай: $\alpha_0 \neq 0$.

Прямая прогонка. Решить две задачи Коши:

$$u'_1 = -u_1^2 q(x) + u_1 p(x) + 1, \quad a \leq x \leq b, \quad u_1(a) = -\frac{\beta_0}{\alpha_0};$$

$$u'_2 = -u_1(x) \cdot [u_2 q(x) + f(x)], \quad a \leq x \leq b, \quad u_2(a) = \frac{A}{\alpha_0}.$$

Обратная прогонка. Решить задачу Коши

$$u_1(x)y' = y - u_2(x), \quad a \leq x \leq b, \quad y(b) = \frac{Bu_1(b) + \beta_1 u_2(b)}{\beta_1 + \alpha_1 u_1(b)}.$$

Обратная прогонка в обоих случаях дает приближенное решение исходной краевой задачи.

Пример 7.7. Найти приближенное решение краевой задачи

$$\begin{aligned} y'' + (x+1)y' - 2y &= 2, \quad 0 \leq x \leq 1; \\ y(0) - y'(0) &= -1, \quad y(1) = 4 \end{aligned}$$

методом дифференциальной прогонки. При прямой и обратной прогонке использовать явный метод Эйлера с шагом $h = 0,02$.

В поставленной краевой задаче $a = 0, b = 1, p(x) = x + 1, q(x) = 2, f(x) = 2, \alpha_0 = 1, \beta_0 = -1, A = -1, \alpha_1 = 1, \beta_1 = 0, B = 4$. Так как $\beta_0 \neq 0$, воспользуемся соотношениями для первого случая.

Таблица 7.6

Значения аргумента	Прямая прогонка		Обратная прогонка
	$\hat{Z}_1(x_i)$	$\hat{Z}_2(x_i)$	
0,0	1,000000	1,000000	0,9856163
0,1	0,996238	0,996238	1,196328
0,2	0,984706	0,984706	1,427262
0,3	0,967711	0,967711	1,678383
0,4	0,947030	0,947030	1,949667
0,5	0,924000	0,924000	2,241091
0,6	0,899611	0,899611	2,552642
0,7	0,574584	0,574584	2,884312
0,8	0,849434	0,849434	3,236095
0,9	0,824526	0,824526	3,607991
1,0	0,8001108	0,8001108	4,000000

На этапе прямой прогонки решим две задачи Коши на отрезке $[0; 1]$:

$$Z'_1 = -Z_1^2 - Z_1(x+1) + 2, \quad Z_1(0) = -\frac{\alpha_0}{\beta_0} = -\frac{1}{-1} = 1;$$

$$Z'_2 = -Z_2 \cdot [Z_1 + x + 1] + 2, \quad Z_2(0) = \frac{A}{\beta_0} = \frac{-1}{-1} = 1.$$

На этапе обратной прогонки решим задачу Коши

$$y' = Z_1(x)y + Z_2(x), \quad 0 \leq x \leq 1, \quad y(1) = \frac{B - \beta_1 Z_2(b)}{\alpha_1 + \beta_1 Z_1(b)} = \frac{4 - 0}{1 + 0} = 4.$$

Результаты вычислений представлены в таблице 7.6 (в целях экономии места численные значения приведены в точках, отстоящих друг от друга на 0,1).

Если при прямой прогонке использовать явный метод Эйлера с шагом $h = 0,01$, а при обратной — метод Эйлера — Коши, получается более точное решение [42]. ■

7.5. МЕТОД КОНЕЧНЫХ ЭЛЕМЕНТОВ

Метод конечных элементов нашел широкое практическое применение при решении уравнений с частными производными. Здесь описывается его применение к решению линейной краевой задачи (7.3), (7.4) для обыкновенного дифференциального уравнения, в которой искомая функция зависит от одной переменной.

Суть метода конечных элементов заключается в следующем:

- а) область определения функции $y(x)$, удовлетворяющей дифференциальному уравнению и краевым условиям, разбивается на конечное число подобластей, называемых *конечными элементами*;
- б) в каждом элементе фиксируется конечное число *узловых точек*, общих с узловыми точками соседних элементов, в которых должны быть найдены узловые значения искомой функции;
- в) искомая функция на каждом элементе аппроксимируется линейной комбинацией базисных функций с коэффициентами, равными узловым значениям искомой функции;
- г) для нахождения коэффициентов применяется метод взвешенных невязок (метод Галеркина).

Разобьем отрезок $[a, b]$ на n конечных элементов (в рассматриваемой задаче ими являются отрезки) e_1, \dots, e_n с $(n + 1)$ нумерованными узлами x_1, x_2, \dots, x_{n+1} (рис. 7.3).

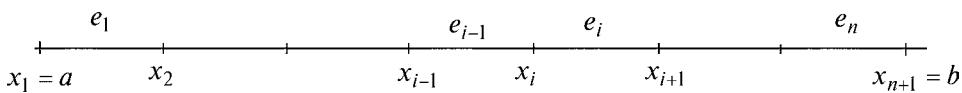


Рис. 7.3

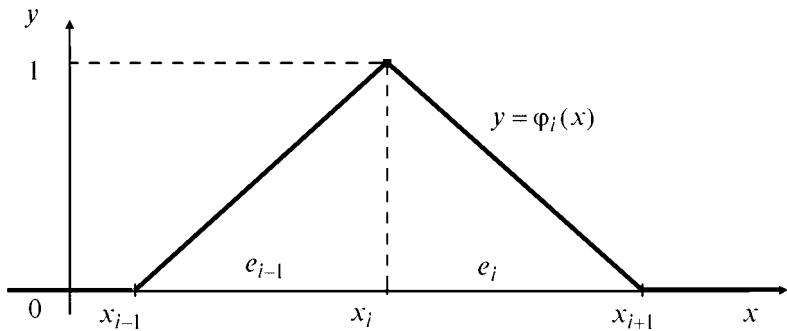


Рис. 7.4

Каждому нумерованному узлу x_i , $i = \overline{1, n+1}$, поставим в соответствие кусочно-линейную глобальную базисную функцию $\varphi_i(x)$, принимающую в данном узле значение единицы, а в остальных — нулевое значение (рис. 7.4):

$$\varphi_i(x) = \begin{cases} \frac{x - x_{i-1}}{x_i - x_{i-1}}, & x_{i-1} \leq x \leq x_i, \\ \frac{x_{i+1} - x}{x_{i+1} - x_i}, & x_i \leq x \leq x_{i+1}, \\ 0, & x < x_{i-1} \text{ или } x > x_{i+1}, \end{cases} \quad i = \overline{1, n+1}. \quad (7.30)$$

Для обеспечения унифицированного определения глобальных базисных функций необходимо ввести в рассмотрение дополнительные узлы x_0 и x_{n+2} . Тогда глобальные базисные функции $\varphi_1(x)$, $\varphi_{n+1}(x)$ будут определены на отрезках $[x_0, x_2]$ и $[x_n, x_{n+2}]$ соответственно. Значения функций $\varphi_1(x)$, $\varphi_{n+1}(x)$ на отрезках $[x_0, x_1]$ и $[x_{n+1}, x_{n+2}]$, примыкающих к отрезку $[a, b]$, в дальнейших расчетах не используются.

Используя глобальные базисные функции, будем искать решение краевой задачи в виде

$$\hat{y}(x) = \sum_{i=1}^{n+1} \hat{y}_i \cdot \varphi_i(x), \quad (7.31)$$

где \hat{y}_i — значения искомой функции в узлах x_i , $i = \overline{1, n+1}$ (коэффициенты).

Требуется найти такие значения \hat{y}_i , $i = \overline{1, n+1}$, чтобы выполнялось условие (7.23) (см. метод Галеркина в п. 7.3).

Функция, аппроксимирующая искомое решение на каждом конечном элементе, называется *функцией элемента*.

Для элемента e_i с узлами x_i и x_{i+1} определим локальные базисные функции (рис. 7.5):

$$\begin{aligned} \varphi_i^{e_i}(x) &= \frac{x_{i+1} - x}{x_{i+1} - x_i}, & x_i \leq x \leq x_{i+1}; \\ \varphi_{i+1}^{e_i}(x) &= \frac{x - x_i}{x_{i+1} - x_i}, & x_i \leq x \leq x_{i+1}, \end{aligned} \quad (7.32)$$

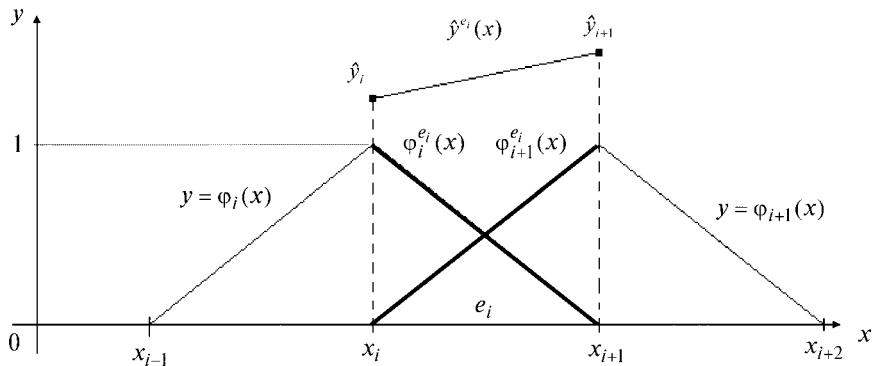


Рис. 7.5

где нижний индекс у базисной функции соответствует номеру узла, а верхний — номеру конечного элемента. На каждом элементе e_i , $i=1, n$, обе локальные базисные функции линейны, причем в узлах этого элемента одна из них равна единице, т. е. $\phi_i^{e_i}(x_i)=1$ и $\phi_{i+1}^{e_i}(x_{i+1})=1$. Функция элемента e_i может быть выражена с помощью локальных базисных функций и узловых значений \hat{y}_i и \hat{y}_{i+1} по формуле

$$\hat{y}^{e_i}(x) = \begin{cases} \hat{y}_i \cdot \phi_i^{e_i}(x) + \hat{y}_{i+1} \cdot \phi_{i+1}^{e_i}(x), & x_i \leq x \leq x_{i+1}; \\ 0, & x < x_i \text{ или } x > x_{i+1}. \end{cases} \quad i = \overline{1, n}, \quad (7.33)$$

Тогда приближенное решение (7.31) краевой задачи можно представить в форме

$$\hat{y}(x) = \sum_{i=1}^n \hat{y}^{e_i}(x). \quad (7.34)$$

В силу свойств локальных базисных функций на элементе e_i функция элемента является линейной и принимает в узле x_i значение \hat{y}_i , а в узле x_{i+1} значение \hat{y}_{i+1} (рис. 7.5). Если в краевой задаче заданы условия первого рода, т. е. $y(a) = A$, $y(b) = B$, то при рассмотрении элемента e_1 следует положить $\hat{y}_1 = A$, так как $x_1 = a$, а при рассмотрении элемента e_n положить $\hat{y}_{n+1} = B$,

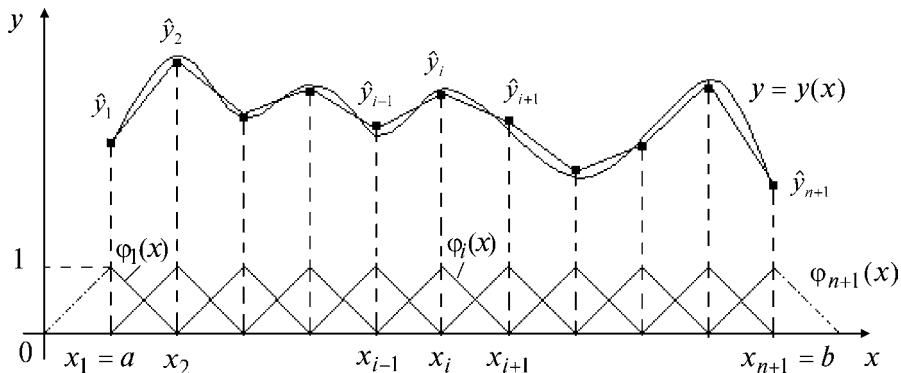


Рис. 7.6

поскольку $x_{n+1} = b$. Тогда краевые условия первого рода для решения, записанного в виде (7.31) или эквивалентном ему (7.34), выполняются автоматически. Для учета условий второго и третьего рода применяется обобщенный подход (см. пример 7.8). В целом аппроксимация (7.31) является кусочно-линейной. На рисунке 7.6 показано применение формулы (7.31) для аппроксимации некоторой произвольной функции $y = y(x)$.

Рассмотрим некоторые теоретические аспекты применения метода конечных элементов на типовом примере линейной краевой задачи с краевыми условиями первого рода.

Пример 7.8. Найти приближенное решение краевой задачи

$$\begin{aligned} y'' - y &= 0, \quad 0 \leq x \leq 1; \\ y(0) &= 0, \quad y(1) = 1 \end{aligned}$$

методом конечных элементов, полагая число конечных элементов, равным трем.

В поставленной задаче $a = 0$, $b = 1$, $p(x) \equiv 0$, $q(x) = 1$, $f(x) \equiv 0$, $\alpha_0 = 1$, $\beta_0 = 0$, $A = 0$, $\alpha_1 = 1$, $\beta_1 = 0$, $B = 1$. Точное решение этой задачи (см. пример 7.1):

$$y(x) = \frac{e}{e^2 - 1} e^x + \frac{e}{1 - e^2} e^{-x}.$$

Применим метод Галеркина. Для этого запишем условие (7.23):

$$\int_0^1 [\hat{y}''(x) - \hat{y}(x)] \cdot \varphi_j(x) dx = 0, \quad j = \overline{1, n+1}. \quad (7.35)$$

Заметим, что базисные функции непрерывны, но в силу кусочной линейности имеют разрывные производные, а при использовании метода Галеркина для дифференциального уравнения второго порядка требуется, чтобы подынтегральные выражения в (7.35) были по крайней мере кусочно-непрерывны. Поэтому рассмотрим так называемую *слабую формулировку метода Галеркина*, в которой требования, предъявляемые к базисным функциям, ослабляются.

Проинтегрируем по частям первое слагаемое в (7.35):

$$\int_0^1 \hat{y}''(x) \cdot \varphi_j(x) dx = \varphi_j(x) \frac{d\hat{y}(x)}{dx} \Big|_0^1 - \int_0^1 \frac{d\hat{y}(x)}{dx} \cdot \frac{d\varphi_j(x)}{dx} dx.$$

Тогда (7.35) запишется в форме

$$-\int_0^1 \left[\frac{d\hat{y}(x)}{dx} \cdot \frac{d\varphi_j(x)}{dx} + \hat{y}(x) \cdot \varphi_j(x) \right] dx + \varphi_j(x) \frac{d\hat{y}(x)}{dx} \Big|_0^1 = 0. \quad (7.36)$$

Из (7.36) видно, что теперь достаточно, чтобы базисные функции $\varphi_i(x)$ принадлежали классу $C([0; 1])$ непрерывных функций.

Задача заключается в следующем: требуется найти такой вектор $\hat{y} = (\hat{y}_1, \dots, \hat{y}_{n+1})^T$, чтобы функция (7.31) удовлетворяла условию (7.36). В этом заключается *слабая формулировка метода Галеркина*.

Для решения задачи подставим (7.31) в (7.36):

$$\int_0^1 \left[\frac{d}{dx} \sum_{i=1}^{n+1} \hat{y}_i \cdot \varphi_i(x) \cdot \frac{d\varphi_j(x)}{dx} + \varphi_j(x) \cdot \sum_{i=1}^{n+1} \hat{y}_i \cdot \varphi_i(x) \right] dx = \varphi_j(x) \cdot \frac{d}{dx} \sum_{i=1}^{n+1} \hat{y}_i \cdot \varphi_i(x) \Big|_0^1. \quad (7.37)$$

Полученное равенство запишем в матричной форме:

$$K\hat{y} = f, \quad (7.38)$$

где K — матрица размеров $(n+1) \times (n+1)$, элемент которой определяется по формуле

$$K_{j,i} = \int_0^1 \left[\frac{d\varphi_j(x)}{dx} \cdot \frac{d\varphi_i(x)}{dx} + \varphi_j(x) \cdot \varphi_i(x) \right] dx, \quad j = \overline{1, n+1}, \quad i = \overline{1, n+1}, \quad (7.39)$$

а $f = (f_1, \dots, f_{n+1})^T$ — вектор, компоненты которого находятся следующим образом:

$$f_j = \varphi_j(x) \cdot \frac{d}{dx} \sum_{i=1}^{n+1} \hat{y}_i \cdot \varphi_i(x) \Big|_0^1, \quad j = \overline{1, n+1}. \quad (7.40)$$

Проанализируем вклад элемента e_i с узлами x_i, x_{i+1} (см. рис. 7.5). Единственными отличными от нуля глобальными базисными функциями для элемента e_i будут $\varphi_i(x)$ и $\varphi_{i+1}(x)$. Поэтому будем использовать соответствующие локальные базисные функции $\varphi_i^{e_i}(x), \varphi_{i+1}^{e_i}(x)$. С учетом (7.32) и (7.39) определим элементы матрицы конечного элемента e_i :

$$\begin{aligned} K_{i,i}^{e_i} &= \int_{x_i}^{x_{i+1}} \left[\left(\frac{d\varphi_i^{e_i}(x)}{dx} \right)^2 + \left(\varphi_i^{e_i}(x) \right)^2 \right] dx; \\ K_{i+1,i}^{e_i} &= \int_{x_i}^{x_{i+1}} \left[\frac{d\varphi_{i+1}^{e_i}(x)}{dx} \cdot \frac{d\varphi_i^{e_i}(x)}{dx} + \varphi_{i+1}^{e_i}(x) \cdot \varphi_i^{e_i}(x) \right] dx; \\ K_{i+1,i+1}^{e_i} &= \int_{x_i}^{x_{i+1}} \left[\left(\frac{d\varphi_{i+1}^{e_i}(x)}{dx} \right)^2 + \left(\varphi_{i+1}^{e_i}(x) \right)^2 \right] dx = K_{i,i}^{e_i}; \\ K_{i,i+1}^{e_i} &= \int_{x_i}^{x_{i+1}} \left[\frac{d\varphi_i^{e_i}(x)}{dx} \cdot \frac{d\varphi_{i+1}^{e_i}(x)}{dx} + \varphi_i^{e_i}(x) \cdot \varphi_{i+1}^{e_i}(x) \right] dx = K_{i+1,i}^{e_i}. \end{aligned} \quad (7.41)$$

В результате можно записать матрицу элемента e_i в форме

$$K^i = \begin{pmatrix} 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \dots & K_{i,i}^{e_i} & K_{i,i+1}^{e_i} & \dots & 0 \\ 0 & \dots & K_{i+1,i}^{e_i} & K_{i+1,i+1}^{e_i} & \dots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 \end{pmatrix}. \quad (7.42)$$

Матрица K системы (7.38) получается суммированием матриц $K^i, i = \overline{1, n}$. Этот процесс называется *ансамблированием элементов*.

*Алгоритм применения метода
конечных элементов для решения
линейных краевых задач*

1. Разбить отрезок $[a, b]$ на n конечных элементов e_1, \dots, e_n и представить решение в виде (7.31).

2. Для каждого конечного элемента составить матрицу K^i , $i = \overline{1, n}$, по формуле (7.42), элементы которой вычисляются согласно (7.41).

3. Сформировать матрицу K в результате суммирования матриц K^i , вектор f с учетом (7.40) и составить систему (7.38).

4. Учесть краевые условия и решить систему (7.38). В результате найти значения $\hat{y}_1, \dots, \hat{y}_{n+1}$ и выписать приближенное решение по формуле (7.31).

Теперь применим описанную методику к решению поставленной задачи.

1. Положим $n = 3$ при равномерной сетке. Тогда

$$h = x_{i+1} - x_i = \frac{1}{3} = \text{const},$$

а глобальные базисные функции, определяемые по формулам (7.30), изображены на рисунке 7.7.

Решение задачи ищется в форме

$$\hat{y}(x) = \hat{y}_1 \cdot \varphi_1(x) + \hat{y}_2 \cdot \varphi_2(x) + \hat{y}_3 \cdot \varphi_3(x) + \hat{y}_4 \cdot \varphi_4(x), \quad (7.43)$$

где $\hat{y}_1, \hat{y}_2, \hat{y}_3, \hat{y}_4$ подлежат определению.

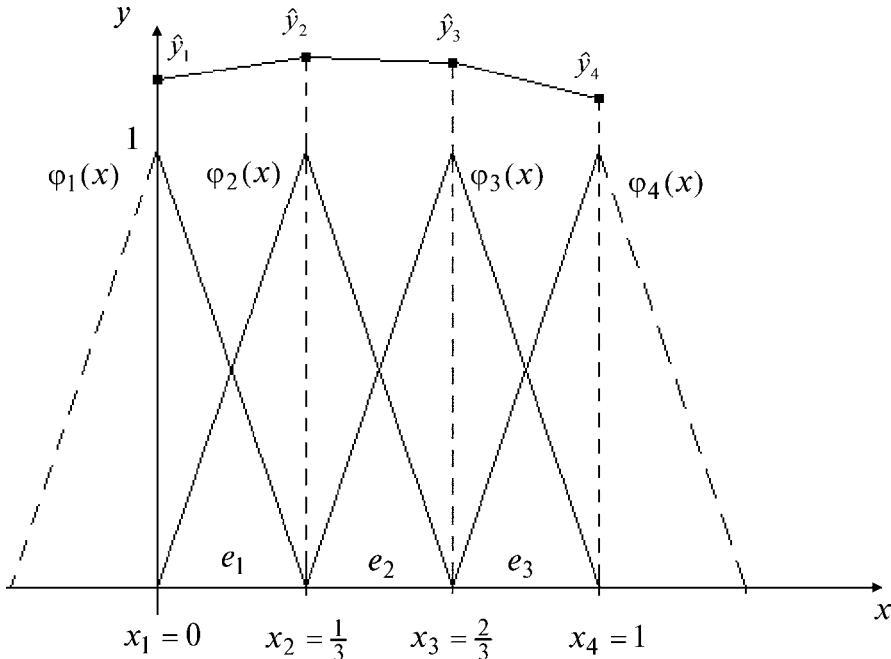


Рис. 7.7

2. Составим матрицу K^i для каждого элемента e_1, e_2, e_3 . С этой целью вычислим отдельные элементы матрицы по формулам (7.41). Для конечного элемента e_1 получим

$$\begin{aligned}
 K_{1,1}^{e_1} &= \int_{x_1}^{x_2} \left[\left(\frac{d\varphi_1^{e_1}(x)}{dx} \right)^2 + (\varphi_1^{e_1}(x))^2 \right] dx = \int_0^{\frac{1}{3}} \left\{ \left[\frac{d}{dx} \left(\frac{x_2 - x}{x_2 - x_1} \right) \right]^2 + \left(\frac{x_2 - x}{x_2 - x_1} \right)^2 \right\} dx = \\
 &= \int_0^{\frac{1}{3}} \left[9 + 9 \cdot \left(\frac{1}{3} - x \right)^2 \right] dx = \left[9x - 9 \frac{\left(\frac{1}{3} - x \right)^3}{3} \right]_0^{\frac{1}{3}} = \frac{28}{9}; \\
 K_{1,2}^{e_1} &= \int_{x_1}^{x_2} \left[\frac{d\varphi_1^{e_1}(x)}{dx} \cdot \frac{d\varphi_2^{e_1}(x)}{dx} + \varphi_1^{e_1}(x) \cdot \varphi_2^{e_1}(x) \right] dx = \\
 &= \int_0^{\frac{1}{3}} \left[\frac{d}{dx} \left(\frac{x_2 - x}{x_2 - x_1} \right) \cdot \frac{d}{dx} \left(\frac{x - x_1}{x_2 - x_1} \right) + \frac{x_2 - x}{x_2 - x_1} \cdot \frac{x - x_1}{x_2 - x_1} \right] dx = \\
 &= \int_0^{\frac{1}{3}} \left[-3 \cdot 3 + 9 \cdot \left(\frac{1}{3} - x \right) (x - 0) \right] dx = -\frac{53}{18}; \\
 K_{2,1}^{e_1} &= K_{1,2}^{e_1}; \quad K_{2,2}^{e_1} = K_{1,1}^{e_1}.
 \end{aligned}$$

Здесь нижние индексы соответствуют связям узлов для элемента, указанного в верхнем индексе.

В результате для конечного элемента e_1 получаем

$$K^1 = \begin{pmatrix} \frac{28}{9} & -\frac{53}{18} & 0 & 0 \\ -\frac{53}{18} & \frac{28}{9} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Так как сетка равномерная, дифференциальное уравнение с постоянными коэффициентами и базисные функции «одинаковые», то для конечных элементов e_2, e_3 в решаемой задаче аналогично имеем

$$K^2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{28}{9} & -\frac{53}{18} & 0 \\ 0 & -\frac{53}{18} & \frac{28}{9} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}; \quad K^3 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{28}{9} & -\frac{53}{18} \\ 0 & 0 & -\frac{53}{18} & \frac{28}{9} \end{pmatrix}.$$

3. Складывая матрицы элементов e_1, e_2, e_3 , находим результат ансамблирования:

$$K = \begin{pmatrix} \frac{28}{9} & -\frac{53}{18} & 0 & 0 \\ -\frac{53}{18} & \frac{56}{9} & -\frac{53}{18} & 0 \\ 0 & -\frac{53}{18} & \frac{56}{9} & -\frac{53}{18} \\ 0 & 0 & -\frac{53}{18} & \frac{28}{9} \end{pmatrix}.$$

По формуле (7.40) с учетом (7.43) определим компоненты вектора f правой части системы (7.38):

$$f_1 = \varphi_1(x) \cdot \frac{d\hat{y}(x)}{dx} \Big|_0 = -\frac{d\hat{y}(x)}{dx} \Big|_{x=0}, \text{ поскольку } \varphi_1(1)=0, \varphi_1(0)=0;$$

$$f_2 = \varphi_2(x) \cdot \frac{d\hat{y}(x)}{dx} \Big|_0 = 0, \text{ так как } \varphi_2(1)=0, \varphi_2(0)=0;$$

$$f_3 = \varphi_3(x) \cdot \frac{d\hat{y}(x)}{dx} \Big|_0 = 0, \text{ так как } \varphi_3(1)=0, \varphi_3(0)=0;$$

$$f_4 = \varphi_4(x) \cdot \frac{d\hat{y}(x)}{dx} \Big|_0 = \frac{d\hat{y}(x)}{dx} \Big|_{x=1}, \text{ поскольку } \varphi_4(1)=0, \varphi_4(0)=0.$$

С учетом полученных результатов запишем систему (7.38):

$$\begin{pmatrix} \frac{28}{9} & -\frac{53}{18} & 0 & 0 \\ -\frac{53}{18} & \frac{56}{9} & -\frac{53}{18} & 0 \\ 0 & -\frac{53}{18} & \frac{56}{9} & -\frac{53}{18} \\ 0 & 0 & -\frac{53}{18} & \frac{28}{9} \end{pmatrix} \begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \\ \hat{y}_4 \end{pmatrix} = \begin{pmatrix} -\frac{d\hat{y}(x)}{dx} \Big|_{x=0} \\ 0 \\ 0 \\ \frac{d\hat{y}(x)}{dx} \Big|_{x=1} \end{pmatrix}. \quad (7.44)$$

Из краевых условий $y(0)=0, y(1)=1$ следует, что $\hat{y}_1=y(0)=0, \hat{y}_4=y(1)=1$. Тогда в системе (7.44) можно исключить первое и четвертое уравнения, положив в остальных $\hat{y}_1=0, \hat{y}_4=1$. В итоге получим систему

$$\begin{pmatrix} \frac{56}{9} & -\frac{53}{18} \\ -\frac{53}{18} & \frac{56}{9} \end{pmatrix} \begin{pmatrix} \hat{y}_2 \\ \hat{y}_3 \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{53}{18} \end{pmatrix}.$$

Отсюда $\hat{y}_2=0,2885; \hat{y}_3=0,6088$. Точные значения $y(x_2)=y\left(\frac{1}{3}\right)=0,2889; y(x_3)=y\left(\frac{2}{3}\right)=0,6102$. Приближенное решение задачи имеет вид

$$\hat{y}(x)=0,2885 \cdot \varphi_2(x)+0,6088 \cdot \varphi_3(x)+\varphi_4(x). \blacksquare$$

Замечание. В случае краевых условий второго или третьего рода следует воспользоваться более общей схемой метода Галеркина с учетом невязки краевых условий, содержащих производные [47]. Например, рассмотрим задачу

$$y'' - y = 0, \quad y(0) = 0, \quad y'(1) = 1.$$

Вместо (7.35) используем следующее условие:

$$\int_0^1 [\hat{y}''(x) - \hat{y}(x)] \cdot \varphi_j(x) dx + [\hat{y}'(1) - 1] \cdot \tilde{\varphi}_j(1) = 0, \quad j = \overline{1, n+1}, \quad (7.45)$$

где $\tilde{\varphi}_j(x)$, $j = \overline{1, n+1}$, — весовые функции, задаваемые в общем случае независимо от функций $\varphi_j(x)$, $j = \overline{1, n+1}$. Как и ранее, краевое условие первого рода на левом конце может быть выполнено автоматически.

Далее интегрируем по частям:

$$-\int_0^1 \left[\frac{d\hat{y}(x)}{dx} \cdot \frac{d\varphi_j(x)}{dx} + \hat{y}(x) \cdot \varphi_j(x) \right] dx + \varphi_j(x) \frac{d\hat{y}(x)}{dx} \Big|_0^1 + \left[\tilde{\varphi}_j(x) \cdot \left(\frac{d\hat{y}(x)}{dx} - 1 \right) \right]_{x=1} = 0.$$

Базисные функции $\tilde{\varphi}_j(x)$, $j = \overline{1, n+1}$, следует подбирать так, чтобы члены, содержащие краевые условия с производными, и члены с производными первого порядка, возникающими в той же точке в процессе интегрирования по частям, уничтожились.

Положим $\tilde{\varphi}_j(x) = -\varphi_j(x)$, $j = \overline{1, n+1}$. Тогда имеем

$$\int_0^1 \left[\frac{d\hat{y}(x)}{dx} \cdot \frac{d\varphi_j(x)}{dx} + \hat{y}(x) \cdot \varphi_j(x) \right] dx = -\varphi_j(x) \frac{d\hat{y}(x)}{dx} \Big|_{x=0} + \varphi_j(1), \quad j = \overline{1, n+1}.$$

Полагая $n = 3$ и рассуждая аналогично случаю краевых условий первого рода с учетом равенств $\varphi_1(0) = 1$, $\varphi_2(0) = 0$, $\varphi_3(0) = 0$, $\varphi_4(0) = 0$, $\varphi_1(1) = 0$, $\varphi_2(1) = 0$, $\varphi_3(1) = 0$, $\varphi_4(1) = 1$, получаем систему

$$\begin{pmatrix} \frac{28}{9} & -\frac{53}{18} & 0 & 0 \\ -\frac{53}{18} & \frac{56}{9} & -\frac{53}{18} & 0 \\ 0 & -\frac{53}{18} & \frac{56}{9} & -\frac{53}{18} \\ 0 & 0 & -\frac{53}{18} & \frac{28}{9} \end{pmatrix} \cdot \begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \\ \hat{y}_4 \end{pmatrix} = \begin{pmatrix} -\frac{d\hat{y}(x)}{dx} \Big|_{x=0} \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

Краевое условие $y(0) = 0$ на левом конце учитывается вычеркиванием первого уравнения и приравниванием $\hat{y}_1 = 0$ в остальных уравнениях.

Задачи для самостоятельного решения

1. Решить краевую задачу

$$\begin{aligned}y'' + y &= -x; \quad x \in [0; 1]; \\y(0) &= 0; \quad y'(1) = 0\end{aligned}$$

методом Галеркина. Аппроксимирующую функцию принять в виде

$$\hat{y}_2(x) = a_1 x + a_2 x^2 \quad (\varphi_1(x) = x, \varphi_2(x) = x^2).$$

Указание. Данная функция тождественно удовлетворяет условию на левой границе. На правой границе необходимо использовать взвешенную неизвестную с весовыми функциями $\tilde{\varphi}_1(x) = -x$, $\tilde{\varphi}_2(x) = -x^2$.

2. Решить краевую задачу

$$\begin{aligned}y'' - y &= 0; \quad x \in [0; 1]; \\y(0) &= 0; \quad y(1) = 1\end{aligned}$$

методом коллокаций и методом Галеркина. Аппроксимирующую функцию принять в виде $\hat{y}_m(x) = x + \sum_{j=1}^m a_j \sin j\pi x$, удовлетворяющем краевым условиям задачи.

3. Решить краевую задачу

$$\begin{aligned}y'' + y + x &= 0; \quad x \in [0; 1]; \\y(0) &= 0, \quad y'(1) = 0\end{aligned}$$

методом Галеркина. Аппроксимирующую функцию принять в виде

$$\hat{y}_1(x) = a_1 x \cdot \left(1 - \frac{x}{2}\right), \text{ где } \varphi_1(x) = x \left(1 - \frac{x}{2}\right),$$

удовлетворяющем краевым условиям.

4. Решить краевую задачу

$$\begin{aligned}y'' + y + kx &= 0; \quad x \in [0; 1]; \\y(0) &= 0, \quad y(1) = 1 - k, \quad k = 1 - N_1 \cdot 0,001 \cdot N_2,\end{aligned}$$

где N_1 — номер группы; N_2 — номер студента по списку.

Указание. Принимая аппроксимирующую функцию в виде

$$\hat{y}_3(x) = x(1 - kx) \cdot (a_1 + a_2 x + a_3 x^2),$$

найти приближенное решение задачи методом Галеркина. Полученное решение сравнить с точным решением $y(x) = \frac{\sin x}{\sin 1} - kx$.

5. Методом конечных элементов, разбивая $[a, b] = [0; 1 - N]$ на 3, 6 и 12 элементов, решить краевую задачу

$$y'' + y = 0; \\ y(0) = 0; \quad \left. \frac{dy}{dx} \right|_{x=1-N} = \operatorname{ctg}(1-N),$$

где $N = N' \cdot N^0 \cdot 0,001$, N' — номер группы; N^0 — номер студента в группе.
Приближенные решения сравнить с точным решением $y(x) = \frac{\sin x}{\sin(1-N)}$.

6. Методом конечных элементов решить краевую задачу

$$y'' - y = 0; \quad x \in [0; 1]; \\ y(0) = 0, \quad y(1) = 1,$$

используя четыре конечных элемента. Полученное решение сравнить с точным решением $y(x) = \frac{\operatorname{sh}x}{\operatorname{sh}1}$.

7. Методом конечных элементов решить краевую задачу

$$y'' - y = 0; \quad x \in [0; 1]; \\ y'(0) = \frac{1}{\operatorname{sh}1}, \quad y'(1) = 1,$$

разбивая отрезок $\Omega = [0; 1]$ на 3 и 4 элемента. Результаты сравнить с точным решением задачи.

ЧАСТЬ ЧЕТВЕРТАЯ

ЧИСЛЕННЫЕ МЕТОДЫ РЕШЕНИЯ
ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ
В ЧАСТНЫХ ПРОИЗВОДНЫХ

ГЛАВА

8

ЧИСЛЕННЫЕ МЕТОДЫ РЕШЕНИЯ УРАВНЕНИЙ МАТЕМАТИЧЕСКОЙ ФИЗИКИ С ДВУМЯ НЕЗАВИСИМЫМИ ПЕРЕМЕННЫМИ

8.1. ПОСТАНОВКА ЗАДАЧИ И ОСНОВНЫЕ ПОЛОЖЕНИЯ

Математические модели современных систем проектирования летательных аппаратов, а также физические явления в таких областях, как динамика жидкости, электричество и магнетизм, механика, оптика, теплопередача, описываются с помощью дифференциальных уравнений в частных производных (ДУЧП). В качестве примеров можно привести уравнения Максвелла, закон теплообмена Ньютона, уравнения Эйлера, уравнения Навье — Стокса в механике жидкости и газа, уравнение Шредингера в квантовой механике и др. Производные появляются в уравнениях потому, что они описывают важнейшие физические величины (такие как скорость, ускорение, сила, поток, трение, электрический ток и т. д.). В отличие от обыкновенных дифференциальных уравнений, в которых неизвестная функция зависит только от одной независимой переменной, в уравнениях с частными производными неизвестная функция зависит от нескольких переменных.

В общем случае *нелинейное уравнение в частных производных* может быть записано в виде

$$F\left(x_1, x_2, \dots, x_k, u, \frac{\partial u}{\partial x_1}, \dots, \frac{\partial u}{\partial x_k}, \frac{\partial^2 u}{\partial x_1^2}, \frac{\partial^2 u}{\partial x_1 \partial x_2}, \dots, \frac{\partial^n u}{\partial x_1^{p_1} \partial x_2^{p_2} \dots \partial x_k^{p_k}}\right) = 0, \quad (8.1)$$

где x_1, x_2, \dots, x_k — независимые переменные; $u(x_1, x_2, \dots, x_k)$ — неизвестная (искомая) функция; $p_1 + p_2 + \dots + p_k = n$, p_i — целые неотрицательные числа. Во многих практических задачах в качестве независимых переменных используется время t и пространственные координаты в декартовой, цилиндрической или сферической системах координат. Задачи с одной пространственной переменной называются *одномерными*, с двумя или более — *многомерными* (двумерными, трехмерными и т. д.).

Классическим решением уравнения (8.1) называется функция $u(x_1, x_2, \dots, x_k)$, имеющая частные производные до требуемого порядка и обращающая это уравнение в тождество. *Общее решение* ДУЧП зависит от произвольных функций с числом аргументов на единицу меньше, чем у функции $u(x_1, x_2, \dots, x_k)$. Например, дифференциальное уравнение

$$\frac{\partial^2 u(x_1, x_2)}{\partial x_1^2} = 1$$

имеет общее решение

$$u(x_1, x_2) = \frac{x_1^2}{2} + \varphi(x_2)x_1 + \psi(x_2),$$

где $\varphi(x_2)$, $\psi(x_2)$ — произвольные функции. Для выделения частного решения необходимо задавать дополнительные условия на поверхностях размерности $k - 1$ в пространстве независимых переменных x_1, x_2, \dots, x_k . К этим условиям относятся начальные и граничные (краевые) условия.

Порядком уравнения называется наивысший порядок частных производных, входящих в уравнение.

Если функция F линейна относительно искомой функции $u(x_1, x_2, \dots, x_k)$ и ее производных, то уравнение (8.1) называется *линейным*.

Если функция F линейна по высшим производным (n -го порядка), т. е. коэффициенты при высших производных зависят только от функции u и ее производных до $(n - 1)$ -го порядка, то дифференциальное уравнение называется *квазилинейным*.

В данной главе будут рассмотрены численные методы решения классических линейных одномерных уравнений математической физики, в которых в качестве независимых переменных, как правило, используется время и одна пространственная координата.

Пусть D — некоторая односвязная область изменения независимых переменных x, t . Напомним принципы классификации линейных дифференциальных уравнений с двумя независимыми переменными на примере *линейного дифференциального уравнения второго порядка*, записанного в канонической форме:

$$\begin{aligned} & A(x, t) \frac{\partial^2 u(x, t)}{\partial x^2} + 2B(x, t) \frac{\partial^2 u(x, t)}{\partial t \partial x} + C(x, t) \frac{\partial^2 u(x, t)}{\partial t^2} + \\ & + D(x, t) \frac{\partial u(x, t)}{\partial x} + E(x, t) \frac{\partial u(x, t)}{\partial t} + F(x, t)u(x, t) = G(x, t), \quad (x, t) \in D, \end{aligned} \quad (8.2)$$

где $u(x, t)$ — искомая функция; $A(x, t), B(x, t), C(x, t), D(x, t), E(x, t), F(x, t)$ — коэффициенты; $G(x, t)$ — свободный член (правая часть). Предполагается, что коэффициенты и правая часть являются заданными дважды непрерывно дифференцируемыми функциями, причем $|A(x, t)| + |B(x, t)| + |C(x, t)| \neq 0$. Если коэффициенты уравнения не зависят от x и t , то уравнение (8.2) называется *уравнением с постоянными коэффициентами* (в противном случае *уравнением с переменными коэффициентами*). Уравнение (8.2) называется *однородным*, если правая часть $G(x, t)$ тождественно равна нулю для всех x и t (в противном случае *неоднородным*). В уравнении (8.2) независимыми переменными являются x и t , но во многих задачах независимыми переменными являются две пространственные переменные x и y , а искомой функцией $u(x, y)$.

Для упрощения записи будем использовать следующие обозначения:

$$u_x \equiv \frac{\partial u(x, t)}{\partial x}, \quad u_t \equiv \frac{\partial u(x, t)}{\partial t}, \quad u_y \equiv \frac{\partial u(x, y)}{\partial y};$$

$$u_{xx} \equiv \frac{\partial^2 u(x, t)}{\partial x^2}, \quad u_{tt} \equiv \frac{\partial^2 u(x, t)}{\partial t^2}, \quad u_{xt} \equiv \frac{\partial^2 u(x, t)}{\partial t \partial x} \dots$$

Тогда уравнение (8.2) записывается в форме

$$Au_{xx} + 2Bu_{xt} + Cu_{tt} + Du_x + Eu_t + Fu = G. \quad (8.3)$$

Уравнение (8.2) называется *уравнением параболического типа* в точке $M(x_0, t_0)$ (в области D), если в этой точке (в области D) выполняется условие $B^2 - AC = 0$.

Уравнение (8.2) называется *уравнением гиперболического типа* в точке $M(x_0, t_0)$ (в области D), если в этой точке (в области D) выполняется условие $B^2 - AC > 0$.

Уравнение (8.2) называется *уравнением эллиптического типа* в точке $M(x_0, t_0)$ (в области D), если в этой точке (в области D) выполняется условие $B^2 - AC < 0$.

Заметим, что в различных точках (областях) тип одного и того же уравнения может быть различным. Если коэффициенты A, B, C постоянные, тип уравнения (8.2) не изменяется для любых значений независимых переменных в области D . Он определяется только коэффициентами при вторых производных и не зависит ни от коэффициентов при первых производных, ни от свободного члена, ни от самой функции.

Среди задач математической физики, включающих уравнения гиперболического и параболического типов, выделяются *эволюционные (маршевые)* задачи. Для этих задач характерно наличие выделенного независимого переменного — *маршевой координаты*, как правило, времени. Если искомая функция зависит от времени, уравнение называется *нестационарным*, а если искомая функция не зависит от времени, то *стационарным*.

Уравнения параболического типа описывают, например, процессы теплопроводности и диффузии, гиперболического — колебательные системы и волновые движения, эллиптического — течение жидкости в стационарных потоках, стационарное распределение напряженности электрического и магнитного полей. Тип уравнения определяет формулировку задачи и численные методы его решения.

Линейное нестационарное уравнение с частными производными первого порядка имеет вид

$$A(x, t)u_x + B(x, t)u_t + C(x, t)u = G(x, t), \quad (8.4)$$

где $u(x, t)$ — искомая функция; $A(x, t), B(x, t), C(x, t)$ — коэффициенты; $G(x, t)$ — правая часть. Предполагается, что коэффициенты и правая часть являются заданными непрерывно дифференцируемыми функциями. Если коэффициенты не зависят от x и t , то уравнение (8.4) называется *уравнением с постоянными коэффициентами* (в противном случае *уравнением с перемен-*

ными коэффициентами). Уравнение (8.4) называется однородным, если правая часть $G(x, t)$ тождественно равна нулю для всех x и t (в противном случае неоднородным).

Пример 8.1. Классифицировать следующие уравнения:

- а) $u_t + u_x = 0$;
- б) $u_t - uu_{xxx} = \sin x$;
- в) $u_{xx} + xu_{tt} = 0$;
- г) $u_{tt} - c^2u_{xx} = 0$;
- д) $u_t - a^2u_{xx} = 0$;
- е) $u_{xx} + u_{yy} = 0$;
- ж) $xu_x + tu_t + u^2 = 0$;
- з) $u_{xy} = 0$;
- и) $3u_{xx} + 7u_{xt} + 2u_{tt} = 0$;
- к) $u_t u_{xx} - 3x^2uu_{xt} + 3u_x - u = 0$.

