

Э. Хайрер  
С. Нёрсетт  
Г. Ваннер

Решение  
обыкновенных  
дифференциальных  
уравнений

Нежесткие задачи

Перевод с английского  
И. А. Кульчицкой и С. С. Филиппова  
под редакцией С. С. Филиппова



Москва «Мир» 1990

ББК 22.193 + 22.161.6

X15

УДК 517.91

Хайрер Э., Нёрсетт С., Ваннер Г.

Х15 Решение обыкновенных дифференциальных уравнений.  
Нежесткие задачи: Пер. с англ. — М.: Мир, 1990. — 512 с.,  
ил.

ISBN 5-03-001179-X

Книга известных математиков (Швейцария, Норвегия), дающая картину современного состояния теории и практики численного решения обыкновенных дифференциальных уравнений. Изложены основные теоретические результаты, приведены наиболее употребительные численные методы, дано большое число примеров практических применений в физике и прикладных науках. Представлены тексты программ на Фортране.

Для математиков-прикладников и всех, кто в своей работе встречается с решением дифференциальных уравнений, для аспирантов и студентов вузов.

X 1602070000-031 31-90  
041 (01)-90

ББК 22.193 + 22.161.6

*Редакция литературы по математическим наукам*

ISBN 5-03-001179-X (русск.)  
ISBN 3-540-17145-2 (англ.)

© Springer-Verlag Berlin Heidelberg 1987.  
All rights reserved. Authorized translation  
from English language edition published  
by Springer-Verlag Berlin Heidelberg New York Tokyo  
© перевод на русский язык, «Мир», 1990

## Предисловие к русскому изданию

Первые численные методы решения обыкновенных дифференциальных уравнений изобрели еще Ньютона и Эйлера. К началу XX века были уже известны ставшие теперь классическими методы Адамса и Рунге—Кутты. Текущий период характерен бурным развитием вычислительной техники и усиленным применением ЭВМ и численных методов для решения резко расширяющегося круга задач. Изменение требований породило новую волну конструирования и исследования численных методов интегрирования обыкновенных дифференциальных уравнений, не ослабевшую и сегодня. Ранее известные методы были детально изучены и обобщены, построены новые классы методов, созданы методы, ориентированные на решение задач со специальными свойствами, например, так называемых жестких систем. Большой прогресс достигнут и в разработке удобных для пользователей, эффективных и надежных программ.

За последние 20—25 лет этим вопросам посвящено огромное число журнальных публикаций, множество монографий, как узкоспециальных, так и общего характера, а также учебных пособий (некоторые из них включены авторами в список литературы в конце книги). Однако почти все они изданы за рубежом. На русский язык были переведены лишь рассчитанная на математиков-профессионалов монография Штеттера (1973) и сборник обзорных лекций, прочитанных на летней школе в Англии (редакторы Холл и Уатт (1976)<sup>1)</sup>), да еще две-три книги по весьма частным вопросам. Важнейшие сведения о численных методах решения задач Коши для обыкновенных дифференциальных уравнений можно, конечно, найти в любом учебном пособии по численным методам. Но систематического и доступного широкому кругу читателей изложения названных выше вопросов на русском языке до сих пор не было.

Предлагаемая вниманию читателей книга призвана восполнить этот пробел. Написана она специалистами, которые внесли значительный вклад в развитие излагаемых в книге вопросов и имеют большой педагогический опыт. Книга доступна студентам университетов и других вузов, специализирующимся в области прикладной математики, и она, несомненно, будет полезна широкому кругу специалистов, работающих в области численных методов, а также всем лицам, занимающимся приложением этих методов.

Н. С. Бахвалов

<sup>1)</sup> См. в конце книги список литературы, добавленной редактором перевода.

## Предисловие

«Насколько я помню, я никогда не встречал авторского предисловия, которое имело бы какую-то иную цель, кроме одной — оправдать публикацию Книги.»  
(Марк Твен)

«Афоризм Гаусса: «Когда здание закончено, никто не должен заметить какие-либо следы строительных лесов», — часто используется математиками для оправдания того, что они не уделяют внимания мотивировке своей работы и истории той области, в которой они работают. К счастью, крепнет противоположное отношение, и многочисленные авторские отступления в этом очерке показывают, на чьей стороне мои симпатии.»  
(Б. Б. Мандельброт, 1982)

«Это дает нам хороший повод закончить большую часть книги до нового года.»

(из письма авторов в издательство Шпрингер, датированного 29 октября 1980 г.)

Книга в двух томах: в законченном сейчас первом томе рассматриваются нежесткие уравнения, а жесткие — в готовящемся втором. Первый том содержит три главы, одна посвящена классической математической теории, вторая — методам Рунге—Кутты и экстраполяционным, третья — многошаговым методам. Приложение содержит несколько программ на фортране, написанных нами для расчетов примеров, приведенных в книге.

Каждая глава делится на разделы, и нумерация формул, теорем, таблиц и рисунков в каждом разделе своя; она включает также номер раздела, но не главы. Ссылки на другие главы редки, они специально оговариваются. Конец каждого доказательства обозначен буквами «QED» (*quod erat demonstrandum* — что и требовалось доказать (лат.)). Поскольку это выглядит довольно нелепо, когда доказательство отсутствует, либо оно предшествует теореме, то мы писали «NED» (*nihil erat demonstrandum*) в таких случаях<sup>1)</sup>. Библиографические ссылки даются указанием автора и (в скобках) года. Список литературы не претендует на полноту, он содержит, в основном, работы, которые обсуждаются в тексте.

В заключение мы хотим поблагодарить всех, кто нам помогал и поддерживал при подготовке этой книги. Первый импульс к написанию этой книги дал чудесный «минисимпозиум», который

■ 1) В русском переводе «QED» и «NED» в конце доказательств или формулировок теорем заменены значком ■. — Прим. ред.

## Предисловие

7

Г. Далквист организовал в Стокгольме в 1979 г. Всю рукопись очень тщательно прочли И. Стайниг и Хр. Любих, которым мы обязаны чрезвычайно цennыми математическими и лингвистическими советами. Мы благодарны также И. П. Экманну за пре-восходные программные средства, с помощью которых окончательно редактировалась и печаталась рукопись. На более ранней стадии мы пользовались программами для работы с текстами, которые написал Р. Менк. Благодарим за помощь также персонал Женевского вычислительного центра. Все компьютерные графики были сделаны на их великолепном прецизионном плоттере. И последнее, хотя не менее важное обстоятельство — это приятное сотрудничество с группой планирования и выпуска издательства Шпрингер, которой мы хотим выразить свою признательность.

29 октября 1986 г.

Авторы

## Глава I

# Классическая математическая теория

«... halte ich es immer für besser, nicht mit dem Anfang anzufangen, der immer das Schwerste ist».

«...я всегда считал, что лучше начинать не с начала, так как оно самое трудное».

(Выписка из Ф. Шиллера в записной книжке Б. Римана)

Первая глава посвящена классической теории дифференциальных уравнений, которую мы считаем полезной и необходимой для глубокого понимания численных процессов и явлений. Кроме того, по ходу изложения удобно привести интересные примеры дифференциальных уравнений и их свойств.

Вначале в разд. I.2—I.6 мы проследим историческое развитие классических методов интегрирования с помощью разложения в ряд, квадратур и элементарных функций с самого их зарождения (Ньютона и Лейбница) и до эпохи Эйлера и Лагранжа. Затем (разд. I.7—I.14) речь пойдет о теоретических свойствах существования, единственности, устойчивости, непрерывности и дифференцируемости решений. Начало этой теории положил Коши в 1824 г., а завершенный вид она приобрела в основном за последующие 100 лет. В конце главы мы дадим краткий обзор еще нескольких тем: краевых задач, периодических решений, предельных циклов и странных аттракторов (разд. I.15 и I.16).

## I.1. Терминология

*Дифференциальное уравнение первого порядка* — это уравнение вида

$$y' = f(x, y) \quad (1.1)$$

с заданной функцией  $f(x, y)$ . Функция  $y(x)$  называется *решением* этого уравнения, если при всех  $x$  выполняется равенство

$$y'(x) = f(x, y(x)). \quad (1.2)$$

Еще Ньютона, Лейбница и Эйлера заметили, что решение обычно содержит свободный параметр и поэтому определяется единственным образом только тогда, когда задано *начальное значение*

$$y(x_0) = y_0. \quad (1.3)$$

В разд. I.7 мы рассмотрим принадлежащее Коши доказательство существования и единственности такого решения. Дифференциальные уравнения возникают во многих приложениях. В разд. I.2 мы увидим, как первые такие уравнения появились в механике, а в разд. I.3 будет показано, как некоторые из них можно решить в явном виде.

*Дифференциальное уравнение второго порядка* для  $y$  имеет вид

$$y'' = f(x, y, y'). \quad (1.4)$$

Здесь решение обычно содержит *два* параметра и определяется однозначно только по *двум* начальным значениям:

$$y(x_0) = y_0, \quad y'(x_0) = y'_0. \quad (1.5)$$

Уравнения второго порядка редко допускают решение в явном виде (см. I.3). Как при их численном решении, так и в теоретических исследованиях обычно вводят переменные  $y_1(x) := y(x)$ ,  $y_2(x) := y'(x)$  и приводят уравнение (1.4) к системе

$$\begin{aligned} y'_1 &= y_2, & y_1(x_0) &= y_0, \\ y'_2 &= f(x, y_1, y_2), & y_2(x_0) &= y'_0. \end{aligned} \quad (1.4')$$

Это пример *системы дифференциальных уравнений первого порядка* размерности  $n$  (см. разд. I.6 и I.9):

$$\begin{aligned} y'_1 &= f_1(x, y_1, \dots, y_n), & y_1(x_0) &= y_{10}, \\ y'_n &= f_n(x, y_1, \dots, y_n), & y_n(x_0) &= y_{n0}. \end{aligned} \quad (1.6)$$

Большая часть этой книги посвящена теории решения задачи Коши для системы (1.6). В конце XIX века [Пeanо (1890)] стало обычным введение векторных обозначений

$$y = (y_1, \dots, y_n)^T, \quad f = (f_1, \dots, f_n)^T,$$

при которых (1.6) принимает вид  $y' = f(x, y)$ , что совпадает с (1.1), но теперь под  $y$  и  $f$  понимаются векторы.

Для уравнения второго порядка (1.4), помимо преобразования его к системе (1.4'), существует и другая возможность — построение методов, специально приспособленных к уравнениям второго порядка (методы Ньюстрёма). Этому будут посвящены специальные разделы книги. Конечно, ничто не мешает нам рассматривать (1.4) как систему уравнений второго порядка размерности  $n$ , интерпретируя  $y$  и  $f$  в (1.4) как векторы.

Если же начальные условия (1.5) заменены условиями типа  $y(x_0) = a$ ,  $y'(x_0) = b$ , т. е. когда не все условия, определяющие частное решение, заданы в одной и той же точке  $x_0$ , то говорят о *краевой задаче*. В этом случае теория существования решения и его численное нахождение значительно усложняются. В разд. I.15 мы приведем ряд соответствующих примеров.

В заключение определим *уравнение в частных производных* как задачу типа

$$\frac{\partial u}{\partial t} = f(t, u, \frac{\partial u}{\partial x}, \frac{\partial^2 u}{\partial x^2}) \quad (1.7)$$

для неизвестной функции  $u(t, x)$  двух независимых переменных. Можно также рассматривать уравнения в частных производных более высокого порядка, задачи с тремя и четырьмя независимыми переменными или системы уравнений в частных производных. Очень часто задачу Коши для таких уравнений удобно привести к системе обыкновенных дифференциальных уравнений, например, используя конечные разности или конечные элементы для приближения по переменной  $x$ . Таким образом, уравнение

$$\frac{\partial u}{\partial t} = a^2 \frac{\partial^2 u}{\partial x^2}$$

приняло бы вид

$$\frac{du_i}{dt} = \frac{a^2}{\Delta x^2} (u_{i+1} - 2u_i + u_{i-1}).$$

Данная процедура называется «методом прямых», или «методом частичной дискретизации» (по пространственной переменной). В разд. I.6 мы увидим, что эта и обратная ей процедура весьма часто употребляется, начиная с работ Лагранжа и Фурье. Аналогичная идея лежит и в основе «метода дискретизации по времени» [Роте (1930)].

## I.2. Наиболее ранние дифференциальные уравнения

«...Такова, например, задача обратного метода касательных; Декарт тоже признавался, что не смог одолеть ее. (Лейбниц, 27 августа 1676 г.)

«...а, как известно, все другие Изобретатели не имеют права на Изобретение.» (Ньютон, 29 мая 1716 г.)

«Вовсе не очевидно, что г-н Ньютон раньше, чем я, получил характеристику и алгоритм для бесконечно малых...» (Лейбниц)

«И этим он признал, что еще не нашел способа, как свести задачи к дифференциальным уравнениям.» (Ньютон)

### Ньютон

Дифференциальные уравнения так же стары, как дифференциальное исчисление. Ньютон рассмотрел их в своем труде о дифференциальном исчислении [Ньютон (1671)] и исследовал их решение с помощью интегрирования и разложения в ряд. Среди изученных Ньютоном уравнений первого порядка одним из самых

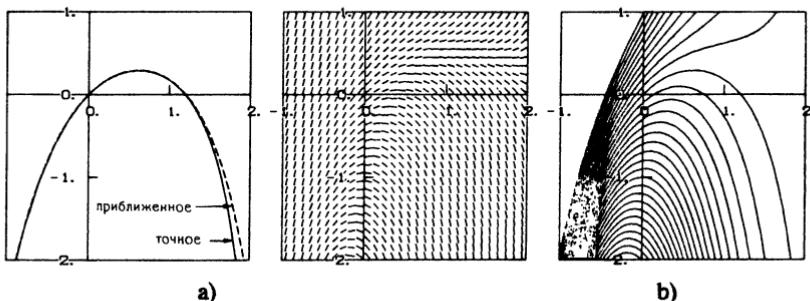


Рис. 2.1. а) Точное решение и приближенное решение; б) векторное поле; в) различные кривые, служащие решениями уравнения (2.1).

ранних был пример [см. работу Ньютона (1671), проблема II, решение второго случая II, пример I]

$$y' = 1 - 3x + y + x^2 + xy. \quad (2.1)$$

Каждому значению  $x$  и  $y$  такое уравнение ставит в соответствие производную решений  $y'$ . Поэтому мы получаем *векторное поле*, которое для данного конкретного уравнения изображено на рис. 2.1. Итак, вопреки распространенному мнению, векторные

поля существовали задолго до Ван Гога. Решением являются кривые, которые всюду отвечают этим заданным направлениям.

Ньютон исследовал решение данного уравнения с помощью бесконечного ряда, члены которого он получал рекурсивно («...и они кидаются к рядам, где г-н Ньютон без труда опередил меня; но...», Лейбниц). Первый член — это начальное значение при  $x = 0$ :

$$y = 0 + \dots .$$

Подставляя его в дифференциальное уравнение (2.1), получим

$$y' = 1 + \dots ,$$

что при интегрировании дает

$$y = x + \dots .$$

Теперь, снова используя (2.1), придем к соотношению

$$y' = 1 - 3x + x + \dots = 1 - 2x + \dots ,$$

интегрируя которое, получим

$$y = x - x^2 + \dots .$$

На следующем этапе

$$y' = 1 - 2x + x^3 + \dots , \quad y = x - x^2 + \frac{x^3}{3} + \dots .$$

Продолжая этот процесс, мы в итоге придем к приближенному решению

$$y = x - xx + \frac{1}{3}x^3 - \frac{1}{6}x^4 + \frac{1}{30}x^5 - \frac{1}{45}x^6 \text{ и т. д., (2.2)}$$

график которого (при отбрасывании членов более высокого порядка) дан на рис. 2.1 вместе с точным решением. На графике можно увидеть, что приближенное решение близко к точному только при малых значениях  $x$ . Другие примеры приведены в упр. 11—13 из разд. I.3.

### Лейбниц

Второй изобретатель дифференциального исчисления, Лейбниц, пришел к дифференциальным уравнениям около 1676 г. при рассмотрении геометрических задач, таких как *обратная задача о касательной* [Лейбниц (1693)]: найти кривую  $y(x)$  по заданной длине касательной  $AB$ , например, если она всюду равна постоянной  $a$  (рис. 2.2). Такая постановка приводит к дифференциальному уравнению первого порядка:

$$y' = -\frac{y}{b} = -\frac{y}{\sqrt{a^2 - y^2}}. \quad (2.3)$$

Лейбниц натолкнулся на эту задачу благодаря серебряным карманным часам (*«horologio portatili suae thecae argenteae»*)  $B$ , которые тянут по столу за цепочку  $AB$  длины  $a$ . При этом он с гордостью заметил, что ни один из лучших ученых не мог раньше решить эту задачу. Отметим интересный факт: для обоснования

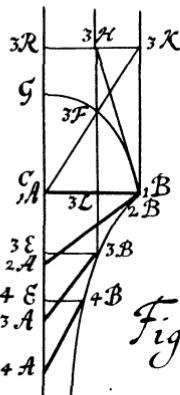


Fig. 1

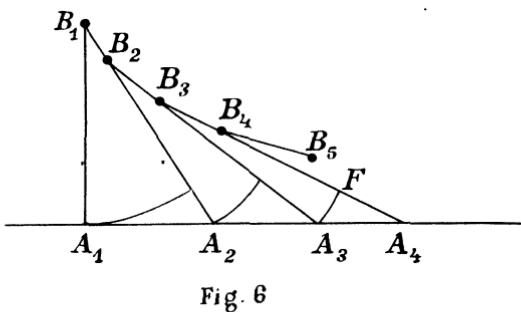


Fig. 6

Рис. 2.2 и 2.3. Иллюстрации из работ Лейбница и Ковалевского.

своего дифференциального уравнения (2.3) для карманных часов Лейбниц представил их движение *прямолинейным* на очень малых участках (рис. 2.3), что сегодня можно интерпретировать как метод Эйлера для численного решения задачи (2.3). В истории часто встречается такое любопытное явление, как с дифференциальными уравнениями, которые первоначально были получены с помощью процедуры, «обратной» к «ломаным Эйлерам». Другим примером этого явления служит вариационное исчисление.

### Вариационное исчисление

В 1696 г. Иоганн Бернулли предложил самим выдающимся математикам мира (*«Profundioris in primitis Mathesos cultori, Salutem!»*) решить задачу о *брахистохроне* (задачу наискорейшего спуска), в основном с целью посрамить своего брата Яакоба, решение которого, как он ожидал, будет неверным. Задача состоит в нахождении кривой  $y(x)$ , соединяющей две точки  $P_0$  и  $P_1$ , такой, что точка, скользящая по ней под действием силы тяжести, достигнет  $P_1$  за возможно кратчайшее время. Для решения этой задачи Иог. Бернулли (1697b) воспользовался оптической аналогией: он представил себе стопку тонких слоев однородных материалов; он знал, что луч света, распространяющийся со скоростью  $v$ , проходит такую стопку за кратчайшее время, если

выполняется закон Снеллиуса

$$\sin \alpha = Kv$$

(принцип Ферма). Поскольку скорость пропорциональна корню квадратному из потери высоты, при переходе ко все более тонким слоям он получил дифференциальное уравнение первого порядка

$$\sin \alpha = \frac{1}{\sqrt{1+y'^2}} = K \sqrt{2g(y-h)}. \quad (2.4)$$

Можно показать, что решениями этого уравнения являются циклоиды (см. упр. 7 из разд. I.3). Якоб в своем ответе также

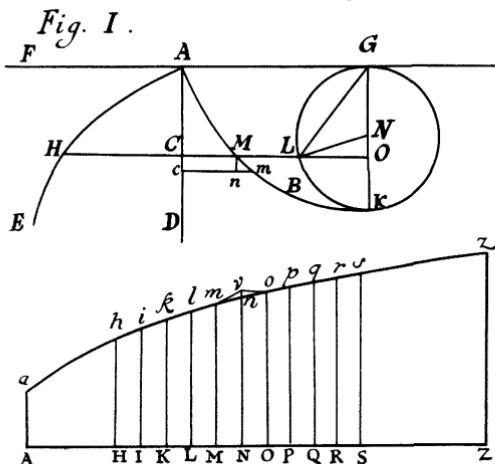


Fig. 4

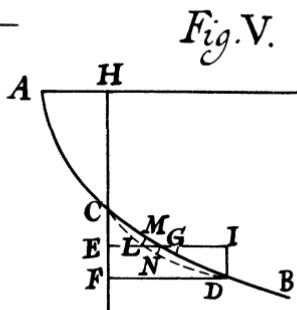


Fig. V.

Рис. 2.4. Решения вариационной задачи (Иоганн Бернулли, Якоб Бернулли, Эйлер).

представил решение, значительно уступающее в изяществе, но при этом, к сожалению, правильное, и теперь уже сам предложил своему брату Иоганну решить пару новых задач. Метод Якоба (см. рис. 2.4) напоминает обращение современного метода «конечных элементов». Он привел к появлению знаменитой работы Эйлера (1744), где общее решение задачи

$$\int_{x_0}^{x_1} f(x, y, y') dx = \min$$

дано с помощью дифференциального уравнения второго порядка

$$f_{y''} y'' + f_{y'} y' + f_{y \cdot x} - f_y = 0 \quad (2.5)$$

и подробно рассмотрены 100 вариационных задач. В частном случае, когда  $f$  не зависит от  $x$ , проинтегрировав уравнение (2.5),

мы получим

$$f - f_y y' = K. \quad (2.6)$$

В своем доказательстве для вывода уравнения (2.5) Эйлер использовал ломаные. И только идеи, высказанные в 1755 г. 19-летним Лагранжем, позволили получить доказательство, которое в настоящее время является общепринятым [Функ (1962), Голдстайн (1980)]. Тогда Эйлер (1756) в честь Лагранжа назвал всю эту область исследований «вариационным исчислением» ("...tamen gloria prima inventionis acutissimo Geometrae Taurinensi La Grange erat reservata").

### Клеро

Клеро открыл новый класс уравнений с интересными свойствами [см. работу Клеро (1734), задача III]. Изучая движение прямого угла (см. рис. 2.5), Клеро получил дифференциальные уравнения вида

$$y - xy' + f(y') = 0. \quad (2.7)$$

Это были первые *неявные* дифференциальные уравнения

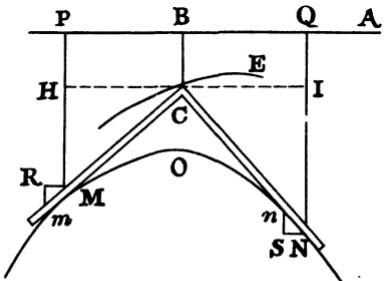


Рис. 2.5. Иллюстрация из работы Клеро (1734).

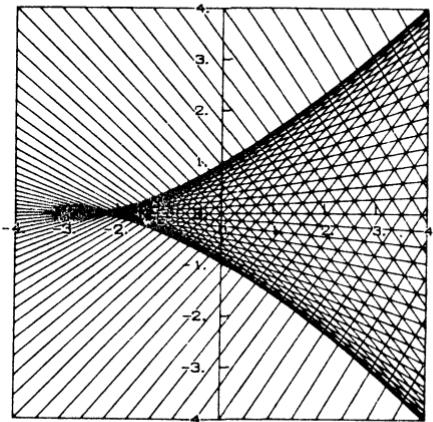


Рис. 2.6. Решения дифференциального уравнения Клеро.

ния. Их особенность состоит в том, что решениями служат не только прямые семейства  $y = Cx - f(C)$ , но также и огибающие это семейство кривые (см. упр. 10 из разд. I.3). На рис. 2.6 показан пример с функцией  $f(C) = 5(C^3 - C)/2$ .

Поскольку уравнение имеет третью степень по  $y'$ , заданное начальное значение допускает три разных решения. Кроме того, в точке, где прямая касается огибающей, решение можно продолжить либо по прямой, либо по огибающей. Это дает огромное разнообразие кривых решения. Ныне мы объясняем это любопытное явление тем, что в этих точках не выполняется условие Липшица [см. также Айнс (1944)].

### I.3. Уравнения, разрешимые в квадратурах и элементарных функциях

Рассмотрим теперь некоторые простейшие типы уравнений, которые можно решить с помощью вычисления интегралов.

#### Уравнение с разделяющимися переменными

$$y' = f(x) g(y). \quad (3.1)$$

Поделив его на  $g(y)$  и проинтегрировав, получим решение

$$\int \frac{dy}{g(y)} = \int f(x) dx + C.$$

(Лейбниц, 1691, из письма Гюйгенсу.)

Частным случаем является линейное уравнение  $y' = f(x)y$  с решением

$$y(x) = CR(x), \quad R(x) = \exp\left(\int f(x) dx\right).$$

#### Неоднородное линейное уравнение

$$y' = f(x)y + g(x). \quad (3.2)$$

Здесь подстановка  $y(x) = c(x)R(x)$  приводит к уравнению  $c'(x) = g(x)/R(x)$  [Иог. Бернулли (1697)]. Таким образом, получим решение

$$y(x) = R(x) \left( \int_{x_0}^x \frac{g(s)}{R(s)} ds + C \right). \quad (3.3)$$

#### Уравнения в полных дифференциалах

Уравнение вида

$$P(x, y) + Q(x, y)y' = 0 \quad (3.4)$$

можно, оказывается, сразу решить, если оно удовлетворяет условию

$$\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x}. \quad (3.5)$$

Для этого с помощью интегрирования находят такую функцию  $F(x, y)$ , что

$$\frac{\partial F}{\partial x} = P, \quad \frac{\partial F}{\partial y} = Q.$$

Тогда (3.4) преобразуется в уравнение  $\frac{d}{dx} F(x, y(x)) = 0$ , решения которого выражаются формулой  $F(x, y(x)) = C$ . В случае когда (3.5) не выполняется, Клеро и Эйлер исследовали способ, который с помощью умножения (3.4) на подходящий множитель  $M(x, y)$  (интегрирующий множитель) позволяет иногда получить уравнение  $MP + MQy' = 0$ , удовлетворяющее условию (3.5).

## Уравнения второго порядка

Уравнения второго порядка допускают решение с помощью интегрирования намного реже, чем уравнения первого порядка. Оно возможно в случае линейных уравнений с постоянными коэффициентами, что было известно еще Ньютону (для уравнений второго порядка). Кроме того существует несколько приемов понижения порядка, например следующий.

*В линейное уравнение*

$$y'' = a(x)y' + b(x)y$$

подставим

$$y = \exp\left(\int p(x) dx\right). \quad (3.6)$$

Производные этой функции содержат только производные  $p$  более низких порядков:

$$y' = p \cdot \exp\left(\int p(x) dx\right), \quad y'' = (p^2 + p') \exp\left(\int p(x) dx\right),$$

а поэтому, подставив их в дифференциальное уравнение, после деления на  $y$  получим уравнение *более низкого порядка*

$$p^2 + p' = a(x)p + b(x), \quad (3.7)$$

которое, однако, является нелинейным [Риккати (1723), Эйлер (1728)].

Если уравнение не зависит от  $y$ , т. е. имеет вид  $y'' = f(x, y')$ , то естественно взять  $y' = v$  и перейти к уравнению  $v' = f(x, v)$ .

Важный случай представляют *уравнения, не зависящие от  $x$* :

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

Здесь мы будем рассматривать  $y'$  как функцию от  $y$ :  $y' = p(y)$ . Тогда по цепному правилу (дифференцирования сложной функции) придет к уравнению первого порядка  $y'' = p'p = f(y, p)$ . После того как функция  $p(y)$  найдена, остается проинтегрировать уравнение  $y' = p(y)$ , относящееся к типу (3.1) (Риккати (1712): «Per liberare la premessa formula dalle seconde differenze, ..., chiamo  $p$  la sunnormale  $BF$  ...», см. также работу Эйлера (1769), задача 96).

Эйлер положил начало изучению всего класса дифференциальных уравнений, допускающих интегрирование аналитическими методами. Эти результаты занимают более 800 страниц в томах XXII и XXIII его труда *Opera Omnia*. Более современное изложение можно найти в работе Айнса (1944). Незаменимым источником по этой теме является книга Камке (1942). Помимо описания методов решения и общих свойств решений, в ней содержится систематически упорядоченный перечень более 1500 дифференциальных уравнений с решениями и ссылками на литературу.

Часто даже для простых на вид уравнений вычисления оказываются очень сложными, и весьма скоро стало ясно, что не всегда можно получить элементарное решение. Лиувилль (1841) первым доказал, что некоторые уравнения, как, например,  $y' = x^2 + y^2$ , нельзя решить с помощью элементарных функций. Поэтому в XIX веке математиков стали все более интересовать общие теоремы существования и численные методы нахождения решений.

## Упражнения

- Найдите с помощью предложенного Ньютоном метода другие решения уравнения (2.1), например при  $y(0) = 0.1$ ,  $y(0) = 0.2$ ,  $y(0) = 0.3$ , и начертите их графики.
- Решите уравнение Ньютона (2.1) в квадратурах.
- Решите уравнение Лейбница (2.3) в терминах элементарных функций.

*Указание.* Вычисление интеграла по  $y$  может вызвать затруднения. Воспользуйтесь заменой  $a^2 - y^2 = u^2$ ,  $-ydy = udu$ .

- Решите уравнение  $y' = f(y)$ , где

$$f(y) = \sqrt{|y|},$$

и начертите график решения.

- Решите задачу о хозяине и собаке: собака бежит со скоростью  $w$  к хозяину, идущему со скоростью  $v$  вдоль оси  $y$ . Задача сводится к дифференциальному уравнению

$$xy'' = \frac{v}{w} (1 + y'^2)^{1/2}.$$

- Решите уравнение  $my'' = -k/y^2$ , описывающее падение тела в соответствии с законом всемирного тяготения Ньютона.
- Проверьте, что циклоида

$$x - x_0 = R(\tau - \sin \tau), \quad y - h = R(1 - \cos \tau), \quad R = \frac{1}{4gK^2}$$

удовлетворяет дифференциальному уравнению (2.4) из задачи о брахистохроне. Решая задачу (2.4) прямым способом, после некоторых упрощений мы придем к интегралу

$$\int \left( \frac{y}{1-y} \right)^{1/2} dy,$$

который вычисляется с помощью замены  $y = (\sin t)^2$ .

8. Покажите, что дифференциальное уравнение

$$x + yy' = y' \sqrt{x^2 + y^2 - 1}$$

имеет решение  $2ay = a^2 + 1 - x^2$  при любом  $a$ . Начертите эти кривые и найдите другое решение (данний пример привел Лагранж (1774) для объяснения «явления Клеро»).

9. Приведите «уравнение Бернулли» [Як. Бернулли (1695)]  $y' + f(x)y = g(x)y^n$  с помощью преобразования координат  $z(x) = (y(x))^q$  с соответствующим выбором  $q$  к линейному уравнению [Лейбниц, Acta Erud. (1696), Иог. Бернулли (1697)].

10. Проверьте, что огибающую решений  $y = Cx - f(C)$  уравнения Клеро (2.7) можно задать параметрически:

$$\begin{aligned} x(p) &= f'(p), \\ y(p) &= pf'(p) - f(p). \end{aligned}$$

Покажите, что эта огибающая также является решением (2.7), и найдите ее при  $f(C) = 5(C^3 - C)/2$  (см. рис. 2.6).

11. [Ньютон (1671), «Problema II, Solutio particularis】. Решите уравнение в полных дифференциалах:

$$3x^2 - 2ax + ay - 3y^2y' + axy' = 0.$$

Ньютон указал решение  $x^3 - ax^2 + axy - y^3 = 0$ . Отметим, что он пропустил произвольную константу интегрирования  $C$ .

12. [Ньютон (1671)]. Решите уравнения

a)  $y' = 1 + \frac{y}{a} + \frac{xy}{a^2} + \frac{x^2y}{a^3} + \frac{x^3y}{a^4} \dots$ ,

б)  $y' = -3x + 3xy + y^2 - xy^2 + y^3 - xy^3 + y^4 - xy^4 + 6x^2y - 6x^2 + 8x^3y - 8x^3 + 10x^4y - 10x^4 + \dots$ .

Ньютон получил следующие результаты:

а)  $y = x + \frac{x^2}{2a} + \frac{x^3}{2a^2} + \frac{x^4}{2a^3} + \frac{x^5}{2a^4} + \frac{x^6}{2a^5} \dots$ ,

б)  $y = -\frac{3}{2}x^2 - 2x^3 - \frac{25}{8}x^4 - \frac{91}{20}x^5 - \frac{111}{16}x^6 - \frac{367}{35}x^7 \dots$ .

13. (Ньютон). Решите уравнение (2.1) с другим начальным значением  $y(0) = 1$ .

Решение Ньютона:  $y = 1 + 2x + x^3 + \frac{1}{4}x^4 + \frac{1}{4}x^5 + \dots$ .

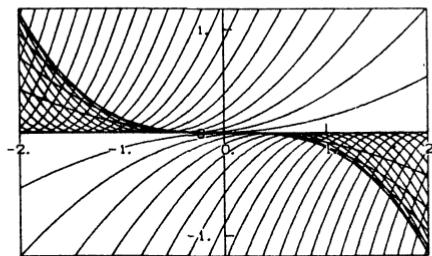


Рис. 3.1. Семейство решений примера Коши, приведенного в упр. 14.

14. (Коши (1824)). Покажите, что семейство  $y = C(x + C)^2$  удовлетворяет дифференциальному уравнению  $(y')^3 = 8y^2 - 4xyy'$ . Найдите еще одно решение, которое не принадлежит указанному семейству (см. рис. 3.1).

*Ответ:*  $y = -\frac{4}{27}x^3$ .

## I.4. Линейные дифференциальные уравнения

«Lisez Euler, lisez Euler, c'est notre maître à tous».

«Читайте Эйлера, читайте Эйлера, он учитель для всех нас.»

(Лаплас)

Следуя за Эйлером (1743), мы хотим изучить общее решение линейного дифференциального уравнения порядка  $n$ . Уравнение

$$L(y) := a_0(x)y^{(n)} + a_1(x)y^{(n-1)} + \dots + a_n(x)y = 0 \quad (4.1)$$

с заданными функциями  $a_0(x), \dots, a_n(x)$  называется однородным. Если известно  $n$  решений  $u_1(x), \dots, u_n(x)$  уравнения (4.1), то любая линейная комбинация

$$y(x) = C_1u_1(x) + C_nu_n(x) \quad (4.2)$$

с постоянными коэффициентами  $C_1, \dots, C_n$  также будет решением, поскольку все производные  $y$  входят в (4.1) только линейно.

### Уравнения с постоянными коэффициентами

Сначала рассмотрим частный случай

$$y^{(n)}(x) = 0. \quad (4.3)$$

После однократного интегрирования получим  $y^{(n-1)}(x) = C_1$ , после двухкратного  $y^{(n-2)}(x) = C_1x + C_2$  и т. д. Заменив произвольные константы  $C_i$  на новые, в итоге придем к формуле

$$y(x) = C_1x^{n-1} + C_2x^{n-2} + \dots + C_n.$$

Таким образом, в «общее решение» уравнения (4.3) входит  $n$  «свободных параметров». Рассмотрев еще ряд примеров, Эйлер интуитивно предположил, что этот же результат верен и для уравнения общего вида (4.1). Однако данный вопрос был полностью выяснен лишь много лет спустя.

Рассмотрим теперь уравнение общего вида с постоянными коэффициентами:

$$y^{(n)} + A_1y^{(n-1)} + \dots + A_ny = 0. \quad (4.4)$$

Нам нужно найти базис, состоящий из  $n$  линейно независимых решений  $u_1(x), \dots, u_n(x)$ . Для этой цели Эйлер взял за основу преобразование (3.6), (3.7): если  $a(x)$  и  $b(x)$  — постоянные вели-

чины, предположим, что в (3.7)  $p$  также является константой и, следовательно,  $p'$  равно нулю, откуда получим квадратное уравнение  $p^2 = ap + b$ . Тогда для любого корня этого уравнения выражение (3.6) принимает вид  $y = e^{px}$ . Таким образом, в общем случае мы предположим, что  $y = e^{px}$ , где  $p$  — неизвестная константа. Поэтому от (4.4) мы придем к *характеристическому уравнению*

$$p^n + A_1 p^{n-1} + \dots + A_n = 0. \quad (4.5)$$

Если его корни  $p_1, \dots, p_n$  различны, то все решения уравнения (4.4) даются формулой

$$y(x) = C_1 e^{p_1 x} + \dots + C_n e^{p_n x}. \quad (4.6)$$

Интересно отметить, что «самые выдающиеся математики мира» многие десятилетия бились над нахождением этого решения, которое современным студентам кажется столь тривиальным.

Сложности при поиске решения (4.6) возникают в том случае, когда (4.5) не имеет  $n$  различных корней. Рассмотрим пример, предложенный Эйлером:

$$y'' - 2qy' + q^2y = 0. \quad (4.7)$$

В нем  $p = q$  — это двукратный корень соответствующего характеристического уравнения. Если мы предположим, что

$$y = e^{qx}u, \quad (4.8)$$

то (4.7) примет вид  $u'' = 0$ , и мы вернемся к уравнению (4.3). Следовательно, общее решение (4.7) имеет вид  $y(x) = e^{qx}(C_1x + C_2)$  (см. также упр. 5 ниже). После нескольких дополнительных примеров такого типа становится ясно, что преобразование (4.8) производит *сдвиг* характеристического многочлена, так что в том случае, когда  $q$  — корень кратности  $k$ , для  $u$  получается уравнение, оканчивающееся членами  $\dots + Bu^{(k+1)} + Cu^{(k)} = 0$ . Поэтому выражение

$$e^{qx}(C_1x^{k-1} + \dots + C_k)$$

дает нам  $k$  независимых решений.

И наконец, для пары комплексных корней  $p = \alpha \pm i\beta$  решения  $e^{(\alpha+i\beta)x}, e^{(\alpha-i\beta)x}$  можно заменить вещественными функциями

$$e^{\alpha x}(C_1 \cos \beta x + C_2 \sin \beta x).$$

Изучение *неоднородного* уравнения

$$L(y) = f(x) \quad (4.9)$$

было начато в работе Эйлера (1750). Приведем из нее случай, когда  $f(x)$  является многочленом, как, например, в уравнении

$$Ay'' + By' + Cy = ax^2 + bx + c. \quad (4.10)$$

Здесь Эйлер представляет решение в виде  $y(x) = Ex^2 + Fx + G + v(x)$ . Подставляя это выражение в (4.10), из условия обращения в нуль всех встречающихся степеней  $x$  получим

$$CE = a, \quad CF + 2BE = b, \quad CG + BF + 2AE = c,$$

$$Av'' + Bv' + Cv = 0.$$

Отсюда при условии, что  $C$  отлично от нуля, мы можем найти  $E$ ,  $F$  и  $G$ . Мы видим, что общее решение линейного однородного уравнения равно сумме его частного решения и общего решения линейного однородного уравнения. Это утверждение верно и в общем случае, что можно проверить с помощью элементарных приемов линейной алгебры.

Описанный метод поиска частного решения с помощью неизвестных коэффициентов работает подобным образом и в том случае, когда  $f(x)$  состоит из экспонент, синусов и косинусов. Его часто называют «быстрым методом». Приятно отметить, что именно этот метод исторически был открыт первым.

### Вариация постоянных

Исследование общего случая неоднородного уравнения

$$a_0(x)y^{(n)} + \dots + a_n(x)y = f(x) \quad (4.11)$$

является заслугой Лагранжа (1775) («...с помощью нового метода, простого, как можно только мечтать.» — См. также работу Лагранжа (1788), вторая часть, разд. V). Пусть нам известны  $n$  линейно независимых решений  $u_1(x), \dots, u_n(x)$  однородного уравнения. Тогда, распространяя на этот случай метод, примененный к уравнению (3.2), представим решение в виде [вместо (4.2)]

$$y(x) = c_1(x)u_1(x) + \dots + c_n(x)u_n(x), \quad (4.12)$$

где  $c_i(x)$  — неизвестные функции («метод вариации постоянных»). Нам надо подставить выражение (4.12) в (4.11). Для этого вычислим первую производную

$$y' = \sum_i c'_i u_i + \sum_i c_i u'_i.$$

Если мы будем слепо продолжать дифференцирование, то скоро получим громоздкие и бесполезные формулы. Здесь Лагранж проявил проницательность, потребовав обращения в нуль первой суммы, т. е. выполнения равенств

$$\sum_i c'_i u_i^{(j)} = 0 \text{ при } j = 0, \text{ а затем при } j = 1, \dots, n - 2. \quad (4.13)$$

Далее, повторяя дифференцирование  $y$  и исключая каждый раз

нежелательные члены (4.13), мы придем к формулам

$$\begin{aligned} y' &= \sum_i c_i u'_i, \quad \dots, \quad y^{(n-1)} = \sum_i c_i u_i^{(n-1)}, \\ y^{(n)} &= \sum_i c'_i u_i^{(n-1)} + \sum_i c_i u_i^{(n)}. \end{aligned}$$

Если подставить их в (4.11), то произойдет замечательное взаимное сокращение членов благодаря тому, что  $u_i(x)$  удовлетворяют однородному уравнению. И с учетом (4.13) мы окончательно получим линейную систему

$$\left( \begin{array}{c} u_1 \dots u_n \\ u'_1 \dots u'_n \\ \vdots \\ u_1^{(n-1)} \dots u_n^{(n-1)} \end{array} \right) \left( \begin{array}{c} c'_1 \\ c'_2 \\ \vdots \\ c'_n \end{array} \right) = \left( \begin{array}{c} 0 \\ 0 \\ \vdots \\ 0 \\ f(x)/a_0(x) \end{array} \right). \quad (4.14)$$

Ее определитель называется «определителем Вронского», а решением являются величины  $c'_1(x), \dots, c'_n(x)$ , проинтегрировав которые, мы найдем  $c_1(x), \dots, c_n(x)$ . Разд. I.11 позволит намного глубже понять эту формулу.

### Упражнения

- Найдите решение «huius, aequationis differentialis quarti gradus» (уравнений четвертого порядка)  $a^4y^{(4)} + y = 0$ ,  $a^4y^{(4)} - y = 0$ ;  
решите уравнение «septimi gradus»  $y^{(7)} + y^{(5)} + y^{(4)} + y^{(3)} + y^{(2)} + y = 0$  [Эйлер (1743), примеры 4, 5, 6].

- Решите с помощью предложенного Эйлером метода уравнения  $y'' - 3y' - 4y = \cos x$ ,  $y'' + y = \cos x$ .

*Указание.* В первом случае частное решение можно искать в виде  $E \cos x + F \sin x$ . Однако во втором случае это представление непригодно из-за резонанса. Тогда надо использовать выражение  $Ex \cos x + Fx \sin x$ , как при решении уравнения (4.7).

- Найдите решение уравнения

$$y'' - 3y' - 4y = g(x), \quad g(x) = \begin{cases} \cos x & 0 \leq x \leq \frac{\pi}{2} \\ 0 & \frac{\pi}{2} \leq x \end{cases}$$

при условии  $y(0) = y'(0) = 0$ :

- а) используя решение из упр. 2;  
 б) методом Лагранжа (вариации постоянных).
4. (Понижение порядка при одном известном решении). Предположим, что нам известно ненулевое решение  $u_1(x)$  уравнения
- $$y'' + a_1(x)y' + a_2(x)y = 0.$$
- Покажите, что второе линейно независимое решение можно найти с помощью представления  $u_2(x) = c(x)u_1(x)$ .
5. Исследуйте случай кратных характеристических корней уравнения (4.7), рассматривая их как предельный случай  $p_2 \rightarrow p_1$  и используя решения

$$u_1(x) = e^{p_1 x}, \quad u_2(x) = \lim_{p_2 \rightarrow p_1} \frac{e^{p_2 x} - e^{p_1 x}}{p_2 - p_1} = \frac{\partial e^{p_1 x}}{\partial p_1} \text{ и т. д.}$$

(Даламбер (1748): «И наконец, если значения  $p$  и  $p'$  равны, мы вместо этого будем считать, что  $p = a + \alpha$ ,  $p' = a - \alpha$ , где  $\alpha$  — бесконечно малая величина...».)

## I.5. Уравнения со слабыми особенностями

Многие встречающиеся в приложениях уравнения имеют особенности, т. е. точки, в которых функция  $f(x, y)$  в дифференциальном уравнении обращается в бесконечность. Мы изучим

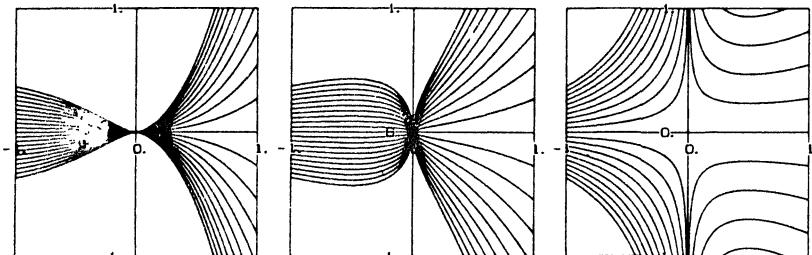


Рис. 5.1. Решения уравнения (5.1) при  $b = 1$ ,  $q = 3/2$ ,  $q = 1/2$ ,  $q = -1/2$ .

некоторые вопросы классической теории таких уравнений, поскольку численные методы, о которых дальше пойдет речь, часто отказывают в особых точках и плохо работают вблизи от них.

В качестве первого примера рассмотрим уравнение

$$y' = \frac{q + bx}{x} y, \quad q \neq 0, \quad (5.1)$$

с особой точкой при  $x = 0$ . Получим это решение с помощью метода разделения переменных (3.1):

$$y(x) = Cx^q e^{bx} = C(x^q + bx^{q+1} + \dots). \quad (5.2)$$

Эти решения изображены на рис. 51 для разных значений  $q$ . В их поведении наблюдаются фундаментальные различия, зависящие от  $q$ .

### Линейные уравнения

Эйлер положил начало систематическому изучению уравнений с особенностями. Он поставил вопрос: как решать уравнений второго порядка с помощью рядов, как в случае (5.2)? [Эйлер (1769), Problema 122, p. 177 «...quas commode per series resolvere licet...»] Он нашел уравнение

$$x^2(a + bx)y'' + x(c + ex)y' + (f + gx)y = 0. \quad (5.3)$$

Предположим, что  $y$  имеет вид

$$y = \sum_{i=0}^{\infty} A_i x^{q+i}, \quad A_0 \neq 0, \quad (5.4)$$

и подставим эту формулу в (5.3), что напоминает процедуру Ньютона (2.2). Мы обнаружим, что степенные множители  $x^2$  и  $x$  при  $y''$  и  $y'$  соответственно в точности восстанавливают то, что утрачено в результате дифференцирования, и получим соотношения

$$q(q-1)a + qc + f = 0, \quad (5.5a)$$

$$\begin{aligned} & [(q+i)(q+i-1)a + (q+i)c + f]A_i = \\ & = [(q+i-1)(q+i-2)b + (q+i-1)e + g]A_{i-1}, \end{aligned} \quad (5.5b)$$

где  $i = 1, 2, 3, \dots$ . Уравнение (5.5a) называется *индексным* и при  $a$ , отличном от нуля, определяет два характеристических корня  $q_1$  и  $q_2$  (расположенных в порядке убывания). Для каждого из них при произвольном  $A_0$  соотношение (5.5b) позволяет рекурсивно вычислять  $A_1, A_2$  и т. д. Радиус сходимости полученного ряда определяется далее в упр. 8. Таким образом, мы находим два линейно независимых решения  $u_1(x)$  и  $u_2(x)$  дифференциального уравнения, что нам и требовалось (см. I.4). Однако существует ситуация, когда определение второго решения вызывает затруднения, а именно: когда коэффициент  $A_i$  в (5.5b) при некотором  $i$  обращается в нуль, т. е. если  $q_1 - q_2$  есть целое число («Problema 123»). В этом случае в результате ряда акробатических приемов [Эйлер (1769): «...quod  $x^0/0$  aequivaleat ipsi  $\log x$ ...»] Эйлер пришел к мысли, что второе решение надо искать с помощью

представления

$$u_2(x) = u_1(x) \log x + \sum_{i=0}^{\infty} A_i^{(2)} x^{q_i+i}. \quad (5.6)$$

Подставив эту формулу в (5.3), он сумел вычислить коэффициенты  $A_i^{(2)}$ . Пример такой ситуации приведен в упр. 1.

Позднее Эйлер (1778) заметил, что рекуррентная формула (5.5b) становится особенно изящной, если вместо (5.3) брать в качестве исходного дифференциальное уравнение

$$x(1-x)y'' + (c - (a + b + 1)x)y' - aby = 0. \quad (5.7)$$

Тогда описанный метод дает соотношения

$$q(q-1) + cq = 0, q_1 = 0, q_2 = 1 - c, \quad (5.8)$$

$$A_{i+1} = \frac{(a+i)(b+i)}{(c+i)(1+i)} A_i \quad \text{при } q_1 = 0. \quad (5.9)$$

Полученные решения, названные позднее гипергеометрическими функциями, стали особенно знаменитыми в XIX веке в связи с работой Гаусса (1812).

Указанный метод применим и в более общем случае дифференциального уравнения

$$x^2y'' + xa(x)y' + b(x)y = 0, \quad (5.10)$$

где  $a(x)$  и  $b(x)$  — регулярные аналитические функции. Тогда говорят, что 0 является *регулярной* (или *слабой*) особой точкой. Аналогично об уравнении  $(x-x_0)^2y'' + (x-x_0)a(x)y' + b(x)y = 0$  говорят, что оно обладает регулярной особой точкой  $x_0$ . В этом случае решения можно получить, используя алгебраические особенности  $(x-x_0)^q$ .

В завершение мы хотим еще изучить поведение решения на бесконечности для уравнения вида

$$a(x)y'' + b(x)y' + c(x)y = 0. \quad (5.10a)$$

Для этого проведем замену переменных  $t = 1/x$ ,  $z(t) = y(x)$ , при которой уравнение примет вид

$$t^4a\left(\frac{1}{t}\right)z'' + \left(2t^3a\left(\frac{1}{t}\right) - t^2b\left(\frac{1}{t}\right)\right)z' + c\left(\frac{1}{t}\right)z = 0; \quad (5.10b)$$

∞ называется *регулярной особой точкой* уравнения (5.10a), если 0 является *регулярной особой точкой* для (5.10b). Соответствующие примеры приведены в упр. 9.

## Нелинейные уравнения

«Der Mathematiker weiss sich ohnedies beim Auftreten von singulären Stellen gegebenenfalls leicht zu helfen». «Если встречаются особенности, математик все равно легко умеет справляться с ними в каждом случае». (К. Хайн, 1900)

Описанный выше метод позволяет иногда получать и для решений нелинейных уравнений если не весь ряд, то по крайней мере пару его членов.

**Пример.** Посмотрим, что произойдет, если мы попытаемся решить классическую задачу о брахистохроне (2.4) с помощью представления решения в виде ряда. Пусть  $h = 0$  и начальное значение  $y(0) = 0$ . Запишем уравнение следующим образом,

$$y'^2 = \frac{L}{y} - 1 \quad \text{или} \quad yy'^2 + y = L. \quad (5.11)$$

В начальной точке  $y(0) = 0$ , а  $y'$  обращается в бесконечность. В этом случае большинство численных методов потерпело бы неудачу. Будем искать решение в виде  $y = A_0x^q$ . Подставив эту формулу в (5.11), придем к уравнению  $q^2A_0^3x^{3q-2} + A_0x^q = L$ . Из начального условия следует, что  $y(x)$  становится пренебрежимо малым при малых значениях  $x$ . Поэтому положим первый член равным  $L$  и получим равенства  $3q - 2 = 0$  и  $q^2A_0^3 = L$ . Таким образом, формула

$$u(x) = \left( \frac{9Lx^2}{4} \right)^{1/3} \quad (5.12)$$

задает первое приближение к решению. Теперь идея состоит в использовании (5.12) для того, чтобы отойти от начальной точки на малую величину  $x$ , а затем продолжить решение с помощью одной из численных пошаговых процедур, описанных в следующих главах.

Попытаемся более точно аппроксимировать решение с помощью представления  $y = A_0x^q + A_1x^{q+r}$ . Подставив его в (5.11), получим уравнение

$$q^2A_0^3x^{3q-2} + q(3q + 2r)A_0^2A_1x^{3q+r-2} + A_0x^q + \dots = L.$$

Используем второй член для нейтрализации третьего, что даст нам равенства  $3q + r - 2 = q$ , или  $r = q = 2/3$ , и  $5q^2A_0A_1 = -1$ . Таким образом, функция

$$v(x) = \left( \frac{9Lx^2}{4} \right)^{1/3} - \left( \frac{9^2x^4}{4^2L5^3} \right)^{1/3} \quad (5.13)$$

лучше аппроксимирует решение. Сравнение следующих численных результатов с точным решением  $y(x)$  из I.3, упр. 7 при  $L = 2$ ,

показывает, что аппроксимации (5.12) и (5.13) являются полезными:

$$\begin{aligned}x &= 0.10 \quad y(x) = 0.342839 \quad u(x) = 0.355689 \quad v(x) = 0.343038 \\x &= 0.01 \quad y(x) = 0.076042 \quad u(x) = 0.076631 \quad v(x) = 0.076044.\end{aligned}$$

### Упражнения

- Найдите общее решение уравнения  $x^2y'' + xy' + gx^n y = 0$  с постоянной  $g$  [Эйлер (1769), Problema 123, Exemplum 1].
- Примените разработанный Эйлером метод к уравнению Бесселя  $x^2y'' + xy' + (x^2 - g^2)y = 0$ . Начертите решения, полученные при  $g = 2/3$  и  $g = 10/3$ .
- Найдите решения уравнений  $x^2y'' - 2xy' + y = 0$ ,  $x^2y'' - 3xy' + 4y = 0$ . Уравнения такого типа часто называют уравнениями Эйлера или Коши. Однако еще Иог. Бернулли было известно их решение.
- [Эйлер (1769), Problema 123, Exemplum 2]. Периметр эллипса с осями 1 и  $x < 1$  равен

$$y(x) = \int_0^{2\pi} (\sin^2 s + x^2 \cos^2 s)^{1/2} ds.$$

- a) Проверьте, что  $y(x)$  удовлетворяет дифференциальному уравнению

$$x(1-x^2)y'' - (1+x^2)y' + xy = 0. \quad (5.14)$$

- b) Найдите решения этого уравнения.  
в) Покажите, что замена переменных  $x^2 = t$ ,  $y(x) = z(t)$  преобразует (5.14) в гипергеометрическое уравнение (5.7).  
*Указание.* В а) расчеты приводят к интегралу

$$\int_0^{2\pi} \frac{1 - 2 \cos^2 s + q^2 \cos^4 s}{(1 - q^2 \cos^2 s)^{3/2}} ds, \quad q^2 = 1 - x^2.$$

Надо показать, что он равен нулю. Разложите его в ряд по степеням  $q^2$ .