□ Для приведения уравнений «е», «з» в соответствие с (8.1), (8.2) обозначим $y = t$. Тогда:

- а) линейное одномерное нестационарное однородное уравнение первого порядка с постоянными коэффициентами, называемое *уравнением переноса*;
- б) квазилинейное одномерное нестационарное уравнение третьего порядка;
- в) линейное одномерное нестационарное однородное уравнение второго порядка с переменными коэффициентами, где $A = 1$, $B = 0$, $C = x$, $D = E = F = G \equiv 0$. Поскольку $B^2 - AC = -x$, то при $x > 0$ это уравнение эллиптического типа, при $x = 0$ — параболического, при $x < 0$ — гиперболического;
- г) линейное одномерное нестационарное однородное уравнение второго порядка с постоянными коэффициентами, где $A = -c^2$, $B = 0$, $C = 1$, $D = E = F = G \equiv 0$. Оно называется одномерным *волновым уравнением* и описывает распространение плоских звуковых волн в покоящейся среде, продольные колебания стержня, поперечные колебания струны. Поскольку $B^2 - AC = c^2 > 0$, то это уравнение гиперболического типа при всех x и t ;
- д) линейное одномерное нестационарное однородное уравнение второго порядка с постоянными коэффициентами, где $A = -a^2$, $B = 0$, $C = 0$, $D = F = G \equiv 0$, $E = 1$. Оно называется одномерным *уравнением теплопроводности*. Поскольку $B^2 - AC = 0$, то это уравнение параболического типа при всех x и t ;
- е) линейное двумерное стационарное однородное уравнение второго порядка с постоянными коэффициентами, где $A = 1$, $B = 0$, $C = 1$, $D = E = F = G \equiv 0$. Оно называется двумерным *уравнением Лапласа*. Поскольку $B^2 - AC = -1 < 0$, то это уравнение эллиптического типа при всех x и y ;
- ж) нелинейное нестационарное уравнение первого порядка;
- з) линейное двумерное стационарное однородное уравнение второго порядка с постоянными коэффициентами, где $A = 0$, $B = 1$, $C = D = E = F = G \equiv 0$. Поскольку $B^2 - AC = 1 > 0$, то это уравнение гиперболического типа при всех x и y ;
- и) линейное одномерное нестационарное однородное уравнение второго порядка с постоянными коэффициентами, где $A = 3$, $B = \frac{7}{2}$, $C = 2$, $D = E = F = G \equiv 0$. Поскольку $B^2 - AC = \frac{49}{4} - 6 > 0$, то это уравнение гиперболического типа при всех x и t ;

к) квазилинейное одномерное нестационарное уравнение второго порядка, так как оно линейно относительно старших производных u_{xx} и u_{xt} , а коэффициенты при этих производных зависят только от u и производных первого порядка. ■

Математические постановки задач с дифференциальными уравнениями в частных производных содержат дифференциальные уравнения и дополнительные условия, позволяющие выделить искомые частные решения среди целого семейства решений.

В математической физике различают три основных типа задач:

1) *задача Коши* (с начальными условиями);

2) *краевые (граничные) задачи* с краевыми (граничными) условиями;

3) *смешанные задачи* (начально-краевые) с начальными и краевыми условиями.

Количество и характер краевых условий в краевых и смешанных задачах определяется типом дифференциального уравнения, его порядком и характером физических процессов, для которых строится математическая модель. Если количество граничных условий превышает нужное число, то задача становится переопределенной и, как правило, не имеет решения. Если их меньше, чем требуется для разрешимости задачи, то ее решение не является единственным. Для анализа разрешимости задач доказываются теоремы о существовании и единственности решения (где это возможно). Для любого заданного дифференциального уравнения целесообразность рассмотрения той или иной задачи определяется корректностью ее постановки. Задача *корректно поставлена*, если ее решение существует, единственно и непрерывно зависит от данных задачи.

Рассмотрим постановки некоторых типовых задач математической физики.

ПОСТАНОВКИ ЗАДАЧ ДЛЯ УРАВНЕНИЙ ПЕРВОГО ПОРЯДКА

Рассмотрим наиболее типичное уравнение первого порядка — одномерное уравнение *переноса*

$$\frac{\partial u(x, t)}{\partial t} + \frac{\partial u(x, t)}{\partial x} = 0, \quad 0 < x < L, \quad 0 < t < T, \quad (8.5)$$

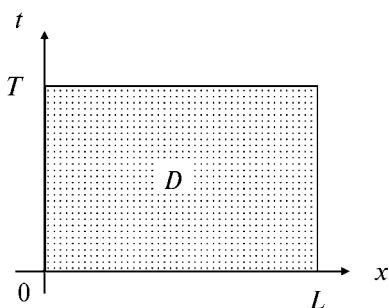


Рис. 8.1

являющееся модельным для некоторых процессов в механике сплошной среды. Оно описывает конвективный одномерный перенос тепла и является эволюционным. Здесь $x = 0$ и $x = L$ — левый и правый концы отрезка изменения пространственной переменной, $t = 0$ и $t = T$ — моменты начала и окончания процесса. Таким образом, задается область $D = (0, L) \times (0, T)$, ограниченная прямоугольником (рис. 8.1) с длиной L и высотой T . Соотношения $t = 0$, $0 \leq x \leq L$ задают *нижнюю границу*

области D , соотношения $t = T$, $0 < x < L$ — верхнюю, $x = 0$, $0 < t \leq T$ — левую границу, $x = L$, $0 < t \leq T$ — правую. Далее на множестве $\bar{D} = D \cup \partial D = [0, L] \times [0, T]$, где ∂D — граница области D , рассматриваются различные начально-краевые задачи для уравнения (8.5).

Начально-краевая задача:

$$\begin{aligned} u_t + u_x &= 0, \quad 0 < x < L, \quad 0 < t < T; \\ u(x, 0) &= \psi(x), \quad 0 \leq x \leq L \text{ (начальное условие);} \\ u(0, t) &= \phi(t), \quad 0 < t \leq T \text{ (краевое условие на левой границе)} \end{aligned}$$

содержит функциональное начальное условие (при $t = 0$) и функциональное краевое условие на левой ($x = 0$) границе области. Здесь и далее $\psi(x)$, $\phi(t)$ — заданные функции, удовлетворяющие условию их согласования в начале координат [7].

Задача Коши:

$$\begin{aligned} u_t + u_x &= 0, \quad -\infty < x < +\infty, \quad 0 < t < T; \\ u(x, 0) &= \psi(x), \quad -\infty < x < +\infty \text{ (начальное условие)} \end{aligned}$$

содержит только функциональное начальное условие (при $t = 0$) и рассматривается в бесконечной области изменения пространственной переменной.

Во всех перечисленных задачах требуется найти функцию $u(x, t)$, которая удовлетворяет дифференциальному уравнению (8.5) в области D и соответствующим условиям на ее границе.

ПОСТАНОВКИ ЗАДАЧ ДЛЯ УРАВНЕНИЙ ПАРАБОЛИЧЕСКОГО ТИПА

Рассмотрим наиболее типичное уравнение параболического типа — уравнение *теплопроводности в однородной среде*, или *диффузии*,

$$\frac{\partial u(x, t)}{\partial t} = a^2 \frac{\partial^2 u(x, t)}{\partial x^2}, \quad 0 < x < L, \quad 0 < t < T, \quad (8.6)$$

где u — температура; $a^2 = \frac{k}{\rho c_p}$; k — коэффициент теплопроводности; c — удельная теплоемкость; ρ — плотность (в задаче диффузии u — концентрация дифундирующего вещества, $a^2 = \frac{d}{c}$, d — коэффициент диффузии, c — коэффициент пористости среды, который определяется отношением объема пор к рассматриваемому объему); $x = 0$ и $x = L$ — левый и правый концы отрезка изменения пространственной переменной; $t = 0$ и $t = T$ — моменты начала и окончания процесса. Таким образом, на множестве $\bar{D} = [0, L] \times [0, T]$ рассматриваются различные начально-краевые задачи для уравнения (8.6) (рис. 8.1).

Первая начально-краевая задача:

$$\begin{aligned} u_t &= a^2 u_{xx}, \quad 0 < x < L, \quad 0 < t < T; \\ u(x, 0) &= \psi(x), \quad 0 \leq x \leq L \text{ (начальное условие);} \end{aligned}$$

$$\begin{aligned} u(0, t) &= \varphi_1(t), \quad 0 < t \leq T \text{ (краевое условие на левой границе);} \\ u(L, t) &= \varphi_2(t), \quad 0 < t \leq T \text{ (краевое условие на правой границе)} \end{aligned}$$

содержит функциональное начальное условие (при $t = 0$) и функциональные краевые условия на левой ($x = 0$) и правой ($x = L$) границах области. Здесь и далее $\psi(x)$, $\varphi_1(t)$, $\varphi_2(t)$ — заданные функции.

Вторая начально-краевая задача:

$$\begin{aligned} u_t &= a^2 u_{xx}, \quad 0 < x < L, \quad 0 < t < T; \\ u(x, 0) &= \psi(x), \quad 0 \leq x \leq L \text{ (начальное условие);} \\ \frac{\partial u(0, t)}{\partial x} &= \varphi_1(t), \quad 0 < t \leq T \text{ (краевое условие на левой границе);} \\ \frac{\partial u(L, t)}{\partial x} &= \varphi_2(t), \quad 0 < t \leq T \text{ (краевое условие на правой границе)} \end{aligned}$$

содержит функциональное начальное условие (при $t = 0$) и дифференциальные краевые условия на левой ($x = 0$) и правой ($x = L$) границах области.

Третья начально-краевая задача:

$$\begin{aligned} u_t &= a^2 u_{xx}, \quad 0 < x < L, \quad 0 < t < T; \\ u(x, 0) &= \psi(x), \quad 0 \leq x \leq L \text{ (начальное условие);} \\ \alpha_0 \frac{\partial u(0, t)}{\partial x} + \beta_0 u(0, t) &= \varphi_1(t), \quad 0 < t \leq T \text{ (краевое условие на левой границе);} \\ \alpha_1 \frac{\partial u(L, t)}{\partial x} + \beta_1 u(L, t) &= \varphi_2(t), \quad 0 < t \leq T \text{ (краевое условие на правой границе)} \end{aligned}$$

содержит функциональное начальное условие (при $t = 0$) и функционально-дифференциальные краевые условия на левой ($x = 0$) и правой ($x = L$) границах области. Здесь α_0 , β_0 , α_1 , β_1 — заданные числа ($\alpha_i^2 + \beta_i^2 > 0$, $i = 1, 2$) либо функции от t в более общем случае [26].

Задача Коши:

$$\begin{aligned} u_t &= a^2 u_{xx}, \quad -\infty < x < +\infty, \quad 0 < t < T; \\ u(x, 0) &= \psi(x), \quad -\infty < x < +\infty \text{ (начальное условие)} \end{aligned}$$

содержит только функциональное начальное условие (при $t = 0$) и рассматривается в бесконечной области изменения пространственной переменной.

Примером физической задачи, приводящей к первой начально-краевой задаче, может быть процесс теплопередачи по длинному тонкому стержню, лежащему вдоль оси Ox от $x = 0$ до $x = L$ (ось стержня совпадает с осью Ox). Предполагается, что в точке $x = 0$ температура изменяется со временем по закону $\varphi_1(t)$, а в точке $x = L$ по закону $\varphi_2(t)$. В начальный момент времени при $t = 0$ функцией $\psi(x)$ задано начальное распределение температуры вдоль стержня. Тогда распределение температуры вдоль него во все последующие моменты времени определяется решением начально-краевой задачи с уравнени-

нием (8.6), где $u(x, t)$ — температура стержня в некоторой точке x в момент времени t .

Во всех перечисленных задачах требуется найти функцию $u(x, t)$, которая удовлетворяет дифференциальному уравнению (8.6) в области D и соответствующим условиям на ее границе.

ПОСТАНОВКИ ЗАДАЧ ДЛЯ УРАВНЕНИЙ ГИПЕРБОЛИЧЕСКОГО ТИПА

Рассмотрим наиболее типичное уравнение гиперболического типа — одномерное *волновое уравнение*, которое описывает, в частности, малые попеченные колебания струны:

$$\frac{\partial^2 u(x, t)}{\partial t^2} = c^2 \frac{\partial^2 u(x, t)}{\partial x^2}, \quad 0 < x < L, \quad 0 < t < T, \quad (8.7)$$

где $u(x, t)$ — отклонение струны от положения равновесия в некоторой точке x в момент времени t ; коэффициент $c^2 = \frac{T}{\rho}$, ρ — линейная плотность струны, T — натяжение; $x = 0$ и $x = L$ — левый и правый концы отрезка изменения пространственной переменной; $t = 0$ и $t = T$ — моменты начала и окончания процесса. Из предположения о малости колебаний следует, что величина натяжения, возникающего в струне, не зависит от времени. При этом также считается, что внешние силы, которые могли бы действовать на струну, отсутствуют.

Волновое уравнение описывает также малые продольные колебания упругого стержня. Тогда $c^2 = \frac{k}{\rho}$, k — модуль Юнга, ρ — плотность. Это же уравнение описывает распространение звуковых волн в сжимаемой среде и называется уравнением акустики, в котором $c^2 = \frac{\gamma p_0}{\rho_0}$; $\gamma = \frac{c_p}{c_v}$; p_0 , ρ_0 — давление и плотность невозмущенной среды; c_v , c_p — коэффициенты теплоемкости при постоянном объеме и давлении соответственно. Для газовой среды это уравнение может быть получено из уравнений газовой динамики путем их линеаризации.

Таким образом, на множестве $\bar{D} = [0, L] \times [0, T]$ рассматриваются различные начально-краевые задачи для уравнения (8.7) (см. рис. 8.1).

Первая начально-краевая задача:

$$u_{tt} = c^2 u_{xx}, \quad 0 < x < L, \quad 0 < t < T;$$

$$u(x, 0) = \psi_1(x), \quad 0 \leq x \leq L \quad (\text{начальное условие});$$

$$\frac{\partial u(x, 0)}{\partial t} = \psi_2(x), \quad 0 \leq x \leq L \quad (\text{начальное условие});$$

$$u(0, t) = \phi_1(t), \quad 0 < t \leq T \quad (\text{краевое условие на левой границе});$$

$$u(L, t) = \phi_2(t), \quad 0 < t \leq T \quad (\text{краевое условие на правой границе})$$

содержит функциональное и дифференциальное начальные условия (при $t = 0$) и функциональные краевые условия на левой ($x = 0$) и правой ($x = L$) границах области. Здесь и далее $\psi_1(x)$, $\psi_2(x)$, $\varphi_1(t)$, $\varphi_2(t)$ — заданные функции.

Вторая начально-краевая задача:

$$u_{tt} = c^2 u_{xx}, \quad 0 < x < L, \quad 0 < t < T;$$

$$u(x, 0) = \psi_1(x), \quad 0 \leq x \leq L \text{ (начальное условие);}$$

$$\frac{\partial u(x, 0)}{\partial t} = \psi_2(x), \quad 0 \leq x \leq L \text{ (начальное условие);}$$

$$\frac{\partial u(0, t)}{\partial x} = \varphi_1(t), \quad 0 < t \leq T \text{ (краевое условие на левой границе);}$$

$$\frac{\partial u(L, t)}{\partial x} = \varphi_2(t), \quad 0 < t \leq T \text{ (краевое условие на правой границе)}$$

содержит функциональное и дифференциальное начальные условия (при $t = 0$) и дифференциальные краевые условия на левой ($x = 0$) и правой ($x = L$) границах области.

Третья начально-краевая задача:

$$u_{tt} = c^2 u_{xx}, \quad 0 < x < L, \quad 0 < t < T;$$

$$u(x, 0) = \psi_1(x), \quad 0 \leq x \leq L \text{ (начальное условие);}$$

$$\frac{\partial u(x, 0)}{\partial t} = \psi_2(x), \quad 0 \leq x \leq L \text{ (начальное условие);}$$

$$\alpha_0 \frac{\partial u(0, t)}{\partial x} + \beta_0 u(0, t) = \varphi_1(t), \quad 0 < t \leq T \text{ (краевое условие на левой границе);}$$

$$\alpha_1 \frac{\partial u(L, t)}{\partial x} + \beta_1 u(L, t) = \varphi_2(t), \quad 0 < t \leq T \text{ (краевое условие на правой границе)}$$

содержит функциональное и дифференциальное начальные условия (при $t = 0$) и функционально-дифференциальные краевые условия на левой ($x = 0$) и правой ($x = L$) границах области. Здесь α_0 , β_0 , α_1 , β_1 — заданные числа, $|\alpha_i| + |\beta_i| > 0$. В более общем случае коэффициенты, входящие в краевые условия, являются функциями времени [26].

Задача Коши:

$$u_{tt} = c^2 u_{xx}, \quad -\infty < x < +\infty, \quad 0 < t < T;$$

$$u(x, 0) = \psi_1(x), \quad -\infty < x < +\infty \text{ (начальное условие);}$$

$$\frac{\partial u(x, 0)}{\partial t} = \psi_2(x), \quad -\infty < x < +\infty \text{ (начальное условие)}$$

содержит только функциональное и дифференциальное начальные условия (при $t = 0$) и рассматривается в бесконечной области изменения пространственной переменной.

Во всех перечисленных задачах требуется найти функцию $u(x, t)$, которая удовлетворяет дифференциальному уравнению (8.7) в области D и соответствующим условиям на ее границе.

ПОСТАНОВКИ ЗАДАЧ ДЛЯ УРАВНЕНИЙ ЭЛЛИПТИЧЕСКОГО ТИПА

Рассмотрим наиболее типичное уравнение эллиптического типа — *уравнение Лапласа*

$$\frac{\partial^2 u(x, y)}{\partial x^2} + \frac{\partial^2 u(x, y)}{\partial y^2} = 0, \quad (x, y) \in \Omega, \quad (8.8)$$

или более общее *уравнение Пуассона*

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

где Ω — некоторая односвязная область на плоскости Oxy с простой (без кратных точек) кусочно-гладкой границей Γ (рис. 8.2). В простейшем случае, рассматриваемом далее, область Ω имеет прямоугольную форму.

Уравнение Пуассона описывает стационарное течение идеальной жидкости, в которой отсутствует вязкость и теплопроводность, стационарное распределение тепла в теле и др. Это же уравнение описывает стационарное распределение напряженности электрического или магнитного поля в электродинамике. Уравнение Лапласа описывает эти явления, когда внутри Ω нет источников и стоков, а уравнение Пуассона — с распределенными по области Ω источниками, задаваемыми правой частью $f(x, y)$.

Поскольку в уравнении (8.8) отсутствуют производные по времени, задача является стационарной, следовательно, начальные условия в ней не задаются. На границе Γ области Ω могут задаваться краевые (граничные) условия трех видов, что порождает три краевые задачи.

Первая краевая задача (задача Дирихле):

$$u_{xx} + u_{yy} = 0, \quad (x, y) \in \Omega; \\ u|_{\Gamma} = \varphi(x, y), \quad (x, y) \in \Gamma \quad (\text{краевое условие})$$

содержит функциональное краевое условие на границе Γ области Ω , $\varphi(x, y)$ — заданная функция.

Вторая краевая задача (задача Неймана):

$$u_{xx} + u_{yy} = 0, \quad (x, y) \in \Omega; \\ \left. \frac{\partial u}{\partial n} \right|_{\Gamma} = \varphi(x, y), \quad (x, y) \in \Gamma \quad (\text{краевое условие})$$

содержит дифференциальное краевое условие, в котором задается производная в направлении внешней нормали n на границе Γ области Ω (рис. 8.2).

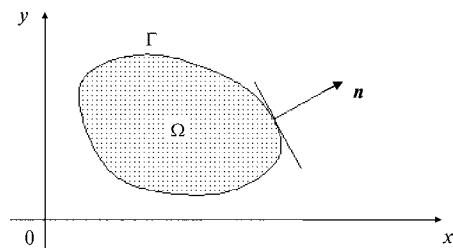


Рис. 8.2

Третья краевая задача:

$$u_{xx} + u_{yy} = 0, \quad (x, y) \in \Omega;$$
$$\alpha \frac{\partial u}{\partial n} \Big|_{\Gamma} + \beta u \Big|_{\Gamma} = \varphi(x, y), \quad (x, y) \in \Gamma \quad (\text{краевое условие})$$

содержит функционально-дифференциальное краевое условие, в котором задается линейная комбинация производной в направлении внешней нормали и функции $u(x, y)$ на границе Γ области Ω . В более общем случае α и β зависят от переменных x и y [26].

В качестве примера задачи Дирихле можно привести задачу о нахождении стационарного распределения температуры внутри области, если задана температура на ее границе. Примером задачи Неймана служит задача стационарной теплопроводности и электростатики, если на границе задан поток (тепла, электронов).

Во всех перечисленных задачах требуется найти функцию $u(x, y)$, которая удовлетворяет дифференциальному уравнению (8.8) в области Ω и соответствующему краевому условию на ее границе Γ .

Замечания

1. Тип неравенств (строгое или нестрогое) в краевых и начальных условиях в каждом конкретном случае определяется особенностями физической задачи.

2. Для существования достаточно гладких решений поставленных задач требуется выполнение условий согласования начальных и краевых условий в углах расчетной области [7], [8], [14].

8.2. ПРИНЦИПЫ ПОСТРОЕНИЯ РАЗНОСТНЫХ СХЕМ

Точное решение задач математической физики (в виде явных формул, рядов и т. п.) можно найти только в редких случаях. Среди приближенных методов, в результате применения которых получается приближенное (неточное) решение, представляемое таблицей чисел, наибольшее распространение получили *разностные методы (методы сеток)*. Сущность разностных методов состоит в том, что исходная область изменения независимых переменных заменяется дискретным множеством точек — сеткой, а производные, входящие в уравнение, аппроксимируются на этой сетке разностными соотношениями. В результате исходная линейная задача заменяется системой конечного числа линейных алгебраических уравнений, называемой *разностной схемой (задачей)*. Аналогично исходная нелинейная задача заменяется нелинейной разностной схемой. За приближенное решение исходной задачи принимается решение разностной схемы. Точность приближения зависит от способа аппроксимации и от густоты сетки, т. е. от того, насколько плотно сетка заполняет исходную область.

Рассмотрим общую запись постановок линейных задач, описанных в п. 8.1. Пусть дана исходная (*дифференциальная*) задача в виде

$$\mathcal{L}u = f, \quad (8.9)$$

где u — искомая функция, определенная на множестве $\bar{D} = D \cup \Gamma$; D — область пространства независимых переменных с границей Γ ; f — заданная функция; \mathcal{L} — линейный дифференциальный оператор. Предполагается, что все производные, входящие в дифференциальное уравнение, перенесены в левую часть, а остальные функции образуют правую часть; дополнительные условия (начальные и краевые) также включены в оператор \mathcal{L} и правую часть f . Например, начально-краевая задача для уравнения переноса (8.5) записывается в виде (8.9), если положить

$$\mathcal{L}u = \begin{cases} u_t + u_x, & 0 < x < L, \quad 0 < t < T, \\ u(x, 0), & 0 \leq x \leq L, \\ u(0, t), & 0 < t \leq T; \end{cases} \quad f = \begin{cases} 0, & 0 < x < L, \quad 0 < t < T, \\ \psi(x), & 0 \leq x \leq L, \\ \phi(t), & 0 < t \leq T. \end{cases}$$

Для численного решения задачи вводится сетка $D_h = \{M_h\}$ — конечное множество точек M_h (узлов сетки), принадлежащих \bar{D} , плотность размещения которых характеризуется параметром h — шагом сетки. В общем случае h — вектор, компонентами которого являются шаги по всем независимым переменным решаемой задачи с длиной $|h|$. Обычно сетка задается так, что при $|h| \rightarrow 0$ множество D_h стремится заполнить множество $\bar{D} = D \cup \Gamma$.

Для определенности далее рассматривается некоторая дифференциальная задача с двумя независимыми переменными x и t . Для простоты изложения будем предполагать, что множество D представляет собой прямоугольник длины L и высоты T , ограниченный отрезками прямых, параллельных осям Ox и Ot (рис. 8.3), т. е. задана двумерная прямоугольная сетка

$$D_h = \{x_i, t^n\},$$

где $x_i = ih$, $0 \leq i \leq I$; $t^n = n\tau$, $0 \leq n \leq N$; N, I — целые положительные числа; h, τ — величины шагов по пространству и времени (для простоты принимаются постоянными); $Ih = L$, $N\tau = T$. Такая сетка называется *равномерной (регулярной)*. Здесь можно положить $|h| = \sqrt{h^2 + \tau^2}$ или $|h| = \sqrt{h^2 + \tau}$. Узлы, принадлежащие промежуткам $\{0 \leq x \leq L, t = 0\}$, $\{x = 0, 0 \leq t \leq T\}$, $\{x = L, 0 \leq t \leq T\}$, $\{t = T, 0 < x < L\}$, называются *граничными*, а остальные — *внутренними*. Слоем $\{x_i, t^n\}$, $i = 0, \dots, I - 1$; $n = \text{const}$ называется множество всех узлов сетки, имеющих одну и ту же временную координату t^n .

Функции, определенные в точках сетки D_h , называются *сеточными*. Введем сеточную функцию $u_h = \{u_i^n = u(x_i, t^n), 0 \leq i \leq I, 0 \leq n \leq N\}$, которая является *сеточным представлением решения исходной (дифференциальной) задачи*, или точным решением дифференциальной задачи в узлах сетки. Как правило, вычислить u_h не удается, поэтому находят другую сеточную функцию

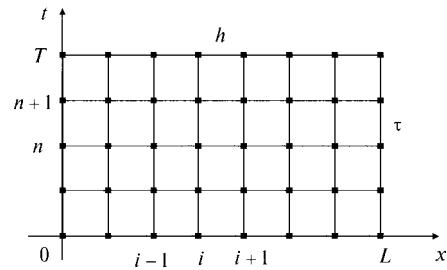


Рис. 8.3

$\hat{u}_h \cong u_h$, приближенно совпадающую с точным решением в узлах сетки. Она вычисляется как решение разностной схемы

$$\mathcal{L}_h \hat{u}_h = f_h, \quad (8.10)$$

в некотором смысле соответствующей задаче (8.9). Здесь \mathcal{L}_h — разностный оператор, аппроксимирующий линейный дифференциальный оператор \mathcal{L} (он формируется в результате аппроксимации частных производных, входящих в \mathcal{L} , соответствующими конечно-разностными соотношениями); f_h — сеточная функция, возникающая в результате замены правой части уравнения (8.9) значениями в узлах сетки. Таким образом, под *разностной схемой* понимается совокупность разностных уравнений, аппроксимирующих основное дифференциальное уравнение во всех внутренних узлах сетки и дополнительные условия (начальные и краевые) — в граничных узлах. Разностную схему по аналогии с дифференциальной задачей будем называть *разностной задачей*.

Обозначим линейное нормированное пространство, образованное совокупностью функций \hat{u}_h , определенных на D_h , через U_h , а пространство, образованное совокупностью функций f_h , через F_h . Пусть в этих пространствах введены нормы $\|\cdot\|_{U_h}$, $\|\cdot\|_{F_h}$.

Если при $|h| \rightarrow 0$ выполняется условие

$$\|\hat{u}_h - u_h\|_{U_h} \rightarrow 0, \quad (8.11)$$

то решение разностной задачи сводится к решению дифференциальной, а разностная схема называется *сходящейся*. Если существуют такие постоянные $p > 0$, $C > 0$, не зависящие от $|h|$, что выполняется неравенство

$$\|u_h - \hat{u}_h\|_{U_h} \leq C|h|^p,$$

то схема имеет *p-й порядок сходимости (порядок точности)*.

Как указано выше, основной идеей разностных методов является преобразование дифференциальной задачи к разностной задаче для нахождения сеточного решения, в некотором смысле близкого к точному решению.

Функция δf_h в выражении $\mathcal{L}_h u_h = f_h + \delta f_h$, получающаяся в результате подстановки сеточного представления u_h в разностную схему (8.10), называется *погрешностью аппроксимации*.

Разностная задача (8.10) *аппроксимирует* дифференциальную задачу (8.9) на решении $u(x, t)$, если

$$\|\delta f_h\|_{F_h} \rightarrow 0 \text{ при } |h| \rightarrow 0.$$

Если существуют такие постоянные $k > 0$, $M > 0$, не зависящие от $|h|$, что выполняется неравенство

$$\|\delta f_h\|_{F_h} \leq M|h|^k, \quad (8.12)$$

то разностная схема имеет *k-й порядок аппроксимации*.

Одновременное выполнение условия $\|\delta f_h\|_{F_h} = O(\tau^p + h^q)$ означает, что схема имеет *p-й порядок аппроксимации по времени и q-й по пространству*. В случае трех независимых переменных аналогичное условие имеет вид

$$\|\delta f_h\|_{F_h} = O(\tau^p + h_x^q + h_y^r).$$

Разностная схема называется *устойчивой*, если существует постоянная $K > 0$, не зависящая от $|h|$, что при любых $f_h \in F_h$ справедливы условия:

- 1) разностная схема (8.10) имеет единственное решение;
- 2) $\|\hat{u}_h\|_{U_h} \leq K \|f_h\|_{F_h}$.

Замечания

1. Свойство устойчивости какой-либо задачи означает, что при небольшом изменении исходных данных решение изменяется мало. Таким образом, для исследования устойчивости необходимо рассматривать уравнение, которому удовлетворяет погрешность, возникающая в результате возмущения исходных данных. Однако в случае линейного оператора структура уравнения для погрешности та же, что и исходного уравнения (8.9). Действительно, рассмотрим уравнение $\mathcal{L}_h \hat{z}_h = \tilde{f}_h$, отличающееся от (8.10) правой частью. Вычитая его из (8.10), получаем уравнение $\mathcal{L}_h \varepsilon_h = \delta f_h$, описывающее изменение погрешности $\varepsilon_h = \hat{u}_h - \tilde{z}_h$ в силу наличия возмущения $\delta f_h = f_h - \tilde{f}_h$ правой части. Тогда если выполнено условие 2) определения устойчивости, то одновременно $\|\varepsilon_h\|_{U_h} \leq K \|\delta f_h\|_{F_h}$ и, следовательно, $\|\varepsilon_h\|_{U_h} \rightarrow 0$ при $\|\delta f_h\|_{F_h} \rightarrow 0$. Это означает, что задача устойчива.

2. Свойство устойчивости связано с понятием норм, вводимых в пространствах U_h , F_h . Возможны случаи, когда условие 2) будет выполняться для одних норм и не выполняться для других. Если это условие не выполняется ни при каком разумном выборе норм, то схема неустойчива.

Разностные схемы, устойчивые при любом соотношении шагов h и τ , называются *абсолютно устойчивыми*.

Разностные схемы, неустойчивые при любом соотношении шагов h и τ , называются *абсолютно неустойчивыми*.

Разностные схемы, устойчивые лишь при некотором ограничении на отношение шагов по пространству и по времени, называются *условно устойчивыми*.

Сходимость и устойчивость связаны теоремой, играющей основную роль при анализе разностных схем.

Теорема 8.1[40]. Пусть дифференциальная задача (8.9) поставлена корректно, разностная схема (8.10) устойчива, аппроксимирует дифференциальную задачу (8.9) и имеет k -й порядок аппроксимации. Тогда решение разностной задачи сходится к решению дифференциальной задачи, причем порядок точности совпадает с порядком аппроксимации.

Установить устойчивость разностной схемы с использованием определения на практике затруднительно. Поэтому предложен ряд методов, позволяющих получить необходимые (иногда необходимые и достаточные) условия устойчивости разностных схем: спектральный признак устойчивости, метод гармоник Фурье, принцип максимума, метод операторных неравенств [14], [35], [40]. В данной книге эти методы не рассматриваются.

Схема называется *явной*, если оператор \mathcal{L} аппроксимируется с использованием известных значений функции $\hat{u}(x_i, t^n)$ на n -м слое, а аппроксимирующее уравнение содержит только одно неизвестное значение функции на следующем $(n+1)$ -м слое, которое нетрудно выразить явно. Схема называется *неявной*, если оператор \mathcal{L} аппроксимируется с использованием нескольких

неизвестных значений искомой функции на $(n + 1)$ -м слое. Узлы сетки D_h , значения в которых используются при аппроксимации оператора \mathcal{L} , образуют *шаблон*. При изображении шаблона светлыми кружочками обозначаются узлы на n -м слое с известными значениями функций, а зачерненными — узлы на $(n + 1)$ -м слое с неизвестными значениями функций, подлежащими определению. Шаблон, содержащий p узлов, называется *p-точечным*.

Получим некоторые разностные формулы для аппроксимации частных производных первого и второго порядков, входящих в уравнения видов (8.3), (8.4) и в соответствующие дифференциальные задачи, рассмотренные далее.

Пусть имеется функция $u(x, t)$, имеющая непрерывные производные по всем переменным до $(k + 1)$ -го порядка включительно. Выберем узел (x_i, t^n) сетки D_h . Сначала рассмотрим разности в направлении x .

Применим формулу Тейлора (B.16)–(B.18) разложения функции $u(x, t^n)$ одной переменной x в окрестности выбранного узла (x_i, t^n) .

При $k = 1$ имеем

$$u(x_i + h, t^n) = u(x_i, t^n) + u_x(x_i, t^n)h + u_{xx}(\xi_1, t^n) \frac{h^2}{2}, \quad \xi_1 \in (x_i, x_i + h); \quad (8.13)$$

$$u(x_i - h, t^n) = u(x_i, t^n) - u_x(x_i, t^n)h + u_{xx}(\xi_2, t^n) \frac{h^2}{2}, \quad \xi_2 \in (x_i - h, x_i). \quad (8.14)$$

Из (8.13) следует

$$u_x(x_i, t^n) = \frac{u(x_i + h, t^n) - u(x_i, t^n)}{h} - u_{xx}(\xi_1, t^n) \frac{h}{2}.$$

Очевидно, справедлива оценка

$$\left| -u_{xx}(\xi_1, t^n) \frac{h}{2} \right| \leq \frac{h}{2} \max_{x \in [x_i, x_i + h]} |u_{xx}(x, t^n)| = \frac{h}{2} M_2,$$

где $M_2 = \max_{x \in [x_i, x_i + h]} |u_{xx}(x, t^n)|$. Отсюда получаем формулу (аппроксимационный оператор)

$$\hat{u}_x(x_i, t^n) = \frac{u(x_i + h, t^n) - u(x_i, t^n)}{h} \left(\frac{h}{2} M_2 \right), \quad (8.15)$$

которая аппроксимирует частную производную u_x с первым порядком аппроксимации на двухточечном шаблоне. Она называется *правой разностью*.

Аналогично из (8.14) получается формула

$$\hat{u}_x(x_i, t^n) = \frac{u(x_i, t^n) - u(x_i - h, t^n)}{h}, \quad (8.16)$$

также аппроксимирующая частную производную u_x с первым порядком аппроксимации на двухточечном шаблоне. Она называется *левой разностью*.

Получим еще одну формулу для аппроксимации частной производной u_x . При $k = 2$ из формулы Тейлора (B.16)–(B.18) получаем

$$u(x_i + h, t^n) = u(x_i, t^n) + u_x(x_i, t^n)h + u_{xx}(x_i, t^n)\frac{h^2}{2} + u_{xxx}(\xi_1, t^n)\frac{h^3}{6},$$

$$\xi_1 \in (x_i, x_i + h); \quad (8.17)$$

$$u(x_i - h, t^n) = u(x_i, t^n) - u_x(x_i, t^n)h + u_{xx}(x_i, t^n)\frac{h^2}{2} - u_{xxx}(\xi_2, t^n)\frac{h^3}{6},$$

$$\xi_2 \in (x_i - h, x_i). \quad (8.18)$$

Вычитая (8.18) из (8.17), имеем

$$u(x_i + h, t^n) - u(x_i - h, t^n) = 2u_x(x_i, t^n)h + [u_{xxx}(\xi_1, t^n) + u_{xxx}(\xi_2, t^n)]\frac{h^3}{6}.$$

Но, согласно утверждению В.1, существует точка $\xi \in (x_i - h, x_i + h)$, что

$$\frac{u_{xxx}(\xi_1, t^n) + u_{xxx}(\xi_2, t^n)}{2} = u_{xxx}(\xi, t^n).$$

Поэтому

$$u_x(x_i, t^n) = \frac{u(x_i + h, t^n) - u(x_i - h, t^n)}{2h} - u_{xxx}(\xi, t^n)\frac{h^2}{6}, \quad \xi \in (x_i - h, x_i + h)$$

или

$$\hat{u}_x(x_i, t^n) = \frac{u(x_i + h, t^n) - u(x_i - h, t^n)}{2h} \left(\frac{h^2}{6} M_3 \right), \quad (8.19)$$

где $M_3 = \max_{x \in [x_i - h, x_i + h]} |u_{xxx}(x, t^n)|$. Эту формулу называют *центральной разностью*, которая аппроксимирует частную производную u_x со вторым порядком аппроксимации на трехточечном шаблоне.

Обозначая $u(x_i, t^n) = u_i^n$, $u(x_i + h, t^n) = u_{i+1}^n$, $u(x_i - h, t^n) = u_{i-1}^n$, перепишем (8.15), (8.16), (8.19) в форме

$$\hat{u}_x(x_i, t^n) = \frac{u_{i+1}^n - u_i^n}{h}; \quad (8.20)$$

$$\hat{u}_x(x_i, t^n) = \frac{u_i^n - u_{i-1}^n}{h}; \quad (8.21)$$

$$\hat{u}_x(x_i, t^n) = \frac{u_{i+1}^n - u_{i-1}^n}{2h}. \quad (8.22)$$

Теперь получим формулы для аппроксимации производных второго порядка. При $k = 3$ формула Тейлора (В.16)–(В.18) имеет вид

$$u(x_i + h, t^n) = u(x_i, t^n) + u_x(x_i, t^n)h + u_{xx}(x_i, t^n)\frac{h^2}{2} + u_{xxx}(x_i, t^n)\frac{h^3}{6} +$$

$$+ u_{xxxx}(\xi_1, t^n)\frac{h^4}{24}, \quad \xi_1 \in (x_i, x_i + h); \quad (8.23)$$

$$u(x_i - h, t^n) = u(x_i, t^n) - u_x(x_i, t^n)h + u_{xx}(x_i, t^n)\frac{h^2}{2} - u_{xxx}(x_i, t^n)\frac{h^3}{6} +$$

$$+ u_{xxxx}(\xi_2, t^n)\frac{h^4}{24}, \quad \xi_2 \in (x_i - h, x_i). \quad (8.24)$$

Сложим (8.23) и (8.24):

$$u(x_i + h, t^n) + u(x_i - h, t^n) = \\ = 2u(x_i, t^n) + u_{xx}(x_i, t^n)h^2 + [u_{xxxx}(\xi_1, t^n) + u_{xxxx}(\xi_2, t^n)]\frac{h^4}{24}.$$

Поскольку, согласно утверждению В.1,

$$\frac{u_{xxxx}(\xi_1, t^n) + u_{xxxx}(\xi_2, t^n)}{2} = u_{xxxx}(\xi, t^n),$$

где $\xi \in (x_i - h, x_i + h)$, то перепишем полученное равенство в виде

$$u_{xx}(x_i, t^n) = \frac{u(x_i + h, t^n) - 2u(x_i, t^n) + u(x_i - h, t^n)}{h^2} - u_{xxxx}(\xi, t^n)\frac{h^2}{12}$$

или окончательно

$$\hat{u}_{xx}(x_i, t^n) = \frac{u(x_i + h, t^n) - 2u(x_i, t^n) + u(x_i - h, t^n)}{h^2} \left(\frac{h^2}{12} M_4 \right), \quad (8.25)$$

где $M_4 = \max_{x \in [x_i - h, x_i + h]} |u_{xxxx}(x, t^n)|$. Формула (8.25), называемая *центальной разностью второго порядка*, аппроксимирует u_{xx} со вторым порядком аппроксимации и может быть записана аналогично (8.20)–(8.22):

$$\hat{u}_{xx}(x_i, t^n) = \frac{u_{i+1}^n - 2u_i^n + u_{i-1}^n}{h^2}. \quad (8.26)$$

Так же получаются разности в узле (x_i, t^n) в направлении t :

$$\hat{u}_t(x_i, t^n) = \frac{u(x_i, t^n + \tau) - u(x_i, t^n)}{\tau}, \quad \hat{u}_t(x_i, t^n) = \frac{u_i^{n+1} - u_i^n}{\tau}; \quad (8.27)$$

$$\hat{u}_t(x_i, t^n) = \frac{u(x_i, t^n) - u(x_i, t^n - \tau)}{\tau}, \quad \hat{u}_t(x_i, t^n) = \frac{u_i^n - u_i^{n-1}}{\tau}; \quad (8.28)$$

$$\hat{u}_t(x_i, t^n) = \frac{u(x_i, t^n + \tau) - u(x_i, t^n - \tau)}{2\tau}, \quad \hat{u}_t(x_i, t^n) = \frac{u_i^{n+1} - u_i^{n-1}}{2\tau}; \quad (8.29)$$

$$\begin{aligned} \hat{u}_{tt}(x_i, t^n) &= \frac{u(x_i, t^n + \tau) - 2u(x_i, t^n) + u(x_i, t^n - \tau)}{\tau^2}, \\ \hat{u}_{tt}(x_i, t^n) &= \frac{u_i^{n+1} - 2u_i^n + u_i^{n-1}}{\tau^2}. \end{aligned} \quad (8.30)$$

Формулы (8.27), (8.28) аппроксимируют производную u_t с первым порядком аппроксимации, а (8.29), (8.30) аппроксимируют соответствующие производные со вторым порядком аппроксимации. Все вышеприведенные формулы записаны на n -м временном слое. При необходимости (например, при конструировании неявных схем) вместо n -го используется $(n + 1)$ -й и $(n - 1)$ -й временные слои.

Теперь получим выражения для аппроксимации смешанной производной. Применим (8.27) к оператору, заданному (8.15):

$$\begin{aligned} \frac{\partial}{\partial t} \left(\frac{\partial u(x_i, t^n)}{\partial x} \right) &= \frac{\frac{u(x_i + h, t^n + \tau) - u(x_i, t^n + \tau)}{h} - \frac{u(x_i + h, t^n) - u(x_i, t^n)}{h}}{\tau} = \\ &= \frac{u(x_i + h, t^n + \tau) - u(x_i, t^n + \tau) - u(x_i + h, t^n) + u(x_i, t^n)}{h\tau} \end{aligned}$$

или

$$\hat{u}_{xt}(x_i, t^n) = \frac{u_{i+1}^{n+1} - u_i^{n+1} - u_{i+1}^n + u_i^n}{h\tau}. \quad (8.31)$$

Получим еще одну формулу. Для этого применим (8.29) к функции, заданной (8.19):

$$\begin{aligned} \frac{\partial}{\partial t} \left(\frac{\partial u(x_i, t^n)}{\partial x} \right) &= \frac{\frac{u(x_i + h, t^n + \tau) - u(x_i - h, t^n + \tau)}{2h} - \frac{u(x_i + h, t^n - \tau) - u(x_i - h, t^n - \tau)}{2\tau}}{2\tau} = \\ &= \frac{u(x_i + h, t^n + \tau) - u(x_i - h, t^n + \tau) - u(x_i + h, t^n - \tau) + u(x_i - h, t^n - \tau)}{4h\tau} \end{aligned}$$

или

$$\hat{u}_{xt}(x_i, t^n) = \frac{u_{i+1}^{n+1} - u_{i-1}^{n+1} - u_{i+1}^{n-1} + u_{i-1}^{n-1}}{4h\tau}. \quad (8.32)$$

Можно показать, что (8.31) аппроксимирует производную u_{xt} с первым порядком, а (8.32) — со вторым порядком аппроксимации [27].

Замечания

1. Поскольку при численных расчетах вместо точного решения $u(x_i, t^n)$ в узлах (x_i, t^n) известно не точное, а приближенное решение $\hat{u}(x_i, t^n)$, то в формулах (8.20)–(8.22), (8.26)–(8.32) следует ставить знак « \wedge » над u .

2. Здесь рассмотрены аппроксимационные формулы на равномерной сетке. С использованием (5.7), (5.10) можно построить аналогичные формулы при неравномерной сетке.

В заключение параграфа сформулируем основные этапы решения задач с линейными дифференциальными уравнениями в частных производных разностными методами:

- 1) выбрать сетку D_h , соответствующую расчетной области \bar{D} ;
- 2) заменить дифференциальные операторы в уравнении и дополнительных условиях разностными, а функции — их сеточными представлениями, т. е. построить разностную схему. Установить ее порядок аппроксимации;
- 3) исследовать сходимость разностной схемы, проверив устойчивость схемы;
- 4) задать величины шагов сетки по всем независимым переменным, обеспечивающие сходимость. Решить разностную задачу одним из известных методов. В результате находятся значения $\hat{u}(x_i, t^n)$ приближенного решения в узлах сетки.

8.3. РАЗНОСТНЫЕ МЕТОДЫ РЕШЕНИЯ УРАВНЕНИЙ ПЕРВОГО ПОРЯДКА

Рассмотрим проблему конструирования разностных схем для решения уравнения (8.4):

$$A(x, t)u_x + B(x, t)u_t + C(x, t)u = G(x, t),$$

где $u(x, t)$ — искомая функция; $A(x, t)$, $B(x, t)$, $C(x, t)$ — коэффициенты; $G(x, t)$ — правая часть; $B(x, t) \neq 0$. Это уравнение относится к эволюционным. Основная идея состоит в том, что после замены дифференциального уравнения его конечно-разностной аппроксимацией получаются формулы, явно или неявно выражающие значения решения для одного момента времени через значения решения в предыдущий момент времени. Таким образом, если известно решение в начальный момент времени, можно шаг за шагом найти решение для всех последующих моментов. Чтобы проиллюстрировать общий подход, рассмотрим типичную задачу для уравнения переноса.

Пример 8.2. Построить разностные схемы для задачи

$$\begin{aligned} u_t + u_x &= 0, \quad -\infty < x < \infty, \quad 0 < t < T; \\ u(x, 0) &= \psi(x), \quad -\infty < x < \infty, \end{aligned}$$

где $\psi(x)$ — заданная функция.

□Решается задача Коши, в которой $A(x, t) = 1$, $B(x, t) = 1$, $C(x, t) = 0$, $G(x, t) = 0$. Для анализа различных разностных схем, используемых для нахождения приближенного решения, найдем сначала аналитическое решение поставленной задачи. Оно может быть найдено различными методами, в частности *методом характеристик*, который опирается на следующий факт: начальное условие в некоторой точке x при $t = 0$ переносится в плоскости Oxt вдоль линии, называемой *характеристикой*. Рассмотрим один из вариантов этого метода [7], [14], [46].

1. Решаем систему двух обыкновенных дифференциальных уравнений (уравнений характеристик):

$$\begin{aligned} \frac{dx}{ds} &= A(x, t), \quad x(0) = \mu; \\ \frac{dt}{ds} &= B(x, t), \quad t(0) = 0, \end{aligned} \tag{8.33}$$

где s, μ — параметры, $0 < s < \infty$, $-\infty < \mu < \infty$.

Поскольку $A(x, t) = 1$, $B(x, t) = 1$, то система имеет вид

$$\begin{aligned} \frac{dx}{ds} &= 1, \quad x(0) = \mu; \\ \frac{dt}{ds} &= 1, \quad t(0) = 0. \end{aligned}$$

Отсюда $x(s) = s + C_1$, $t(s) = s + C_2$ и $x(0) = C_1 = \mu$, $t(0) = C_2 = 0$. Следовательно, характеристики описываются уравнениями

$$x(s) = s + \mu, \quad t = s.$$

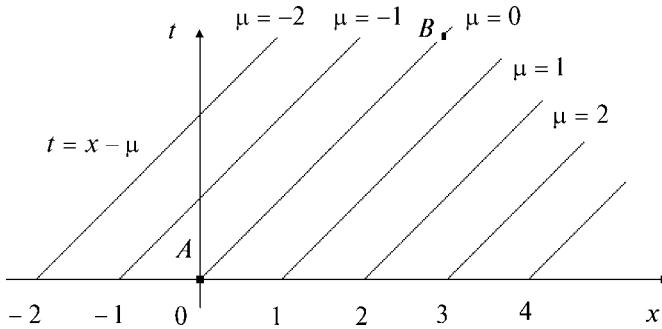


Рис. 8.4

Заметим, что в общем случае результатом п. 1 будут формулы связи переменных x , t и s , μ :

$$x = x(s, \mu), \quad t = t(s, \mu),$$

где параметр s изменяется вдоль характеристик, а параметр μ изменяется вдоль прямой линии $t = 0$.

Исключая s , получаем $x = t + \mu$, $-\infty < \mu < \infty$, т. е. каждому значению μ соответствует прямая линия (рис. 8.4).

2. Используя координаты s , μ и то, что с учетом (8.33) справедливо

$$\frac{du}{ds} = u_x \frac{dx}{ds} + u_t \frac{dt}{ds} = A(x, t)u_x + B(x, t)u_t,$$

сведем исходное уравнение в частных производных к обыкновенному дифференциальному уравнению

$$\frac{du}{ds} + C(x(s, \mu), t(s, \mu))u = G(x(s, \mu), t(s, \mu)), \quad u(0) = \psi(\mu). \quad (8.34)$$

Для рассматриваемого примера уравнение (8.34) имеет вид

$$\frac{du}{ds} = 0, \quad u(0) = \psi(\mu).$$

Отсюда $u(s, \mu) = C$, $u(0, \mu) = C = \psi(\mu)$, следовательно, $u(s, \mu) = \psi(\mu)$.

3. Найдем решение $u(x, t)$ исходной задачи, выражая s и μ через x и t (см. п. 1). Так как $s = t$, $\mu = x - t$, то $u(x, t) = \psi(x - t)$. На каждой характеристике значение решения постоянно, определяется значением параметра μ и равно $\psi(\mu)$. Например, значение решения в точке B то же, что и в точке A , и равно $\psi(0)$ (рис. 8.4).

Заметим, что если функция $\psi(x)$ имеет вид

$$\psi(x) = \begin{cases} \psi_0(x) \neq 0, & x < 0; \\ 0, & x \geq 0, \end{cases}$$

решение задачи в точках, расположенных на характеристике, проходящей через точки A и B , и правее этой характеристики, равно нулю, а в точках, расположенных левее этой характеристики, не равно нулю.

Перейдем к конструированию различных явных и неявных разностных схем. В качестве сетки D_h возьмем совокупность точек (x_i, t^n) , где $x_i = ih$, $t^n = n\tau$, $i = 0, \pm 1, \pm 2, \dots$; $n = 0, 1, 2, \dots, N$; $T = N\tau$. При построении разностных схем производную u_x можно аппроксимировать по формулам (8.20)–(8.22), а производную u_t — по формулам (8.27)–(8.29). Сформируем некоторые из них.

Для получения *первой схемы* применим (8.20), (8.27):

$$\begin{aligned} \frac{\hat{u}_i^{n+1} - \hat{u}_i^n}{\tau} + \frac{\hat{u}_{i+1}^n - \hat{u}_i^n}{h} &= 0, \quad i = 0, \pm 1, \pm 2, \dots, \quad n = 0, 1, 2, \dots, N-1; \\ \hat{u}_i^0 &= \psi(x_i), \quad i = 0, \pm 1, \pm 2, \dots \end{aligned}$$

Полученную схему здесь и далее можно представить в форме $\mathcal{L}_h \hat{u}_h = f_h$, где

$$\begin{aligned} \mathcal{L}_h &= \begin{cases} \frac{\hat{u}_i^{n+1} - \hat{u}_i^n}{\tau} + \frac{\hat{u}_{i+1}^n - \hat{u}_i^n}{h}, & i = 0, \pm 1, \pm 2, \dots, \quad n = 0, 1, 2, \dots, N-1; \\ \hat{u}_i^0, & i = 0, \pm 1, \pm 2, \dots; \end{cases} \\ f_h &= \begin{cases} 0, & i = 0, \pm 1, \pm 2, \dots; \quad n = 0, 1, 2, \dots, N-1; \\ \psi(x_i), & i = 0, \pm 1, \pm 2, \dots \end{cases} \end{aligned}$$

Решение на $(n+1)$ -м слое запишется в виде

$$\hat{u}_i^{n+1} = \hat{u}_i^n \left(1 + \frac{\tau}{h}\right) - \hat{u}_{i+1}^n \frac{\tau}{h}, \quad i = 0, \pm 1, \pm 2, \dots, \quad n = 0, 1, 2, \dots, N-1. \quad (8.35)$$

Узлы, участвующие в расчетах, образуют трехточечный шаблон $\Pi_3(x_i, t^n) = \{(x_i, t^n), (x_{i+1}, t^n), (x_i, t^{n+1})\}$, изображенный на рисунке 8.5а. Проанализируем эту схему, используя выводы, полученные при применении метода характеристик.

Если функция $\psi(x)$ имеет вид

$$\psi(x) = \begin{cases} \psi_0(x) \neq 0, & x < 0, \\ 0, & x \geq 0, \end{cases} \text{ то } \hat{u}_i^0 = 0, \quad i = 0, 1, 2, \dots,$$

тогда по формуле (8.35) при любом соотношении между τ и h получаем $\hat{u}_i^0 = 0$, $i = 0, 1, 2, \dots; n = 1, 2, \dots$, т. е. в первом квадранте решение уравнения тождественно равно нулю. Это противоречит точному решению (рис. 8.4), следовательно, сходимость решения разностной задачи к решению дифференциальной задачи отсутствует. Таким образом, схема (8.35) имеет первый порядок аппроксимации по τ и по h и является абсолютно неустойчивой [35], [42]. Следовательно, она не может быть использована на практике и приведена здесь из методических соображений.

Для получения *второй схемы* применим (8.21) и (8.27), т. е. в отличие от предыдущего случая производная u_x аппроксимируется левой разностью:

$$\begin{aligned} \frac{\hat{u}_i^{n+1} - \hat{u}_i^n}{\tau} + \frac{\hat{u}_i^n - \hat{u}_{i-1}^n}{h} &= 0, \quad i = 0, \pm 1, \pm 2, \dots, \quad n = 0, 1, 2, \dots, N-1; \\ \hat{u}_i^0 &= \psi(x_i), \quad i = 0, \pm 1, \pm 2, \dots \end{aligned}$$

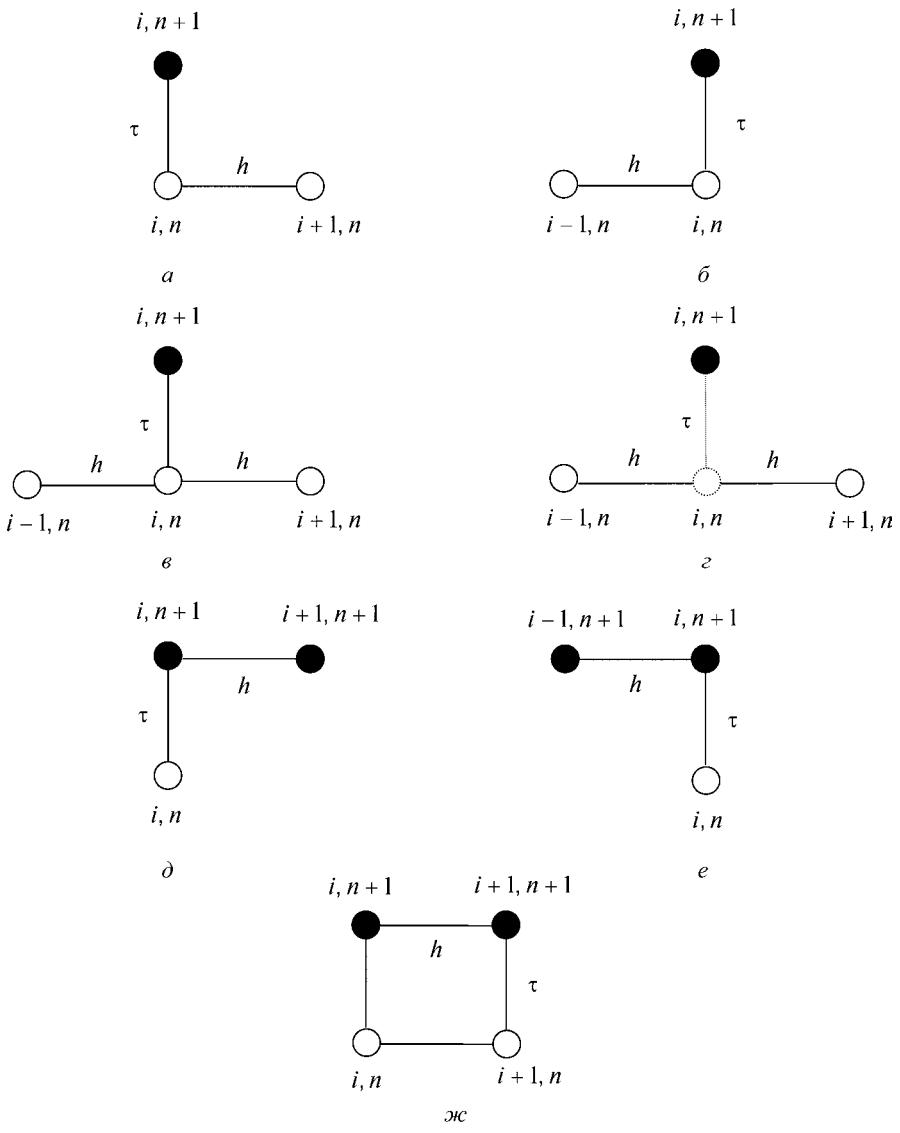


Рис. 8.5

Решение на $(n + 1)$ -м слое запишется в виде

$$\hat{u}_i^{n+1} = \hat{u}_i^n \left(1 - \frac{\tau}{h}\right) + \hat{u}_{i-1}^n \frac{\tau}{h}, \quad i = 0, \pm 1, \pm 2, \dots, \quad n = 0, 1, 2, \dots, N - 1. \quad (8.36)$$

Узлы, участвующие в расчетах, образуют трехточечный шаблон $\mathbb{W}_3(x_i, t^n) = \{(x_{i-1}, t^n), (x_i, t^n), (x_{i+1}, t^{n+1})\}$, изображенный на рисунке 8.5б. Схема имеет первый порядок аппроксимации по τ и по h и является условно устойчивой (является устойчивой при $\frac{\tau}{h} \leq 1$) [35], [42].

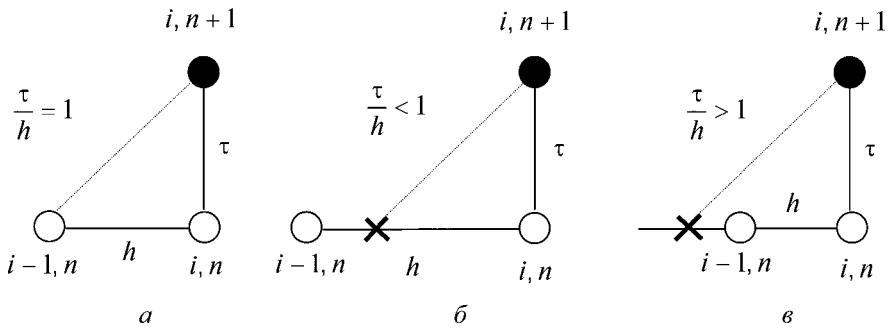


Рис. 8.6

Правая часть формулы (8.36) является формулой линейной интерполяции. Она определяет значение функции, отстоящей от точки (x_i, t^n) на расстояние, равное τ . Это значение переносится по характеристике в точку (x_i, t^{n+1}) . Поведение характеристик в данной задаче изображено на рисунке 8.6. Проанализируем более подробно свойства этой схемы для трех случаев соотношения шагов h и τ сетки D_h .

При $\frac{\tau}{h} = 1$ разностная схема в данной задаче дает точное решение, так как $\hat{u}_i^{n+1} = \hat{u}_{i-1}^n$, т. е. решение на следующем временном слое, передающееся вдоль характеристик (помечены на рисунке пунктирной линией), остается постоянным (рис. 8.6а).

При $\frac{\tau}{h} < 1$ и $\frac{\tau}{h} > 1$ по характеристике переносится значение, соответствующее точке, отмеченной крестиком (рис. 8.6б, в). При $\frac{\tau}{h} < 1$ разностная схема дает решение, соответствующее истинному с точностью, которой обладает линейная интерполяция. При $\frac{\tau}{h} > 1$ можно получить решение, существенно отличающееся от точного. Например, пусть начальное условие задано в виде

$$\psi(x) = \begin{cases} \psi_0(x) \neq 0, & x < 0; \\ 0, & x \geq 0, \end{cases}$$

тогда $\hat{u}_i^0 = 0$, $i = 0, 1, 2, \dots, n$, и, следовательно, $\hat{u}_i^1 = 0$ при $i = 1$. Точное решение в этом узле определяется значением $\psi_0(h - \mu)$ (см. п. 3 метода характеристик), поэтому погрешность может быть практически любой.

Для получения *третьей схемы* применим (8.22) и (8.27):

$$\begin{aligned} \frac{\hat{u}_i^{n+1} - \hat{u}_i^n}{\tau} + \frac{\hat{u}_{i+1}^n - \hat{u}_{i-1}^n}{2h} &= 0, \quad i = 0, \pm 1, \pm 2, \dots, n = 0, 1, 2, \dots, N-1; \\ \hat{u}_i^0 &= \psi(x_i), \quad i = 0, \pm 1, \pm 2, \dots \end{aligned} \tag{8.37}$$

Узлы, участвующие в расчетах, образуют четырехточечный шаблон $\mathbb{W}_4(x_i, t^n) = \{(x_{i-1}, t^n), (x_i, t^n), (x_{i+1}, t^n), (x_i, t^{n+1})\}$, изображенный на рисунке 8.5в. Схема является абсолютно неустойчивой [35] и имеет первый порядок аппроксимации по τ и второй — по h .

Для получения четвертой схемы (схемы Лакса) применим (8.22) и модификацию (8.27), где вместо u_i^n берется среднее $\frac{u_{i+1}^n + u_{i-1}^n}{2}$:

$$\begin{aligned} \frac{\hat{u}_i^{n+1} - \frac{\hat{u}_{i+1}^n + \hat{u}_{i-1}^n}{2}}{\tau} + \frac{\hat{u}_{i+1}^n - \hat{u}_{i-1}^n}{2h} &= 0, \quad i = 0, \pm 1, \pm 2, \dots, \quad n = 0, 1, 2, \dots, N-1; \\ \hat{u}_i^0 &= \psi(x_i), \quad i = 0, \pm 1, \pm 2, \dots \end{aligned} \quad (8.38)$$

Узлы, участвующие в расчетах, образуют трехточечный шаблон $\Pi_3(x_i, t^n) = \{(x_{i-1}, t^n), (x_{i+1}, t^n), (x_i, t^{n+1})\}$, изображенный на рисунке 8.5г. Схема является условно устойчивой (является устойчивой при $\frac{\tau}{h} \leq 1$) [35]. Заметим, что простейшая модификация схемы (8.37), а именно использование среднего значения, привела к тому, что разностная схема из абсолютно неустойчивой превратилась в условно устойчивую.

Формулы (8.35)–(8.38) явно выражают решение на $(n+1)$ -м временном слое через решение на n -м слое. Поэтому соответствующие им схемы называются *явными двухслойными*.

Для получения пятой схемы применим формулу (8.20), записанную на $(n+1)$ -м временном слое, и (8.27):

$$\begin{aligned} \frac{\hat{u}_i^{n+1} - \hat{u}_i^n}{\tau} + \frac{\hat{u}_{i+1}^{n+1} - \hat{u}_i^{n+1}}{h} &= 0, \quad i = 0, \pm 1, \pm 2, \dots, \quad n = 0, 1, 2, \dots, N-1; \\ \hat{u}_i^0 &= \psi(x_i), \quad i = 0, \pm 1, \pm 2, \dots \end{aligned} \quad (8.39)$$

Эта схема относится к неявным, использует трехточечный шаблон $\Pi_3(x_i, t^n) = \{(x_i, t^n), (x_i, t^{n+1}), (x_{i+1}, t^{n+1})\}$ (рис. 8.5д) и является абсолютно устойчивой [35]. Схема (8.39) имеет первый порядок аппроксимации по τ и по h .

Для получения шестой схемы применим (8.21), записанную на $(n+1)$ -м временном слое, и (8.27):

$$\begin{aligned} \frac{\hat{u}_i^{n+1} - \hat{u}_i^n}{\tau} + \frac{\hat{u}_i^{n+1} - \hat{u}_{i-1}^{n+1}}{h} &= 0, \quad i = 0, \pm 1, \pm 2, \dots, \quad n = 0, 1, 2, \dots, N-1; \\ \hat{u}_i^0 &= \psi(x_i), \quad i = 0, \pm 1, \pm 2, \dots \end{aligned} \quad (8.40)$$

Эта схема также относится к неявным, использует трехточечный шаблон $\Pi_3(x_i, t^n) = \{(x_{i-1}, t^{n+1}), (x_i, t^{n+1}), (x_i, t^n)\}$ (рис. 8.5е) и является абсолютно устойчивой [35]. Схема имеет первый порядок аппроксимации по τ и по h .

Для получения седьмой схемы будем использовать средние значения выражений, определяемых формулой (8.20), записанной на $(n+1)$ -м и n -м временных слоях, и формулой (8.27), записанной при $x = x_i$ и $x = x_{i+1}$:

$$\begin{aligned} \frac{\hat{u}_i^{n+1} - \hat{u}_i^n}{\tau} + \frac{\hat{u}_{i+1}^{n+1} - \hat{u}_{i+1}^n}{\tau} + \frac{\hat{u}_{i+1}^n - \hat{u}_i^n}{h} + \frac{\hat{u}_{i+1}^{n+1} - \hat{u}_i^{n+1}}{h} &= 0, \\ i &= 0, \pm 1, \pm 2, \dots, \quad n = 0, 1, 2, \dots, N-1; \\ \hat{u}_i^0 &= \psi(x_i), \quad i = 0, \pm 1, \pm 2, \dots \end{aligned}$$

Эта схема относится к неявным, использует четырехточечный шаблон $\Pi_3(x_i, t^n) = \{(x_i, t^n), (x_i, t^{n+1}), (x_{i+1}, t^n), (x_{i+1}, t^{n+1})\}$ (рис. 8.5ж) и является абсолютно устойчивой. Схема имеет второй порядок аппроксимации по t и по h [14]. Второй порядок аппроксимации обеспечивается путем применения оператора осреднения аппроксимационных операторов на $(n+1)$ -м и n -м временных слоях.

Пятая, шестая и седьмая схемы относятся к *неявным двухслойным*. ■

8.4. РАЗНОСТНЫЕ МЕТОДЫ РЕШЕНИЯ УРАВНЕНИЙ ВТОРОГО ПОРЯДКА

8.4.1. РАЗНОСТНЫЕ МЕТОДЫ РЕШЕНИЯ УРАВНЕНИЙ ПАРАБОЛИЧЕСКОГО ТИПА

Сначала рассмотрим проблему конструирования явных разностных схем. Основная идея состоит в том, что после замены дифференциального уравнения параболического типа его конечно-разностной аппроксимацией получаются формулы, явно выражающие значения решения для одного расчетного временного слоя через значения решения на предыдущем временном слое. Таким образом, если известно решение в начальный момент времени, можно шаг за шагом (послойно) найти решение для всех последующих моментов. Чтобы проиллюстрировать общий подход, рассмотрим типичную задачу теплопроводности.

Пример 8.3. Построить явную разностную схему для решения задачи теплопроводности в стержне, начальная температура которого равна нулю. Пусть температура левого конца ($x = 0$) фиксирована, а на правом конце ($x = L$) происходит теплообмен с окружающей средой, так что тепловой поток пропорционален разности температур конца стержня и среды (определяется функцией $g(t)$).

□Решается третья начально-краевая задача:

$$\begin{aligned} u_t &= u_{xx}, \quad 0 < x < L, \quad 0 < t < T; \\ u(x, 0) &= 0, \quad 0 \leq x \leq L \text{ (начальное условие);} \\ u(0, t) &= 1, \quad 0 < t < T \text{ (краевое условие на левой границе);} \\ u_x(L, t) &= -[u(L, t) - g(t)], \quad 0 < t \leq T \text{ (краевое условие на правой границе).} \end{aligned}$$

Сравнивая с общей постановкой задачи (см. п. 8.2), получаем $a^2 = 1$, $\psi(x) \equiv 0$, $\alpha_0 = 0$, $\beta_0 = 1$, $\varphi_1(t) = 1$, $\alpha_1 = 1$, $\beta_1 = 1$, $\varphi_2(t) = g(t)$; L , T — заданные числа.

Для аппроксимации производных в решаемом уравнении будем использовать (8.27), (8.26):

$$\frac{\hat{u}_i^{n+1} - \hat{u}_i^n}{\tau} = \frac{\hat{u}_{i+1}^n - 2\hat{u}_i^n + \hat{u}_{i-1}^n}{h^2}.$$

Отсюда

$$\hat{u}_i^{n+1} = \frac{\tau}{h^2} \hat{u}_{i-1}^n + \left(1 - 2 \frac{\tau}{h^2}\right) \hat{u}_i^n + \frac{\tau}{h^2} \hat{u}_{i+1}^n, \quad i = 1, \dots, I-1. \quad (8.41)$$

Эта формула при фиксированном значении i выражает решение на $(n+1)$ -м временном слое через решение на n -м слое. Ей соответствует четырехточечный шаблон, изображенный на рисунке 8.5в.

При записи в разностной форме начальное условие $u(x, 0) = 0$, $0 \leq x \leq L$, запишется в следующем виде:

$$\hat{u}_i^0 = 0, \quad i = 0, 1, \dots, I, \quad (8.42)$$

краевые условия на левом конце ($x = 0$):

$$\hat{u}_0^n = 1, \quad n = 1, \dots, N, \quad (8.43)$$

краевые условия на правом конце ($x = L$) определяются левой разностью по формуле (8.21):

$$\frac{\hat{u}_I^n - \hat{u}_{I-1}^n}{h} = -[\hat{u}_I^n - g(n\tau)].$$

Отсюда

$$\hat{u}_I^n = \frac{\hat{u}_{I-1}^n - hg(n\tau)}{1+h}, \quad n = 1, \dots, N. \quad (8.44)$$

Заметим, что соотношения (8.42), (8.43) точно аппроксимируют начальное и соответствующее краевое условие. Соотношения (8.41)–(8.44) образуют явную двухслойную разностную схему, имеющую первый порядок аппроксимации по τ и второй — по h .

Замечания

1. Данная схема условно устойчива (устойчива при выполнении условия $\frac{\tau}{h^2} \leq \frac{1}{2}$, т. е. при $\tau \leq \frac{h^2}{2}$). Тем самым накладываются серьезные ограничения на выбор шага по времени. Например, если шаг сетки по переменной x выбрать равным $h = 0,1$, то шаг сетки по времени не должен быть больше, чем $\tau = 0,5 \cdot (0,1)^2 = 0,005$ (это означает, что отрезок $[0; 1]$ будет пройден за 200 шагов). Наилучшим с точки зрения минимизации ошибок является соотношение $\tau = \frac{h^2}{6}$ [27].

2. Разностная схема для уравнения $u_t = a^2(x, t)u_{xx}$ строится аналогично и является устойчивой при $\frac{\tau}{h^2} \leq \frac{1}{2A}$, где $A = \max_{(x,t) \in D} a^2(x, t)$ [42].

Методика вычислений по явной схеме

1. Задать значения шагов сетки h, τ так, чтобы $Ih = L, N\tau = T$, где I, N — целые положительные числа, причем значение шага по времени выбрать из условия устойчивости $\tau \leq \frac{h^2}{2}$. Вычислить $\hat{u}_i^0 = 0$, $i = 0, 1, \dots, I$; $\hat{u}_0^n = 1$, $n = 1, \dots, N$. Положить $n = 0$.

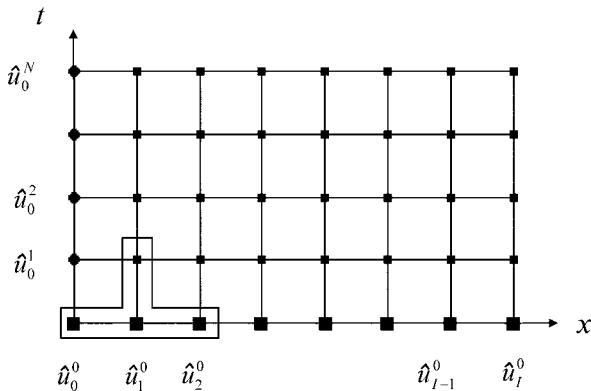


Рис. 8.7

2. Найти решение на $(n + 1)$ -м слое (рис. 8.7):

$$\begin{aligned}\hat{u}_i^{n+1} &= \frac{\tau}{h^2} \hat{u}_{i-1}^n + \left(1 - 2\frac{\tau}{h^2}\right) \hat{u}_i^n + \frac{\tau}{h^2} \hat{u}_{i+1}^n, \quad i = 1, \dots, I-1; \\ \hat{u}_I^n &= \frac{\hat{u}_{I-1}^n - hg(n\tau)}{1+h}.\end{aligned}$$

3. Если $n = N$, вычисления завершить. Иначе положить $n = n + 1$ и перейти к п. 2. ■

Пример 8.4. Пользуясь явной схемой, используемой в примере 8.3, найти приближенное решение первой начально-краевой задачи вида [10]:

$$\begin{aligned}u_t &= u_{xx}, \quad 0 < x < 1, \quad 0 < t < 0,01; \\ u(x, 0) &= 4x(1-x), \quad 0 \leq x \leq 1 \text{ (начальное условие);} \\ u(0, t) &= 0, \quad 0 < t \leq 0,01 \text{ (краевое условие на левой границе);} \\ u(1, t) &= 0, \quad 0 < t \leq 0,01 \text{ (краевое условие на правой границе).}\end{aligned}$$

□ Сравнивая с общей постановкой задачи (см. п. 8.2), получаем $a^2 = 1$, $\psi(x) = 4x(1-x)$, $\varphi_1(t) \equiv 0$, $\varphi_2(t) \equiv 0$, $L = 1$, $T = 0,01$.

1. Зададим шаг по пространству: $h = 0,1$, следовательно,

$$I = \frac{L}{0,1} = \frac{1}{0,1} = 10.$$

Выберем шаг по времени из условия обеспечения устойчивости и минимизации ошибок

$$\tau = \frac{h^2}{6} = \frac{0,01}{6} = \frac{1}{600},$$

поэтому

$$N = \frac{T}{\tau} = \frac{0,01}{\frac{1}{600}} = 6.$$

Таким образом, получаем $x_i = 0,1i$, $i = 0, 1, \dots, 10$; $t^n = \frac{n}{600}$, $n = 0, 1, \dots, 6$.

$n \backslash i$	0	1	2	3	4	5	6	7	8	9	10
0	0	0,360	0,640	0,840	0,960	1,000	0,960	0,840	0,640	0,360	0
1	0	0,347	0,627	0,827	0,947	0,987	0,947	0,827	0,627	0,347	0
2	0	0,336	0,613	0,813	0,933	0,973	0,933	0,813	0,613	0,336	0
3	0	0,326	0,600	0,800	0,920	0,960	0,920	0,800	0,600	0,326	0
4	0	0,317	0,588	0,787	0,907	0,947	0,907	0,787	0,588	0,317	0
5	0	0,309	0,576	0,774	0,894	0,934	0,894	0,774	0,576	0,309	0
6	0	0,302	0,564	0,761	0,881	0,921	0,881	0,761	0,564	0,302	0

Вычислим значения решения на нулевом слое:

$$\hat{u}_i^0 = 4x_i(1-x_i) = 4 \cdot 0,1 \cdot i(1 - 0,1 \cdot i), \quad i = 0, 1, \dots, 10 \quad (I = 10).$$

Краевые условия на левой и правой границах записываются в виде:

$$\hat{u}_0^n = 0, \quad n = 1, \dots, N = 6; \quad \hat{u}_I^n = 0, \quad n = 1, \dots, 6 \quad (N = 6).$$

2, 3. Поскольку $\frac{\tau}{h^2} = \frac{1/600}{0,01} = \frac{1}{6}$, решение на следующих слоях при $n = 0, 1, \dots, 5$ находится по формуле, вытекающей из (8.41):

$$\hat{u}_i^{n+1} = \frac{1}{6}\hat{u}_{i-1}^n + \frac{2}{3}\hat{u}_i^n + \frac{1}{6}\hat{u}_{i+1}^n, \quad i = 1, \dots, 9,$$

и занесено в таблицу 8.1.

Подчеркнем, что полученное решение на каждом из временных слоев является симметричным. ■

Теперь рассмотрим проблему конструирования неявных разностных схем. Для их получения все производные, входящие в дифференциальное уравнение, заменяются конечно-разностными аппроксимациями. Однако значения решения на $(n+1)$ -м слое уже не выражаются в явном виде через значения на предыдущих слоях. В неявных схемах для нахождения решения на следующем временном слое необходимо решать систему линейных алгебраических уравнений.

Для иллюстрации общего подхода решим следующую задачу теплопроводности.

Пример 8.5. Построить неявную разностную схему решения задачи теплопроводности в стержне, начальная температура которого равна 1, а температура левого и правого концов равна нулю.

□Данным физическим условиям соответствует первая начально-краевая задача:

$$u_t = u_{xx}, \quad 0 < x < L, \quad 0 < t < T; \\ u(x, 0) = 1, \quad 0 \leq x \leq L \quad (\text{начальное условие});$$

$$u(0, t) = 0, \quad 0 < t \leq T \text{ (краевое условие на левой границе);}$$

$$u(L, t) = 0, \quad 0 < t \leq T \text{ (краевое условие на правой границе).}$$

Сравнивая с общей постановкой задачи (см. п. 8.2), получаем $a^2 = 1$, $\psi(x) \equiv 1$, $\varphi_1(t) = \varphi_2(t) \equiv 0$; L, T — заданные числа.

Для аппроксимации производных в решаемом уравнении будем использовать формулу (8.27) и формулу (8.26), записанную на $(n+1)$ -м временном слое:

$$\frac{\hat{u}_i^{n+1} - \hat{u}_i^n}{\tau} = \frac{\hat{u}_{i+1}^{n+1} - 2\hat{u}_i^{n+1} + \hat{u}_{i-1}^{n+1}}{h^2}.$$

Отсюда

$$-\frac{\tau}{h^2} \hat{u}_{i-1}^{n+1} + \left(1 + 2\frac{\tau}{h^2}\right) \hat{u}_i^{n+1} - \frac{\tau}{h^2} \hat{u}_{i+1}^{n+1} = \hat{u}_i^n, \quad i = 1, \dots, I-1. \quad (8.45)$$

Эта формула выражает решение на $(n+1)$ -м временном слое через решение на n -м слое. Ей соответствует четырехточечный шаблон, изображенный на рисунке 8.8а.

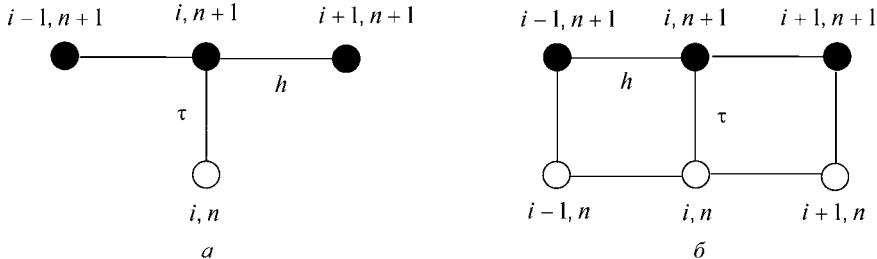


Рис. 8.8

При записи в разностной форме начальное условие $u(x, 0) = 1$, $0 \leq x \leq L$, запишется в следующем виде:

$$\hat{u}_i^0 = 1, \quad i = 0, 1, \dots, I, \quad (8.46)$$

краевые условия:

$$\begin{aligned} \hat{u}_0^n &= 0, \quad n = 1, \dots, N; \\ \hat{u}_I^n &= 0, \quad n = 1, \dots, N. \end{aligned} \quad (8.47)$$

Соотношения (8.45)–(8.47) образуют *неявную двухслойную* разностную схему, имеющую первый порядок аппроксимации по τ и второй — по h . Она является абсолютно устойчивой и сходящейся [35].

Методика вычислений по неявной схеме

1. Задать значения шагов сетки h, τ из условия обеспечения требуемой точности решения так, что $Ih = L, N\tau = T$, где I, N — целые положительные числа. Вычислить $\hat{u}_i^0 = 1$, $i = 0, 1, \dots, I$; $\hat{u}_0^n = 0$, $n = 1, \dots, N$; $\hat{u}_I^n = 0$, $n = 1, \dots, N$. Положить $n = 0$.

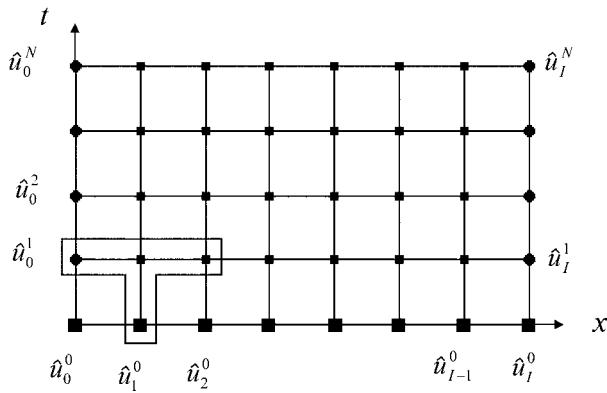


Рис. 8.9

2. Сформируем систему линейных алгебраических уравнений для некоторого временного слоя, записывая уравнение (8.45) при $i = 1, \dots, I - 1$ (рис. 8.9):

$$\begin{aligned} -\frac{\tau}{h^2} \hat{u}_0^{n+1} + \left(1 + 2 \frac{\tau}{h^2}\right) \hat{u}_1^{n+1} - \frac{\tau}{h^2} \hat{u}_2^{n+1} &= \hat{u}_1^n; \\ -\frac{\tau}{h^2} \hat{u}_1^{n+1} + \left(1 + 2 \frac{\tau}{h^2}\right) \hat{u}_2^{n+1} - \frac{\tau}{h^2} \hat{u}_3^{n+1} &= \hat{u}_2^n; \\ \dots \\ -\frac{\tau}{h^2} \hat{u}_{I-2}^{n+1} + \left(1 + 2 \frac{\tau}{h^2}\right) \hat{u}_{I-1}^{n+1} - \frac{\tau}{h^2} \hat{u}_I^{n+1} &= \hat{u}_{I-1}^n. \end{aligned}$$

Согласно п. 1, $\hat{u}_0^n = 1$, $n = 1, \dots, N$; $\hat{u}_I^n = 0$, $n = 1, \dots, N$. Поэтому для нахождения решения на следующем временном слое требуется решить трехдиагональную систему ($I - 1$) линейных алгебраических уравнений с ($I - 1$) неизвестными $\hat{u}_1^{n+1}, \dots, \hat{u}_{I-1}^{n+1}$:

$$\begin{aligned} \left(1 + 2 \frac{\tau}{h^2}\right) \hat{u}_1^{n+1} - \frac{\tau}{h^2} \hat{u}_2^{n+1} &= \hat{u}_1^n; \\ -\frac{\tau}{h^2} \hat{u}_1^{n+1} + \left(1 + 2 \frac{\tau}{h^2}\right) \hat{u}_2^{n+1} - \frac{\tau}{h^2} \hat{u}_3^{n+1} &= \hat{u}_2^n; \\ \dots \\ -\frac{\tau}{h^2} \hat{u}_{I-2}^{n+1} + \left(1 + 2 \frac{\tau}{h^2}\right) \hat{u}_{I-1}^{n+1} &= \hat{u}_{I-1}^n. \end{aligned}$$

Методы решения таких систем рассматривались в главе 1.

3. Если $n = N$, вычисления завершить. Иначе положить $n = n + 1$ и перейти к п. 2.

Методику вычислений по неявной схеме можно проследить на рисунке 8.9, где выделен используемый четырехточечный шаблон. Заметим, что в неявной схеме объем вычислений на каждом шаге больше, чем в явных, но хорошую точность можно получить при гораздо большем шаге по времени.

Получим еще одну *неявную схему*. Для аппроксимации производной u_t применим (8.27), а для аппроксимации u_{xx} взвешенное среднее выражения (8.26), записанного для узлов (x_i, t^n) , (x_i, t^{n+1}) на n -м и $(n+1)$ -м временных слоях (рис. 8.8б):

$$\hat{u}_{xx} = \frac{\hat{u}_{i+1}^n - 2\hat{u}_i^n + \hat{u}_{i-1}^n}{h^2}; \quad \hat{u}_{xx} = \frac{\hat{u}_{i+1}^{n+1} - 2\hat{u}_i^{n+1} + \hat{u}_{i-1}^{n+1}}{h^2},$$

т. е.

$$u_{xx} = \lambda \frac{\hat{u}_{i+1}^{n+1} - 2\hat{u}_i^{n+1} + \hat{u}_{i-1}^{n+1}}{h^2} + (1-\lambda) \frac{\hat{u}_{i+1}^n - 2\hat{u}_i^n + \hat{u}_{i-1}^n}{h^2},$$

где $\lambda \in [0; 1]$. При $\lambda = \frac{1}{2}$ получается обычное среднее, при $\lambda = \frac{3}{4}$ одна из разностных производных берется с весом $\frac{3}{4}$, а другая — с весом $\frac{1}{4}$. В результате получаем разностную схему:

$$\begin{aligned} \frac{\hat{u}_i^{n+1} - \hat{u}_i^n}{\tau} &= \lambda \frac{\hat{u}_{i+1}^{n+1} - 2\hat{u}_i^{n+1} + \hat{u}_{i-1}^{n+1}}{h^2} + (1-\lambda) \frac{\hat{u}_{i+1}^n - 2\hat{u}_i^n + \hat{u}_{i-1}^n}{h^2}, \quad i=1, \dots, I-1; \\ \hat{u}_i^0 &= 1, \quad i=0, 1, \dots, I; \\ \hat{u}_0^n &= 0, \quad n=1, \dots, N; \quad \hat{u}_I^n = 0, \quad n=1, \dots, N. \end{aligned}$$

Перенесем в первых $(I-1)$ уравнениях все неизвестные в левую часть. Тогда получаем

$$\begin{aligned} -\lambda \frac{\tau}{h^2} \hat{u}_{i-1}^{n+1} + \left(1 + 2\lambda \frac{\tau}{h^2}\right) \hat{u}_i^{n+1} - \lambda \frac{\tau}{h^2} \hat{u}_{i+1}^{n+1} &= \\ = (1-\lambda) \frac{\tau}{h^2} \hat{u}_{i-1}^n + \left(1 - 2(1-\lambda) \frac{\tau}{h^2}\right) \hat{u}_i^n + (1-\lambda) \frac{\tau}{h^2} \hat{u}_{i+1}^n, \quad i=1, \dots, I-1; \\ \hat{u}_i^0 &= 1, \quad i=0, 1, \dots, I; \\ \hat{u}_0^n &= 0, \quad n=1, \dots, N; \quad \hat{u}_I^n = 0, \quad n=1, \dots, N. \end{aligned} \tag{8.48}$$

Эта схема абсолютно устойчива и сходится [46]. Для нахождения решения на следующем слое, как и ранее, требуется решать трехдиагональную систему $(I-1)$ линейных алгебраических уравнений. При $\lambda = \frac{1}{2}$ схема называется *схемой Кранка — Николсон*. При $\lambda = 0$ из (8.48) следует явная схема (8.41), а при $\lambda = 1$ — неявная схема (8.45). Шеститочечный шаблон, соответствующий схеме (8.48), изображен на рисунке 8.8б. При $\lambda = \frac{1}{2}$ схема имеет второй порядок аппроксимации по τ и по h , а при $\lambda \neq \frac{1}{2}$ — первый порядок аппроксимации по τ и второй — по h [14].

Методику вычислений, совпадающую с уже описанной для схемы (8.45), продемонстрируем на примере.

1. Зададим $h = 0,2$; $\tau = 0,08$; $L = 1$; $T = 0,8$. Тогда $I \cdot 0,2 = 1$, $N \cdot 0,08 = 0,8$, т. е. $I = 5$, $N = 10$. Тогда начальные и краевые условия примут вид:

$$\begin{aligned}\hat{u}_i^0 &= 1, \quad i = 0, 1, \dots, 5; \\ \hat{u}_0^n &= 0, \quad n = 1, \dots, 10; \quad \hat{u}_5^n = 0, \quad n = 1, \dots, 10.\end{aligned}$$

Положим $n = 0$, $\lambda = \frac{1}{2}$. Вычислим $\frac{\tau}{h^2} = \frac{0,08}{0,04} = 2$.

2. Запишем систему (8.48), принимая $i = 1, 2, 3, 4$ ($I - 1 = 4$) с учетом п. 1:

$$\begin{aligned}-\hat{u}_0^1 + 3\hat{u}_1^1 - \hat{u}_2^1 &= \hat{u}_0^0 - \hat{u}_1^0 + \hat{u}_2^0; \\ -\hat{u}_1^1 + 3\hat{u}_2^1 - \hat{u}_3^1 &= \hat{u}_1^0 - \hat{u}_2^0 + \hat{u}_3^0; \\ -\hat{u}_2^1 + 3\hat{u}_3^1 - \hat{u}_4^1 &= \hat{u}_2^0 - \hat{u}_3^0 + \hat{u}_4^0; \\ -\hat{u}_3^1 + 3\hat{u}_4^1 - \hat{u}_5^1 &= \hat{u}_3^0 - \hat{u}_4^0 + \hat{u}_5^0; \\ \hat{u}_i^0 &= 1, \quad i = 0, 1, \dots, 5; \quad \hat{u}_0^1 = 0; \quad \hat{u}_5^1 = 0.\end{aligned}$$

Тогда получаем

$$\begin{aligned}3\hat{u}_1^1 - \hat{u}_2^1 &= 1; \\ -\hat{u}_1^1 + 3\hat{u}_2^1 - \hat{u}_3^1 &= 1; \\ -\hat{u}_2^1 + 3\hat{u}_3^1 - \hat{u}_4^1 &= 1; \\ -\hat{u}_3^1 + 3\hat{u}_4^1 &= 1\end{aligned}$$

или в матричной форме

$$\begin{pmatrix} 3 & -1 & 0 & 0 \\ -1 & 3 & -1 & 0 \\ 0 & -1 & 3 & -1 \\ 0 & 0 & -1 & 3 \end{pmatrix} \begin{pmatrix} \hat{u}_1^1 \\ \hat{u}_2^1 \\ \hat{u}_3^1 \\ \hat{u}_4^1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}.$$

В результате имеем $\hat{u}_1^1 = 0,6$; $\hat{u}_2^1 = 0,8$; $\hat{u}_3^1 = 0,8$; $\hat{u}_4^1 = 0,6$.

3. Далее, полагая $n = 1$, можно сделать следующий шаг по времени, решая новую систему алгебраических уравнений. Аналогичный процесс продолжается до достижения значения $N = 10$. ■

8.4.2. РАЗНОСТНЫЕ МЕТОДЫ РЕШЕНИЯ УРАВНЕНИЙ ГИПЕРБОЛИЧЕСКОГО ТИПА

Методы сеток для решения гиперболических уравнений имеют много общего с соответствующими методами для параболических уравнений. Однако они имеют и свои особенности, связанные с типом уравнения.

Сначала рассмотрим проблему конструирования явных разностных схем. Основная идея состоит в том, что после замены дифференциального уравнения гиперболического типа его конечно-разностной аппроксимацией получаются формулы, явно выражающие значения решения для одного расчетного временного слоя через значения решения на предыдущем временном слое. Таким образом, если известно решение в начальный момент времени, можно шаг за шагом (послойно) найти решение для всех последующих

моментов. Чтобы проиллюстрировать общий подход, рассмотрим типичную краевую задачу для волнового уравнения.

Пример 8.6. Построить явную и неявную разностные схемы для решения задачи, в которой имеется струна длиной L , натянутая между двумя точками оси Ox , точкой $x = 0$ и точкой $x = L$. Концы струны закреплены, начальное смещение струны описывается функцией $f(x)$, а начальная скорость — функцией $g(x)$.