- Попытайтесь решить уравнение

$$x^2y'' + (3x - 1)y' + y = 0$$

с помощью ряда (5.4) и изучите его сходимость.

- Найдите ряд вида

$$y = A_0 x^q + A_1 x^{q+s} + A_2 x^{q+2s} + \dots,$$

который в окрестности точки  $x = 0$  служит решением нелинейного «уравнения Эмдена — Фаулера» из астрофизики:  $(x^2y')' + y^2x^{-\frac{4}{3}} = 0$ .

7. Аппроксимируйте решение уравнения Лейбница (2.3) в окрестности сингулярного начального значения  $y(0) = a$  с помощью функции вида  $y(x) = a - Cx^a$ . Сравните результат с точным решением, полученным в упр. 3 из разд. I.3.
8. Покажите, что радиус сходимости ряда (5.4) с коэффициентами, заданными в (5.5) и (5.9), равен соответственно
  - i)  $r = \left| \frac{a}{b} \right|$ ; ii)  $r = 1$ .
9. Покажите, что для гипергеометрического уравнения (5.7)  $\infty$  является регулярной особой точкой, а для уравнения Бесселя из упр. 2 не является.
10. Рассмотрим задачу Коши

$$y' = \frac{\lambda}{x} y + g(x), \quad y(0) = 0. \quad (5.15)$$

- a) Докажите, что для  $\lambda \leq 0$  задача (5.15) имеет единственное решение при  $x \geq 0$ .
- б) Если  $g(x)$  дифференцируема  $k$  раз и  $\lambda \leq 0$ , то решение  $y(x)$  при  $x \geq 0$  дифференцируемо  $k + 1$  раз и выполняется равенство

$$y^{(j)}(0) = \left(1 - \frac{\lambda}{j}\right)^{-1} g^{(j-1)}(0), \quad j = 1, 2, \dots .$$

## I.6. Системы уравнений

### Лагранж

Первые системы уравнений появились в работе Лагранжа (1759). Одним из вопросов, интересовавших его в самом начале, было распространение звука. Считая подход Ньютона недостаточным, Лагранж попытался объяснить это явление с помощью известных результатов теории упругости. Если мы будем рассматривать среду как последовательность материальных точек и обозначим через  $y_1(t)$ ,  $y_2(t)$ ,  $y_3(t)$ , ... их отклонения от положения

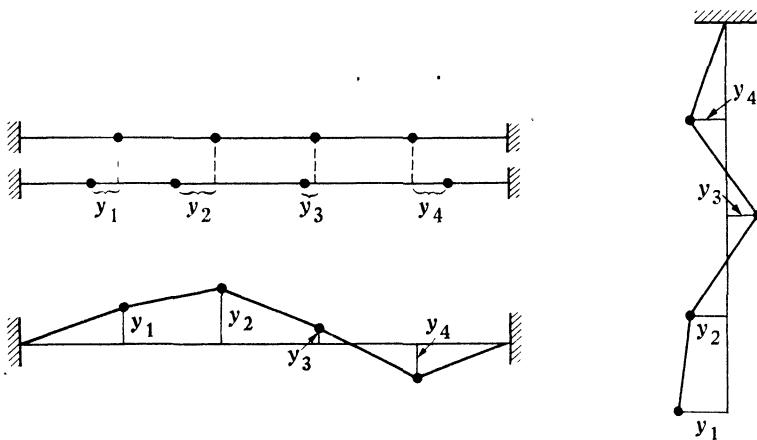


Рис. 6.1. Модели распространения звука, колеблющейся и висящей струны равновесия (рис. 6.1), то по закону Гука из теории упругости получим

$$\begin{aligned} y_1'' &= K^2 (-2y_1 + y_2), \\ y_2'' &= K^2 (y_1 - 2y_2 + y_3), \\ y_n'' &= K^2 (y_{n-1} - 2y_n). \end{aligned} \quad (6.1)$$

Это система  $n$  линейных дифференциальных уравнений. В данном разделе мы будем обозначать через  $t$  независимую переменную (время), чтобы не перепутать ее с пространственной переменной  $x$  в соответствующем уравнении в частных производных. Лагранж заметил, что такие же уравнения получаются из уравнения Даламбера (1748)

$$\frac{\partial^2 u}{\partial t^2} = a^2 \frac{\partial^2 u}{\partial x^2}$$

для колеблющейся струны (*cordis vibrantibus*), если ее моделировать с помощью конечного числа материальных точек, расположенных на струне, не имеющей массы (рис. 6.1).

В качестве другого примера Лагранж [(1762), № 36] рассмотрел материальные точки, присоединенные к *висящей* струне (рис. 6.1). В этом случае натяжение увеличивается в верхней части струны, и мы имеем следующие уравнения движения:

$$\begin{aligned} y_1'' &= K^2 (-y_1 + y_2), \\ y_2'' &= K^2 (y_1 - 3y_2 + 2y_3), \\ y_3'' &= K^2 (2y_2 - 5y_3 + 3y_4), \\ &\dots \\ y_n'' &= K^2 ((n-1)y_{n-1} - (2n-1)y_n). \end{aligned} \quad (6.2)$$

Конечно, во всех этих примерах предполагается, что отклонения  $y_i$  «бесконечно» малы, и поэтому линейные модели реалистичны.

Используя обозначения, которые получили распространение лишь веком позже, запишем эти уравнения в виде

$$y_i'' = \sum_{j=1}^n a_{ij}y_j \quad (i = 1, \dots, n). \quad (6.3)$$

Данные формулы представляют *систему линейных уравнений 2-го порядка с постоянными коэффициентами*. Лагранж решил систему (6.3) с помощью подстановки  $y_i = c_i e^{pt}$ , которая приводит к уравнениям

$$p^2 c_i = \sum_{j=1}^n a_{ij}c_j \quad (i = 1, \dots, n). \quad (6.4)$$

Таким образом,  $p^2$  должно быть *собственным значением* матрицы  $A$ , а  $c = (c_1, \dots, c_n)^T$  — соответствующим *собственным вектором*. Здесь, как мы видим, впервые появляется задача на собственные значения; впервые используется также «метод прямых» для волнового уравнения.

Лагранж [(1762), № 30] объясняет, что для решения (6.4) из  $n - 1$  уравнений вычисляются величины  $c_2/c_1, \dots, c_n/c_1$  как функции от  $p$ , а затем полученный результат подставляется в последнее уравнение. Это приводит нас к многочлену степени  $n$  для величины  $p^2$  (*характеристическому многочлену*), для которого надо найти  $n$  различных корней. Таким образом, мы получаем  $2n$  решений  $y_i^{(j)} = c_i^{(j)} \exp(\pm p_j t)$  и общее решение как их линейную комбинацию.

Ситуация усложняется, если характеристический многочлен имеет кратные корни. В этом случае Лагранж в своей знаменитой «Аналитической Механике» [(1788), вторая часть, разд. VI, № 7] утверждает, что в решение входят «вековые» члены, аналогичные

тем, которые получаются из формулы (4.8). Однако веком позже стало понятно, что это не совсем верно. (См., например, работу Вейерштрасса (1858): «...воспользуясь случаем, чтобы устраниТЬ недоразумение, содержащееся в лагранжевой теории малых колебаний и во всех известных мне более поздних изложениях последней».) Поэтому отложим выяснение данного вопроса до разд. I.12.

В завершение раздела рассмотрим подробное решение уравнений (6.1), так как полученные при этом результаты чрезвычайно важны [Лагранж (1759)]. Соответствующая задача на собственные значения (6.4) принимает в этом случае вид  $p^2 c_1 = K^2 (-2c_1 + c_2)$ ,  $p^2 c_i = K^2 (c_{i-1} - 2c_i + c_{i+1})$  при  $i = 2, \dots, n-1$  и  $p^2 c_n = K^2 (c_{n-1} - 2c_n)$ . Введем обозначение  $p^2/K^2 + 2 = q$ . Тогда

$$c_{j+1} - qc_j + c_{j-1} = 0, \quad c_0 = 0, \quad c_{n+1} = 0. \quad (6.5)$$

Это означает, что  $c_i$  являются решениями *разностного уравнения*, а поэтому <sup>1)</sup>  $c_j = Aa^j + Bb^j$ , где  $a$  и  $b$  — корни соответствующего характеристического уравнения  $z^2 - qz + 1 = 0$ , и, следовательно, для них выполняются равенства

$$a + b = q, \quad ab = 1. \quad (6.6)$$

Условие  $c_0 = 0$  из (6.5), приводящее к равенству  $A + B = 0$ , показывает, что  $c_j = A(a^j - b^j)$  при  $A \neq 0$ . Из второго условия  $c_{n+1} = 0$  или эквивалентного ему условия  $(a/b)^{n+1} = 1$  с учетом соотношения  $ab = 1$  следует, что

$$a = \exp\left(\frac{k\pi i}{n+1}\right), \quad b = \exp\left(-\frac{k\pi i}{n+1}\right),$$

где  $k$  может иметь любое из значений  $k = 1, \dots, n$ . Таким образом, получим

$$q_k = 2 \cos \frac{\pi k}{n+1}, \quad k = 1, \dots, n, \quad (6.7a)$$

$$p_k^2 = 2K^2 \left( \cos \frac{\pi k}{n+1} - 1 \right) = -4K^2 \left( \sin \frac{\pi k}{2n+2} \right)^2. \quad (6.7b)$$

И наконец, формула Эйлера, открытая им в 1740 г.,  $e^{ix} - e^{-ix} = 2i \sin x$  («...такая привычная сегодня для геометров»), дает для собственных векторов (при  $A = -i/2$ ) значения

$$c_j^{(k)} = \sin \frac{jk\pi}{n+1}, \quad j, k = 1, \dots, n. \quad (6.8)$$

Поскольку  $p_k$  — чисто мнимые величины, мы используем «привычную» формулу также и для  $\exp(\pm p_k t)$  и получим, наконец,

<sup>1)</sup> См., например: Н. С. Бахвалов, Численные методы. — М.: Наука, 1973, гл. II., § 7. — Прим. ред.

общее решение

$$y_j(t) = \sum_{k=1}^n \sin \frac{jk\pi}{n+1} (a_k \cos r_k t + b_k \sin r_k t), \quad r_k = 2K \sin \frac{\pi k}{2n+2}. \quad (6.9)$$

В результате длительных расчетов Лагранж получил, что значения

$$a_k = \frac{2}{n+1} \sum_{l=1}^n \sin \frac{kj\pi}{n+1} y_l(0),$$

$$b_k = \frac{1}{r_k} \frac{2}{n+1} \sum_{l=1}^n \sin \frac{kj\pi}{n+1} y'_l(0)$$

определяются начальными положениями и скоростями материальных точек. В настоящее время эти выражения получаются с помощью соотношений ортогональности

$$\sum_{l=1}^n \sin \frac{l j \pi}{n+1} \sin \frac{l k \pi}{n+1} = \begin{cases} 0 & , j \neq k, \\ \frac{n+1}{2} & , j = k, \end{cases} \quad j, k = 1, \dots, n.$$

Он изучил также случай, когда число материальных точек  $n$  стремится к бесконечности (при этом в формуле для  $r_k$  можно заменить  $\sin x$  на  $x$ ), и, опередив на 50 лет Фурье, оказался у порога открытия теории рядов Фурье. «Первую часть этого исследования он провел с большим умением, применяя много искусственных аналитических приемов, однако переход от конечных сумм к бесконечным оставляет желать много лучшего...» [Риман (1854)].

## Фурье

«J'ajouterai que le livre de Fourier a une importance capitale dans l'histoire des mathématiques».

«Добавлю, что книга Фурье имеет основополагающее значение в истории математики».

(А. Пуанкаре, 1893)

Первые системы первого порядка появились благодаря задаче о теплопроводности [Био (1804), Фурье (1807)]. Фурье представил стержень последовательностью молекул, температуру которых мы обозначим через  $y_i$ , и вывел из закона Ньютона, что энергия, передаваемая данной частицей соседним, пропорциональна раз-

ности их температур, т. е.  $y_{i-1} - y_i$  для левой пары и  $y_{i+1} - y_i$  — для правой. («Поскольку две молекулы одного тела расположены чрезвычайно близко друг к другу и имеют разные температуры, более разогретая молекула сообщает менее разогретой количеством тепла, которое точно выражается через произведение, образованное длительностью промежутка времени, очень малой разницей температур и некоторой функцией расстояния между молекулами.») На языке формул эта длинная фраза означает, что полное приращение энергии молекулы  $i$  выражается следующим образом:

$$y'_i = K^2 (y_{i-1} - 2y_i + y_{i+1}), \quad (6.10)$$

или, в более общем виде — системой первого порядка с постоянными коэффициентами:

$$y'_i = \sum_{j=1}^n a_{ij} y_j, \quad i = 1, \dots, n. \quad (6.11)$$

Подставив теперь выражение  $y_i = c_i e^{pt}$ , получим задачу на собственные значения

$$pc_i = \sum_{j=1}^n a_{ij} c_j, \quad i = 1, \dots, n. \quad (6.12)$$

Если предположить, что стержень охлажден на обоих концах до нуля ( $y_0 = y_{n+1} = 0$ ), то можно использовать введенные выше собственные векторы Лагранжа, с помощью которых мы получим решение

$$y_j(t) = \sum_{k=1}^n \sin \frac{jk\pi}{n+1} a_k \exp(-\omega_k t), \quad \omega_k = 4K^2 \left( \sin \frac{\pi k}{2n+2} \right)^2. \quad (6.13)$$

Неограниченно увеличивая  $n$  (и тем самым обращая «метод прямых», Фурье от (6.10) пришел к своему знаменитому уравнению теплопроводности

$$\frac{\partial u}{\partial t} = a^2 \frac{\partial^2 u}{\partial x^2}. \quad (6.14)$$

Так было положено начало теории рядов Фурье.

## Упражнения

1. Проверьте, что для функции  $u(x)$ , дифференцируемой достаточное число раз, выполняется равенство

$$\frac{u(x-\delta) - 2u(x) + u(x+\delta)}{\delta^2} = u''(x) + \frac{\delta^2 u^4(x)}{12} + O(\delta^4).$$

**Указание.** Используйте разложение в ряд Тейлора для  $u(x+\delta)$  и  $u(x-\delta)$ . Это соотношение устанавливает связь между (6.10) и (6.14), а также между (6.1) и волновым уравнением.

2. Решите уравнение (6.2) при  $n = 2$  и  $n = 3$ , следуя описанной выше схеме Лагранжа (1762). При этом естественным образом возникает характеристический многочлен матрицы.
3. Решите систему первого порядка (6.11) с начальными значениями  $y_i(0) = (-1)^i$  и той же матрицей  $A$ , что и в упр. 2, и начертите решения. Физически такое уравнение могло бы описывать струну с грузами, висящую, например, в мёде.
4. Найдите первые члены разложения в особой точке  $x = 0$  для следующей системы нелинейных уравнений:

$$\begin{aligned} x^2y'' + 2xy' &= 2yz^2 + \lambda x^2y(y^2 - 1), \quad y(0) = 0, \\ x^2z'' &= z(z^2 - 1) + x^2y^2z, \quad z(0) = 1, \end{aligned} \quad (6.15)$$

с постоянным параметром  $\lambda$ . Это уравнение Эйлера для вариационной задачи

$$I = \int_0^\infty \left( z'^2 + \frac{x^2y'^2}{2} + \frac{(z^2 - 1)}{2x^2} + y^2z^2 + \frac{\lambda}{4} x^2 (y^2 - 1)^2 \right) dx,$$

$$y(\infty) = 1, \quad z(\infty) = 0,$$

о массе «монополя» из ядерной физики [см. работу Хуфта (1974)].

## I.7. Общая теорема существования

«M. Cauchy annonce, que, pour se conformer au voeu du Conseil, il ne s'attachera plus à donner, comme il a fait jusqu'à présent, des démonstrations parfaitement rigoureuses». (Conseil d'instruction de l'Ecole polytechnique, 24 nov. 1825)

«Г-н Коши объявляет о том, что, выполняя положение Совета, он больше не будет стремиться давать совершенно строгие доказательства, как он это делал до сих пор».

(Совет по обучению Политехнической школы, 24 ноября 1825 г.)

«У всех вас профессиональная деформация сознания; здесь сходимость совершенно не важна...»  
(П. Хенричи, 1985)

Теперь мы вступаем в новую эру развития предмета нашей книги, более теоретическую, чем предыдущая. Ее открывают исследования Коши, который не был столь очарован долгими численными расчетами, как, например, Эйлер, зато фанатично добивался безупречной математической строгости и точности. Он критиковал своих предшественников за использование в работе бесконечных рядов и других бесконечных процессов без должного внимания к оценкам погрешности или результатам о сходимости. Около 1820 г. Коши сформулировал теорему сходимости для метода ломаных Эйлера, а лет через 15 — для метода степенных рядов Ньютона (см. разд. I.8). Кроме оценки погрешности эти результаты позволяют также получить общие теоремы существования решений произвольных дифференциальных уравнений («любого дифференциального уравнения»), для которых решения были известны только в весьма немногих случаях. Другим важным следствием стала возможность судить о единственности решения, позволяющей делать вывод о том, что только найденное нами (численно или аналитически) решение отвечает данному начальному значению и других таких решений не существует. Лишь в этом случае мы вправе сказать, что решили задачу.

Самое первое доказательство Коши было недавно найдено в отрывочных заметках [Коши (1824)], которые никогда при его жизни не публиковались (может быть, эти заметки не удовлетворили министра образования: «...что второй профессор г-н Коши представил только отдельные листки, которые не могли удовлетворить комиссию, и что до сих пор было невозможно заставить его подчиниться требованию Совета и выполнить постановление министра»).

## Сходимость метода Эйлера

Приступим теперь, «обнажив голову и преклонив колени», к изложению идей этого исторического доказательства. Сформулируем его так, чтобы оно допускало прямое обобщение на системы большей размерности.

Начнем со скалярного дифференциального уравнения

$$y' = f(x, y), \quad y(x_0) = y_0, \quad y(X) = ? \quad (7.1)$$

и воспользуемся методом, который описан Эйлером (1768) в последнем разделе его работы «Institutiones Calculi Integralis» (Caput VII), т. е. рассмотрим разбиение отрезка интегрирования

$$x_0, x_1, \dots, x_{n-1}, x_n = X, \quad (7.2)$$

и на каждом отрезке заменим решение первым членом его ряда Тейлора:

$$\begin{aligned} y_1 - y_0 &= (x_1 - x_0) f(x_0, y_0), \\ y_2 - y_1 &= (x_2 - x_1) f(x_1, y_1), \\ &\dots \\ y_n - y_{n-1} &= (x_n - x_{n-1}) f(x_{n-1}, y_{n-1}). \end{aligned} \quad (7.3)$$

Для указанного разбиения будем использовать также обозначение

$$h = (h_0, h_1, \dots, h_{n-1}),$$

где  $h_i = x_{i+1} - x_i$ . Если соединить  $y_0$  и  $y_1$ ,  $y_1$  и  $y_2$  и т. д. прямыми, то получится ломаная Эйлера

$$y_h(x) = y_0 + (x - x_0) f(x_0, y_0) \text{ при } x_0 \leq x \leq x_1. \quad (7.3a)$$

**Лемма 7.1.** Предположим, что  $|f|$  не превосходит  $A$  на множестве

$$D = \{(x, y) \mid x_0 \leq x \leq X, |y - y_0| \leq b\}.$$

Если  $X - x_0 \leq b/A$ , то заданное формулами (7.3) численное решение  $(x_i, y_i)$  останется в  $D$  при любом разбиении (7.2) и будут выполняться неравенства

$$|y_h(x) - y_0| \leq A \cdot |x - x_0|, \quad (7.4)$$

$$|y_h(x) - (y_0 + (x - x_0) f(x_0, y_0))| \leq \varepsilon |x - x_0|, \quad (7.5)$$

если  $|f(x, y) - f(x_0, y_0)| \leq \varepsilon$  на  $D$ .

**Доказательство.** Оба неравенства получаются в результате суммирования строк (7.3) и использования неравенства треугольника. Из (7.4) сразу следует, что при  $A(x - x_0) \leq b$  ломаная остается в  $D$ . ■

Наша следующая задача состоит в оценке изменения  $y_h(x)$  при изменении  $y_0$ . Возьмем другое начальное значение  $z_0$  и вы-

числим для него

$$z_1 - z_0 = (x_1 - x_0) f(x_0, z_0). \quad (7.6)$$

Нам надо оценить  $|z_1 - y_1|$ . Вычитая (7.6) из первой формулы в (7.3), получим

$$z_1 - y_1 = z_0 - y_0 + (x_1 - x_0) (f(x_0, z_0) - f(x_0, y_0)).$$

Отсюда видно, что нам нужна оценка для  $f(x_0, z_0) - f(x_0, y_0)$ . Если предположить, что

$$|f(x, z) - f(x, y)| \leq L |z - y|, \quad (7.7)$$

то мы приедем к неравенству

$$|z_1 - y_1| \leq (1 + (x_1 - x_0) L) |z_0 - y_0|. \quad (7.8)$$

**Лемма 7.2.** Пусть при фиксированном разбиении  $h$  начальными значениями  $y_0$  и  $z_0$  отвечают ломаные Эйлера  $y_h(x)$  и  $z_h(x)$  соответственно. Если в выпуклой области, содержащей  $(x, y_h(x))$  и  $(x, z_h(x))$  при всех  $x_0 \leq x \leq X$ , выполняется соотношение

$$\left| \frac{\partial f}{\partial y}(x, y) \right| \leq L, \quad (7.9)$$

то

$$|z_h(x) - y_h(x)| \leq e^{L(x-x_0)} |z_0 - y_0|. \quad (7.10)$$

*Доказательство.* Из (7.9) следует (7.7), из (7.7) — (7.8), а из (7.8) — неравенство

$$|z_1 - y_1| \leq e^{L(x_1-x_0)} |z_0 - y_0|.$$

Если повторить те же рассуждения для  $z_2 - y_2$ ,  $z_3 - y_3$  и так далее, то в итоге мы получим (7.10). ■

*Замечание.* Условие (7.7) называется «условием Липшица». Именно Липшиц (1876) заново открыл эту теорию (примечание к статье Липшица: «Автор, очевидно, не знаком с работами Коши...») и отстаивал использование условия (7.7) вместо более сильного предположения (7.9). Доказательство Липшица излагается также в классическом труде Пикара [(1891—1896), том II, гл. XI, разд. I].

При переходе ко все более мелкому разбиению (7.2), т. е. при

$$|h| := \max_{i=0, \dots, n-1} h_i \rightarrow 0,$$

мы надеемся, что ломаная Эйлера сойдется к решению задачи (7.1). И действительно, мы имеем следующий результат.

**Теорема 7.3.** Пусть на множестве

$$D = \{(x, y) | x_0 \leq x \leq X, |y - y_0| \leq b\}$$

$f(x, y)$  — непрерывная функция,  $|f|$  не превосходит  $A$  и удовлетворено условие Липшица (7.7). Если  $X - x_0 \leq b/A$ , то:

- при  $|h| \rightarrow 0$  ломаные Эйлера  $y_h(x)$  сходятся равномерно к непрерывной функции  $\varphi(x)$ ;
- $\varphi(x)$  непрерывно дифференцируема и является решением задачи (7.1) на отрезке  $[x_0, X]$ ;
- при  $x_0 \leq x \leq X$  задача (7.1) не имеет других решений.

*Доказательство.* а) Выберем  $\varepsilon > 0$ . Поскольку  $f$  равномерно ограничена на компактном множестве  $D$ , существует  $\delta > 0$ , такое что из условий

$$|u_1 - u_2| \leq \delta \text{ и } |v_1 - v_2| \leq A\delta$$

следует неравенство

$$|f(u_1, v_1) - f(u_2, v_2)| \leq \varepsilon. \quad (7.11)$$

Теперь преположим, что разбиение (7.3) отвечает условию

$$|x_{i+1} - x_i| \leq \delta, \text{ т. е. } |h| \leq \delta. \quad (7.12)$$

Сначала изучим, что произойдет при добавлении точек сетки. На первом шаге рассмотрим разбиение  $h(1)$ , которое получается при добавлении новых точек только в первом отрезку разбиения. Из неравенства (7.5) (относящегося к этому первому отрезку) следует, что для нового уточненного решения  $y_{h(1)}(x_1)$  выполнняется оценка  $|y_{h(1)}(x_1) - y_h(x_1)| \leq \varepsilon |x_1 - x_0|$ . Поскольку разбиения  $h$  и  $h(1)$  совпадают на отрезке  $x_1 \leq x \leq X$ , мы можем применить лемму 7.2, с помощью которой получим неравенство

$$|y_{h(1)}(x) - y_h(x)| \leq e^{L(x-x_1)} (x_1 - x_0) \varepsilon \text{ при } x_1 \leq x \leq X.$$

Затем добавим дополнительные точки на интервале  $(x_1, x_2)$  и обозначим новое разбиение через  $h(2)$ . Тем же способом придем к оценкам

$$|y_{h(2)}(x_2) - y_{h(1)}(x_2)| \leq \varepsilon |x_2 - x_1|$$

и

$$|y_{h(2)}(x) - y_{h(1)}(x)| \leq e^{L(x-x_2)} (x_2 - x_1) \varepsilon \text{ при } x_2 \leq x \leq X.$$

Полностью вся ситуация представлена на рис. 7.1. Обозначив через  $\hat{h}$  окончательное разбиение, при  $x_l \leq x \leq x_{l+1}$  мы получим

$$\begin{aligned} & |y_{\hat{h}}(x) - y_h(x)| \leq \\ & \leq \varepsilon (e^{L(x-x_1)} (x_1 - x_0) + \dots + e^{L(x-x_l)} (x_l - x_{l-1})) + \varepsilon (x - x_l) \leq \\ & \leq \varepsilon \int_{x_0}^x e^{L(s-x)} ds = \frac{\varepsilon}{L} (e^{L(x-x_0)} - 1). \end{aligned} \quad (7.13)$$

Теперь, если мы имеем два различных разбиения  $h$  и  $\tilde{h}$  и они оба удовлетворяют условию (7.12), введем *третье* разбиение  $\tilde{h}$ , которое получается измельчением каждого из них (как это обычно делается при доказательстве существования интеграла Римана), и дважды применим (7.13). Тогда из неравенства треугольника и из (7.13) следует, что

$$|y_h(x) - y_{\tilde{h}}(x)| \leq 2 \frac{\varepsilon}{L} (e^{L(x-x_0)} - 1).$$

При достаточно малых  $\varepsilon > 0$  эта величина становится сколь угодно малой, что говорит о равномерной сходимости ломаных Эйлера к некоторой непрерывной функции  $\varphi(x)$ .

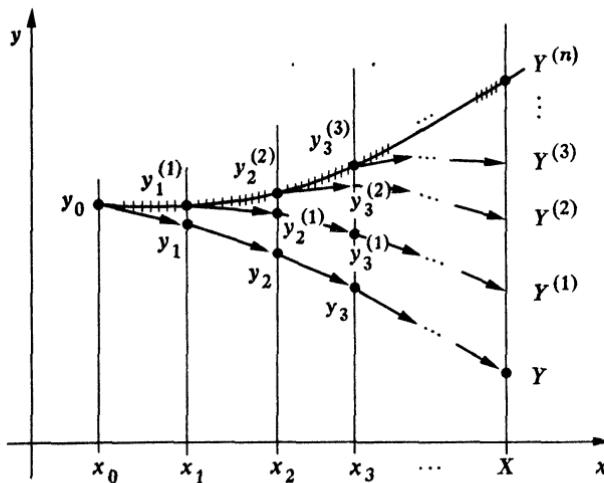


Рис. 7.1. Веер леди Уиндермир [О. Уайльд (1892)].

b) Введем модуль непрерывности

$$\varepsilon(\delta) := \sup \{|f(u_1, v_1) - f(u_2, v_2)|;$$

$$|u_1 - u_2| \leq \delta, \quad |v_1 - v_2| \leq A\delta, \quad (u_i, v_i) \in D\}.$$

Если  $x$  принадлежит разбиению  $h$ , то из (7.5) [заменив  $(x_0, y_0)$  на  $(x, y_h(x))$  и  $x$  на  $x + \delta$ ] мы получим

$$|y_h(x + \delta) - g_h(x) - \delta f(x, y_h(x))| \leq \varepsilon(\delta) \delta. \quad (7.14)$$

Переходя к пределу при  $|h| \rightarrow 0$ , мы придем к соотношению

$$\begin{aligned} |\varphi(x + \delta) - \varphi(x) - \delta f(x, \varphi(x))| &\leq \\ &\leq \varepsilon(\delta) \delta. \end{aligned} \quad (7.15)$$

Поскольку  $\varepsilon(\delta) \rightarrow 0$  при  $\delta \rightarrow 0$ , отсюда следует, что  $\varphi(x)$  дифференцируема и  $\varphi'(x) = f(x, \varphi(x))$ .

с) Предположим, что задача (7.1) имеет второе решение  $\psi(x)$ , и пусть разбиение  $h$  отвечает условию (7.12). Обозначим через  $y_h^{(i)}(x)$  ломаную Эйлера при начальном значении  $(x_i, \psi(x_i))$  (она определена при  $x_i \leq x \leq X$ ). Из соотношения

$$\psi(x) = \psi(x_i) + \int_{x_i}^x f(s, \psi(s)) ds$$

и (7.11) следует, что

$$|\psi(x) - y_h^{(i)}(x)| \leq \varepsilon |x - x_i| \text{ при } x_i \leq x \leq x_{i+1}.$$

Используя лемму 7.2, мы тем же способом, что и в части а), выведем оценку

$$|\psi(x) - y_h(x)| \leq \frac{\varepsilon}{L} (e^{L(x-x_0)} - 1). \quad (7.16)$$

Переходя к пределу при  $|h| \rightarrow 0$  и  $\varepsilon \rightarrow 0$ , получим неравенство  $|\psi(x) - \varphi(x)| \leq 0$ , доказывающее единственность. ■

Теорема 7.3 о существовании и единственности решения — это локальный результат. Однако если рассматривать конечную точку решения как новое начальное значение, то можно снова применить теорему 7.3 и продолжить решение. Повторяя эту процедуру, мы придем к следующему результату.

**Теорема 7.4.** Пусть  $U$  является открытым множеством в  $\mathbb{R}^2$ ,  $a$  и  $f/\partial y$  непрерывны в  $U$ . Тогда для любой точки  $(x_0, y_0) \in U$  существует единственное решение задачи (7.1), которое можно продолжить до границы  $U$  (в обоих направлениях).

*Доказательство.* Теорему 7.3, очевидно, можно переписать так, чтобы она утверждала существование и единственность решения для интервала  $(X, x_0)$  слева от  $x_0$ . Остальное следует из того факта, что любая точка в  $U$  имеет окрестность, удовлетворяющую условиям теоремы 7.3. ■

Интересно упомянуть о том, что формула (7.13) при  $|h| \rightarrow 0$  дает оценку погрешности

$$|y(x) - y_h(x)| \leq \frac{\varepsilon}{L} (e^{L(x-x_0)} - 1) \quad (7.17)$$

для ломаной Эйлера ( $|h| \leq \delta$ ). Здесь  $y(x)$  означает точное решение задачи (7.1). Следующая теорема уточняет приведенные выше оценки в случае, когда  $f(x, y)$  является также дифференцируемой по  $x$ .

**Теорема 7.5.** Предположим, что в окрестности решения выполняются неравенства

$$|f| \leq A, \quad \left| \frac{\partial f}{\partial y} \right| \leq L, \quad \left| \frac{\partial f}{\partial x} \right| \leq M.$$

Тогда для ломаных Эйлера при достаточно малых  $|h|$  справедлива следующая оценка погрешности:

$$|y(x) - y_h(x)| \leq \frac{M + AL}{L} (e^{L|x-x_0|} - 1) \cdot |h|. \quad (7.18)$$

**Доказательство.** При  $|u_1 - u_2| \leq |h|$  и  $|v_1 - v_2| \leq A|h|$  в силу дифференцируемости  $f$  мы получим вместо (7.11) оценку

$$|f(u_1, v_1) - f(u_2, v_2)| \leq (M + AL)|h|.$$

Подставив эту величину в качестве  $\epsilon$  в (7.16), мы придем к требуемому результату. ■

Оценка (7.18) показывает, что глобальная погрешность метода Эйлера пропорциональна максимальной длине шага  $|h|$ . Тогда для достижения точности, например, в три десятичных знака нам понадобилось бы порядка тысячи шагов; точность в шесть десятичных знаков требует, как правило, миллиона шагов и т. д. Таким образом, мы видим, что данный метод нельзя рекомендовать для расчетов с высокой точностью. Поиск методов, которые сходятся быстрее, составит основное содержание гл. II.

### Теорема существования Пеано

«Si  $a$  est un complexe d'ordre  $n$ , et  $b$  un nombre réel, alors on peut déterminer  $b'$  et  $f$ , où  $b'$  est une quantité plus grande que  $b$ , et  $f$  est un signe de fonction qui à chaque nombre de l'intervalle de  $b$  à  $b'$  fait correspondre un complexe (en d'autres mots,  $ft$  est un complexe fonction de la variable réelle  $t$ , définie pour toutes les valeurs de l'intervalle  $(b, b')$ ; la valeur de  $ft$  pour  $t=b$  est  $a$ ; et dans tout l'intervalle  $(b, b')$  cette fonction  $ft$  satisfait à l'équation différentielle donnée».

«Если  $a$  — комплексное число порядка  $n$ , а  $b$  — вещественное, тогда можно определить  $b'$  и  $f$ , где  $b'$  — величина, превосходящая  $b$ , а  $f$  обозначает функцию, которая каждому числу из интервала от  $b$  до  $b'$  ставит в соответствие комплексное число (другими словами,  $ft$  — это комплексная функция вещественной переменной  $t$ , определенная для всех величин из интервала  $(b, b')$ ); значение  $ft$  при  $t = b$  равно  $a$ , и на всем интервале  $(b, b')$  эта функция  $ft$  удовлетворяет заданному дифференциальному уравнению».

(Первоначальная формулировка теоремы Пеано)

Условие Липшица (7.7) — это решающее средство при доказательстве соотношения (7.10), а в итоге и теоремы сходимости. Если мы полностью откажемся от условия (7.7) и потребуем только непрерывности  $f(x, y)$ , то уже нельзя будет гарантировать сходимость ломаных Эйлера.

Пусть, например, задано уравнение

$$y' = 4 \left( \operatorname{sign}(y) |y|^{1/2} + \max \left( 0, x - \frac{|y|}{x} \right) \cdot \cos \left( \frac{\pi \log x}{\log 2} \right) \right). \quad (7.19)$$

Оно подобрано таким образом, чтобы выполнялись равенства

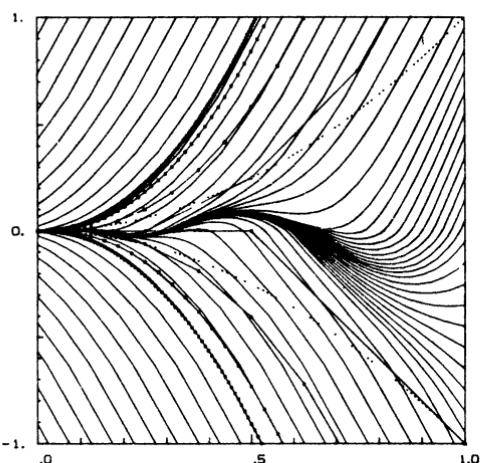


Рис. 7.2. Кривые решений и ломаные Эйлера для уравнения (7.19).

$$\begin{aligned} f(h, 0) &= 4h(-1)^i \\ \text{при } h &= 2^{-i}, \\ f(x, y) &= 4 \operatorname{sign}(y) \cdot |y|^{1/2} \\ \text{при } |y| &\geq x^2. \end{aligned}$$

Начальному значению  $y(0) = 0$  соответствует бесконечное множество решений; часть из них изображена на рис. 7.2. Ломаные Эйлера сходятся при  $h = 2^{-i}$  и четных  $i$  к максимальному решению  $y = 4x^2$ , а при нечетных  $i$  — к  $y = -4x^2$ . При других последовательностях  $h$  можно получить и все промежуточные решения.

**Теорема 7.6.** [Пeanо (1890)] Пусть  $f(x, y)$  непрерывна, а  $|f|$  не превосходит  $A$  на множестве

$$D = \{(x, y) \mid x_0 \leq x \leq X, |y - y_0| \leq b\}.$$

Если  $X - x_0 \leq b/A$ , то существует подпоследовательность последовательности ломаных Эйлера, сходящаяся к решению дифференциального уравнения.

Доказательство самого Пеано в главной части, касающейся сходимости, является очень кратким и неясным для неискушенных читателей. Арцела (1895), вернувшись к этому вопросу, объясняет свою идею более подробно и подчеркивает необходимость *равностепенной* непрерывности рассматриваемой последовательности. Доказательство того, что называют теоремой Арцела — Асколи, которое обычно приводят в настоящее время, появилось

лишь позднее (см., например, работы Перрона (1918), Хана (1921)). Вкратце его можно изложить следующим образом.

*Доказательство.* Рассмотрим последовательность ломаных Эйлера при уменьшении длины шага:

$$v_1(x), v_2(x), v_3(x), \dots . \quad (7.20)$$

Из (7.4) следует, что при фиксированном  $x$  она ограничена. Выберем последовательность чисел  $r_1, r_2, r_3, \dots$ , плотную на интервале  $(x_0, X)$ . Тогда в (7.20) существует последовательность, сходящаяся при  $x = r_1$  (Больцано—Вейерштрасс). Обозначим ее через

$$v_1^{(1)}(x), v_2^{(1)}(x), v_3^{(1)}(x), \dots \quad (7.21)$$

Далее выберем из (7.21) подпоследовательность, сходящуюся при  $x = r_2$ :

$$v_1^{(2)}(x), v_2^{(2)}(x), v_3^{(2)}(x), \dots \quad (7.22)$$

и т. д. Затем возьмем «диагональную» последовательность

$$v_1^{(1)}(x), v_2^{(2)}(x), v_3^{(3)}(x), \dots, \quad (7.23)$$

которая, за исключением конечного числа членов, является подпоследовательностью всех этих последовательностей и поэтому сходится при всех  $r_j$ . В завершение с помощью оценки

$$|v_n^{(n)}(x) - v_n^{(n)}(r_j)| \leq A|x - r_j|$$

[см. (7.4)], которая выражает равностепенную непрерывность последовательности, мы получим

$$\begin{aligned} & |v_n^{(n)}(x) - v_m^{(m)}(x)| \leq \\ & \leq |v_n^{(n)}(x) - v_n^{(n)}(r_j)| + |v_n^{(n)}(r_j) - v_m^{(m)}(r_j)| + \\ & + |v_m^{(m)}(r_j) - v_m^{(m)}(x)| \leq \\ & \leq 2A|x - r_j| + |v_n^{(n)}(r_j) - v_m^{(m)}(r_j)|. \end{aligned}$$

Для фиксированного  $\varepsilon > 0$  мы выберем из  $\{r_1, r_2, \dots\}$  конечное подмножество  $R$ , отвечающее условию  $\min \{|x - r_j| \mid r_j \in R, x_0 \leq x \leq X\} \leq \varepsilon/A$ , а затем выберем такое  $N$ , что

$$|v_n^{(n)}(r_j) - v_m^{(m)}(r_j)| \leq \varepsilon \text{ при } n, m \in R \text{ и } r_j \in R.$$

Это неравенство указывает на равномерную сходимость (7.23). Таким же способом, как и в части б) доказательства теоремы 7.3, мы придем к выводу, что получаемая в пределе функция — это решение задачи (7.1). Надо только добавить в (7.14) член  $O(|h|)$ , если  $x$  не является точкой разбиения. ■

## Упражнения

1. Примените метод Эйлера с постоянным шагом  $x_{i+1} - x_i = 1/n$  к дифференциальному уравнению  $y' = ky$ ,  $y(0) = 1$  и получите классическую аппроксимацию решения  $y(1) = e^k$ . Оцените погрешность.
2. Примените метод Эйлера с постоянным шагом к задачам

a)  $y' = y^2$ ,  $y(0) = 1$ ,  $y\left(\frac{1}{2}\right) = ?$

b)  $y' = x^2 + y^2$ ,  $y(0) = 0$ ,  $y\left(\frac{1}{2}\right) = ?$

С помощью теоремы 7.5 дайте строгие оценки погрешности и сравните их с истинными погрешностями. Основная трудность состоит в нахождении нужной области, в которой выполнялись бы оценки теоремы 7.5 и при этом постоянные  $A$ ,  $L$  и  $M$  не были бы слишком большими, но в то же время гарантировали, что графики решений останутся внутри области (см. также I.8, упр. 3).

3. Докажите следующий результат: если дифференциальное уравнение  $y' = f(x, y)$ ,  $y(x_0) = y_0$  с непрерывной функцией  $f$  имеет единственное решение, то ломаные Эйлера сходятся к нему.
4. «Существует элементарное доказательство теоремы существования Пеано» [Уолтер (1971)]. Предположим, что  $|f|$  не пре-  
восходит  $A$ . Тогда последовательность

$$y_{i+1} = y_i + h \max \{f(x, y) | x_i \leq x \leq x_{i+1}, y_i - 3Ah \leq y \leq y_i + Ah\}$$

сходится для всех непрерывных  $f$  к (максимальному) решению. Попытайтесь доказать это. К сожалению, такое доказательство нельзя распространить на системы уравнений, если они не являются «квазимонотонными» (см. разд. I.10, упр. 3).

## I.8. Теория существования решения, основанная на итерационных методах и рядах Тейлора

Возможен и другой подход к теории существования решения — на основе итерационного уточнения приближенных решений. К этой мысли пришли уже очень давно. В частности, много примеров такого типа можно найти в работе Лагранжа, главным образом в его астрономических расчетах. В качестве иллюстрации рассмотрим уравнение Риккетти

$$y' = x^2 + y + 0.1y^2, \quad y(0) = 0. \quad (8.1)$$

Оно не имеет элементарного решения из-за вхождения в правую часть квадратичного члена. Поэтому совершенно естественно возникает мысль пренебречь данным членом, который действительно вначале очень мал, и решать пока задачу

$$\tilde{y}_1' = x^2 + \tilde{y}_1, \quad \tilde{y}_1(0) = 0. \quad (8.2)$$

С помощью формулы (3.3) мы получим отсюда первое решение

$$y_1(x) = 2e^x - (x^2 + 2x + 2). \quad (8.3)$$

Благодаря ему мы теперь больше знаем о члене  $0.1 y^2$ , которым первоначально пренебрегли: он будет близок к  $0.1 \tilde{y}_1^2$ . Поэтому напрашивается мысль о том, чтобы данное решение снова подставить в (8.1) и решать уже дифференциальное уравнение

$$\tilde{y}_2' = x^2 + \tilde{y}_2 + 0.1(y_1(x))^2, \quad \tilde{y}_2(0) = 0. \quad (8.4)$$

Можно еще раз использовать формулу (3.3), и после некоторых вычислений мы получим выражение

$$\begin{aligned} y_2(x) = y_1(x) + \frac{2}{5}e^{2x} - \frac{2}{15}e^x(x^3 + 3x^2 + 6x - 54) - \\ - \frac{1}{10}(x^4 + 8x^3 + 32x^2 + 72x + 76). \end{aligned}$$

Оно уже намного ближе к точному решению, что видно из сравнения погрешностей  $e_1 = y(x) - y_1(x)$  и  $e_2 = y(x) - y_2(x)$ :

$$x = 0.2 \quad e_1 = 0.228 \times 10^{-07} \quad e_2 = 0.233 \times 10^{-12}$$

$$x = 0.4 \quad e_1 = 0.327 \times 10^{-05} \quad e_2 = 0.566 \times 10^{-09}$$

$$x = 0.8 \quad e_1 = 0.534 \times 10^{-03} \quad e_2 = 0.165 \times 10^{-05}.$$

Продолжение данного процесса выглядит многообещающим, но расчеты вскоре становятся чрезвычайно утомительными.

## Метод последовательных приближений Пикара

Общая формулировка этого метода заключается в следующем: мы попытаемся, если это возможно, таким образом расщепить функцию  $f(x, y)$  из дифференциального уравнения

$$y' = f(x, y) = f_1(x, y) + f_2(x, y); \quad y(x_0) = y_0, \quad (8.5)$$

чтобы любое дифференциальное уравнение вида  $y' = f_1(x, y) + g(x)$  можно было решить аналитически, а  $f_2(x, y)$  была мала. Тогда мы начнем с первого приближения  $y_0(x)$  и, решая уравнения

$$y_{i+1}' = f_1(x, y_{i+1}) + f_2(x, y_i(x)), \quad y_{i+1}(x_0) = y_0, \quad (8.6)$$

вычислим последовательно  $y_1(x), y_2(x), \dots$ . Самая примитивная разновидность этого процесса получается при выборе  $f_1 = 0, f_2 = f$ ; тогда, проинтегрировав (8.6), мы сразу придем к формуле

$$y_{i+1}(x) = y_0 + \int_{x_0}^x f(s, y_i(s)) ds. \quad (8.7)$$

Она называется методом *последовательных приближений Пикара*. Этот метод появлялся в литературе несколько раз, в частности в работах Лиувилля (1838), Коши, Пеано (1888), Линделёфа (1894), Бендикссона (1893). Пикар (1890) считал его лишь побочным продуктом аналогичного процесса для уравнений в частных производных и досконально проанализировал его в своем знаменитом трактате [Пикар (1891—96), том II, гл. XI, разд. III].

Основной практический недостаток метода состоит в необходимости многократного вычисления интегралов, что обычно не очень удобно, если вообще возможно, и что быстро становится чрезвычайно утомительным занятием. С другой стороны, метод является хорошим инструментом для теоретических исследований.

Легко увидеть его быструю *сходимость* при малых  $|x - x_0|$ : если мы вычтем из (8.7) такую же формулу, но с заменой  $i$  на  $i-1$ , то придем к выражению

$$y_{i+1}(x) - y_i(x) = \int_{x_0}^x (f(s, y_i(s)) - f(s, y_{i-1}(s))) ds. \quad (8.8)$$

Теперь, применив условие Липшица (7.7) и неравенство треугольника, получим

$$|y_{i+1}(x) - y_i(x)| \leq L \int_{x_0}^x |y_i(s) - y_{i-1}(s)| ds. \quad (8.9)$$

Положив  $y_0(x) \equiv y_0$  и применив неравенство треугольника к формуле (8.7) с индексом  $i = 0$ , выведем оценку

$$|y_1(x) - y_0(x)| \leq A|x - x_0|,$$

где  $A$ , как и в разд. 1.7, ограничивает  $|f|$ . Далее, подставляя это выражение в правую часть (8.9) и многократно повторяя эту подстановку, окончательно приедем к оценке

$$|y_i(x) - y_{i-1}(x)| \leq AL^{i-1} \frac{|x - x_0|^i}{i!}. \quad (8.10)$$

Ее правая часть — это (с точностью до постоянного множителя  $A/L$ ) член ряда Тейлора для  $e^{L|x-x_0|}$ , сходящегося при всех  $x$ . Отсюда следует, что  $|y_{i+k} - y_i|$  становится сколь угодно малым при больших  $i$ . Погрешность ограничена остаточным членом данного ряда. Поэтому последовательность  $y_i(x)$  равномерно сходится к решению  $y(x)$ . Например, если  $L|x - x_0| \leq 1/10$  и постоянная  $A$  невелика (порядка 1. — Ред.), то 10 итераций дали бы численное решение примерно с 17 правильными знаками.

### Метод рядов Тейлора

«Après avoir montré l'insuffisance des méthodes d'intégration fondées sur le développement en séries, il me reste à dire en peu de mots ce qu'on peut leur substituer».

«После того как я показал недостаточность методов интегрирования, основанных на разложении в ряд, мне остается кратко сказать, чем их можно заменить». (Коши)

При доказательстве существования решения возможен и третий способ — на основе изучения сходимости рядов Тейлора для решений. О нем упоминает в примечании Лиувилль (1836), а свое окончательное развитие этот способ получил в работе Коши (1839—1842).

Мы уже встречались с рекурсивным вычислением коэффициентов ряда Тейлора в исследованиях Ньютона (см. разд. I.2). Позднее Эйлер (1768) сформулировал общую процедуру для старших производных решения задачи

$$y' = f(x, y), \quad y(x_0) = y_0, \quad (8.11)$$

которые получаются последовательным дифференцированием:

$$\begin{aligned} y'' &= f_x + f_y y' = f_x + f_y f, \\ y''' &= f_{xx} + 2f_{xy} f + f_{yy} f^2 + f_y (f_x + f_y f) \end{aligned} \quad (8.12)$$

и т. д. Тогда решение имеет вид

$$y(x_0 + h) = y(x_0) + y'(x_0)h + y''(x_0) \frac{h^2}{2!} + \dots \quad (8.13)$$

Формулы (8.12) для старших производных быстро становятся очень сложными. Поэтому Эйлер предложил использовать только несколько членов этого ряда при достаточно малом  $h$  и повторить вычисления от точки  $x_1 = x_0 + h$  («аналитическое продолжение»).

### Доказательство сходимости

Теперь мы наметим основные этапы принадлежащего Коши доказательства сходимости ряда (8.13). Предположим, что  $f(x, y)$  является *аналитической* в окрестности начального значения  $x_0, y_0$ , которое для простоты обозначений поместим в начало координат  $x_0 = y_0 = 0$ :

$$\tilde{f}(x, y) = \sum_{i, l \geq 0} a_{ij} x^i y^l, \quad (8.14)$$

где  $a_{ij}$  — кратные частных производных, входящих в (8.12). Если предположить, что ряд (8.14) сходится при  $|x| \leq r, |y| \leq r$ , то неравенства Коши из классического комплексного анализа дают оценку

$$|a_{ij}| \leq \frac{M}{r^{i+l}}, \text{ где } M = \max_{|x| \leq r, |y| \leq r} |\tilde{f}(x, y)|. \quad (8.15)$$

Теперь идея состоит в следующем: поскольку в (8.12) все знаки положительны, мы получим наихудший из возможных результатов, если заменим в (8.14) все  $a_{ij}$  на наибольшие возможные значения (8.15) («метод мажорант»):

$$\tilde{f}(x, y) \rightarrow \sum_{i, l \geq 0} M \frac{x^i y^l}{r^{i+l}} = \frac{M}{(1 - x/r)(1 - y/r)}.$$

А мажорирующее дифференциальное уравнение

$$g' = \frac{M}{(1 - x/r)(1 - y/r)}, \quad g(0) = 0,$$

легко интегрируется с помощью разделения переменных (см. разд. I.3) и имеет решение

$$y = r \left( 1 - \left( 1 + 2M \log \left( 1 - \frac{x}{r} \right) \right)^{1/2} \right). \quad (8.16)$$

Это решение представляется степенным рядом, сходящимся при всех  $x$ , таких, что  $|2M \log(1 - x/r)| < 1$ . Поэтому ряд (8.13) также сходится, по крайней мере при всех

$$|h| < r(1 - \exp(-1/2 M)). \blacksquare$$

## Рекуррентное вычисление коэффициентов ряда Тейлора

«...dieses Verfahren praktisch nicht in Frage kommen kann».

«...об этом способе практически не может быть и речи». (Рунге—Кёниг, 1924)

«Если мы используем правильный подход, то верным будет в точности обратное...».

(Р. Е. Мур, 1979)

«Правильный подход» в действительности является развитием подхода Ньютона. Его открывали заново несколько раз [см., например, работу Стеффенсена (1956)], а программная реализация его общей формы была осуществлена Гиббонсом (1960) и Муром (1966). Более широко библиография представлена в работе Ваннера (1969).

Идея подхода состоит в следующем: пусть величины

$$Y_i = \frac{1}{i!} y^{(i)}(x_0), \quad F_i = \frac{1}{i!} (f(x, y(x)))^{(i)}|_{x=x_0} \quad (8.17)$$

обозначают коэффициенты ряда Тейлора для  $y(x)$  и  $f(x, y(x))$ . Тогда (8.13) примет вид

$$y(x_0 + h) = \sum_{i=0}^{\infty} h^i Y_i.$$

Из (8.11) получим равенство

$$Y_{i+1} = \frac{1}{i+1} F_i. \quad (8.18)$$

Теперь предположим, что  $f(x, y)$  состоит из последовательности алгебраических операций и элементарных функций. Это приводит нас к последовательности элементов

$$x, y, p, q, r, \dots \text{ и, наконец, } f. \quad (8.19)$$

Для каждого из них найдем формулы для определения  $i$ -го коэффициента ряда Тейлора по предыдущим:

a)  $r = p \pm q$ :

$$R_i = P_i \pm Q_i, \quad i = 0, 1, \dots, \quad (8.20a)$$

b)  $r = pq$ :

$$R_i = \sum_{j=0}^i P_j Q_{i-j}, \quad i = 0, 1, \dots, \quad (8.20b)$$

(произведение Коши).

с)  $r = p/q$ : запишем  $p = rq$ , используем формулу б) и разрешим ее относительно  $R_t$ :

$$R_t = \frac{1}{Q_0} \left( P_t - \sum_{j=0}^{t-1} R_j Q_{t-j} \right), \quad t = 0, 1, \dots \quad (8.20c)$$

Такие формулы существуют и для многих элементарных функций (по той причине, что в действительности эти функции сами являются решениями рациональных дифференциальных уравнений). Примеры:

д)  $r = \exp(p)$ : используем равенство  $r' = p'r$  и применим (8.20b). Отсюда получим (при  $t = 1, 2, \dots$ )

$$R_0 = \exp(P_0), \quad R_t = \frac{1}{t} \sum_{j=0}^{t-1} (t-j) R_j P_{t-j}. \quad (8.20d)$$

е)  $r = \log(p)$ : представим формулу как  $p = \exp(r)$  и проведем перестановку в д), откуда следует, что

$$R_0 = \log(P_0), \quad R_t = \frac{1}{P_0} \left( P_t - \frac{1}{t} \sum_{j=1}^{t-1} (t-j) P_j R_{t-j} \right). \quad (8.20e)$$

ф)  $r = p^c$ ,  $c \neq 1$  — постоянная. Используем соотношение  $r' = cp^{c-1}p'$ , т. е.  $pr' = crp'$ , и применим (8.20b):

$$R_0 = P_0^c, \quad R_t = \frac{1}{tP_0} \left( \sum_{j=0}^{t-1} (ci - (c+1)j) R_j P_{t-j} \right). \quad (8.20f)$$

г)  $r = \cos(p)$ ,  $s = \sin(p)$ : как и в д), получим:

$$\begin{aligned} R_0 &= \cos P_0, \quad R_t = -\frac{1}{t} \sum_{j=0}^{t-1} (i-j) S_j P_{t-j}, \\ S_0 &= \sin P_0, \quad S_t = \frac{1}{t} \sum_{j=0}^{t-1} (i-j) R_j P_{t-j}. \end{aligned} \quad (8.20g)$$

Попеременное использование (8.20) и (8.18) позволяет нам очень экономичным способом вычислить коэффициенты ряда Тейлора для (8.17) до любого требуемого порядка. Нетрудно написать для приведенных формул подпрограммы, к которым надо обращаться в том же порядке, в каком элементарные операции входят в дифференциальное уравнение (8.11). Существуют и такие программы, которые для  $f(x, y)$  «компилируют» операторы Фортрана в список вызовов подпрограмм. Одну из них написали Т. Шиманский и Дж. Х. Грей [см. работу Кнаппа—Ваннера (1969)].

*Пример.* Дифференциальное уравнение  $y' = x^2 + y^2$  приводит к рекуррентным соотношениям

$$Y_{t+1} = \frac{1}{t+1} \left( P_t + \sum_{j=0}^t Y_j Y_{t-j} \right), \quad t = 0, 1, \dots, Y_0 = y(0),$$

где  $P_t = 1$  при  $t = 2$  и  $P_t = 0$  при  $t \neq 2$  — коэффициенты для  $x^2$ . Можно представить, насколько это проще, чем формулы (8.12).

Важное свойство данного подхода состоит в том, что его можно применять в *интервальном анализе* и получать с помощью остаточного члена ряда Тейлора в форме Лагранжа *надежные оценки погрешности*. Более подробно этот вопрос изложен в книгах Р. Е. Мура (1966) и (1979).

### Упражнения

1. Получите из (8.10) оценку

$$|y_t(x) - y_0| \leq \frac{1}{L} (e^{L(x-x_0)} - 1) A$$

и объясните, почему этот результат похож на соотношение (7.16).

2. Примените метод Пикара к задаче  $y' = Ky$ ,  $y(0) = 1$ .
3. Выполните три последовательных приближения Пикара для задачи  $y' = x^2 + y^2$ ,  $y(0) = 0$ ,  $y(1/2) = ?$  и дайте строгую оценку погрешности. Сравните результат с точным решением  $y(1/2) = 0.041791146154681863220768806849179$ .
4. Найдите с помощью метода последовательных приближений решение задачи

$$y' = \sqrt{x} + \sqrt{y}, \quad y(0) = 0.$$

Обратите внимание, что метод может хорошо работать и для тех уравнений, которые у других методов вызывают серьезные затруднения. Еще большее различие возникает при решении уравнений

$$y' = \sqrt{x} + y^2, \quad y(0) = 0 \quad \text{и} \quad y' = \frac{1}{\sqrt{x}} + y^2, \quad y(0) = 0.$$

5. Определим  $f(x, y)$  следующим образом:

$$f(x, y) = \begin{cases} 0 & \text{при } x \leq 0, \\ 2x & \text{при } x > 0, y < 0, \\ 2x - \frac{4y}{x} & \text{при } x > 0, 0 \leq y \leq x^2, \\ -2x & \text{при } x > 0, x^2 < y. \end{cases}$$

- а) Покажите, что  $f(x, y)$  непрерывна, но не удовлетворяет условию Липшица.
- б) Покажите, что для задачи  $y' = f(x, y)$ ,  $y(0) = 0$  метод последовательных приближений не сходится.
- в) Покажите, что существует единственное решение и метод ломаных Эйлера сходится.

6. Используйте метод последовательных приближений Пикара для доказательства следующего утверждения:  
если  $f(x, y)$  непрерывна и удовлетворяет условию Липшица (7.7) в бесконечной полосе  $D = \{(x, y); x_0 \leq x \leq X\}$ , то задача Коши  $y' = f(x, y)$ ,  $y(x_0) = y_0$  имеет единственное решение при  $x_0 \leq x \leq X$ .

Сравните этот общий результат с теоремой 7.3.

7. Определим функцию  $y(x)$  (обратную «интегралу ошибок») соотношением

$$x = \frac{2}{V\pi} \int_0^y e^{-t^2} dt.$$

Покажите, что она удовлетворяет дифференциальному уравнению

$$y' = \frac{\sqrt{\pi}}{2} e^{y^2}, \quad y(0) = 0.$$

Получите рекуррентные формулы для ее коэффициентов ряда Тейлора.

## I.9. Теория существования решения для систем уравнений

Первое обращение к теории существования решений дифференциальных уравнений, рассматриваемых совместно, содержится на последних сохранившихся страницах работы Коши (1824). Используя индексные обозначения, которые еще не были приняты в те времена, мы запишем уравнения в виде

$$\begin{aligned} y'_1 &= f_1(x, y_1, \dots, y_n), \quad y_1(x_0) = y_{10}, \quad y_1(X) = ? \\ &\quad \ddots \\ y'_n &= f_n(x, y_1, \dots, y_n), \quad y_n(x_0) = y_{n0}, \quad y_n(X) = ? \end{aligned} \tag{9.1}$$

и поставим вопрос о существовании  $n$  решений  $y_1(x), \dots, y_n(x)$ . Поскольку любым заданным функциям  $y_1, \dots, y_n$  уравнения (9.1) ставят в соответствие производные этих функций, естественно снова рассмотреть по аналогии с (7.3) метод Эйлера

$$\begin{aligned} y_{k, i+1} &= y_{ki} + (x_{i+1} - x_i) \cdot f_k(x_i, y_{1i}, \dots, y_{ni}) \\ k &= 1, \dots, n, \quad i = 0, 1, 2, \dots. \end{aligned} \tag{9.2}$$

Здесь подразумевается, что  $y_{ki}$  аппроксимируют  $y_k(x_i)$ , где  $x_0 < x_1 < x_2 < \dots$  — разбиение отрезка интегрирования, такое, как в (7.2).

Теперь попытаемся перенести все, что мы сделали в разд. I.7, на новую ситуацию. Хотя при обобщении оценки (7.4) не возникает проблем, и мы получаем

$$|y_{ki} - y_{k0}| \leq A_k |x_i - x_0|, \text{ если } |f_k(x, y_1, \dots, y_n)| \leq A_k, \tag{9.3}$$

но положение несколько осложняется для (7.7). Нам надо оценить

$$\begin{aligned} f_k(x, z_1, \dots, z_n) - f_k(x, y_1, \dots, y_n) &= \\ &= \frac{\partial f_k}{\partial y_1} \cdot (z_1 - y_1) + \dots + \frac{\partial f_k}{\partial y_n} (z_n - y_n), \end{aligned} \tag{9.4}$$

где производные  $\partial f_k / \partial y_i$  взяты в подходящих промежуточных точках. Здесь Коши использовал неравенство, которое он открыл незадолго до этого и которое теперь называют неравенством Коши—Шварца<sup>1</sup>. («И наконец, из формулы (13) 11-го урока по дифференциальному исчислению следует ...».) С его помощью

<sup>1</sup>) Или неравенством Коши—Буняковского. — Прим. ред.

получается соотношение

$$|f_k(x, z_1, \dots, z_n) - f_k(x, y_1, \dots, y_n)| \leqslant \quad (9.5)$$

$$\leqslant \left( \left( \frac{\partial f_k}{\partial y_1} \right)^2 + \dots + \left( \frac{\partial f_k}{\partial y_n} \right)^2 \right)^{1/2} ((z_1 - y_1)^2 + \dots + (z_n - y_n)^2)^{1/2}.$$

На этом этапе уже чувствуется, что для дальнейшего развития теории целесообразно ввести векторные обозначения.

### Векторные обозначения

В интересующей нас области эти обозначения были введены в работах Пеано (1888) и (1890). Как он сам говорил, на него оказали влияние знаменитый трактат Грассмана «Учение о протяженности» и исследования Гамильтона, Кели и Сильвестра. Введем векторы (Пеано называл их «комплексами»)

$$y = (y_1, \dots, y_n)^T, \quad y_i = (y_{1i}, \dots, y_{ni})^T, \quad z = (z_1, \dots, z_n)^T \text{ и т. д.}$$

При этом мы надеемся, что читатель не будет путать компоненты  $y_i$  вектора  $y$  с векторами, помеченными индексом. Рассмотрим «векторную функцию»  $f(x, y) = (f_1(x, y), \dots, f_n(x, y))^T$ ; тогда уравнения (9.1) примут вид

$$y' = f(x, y), \quad y(x_0) = y_0, \quad y(X) = ?, \quad (9.1')$$

а метод Эйлера (9.2) будет задан формулой

$$y_{i+1} = y_i + (x_{i+1} - x_i) f(x_i, y_i), \quad i = 0, 1, 2 \text{ и т. д.} \quad (9.2')$$

Теперь обозначения ничем не отличаются от одномерных случаев (7.1) и (7.3).

Принимая во внимание оценку (9.5), введем для вектора  $y = (y_1, \dots, y_n)^T$  норму (в оригинале «модуль»)

$$\|y\| = (y_1^2 + \dots + y_n^2)^{1/2}, \quad (9.6)$$

которая удовлетворяет всем обычным свойствам нормы, в частности неравенству треугольника

$$\|y + z\| \leqslant \|y\| + \|z\|, \quad \left\| \sum_{i=1}^m y_i \right\| \leqslant \sum_{i=1}^m \|y_i\|. \quad (9.7)$$

Теперь мы можем записать оценки (9.3) и (9.5) вполне аналогично (7.4) и (7.7).

**Теорема 9.1.** Предположим, что  $\|f(x, y)\| \leqslant A$ ; тогда для определенного с помощью (9.2') вектора  $y_i$  выполняется оценка

$$\|y_i - y_0\| \leqslant A |x_i - x_0|. \quad (9.8)$$

Если  $\left| \frac{\partial f_k}{\partial y_t} \right| \leq q_{kt}$ , то

$$\|f(x, z) - f(x, y)\| \leq L \|z - y\|, \quad (9.9)$$

$$\text{где } L = \left( \sum_{t,k} q_{kt}^2 \right)^{1/2}. \quad (9.10)$$

*Доказательство.* Оценка (9.9) получается в результате возвведения в квадрат неравенства (9.5) и суммирования по  $k = 1, \dots, n$ . ■

Оценка (9.9) называется «теоремой о среднем»<sup>1)</sup>. Ее можно найти во многих книгах по математическому анализу.

Евклидова норма (9.6) — это не единственная возможная норма, мы будем также использовать норму

$$\|y\| = \max(|y_1|, \dots, |y_n|). \quad (9.6')$$

(«Можно было бы также определить  $mx$  — наибольшее из абсолютных значений элементов  $x$ ; для него свойства модулей почти очевидны» (Пeano).) В этом случае, чтобы установить соотношение (9.9), мы оценим (9.4) следующим образом:

$$|f_k(x, z) - f_k(x, y)| \leq \left( \left| \frac{\partial f_k}{\partial y_1} \right| + \dots + \left| \frac{\partial f_k}{\partial y_n} \right| \right) \cdot \max_t (|z_t - y_t|).$$

Поэтому теорема 9.1 остается справедливой, если заменить (9.10) выражением

$$L = \max_k \left( \sum_t |q_{kt}| \right). \quad (9.10')$$

И наконец, если мы хотим использовать норму

$$\|y\| = |y_1| + \dots + |y_n|, \quad (9.6'')$$

то в теореме 9.1 вместо (9.10) надо подставить

$$L = \max_t \left( \sum_k |q_{kt}| \right). \quad (9.10'')$$

Во всех трех случаях мы установили полную аналогию с (7.5) и (7.7) из разд. I.7. Остается только еще раз перечитать разд. I.7 и I.8: теоремы 7.2, 7.3, 7.4, 7.5 и 7.6 вместе с доказательствами и оценки (7.10), (7.13), (7.15), (7.16), (7.17) и (7.18) переносятся на более общий случай с единственным изменением — вместо абсолютных величин надо подставить нормы.

<sup>1)</sup> В русскоязычной математической литературе этот термин используется в ином смысле, а соотношение (9.9) называют условием Липшица. — Прим. ред.

### Подчиненные матричные нормы

Обозначим через  $Q$  матрицу с элементами  $q_{ki}$ . Определенная в (9.10') константа, которую мы теперь обозначим через

$$\|Q\| = \max_k \left( \sum_i |q_{ki}| \right), \quad (9.11')$$

называется *матричной нормой, подчиненной векторной норме* (9.6'). Это наименьшее число, при котором для всех векторов  $y$  справедлива оценка

$$\|Qy\| \leq \|Q\| \cdot \|y\|, \quad (9.12)$$

или, другими словами,

$$\|Q\| = \sup_{y \neq 0} \frac{\|Qy\|}{\|y\|}. \quad (9.13)$$

Аналогично в (9.10'') входит матричная норма, подчиненная норме (9.6''), т. е.

$$\|Q\| = \max_i \left( \sum_k |q_{ki}| \right). \quad (9.11'')$$

Однако выражение (9.10) (которое называется «нормой Шура») не подчинено норме (9.6), так как оно, хотя и удовлетворяет (9.12), но не является наименьшей возможной оценкой, т. е. точной верхней гранью. В этом случае подчиненная норма задается формулой

$$\|Q\| = \sqrt{\lambda_{\max}}, \quad (9.11)$$

где  $\lambda_{\max}$  — максимальное собственное значение произведения матриц  $Q^T Q$ . Это можно усмотреть из равенства

$$\|Qy\|^2 = y^T Q^T Q y,$$

так как  $Q^T Q$  всегда можно привести к диагональному виду с помощью ортогонального преобразования.

Все указанные формулы остаются справедливыми и для *комплексных матриц*. Надо только заменить  $Q^T$  на  $Q^*$  (транспонированная и комплексно сопряженная матрица). Об этом можно прочитать, например, в книгах Уилкинсона (1965) или Бахвалова (1973), гл. VI, § 3.

### Последовательные приближения Пикара для систем

Метод последовательных приближений Пикара также распространяется на системы, если в (8.7)  $y_{i+1}(x)$ ,  $y_0$  и  $f(s, y_i(s))$  интерпретировать как векторы, интегрируемые покомпонентно. Результат о сходимости с оценкой (8.10) тоже остается в силе; для

его доказательства надо использовать после (8.8) и перед (8.9) неравенство

$$\left\| \int_a^b g(x) dx \right\| \leq \int_a^b \|g(x)\| dx \quad (9.14)$$

для векторной функции  $g(x) = (g_1(x), \dots, g_n(x))^T$ . Соотношение (9.14) доказывается применением неравенства треугольника (9.7) к частичным суммам Римана, определяющим оба интеграла.

Метод рядов Тейлора, доказательство сходимости и рекуррентное вычисление коэффициентов ряда Тейлора также непосредственно обобщаются на системы уравнений.

## Упражнения

### 1. Решите систему

$$\begin{aligned} y'_1 &= -y_2, & y_1(0) &= 1, \\ y'_2 &= +y_1, & y_2(0) &= 0 \end{aligned}$$

методами Эйлера и Пикара, установите точные оценки погрешности во всех трех упоминавшихся нормах. Проверьте результаты, используя точное решение  $y_1(x) = \cos x$ ,  $y_2(x) = \sin x$ .

### 2. Рассмотрим дифференциальные уравнения

$$\begin{aligned} y'_1 &= -100y_1 + y_2, & y_1(0) &= 1, & y_1(1) &=? \\ y'_2 &= y_1 - 100y_2, & y_2(0) &= 0, & y_2(1) &=? \end{aligned}$$

- a) Найдите точное решение  $y(x)$  с помощью метода, изложенного в разд. I.6.
- b) С помощью (7.10) найдите оценку погрешности для  $\|z(x) - y(x)\|$ , где  $z(x) = 0$ .
- c) Примените к этому уравнению метод Эйлера с шагом  $h = 1/10$ .
- d) Примените метод последовательных приближений Пикара.

### 3. Вычислите решение в виде ряда Тейлора для системы с постоянными коэффициентами $y' = Ay$ , $y(0) = y_0$ . Докажите, что этот ряд сходится при всех $x$ . Используйте его для уравнения из упр. 1.

*Ответ:*

$$y(x) = \sum_{l=0}^{\infty} \frac{x^l}{l!} A^l y_0 = :e^{Ax} y_0.$$

## I.10. Дифференциальные неравенства

Дифференциальные неравенства — это изящный инструмент, который позволяет совершенно по-новому взглянуть на уравнения (7.10), (7.17) и лучше понять их. И в этой области начало исследования положила работа Пеано (1890), а затем они были продолжены Перроном (1915), Мюллером (1926), Камке (1930). Классическим трудом по этому предмету является книга Вальтера (1970).

### Введение

Основная идея состоит в следующем: обозначим через  $v(x)$  ломаную Эйлера, определенную в (7.3) или в (9.2); тогда

$$v'(x) = f(x_i, y_i) \quad \text{при } x_i < x < x_{i+1}. \quad (10.1)$$

В произвольно выбранной норме исследуем *погрешность*

$$m(x) = \|v(x) - y(x)\| \quad (10.2)$$

как функцию от  $x$  и, естественно, попытаемся оценить ее рост.

К сожалению,  $m(x)$  не обязательно будет дифференцируемой из-за угловых точек, которые, во-первых, содержатся в ломаной Эйлера, а во-вторых, порождены нормами, особенно (9.6') и (9.6''). Поэтому рассмотрим так называемые *производные Дини*, определенные следующим образом:

$$D^+m(x) = \limsup_{h>0} \frac{m(x+h) - m(x)}{h},$$

$$D^-m(x) = \limsup_{h<0} \frac{m(x+h) - m(x)}{h},$$

$$D_+m(x) = \liminf_{h>0} \frac{m(x+h) - m(x)}{h},$$

$$D_-m(x) = \liminf_{h<0} \frac{m(x+h) - m(x)}{h};$$

см., например, работы Шеффера (1884), Хобсона (1921), гл. V, § 260, § 280. Свойства

$$\begin{aligned} \|w(x+h)\| - \|w(x)\| &\leq \|w(x+h) - w(x)\|, \\ \|w(x)\| - \|w(x+h)\| &\leq \|w(x+h) - w(x)\| \end{aligned} \quad (10.3)$$

являются простым следствием неравенства треугольника (9.7). Разделив (10.3) на  $h > 0$ , мы получим оценки

$$D_+ \|w(x)\| \leq \|w'(x)\|, \quad D^+ \|w(x)\| \leq \|w'(x)\|, \quad (10.4)$$

где  $w'(x)$  — правая производная вектор-функции  $w(x)$ . Если мы применим их к  $m(x)$  из (10.2), то придем к соотношению

$$D_+ m(x) \leq \|v'(x) - y'(x)\| = \|v'(x) - f(x, v(x)) + f(x, v(x)) - f(x, g(x))\|,$$

из которого, применяя неравенство треугольника и условие Липшица (9.9), получим, что

$$D_+ m(x) \leq \delta(x) + Lm(x). \quad (10.5)$$

Здесь мы ввели функцию

$$\delta(x) = \|v'(x) - f(x, v(x))\|, \quad (10.6)$$

которая называется *дефектом*<sup>1)</sup> приближенного решения  $v(x)$ . Эта фундаментальная величина служит мерой того, насколько функция  $v(x)$  не удовлетворяет исходному дифференциальному уравнению. Неравенство (7.11) совместно с (10.1) указывает на то, что  $\delta(x) \leq \varepsilon$ , а поэтому мы можем продолжить оценку (10.5):

$$D_+ m(x) \leq Lm(x) + \varepsilon, \quad m(x_0) = 0. \quad (10.7)$$

Соотношение (10.7) [или (10.5)] называется *дифференциальным неравенством*. Вопрос заключается в том, допускается ли замена знака « $\leq$ » на « $=$ », т. е. можно ли вместо (10.7) решить уравнение

$$u' = Lu + \varepsilon, \quad u(x_0) = 0, \quad (10.8)$$

и сделать вывод, что  $m(x) \leq u(x)$ ? С учетом формул из разд. I.3 или I.5 это означало бы, что

$$m(x) \leq \frac{\varepsilon}{L} (e^{L(x-x_0)} - 1). \quad (10.9)$$

Таким образом, мы получили бы оценку погрешности (7.17) естественным способом и, кроме того, открыли бы изящное и мощное средство для новых оценок различного рода.

## Фундаментальные теоремы

Общая теорема типа

$$\left. \begin{array}{l} D_+ m(x) \leq g(x, m(x)) \\ D_+ u(x) \geq g(x, u(x)) \\ m(x_0) \leq u(x_0) \end{array} \right\} \Rightarrow \begin{array}{l} m(x) \leq u(x) \\ \text{при } x_0 \leq x \end{array} \quad (10.10)$$

<sup>1)</sup> В отечественной литературе  $\delta(x)$  принято называть *нормой невязки* исходного дифференциального уравнения (9.1') для функции  $v(x)$ . — Прим. ред.

не может быть верна. Контрпримером может служить любое дифференциальное уравнение с неединственным решением, скажем  $y' = g(x, y) = \sqrt{y}$ , для которого

$$m(x) = \frac{x^2}{4}, \quad u(x) = 0. \quad (10.11)$$

Пеано и Перрон сделали важное наблюдение, позволяющее нам преодолеть эту трудность. Оно заключается в том, что одно из первых двух неравенств надо сделать *строгим* (см. работу Пеано (1890), § 3, лемма 1).

**Теорема 10.1.** *Предположим, что функции  $m(x)$  и  $u(x)$  непрерывны и при  $x_0 \leq x \leq X$  удовлетворяют условиям*

- a)  $D_+m(x) \leq g(x, m(x))$ ,
- b)  $D_+u(x) > g(x, u(x))$ ,
- c)  $m(x_0) \leq u(x_0)$ .

Тогда

$$m(x) \leq u(x) \text{ при } x_0 \leq x \leq X. \quad (10.13)$$

Это утверждение справедливо и в том случае, если в а) и б)  $D_+$  заменить на  $D^+$ .

**Доказательство.** Чтобы можно было сравнивать производные  $D_+m$  и  $D_+u$  в (10.12), рассмотрим точки, в которых  $m(x) = u(x)$ . В этом состоит основная идея доказательства.

Если бы (10.13) не выполнялось, мы могли бы выбрать точку  $x_2$ , в которой  $m(x_2) > u(x_2)$ , и искать первую слева от  $x_2$  точку  $x_1$ , такую что  $m(x_1) = u(x_1)$ . Тогда при малых  $h > 0$  будет выполняться неравенство

$$\frac{m(x_1 + h) - m(x_1)}{h} > \frac{u(x_1 + h) - u(x_1)}{h}.$$

Переходя в нем к пределу, получим соотношение  $D_+m(x_1) \geq D_+u(x_1)$ . Однако оно противоречит условиям а) и б), в силу которых

$$\begin{aligned} D_+m(x_1) &\leq g(x_1, m(x_1)) = \\ &= g(x_1, u(x_1)) < D_+u(x_1). \end{aligned} \quad \blacksquare$$

Эту теорему можно сформулировать многими способами, например с использованием левых производных Дини [Вальтер (1970), гл. II, § 8, теорема V]. Опираясь на рассуждения из введения, мы придем к следующему результату.

**Теорема 10.2.** *(«Фундаментальная лемма».) Предположим, что  $v(x)$  — приближенное решение системы дифференциальных урав-*

нений  $y' = f(x, y)$ ,  $y(x_0) = y_0$ , которое удовлетворяет следующим условиям:

- $\|v(x_0) - y(x_0)\| \leq \rho$ ,
- $\|v'(x) - f(x, v(x))\| \leq \varepsilon$ ,
- $\|f(x, v) - f(x, y)\| \leq L \|v - y\|$ .

Тогда при  $x \geq x_0$  справедлива оценка погрешности

$$\|y(x) - v(x)\| \leq \rho e^{L(x-x_0)} + \frac{\varepsilon}{L} (e^{L(x-x_0)} - 1). \quad (10.14)$$

*Замечание.* Два члена в (10.14) выражают влияние на погрешность приближенного решения погрешности  $\rho$  в начальных значениях и невязки  $\varepsilon$  соответственно. Это соотношение означает непрерывную зависимость погрешности от обеих величин и единственность решения, поскольку  $y(x) = v(x)$  при  $\rho = \varepsilon = 0$ .

*Доказательство.* Положив  $m(x) = \|y(x) - v(x)\|$ , как и в (10.7), получим

$$D_+ m(x) \leq Lm(x) + \varepsilon, \quad m(x_0) \leq \rho.$$

Попытаемся сравнить это неравенство с дифференциальным уравнением

$$u' = Lu + \varepsilon, \quad u(x_0) = \rho. \quad (10.15)$$

Теорему 10.1 непосредственно применить к этому случаю нельзя. Поэтому в (10.15) заменим  $\varepsilon$  на  $\varepsilon + \eta$ ,  $\eta > 0$  и решим другое уравнение:

$$u' = Lu + \varepsilon + \eta > Lu + \varepsilon, \quad u(x_0) = \rho.$$

Теперь теорема 10.1 дает оценку (10.14), где вместо  $\varepsilon$  стоит  $\varepsilon + \eta$ . Поскольку эта оценка справедлива при всех  $\eta > 0$ , она выполняется и для  $\eta = 0$ . ■

*Другая формулировка теоремы 10.2. Из условий*

- $\|v(x_0) - y(x_0)\| \leq \rho$ ,
- $\|v'(x) - f(x, v(x))\| \leq \delta(x)$ ,
- $\|f(x, v) - f(x, y)\| \leq l(x) \|v - y\|$

следует, что при  $x \geq x_0$  имеет место оценка

$$\|y(x) - v(x)\| \leq e^{L(x)} \left[ \rho + \int_{x_0}^x e^{-L(s)} \delta(s) ds \right], \quad L(x) = \int_{x_0}^x l(s) ds.$$

*Доказательство.* Достаточно применить формулы (3.3). ■

**Теорема 10.3.** Если функция  $g(x, y)$  непрерывна и удовлетворяет условию Липшица, то импликация (10.10) справедлива для непрерывных функций  $t(x)$  и  $u(x)$ .

*Доказательство.* Определим функции  $w_n(x)$ ,  $v_n(x)$  как решения начальных задач

$$w'_n(x) = g(x, w_n(x)) + \frac{1}{n}, \quad w_n(x_0) = m(x_0),$$

$$v'_n(x) = g(x, v_n(x)) - \frac{1}{n}, \quad v_n(x_0) = u(x_0).$$

Тогда по теореме 10.1

$$m(x) \leq w_n(x), \quad v_n(x) \leq u(x) \text{ при } x_0 \leq x \leq X. \quad (10.16)$$

Из теоремы 10.2 следует, что функции  $w_n(x)$  и  $v_n(x)$  сходятся при  $n \rightarrow \infty$  к решениям уравнений

$$w'(x) = g(x, w(x)), \quad w(x_0) = m(x_0),$$

$$v'(x) = g(x, v(x)), \quad v(x_0) = u(x_0),$$

так как дефект равен  $\pm 1/n$ . И наконец, вследствие неравенства  $m(x_0) \leq u(x_0)$  и единственности решений мы придем к соотношению  $w(x) \leq v(x)$ . Переходя в (10.16) к пределу при  $n \rightarrow \infty$ , мы получим, что  $m(x) \leq u(x)$ . ■

Можно еще больше обобщить теорему 10.2, если заменить условие Липшица b) некоторым нелинейным выражением типа

$$\|f(x, v) - f(x, y)\| < \omega(x, \|v - y\|).$$

Тогда дифференциальное неравенство для погрешности  $m(x)$  надо сравнивать с решением уравнения

$$u' = \omega(x, u) + \delta(x) + \eta, \quad u(x_0) = \rho, \quad \eta \gg 0.$$

Более подробно этот вопрос изложен в работе Вальтера (1970), гл. II, § 11.

### Оценки с использованием односторонних условий Липшица

В упр. 2 из разд. I.9 мы уже отметили давно известный факт, что при использовании в оценках (7.16), (7.17) или (7.18) положительных констант Липшица  $L$  (например (9.10), (9.10'), (9.10'')) можно упустить много информации о погрешностях. Все оценки растут экспоненциально по  $x$ , даже если решения и погрешности затухают. Поэтому было предпринято много усилий для улучшения оценок погрешности, как, например, в статьях Эльтерманна (1955), Ульманна (1957), Далквиста (1959) и в работах, на которые ссылаются эти авторы. Мы с большим удовольствием будем следовать особо ясному изложению этого вопроса в работе Далквиста.

Оценим производную функции  $m(x) = \|v(x) - y(x)\|$  более тщательно по сравнению с (10.5): при  $h > 0$  с использованием (10.6) и (9.7) мы получим

$$\begin{aligned} m(x+h) &= \|v(x+h) - y(x+h)\| \\ &= \|v(x) - y(x) + h(v'(x) - y'(x))\| + O(h^2) \\ &\leq \|v(x) - y(x) + h(f(x, v(x)) - f(x, y(x)))\| \\ &\quad + h\delta(x) + O(h^2). \end{aligned} \quad (10.17)$$

Теперь, применив к функции  $y + hf(x, y)$  формулу Коши для конечных приращений, придем к неравенству

$$m(x+h) \leq \left( \max_{\eta \in [y(x), v(x)]} \left\| I + h \frac{\partial f}{\partial y}(x, \eta) \right\| \right) m(x) + h\delta(x) + O(h^2),$$

из которого окончательно получим для  $h > 0$

$$\begin{aligned} \frac{m(x+h) - m(x)}{h} &\leq \max_{\eta \in [y(x), v(x)]} \frac{\left\| I + h \frac{\partial f}{\partial y}(x, \eta) \right\| - 1}{h} m(x) + \\ &\quad + \delta(x) + O(h). \end{aligned} \quad (10.18)$$

Выражение в правой части (10.18) приводит нас к следующему определению.

**Определение 10.4.** Пусть  $Q$  является квадратной матрицей. Назовем выражение

$$\mu(Q) = \lim_{h \rightarrow 0, h > 0} \frac{\|I + hQ\| - 1}{h} \quad (10.19)$$

логарифмической нормой  $Q$ .

Приведем далее формулы для ее вычисления [Далквист (1959), Эльтерманн (1955)].

**Теорема 10.5.** Логарифмическая норма (10.19) получается по следующим формулам: для евклидовой нормы (9.6)

$\mu(Q) = \lambda_{\max}$  = наибольшее собственное значение матрицы

$$\frac{1}{2}(Q^T + Q). \quad (10.20)$$

Для так-нормы (9.6')

$$\mu(Q) = \max_k \left( q_{kk} + \sum_{i \neq k} |q_{ki}| \right). \quad (10.20')$$

Для нормы (9.6'')

$$\mu(Q) = \max_i \left( q_{ii} + \sum_{k \neq i} |q_{ki}| \right). \quad (10.20'')$$

*Доказательство.* Соотношения (10.20') и (10.20'') являются почти тривиальным следствием формул (9.11'), (9.11'') и определения (10.19). Дело в том, что присутствие единичной матрицы  $I$  подавляет при достаточно малых  $h$  абсолютные величины диагональных элементов. (10.20) следует из того факта, что собственные значения матрицы

$$(I + hQ)^T (I + hQ) = I + h(Q^T + Q) + h^2 Q^T Q$$

при  $h \rightarrow 0$  сходятся к  $1 + h\lambda_i$ , где  $\lambda_i$  — собственные значения матрицы  $Q^T + Q$ . ■

*Замечание.* Для комплекснозначных матриц приведенные выше формулы принимают вид:

$\mu(Q)$  — наибольшее собственное значение матрицы  $\frac{1}{2}(Q^* + Q)$ ,

$$\mu(Q) = \max_k \left( \operatorname{Re} q_{kk} + \sum_{i \neq k} |q_{ki}| \right),$$

$$\mu(Q) = \max_t \left( \operatorname{Re} q_{tt} + \sum_{k \neq t} |q_{kt}| \right).$$

С помощью неравенства (10.18) можно весьма существенно усилить теорему 10.2.

**Теорема 10.6.** Предположим, что справедливы оценки

$$B \left( \frac{\partial f}{\partial y}(x, \eta) \right) \leq l(x) \text{ при } \eta \in [y(x), v(x)]$$

и

$$\|v'(x) - f(x, v(x))\| \leq \delta(x), \quad \|v(x_0) - y(x_0)\| \leq \rho. \quad (10.21)$$

Тогда при  $x > x_0$  выполняется соотношение

$$\|y(x) - v(x)\| \leq e^{L(x)} \left( \rho + \int_{x_0}^x e^{-L(s)} \delta(s) ds \right), \quad (10.22)$$

$$\text{где } L(x) = \int_{x_0}^x l(s) ds.$$

*Доказательство.* Поскольку при фиксированном  $x$  отрезок  $[v(x), y(x)]$  является компактным, величина

$$K = \max_t \max_{[v(x), y(x)]} \left| \frac{\partial f_t}{\partial y_t} \right|$$

конечна. Тогда (см. доказательство теоремы 10.5) выполняется соотношение

$$\underbrace{\left\| I + h \frac{\partial f}{\partial y}(x, \eta) \right\| - 1}_h = B \left( \frac{\partial f}{\partial y}(x, \eta) \right) + Q(h),$$

где член  $O(h)$  равномерно ограничен по  $\eta$ . (Для норм (9.6') и (9.6'') этот член в действительности равен нулю при  $h < 1/K$ .) Таким образом, подстановка условия (10.21) в (10.18) дает неравенство

$$D_+ m(x) \leq l(x)m(x) + \delta(x).$$

Теперь оценка (10.22) получается тем же способом, что и в теореме 10.2.

### Упражнения

1. Примените теорему 10.5 к примеру из упр. 2 в разд. I.9. Обратите внимание на значительное улучшение оценок.
2. Докажите следующее утверждение (одна из формулировок знаменитой «леммы Гронуолла», Гронуолл (1919)): предположим, что положительная функция  $m(x)$  удовлетворяет неравенству

$$m(x) \leq \rho + \varepsilon(x - x_0) + L \int_{x_0}^x m(s) ds = : w(x). \quad (10.23)$$

Тогда имеет место оценка

$$m(x) \leq \rho e^{L(x-x_0)} + \frac{\varepsilon}{L}(e^{L(x-x_0)} - 1). \quad (10.24)$$

Примените при доказательстве:

- а) прямой способ, вычитая из (10.23) равенство

$$u(x) = \rho + \varepsilon(x - x_0) + L \int_{x_0}^x u(s) ds;$$

б) дифференцирование  $w(x)$  в (10.23) и теорему 10.1.

с) Докажите теорему 10.2 с помощью приведенной выше леммы Гронуолла. Такие же взаимосвязи имеются, конечно, и в более общих случаях.

3. Воспользовавшись примером

$$y'_1 = y_2, \quad y'_2 = -y_1,$$

установите, что распространение теоремы 10.1 на *системы* дифференциальных уравнений первого порядка при покомпонентной интерпретации неравенств в общем случае неверно [Мюллер (1926)]. Однако такого рода обобщение теоремы 10.1 можно доказать при дополнительном предположении о *квазимонотонности* функций  $g_i(x, y_1, \dots, y_n)$ , т. е. при условии

$$g_i(x, y_1, \dots, y_j, \dots, y_n) \leq g_i(x, y_1, \dots, z_j, \dots, y_n),$$

если  $y_j < z_j$  при всех  $j \neq i$ .

Попытайтесь доказать это.

Важно отметить, что многие системы, полученные из параболических дифференциальных уравнений, вроде (6.10), *действительно* квазимонотонны. Поэтому изложенные в данном разделе идеи допускают много интересных применений [см. работу Вальтера (1970), гл. IV].

4. Докажите следующие свойства логарифмической нормы:
  - а)  $\mu(\alpha Q) = \alpha \mu(Q)$  при  $\alpha \geq 0$ ,
  - б)  $-\|Q\| \leq \mu(Q) \leq \|Q\|$ ,
  - в)  $\mu(Q + P) \leq \mu(Q) + \mu(P)$ ,  $\mu\left(\int Q(t) dt\right) \leq \int \mu(Q(t)) dt$ ,
  - г)  $|\mu(Q) - \mu(P)| \leq \|Q - P\|$ .
5. Для евклидовой нормы  $\mu(Q)$  (10.20) является наименьшим числом, удовлетворяющим неравенству

$$\langle v, Qv \rangle \leq \mu(Q) \|v\|^2.$$

Данное свойство справедливо для всех норм, порожденных скалярным произведением. Докажите это.

6. Покажите, что для евклидовой нормы условие (10.21) эквивалентно неравенству

$$\langle y - z, f(x, y) - f(x, z) \rangle \leq l(x) \|y - z\|^2.$$

## I.11. Системы линейных дифференциальных уравнений

Приобретенные знания о существовании и единственности решений и сведения из линейной алгебры помогут нам, как и математикам XIX века, лучше понять многие вопросы, касающиеся линейных дифференциальных уравнений высокого порядка, которые остались не вполне выясненными в разд. I.4 и I.6.

Уравнение (4.9) в результате деления на  $a_0(x)$  ( $a_0(x) \neq 0$  вне особых точек) принимает вид

$$y^{(n)} + a_1(x)y^{(n-1)} + \dots + a_n(x)y = f(x). \quad (11.1)$$

Вводя переменные  $y = y_1, y' = y_2, \dots, y^{(n-1)} = y_n$ , мы получим

$$\begin{pmatrix} y'_1 \\ y'_2 \\ \vdots \\ y'_n \end{pmatrix} = \begin{pmatrix} 0 & 1 & & & \\ 0 & 0 & & & \\ \vdots & \vdots & \ddots & & 1 \\ -a_n(x) & -a_{n-1}(x) & \dots & -a_1(x) & \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ f(x) \end{pmatrix}. \quad (11.1')$$

Как и в разд. I.9, обозначим через  $y$  вектор  $(y_1, \dots, y_n)^T$ , а через  $f(x)$  — неоднородную часть. Тогда (11.1') будет частным случаем *системы линейных дифференциальных уравнений общего вида*

$$y' = A(x)y + f(x), \quad (11.2)$$

$$A(x) = (a_{ij}(x)), \quad f(x) = (f_i(x)), \quad i, j = 1, \dots, n.$$

К ней можно непосредственно применить теоремы из разд. I.9 и I.10. Поскольку частные производные правой части (11.2) по  $y_i$  равны  $a_{ki}(x)$ , выполняется оценка Липшица (см. условие (с) из второй формулировки теоремы 10.2), где  $l(x) = \|A(x)\|$  в любой подчиненной матричной норме [(9.11), (9.11'), (9.11")]. Применим теорему 7.4 и вторую формулировку теоремы 10.2, где в качестве «приближенного решения» возьмем  $v(x) = 0$ . Мы можем также выбрать  $l(x) = \mu(A(x))$  [см. (10.22), (10.22'), (10.22'')] и использовать теорему 10.6.

**Теорема 11.1.** Пусть  $A(x)$  непрерывна на отрезке  $[x_0, X]$ . Тогда при любых начальных значениях  $y_0 = (y_{10}, \dots, y_{n0})^T$  существует единственное решение системы (11.2) для всех  $x_0 \leqslant$

$\leqslant x \leqslant X$ , и оно удовлетворяет оценке

$$\|y(x)\| \leqslant e^{L(x)} \left[ \|y_0\| + \int_{x_0}^x e^{-L(s)} \|f(s)\| ds \right], \quad (11.3)$$

$$L(x) = \int_{x_0}^x l(s) ds, \quad l(x) = \|A(x)\| \text{ или } l(x) = \mu(A(x)).$$

При  $f(x) \equiv 0$  решение  $y(x)$  линейно зависит от начальных значений, т. е. существует матрица  $R(x, x_0)$  («резольвента»), такая что

$$y(x) = R(x, x_0) y_0. \quad (11.4)$$

*Доказательство.* Поскольку  $l(x)$  непрерывна и поэтому ограничена на любом компактном отрезке  $[x_0, X]$ , оценка (11.3) показывает, что решение можно продолжить до конца отрезка. Линейная зависимость следует из того факта, что при  $f \equiv 0$  линейные комбинации решений снова будут решениями, и из единственности решения. ■

В силу единственности решения с начальными условиями  $y_0$  в  $x_0$  и  $y_1 = R(x_1, x_0) y_0$  в  $x_1$  должны совпадать. Поэтому при  $x_0 \leqslant x_1 \leqslant x_2$  справедливо равенство

$$R(x_2, x_0) = R(x_2, x_1) R(x_1, x_0). \quad (11.5)$$

И наконец, проинтегрировав назад от  $x_1$ ,  $y_1$ , т. е. выполнив преобразование координат  $x = x_1 - t$ ,  $0 \leqslant t \leqslant x_1 - x_0$ , мы должны прийти, снова вследствие единственности, к начальным значениям. Таким образом,

$$R(x_0, x_1) = (R(x_1, x_0))^{-1}, \quad (11.6)$$

так что (11.5) справедливо и в общем случае.

### Матрица Вронского (вронсиан)

Рассмотрим  $y_i(x) = (y_{1i}(x), \dots, y_{ni}(x))^T$  ( $i = 1, \dots, n$ ) — множество  $n$  решений однородного дифференциального уравнения

$$y' = A(x)y, \quad (11.7)$$

которые в точке  $x = x_0$  являются линейно независимыми (т. е. образуют фундаментальную систему). Составим матрицу Вронского [Вронский (1810)]

$$W(x) = \begin{pmatrix} y_{11}(x) & \dots & y_{1n}(x) \\ \vdots & & \vdots \\ y_{n1}(x) & \dots & y_{nn}(x) \end{pmatrix}.$$

Тогда

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

а любое решение можно записать в виде

$$c_1 y_1(x) + \cdots + c_n y_n(x) = W(x) \cdot c, \quad \text{где } c = (c_1, \dots, c_n)^T. \quad (11.8)$$

Если это решение должно удовлетворять начальным условиям  $y(x_0) = y_0$ , то мы получим  $c = W^{-1}(x_0) y_0$  и формулу

$$R(x, x_0) = W(x) W^{-1}(x_0). \quad (11.9)$$

Таким образом, если найдено  $n$  линейно независимых решений, то и все решения известны. Однако в общем случае при непостоянной матрице  $A(x)$  почти ничего не известно о том, как их построить.

### Тождество Абеля — Лиувилля — Якоби — Остроградского

Из (11.6) уже видно, что  $W(x)$  остается регулярным при всех  $x$ . Теперь покажем, что определитель  $W(x)$  можно задать явно следующим соотношением:

$$\det W(x) = \det(W(x_0)) \cdot \exp \left( \int_{x_0}^x \operatorname{Tr}(A(s)) ds \right),$$

$$\operatorname{Tr}(A(x)) = a_{11}(x) + a_{22}(x) + \dots + a_{nn}(x), \quad (11.10)$$

которое связывает определитель  $W(x)$  со следом матрицы  $A$  [Абель (1827), Лиувилль (1838), Якоби (1845), § 17].

Для доказательства (см. также упр. 2) мы найдем производную  $\frac{d}{dx} \det(W(x))$ . Поскольку  $\det(W(x))$  — полилинейная форма, ее производная (по правилу Лейбница) является суммой  $n$  членов, первый из которых равен

$$T_1 = \det \begin{pmatrix} y'_{11} & y'_{12} & \dots & y'_{1n} \\ y_{21} & y_{22} & & y_{2n} \\ \vdots & \vdots & & \vdots \\ y_{n1} & y_{n2} & \dots & y_{nn} \end{pmatrix}.$$

Подставим сюда  $y'_{it} = a_{1i} y_{1t} + \dots + a_{ni} y_{nt}$  в соответствии с (11.7). Члены  $a_{12} y_{2i}, \dots, a_{in} y_{ni}$  исчезают, так как они кратны строкам со 2-й по  $n$ -ю. Поэтому  $T_1 = a_{11} \det(W(x))$ . Сложив все  $T_i$ , мы окончательно получим формулу

$$\frac{d}{dx} \det(W(x)) = (a_{11}(x) + \dots + a_{nn}(x)) \cdot \det(W(x)), \quad (11.11)$$

проинтегрировав которую, придем к (11.10). ■

## Неоднородные линейные уравнения

Обобщив идеи Иог. Бернулли для (3.2) и Лагранжа для (4.9), мы теперь найдем решение неоднородного уравнения (11.2) посредством «вариации» постоянной  $c$ , т. е. сделав с «переменной» в «общем решении» (11.8) однородного уравнения  $y(x) = W(x) \times c(x)$  [Лиувилль (1838)]. Точно так же, как в разд. I.3 для (3.2), из (11.2) и (11.7) с помощью дифференцирования мы получим

$$y' = W'c + Wc' = AWc + Wc' = AWc + f.$$

Следовательно,  $c' = W^{-1}f$ . В результате интегрирования, взяв постоянные интегрирования  $c$ , мы придем к равенству

$$y(x) = W(x) \int_{x_0}^x W^{-1}(s) f(s) ds + W(x) \cdot c.$$

Из начальных условий  $y(x_0) = y_0$  определим  $c = W^{-1}(x_0)y_0$ . Итак, мы получили следующий результат.

**Теорема 11.2.** («Формула вариации постоянных».) Пусть  $A(x)$  и  $f(x)$  являются непрерывными. Тогда отвечающее начальным условиям  $y(x_0) = y_0$  решение неоднородного дифференциального уравнения  $y' = A(x)y + f(x)$  вычисляется по формуле

$$\begin{aligned} y(x) &= W(x) \left[ W^{-1}(x_0)y_0 + \int_{x_0}^x W^{-1}(s) f(s) ds \right] = \\ &= R(x, x_0)y_0 + \int_{x_0}^x R(x, s)f(s) ds. \end{aligned} \quad (11.12)$$

## Упражнения

1. Найдите резольвентную матрицу  $R(x, x_0)$  для систем

$$\begin{array}{ll} a) \quad y'_1 = y_1, & b) \quad y'_1 = y_2, \\ y'_2 = 3y_2, & y'_2 = -y_1 \end{array}$$

и проверьте справедливость соотношений (11.5), (11.6), а также (11.10).

2. Восстановите оригинальное доказательство Абеля для (11.10), которое относилось к случаю

$$y''_1 + py'_1 + qy_1 = 0, \quad y''_2 + py'_2 + qy_2 = 0;$$

для этого надо умножить уравнения на  $y_2$  и  $y_1$  соответственно и вычесть одно из другого, чтобы исключить  $q$ . Затем проин-

тегрируйте полученное равенство и используйте результат для вывода тождества, связывающего два интеграла,

$$y_1(a) = \int_0^\infty e^{ax-x^2} x^{\alpha-1} dx \quad \text{и} \quad y_2(a) = \int_0^\infty e^{-ax-x^2} x^{\alpha-1} dx,$$

каждый из которых удовлетворяет уравнению

$$\frac{d^2 y_t}{da^2} - \frac{a}{2} \frac{dy_t}{da} - \frac{\alpha}{2} y_t = 0. \quad (11.13)$$

*Указание.* Для проверки (11.13) проинтегрируйте от нуля до бесконечности выражение  $\frac{d}{dx} (\exp(ax - x^2) x^\alpha)$  [Абель (1827), случай IV].

3. [Куммер (1839)] Покажите, что общее решение уравнения

$$y^{(n)}(x) = x^m y(x) \quad (11.14)$$

можно получить в квадратурах.

*Указание.* Дифференцирование (11.14) дает

$$y^{(n+1)} = x^m y' + mx^{m-1} y. \quad (11.15)$$

Для нахождения рекуррентного соотношения предположим что общее решение уравнения

$$\psi^{(n+1)} = x^{m-1} \psi, \quad \text{т. е.} \quad \frac{d^{(n+1)}}{dx^{n+1}} \psi(xu) = x^{m-1} u^{m+n} \psi(xu), \quad (11.16)$$

уже известно. Покажите, что тогда функция

$$y(x) = \int_0^\infty u^{m-1} \exp\left(-\frac{u^{m+n}}{m+n}\right) \psi(xu) du$$

является общим решением уравнения (11.15), а при некоторых условиях на параметры — и (11.14). Для того чтобы упростить расчеты, рассмотрите функцию

$$g(u) = u^m \exp\left(-\frac{u^{m+n}}{m+n}\right) \psi(xu),$$

найдите производную по  $u$ , умножьте ее на  $x^{m-1}$  и проинтегрируйте от 0 до бесконечности.

4. (Слабые особенности в системах.) Покажите, что линейная система

$$y' = \frac{1}{x} (A_0 + A_1 x + A_2 x^2 + \dots) y \quad (11.17)$$

имеет решения вида

$$y(x) = x^p (v_0 + v_1 x + v_2 x^2 + \dots), \quad (11.18)$$

где  $v_0, v_1, \dots$  — векторы. Определите сначала  $p$  и  $v_0$ , а затем рекурсивно  $v_1, v_2$  и т. д. (Убедитесь, что существует  $n$  независимых решений вида (11.18), если собственные значения  $A_0$  удовлетворяют условию  $\lambda_i \neq \lambda_j \pmod{(\mathbb{Z})^{-1}}$  [Фукс (1866)].)

5. Найдите общее решение систем со слабой особенностью

$$y' = \frac{1}{x} \begin{pmatrix} \frac{3}{4} & 1 \\ \frac{1}{4} & -\frac{1}{4} \end{pmatrix} y \quad \text{и} \quad y' = \frac{1}{x} \begin{pmatrix} \frac{3}{4} & 1 \\ -\frac{1}{4} & -\frac{1}{4} \end{pmatrix} y. \quad (11.19)$$

*Указание.* Первую систему легко решить на основании упр. 4, а для второй требуются дополнительные приемы (см. формулу (5.6)). Возможен и другой способ: использовать преобразование  $x = e^t$ ,  $y(x) = z(t)$  и применить методы из разд. I.12.

---

<sup>1)</sup> Т. е. разность  $\lambda_i - \lambda_j$  не равна целому числу. — Прим. ред.

## I.12. Системы с постоянными коэффициентами

«Техника интегрирования линейных дифференциальных уравнений с постоянными коэффициентами развита здесь на самом высоком уровне.»

(Примечание Ф. Клейна в книге Рауса (1898))

### Линеаризация

Системы линейных дифференциальных уравнений с постоянными коэффициентами образуют класс уравнений, для которых резольвенту  $R(x, x_0)$  можно вычислить явно. Обычно они встречаются при линеаризации не зависящих от времени (т. е. *автономных*, или *перманентных*) нелинейных дифференциальных уравнений

$$y'_i = f_i(y_1, \dots, y_n) \text{ или } y''_i = f_i(y_1, \dots, y_n) \quad (12.1)$$

в окрестности стационарной точки [Лагранж (1788), см. также работы Рауса (1860), гл. IX, Томсона и Тейта (1879)]. Выберем координаты таким образом, чтобы эта стационарная точка попала в начало координат, т. е. выполнялось равенство  $f_i(0, \dots, 0) = 0$ . Затем разложим  $f_i$  в ряд Тейлора, пренебрегая всеми нелинейными членами:

$$y'_i = \sum_k \frac{\partial f_i}{\partial y_k}(0) y_k \text{ или } y''_i = \sum_k \frac{\partial^2 f_i}{\partial y_k^2}(0) y_k. \quad (12.1')$$

Эта система уравнений с постоянными коэффициентами, введенная в разд. I.6 (см. (6.3), (6.11)):

$$y' = Ay \text{ или } y'' = Ay. \quad (12.1'')$$

Автономные системы инвариантны относительно *сдвига*  $x \rightarrow x + C$ . Поэтому мы всегда можем предполагать, что  $x_0 = 0$ . Для произвольного  $x_0$  резольвента задается формулой

$$R(x, x_0) = R(x - x_0, 0). \quad (12.2)$$

### Приведение к диагональному виду

В разд. I.6 мы уже видели, что из предположения  $y(x) = ve^{\lambda x}$  следует равенство

$$Av = \lambda v \text{ или } Av = \lambda^2 v, \quad (12.3)$$

и поэтому  $v \neq 0$  должен быть *собственным вектором*  $A$ , а  $\lambda$  — соответствующим *собственным числом* (в первом случае, а во втором

ром случае, который мы больше не будем рассматривать, — квадратным корнем из собственного числа). Из (12.3) следует, что такой вектор  $v \neq 0$  существует тогда и только тогда, когда определитель, который называется *характеристическим многочленом матрицы A*, равен нулю:

$$\begin{aligned}\chi_A(\lambda) &:= \det(\lambda I - A) = \\ &= (\lambda - \lambda_1)(\lambda - \lambda_2) \dots (\lambda - \lambda_n) = 0.\end{aligned}\quad (12.4)$$

Теперь предположим, что для  $n$  собственных значений  $\lambda_i$  можно выбрать  $n$  линейно независимых собственных векторов  $v_i$ . Тогда из (12.3) мы получим равенство

$$A(v_1 v_2 \dots v_n) = (v_1 v_2 \dots v_n) \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_n),$$

или

$$T^{-1}AT = \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_n), \quad (12.5)$$

если через  $T$  обозначить матрицу, столбцами которой являются собственные векторы  $A$ . При сравнении (12.5) с (12.1'') мы видим, что дифференциальное уравнение значительно упрощается за счет использования преобразования координат

$$y(x) = Tz(x), \quad y'(x) = Tz'(x), \quad (12.6)$$

которое приводит систему к виду

$$z'(x) = \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) z(x). \quad (12.7)$$

Таким образом, исходная система дифференциальных уравнений распадается на  $n$  скалярных уравнений, для которых легко найти решение

$$z(x) = \operatorname{diag}(\exp(\lambda_1 x), \exp(\lambda_2 x), \dots, \exp(\lambda_n x)) z_0,$$

откуда с помощью (12.6) определяется

$$\begin{aligned}y(x) &= T \operatorname{diag}(\exp(\lambda_1 x), \exp(\lambda_2 x), \dots, \\ &\quad \exp(\lambda_n x)) \cdot T^{-1} y_0.\end{aligned}\quad (12.8)$$

### Разложение Шура

«Der Beweis ist leicht zu erbringen.»