□ Решается первая начально-краевая задача (см. п. 8.1):

$$u_{tt} = u_{xx}, \quad 0 < x < L, \quad 0 < t < T;$$

$u(x, 0) = f(x), \quad 0 < x < L$ (начальное условие);

$u_t(x, 0) = g(x), \quad 0 < x < L$ (начальное условие);

$u(0, t) = 0, \quad 0 \leq t \leq T$ (краевое условие на левой границе);

$u(L, t) = 0, \quad 0 \leq t \leq T$ (краевое условие на правой границе).

Сравнивая с общей постановкой задачи (см. п. 8.1), получаем $c^2 = 1$, $\psi_1(x) = f(x)$, $\psi_2(x) = g(x)$, $\varphi_1(t) = \varphi_2(t) \equiv 0$; L , T — заданные числа.

Для аппроксимации производных в решаемом уравнении будем использовать (8.30) и (8.26):

$$\frac{\hat{u}_i^{n+1} - 2\hat{u}_i^n + \hat{u}_i^{n-1}}{\tau^2} = \frac{\hat{u}_{i+1}^n - 2\hat{u}_i^n + \hat{u}_{i-1}^n}{h^2}.$$

Отсюда

$$\hat{u}_i^{n+1} = \frac{\tau^2}{h^2} \hat{u}_{i-1}^n + 2\left(1 - \frac{\tau^2}{h^2}\right) \hat{u}_i^n + \frac{\tau^2}{h^2} \hat{u}_{i+1}^n - \hat{u}_{i-1}^{n-1}, \quad i = 1, \dots, I-1. \quad (8.49)$$

Эта формула выражает решение на $(n + 1)$ -м временном слое через решение на n -м и $(n - 1)$ -м слоях. В отличие от схем для уравнения теплопроводности, в которых использовались только два временных слоя (n -й и $(n + 1)$ -й), здесь требуется использовать три слоя: $(n - 1)$ -й, n -й и $(n + 1)$ -й. Ей соответствует пятиточечный шаблон, изображенный на рисунке 8.10а, называемый «крест».

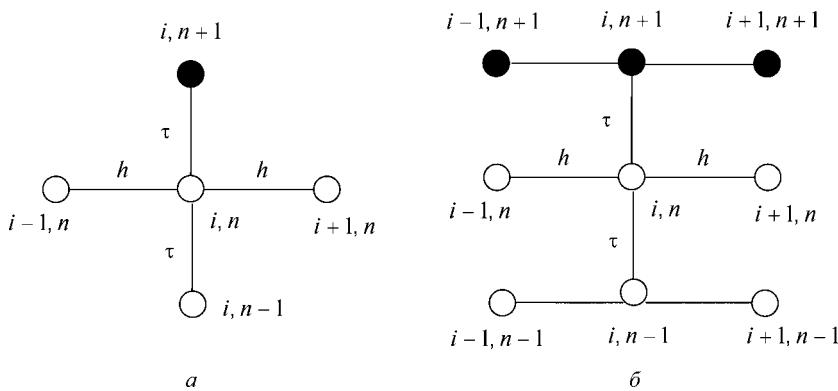


Рис. 8.10

При записи в разностной форме начальное условие $u(x, 0) = f(x)$, $0 < x < L$, запишется в следующем виде:

$$\hat{u}_i^0 = f(ih), \quad i = 1, \dots, I - 1, \quad (8.50)$$

начальное условие $u_i(x, 0) = g(x)$, $0 < x < L$, с применением (8.27) при $n = 0$ в виде $\frac{\hat{u}_i^1 - \hat{u}_i^0}{\tau} = g(ih)$, $i = 1, \dots, I - 1$, или, используя предыдущее соотношение,

$$\hat{u}_i^1 = f(ih) + \tau g(ih), \quad i = 1, \dots, I - 1, \quad (8.51)$$

краевые условия представляются в форме:

$$\hat{u}_0^n = 0, \quad \hat{u}_I^n = 0, \quad n = 0, \dots, N. \quad (8.52)$$

Соотношения (8.49)–(8.52) образуют явную трехслойную разностную схему. Эта схема имеет первый порядок аппроксимации по t и второй — по x , так как соотношение (8.51) аппроксимирует дифференциальное начальное условие с первым порядком.

Замечания

1. Данная схема условно устойчива (устойчива при выполнении условия $\frac{\tau}{h} \leq 1$) [27].

2. Разностная схема для уравнения $u_{tt} = c^2(x, t)u_{xx}$ строится аналогично и является устойчивой при $\frac{\tau^2}{h^2} < \frac{1}{C}$, где $C = \max_{(x, t) \in D} c^2(x, t)$ [42].

3. Простейшая замена (8.51) имеет первый порядок аппроксимации. Поскольку разностная схема (8.49) аппроксимирует дифференциальное уравнение со вторым порядком, желательно, чтобы разностное начальное условие также имело второй порядок аппроксимации. Для этого запишем аналог формулы Тейлора (8.17) в направлении t при $k = 2$:

$$u(x, t^0 + \tau) = u(x, t^0) + u_t(x, t^0)\tau + u_{tt}(x, t^0)\frac{\tau^2}{2} + u_{ttt}(x, \xi)\frac{\tau^3}{6}; \\ \xi \in (t^0, t^0 + \tau), \quad 0 < x < L.$$

Если функция $f(x)$ имеет ограниченную вторую производную, то в силу исходного дифференциального уравнения $u_{tt} = u_{xx}$ получаем $u_{tt}(x, t^0) = u_{xx}(x, t^0) = f''(x)$. Из предыдущего соотношения получаем

$$u_t(x, t^0) = \frac{u(x, t^0 + \tau) - u(x, t^0)}{\tau} + f''(x)\frac{\tau^2}{2} - u_{ttt}(x, \xi)\frac{\tau^2}{6}$$

или

$$u_t(x, t^0) \cong \frac{u(x, t^0 + \tau) - u(x, t^0)}{\tau} + f''(x)\frac{\tau^2}{2} - \left(\frac{\tau^2}{6}M_3\right),$$

где $M_3 = \max_{t \in [t^0, t^0 + \tau]} |u_{ttt}(x, t)|$. В результате соответствующее начальное условие в разностной форме имеет вид

$$\frac{\hat{u}_i^1 - \hat{u}_i^0}{\tau} - f''(ih) \frac{\tau}{2} = g(ih), \quad i=1, \dots, I-1.$$

Так как $\hat{u}_i^0 = f(ih)$, $i = 1, \dots, I-1$, то

$$\hat{u}_i^1 = f(ih) + \tau g(ih) + f''(ih) \frac{\tau^2}{2}, \quad i=1, \dots, I-1. \quad (8.53)$$

Если в разностной схеме используется (8.51), схема имеет первый порядок аппроксимации по τ и второй — по h , а если (8.53), то имеет второй порядок аппроксимации по τ и по h .

Методика вычислений по явной схеме

1. Задать значения шагов сетки h, τ так, что $Ih = L, N\tau = T$, где I, N — целые положительные числа, причем значение шага по времени выбрать из условия устойчивости $\tau \leq h$. Вычислить

$$\begin{aligned} \hat{u}_i^0 &= f(ih), \quad i=1, \dots, I-1; \quad \hat{u}_i^1 = f(ih) + \tau g(ih), \quad i=1, \dots, I-1; \\ \hat{u}_0^n &= 0; \quad \hat{u}_I^n = 0, \quad n=0, \dots, N. \end{aligned}$$

Приведенные выражения дают решение задачи для первых двух слоев сетки (рис. 8.8). Положить $n = 1$.

2. Найти решение на $(n+1)$ -м слое (рис. 8.11):

$$\hat{u}_i^{n+1} = \frac{\tau^2}{h^2} \hat{u}_{i-1}^n + 2 \left(1 - \frac{\tau^2}{h^2}\right) \hat{u}_i^n + \frac{\tau^2}{h^2} \hat{u}_{i+1}^n - \hat{u}_i^{n-1}, \quad i=1, \dots, I-1.$$

3. Если $n = N$, вычисления завершить. Иначе положить $n = n + 1$ и перейти к п. 2.

Теперь рассмотрим проблему конструирования *неявных разностных схем*. Для их получения все производные, входящие в дифференциальное уравнение, заменяются конечно-разностными аппроксимациями с использованием $(n+1)$ -го слоя для аппроксимации u_{xx} . Однако значения решения

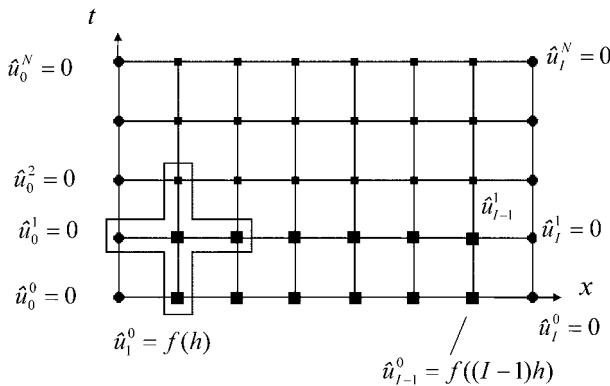


Рис. 8.11

на $(n + 1)$ -м слое уже не выражаются в явном виде через значения на предыдущих слоях. В связи с этим в неявных схемах для нахождения решения на следующем временном слое необходимо решать систему линейных алгебраических уравнений.

Для иллюстрации общего подхода получим неявную схему для волнового уравнения. При аппроксимации производных в уравнении будем использовать (8.30) и формулу (8.26), записанную на $(n - 1)$ -м, n -м и $(n + 1)$ -м слоях:

$$\begin{aligned} \frac{\hat{u}_i^{n+1} - 2\hat{u}_i^n + \hat{u}_i^{n-1}}{\tau^2} &= \lambda \frac{\hat{u}_{i+1}^{n+1} - 2\hat{u}_i^{n+1} + \hat{u}_{i-1}^{n+1}}{h^2} + (1 - 2\lambda) \frac{\hat{u}_{i+1}^n - 2\hat{u}_i^n + \hat{u}_{i-1}^n}{h^2} + \\ &+ \lambda \frac{\hat{u}_{i+1}^{n-1} - 2\hat{u}_i^{n-1} + \hat{u}_{i-1}^{n-1}}{h^2}, \quad i = 1, \dots, I - 1, \end{aligned} \quad (8.54)$$

где λ — параметр, $0 \leq \lambda \leq \frac{1}{2}$. Значения приближенного решения на нулевом и первом слоях вычисляются по формулам (8.50)–(8.52), как и в явной схеме. На остальных слоях схема представляет собой линейную систему алгебраических уравнений с трехдиагональной матрицей, в которой диагональные элементы преобладают. Решение этой системы существует, единственno и вычисляется методом прогонки (см. главу 1).

Формула (8.54) выражает три неизвестных значения решения на $(n + 1)$ -м временном слое через решения на n -м и $(n - 1)$ -м слоях. В отличие от схем для уравнения теплопроводности, в которых использовались только два временных слоя (n -й и $(n + 1)$ -й), здесь требуется использовать три слоя: $(n - 1)$ -й, n -й и $(n + 1)$ -й. Поэтому такая схема называется *неявной трехслойной*. Ей соответствует девятиточечный шаблон, изображенный на рисунке 8.10б. При $\lambda = 0$ схема (8.50)–(8.52), (8.54) переходит в явную схему «крест».

При $\frac{1}{4} \leq \lambda \leq \frac{1}{2}$ схема абсолютно устойчива, а при $0 \leq \lambda < \frac{1}{4}$ схема условно устойчива (устойчива при выполнении условия $\frac{\tau}{h} \leq (1 - 4\lambda)^{\frac{1}{2}}$). При $\frac{1}{4} \leq \lambda \leq \frac{1}{2}$ схема имеет второй порядок аппроксимации по τ и по h [14]. ■

Пример 8.7. Пользуясь явной схемой, найти приближенное решение первой начально-краевой задачи [10]:

$$u_{tt} = u_{xx}, \quad 0 < x < \pi, \quad 0 < t < \frac{5\pi}{18};$$

$$u(x, 0) = x(\pi - x), \quad 0 < x < \pi \text{ (начальное условие);}$$

$$u_t(x, 0) = 0, \quad 0 < x < \pi \text{ (начальное условие);}$$

$$u(0, t) = 0, \quad 0 \leq t \leq \frac{5\pi}{18} \text{ (краевое условие на левой границе);}$$

$$u(\pi, t) = 0, \quad 0 \leq t \leq \frac{5\pi}{18} \text{ (краевое условие на правой границе).}$$

□Сравнивая с общей постановкой задачи (см. п. 8.1), получаем $c^2 = 1$, $\psi_1(x) = x(\pi - x)$, $\psi_2(x) \equiv 0$, $\varphi_1(t) = \varphi_2(t) \equiv 0$; $L = \pi$, $T = \frac{5\pi}{18}$.

1. Зададим шаг по пространству, равным $h = \frac{\pi}{18}$, шаг по времени выберем из условия устойчивости: $\tau \leq h$. Положим $\tau = h = \frac{\pi}{18}$. Тогда

$$x_i = \frac{\pi i}{18}, \quad t^n = \frac{\pi n}{18}, \quad i = 0, 1, \dots, I = \frac{L}{h} = 18, \quad n = 0, 1, \dots, N = \frac{T}{\tau} = 5.$$

По формулам (8.50), (8.52), (8.53), учитывая, что $f(x) = x(\pi - x)$, $f''(x) = -2$, $g(x) \equiv 0$, получаем значения приближенного решения на нулевом и первом слоях:

$$\begin{aligned}\hat{u}_i^0 &= f(ih) = x_i(\pi - x_i) = \frac{\pi i}{18} \left(\pi - \frac{\pi i}{18} \right), \quad i = 1, \dots, 17; \\ \hat{u}_0^n &= 0; \quad \hat{u}_I^n = 0, \quad n = 0, \dots, 5; \\ \hat{u}_i^1 &= f(ih) + \tau g(ih) + f''(ih) \frac{\tau^2}{2} = x_i(\pi - x_i) + 0 - 2 \frac{\tau^2}{2} = \\ &= \frac{\pi i}{18} \left(\pi - \frac{\pi i}{18} \right) - \left(\frac{\pi}{18} \right)^2, \quad i = 1, \dots, 17.\end{aligned}$$

2, 3. Поскольку $\frac{\tau^2}{h^2} = 1$, решение на последующих слоях вычисляется по формуле, следующей из (8.49):

$$\hat{u}_i^{n+1} = \hat{u}_{i-1}^n + \hat{u}_{i+1}^n - \hat{u}_i^{n-1}, \quad i = 1, \dots, 17.$$

Таблица 8.2

$n \backslash i$	0	1	2	3	4	5	6	7	8	9
0	0	0,518	0,975	1,371	1,706	1,980	2,193	2,346	2,437	2,467
1	0	0,487	0,944	1,340	1,675	1,950	2,163	2,315	2,406	2,437
2	0	0,426	0,853	1,249	1,584	1,858	2,071	2,224	2,315	2,346
3	0	0,366	0,731	1,097	1,432	1,706	1,919	2,071	2,163	2,193
4	0	0,305	0,609	0,914	1,218	1,493	1,706	1,858	1,950	1,980
5	0	0,244	0,487	0,731	0,975	1,218	1,432	1,584	1,675	1,706

В таблице 8.2 приведены результаты расчетов при $i = 0, 1, \dots, 9$, так как решение $u(x, t)$ симметрично относительно $x = \frac{\pi}{2}$ ($i = 9$) при фиксированном t . ■

Замечания. В практике применения численных методов существует полезный прием, позволяющий на основе результатов расчетов судить о том, с какой точностью они получены. С этой целью используется правило Рунге, изложенное выше применительно к уточнению результатов численного дифференцирования и интегрирования (см. п. 5.4). Аналогичный подход можно использовать для оценки точности решения задач математической физики [26]. Опишем идею его реализации:

а) находится численное решение с шагом h по пространственной переменной x и шагом τ по времени t ;

- б) расчет повторяется с шагами $\frac{h}{2}$ и $\frac{\tau}{2}$ с тем, чтобы образовывались слои, некоторые из которых совпадали бы со слоями, полученными ранее;
- в) сравниваются оба численных решения в одинаковых узлах сетки (им соответствуют одни и те же значения пространственной координаты и времени). Если отклонение удовлетворяет заданным требованиям, то процесс завершается и в качестве приближенного решения задачи принимается полученное в п. б. Иначе шаги снова делятся пополам и процесс продолжается.

8.4.3. РАЗНОСТНЫЕ МЕТОДЫ РЕШЕНИЯ УРАВНЕНИЙ ЭЛЛИПТИЧЕСКОГО ТИПА

Рассмотрим проблему решения краевых задач для уравнений эллиптического типа. Эти задачи являются стационарными, так как в них отсутствует временная переменная. В них требуется найти решение уравнения с частными производными в данной области пространства, если на границе области решение или его производная заданы. Разностные схемы получаются путем замены производных их конечно-разностными аппроксимациями. В результате приближенное решение эллиптических задач сводится к решению системы линейных алгебраических уравнений для значений искомой функции во внутренних узлах сетки.

Чтобы проиллюстрировать общий подход, рассмотрим задачу Дирихле для уравнения Лапласа.

Пример 8.8. Построить разностную схему для решения задачи Дирихле:

$$\begin{aligned} u_{xx} + u_{yy} &= 0, \quad 0 < x < L, \quad 0 < y < M; \\ u(0, y) &= \phi(0, y), \quad 0 < y < M; \\ u(L, y) &= \phi(L, y), \quad 0 < y < M; \\ u(x, M) &= \phi(x, M), \quad 0 \leq x \leq L; \\ u(x, 0) &= \phi(x, 0), \quad 0 \leq x \leq L, \end{aligned}$$

где $\phi(x, y)$ — заданная функция; L, M — заданные числа.

В данном случае задача решается в области Ω прямоугольной формы, краевые условия заданы на сторонах прямоугольника. В плоскости Oxy построим сетку $D_h = \{(x_i, y_j)\}$, где $x_i = ih_x$, $y_j = jh_y$, $i = 0, 1, \dots, I$; $j = 0, 1, \dots, J$; h_x , h_y — величины шагов по x и y ; $Ih_x = L$, $Jh_y = M$, I и J — целые положительные числа.

Заменим дифференциальные операторы в уравнении Лапласа по формулам, следующим из (8.25) при замене переменной t^n на переменную y_j :

$$\begin{aligned} \hat{u}_{xx}(x_i, y_j) &= \frac{u(x_i + h_x, y_j) - 2u(x_i, y_j) + u(x_i - h_x, y_j)}{h_x^2}; \\ \hat{u}_{yy}(x_i, y_j) &= \frac{u(x_i, y_j + h_y) - 2u(x_i, y_j) + u(x_i, y_j - h_y)}{h_y^2}. \end{aligned}$$

Применяя краткую запись $u(x_i, y_j) = u_{i,j}$, получаем

$$\frac{\hat{u}_{i+1,j} - 2\hat{u}_{i,j} + \hat{u}_{i-1,j}}{h_x^2} + \frac{\hat{u}_{i,j+1} - 2\hat{u}_{i,j} + \hat{u}_{i,j-1}}{h_y^2} = 0, \quad i=1, \dots, I-1, \quad j=1, \dots, J-1$$

или

$$\frac{h_y^2}{h_x^2} \hat{u}_{i+1,j} + \frac{h_y^2}{h_x^2} \hat{u}_{i-1,j} + \hat{u}_{i,j+1} + \hat{u}_{i,j-1} - 2 \left(1 + \frac{h_y^2}{h_x^2}\right) \hat{u}_{i,j} = 0, \quad i=1, \dots, I-1, \quad j=1, \dots, J-1. \quad (8.55)$$

Полагая $\phi_{i,j} = \phi(ih_x, jh_y)$, краевые условия можно записать в виде:

$$\begin{aligned} \hat{u}_{0,j} &= \phi_{0,j}, \quad j=1, \dots, J-1; \\ \hat{u}_{J,j} &= \phi_{I,j}, \quad j=1, \dots, J-1; \\ \hat{u}_{i,J} &= \phi_{i,J}, \quad i=0, 1, \dots, I; \\ \hat{u}_{i,0} &= \phi_{i,0}, \quad i=0, 1, \dots, I. \end{aligned} \quad (8.56)$$

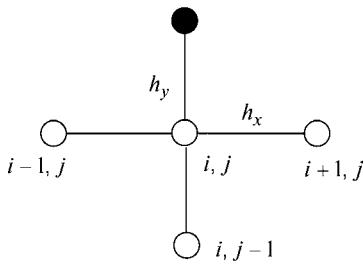


Рис. 8.12

Для нахождения искомого решения требуется решить систему $(I-1) \times (J-1)$ линейных алгебраических уравнений. Схема (8.55), (8.56) имеет второй порядок аппроксимации по τ и по h , является абсолютно устойчивой и сходится [14]. Ей соответствует пятиточечный шаблон, изображенный на рисунке 8.12.

При $h_x = h_y$, т. е. одинаковых шагах в горизонтальном и вертикальном направлениях, получаем

$$\hat{u}_{i+1,j} + \hat{u}_{i-1,j} + \hat{u}_{i,j+1} + \hat{u}_{i,j-1} - 4\hat{u}_{i,j} = 0, \quad i=1, \dots, I-1, \quad j=1, \dots, J-1. \quad (8.57)$$

Разрешив относительно $\hat{u}_{i,j}$, имеем

$$\hat{u}_{i,j} = \frac{1}{4}(\hat{u}_{i+1,j} + \hat{u}_{i-1,j} + \hat{u}_{i,j+1} + \hat{u}_{i,j-1}), \quad i=1, \dots, I-1, \quad j=1, \dots, J-1. \quad (8.58)$$

Это означает, что решение $\hat{u}_{i,j}$ аппроксимируется средним значением по четырем соседним узлам. Как правило, для решения системы (8.58) применяются различные итерационные методы, например, метод простой итерации или метод Зейделя (см. главу 1).

Рассмотрим частный случай поставленной задачи при

$$L=1, \quad M=1, \quad I=J=3, \quad \phi(x, y) = \begin{cases} \sin \pi x, & y=0, \quad 0 \leq x \leq 1, \\ 0, & \text{на остальных границах} \\ & \text{прямоугольной области } \Omega. \end{cases}$$

Так как $h_x = h_y = \frac{1}{3}$, то запишем систему (8.57) для четырех внутренних точек ($i=1, 2; j=1, 2$):

$$\begin{aligned}\hat{u}_{2,1} + \hat{u}_{0,1} + \hat{u}_{1,2} + \hat{u}_{1,0} - 4\hat{u}_{1,1} &= 0 \quad (i=1, j=1); \\ \hat{u}_{2,2} + \hat{u}_{0,2} + \hat{u}_{1,3} + \hat{u}_{1,1} - 4\hat{u}_{1,2} &= 0 \quad (i=1, j=2); \\ \hat{u}_{3,1} + \hat{u}_{1,1} + \hat{u}_{2,2} + \hat{u}_{2,0} - 4\hat{u}_{2,1} &= 0 \quad (i=2, j=1); \\ \hat{u}_{3,2} + \hat{u}_{1,2} + \hat{u}_{2,3} + \hat{u}_{2,1} - 4\hat{u}_{2,2} &= 0 \quad (i=2, j=2).\end{aligned}$$

Краевые условия (8.56) запишутся в форме

$$\begin{aligned}\hat{u}_{0,j} &= 0, \quad j=1, 2; \quad \hat{u}_{3,j} = 0, \quad j=1, 2; \\ \hat{u}_{i,3} &= 0, \quad i=0, 1, 2, 3; \quad \hat{u}_{i,0} = \sin(\pi i h_x) = \sin \frac{\pi i}{3}, \quad i=0, 1, 2, 3.\end{aligned}$$

С учетом краевого условия система может быть записана в следующей матричной форме

$$\begin{pmatrix} -4 & 1 & 1 & 0 \\ 1 & -4 & 0 & 1 \\ 1 & 0 & -4 & 1 \\ 0 & 1 & 1 & -4 \end{pmatrix} \begin{pmatrix} \hat{u}_{1,1} \\ \hat{u}_{1,2} \\ \hat{u}_{2,1} \\ \hat{u}_{2,2} \end{pmatrix} = \begin{pmatrix} -\sin \frac{\pi}{3} \\ 0 \\ -\sin \frac{2\pi}{3} \\ 0 \end{pmatrix}.$$

В результате имеем

$$\hat{u}_{1,1} = 0,325; \quad \hat{u}_{1,2} = 0,108; \quad \hat{u}_{2,1} = 0,325; \quad \hat{u}_{2,2} = 0,108.$$

Методика вычислений

1. Задать сетку $D_h = \{(x_i, y_j)\}$, где $x_i = ih_x$, $y_j = jh_y$, $i = 0, 1, \dots, I$; $j = 0, 1, \dots, J$; h_x, h_y — величины шагов по x и y ; $Ih_x = L$, $Jh_y = M$; I и J — целые положительные числа; $\varepsilon > 0$ — желаемую точность.

Вычислить

$$\begin{aligned}\hat{u}_{0,j} &= \varphi_{0,j}, \quad j=1, \dots, J-1; \quad \hat{u}_{I,j} = \varphi_{I,j}, \quad j=1, \dots, J-1; \\ \hat{u}_{i,J} &= \varphi_{i,J}, \quad i=0, 1, \dots, I; \quad \hat{u}_{i,0} = \varphi_{i,0}, \quad i=0, 1, \dots, I.\end{aligned}$$

Положить $k = 0$; задать начальное приближение искомого решения во внутренних узлах: $\hat{u}_{i,j}^{(0)}$, $i = 1, \dots, I-1$; $j = 1, \dots, J-1$.

2. Вычислить $\hat{u}_{i,j}^{(k+1)}$ одним из двух методов (рис. 8.13):

а) методом простых итераций:

$$\hat{u}_{i,j}^{(k+1)} = \frac{1}{4} \left(\hat{u}_{i+1,j}^{(k)} + \hat{u}_{i-1,j}^{(k)} + \hat{u}_{i,j+1}^{(k)} + \hat{u}_{i,j-1}^{(k)} \right), \quad i=1, \dots, I-1; \quad j=1, \dots, J-1;$$

б) методом Зейделя:

Шаг 1. Положить $j = 1$.

Шаг 2. Вычислить

$$\hat{u}_{i,j}^{(k+1)} = \frac{1}{4} \left(\hat{u}_{i+1,j}^{(k)} + \hat{u}_{i-1,j}^{(k+1)} + \hat{u}_{i,j+1}^{(k)} + \hat{u}_{i,j-1}^{(k+1)} \right), \quad i=1, \dots, I-1.$$

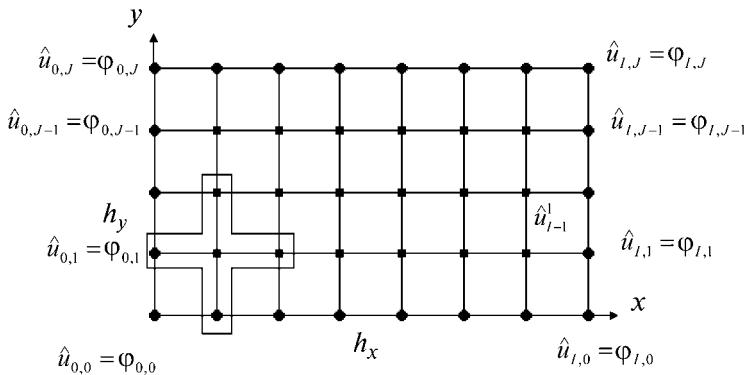


Рис. 8.13

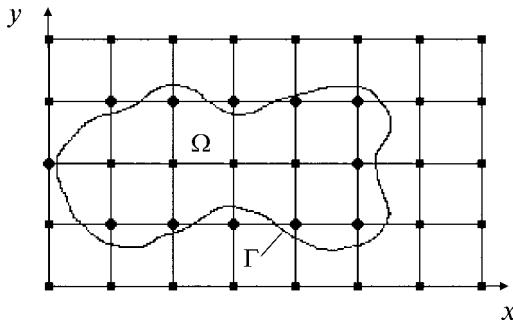


Рис. 8.14

Шаг 3. Если $j = J - 1$, процесс завершить. Иначе положить $j = j + 1$ и перейти к шагу 2.

3. Если выполняется условие окончания $\max_{i,j} |u_{i,j}^{(k+1)} - \hat{u}_{i,j}^{(k)}| \leq \varepsilon$, процесс завершить. Иначе положить $k = k + 1$ и перейти к п. 2.

Замечания

1. Метод Зейделя обеспечивает движение по узлам слева направо, начиная с первого слоя сетки по переменной x . Использование значений с $(k + 1)$ -й итерации, как правило, улучшает сходимость.

2. Вместо метода Зейделя можно использовать более эффективные релаксационные методы:

$$\hat{u}_{i,j}^{(k+1)} = \omega \hat{u}_{i,j}^{(k)} + \frac{1-\omega}{4} (\hat{u}_{i+1,j}^{(k)} + \hat{u}_{i-1,j}^{(k+1)} + \hat{u}_{i,j+1}^{(k)} + \hat{u}_{i,j-1}^{(k+1)}),$$

где ω — параметр релаксации. Член $\omega \hat{u}_{i,j}^{(k)}$ характеризует «память» алгоритма о значении решения на предыдущей итерации. При $\omega > 1$ итерационный метод называется методом *верхней релаксации*, при $0 < \omega < 1$ — методом *нижней релаксации*, при $\omega = 0$ совпадает с методом Зейделя. Существует оптимальное значение ω^* параметра релаксации, зависящее от шагов сетки и способа получения $\hat{u}_{i,j}^{(k+1)}$, при котором достигается максимальная скорость сходимости метода.

3. При решении задачи Неймана или третьей краевой задачи производные, входящие в краевые условия, следует также заменить их конечно-разностными аппроксимациями.

4. Если область Ω , в которой решается задача, имеет неправильную форму, можно покрыть ее сеткой (рис. 8.14) и найти решение в ближайших к границе точках, а затем рассчитать значения искомой функции на границе, например, с помощью линейной интерполяции [10].

В данной главе рассмотрены принципы конструирования разностных схем для простейших линейных начальных, краевых и начально-краевых задач. Для нелинейных задач, которые решаются при расчетно-теоретической проработке современных технических систем, применяют те же разностные схемы, что и для линейных задач.

8.5. МЕТОД ПРЯМЫХ

Основная идея метода состоит в сведении проблемы решения уравнений в частных производных к решению системы обыкновенных дифференциальных уравнений. Он используется в основном для решения уравнений эллиптического и параболического типов. В методе прямых, в отличие от метода сеток, не все дифференциальные операторы, входящие в исходную дифференциальную задачу, заменяются разностными. Обычно для аппроксимации выбирается оператор, соответствующий пространственной переменной в задаче теплопроводности или соответствующий неравноправному направлению в задачах с эллиптическими уравнениями. Тем самым дифференциальная задача заменяется дифференциально-разностной с меньшим числом непрерывных переменных (понижается размерность дифференциальной задачи). В связи с этим данный метод можно рассматривать как предельный случай метода сеток, когда шаги сетки по определенным независимым переменным стремятся к нулю.

Основную идею конструирования метода прямых рассмотрим на примере нахождения решения $u(x, y)$ уравнения:

$$A(x, y) \frac{\partial^2 u(x, y)}{\partial x^2} + 2B(x, y) \frac{\partial^2 u(x, y)}{\partial y \partial x} + C(x, y) \frac{\partial^2 u(x, y)}{\partial y^2} + D(x, y) \frac{\partial u(x, y)}{\partial x} + E(x, y) \frac{\partial u(x, y)}{\partial y} + F(x, y)u(x, y) = G(x, y), \quad (x, y) \in \Omega = (0, L) \times (0, M), \quad (8.59)$$

с краевыми условиями

$$\begin{aligned} u(x, 0) &= \varphi(x), \quad u(x, M) = \psi(x), \\ u(0, y) &= \Phi(y), \quad u(L, y) = \Psi(y) \end{aligned} \quad (8.60)$$

на границах области прямоугольной формы (см. рис. 8.15), где $\varphi(x)$, $\psi(y)$, $\Phi(y)$, $\Psi(y)$ — заданные функции. Пусть выполнено условие $B^2 - AC < 0$ эллиптичности уравнения для всех $(x, y) \in \Omega$. Отсюда следует, что $A(x, y) \neq 0$, $C(x, y) \neq 0$.

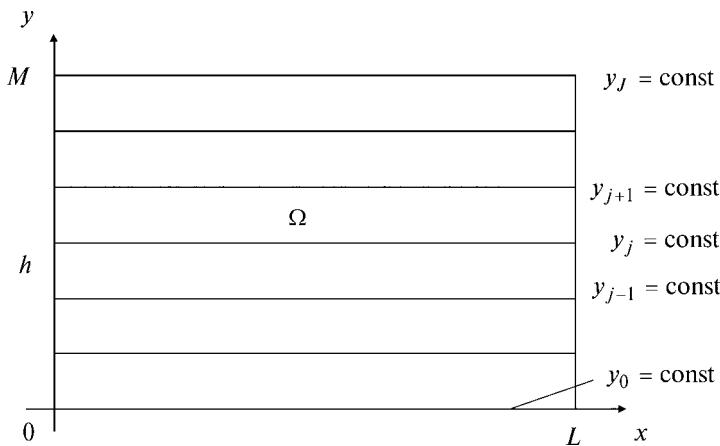


Рис. 8.15

Для решения задачи выберем на отрезке $[0, M]$ точки $y_j = jh$, $0 \leq j \leq J$, $Jh = M$, и проведем через них прямые, параллельные оси OX . На каждой прямой дифференциальное уравнение в частных производных (8.59) приближенно заменим обыкновенным дифференциальным уравнением для функций $u(x, y_j)$. Для этого аппроксимируем частные производные по y (см. п. 8.2):

$$\begin{aligned} \frac{\partial \hat{u}(x, y)}{\partial y} \Big|_{y=y_j} &= \frac{\hat{u}(x, y_{j+1}) - \hat{u}(x, y_{j-1})}{2h}; \\ \frac{\partial^2 \hat{u}(x, y)}{\partial y^2} \Big|_{y=y_j} &= \frac{\hat{u}(x, y_{j+1}) - 2\hat{u}(x, y_j) + \hat{u}(x, y_{j-1})}{h^2}; \\ \frac{\partial^2 \hat{u}(x, y)}{\partial y \partial x} \Big|_{y=y_j} &= \frac{\hat{u}_x(x, y_{j+1}) - \hat{u}_x(x, y_{j-1})}{2h}. \end{aligned} \quad (8.61)$$

Введем обозначения

$$\begin{aligned} \hat{u}(x, y_j) &= \hat{u}_j(x), \quad \frac{\partial \hat{u}(x, y_j)}{\partial x} = \hat{u}'_j(x), \quad \frac{\partial^2 \hat{u}(x, y_j)}{\partial x^2} = \hat{u}''_j(x), \\ A(x, y_j) &= A_j(x), \quad B(x, y_j) = B_j(x), \quad C(x, y_j) = C_j(x) \text{ и т. д.} \end{aligned}$$

На прямых, описываемых уравнениями $y_j = \text{const}$, $j = 1, \dots, J-1$, получаем

$$\begin{aligned} A_j(x)\hat{u}''_j(x) + B_j(x)\frac{\hat{u}'_{j+1}(x) - \hat{u}'_{j-1}(x)}{h} + C_j(x)\frac{\hat{u}_{j+1}(x) - 2\hat{u}_j + \hat{u}_{j-1}(x)}{h^2} + \\ + D_j(x)\hat{u}'_j(x) + E_j(x)\frac{\hat{u}_{j+1}(x) - \hat{u}_{j-1}(x)}{2h} + F_j(x)\hat{u}_j(x) = G_j(x), \end{aligned} \quad (8.62)$$

$j = 1, \dots, J-1; \quad 0 < x < L.$

Здесь $\hat{u}_1(x), \dots, \hat{u}_{J-1}(x)$ — приближенное решение задачи на прямых $y = y_1 = \text{const}, \dots, y = y_{J-1} = \text{const}$, $A_j(x), \dots, G_j(x)$ — коэффициенты уравнения (8.59) на этих прямых.

В силу краевых условий находим

$$\begin{aligned}\hat{u}_0(x) &= u(x, 0) = \varphi(x), \quad \hat{u}_J(x) = u(x, M) = \psi(x), \\ \hat{u}_j(0) &= \Phi(y_j) = \Phi(jh), \quad \hat{u}_j(L) = \Psi(y_j) = \Psi(jh), \quad j = 1, \dots, J-1.\end{aligned}\tag{8.63}$$

Таким образом, от линейного дифференциального уравнения в частных производных (8.59) и краевыми условиями (8.60) переходим к дифференциально-разностной краевой задаче (8.62), (8.63) для системы из $(n-1)$ обыкновенных дифференциальных уравнений (8.62) относительно $(n-1)$ неизвестных функций $\hat{u}_1(x), \dots, \hat{u}_{J-1}(x)$. В результате ее решения можно найти законы изменения искомой функции $u(x, y)$ вдоль прямых $y_j = \text{const}, j = 0, 1, \dots, J$, т. е. $\hat{u}(x, y_j), j = 0, 1, \dots, J$. Значения функции в промежуточных точках по y может быть найдено с помощью интерполяции. Система (8.62), (8.63) называется *системой уравнений метода прямых*.

Замечания

1. В зависимости от выбора аппроксимационных формул для производных по y можно получить различные системы метода прямых, отличающиеся порядком аппроксимации.

2. Метод прямых удобнее применять в том случае, когда коэффициенты в уравнении (8.59) не зависят от x . Тогда система (8.62) будет системой с постоянными коэффициентами.

3. Если расчетная область отлична от прямоугольника, при реализации метода возникают трудности, связанные с тем, что функции $u_j(x)$ определены на различных множествах [10].

4. При решении нестационарных уравнений дискретизация проводится по пространственным переменным, а время t остается непрерывной переменной.

На практике система дифференциальных уравнений (8.62) из-за большого числа J должна решаться численно, поэтому потребуется вводить шаг не только по y .

Если изменение искомой функции внутри множества Ω в направлении y не велико, то можно ограничиться одной или двумя прямыми.

Пример 8.9. Составить уравнения метода прямых для решения одномерной задачи теплопроводности [4]

$$\begin{aligned}u_t &= a^2 u_{xx}, \quad 0 < x < 1, \quad 0 < t < 1; \\ u(x, 0) &= \psi(x), \quad 0 \leq x \leq 1; \\ u(0, t) &= u(1, t) = 0, \quad 0 < t \leq 1,\end{aligned}$$

где $\psi(x)$ — заданная функция.

□ Для решения задачи выберем на отрезке $[0, 1]$ точки $x_i = ih, i = 0, 1, \dots, I$; $Ih = 1$ и проведем через них прямые, параллельные оси Ot (см. рис. 8.16). На каждой прямой приближенно заменим уравнение теплопроводности

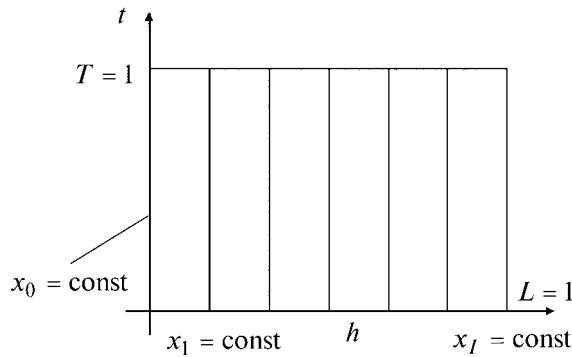


Рис. 8.16

обыкновенным дифференциальным уравнением для функции $u(x_i, t)$. Для этого заменим частную производную u_{xx} по формуле

$$\frac{\partial^2 \hat{u}}{\partial x^2} \Big|_{x=x_i} = \frac{\hat{u}(x_{i-1}, t) - 2\hat{u}(x_i, t) + \hat{u}(x_{i+1}, t)}{h^2}$$

и введем обозначения $\hat{u}(x_i, t) = \hat{u}_i(t)$, $\frac{\partial \hat{u}(x_i, t)}{\partial t} = \hat{u}'_i(t)$.

Тогда на прямых $x_i = \text{const}$, $i = 1, \dots, I - 1$, получаем систему

$$\begin{aligned} \hat{u}'_i(t) &= a^2 \frac{\hat{u}(x_{i-1}, t) - 2\hat{u}(x_i, t) + \hat{u}(x_{i+1}, t)}{h^2}, \quad 1 \leq i \leq I - 1; \\ \hat{u}_0(t) &= u(0, t) = 0, \quad 0 < t \leq 1; \\ \hat{u}_I(t) &= u(1, t) = 0, \quad 0 < t \leq 1 \end{aligned} \tag{8.64}$$

с начальными условиями

$$\hat{u}_i(0) = \psi(x_i) = \psi(ih), \quad i = 1, \dots, I - 1.$$

Полученную систему дифференциальных уравнений можно записать в векторно-матричной форме:

$$\frac{d}{dt} \begin{pmatrix} \hat{u}_1 \\ \vdots \\ \hat{u}_{I-1} \end{pmatrix} = \frac{a^2}{h^2} \underbrace{\begin{pmatrix} -2 & 1 & \cdots & 0 & 0 \\ 1 & -2 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & 1 & -2 & 1 \\ 0 & 0 & \cdots & 1 & -2 \end{pmatrix}}_A \cdot \begin{pmatrix} \hat{u}_1 \\ \vdots \\ \hat{u}_{I-1} \end{pmatrix}.$$

Задача имеет аналитическое решение [4]

$$\hat{u}(t) = H^{-1} \cdot \begin{pmatrix} e^{\frac{\lambda_1 a^2 t}{h^2}} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & e^{\frac{\lambda_{I-1} a^2 t}{h^2}} \end{pmatrix} \cdot Hu(0),$$

где $u(0) = \begin{pmatrix} \psi(h) \\ \psi(2h) \\ \vdots \\ \psi((I-1)h) \end{pmatrix}$, $\lambda_i = -\left(2\cos\frac{\pi i}{n} + 2\right)$, $i = 1, \dots, I-1$, — собственные значения матрицы системы, элементы матрицы H (собственные векторы матрицы A) равны $H_{ik} = \sin\frac{\pi i k}{n}$, $i = 1, \dots, I-1$, $k = 1, \dots, I-1$.

Заметим, что при численном решении системы обыкновенных дифференциальных уравнений при увеличении числа I (улучшении точности) коэффициент жесткости S системы увеличивается [4], [33]. При больших значениях n справедливо.

$$S = \frac{\max_i |\lambda_i|}{\min_i |\lambda_i|} \approx \frac{4I^2}{\pi^2}. \blacksquare$$

8.6. МЕТОД ХАРАКТЕРИСТИК

Данный метод применяется при численном решении уравнений математической физики, описывающих различные процессы в газовой динамике, теории упругости, прочности и других прикладных областях. При его использовании в процессе решения задачи имеется возможность выделять в расчетных областях особые поверхности или линии, отделяющие области с одними свойствами от областей с другими свойствами.

Рассмотрим основные идеи метода на примере задачи Коши для квазилинейной системы дифференциальных уравнений m -го порядка с двумя независимыми переменными x, t :

$$\begin{aligned} a_{11} \frac{\partial u_1(x, t)}{\partial t} + \dots + a_{1m} \frac{\partial u_m(x, t)}{\partial t} + b_{11} \frac{\partial u_1(x, t)}{\partial x} + \dots + b_{1m} \frac{\partial u_m(x, t)}{\partial x} &= f_1(x, t, u), \quad (x, t) \in D, \\ \vdots \\ a_{m1} \frac{\partial u_1(x, t)}{\partial t} + \dots + a_{mm} \frac{\partial u_m(x, t)}{\partial t} + b_{m1} \frac{\partial u_1(x, t)}{\partial x} + \dots + b_{mm} \frac{\partial u_m(x, t)}{\partial x} &= f_m(x, t, u), \\ u_1(x, 0) &= \varphi_1(x), \quad x \in R, \\ \vdots \\ u_m(x, 0) &= \varphi_m(x), \quad x \in R, \end{aligned}$$

где $a_{ij} = a_{ij}(x, t, u)$, $b_{ij} = b_{ij}(x, t, u)$, $i = 1, \dots, m$; $j = 1, \dots, m$; $D = R \times (0, T)$, $T > 0$ — заданный момент времени.

Перепишем систему в векторно-матричном виде

$$\begin{aligned} A(x, t, u) \frac{\partial u(x, t)}{\partial t} + B(x, t, u) \frac{\partial u(x, t)}{\partial x} &= f(x, t, u), \quad (x, t) \in D, \\ u(x, 0) &= \varphi(x), \quad x \in R, \end{aligned} \tag{8.65}$$

где $A(x, t, u) = (a_{ij}(x, t, u))$, $B(x, t, u) = (b_{ij}(x, t, u))$ — матрицы размеров $(m \times m)$;

$f(x, t, u) = \begin{pmatrix} f_1(x, t, u) \\ \vdots \\ f_m(x, t, u) \end{pmatrix}$ — вектор-функция правых частей; $u(x, t) = \begin{pmatrix} u_1(x, t) \\ \vdots \\ u_m(x, t) \end{pmatrix}$ —

искомая вектор-функция; $\varphi(x) = \begin{pmatrix} \varphi_1(x) \\ \vdots \\ \varphi_m(x) \end{pmatrix}$ — заданная вектор-функция.

Пусть известно, что система (8.65) имеет гладкое решение в области D и на некоторой кривой γ оно задано и равно $u(x, t)|_{\gamma}$ (рис. 8.17). Рассмотрим задачу о нахождении решения в окрестности кривой γ .

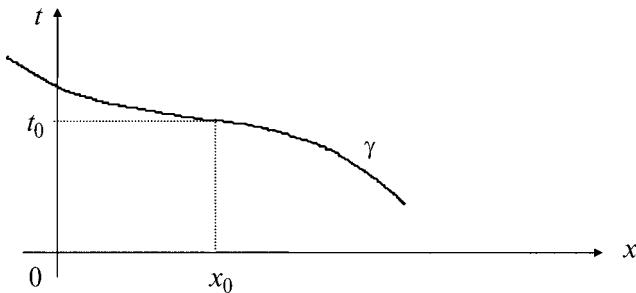


Рис. 8.17

Выберем на кривой γ произвольную точку (x_0, t_0) , а вектор бесконечно малого смещения вдоль этой кривой из точки (x_0, t_0) обозначим (dx, dt) .

Присоединим к уравнениям (8.65) выражение для полного дифференциала (для упрощения записи опустим аргументы):

$$\begin{aligned} A \frac{\partial u}{\partial t} + B \frac{\partial u}{\partial x} &= f; \\ du &= \frac{\partial u}{\partial x} dx + \frac{\partial u}{\partial t} dt. \end{aligned}$$

Умножим первое уравнение на dt , а из второго выразим $\frac{\partial u}{\partial t} dt$:

$$\begin{aligned} A \frac{\partial u}{\partial t} dt + B \frac{\partial u}{\partial x} dt &= f dt; \\ \frac{\partial u}{\partial t} dt &= du - \frac{\partial u}{\partial x} dx. \end{aligned}$$

Подставляя второе уравнение в первое, имеем

$$Adu - A \frac{\partial u}{\partial x} dx + B \frac{\partial u}{\partial x} dt = f dt$$

или

$$(Bdt - Adx) \frac{\partial u}{\partial x} = fdt - Adu. \quad (8.66)$$

Если определитель матрицы $(Bdt - Adx)$ отличен от нуля в каждой точке кривой γ , то частная производная $\frac{\partial u}{\partial x}$ и, следовательно, частная производная $\frac{\partial u}{\partial t}$ определяются однозначно. В этом случае можно найти решение в окрестности кривой γ : $u(x, t) = u(x, t)|_{\gamma} + du$, что соответствует разложению искомого решения по формуле Тейлора до членов первого порядка.

Если определитель матрицы $(Bdt - Adx)$ в каждой точке кривой γ равен нулю, то в силу предположения о существовании решения производные $\frac{\partial u}{\partial x}$, $\frac{\partial u}{\partial t}$ и, следовательно, функция $u(x, t)$ находятся неоднозначно. В этом случае кривая γ называется *характеристикой*.

Приравнивая определитель матрицы $(Bdt - Adx)$ к нулю и предполагая, что $dt \neq 0$ вдоль кривой γ , получаем соотношение

$$\left| B - A \frac{dx}{dt} \right| = 0 \text{ или } |B - A\lambda| = 0, \quad (8.67)$$

где $\lambda(x, t, u) = \frac{dx}{dt}$. Линии, задаваемые дифференциалами смещения dx , dt , вдоль которых справедливо (8.67), являются *характеристиками*. Анализ разрешимости уравнения (8.67) позволяет ввести классификацию систем вида (8.65).

Если уравнение (8.67) имеет m различных действительных корней, то система (8.65) называется *гиперболической*; если их число меньше m , то система называется *параболической*; если уравнение не имеет действительных корней, то — *эллиптической*.

В дальнейшем будем рассматривать только *гиперболические системы*, т. е. имеющие m различных *характеристик*:

$$\left(\frac{dx}{dt} \right)_i = \lambda_i(x, t, u), \quad i = 1, \dots, m, \quad (8.68)$$

где функции $\lambda_i(x, t, u)$ задают *характеристические направления* в каждой точке кривой γ .

Вдоль характеристик решение системы должно удовлетворять *характеристическим соотношениям* (соотношениям совместности), которые получаются в результате приравнивания к нулю определителя m -го порядка расширенной матрицы, т. е.

$$(Bdt - Adx, fdt - Adu). \quad (8.69)$$

При этом следует избегать случаев получения тривиальных тождеств. Характеристические соотношения связывают дифференциалы искомых функций с дифференциалами независимых переменных вдоль характеристик.

Этот способ получения характеристических соотношений основан на применении теоремы Кронекера — Капелли: решение системы (8.66) существует тогда и только тогда, когда ранг матрицы $Bdt - Adx$ этой системы равен рангу расширенной матрицы. Предполагается, что решение существует и определитель матрицы системы (m -го порядка) вдоль характеристик равен

нулю. Поэтому должен быть равен нулю определитель расширенной матрицы, составленный из произвольных m столбцов.

В результате описан принцип получения характеристических уравнений (уравнений характеристик и характеристических соотношений). Их вывод составляет первый этап метода характеристик. Основные идеи первого, а также второго этапа, связанного с конструированием численной схемы, рассмотрим на примере линейной задачи Коши для одномерного волнового уравнения, описывающего распространение плоских звуковых волн в покоящейся среде.

Пример 8.10. Получить соотношения метода характеристик для решения линейной задачи Коши

$$\begin{aligned}\frac{\partial^2 \varphi(x, t)}{\partial t^2} &= \frac{\partial^2 \varphi(x, t)}{\partial x^2}, \quad (x, t) \in R \times (0, T); \\ \varphi(x, 0) &= x^2 + x, \quad x \in R; \\ \varphi_t(x, 0) &= \cos x, \quad x \in R.\end{aligned}$$

□ Данная задача, в частности, описывает процесс распространения звуковых волн в спокойной среде (отсутствуют источники и стоки), следовательно, решаемое уравнение можно рассматривать как уравнение акустики.

Для применения метода характеристик сначала сведем задачу к виду (8.65), т. е. к системе двух уравнений первого порядка. Для этого введем новые искомые функции

$$v(x, t) = \frac{\partial \varphi(x, t)}{\partial t}, \quad p(x, t) = \frac{\partial \varphi(x, t)}{\partial x}.$$

Здесь $v(x, t)$, $p(x, t)$ — отклонения скорости и давления от их значений в невозмущенной среде, вызванные распространением звуковых волн, $\varphi(x, t)$ — потенциал функции $v(x, t)$. Тогда уравнение можно представить в форме

$$\frac{\partial}{\partial t} \left(\frac{\partial \varphi(x, t)}{\partial t} \right) = \frac{\partial}{\partial x} \left(\frac{\partial \varphi(x, t)}{\partial x} \right) \text{ или } \frac{\partial v(x, t)}{\partial t} - \frac{\partial p(x, t)}{\partial x} = 0.$$

Предполагая, что задача имеет гладкое решение (т. е. рассматриваемая функция имеет столько непрерывных производных, сколько нужно для законности приводимых преобразований), воспользуемся свойством, справедливым для дважды непрерывно дифференцируемых функций:

$$\frac{\partial^2 \varphi(x, t)}{\partial x \partial t} = \frac{\partial^2 \varphi(x, t)}{\partial t \partial x} \text{ или } \frac{\partial p(x, t)}{\partial t} = \frac{\partial v(x, t)}{\partial x}.$$

В результате от исходной задачи для волнового уравнения переходим к задаче для системы двух дифференциальных уравнений с начальными условиями относительно новых искомых функций:

$$\begin{aligned}\frac{\partial v(x, t)}{\partial t} - \frac{\partial p(x, t)}{\partial x} &= 0, \quad v(x, 0) = \cos x; \\ \frac{\partial p(x, t)}{\partial t} - \frac{\partial v(x, t)}{\partial x} &= 0, \quad p(x, 0) = 2x + 1.\end{aligned}$$

Запишем систему в виде (8.65)

$$A \frac{\partial u}{\partial t} + B \frac{\partial u}{\partial x} = 0, \quad u(x, 0) = \begin{pmatrix} \cos x \\ 2x+1 \end{pmatrix}, \quad (8.70)$$

где $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $B = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}$, $u = \begin{pmatrix} v \\ p \end{pmatrix}$, $m = 2$.

Для решения задачи (8.70) применим метод характеристик, содержащий два этапа.

Первый этап. Нахождение характеристик и вывод характеристических соотношений. Для этого решим уравнение (8.67), записанное для (8.70):

$$|B - A\lambda| = \begin{vmatrix} -\lambda & -1 \\ -1 & -\lambda \end{vmatrix} = 0.$$

Получаем $\lambda^2 - 1 = 0$, $\lambda_{1,2} = \pm 1$, т. е. два направления характеристик:

$$\left(\frac{dx}{dt} \right)_1 = \lambda_1 = 1; \quad \left(\frac{dx}{dt} \right)_2 = \lambda_2 = -1. \quad (8.71)$$

Интегрируя, найдем общие решения полученных дифференциальных уравнений. Для первого уравнения $x = t + \text{const}$, а для второго $x = -t + \text{const}$. Им соответствуют два семейства характеристик (в данной задаче прямых линий). Их принято обозначать буквой « C^+ » с верхним индексом «+» для семейства характеристик, имеющих меньший угол с положительным направлением оси Ox по сравнению с соответствующим углом для характеристик семейства « C^- ». В данной задаче характеристики семейства « C^+ » имеют положительный наклон, а характеристики семейства « C^- » — отрицательный наклон (рис. 8.18). Условно их также называют характеристиками первого и второго семейств соответственно.

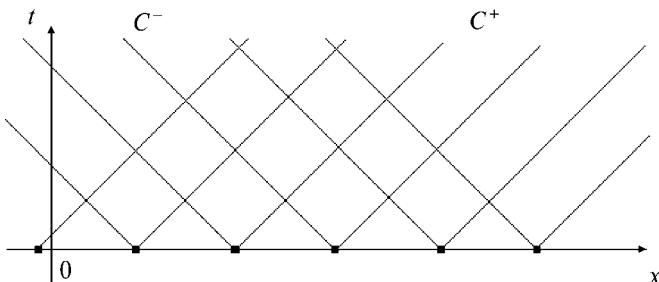


Рис. 8.18

Для получения характеристических соотношений запишем расширенную матрицу (в решаемой задаче $f = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$):

$$(Bdt - Adx, fdt - Adu) = \\ = \underbrace{\begin{pmatrix} 0 & -dt \\ -dt & 0 \end{pmatrix}}_{Bdt} - \underbrace{\begin{pmatrix} dx & 0 \\ 0 & dx \end{pmatrix}}_{Adx} - \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} dv \\ dp \end{pmatrix}}_{Adu} = \begin{pmatrix} -dx & -dt & -dv \\ -dt & -dx & -dp \end{pmatrix}.$$

Приравнивая определитель матрицы, составленной из последних двух столбцов, к нулю, получаем

$$\begin{vmatrix} -dt & -dv \\ -dx & -dp \end{vmatrix} = dpdt - dvdx = 0.$$

Поделив на dt (предполагается, что $dt \neq 0$), имеем

$$dp - dv \frac{dx}{dt} = 0.$$

Так как на прямых семейства C^+ выполняется равенство $\left(\frac{dx}{dt}\right)_1 = \lambda_1 = 1$, то соответствующие характеристическое соотношение имеет вид

$$dp - dv = 0. \quad (8.72)$$

Аналогично на прямых семейства C^- справедливо $\left(\frac{dx}{dt}\right)_2 = \lambda_2 = -1$ и

$$dp + dv = 0. \quad (8.73)$$

Обыкновенные дифференциальные уравнения (8.71)–(8.73) образуют систему уравнений метода характеристик. Полученные характеристические соотношения (8.72), (8.73) в данной задаче могут быть проинтегрированы.

В результате получаются конечные соотношения

$$\begin{aligned} p - v &= \text{const} \text{ (вдоль семейства } C^+); \\ p + v &= \text{const} \text{ (вдоль семейства } C^-), \end{aligned}$$

называемые *инвариантами Римана*. Далее они будут использованы при построении численной схемы.

Результатом первого этапа является получение уравнений характеристик и характеристических соотношений (следствий из них):

$$\begin{aligned} \left(\frac{dx}{dt}\right)_1 &= 1 \text{ (или } x = t + \text{const}); \quad p - v = \text{const}; \\ \left(\frac{dx}{dt}\right)_2 &= -1 \text{ (или } x = -t + \text{const}); \quad p + v = \text{const}. \end{aligned} \quad (8.74)$$

К системе (8.74) следует добавить начальные условия

$$v(x, 0) = \cos x, \quad p(x, 0) = 2x + 1.$$

Второй этап. Конструирование численной схемы и вычисление значений искомых функций.

1. Для простоты изложения выделим на оси Ox отрезок $[0; 1]$ и на нем зададим три точки $0; 0,5; 1$ (рис. 8.19a). Они соответствуют нулевому временному слою. Значения функций $p(x, t)$, $v(x, t)$ в этих точках будем обозначать p_i^0 , v_i^0 , $i = 0, 1, 2$, а сами точки (x_i, t^n) парами индексов (i, n) , где i — номер точки по переменной x , а n — номер временного слоя.

Для $n = 0$ с учетом начальных условий $v(x, 0) = \cos x$, $p(x, 0) = 2x + 1$ получаем

$$v_0^0 = v(0; 0) = 1; \quad v_1^0 = v(0,5; 0) = 0,8776; \quad v_2^0 = v(1; 0) = 0,5403;$$

$$p_0^0 = p(0; 0) = 1; \quad p_1^0 = p(0,5; 0) = 2; \quad p_2^0 = p(1; 0) = 3.$$

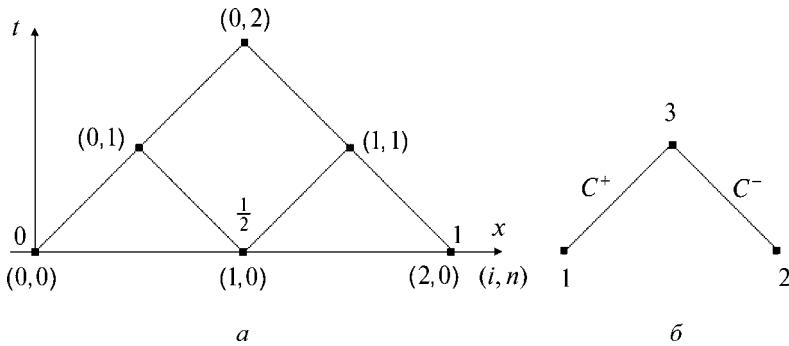


Рис. 8.19

2. Выведем разностные уравнения, определяющие положение точек на следующем временном слое и значения искомых функций в них. Система (8.74) заменяется системой разностных уравнений. Численное решение задачи строится на *характеристической сетке*, которая формируется одновременно с процессом нахождения решения. При этом обычно рассматривается *трехточечный угловой шаблон* с узлами «1», «2», «3», составленный отрезками характеристик (из семейств C^+ и C^-), пересекающихся в рассчитываемой точке «3» (рис. 8.19б). Назовем этот шаблон базовым. Предполагается, что положение узлов «1», «2» и значения решения в узлах известны из начальных условий или рассчитаны на предыдущем шаге. Положение узла «3» и значение решения в нем подлежат определению.

Аппроксимируя производную по формулам:

$$\begin{aligned} \left(\frac{dx}{dt} \right)_1 &\cong \frac{x_3 - x_1}{t_3 - t_1} \text{ (вдоль характеристики из семейства } C^+ \text{ с } \lambda_1 = 1); \\ \left(\frac{dx}{dt} \right)_2 &\cong \frac{x_3 - x_2}{t_3 - t_2} \text{ (вдоль характеристики из семейства } C^- \text{ с } \lambda_2 = -1), \end{aligned}$$

получаем

$$\begin{aligned} x_3 - x_1 &= t_3 - t_1; \\ x_3 - x_2 &= t_2 - t_3. \end{aligned} \tag{8.75}$$

Условие $p - v = \text{const}$ вдоль характеристик из семейства C^+ записывается для лежащих на ней узлов «1» и «3»:

$$p_3 - v_3 = p_1 - v_1, \tag{8.76}$$

а условие $p + v = \text{const}$ вдоль характеристики из семейства C^- записывается для лежащих на ней узлов «2» и «3»:

$$p_3 + v_3 = p_2 + v_2. \tag{8.77}$$

Система (8.75)–(8.77) является замкнутой относительно искомых значений x_3, t_3, p_3, v_3 при известных $x_1, t_1, x_2, t_2, p_1, v_1, p_2, v_2$. Формулы (8.75)–(8.77) являются базовыми для расчета всех точек, принадлежащих области D . После

совмещения узлов «1» и «2» базового шаблона с выбранными очередными узлами сетки в (8.75)–(8.77) подставляются конкретные величины в соответствии с принятыми обозначениями.

3. Производится расчет первого временного слоя.

В решаемой задаче на нулевом временном слое угловой шаблон может занимать два положения (рис. 8.20 a , b).

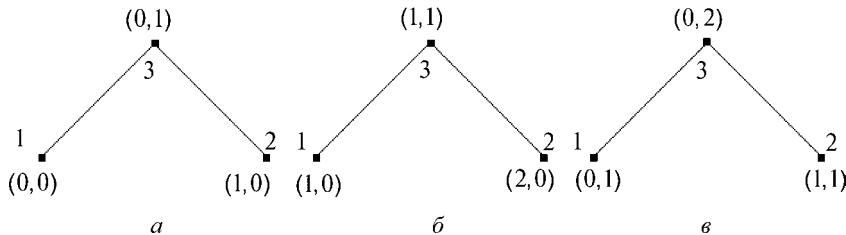


Рис. 8.20

Для точки с номерами (0, 1) первого рассчитываемого слоя узлы «1» и «2» углового шаблона совмещаются с узлами (0, 0) и (1, 0) (рис. 8.20 a). Поэтому система (8.75)–(8.77) имеет вид

$$\begin{aligned}x_{0,1} - x_{0,0} &= t_{0,1} - t_{0,0}; \quad x_{0,1} - x_{1,0} = t_{1,0} - t_{0,1}; \\p_0^1 - v_0^1 &= p_0^0 - v_0^0; \quad p_0^1 + v_0^1 = p_1^0 + v_1^0.\end{aligned}$$

Подставляя $t_{0,0} = 0$, $t_{1,0} = 0$, $x_{0,0} = 0$, $x_{1,0} = 0,5$ и известные значения p_0^0 , p_1^0 , v_0^0 , v_1^0 на нулевом слое (см. п. 1), получаем

$$\begin{aligned}x_{0,1} - 0 &= t_{0,1} - 0; \quad x_{0,1} - 0,5 = 0 - t_{0,1}; \\p_0^1 - v_0^1 &= 1 - 1 = 0; \quad p_0^1 + v_0^1 = 2 + 0,8776 = 2,8776.\end{aligned}$$

Решение этой системы:

$$x_{0,1} = 0,25; \quad t_{0,1} = 0,25; \quad p_0^1 = 1,4388; \quad v_0^1 = 1,4388.$$

Для второй точки первого слоя (рис. 8.20 b) с номерами (1, 1) узлы «1» и «2» углового шаблона совмещаются с узлами (1, 0) и (2, 0). Тогда система (8.75)–(8.77) имеет вид:

$$\begin{aligned}x_{1,1} - x_{1,0} &= t_{1,1} - t_{1,0}; \quad x_{1,1} - x_{2,0} = t_{2,0} - t_{1,1}; \\p_1^1 - v_1^1 &= p_1^0 - v_1^0; \quad p_1^1 + v_1^1 = p_2^0 + v_2^0.\end{aligned}$$

Подставляя $t_{1,0} = 0$, $t_{2,0} = 0$, $x_{1,0} = 0,5$, $x_{2,0} = 1$ и известные значения p_1^0 , v_1^0 , p_2^0 , v_2^0 на нулевом слое (см. п. 1), получаем

$$\begin{aligned}x_{1,1} - 0,5 &= t_{1,1} - 0; \quad x_{1,1} - 1 = 0 - t_{1,1}; \\p_1^1 - v_1^1 &= 2 - 0,8776 = 1,1224; \quad p_1^1 + v_1^1 = 3 + 0,5403 = 3,5403.\end{aligned}$$

Отсюда $x_{1,1} = 0,75$; $t_{1,1} = 0,25$; $p_1^1 = 2,33135$; $v_1^1 = 1,20895$.

4. Для расчета точки $(0, 2)$ следующего временного слоя узлы «1» и «2» углового шаблона совмещаются с узлами $(0, 1)$ и $(1, 1)$ (рис. 8.20e). Тогда система (8.75)–(8.77) принимает вид:

$$\begin{aligned}x_{0,2} - x_{0,1} &= t_{0,2} - t_{0,1}; \quad x_{0,2} - x_{1,1} = t_{1,1} - t_{0,2}; \\p_0^2 - v_0^2 &= p_0^1 - v_0^1; \quad p_0^2 + v_0^2 = p_1^1 + v_1^1.\end{aligned}$$

Подставляя значения с предыдущего слоя, имеем

$$\begin{aligned}x_{0,2} - 0,25 &= t_{0,2} - 0,25; \quad x_{0,2} - 0,75 = 0,25 - t_{0,2}; \\p_0^2 - v_0^2 &= 1,4388 - 1,4388; \quad p_0^2 + v_0^2 = 2,33135 + 1,20895.\end{aligned}$$

Отсюда $x_{0,2} = 0,5$; $t_{0,2} = 0,5$; $p_0^2 = v_0^2 = 1,77015$.

Очевидно (см. рис. 8.19a), количество узлов на каждом последующем слое уменьшается на единицу. Расчет продолжается до замыкания области, так как на последнем слое останется один узел. В данном примере этим узлом является точка $(0, 2)$. ■

Замечание. Рассмотренная в примере 8.10 задача является всего лишь модельной, выбранной для пояснения основных принципов конструирования метода характеристик и его численной реализации. В вычислительной практике, как правило, ставятся нелинейные задачи. Проиллюстрируем некоторые особенности их решения на следующем примере.

Пример 8.11. Получить соотношения метода характеристик для решения нелинейной задачи Коши

$$\begin{aligned}a_{11} \frac{\partial u_1(x, t)}{\partial t} + a_{12} \frac{\partial u_2(x, t)}{\partial t} + b_{11} \frac{\partial u_1(x, t)}{\partial x} + b_{12} \frac{\partial u_2(x, t)}{\partial x} &= f_1; \\a_{21} \frac{\partial u_1(x, t)}{\partial t} + a_{22} \frac{\partial u_m(x, t)}{\partial t} + b_{21} \frac{\partial u_1(x, t)}{\partial x} + b_{22} \frac{\partial u_2(x, t)}{\partial x} &= f_2, \quad (x, t) \in D; \\u_1(x, 0) &= \varphi_1(x), \quad u_2(x, 0) = \varphi_2(x), \quad x \in R,\end{aligned}$$

где $a_{ij} = a_{ij}(x, t, u)$; $b_{ij} = b_{ij}(x, t, u)$; $f_i = f_i(x, t, u)$; $i = 1, 2$; $j = 1, 2$; $\varphi_1(x)$, $\varphi_2(x)$ — заданные функции; $D = R \times (0, T)$, $T > 0$ — заданный момент времени.

Поставленная задача является частным случаем задачи (8.65) при $m = 2$. Применим метод характеристик, включающий два этапа.

Первый этап. Нахождение характеристик и характеристических соотношений. Уравнение (8.67) имеет вид

$$|B - A\lambda| = \begin{vmatrix} b_{11} - \lambda a_{11} & b_{12} - \lambda a_{12} \\ b_{21} - \lambda a_{21} & b_{22} - \lambda a_{22} \end{vmatrix} = 0$$

или

$$a\lambda^2 + b\lambda + c = 0,$$

где $a = a_{11}a_{22} - a_{12}a_{21}$; $b = a_{21}b_{12} + a_{12}b_{21} - a_{11}b_{22} - a_{22}b_{11}$; $c = b_{11}b_{22} - b_{12}b_{21}$.

Решая полученное уравнение, находим выражения для двух характеристических направлений в каждой точке расчетной области:

$$\begin{aligned} \left(\frac{dx}{dt} \right)_1 &= \lambda_1(x, t, u_1, u_2) = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \quad (\text{определяет семейство } C^+); \\ \left(\frac{dx}{dt} \right)_2 &= \lambda_2(x, t, u_1, u_2) = \frac{-b - \sqrt{b^2 - 4ac}}{2a} \quad (\text{определяет семейство } C^-). \end{aligned} \quad (8.78)$$

Сформируем расширенную матрицу

$$(Bdt - Adx, fdt - Adu) = \begin{pmatrix} b_{11}dt - a_{11}dx & b_{12}dt - a_{12}dx & f_1dt - a_{11}du_1 - a_{12}du_2 \\ b_{21}dt - a_{21}dx & b_{22}dt - a_{22}dx & f_2dt - a_{21}du_1 - a_{22}du_2 \end{pmatrix}.$$

Характеристические соотношения получаются в результате приравнивания нулю определителя второго порядка ($m = 2$), составленного из второго и третьего столбцов расширенной матрицы:

$$\begin{vmatrix} b_{12}dt - a_{12}dx & f_1dt - a_{11}du_1 - a_{12}du_2 \\ b_{22}dt - a_{22}dx & f_2dt - a_{21}du_1 - a_{22}du_2 \end{vmatrix} = 0.$$

Раскрыв определитель и поделив на $dt \neq 0$, получим

$$\left(e - a \frac{dx}{dt} \right) du_1 + g du_2 + h dt + q dx = 0,$$

где $e = a_{11}b_{22} - a_{21}b_{12}$; $g = a_{12}b_{22} - a_{22}b_{12}$; $h = b_{12}f_2 - b_{22}f_1$; $q = a_{22}f_1 - a_{12}f_2$.

Последние два слагаемых перепишем в эквивалентной форме

$$h dt + q dx = \left(h + q \frac{dx}{dt} \right) dt.$$

Поскольку для семейства характеристик C^+ выполняется равенство $\left(\frac{dx}{dt} \right)_1 = \lambda_1$, а для семейства C^- — равенство $\left(\frac{dx}{dt} \right)_2 = \lambda_2$, то характеристические соотношения имеют вид

$$\begin{aligned} (e - a\lambda_1)du_1 + gdu_2 + (h + q\lambda_1)dt &= 0 \quad (\text{для семейства } C^+); \\ (e - a\lambda_2)du_1 + gdu_2 + (h + q\lambda_2)dt &= 0 \quad (\text{для семейства } C^-). \end{aligned} \quad (8.79)$$

В результате получена характеристическая система (8.78), (8.79) из четырех обыкновенных дифференциальных уравнений. Они позволяют найти значения t , x , u_1 , u_2 вдоль характеристик.

Второй этап. Конструирование численной схемы.

Как и в предыдущем примере, будем использовать трехточечный угловой шаблон (рис. 8.19б). Предполагается, что в точках «1» и «2» известны значения t , x , u_1 , u_2 из начальных данных или из расчета предыдущего слоя, т. е. t_1 , x_1 , u_{11} , u_{21} для точки «1» и t_2 , x_2 , u_{12} , u_{22} для точки «2». Здесь при обозначении значений функций u_1 , u_2 , а далее и при обозначении других функций в точках шаблона первый нижний индекс соответствует номеру самой функции, а второй индекс — номеру точки в шаблоне. Требуется найти значения t_3 , x_3 , u_{13} , u_{23} , определяющие положение точки «3» и значения искомых функций в ней.

Предполагается, что точки «1» и «2» близки друг к другу, поэтому заменим дифференциалы конечными приращениями, а производную $\frac{dx}{dt}$ — конечно-разностным отношением:

$$dt \cong \Delta t = t_3 - t_1, \quad dx \cong \Delta x = x_3 - x_1, \quad du_1 \cong \Delta u_1 = u_{13} - u_{11}, \quad du_2 \cong \Delta u_2 = u_{23} - u_{21},$$

$$\left(\frac{dx}{dt}\right)_1 \cong \frac{x_3 - x_1}{t_3 - t_1} \text{ (для семейства } C^+);$$

$$dt \cong \Delta t = t_3 - t_2, \quad dx \cong \Delta x = x_3 - x_2, \quad du_1 \cong \Delta u_1 = u_{13} - u_{12}, \quad du_2 \cong \Delta u_2 = u_{23} - u_{22},$$

$$\left(\frac{dx}{dt}\right)_2 \cong \frac{x_3 - x_2}{t_3 - t_2} \text{ (для семейства } C^-).$$

В итоге уравнения характеристик являются уравнениями прямых, соединяющих точки «1» и «2» с точкой «3», а характеристическая система (8.78), (8.79) преобразуется к виду

$$\begin{aligned} x_3 - x_1 &= (\lambda_1)_1(t_3 - t_1); \\ x_3 - x_2 &= (\lambda_2)_2(t_3 - t_2); \\ (e - a\lambda_1)_1(u_{13} - u_{11}) + (g)_1 \cdot (u_{23} - u_{21}) + (h + q\lambda_1)_1(t_3 - t_1) &= 0; \\ (e - a\lambda_2)_2(u_{13} - u_{12}) + (g)_2 \cdot (u_{23} - u_{22}) + (h + q\lambda_2)_2(t_3 - t_2) &= 0. \end{aligned} \quad (8.80)$$

Здесь нижние индексы 1 и 2 указывают, что значения всех функций, входящих в соответствующие выражения, подсчитываются в точке «1» или «2».

Решая систему (8.80), можно найти значения t_3 , x_3 , u_{13} , u_{23} , определяющие положение точки «3» и значения искомых функций в ней. Сначала обычно решаются первые два уравнения и определяются значения t_3 , x_3 , а затем — последние два уравнения, откуда находятся u_{13} , u_{23} .

Опишем теперь общую схему численного решения поставленной задачи Коши.

1. При $t = 0$ на оси абсцисс задаются точки с некоторым шагом. В этих точках (на рис. 8.21 выделен фрагмент, содержащий точки 1–6) значения

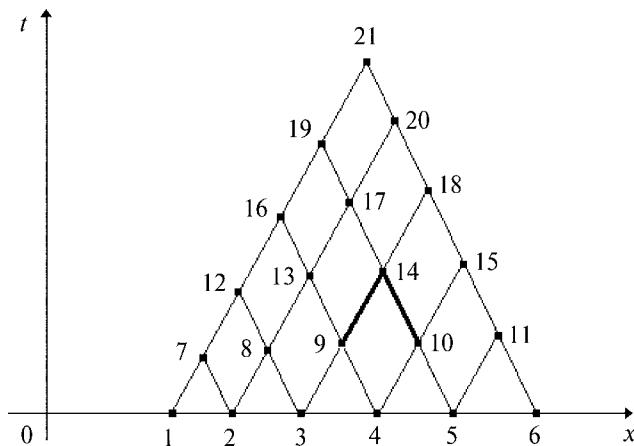


Рис. 8.21

функций u_1 , u_2 определяются начальными данными, т. е. функциями $\varphi_1(x)$, $\varphi_2(x)$.

2. В соответствии с изложенной методикой последовательно рассчитываются координаты точек 7–11 и приближенные значения решения в них. При этом точки «1» и «2» углового шаблона последовательно совмещаются с точками 1 и 2, затем с точками 2 и 3, 3 и 4, 4 и 5, 5 и 6. Узлами выстраиваемой характеристической сетки служат точки пересечения характеристик (они образуют следующий временной слой).

3. Расчет слоев продолжается до достижения последней точки (точки 21) треугольной области.

Отметим, что дифференциальным уравнениям (8.78) соответствуют криволинейные характеристики, а точка «3» должна быть точкой пересечения этих криволинейных характеристик. При переходе от дифференциальных уравнений (8.78), (8.79) к разностным дифференциалам были заменены конечными разностями, а производные — конечно-разностными отношениями. Вследствие этого криволинейные характеристики были аппроксимированы отрезками прямых. Решение системы (8.80) объявляется первым приближением, т. е. $t_3^{(1)} = t_3$, $x_3^{(1)} = x_3$, $u_{13}^{(1)} = u_{13}$, $u_{23}^{(1)} = u_{23}$, и может возникнуть необходимость в уточнении координат точки «3» и значений искомых функций в этой точке. Ограничимся изложением одного из возможных способов уточнения [3], [26].

После нахождения первого приближения следует найти средние арифметические значения:

$$\begin{aligned}\bar{\lambda}_1 &= \frac{(\lambda_1)_1 + (\lambda_1)_3}{2}, \quad \bar{\lambda}_2 = \frac{(\lambda_2)_2 + (\lambda_2)_3}{2}; \\ K_1 &= \frac{(e - a\lambda_1)_1 + (e - a\lambda_1)_3}{2}, \quad K_2 = \frac{(e - a\lambda_2)_2 + (e - a\lambda_2)_3}{2}, \\ L_1 &= \frac{(g)_1 + (g)_3}{2}, \quad L_2 = \frac{(g)_2 + (g)_3}{2}; \\ M_1 &= \frac{(h + q\lambda_1)_1 + (h + q\lambda_1)_3}{2}, \quad M_2 = \frac{(h + q\lambda_2)_2 + (h + q\lambda_2)_3}{2},\end{aligned}$$

где нижний индекс 3 указывает, что соответствующее выражение подсчитывается в найденной точке «3». Эти соотношения получаются, если при интегрировании дифференциальных уравнений применять правило трапеций [26].

Затем решается система

$$\begin{aligned}x_3^{(2)} - x_1 &= \bar{\lambda}_1(t_3^{(2)} - t_1); \\ x_3^{(2)} - x_2 &= \bar{\lambda}_2(t_3^{(2)} - t_2); \\ K_1(u_{13}^{(2)} - u_{11}) + L_1 \cdot (u_{23}^{(2)} - u_{21}) + M_1(t_3^{(2)} - t_1) &= 0; \\ K_2(u_{13}^{(2)} - u_{12}) + L_2 \cdot (u_{23}^{(2)} - u_{22}) + M_2(t_3^{(2)} - t_2) &= 0\end{aligned}$$

относительно нового (второго) приближения $t_3^{(2)}$, $x_3^{(2)}$, $u_{13}^{(2)}$, $u_{23}^{(2)}$. Описанный процесс продолжают до тех пор, пока решения, полученные при двух последовательных приближениях, не будут совпадать с заданной точностью. Как

правило, если расстояние между точками «1» и «2» невелико, достаточно сделать два уточнения.

Замечания

1. Приведенные примеры иллюстрируют важное свойство метода характеристик, отличающее его от ранее рассмотренных разностных методов. Оно заключается в том, что расчеты осуществляются вдоль характеристик — линий, по которым распространяется информация об исходных данных, что придает методу четкий физический смысл.

2. Описанный метод расчета, когда точки следующего слоя заранее неизвестны, а вычисления ведутся вдоль характеристик семейств C^+ и C^- , выделяемых в процессе расчета, называется *классическим методом характеристик*. Существуют модификации этого метода, когда точки следующего слоя задаются заранее, а для нахождения значений искомых функций в них характеристики выпускаются в направлении к известному слою. Такие методы называются *сеточно-характеристическими* [22].

8.7. МЕТОД ГОДУНОВА

В механике сплошных сред при решении задач, описывающих квазилинейными дифференциальными уравнениями в частных производных, в расчетных областях могут возникать *поверхности сильного разрыва*, к которым относятся ударные волны, центрированные волны разрежения, поверхности контактного разрыва. На этих поверхностях терпят разрыв как искомые функции, характеризующие исследуемые процессы, так и их производные. При решении таких задач используются два класса численных схем, в одном из которых с использованием законов сохранения явно выделяются поверхности разрыва, а в другом эти поверхности явно не выделяются, а рассчитываются «насквозь» вместе с областями непрерывности. К последнему классу методов относится метод С. К. Годунова [50]. Идею его конструирования рассмотрим на примере решения смешанной краевой задачи для уравнений акустики.

Пример 8.12. Получить соотношения метода Годунова для решения задачи

$$\begin{aligned} \frac{\partial u(x, t)}{\partial t} + \frac{1}{\rho_0} \frac{\partial p(x, t)}{\partial x} &= 0; \\ \frac{\partial p(x, t)}{\partial t} + \rho_0 a_0^2 \frac{\partial u(x, t)}{\partial x} &= 0 \end{aligned}$$

с начальными условиями

$$u(x, 0) = \psi_1(x), \quad p(x, 0) = \psi_2(x), \quad 0 \leq x \leq 1,$$

и краевыми условиями

$$p(0, t) = \varphi_1(t), \quad p(1, t) = \varphi_2(t), \quad 0 < t \leq T.$$

Система двух дифференциальных уравнений описывает распространение плоских звуковых волн (малых возмущений) в покоящейся среде. Здесь $u(x, t)$,

$p(x, t)$ — малые отклонения скорости и давления от их значений в невозмущенной среде, вызванные распространением звуковых волн; ρ_0 — плотность невозмущенной среды; a_0 — скорость звука.

□ Рассмотрим основные этапы конструирования метода Годунова.

1. Введем разбиение отрезка $[0; 1]$ с шагом $h = \frac{1}{n}$, где n — число промежутков. В результате получим узлы $x_i, i = 0, 1, \dots, n$, определяемые соотношением $x_{i+1} = x_i + h, i = 0, 1, \dots, n-1; x_0 = 0, x_n = 1$.

2. Заменим функции $\psi_1(x), \psi_2(x)$ на начальном слое $t = 0$ кусочно-постоянными (пример описанной замены для некоторой произвольной функции $\psi_1(x)$ изображен на рис. 8.22) путем их интегрального осреднения по формулам

$$u_{i-\frac{1}{2}} = \frac{1}{h} \int_{x_{i-1}}^{x_i} \psi_1(x) dx, \quad p_{i-\frac{1}{2}} = \frac{1}{h} \int_{x_{i-1}}^{x_i} \psi_2(x) dx, \quad i = 1, 2, \dots, n. \quad (8.81)$$

Здесь и далее все заданные или ранее рассчитанные функции снабжаются нижними полуцелыми индексами. Из рисунка 8.22 видно, что полученная кусочно-постоянная функция имеет в узлах разрывы первого рода.

3. Выберем шаг τ по времени t и образуем следующий расчетный слой ($t_1 = \tau$) с теми же узлами x_i (рис. 8.23). Таким образом, образуется совокуп-

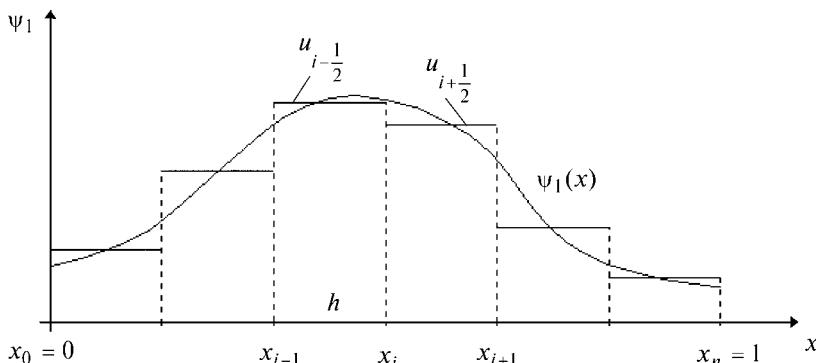


Рис. 8.22

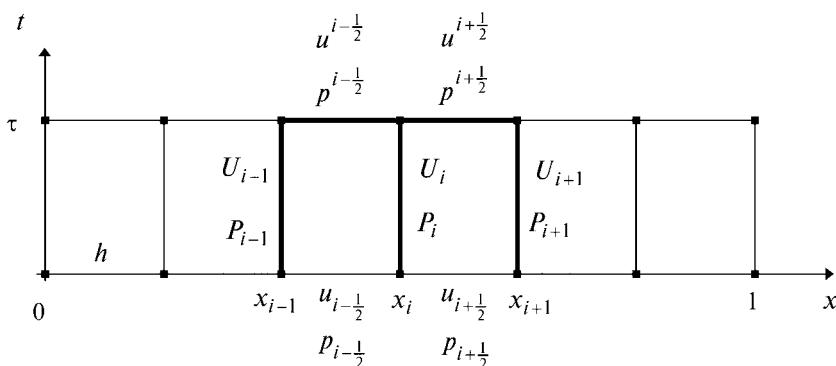


Рис. 8.23

ность прямоугольных ячеек. Предполагается, что искомые функции на горизонтальных и вертикальных границах образованных ячеек также являются постоянными.

На рассчитываемом слое при $t = \tau$ приближенное решение задачи также представляется в виде кусочно-постоянной функции, имеющей разрывы в узлах сетки, совпадающих с узлами на предыдущем слое. Это означает, что во времени на слоях $t = \text{const}$ сохраняется структура решения, изображенная на рисунке 8.22. Значения искомых функций на рассчитываемом слое (верхних границах ячеек) обозначаются верхними полуцелыми индексами: $u^{i-\frac{1}{2}}, p^{i-\frac{1}{2}}$, $i = 1, 2, \dots, n$. На вертикальных границах ячеек значения функций обозначаются прописными буквами с нижними целыми индексами: U_i, P_i , $i = 0, \dots, n$.

При $i = 0, i = n$ значения P_0 и P_n определяются краевыми условиями по формулам, аналогичным (8.81):

$$P_0 = \frac{1}{\tau} \int_0^\tau \phi_1(t) dt, \quad P_n = \frac{1}{\tau} \int_0^\tau \phi_2(t) dt. \quad (8.82)$$

Значения $U_1, \dots, U_{n-1}, P_1, \dots, P_{n-1}$ находятся в результате решения вспомогательной задачи о распаде произвольного разрыва, описанного в п. 4, а значения U_0, U_n определяются в п. 5.

4. Выберем две соседние ячейки с узлами x_{i-1}, x_i и x_i, x_{i+1} (эти ячейки выделены на рис. 8.23 жирными линиями). С целью нахождения значений U_i, P_i на внутренней границе ячеек при $x = x_i$ рассмотрим решение вспомогательной задачи, называемой *задачей о распаде произвольного разрыва* [7], [50].

Дана система уравнений

$$\begin{aligned} \frac{\partial u(x, t)}{\partial t} + \frac{1}{\rho_0} \frac{\partial p(x, t)}{\partial x} &= 0; \\ \frac{\partial p(x, t)}{\partial t} + \rho_0 a_0^2 \frac{\partial u(x, t)}{\partial x} &= 0 \end{aligned} \quad (8.83)$$

с начальными условиями вида

$$u(x, 0) = \begin{cases} u_{i-\frac{1}{2}}, & x_{i-1} \leq x < x_i; \\ u_{i+\frac{1}{2}}, & x_i \leq x < x_{i+1}; \end{cases} \quad p(x, 0) = \begin{cases} p_{i-\frac{1}{2}}, & x_{i-1} \leq x < x_i; \\ p_{i+\frac{1}{2}}, & x_i \leq x < x_{i+1}. \end{cases}$$

Требуется найти решение $u(x, t), p(x, t)$ при $t > 0$ и $x = x_i$, т. е. $u(x_i, t) = U_i, p(x_i, t) = P_i$.

Заметим, что в данной задаче начальные данные имеют разрыв первого рода в точке x_i . При $t > 0$ они трансформируются, т. е. происходит распад разрыва, что обусловливает название рассматриваемой задачи.

Для ее решения применим метод характеристик (см. п. 8.6). Перепишем систему (8.83) в матричной форме

$$A \frac{\partial u(x, t)}{\partial t} + B \frac{\partial u(x, t)}{\partial x} = 0,$$

где

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & \frac{1}{\rho_0} \\ \rho_0 a_0^2 & 0 \end{pmatrix}, \quad \mathbf{u} = \begin{pmatrix} u \\ p \end{pmatrix}.$$

Для нахождения характеристических направлений используем условие (8.67):

$$|B - A\lambda| = \begin{vmatrix} -\lambda & \frac{1}{\rho_0} \\ \rho_0 a_0^2 & -\lambda \end{vmatrix} = 0.$$

Отсюда $\lambda = \pm a_0$ и

$$\left(\frac{dx}{dt} \right)^+ = \lambda^+ = a_0, \quad \left(\frac{dx}{dt} \right)^- = \lambda^- = -a_0.$$

Интегрируя последние уравнения, имеем

$$\begin{aligned} x - a_0 t &= \text{const} \text{ (для семейства } C^+); \\ x + a_0 t &= \text{const} \text{ (для семейства } C^-). \end{aligned}$$

Если при $t = 0$ выполняется условие $x = x_i$, то из последних соотношений выделяются две прямые, исходящие из точки x_i на начальном слое:

$$\begin{aligned} x - a_0 t &= x_i \text{ (прямая семейства } C^+); \\ x + a_0 t &= x_i \text{ (прямая семейства } C^-). \end{aligned}$$

Для получения характеристических соотношений запишем расширенную матрицу (8.69) (учитывая, что $f \equiv 0$):

$$\begin{pmatrix} B dt - A dx, -A \begin{pmatrix} du \\ dp \end{pmatrix} \end{pmatrix} = \begin{pmatrix} -dx & \frac{1}{\rho_0} dt & -du \\ \rho_0 a_0^2 dt & -dx & -dp \end{pmatrix}.$$

Приравнивая определитель матрицы, составленной из двух последних столбцов, к нулю, получаем

$$-\frac{1}{\rho_0} dp dt - dx du = 0.$$

Полагая $dt \neq 0$, разделим уравнение на dt :

$$-\frac{1}{\rho_0} dp - \frac{dx}{dt} du = 0.$$

Подставляя вместо $\frac{dx}{dt}$ сначала $\lambda^+ = a_0$, а затем $\lambda^- = -a_0$, получаем

$$du + \frac{1}{\rho_0 a_0} dp = 0 \text{ (соотношение вдоль семейства } C^+);$$

$$du - \frac{1}{\rho_0 a_0} dp = 0 \text{ (соотношение вдоль семейства } C^-).$$

Так как коэффициенты обоих уравнений постоянны, то их можно переписать в эквивалентной форме

$$d\left(u + \frac{1}{\rho_0 a_0} p\right) = 0 \text{ (соотношение вдоль семейства } C^+);$$

$$d\left(u - \frac{1}{\rho_0 a_0} p\right) = 0 \text{ (соотношение вдоль семейства } C^-).$$

Интегрируя и обозначая постоянные интегрирования I^+ , I^- , имеем

$$u + \frac{1}{\rho_0 a_0} p = I^+ \text{ (соотношение вдоль семейства } C^+);$$

$$u - \frac{1}{\rho_0 a_0} p = I^- \text{ (соотношение вдоль семейства } C^-). \quad (8.84)$$

Полученные соотношения называются *инвариантами Римана*. Они показывают, что величина $u + \frac{1}{\rho_0 a_0} p = I^+$ остается постоянной вдоль семейства C^+ , а величина $u - \frac{1}{\rho_0 a_0} p = I^-$ остается постоянной вдоль семейства C^- .