«Это легко доказать.»

(Шур, 1909)

Предыдущая теория, какой бы красивой она ни казалась, имеет несколько недостатков:

- a) не все  $n \times n$ -матрицы имеют  $n$  линейно независимых собственных векторов;
- b) даже если исходная матрица обратима, матрица  $T$  может иметь очень плохое поведение (см. упр. 1).

Однако, как утверждает классическая теория, *симметричные* матрицы  $A$  всегда можно привести к диагональному виду с помощью *ортогональных* преобразований. Следуя работе Шура (1909), обобщим эту классическую теорию на несимметричные матрицы. Вещественная матрица  $Q$  называется *ортогональной*, если ее столбцы являются попарно ортогональными векторами с нормой 1, т. е. если  $Q^T Q = I$  или  $Q^T = Q^{-1}$ . Комплексная матрица  $Q$  называется *унитарной*, если  $Q^* Q = I$  или  $Q^* = Q^{-1}$ , где  $Q^*$  — сопряженная матрица для  $Q$ , т. е. транспонированная и комплексно сопряженная.

### Теорема 12.1.

а) (Шур) Для каждой комплексной матрицы  $A$  существует унитарная матрица  $Q$ , такая что

$$Q^* A Q = \begin{pmatrix} \lambda_1 & \times & \times & \dots & \times \\ \lambda_2 & \times & \dots & & \times \\ \vdots & & & & \vdots \\ & & & & \lambda_n \end{pmatrix}; \quad (12.9)$$

б) [Уинтнер—Мурнаган (1931)] Для вещественной матрицы  $A$  можно выбрать вещественную и ортогональную матрицу  $Q$ , если для каждой пары комплексно сопряженных собственных значений  $\lambda, \bar{\lambda} = \alpha \pm i\beta$  можно заменить блок

$$\begin{pmatrix} \lambda & \times \\ \bar{\lambda} & \end{pmatrix} \quad \text{на} \quad \begin{pmatrix} \times & \times \\ \times & \times \end{pmatrix}.$$

*Доказательство.* а) Матрица  $A$  имеет по крайней мере один собственный вектор с собственным значением  $\lambda_1$ . Используем этот нормированный вектор в качестве первого столба матрицы  $Q_1$ . Ее остальные столбцы выбираются произвольно среди векторов, дополняющих первый столбец до ортонормированного базиса. Тогда

$$AQ_1 = Q_1 \left( \begin{array}{c|cccc} \lambda_1 & x_2 & \dots & x_n \\ \hline 0 & \tilde{A} \end{array} \right). \quad (12.10)$$

Далее применим тот же прием к  $(n - 1)$ -мерной матрице  $\tilde{A}$  и придем к соотношению

$$\tilde{A} \tilde{Q}_2 = \tilde{Q}_2 \left( \begin{array}{c|cccc} \lambda_2 & \times & \dots & \times \\ \hline 0 & \times & \dots & \times \\ \vdots & & & \vdots \\ \times & & & \times \end{array} \right).$$

С помощью унитарной матрицы

$$Q_2 = \left( \begin{array}{c|c} 1 & 0 \\ \hline 0 & \tilde{Q}_2 \end{array} \right)$$

получим

$$Q_1^* A Q_1 Q_2 = Q_2 \left( \begin{array}{c|c} \lambda_1 & x & \times \dots \times \\ \lambda_2 & x & \dots \times \\ \hline & & \times \dots \times \\ & & \vdots \quad \vdots \\ & & \times \dots \times \end{array} \right)$$

Продолжение этого процесса в итоге приведет нас к треугольной матрице типа (12.9), причем  $Q = Q_1 Q_2 \dots Q_{n-1}$ .

б) Пусть  $A$  — вещественная матрица. Если  $\lambda_i$  — вещественное число, то в качестве  $Q_1$  можно выбрать вещественную и ортогональную матрицу. Теперь предположим, что  $\lambda_1 = \alpha + i\beta$  ( $\beta \neq 0$ ) — невещественное собственное число с соответствующим собственным вектором  $u + iv$ , т. е.

$$A(u \pm iv) = (\alpha \pm i\beta)(u \pm iv) \quad (12.11)$$

или

$$Au = \alpha u - \beta v, \quad Av = \beta u + \alpha v. \quad (12.11')$$

Поскольку  $\beta \neq 0$ ,  $u$  и  $v$  линейно независимы. Выберем ортогональный базис  $\hat{u}$ ,  $\hat{v}$  в подпространстве, натянутом на  $u$  и  $v$ , и возьмем  $\hat{u}$  и  $\hat{v}$  в качестве первых двух столбцов ортогональной матрицы  $Q_1$ . Тогда из (12.11') следует, что

$$AQ_1 = Q_1 \left( \begin{array}{c|c} \times \times & \times \dots \times \\ \times \times & \times \dots \times \\ \hline 0 & \tilde{A} \end{array} \right).$$

■

Сам Шур не придавал большого значения «своему» разложению, он просто вывел его как средство для доказательства интересных свойств собственных значений (см., например, упр. 2).

Ясно, что при вещественной и симметричной  $A$  матрица  $Q^T A Q$  тоже будет симметричной, а поэтому диагональной (см. также упр. 3).

### Численные расчеты

Приведенное выше теоретическое доказательство не слишком полезно практически: оно требует знания собственных значений, а их вычисление из характеристического многочлена является

одной из хорошо известных глупостей в численном анализе. При разумном применении численного анализа идут по другому пути: вещественную матрицу  $A$  сразу приводят сначала к форме Хессенберга, а затем с помощью последовательности ортогональных преобразований — вещественной форме Шура, предложенной Уинтнером — Мурнаганом (QR-алгоритм Фрэнсиса, реализованный в виде программы Мартином, Петерсон и Уилкинсоном, алгоритм II/14 в книге Уилкинсона—Райнша (1970)). Тогда собственные значения получаются без дополнительных усилий. Однако программа, названная «HQR2», не дает форму Шура для  $A$ , так как она продолжает работу до нахождения собственных векторов. Поэтому надо произвести некоторые действия, чтобы прервать программу в нужный момент (например, в версии HQR2 на Фортране из сборника Эйспак (1974) в операторе с меткой 60 надо заменить «340» на «1001»). Счастливые обладатели пакета «Matlab» просто вызывают процедуру SCHUR.

Когда форма Шура найдена, преобразование  $y(x) = Qz(x)$ ,  $y'(x) = Qz'(x)$  [см. (12.6)] приводит к системе

$$\begin{pmatrix} z'_1 \\ \vdots \\ z'_{n-1} \\ z'_n \end{pmatrix} = \begin{pmatrix} \lambda_1 b_{12} & \dots & b_{1,n-1} & b_{1n} \\ & \ddots & \ddots & \vdots \\ & & \lambda_{n-1} & b_{n-1,n} \\ & & & \lambda_n \end{pmatrix} \begin{pmatrix} z_1 \\ \vdots \\ z_{n-1} \\ z_n \end{pmatrix}. \quad (12.12)$$

Для ее последнего уравнения  $z'_n = \lambda_n z_n$  можно найти решение  $z_n = \exp(\lambda_n x) z_{n0}$ . Далее, уравнение для  $z_{n-1}$  имеет вид

$$z'_{n-1} = \lambda_{n-1} z_{n-1} + b_{n-1,n} z_n, \quad (12.12')$$

где  $z_n$  известно. Это линейное уравнение (неоднородное, если  $b_{n-1,n} \neq 0$ ), которое можно решить с помощью метода Эйлера, изложенного в разд. I.4. Здесь возможны два случая:

а) если  $\lambda_{n-1} \neq \lambda_n$ , то представим  $z_{n-1}$  в виде  $z_{n-1} = E \exp(\lambda_{n-1} x) + F \exp(\lambda_n x)$ , подставим эту формулу в (12.12') и сравним коэффициенты. В результате получим:  $F = b_{n-1,n} z_{n0} / (\lambda_n - \lambda_{n-1})$  и  $E = z_{n-1,0} - F$ .

б) Если  $\lambda_{n-1} = \lambda_n$ , то будем искать  $z_{n-1}$  в виде  $z_{n-1} = (E + Fx) \exp(\lambda_n x)$  и получим  $F = b_{n-1,n} z_{n0}$  и  $E = z_{n-1,0}$ .

На следующем этапе, действуя по той же схеме, найдем  $z_{n-2}$  и т. д. Для элементов резольвенты в случае  $\lambda_i \neq \lambda_j$  можно полу-

чить простые рекуррентные формулы [Парлетт (1976)]. Предположим, что

$$z_i(x) = \sum_{j=i}^n E_{ij} \exp(\lambda_j x),$$

и подставим это выражение в (12.12). Из сравнения коэффициентов мы получим для  $i = n, n - 1, n - 2$  и т. д. следующие значения:

$$E_{ih} = \frac{1}{\lambda_h - \lambda_i} \left( \sum_{j=i+1}^k b_{ij} E_{jh} \right), \quad k = i + 1, i + 2, \dots, \quad (12.13)$$

$$E_{ii} = z_{i,0} - \sum_{j=i+1}^n E_{ij}.$$

### Каноническая форма Жордана

«Проще, чем вы думали». (Американский математический журнал, 1980)

Если мы, как и многие математики, не будем чураться плохо обусловленных матриц (см. упр. 1), то полученную выше форму Шура мы сможем и дальше преобразовать к знаменитой *жордановой канонической форме*.

**Теорема 12.2.** [Жордан (1870), книга вторая, § 5 и 6] Для любой матрицы  $A$  существует невырожденная матрица  $T$ , такая, что

$$T^{-1}AT = \text{diag} \left\{ \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_1 \end{pmatrix}, \begin{pmatrix} \lambda_2 & & \\ & \ddots & \\ & & \lambda_2 \end{pmatrix}, \dots \right\}. \quad (12.14)$$

(Блоки могут иметь разную размерность, а  $\lambda_i$  не обязательно различны.)

**Доказательство.** [Флетчер—Соренсен (1983)] Мы можем считать, что матрица уже имеет форму Шура. Можно, конечно, также считать, что одинаковые собственные значения располагаются рядом на главной диагонали.

Следующий шаг состоит в том, чтобы убрать все ненулевые элементы, не принадлежащие верхним треугольным блокам,

содержащим одинаковые собственные значения. Продемонстрируем эту процедуру на следующем примере: пусть  $\lambda_1 \neq \lambda_2$  и

$$A = \begin{pmatrix} \lambda_1 & a & b & c \\ & \lambda_1 & d & e \\ & & \lambda_2 & f \\ & & & \lambda_2 \end{pmatrix}.$$

Наша задача — найти такую матрицу преобразования, чтобы исчезли  $b, c, d$  и  $e$ . Будем искать ее в виде

$$\left( \begin{array}{cc|cc} \lambda_1 & a & b & c \\ \lambda_1 & & d & e \\ \hline & & \lambda_2 & f \\ & & & \lambda_2 \end{array} \right) \left( \begin{array}{c|cc} I & p & q \\ & r & s \\ \hline 0 & & I \end{array} \right) = \left( \begin{array}{c|cc} I & p & q \\ & r & s \\ \hline 0 & & I \end{array} \right) \left( \begin{array}{cc|cc} \lambda_1 & a & 0 & 0 \\ & \lambda_1 & 0 & 0 \\ \hline & & \lambda_2 & f \\ & & & \lambda_2 \end{array} \right).$$

Все элементы  $c_{ij}$  этих матриц уже совпадают, за исключением правого верхнего блока  $2 \times 2$  ( $i = 1, 2$  и  $j = 3, 4$ ). Перемножая матрицы, получим  $\lambda_1 p + ar + b = \lambda_2 p$ ,  $\lambda_1 r + d = \lambda_2 r$  и аналогичные уравнения для  $q$  и  $s$ . Матрицы этих систем имеют треугольную структуру, и, поскольку  $\lambda_1 \neq \lambda_2$ , легко найти  $r$  и  $p$ , а затем  $s$  и  $q$ . Теперь ясно, что нужно делать в более общей ситуации (попытайтесь построить пример!).

На последующих шагах мы обрабатываем оставшиеся блоки раздельно. Таким образом, мы предполагаем, что все диагональные элементы блока равны, и рекурсивно преобразуем блок к форме, указанной в утверждении теоремы.

Будем считать, что матрица  $A$  имеет вид

$$A = \begin{pmatrix} \lambda & | & a^T \\ \hline 0 & | & A_1 \end{pmatrix},$$

и в качестве индуктивного предположения примем, что  $A_1$  можно привести к жордановой форме с помощью матрицы  $T_1$ . Тогда выполняется равенство

$$\left( \begin{array}{c|cc} \lambda & | & a^T \\ \hline 0 & | & A_1 \end{array} \right) \left( \begin{array}{c|cc} 1 & | & 0 \\ \hline 0 & | & T_1 \end{array} \right) = \left( \begin{array}{c|cc} 1 & | & 0 \\ \hline 0 & | & T_1 \end{array} \right) \left( \begin{array}{c|cc|cc} \lambda & | & b_1 & b_2 & b_3 \dots \\ \hline 0 & | & \lambda & 1 & \\ 0 & | & \lambda & & \\ \vdots & | & & \lambda & \dots \end{array} \right),$$

где вектор  $b$  имеет значение  $b = T_1^T a$ . Затем стоящую справа матрицу можно упростить следующим образом:

$$\left( \begin{array}{c|cc|c} \lambda & b_1 & b_2 & b_3 \dots \\ \hline 0 & \lambda & 1 & \\ 0 & & \lambda & \\ \hline \vdots & & \lambda \dots & \end{array} \right) \left( \begin{array}{c|cc|c} 1 & b_2 & 0 & 0 \\ \hline & 1 & & \\ & & 1 & \\ \hline & & & I \end{array} \right) =$$

$$= \left( \begin{array}{c|cc|c} 1 & b_2 & 0 & 0 \\ \hline & 1 & & \\ & & 1 & \\ \hline & & & I \end{array} \right) \left( \begin{array}{c|cc|c} \lambda & b_1 & 0 & b_3 \dots \\ \hline 0 & \lambda & 1 & \\ 0 & & \lambda & \\ \hline \vdots & & \lambda \dots & \end{array} \right).$$

Здесь мы будем различать два случая:

a)  $b_1 = 0$ . Тогда мы поменяем местами первую пару (блочных) строк и столбцов и снова получим индуктивное предположение.

b)  $b_1 \neq 0$ . В этом случае для замены  $b_1$  на 1 используем масштабирующую диагональную матрицу (упр. 4). Наша последняя задача — убрать  $b_3, b_4, \dots$ . Эти члены выталкиваются вправо следующим образом:

$$\left( \begin{array}{c|cc} \lambda & 1 & b_3 \ b_4 \\ \hline \lambda & 1 & \\ \lambda & & \\ \hline & \lambda & 1 \\ & & \lambda \end{array} \right) \left( \begin{array}{c|cc} 0 & 0 \\ \hline I & -b_3 \ -b_4 \\ 0 & 0 \\ \hline 1 & 0 \\ 0 & 1 \end{array} \right) =$$

$$= \left( \begin{array}{c|cc} 0 & 0 \\ \hline I & -b_3 \ -b_4 \\ 0 & 0 \\ \hline 1 & 0 \\ 0 & 1 \end{array} \right) \left( \begin{array}{c|cc} \lambda & 1 & \\ \hline \lambda & 1 & \\ \lambda & & \\ \hline & \lambda & 1 \\ & & \lambda \end{array} \right)$$

и т. д. Чтобы с помощью такой процедуры окончательно избавиться от всех  $b_i$ , мы должны предположить, что первый блок не меньше последующих. Это упорядочение надо провести в начале с помощью перестановок.

Произведением матриц всех этих преобразований и будет матрица  $T$  из (12.14). ■

**Следствие 12.3.** Для любой матрицы  $A$  и любого числа  $\varepsilon \neq 0$  существует невырожденная матрица  $T$  (зависящая от  $\varepsilon$ ), такая что

$$T^{-1}AT = \text{diag} \left\{ \begin{pmatrix} \lambda_1 & \varepsilon \\ & \ddots \\ & & \ddots & \varepsilon \\ & & & \lambda_1 \end{pmatrix}, \begin{pmatrix} \lambda_2 & \varepsilon \\ & \ddots \\ & & \ddots & \varepsilon \\ & & & \lambda_2 \end{pmatrix}, \dots \right\}. \quad (12.14')$$

*Доказательство.* Умножьте уравнение (12.14) справа на матрицу  $D = \text{diag}(1, \varepsilon, \varepsilon^2, \varepsilon^3, \dots)$ , а слева — на  $D^{-1}$ . ■

Численные трудности, возникающие при определении жордановой канонической формы, описаны в работе Голуба—Уилкинсона (1976). Для этой цели создано несколько программ, например программа, описанная в статье Когстрёма—Руэ (1980).

Когда матрица  $A$  приведена к жордановой канонической форме (12.14), решение дифференциального уравнения  $y' = Ay$  можно вычислить с помощью метода, изложенного в связи с (12.12'), случай б):

$$y(x) = TDT^{-1}y_0, \quad (12.15)$$

где  $D$  — блочно-диагональная матрица с блоками вида

$$\begin{pmatrix} e^{\lambda_1 x} & xe^{\lambda_1 x} & \dots & \frac{x^k}{k!} e^{\lambda_1 x} \\ & e^{\lambda_1 x} & & \cdot \\ & & \ddots & \cdot \\ & & & xe^{\lambda_1 x} \\ & & & e^{\lambda_1 x} \end{pmatrix}.$$

Это обобщение формулы (12.8).

### Геометрическое представление

На рис. 12.1 представлены геометрические формы решений системы  $y' = Ay$  размерности  $n = 2$ . Они изображены в виде траекторий в фазовом пространстве  $(y_1, y_2)$ . Случаи а), б), с) и е) относятся к уравнениям, полученным при линеаризации (12.16) в четырех критических точках (см. рис. 12.2).

Многое в их структуре сохраняется и для *нелинейных* систем (12.1) в окрестности точек равновесия. Исключение могут составлять «структурно неустойчивые» случаи, как, например, при

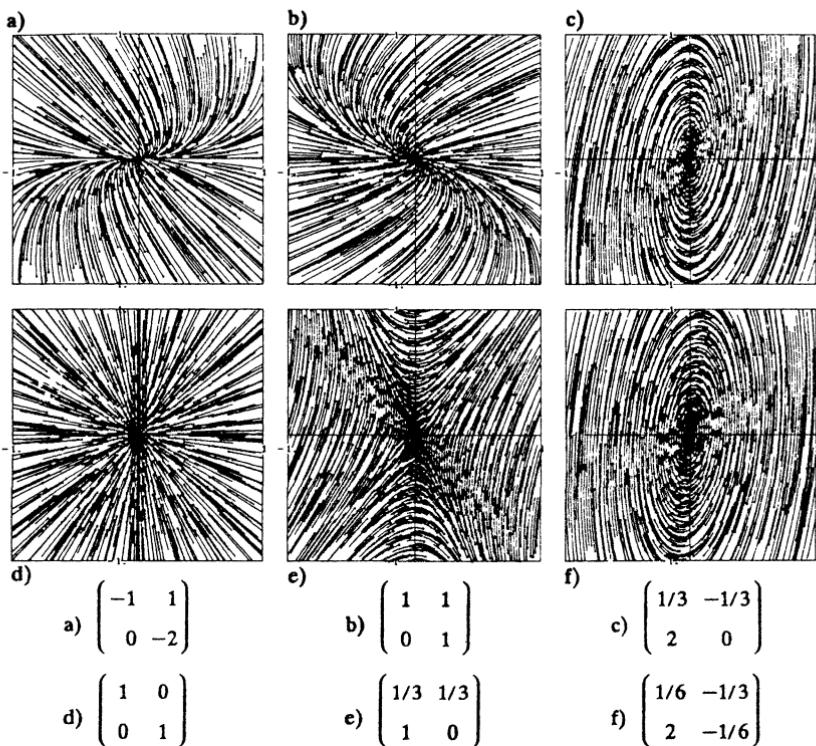


Рис. 12.1. Решения линейных двумерных систем.

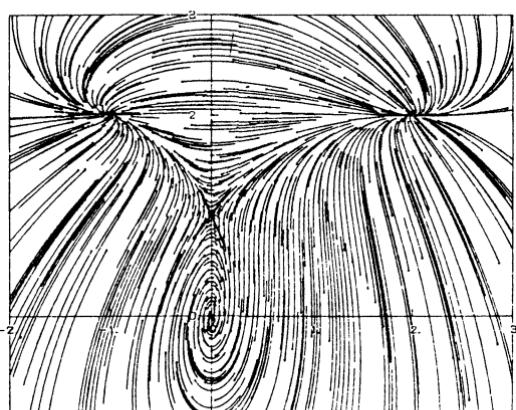


Рис. 12.2. Поток решений системы (12.16).

наличии комплексных собственных значений, у которых  $\alpha = \operatorname{Re}(\lambda) = 0$ . По этому вопросу написано много статей, где обсуждаются «критические» или «особые» точки (см., например, знаменитый труд Пуанкаре (1881, 1882, 1885)).

На рис. 12.2 мы показали решения квадратичной системы

$$\begin{aligned} y'_1 &= \frac{1}{3} (y_1 - y_2)(1 - y_1 - y_2), \\ y'_2 &= y_1(2 - y_2), \end{aligned} \quad (12.16)$$

которая имеет четыре критические точки всех четырех возможных структурно устойчивых типов (упр. 4).

### Упражнения

1. а) С помощью решения уравнений  $(A - \lambda_i I) v_i = 0$  найдите собственные векторы матрицы

$$A = \begin{pmatrix} -1 & 20 & & & & \\ -2 & 20 & & & & \\ -3 & 20 & & & & \\ . & . & & & & \\ . & . & & & & \\ -19 & 20 & & & & \\ & -20 & & & & \end{pmatrix}. \quad (12.17)$$

*Ответ.*  $v_1 = (1, 0, \dots)^T$ ,  $v_2 = (1, -1/20, 0, \dots)^T$ ,  $v_3 = (1, -2/20, 2/400, 0, \dots)^T$ ,  $v_4 = (1, -3/20, 6/400, -6/8000, 0, \dots)^T$

и т. д.

б) Численно найдите матрицу, обратную к  $T = (v_1, v_2, \dots, v_n)$ , и определите ее наибольший элемент (ответ:  $4.5 \times 10^{12}$ ). Таким образом, матрица  $T$  очень плохо обусловлена.

с) Найдите численно или аналитически из (12.13) решения системы

$$y' = Ay, \quad y_i(0) = 1, \quad i = 1, \dots, 20. \quad (12.18)$$

Обратите внимание на наличие «горба» [Молер и ван Лоан 1978]: хотя все собственные значения  $A$  отрицательны, решения сначала очень сильно растут и только после этого затухают, стремясь к нулю. Это типично для несимметричных матриц и связано с плохой обусловленностью  $T$ . (См. рис. 12.3, где решения показаны в логарифмическом масштабе (!).)

*Ответ.*

$$y_1 = -e^{-20x} \frac{20^{18}}{19!} + e^{-19x} \frac{(1+20) 20^{18}}{18!}$$

$$-e^{-18x} \frac{(1+20+20^2/2!) 20^{18}}{17!} \pm \dots$$

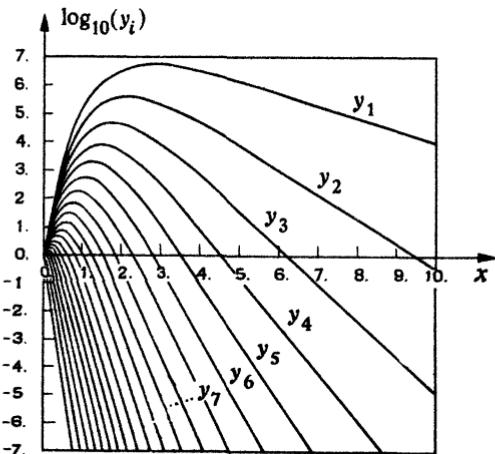


Рис. 12.3. Решения уравнения (12.18) с матрицей (12.17).

2. (Шур.) Докажите, что собственные значения матрицы  $A$  удовлетворяют оценке

$$\sum_{i=1}^n |\lambda_i|^2 \leq \sum_{i,j=1}^n |a_{ij}|^2$$

и равенство выполняется тогда и только тогда, когда  $A$  приводится к диагональному виду с помощью ортогонального преобразования (см. также упр. 3).

*Указание.*  $\sum_{i,j} |a_{ij}|^2$  — это след матрицы  $A^*A$ , т. е. инвариант унитарных преобразований  $Q^*AQ$ .

3. Покажите, что разложение Шура  $S = Q^*AQ$  имеет диагональный вид тогда и только тогда, когда  $A^*A = AA^*$ . Такие матрицы называются *нормальными*.

*Указание.* Это условие эквивалентно равенству  $S^*S = SS^*$ . Покажите, что кососимметричные матрицы являются нормальными.

4. Пусть задано  $b \neq 0$ . Определите такое  $\alpha$ , при котором выполняется равенство

$$\begin{pmatrix} \lambda & b \\ \lambda & \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \alpha & \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \alpha & \end{pmatrix} \begin{pmatrix} \lambda & 1 \\ \lambda & \lambda \end{pmatrix}.$$

5. Найдите четыре критические точки системы (12.16) и для каждой из них — собственные значения и векторы матрицы  $df/dy$ . Сравните свои результаты с рис. 12.2 и 12.1.
6. Определите разложение Шура и жорданову каноническую форму матрицы

$$A = \frac{1}{9} \begin{pmatrix} 14 & 4 & 2 \\ -2 & 20 & 1 \\ -4 & 4 & 20 \end{pmatrix}.$$

*Ответ.* Жорданова каноническая форма имеет вид

$$\begin{pmatrix} 2 & 1 & \\ & 2 & \\ & & 2 \end{pmatrix}.$$

7. Приведите матрицы

$$A = \begin{pmatrix} \lambda & 1 & b & c \\ & \lambda & 1 & d \\ & & \lambda & 1 \\ & & & \lambda \end{pmatrix}, \quad A = \begin{pmatrix} \lambda & 1 & b & c \\ & \lambda & 0 & d \\ & & \lambda & 1 \\ & & & \lambda \end{pmatrix}$$

к жордановой канонической форме. Во втором случае рассмотрите две возможности:  $b + d = 0$  и  $b + d \neq 0$ .

## I.13. Устойчивость

«Жюри конкурса извещает о том, что Премия 1877 года будет присуждаться за исследование по теме: *Критерий динамической устойчивости.*»

(С. Дж. Фир (вице-председатель), Дж. Каллис, Дж. Дж. Стокс, Дж. Клерк Максвелл)

### Введение

«Чтобы проиллюстрировать значение данного вопроса, представим частицу, скользящую вниз внутри гладкого наклонного цилиндра вдоль его нижней образующей либо по внешней стороне цилиндра вдоль верхней образующей. В первом случае слабое нарушение движения вызвало бы лишь осцилляции частицы около образующей, а во втором случае частица совсем удалилась бы от образующей. В первом случае движение было бы устойчивым в смысле поставленного вопроса, а во втором — неустойчивым... желательно найти условие, позволяющее нам судить, когда движение системы, возможное с точки зрения динамики, отвечает требованию: *при слабом нарушении движения оно продолжится лишь с малыми отклонениями.*» («Жюри конкурса» в книге Рауса (1877)).

Всякий раз, когда аналитическое решение задачи неизвестно, численные решения можно получить только для заданных начальных значений. Но часто нам нужна информация об устойчивости решений при *всех* начальных значениях в окрестности некоторой точки равновесия. Как и в разд. I.12, мы поместим точку равновесия в начало координат и дадим следующее определение.

**Определение 13.1.** Пусть имеется система уравнений

$$y'_i = f_i(y_1, \dots, y_n), \quad i = 1, \dots, n, \quad (13.1)$$

в которой  $f_i(0, \dots, 0) = 0$ ,  $i = 1, \dots, n$ . Тогда точка 0 (начало координат) называется *устойчивой по Ляпунову*, если для любого  $\varepsilon > 0$  существует  $\delta > 0$ , такое, что для всех решений системы из неравенства  $\|y(x_0)\| < \delta$  следует, что  $\|y(x)\| < \varepsilon$  при всех  $x > x_0$ .

В своей знаменитой работе, получившей премию Адамса, Раус (1877) сделал первый шаг: он исследовал устойчивость линеаризованного уравнения

$$y'_i = \sum_j a_{ij} y_j, \quad a_{ii} = \frac{\partial f_i}{\partial y_j}(0). \quad (13.2)$$

(«Будем говорить, что величины  $x, y, z$  и т. д. *малы*, если можно пренебречь их квадратами.») Из вида полученного в разд. I.12 общего решения (13.2) непосредственно вытекает следующий результат.

**Теорема 13.1.** *Линеаризованное уравнение (13.2) устойчиво по Ляпунову тогда и только тогда, когда все корни характеристического уравнения*

$$\det(\lambda I - A) = a_0\lambda^n + a_1\lambda^{n-1} + \dots + a_{n-1}\lambda + a_n = 0 \quad (13.3)$$

*удовлетворяют условию  $\operatorname{Re}(\lambda) \leq 0$ , а для кратных корней, порождающих жордановы клетки, выполняется неравенство  $\operatorname{Re}(\lambda) < 0$ .*

*Доказательство.* Непосредственно усматривается из (12.12) и (12.15). Для жордановых клеток «вековой» член [т. е.  $E + Fx$  в решении уравнения (12.12), случай б)], стремящийся к бесконечности с ростом  $x$ , должен быть «убит» с помощью экспоненты со строго отрицательным показателем. ■

### Критерий Рауса — Гурвица

Следующим шагом, который привел к знаменитому критерию Рауса—Гурвица, была проверка условий  $\operatorname{Re}(\lambda) < 0$  *прямо по коэффициентам уравнения* (13.3) без вычисления его корней. Для решения данной задачи Раус объединил две известные идеи; первая — это принцип аргумента (Коши), согласно которому число корней многочлена  $p(z) = u(z) + iv(z)$  внутри замкнутого контура равно числу (положительных) оборотов вектора  $(u(z), v(z))$  при обходе переменной  $z$  границы в положительном направлении (см., например, книгу Хенричи (1974)).<sup>1)</sup> На рис. 13.1 представлен пример для многочлена

$$\begin{aligned} z^6 + 6z^5 + 16z^4 + 25z^3 + 24z^2 + 14z + 4 &= \\ = (z+1)(z+2)(z^2+z+1)(z^2+2z+2). \end{aligned} \quad (13.4)$$

На полуокружности  $z = Re^{i\theta}$  ( $\pi/2 \leq \theta \leq 3\pi/2$ ,  $R$  достаточно велико) аргумент  $p(z)$  благодаря доминирующему члену  $z^n$  делает  $n/2$  положительных оборотов. Чтобы все нули  $p$  принадлежали отрицательной полуплоскости, дополнительно требуется  $n/2$  положительных оборотов вдоль мнимой оси. Отсюда вытекает

**Лемма 13.2.** *Пусть  $p(z)$  является многочленом степени  $n$  и пусть  $p(iy) \neq 0$  при  $y \in \mathbb{R}$ . В этом случае все корни  $p(z)$  лежат*

<sup>1)</sup> На русском языке см., например, М. А. Лаврентьев и Б. В. Шабат, Методы теории функций комплексного переменного (изд. 3-е) — М.: Наука, 1965, с. 88. — Прим. ред.

в отрицательной полуплоскости тогда и только тогда, когда вдоль мнимой оси  $\arg(p(iy))$  делает  $n/2$  положительных оборотов при  $y$ , изменяющемся от  $-\infty$  до  $+\infty$ . ■

Вторая идея состоит в использовании теоремы Штурма [Штурм (1829)], которая тесно связана с алгоритмом Евклида для многочленов. Штурм сделал открытие, что при делении многочлена

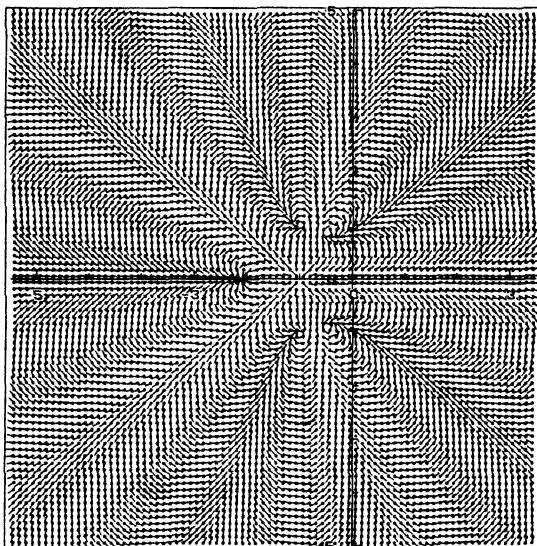


Рис. 13.1. Векторное поле функции  $\arg(p(z))$  для многочлена  $p(z)$ , определенного формулой (13.4).

$p_{i-1}(y)$  на  $p_i(y)$  остаток  $p_{i+1}(y)$  лучше брать с *отрицательным* знаком:

$$p_{i-1}(y) = p_i(y) q_i(y) - p_{i+1}(y). \quad (13.5)$$

Тогда в силу свойства «системы Штурма», состоящего в том, что

$$\operatorname{sign}(p_{i+1}(y)) \neq \operatorname{sign}(p_{i-1}(y)), \text{ если } p_i(y) = 0, \quad (13.6)$$

величина

$w(y)$  — число перемен знака в ряду чисел

$$(p_0(y), p_1(y), \dots, p_m(y)) \quad (13.7)$$

не меняется в нулях многочленов  $p_1(y), \dots, p_{m-1}(y)$ . Из этого вытекает следующий результат.

**Лемма 13.3.** Предположим, что система вещественных многочленов  $p_0(y)$ ,  $p_1(y)$ , ...,  $p_m(y)$  удовлетворяет условиям:

- i)  $\deg(p_0) > \deg(p_1)$ ;
- ii)  $p_0(y)$  и  $p_1(y)$  не обращаются одновременно в нуль;
- iii)  $p_m(y) \neq 0$  при всех  $y \in \mathbb{R}$ ;
- iv) выполняется свойство (13.6) системы Штурма.

Тогда величина

$$\frac{w(\infty) - w(-\infty)}{2} \quad (13.8)$$

равна отсчитываемому в положительном направлении числу оборотов вектора  $(p_0(y), p_1(y))$  при  $y$ , стремящемся от  $-\infty$  к  $+\infty$ .

**Доказательство.** В силу свойства системы Штурма  $w(y)$  не меняется в нулях  $p_1(y), \dots, p_{m-1}(y)$ . По предположению iii),

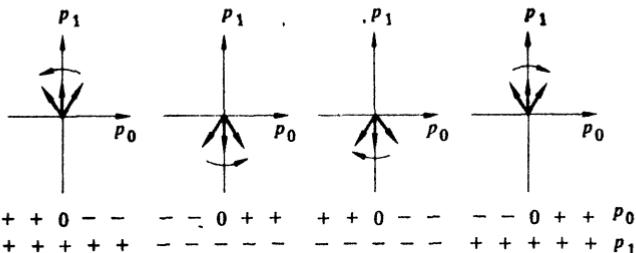


Рис. 13.2. Повороты вектора  $(p_0(y), p_1(y))$  и соответствующая функция  $w$ .

$p_m(y)$  также не оказывает влияния. Поэтому  $w(y)$  может изменяться только в нулях  $p_0(y)$ . Если  $w(y)$  в точке  $\hat{y}$  возрастает на единицу, то  $p_0(y)$  либо изменяет знак с «+» на «-» и  $p_1(\hat{y}) > 0$ , либо с «-» на «+» и  $p_1(\hat{y}) < 0$  (по условию ii) равенство  $p_i(\hat{y}) = 0$  невозможно). В обоих случаях вектор  $(p_0(y), p_1(y))$ , поворачиваясь, пересекает мнимую ось в положительном направлении (см. рис. 13.2). Если  $w(y)$  уменьшается на единицу, то этот вектор пересекает мнимую ось в отрицательном направлении. Утверждение теоремы следует теперь из условия i), так как вектор  $(p_0(y), p_1(y))$  направлен горизонтально при  $y \rightarrow -\infty$  и  $y \rightarrow +\infty$ . ■

Объединение двух предыдущих лемм дает нам требуемый критерий устойчивости: пусть задан характеристический многочлен (13.3)

$$p(z) = a_0 z^n + a_1 z^{n-1} + \dots + a_n = 0, \quad a_0 > 0.$$

Разделим  $p(iy)$  на  $i^n$  и отделим вещественную и мнимую части:

$$p_0(y) = \operatorname{Re} \frac{p(iy)}{i^n} = a_0 y^n - a_2 y^{n-2} + a_4 y^{n-4} \pm \dots, \quad (13.9)$$

$$p_1(y) = -\operatorname{Im} \frac{p(iy)}{i^n} = a_1 y^{n-1} - a_3 y^{n-3} + a_5 y^{n-5} \pm \dots.$$

Благодаря специальной структуре этих многочленов алгоритм Евклида (13.5) для них чрезвычайно прост: запишем многочлен

$$p_i(y) = c_{i0}y^{n-i} + c_{i1}y^{n-i-2} + c_{i2}y^{n-i-4} + \dots, \quad (13.10)$$

и для частного в (13.5) имеем формулу  $q_i(y) = (c_{i-1,0}/c_{i0}) \cdot y$ , при условии, что  $c_{i0} \neq 0$ . Теперь, подставив (13.10) в (13.5), придем к следующим рекуррентным формулам для вычисления коэффициентов  $c_{ij}$ :

$$c_{i+1,j} = c_{i,j+1} \frac{c_{i-1,0}}{c_{i0}} - c_{i-1,j+1} = \frac{1}{c_{i0}} \det \begin{pmatrix} c_{i-1,0} & c_{i-1,j+1} \\ c_{i,0} & c_{i,j+1} \end{pmatrix}. \quad (13.11)$$

Если при некотором  $i$   $c_{i0} = 0$ , то частное  $q_i(y)$  будет многочленом степени выше первой и алгоритм Евклида будет оканчиваться на  $p_m(y)$ , где  $m < n$ .

Очевидно, что полученная таким способом система многочленов ( $p_i(y)$ ) удовлетворяет условиям i) и iv) леммы 13.3. Условие ii) эквивалентно неравенству  $p(iy) \neq 0$  при  $y \in \mathbb{R}$ , а iii) следует из ii), так как  $p_m(y)$  — это наибольший общий делитель многочленов  $p_0(y)$  и  $p_1(y)$ .

**Теорема 13.4.** [Раус (1877)] Все корни вещественного многочлена (13.3) с коэффициентом  $a_0 > 0$  лежат в отрицательной полуплоскости  $\operatorname{Re} \lambda < 0$  тогда и только тогда, когда

$$c_{i0} > 0 \quad \text{при } i = 0, 1, 2, \dots, n. \quad (13.12)$$

**Замечание.** Благодаря условию  $c_{i0} > 0$  деление на  $c_{i0}$  в формуле (13.11) можно опустить (общий положительный множитель в  $p_{i+1}(y)$ ), откуда следует та же теорема [Раус (1877)]: «...таким образом, вспомнив это простое правило перекрестного умножения, мы можем записать...». Однако при больших  $n$  это не рекомендуется из-за возможного переполнения.

**Доказательство.** Системы координат  $(p_0, p_1)$  и  $(\operatorname{Re}(p), \operatorname{Im}(p))$  имеют противоположную ориентацию. Поэтому  $n/2$  положительных оборотов  $p(iy)$  соответствуют  $n/2$  отрицательным оборотам  $(p_0(y), p_1(y))$ . Если все корни  $p(\lambda)$  лежат в отрицательной полуплоскости  $\operatorname{Re} \lambda < 0$ , то из лемм 13.2 и 13.3 следует равенство  $w(\infty) - w(-\infty) = -n$ , которое возможно только в том случае, когда  $w(\infty) = 0$ ,  $w(-\infty) = n$ . Это означает, что у всех многочленов  $p_i(y)$  коэффициенты при старших членах положительны.

С другой стороны, если выполнено условие (13.12), то  $p_n(y) \equiv c_{n0}$ . Следовательно, многочлены  $p_0(y)$  и  $p_1(y)$  не могут иметь общих множителей и  $p(\lambda) \neq 0$  на мнимой оси. Теперь, снова применяя леммы 13.2 и 13.3, мы придем к утверждению теоремы. ■

Эту теорему можно доказать и по-другому, используя «положительные функции» (см. том II).

*Пример 1.* В табл. 13.1 приведена таблица коэффициентов Рауса (13.11) для многочлена (13.4). Ясно, что он удовлетворяет условиям устойчивости.

*Пример 2.* [Раус (1877)] Выразим условия устойчивости для уравнения четвертой степени

$$z^4 + pz^3 + qz^2 + rz + s = 0. \quad (13.13)$$

Значения  $c_{ij}$  (без деления) приведены в табл. 13.2.

Таблица 13.2. Таблица Рауса для (13.13)

	$j = 0$	$j = 1$	$j = 2$
$i = 0$	1	$-16$	$24$
$i = 1$	$p$	$-25$	$14$
$i = 2$	$pq - r$	$11.83$	$-21.67$
$i = 3$	$(pq - r)r - p^2s$	$14.01$	$-11.97$
$i = 4$	$((pq - r)r - p^2s)ps$	$11.56$	$-4$
$i = 5$		$7.12$	
$i = 6$		$4$	

Мы имеем устойчивость тогда и только тогда, когда выполняются неравенства

$$p > 0, \quad pq - r > 0, \quad ((pq - r)r - p^2s)ps > 0, \quad s > 0.$$

### Вопросы численной реализации

Несмотря на большое историческое значение и математическое изящество критерия Рауса, при больших размерностях для его применения в вычислительной практике в настоящее время существует два препятствия:

- 1) для больших матриц нелегко вычислить характеристический многочлен;
- 2) этот многочлен очень опасно использовать при наличии погрешностей округления.

Таким образом, если мы не занимаемся точными символьными преобразованиями или не производим приближенные расчеты с высокой степенью точности, рекомендуется избегать характеристических многочленов и использовать численно устойчивые алгоритмы для задачи на собственные значения (см., например, книгу Эйспака (1974)).

Таблица 13.1. Таблица Рауса для (13.4)

	$j = 0$	$j = 1$	$j = 2$	$j = 3$
$i = 0$	1	$-16$	$24$	$-4$
$i = 1$	$6$	$-25$	$14$	
$i = 2$	$11.83$	$-21.67$	$4$	
$i = 3$	$14.01$	$-11.97$		
$i = 4$	$11.56$	$-4$		
$i = 5$	$7.12$			
$i = 6$	$4$			

### Численные эксперименты. 1. Матрица

$$A = \left[ \begin{array}{cccc|cc} -0.5 & & & & -1 & & \\ & \ddots & & & & \ddots & \\ & & -0.5 & & & & -n \\ \hline & & & 1 & & -0.5 & \\ & & & & n & & -0.5 \\ & & & & & & \end{array} \right]$$

имеет характеристический многочлен

$$p(z) = \prod_{j=1}^n (z^2 + 0.1z + j^2 + 0.0025).$$

Мы вычислили коэффициенты  $p$  с двойной точностью, а затем применили алгоритм Рауса с обычной точностью (которая у нашей машины равна  $1.5 \times 10^{-8}$ ). Результаты вычислений указывают на устойчивость только при  $n \leq 15$ , а при  $n \geq 16$  она не обнаружена, хотя собственные числа матрицы  $-0.05 \pm ki$  всегда лежат в отрицательной полуплоскости. С другой стороны, при прямом вычислении собственных значений  $A$  с помощью подпрограмм Эйспака осложнения не возникали ни при каких  $n$ .

2. Мы испытывали алгоритм Рауса также и на (масштабированных) *числителях диагональных аппроксимаций Паде* для  $\exp(z)$ ,

$$1 + \frac{n}{2n} (nz) + \frac{n(n-1)}{(2n)(2n-1)} \frac{(nz)^3}{2!} + \frac{n(n-1)(n-2)}{(2n)(2n-1)(2n-2)} \frac{(nz)^5}{3!} + \dots, \quad (13.14)$$

у которых, как известно, все собственные значения принадлежат  $\mathbb{C}^-$ . В этом примере результаты были правильными лишь при  $n \leq 21$ , а при больших  $n$  — неверными из-за погрешностей округления.

### Функции Ляпунова

Продолжим рассмотрение вопроса об устойчивости нелинейной системы (13.1): «можно ли ее действительно определить, изучая только члены первого порядка?» [Раус (1877), гл. VII]. Начало этой теории положили Раус и Пуанкаре, а свое завершение она получила в знаменитой работе Ляпунова (1892). В качестве общей ссылки на обширную теорию, которая была разработана с тех пор, приведем книги Руша, Хабетса и Лалуа (1977) и В. Хана (1967).

Ляпунову (и Раусу) основным инструментом служили так называемые функции Ляпунова  $V(y_1, \dots, y_n)$ , которые должны удовлетворять условиям

$$\begin{aligned} V(y_1, \dots, y_n) &\geq 0, \\ V(y_1, \dots, y_n) = 0 &\text{ тогда и только тогда,} \\ \text{когда } y_1 = \dots = y_n = 0, \end{aligned} \quad (13.15)$$

и на решениях задачи (13.1)

$$\frac{d}{dx} V(y_1(x), \dots, y_n(x)) \leq 0. \quad (13.16)$$

Обычно при малых  $y$  функция  $V(y)$  имеет квадратичное поведение, и условие (13.15) означает, что

$$c \|y\|^2 \leq V(y) \leq C \|y\|^2, \quad C \geq c > 0. \quad (13.17)$$

Существование такой функции Ляпунова является достаточным условием устойчивости начала координат.

Начнем с построения функций Ляпунова в линейном случае:

$$y' = Ay. \quad (13.18)$$

Лучше всего это делать в базисе, который естественно задается собственными векторами (или жордановыми цепочками векторов) матрицы  $A$ . Поэтому введем преобразование  $y = Tz$ ,  $z = T^{-1}y$ , которое приводит  $A$  к жордановой канонической форме (12.14')  $J = T^{-1}AT$ , а систему (13.18) к виду

$$z' = Jz. \quad (13.19)$$

Если мы положим

$$V_0(z) = \|z\|^2 \text{ и } V(y) = V_0(T^{-1}y) = V_0(z), \quad (13.20)$$

то производная  $V(y(x))$  будет вычисляться по формулам

$$\begin{aligned} \frac{d}{dx} V(y(x)) &= \frac{d}{dx} V_0(z(x)) = 2 \operatorname{Re} \langle z(x), z'(x) \rangle \\ &= 2 \operatorname{Re} \langle z(x), Jz(x) \rangle \leq 2\mu(J) V(y(x)). \end{aligned} \quad (13.21)$$

Согласно формуле (10.20) для логарифмической нормы,

$2\mu(J)$  = наибольшее собственное значение матрицы  $J + J^*$ .

Матрица  $J + J^*$  является блочно-диагональной и состоит из трехдиагональных блоков:

$$\left[ \begin{array}{cccccc} 2 \operatorname{Re} \lambda_1 & & & & & \\ & \ddots & & & & \\ & & 2 \operatorname{Re} \lambda_2 & & & \\ & & & \ddots & & \\ & & & & \ddots & \\ & & & & & 2 \operatorname{Re} \lambda_n \end{array} \right]. \quad (13.22)$$

Вычтя диагональ и применив формулу (6.7а), мы увидим, что собственные значения матрицы (13.22) размера  $m \times m$  равны

$$2 \left( \operatorname{Re} \lambda_i + \varepsilon \cos \frac{\pi k}{m+1} \right), \quad k = 1, \dots, m. \quad (13.23)$$

Используя эту формулу или с помощью упр. 4 мы получим следующую лемму.

**Лемма 13.5.** *Если все собственные значения  $A$  удовлетворяют условию  $\operatorname{Re} \lambda_i < -\rho < 0$ , то для уравнения (13.18) существует квадратичная функция Ляпунова, удовлетворяющая неравенству*

$$\frac{d}{dx} V(y(x)) \leq -\rho V(y(x)). \quad \blacksquare \quad (13.24)$$

Последнее дифференциальное неравенство приводит к оценке (теорема 10.1)

$$V(y(x)) \leq V(y_0) \cdot \exp(-\rho(x - x_0))$$

и обеспечивает выполнение соотношения  $\lim_{x \rightarrow \infty} \|y(x)\| = 0$ , т. е. асимптотическую устойчивость.

### Устойчивость нелинейных систем

Теперь легко распространить те же идеи на *нелинейные уравнения*. Примером служит следующая теорема.

**Теорема 13.6.** *Пусть задана нелинейная система*

$$y' = Ay + g(x, y), \quad (13.25)$$

*у которой все собственные значения матрицы  $A$  отвечают условию  $\operatorname{Re} \lambda_i < -\rho < 0$ . Предположим далее, что для любого  $\varepsilon > 0$  существует такое  $\delta > 0$ , что*

$$\|g(x, y)\| \leq \varepsilon \|y\| \text{ при } \|y\| < \delta, \quad x \geq x_0. \quad (13.26)$$

*Тогда начало координат асимптотически устойчиво по Ляпунову.*

**Доказательство.** Используя построенную в лемме 13.5 функцию Ляпунова  $V(y)$ , из (13.25) получим

$$\frac{d}{dx} V(y(x)) \leq -\rho V(y(x)) +$$

$$+ 2 \operatorname{Re} \langle T^{-1}y(x), T^{-1}g(x, y(x)) \rangle. \quad (13.27)$$

Из неравенства Коши с учетом (13.26) следует, что

$$\frac{d}{dx} V(y(x)) \leq (-\rho + \|T\| \cdot \|T^{-1}\| \varepsilon) V(y(x)). \quad (13.28)$$

При достаточно малых  $\varepsilon$  выражение в правой части отрицательно, и мы получим асимптотическую устойчивость. ■

Мы видим, что для нелинейных систем *устойчивость обеспечивается только в окрестности начала координат*. Это можно также заметить и на рис. 12.2. Другое отличие состоит в том, что *может быть нарушена устойчивость для собственных значений, лежащих на мнимой оси*. Примером служит система [Раус (1877)]

$$\begin{aligned} y_1' &= -y_2 + y_1^3, \\ y_2' &= y_1. \end{aligned} \quad (13.29)$$

Здесь, взяв функцию Ляпунова  $V = (y_1^2 + y_2^2)/2$ , мы получим производную  $V' = y_1^4$ , которая положительна при  $y_1 \neq 0$ . Поэтому все решения с ненулевыми начальными значениями возрастают. В работе Ваннера (1983) дан обзор по этому вопросу («проблема центра» и ее связь с предельными циклами).

### Устойчивость неавтономных систем

Если коэффициенты системы

$$y' = A(x) y \quad (13.30)$$

не являются постоянными, проверка условия устойчивости для собственных значений  $A$  при каждом значении  $x$  становится недостаточной для утверждения об устойчивости.

*Примеры:* 1. [Раус (1877)] Системе

$$y_1' = y_2, \quad y_2' = -\frac{1}{4x^2} y_1 \quad (13.31)$$

удовлетворяет функция  $y_1(x) = a\sqrt{x}$ .

2. Пример со строго отрицательными собственными значениями: начнем с задачи

$$B = \begin{pmatrix} -1 & 0 \\ 4 & -1 \end{pmatrix}, \quad y' = By.$$

Рассмотрение производной функции  $V = (y_1^2 + y_2^2)/2$  показывает, что  $V$  *возрастает* в секторе

$$\frac{1}{2 + \sqrt{3}} < \frac{y_2}{y_1} < 2 + \sqrt{3}.$$

Идея состоит в том, чтобы взять начальные значения из этой области и с ростом  $x$  поворачивать систему координат с той же скоростью, с которой вращается решение:

$$y' = T(x)BT(-x)y = A(x)y, \quad T(x) = \begin{pmatrix} \cos ax & -\sin ax \\ \sin ax & \cos ax \end{pmatrix}. \quad (13.32)$$

При  $y(0) = (1, 1)^T$  подходит выбор  $a = 2$ , и (13.32) имеет решение

$$y(x) = ((\cos 2x - \sin 2x)e^x, (\cos 2x + \sin 2x)e^x)^T. \quad (13.33)$$

Ясно, что оно неустойчиво, хотя при всех  $x$  величина  $-1$  остается двукратным собственным значением матрицы  $A(x)$ .

Третий пример можно найти в приведенном ниже упр. 6.

Мы видим, что для неавтономных систем теория устойчивости является более сложной. Отметим два частных случая, когда устойчивость может быть установлена:

1.  $a_{ii}(x) < 0$  и  $A(x)$  отвечает условию диагонального доминирования; тогда  $\mu(A(x)) \leq 0$ , и устойчивость следует из теоремы 10.5.

2.  $A(x) = B + C(x)$ , где  $B$  — постоянная матрица, собственные значения которой отвечают неравенствам  $\operatorname{Re} \lambda_i < -\rho < 0$ , а  $\|C(x)\| < \varepsilon$ , причем  $\varepsilon$  настолько мало, что можно применить доказательство теоремы 13.6.

### Упражнения

1. Напишите условия устойчивости для многочленов  $z^2 + pz + q = 0$  и  $z^3 + pz^2 + qz + r = 0$ .

*Ответ.* а)  $p > 0$  и  $q > 0$ ; б)  $r > 0$  и  $pq - r > 0$ .

2. [Гурвиц (1895)] Проверьте, что условие (13.12) эквивалентно положительности главных миноров матрицы

$$H = \begin{pmatrix} a_1 & a_3 & a_5 & \dots \\ a_0 & a_2 & a_4 & \dots \\ a_1 & a_3 & \dots \\ a_0 & a_2 & \dots \\ \dots & \dots \end{pmatrix} = (a_{2j-i})_{i=1}^n$$

( $a_k = 0$  при  $k < 0$  и  $k > n$ ). Убедитесь, что алгоритм Райса (13.11) — это разновидность метода исключения Гаусса, преобразующего  $H$  к треугольной форме.