Теперь, используя направления характеристик и сохраняющиеся вдоль них инварианты, построим численную схему решения задачи о распаде разрыва.

Рассмотрим две характеристики $x - a_0 t = x_{i-1}$ и $x - a_0 t = x_i$ из семейства C^+ , исходящие из узлов x_{i-1} , x_i , а также две характеристики $x + a_0 t = x_i$ и $x + a_0 t = x_{i+1}$ из семейства C^- , исходящие из узлов x_i , x_{i+1} (рис. 8.24). В результате пересечения указанных характеристик образуются области I, II, III, первые две из которых представлены треугольниками, а третья — четырехугольником. Вертикальная граница ($x_i = \text{const}$), разделяющая ячейки, на ко-

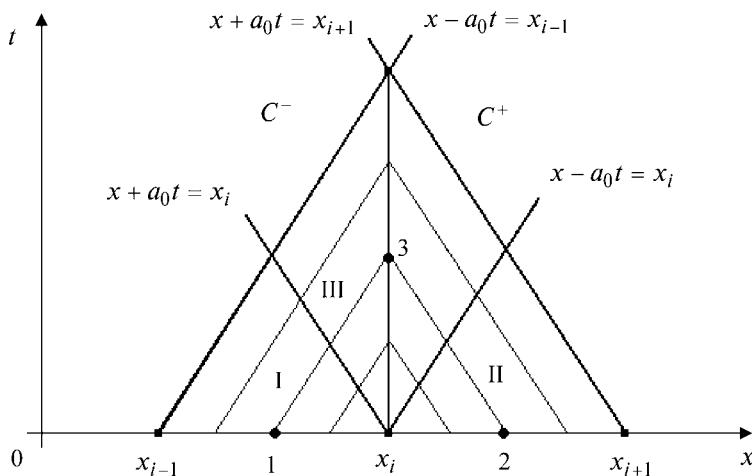


Рис. 8.24

торой ищется решение задачи, располагается в области III. Покажем, что на этой границе искомое решение постоянно. Для этого выберем на ней произвольную точку (назовем ее точка 3), а решение в ней обозначим U_i, P_i . В эту точку приходит характеристика семейства C^+ , исходящая из точки 1 промежутка (x_{i-1}, x_i) . Точке 1 соответствуют заданные значения $u_{i-\frac{1}{2}}, p_{i-\frac{1}{2}}$. Вдоль этой характеристики остается постоянной величина $u + \frac{1}{\rho_0 a_0} p = I^+$, т. е.

$$U_i + \frac{1}{\rho_0 a_0} P_i = u_{i-\frac{1}{2}} + \frac{1}{\rho_0 a_0} p_{i-\frac{1}{2}}.$$

Кроме того, в точку 3 приходит характеристика семейства C^- , исходящая из точки 2 промежутка (x_i, x_{i+1}) . Точке 2 соответствуют заданные значения $u_{i+\frac{1}{2}}, p_{i+\frac{1}{2}}$. Вдоль этой характеристики остается постоянной величина

$$u - \frac{1}{\rho_0 a_0} p = I^-, \text{ т. е.}$$

$$U_i - \frac{1}{\rho_0 a_0} P_i = u_{i+\frac{1}{2}} - \frac{1}{\rho_0 a_0} p_{i+\frac{1}{2}}.$$

Разрешая систему из двух последних уравнений относительно U_i, P_i , получаем

$$\begin{aligned} U_i &= \frac{u_{i-\frac{1}{2}} + u_{i+\frac{1}{2}}}{2} - \frac{p_{i+\frac{1}{2}} - p_{i-\frac{1}{2}}}{2\rho_0 a_0}; \\ P_i &= \frac{p_{i-\frac{1}{2}} + p_{i+\frac{1}{2}}}{2} - \rho_0 a_0 \frac{u_{i+\frac{1}{2}} - u_{i-\frac{1}{2}}}{2}, \quad i=1, \dots, n-1. \end{aligned} \tag{8.85}$$

Поскольку точка 3 была выбрана произвольно на вертикальной границе ячейки, то полученные выражения определяют решение в любой точке этой границы (очевидно, решение постоянно вдоль границы), за исключением самой точки x_i и точки пересечения характеристик, исходящих из точек x_{i-1}, x_{i+1} . Таким образом, формулы (8.85) определяют «большие» величины U_i, P_i на всех внутренних вертикальных границах ячеек.

Результаты решения данной вспомогательной задачи определяют шаг τ по времени. Действительно, влияние каждого из разрывов, относящихся к соседним узлам x_{i-1}, x_i на слое $t = 0$, ограничивается характеристиками, исходящими из этих узлов. Физически это означает, что разрывы, возникшие при $t = 0$, распространяются со скоростью звука a_0 . Характеристики, исходящие из соседних узлов навстречу друг другу, пересекутся в точке с координатами $t = \frac{h}{2a_0}, x = \frac{x_{i-1} + x_i}{2}$. От этой встречи разрывов возникнут новые волны, которые через следующий отрезок времени длительностью $\frac{h}{2a_0}$ достигнут вертикальных границ ($x_{i-1} = \text{const}, x_i = \text{const}$) соседних ячеек. Именно до этого момента, т. е. в течение времени $\frac{h}{a_0} = \frac{h}{2a_0} + \frac{h}{2a_0}$, значения U_i, P_i будут сохраняться неизменными и определяться по формулам (8.85). Таким образом, величина шага по времени должна выбираться из условия $\tau < \frac{h}{a_0}$.

5. При $i = 0$ для нахождения величины U_0 при известном P_0 используется инвариант Римана $u - \frac{1}{\rho_0 a_0} p = I^-$ вдоль семейства C^- (рис. 8.25). Поскольку на промежутке $(0, x_1)$ заданы значения $u_{\frac{1}{2}}, p_{\frac{1}{2}}$, то $U_0 - \frac{1}{\rho_0 a_0} P_0 = u_{\frac{1}{2}} - \frac{1}{\rho_0 a_0} p_{\frac{1}{2}}$. Отсюда получаем

$$U_0 = \frac{1}{\rho_0 a_0} P_0 + u_{\frac{1}{2}} - \frac{1}{\rho_0 a_0} p_{\frac{1}{2}}. \quad (8.86)$$

При $i = n$ для нахождения величины U_n при известном P_n используется инвариант Римана $u + \frac{1}{\rho_0 a_0} p = I^+$ вдоль семейства C^+ (рис. 8.25). Поскольку на последнем промежутке $(x_{n-1}, 1)$ заданы значения $u_{\frac{n-1}{2}}, p_{\frac{n-1}{2}}$, то $U_n + \frac{1}{\rho_0 a_0} P_n = u_{\frac{n-1}{2}} + \frac{1}{\rho_0 a_0} p_{\frac{n-1}{2}}$. Отсюда получаем

$$U_n = -\frac{1}{\rho_0 a_0} P_n + u_{\frac{n-1}{2}} + \frac{1}{\rho_0 a_0} p_{\frac{n-1}{2}}. \quad (8.87)$$

На рисунке 8.25 показаны две крайние ячейки с тремя характеристика- ми, одна из которых (верхняя) ограничивает область влияния начальных данных, заданных на промежутках (x_0, x_1) и (x_{n-1}, x_n) соответственно.

6. Переидем к задаче расчета искомых величин на следующем временном слое. Для этого рассмотрим одну ячейку (рис. 8.26), для которой известны

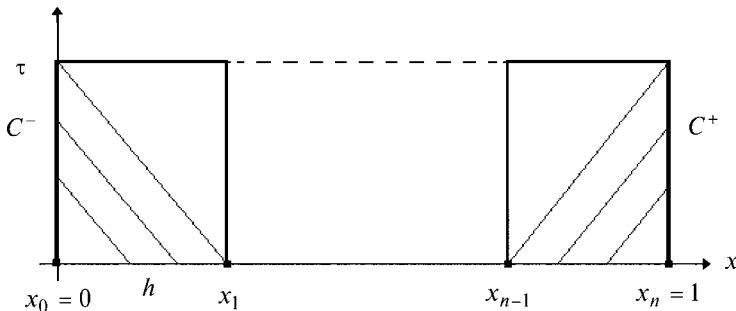


Рис. 8.25

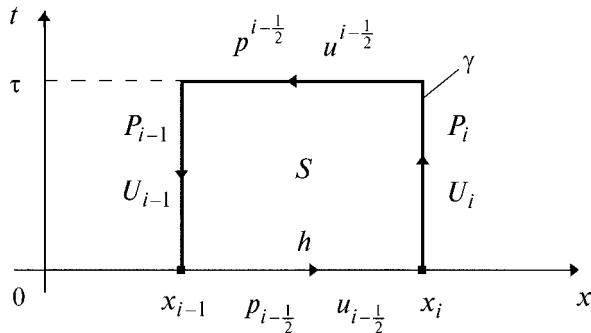


Рис. 8.26

значения приближенного решения $u_{i-\frac{1}{2}}, p_{i-\frac{1}{2}}$ на нижней границе и U_{i-1}, P_{i-1} , U_i, P_i , $i = 1, \dots, n$, на левой и правой вертикальных границах (см. п. 4). Требуется определить значения $p^{i-\frac{1}{2}}, u^{i-\frac{1}{2}}$ на верхней границе ячейки.

Для решения поставленной задачи перепишем уравнения (8.83) в так называемом дивергентном виде:

$$\frac{\partial u(x, t)}{\partial t} + \frac{\partial \left(\frac{p(x, t)}{\rho_0} \right)}{\partial x} = 0, \quad \frac{\partial p(x, t)}{\partial t} + \frac{\partial (\rho_0 a_0^2 u(x, t))}{\partial x} = 0. \quad (8.88)$$

Запишем двойной интеграл от левой и правой частей уравнений по поверхности ячейки. Для его нахождения применим формулу Грина перехода от двойного интеграла к криволинейному интегралу второго рода:

$$\iint_S \left(\frac{\partial \Phi_1}{\partial t} + \frac{\partial \Phi_2}{\partial x} \right) dx dt = \oint_{\gamma} (\Phi_1 dx - \Phi_2 dt),$$

где S — поверхность ячейки, а γ — ее граница, обходимая против часовой стрелки. Таким образом, вместо системы (8.88) получается система интегральных уравнений

$$\oint_{\gamma} u dx - \frac{p}{\rho_0} dt = 0, \quad \oint_{\gamma} p dx - \rho_0 a_0^2 u dt = 0.$$

Проводя интегрирование, имеем

$$u_{i-\frac{1}{2}} h - \frac{P_i}{\rho_0} \tau - u^{i-\frac{1}{2}} h + \frac{P_{i-1}}{\rho_0} \tau = 0; \\ p_{i-\frac{1}{2}} h - \rho_0 a_0^2 U_i \tau - p^{i-\frac{1}{2}} h + \rho_0 a_0^2 U_{i-1} \tau = 0.$$

Из этих уравнений получаем

$$u^{i-\frac{1}{2}} = u_{i-\frac{1}{2}} - \frac{\tau}{h \rho_0} (P_i - P_{i-1}); \\ p^{i-\frac{1}{2}} = p_{i-\frac{1}{2}} - \frac{\tau}{h} \rho_0 a_0^2 (U_i - U_{i-1}), \quad i = 1, \dots, n. \quad (8.89)$$

Таким образом, получены формулы для расчета первого временного слоя ($t = \tau$). После нахождения значений $u^{\frac{1}{2}}, p^{\frac{1}{2}}, u^{\frac{3}{2}}, p^{\frac{3}{2}}, \dots, u^{\frac{n-1}{2}}, p^{\frac{n-1}{2}}$ из условия $\tau < \frac{h}{a_0}$ выбирается очередной шаг и формируется новый временной слой ($t_2 = t_1 + \tau$). Найденные значения используются в качестве опорных (они переобозначаются путем присвоения им соответствующих нижних индексов) для нахождения приближенного решения на следующем слое по формулам (8.85)–(8.87), (8.89). Заметим, что для вычисления P_0, P_n вместо (8.82) используются формулы интегрального осреднения, аналогичные (8.81). Для расчета, например, второго временного слоя они имеют вид

$$P_0 = \frac{1}{\tau} \int_{t_1}^{t_2} \varphi_1(t) dt, \quad P_n = \frac{1}{\tau} \int_{t_1}^{t_2} \varphi_2(t) dt. \quad (8.90)$$

Изложенная схема метода Годунова имеет первый порядок аппроксимации по τ и h [7], [50].

Замечание. Из процедуры интегрирования следует, что искомые значения $u^{i-\frac{1}{2}}$, $p^{i-\frac{1}{2}}$ соответствуют среднеинтегральным величинам

$$u^{i-\frac{1}{2}} = \frac{1}{h} \int_{x_{i-1}}^{x_i} u(x, \tau) dx, \quad p^{i-\frac{1}{2}} = \frac{1}{h} \int_{x_{i-1}}^{x_i} p(x, \tau) dx.$$

Приведем пример расчета по описанной схеме для частного случая рассматриваемой задачи:

$$\begin{aligned} \frac{\partial u(x, t)}{\partial t} + \frac{1}{\rho_0} \frac{\partial p(x, t)}{\partial x} &= 0, \quad \frac{\partial p(x, t)}{\partial t} + \rho_0 a_0^2 \frac{\partial u(x, t)}{\partial x} = 0, \quad 0 < x < 1, \quad 0 < t < 1; \\ u(x, 0) &= \psi_1(x) = \sin x, \quad p(x, 0) = \psi_2(x) = \rho_0 x \sin x, \quad 0 \leq x \leq 1; \\ p(0, t) &= \varphi_1(t) = 0,1t + \rho_0 a_0^2 \sin t, \quad p(1, t) = \varphi_2(t) = 0,1t + \rho_0 a_0^2 \sin t, \quad 0 < t \leq 1, \end{aligned}$$

где $\rho_0 = 0,123$; $a_0 = 0,25$.

А. Выберем число промежутков $n = 2$.

Тогда находится шаг $h = \frac{1}{2} = 0,5$ и вводятся узлы $x_0 = 0$; $x_1 = 0,5$; $x_2 = 1$. Зададим шаг $\tau = 0,5$ по времени из условия $\tau < \frac{h}{a_0} = \frac{0,5}{0,25} = 2$. Сформируем две ячейки «A» и «B» первого временного слоя $t_1 = \tau = 0,5$ (рис. 8.27).

Б. Заменим функции $\psi_1(x) = \sin x$, $\psi_2(x) = \rho_0 x \sin x$ кусочно-постоянными по формулам (8.81). Для вычисления интегралов используются квадратурные формулы трапеций (см. п. 5.3.1):

$$u_{i-\frac{1}{2}} = \frac{1}{h} \cdot \frac{u_{i-1} + u_i}{2} h = \frac{u_{i-1} + u_i}{2}, \quad p_{i-\frac{1}{2}} = \frac{1}{h} \cdot \frac{p_{i-1} + p_i}{2} h = \frac{p_{i-1} + p_i}{2}, \quad i = 1, 2,$$

где $u_i = \sin x_i$, $p_i = \rho_0 x_i \sin x_i$. Поэтому

$$u_{\frac{1}{2}} \approx \frac{0 + 0,479}{2} \approx 0,239, \quad u_{\frac{3}{2}} \approx \frac{0,479 + 0,841}{2} \approx 0,660;$$

$$p_{\frac{1}{2}} \approx \frac{0 + 0,0294}{2} \approx 0,0147, \quad p_{\frac{3}{2}} \approx \frac{0,0294 + 0,1035}{2} \approx 0,0664.$$

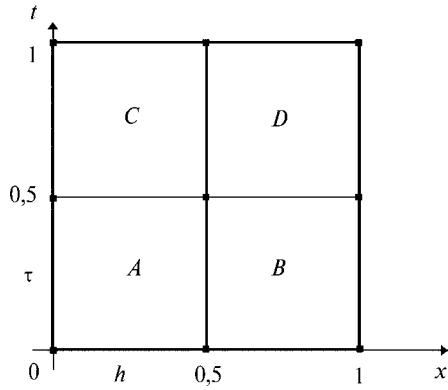


Рис. 8.27

В. Найдем значения P_0 , P_2 по формулам (8.82). Для вычисления интегралов, как и ранее, применяется формула трапеций:

$$P_0 \approx \frac{0+0,0537}{2} \approx 0,0269, \quad P_2 = \frac{0+0,0463}{2} \approx 0,0231.$$

Рассчитаем значение U_0 на левой границе ячейки «А» по формуле (8.86):

$$U_0 = \frac{1}{\rho_0 a_0} P_0 + u_{\frac{1}{2}} - \frac{1}{\rho_0 a_0} p_{\frac{1}{2}} \approx 0,635$$

и значение U_2 на правой границе ячейки «В» по формуле (8.87)

$$U_2 = -\frac{1}{\rho_0 a_0} P_2 + u_{\frac{3}{2}} + \frac{1}{\rho_0 a_0} p_{\frac{3}{2}} \approx 1,476.$$

Г. Вычислим значения U_1 , P_1 на правой границе ячейки «А» (левой границе ячейки «В»). По формулам (8.85) получаем

$$U_1 = \frac{u_{\frac{1}{2}} + u_{\frac{3}{2}}}{2} - \frac{p_{\frac{3}{2}} - p_{\frac{1}{2}}}{2\rho_0 a_0} \approx -0,392, \quad P_1 = \frac{p_{\frac{1}{2}} + p_{\frac{3}{2}}}{2} - \rho_0 a_0 \frac{u_{\frac{3}{2}} - u_{\frac{1}{2}}}{2} \approx 0,0341.$$

Д. Вычислим значения $u^{\frac{1}{2}}$, $p^{\frac{1}{2}}$; $u^{\frac{3}{2}}$, $p^{\frac{3}{2}}$ на верхних границах ячеек:

$$\begin{aligned} u^{\frac{1}{2}} &= u_{\frac{1}{2}} - \frac{\tau}{h\rho_0} (P_1 - P_0) \approx 0,1805, & p^{\frac{1}{2}} &= p_{\frac{1}{2}} - \frac{\tau}{h} \rho_0 a_0^2 (U_1 - U_0) \approx 0,0226; \\ u^{\frac{3}{2}} &= u_{\frac{3}{2}} - \frac{\tau}{h\rho_0} (P_2 - P_1) \approx 0,749, & p^{\frac{3}{2}} &= p_{\frac{3}{2}} - \frac{\tau}{h} \rho_0 a_0^2 (U_2 - U_1) = 0,0521. \end{aligned}$$

Таким образом, первый слой полностью рассчитан.

Е. Для расчета второго слоя ($t_2 = t_1 + \tau$) с шагом $\tau = 0,5$ полученные значения $u^{\frac{1}{2}}$, $p^{\frac{1}{2}}$; $u^{\frac{3}{2}}$, $p^{\frac{3}{2}}$ переобозначим, для чего верхние индексы заменяются нижними: $u_{\frac{1}{2}} = 0,1805$; $p_{\frac{1}{2}} = 0,0226$; $u_{\frac{3}{2}} = 0,749$; $p_{\frac{3}{2}} = 0,0521$. Сформируем две ячейки «С» и «Д» второго временного слоя (см. рис. 8.27).

Ж. Найдем значения P_0 , P_2 по формулам (8.82). Для вычисления интегралов, как и ранее, применяется формула трапеций:

$$P_0 \approx \frac{0,0537 + 0,1065}{2} \approx 0,0801, \quad P_2 = \frac{0,0463 + 0,00935}{2} \approx 0,0278.$$

Рассчитаем значение U_0 на левой границе ячейки «С» по формуле (8.86):

$$U_0 = \frac{1}{\rho_0 a_0} P_0 + u_{\frac{1}{2}} - \frac{1}{\rho_0 a_0} p_{\frac{1}{2}} \approx 2,0504$$

и значение U_2 на правой границе ячейки «Д» по формуле (8.87)

$$U_2 = -\frac{1}{\rho_0 a_0} P_2 + u_{\frac{3}{2}} + \frac{1}{\rho_0 a_0} p_{\frac{3}{2}} \approx 0,8651.$$

3. Вычислим значения U_1 , P_1 на правой границе ячейки «С» (левой границе ячейки «D»). По формулам (8.85) получаем

$$U_1 = \frac{u_{\frac{1}{2}} + u_{\frac{3}{2}}}{2} - \frac{p_{\frac{3}{2}} - p_{\frac{1}{2}}}{2\rho_0 a_0} \approx -0,0150, \quad P_1 = \frac{p_{\frac{1}{2}} + p_{\frac{3}{2}}}{2} - \rho_0 a_0 \frac{u_{\frac{3}{2}} - u_{\frac{1}{2}}}{2} \approx 0,02861.$$

И. Вычислим значения $u^{\frac{1}{2}}, p^{\frac{1}{2}}$; $u^{\frac{3}{2}}, p^{\frac{3}{2}}$ на верхних границах ячеек «С» и «D»:

$$u^{\frac{1}{2}} = u_{\frac{1}{2}} - \frac{\tau}{h\rho_0}(P_1 - P_0) \approx 0,599, \quad p^{\frac{1}{2}} = p_{\frac{1}{2}} - \frac{\tau}{h}\rho_0 a_0^2(U_1 - U_0) \approx 0,0385,$$

$$u^{\frac{3}{2}} = u_{\frac{3}{2}} - \frac{\tau}{h\rho_0}(P_2 - P_1) \approx 0,7556, \quad p^{\frac{3}{2}} = p_{\frac{3}{2}} - \frac{\tau}{h}\rho_0 a_0^2(U_2 - U_1) \approx 0,04533.$$

Таким образом, второй слой полностью рассчитан и найдено приближенное решение поставленной задачи. ■

Замечание. Здесь рассмотрен пример методики конструирования метода С. К. Годунова и ее реализации для относительно простой линейной задачи. В нелинейном случае метод существенно усложняется. Для более глубокого изучения проблемы авторы рекомендуют [22], [36], [44], где данный метод применяется для различных классов нелинейных задач и задач с тремя и четырьмя независимыми переменными.

Задачи для самостоятельного решения

Составить явную и неявную схемы для решения задач 1–4.

1. $u_t = u_{xx}$, $0 < x < 1$, $0 < t < 0,08$;

$$u(0, t) = t, \quad u(1, t) = 0,5 + t, \quad 0 < t \leq 0,08;$$

$$u(x, 0) = \frac{x^2}{2}, \quad 0 \leq x \leq 1.$$

2. $u_t = u_{xx} + x^2 - 2t$, $0 < x < 1$, $0 < t < 0,02$;

$$u(0, t) = 0, \quad u(1, t) = t, \quad 0 < t \leq 0,02;$$

$$u(x, 0) = 0, \quad 0 \leq x \leq 1.$$

3. $u_{tt} = u_{xx}$, $0 < x < 1$, $0 < t < 1$;

$$u(0, t) = 0, \quad u(1, t) = 0, \quad 0 < t \leq 1;$$

$$u(x, 0) = 0, \quad u_t(x, 0) = \pi \sin \pi x, \quad 0 \leq x \leq 1.$$

4. $u_{tt} = u_{xx} + e^{xt}$, $0 < x < 1$, $0 < t < 0,5$;

$$u(0, t) = \operatorname{arctg} t, \quad u(1, t) = 1, \quad 0 < t \leq 1;$$

$$u(x, 0) = x, \quad u_t(x, 0) = \sqrt{x}, \quad 0 \leq x \leq 1.$$

5. Составить явную схему решения задачи.

$$\begin{aligned} u_{xx} + u_{yy} &= 0, \quad 0 < x < 0,8, \quad 0 < y < 0,8; \\ u(x, 0) &= x^2, \quad u(x, 0,8) = x^2 - 0,64, \quad 0 \leq x \leq 0,8; \\ u(0, y) &= -y^2, \quad u(0,8, y) = 0,64 - y^2, \quad 0 \leq y \leq 0,8. \end{aligned}$$

6. $u_{xx} + u_{yy} = 4$, $0 < x < 0,8$, $0 < y < 0,8$;

$$\begin{aligned} u(x, 0) &= x^2, \quad u(x, 0,8) = x^2 + 0,64, \quad 0 \leq x \leq 0,8; \\ u(0, y) &= y^2, \quad u(0,8, y) = 0,64 + y^2, \quad 0 \leq y \leq 0,8. \end{aligned}$$

7. Методом сеток найти решение задачи

$$\begin{aligned} u_{tt} &= u_{xx}, \quad 0 < x < 1, \quad 0 < t < 0,5; \\ u(0, t) &= 0, \quad u(1, t) = 0, \quad 0 < t \leq 1; \\ u(x, 0) &= 0,2x(1 - x)\sin\pi x, \quad u_t(x, 0) = 0, \quad 0 \leq x \leq 1. \end{aligned}$$

8. Методом характеристик решить задачу Коши

$$\begin{aligned} u_{tt} &= u_{xx}, \quad 0 < x < \pi, \quad 0 < t < 0,5; \\ u(x, 0) &= x(\pi - x), \quad u_t(x, 0) = 0, \quad 0 \leq x \leq 1. \end{aligned}$$

ГЛАВА
9**ЧИСЛЕННЫЕ МЕТОДЫ
РЕШЕНИЯ УРАВНЕНИЙ
МАТЕМАТИЧЕСКОЙ ФИЗИКИ
С ТРЕМЯ НЕЗАВИСИМЫМИ
ПЕРЕМЕННЫМИ****9.1. ПОСТАНОВКА ЗАДАЧИ
И ОСНОВНЫЕ ПОЛОЖЕНИЯ**

Рассмотрим задачи с тремя независимыми переменными x , y , t , где x , y — пространственные переменные, а t — время, $t = 0$ и $t = T$ — моменты начала и окончания процесса. В качестве примера расчетного множества, в котором ищется решение задачи, используем цилиндр (рис. 9.1). В начальный момент времени $t = 0$ множество задано односвязной областью Ω с кусочно-гладкой границей Γ .

Приведем обобщения задач, поставленных в п. 8.1, на случай трех независимых переменных.

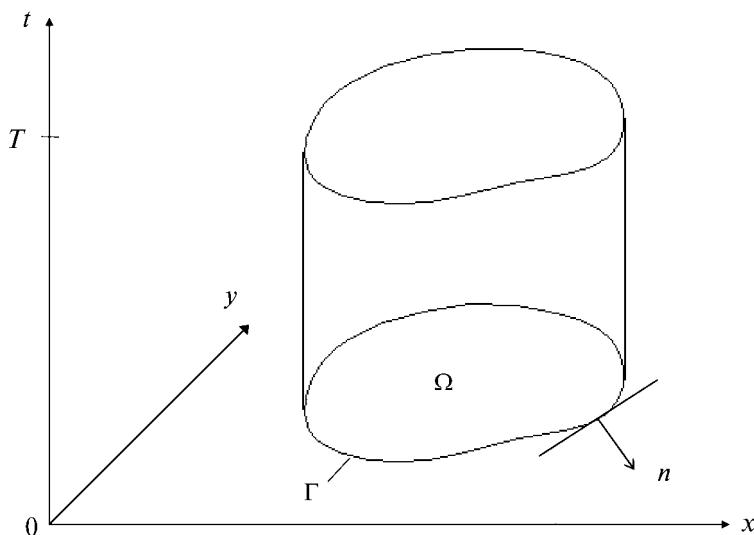


Рис. 9.1

ПОСТАНОВКА ЗАДАЧИ ДЛЯ ДВУМЕРНОГО УРАВНЕНИЯ ПЕРЕНОСА

Рассмотрим *двуумерное уравнение переноса*

$$\frac{\partial u(x, y, t)}{\partial t} + c_1 \frac{\partial u(x, y, t)}{\partial x} + c_2 \frac{\partial u(x, y, t)}{\partial y} = f(x, y, t), \quad (x, y, t) \in \Omega \times (0, T), \quad (9.1)$$

где c_1, c_2 — скорости переноса по осям Ox, Oy ; $f(x, y, t)$ — заданная функция. Оно является модельным для математического описания многих физических процессов (нейтронных потоков в реакторе; теплопроводности в газах, обусловленной диффузией атомов и электронов, и др.).

Начально-краевая задача:

$$\begin{aligned} u_t + u_x + u_y &= f, \quad (x, y, t) \in \Omega \times (0, T); \\ u(x, y, 0) &= \psi(x, y), \quad (x, y) \in \Omega \cup \Gamma \text{ (начальное условие);} \\ u(x, y, t) &= \phi(x, y, t), \quad (x, y) \in \Gamma, \quad 0 < t \leq T \text{ (краевое условие)} \end{aligned}$$

содержит функциональное начальное условие (при $t = 0$) и функциональное краевое условие. Здесь и далее $\psi(x, y), \phi(x, y, t)$ — заданные функции.

Требуется найти функцию $u(x, y, t)$, удовлетворяющую дифференциальному уравнению (9.1), начальному и краевому условиям.

ПОСТАНОВКИ ЗАДАЧ ДЛЯ ДВУМЕРНОГО УРАВНЕНИЯ ТЕПЛОПРОВОДНОСТИ

Рассмотрим *двуумерное уравнение теплопроводности*

$$\frac{\partial u(x, y, t)}{\partial t} = a^2 \left(\frac{\partial^2 u(x, y, t)}{\partial x^2} + \frac{\partial^2 u(x, y, t)}{\partial y^2} \right), \quad (x, y, t) \in \Omega \times (0, T), \quad (9.2)$$

где u — температура; a^2 — коэффициент температуропроводности, $a^2 = \frac{k}{\rho c}$, k — коэффициент теплопроводности, c — удельная теплоемкость, ρ — плотность. Оно описывает, например, распределение температуры на плоской пластине при отсутствии тепловых источников.

Первая начально-краевая задача:

$$\begin{aligned} u_t &= a^2(u_{xx} + u_{yy}), \quad (x, y, t) \in \Omega \times (0, T); \\ u(x, y, 0) &= \psi(x, y), \quad (x, y) \in \Omega \cup \Gamma \text{ (начальное условие);} \\ u(x, y, t) &= \phi(x, y, t), \quad (x, y) \in \Gamma, \quad 0 < t \leq T \text{ (краевое условие)} \end{aligned}$$

содержит функциональное начальное условие (при $t = 0$) и функциональное краевое условие на границе Γ . Здесь и далее $\psi(x, y), \phi(x, y, t)$ — заданные функции.

Вторая начально-краевая задача:

$$\begin{aligned} u_t &= a^2(u_{xx} + u_{yy}), \quad (x, y, t) \in \Omega \times (0, T); \\ u(x, y, 0) &= \psi(x, y), \quad (x, y) \in \Omega \cup \Gamma \text{ (начальное условие);} \end{aligned}$$

$$\frac{\partial u(x, y, t)}{\partial n} = \varphi(x, y, t), \quad (x, y) \in \Gamma, \quad 0 < t \leq T \quad (\text{краевое условие})$$

содержит функциональное начальное условие (при $t = 0$) и дифференциальное краевое условие на границе Γ , задающее производную в направлении внешней нормали.

Третья начально-краевая задача:

$$u_t = a^2(u_{xx} + u_{yy}), \quad (x, y, t) \in \Omega \times (0, T); \\ u(x, y, 0) = \psi(x, y), \quad (x, y) \in \Omega \cup \Gamma \quad (\text{начальное условие});$$

$$\alpha \frac{\partial u(x, y, t)}{\partial n} + \beta u = \varphi(x, y, t), \quad (x, y) \in \Gamma, \quad 0 < t \leq T \quad (\text{краевое условие})$$

содержит функциональное начальное условие (при $t = 0$) и функционально-дифференциальное краевое условие, задающее линейную комбинацию производной в направлении внешней нормали и функции на границе Γ . Здесь α , β — заданные числа ($\alpha^2 + \beta^2 > 0$) либо функции от t в более общем случае [26].

Во всех перечисленных задачах требуется найти функцию $u(x, y, t)$, которая удовлетворяет дифференциальному уравнению (9.2) в области $D = \Omega \times (0, T)$ и соответствующим условиям на ее границе.

Примером физической задачи, приводящей к первой начально-краевой задаче, может быть процесс изменения распределения температуры в плоской пластине, на границе которой поддерживается заданная температура, определенная функцией $\varphi(x, y, t)$. В начальный момент времени при $t = 0$ функцией $\psi(x, y)$ задано начальное распределение температуры на поверхности пластины. Искомая функция $u(x, y, t)$ характеризует температуру пластины в точке с координатами (x, y) в некоторый момент времени t .

ПОСТАНОВКИ ЗАДАЧ ДЛЯ ДВУМЕРНОГО ВОЛНОВОГО УРАВНЕНИЯ

Рассмотрим *двумерное волновое уравнение*:

$$\frac{\partial^2 u(x, y, t)}{\partial t^2} = c^2 \left(\frac{\partial^2 u(x, y, t)}{\partial x^2} + \frac{\partial^2 u(x, y, t)}{\partial y^2} \right), \quad (x, y, t) \in \Omega \times (0, T), \quad (9.3)$$

где $u(x, y, t)$ — величина смещения точки (x, y) в момент времени t ; $c^2 = \frac{\tilde{T}}{\rho}$, \tilde{T} — проекция вектора натяжения на плоскость Oxy , ρ — поверхностная плотность мембранны. Волновое уравнение описывает малые поперечные колебания мембранны.

Первая начально-краевая задача:

$$u_{tt} = c^2(u_{xx} + u_{yy}), \quad (x, y, t) \in \Omega \times (0, T); \\ u(x, y, 0) = \psi_1(x, y), \quad (x, y) \in \Omega \cup \Gamma \quad (\text{начальное условие}); \\ \frac{\partial u(x, y, 0)}{\partial t} = \psi_2(x, y), \quad (x, y) \in \Omega \cup \Gamma \quad (\text{начальное условие}); \\ u(x, y, t) = \varphi(x, y, t), \quad (x, y) \in \Gamma, \quad 0 < t \leq T \quad (\text{краевое условие})$$

содержит функциональное и дифференциальное начальные условия (при $t = 0$) и функциональное краевое условие. Здесь и далее $\psi_1(x, y)$, $\psi_2(x, y)$, $\phi(x, y, t)$ — заданные функции.

Вторая начально-краевая задача:

$$\begin{aligned} u_{tt} &= c^2(u_{xx} + u_{yy}), \quad (x, y, t) \in \Omega \times (0, T); \\ u(x, y, 0) &= \psi_1(x, y), \quad (x, y) \in \Omega \cup \Gamma \text{ (начальное условие);} \\ \frac{\partial u(x, y, 0)}{\partial t} &= \psi_2(x, y), \quad (x, y) \in \Omega \cup \Gamma \text{ (начальное условие);} \\ \frac{\partial u(x, y, t)}{\partial n} &= \phi(x, y, t), \quad (x, y) \in \Gamma, \quad 0 < t \leq T \text{ (краевое условие)} \end{aligned}$$

содержит функциональное и дифференциальное начальные условия (при $t = 0$) и дифференциальное краевое условие на границе Γ , задающее производную в направлении внешней нормали.

Третья начально-краевая задача:

$$\begin{aligned} u_{tt} &= c^2(u_{xx} + u_{yy}), \quad (x, y, t) \in \Omega \times (0, T); \\ u(x, y, 0) &= \psi_1(x, y), \quad (x, y) \in \Omega \cup \Gamma \text{ (начальное условие);} \\ \frac{\partial u(x, y, 0)}{\partial t} &= \psi_2(x, y), \quad (x, y) \in \Omega \cup \Gamma \text{ (начальное условие);} \\ \alpha \frac{\partial u(x, y, t)}{\partial n} + \beta u &= \phi(x, y, t), \quad (x, y) \in \Gamma, \quad 0 < t \leq T \text{ (краевое условие)} \end{aligned}$$

содержит функциональное и дифференциальное начальные условия (при $t = 0$) и функционально-дифференциальное краевое условие, задающее линейную комбинацию производной в направлении внешней нормали и функции на границе Γ . Здесь α, β — заданные числа ($\alpha^2 + \beta^2 > 0$) либо функции от t более общем случае [26].

Во всех перечисленных задачах требуется найти функцию $u(x, y, t)$, которая удовлетворяет дифференциальному уравнению (9.3) в области $D = \Omega \times (0, T)$ и соответствующим условиям на ее границе.

ПОСТАНОВКИ ЗАДАЧ ДЛЯ ТРЕХМЕРНОГО УРАВНЕНИЯ ЛАПЛАСА

Рассмотрим *трехмерное уравнение Лапласа*

$$\frac{\partial^2 u(x, y, z)}{\partial x^2} + \frac{\partial^2 u(x, y, z)}{\partial y^2} + \frac{\partial^2 u(x, y, z)}{\partial z^2} = 0, \quad (x, y, z) \in \Omega, \quad (9.4)$$

или более общее *уравнение Пуассона*

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

где Ω — некоторая односвязная область пространства $Oxyz$ с кусочно-гладкой границей Γ ; $f(x, y, z)$ — заданная функция.

Поскольку в уравнении (9.4) отсутствуют производные по времени, задача является стационарной, следовательно, начальные условия в ней не задаются. На границе Γ области Ω могут задаваться краевые (граничные) условия трех видов.

Первая краевая задача (задача Дирихле):

$$\begin{aligned} u_{xx} + u_{yy} + u_{zz} &= 0, \quad (x, y, z) \in \Omega; \\ u|_{\Gamma} &= \varphi(x, y, z), \quad (x, y, z) \in \Gamma \text{ (краевое условие)} \end{aligned}$$

содержит функциональное краевое условие на границе области, $\varphi(x, y, z)$ — заданная функция.

Вторая краевая задача (задача Неймана):

$$\begin{aligned} u_{xx} + u_{yy} + u_{zz} &= 0, \quad (x, y, z) \in \Omega; \\ \frac{\partial u}{\partial n}|_{\Gamma} &= \varphi(x, y, z), \quad (x, y, z) \in \Gamma \text{ (краевое условие)} \end{aligned}$$

содержит дифференциальное краевое условие, в котором задается производная в направлении внешней нормали на границе Γ области Ω .

Третья краевая задача:

$$\begin{aligned} u_{xx} + u_{yy} + u_{zz} &= 0, \quad (x, y, z) \in \Omega; \\ \alpha \frac{\partial u}{\partial n}|_{\Gamma} + \beta u|_{\Gamma} &= \varphi(x, y, z), \quad (x, y, z) \in \Gamma \text{ (краевое условие)} \end{aligned}$$

содержит функционально-дифференциальное краевое условие, в котором задается линейная комбинация производной в направлении внешней нормали и функции $u(x, y)$ на границе Γ области Ω .

Во всех перечисленных задачах требуется найти функцию $u(x, y)$, которая удовлетворяет дифференциальному уравнению (9.5) внутри области Ω и соответствующим условиям на ее границе.

В качестве примера задачи Дирихле можно привести задачу о нахождении стационарного распределения температуры внутри объемной области, если задана температура на ее границе.

9.2. РАЗНОСТНЫЙ МЕТОД

Принципы построения разностных схем изложены в п. 8.2. Для простоты изложения будем предполагать, что в двумерных начально-краевых задачах для уравнения переноса, уравнения теплопроводности и волнового уравнения множество D представляет собой параллелепипед длины L , ширины M и высоты T , ограниченный отрезками прямых, параллельных осям Ox , Oy и Ot в пространстве трех независимых переменных x , y и t (см. рис. 9.2). Поэтому зададим сетку $D_h = \{x_i, y_j, t^n\}$, где $x_i = ih_x$, $0 \leq i \leq I$; $y_j = jh_y$, $0 \leq j \leq J$; $t^n = n\tau$, $0 \leq n \leq N$; N, J, I — целые положительные числа; h_x , h_y , τ — величины шагов по пространственным координатам и времени (для простоты принимаются постоянными); $Ih_x = L$, $Jh_y = M$, $N\tau = T$. Такая сетка называется *равномерной (регулярной)*.

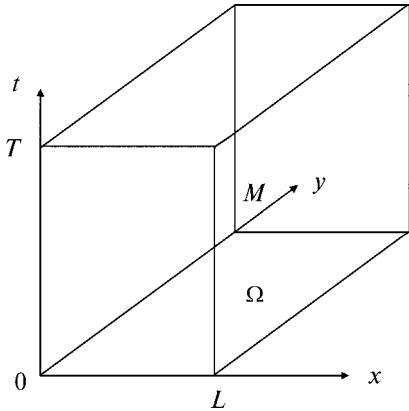


Рис. 9.2

Введем сеточную функцию

$$u_h = \begin{cases} u_{i,j}^n = u(x_i, y_j, t^n), \\ 0 \leq i \leq I, 0 \leq j \leq J, 0 \leq n \leq N \end{cases},$$

которая является *сеточным представлением решения исходной (дифференциальной) задачи* или точным решением дифференциальной задачи в узлах сетки. Как правило, вычислить u_h не удается, поэтому находят другую сеточную функцию $\hat{u}_h \approx u_h$, приближенно совпадающую с точным решением в узлах сетки. Она ищется как решение *разностной схемы*. Под *разностной схемой*, как и ранее, понимается совокупность разностных уравнений, аппроксимирующих основное дифференциальное уравнение во всех внутренних узлах сетки и дополнительные условия (начальные и краевые) — в граничных узлах.

Для аппроксимации частных производных можно использовать естественные обобщения формул (8.20)–(8.22), (8.26)–(8.32):

$$\begin{aligned} \hat{u}_t(x_i, y_j, t^n) &= \frac{u_{i,j}^{n+1} - u_{i,j}^n}{\tau}, \quad \hat{u}_t(x_i, y_j, t^n) = \frac{u_{i,j}^n - u_{i,j}^{n-1}}{\tau}, \\ \hat{u}_t(x_i, y_j, t^n) &= \frac{u_{i,j}^{n+1} - u_{i,j}^{n-1}}{2\tau}, \quad \hat{u}_{tt}(x_i, y_j, t^n) = \frac{u_{i,j}^{n+1} - 2u_{i,j}^n + u_{i,j}^{n-1}}{\tau^2}; \end{aligned} \quad (9.5)$$

$$\begin{aligned} \hat{u}_x(x_i, y_j, t^n) &= \frac{u_{i+1,j}^n - u_{i,j}^n}{h_x}, \quad \hat{u}_x(x_i, y_j, t^n) = \frac{u_{i,j}^n - u_{i-1,j}^n}{h_x}, \\ \hat{u}_x(x_i, y_j, t^n) &= \frac{u_{i+1,j}^n - u_{i-1,j}^n}{2h_x}, \quad \hat{u}_{xx}(x_i, y_j, t^n) = \frac{u_{i+1,j}^n - 2u_{i,j}^n + u_{i-1,j}^n}{h_x^2}; \end{aligned} \quad (9.6)$$

$$\begin{aligned} \hat{u}_y(x_i, y_j, t^n) &= \frac{u_{i,j+1}^n - u_{i,j}^n}{h_y}, \quad \hat{u}_y(x_i, y_j, t^n) = \frac{u_{i,j}^n - u_{i,j-1}^n}{h_y}, \\ \hat{u}_y(x_i, y_j, t^n) &= \frac{u_{i,j+1}^n - u_{i,j-1}^n}{2h_y}, \quad \hat{u}_{yy}(x_i, y_j, t^n) = \frac{u_{i,j+1}^n - 2u_{i,j}^n + u_{i,j-1}^n}{h_y^2}. \end{aligned} \quad (9.7)$$

Заметим, что приведенные в (9.5)–(9.7) операторы, в знаменателе которых находятся τ , h_x , h_y , имеют первый порядок аппроксимации по соответствующим шагам, а операторы с τ^2 , h_x^2 , h_y^2 , 2τ , $2h_x$, $2h_y$ в знаменателе — второй порядок.

Пример 9.1. Построить разностную схему для задачи:

$$\begin{aligned} u_t + u_x + u_y &= f(x, y, t), \quad (x, y, t) \in (0, L) \times (0, M) \times (0, T); \\ u(x, y, 0) &= \psi(x, y), \quad (x, y) \in [0, L] \times [0, M] \text{ (начальное условие);} \end{aligned}$$

$$\begin{aligned} u(0, y, t) &= \varphi(y, t), \quad y \in [0, M], \quad 0 < t \leq T \text{ (краевое условие);} \\ u(x, 0, t) &= \Phi(x, t), \quad x \in [0, L], \quad 0 < t \leq T \text{ (краевое условие)}, \end{aligned}$$

где $f(x, y, t)$, $\psi(x, y)$, $\varphi(y, t)$, $\Phi(x, t)$ — заданные функции.

□Решается первая начально-краевая задача для двумерного уравнения переноса (9.1), где $c_1 = c_2 = 1$. Построим аналог абсолютно устойчивой схемы (8.40). Для этого используем сетку D_h и четырехточечный шаблон (рис. 9.3).