### 3. Многочлен

$$\frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6} \frac{z^5}{5!} + \frac{5 \cdot 4 \cdot 3 \cdot 2}{10 \cdot 9 \cdot 8 \cdot 7} \frac{z^4}{4!} + \frac{5 \cdot 4 \cdot 3}{10 \cdot 9 \cdot 8} \frac{z^3}{3!} + \frac{5 \cdot 4}{10 \cdot 9} \frac{z^2}{2!} + \frac{5}{10} z + 1$$

является числителем аппроксимации Паде 5—5 для  $\exp z$ . Проверьте, что все его корни удовлетворяют условию  $\operatorname{Re} z < 0$ . Попытайтесь установить этот результат для произвольных  $n$  (см., например, работу Биркгофа и Варги (1965), лемма 7).

4. (Гершгорин) Докажите, что собственные значения матрицы  $A = (a_{ij})$  лежат в объединении кругов

$$\left\{ z \mid |z - a_{ii}| \leq \sum_{j \neq i} |a_{ij}| \right\}.$$

*Указание.* Запишите формулу  $Ax = \lambda x$  в координатах

$$\sum_j a_{ij} x_j = \lambda x_i,$$

перенесите диагональные элементы в правую часть и выберите такое  $i$ , чтобы  $|x_i|$  был максимальным.

5. Определите устойчивость начала координат для системы

$$\begin{aligned} y_1' &= -y_2 - y_1^2 - y_1 y_2, \\ y_2' &= y_1 + 2y_1 y_2. \end{aligned}$$

*Указание.* Исходя из выражения  $V = (y_1^2 + y_2^2)/2 + \dots$ , найдите такую функцию Ляпунова 4-й степени, чтобы выполнялось равенство  $V' = K(y_1^2 + y_2^2)^2 + \dots$ , и определите знак  $K$ .

6. (Дж. Ламберт, 1986). Рассмотрите систему

$$y' = A(x) \cdot y, \text{ где } A(x) = \begin{pmatrix} -\frac{1}{4x} & \frac{1}{x^2} \\ -\frac{1}{4} & -\frac{1}{4x} \end{pmatrix}. \quad (13.34)$$

a) Покажите, что оба собственных значения матрицы  $A(x)$  при всех  $x > 0$  удовлетворяют условию  $\operatorname{Re} \lambda < 0$ .

b) Вычислите  $\mu(A)$  по формуле (10.20) и покажите, что  $\mu(A) \leqslant 0$  тогда и только тогда, когда  $\sqrt{5} - 1 \leqslant x \leqslant \sqrt{5} + 1$ .

c) Найдите общее решение системы (13.34).

*Указание.* Введите новые функции

$$z_2(x) = y_2(x), \quad z_1(x) = xy_1(x),$$

с помощью которых вы придетете ко второму уравнению из (11.19) (упр. 5 в разд. I.11).

*Решение:*

$$y_1(x) = x^{-3/4} (a + b \log x), \quad (13.35)$$

$$y_2(x) = x^{1/4} \left( -\frac{a}{2} + b \left( 1 - \frac{1}{2} \log x \right) \right).$$

d) Определите такие  $a$  и  $b$ , чтобы функция  $\|y(x)\|_2^2$  была возрастающей при  $0 < x < \sqrt{5} - 1$ .

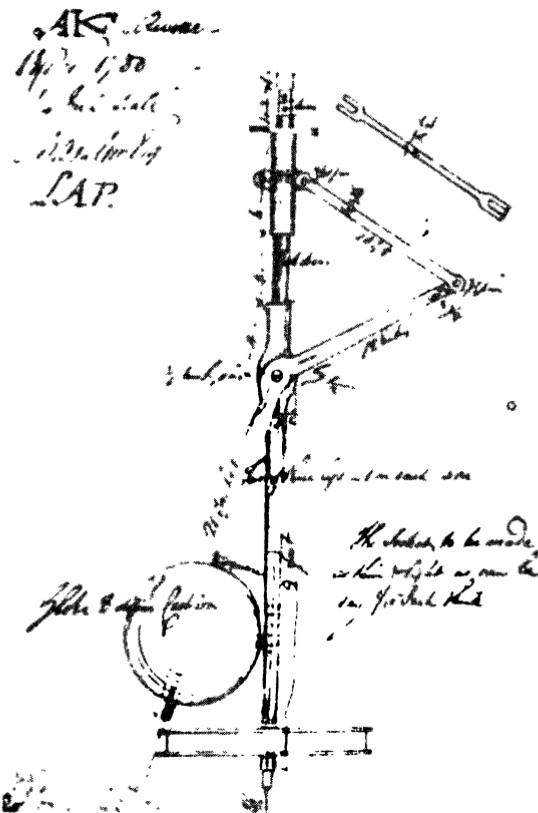


Рис. 13.3. Регулятор скорости паровой машины Джеймса Уатта.

e) Определите такие  $a$  и  $b$ , чтобы функция  $\|y(x)\|_2^2$  была возрастающей при  $\sqrt{5} + 1 < x < \infty$ .

Ответы:  $b = 1.8116035a$  для d) и

$b = 0.2462015a$  для e).

7. Помогите Джеймсу Уатту (см. копию оригинального чертежа 1788 г. на рис. 13.3) решить задачу устойчивости для регуля-

тора его паровой машины: если  $\omega$  — угловая скорость вращения вала двигателя, его ускорение зависит от подачи пара и внешней работы следующим образом:

$$\omega' = (k \cos(\varphi + \alpha) - F), \quad k, F > 0.$$

Здесь  $\varphi$  и  $\alpha$  — углы, показанные на рис. 13.3. Вторая производная  $\varphi$  определяется через центробежную силу, вес и трение по следующей формуле:

$$\varphi'' = \omega^2 \sin \varphi \cos \varphi - g \sin \varphi - b\varphi', \quad g, b > 0.$$

Найдите точку равновесия  $\varphi'' = \varphi' = \omega' = 0$  и определите, при каких условиях она устойчива (решение упрощается при  $\alpha = 0$ ).

Правильные решения следует посыпать по адресу: Джеймсу Уатту, знаменитому изобретателю паровой машины, Вестминстерское аббатство, 6HQ, 1FX, Лондон.

*Замечание.* Статья Гурвица (1895) появилась благодаря изучению похожей практической задачи, а именно о «...регулировании турбин курорта на водах Давос».

## I.14 Производные по параметрам и начальным значениям

«Для одного уравнения д-р Ритт решил указанную в заголовке задачу очень простым прямым методом... Доказательство д-ра Ритта нельзя непосредственно обобщить на системы уравнений».

(Т. Х. Гронуолл (1919))

В данном разделе мы рассмотрим вопрос о том, являются ли решения дифференциальных уравнений дифференцируемыми:

а) по начальным значениям;

б) по постоянным параметрам, входящим в уравнение,

и о том, как вычислить эти производные. Оба вопроса, конечно, чрезвычайно важны: если для заданных начальных значений решение уже найдено (численно), часто требуется узнать, как повлияют на него малые изменения этих начальных значений. Такой вопрос встает, например, в том случае, когда некоторые начальные значения не известны точно и должны быть определены из других условий, таких как фиксированные граничные условия. Кроме того, начальные значения могут содержать погрешности, влияние которых требуется изучить. Те же задачи возникают в том случае, когда в дифференциальные уравнения входят неизвестные или содержащие погрешности постоянные параметры.

Задачи а) и б) эквивалентны: пусть система дифференциальных уравнений

$$y' = f(x, y, p), \quad y(x_0) = y_0 \quad (14.1)$$

содержит параметр  $p$  (или несколько параметров). Мы можем расширить исходную систему:

$$\begin{pmatrix} y' \\ p' \end{pmatrix} = \begin{pmatrix} f(x, y, p) \\ 0 \end{pmatrix}, \quad \begin{matrix} y(x_0) = y_0, \\ p(x_0) = p. \end{matrix} \quad (14.1')$$

Тогда параметр станет начальным значением для уравнения  $p' = 0$ . И обратно: для системы

$$y' = f(x, y), \quad y(x_0) = y_0 \quad (14.2)$$

можно записать равенство  $y(x) = z(x) + y_0$ , из которого получим

$$z' = f(x, z + y_0) = F(x, z, y_0), \quad z(x_0) = 0. \quad (14.2')$$

В этом случае начальное значение становится параметром. Поэтому из двух задач а) и б) выберем сначала б) (как это сделал Гронуолл), так как она нам кажется более простой.

### Производная по параметру

Обычно задача содержит *несколько* параметров. Но поскольку нас интересуют частные производные, мы можем рассматривать параметры по очереди, сохраняя остальные фиксированными. Поэтому в излагаемой теории достаточно считать, что  $f(x, y, p)$  зависит только от *одного* скалярного параметра  $p$ .

Заменив параметр  $p$  в (14.1) на  $q$ , мы получим другое решение. Обозначим его через  $z(x)$ :

$$z' = f(x, z, q), \quad z(x_0) = y_0. \quad (14.3)$$

Тогда будет естественным вычесть (14.1) из (14.3) и линеаризовать разность:

$$\begin{aligned} z' - y' &= f(x, z, q) - f(x, y, p) = \\ &= \frac{\partial f}{\partial y}(x, y, p)(z - y) + \rho_1(z - y) + \\ &\quad + \frac{\partial f}{\partial p}(x, y, p)(q - p) + \rho_2(q - p). \end{aligned} \quad (14.4)$$

Если ввести обозначение  $(z(x) - y(x))/(q - p) = \psi(x)$  и отбросить остаточные члены, то мы придем к уравнению вида

$$\psi' = \frac{\partial f}{\partial y}(x, y(x), p)\psi + \frac{\partial f}{\partial p}(x, y(x), p), \quad \psi(x_0) = 0. \quad (14.5)$$

Оно является ключевым при решении поставленной задачи. Чтобы это уравнение имело смысл, нам необходимо сделать предположения о существовании входящих в него производных.

**Теорема 14.1.** [Гронуолл (1919)] Предположим, что при  $x_0 \leqslant x \leqslant X$  частные производные

$$\frac{\partial f}{\partial y} \text{ и } \frac{\partial f}{\partial p}$$

существуют и являются непрерывными в окрестности решения  $y(x)$ . Тогда существуют частные производные

$$\frac{dy(x)}{dp} = \psi(x),$$

причем они непрерывны и удовлетворяют дифференциальному уравнению (14.5).

*Доказательство.* Эта теорема породила знаменитую лемму Гронуолла (см. I.10, упр. 2). Докажем ее здесь с помощью эквивалентной теоремы 10.2. Введем обозначения

$$L = \max \left\| \frac{\partial f}{\partial y} \right\|, \quad A = \max \left\| \frac{\partial f}{\partial p} \right\|, \quad (14.6)$$

где  $\max$  взят по всей рассматриваемой области. Если считать  $z(x)$  приближенным решением задачи (14.1), то, используя первый раз теорему 10.2, мы получим для нормы невязки соотношение

$$\begin{aligned} \|z'(x) - f(x, z(x), p)\| &= \|f(x, z(x), q) - f(x, z(x), p)\| \leqslant \\ &\leqslant A |q - p|, \end{aligned}$$

из которого следует оценка

$$\|z(x) - y(x)\| \leqslant \frac{A}{L} |q - p| (e^{L(x-x_0)} - 1). \quad (14.7)$$

Поэтому при достаточно малом значении  $|q - p|$  и  $x_0 \leqslant x \leqslant X$  мы можем сделать  $\|z(x) - y(x)\|$  сколь угодно малой. По определению дифференцируемости и в силу оценки (14.7) для любого  $\varepsilon > 0$  существует такое  $\delta$ , при котором для остаточных членов в (14.4) выполняется неравенство

$$\|\rho_1(z - y) + \rho_2(q - p)\| \leqslant \varepsilon |q - p|, \text{ если } |q - p| < \delta. \quad (14.8)$$

(В действительности ситуация несколько сложнее: для оценок  $\|\rho_1\| < \varepsilon$  и  $\|\rho_2\| < \varepsilon$  величина  $\delta$  может зависеть от  $x$ , но в силу компактности и непрерывности ее можно заменить равномерной оценкой. Это небольшое препятствие можно было бы преодолеть и другим способом — ограничив вторые производные. Но почему мы должны заботиться о таких деталях? Сам Гронуолл их не упоминает.)

Рассмотрим теперь выражение  $(z(x) - y(x))/(q - p)$  как приближенное решение задачи (14.5) и применим теорему 10.2 во второй раз. В силу (14.8) и (14.4) норма невязки уравнения (14.5) ограничена величиной  $\varepsilon$ , а константа Липшица линейного дифференциального уравнения (14.5) также равна  $L$  (см. (11.2)). Поэтому из (10.14) мы получим оценку

$$\left\| \frac{z(x) - y(x)}{q - p} - \psi(x) \right\| \leqslant \frac{\varepsilon}{L} (e^{L(x-x_0)} - 1),$$

которую можно сделать сколь угодно малой; этим доказывается тот факт, что  $\psi(x)$  — производная функции  $y(x)$  по  $p$ .

*Непрерывность.* Частные производные  $dy/dp = \psi(x)$  являются решениями дифференциального уравнения (14.5), которое мы запишем в виде

$$\psi' = g(x, \psi, p),$$

где по предположению  $g$  непрерывно зависит от  $p$ . Поэтому непрерывная зависимость  $\psi$  от  $p$  снова следует из теоремы 10.2. ■

**Теорема 14.2.** Пусть  $y(x)$  является решением уравнения (14.1); рассмотрим матрицу Якоби

$$A(x) = \frac{\partial f}{\partial y}(x, y(x), p). \quad (14.9)$$

Предположим, что  $R(x, x_0)$  является резольвентой уравнения  $y' = A(x)y$  (см. (11.4)). Тогда решение уравнения (14.3) со слабо возмущенным параметром  $q$  задается формулой

$$z(x) = y(x) + (q - p) \int_{x_0}^x R(x, s) \frac{\partial f}{\partial p}(s, y(s), p) ds + o(|q - p|). \quad (14.10)$$

**Доказательство.** Это соотношение — формула вариации постоянных (11.12), примененная к (14.5). ■

Итак, мы видим, что чувствительность решений к изменениям параметров зависит, во-первых, от частных производных  $\partial f / \partial p$  (что вполне естественно), а во-вторых, от величины  $R(x, s)$ , т. е. от устойчивости дифференциального уравнения с матрицей (14.9).

### Производные по начальным значениям

**Обозначения.** Обозначим через  $y(x, x_0, y_0)$  решение  $y(x)$  в точке  $x$ , удовлетворяющее начальным значениям  $y(x_0) = y_0$ . Мы надеемся, что использование одной и той же буквы  $y$  для двух различных функций не вызовет никаких недоразумений.

Следующие тождества являются тривиальными следствиями определения или вытекают из соображений единственности, как в случае (11.6):

$$\frac{\partial y(x, x_0, y_0)}{\partial x} = f(x, y(x, x_0, y_0)), \quad (14.11)$$

$$y(x_0, x_0, y_0) = y_0, \quad (14.12)$$

$$y(x_2, x_1, y(x_1, x_0, y_0)) = y(x_2, x_0, y_0). \quad (14.13)$$

**Теорема 14.3.** Предположим, что частная производная  $f$  по  $y$  существует и непрерывна. Тогда решения  $y(x, x_0, y_0)$  дифференцируемы по  $y_0$ , и производная задается матрицей

$$\frac{\partial y(x, x_0, y_0)}{\partial y_0} = R(x, x_0), \quad (14.14)$$

где  $R$  — решение так называемого «уравнения в вариациях»

$$\frac{\partial R(x, x_0)}{\partial x} = \frac{\partial f}{\partial y}(x, y(x, x_0, y_0)) \cdot R(x, x_0),$$

$$R(x_0, x_0) = I. \quad (14.15)$$

*Доказательство.* Формулы (14.2) и (14.2') показывают, что обе производные  $\partial F/\partial z$  и  $\partial F/\partial y_0$  равны  $\partial f/\partial y$ , а поэтому в силу теоремы 14.1 производная  $\partial y/\partial y_0$  существует. Чтобы получить формулу (14.15), нам надо продифференцировать (14.11) и (14.12) по  $y_0$ . ■

В завершение мы вычислим производную  $y(x, x_0, y_0)$  по  $x_0$ .

**Теорема 14.4.** При условиях теоремы 14.3 решения дифференцируемы также и по  $x_0$ , а производная вычисляется по формуле

$$\frac{\partial y(x, x_0, y_0)}{\partial x_0} = -R(x, x_0) \cdot f(x_0, y_0). \quad (14.16)$$

*Доказательство.* Продифференцируем по  $x_0$  тождество

$$y(x_1, x_0, y(x_0, x_1, y_1)) = y_1,$$

которое следует из (14.13), и применим (14.14) и (14.11) (см. упр. 3). ■

### Нелинейная формула вариации постоянных

Завершим данный раздел обобщением теоремы 11.2 на системы нелинейных дифференциальных уравнений.

**Теорема 14.5.** [Алексеев (1961), Грёбнер (1960)] Обозначим через  $y$  и  $z$  решения систем

$$y' = f(x, y), \quad y(x_0) = y_0, \quad (14.17a)$$

$$z' = f(x, z) + g(x, z), \quad z(x_0) = y_0 \quad (14.17b)$$

соответственно и предположим, что  $\partial f/\partial y$  существует и является непрерывной. Тогда решения уравнения (14.17a) и возмущенного уравнения (14.17b) связаны соотношением

$$z(x) = y(x) + \int_{x_0}^x \frac{\partial y}{\partial z}(x, s, z(s)) g(s, z(s)) ds. \quad (14.18)$$

*Доказательство.* Выберем разбиение  $x_0 = s_0 < s_1 < s_2 < \dots < s_N = x$  (см. рис. 14.1). Ниспадающие кривые представляют решения невозмущенного уравнения (14.17a) с начальными значениями  $s_i, z(s_i)$ . Разности  $d_i$ , возникающие вследствие различных наклонов  $z(s)$  и  $y(s)$  [из (14.17b)] надо вычесть (14.17a)], равны  $d_i = g(s_i, z(s_i)) \cdot \Delta s_i + o(\Delta s_i)$ . Эта «погрешность» в точке  $s_i$

«переносится» затем на конечное значение  $x$ , составляя в точке  $x$  величину, определенную теоремой 14.3. В итоге имеем

$$D_i = \frac{\partial y}{\partial z} (x, s_i, z(s_i)) \cdot g(s_i, z(s_i)) \cdot \Delta s_i + o(\Delta s_i). \quad (14.19)$$

Поскольку

$$z(x) - y(x) = \sum_{i=1}^N D_i,$$

в результате подстановки (14.19) и перехода к пределу  $\Delta s_i \rightarrow 0$  мы получим интеграл в (14.18). ■

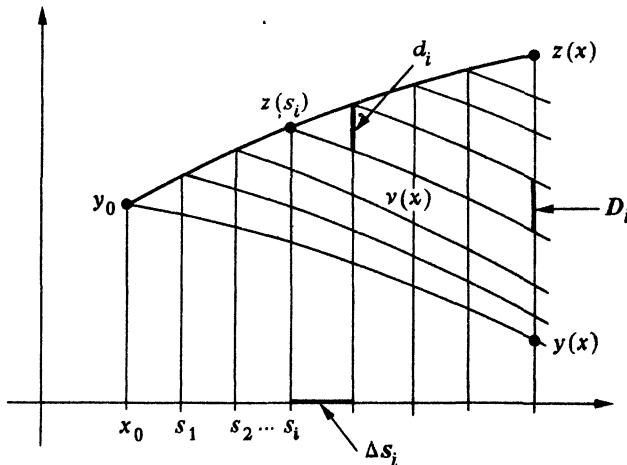


Рис. 14.1. Веер леди Уиндермир, действие второе.

Если нам нужно учесть также возможное различие в начальных значениях, мы можем сформулировать

**Следствие 14.6.** Пусть  $y(x)$  и  $z(x)$  являются решениями уравнений

$$y' = f(x, y), \quad y(x_0) = y_0,$$

$$z' = f(x, z) + g(x, z), \quad z(x_0) = z_0.$$

Тогда выполняется соотношение

$$\begin{aligned} z(x) &= y(x) + \int_0^1 \frac{\partial y}{\partial z} (x, x_0, y_0 + s(z_0 - y_0)) \cdot (z_0 - y_0) ds + \\ &+ \int_{x_0}^x \frac{\partial y}{\partial z} (x, s, z(s)) \cdot g(s, z(s)) ds. \quad ■ \end{aligned} \quad (14.20)$$

Эти две теоремы позволяют получить множество оценок устойчивости общих нелинейных систем. В случае *линейных* систем  $\frac{dy}{dz}(x, s, z)$  не зависит от  $z$ , и формулы (14.20), (14.17) сводятся к формуле вариации постоянных (11.12). Кроме того, оценивая сверху тривиальным образом интегралы в (14.20), мы приедем к фундаментальной лемме (10.14), а также ко второй формулировке теоремы 10.2.

### Упражнения

1. Изучите зависимость от начальных значений в задаче

$$y' = y^2, \quad x_0 = 0, \quad y_0 = 1, \quad X = 0.999. \quad (14.21)$$

2. Решите численно уравнение (14.21), используя 999 шагов метода Эйлера (7.3) при  $x_{i+1} - x_i = 0.001$ . Оцените влияние на конечный результат погрешности на первом шаге (равной  $10^{-6}$ ) и объясните низкую точность.
3. Докажите следующую лемму элементарного анализа, примененную в доказательстве теоремы 14.4. Если для функции  $F(x, y)$  существует  $\partial F / \partial y$ , а  $y(x)$  дифференцируема и отвечает условию

$$F(x, y(x)) = \text{const}, \quad (14.22)$$

то в точке  $(x, y(x))$  существует производная  $\partial F / \partial x$ , которая вычисляется по формуле

$$\frac{\partial F}{\partial x}(x, y(x)) = -\frac{\partial F}{\partial y}(x, y(x)) \cdot y'(x).$$

*Указание.* Из (14.22) выведите тождество

$$F(x_1, y(x_1)) - F(x_0, y(x_1)) = F(x_0, y(x_0)) - F(x_0, y(x_1)).$$

## I.15. Краевые задачи и задачи на собственные значения

Хотя наша книга в основном посвящена задачам с начальными условиями, или задачам Коши, мы хотим в этой первой главе осветить также некоторые свойства краевых задач и задач на собственные значения.

### Краевые задачи

Такие задачи возникают в системах дифференциальных уравнений, например:

$$\begin{aligned}y'_1 &= f_1(x, y_1, y_2), \\y'_2 &= f_2(x, y_1, y_2),\end{aligned}\quad (15.1)$$

когда нет начальной точки  $x_0$ , в которой были бы известны одновременно  $y_1(x_0)$  и  $y_2(x_0)$ . В этом случае вопросы существования и единственности становятся намного сложнее.

*Пример 1.* Рассмотрим дифференциальное уравнение

$$y'' = \exp(y) \text{ или } y'_1 = y_2, \quad y'_2 = \exp(y_1) \quad (15.2a)$$

с граничными условиями

$$y_1(0) = a, \quad y_1(1) = b. \quad (15.2b)$$

Чтобы применить доказанные нами теоремы существования или провести численные расчеты [например, с помощью метода Эйлера (7.3)], мы можем поступить следующим образом: предположим, что недостающее начальное значение равно  $y_{20}$ . Тогда можно вычислить решение и проверить, равно ли найденное значение для  $y_1(1)$  числу  $b$ . Таким образом, наша задача состоит в том, чтобы определить, есть ли нуль у функции одной переменной

$$F(y_{20}) := y_1(1) - b. \quad (15.3)$$

Уравнение (15.2a) является *квазимонотонным*, что означает монотонную зависимость  $F(y_{20})$  от  $y_{20}$  (рис. 15.1a, см. упр. 3 в разд. I.10). Кроме того, при очень малых или очень больших значениях  $y_{20}$  величина  $y_1(1)$  сколь угодно мала или велика или даже обращается в бесконечность. Поэтому (15.2) *при всех*  $a, b$  имеет единственное решение (см. рис. 15.1b).

Корень уравнения  $F(y_{20}) = 0$  можно вычислить итерационным методом (методом деления отрезка пополам, методом ложного положения и т. д.; а если производную  $y_1(1)$  по  $y_{20}$  найти из тео-

ремы 14.3 или определить численно с помощью конечных разностей, то также методом Ньютона). Тогда задачу Коши нужно решать несколько раз. Для небольших задач, таких, как в приведенном выше примере, это можно просто сделать на компьютере в диалоговом режиме. Более сложные задачи с большим числом неизвестных начальных значений требуют более высокой техники программирования. Описанный метод принадлежит к наиболее широко распространенным методам и называется *методом стрельбы* (или *пристрелки*).

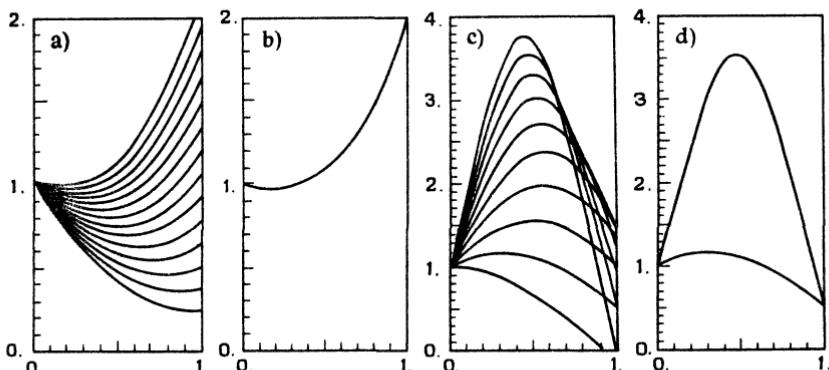


Рис. 15.1. a) Решения уравнения (15.2а) при различных начальных значениях  $y_{20} = -1.7, \dots, -0.4$ .

b) Единственное решение уравнения (15.2а) при  $a = 1, b = 2, y_{20} = -0.476984656$ .

c) Решения уравнения (15.4а) при  $y(0) = 1$  и  $y_{20} = 1, 2, \dots, 8$ .

d) Два решения уравнения (15.4а),  $y(0) = 1, y(1) = 0.5, y_{20} = 7.93719, y_{20} = 0.97084$ .

*Пример 2.* Для дифференциального уравнения

$$y'' = -\exp(y) \text{ или } y'_1 = y_2, y'_2 = -\exp(y_1) \quad (15.4a)$$

с граничными условиями

$$y_1(0) = a, \quad y_1(1) = b. \quad (15.4b)$$

$F(y_{20})$  уже не удовлетворяет требованию монотонности, что осложняет ситуацию: на рис. 15.1с изображены решения при  $a = 1$  и различных начальных значениях  $y_{20}$ . Из рисунка можно увидеть, что если  $b$  расположено выше некоторого значения (равного 1.499719998), то задача *совсем не имеет* решений, а при меньших  $b$  существует *два* решения (рис. 15.1д).

*Пример 3.*

$$\dot{y}_1 = y_2, \quad \dot{y}_2 = y_1^3, \quad y_1(0) = 1, \quad y_1(100) = 2. \quad (15.5)$$

Это уравнение похоже на (15.2), и для него справедливо то же, что и прежде, утверждение о существовании и единственности решения. Однако если мы попытаемся найти решения методом стрельбы, то окажемся в затруднении из-за длины интервала: *решение почти никогда не существует на всем интервале*. В действительности правильное решение равно  $y_{20} = -0.70710616655$ . Но уже при  $y_{20} = -0.7071061$   $y_1(x)$  стремится к  $+\infty$  при  $x \rightarrow \rightarrow 98.2$ . С другой стороны, при  $y_{20} = -0.70711$  мы имеем  $y_1(94.1) = = -\infty$ . Таким образом, область *существования*  $F(y_{20})$  из (15.3) по длине меньше чем  $4 \times 10^{-6}$ .

В подобных случаях используют *метод параллельной стрельбы*: отрезок разбивают на *несколько* подотрезков, на каждом из которых находят решение при хорошем выборе начальных значений. В конечных точках подотрезков производится согласование решений. Таким образом, уравнение (15.3) превращается в систему большей размерности. Параллельной стрельбе посвящено множество публикаций. В качестве общей ссылки назовем работу Дойфлхарда (1980).

### Задачи Штурма — Лиувилля на собственные значения

Начало изучению данного вопроса положила замечательная статья Штурма [Штурм (1836)] в журнале, незадолго до этого основанном Лиувиллем. Ее продолжил ряд статей Лиувилля и Штурма, опубликованных в последующих номерах. В настоящее время эти работы считаются отправной точкой «геометрической теории», где главные усилия направлены не на интегрирование уравнения, а только на получение *прямо из дифференциального уравнения* геометрических свойств решения, таких как его форма, осцилляции, изменения знака, нули, существование максимумов и минимумов и т. д. («Однако к этой цели можно прийти, рассматривая лишь сами дифференциальные уравнения, и нет никакой необходимости их интегрировать».)

Физические задачи, которые привели к такой постановке проблемы, как в разд. I.6, связаны с изучением теплоты и малых колебаний упругой среды. Рассмотрим уравнение теплопроводности с переменным коэффициентом теплопроводности:

$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial x} \left( k(x) \frac{\partial u}{\partial x} \right) - l(x)u, \quad k(x) > 0, \quad (15.6)$$

которое подробно изучено в «Теории теплоты» Пуассона. Этот автор предположил, что  $u(x, t) = y(x) e^{-\lambda t}$ . Тогда (15.6) принимает вид

$$\frac{d}{dx} \left( k(x) \frac{dy}{dx} \right) - l(x)y = -\lambda y. \quad (15.7)$$

Перепишем (15.7) в виде

$$(k(x)y')' + G(x)y = 0 \quad (15.8)$$

и сформулируем теорему сравнения Штурма.

**Теорема 15.1.** Рассмотрим дополнительно к (15.8) дифференциальное уравнение

$$(\hat{k}(x)\hat{y}')' + \hat{G}(x)\hat{y} = 0 \quad (15.9)$$

и предположим, что  $k, \hat{k}$  дифференцируемы,  $G, \hat{G}$  непрерывны и при всех  $x$  выполняются соотношения

$$0 < \hat{k}(x) \leq k(x), \quad \hat{G}(x) \geq G(x). \quad (15.10)$$

Пусть  $y(x), \hat{y}(x)$  являются линейно независимыми решениями уравнений (15.8) и (15.9) соответственно. Тогда между любыми двумя нулями  $y(x)$  находится по крайней мере один нуль  $\hat{y}(x)$ , т. е. если  $y(x_1) = y(x_2) = 0$  и  $x_1 < x_2$ , то на открытом интервале  $(x_1, x_2)$  существует такое  $x_3$ , что  $\hat{y}(x_3) = 0$ .

**Доказательство.** Доказательство самого Штурма основано на отношении

$$q(x) = \frac{y(x)}{k(x)y'(x)},$$

которое равно тангенсу угла наклона прямой, соединяющей начало координат с точкой решений на плоскости  $(ky', y)$ , и удовлетворяет дифференциальному уравнению первого порядка. Чтобы избежать сингулярностей, вызванных обращением  $y'(x)$  в нуль, мы предпочтетм пользоваться полярными координатами [Прюфера (1926)]:

$$k(x)y'(x) = \rho(x) \cos \varphi(x), \quad y(x) = \rho(x) \sin \varphi(x). \quad (15.11)$$

Дифференцируя (15.11), мы придем к следующему дифференциальному уравнению для  $\varphi$ :

$$\varphi' = \frac{1}{k(x)} \cos^2 \varphi + G(x) \sin^2 \varphi. \quad (15.12)$$

Тем же способом мы введем функции  $\hat{\rho}(x)$  и  $\hat{\varphi}(x)$  для второго дифференциального уравнения (15.9) и получим

$$\hat{\varphi}' = \frac{1}{\hat{k}(x)} \cos^2 \hat{\varphi} + \hat{G}(x) \sin^2 \hat{\varphi}. \quad (15.13)$$

Теперь предположим, что  $x_1, x_2$  — два последовательных нуля  $y(x)$ . Тогда  $\varphi(x_1)$  и  $\varphi(x_2)$  должны быть кратны  $\pi$ , так как  $\rho(x)$  всегда отлична от нуля (единственность в задаче Коши). В силу

(15.12)  $\varphi'(x)$  положительна в  $x_1$  и  $x_2$ . Поэтому мы можем считать, что

$$\varphi(x_1) = 0, \varphi(x_2) = \pi, \hat{\varphi}(x_1) \in [0, \pi]. \quad (15.14)$$

Вследствие того, что уравнения (15.12) и (15.13) имеют первый порядок, а также в силу неравенств (15.10) можно применить теорему 10.3, которая приведет к соотношению

$$\hat{\varphi}(x) \geq \varphi(x) \text{ при } x_1 \leq x \leq x_2.$$

Равенство  $\hat{\varphi}(x) = \varphi(x)$  не может выполняться всюду, так как в силу (15.12) и (15.13) мы получили бы тогда соотношения  $k(x) = \hat{k}(x)$ ,  $G(x) = \hat{G}(x)$  и линейную зависимость решений  $y(x)$ ,  $\hat{y}(x)$ . Следовательно, существует такое  $x_0 \in (x_1, x_2)$ , при котором  $\varphi(x_0) > \hat{\varphi}(x_0)$ . В этом случае  $\hat{\varphi}(x) > \varphi(x)$  при всех  $x \geq x_0$ , что обеспечивает существование точки  $x_3 \in (x_1, x_2)$ , в которой  $\varphi(x_3) = \pi$ . ■

Следующая теорема показывает, что у нашей задачи на собственные значения имеется бесконечное множество решений. Добавим к (15.7) граничные условия

$$y(x_0) = y(x_1) = 0. \quad (15.15)$$

**Теорема 15.2.** Задача на собственные значения (15.7), (15.15) имеет бесконечную последовательность собственных значений  $\lambda_1 < \lambda_2 < \lambda_3 < \dots$ , которым отвечают решения  $y_j(x)$  («собственные функции») с  $0, 1, 2, \dots$  нулями соответственно внутри интервала  $(x_0, x_1)$ . Нули  $y_j(x)$  разделяются нулями  $y_{j+1}(x)$ . Если  $0 < K_1 \leq k(x) \leq K_2$  и  $L_1 \leq l(x) \leq L_2$ , то справедливо соотношение

$$L_1 + K_1 \frac{j^2\pi^2}{(x_1 - x_0)^2} \leq \lambda_j \leq L_2 + K_2 \frac{j^2\pi^2}{(x_1 - x_0)^2}. \quad (15.16)$$

**Доказательство.** Пусть  $y(x, \lambda)$  является решением (15.7) с начальными значениями  $y(x_0) = 0$ ,  $y'(x_0) = 1$ . Из теоремы 15.1 (при  $\hat{k}(x) = k(x)$ ,  $\hat{G}(x) = G(x) + \Delta\lambda$ ) следует, что с ростом  $\lambda$  нули  $y(x, \lambda)$  перемещаются к  $x_0$ , а поэтому число нулей на  $(x_0, x_1)$  представляет собой неубывающую функцию от  $\lambda$ .

Сравнивая теперь (15.7) с решением

$$\sin(\sqrt{(\lambda - L_1)/K_1}(x - x_0))$$

уравнения  $K_1 y'' + (\lambda - L_1) y = 0$  при ( $\lambda > L_1$ ), мы увидим, что  $y(x, \lambda)$  при  $\lambda < L_1 + K_1 j^2\pi^2/(x_1 - x_0)^2$  имеет на  $(x_0, x_1]$  не более  $j - 1$  нулей. Подобным образом, сравнение  $y(x, \lambda)$  с функцией

$$\sin(\sqrt{(\lambda - L_2)/K_2}(x - x_0)),$$

которая служит решением уравнения  $K_2 y'' + (\lambda - L_2) y = 0$ , показывает, что при  $\lambda > L_2 + K_2 j^2\pi^2/(x_1 - x_0)^2$  функция  $y(x, \lambda)$

имеет по крайней мере  $j$  нулей на  $(x_0, x_1)$ . Теперь утверждения теоремы получаются как простые следствия этих трех свойств. ■

*Пример.* На рис. 15.2 показаны 5 первых решений задачи

$$((1 - 0.8 \sin^2 x) y')' - (x - \lambda) y = 0, \quad y(0) = y(\pi) = 0. \quad (15.17)$$

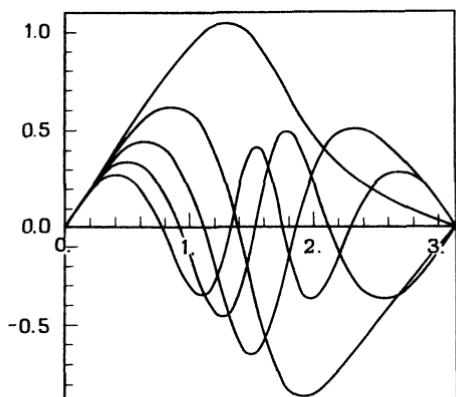


Рис. 15.2. Решения задачи Штурма—Лиувилля (задачи на собственные значения) (15.17).

Первые собственные значения равны 2.1224, 3.6078, 6.0016, 9.3773, 13.7298, 19.053, 25.347, 32.609, 40.841, 50.041 и т. д.

С этой теорией, представляющей очень важную страницу истории математики, более обстоятельно можно познакомиться в книге Рейда (1980).

### Упражнения

1. Для уравнения

$$L(x)y'' + M(x)y' + N(x)y = 0$$

подберите такую функцию  $\varphi(x)$ , чтобы полученное в результате умножения на нее уравнение имело вид (15.7) [Штурм (1836)].

2. Докажите, что два решения задачи (15.7), (15.15) удовлетворяют условию ортогональности

$$\int_{x_0}^{x_1} y_j(x) y_k(x) dx = 0 \quad \text{при } \lambda_j \neq \lambda_k.$$

*Указание.* Умножьте это соотношение на  $\lambda_j$ , замените  $\lambda_j y_j(x)$  выражением из левой части (15.7) и выполните интегрирование по частям [Лиувиль (1836)].

3. Найдите решение задачи (15.5) в элементарных функциях. Объясните, почему указанное для  $y_{20}$  значение столь близко к  $-\sqrt{2}/2$ .

4. (Коллатц) Покажите, что краевая задача

$$y'' = y^3, \quad y(0) = 0, \quad y(A) = B \quad (15.18)$$

имеет бесконечно много решений для каждой пары  $(A, B)$  при  $A \neq 0$ .

*Указание.* Получите решение  $y(x)$  задачи (15.18) при  $y(0) = 0$ ,  $y'(0) = 1$ . Покажите, что для любой константы  $a$  функция  $z(x) = ay(ax)$  также будет решением.

## I.16. Периодические решения, предельные циклы, странные атTRACTоры

Явление предельных циклов впервые было теоретически описано Пуанкаре (1882) и Бендикссоном (1901) и с тех пор нашло много применений в физике, химии и биологии.

### Уравнение Ван-дер-Поля

«У меня есть теория: если вы хотите какой-то метод скомпрометировать—ищите решение уравнения Ван-дер-Поля».

(П. Э. Задунайский, 1982)

Первые примеры такого типа из практики были изучены Релеем (1883), а позднее Ван-дер-Полем (1920—1926) в серии статей по нелинейным осцилляциям: решения уравнения

$$y'' + \alpha y' + y = 0$$

являются затухающими при  $\alpha > 0$  и неустойчивыми при  $\alpha < 0$ . Идея состоит в изменении  $\alpha$  (например, с помощью включенного в цепь триода) таким образом, чтобы при малых  $y$  выполнялось  $\alpha < 0$ , а при больших  $y$  — неравенство  $\alpha > 0$ . В качестве простейшего выражения, описывающего физическую ситуацию в несколько идеализированной форме, можно взять  $\alpha = \varepsilon(y^2 - 1)$ ,  $\varepsilon > 0$ . Тогда данное уравнение примет вид

$$y'' + \varepsilon(y^2 - 1)y' + y = 0, \quad (16.1)$$

или, если записать его как систему,

$$\begin{aligned} y'_1 &= y_2, \\ y'_2 &= \varepsilon(1 - y_1^2)y_2 - y_1, \quad \varepsilon > 0. \end{aligned} \quad (16.2)$$

В этом уравнении малые колебания усиливаются, а большие — затухают. Поэтому можно ожидать, что существует устойчивое периодическое решение, к которому сходятся все остальные решения. Его называют *предельным циклом*. Иллюстрации из статьи Ван-дер-Поля в их первоначальном виде воспроизведены на рис. 16.1.

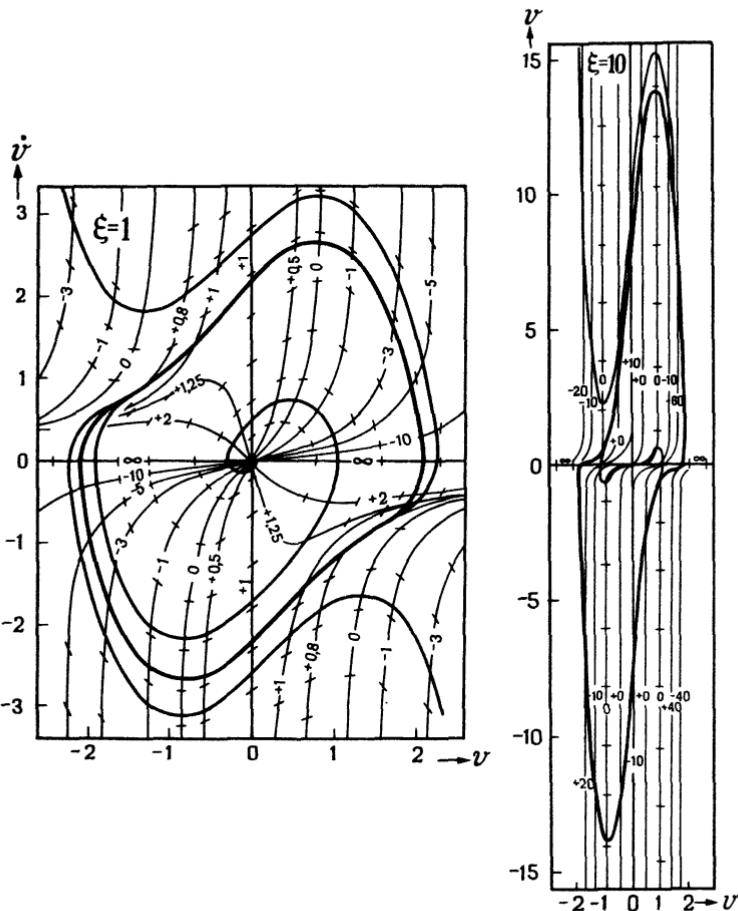


Рис. 16.1. Иллюстрации из работы Ван-дер-Поля (1926).

### Доказательство существования

Существование предельных циклов изучается с помощью *метода сечений Пуанкаре*. Его идея состоит в том, чтобы пересечь решения гиперплоскостью  $P$  и при начальном значении  $y_0 \in P$  рассмотреть первую точку  $\Phi(y_0)$ , в которой решение снова пересекает  $P$  в том же направлении.

Для нашего примера (16.2) выберем в качестве  $P$  луч  $y_2 = 0$ ,  $y_1 > 0$  и исследуем знаки производных  $y'_1$  и  $y'_2$  в (16.2). Знак  $y'_2$  меняется на кривой

$$y_2 = \frac{y_1}{\varepsilon(1-y_1^2)}, \quad (16.3)$$

которая на рис. 16.2 изображена штриховой линией. Отсюда следует (см. рис. 16.2), что  $\Phi(y_0)$  существует для всех  $y_0 \in P$ . Поскольку два различных решения не могут пересекаться вследствие единственности, отображение  $\Phi$  является монотонным. Кроме того,  $\Phi$  ограничено [например, любым решением, начинаящимся на кривой (16.3)], а поэтому  $\Phi(y_0) < y_0$  при больших  $y_0$ . И наконец, в силу неустойчивости начала координат  $\Phi(y_0) > y_0$  при малых  $y_0$ . Следовательно, должна существовать неподвижная точка отображения  $\Phi(y_0)$ , т. е. предельный цикл. ■

**Замечание.** На самом деле предельный цикл является единственным. Этот факт доказывается сложнее; идею доказательства мы укажем ниже, в упр. 8 [Льенар (1928)].

С помощью подобных рассуждений доказывается следующий общий результат.

**Теорема 16.1.** [Пуанкаре (1882), Бендикуссон (1901)] Любое ограниченное решение системы размерности 2

$$y'_1 = f_1(y_1, y_2), \quad y'_2 = f_2(y_1, y_2) \quad (16.4)$$

должно:

- i) либо стремиться к критической точке  $f_1 = f_2 = 0$  для бесконечного множества точек  $x_i \rightarrow \infty$ ;
- ii) либо быть периодическим;
- iii) либо стремиться к предельному циклу. ■

**Замечание.** Приводимое ниже упр. 1 объясняет, почему первая возможность записана в несколько более сложной форме, чем это кажется необходимым.

### Стационарные приближения при больших $\varepsilon$

Стационарные приближения представляют собой важное средство упрощения сложных нелинейных систем. Рассмотрим (16.2) при очень большом  $\varepsilon$ . Тогда в окрестности кривой  $f_2(y_1, y_2) = 0$  производная  $y'_2 = f_2$  по  $y_2$  имеет очень большое по модулю отри-

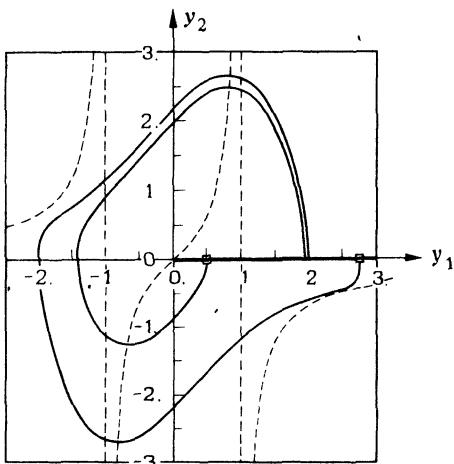


Рис. 16.2. Отображение Пуанкаре для уравнения Ван-дер-Поля,  $\varepsilon = 1$ .

цательное значение. Поэтому решение будет очень быстро приближаться к положению равновесия в окрестности кривой  $y_2' = -f_2(y_1, y_2) = 0$ , которая в нашем примере задана формулой (16.3). Этую формулу можно подставить в (16.2), что даст нам уравнение более низкой размерности

$$y_1' = \frac{y_1}{\varepsilon(1-y_1^2)}. \quad (16.5)$$

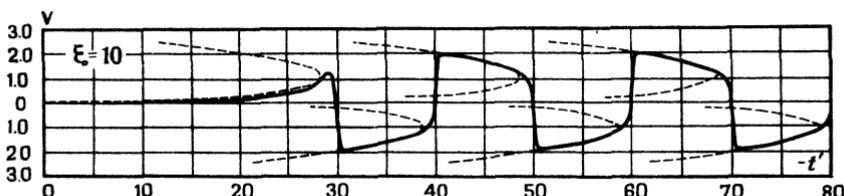


Рис. 16.3. Решение уравнения Ван-дер-Поля при  $\varepsilon = 10$  в сравнении со стационарными приближениями.

Оно легко решается с помощью формул из разд. I.3, и в результате мы получаем

$$\log(y_1) - \frac{y_1^2}{2} = \frac{x - x_0}{\varepsilon} + \text{const.}$$

Эти кривые показаны пунктиром на рис. 16.3 из работы Ван-дер-Поля, где они соответствуют значению  $\varepsilon = 10$  и дают, как видно, хорошее приближение к решению.

### Асимптотические решения при малых $\varepsilon$

Основы метода вычисления периодических решений при малых параметрах заложили астрономы, в частности Ньюкомб и Линдстедт, а в работе Пуанкаре (1893) он был доведен до совершенства. Продемонстрируем метод малого параметра на примере уравнения Ван-дер-Поля (16.1). Идея состоит в том, чтобы искать решение в виде ряда по степеням  $\varepsilon$ . Поскольку период также будет изменяться, сделаем замену переменной

$$t = x(1 + \gamma_1\varepsilon + \gamma_2\varepsilon^2 + \dots) \quad (16.6)$$

и положим

$$y(x) = z(t) = z_0(t) + \varepsilon z_1(t) + \varepsilon^2 z_2(t) + \dots \quad (16.7)$$

Теперь, подставив в (16.1)  $y'(x) = z'(t)(1 + \gamma_1\varepsilon + \dots)$ ,  $y''(x) = z''(t)(1 + \gamma_1\varepsilon + \dots)^2$ , мы получим уравнение

$$\begin{aligned} & (\ddot{z}_0 + \varepsilon \ddot{z}_1 + \varepsilon^2 \ddot{z}_2 + \dots)(1 + 2\gamma_1\varepsilon + (2\gamma_2 + \gamma_1^2)\varepsilon^2 + \dots) + \\ & + \varepsilon((z_0 + \varepsilon z_1 + \dots)^2 - 1)(z_0' + \varepsilon z_1' + \dots)(1 + \gamma_1\varepsilon + \dots) + \\ & + (z_0 + \varepsilon z_1 + \varepsilon^2 z_2 + \dots) = 0. \end{aligned} \quad (16.8)$$

Сравнивая коэффициенты при  $\varepsilon^0$ , получим уравнение

$$z_0'' + z_0 = 0. \quad (16.8; 0)$$

Фиксируем произвольно начальное значение  $z'(0) = 0$ , тогда  $z_0 = A \cos t$ . Далее, для коэффициентов при  $\varepsilon$  получим

$$\begin{aligned} z_1'' + z_1 &= -2\gamma_1 z_0'' - (z_0^2 - 1) z_1' = \\ &= 2\gamma_1 A \cos t + \left( \frac{A^3}{4} - A \right) \sin t + \frac{A^3}{4} \sin 3t. \end{aligned} \quad (16.8; 1)$$

Главная идея здесь в том, что, поскольку мы ищем *периодические* решения, из правой части (16.8; 1) надо удалить члены, содержащие  $\cos t$  и  $\sin t$ , чтобы избежать появления в  $z_1(t)$  членов вида  $t \cdot \cos t$  и  $t \cdot \sin t$  («...и таким образом убрать члены, называемые вековыми...»). Отсюда мы получим значения  $\gamma_1 = 0$  и  $A = 2$ . Тогда можно найти решение (16.8; 1), которое при начальном условии  $z_1(0) = 0$  равно

$$z_1 = B \cos t + \frac{3}{4} \sin t - \frac{1}{4} \sin 3t. \quad (16.9)$$

Теперь ясно, как следует продолжить этот процесс: после подстановки (16.9) и упрощения члены при  $\varepsilon^2$  в (16.8) дают уравнение

$$\begin{aligned} z_2'' + z_2 &= \left( 2\gamma_2 + \frac{1}{4} \right) \cos t + 2B \sin t + 3B \sin 3t \\ &\quad - \frac{3}{2} \cos 3t + \frac{5}{4} \cos 5t. \end{aligned} \quad (16.8; 2)$$

Чтобы избежать вековых членов, надо взять  $B = 0$  и  $\gamma_2 = -1/8$ , и тогда

$$z_2 = C \cos t + \frac{3}{16} \cos 3t - \frac{5}{96} \cos 5t.$$

На следующем этапе будут определены значения  $C = -1/8$  и  $\gamma_s = 0$ , и т. д. В итоге мы придем к следующему результату: *периодическая траектория уравнения Ван-дер-Поля (16.1) при малых  $\varepsilon$  задается формулами*

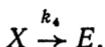
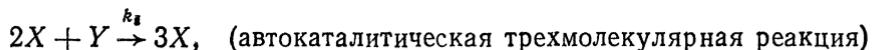
$$\begin{aligned} y(x) &= z(t), \quad t = x(1 - \varepsilon^2/8 + \dots), \\ z(t) &= 2 \cos t + \varepsilon \left( \frac{3}{4} \sin t - \frac{1}{4} \sin 3t \right) + \\ &\quad + \varepsilon^2 \left( -\frac{1}{8} \cos t + \frac{3}{16} \cos 3t - \frac{5}{96} \cos 5t \right) + \dots \end{aligned} \quad (16.10)$$

и имеет период  $2\pi(1 + \varepsilon^2/8 + \dots)$ .

## Химические реакции

Законы химической кинетики порождают дифференциальные уравнения, которые для реакций с участием двух и более молекул становятся нелинейными и обладают интересными свойствами. Некоторые из этих уравнений имеют периодические решения (например, в случае реакции Белоусова—Жаботинского), они находят также важные применения для объяснения биологических явлений (см., например, работу Пригожина, Лефевера (1969)).

Рассмотрим подробно модель Лефевера и Николиса (1971), так называемый «брюсселятор»: предположим, что шесть веществ,  $A, B, D, E, X, Y$ , участвуют в следующих реакциях:



Если обозначить через  $A(x), B(x), \dots$  концентрации веществ  $A, B, \dots$  как функции времени  $x$ , то реакции (16.11), согласно закону действующих масс, описываются следующими дифференциальными уравнениями:

$$A' = -k_1 A,$$

$$B' = -k_2 BX,$$

$$D' = k_2 BX,$$

$$E' = k_4 X,$$

$$X' = k_1 A - k_2 BX + k_3 X^2 Y - k_4 X,$$

$$Y' = k_2 BX - k_3 X^2 Y.$$

Теперь упростим эту систему: исключим из рассмотрения уравнения для  $D$  и  $E$ , так как они не влияют на остальные; предположим, что  $A$  и  $B$  поддерживаются постоянными (положительными) и возьмем все скорости реакций  $k_i$  равными единице. Далее, введем обозначения  $X(x) =: y_1(x)$ ,  $Y(x) =: y_2(x)$  и в результате получим систему

$$\begin{aligned} y_1' &= A + y_1^2 y_2 - (B + 1) y_1, \\ y_2' &= B y_1 - y_1^2 y_2. \end{aligned} \quad (16.11')$$

Она имеет одну особую точку  $y_1' = y_2' = 0$  при  $y_1 = A$ ,  $y_2 = B/A$ . В окрестности этой точки линеаризованное уравнение неустойчиво тогда и только тогда, когда  $B > A^2 + 1$ . Далее, изучая об-

ласти, где  $y_1$ ,  $y_2$  или  $(y_1 + y_2)'$  положительны или отрицательны, мы придем к выводу, что все решения остаются ограниченными. Таким образом, при  $B > A^2 + 1$  должен существовать предельный цикл, который, как видно из численных расчетов, является единственным (рис. 16.4).

Когда  $B$  приближается к  $A^2 + 1$ , наблюдается интересное явление (бифуркация Хопфа). Оно состоит в том, что предельный цикл становится все меньше и меньше и, наконец, исчезает в критическую точку. Другой пример такого типа приведен в упр. 2.

### Предельные циклы в системах больших размерностей, бифуркация Хопфа

Теорема Пуанкаре—Бендиксона, очевидно, верна только в том случае, когда размерность системы равна двум. Контрпримером при большей размерности может служить почти каждое механическое движение без трения, как, например, сферический маятник (рис. 16.5), к изучению которого обращались многие ученые прошлого (см., например, работу Лагранжа (1788), вторая часть, разд. VIII, гл. II, § 1). В этом случае уравнения движения в сферических координатах принимают вид

$$\begin{aligned}\varphi'' &= -2 \frac{\cos \psi}{\sin \psi} \cdot \varphi' \cdot \psi', \\ \psi'' &= \sin \psi \cos \psi (\varphi')^2 - \sin \psi,\end{aligned}\tag{16.12}$$

а кривая решения представлена на рис. 16.5. Если увеличение  $\varphi/(2\pi)$  за одну осцилляцию — иррациональное число, то решение не является периодическим и не стремится к периодическому или стационарному решению.

Поэтому в случае больших размерностей предельные циклы обычно находят с помощью численного исследования определенного выше отображения Пуанкаре  $\Phi$  в методе сечений. Периодическое решение (т. е. неподвижная точка отображения  $\Phi$ ) называется аттрактором, если все собственные значения  $\Phi'(y_0)$  по модулю меньше 1. Все решения, начинающиеся в окрестности аттрактора, будут стремиться к нему.

Однако существует одна ситуация, в которой предельные циклы возникают совершенно естественно [Хопф (1942)], а именно,

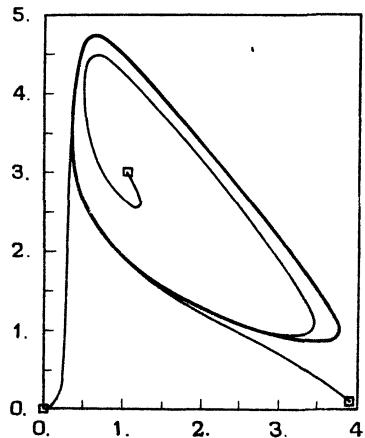


Рис. 16.4. Решения уравнений «брюс селектора»,  $A = 1$ ,  $B = 3$ .

когда в особой точке системы  $y' = f(y)$ ,  $y, f \in \mathbb{R}^n$ , все собственные значения  $f'(y_0)$  имеют строго отрицательную вещественную часть, за исключением одной пары, пересекающей мнимую ось. При этом подпространство, натянутое на собственные векторы, отвечающие устойчивым собственным значениям, продолжается в двумерное аналитическое многообразие, внутри которого появляется предельный цикл. Это явление называется «бифуркацией Хопфа». Доказательство этого факта аналогично методу Пуанкаре разложения по малому параметру (16.7) [см. ниже упр. 6 и 7], и поэтому Хопф даже сомневался, стоит ли его публиковать («...мне трудно поверить, что в приведенном выше предложении содержится что-либо существенно новое...»).

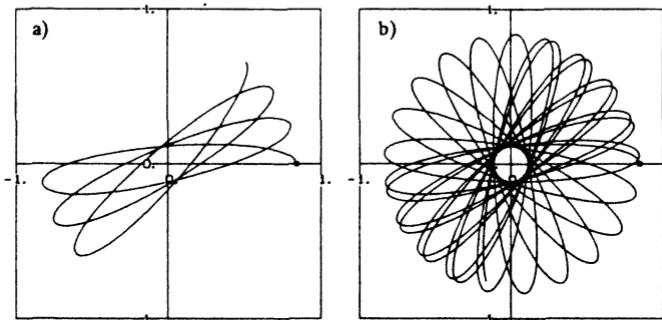


Рис. 16.5. Одно из решений для сферического маятника:  
a)  $0 \leq x \leq 20$ , b)  $0 \leq x \leq 100$ .

В качестве примера рассмотрим «полный брюсселятор» (16.11): мы больше не будем считать  $B$  постоянным, а предположим, что  $B$  добавляется к смеси с постоянной скоростью  $\alpha$ . Введем обозначение  $B(x) =: y_3(x)$  и вместо (16.11') (при  $A = 1$ ) получим систему

$$\begin{aligned} \dot{y}_1 &= 1 + y_1^2 y_2 - (y_3 + 1)y_1, \\ \dot{y}_2 &= y_1 y_3 - y_1^2 y_2, \\ \dot{y}_3 &= -y_1 y_3 + \alpha. \end{aligned} \quad (16.13)$$

Она имеет особую точку  $y_1 = 1$ ,  $y_2 = y_3 = \alpha$  с производной

$$\frac{\partial f}{\partial y} = \begin{pmatrix} \alpha - 1 & 1 & -1 \\ -\alpha & -1 & 1 \\ -\alpha & 0 & -1 \end{pmatrix}. \quad (16.14)$$

Характеристический многочлен этой матрицы равен  $\lambda^3 + (3 - \alpha)\lambda^2 + (3 - 2\alpha)\lambda + 1 = 0$  и удовлетворяет условию устойчивости (см. I.13, упр. 1) тогда и только тогда, когда  $\alpha < (9 - \sqrt{17})/4 = 1.21922$ . Поэтому, если  $\alpha$  возрастает сверх этого

значения, появляется предельный цикл, который существует вплоть до значения  $\alpha$ , приблизительно равного 1.5 (см. рис. 16.6).

Если  $\alpha$  продолжает расти, то предельный цикл «взрывается» и  $y_1 \rightarrow 0$ , в то время как  $y_2$  и  $y_3 \rightarrow \infty$ . Таким образом, поведение системы (16.13) полностью отличается от поведения упрощенной модели (16.11').

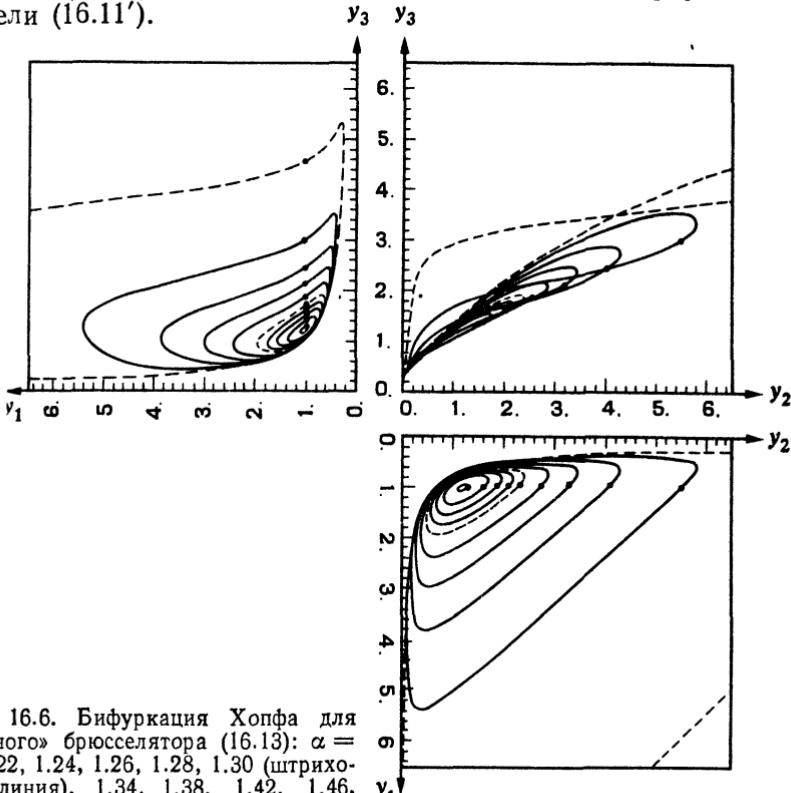


Рис. 16.6. Бифуркация Хопфа для «полного» брюсселятора (16.13):  $\alpha = 1.22, 1.24, 1.26, 1.28, 1.30$  (штрихованная линия),  $1.34, 1.38, 1.42, 1.46, 1.50$  (сплошная линия).

Знаменитой химической реакцией с предельным циклом в трехмерном случае является так называемый «орегонатор» — реакция между  $\text{HBrO}_2$ ,  $\text{Br}^-$  и  $\text{Ce}(\text{IV})$  [Филд и Нойес (1974)], описываемая системой

$$\begin{aligned} y'_1 &= 77.27(y_2 + y_1(1 - 8.375 \times 10^{-6}y_1 - y_2)), \\ y'_2 &= \frac{1}{77.27}(y_3 - (1 + y_1)y_2), \\ y'_3 &= 0.161(y_1 - y_3), \end{aligned} \quad (16.15)$$

решения которой показаны на рис. 16.7. Это пример «жесткого» дифференциального уравнения, решения которого быстро изменяются по величине на много порядков. Поэтому данный пример служит серьезным испытанием для программ численного интегрирования, и мы обратимся к нему еще раз во II томе нашей книги.

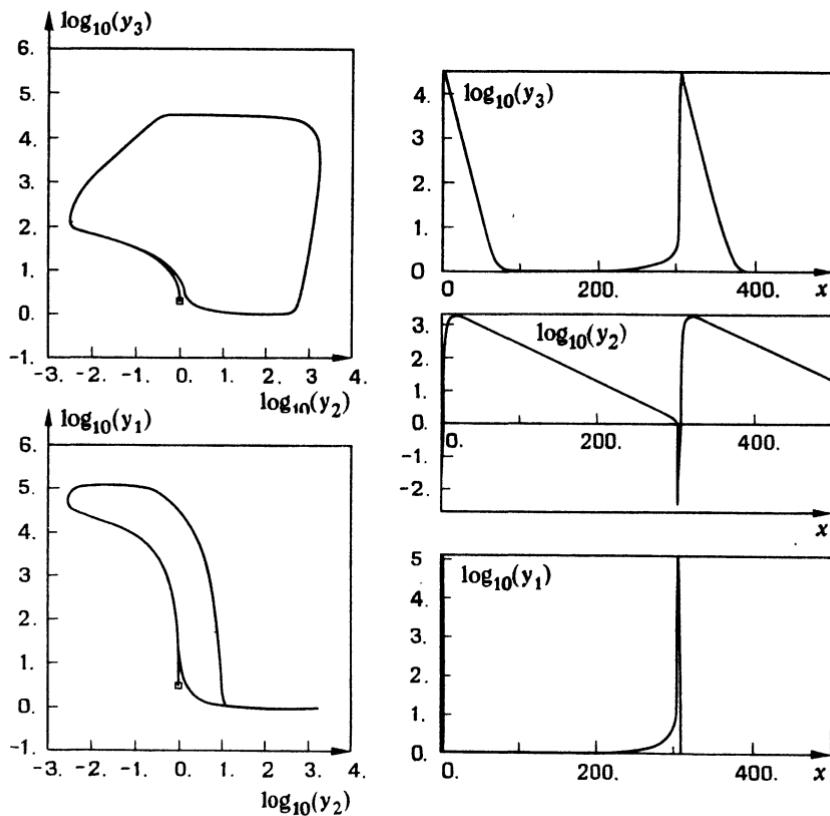


Рис. 16.7. Предельный цикл «орегонатора».

Следующий пример мы взяли из теории сверхпроводимости: два перехода Джозефсона с емкостной связью между ними. Получаемые в этом случае уравнения имеют вид [см., например, работу Джiovанини—Вайса—Ульриха (1978)]:

$$\begin{aligned} c(y_1'' - \alpha y_2'') &= i_1 - \sin(y_1) - y_1', \\ c(y_2'' - \alpha y_1'') &= i_2 - \sin(y_2) - y_2'. \end{aligned} \quad (16.16)$$

Здесь  $y_1$  и  $y_2$  — углы («разности квантовых фаз на переходах»), рассматриваемые по модулю  $2\pi$ . Таким образом, уравнение

(16.16) представляет собой систему на торе  $T^2$  для  $(y_1, y_2)$  и в  $\mathbb{R}^2$  для разностей потенциалов  $(y'_1, y'_2)$ . Из численных расчетов видно, что система (16.16) имеет притягивающий предельный цикл, который описывает явление «запирания фаз» (см. рис. 16.8).

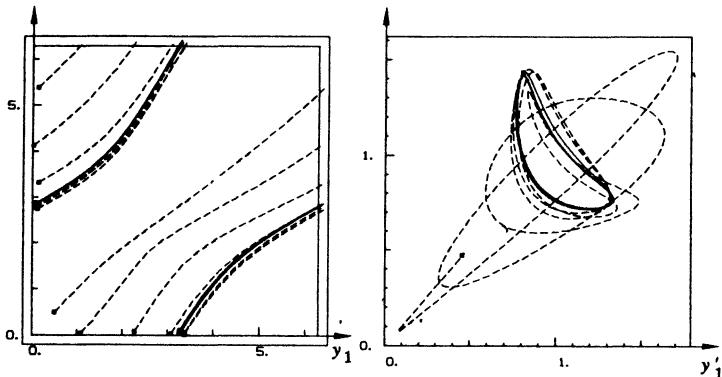


Рис. 16.8. Переходы Джозефсона (16.16) при  $c = 2$ ,  $\alpha = 0.5$ ,  $i_1 = 1.11$ ,  $i_2 = 1.08$ .

### Странные аттракторы

«Г-н Далквист, когда наступит весна?» «Завтра в два часа».

(Прогноз погоды, Стокгольм, 1955)

«Мы были так наивны...»

(Х. О. Крайс, Стокгольм, 1985)

Когда речь идет об открытии знаменитой «модели Лоренца», лучше всего процитировать самого автора [Лоренц (1979)]:

«К середине 1950-х годов, несмотря на весьма низкое качество получаемых результатов, вошло в моду «численное предсказание погоды», т. е. прогноз на основании численного интегрирования доступных для обработки приближений к уравнениям, описывающим атмосферу. Маленькая, но решительная группа ученых поддерживала статистическое предсказание (...), вероятно, вследствие неправильного истолкования одной статьи Винера (...). Я скептически относился к этой идеи и решил проверить ее, применив статистический метод к множеству искусственных данных, полученных при численном решении некоторой системы уравнений (...). Первая задача состояла в нахождении подходящей для решения системы (...). Требовалось, чтобы она была достаточно простой и имела непериодическое общее решение, так как при выявлении периодичности статистическое предсказание периодического ряда

стало бы тривиальным делом. Возможность соединения этих условий была далеко не очевидной (...). Прорыв наступил во время моего посещения д-ра Барри Зальцмана, работающего теперь в Йельском университете. Во время наших бесед он показал мне одну работу по тепловой конвекции, где им использована система из семи обыкновенных дифференциальных уравнений. Большинство полученных им численных решений быстро приобретало периодический характер, но одно решение отказывалось устанавливаться. Кроме того, оказалось, что в этом решении четыре переменных приближаются к нулю. Можно было предположить, что уравнения, которые управляют остальными тремя неизвестными, при исключении из них членов, содержащих первые четыре переменные, также будут иметь непериодические решения. По возращении я решил эти три уравнения на нашем компьютере и подтвердил непериодичность, на которую указывал Зальцман. В конце концов дело пошло».

В измененных обозначениях три уравнения с непериодическими решениями имеют вид:

$$\begin{aligned} y'_1 &= -\sigma y_1 + \sigma y_2, \\ y'_2 &= -y_1 y_3 + r y_1 - y_2, \\ y'_3 &= y_1 y_2 - b y_3, \end{aligned} \quad (16.17)$$

где  $\sigma$ ,  $r$  и  $b$  — положительные константы. Обозначим через  $\Phi_x$ :  $\mathbb{R}^3 \rightarrow \mathbb{R}^3$  отображение

$$\Phi_x(y_0) = y(x, 0, y_0), \quad (16.18)$$

которое является фазовым потоком дифференциального уравнения. Поскольку (16.17) — автономная система, выбор  $x_0 = 0$  не ограничивает общности, и тождество (14.13) принимает вид

$$\Phi_{x_1+x_2} = \Phi_{x_1} \circ \Phi_{x_2}. \quad (16.19)$$

Из (16.17) следует, что

$$\begin{aligned} \frac{1}{2} \frac{d}{dx} (y_1^2 + y_2^2 + (y_3 - \sigma - r)^2) &- \\ = - &\left( \sigma y_1^2 + y_2^2 + b \left( y_3 - \frac{\sigma}{2} - \frac{r}{2} \right)^2 \right) + b \left( \frac{\sigma}{2} + \frac{r}{2} \right)^2. \end{aligned} \quad (16.20)$$

Поэтому шар

$$R_0 = \{ (y_1, y_2, y_3) \mid y_1^2 + y_2^2 + (y_3 - \sigma - r)^2 \leq c^2 \}$$

отображается потоком  $\Phi_1$  в себя при условии, что  $c$  достаточно велико, так что  $R_0$  полностью содержит эллипсоид, определяемый приравниванием нулю правой части (16.20). Следовательно, если  $x$  принимает возрастающие значения  $1, 2, 3, \dots$ , то  $R_0$  отображается в области  $R_1 = \Phi_1(R_0)$ ,  $R_2 = \Phi_2(R_0) = \Phi_1(R_1)$  [в силу (16.19)] и т. д., причем  $R_0 \supset R_1 \supset R_2 \supset R_3 \dots$ .

Область  $R_k$  имеет объем

$$V_k = \int_{R_0} \left| \det \left( \frac{\partial y}{\partial y_0} (k, 0, y_0) \right) \right| dy_0,$$

который экспоненциально стремится к нулю при  $k \rightarrow \infty$ . В этом можно убедиться, применив к уравнению в вариациях (14.15) тождество Абеля—Якоби (11.10). След матрицы  $\partial f / \partial y$  для системы

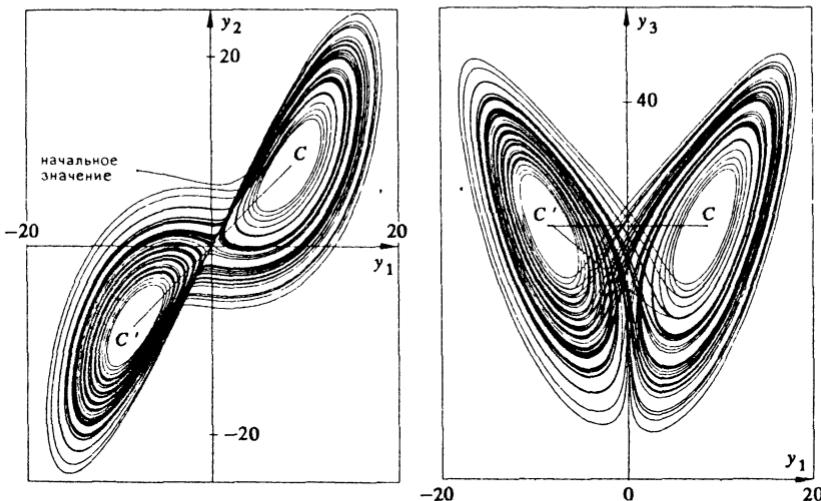


Рис. 16.9. Две проекции решения уравнения (16.17).

(16.17) имеет постоянное значение —  $(\sigma + b + 1)$ , откуда следует, что

$$V_k = e^{-k(\sigma+b+1)} V_0.$$

Таким образом, каждая траектория в конечном счете попадает в множество  $R_\infty = R_0 \cap R_1 \cap R_2 \cap \dots$  нулевого объема.

Система (16.17) имеет очевидную особую точку  $y_1 = y_2 = y_3 = 0$ , которая становится неустойчивой при  $r > 1$ . В этом случае есть еще две дополнительные особые точки  $C$  и  $C'$ , задаваемые соответственно значениями

$$y_1 = y_2 = \pm \sqrt{b(r-1)}, \quad y_3 = r - 1.$$

Они становятся неустойчивыми (например, в силу критерия Рауса, приведенного в упр. 1 из разд. I.13) при  $\sigma > b + 1$  и

$$r \geq r_c = \frac{\sigma(\sigma + b + 3)}{\sigma - b - 1}. \quad (16.21)$$

В первом примере мы используем значения Зальцмана  $b = 8/3$ ,  $\sigma = 10$  и  $r = 28$ . («Здесь мы отметим еще одно счастливое обстоятельство: Зальцман использовал  $\sigma = 10$  в качестве грубого приближения числа Прандтля (около 6) для воды. Если бы он выбрал для изучения воздух, то, вероятно, взял бы  $\sigma = 1$ , и непериодич-

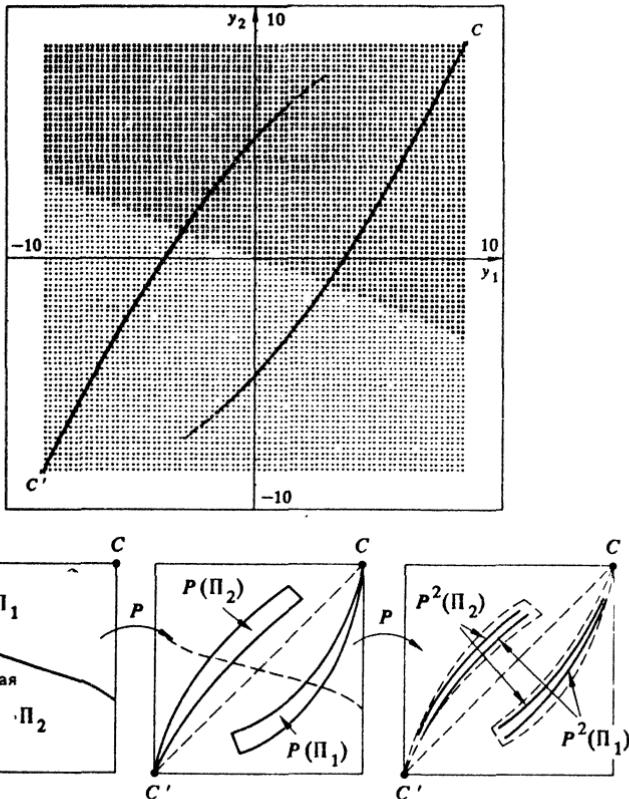


Рис. 16.10. Отображение Пуанкаре для уравнения (16.17).

ность не была бы открыта» [Лоренц (1979)].) На рис. 16.9 представлена кривая решения системы (16.17) с начальными значениями  $y_1 = -8$ ,  $y_2 = 8$ ,  $y_3 = r - 1$ , которая действительно выглядит довольно хаотической.

Чтобы яснее понять, что происходит, отметим следующий факт: любое решение, начинающееся на плоскости  $y_3 = r - 1$  между двумя особыми точками, окружает одну из них и возвращается с обратной стороны. Поэтому выберем указанную плоскость  $\Pi$

в качестве сечения и рассмотрим отображение Пуанкаре  $P: \Pi \rightarrow \Pi$ . Численное изучение данного отображения показывает, что существует критическая линия, выше которой решения идут вправо, а ниже — влево. Следовательно, образ  $P(\Pi)$  состоит из двух множеств, как показано на рис. 16.10. Отсюда вытекает, что образ  $P^2$  включает 4 слоя, образ  $P^3$  — 8 слоев и т. д., причем слои размещаются все ближе друг к другу. Поэтому аттрактор состоит из бесконечного множества листов, расположенных наподобие точек канторова совершенного множества. Такие аттракторы называют теперь *странными аттракторами*. («Следовательно, аттрактор описанного типа нельзя отбрасывать как частный патологический случай». — Рюэлль—Такенс (1971).)

### Каскады Фейгенбаума

Хотя начало статьи Лоренца (1979) написано очень хорошо, утверждения последнего ее раздела верны лишь отчасти. Следуя Лоренцу, будем теперь изменять в (16.17) параметр  $b$ , полагая одновременно  $r = r_c$  (см. (16.21)) и

$$\sigma = b + 1 + \sqrt{2(b+1)(b+2)}. \quad (16.22)$$

При таком значении  $\sigma$  параметр  $r_c$  достигает минимума. Численное интегрирование показывает, что при очень малых  $b$  (точнее,  $b < 0.139$ ) решения (16.17), очевидно, сходятся к устойчивому предельному циклу, который дважды пересекает сечение Пуанкаре  $y_3 = r - 1$  в двух различных местах и окружает обе критические точки  $C$  и  $C'$ . При больших  $b$  (например,  $b = 8/3$ ) коэффициенты  $\sigma$  и  $r$  слабо отличаются от изученного выше случая и существует странный аттрактор. Но что происходит внутри интервала изменения  $b$ ? Мы вычислили решения модели Лоренца (16.17) для 687 промежуточных значений  $b$ , расположенных от 0.138 до 0.148. Для каждого из них мы рассчитали 800 пересечений сечения Пуанкаре. На рис. 16.11 представлены значения  $y_1$ , при которых решение пересекает плоскость Пуанкаре  $y_3 = r - 1$ . При этом после каждого изменения  $b$  первые 40 итераций не показаны на рисунке, так что виден только аттрактор.

Для малых  $b$  существует одна периодическая траектория; потом при  $b = b_1 = 0.1397$  она неожиданно расщепляется на траекторию с периодом два, которая в свою очередь расщепляется при  $b = b_2 = 0.1433$  на траекторию с периодом четыре, затем восемь и т. д. Существует точка  $b_\infty$ , после которой движение становится хаотическим. Но и за этим значением снова и снова встречаются интервалы устойчивых аттракторов периодов 5, 3, ... и т. д. Полная картина напоминает результат вычислений по рекуррентной формуле

$$x_{n+1} = a(x_n - x_n^2), \quad (16.23)$$

которая рассматривалась во многих статьях [см., например, Мей (1976), Фейгенбаум (1978), Колле—Экманн (1980)].

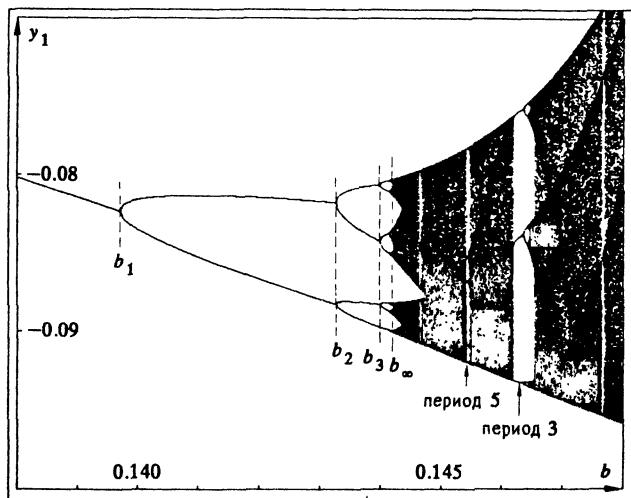


Рис. 16.11. Значения  $y_1$  сечений Пуанкаре для уравнения (16.17) в зависимости от  $b$ .

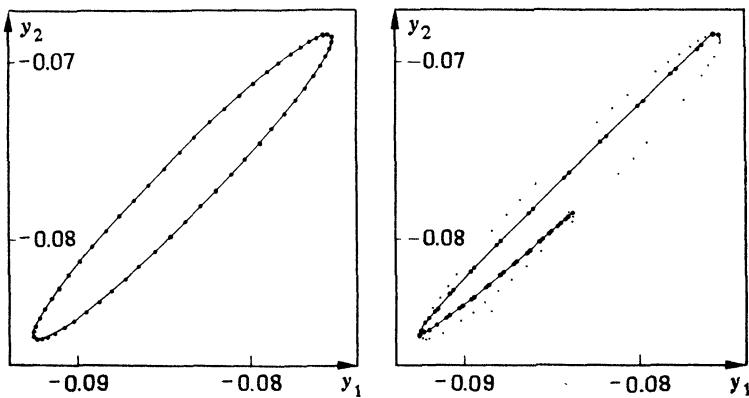


Рис. 16.12. Отображение Пуанкаре эллипса (16.24).

Но отчего возникает это сходство? С помощью рис. 16.12 изучим отображение Пуанкаре для системы (16.17) со значением  $b = 0.146$ . Мы выбрали эллипс

$$\begin{aligned} y_1 &= -0.084 + (u - v) \cdot 0.7071, \\ y_2 &= -0.077 + (u + v) \cdot 0.7071, \end{aligned} \quad (16.24)$$

где  $u = 0.012 \cos \varphi$ ,  $v = 0.0015 \sin \varphi$ ,  $0 \leq \varphi \leq 2\pi$ . Что происходит с этим эллипсом после одной итерации? Он сожмется в тонкую линию, изогнутую в форме буквы  $U$  и отображающуюся в себя.

Таким образом, отображение Пуанкаре — это, по существу, отображение отрезка  $[0, 1]$  в себя, подобное тому, которое описывается соотношением (16.23). Замечательное открытие Фейгенбаума состоит в том, что для всех отображений подобной формы всегда наблюдаются одни и те же явления, и, в частности, величина

$$\lim_{i \rightarrow \infty} \frac{b_i - b_{i-1}}{b_{i+1} - b_i} = 4.6692016091029906715\dots$$

является универсальной константой — числом Фейгенбаума. Повторяющиеся удвоения периодов при значениях  $b_1, b_2, b_3, \dots$  называются каскадами Фейгенбаума.

### Упражнения

1. Уравнение Ван-дер-Поля (16.2) при  $\varepsilon = 1$  имеет предельный цикл с периодом  $T = 6.6632868593231301896996820305$ , проходящий через  $y_2 = 0$ ,  $y_1 = A$ , где  $A = 2.00861986087484313650940188$ .

Теперь замените (16.2) системой

$$\begin{aligned} y' &= y_2(A - y_1), \\ y_2' &= ((1 - y_1^2)y_2 - y_1)(A - y_1), \end{aligned}$$

у которой на предельном цикле появилась стационарная точка. Изучите поведение решения, начинающегося внутри него, например, со значения  $y_{10} = 1$ ,  $y_{20} = 0$ .

2. [Фроммер (1934)] Рассмотрите систему

$$\ddot{y}_1 = y_2 + 2y_1y_2 - y_2^2, \quad \ddot{y}_2 = y_1 + (1 + \varepsilon)y_1^2 + 2y_1y_2 - y_2^2. \quad (16.25)$$

Исследуя устойчивость, как в упр. 5 из разд. I.13, либо с помощью численных расчетов покажите, что при  $\varepsilon > 0$  система (16.25) имеет предельный цикл асимптотического радиуса  $r = \sqrt{6\varepsilon/7}$ . (См. также статью Ваннера (1983) и разд. I.13, упр. 5.)

3. Решите 16-ю проблему Гильберта: какое максимальное число предельных циклов может иметь квадратичная система

$$\ddot{y}_1 = \alpha_0 + \alpha_1 y_1 + \alpha_2 y_2 + \alpha_3 y_1^2 + \alpha_4 y_1 y_2 + \alpha_5 y_2^2,$$

$$\ddot{y}_2 = \beta_0 + \beta_1 y_1 + \beta_2 y_2 + \beta_3 y_1^2 + \beta_4 y_1 y_2 + \beta_5 y_2^2?$$

Математическая общественность ждет Вас: более 80 лет никому не удается решить эту задачу. В настоящее время наибольшее

известное число равно 4, как, например, для системы

$$\dot{y}_1 = \lambda y_1 - y_2 - 10y_1^2 + (5 + \delta) y_1 y_2 + y_2^2,$$

$$\dot{y}_2 = y_1 + y_1^2 + (-25 + 8\epsilon - 9\delta) y_1 y_2,$$

$$\delta = -10^{-13}, \quad \epsilon = -10^{-52}, \quad \lambda = -10^{-200}$$

(См. работы Ши Сонглина (1980), Ваннера (1983), Перко (1984).)

4. Найдите замену координат, при которой уравнение

$$my'' + (-A + B(y')^2)y' + ky = 0$$

становится уравнением Ван-дер-Поля (16.2), см. книгу Крылова—Боголюбова (1947).

5. Примените к уравнению математического маятника

$$y'' + \sin y = y'' + y - \frac{y^3}{6} + \frac{y^5}{120} \pm \dots = 0,$$

$$y(0) = \epsilon, \quad y'(0) = 0$$

метод асимптотических разложений (16.6) и (16.7) и изучите период как функцию от  $\epsilon$ .

*Ответ.* Период равен  $2\pi(1 + \epsilon^2/16 + \dots)$ .

6. Вычислите предельный цикл (биfurкацию Хопфа) для уравнения

$$y'' + y = \epsilon^2 y' - (y')^3, \quad y'(0) = 0$$

при малых  $\epsilon$ , используя для этого метод Пуанкаре (16.6), (16.7) со значением  $z_0(t) = 0$ .

7. Аналогичным способом (как в упр. 6) изучите брюсселятор (6.11') при  $A = 1$  и  $B = 2 + \epsilon^2$ .

*Указание.* Для упрощения расчетов перенесите особую точку (1,2) в начало координат, а затем выберите новые координаты, используя собственные векторы линеаризованного уравнения (см. (12.22) и (12.23)). В результате получится уравнение

$$y'' + y = -\epsilon^2(y' - 1) - (y')^2(y + y') + 2yy'.$$

*Ответ.*  $z(t) = \epsilon(2/\sqrt{3}) \cos t + \dots$ ,  $t = x(1 - \epsilon^2/18 + \dots)$ , а поэтому период асимптотически равен  $2\pi(1 + \epsilon^2/18 + \dots)$ .

8. [Лъенар (1928)] Докажите, что предельный цикл уравнения Ван-дер-Поля (16.1) для каждого  $\epsilon > 0$  является единственным.

*Указание.* Тождество

$$y'' + \epsilon(y^2 - 1)y' = \frac{d}{dx}(y' + \epsilon(y^3/3 - y))$$

подсказывает выбор координатной системы  $y_1(x) = y(x)$ ,  $y_2(x) = y' + \epsilon(y^3/3 - y)$ . Запишите соответствующую систему первого порядка, исследуйте знаки  $\dot{y}_1$ ,  $\dot{y}_2$  и рост функции

«энергии»  $V(x) = (y_1^2 + y_2^2)/2$ . Кроме того, обобщите результат на уравнение вида  $y'' + f(y)y' + g(y) = 0$ . Подробнее этот вопрос изложен, например, в книге Симмонса (1972).

9. [Рэлей (1883)]. При малых  $\kappa$  и  $\lambda$  найдите периодическое решение уравнения

$$y'' + \kappa y' + \lambda (y')^3 + n^2 y = 0.$$

*Ответ.*  $y = A \sin(nx) + (\lambda n A^3/32) \cos(3nx)$ , где значение  $A$  определяется из формулы  $\kappa + (3/4)\lambda \cdot n^2 A^2 = 0$ .

10. [Бендикссон (1901)]. Если в некоторой области  $\Omega$  на плоскости выражение

$$\frac{\partial f_1}{\partial y_1} + \frac{\partial f_2}{\partial y_2}.$$

всегда отрицательно или всегда положительно, система (16.4) не может иметь замкнутых решений в  $\Omega$ .

*Указание.* Примените формулу Грина

$$\int \int \left( \frac{\partial f_1}{\partial y_1} + \frac{\partial f_2}{\partial y_2} \right) dy_1 dy_2 = \int (f_1 dy_2 - f_2 dy_1).$$

## Глава II

# Методы Рунге — Кутты и экстраполяционные методы

Численные методы для обыкновенных дифференциальных уравнений естественным образом разделяются на два класса. В один из них входят методы, использующие одно стартовое значение на каждом шаге («одношаговые методы»), а другой образуют методы, опирающиеся на *несколько* значений решения («многошаговые методы»). В этой главе изучаются одношаговые методы, а многошаговые методы рассмотрены в гл. III. Обе главы в значительной мере независимы, и каждую из них можно читать, по крайней мере их начальные разделы, не зная содержания другой.

Глава начинается с изложения теории методов Рунге—Кутты, включая вывод (с помощью помеченных деревьев) условий, определяющих порядок аппроксимации, оценки погрешностей, доказательства сходимости, вопросы программной реализации, методы высших порядков. В разд. II.7 вводятся неявные методы Рунге—Кутты. Им будет уделено больше внимания во втором томе, посвященном жестким дифференциальным уравнениям. Следующие несколько разделов посвящены элегантной идее *экстраполяции* (Ричардсон, Ромберг и др.) и ее использованию для построения программ, интегрирующих с высоким порядком точности. Затем рассматривается алгебраическая теория композиции методов, на основании которой в дальнейших главах будут изучены свойства порядка аппроксимации для многих общих классов методов. Глава заканчивается специальными методами для дифференциальных уравнений второго порядка  $y'' = f(x, y)$  и для дифференциальных уравнений с запаздывающим аргументом.

Для иллюстрации работы методов, описанных в этой главе, мы выбрали пример из небесной механики — ограниченную задачу трех тел. Рассматриваются два тела с массами  $1 - \mu$  и  $\mu$ , участвующие в совместном круговом движении в некоторой плоскости, и движущееся вблизи них в той же плоскости третье тело пренебрежимо малой массы. Уравнения имеют вид

$$y_1'' = y_1 + 2y_2' - \mu' \frac{y_1 + \mu}{D_1} - \mu \frac{y_1 - \mu'}{D_2}, \quad (0.1)$$

$$y_2'' = y_2 - 2y_1' - \mu' \frac{y_2}{D_1} - \mu \frac{y_2}{D_2},$$

$$D_1 = ((y_1 + \mu)^2 + y_2^2)^{3/2}, \quad D_2 = ((y_1 - \mu')^2 + y_2^2)^{3/2}, \\ \mu = 0.012277471, \quad \mu' = 1 - \mu.$$

Мы тщательно определили начальные условия:

$$\begin{aligned}y_1(0) &= 0.994, \quad y'_1(0) = 0, \quad y_2(0) = 0, \\y'_2(0) &= -2.00158510637908252240537862224, \\X_{end} &= 17.0652165601579625588917206249,\end{aligned}$$

чтобы решение получилось периодическим с периодом  $X_{end}$ . Такие периодические решения много десятилетий привлекали

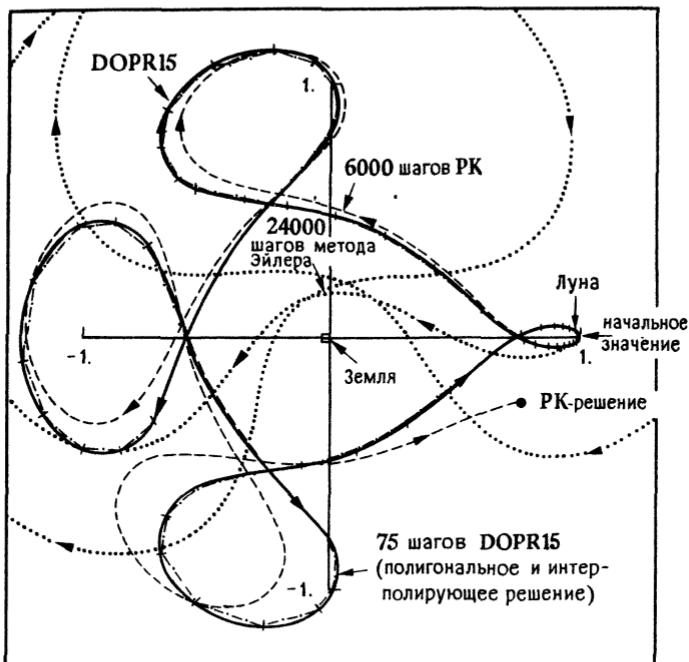


Рис. 0.1. Орбита Аренсторфа, вычисленная эквидистантным методом Эйлера, эквидистантным методом Рунге — Кутты и методом Дормана и Принса с переменной длиной шага.

внимание астрономов и математиков (Пуанкаре). Теперь их часто называют «орбитами Аренсторфа» (см. Аренсторф (1963); этот автор также провел много расчетов на «быстро действующих электронных вычислительных машинах»). Задача принадлежит классу  $C^\infty$  (т. е. решения бесконечно дифференцируемы), за исключением двух особых точек:  $y_1 = -\mu$  и  $y_1 = 1 - \mu$ ,  $y_2 = 0$ . Поэтому, как мы знаем из разд. I.7, метод ломанных Эйлера сходится к решению. Но будет ли он действительно полезным для вычислений в нашем случае? Мы провели вычисления, взяв 24000 шагов длины

$h = X_{end}/24000$ , и построили орбиту, приведенную на рис. 0.1. Результат нельзя назвать замечательным.

Метод Рунге—Кутты (приведенный в табл. 1.2 слева) работает уже гораздо лучше, он быстрее сходится к решению. Мы взяли 6000 шагов длины  $X_{end}/6000$ , чтобы объем вычислительной работы остался прежним. Ясно, что точность теряется больше всего в тех частях орбиты, которые близки к особой точке. Поэтому программы с автоматическим выбором длины шага, описанные в разд. II.4, работают гораздо лучше, и программа DOPRI5 (табл. 4.6) вычисляет орбиту с точностью до  $10^{-3}$  за 75 шагов. На некоторых участках длина шага становится очень большой, и графическое представление результата в виде ломаной, соединяющей точки решения, становится неудовлетворительным. Сплошная кривая на рис. 0.1 — интерполированное решение (5.7), точное и для всех промежуточных значений. Оно полезно также во многих других случаях, например при решении дифференциальных уравнений с запаздывающим аргументом, для вывода результатов в большом числе промежуточных точек («плотная выдача») или в точке, заранее не известной («неявная выдача»).

Для достижения еще более высокой точности нужны методы более высокого порядка. Например, программа DOPRI8 (табл. 6.4) вычисляет орбиту быстрее, чем DOPRI5, и при более строгих ограничениях на допустимую погрешность (скажем, не превосходящую  $10^{-6}$ ). Наивысший порядок точности достигается применением экстраполяционных методов (разд. II.9). Программа ODEX (при  $K_{max} = 15$ ) вычисляет орбиту с точностью до  $10^{-30}$ , делая около 25000 вычислений функций, стоящих в правых частях уравнений — как раз тот объем вычислений, что и в упомянутом выше решении методом Эйлера.

## II.1. Первые методы Рунге — Кутты

«Численное определение какого-то решения заданного дифференциального уравнения, аналитическое решение которого неизвестно, кажется, до сих пор мало занимало внимание математиков...»

(К. Рунге, 1895)

Метод Эйлера для решения начальной задачи

$$y' = f(x, y), \quad y(x_0) = y_0 \quad (1.1)$$

был описан Эйлером (1768) в его «Интегральном исчислении» (раздел второй, гл. VII). Метод прост для понимания и программирования. Мы подробно исследовали его сходимость в разд. I.7 и убедились, что глобальная погрешность имеет вид  $Ch$ , где  $C$  — постоянная, зависящая от задачи, и  $h$  — максимальная длина шага. Если желательно, скажем, получить 6 точных десятичных знаков, то требуется, следовательно, порядка миллиона шагов, что не слишком удовлетворительно. С другой стороны, еще со времен Ньютона известно, что можно найти гораздо более точные методы, если  $f$  в (1.1) не зависит от  $y$ , т. е. если мы имеем задачу

$$y' = f(x), \quad y(x_0) = y_0, \quad (1.1')$$

решаемую квадратурой:

$$y(X) = y_0 + \int_{x_0}^X f(x) dx. \quad (1.2)$$

В качестве примера рассмотрим «правило средней точки» (или первую квадратурную формулу Гаусса):

$$y(x_0 + h_0) \approx y_1 = y_0 + h_0 f\left(x_0 + \frac{h_0}{2}\right),$$

$$y(x_1 + h_1) \approx y_2 = y_1 + h_1 f\left(x_1 + \frac{h_1}{2}\right), \quad (1.3')$$

• • • • • • • • • • • • •

$$y(X) \approx Y = y_{n-1} + h_{n-1} f\left(x_{n-1} + \frac{h_{n-1}}{2}\right),$$

где  $h_i = x_{i+1} - x_i$  и  $x_0, x_1, \dots, x_{n-1}, x_n = X$  — граничные точки подынтервалов, на которые разбит интервал интегрирования. Известно, что оценка глобальной погрешности этой формулы  $y(X) - Y$  имеет вид  $Ch^2$ . Таким образом, если желаемая точность составляет 6 десятичных знаков, ее обычно можно получить приблизительно за тысячу шагов, т. е. этот метод в тысячу раз быстрее. Поэтому Рунге (1895) поставил следующий вопрос: нельзя ли распространить метод (1.3') и на задачу (1.1)? Первый шаг длины  $h = h_0$  должен иметь вид

$$y(x_0 + h) \approx y_0 + h f\left(x_0 + \frac{h}{2}, y\left(x_0 + \frac{h}{2}\right)\right). \quad (1.3)$$

Но какое значение взять для  $y(x_0 + h/2)$ ? За неимением лучшего естественно использовать один малый шаг метода Эйлера длины  $h/2$ . Тогда из (1.3) получим

$$\begin{aligned} k_1 &= f(x_0, y_0), \\ k_2 &= f\left(x_0 + \frac{h}{2}, y_0 + \frac{h}{2}k_1\right), \\ y_1 &= y_0 + hk_2. \end{aligned} \quad (1.4)$$

Может показаться странным, что для вычисления  $k_2$  мы предлагаем сделать шаг методом Эйлера, о неэффективности которого говорилось полстраницей выше. Однако решающим обстоятельством является умножение  $k_2$  в третьем выражении на  $h$ , в результате чего влияние погрешности становится менее существенным. Точнее говоря, вычислим для  $y_1$  в (1.4) разложение Тейлора по степеням  $h$ :

$$\begin{aligned} y_1 &= y_0 + hf\left(x_0 + \frac{h}{2}, y_0 + \frac{h}{2}f_0\right) = \\ &= y_0 + hf(x_0, y_0) + \frac{h^2}{2}(f_x + f_yf)(x_0, y_0) + \\ &\quad + \frac{h^3}{8}(f_{xx} + 2f_{xy}f + f_{yy}f^2)(x_0, y_0) + \dots \end{aligned} \quad (1.5)$$

Его можно сравнить с рядом Тейлора для точного решения, который получается из (1.1) повторным дифференцированием с заменой  $y'$  на  $f$  каждый раз, когда оно появляется [(Эйлер (1768), задача 86, пар. 656; см. также формулы (8.12) гл. I):

$$\begin{aligned} y(x_0 + h) &= y_0 + hf(x_0, y_0) + \frac{h^2}{2}(f_x + f_yf)(x_0, y_0) + \\ &\quad + \frac{h^3}{6}(f_{xx} + 2f_{xy}f + f_{yy}f^2 + f_yf_x + f_y^2f)(x_0, y_0) + \dots \end{aligned} \quad (1.6)$$

Вычитая из последнего равенства предыдущее, получим для погрешности первого шага выражение

$$\begin{aligned} y(x_0 + h) - y_1 &= \\ &= \frac{h^3}{24}(f_{xx} + 2f_{xy}f + f_{yy}f^2 + 4(f_yf_x + f_y^2f))(x_0, y_0) + \dots \end{aligned} \quad (1.7)$$

Таким образом, если все частные производные  $f$  второго порядка ограничены, то  $\|y(x_0 + h) - y_1\| \leq K h^3$ .

Чтобы получить приближенное значение решения задачи (1.1) в конечной точке  $X$ , будем применять формулы (1.4) последовательно к интервалам  $(x_0, x_1), (x_1, x_2), \dots, (x_{n-1}, X)$ , подобно тому как применялся метод Эйлера в разд. I.7. По аналогии с доказательством сходимости в разд. I.7, мы покажем в разд. II.3, что, как и в случае задачи (1.1'), погрешность численного решения ограничена величиной вида  $Ch^2$  ( $h$  — максимальная длина шага).

Таким образом, (1.4) является усовершенствованием метода Эйлера. Для вычислений с высокой точностью надо найти еще лучшие методы, и этим мы будем заниматься главным образом в следующих разделах.

## Общая формулировка методов Рунге — Кутты

Рунге (1895) и Хайн (1900) построили новые методы, включив в (1.4) один или два добавочных шага по Эйлеру. Но именно Кутта (1901) сформулировал общую схему того, что теперь называется методом Рунге—Кутты.

**Определение 1.1.** Пусть  $s$  — целое положительное число («число стадий», или «этапов») и  $a_{21}, a_{31}, a_{32}, \dots, a_{s1}, a_{s2}, a_{s-s-1}, \dots, b_1, \dots, b_s, c_2, \dots, c_s$  — вещественные коэффициенты. Тогда метод

называется *s-стадийным* (*s*-этапным) явным методом Рунге—Кутты (ЯМРК) для задачи (1.1).

Обычно коэффициенты  $c_i$  удовлетворяют условиям

$$c_2 = a_{21}, \quad c_3 = a_{31} + a_{32}, \dots, \quad c_s = a_{s1} + \dots + a_{s,s-1} \quad (1.9)$$

или, короче,

$$c_i = \sum_j a_{ij}. \quad (1.9')$$

Эти условия были приняты Куттой без каких-либо комментариев. Смысл их в том, что все точки, в которых вычисляется  $f$ , являются приближениями первого порядка к решению. Эти условия сильно упрощают вывод условий, определяющих порядок аппроксимации для методов высокого порядка. Однако для методов низких порядков эти предположения не являются необходимыми (см. упр. 6).

**Определение 1.2.** Метод Рунге—Кутты (1.8) имеет порядок  $p$ , если для достаточно гладких задач (1.1)

$$\|y(x_0 + h) - y_1\| \leq K h^{p+1}, \quad (1.10)$$

т. е. если ряды Тейлора для точного решения  $y(x_0 + h)$  и для  $y_i$  совпадают до члена  $h^p$  включительно.

После статьи Бутчера (1964b) вошло в обычай символически представлять метод (1.8) посредством следующей таблицы:

$$\begin{array}{c|ccccc} 0 & & & & & \\ c_2 & a_{21} & & & & \\ c_3 & a_{31} & a_{32} & & & \\ \vdots & \vdots & \vdots & & & \\ \vdots & \vdots & \vdots & & & \\ c_s & a_{s1} & a_{s2} & \dots & a_{s,s-1} & \\ \hline & b_1 & b_2 & \dots & b_{s-1} & b_s \end{array} \quad (1.8')$$

**Примеры.** Рассмотренный выше метод Рунге, а также методы Рунге и Хойна порядка 3 приведены в табл. 1.1.

Таблица 1.1. Методы Рунге—Кутты низших порядков

Рунге, порядок 2		Рунге, порядок 3				Хойн, порядок 3		
0		0				0		
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$			$\frac{1}{3}$	$\frac{1}{3}$	
$\frac{1}{2}$	$\frac{1}{2}$	1	0	1		$\frac{2}{3}$	0	$\frac{2}{3}$
		1	0	0	1			
			$\frac{1}{6}$	$\frac{2}{3}$	0	$\frac{1}{6}$		
	0	1					$\frac{1}{4}$	0
								$\frac{3}{4}$

### Обсуждение методов порядка 4

«Из более новых методов я считаю наилучшим следующий, предложенный г-ном Куттой».  
(К. Рунге, 1905)

Займемся теперь определением 4-стадийных методов Рунге—Кутты (1.8) с таким расчетом, чтобы они имели порядок 4. Выше мы видели, что надо делать: вычислить производные порядков 1, 2, 3 и 4 от  $y_1 = y_1(h)$  при  $h = 0$  и сравнить их с производными точного решения. Теоретически при известных правилах дифференциального исчисления это совершенно тривиальная задача. С использованием (1.9) получаются следующие условия:

$$\sum_i b_i = b_1 + b_2 + b_3 + b_4 = 1, \quad (1.11a)$$

$$\sum_i b_i c_i = b_2 c_2 + b_3 c_3 + b_4 c_4 = \frac{1}{2}, \quad (1.11b)$$

$$\sum_i b_i c_i^2 = b_2 c_2^2 + b_3 c_3^2 + b_4 c_4^2 = \frac{1}{3}, \quad (1.11\text{c})$$

$$\sum_{i, l} b_i a_{ij} c_j = b_3 a_{32} c_2 + b_4 (a_{42} c_2 + a_{43} c_3) = \frac{1}{6}, \quad (1.11\text{d})$$

$$\sum_i b_i c_i^3 = b_2 c_2^3 + b_3 c_3^3 + b_4 c_4^3 = \frac{1}{4}, \quad (1.11\text{e})$$

$$\sum_{i, l} b_i c_i a_{ij} c_j = b_3 c_3 a_{32} c_2 + b_4 c_4 (a_{42} c_2 + a_{43} c_3) = \frac{1}{8}, \quad (1.11\text{f})$$

$$\sum_{i, l} b_i a_{il} c_l^2 = b_3 a_{32} c_2^2 + b_4 (a_{42} c_2^2 + a_{43} c_3^2) = \frac{1}{12}, \quad (1.11\text{g})$$

$$\sum_{i, l, k} b_i a_{ij} a_{jk} c_k = b_4 a_{43} a_{32} c_2 = \frac{1}{24}. \quad (1.11\text{h})$$

Эти вычисления, не приведенные в статье Кутты (они даны, однако, в работе Хойна (1900)), очень утомительны. И громоздкость их очень быстро растет для более высоких порядков. В разд. II.2 мы увидим, что при использовании соответствующих обозначений они могут стать очень изящными.

Кутта дал общее решение уравнений (1.11) без объяснений. Ясное изложение вывода решений приведено в книге Рунге и Кёнига (1924). Здесь мы будем следовать идеям Дж. К. Бутчера, которые проясняют роль так называемых *упрощающих предположений* и применимы также в случае более высоких порядков.

**Лемма 1.3.** *Если*

$$\sum_i b_i a_{ij} = b_j (1 - c_j), \quad j = 1, \dots, 4, \quad (1.12)$$

*то уравнения d), g) и h) в (1.11) являются следствиями остальных.*

*Доказательство.* Покажем это для g). С помощью уравнений с) и e) получим:

$$\sum_{i, l} b_i a_{il} c_l^2 = \sum_l b_l c_l^2 - \sum_l b_l c_l^3 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}.$$

Для уравнений d) и h) процедура аналогична. ■

А теперь мы покажем, что в нашем случае условие (1.12) является и необходимым.

**Лемма 1.4.** *При  $s = 4$  (1.12) следует из уравнений (1.11) и (1.9).*

При доказательстве этой леммы используется следующая

**Лемма 1.5.** Пусть  $U$  и  $V$  суть  $3 \times 3$ -матрицы, такие, что

$$UV = \begin{pmatrix} a & b & 0 \\ c & d & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} \neq 0. \quad (1.13)$$

Тогда либо  $Ve_3 = 0$ , либо  $U^T e_3 = 0$ , где  $e_3 = (0, 0, 1)^T$ .