Для аппроксимации производной u_t используем первую формулу в (9.5) и вторые формулы в (9.6) и (9.7), записанные на $(n+1)$ -м временном слое:

$$\frac{\hat{u}_{i,j}^{n+1} - \hat{u}_{i,j}^n}{\tau} + \frac{\hat{u}_{i,j}^{n+1} - \hat{u}_{i-1,j}^{n+1}}{h_x} + \frac{\hat{u}_{i,j}^{n+1} - \hat{u}_{i,j-1}^{n+1}}{h_y} = f_{i,j}^n, \quad 1 \leq i \leq I-1, \quad 1 \leq j \leq J-1, \quad 1 \leq n \leq N-1, \quad (9.8)$$

где $f_{i,j}^n = f(x_i, y_j, t^n)$ — сеточное представление правой части.

Начальные и краевые условия представляются в виде

$$\begin{aligned} \hat{u}_{i,j}^0 &= \psi_{i,j}, \quad 0 \leq i \leq I, \quad 0 \leq j \leq J; \\ \hat{u}_{0,j}^n &= \varphi_j^n, \quad 0 \leq j \leq J, \quad 1 \leq n \leq N; \\ \hat{u}_{i,0}^n &= \Phi_i^n, \quad 0 \leq i \leq I, \quad 1 \leq n \leq N, \end{aligned} \quad (9.9)$$

где $\varphi_j^n = \varphi(y_j, t^n)$, $\Phi_i^n = \Phi(x_i, t^n)$, $\psi_{i,j} = \psi(x_i, y_j)$.

Совокупность соотношений (9.8), (9.9) составляет *неявную двухслойную* разностную схему. Можно показать, что она абсолютно устойчивая и имеет первый порядок аппроксимации. Следовательно, она сходится с первым порядком точности [14]. Последовательность вычислений по разностной схеме может быть различной. Достаточно, чтобы она соответствовала какому-либо порядку заполнения первого координатного угла в пространстве (x, y, t) ячейками, при котором новая ячейка прикладывается тремя гранями к ранее уложенным ячейкам или координатным плоскостям. ■

Пример 9.2. Построить разностные схемы для решения первой начально-краевой задачи для двумерного уравнения теплопроводности:

$$\begin{aligned} u_t &= u_{xx} + u_{yy}, \quad (x, y, t) \in (0, L) \times (0, M) \times (0, T); \\ u(x, y, 0) &= \psi(x, y), \quad (x, y) \in [0, L] \times [0, M] \text{ (начальное условие);} \\ u(x, y, t) &= \varphi(x, y, t), \quad (x, y) \in \Gamma, \quad 0 < t \leq T \text{ (краевое условие)}, \end{aligned}$$

где $\Omega = (0, L) \times (0, M)$; Γ — граница области Ω ; $\psi(x, y)$, $\varphi(y, t)$ — заданные функции; L, M, T — заданные числа.

□Для решения задачи зададим сетку D_h . Сначала построим *явную схему*. Для аппроксимации уравнения используем шеститочечный шаблон (см. рис. 9.4a).

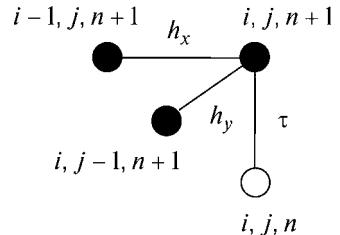


Рис. 9.3

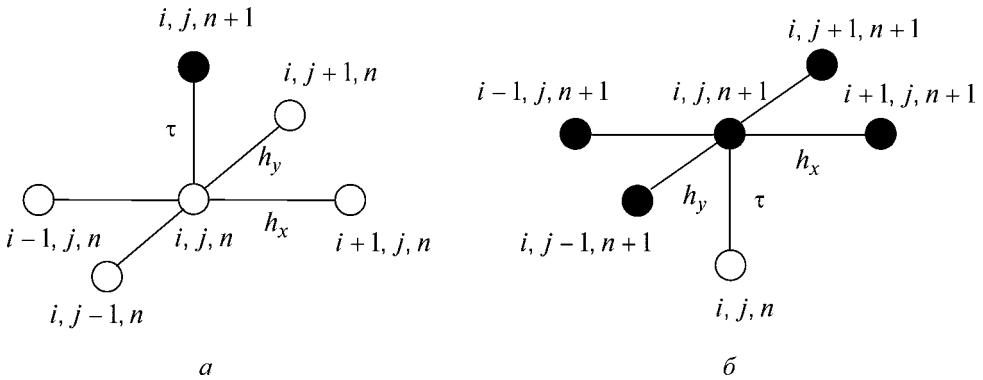


Рис. 9.4

Для аппроксимации производных используем первую формулу в (9.5) и четвертые формулы в (9.6), (9.7):

$$\frac{\hat{u}_{i,j}^{n+1} - \hat{u}_{i,j}^n}{\tau} = \frac{\hat{u}_{i+1,j}^n - 2\hat{u}_{i,j}^n + \hat{u}_{i-1,j}^n}{h_x^2} + \frac{\hat{u}_{i,j+1}^n - 2\hat{u}_{i,j}^n + \hat{u}_{i,j-1}^n}{h_y^2}, \quad (9.10)$$

$1 \leq i \leq I-1, \quad 1 \leq j \leq J-1, \quad 1 \leq n \leq N-1,$

Начальные и краевые условия представляются в виде

$$\begin{aligned} \hat{u}_{i,j}^0 &= \psi_{i,j}, \quad 0 \leq i \leq I, \quad 0 \leq j \leq J; \\ \hat{u}_{0,j}^n &= \varphi_{0,j}^n, \quad \hat{u}_{I,j}^n = \varphi_{I,j}^n, \quad 0 \leq j \leq J, \quad 1 \leq n \leq N; \\ \hat{u}_{i,0}^n &= \varphi_{i,0}^n, \quad \hat{u}_{i,J}^n = \varphi_{i,J}^n, \quad 1 \leq i \leq I, \quad 1 \leq n \leq N, \end{aligned} \quad (9.11)$$

где $\varphi_{i,j}^n = \varphi(x_i, y_j, t^n)$, $\psi_{i,j} = \psi(x_i, y_j)$. Начальные условия задают значения искомой функции на нижнем основании параллелепипеда, а краевые условия — на боковых гранях.

Соотношения (9.10), (9.11) образуют явную двухслойную разностную схему. Можно показать, что она условно устойчива (устойчива при выполнении условия $\tau \left(\frac{1}{h_x^2} + \frac{1}{h_y^2} \right) \leq \frac{1}{2}$) [14]. При $h_x = h_y = h$ имеем условие $\frac{\tau}{h^2} \leq \frac{1}{4}$. В этом случае (9.10) принимает вид

$$\begin{aligned} \frac{\hat{u}_{i,j}^{n+1} - \hat{u}_{i,j}^n}{\tau} &= \frac{\hat{u}_{i+1,j}^n - 2\hat{u}_{i,j}^n + \hat{u}_{i-1,j}^n}{h^2} + \frac{\hat{u}_{i,j+1}^n - 2\hat{u}_{i,j}^n + \hat{u}_{i,j-1}^n}{h^2} = \\ &= \frac{\hat{u}_{i+1,j}^n + \hat{u}_{i-1,j}^n + \hat{u}_{i,j-1}^n + \hat{u}_{i,j+1}^n - 4\hat{u}_{i,j}^n}{h^2}, \end{aligned}$$

$1 \leq i \leq I-1, \quad 1 \leq j \leq J-1, \quad 1 \leq n \leq N-1.$

Из этого соотношения следует, что при фиксированных i, j, n искомое значение $\hat{u}_{i,j}^{n+1}$ определяется явным образом через известные значения на n -м слое:

$$\hat{u}_{i,j}^{n+1} = \hat{u}_{i,j}^n + \frac{\tau}{h^2} [\hat{u}_{i+1,j}^n + \hat{u}_{i-1,j}^n + \hat{u}_{i,j-1}^n + \hat{u}_{i,j+1}^n - 4\hat{u}_{i,j}^n].$$

Построим *неявную схему*, применяя шеститочечный шаблон, изображенный на рисунке 9.4б. Для аппроксимации производных используем первую формулу в (9.5) и четвертые в (9.6), (9.7), записанные на $(n+1)$ -м временном слое.

$$\frac{\hat{u}_{i,j}^{n+1} - \hat{u}_{i,j}^n}{\tau} = \frac{\hat{u}_{i+1,j}^{n+1} - 2\hat{u}_{i,j}^{n+1} + \hat{u}_{i-1,j}^{n+1}}{h_x^2} + \frac{\hat{u}_{i,j+1}^{n+1} - 2\hat{u}_{i,j}^{n+1} + \hat{u}_{i,j-1}^{n+1}}{h_y^2}, \quad (9.12)$$

$$1 \leq i \leq I-1, \quad 1 \leq j \leq J-1, \quad 1 \leq n \leq N-1.$$

Соотношения (9.11), (9.12) образуют неявную двухслойную схему. Можно показать, что она является абсолютно устойчивой [35]. Соотношение (9.12), записанное для всех внутренних узлов $(n+1)$ -го слоя, порождает систему линейных алгебраических уравнений для определения значений искомой функции в узлах. Каждое уравнение этой системы содержит только пять неизвестных. Следовательно, матрица системы является пятидиагональной и сильно разреженной. Для решения системы применяются, как правило, итерационные методы и метод прогонки.

Получим еще одну *неявную схему*, используя десятиточечный шаблон, изображенный на рисунке 9.5. Он является обобщением шеститочечного шаблона (см. рис. 8.8б) на двумерный случай.

Аппроксимация уравнения имеет вид

$$\frac{\hat{u}_{i,j}^{n+1} - \hat{u}_{i,j}^n}{\tau} = \lambda \left[\frac{\hat{u}_{i+1,j}^{n+1} - 2\hat{u}_{i,j}^{n+1} + \hat{u}_{i-1,j}^{n+1}}{h_x^2} + \frac{\hat{u}_{i,j+1}^{n+1} - 2\hat{u}_{i,j}^{n+1} + \hat{u}_{i,j-1}^{n+1}}{h_y^2} \right] +$$

$$+ (1-\lambda) \left[\frac{\hat{u}_{i+1,j}^n - 2\hat{u}_{i,j}^n + \hat{u}_{i-1,j}^n}{h_x^2} + \frac{\hat{u}_{i,j+1}^n - 2\hat{u}_{i,j}^n + \hat{u}_{i,j-1}^n}{h_y^2} \right], \quad (9.13)$$

$$1 \leq i \leq I-1, \quad 1 \leq j \leq J-1, \quad 1 \leq n \leq N-1,$$

где $0 \leq \lambda \leq 1$ — параметр схемы. При $\lambda = 0$ схема становится явной и значение решения на $(n+1)$ -м слое непосредственно вычисляется по известным значениям с предыдущего слоя, а при $\lambda = 1$ совпадает с неявной схемой (9.11), (9.12). Начальные и краевые условия записываются в виде (9.11). Соотношения (9.11), (9.13) образуют *неявную двухслойную разностную схему*. Можно доказать, что схема имеет второй порядок аппроксимации по h_x, h_y , а по τ — первый порядок при $\lambda \neq \frac{1}{2}$ и второй при $\lambda = \frac{1}{2}$. Схема является условно устойчивой (устойчивой при $\lambda \geq \frac{1}{2} - \frac{1}{4\tau} \left(\frac{1}{h_x^2} + \frac{1}{h_y^2} \right)^{-1}$) [14]. ■

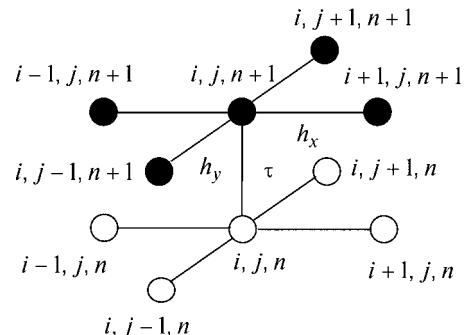


Рис. 9.5

Пример 9.3. Построить разностные схемы для первой начально-краевой задачи для двумерного волнового уравнения:

$$u_{tt} = c^2(u_{xx} + u_{yy}), \quad (x, y, t) \in \Omega \times (0, T);$$

$$u(x, y, 0) = \psi_1(x, y), \quad (x, y) \in \Omega \cup \Gamma \text{ (начальное условие);}$$

$$\frac{\partial u(x, y, 0)}{\partial t} = \psi_2(x, y), \quad (x, y) \in \Omega \cup \Gamma \text{ (начальное условие);}$$

$$u(x, y, t) = \varphi(x, y, t), \quad (x, y) \in \Gamma, \quad 0 < t \leq T \text{ (краевое условие),}$$

где $\Omega = (0, L) \times (0, M)$; Γ — граница области Ω ; $\psi_1(x, y)$, $\psi_2(x, y)$, $\varphi(x, y, t)$ — заданные функции.

□ Сначала построим явную схему. Для этого воспользуемся шаблоном, изображенным на рисунке 9.6a, обобщаяющим шаблон «крест», применяемый для решения одномерного волнового уравнения (см. п. 8.3.2). Для аппроксимации частных производных будем использовать четвертые формулы в (9.5)–(9.7):

$$\frac{u_{i,j}^{n+1} - 2u_{i,j}^n + u_{i,j}^{n-1}}{\tau^2} = \frac{\hat{u}_{i+1,j}^n - 2\hat{u}_{i,j}^n + \hat{u}_{i-1,j}^n}{h_x^2} + \frac{\hat{u}_{i,j+1}^n - 2\hat{u}_{i,j}^n + \hat{u}_{i,j-1}^n}{h_y^2}, \quad (9.14)$$

$$1 \leq i \leq I-1, \quad 1 \leq j \leq J-1, \quad 1 \leq n \leq N-1.$$

Функциональные начальное и краевое условия представляются так же, как и в примере 9.1. Дифференциальное начальное условие следует аппроксимировать со вторым порядком, чтобы не ухудшать свойства схемы (см. пример 8.6). В результате имеем явную трехслойную схему, имеющую второй порядок аппроксимации и условно устойчивую (устойчивую при выполнении условия $\tau < \left(\frac{1}{h_x^2} + \frac{1}{h_y^2}\right)^{\frac{1}{2}}$) [14].

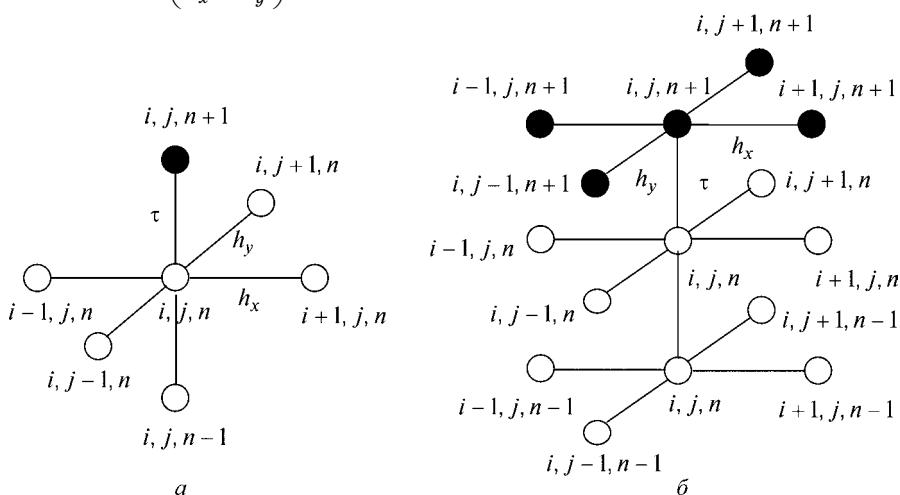


Рис. 9.6

Построим неявную схему, используя пятнадцатиточечный шаблон, изображенный на рисунке 9.6б.

$$\begin{aligned} \frac{u_{i,j}^{n+1} - 2u_{i,j}^n + u_{i,j}^{n-1}}{\tau^2} = & \lambda \left[\frac{\hat{u}_{i+1,j}^{n+1} - 2\hat{u}_{i,j}^{n+1} + \hat{u}_{i-1,j}^{n+1}}{h_x^2} + \frac{\hat{u}_{i,j+1}^{n+1} - 2\hat{u}_{i,j}^{n+1} + \hat{u}_{i,j-1}^{n+1}}{h_y^2} \right] + \\ & + (1-\lambda) \left[\frac{\hat{u}_{i+1,j}^n - 2\hat{u}_{i,j}^n + \hat{u}_{i-1,j}^n}{h_x^2} + \frac{\hat{u}_{i,j+1}^n - 2\hat{u}_{i,j}^n + \hat{u}_{i,j-1}^n}{h_y^2} \right] + \\ & + \lambda \left[\frac{\hat{u}_{i+1,j}^{n-1} - 2\hat{u}_{i,j}^{n-1} + \hat{u}_{i-1,j}^{n-1}}{h_x^2} + \frac{\hat{u}_{i,j+1}^{n-1} - 2\hat{u}_{i,j}^{n-1} + \hat{u}_{i,j-1}^{n-1}}{h_y^2} \right], \end{aligned} \quad (9.15)$$

$1 \leq i \leq I-1, \quad 1 \leq j \leq J-1, \quad 1 \leq n \leq N-1.$

Схема (9.15), (9.11) имеет второй порядок аппроксимации и является трехслойной. При $\lambda = 0$ она переходит в явную. При $\frac{1}{4} \leq \lambda \leq \frac{1}{2}$ схема абсолютно устойчива, но относится к неэкономичным [14]. ■

9.3. МЕТОД РАСПЩЕПЛЕНИЯ

При решении многомерных задач теплопроводности, в которых число независимых переменных больше двух, объем вычислений методом сеток существенно возрастает. Для его снижения Н. Н. Яненко предложил весьма эффективный метод расщепления, который наряду с устойчивостью обладает свойством минимальности объема вычислений [51]. Такие методы принято называть экономичными.

Рассмотрим методику конструирования одной схемы расщепления на примере модельной задачи Коши для двумерного уравнения теплопроводности.

Пример 9.4. Построить схему метода расщепления для решения задачи Коши:

$$\begin{aligned} u_t &= u_{xx} + u_{yy}, \quad (x, y, t) \in R \times R \times (0, T); \\ u(x, y, 0) &= \psi(x, y), \quad (x, y) \in R^2 \text{ (начальное условие)}, \end{aligned}$$

где $\psi(x, y)$ — заданная функция.

□ Следуя [51], введем в рассмотрение дифференциальный оператор $\mathcal{A}u \equiv u_{xx} + u_{yy}$. Тогда решаемое уравнение записывается в оперативной форме: $u_t = \mathcal{A}u$. Предположим, что решение $u(x, y, t^n)$ поставленной задачи известно в момент времени $t = t^n$. Выразим решение $u(x, y, t^{n+1})$ при $t = t^{n+1}$ для фиксированных x и y через решение $u(x, y, t^n)$, пользуясь формулой Тейлора:

$$\begin{aligned} u(x, y, t^{n+1}) &= u(x, y, t^n + \tau) = u(x, y, t^n) + \frac{\tau}{1!} \frac{\partial u(x, y, t^n)}{\partial t} + O(\tau^2) = \\ &= u(x, y, t^n) + \frac{\tau}{1!} \mathcal{A}u(x, y, t^n) + O(\tau^2) = (\mathcal{E} + \tau \mathcal{A})u(x, y, t^n) + O(\tau^2), \end{aligned}$$

где производная $\frac{\partial u(x, y, t^n)}{\partial t}$ во втором слагаемом заменена правой частью исходного дифференциального уравнения, записанного в операторной форме, а \mathcal{E} — тождественный оператор ($\mathcal{E}u(x, y, t^n) = u(x, y, t^n)$).

Зададим сетку $D_h = \{x_i, y_j, t^n\}$, где $x_i = ih, y_j = jh, t^n = n\tau, i, j = 0, \pm 1, \pm 2, \dots, n = 0, 1, \dots, N; T = N\tau$.

Оператор правой части исходного уравнения представим в виде суммы двух операторов:

$$\mathcal{A} = \mathcal{A}_1 + \mathcal{A}_2, \quad \mathcal{A}_1 \equiv u_{xx}, \quad \mathcal{A}_2 \equiv u_{yy}.$$

Наряду с исходной задачей рассмотрим две вспомогательные задачи Коши:

- *задача 1*, в которой решается уравнение

$$\frac{\partial v(x, y, t)}{\partial t} = \frac{\partial^2 v(x, y, t)}{\partial x^2}, \quad (x, y, t) \in R \times R \times (t^n, t^{n+1})$$

с начальным условием

$$\begin{aligned} v(x, y, t^n) &= u(x, y, t^n), \quad (x, y) \in R^2, \text{ при } n = 1, 2, \dots, N - 1; \\ v(x, y, t^0) &= \psi(x, y), \quad (x, y) \in R^2, \text{ при } n = 0; \end{aligned}$$

- *задача 2*, в которой решается уравнение

$$\frac{\partial w(x, y, t)}{\partial t} = \frac{\partial^2 w(x, y, t)}{\partial y^2}, \quad (x, y, t) \in R \times R \times (t^n, t^{n+1})$$

с начальным условием

$$w(x, y, t^n) = v(x, y, t^{n+1}), \quad (x, y) \in R^2, \quad n = 0, 1, \dots, N - 1.$$

Поставленные задачи могут быть решены последовательно: сначала задача 1, а затем задача 2. Решая задачу 1, можно найти функцию $v(x, y, t^{n+1})$ на $(n+1)$ -м слое, которая определяет начальное условие задачи 2. Решая задачу 2, можно найти функцию $w(x, y, t^{n+1})$ на том же слое. Задача 1 при фиксированном y и задача 2 при фиксированном x фактически являются одномерными. Установим связь между функциями $v(x, y, t^{n+1}), w(x, y, t^{n+1})$ и решением $u(x, y, t^{n+1})$ исходной задачи на $(n+1)$ -м слое.

Пользуясь формулой Тейлора, для функции $v(x, y, t^{n+1})$ получим соотношение $v(x, y, t^{n+1}) = (\mathcal{E} + \tau\mathcal{A}_1)v(x, y, t^n) + O(\tau^2) = (\mathcal{E} + \tau\mathcal{A}_1)u(x, y, t^n) + O(\tau^2)$, где функция $v(x, y, t^n)$ заменена на $u(x, y, t^n)$ в силу начального условия в задаче 1.

Для функции $w(x, y, t^{n+1})$ можно получить аналогичное выражение

$$\begin{aligned} w(x, y, t^{n+1}) &= (\mathcal{E} + \tau\mathcal{A}_2)w(x, y, t^n) + O(\tau^2) = (\mathcal{E} + \tau\mathcal{A}_2)v(x, y, t^{n+1}) + O(\tau^2); \\ w(x, y, t^{n+1}) &= (\mathcal{E} + \tau\mathcal{A}_2)[(\mathcal{E} + \tau\mathcal{A}_1)u(x, y, t^n) + O(\tau^2)] + O(\tau^2) = \\ &= \mathcal{E}^2u(x, y, t^n) + \tau(\mathcal{A}_1 + \mathcal{A}_2)u(x, y, t^n) + \tau^2\mathcal{A}_2\mathcal{A}_1u(x, y, t^n) + O(\tau^2) = \\ &= (\mathcal{E} + \tau\mathcal{A})u(x, y, t^n) + O(\tau^2) = u(x, y, t^{n+1}) + O(\tau^2), \end{aligned} \tag{9.16}$$

так как $\mathcal{E}^2 u(x, y, t^n) = \mathcal{E}u(x, y, t^n)$, $\tau^2 \mathcal{A}_2 \mathcal{A}_1 u(x, y, t^n) = O(\tau^2)$. Таким образом, в результате последовательного решения задач 1 и 2 при $t = t^{n+1}$ получаем функцию $w(x, y, t^{n+1})$, которая отличается от решения исходной задачи $u(x, y, t^{n+1})$ лишь на величину $O(\tau^2)$. Поэтому можно положить

$$u(x, y, t^{n+1}) \cong w(x, y, t^{n+1}).$$

Следовательно, здесь процесс решения исходной двумерной задачи заменен процессом решения двух задач в силу *расщепления* дифференциального оператора \mathcal{A} на сумму двух операторов $\mathcal{A}_1 + \mathcal{A}_2$.

Замечания

1. В описываемом методе расщепление произведено на уровне задач. В методе переменных направлений (см. п. 9.4) расщепление на уровне задач не делается, но слагаемые оператора \mathcal{A} аппроксимируются на различных временных слоях.

2. Имеется модификация данного метода, в которой расщепление осуществляется по физическим процессам, когда отдельные слагаемые оператора \mathcal{A} соответствуют различным физическим явлениям (например, неравновесным химическим или колебательным процессам, процессам подогрева, двухфазным процессам и т. д.).

Сформируем две схемы расщепления, заменяя задачи 1 и 2 разностными аппроксимациями. При построении явной схемы применим шаблоны, изображенные на рисунке 9.7. Для аппроксимации частных производных по времени применим первую формулу в (9.5), а для аппроксимации пространственных производных — последние формулы в (9.6) и (9.7), имеющие второй порядок аппроксимации.

Тогда для задачи 1 получаем (шаблон на рис. 9.7a)

$$\begin{aligned} \frac{\hat{v}_{i,j}^{n+1} - \hat{v}_{i,j}^n}{\tau} &= \frac{\hat{v}_{i+1,j}^n - 2\hat{v}_{i,j}^n + \hat{v}_{i-1,j}^n}{h^2}, \quad i, j = 0, \pm 1, \pm 2, \dots, \quad n = 0, 1, \dots, N-1; \\ \hat{v}_{i,j}^0 &= \psi_{i,j}, \quad i, j = 0, \pm 1, \pm 2, \dots; \\ \hat{v}_{i,j}^n &= \hat{u}_{i,j}^n, \quad i, j = 0, \pm 1, \pm 2, \dots, \quad n = 1, \dots, N-1, \end{aligned} \quad (9.17)$$

где $\psi_{i,j} = \psi(x_i, y_j)$.

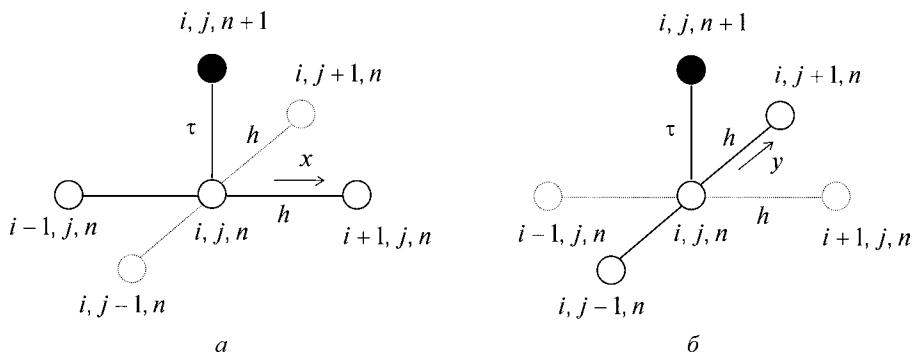


Рис. 9.7

Для задачи 2 аналогично получаем (шаблон на рис. 9.7б)

$$\begin{aligned}\frac{\hat{w}_{i,j}^{n+1} - \hat{w}_{i,j}^n}{\tau} &= \frac{\hat{w}_{i,j+1}^n - 2\hat{w}_{i,j}^n + \hat{w}_{i,j-1}^n}{h^2}, \quad i, j = 0, \pm 1, \pm 2, \dots, n = 0, 1, \dots, N-1; \\ \hat{w}_{i,j}^n &= \hat{v}_{i,j}^{n+1}, \quad i, j = 0, \pm 1, \pm 2, \dots, n = 0, 1, \dots, N-1.\end{aligned}\quad (9.18)$$

Заметим, что (9.17) с учетом условия $\hat{v}_{i,j}^n = \hat{w}_{i,j}^n$ можно переписать в форме

$$\hat{v}_{i,j}^{n+1} = \hat{u}_{i,j}^n + \tau \frac{\hat{u}_{i+1,j}^n - 2\hat{u}_{i,j}^n + \hat{u}_{i-1,j}^n}{h^2},$$

а (9.18) — в виде

$$\hat{w}_{i,j}^{n+1} = \hat{v}_{i,j}^{n+1} + \tau \frac{\hat{w}_{i,j+1}^n - 2\hat{w}_{i,j}^n + \hat{w}_{i,j-1}^n}{h^2}, \quad \hat{w}_{i,j}^n = \hat{v}_{i,j}^{n+1}.$$

С учетом (9.16) и того, что начальные условия при $n = 0$ аппроксимируются точно, эта явная двухслойная разностная схема аппроксимирует исходную задачу с порядком $O(\tau^2 + h^2)$. Схема является условно устойчивой (устойчивой при $\tau \leq \frac{h^2}{2}$) [26].

Методика вычислений по явной схеме

1. Задать значения шагов сетки h, τ так, что $N\tau = T$, где N — целое положительное число, причем значение шага по времени выбрать из условия устойчивости $\tau \leq \frac{h^2}{2}$. Вычислить $\hat{v}_{i,j}^0 = \hat{u}_{i,j}^0 = \psi_{i,j}$, $i, j = 0, \pm 1, \pm 2, \dots$, где $\psi_{i,j} = \psi(x_i, y_j)$. Положить $n = 0$.

2. Найти решение на $(n+1)$ -м слое.

Первый этап. Решить задачу 1 при каждом $j = 0, \pm 1, \pm 2, \dots$:

$$\hat{v}_{i,j}^{n+1} = \hat{u}_{i,j}^n + \tau \frac{\hat{u}_{i+1,j}^n - 2\hat{u}_{i,j}^n + \hat{u}_{i-1,j}^n}{h^2}, \quad i = 0, \pm 1, \pm 2, \dots$$

Второй этап. Решить задачу 2 при каждом $i = 0, \pm 1, \pm 2, \dots$:

$$\hat{w}_{i,j}^{n+1} = \hat{v}_{i,j}^{n+1} + \tau \frac{\hat{w}_{i,j+1}^n - 2\hat{w}_{i,j}^n + \hat{w}_{i,j-1}^n}{h^2}, \quad i = 0, \pm 1, \pm 2, \dots, \quad \hat{w}_{i,j}^n = \hat{v}_{i,j}^{n+1}.$$

Положить $\hat{u}_{i,j}^{n+1} = \hat{w}_{i,j}^{n+1}$, $i, j = 0, \pm 1, \pm 2, \dots$

3. Если $n = N$, вычисления завершить. Иначе положить $n = n + 1$ и перейти к п. 2.

При построении *неявной схемы* применим шаблоны, изображенные на рисунке 9.8. Для аппроксимации частных производных по времени применим первую формулу в (9.5), а для аппроксимации пространственных производных — последние формулы в (9.6) и (9.7), записанные на $(n+1)$ -м временном слое.

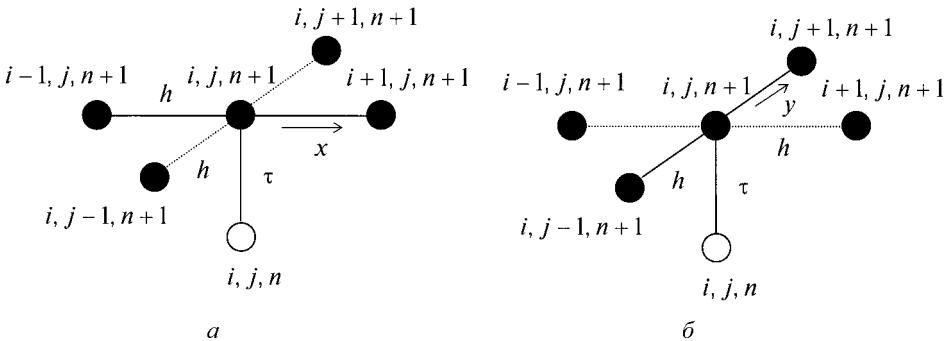


Рис. 9.8

Тогда для задачи 1 получаем (шаблон на рис. 9.8а)

$$\begin{aligned} \frac{\hat{v}_{i,j}^{n+1} - \hat{v}_{i,j}^n}{\tau} &= \frac{\hat{v}_{i+1,j}^{n+1} - 2\hat{v}_{i,j}^{n+1} + \hat{v}_{i-1,j}^{n+1}}{h^2}, \quad i, j = 0, \pm 1, \pm 2, \dots, \quad n = 0, 1, \dots, N-1; \\ \hat{v}_{i,j}^0 &= \psi_{i,j}, \quad i, j = 0, \pm 1, \pm 2, \dots; \\ \hat{v}_{i,j}^n &= \hat{u}_{i,j}^n, \quad i, j = 0, \pm 1, \pm 2, \dots, \quad n = 0, 1, \dots, N-1, \end{aligned} \quad (9.19)$$

где $\psi_{i,j} = \psi(x_i, y_j)$.

Для задачи 2 аналогично получаем (шаблон на рис. 9.8б)

$$\begin{aligned} \frac{\hat{w}_{i,j}^{n+1} - \hat{w}_{i,j}^n}{\tau} &= \frac{\hat{w}_{i,j+1}^{n+1} - 2\hat{w}_{i,j}^{n+1} + \hat{w}_{i,j-1}^{n+1}}{h^2}, \quad i, j = 0, \pm 1, \pm 2, \dots, \quad n = 0, 1, \dots, N-1; \\ \hat{w}_{i,j}^n &= \hat{v}_{i,j}^{n+1}, \quad i, j = 0, \pm 1, \pm 2, \dots, \quad n = 0, 1, \dots, N-1. \end{aligned} \quad (9.20)$$

Приведенные соотношения образуют абсолютно устойчивую двухслойную неявную разностную схему метода расщепления, имеющую порядок аппроксимации $O(\tau^2 + h^2)$ [26].

Методика вычислений по неявной схеме

1. Задать значения шагов сетки h, τ так, что $N\tau = T$, где N — целое положительное число, причем значения шагов выбрать из условия достижения заданной точности. Вычислить $\hat{v}_{i,j}^0 = \hat{u}_{i,j}^0 = \psi_{i,j}$, $i, j = 0, \pm 1, \pm 2, \dots$, где $\psi_{i,j} = \psi(x_i, y_j)$. Положить $n = 0$.

2. Найти решение на $(n+1)$ -м слое.

Первый этап. Решить задачу 1 при каждом $j = 0, \pm 1, \pm 2, \dots$:

$$\frac{1}{h^2} \hat{v}_{i+1,j}^{n+1} - \left(\frac{2}{h^2} + \frac{1}{\tau} \right) \hat{v}_{i,j}^{n+1} + \frac{1}{h^2} \hat{v}_{i-1,j}^{n+1} = -\frac{\hat{u}_{i,j}^n}{\tau}, \quad i = 0, \pm 1, \pm 2, \dots,$$

применяя метод прогонки для системы линейных алгебраических уравнений с трехдиагональной матрицей (система получена из (9.19) при $\hat{v}_{i,j}^n = \hat{u}_{i,j}^n$).

Второй этап. Решить задачу 2 при каждом $i = 0, \pm 1, \pm 2, \dots$:

$$\frac{1}{h^2} \hat{w}_{i,j+1}^{n+1} - \left(\frac{2}{h^2} + \frac{1}{\tau} \right) \hat{w}_{i,j}^{n+1} + \frac{1}{h^2} \hat{w}_{i,j-1}^{n+1} = -\frac{1}{\tau} \hat{v}_{i,j}^{n+1}, \quad j = 0, \pm 1, \pm 2, \dots,$$

применяя метод прогонки для системы линейных алгебраических уравнений с трехдиагональной матрицей (система получена из (9.20) при $\hat{w}_{i,j}^n = \hat{v}_{i,j}^{n+1}$).

Положить $\hat{u}_{i,j}^{n+1} = \hat{w}_{i,j}^{n+1}$, $i, j = 0, \pm 1, \pm 2, \dots$

3. Если $n = N$, вычисления завершить. Иначе положить $n = n + 1$ и перейти к п. 2. ■

Замечания

1. Для решаемых на первом и втором этапах методики систем линейных алгебраических уравнений выполняется условие преобладания диагональных элементов, следовательно, метод прогонки устойчив.

2. Для пояснения общей схемы вычислений здесь рассматривалась задача Коши, поэтому индексы i и j были не ограничены. Однако эти схемы могут быть легко приспособлены для решения смешанных краевых задач с ограниченным множеством D [26].

9.4. МЕТОД ПЕРЕМЕННЫХ НАПРАВЛЕНИЙ

Рассмотрим идею метода переменных направлений, относящегося к классу экономичных схем, применительно к решению первой начально-краевой задачи для двумерного уравнения теплопроводности.

Пример 9.5. Построить разностную схему метода переменных направлений для решения первой начально-краевой задачи для уравнения теплопроводности:

$$\begin{aligned} u_t &= u_{xx} + u_{yy} + f(x, y, t), \quad (x, y, t) \in (0, L) \times (0, M) \times (0, T); \\ u(x, y, 0) &= \psi(x, y), \quad (x, y) \in [0, L] \times [0, M] \text{ (начальное условие);} \\ u(x, y, t) &= \varphi(x, y, t), \quad (x, y) \in \Gamma, \quad 0 < t \leq T \text{ (краевое условие)}, \end{aligned}$$

где $\Omega = (0, L) \times (0, M)$; Γ — граница области Ω (прямоугольника); $f(x, y, t)$ — плотность источников тепла; $\psi(x, y)$, $\varphi(y, t)$ — заданные функции; L, M, T — заданные числа.

□Как и в методе расщепления, исходное уравнение аппроксимируется совокупностью двух разностных схем, каждая из которых соответствует только одному пространственному направлению. Однако оператор A здесь не расщепляется на сумму $A_1 + A_2$ двух операторов на уровне задач, а каждый элемент суммы аппроксимируется с помощью явных и неявных конструкций. Для этого наряду со слоем $t = t^{n+1}$, расчет которого осуществляется на данном этапе, вводится один дополнительный (промежуточный) слой $t = t^{\frac{n+1}{2}}$. Тогда для перехода от слоя $t = t^n$ к слою $t = t^{n+1}$ с использованием слоя $t = t^{\frac{n+1}{2}}$ исходное дифференциальное уравнение аппроксимируется двумя разностными уравнениями, одно из которых связывает слои $t = t^n$ и $t = t^{\frac{n+1}{2}}$, а второе — слои $t = t^{\frac{n+1}{2}}$ и $t = t^{n+1}$.

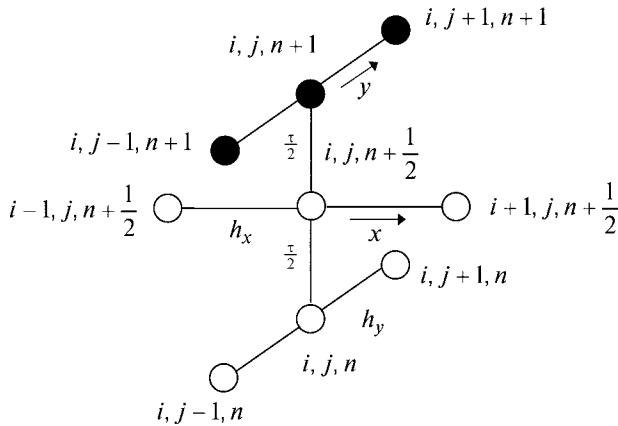


Рис. 9.9

При реализации метода используем девятиточечный шаблон, изображенный на рисунке 9.9. Переход с n -го временного слоя на $(n + 1)$ -й разбивается на два этапа.

Первый этап. Переход на дополнительный промежуточный $\left(n + \frac{1}{2}\right)$ -й слой с шагом $\frac{\tau}{2}$. Частная производная по времени аппроксимируется по первой формуле в (9.5), производная по x аппроксимируется на $\left(n + \frac{1}{2}\right)$ -м слое, производная по y — на n -м слое. Входящая в правую часть исходного уравнения функция $f(x, y, t)$ заменяется своим сеточным представлением. Соответствующая разностная схема имеет вид:

$$\frac{\hat{u}_{i,j}^{n+\frac{1}{2}} - \hat{u}_{i,j}^n}{\frac{\tau}{2}} = \frac{\hat{u}_{i+1,j}^{n+\frac{1}{2}} - 2\hat{u}_{i,j}^{n+\frac{1}{2}} + \hat{u}_{i-1,j}^{n+\frac{1}{2}}}{h_x^2} + \frac{\hat{u}_{i,j+1}^n - 2\hat{u}_{i,j}^n + \hat{u}_{i,j-1}^n}{h_y^2} + f_{i,j}^{n+\frac{1}{2}}, \quad (9.21)$$

$$1 \leq i \leq I-1, \quad 1 \leq j \leq J-1,$$

где $f_{i,j}^{n+\frac{1}{2}} = f\left(x_i, y_j, t^n + \frac{1}{2}\tau\right)$, или в традиционной записи:

$$\frac{\tau}{2h_x^2} \hat{u}_{i-1,j}^{n+\frac{1}{2}} - \left(1 + \frac{\tau}{h_x^2}\right) \hat{u}_{i,j}^{n+\frac{1}{2}} + \frac{\tau}{2h_x^2} \hat{u}_{i+1,j}^{n+\frac{1}{2}} = F_{i,j}^n, \quad 1 \leq i \leq I-1, \quad 1 \leq j \leq J-1,$$

где $F_{i,j}^n = -\frac{\tau}{2} \hat{f}_{i,j}^{n+\frac{1}{2}} - \frac{\tau}{2h_y^2} (\hat{u}_{i,j+1}^n - 2\hat{u}_{i,j}^n + \hat{u}_{i,j-1}^n) - \hat{u}_{i,j}^n$. Заметим, что вместо значения $\hat{f}_{i,j}^{n+\frac{1}{2}}$ в (9.21) можно использовать $f_{i,j}^n$ [14].

Для каждого $j = 1, \dots, J-1$ требуется решить трехдиагональную систему линейных алгебраических уравнений, так как каждое уравнение содержит три неизвестных значения $\hat{u}_{i-1,j}^{n+\frac{1}{2}}, \hat{u}_{i,j}^{n+\frac{1}{2}}, \hat{u}_{i+1,j}^{n+\frac{1}{2}}$, а остальные значения берутся с n -го слоя или вычисляются, например $\hat{f}_{i,j}^{n+\frac{1}{2}}$. Иными словами, при таком

переходе схема является неявной по направлению x и явной по направлению y . Искомые значения на промежуточном слое вычисляются методом прогонки по направлению x , т. е. продольному направлению.

Второй этап. Переход на $(n+1)$ -й временной слой с промежуточного $\left(n+\frac{1}{2}\right)$ -го с шагом $\frac{\tau}{2}$. Частная производная по времени аппроксимируется по первой формуле в (9.5), производная по x аппроксимируется на $\left(n+\frac{1}{2}\right)$ -м слое, производная по y — на $(n+1)$ -м слое. Соответствующая разностная схема имеет вид:

$$\frac{\hat{u}_{i,j}^{n+1} - \hat{u}_{i,j}^{n+\frac{1}{2}}}{\frac{\tau}{2}} = \frac{\hat{u}_{i+1,j}^{n+\frac{1}{2}} - 2\hat{u}_{i,j}^{n+\frac{1}{2}} + \hat{u}_{i-1,j}^{n+\frac{1}{2}}}{h_x^2} + \frac{\hat{u}_{i,j+1}^{n+1} - 2\hat{u}_{i,j}^{n+1} + \hat{u}_{i,j-1}^{n+1}}{h_y^2} + f_{i,j}^{n+1}, \quad (9.22)$$

$1 \leq i \leq I-1, \quad 1 \leq j \leq J-1,$

или в традиционной записи:

$$\frac{\tau}{2h_y^2} \hat{u}_{i,j-1}^{n+1} - \left(1 + \frac{\tau}{h_y^2}\right) \hat{u}_{i,j}^{n+1} + \frac{\tau}{2h_y^2} \hat{u}_{i,j+1}^{n+1} = \Phi_{i,j}^{n+\frac{1}{2}}, \quad 1 \leq i \leq I-1, \quad 1 \leq j \leq J-1,$$

где

$$\Phi_{i,j}^{n+\frac{1}{2}} = -\frac{\tau}{2} f_{i,j}^{n+1} - \frac{\tau}{2h_x^2} \left(\hat{u}_{i-1,j}^{n+\frac{1}{2}} - 2\hat{u}_{i,j}^{n+\frac{1}{2}} + \hat{u}_{i+1,j}^{n+\frac{1}{2}} \right) - \hat{u}_{i,j}^{n+\frac{1}{2}}.$$

Заметим, что вместо значения $f_{i,j}^{n+1}$ в (9.22) можно использовать $f_{i,j}^{n+\frac{1}{2}}$ [14].

Для каждого значения $i = 1, \dots, I-1$ требуется решить трехдиагональную систему линейных алгебраических уравнений, так как каждое уравнение содержит три неизвестных значения $\hat{u}_{i,j-1}^{n+1}$, $\hat{u}_{i,j}^{n+1}$, $\hat{u}_{i,j+1}^{n+1}$, а остальные значения берутся с $\left(n+\frac{1}{2}\right)$ -го слоя. Иными словами, при таком переходе схема является явной по направлению x и неявной по направлению y . Искомые значения на целом слое вычисляются методом прогонки по направлению y , т. е. в поперечном направлении. Диагональные элементы в системах (9.21), (9.22) преобладают, следовательно, процесс прогонки устойчив, а решение разностной задачи существует и единственno. К достоинствам метода, достигнутым благодаря введению промежуточного слоя, следует отнести расщепление исходной задачи на две более простые, решаемые с помощью алгоритма скалярной прогонки.