**Доказательство леммы 1.5.** Если  $\det U \neq 0$ , то из  $UVe_3 = 0$  следует  $Ve_3 = 0$ . Если же  $\det U = 0$ , то существует вектор  $x = (x_1, x_2, x_3)^T \neq 0$ , такой, что  $U^T x = 0$ , и поэтому  $V^T U^T x = 0$ . Но тогда из (1.13) следует, что  $x$  должен быть пропорционален вектору  $e_3$ . ■

**Доказательство леммы 1.4.** Введем величины

$$d_j = \sum_i b_i a_{ij} - b_j(1 - c_j) \quad \text{для } j = 1, \dots, 4.$$

Итак, надо доказать, что  $d_j = 0$ . Введем теперь матрицы

$$U = \begin{pmatrix} b_2 & b_3 & b_4 \\ b_2c_2 & b_3c_3 & b_4c_4 \\ d_2 & d_3 & d_4 \end{pmatrix}, \quad V = \begin{pmatrix} c_2 & c_2^2 & \sum_l a_{2l}c_l - \frac{c_2^2}{2} \\ c_3 & c_3^2 & \sum_l a_{3l}c_l - \frac{c_3^2}{2} \\ c_4 & c_4^2 & \sum_l a_{4l}c_l - \frac{c_4^2}{2} \end{pmatrix}. \quad (1.14)$$

Перемножение этих матриц с использованием условий (1.11) дает

$$UV = \begin{pmatrix} \frac{1}{2} & \frac{1}{3} & 0 \\ \frac{1}{3} & \frac{1}{4} & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \text{причем} \quad \det \begin{pmatrix} \frac{1}{2} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{4} \end{pmatrix} \neq 0.$$

Далее, последний столбец  $V$  не может быть нулевым, так как из того, что  $c_1 = 0$ , следует

$$\sum_l a_{2l}c_2 - \frac{c_2^2}{2} = -\frac{c_2^2}{2} \neq 0$$

в силу условия h). Таким образом, из леммы 1.5 следует, что  $d_2 = d_3 = d_4 = 0$ . Последнее тождество  $d_1 = 0$  вытекает из равенства  $d_1 + d_2 + d_3 + d_4 = 0$ , которое является следствием (1.11a, b) и (1.9). ■

Из лемм 1.3 и 1.4 получается следующая

**Теорема 1.6.** Если выполнены предположения (1.9), то уравнения (1.11) эквивалентны следующим:

$$b_1 + b_2 + b_3 + b_4 = 1, \quad (1.15a)$$

$$b_2 c_2 + b_3 c_3 + b_4 c_4 = \frac{1}{2}, \quad (1.15b)$$

$$b_2 c_2^2 + b_3 c_3^2 + b_4 c_4^2 = \frac{1}{3}, \quad (1.15c)$$

$$b_2 c_2^3 + b_3 c_3^3 + b_4 c_4^3 = \frac{1}{4}, \quad (1.15e)$$

$$b_3 c_3 a_{32} c_2 + b_4 c_4 (a_{42} c_2 + a_{43} c_3) = \frac{1}{8}, \quad (1.15f)$$

$$b_3 a_{32} + b_4 a_{42} = b_2 (1 - c_2), \quad (1.15i)$$

$$b_4 a_{43} = b_3 (1 - c_3), \quad (1.15j)$$

$$0 = b_4 (1 - c_4), \quad (1.15k)$$

Доказательство очевидно. ■

Из (1.15j) и (1.11h) следует, что

$$b_3 b_4 c_2 (1 - c_3) \neq 0. \quad (1.16)$$

Отсюда, в частности, вытекает, что в силу (1.15k)  $c_4 = 1$ .

*Решение уравнений (1.15).* Уравнения а)—е) и к) выражают тот факт, что коэффициенты  $b_i$  и  $c_i$  являются весами и узлами квадратурной формулы четвертого порядка при  $c_1 = 0$  и  $c_4 = 1$ . В силу (1.16) возможны следующие четыре случая:

$$1) \quad c_2 = u, \quad c_3 = v, \quad \text{причем } u \neq v, \quad u \neq \frac{1}{2}, \quad u \text{ и } v \neq 0, \neq 1. \quad (1.17)$$

Тогда уравнения а)—е) образуют невырожденную линейную систему для определения  $b_1, b_2, b_3, b_4$ . Эта система имеет решение

$$b_2 = \frac{2v - 1}{12u(1-u)(v-u)}, \quad b_3 = \frac{1 - 2u}{12v(1-v)(v-u)},$$

$$b_4 = \frac{6uv + 3 - 4u - 4v}{12(1-u)(1-v)}, \quad b_1 = 1 - b_2 - b_3 - b_4.$$

Остальные три случая с двойными узлами основаны на правиле Симпсона:

$$2) \quad c_3 = 0, \quad c_2 = \frac{1}{2}, \quad b_3 = u \neq 0, \quad b_1 = \frac{1}{6} - u, \quad b_2 = \frac{4}{6}, \quad b_4 = \frac{1}{6};$$

$$3) \quad c_2 = c_3 = \frac{1}{2}, \quad b_1 = \frac{1}{6}, \quad b_3 = u \neq 0, \quad b_2 = \frac{4}{6} - u, \quad b_4 = \frac{1}{6};$$

$$4) \quad c_2 = 1, \quad c_3 = \frac{1}{2}, \quad b_4 = u \neq 0, \quad b_2 = \frac{1}{6} - u, \quad b_1 = \frac{1}{6}, \quad b_3 = \frac{4}{6}.$$

После того как выбраны  $b_i$  и  $c_i$ , получаем  $a_{43}$  из уравнения j), и тогда два уравнения f) и i) образуют линейную систему для определения  $a_{32}$  и  $a_{42}$ . Определитель этой системы

$$\det \begin{pmatrix} b_3 & b_4 \\ b_3 c_3 c_2 & b_4 c_4 c_2 \end{pmatrix} = b_3 b_4 c_2 (c_4 - c_3),$$

согласно (1.16), не равен нулю. Наконец, из (1.9) находим  $a_{21}$ ,  $a_{31}$  и  $a_{41}$ .

Особенно популярными стали два варианта, которые выбрал Кутта (1901). Это случай 3) при  $\mu = 2/6$  и случай 1) при  $\mu = 1/3$ ,  $v = 2/3$ . Они приведены в табл. 1.2. Оба метода обобщают классические квадратурные формулы, сохраняя их порядок. Первый из них более популярен, второй более точен («Мы будем рассматривать это приближение как лучшее в общем случае...» —Кутта).

Таблица 1.2. Методы Кутты

«Классический»  
метод Рунге—  
Кутты

Правило 3/8

$0$ $\frac{1}{2}$ $\frac{1}{2}$ $\frac{1}{2}$ $1$	$\frac{1}{2}$ $0$ $\frac{1}{2}$ $0$ $0$ $1$	$0$ $\frac{1}{3}$ $\frac{2}{3}$ $\frac{1}{3}$ $1$	$\frac{1}{3}$ $-\frac{1}{3}$ $1$ $1$ $-1$ $1$
$\frac{1}{6}$ $\frac{2}{6}$ $\frac{2}{6}$ $\frac{1}{6}$	$\frac{1}{8}$ $\frac{3}{8}$ $\frac{3}{8}$ $\frac{1}{8}$		

### «Оптимальные» формулы

Предпринималось много исследований, чтобы выбрать возможно «лучшие» из множества различных формул РК 4-го порядка.

Первой попыткой в этом направлении был очень популярный метод, который предложил Гилл (1951). Он преследовал цель уменьшить, насколько возможно, количество требуемой машинной памяти («регистров»). Этот метод широко использовался на первых компьютерах в пятидесятых годах и представляет поэтому исторический интерес. Гилл заметил, что больше всего машинной памяти нужно при вычислении  $k_3$ , когда «требуются регистры для хранения в какой-либо форме» величин

$$y_0 + a_{31}hk_1 + a_{32}hk_2, \quad y_0 + a_{41}hk_1 + a_{42}hk_2, \\ y_0 + b_1hk_1 + b_2hk_2, \quad hk_3.$$

«Ясно, что для третьей стадии будет достаточно трех регистров, если подлежащие хранению величины линейно зависимы, т. е. если»

$$\det \begin{pmatrix} 1 & a_{31} & a_{32} \\ 1 & a_{41} & a_{42} \\ 1 & b_1 & b_2 \end{pmatrix} = 0.$$

Гилл заметил, что это условие удовлетворяется для методов типа 3), если  $u = (1 + \sqrt{0,5})/3$ . Получающийся метод можно тогда переформулировать следующим образом («Каждая вычисленная величина засыпается в регистр, ранее содержавший соответствующую величину предыдущей стадии, больше не нужную»):

$$\begin{aligned} y &:= \text{начальное значение}, \quad k := h f(y), \quad y := y + 0.5k, \quad q := k, \\ k &:= h f(y), \quad y := y + (1 - \sqrt{0.5})(k - q), \\ q &:= (2 - \sqrt{2})k + (-2 + 3\sqrt{0.5})q, \\ k &:= h f(y), \quad y := y + (1 + \sqrt{0.5})(k - q), \\ q &:= (2 + \sqrt{2})k + (-2 - 3\sqrt{0.5})q, \\ k &:= h f(y), \quad y := y + \frac{k}{6} - \frac{q}{3}, \quad (\rightarrow \text{переход к следующему шагу}). \end{aligned} \tag{1.18}$$

Ныне на больших компьютерах с высокой скоростью счета этот метод больше не употребляется, но он может все же представлять интерес для систем уравнений очень высокой размерности.

Делались и другие попытки: для случая 1) подобрать  $u$  и  $v$  в (1.17) так, чтобы погрешность (члены порядка  $h^6$ , см. разд. II.3) стала, насколько возможно, малой. Мы обсудим этот вопрос в разд. II.3.

### Численный пример

«Zu grossen Gewicht darf man natürlich solchen Beispielen nicht beilegen...»

«Конечно, таким примерам нельзя придавать слишком большого веса...»

(В. Кутта, 1901)

Сравним результаты применения пяти различных методов к двум системам уравнений, которые возьмем в качестве тестов. Выберем решение вида

$$\begin{aligned} y_1 &= \exp(\sin(x^2)), \quad y_2 = \exp(5\sin(x^2)), \\ y_3 &= \sin(x^2) + 1, \quad y_4 = \cos(x^2), \end{aligned}$$

все компоненты которого удовлетворяют начальным условиям  $y_i(0) = 1$ . Можно построить несколько систем дифференциальных уравнений, которым удовлетворяет это решение, например:

$$\begin{aligned} y'_1 &= 2xy_4y_1, & y'_2 &= 10xy_4y_1^5, \\ y'_3 &= 2xy_4, & y'_4 &= -2x(y_3 - 1), \end{aligned} \quad (1.19)$$

или более сложную систему:

$$\begin{aligned} y'_1 &= 2xy_2^{1/5}y_4, & y'_2 &= 10x \exp(5(y_3 - 1)) \cdot y_4, \\ y'_3 &= 2xy_4, & y'_4 &= -2x \ln(y_1). \end{aligned} \quad (1.20)$$

Мы применили для их решения пять программ при нескольких различных фиксированных длинах шага. На рис. 1.1 по осям

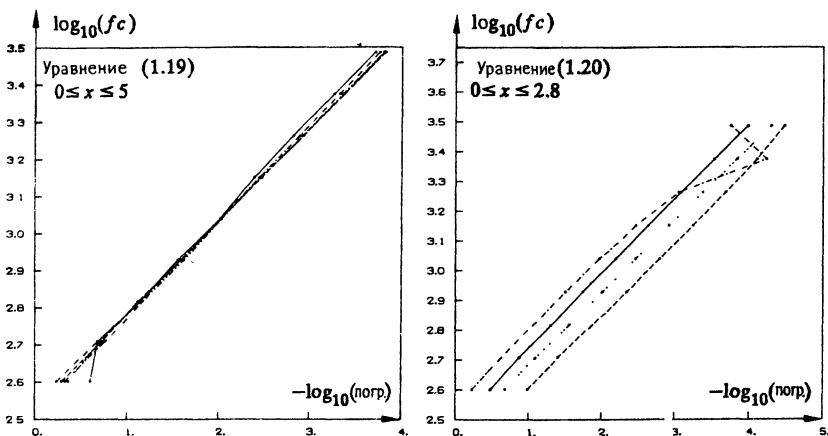


Рис. 1.1. Зависимость максимальной глобальной погрешности от числа обращений к подпрограмме вычисления значения функции (обозначенного символом  $fc$ ).

отложены в логарифмическом масштабе максимальные погрешности и число обращений к вычислению функций, стоящих в правых частях уравнений (которое для четырехстадийных методов в 4 раза больше числа шагов). Масштабы выбраны так, чтобы теоретический наклон графиков (который по теореме 3.4 равен  $1/4$ ) составлял  $45^\circ$ . Эти тесты обычно показывают для классической схемы РК слегка худшие результаты, чем для других схем РК,

но различия оказываются очень малыми. Здесь сравнивались следующие методы:

- классический метод РК (в табл. 1.2 — левая таблица)
- · · · · правило 3/8 Кутты (в табл. 1.2 — правая таблица)
- — — — оптимальная формула (см. упр. 3а из разд. II.3),  $u = 0.3587$ ,  $v = 0.6346$
- . — . — . — . — . Ральстон (1962), Халл (1967),  $u = 0.4$ ,  $v = 0.45$
- . — . — . — . — . — . формула Гилла (1.18)

### Упражнения

- Покажите, что любой  $s$ -стадийный явный метод РК порядка  $s$  при применении к задаче  $\dot{y}' = \lambda y$  ( $\lambda$  — комплексная постоянная) дает

$$y_1 = \left\{ \sum_{j=0}^s \frac{z^j}{j!} \right\} y_0, \quad z = h\lambda.$$

*Указание.* Покажите сначала, что отношение  $y_1/y_0$  должно быть многочленом степени  $s$  от  $z$ , а затем определите его коэффициенты, сравнивая производные  $y_1$  по  $h$  с производными точного решения.

- [Рунге (1895); см. также введение к методам Адамса в разд. III.1.] Теоретическая форма капель жидкости определяется дифференциальным уравнением Лапласа (1805):

$$-z = \alpha^2 \frac{(K_1 + K_2)}{2}, \quad (1.21)$$

где  $\alpha$  — постоянная,  $(K_1 + K_2)/2$  — средняя кривизна и  $z$  — высота (см. рис. 1.2). Если подставим сюда  $1/K_1 = r/\sin \varphi$  и  $K_2 = d\varphi/ds$  (кривизну меридиональной кривой), то получим

$$-2z = \alpha^2 \left( \frac{\sin \varphi}{r} + \frac{d\varphi}{ds} \right), \quad (1.22)$$

где мы положим  $\alpha = 1$ . Добавив уравнения

$$\frac{dr}{ds} = \cos \varphi, \quad \frac{dz}{ds} = -\sin \varphi, \quad (1.22')$$

получим систему трех дифференциальных уравнений для  $\varphi(s)$ ,  $r(s)$ ,  $z(s)$ , где  $s$  — длина дуги. С помощью метода Рунге (1.4) вычислите и постройте различные кривые, соответствующие начальным значениям  $\varphi(0) = 0$ ,  $r(0) = 0$  и  $z(0) = z_0$  ( $z_0 < 0$  для лежащих капель; рассчитайте также форму висящих

капель, изменив соответствующим образом знаки в (1.22)). Проделайте вычисления с различными длинами шага и сравните результаты.

*Указание.* Не забудьте про особенность в начале: из (1.22) и (1.22') при малых  $s$  имеем  $r = s$ ,  $\varphi = As$ , где  $-A = z_0$ , откуда  $(\sin \varphi)/r \rightarrow -z_0$ .

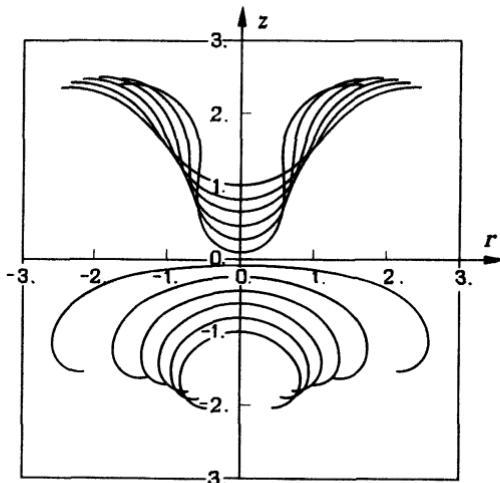


Рис. 1.2. Формы капель.

3. Найдите условия, при которых двухстадийный явный метод РК должен иметь порядок два, и определите все такие методы («...wozu eine weitere Erörterung nicht mehr nötig ist»(Kutta))<sup>1)</sup>.
4. Найдите все трехстадийные методы порядка три (т. е. решите уравнения (1.11, a—d) при  $b_4 = 0$ ).  
*Ответ.*  $c_2 = u$ ,  $c_3 = v$ ,  $a_{32} = v(v-u)/(u(2-3u))$ ,  
 $b_2 = (2-3v)/(6u(u-v))$ ,  $b_3 = (2-3u)/(6v(v-u))$ ,  
 $b_1 = 1 - b_2 - b_3$ ,  $a_{31} = c_3 - a_{32}$ ,  $a_{21} = c_2$   
[Кутта (1901)].
5. Постройте все методы порядка 2 вида

$$\begin{array}{c|cc} 0 & & \\ c_2 & c_2 & \\ c_3 & 0 & c_3 \\ \hline & 0 & 0 & 1 \end{array}$$

<sup>1)</sup>) «... далее обсуждать здесь больше нечего» (Кутта). — *Нем.*

Такие методы «имеют свойство, заключающееся в том, что соответствующий процесс Рунге—Кутты требует относительно мало памяти в компьютере» [Ван дер Хаувен (1977), § 2.7.2]. Примените эти методы к уравнению  $y' = \lambda y$  и сравните с результатом упр. 1.

6. Определите условия, при которых двухстадийные методы РК, не удовлетворяющие условиям (1.9):

$$k_1 = f(x_0 + c_1 h, y_0),$$

$$k_2 = f(x_0 + c_2 h, y_0 + a_{21} h k_1),$$

$$y_1 = y_0 + h(b_1 k_1 + b_2 k_2),$$

будут иметь порядок два. Обсудите возможности использовать эту дополнительную свободу выбора  $c_1$  и  $c_2$  [Оливер (1975)].

## II.2. Условия порядка для методов Рунге—Кутты

«...я слышал сообщение Мерсона...».  
(первое знакомство Дж. Бутчера с методами  
Рунге—Кутты)

В этом разделе мы изучим общую структуру условий, определяющих порядок метода, или условий порядка, как их называют для краткости [Мерсон (1957), Бутчера (1963)]. Со временем появления этих первых работ способ вывода условий порядка прошел большую эволюцию. Он совершенствовался главным образом под влиянием более поздних работ Бутчера, с которым авторы много раз обсуждали эти вопросы. В нашем изложении использовано, кроме того, доказательство «теоремы 6» из работы Хайрера и Баннера (1974); мы опирались также на наш опыт преподавания. В разд. II.11 мы увидим, что те же самые идеи позволяют доказать общую теорему о композиции методов (В-рядов), открывающую подход к условиям порядка для гораздо более широкого класса методов.

Для упрощения вывода целесообразно преобразовать уравнение (1.1) к *автономной форме* путем добавления  $x$  к зависимым переменным:

$$\begin{pmatrix} y \\ x \end{pmatrix}' = \begin{pmatrix} 1 \\ f(x, y) \end{pmatrix}. \quad (2.1)$$

Основная трудность вывода условий порядка состоит в уяснении связи между формулами и соответствующими корневыми помеченными деревьями. Эта связь выявляется наиболее естественным образом, если воспользоваться удачно выбранной системой индексных и тензорных обозначений [как в работах Гилла (1951), Хенричи (1962), Гира (1971)]. Как это принято в тензорной нотации, мы будем обозначать компоненты векторов верхними индексами, причем, чтобы избежать путаницы, будем использовать для этих индексов *заглавные* буквы. Тогда автономную систему общего вида можно записать так:

$$(y^J)' = f^J(y^1, \dots, y^n), \quad J = 1, \dots, n. \quad (2.2)$$

Перепишем прежде всего формулы метода (1.8) для автономного дифференциального уравнения (2.2). Чтобы сделать запись всех

формул (1.8) более симметричной, мы перейдем от функций  $k_i = f(g_i)$  к их аргументам  $g_i$ . Тогда формулы (1.8) примут вид

$$\begin{aligned} g_i^J &= y_0^J + \sum_{l=1}^{i-1} a_{il} h f^J(g_1^1, \dots, g_i^n), \quad i = 1, \dots, s, \\ y_1^J &= y_0^J + \sum_{l=1}^s b_{il} h f^J(g_1^1, \dots, g_i^n). \end{aligned} \quad (2.3)$$

В частности, если система (2.2) происходит из (2.1), то при  $J = 1$

$$g_i^1 = y_0^1 + \sum_{l=1}^{i-1} a_{il} h = x_0 + c_i h.$$

Второе равенство вытекает из условия (1.9), которое, как видим, становится вполне естественным. Если оно удовлетворено, то для вывода условий порядка достаточно рассмотреть только автономное уравнение (2.2).

Как указывалось в разд. II.1, нам надо сравнить ряды Тейлора для  $y_1^J$  и для точного решения. Для этой цели вычислим сначала значения производных  $y_1^J$  и  $g_i^J$  по  $h$  при  $h = 0$ . Ввиду сходства обеих формул достаточно проделать это для  $g_i^J$ . В правые части формул (2.3) входят выражения вида  $h\varphi(h)$ , и мы воспользуемся формулой Лейбница:

$$(h\varphi(h))^{(q)}|_{h=0} = q \cdot (\varphi(h))^{(q-1)}|_{h=0}. \quad (2.4)$$

Теперь приглашаем читателя сделать глубокий вдох, вспомнить основные правила дифференциального исчисления, взять пятачок листов машинных распечаток и на обратной стороне их приступить к следующим вычислениям:

$q = 0$ : согласно (2.3),

$$(g_i^J)^{(0)}|_{h=0} = y_0^J. \quad (2.5; 0)$$

$q = 1$ : из (2.3) и (2.4) получим

$$(g_i^J)^{(1)}|_{h=0} = \sum_l a_{il} f^J|_{y=y_0}. \quad (2.5; 1)$$

$q = 2$ : чтобы применить формулу (2.4), нам понадобится первая производная от  $f^J(g_j)$ :

$$(f^J(g_i))^{(1)} = \sum_K f_K^J(g_i) \cdot (g_i^K)^{(1)}, \quad (2.6; 1)$$

где, как это принято,  $f_K^J$  означает  $\partial f^J / \partial y^K$ . Подставим сюда выражение для  $(g_i^K)^{(1)}$  из формулы (2.5; 1), заменив в ней индексы  $i, j, J$  на  $j, k, K$ . Применяя теперь формулу (2.4) к правой части

равенства (2.3), найдем, что

$$(g_i^J)^{(2)}|_{h=0} = 2 \sum_{l, k} a_{il} a_{jk} \sum_k f_K^J f^K|_{y=y_0}. \quad (2.5; 2)$$

$b = 3$ : сначала продифференцируем выражение (2.6; 1), чтобы найти

$$(f^J(g_i))^{(2)} = \sum_{K, L} f_{KL}^J(g_i) \cdot (g_i^K)^{(1)} (g_i^L)^{(1)} + \sum_K f_K^J(g_i) (g_i^K)^{(2)}. \quad (2.6; 2)$$

Подставим сюда значения производных  $(g_i^K)^{(1)}$  и  $(g_i^K)^{(2)}$  при  $h = 0$ , которые уже найдены (формулы (2.5; 1) и (2.5; 2)), затем опять из (2.3) с помощью (2.4) получим

$$\begin{aligned} (g_i^J)^{(3)}|_{h=0} = & 3 \sum_{l, k, l} a_{il} a_{jk} a_{kl} \sum_{K, L} f_{KL}^J f^K f^L|_{y=y_0} + \\ & + 3 \cdot 2 \sum_{l, k, l} a_{il} a_{jk} a_{kl} \sum_{K, L} f_K^J f_L^K f^L|_{y=y_0}. \end{aligned} \quad (2.5; 3)$$

Для величин  $(y_i^J)^{(q)}|_{h=0}$  ( $q = 1, 2, 3$ ) справедливы те же формулы (2.5; 1—3), если в них заменить  $a_{ij}$  на  $b_j$ .

### Производные точного решения

Производные точного решения получить гораздо легче. Для этого надо взять уравнение (2.2):

$$(y^J)^{(1)} = f^J(y) \quad (2.7; 1)$$

и просто дифференцировать его, подставляя вместо производных правые части (2.2). Для второй производной получим

$$(y^J)^{(2)} = \sum_K f_K^J(y) \cdot (y^K)^{(1)} = \sum_K f_K^J(y) f^K(y). \quad (2.7; 2)$$

Дифференцируя (2.7; 2) еще раз, найдем

$$(y^J)^{(3)} = \sum_{K, L} f_{KL}^J(y) f^K(y) f^L(y) + \sum_{K, L} f_K^J(y) f_L^K(y) f^L(y). \quad (2.7; 3)$$

### Условия для порядка 3

Чтобы порядок метода был равен 3, производные, определяемые выражениями (2.5; 1—3) с заменой  $a_{ij}$  на  $b_j$ , должны быть равны производным (2.7; 1—3), причем для любого дифференци-

ального уравнения. Сопоставляя соответствующие выражения, убеждаемся, что имеет место

**Теорема 2.1.** *Метод Рунге—Кутты (2.3) [или (1.8)] имеет порядок 3 тогда и только тогда, если выполнены следующие равенства:*

$$\sum_l b_l = 1,$$

$$2 \sum_{l, k} b_j a_{jk} = 1,$$

$$3 \sum_{l, k, l} b_j a_{jk} a_{kl} = 1, \quad (2.8)$$

$$6 \sum_{l, k, l} b_j a_{jk} a_{kl} = 1. \quad \blacksquare$$

Эти выражения можно еще более упростить, подставив в них  $\sum_k a_{jk} = c_j$  согласно (1.9), что приводит к формулам (1.11a — d).

### Деревья и элементарные дифференциалы

Теоретически ясно, как продолжать этот процесс, но на практике оно скоро приводит к очень сложным формулам. Поэтому целесообразно перейти к графическому представлению. Для этого заметим, что в каждом члене формулы (2.5; 3) индексы  $j, k, l$  связаны в пары в качестве нижних индексов в  $a_{jk}, a_{jl}, a_{kl}$ , и точно таким же образом индексы  $J, K, L$  связаны попарно в качестве верхних и нижних индексов в выражениях  $f_{KL}^J, f_K^J, f_L^K$ . Графически представить эти связи можно так:

$$t_{31} = \begin{array}{c} l \\ \backslash \\ j \end{array} \quad t_{32} = \begin{array}{c} l \\ \diagup \\ k \\ \diagdown \\ j \end{array} \quad (2.9)$$

(для первого и второго члена соответственно). Назовем эти объекты «помеченными деревьями», потому что это связные графы (деревья), вершины которых помечены индексами суммирования. Их можно

представить и как отображения. Например, помеченным деревьям  $t_{31}$  и  $t_{32}$  соответствуют отображения

$$l \rightarrow j, k \rightarrow f \quad \text{и} \quad l \rightarrow k, k \rightarrow j. \quad (2.9')$$

Эти отображения каждой вершине графа (кроме одной) ставят в соответствие другую вершину, связанную с ней направленным вниз ребром.

**Определение 2.2.** Пусть  $A$  — упорядоченное множество индексов:  $A = \{j < k < l < m < \dots\}$ , и  $A_q$  — его подмножество, состоящее из первых  $q$  индексов. Назовем (корневым) помеченным деревом порядка  $q$  ( $q \geq 1$ ) отображение

$$t: A_q - \{j\} \rightarrow A_q,$$

такое, что  $t(z) < z$  для всех  $z \in A_q - \{j\}$ . Множество всех помеченных деревьев порядка  $q$  обозначается  $LT_q$ . Будем называть  $z$  «сыном»  $t(z)$ , а  $t(z)$  — «отцом»  $z$  ( $t$  — отображение «сыновей» на «отцов»). Вершина  $j$  («праотец» всей «династии») называется корнем  $t$ . Порядок  $q$  помеченного дерева равен числу всех его вершин и обычно обозначается  $\rho(t)$ .

**Определение 2.3.** Назовем выражение

$$F^J(t)(y) = \sum_{K, L, \dots} f_K^J(y) f_L^K(y) f_{\dots}^L(y) \dots$$

элементарным дифференциалом, соответствующим помеченному дереву  $t \in LT_q$ . Здесь суммирование производится по  $q-1$  индексам  $K, L, \dots$ , (которые соответствуют множеству  $A_q - \{j\}$ ), и каждое слагаемое является произведением  $q$  символов  $f$ , верхние индексы которых пробегают все множество  $A_q$  вершин  $t$ , а нижние индексы — множество соответствующих «сыновей».

Если множество  $A_q$  записано в виде

$$A_q = \{j_1 < j_2 < \dots < j_q\}, \quad (2.10)$$

то определение  $F(t)$  можно переписать следующим образом:

$$F^{J_1}(t) = \sum_{J_2, \dots, J_q} \prod_{i=1}^q f_t^{J_i}(J_i), \quad (2.11)$$

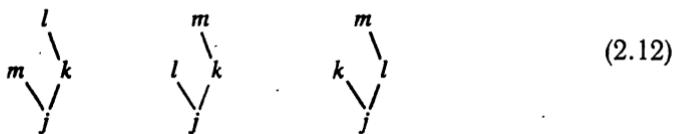
поскольку «сыновья» индекса составляют его полный прообраз при отображении  $t$ .

Примерами элементарных дифференциалов являются

$$\sum_{K, L} f_K^J f_L^K f^L \quad \text{и} \quad \sum_{K, L} f_K^J f_L^K f^L,$$

соответствующие помеченным деревьям  $t_{31}$  и  $t_{32}$  (2.9). Эти выражения входят в формулы (2.5; 3) и (2.7; 3).

Следующие три помеченных дерева



имеют одинаковую топологическую структуру; более того, их элементарные дифференциалы

$$\sum_{K, L, M} f_{KM}^J f_L^K f^L, \quad \sum_{K, L, M} f_{KL}^J f_L^K f_M^M, \quad \sum_{K, L, M} f_{LK}^J f_K^K f_M^L f^M \quad (2.12')$$

совпадают, поскольку они отличаются лишь обозначениями индексов суммирования. Этот пример подсказывает дать следующее.

**Определение 2.4.** Два помеченных дерева  $t$  и  $u$  назовем **эквивалентными**, если они имеют одинаковый порядок  $q$  и существует подстановка  $\sigma: A_q \rightarrow A_q$ , такая, что  $\sigma(j) = j$  и  $t\sigma = \sigma u$  на множестве  $A_q - \{j\}$ .

Легко видеть, что этим мы определили отношение эквивалентности на множестве  $LT_q$ .

**Определение 2.5.** Класс эквивалентности помеченных деревьев порядка  $q$  называют (*корневым*) деревом порядка  $q$ . Множество всех деревьев порядка  $q$  обозначается  $T_q$ . Порядок дерева определяется как порядок его представителя из  $LT_q$  и тоже обозначается  $p(t)$ . Обозначим, далее, через  $\alpha(t)$  (для  $t \in T_q$ ) число элементов в классе эквивалентности  $t$ , т. е. число различных возможных монотонных индексаций  $t$ .<sup>1)</sup>

Геометрически дерево отличается от помеченного дерева тем, что у него опущены все индексы. Часто бывает целесообразно включить в рассмотрение пустое дерево  $\emptyset$  как единственное дерево порядка 0. Единственное дерево порядка 1 обозначают  $\tau$ . Представители всех деревьев порядка  $\leq 5$  приведены в табл. 2.1. Число деревьев для порядков 1, 2, ..., 10 представлено в табл. 2.2.

<sup>1)</sup> Речь идет о таких расстановках индексов, при которых сохраняется отношение порядка, заданное на множестве  $A_q$ , т. е. по существу о монотонных (изотоничных) отображениях. — Прим. ред.

Таблица 2.1. Деревья и элементарные дифференциалы вплоть до порядка пять

$q$	$t$	граф	$\gamma(t)$	$\alpha(t)$	$F^J(t)(y)$	$\Phi_j(t)$
0	$\emptyset$	$\emptyset$	1	1	$y^J$	
1	$\tau$	$\bullet_j$	1	1	$f^J$	1
2	$t_{21}$		2	1	$\sum_K f_K^J f^K$	$\sum_k a_{jk}$
3	$t_{31}$		3	1	$\sum_{K,L} f_{KL}^J f^K f^L$	$\sum_{k,l} a_{jk} a_{jl}$
	$t_{32}$		6	1	$\sum_{K,L} f_K^J f_L^K f^L$	$\sum_{k,l} a_{jk} a_{kl}$
4	$t_{41}$		4	1	$\sum_{K,L,M} f_{KLM}^J f^K f^L f^M$	$\sum_{k,l,m} a_{jk} a_{jl} a_{jm}$
	$t_{42}$		8	3	$\sum_{K,L,M} f_{KM}^J f_L^K f^L f^M$	$\sum_{k,l,m} a_{jk} a_{kl} a_{jm}$
	$t_{43}$		12	1	$\sum_{K,L,M} f_K^J f_{LM}^K f^L f^M$	$\sum_{k,l,m} a_{jk} a_{kl} a_{km}$
	$t_{44}$		24	1	$\sum_{K,L,M} f_K^J f_L^K f_M^L f^M$	$\sum_{k,l,m} a_{jk} a_{kl} a_{lm}$
5	$t_{51}$		5	1	$\sum f_{KLMP}^J f^K f^L f^M f^P$	$\sum a_{jk} a_{jl} a_{jm} a_{jp}$
	$t_{52}$		10	6	$\sum f_{KMP}^J f_L^K f^L f^M f^P$	$\sum a_{jk} a_{kl} a_{jm} a_{jp}$
	$t_{53}$		15	4	$\sum f_{KPL}^J f_M^K f^L f^M f^P$	$\sum a_{jk} a_{kl} a_{km} a_{jp}$
	$t_{54}$		30	4	$\sum f_{KPL}^J f_M^K f^L f^M f^P$	$\sum a_{jk} a_{kl} a_{lm} a_{jp}$
	$t_{55}$		20	3	$\sum f_{KMF}^J f_L^K f_P^L f^M f^P$	$\sum a_{jk} a_{kl} a_{jm} a_{mp}$
	$t_{56}$		20	1	$\sum f_K^J f_{LMP}^K f^L f^M f^P$	$\sum a_{jk} a_{kl} a_{km} a_{kp}$
	$t_{57}$		40	3	$\sum f_K^J f_{LP}^K f_M^L f^M f^P$	$\sum a_{jk} a_{kl} a_{bm} a_{kp}$
	$t_{58}$		60	1	$\sum f_K^J f_L^K f_{MP}^L f^M f^P$	$\sum a_{jk} a_{kl} a_{lm} a_{lp}$
	$t_{59}$		120	1	$\sum f_K^J f_L^K f_M^L f_P^M f^P$	$\sum a_{jk} a_{kl} a_{lm} a_{mp}$

Таблица 2.2. Число деревьев до 10-го порядка

$q$	1	2	3	4	5	6	7	8	9	10
card ( $T_q$ )	1	1	2	4	9	20	48	115	286	719

### Разложение Тейлора для точного решения

Теперь мы можем для  $q$ -й производной точного решения сформулировать результат в общем виде.

**Теорема 2.6.** Для точного решения уравнения (2.2) справедлива формула

$$(y^J)^{(q)}|_{x=x_0} = \sum_{t \in LT_q} F^J(t)(y_0) = \sum_{t \in T_q} \alpha(t) F^J(t)(y_0). \quad (2.7; q)$$

*Доказательство.* Теорема справедлива при  $q = 1, 2, 3$  [см. формулы (2.7; 1—3)]. Чтобы вычислить, например, четвертую

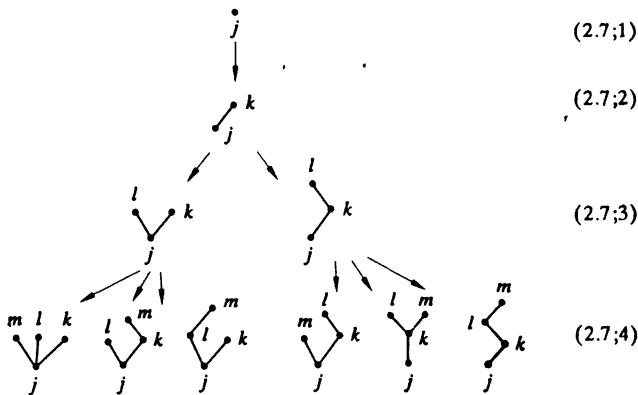


Рис. 2.1. Производные точного решения.

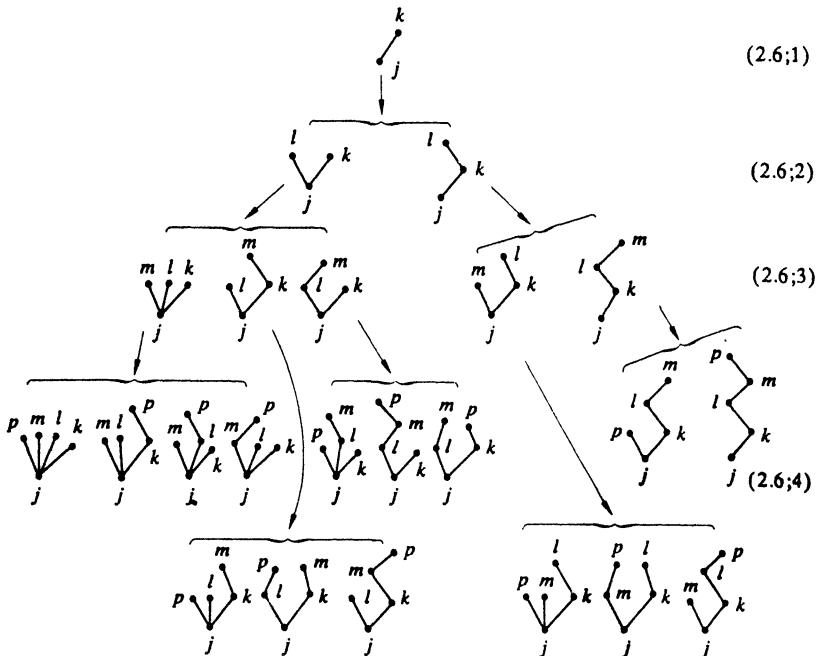
производную, надо продифференцировать формулу (2.7; 3), в которой два члена соответствуют двум деревьям (2.9) и каждый из них содержит три множителя  $f_{...}$ , отвечающие трем вершинам этих деревьев. Дифференцирование этих членов по правилу Лейбница и подстановка правой части уравнения (2.2) вместо производных  $y'$  геометрически интерпретируется как добавление к каждой вершине нового ребра с вершиной, помеченной новым индексом суммирования (рис. 2.1).

Очевидно, что в этом процессе для  $q$ -й производной появляются все помеченные деревья порядка  $q$ , и каждое из них только один раз.

Если сгруппировать теперь все члены с одинаковыми элементарными дифференциалами, мы получим второе выражение в формуле (2.7;  $q$ ). ■

## Формула Фаа ди Бруно

Займемся теперь вычислением  $q$ -й производной численного решения  $y_1$  и величин  $g_j$ . Для этого надо сначала получить обобщение формул (2.6; 1) («цепного правила») и (2.6; 2) на случай  $q$ -й производной от суперпозиции двух функций. На рис. 2.2 приведено представление этих формул в виде графов.

Рис. 2.2. Производные функции  $f^J(g)$ .

Формула (2.6; 2) состоит из двух членов, причем первый член содержит три множителя, а второй — только два, так что в его графике индекс  $l$  — «пустой», в формуле он отсутствует, его назначение — показать, что надо взять вторую производную. Следовательно, дифференцируя выражение (2.6; 2), мы получим пять членов, которые выпишем здесь для удобства читателя (и невзирая на неудобства при их наборе!..):

$$\begin{aligned}
 (f^J(g))^{(3)} = & \sum_{K, L, M} f_{KLM}^J(g) \cdot (g^K)^{(1)} (g^L)^{(1)} (g^M)^{(1)} + \\
 & + \sum_{K, L} f_{KL}^J(g) \cdot (g^K)^{(2)} (g^L)^{(1)} + \sum_{K, L} f_{KL}^J(g) \cdot (g^K)^{(1)} (g^L)^{(2)} + \\
 & + \sum_{K, M} f_{KM}^J(g) \cdot (g^K)^{(2)} (g^M)^{(1)} + \sum_K f_K^J(g) \cdot (g^K)^{(3)}. \quad (2.6; 3)
 \end{aligned}$$

Соответствующие деревья изображены в третьем ряду рис. 2.2. При каждом дифференцировании мы производим следующие операции:

1) дифференцируем первый множитель  $f'_{k_1 \dots k_m}$ , т. е. добавляем к дереву новую ветвь, идущую от корня  $j$ ;

2) увеличиваем на единицу порядок производной каждого из множителей  $g$ , что графически представляется удлинением соответствующей ветви.

Каждый раз мы добавляем новый индекс. Все получаемые при этом деревья имеют «специальный» вид: разветвления у них возможны только у корня.

**Определение 2.7.** Обозначим  $LS_q$  множество *специальных помеченных деревьев порядка  $q$* , т. е. таких помеченных деревьев, у которых разветвления встречаются только у корня.

**Лемма 2.8** (формула Фаа ди Бруно). *При  $q \geq 1$  справедлива формула*

$$(f'(g))^{(q-1)} = \sum_{u \in LS_q} \sum_{K_1, \dots, K_m} f'_{K_1, \dots, K_m}(g) \cdot (g^{K_1})^{(\delta_1)} \dots (g^{K_m})^{(\delta_m)}. \quad (2.6; q-1)$$

Для каждого  $u \in LS_q$  здесь  $m$  — число выходящих из корня ветвей,  $\delta_1, \dots, \delta_m$  — числа вершин на этих ветвях, так что  $q = 1 + \delta_1 + \dots + \delta_m$ . ■

*Замечание.* Здесь отсутствуют привычные коэффициенты, указывающие кратности членов, поскольку мы пользуемся помеченными деревьями.

### Производные численного решения

«Трудно сохранить хладнокровие, когда обсуждаешь различные производные...» (С. Гилл, 1956)

Чтобы обобщить формулы (2.5; 1—3), нам потребуются следующие определения:

**Определение 2.9.** Пусть  $t$  — помеченное дерево с корнем  $j$ . Введем величину

$$\Phi_j(t) = \sum_{k, l, \dots} a_{jk} a_{l\dots},$$

где суммирование производится по  $q - 1$  индексам  $k, l, \dots$  (как в определении 2.3), каждое слагаемое является произведением  $q - 1$  коэффициентов  $a$ , у каждого из которых индексы соста-

вляют одну из пар «отец» — «сын» дерева  $t$ . Если множество  $A_q$  записано в виде (2.10), то

$$\Phi_{j_1}(t) = \sum_{j_2, \dots, j_q} a_t(j_2), j_2 \dots a_t(j_q), j_q. \quad (2.13)$$

**Определение 2.10.** Для каждого  $t \in LT_q$  определим величину  $\gamma(t)$  как произведение  $\rho(t)$  на порядки всех деревьев, которые получаются из  $t$ , если последовательно удалять их корни [вместе с инцидентными им ребрами, см. рис. 2.3 или формулу (2.17)].

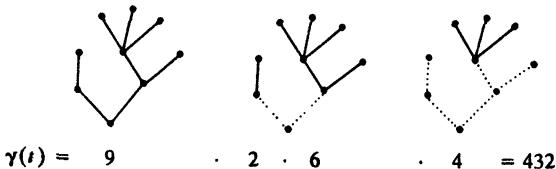


Рис. 2.3. Пример определения величины  $\gamma(t)$ .

Обе введенные величины, разумеется, не зависят от индексации, поэтому как  $\Phi_j(t)$ , так и  $\gamma(t)$  имеют смысл и для  $t \in T_q$ . Примеры приведены в табл. 2.1.

**Теорема 2.11.** Для производных  $g_i^J$  имеет место равенство

$$(g_i^J)^{(q)}|_{h=0} = \sum_{t \in LT_q} \gamma(t) \sum_l a_{lj} \Phi_l(t) F^J(t)(y_0). \quad (2.5; q)$$

Для производных определяемого методом (2.3) численного решения  $y_1^J$  справедлива формула

$$\begin{aligned} (y_1^J)^{(q)}|_{h=0} &= \sum_{t \in LT_q} \gamma(t) \sum_l b_l \Phi_l(t) F^J(t)(y_0) = \\ &= \sum_{t \in T_q} \alpha(t) \gamma(t) \sum_l b_l \Phi_l(t) F^J(t)(y_0). \end{aligned} \quad (2.14)$$

**Доказательство.** Благодаря сходству выражений (2.3) для  $y_1$  и  $g_i$  достаточно доказать первое равенство. Мы докажем его индукцией по  $q$ , следуя тем же путем, на котором были получены равенства (2.5; 1—3). Сначала применим к (2.3) формулу Лейбница (2.4), что дает

$$(g_i^J)^{(q)}|_{h=0} = q \sum_l a_{lj} (f^J(g_i))^{(q-1)}|_{y=y_0}. \quad (2.15)$$

Далее воспользуемся формулой Фаа ди Бруно (лемма 2.8). Наконец, используем индуктивное предположение о справедливости

формул (2.5; 1)—(2.5;  $q-1$ ) и подставим соответствующие выражения в формулу (2.6;  $q-1$ ) вместо производных  $(g_j^{K_s})^{(\delta_s)}$ , приняв во внимание, что всегда  $\delta_s < q$ . После несложных преобразований, изменив порядок суммирования, получим

$$\begin{aligned} (g_i^J)^{(q)}|_{h=0} = & q \sum_{u \in LS_q} \sum_{t_1 \in LT_{\delta_1}} \dots \sum_{t_m \in LT_{\delta_m}} \gamma(t_1) \dots \gamma(t_m) \times \\ & \times \sum_l a_{ij} \sum_k a_{j k_1} \Phi_{k_1}(t_1) \dots \sum_{k_m} a_{j k_m} \Phi_{k_m}(t_m) \times \\ & \times \sum_{K_1, \dots, K_m} f_{K_1, \dots, K_m}^J(y_0) F^{K_1}(t_1)(y_0) \dots F^{K_m}(t_m)(y_0). \end{aligned} \quad (2.16)$$

Теперь мы подошли к самому трудному месту доказательства: надо уяснить себе, что каждой совокупности деревьев

$$(u, t_1, \dots, t_m), \text{ где } u \in LS_q, \quad t_s \in LT_{\delta_s},$$

соответствует определенное  $t \in LT_q$ , такое, что

$$\gamma(t) = q \cdot \gamma(t_1) \dots \gamma(t_m), \quad (2.17)$$

$$F^J(t)(y) = \sum_{K_1, \dots, K_m} f_{K_1, \dots, K_m}^J(y) F^{K_1}(t_1)(y) \dots F^{K_m}(t_m)(y), \quad (2.18)$$

$$\Phi_J(t) = \sum_{k_1, \dots, k_m} a_{j k_1} \dots a_{j k_m} \Phi_{k_1}(t_1) \dots \Phi_{k_m}(t_m). \quad (2.19)$$

Это помеченное дерево  $t$  получается заменой ветвей  $u$  деревьями  $t_1, \dots, t_m$ , и соответствующие индексы переносятся с  $u$  на  $t$  в естественном порядке, т. е. сохраняется их упорядоченность (примеры приведены на рис. 2.4). Таким способом оказываются построенными все деревья  $t \in LT_q$ , причем каждое возникает только один раз. Итак, после подстановки (2.17), (2.18) и (2.19) в формулу (2.16) последняя перейдет в формулу (2.5;  $q$ ). ■

Описанное выше построение  $t$  можно использовать также для индуктивного определения деревьев. Заметим сначала, что класс эквивалентности  $t$  зависит только от классов эквивалентности  $t_1, \dots, t_m$  (это видно на рис. 2.4).

**Определение 2.12.** Будем обозначать

$$t = [t_1, \dots, t_m] \quad (2.20)$$

дерево, после удаления корня которого (вместе с инцидентными ему ребрами) останутся деревья  $t_1, \dots, t_m$  (рис. 2.5).

С помощью (2.20) все деревья могут быть выражены в конечном счете через  $\tau$ . Например,  $t_{21} = [\tau]$ ,  $t_{31} = [\tau, \tau]$ ,  $t_{32} = [[\tau]]$ , и т. д.

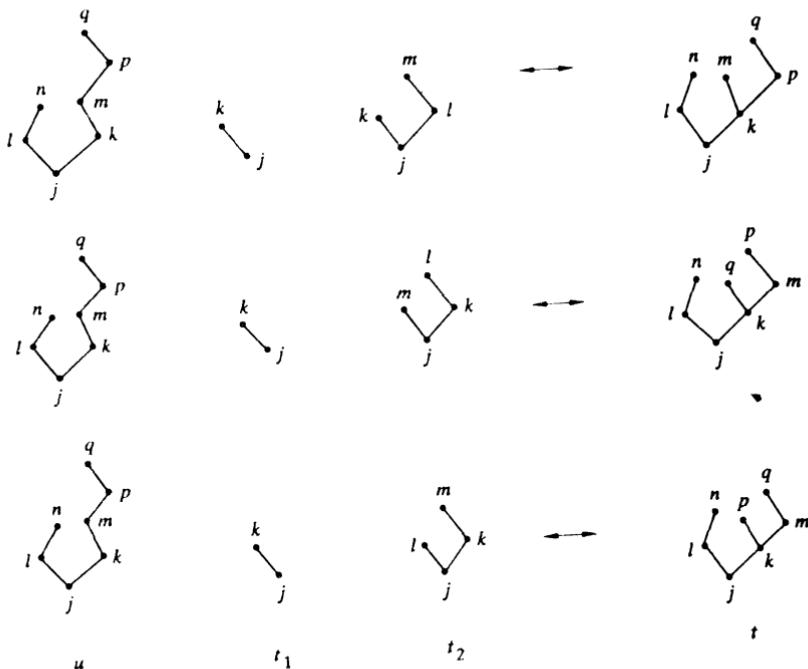
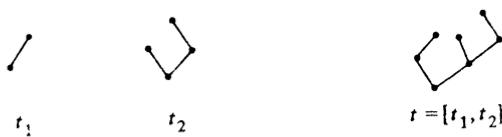
Рис. 2.4. Пример взаимно однозначного соответствия  $(u, t_1, \dots, t_m) \leftrightarrow t$ .

Рис. 2.5. Рекурсивное определение деревьев.

### Условия порядка

Теперь, сопоставляя теоремы 2.6 и 2.11, мы получим окончательный результат:

**Теорема 2.13.** Чтобы метод Рунге—Кутты (1.8) имел порядок  $p$ , необходимо и достаточно выполнения равенств

$$\sum_l b_j \Phi_j(t) = \frac{1}{\gamma(t)} \quad (2.21)$$

для всех деревьев  $t$  порядка, меньшего или равного  $p$ .

**Доказательство.** Достаточность вытекает из проведенных выше рассуждений. Для доказательства необходимости нужно, чтобы

элементарные дифференциалы, соответствующие различным деревьям, были действительно независимыми (см. ниже упр. 3 и 4). ■

Эта теорема позволяет легко сосчитать число условий порядка для любого  $p$ . Результаты подсчета, сделанного с помощью табл. 2.2, приведены в табл. 2.3. Читателю, наверно, теперь понятно, что конструирование формул Рунге—Кутты высоких порядков — нелегкая задача.

Таблица 2.3. Число условий порядка

Порядок $p$	1	2	3	4	5	6	7	8	9	10
Число условий	1	2	4	8	17	37	85	200	486	1205

Пример. Для дерева  $t_{42}$  из табл. 2.1 имеем

$$\sum_{j, k, l, m} b_j a_{jk} a_{jl} a_{km} = \sum_{j, k} b_j a_{jk} c_j c_k = \frac{1}{8}$$

(второе выражение получено с помощью условия (1.9)). Получилась как раз формула (1.11f). Все остальные условия (1.11) соответствуют другим деревьям, у которых порядок  $p \leq 4$ .

### Упражнения

1. Найдите все деревья порядка 6 и порядка 7.

*Указание.* Рассмотрите все представления  $p - 1$  в виде суммы положительных целых чисел и подставьте вместо каждого члена этих сумм все известные деревья соответствующих низших порядков. Вы можете также написать программу для произвольного порядка  $p$  и воспользоваться компьютером.

2. Обозначим число деревьев порядка  $q$  через  $a_q$ . Докажите, что

$$a_1 + a_2 x + a_3 x^2 + a_4 x^3 + \dots = (1 - x)^{-a_1} (1 - x^2)^{-a_2} (1 - x^3)^{-a_3} \dots$$

Сравните результат с табл. 2.2.

3. Вычислите элементарные дифференциалы из табл. 2.1 для случая скалярного неавтономного уравнения первого порядка, т. е. уравнения (2.1), где  $f^1 = 1$  и  $f^2 = f(x, y)$ . Это поможет вам понять трудности, с которыми встретились первые авторы (Кутта, Нюстрём, Гута) при поисках условий для высших порядков. Обратите внимание, что в этом случае выражения для  $t_{54}$  и  $t_{57}$  совпадают, так что здесь условие теоремы 2.13 для порядка 5 является достаточным, но не необходимым.

**Указание.** Для  $t_{54}$ , к примеру, мы имеем ненулевые производные, если только  $K = L = 2$ . Заставляя индексы  $M$  и  $P$  пребегать значения от 1 до 2, получим тогда

$$F^2(t) = (f_x + ff_y)(f_{yy} + ff_{yy})f_y$$

(см. также Бутчер (1963а)).

4. Покажите, что для любого  $t \in T_q$  существует дифференциальная система, такая, что для нее  $F^1(t)(y_0) = 1$ , а для всех остальных деревьев и выполняется равенство  $F^1(u)(y_0) = 0$ .