Начальные и краевые условия представляются в виде

$$\begin{aligned} \hat{u}_{i,j}^0 &= \psi_{i,j}, \quad 0 \leq i \leq I, \quad 0 \leq j \leq J \quad (\text{нулевой слой}); \\ \hat{u}_{0,j}^n &= \varphi_{0,j}^n, \quad \hat{u}_{I,j}^n = \varphi_{I,j}^n, \quad 0 \leq j \leq J, \quad 0 \leq n \leq N \quad (\text{на плоскостях,} \\ &\quad \text{перпендикулярных оси } Ox); \\ \hat{u}_{i,0}^n &= \varphi_{i,0}^n, \quad \hat{u}_{i,J}^n = \varphi_{i,J}^n, \quad 1 \leq i \leq I, \quad 1 \leq n \leq N \quad (\text{на плоскостях,} \\ &\quad \text{перпендикулярных оси } Oy), \end{aligned} \quad (9.23)$$

где $\varphi_{i,j}^n = \varphi(x_i, y_j, t^n)$, $\psi_{i,j} = \psi(x_i, y_j)$.

Для промежуточного слоя требуются значения $\hat{u}_{i,j}^{n+\frac{1}{2}}$ на сторонах расчетной области, определяемых уравнениями $x = 0$ и $x = L$, т. е. $\hat{u}_{0,j}^{n+\frac{1}{2}}$, $\hat{u}_{I,j}^{n+\frac{1}{2}}$. Вычитая (9.22) из (9.21), получаем

$$\hat{u}_{i,j}^{n+\frac{1}{2}} = \frac{1}{2}(\hat{u}_{i,j}^{n+1} + \hat{u}_{i,j}^n) - \frac{\tau}{4h_y^2}(\hat{u}_{i,j+1}^{n+1} - 2\hat{u}_{i,j}^{n+1} + \hat{u}_{i,j-1}^{n+1}) + \frac{\tau}{4h_y^2}(\hat{u}_{i,j+1}^n - 2\hat{u}_{i,j}^n + \hat{u}_{i,j-1}^n).$$

Отсюда при $i = 0$ ($x = 0$) имеем

$$\hat{u}_{0,j}^{n+\frac{1}{2}} = \frac{1}{2}(\hat{u}_{0,j}^{n+1} + \hat{u}_{0,j}^n) - \frac{\tau}{4h_y^2}(\hat{u}_{0,j+1}^{n+1} - 2\hat{u}_{0,j}^{n+1} + \hat{u}_{0,j-1}^{n+1}) + \frac{\tau}{4h_y^2}(\hat{u}_{0,j+1}^n - 2\hat{u}_{0,j}^n + \hat{u}_{0,j-1}^n).$$

Такой способ обеспечивает второй порядок аппроксимации. Аналогично получается соотношение при $i = I$ ($x = L$).

Схема (9.21)–(9.23) метода переменных направлений имеет второй порядок аппроксимации и является абсолютно устойчивой [14].

Методика расчетов

1. Задать сетку $D_h = \{x_i, y_j, t^n\}$, где $x_i = ih_x$, $0 \leq i \leq I$; $y_j = jh_y$, $0 \leq j \leq J$; $t^n = n\tau$, $0 \leq n \leq N$; N, J, I — целые положительные числа; h_x, h_y, τ — величины шагов по пространственным координатам и времени (для простоты принимаются постоянными); $Ih_x = L$, $Jh_y = M$, $N\tau = T$.

2. Вычислить

$$\begin{aligned}\hat{u}_{i,j}^0 &= \psi_{i,j}, \quad 0 \leq i \leq I, \quad 0 \leq j \leq J; \\ \hat{u}_{0,j}^n &= \varphi_{0,j}^n, \quad \hat{u}_{I,j}^n = \varphi_{I,j}^n, \quad 0 \leq j \leq J, \quad 0 \leq n \leq N; \\ \hat{u}_{i,0}^n &= \varphi_{i,0}^n, \quad \hat{u}_{i,J}^n = \varphi_{i,J}^n, \quad 0 \leq i \leq I, \quad 0 \leq n \leq N,\end{aligned}$$

где $\varphi_{i,j}^n = \varphi(x_i, y_j, t^n)$, $\psi_{i,j} = \psi(x_i, y_j)$.

Положить $n = 0$.

3. Найти решение на $(n + 1)$ -м слое.

Первый этап. Перейти на дополнительный промежуточный $\left(n + \frac{1}{2}\right)$ -й слой с шагом $\frac{\tau}{2}$: для каждого $j = 1, \dots, J - 1$ решить трехдиагональную систему линейных алгебраических уравнений

$$\frac{\tau}{2h_x^2} \hat{u}_{i-1,j}^{n+\frac{1}{2}} - \left(1 + \frac{\tau}{h_x^2}\right) \hat{u}_{i,j}^{n+\frac{1}{2}} + \frac{\tau}{2h_x^2} \hat{u}_{i+1,j}^{n+\frac{1}{2}} = F_{i,j}^{n+\frac{1}{2}}, \quad 1 \leq i \leq I - 1,$$

где

$$\begin{aligned}F_{i,j}^{n+\frac{1}{2}} &= -\frac{\tau}{2} f_{i,j}^{n+\frac{1}{2}} - \frac{\tau}{2h_y^2} (\hat{u}_{i,j+1}^n - 2\hat{u}_{i,j}^n + \hat{u}_{i,j-1}^n) - \hat{u}_{i,j}^n; \\ \hat{u}_{0,j}^{n+\frac{1}{2}} &= \frac{1}{2}(\hat{u}_{0,j}^{n+1} + \hat{u}_{0,j}^n) - \frac{\tau}{4h_y^2} (\hat{u}_{0,j+1}^{n+1} - 2\hat{u}_{0,j}^{n+1} + \hat{u}_{0,j-1}^{n+1}) + \frac{\tau}{4h_y^2} (\hat{u}_{0,j+1}^n - 2\hat{u}_{0,j}^n + \hat{u}_{0,j-1}^n); \\ \hat{u}_{I,j}^{n+\frac{1}{2}} &= \frac{1}{2}(\hat{u}_{I,j}^{n+1} + \hat{u}_{I,j}^n) - \frac{\tau}{4h_y^2} (\hat{u}_{I,j+1}^{n+1} - 2\hat{u}_{I,j}^{n+1} + \hat{u}_{I,j-1}^{n+1}) + \frac{\tau}{4h_y^2} (\hat{u}_{I,j+1}^n - 2\hat{u}_{I,j}^n + \hat{u}_{I,j-1}^n).\end{aligned}$$

Искомые значения $\hat{u}_{i-1,j}^{n+\frac{1}{2}}$, $\hat{u}_{i,j}^{n+\frac{1}{2}}$, $\hat{u}_{i+1,j}^{n+\frac{1}{2}}$ на промежуточном слое вычисляются методом прогонки по направлению x .

Второй этап. Перейти на $(n + 1)$ -й временной слой с промежуточного $\left(n + \frac{1}{2}\right)$ -го с шагом $\frac{\tau}{2}$: для каждого значения $i = 1, \dots, I - 1$ решить трехдиагональную систему линейных алгебраических уравнений

$$\frac{\tau}{2h_y^2} \hat{u}_{i,j-1}^{n+1} - \left(1 + \frac{\tau}{h_y^2}\right) \hat{u}_{i,j}^{n+1} + \frac{\tau}{2h_x^2} \hat{u}_{i,j+1}^{n+1} = \Phi_{i,j}^{n+\frac{1}{2}}, \quad 1 \leq j \leq J - 1,$$

где

$$\Phi_{i,j}^{n+\frac{1}{2}} = -\frac{\tau}{2} f_{i,j}^{n+1} - \frac{\tau}{2h_x^2} \left(\hat{u}_{i-1,j}^{n+\frac{1}{2}} - 2\hat{u}_{i,j}^{n+\frac{1}{2}} + \hat{u}_{i+1,j}^{n+\frac{1}{2}} \right) - \hat{u}_{i,j}^{n+\frac{1}{2}}.$$

Искомые значения $\hat{u}_{i,j-1}^{n+1}$, $\hat{u}_{i,j}^{n+1}$, $\hat{u}_{i,j+1}^{n+1}$ на целом слое вычисляются методом прогонки по направлению y .

4. Если $n = N$, вычисления завершить. Иначе положить $n = n + 1$ и перейти к п. 3. ■

Замечания

1. При решении двумерных задач со смешанными производными по пространственным переменным и трехмерных задач изложенная схема становится условно устойчивой.

2. Описанные в п. 9.2 и 9.3 алгоритмы решения двумерного уравнения теплопроводности могут применяться для решения двумерных уравнений эллиптического типа.

Рассмотрим задачу Дирихле для уравнения Лапласа с двумя независимыми переменными (см. п. 8.1):

$$u_{xx} + u_{yy} = 0, \quad (x, y) \in \Omega; \\ u|_{\Gamma} = \phi(x, y), \quad (x, y) \in \Gamma \text{ (краевое условие)},$$

где $\phi(x, y)$ — заданная функция.

Наряду с этой стационарной задачей рассмотрим нестационарную задачу для двумерного уравнения теплопроводности с теми же краевыми условиями и произвольно заданными начальными условиями:

$$u_t = u_{xx} + u_{yy}, \quad (x, y, t) \in \Omega \times (0, +\infty); \\ u(x, y, 0) = \psi(x, y), \quad (x, y) \in \Omega \cup \Gamma \text{ (начальное условие)}; \\ u(x, y, t)|_{\Gamma} = \phi(x, y), \quad (x, y) \in \Gamma, \quad 0 < t < +\infty \text{ (краевое условие)}.$$

Таким образом, вводится дополнительная независимая переменная t и исходная задача усложняется, так как увеличивается ее размерность. Однако при этом становится возможным выбор наиболее экономичной разностной схемы решения уравнения теплопроводности. Можно доказать, что решение нестационарной задачи при $t \rightarrow +\infty$ сходится к решению стационарной задачи, а при достаточно большом $t = T$ может быть принято за приближенное

решение задачи Дирихле [14], [35]. Такой метод называется *методом усташновления*. Для прекращения вычислений можно использовать, например, условие $\max_{i,j} |\hat{u}_{i,j}^{n+l} - \hat{u}_{i,j}^n| \leq \varepsilon$, где $\varepsilon > 0$, l — положительное целое число (порядка нескольких десятков).

Для решения нестационарной задачи можно использовать схемы (9.21), (9.22) и метод переменных направлений. В последнем случае существует наилучший шаг по времени:

$$\tau \cong \frac{1}{\pi} \sqrt{h_x^2 + h_y^2} \left(\frac{1}{L^2} + \frac{1}{M^2} \right)^{-\frac{1}{2}},$$

при котором нестационарное решение выйдет на стационарное за наименьшее число шагов [14].

9.5. МЕТОД ДРОБНЫХ ШАГОВ

Рассмотрим методику конструирования схемы метода дробных шагов на примере модельной задачи Коши для двумерного уравнения теплопроводности.

Пример 9.6. Построить схему метода дробных шагов для решения задачи Коши:

$$u_t = u_{xx} + u_{yy}, \quad (x, y, t) \in R \times R \times (0, T); \\ u(x, y, 0) = \psi(x, y), \quad (x, y) \in R^2 \text{ (начальное условие)},$$

где $\psi(x, y)$ — заданная функция.

Зададим сетку $D_h = \{x_i, y_j, t^n\}$, где $x_i = ih_x$, $y_j = jh_y$, $t^n = n\tau$, $i, j = 0, \pm 1, \pm 2, \dots, n = 0, 1, \dots, N; T = N\tau$.

Оператор правой части исходного уравнения, как в методе расщепления, представим в виде суммы двух операторов:

$$\mathcal{A} = \mathcal{A}_1 + \mathcal{A}_2, \quad \mathcal{A}_1 \equiv u_{xx}, \quad \mathcal{A}_2 \equiv u_{yy}.$$

Далее они аппроксимируются отдельно друг от друга. В отличие от метода расщепления вводится вспомогательный (промежуточный) слой, соответствующий значению $t = t^{n+\frac{1}{2}}$. Как в методе переменных направлений, сначала осуществляется переход на промежуточный $\left(n + \frac{1}{2}\right)$ -й слой с дробным шагом $\frac{\tau}{2}$, а затем с тем же шагом на целый $(n+1)$ -й слой.

Наряду с исходной задачей рассмотрим две вспомогательные задачи Коши:

- *задача 1*, в которой решается уравнение

$$\frac{\partial v(x, y, t)}{\partial t} = \frac{\partial^2 v(x, y, t)}{\partial x^2}, \quad (x, y, t) \in R \times R \times \left(t^n, t^{n+\frac{1}{2}}\right),$$

с начальным условием

$$v(x, y, t^n) = u(x, y, t^n), \quad (x, y) \in R^2, \text{ при } n = 1, 2, \dots, N-1; \\ v(x, y, t^0) = \psi(x, y), \quad (x, y) \in R^2, \text{ при } n = 0;$$

- задача 2, в которой решается уравнение

$$\frac{\partial w(x, y, t)}{\partial t} = \frac{\partial^2 w(x, y, t)}{\partial y^2}, \quad (x, y, t) \in R \times R \times \left(t^{n+\frac{1}{2}}, t^{n+1}\right),$$

с начальным условием

$$w\left(x, y, t^{n+\frac{1}{2}}\right) = v\left(x, y, t^{n+\frac{1}{2}}\right), \quad (x, y) \in R^2, \quad n = 0, 1, \dots, N-1.$$

Поставленные задачи могут быть решены последовательно: сначала задача 1, а затем задача 2. Решая задачу 1, можно найти функцию $v(x, y, t^{n+\frac{1}{2}})$, которая определяет начальное условие задачи 2. Решая задачу 2, можно найти функцию $w(x, y, t^{n+1})$. Задача 1 при фиксированном y и задача 2 при фиксированном x фактически являются одномерными. Как и в методе расщепления, функция $w(x, y, t^{n+1})$ является приближенным решением исходной задачи на $(n+1)$ -м слое, т. е. можно положить $u(x, y, t^{n+1}) \approx w(x, y, t^{n+1})$.

При реализации метода используем семиточечный шаблон, изображенный на рисунке 9.10. Переход с n -го временного слоя на $(n+1)$ -й разбивается на два этапа.

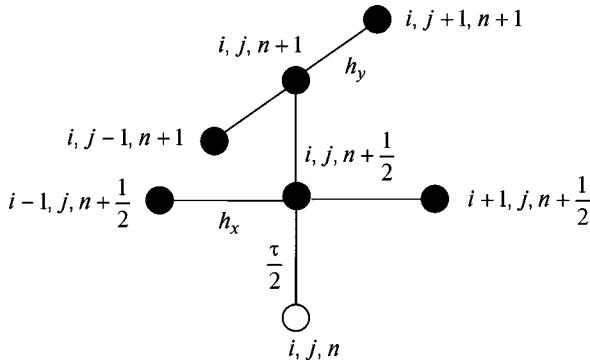


Рис. 9.10

Первый этап. Переход на дополнительный промежуточный $\left(n+\frac{1}{2}\right)$ -й слой с шагом $\frac{\tau}{2}$. В задаче 1 частная производная по времени аппроксимируется по первой формуле в (9.5), производная по x аппроксимируется по последней формуле в (9.6) на $\left(n+\frac{1}{2}\right)$ -м слое. Соответствующая разностная схема имеет вид:

$$\begin{aligned} \frac{\hat{v}_{i,j}^{n+\frac{1}{2}} - \hat{v}_{i,j}^n}{\frac{\tau}{2}} &= \frac{\hat{v}_{i+1,j}^{n+\frac{1}{2}} - 2\hat{v}_{i,j}^{n+\frac{1}{2}} + \hat{v}_{i-1,j}^{n+\frac{1}{2}}}{h_x^2}, \quad i, j = 0, \pm 1, \pm 2, \dots, \quad n = 0, 1, \dots, N-1; \\ \hat{v}_{i,j}^0 &= \psi_{i,j}, \quad i, j = 0, \pm 1, \pm 2, \dots; \\ \hat{v}_{i,j}^n &= \hat{u}_{i,j}^n, \quad i, j = 0, \pm 1, \pm 2, \dots, \quad n = 1, \dots, N-1, \end{aligned} \tag{9.24}$$

где $\psi_{i,j} = \psi(x_i, y_j)$, или в традиционной записи:

$$\frac{\tau}{2h_x^2} \hat{v}_{i-1,j}^{n+\frac{1}{2}} - \left(1 + \frac{\tau}{h_x^2}\right) \hat{v}_{i,j}^{n+\frac{1}{2}} + \frac{\tau}{2h_x^2} \hat{v}_{i+1,j}^{n+\frac{1}{2}} = -\hat{u}_{i,j}^n, \quad i, j = 0, \pm 1, \pm 2, \dots$$

Для каждого j требуется решить трехдиагональную систему линейных алгебраических уравнений, так как каждое уравнение содержит три неизвестных значения $\hat{v}_{i-1,j}^{n+\frac{1}{2}}$, $\hat{v}_{i,j}^{n+\frac{1}{2}}$, $\hat{v}_{i+1,j}^{n+\frac{1}{2}}$, а остальные значения берутся с n -го слоя. Иными словами, при таком переходе схема является неявной. Искомые значения на промежуточном слое вычисляются методом прогонки по направлению x .

Второй этап. Переход на $(n+1)$ -й временной слой с промежуточного $\left(n + \frac{1}{2}\right)$ -го с шагом $\frac{\tau}{2}$. Частная производная по времени аппроксимируется по первой формуле в (9.5), производная по y аппроксимируется по последней формуле в (9.7) на $(n+1)$ -м слое. Соответствующая разностная схема имеет вид:

$$\begin{aligned} \frac{\hat{w}_{i,j}^{n+1} - \hat{w}_{i,j}^{n+\frac{1}{2}}}{\frac{\tau}{2}} &= \frac{\hat{w}_{i,j+1}^{n+1} - 2\hat{w}_{i,j}^{n+1} + \hat{w}_{i,j-1}^{n+1}}{h_y^2}, \quad i, j = 0, \pm 1, \pm 2, \dots, n = 0, 1, \dots, N-1; \\ \hat{w}_{i,j}^{n+\frac{1}{2}} &= \hat{v}_{i,j}^{n+\frac{1}{2}}, \quad i, j = 0, \pm 1, \pm 2, \dots, n = 0, 1, \dots, N-1, \end{aligned} \quad (9.25)$$

или в традиционной записи:

$$\frac{\tau}{2h_y^2} \hat{w}_{i,j-1}^{n+1} - \left(1 + \frac{\tau}{h_y^2}\right) \hat{w}_{i,j}^{n+1} + \frac{\tau}{2h_y^2} \hat{w}_{i,j+1}^{n+1} = -\hat{v}_{i,j}^{n+\frac{1}{2}}, \quad i, j = 0, \pm 1, \pm 2, \dots, n = 0, 1, \dots, N-1.$$

Для каждого значения i требуется решить трехдиагональную систему линейных алгебраических уравнений, так как каждое уравнение содержит три неизвестных значения $\hat{w}_{i,j-1}^{n+1}$, $\hat{w}_{i,j}^{n+1}$, $\hat{w}_{i,j+1}^{n+1}$, а остальные значения берутся с $\left(n + \frac{1}{2}\right)$ -го слоя. Иными словами, при таком переходе схема является неявной. Искомые значения на целом слое вычисляются методом прогонки по направлению y . Далее полагают $\hat{u}_{i,j}^{n+1} = \hat{w}_{i,j}^{n+1}$, $i, j = 0, \pm 1, \pm 2, \dots$

Можно показать, что схема (9.24), (9.25) метода дробных шагов имеет первый порядок аппроксимации по t и второй по пространственной переменной, а также является абсолютно устойчивой.

Диагональные элементы в системах (9.24), (9.25) преобладают, следовательно, процесс прогонки устойчив. К достоинствам метода, достигнутым благодаря введению промежуточного слоя, следует отнести расщепление исходной задачи на две более простые, решаемые с помощью алгоритма скалярной прогонки. К недостаткам метода относится сравнительно низкая точность из-за того, что на каждом дробном шаге аппроксимируется только один пространственный оператор.

Методика расчетов

1. Задать сетку $D_h = \{x_i, y_j, t^n\}$, где $x_i = ih_x, y_j = jh_y, t^n = n\tau, i, j = 0, \pm 1, \pm 2, \dots, n = 0, 1, \dots, N; T = N\tau$. Величины шагов в силу абсолютной устойчивости метода выбрать из условия удовлетворения точности расчетов.

2. Вычислить

$$\hat{u}_{i,j}^0 = \hat{v}_{i,j}^0 = \psi_{i,j}, \quad i, j = 0, \pm 1, \pm 2, \dots,$$

где $\psi_{i,j} = \psi(x_i, y_j)$.

Положить $n = 0$.

3. Найти решение на $(n + 1)$ -м слое.

Первый этап. Перейти на дополнительный промежуточный $\left(n + \frac{1}{2}\right)$ -й слой с шагом $\frac{\tau}{2}$: для каждого j решить трехдиагональную систему линейных алгебраических уравнений с учетом п. 2:

$$\frac{\tau}{2h_x^2} \hat{v}_{i-1,j}^{n+\frac{1}{2}} - \left(1 + \frac{\tau}{h_x^2}\right) \hat{v}_{i,j}^{n+\frac{1}{2}} + \frac{\tau}{2h_x^2} \hat{v}_{i+1,j}^{n+\frac{1}{2}} = -\hat{u}_{i,j}^n, \quad i, j = 0, \pm 1, \pm 2, \dots$$

Искомые значения $\hat{v}_{i-1,j}^{n+\frac{1}{2}}, \hat{v}_{i,j}^{n+\frac{1}{2}}, \hat{v}_{i+1,j}^{n+\frac{1}{2}}$ на промежуточном слое вычисляются методом прогонки по направлению x .

Второй этап. Перейти на $(n + 1)$ -й временной слой с промежуточного $\left(n + \frac{1}{2}\right)$ -го с шагом $\frac{\tau}{2}$: для каждого значения i решить трехдиагональную систему линейных алгебраических уравнений

$$\frac{\tau}{2h_y^2} \hat{w}_{i,j-1}^{n+1} - \left(1 + \frac{\tau}{h_y^2}\right) \hat{w}_{i,j}^{n+1} + \frac{\tau}{2h_y^2} \hat{w}_{i,j+1}^{n+1} = -\hat{v}_{i,j}^{n+\frac{1}{2}}, \quad i, j = 0, \pm 1, \pm 2, \dots$$

Искомые значения $\hat{w}_{i,j-1}^{n+1}, \hat{w}_{i,j}^{n+1}, \hat{w}_{i,j+1}^{n+1}$ на целом слое вычисляются методом прогонки по направлению y .

Положить $\hat{u}_{i,j}^{n+1} = \hat{w}_{i,j}^{n+1}, i, j = 0, \pm 1, \pm 2, \dots$

4. Если $n = N$, вычисления завершить. Иначе положить $n = n + 1$ и перейти к п. 3. ■

Замечания

1. В приведенной методике для простоты можно сразу положить $\hat{v}_{i,j}^{n+\frac{1}{2}} = \hat{u}_{i,j}^{n+\frac{1}{2}}$, $\hat{w}_{i,j}^{n+1} = \hat{u}_{i,j}^{n+1}$.

2. При применении метода дробных шагов к решению уравнения с тремя пространственными переменными

$$u_t = u_{xx} + u_{yy} + u_{zz}$$

используются два промежуточных слоя: $t = t^{n+\frac{1}{3}}$ и $t = t^{n+\frac{2}{3}}$. Тогда на дробных шагах реализуется следующая аппроксимация дифференциальных операторов:

$$\begin{aligned} \frac{\hat{v}_{i,j,k}^{n+\frac{1}{3}} - \hat{v}_{i,j,k}^n}{\frac{\tau}{3}} &= \frac{\hat{v}_{i+1,j,k}^{n+\frac{1}{3}} - 2\hat{v}_{i,j,k}^{n+\frac{1}{3}} + \hat{v}_{i-1,j,k}^{n+\frac{1}{3}}}{h_x^2}, \quad \hat{v}_{i,j,k}^n = \hat{u}_{i,j,k}^n; \\ \frac{\hat{w}_{i,j}^{n+\frac{2}{3}} - \hat{w}_{i,j}^{n+\frac{1}{3}}}{\frac{\tau}{3}} &= \frac{\hat{w}_{i,j+1,k}^{n+\frac{2}{3}} - 2\hat{w}_{i,j,k}^{n+\frac{2}{3}} + \hat{w}_{i,j-1,k}^{n+\frac{2}{3}}}{h_y^2}, \quad \hat{w}_{i,j}^{n+\frac{1}{3}} = \hat{v}_{i,j,k}^{n+\frac{1}{3}}; \\ \frac{\hat{u}_{i,j,k}^{n+1} - \hat{u}_{i,j,k}^{n+\frac{2}{3}}}{\frac{\tau}{3}} &= \frac{\hat{u}_{i,j,k+1}^{n+1} - 2\hat{u}_{i,j,k}^{n+1} + \hat{u}_{i,j,k-1}^{n+1}}{h_z^2}, \quad \hat{u}_{i,j,k}^{n+\frac{2}{3}} = \hat{w}_{i,j,k}^{n+\frac{2}{3}}. \end{aligned}$$

3. Метод дробных шагов имеет недостаток, состоящий в том, что в нем имеет место только частичная аппроксимация пространственных операторов на каждом дробном временном слое.

Задачи для самостоятельного решения

Построить разностную схему для решения начально-краевых задач 1–3.

1. $u_t + u_x + u_y = t + 1$, $(x, y, t) \in (0, L) \times (0, M) \times (0, T)$;

$$\begin{aligned} u(x, y, 0) &= x + y, \quad (x, y) \in [0, L] \times [0, M]; \\ u(0, y, t) &= y + t^2, \quad y \in [0, M], \quad 0 < t \leq T; \\ u(x, 0, t) &= xt, \quad x \in [0, L], \quad 0 < t \leq T, \end{aligned}$$

где L, M, T — заданные числа.

2. $u_t = u_{xx} + 3u_{yy}$, $(x, y, t) \in (0, L) \times (0, M) \times (0, T)$;

$$\begin{aligned} u(x, y, 0) &= 2x + y, \quad (x, y) \in [0, L] \times [0, M]; \\ u(x, y, t) &= 2, \quad (x, y) \in \Gamma, \quad 0 < t \leq T, \end{aligned}$$

где $\Omega = (0, L) \times (0, M)$; Γ — граница области Ω ; L, M, T — заданные числа.

3. $u_{tt} = 4(u_{xx} + u_{yy})$, $(x, y, t) \in \Omega \times (0, T)$;

$$\begin{aligned} u(x, y, 0) &= x^2 + y^2, \quad (x, y) \in \Omega \cup \Gamma; \\ \frac{\partial u(x, y, 0)}{\partial t} &= 1, \quad (x, y) \in \Omega \cup \Gamma; \\ u(x, y, t) &= t, \quad (x, y) \in \Gamma, \quad 0 < t \leq T, \end{aligned}$$

где $\Omega = (0, L) \times (0, M)$, Γ — граница области Ω ; L, M, T — заданные числа.

4. Построить схему метода расщепления для решения задачи Коши:

$$\begin{aligned} u_t &= 2u_{xx} + 3u_{yy}, \quad (x, y, t) \in R \times R \times (0, T); \\ u(x, y, 0) &= x + 2y, \quad (x, y) \in R^2, \end{aligned}$$

где T — заданное положительное число.

5. Построить разностную схему метода переменных направлений для решения первой начально-краевой задачи:

$$\begin{aligned} u_t &= u_{xx} + u_{yy} + 3t, \quad (x, y, t) \in (0, L) \times (0, M) \times (0, T); \\ u(x, y, 0) &= 2x + 5y, \quad (x, y) \in [0, L] \times [0, M]; \\ u(x, y, t) &= \sin t, \quad (x, y) \in \Gamma, \quad 0 < t \leq T, \end{aligned}$$

где $\Omega = (0, L) \times (0, M)$; Γ — граница области Ω ; L, M, T — заданные числа.

6. Построить схему метода дробных шагов для решения задачи Коши:

$$\begin{aligned} u_t &= 2u_{xx} + u_{yy}, \quad (x, y, t) \in R \times R \times (0, T); \\ u(x, y, 0) &= x^2 + y^2, \quad (x, y) \in R^2, \end{aligned}$$

где T — заданное положительное число.

ЛИТЕРАТУРА

1. Алберг, Дж. Теория сплайнов и ее приложения / Дж. Алберг, Э. Нильсон, Дж. Уолш. — М. : Мир, 1972.
2. Бахвалов, Н. С. Численные методы / Н. С. Бахвалов, Н. П. Жидков, Г. М. Кобельков. — М. : Наука, 1983.
3. Березин, И. С. Методы вычислений / И. С. Березин, Н. П. Жидков. — Т. 1, 2. — М. : Наука, 1960.
4. Боглаев, Ю. П. Вычислительная математика и программирование. — М. : Высш. шк., 1990.
5. Вержбицкий, В. М. Основы численных методов. — М. : Высш. шк., 2002.
6. Волков, Е. А. Численные методы. — СПб. : Лань, 2008.
7. Годунов, С. К. Уравнения математической физики. — М. : Наука, 1971.
8. Годунов, С. К. Введение в теорию разностных схем / С. К. Годунов, В. С. Рябенький. — М. : Физматгиз, 1962.
9. Демидович, Б. П. Основы вычислительной математики / Б. П. Демидович, И. А. Марон. — СПб. : Лань, 2008.
10. Демидович, Б. П. Численные методы анализа / Б. П. Демидович, И. А. Марон, Э. З. Шувалова. — М. : Физматгиз, 1963.
11. Денис, Дж. Численные методы безусловной оптимизации и решения нелинейных уравнений / Дж. Денис, Р. Шнабель. — М. : Мир, 1988.
12. Дьяконов, В. П. Справочник по алгоритмам и программам на языке Бейсик для персональных ЭВМ. — М. : Наука, 1987.
13. Завьялов, Ю. С. Методы сплайн-функций / Ю. С. Завьялов, Б. И. Квасов, В. Л. Мирошниченко. — М. : Наука, 1980.
14. Калиткин, Н. Н. Численные методы. — М. : Наука, 1978.
15. Киреев, В. И. Интегродифференциальный метод приближения функций алгебраическими многочленами // Вычислительные технологии. — Новосибирск, ИВТ СО РАН. — 1993. — Т. 2. — № 6.
16. Киреев, В. И. Интегральный метод приближения функций алгебраическими многочленами и биквадратными сплайнами // Вестник МАИ. — 1994. — Т. 1. — № 1.
17. Киреев, В. И. Интегродифференциальные сплайны и их приложение к численному дифференцированию и интегрированию. — Деп. в ВИНИТИ : № 1675–В93, 1993.
18. Киреев, В. И. Многошаговые дискретные схемы и явные, последовательные сплайн-методы решения обыкновенных дифференциальных уравнений на нерегулярном шаблоне // Вычислительные технологии. — Новосибирск, ИВТ СО РАН. — 1996. — Т. 1. — № 2.

19. Киреев, В. И. Численные методы решения задач алгебры и теории приближений. — М. : Изд-во МАИ, 1991.
20. Киреев, В. И. Интегродифференциальные сплайны и их применение в интерполяции, численном дифференцировании и интегрировании / В. И. Киреев, Т. К. Бирюкова // Вычислительные технологии. — Новосибирск, ИВТ СО РАН. — 1995. — Т. 4. — № 10.
21. Киреев, В. И. Полиномиальные интегродифференциальные одномерные и двумерные сплайны в задачах интерполяирования и сглаживания одномерных и двумерных функций / В. И. Киреев, Т. К. Бирюкова // Вестник МАИ. — 1997. — Т. 4. — № 2.
22. Киреев, В. И. Численное моделирование газодинамических течений / В. И. Киреев, А. С. Войновский. — М. : Изд-во МАИ, 1991.
23. Киреев, В. И. Численные методы в примерах и задачах / В. И. Киреев, А. В. Пантелеев. — М. : Высш. шк., 2008.
24. Киреев, В. И. Вычисление интегралов по явным квадратурным формулам и неявным алгоритмам, построенным на основе интегродифференциальных сплайн-функций на нерегулярной сетке / В. И. Киреев, Г. В. Цирков. — Деп. в ВИНТИ : № 3206-В97, 1997.
25. Корнейчук, Н. П. Сплайны в теории приближения. — М. : Наука, 1984.
26. Крылов, В. И. Вычислительные методы : в 2-х т. / В. И. Крылов, В. В. Бобков, П. И. Монастырный. — М. : Наука, 1976–1977.
27. Мак-Кракен, Д. Численные методы и программирование на ФОРТРАНе / Д. Мак-Кракен, У. Дорн. — М. : Мир, 1977.
28. Мысовских, И. П. Интерполяционные кубатурные формулы. — М. : Наука, 1981.
29. Никольский, С. М. Квадратурные формулы. — М. : Наука, 1979.
30. Пантелеев, А. В. Теория управления в примерах и задачах / А. В. Пантелеев, А. С. Бортаковский. — М. : Высш. шк., 2003.
31. Пантелеев, А. В. Методы оптимизации в примерах и задачах / А. В. Пантелеев, Т. А. Летова. — М. : Высш. шк., 2008.
32. Пантелеев, А. В. Теория функций комплексного переменного и операционное исчисление в примерах и задачах / А. В. Пантелеев, А. С. Якимова. — М. : Высш. шк., 2007.
33. Пантелеев, А. В. Обыкновенные дифференциальные уравнения в примерах и задачах / А. В. Пантелеев, А. С. Якимова, А. В. Босов. — М. : Высш. шк., 2001.
34. Плис, А. И. Лабораторный практикум по высшей математике / А. И. Плис, Н. А. Сливина. — М. : Высш. шк., 1994.
35. Пирумов, У. Г. Численные методы. — М. : Изд-во МАИ, 1998.
36. Пирумов, У. Г. Численные методы газовой динамики / У. Г. Пирумов, Г. С. Ростяков. — М. : Высш. шк., 1987.
37. Ракитин, В. И. Практическое руководство по методам вычислений с применением программ для персональных компьютеров / В. И. Ракитин, В. Е. Первушкин. — М. : Высш. шк., 1998.
38. Рябенький, В. С. Введение в вычислительную математику. — М. : Наука, 1994.
39. Самарский, А. А. Теория разностных схем. — М. : Наука, 1983.
40. Самарский, А. А. Численные методы / А. А. Самарский, А. В. Гулин. — М. : Наука, 1989.
41. Саульев, В. К. Современные вычислительные методы. — М. : Изд-во МАИ, 1977.
42. Сборник задач по методам вычислений / под ред. П. И. Монастырного. — М. : Физматлит, 1994.
43. Сигорский, В. П. Математический аппарат инженера. — Киев : Техніка, 1977.
44. Стечкин, С. Б. Сплайны в вычислительной математике / С. Б. Стечкин, Ю. Н. Субботин. — М. : Наука, 1976.
45. Фаддеев, Д. К. Вычислительные методы линейной алгебры / Д. К. Фаддеев, В. Н. Фаддеева. — СПб. : Лань, 2009.
46. Фарлоу, С. Уравнения с частными производными для научных работников и инженеров. — М. : Мир, 1985.

47. *Формалев, В. Ф.* Численные методы / В. Ф. Формалев, Д. Л. Ревизников. — М. : Физматлит, 2006.
48. *Хайрер, Э.* Решение обыкновенных дифференциальных уравнений. Нежесткие задачи / Э. Хайрер, С. Нерсетт, Г. Ваннер. — М. : Мир, 1990.
49. *Холл, Дж.* Современные численные методы решения обыкновенных дифференциальных уравнений / Дж. Холл, Дж. Уатт. — М. : Мир, 1979.
50. Численное решение многомерных задач газовой динамики / под ред. С. К. Годунова. — М. : Наука, 1976.
51. *Яненко, Н. Н.* Метод дробных шагов решения многомерных задач математической физики. — Новосибирск : Наука, 1967.
52. *Butcher, J. Ch.* The numerical analysis of ordinary differential equations: Runge–Kutta and general linear methods. — Chichester : Wiley, 1987.
53. *Fatunla, S. O.* Numerical methods for initial value problems in ordinary differential equations. — Academic Press, 1988.
54. *Gear, C. W.* Numerical initial value problems in ordinary differential equations. — Prentice-Hall, 1971.
55. *Hairer, E.* Solving ordinary differential equations. Nonstiff problems / E. Hairer, S. P. Norsett, G. Wanner. — Springer-Verlag, 1991.
56. *Henrici, P.* Elements of numerical analysis. — New York : McGraw-Hill Book Co., 1964.
57. *Lambert, J. D.* Computational methods in ODEs. — New York : John Wiley, 1973.
58. *Lapidus, L.* Numerical solution of ordinary differential equations / L. Lapidus, J. Seinfeld. — New York : Academic Press, 1971.
59. *Schoenberg, I. J.* Cardinal interpolation and spline functions // J. Approximation Theory. — 1969. — V. 2. — № 2. — P. 167–206.

ОГЛАВЛЕНИЕ

Введение	3
B.1. Понятие о численных методах	3
B.2. Погрешности вычислений	6
B.3. Некоторые понятия математического анализа, использующиеся в численных методах	12
B.4. Дискретизация и принцип соответствия порядков аппроксимации дискретных моделей для интегралов, функций и производных	14

ЧАСТЬ ПЕРВАЯ ЧИСЛЕННЫЕ МЕТОДЫ АЛГЕБРЫ

<i>Г л а в а п е р в а я</i>	
Методы решения систем линейных алгебраических уравнений	20
1.1. Постановка задачи	20
1.2. Прямые методы	22
1.2.1. Метод Гаусса	22
1.2.2. Метод прогонки	27
1.2.3. Метод LU-разложения	32
1.3. Итерационные методы	37
1.3.1. Метод простых итераций	37
1.3.2. Метод Зейделя	45

<i>Г л а в а в т о р а я</i>	
Методы решения задач о собственных значениях и собственных векторах матриц	55
2.1. Постановка задачи	55
2.2. Метод непосредственного развертывания	56
2.3. Метод итераций	59
2.4. Метод вращений	62

Г л а в а т р е т ъ я

Методы решения нелинейных уравнений и систем	70
3.1. Методы решения нелинейных уравнений	70
3.1.1. Постановка задачи	70
3.1.2. Отделение корней	72
3.1.3. Метод половинного деления	77
3.1.4. Метод хорд	80
3.1.5. Метод простых итераций	83
3.1.6. Метод Ньютона	89
3.1.7. Модификации метода Ньютона	95
3.2. Методы решения систем нелинейных уравнений	99
3.2.1. Постановка задачи	99
3.2.2. Метод простых итераций	100
3.2.3. Метод Зейделя	104
3.2.4. Метод Ньютона	105
3.2.5. Модификации метода Ньютона	111

Ч А С Т Ъ В Т О Р А Я

ЧИСЛЕННЫЕ МЕТОДЫ ТЕОРИИ ПРИБЛИЖЕНИЙ

Г л а в а ч е т в е р т а я

Методы приближения сеточных функций	118
4.1. Общая постановка задачи и классификация методов	118
4.2. Методы функциональной интерполяции	121
4.2.1. Постановка задачи	121
4.2.2. Многочлен Лагранжа	123
4.2.3. Многочлены Ньютона	133
4.3. Методы интегрально-дифференциальной интерполяции	144
4.3.1. Постановка задачи	144
4.3.2. Интерполяционный параболический интегрально-функциональный многочлен	144
4.3.3. Интерполяционный параболический интегрально-дифференциальный многочлен	149
4.4. Методы интегрального сглаживания	150
4.4.1. Постановка задачи	150
4.4.2. Метод наименьших квадратов	153
4.4.3. Метод наилучшего интегрального приближения	165
4.5. Методы интерполяции и сглаживания на основе сплайнов	170
4.5.1. Постановка задачи и основные положения	170
4.5.2. Интерполяционные дифференциальные кубические сплайны	174
4.5.3. Интерполяционные дифференциальные параболические сплайны	181
4.5.4. Восстанавливающие, интерполяционные и сглаживающие интегрально-дифференциальные параболические сплайны	186
4.5.5. Слабо сглаживающие интерполяционные интегрально-дифференциальные параболические сплайны	192

Г л а в а п я т а я

Методы численного дифференцирования и интегрирования	198
5.1. Постановка задачи и принципы конструирования аппроксимационных формул	198
5.2. Методы численного дифференцирования	201
5.2.1. Формулы, полученные на основе разложения функций по формуле Тейлора	201
5.2.2. Формулы, полученные на основе разложения первообразных по формуле Тейлора	214
5.2.3. Формулы, полученные на основе сплайнов	218
5.3. Методы численного интегрирования	226
5.3.1. Формулы, полученные на основе интерполяционных многочленов	226
5.3.2. Формулы, полученные на основе сплайнов	237
5.3.3. Формулы, полученные на основе разложения первообразных по формуле Тейлора	244
5.4. Метод Рунге уточнения результатов численного дифференцирования и интегрирования	247

ЧАСТЬ ТРЕТЬЯ

ЧИСЛЕННЫЕ МЕТОДЫ РЕШЕНИЯ ОБЫКНОВЕННЫХ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ

Г л а в а ш е с т а я

Методы решения задачи Коши	252
6.1. Постановка задачи и основные положения	252
6.2. Принципы построения разностных схем	261
6.2.1. Принцип аппроксимаций	261
6.2.2. Интегрально-интерполяционный принцип	268
6.2.3. Принцип согласования с разложением по формуле Тейлора	278
6.2.4. Принцип аналогий	285
6.3. Составные схемы	290
6.4. Экстраполяционные методы	297
6.5. Непрерывно-дискретные методы	303
6.5.1. Конструирование последовательных сплайн-методов	303
6.5.2. Схема второго порядка	304
6.5.3. Схема третьего порядка	306

Г л а в а с е д ь м а я

Методы решения краевых задач	311
7.1. Постановка задачи и основные положения	311
7.2. Метод сеток	314
7.3. Методы минимизации невязки	319
7.4. Методы сведения краевой задачи к задаче Коши	327
7.4.1. Метод стрельбы	327
7.4.2. Метод дифференциальной прогонки	331
7.5. Метод конечных элементов	333

ЧАСТЬ ЧЕТВЕРТАЯ
ЧИСЛЕННЫЕ МЕТОДЫ РЕШЕНИЯ
ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ
В ЧАСТНЫХ ПРОИЗВОДНЫХ

Г л а в а в о с ь м а я

Численные методы решения уравнений математической физики с двумя независимыми переменными	346
8.1. Постановка задачи и основные положения	346
8.2. Принципы построения разностных схем	356
8.3. Разностные методы решения уравнений первого порядка	364
8.4. Разностные методы решения уравнений второго порядка	370
8.4.1. Разностные методы решения уравнений параболического типа	370
8.4.2. Разностные методы решения уравнений гиперболического типа	377
8.4.3. Разностные методы решения уравнений эллиптического типа	383
8.5. Метод прямых	387
8.6. Метод характеристик	391
8.7. Метод Годунова	403

Г л а в а д е в я т а я

Численные методы решения уравнений математической физики с тремя независимыми переменными	415
9.1. Постановка задачи и основные положения	415
9.2. Разностный метод	419
9.3. Метод расщепления	425
9.4. Метод переменных направлений	430
9.5. Метод дробных шагов	435

Литература	441
-----------------------------	------------

*Владимир Иванович КИРЕЕВ,
Андрей Владимирович ПАНТЕЛЕЕВ*

ЧИСЛЕННЫЕ МЕТОДЫ В ПРИМЕРАХ И ЗАДАЧАХ

Учебное пособие

Издание четвертое, исправленное

Зав. редакцией
физико-математической литературы *Н. Р. Нигмадзянова*
Ответственный редактор *Н. В. Черезова*
Технический редактор *О. О. Николаева*
Корректор *Т. С. Симонова*
Подготовка иллюстраций *А. П. Маркова*
Верстка *М. И. Хетерели*
Выпускающие *Н. А. Крылова, О. В. Шилкова*

ЛР № 065466 от 21.10.97
Гигиенический сертификат 78.01.07.953.П.007216.04.10
от 21.04.2010 г., выдан ЦГСЭН в СПб

Издательство «ЛАНЬ»
lan@lanbook.ru; www.lanbook.com
192029, Санкт-Петербург, Общественный пер., 5.
Тел./факс: (812) 412-29-35, 412-05-97, 412-92-72.
Бесплатный звонок по России: 8-800-700-40-71

ГДЕ КУПИТЬ
по России и зарубежью
«ЛАНЬ-ТРЕЙД». 192029, Санкт-Петербург, ул. Крупской, 13
тел.: (812) 412-85-78, 412-14-45, 412-85-82; тел./факс: (812) 412-54-93
e-mail: trade@lanbook.ru; ICQ: 446-869-967; www.lanpb.spb.ru/price.htm

в Москве и в Московской области
«ЛАНЬ-ПРЕСС». 109263, Москва, 7-я ул. Текстильщиков, д. 6/19
тел.: (499) 178-65-85; e-mail: lanpress@lanbook.ru

в Краснодаре и в Краснодарском крае
«ЛАНЬ-ЮГ». 350901, Краснодар, ул. Жлобы, д. 1/1
тел.: (861) 274-10-35; e-mail: lankrd98@mail.ru

ДЛЯ РОЗНИЧНЫХ ПОКУПАТЕЛЕЙ:
интернет-магазины:
Издательство «Лань»: <http://www.lanbook.com>
«Сова»: <http://www.symplex.ru>; «Ozon.ru»: <http://www.ozon.ru>
«Библион»: <http://www.biblion.ru>

Подписано в печать 20.05.15.
Бумага офсетная. Гарнитура Школьная. Формат 70×100¹/₁₆.
Печать офсетная. Усл. п. л. 36,40. Тираж 700 экз.

Заказ № .

Отпечатано в ОАО «Первая образцовая типография»,
филиал «Чеховский Печатный Двор»
в полном соответствии с качеством предоставленного оригинал-макета.
142300, Московская обл., г. Чехов, ул. Полиграфистов, д. 1.
Тел.: (495) 988-63-76, факс: 8 (496) 726-54-10.