**Указание.** Для дерева  $t_{54}$  такой системой будет

$$y'_1 = y_2 y_5, \quad y'_2 = y_3, \quad y'_3 = y_4, \quad y'_4 = 1, \quad y'_5 = 1$$

со всеми начальными значениями, равными нулю. Уясните себе этот факт, а также общую формулу

$$y'_{\text{«отца»}} = \prod y_{\text{«сыновей»}}.$$

5. Кутта (1901) утверждал, что схема, приведенная в табл. 2.4, имеет порядок 5. Был ли он прав? Попытайтесь исправить приведенные значения.

Таблица 2.4. Один из методов Кутты

0						
$\frac{1}{3}$		$\frac{1}{3}$				
$\frac{2}{5}$	$\frac{4}{25}$	$\frac{6}{25}$				
1	$\frac{1}{4}$	$-3$	$\frac{15}{4}$			
$\frac{2}{3}$	$\frac{6}{81}$	$\frac{90}{81}$	$-\frac{50}{81}$	$\frac{8}{81}$		
$\frac{4}{5}$	$\frac{7}{30}$	$\frac{18}{30}$	$-\frac{5}{30}$	$\frac{4}{30}$	0	
	$\frac{48}{192}$	0	$\frac{125}{192}$	0	$-\frac{81}{192}$	$\frac{100}{192}$

**Ответ.** Значения коэффициентов  $a_{6j}$  ( $j = 1, \dots, 5$ ) должны быть  $(6, 36, 10, 8, 0)/75$ ; правильные значения коэффициентов  $b_j$  равны  $(23, 0, 125, 0, -81, 125)/192$  [Нюстрэм (1925)].

6. Проверьте равенство

$$\sum_{\rho(t)=\rho} \alpha(t) = (\rho - 1)!$$

7. Докажите, что при применении метода РК к линейной системе

$$y' = A(x)y + g(x) \quad (2.22)$$

локальная погрешность  $y(x_0 + h) - y_1$  будет величиной  $O(h^{p+1})$  в том и только том случае, если

$$\sum_l b_l c_l^{q-1} = \frac{1}{q} \quad \text{для } q \leq p,$$

$$\sum_{j, k} b_j c_j^{q-1} a_{jk} c_k^{r-1} = \frac{1}{(q+r)r} \quad \text{для } q+r \leq p,$$

$$\sum_{j, k, l} b_j c_j^{q-1} a_{jk} c_k^{r-1} a_{kl} c_l^{s-1} = \frac{1}{(q+r+s)(r+s)s} \quad \text{для } q+r+s \leq p,$$

и т. д.

*Указание.* Запишите систему (2.22) в автономной форме и исследуйте, какие элементарные дифференциалы равны нулю тождественно. [См. также Крузé (1975)].

## II.3. Оценка погрешности и сходимость методов Рунге — Кутты

«Es fehlt indessen noch der Beweis dass diese Näherungs—Verfahren convergent sind oder, was praktisch wichtiger ist, es fehlt ein Kriterium, um zu ermitteln, wie klein die Schritte gemacht werden müssen, um eine vorgeschriebene Genauigkeit zu erreichen».

«Между тем еще нет доказательства, что эти приближенные методы сходятся, или, что практически важнее, нет критерия, определяющего, сколь малыми надо сделать шаги, чтобы достичь предписанной точности».

(Рунге, 1905).

Со времен работы Лагранжа (1797) и особенно Коши всякий установленный численно результат принято сопровождать надежной оценкой погрешности («...l'erreur commise sera inférieure à...»)<sup>1)</sup>. Лагранж дал известные оценки погрешности многочленов Тейлора, а Коши вывел оценки для погрешности метода ломаных Эйлера (см. разд. I.7). Через несколько лет после первых успехов методов Рунге—Кутты Рунге (1905) также пришел к заключению, что для этих методов нужны оценки погрешностей.

### Строгие оценки погрешности

Способ, которым Рунге получил оценку погрешности, делаемой на одном шаге («локальной погрешности»), может быть описан в нескольких строках (далее следует свободный перевод):

«Для метода порядка  $p$  рассмотрим локальную погрешность

$$e(h) = y(x_0 + h) - y_1 \quad (3.1)$$

и воспользуемся ее тейлоровским разложением:

$$e(h) = e(0) + he'(0) + \dots + \frac{h^p}{p!} e^{(p)}(0h), \quad (3.2)$$

где  $0 < \theta < 1$  и  $e(0) = e'(0) = \dots = e^{(p)}(0) = 0$ .

Явное вычисление  $e^{(p)}(h)$  дает выражение вида

$$e^{(p)}(h) = E_1(h) + hE_2(h), \quad (3.3)$$

<sup>1)</sup> «... совершаемая ошибка была бы меньше, чем ...». — Франц.

где  $E_1(h)$  и  $E_2(h)$  содержат частные производные  $f$  до порядков  $p-1$  и  $p$  соответственно. Далее, поскольку  $e^{(p)}(0) = 0$ , имеем  $E_1(0) = 0$ . Таким образом, если ограничены все частные производные  $f$  до порядка  $p$  включительно, имеем  $E_1(h) = O(h)$  и  $E_2(h) = O(1)$ . Следовательно, существует постоянная  $C$ , такая что  $|e^{(p)}(h)| \leq C h$  и

$$|e(h)| \leq C \frac{h^{p+1}}{p!}. \quad (3.4)$$

Бибербах (1923, разд. 1-й, гл. II, § 7) использовал несколько иной подход [более подробное объяснение см. Бибербах (1951)]. Запишем

$$e(h) = y(x_0 + h) - y_1 = y(x_0 + h) - y_0 - h \sum_{i=1}^s b_i k_i \quad (3.5)$$

и воспользуемся тейлоровскими разложениями

$$\begin{aligned} y(y_0 + h) &= y_0 + y'(x_0)h + y''(x_0) \frac{h^2}{2!} + \dots + y^{(p+1)}(\theta h) \frac{h^{p+1}}{(p+1)!}, \\ k_i(h) &= k_i(0) + k'_i(0)h + \dots + k_i^{(p)}(\theta_i h) \frac{h^p}{p!}. \end{aligned} \quad (3.6)$$

Для векторных функций эти формулы справедливы покомпонентно (возможно, с различными  $\theta$ ). В силу условий порядка первые члены разложения (3.5) по степеням  $h$  обращаются в нуль. Таким образом, справедлива

**Теорема 3.1.** *Если метод Рунге—Кутты (1.8) имеет порядок  $p$  и если все частные производные  $f(x, y)$  до порядка  $p$  включительно существуют (и непрерывны), то локальная погрешность метода (1.8) допускает следующую строгую оценку:*

$$\begin{aligned} \|y(x_0 + h) - y_1\| &\leq h^{p+1} \left( \frac{1}{(p+1)!} \max_{t \in [0, 1]} \|y^{(p+1)}(x_0 + th)\| + \right. \\ &\quad \left. + \frac{1}{p!} \sum_{i=1}^s |b_i| \max_{t \in (0, 1]} \|k_i^{(p)}(th)\| \right), \end{aligned} \quad (3.7)$$

или

$$\|y(x_0 + h) - y_1\| \leq Ch^{p+1}. \quad \blacksquare \quad (3.8)$$

Продемонстрируем этот результат, применяя к скалярному дифференциальному уравнению первый метод Рунге (1.4), который имеет порядок  $p = 2$ . Дифференцируя (1.1), получим

$$y^{(3)}(x) = (f_{xx} + 2f_{xy}f + f_{yy}f^2 + f_y(f_x + f_yf))(x, y(x)). \quad (3.9)$$

Вторая производная величины  $k_2(h) = f\left(x_0 + \frac{h}{2}, y_0 + \frac{h}{2}f_0\right)$  имеет вид

$$k_2^{(2)}(h) = \frac{1}{4} \left( f_{xx}\left(x_0 + \frac{h}{2}, y_0 + \frac{h}{2}f_0\right) + 2f_{xy}(\dots)f_0 + f_{yy}(\dots)f_0^2 \right) \quad (3.10)$$

(здесь  $f_0$  означает  $f(x_0, y_0)$ ). Если условия теоремы 3.1 выполнены, то легко видеть, что выражения (3.9) и (3.10) ограничены постоянной, которая не зависит от  $h$ , что и дает оценку (3.8).

### Главный член погрешности

Для методов высших порядков строгие оценки погрешностей, подобные (3.7), становятся очень непрактичными. Поэтому гораздо более реалистично рассматривать первый ненулевой член в тейлоровском разложении погрешности. Для автономных систем уравнений (2.2) член погрешности проще всего получить, вычитая друг из друга ряды Тейлора и пользуясь формулами (2.14) и (2.7;  $q$ ).

**Теорема 3.2.** *Если метод Рунге—Кутты имеет порядок  $p$  и если  $f$  непрерывно дифференцируема  $p+1$  раз, то для главного члена погрешности имеем:*

$$y^J(x_0 + h) - y_1^J = \frac{h^{p+1}}{(p+1)!} \sum_{t \in T_{p+1}} \alpha(t) e(t) F^J(t)(y_0) + O(h^{p+2}). \quad (3.18)$$

где

$$e(t) = 1 - \gamma(t) \sum_j b_j \Phi_j(t). \quad ■ \quad (3.19)$$

По поводу величин  $\gamma(t)$  и  $\Phi_j(t)$  см. определения 2.9 и 2.10, а также формулы (2.17) и (2.19). Выражения  $e(t)$  называются коэффициентами погрешности.

**Пример 3.3.** Для двухпараметрического семейства методов РК (1.17) 4-го порядка коэффициенты погрешности для девяти деревьев из табл. 2.1 таковы ( $c_2 = u$ ,  $c_3 = v$ ):

$$\begin{aligned} e(t_{51}) &= -\frac{1}{4} + \frac{5}{12}(u+v) - \frac{5}{6}uv, & e(t_{52}) &= \frac{5}{12}v - \frac{1}{4}, \\ e(t_{53}) &= \frac{5}{8}u - \frac{1}{4}, & e(t_{54}) &= -\frac{1}{4}, \\ e(t_{55}) &= 1 - \frac{5(b_4 + b_3(3-4v)^2)}{144b_3b_4(1-v)^2}, & & \\ e(t_{56}) &= -4e(t_{51}), & e(t_{57}) &= -4e(t_{52}), \\ e(t_{58}) &= -4e(t_{53}), & e(t_{59}) &= -4e(t_{54}). \end{aligned} \quad (3.20)$$

*Доказательство.* Последние четыре формулы следуют из (1.12). Выражение для  $e(t_{59})$  тривиально, а выражения для  $e(t_{58})$  и  $e(t_{57})$  следуют из (1.11h). Далее,

$$e(t_{51}) = 5 \int_0^1 t(t-1)(t-u)(t-v) dt$$

выражает погрешность квадратуры. Для  $e(t_{55})$  удобнее всего ввести величину  $c'_i = \sum_j a_{ij} c_j$ , тогда  $e(t_{55}) = 1 - 20 \sum_i b_i c'_i c'_i$ . Теперь из (1.11d, f) получаем:

$$c'_1 = c'_2 = 0 \quad b_3 c'_3 = \frac{1}{24(1-v)}, \quad b_4 c'_4 = \frac{3-4v}{24(1-v)}. \quad \blacksquare$$

Для «классического» метода 4-го порядка (табл. 1.2a) эти коэффициенты погрешности нашел Кутта (1901) [см. также Лоткин (1951)]:

$$\left( -\frac{1}{24}, -\frac{1}{24}, \frac{1}{16}, -\frac{1}{4}, -\frac{2}{3}, \frac{1}{6}, \frac{1}{6}, -\frac{1}{4}, 1 \right).$$

Кутта отметил, что для второго метода (табл. 1.2b) коэффициенты погрешности

$$\left( -\frac{1}{54}, \frac{1}{36}, -\frac{1}{24}, -\frac{1}{4}, -\frac{1}{9}, \frac{2}{27}, -\frac{1}{9}, \frac{1}{6}, 1 \right)$$

таковы, что, за исключением 4-го и 9-го коэффициентов, все они меньше, чем у предыдущего метода («Als besser noch erweist sich...»<sup>1)</sup>). Для определения оптимальных коэффициентов (1.17) Ральстоном (1962) (и многими другими авторами) были проведены весьма кропотливые вычисления. Решения, минимизирующие константы (3.20), рассматриваются в упр. 3 ниже.

### Оценка глобальной погрешности

«Das war auch eine aufregende Zeit...»

«Да, это было волнующее время...»

(П. Хенричи, 1983)

Глобальной (накопленной) погрешностью называется погрешность численного решения после выполнения *нескольких* шагов. Пусть мы имеем некоторый одношаговый метод, с помощью которого при заданных начальных данных ( $x_0, y_0$ ) и длине шага  $h$  мы определяем численное решение  $y_1$ , аппроксимирующее  $y(x_0 + h)$ . Воспользуемся обозначениями Хенричи для этого процесса:

$$y_1 = y_0 + h\Phi(x_0, y_0, h), \quad (3.21)$$

и назовем  $\Phi$  функцией приращения для данного метода.

<sup>1)</sup> «Но еще лучше оказывается ...». — *Nem.*

Тогда численное решение в точке  $X > x_0$  получается с помощью пошаговой процедуры

$$y_{i+1} = y_i + h\Phi(x_i, y_i, h_i), \quad h_i = x_{i+1} - x_i, \quad x_N = X, \quad (3.22)$$

и наша задача состоит в оценке *глобальной погрешности*

$$E = y(X) - y_N. \quad (3.23)$$

Эта оценка находится простым способом, очень похожим на данное Коши доказательство теоремы сходимости (теорема 7.3 из гл. I): локальные погрешности переносятся в конечную точку  $x_N$ .

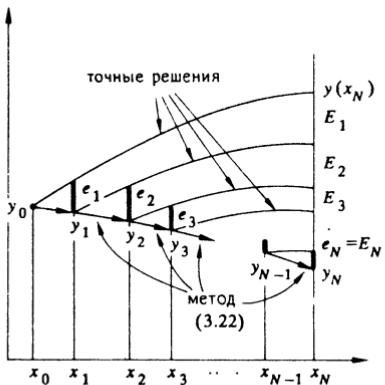


Рис. 3.1. Оценивание глобальной погрешности, метод а.

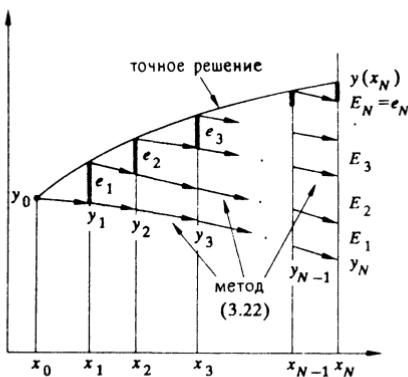


Рис. 3.2. Оценивание глобальной погрешности, метод б).

и затем складываются. Этот «перенос погрешностей» можно выполнить двумя разными способами:

а) перенося погрешности вдоль кривых точных решений (см. рис. 3.1); этот способ может дать хорошие результаты, если известны хорошие оценки распространения погрешности для точных решений, например из теоремы 10.6 из гл. I, использующей понятие логарифмической нормы  $\mu(\partial f / \partial y)$ ;

б) перенося погрешность  $i$ -го шага посредством выполнения  $N - i$  шагов численного метода (см. рис. 3.2); этот способ использовали в своих доказательствах Коши (1824) и Рунге (1905), он легко обобщается на многошаговые методы (см. гл. III) и будет для нас важным инструментом при доказательстве существования асимптотических разложений (см. разд. II.8).

В обоих случаях оценим сначала с помощью теоремы 3.1 *локальные погрешности*:

$$\|e_i\| \leq C \cdot h_i^{p+1}. \quad (3.24)$$

**Предупреждение:** величины  $e_i$  на рис. 3.1 и 3.2 при  $i \neq 1$  не совпадают, но они допускают сходные оценки.

Займемся теперь оценкой перенесенных погрешностей  $E_i$ .

а) При первом способе переноса мы воспользуемся известными результатами из гл. I, в частности теоремами I.10.6 и I.10.2 или формулой (I.7.17).

**Теорема 3.4.** Обозначим  $U$  окрестность точки  $\{(x, y(x)) \mid x_0 \leqslant x \leqslant X\}$ , где  $y(x)$  — точное решение уравнения (2.2). Пусть в  $U$  справедливы оценки локальных погрешностей (3.24) и выполнено одно из условий:

$$\left\| \frac{df}{dy} \right\| \leqslant L \text{ или } \mu \left( \frac{df}{dy} \right) \leqslant L. \quad (3.25)$$

Тогда имеет место следующая оценка глобальной погрешности (3.23):

$$\|E\| \leqslant h^p \frac{C'}{L} (\exp(L(x_N - x_0)) - 1), \quad (3.26)$$

где  $h = \max h_i$ ,

$$C' = \begin{cases} C & \text{при } L \geqslant 0, \\ C \exp(-Lh) & \text{при } L < 0, \end{cases}$$

и  $h$  достаточно мало для того, чтобы численное решение оставалось в  $U$ .

**Замечание.** При  $L \rightarrow 0$  оценка (3.26) переходит в  $h^p C (x_N - x_0)$ .

**Доказательство.** Из теоремы I.10.2 (при  $\varepsilon = 0$ ) или теоремы I.10.6 (при  $\delta = 0$ ) получаем

$$\|E_i\| \leqslant \exp(L(x_N - x_i)) \|e_i\|. \quad (3.27)$$

Подставляя в неравенство

$$\|E\| \leqslant \sum_{i=1}^N \|E_i\|$$

выражение (3.27) с учетом (3.24) и принимая во внимание, что  $h_i^{p+1} \leqslant h^p \cdot h_i$ , приходим к такому неравенству:

$$\|E\| \leqslant h^p C [h_0 \exp(L(x_N - x_1)) + h_1 \exp(L(x_N - x_2)) + \dots].$$

Выражение в квадратных скобках мажорируется следующими интегралами:

$$\int_{x_0}^{x_N} \exp(L(x_N - x)) dx \quad \text{при } L \geqslant 0, \quad (3.28)$$

$$\int_{x_0}^{x_N} \exp(L(x_N - h - x)) dx \quad \text{при } L < 0 \quad (3.29)$$

(см. рис. 3.3). Отсюда и получается оценка (3.26). ■

б) При втором способе переноса погрешностей рассмотрим кроме (3.22) еще одно численное решение, значения которого в соседних узлах связаны равенством

$$z_{i+1} = z_i + h_i \Phi(x_i, z_i, h_i).$$

Нам нужно оценить норму разности  $\|z_{i+1} - y_{i+1}\|$  через  $\|z_i - y_i\|$ .

Для  $z_{i+1}$  формулы метода РК запишем в таких обозначениях:

$$l_1 = f(x_i, z_i),$$

$$l_2 = f(x_i + c_2 h_i, z_i + h_i a_{21} l_1) \quad \text{и т. д.}$$

Вычитая из этих формул соответствующие формулы (1.8), получим для норм разностей такие оценки:

$$\|k_1 - l_1\| \leq L \|y_i - z_i\|,$$

$$\|k_2 - l_2\| \leq L (1 + |a_{21}| h L) \|y_i - z_i\| \quad \text{и т. д.}$$

Это приводит нас к следующему предложению.

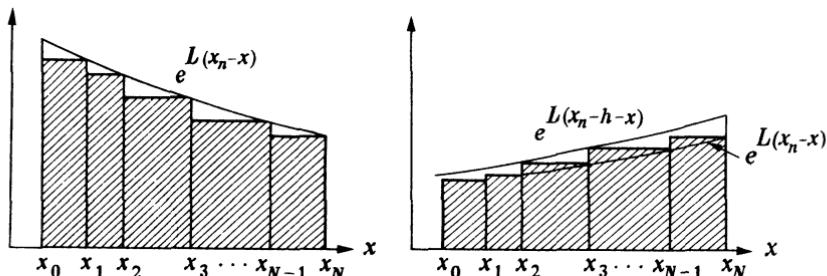


Рис. 3.3. Оценивание римановых сумм.

**Лемма 3.5.** Пусть  $L$  — постоянная Липшица для функции  $f$  и пусть  $h_i \leq h$ . Тогда функция приращения  $\Phi$  для метода (1.8) удовлетворяет неравенству

$$\|\Phi(x_i, z_i, h_i) - \Phi(x_i, y_i, h_i)\| \leq \Lambda \|z_i - y_i\|, \quad (3.30)$$

где

$$\Lambda = L \left( \sum_i |b_i| + hL \sum_{i,j} |b_i a_{ij}| + h^2 L^2 \sum_{i,j,k} |b_i a_{ij} a_{jk}| + \dots \right). \quad (3.31)$$

■ Из (3.30) мы получаем искомую оценку:

$$\|z_{i+1} - y_{i+1}\| \leq (1 + h_i \Lambda) \|z_i - y_i\| \leq \exp(h_i \Lambda) \|z_i - y_i\|, \quad (3.32)$$

и с ее помощью оценку перенесенных погрешностей на рис. 3.2

$$\|E_i\| \leq \exp(\Lambda(x_N - x_i)) \|e_i\| \quad (3.33)$$

вместо оценки (3.27), соответствующей рис. 3.1.

Повторяя доказательство теоремы 3.4, придем к следующему результату.

**Теорема 3.6.** Предположим, что для начальных значений, лежащих на точном решении, локальная погрешность удовлетворяет оценке

$$\|y(x+h) - y(x) - h\Phi(x, y(x), h)\| \leq Ch^{p+1} \quad (3.34)$$

и что в окрестности решения функция приращения  $\Phi$  удовлетворяет неравенству

$$\|\Phi(x, z, h) - \Phi(x, y, h)\| \leq \Lambda \|z - y\|. \quad (3.35)$$

Тогда для глобальной погрешности (3.23) справедлива следующая оценка:

$$\|E\| \leq h^p \frac{C}{\Lambda} (\exp(\Lambda(x_N - x_0)) - 1), \quad (3.36)$$

где  $h = \max h_i$ . ■

### Упражнения

1. [Рунге (1905)] Покажите, что для явных методов Рунге—Кутты порядка  $p = s$ , у которых  $b_i \geq 0$ ,  $a_{ij} \geq 0$  (для всех  $i, j$ ), постоянная  $\Lambda$  для  $\Phi$  удовлетворяет неравенству

$$1 + h\Lambda < \exp(hL)$$

и что оценка (3.36) останется справедливой, если  $\Lambda$  заменить на  $L$ .

2. Покажите, что формула (3.20) для  $e(t_{55})$  после подстановки выражений для  $b_3$  и  $b_4$ , приведенных ниже формулы (1.17), примет вид

$$e(t_{55}) = 1 - 5 \frac{(4v^2 - 15v + 9) - u(6v^2 - 42v + 27) - u^2(26v - 18)}{12(1 - 2u)(6uv - 4(u + v) + 3)}.$$

3. Определите значения  $u$  и  $v$  в (1.17) так, чтобы в формулах (3.20) было:

a)  $\max_{t=5, 6, 7, 8} |e(t_{5t})| = \min$ ,      b)  $\sum_{t=1}^9 |e(t_{5t})| = \min$ ,

c)  $\max_{t=5, 6, 7, 8} |\alpha(t)| |e(t_{5t})| = \min$ , d)  $\sum_{t=1}^9 |\alpha(t)| |e(t_{5t})| = \min$ .

Ответ:

a)  $u = 0.3587$ ,  $v = 0.6346$ ,  $\min = 0.1033$ ;

b)  $u = 0.3995$ ,  $v = 0.6$ ,  $\min = 1.55$ ;

c)  $u = 0.3501$ ,  $v = 0.5839$ ,  $\min = 0.1248$ ;

d)  $u = 0.3716$ ,  $v = 0.6$ ,  $\min = 2.53$ .

4. Примените какой-либо явный метод Рунге—Кутты к задаче  $y' = f(x, y)$ ,  $y(0) = 0$ , где

$$f(x, y) = \begin{cases} \frac{\lambda}{x} y + g(x) & \text{при } x > 0, \\ (1 - \lambda)^{-1} g(0) & \text{при } x = 0, \end{cases}$$

причем  $\lambda \leq 0$  и  $g(x)$  — функция, дифференцируемая достаточноное число раз (см. упр. 10 в разд. I.5).

а) Покажите, что погрешность после первого шага имеет вид

$$y(h) - y_1 = C_2 h^2 g'(0) + O(h^3),$$

где  $C_2$  — постоянная, зависящая от коэффициентов метода.

Вообще говоря,  $C_2 \neq 0$  и для методов высших порядков.

б) Вычислите  $C_2$  для классического метода 4-го порядка (см. табл. 1.2).

## II.4. Практическая оценка погрешности и выбор длины шага

«Ich glaube indessen, dass ein praktischer Rechner sich meistens mit der geringeren Sicherheit begnügen wird, die er aus der Uebereinstimmung seiner Resultate für grössere und kleinere Schritte gewinnt».

«Между тем я верю, что вычислитель-практик в большинстве случаев будет довольствоваться меньшими гарантиями, которые он извлекает из совпадения своих результатов при больших и меньших шагах». (К. Рунге, 1895)

Даже упрощенные оценки погрешностей из разд. II.3, которые довольствуются лишь главным членом погрешности, не представляют практического интереса, потому что они требуют вычисления и оценивания верхних границ ряда частных производных высоких порядков. Между тем главное преимущество методов Рунге—Кутты по сравнению с методом ряда Тейлора как раз в том и состоит, что они не требуют вычисления этих производных. Для практики, однако, оценки погрешностей *необходимы*, — с одной стороны, чтобы обеспечить длину шага  $h_i$ , достаточно *малую* для достижения требуемой точности вычисляемых результатов, а с другой стороны — чтобы гарантировать достаточно *большую* длину шага во избежание бесполезной вычислительной работы. Поэтому мы обсудим теперь альтернативные методы оценивания погрешностей.

Самый старый способ, который использовал Рунге в своих численных примерах, состоит в повторении вычислений с *уменьшенной вдвое* длиной шага и сравнении результатов: те десятичные знаки, которые не изменились, считаются правильными («...woraus ich schliessen zu dürfen glaube...»<sup>1)</sup>).

### Экстраполяция по Ричардсону

«...ее полезность при практических вычислениях вряд ли может быть переоценена».  
(Г. Биркгофф—Г. К. Рота)

Идея Ричардсона, которая была опубликована в его классической работе [Ричардсон (1910)], посвященной главным образом дифференциальным уравнениям в частных производных, и разъяснена во всех подробностях в более поздней работе [Ричард-

<sup>1)</sup> «... я надеюсь, что отсюда можно заключить ...». — *Nem.*

сон (1927)], состоит в более осторожном использовании известного поведения погрешности как функции  $h$ .

Предположим, что, пользуясь некоторым определенным методом Рунге—Кутты порядка  $p$  и исходя из заданной начальной точки  $(x_0, y_0)$ , мы провели расчет для двух шагов при длине шага  $h$  и получили при этом численные результаты  $y_1$  и  $y_2$ . Исходя из той же начальной точки  $(x_0, y_0)$ , сделаем теперь расчет для одного большого шага длины  $2h$  и обозначим полученное численное решение  $w$ . Как мы знаем из теоремы 3.2, погрешность  $y_1$  представима в виде

$$e_1 = y(x_0 + h) - y_1 = C \cdot h^{p+1} + O(h^{p+2}), \quad (4.1)$$

где  $C$  выражается через коэффициенты погрешности метода и элементарные дифференциалы  $F^j(t)(y_0)$  порядка  $p + 1$ . Погрешность  $y_2$  состоит из двух частей: из перенесенной погрешности первого шага, имеющей вид

$$\left( I + h \frac{\partial f}{\partial y} + Q(h^2) \right) e_1,$$

и локальной погрешности второго шага, тоже выражаемой формулой (4.1), но с элементарными дифференциалами, вычисленными при  $y_1 = y_0 + O(h)$ . Таким образом, получаем, что

$$e_2 = y(x_0 + 2h) - y_2 = (I + O(h)) Ch^{p+1} + (C + O(h)) h^{p+1} + O(h^{p+2}), \quad (4.2)$$

откуда

$$e_2 = 2Ch^{p+1} + O(h^{p+2}).$$

По аналогии с (4.1) для большого шага имеем

$$y(x_0 + 2h) - w = C(2h)^{p+1} + O(h^{p+2}). \quad (4.3)$$

Если пренебречь членами  $O(h^{p+2})$ , формулы (4.2) и (4.3) позволяют исключить неизвестную постоянную  $C$  и «экстраполировать» значение  $\hat{y}_2$ , лучше аппроксимирующее  $y(x_0 + 2h)$ . Итак, мы получили следующий результат.

**Теорема 4.1.** Пусть некоторым методом РК порядка  $p$  в результате выполнения двух шагов длины  $h$  найдено численное значение  $y_2$ , а в результате выполнения одного большого шага длины  $2h$  получено значение  $w$ . Тогда погрешность  $y_2$  может быть оценена по формуле

$$y(x_0 + 2h) - y_2 = \frac{y_1 - w}{2^p - 1} + O(h^{p+2}), \quad (4.4)$$

а выражение

$$\hat{y}_2 = y_2 + \frac{y_1 - w}{2^p - 1} \quad (4.5)$$

аппроксимирует величину  $y(x_0 + 2h)$  с порядком  $p + 1$ . ■

Формула (4.4) дает очень простой способ оценки погрешности, а (4.5) позволяет дополнительно увеличить точность на один порядок<sup>1)</sup>. «...Развиваемая в следующих разделах улучшенная теория сложна и наводит на мысль, что практика может быть столь же сложной, тогда как в действительности она проста». (Ричардсон.)

### Автоматическое управление длиной шага

«D'ordinaire, on se contente de multiplier ou de diviser par 2 la valeur du pas...»

«Обычно довольствуются умножением или делением длины шага на 2...»

(Ческино, 1961)

Мы хотим теперь написать программу, которая автоматически подбирала бы длину шага так, чтобы локальная погрешность не превышала предписанной допустимой величины (допуска).

Если начальная длина шага была выбрана равной  $h$ , программа выполняет вычисление двух шагов длины  $h$  и одного шага длины  $2h$ , как описано выше. Затем на основе формулы (4.4) вычисляется погрешность:

$$err = \frac{1}{2^p - 1} \max_{i=1, \dots, n} \frac{|y_{2,i} - w_i|}{d_i}. \quad (4.6)$$

Часто используют и другие нормы, например евклидову. В (4.6)  $d_i$  — масштабирующий множитель. Для вычисления абсолютной погрешности полагают  $d_i = 1$ , а для относительной  $d_i = |\hat{y}_{2,i}|$ . Часто используют и смешанное масштабирование типа

$$d_i = \max(|\hat{y}_{2,i}|, |y_{0,i}|, 1), \text{ или } d_i = \max(|\hat{y}_{2,i}|, |y_{0,i}|, 10^{-6}),$$

или что-нибудь еще в этом роде. Затем величина  $err$  сравнивается с заданной величиной допустимой погрешности  $tol$ , что позволяет, зная зависимость погрешности от длины шага  $Ch^{p+1}$ , вычислить оптимальную длину шага («...le procédé connu», Ceschino (1961))<sup>2)</sup>

$$h \cdot (tol/err)^{1/(p+1)}. \quad (4.7)$$

<sup>1)</sup> В отечественной литературе по численному анализу эти формулы рассматриваются как частный случай правила Рунге [см., напр., Бахвалов (1975)]. — Прим. ред.

<sup>2)</sup> «Известный прием» Ческино (1961). — Франц.

Далее, для получения хорошей программы требуется известная осторожность. Умножим (4.7) на гарантийный фактор  $fac$ . Обычно выбирают  $fac$  равным 0.8, 0.9,  $(0.25)^{1/(p+1)}$  или  $(0.38)^{1/(p+1)}$ , чтобы в ближайшее время погрешность была приемлемой с высокой вероятностью. Кроме того, нельзя допускать, чтобы длина шага  $h$  возрастала или убывала слишком быстро. Для этого можно выбирать новую длину шага, например следующим образом:

$$h_{new} = h \cdot \min(facmax, \max(facmin, fac \cdot (tol/err)^{1/(p+1)})). \quad (4.7')$$

Тогда, если  $err \leq tol$ , два вычисленных шага считаются *принятыми*, и решение продолжается исходя из  $y_2$  или  $\hat{y}_2$  (в последнем случае говорят о методе «локальной экстраполяции»), причем в качестве длины нового шага берется  $h_{new}$ . В противном случае оба шага *отбрасываются*, и вычисления *повторяются* с новой длиной шага  $h_{new}$ . Максимальный коэффициент увеличения шага  $facmax$  обычно выбирают между 1.5 и 5. Он предотвращает слишком сильное увеличение длины шага программой и способствует ее надежности. Понятно, что, если его выбрать слишком малым, это может привести и к излишнему увеличению вычислительной работы. Наконец, для шагов, выполняемых непосредственно после отбрасывания забракованных шагов, рекомендуется положить  $facmax = 1$  [Шампайн и Уоттс (1979)].

### Вложенные формулы Рунге—Кутты

«Мне потребовалось много времени, чтобы понять это, но объяснение с трудом поддается описанию».

(Мерсон, 1957)

«Скрайтон правильно критикует процесс Мерсона: некоторые ожидали от него очень многое, хотя сам Мерсон не претендовал на это».

(Р. Ингланд, 1969)

Идея их состоит в том, чтобы вместо пользования экстраполяцией Ричардсона построить такие формулы РК, которые сами содержали бы кроме численного приближенного значения  $y_1$  некоторое выражение  $\hat{y}_1$  более высокого порядка (или каким-либо иным образом более точное, чем  $y_1$ ). Последнее могло бы тогда служить для управления погрешностью и длиной шага на каждом шаге. В частности, это удешевило бы выбраковку шагов.

Итак, нам надо найти такую таблицу коэффициентов (см. (1.8'))

$0$					
$c_2$	$a_{21}$				
$c_3$	$b_{31}$	$a_{32}$			
$\vdots$	$\vdots$	$\vdots$	$\vdots$		
$c_s$	$a_{s1}$	$a_{s2}$	$\dots$	$a_{s, s-1}$	
	$b_1$	$b_2$	$\dots$	$b_{s-1}$	$b_s$
	$\widehat{b}_1$	$\widehat{b}_2$	$\dots$	$\widehat{b}_{s-1}$	$\widehat{b}_s$

(4.8)

чтобы величина

$$y_1 = y_0 + h(b_1 k_1 + \dots + b_s k_s) \quad (4.9)$$

имела порядок  $p$ , а

$$\widehat{y}_1 = y_0 + h(\widehat{b}_1 k_1 + \dots + \widehat{b}_s k_s) \quad (4.9')$$

— порядок  $q$  (обычно  $q = p - 1$  или  $q = p + 1$ ).

Согласно теореме 2.13, мы должны удовлетворить условиям

$$\sum_t b_j \Phi_j(t) = \frac{1}{\gamma(t)} \quad \text{для всех деревьев порядка } \leq p \quad (4.10)$$

и

$$\sum_t \widehat{b}_j \Phi_j(t) = \frac{1}{\gamma(t)} \quad \text{для всех деревьев порядка } \leq q. \quad (4.10')$$

Первые методы такого типа предложили Мерсон (1957), Ческино (1962) и Зонневельд (1963) (см. табл. 4.1, 4.2 и 4.3, в которых заголовок вида «фамилия  $p$  ( $q$ )» указывает, что порядок  $y_1$  равен  $p$ , а порядок «оценщика погрешности»  $\widehat{y}_1$  равен  $q$ ).

Однако эти методы еще не полностью удовлетворяют сформулированным нами выше требованиям. У Мерсона  $\widehat{y}_1$  имеет порядок 5 только для линейных уравнений с постоянными коэффициентами. Поэтому этот метод переоценивает погрешность при малых  $h$ , но все же работает вполне хорошо и применяется весьма широко, особенно пользователями библиотеки NAG («команда, приносящая результаты»). Точно так же у Зонневельда  $\widehat{y}_1$  оценивает не погрешность аппроксимации, а «последний учтенный член». Метод

Таблица 4.1. Ческино 2 (4)

$0$					
$\frac{1}{4}$	$\frac{1}{4}$				
$\frac{1}{2}$	$0$	$\frac{1}{2}$			
$1$	$1$	$-2$	$2$		
$y_1$	$1$	$-2$	$2$	$0$	
$\widehat{y}_1$	$\frac{1}{6}$	$0$	$\frac{4}{6}$	$\frac{1}{6}$	

Таблица 4.2. Мерсон 4(«5»)

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
<hr/>				
$y_1$	$\frac{1}{2}$	0	$-\frac{3}{2}$	2
<hr/>				
$\hat{y}_1$	$\frac{1}{6}$	0	0	$-\frac{1}{6}$

Таблица 4.3. Зонневельд 4(3)

0					
$\frac{1}{2}$	$\frac{1}{2}$				
$\frac{1}{2}$	0	$\frac{1}{2}$			
1	0	0	1		
$\frac{3}{4}$	$\frac{5}{32}$	$\frac{7}{32}$	$\frac{13}{32}$	$-\frac{1}{32}$	
<hr/>					
$y_1$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{6}$	0
<hr/>					
$\hat{y}_1$	$-\frac{1}{2}$	$\frac{7}{3}$	$\frac{7}{3}$	$\frac{13}{6}$	$-\frac{16}{3}$

Ческино не экономичен, потому что он оценивает погрешность со слишком высокой точностью (см. ниже упр. 1—3).

Формулы, которые удовлетворяют условиям (4.10) и (4.10'), вывели Сарафян (1966), Ингланд (1969) и Фельберг (1968, 1969). Продемонстрируем вывод формулы Фельберга второго порядка:

0			
$c_2$	$a_{21}$		
$c_3$	$a_{31}$	$a_{32}$	
<hr/>			
	$b_1$	$b_2$	$b_3$
<hr/>			
	$\hat{b}_1$	$\hat{b}_2$	$\hat{b}_3$

Согласно (4.10) и (4.10'), должны удовлетворяться уравнения

$$b_1 + b_2 + b_3 = 1, \quad \hat{b}_1 + \hat{b}_2 + \hat{b}_3 = 1,$$

$$b_2 c_2 + b_3 c_3 = \frac{1}{2}, \quad \hat{b}_2 c_2 + \hat{b}_3 c_3 = \frac{1}{2},$$

$$\hat{b}_2 c_2^2 + \hat{b}_3 c_3^2 = \frac{1}{3}, \quad (4.12)$$

$$\hat{b}_3 c_3 c_2 = \frac{1}{6}.$$

Выбрав  $c_2 = 1$  и  $b_3 = 0$ , из первых двух уравнений получим  $b_2 = b_1 = \frac{1}{2}$ . Осталось 4 уравнения с 5 неизвестными. Если положить  $c_3 = \frac{1}{2}$ , то  $b_1 = \frac{1}{6}$ ,  $\hat{b}_2 = \frac{1}{6}$ ,  $\hat{b}_3 = \frac{4}{6}$  оказываются коэффициентами правила Симпсона и, наконец,  $a_{32} = \frac{1}{4}$ . Получившийся метод приведен в табл. 4.4. Второй метод из табл. 4.4 тоже предложил Фельберг (1969). Он содержит одну дополнительную строку:

тельную стадию, но коэффициенты выбраны так, что  $a_{4i} = b_i$  для всех  $i$ , поэтому вычисленное последним значение  $f$  на текущем шаге может быть снова использовано, так как оно совпадает с первым значением  $f$ , которое надо вычислить на следующем шаге. Поскольку коэффициенты погрешности для  $y_1$  у метода RKF2 (3)В в сто с лишним раз меньше, последний претендует на лучшее качество, чем метод RKF2 (3) (ниже мы увидим, что это, конечно, не совсем верно).

Таблица 4.4. Вложенные методы Рунге—Кутты

RKF 2(3)			RKF 2(3)В		
0			0		
			$\frac{1}{4}$	$\frac{1}{4}$	
1	1		$\frac{27}{40}$	$\frac{189}{800}$	$\frac{729}{800}$
$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	1	$\frac{214}{891}$	$\frac{1}{33}$
$y_1$	$\frac{1}{2}$	$\frac{1}{2}$	$y_1$	$\frac{214}{891}$	$\frac{650}{891}$
$y_1$	$\frac{1}{6}$	$\frac{1}{6}$	$y_1$	$\frac{533}{2106}$	$0$
	$\frac{4}{6}$	$\frac{4}{6}$		$\frac{800}{1053}$	$-\frac{1}{78}$

Фельберг вывел много других формул различных порядков. В табл. 4.5 приведен метод 4-го порядка с 6 стадиями — один из наиболее используемых в настоящее время.

Таблица 4.5. Фельберг 4(5)

0							
$\frac{1}{4}$		$\frac{1}{4}$					
$\frac{3}{8}$		$\frac{3}{32}$	$\frac{9}{32}$				
$\frac{12}{13}$		$\frac{1932}{2197}$	$\frac{7296}{2197}$				
1	$\frac{439}{216}$	-8	$\frac{3680}{513}$	$-\frac{845}{4104}$			
$\frac{1}{2}$	$-\frac{8}{27}$	2	$-\frac{3544}{2565}$	$\frac{1859}{4104}$	$-\frac{11}{40}$		
$y_1$	$\frac{25}{216}$	0	$\frac{1408}{2565}$	$\frac{2197}{4104}$	$-\frac{1}{5}$	0	
$y_1$	$\frac{16}{135}$	0	$\frac{6656}{12825}$	$\frac{28561}{56430}$	$-\frac{9}{50}$	$\frac{2}{55}$	

### Формула Дормана и Принса

Возникает естественный вопрос: а почему бы не использовать лучшее из полученных значений, т. е. выражение старшего порядка  $\hat{y}_1$ , в качестве численного результата и начального значения для следующего шага? Естественно ответить на этот вопрос так: но тогда наша «оценка погрешности»  $\hat{y}_1 - y_1$  уже не будет оценивать погрешность, а ведь мы хотим оценивать погрешности, не так ли? Однако на такой ответ имеются два естественных возражения:

а) поскольку обычно свойства устойчивости или неустойчивости решаемой дифференциальной системы неизвестны, локальные погрешности обычно имеют мало общего с глобальными погрешностями;

б) Фельберг, чтобы сделать свои методы оптимальными, пытался минимизировать коэффициенты погрешности для результата *низшего* порядка  $y_1$ . Как следствие этого возникает опасность, что разность  $y_1 - \hat{y}_1$ , состоящая главным образом из минимизированных коэффициентов погрешности, недооценивает локальную погрешность.

Поэтому интересно вывести методы, у которых члены погрешности для результата старшего порядка были бы минимизированы, а результат младшего порядка вычислялся бы только для использования в механизме управления длиной шага. Такой

Таблица 4.6. Дорман—Принс 5(4) (DOPRI5)

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	$\frac{500}{1113}$	$\frac{125}{192}$	$-\frac{2187}{6784}$	$\frac{11}{84}$	
$y_1$	$\frac{35}{384}$	0	$\frac{500}{1113}$	$\frac{125}{192}$	$-\frac{2187}{6784}$	$\frac{11}{84}$	0
$\hat{y}_1$	$\frac{5179}{57600}$	0	$\frac{7571}{16695}$	$\frac{393}{640}$	$-\frac{92097}{339200}$	$\frac{187}{2100}$	$\frac{1}{40}$

метод, в котором к тому же  $a_{st} = b_t$ , так что результат последнего вычисления производной может снова использоваться на следующем шаге, построили Дорман и Принс (1980). Он дает великолепные результаты. Коэффициенты этого метода приведены в табл. 4.6.

### Численное исследование механизма управления длиной шага

На рис. 4.1 и 4.2 представлены результаты действия описанного выше механизма управления длиной шага (4.7'). В качестве примера мы выбрали модель брюсселятера (см. разд. I.16)

$$\begin{aligned} \dot{y}_1 &= 1 + y_1^2 y_2 - 4y_1, \\ \dot{y}_2 &= 3y_1 - y_1^2 y_2 \end{aligned} \quad (4.15)$$

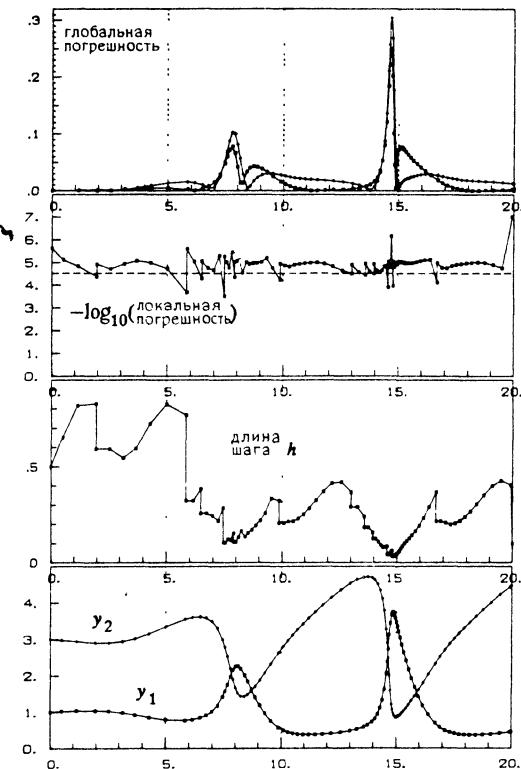


Рис. 4.1. Метод RKF2(3) с управлением длиной шага,  $tol = 5 \times 10^{-4}$ , 178 шагов + 5 отброшенных.

с начальными значениями  $y_1(0) = 1.01$ ,  $y_2(0) = 3$ , очень близкими к особой точке; интервал интегрирования взят  $0 \leq x \leq 20$ . На этих рисунках представлены следующие графики:

1) внизу — решение  $y_1(x)$  и  $y_2(x)$  со всеми принятами шагами интегрирования;

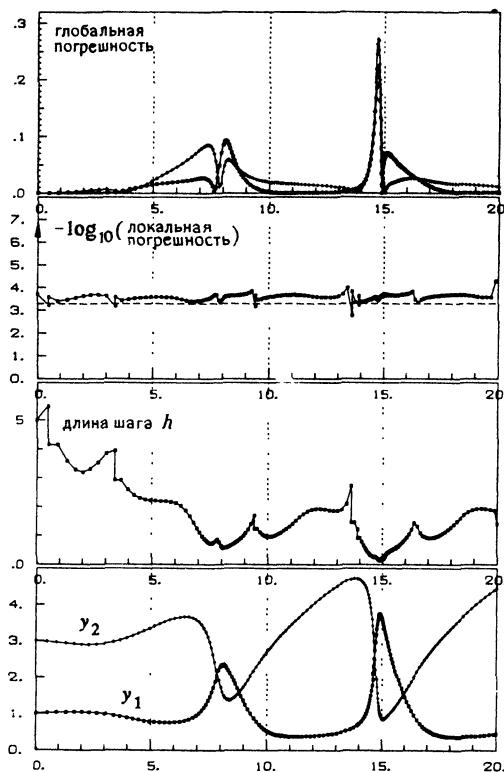


Рис. 4.2. Метод RKF2(3)В с управлением длиной шага,  $tol = 3 \times 10^{-5}$ , 101 шаг + 12 отброшенных.

2) выше — все длины шагов, как принятых, так и отброшенных;

3) третий снизу график показывает оцененную погрешность  $-\lg(err)$ ; желаемая точность (допуск погрешности  $tol$ ) показана прерывистой горизонтальной линией;

4) вверху рисунков изображены *глобальные* погрешности обеих компонент численного решения. Видно, что из-за неустойчивости решений относительно возмущений начальных значений появ-

ляются довольно большие глобальные погрешности, хотя допускаемые локальные погрешности всюду малы.

Мы сравнивали здесь две формулы Фельберга RKF2(3) и RKF2(3)В из табл. 4.4. Можно заметить, что вторая формула требует много меньшего допуска на локальные погрешности, чтобы получаемые в действительности глобальные погрешности были сравнимыми. Этот метод с его малыми коэффициентами погрешности для  $y_1$  недооценивает действительные погрешности.

Можно заметить также, что гарантийный фактор в формуле (4.7') (здесь выбран  $fac = 0.8$ ) обычно хорошо поддерживает локальную погрешность ниже желаемого уровня.

### Численное сравнение методов 4-го порядка

Интересно посмотреть, как ведут себя различные методы 4-го порядка с управлением длиной шага. Следующий пример мы выбрали из отчета Фельберга (1968):

$$\begin{aligned} y'_1 &= 2xy_1 \ln(\max(y_2, 10^{-3})), \quad y_1(0) = 1, \\ y'_2 &= -2xy_2 \ln(\max(y_1, 10^{-3})), \quad y_2(0) = e, \end{aligned} \quad (4.16)$$

правильным решением которого является  $y_1 = \exp(\sin(x^2))$ ,  $y_2 = \exp(\cos(x^2))$ . В серии расчетов мы изменяли допустимую погрешность от  $10^{-3}$  до  $10^{-7}$ , причем каждое следующее значение было в  $\sqrt{10}$  раз меньше предыдущего. Интегрирование производилось на интервале  $0 \leq x \leq 5$ . Все полученные результаты представлены на рис. 4.3, где по оси абсцисс отложены (со знаком минус) десятичные логарифмы максимальной глобальной погрешности («точность»), а по оси ординат — десятичные логарифмы количества обращений к вычислению функций («работа»). Использовались те же методы, что и представленные на рис. 1.1, и им соответствуют такие же линии. К ним добавлены здесь метод Фельберга (табл. 4.5), метод Дормана и Принса (табл. 4.6) и метод Бутчера шестого порядка (табл. 6.1). Метод Фельберга RKF4(5) оказался лучше, чем классический без локальной экстраполяции. Однако имеющееся в работе Фельберга (1968) утверждение, что его метод якобы «...экономит от 40 до 60% машинного времени...», вряд ли правомерно. Как видим, и этот метод недооценивает локальную погрешность. Кроме того, применение локальной экстраполяции в значительной степени лишает метод RKF4(5) его преимущества.

Лучшим из всех этих методов, несомненно, является метод, использующий набор коэффициентов, предложенный Дорманом и Принсом.

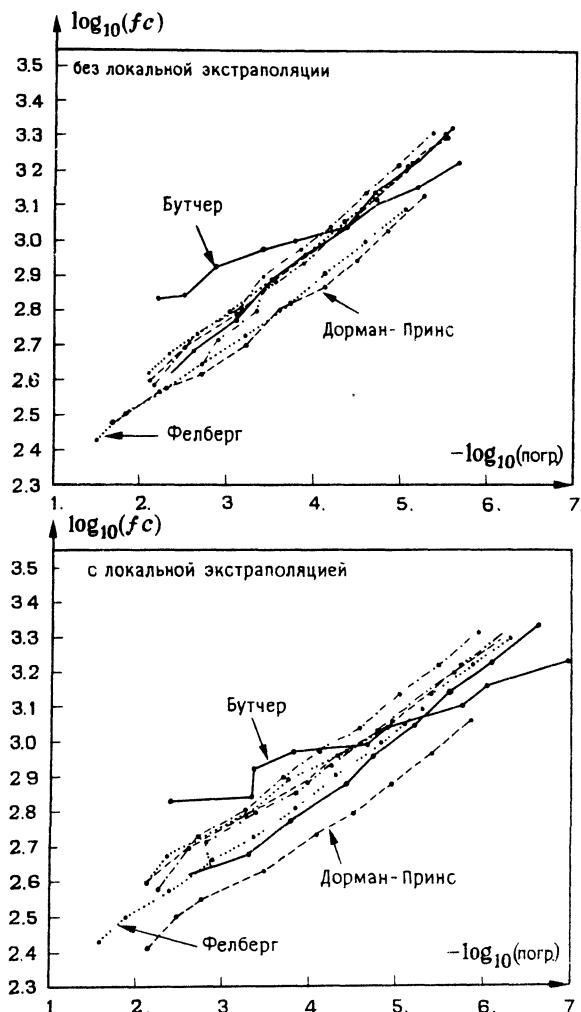


Рис. 4.3. Диаграмма «точность — объем вычислений» (последний оценивается числом обращений к подпрограмме вычисления функции).

## Упражнения

- Покажите, что метод Рунге (1.4) можно интерпретировать как два шага методом Эйлера (длины  $h/2$ ), сопровождаемые экстраполяцией по Ричардсону.
- Постройте вложенные формулы RK1(2) с двумя и тремя стадиями, удовлетворяющие условию  $a_{si} = b_i$ . (Фельберг: «Der Vollständigkeit halber sei noch erwähnt...»<sup>1)</sup>.) Примеры приведены в табл. 4.7 (второй из методов определяется, конечно, неоднозначно).

Таблица 4.7. Вложенные методы низкого порядка

RK1(2)			RK1(2)B		
0			0		
			$\frac{1}{2}$	$\frac{1}{2}$	
1	1		1	$\frac{1}{256}$	$\frac{255}{256}$
$y_1$	1	0	$y_1$	$\frac{1}{256}$	$\frac{255}{256}$
$\theta_1$	$\frac{1}{2}$	$\frac{1}{2}$	$\theta_1$	$\frac{1}{512}$	$\frac{255}{256}$
					$\frac{1}{512}$

<sup>1)</sup> «Ради полноты следует еще упомянуть...» — *Nem.*

## II.5. Дальнейшие вопросы практических вычислений

### Плотная выдача

Механизм управления длиной шага, описанный в разд. II.4, выбирает точки, в которых вычисляется решение, в соответствии с требованиями к допустимой локальной погрешности. Часто, однако, требуется вычислить и выдать решение в *заданных* точках, которые к тому же могут быть расположены достаточно плотно. Одна из возможностей получения решения в таких заданных точках состоит в уменьшении длины шага всякий раз, когда это нужно. Такое повторное уменьшение длины шага от почти оптимальной величины нарушает управление длиной шага и может привести к очень сильному росту времени счета и погрешностей округления (Шампайн, Уоттс и Давенпорт (1976)). Поэтому представляет интерес вывод таких методов РК, которые были бы способны выдавать значения решения во всех промежуточных точках  $x^* = x_0 + \theta h$ , где  $0 < \theta \leq 1$  (так называемые *непрерывные* методы РК). Разумеется, такие непрерывные методы представляют особенный интерес, если их коэффициенты  $a_{ij}$  не зависят от  $\theta$ : это позволяет на каждом шаге интегрирования вычислить решение в сотнях точек без особых дополнительных затрат.

Пример 5.1. В случае трехстадийного непрерывного метода РК третьего порядка условия порядка примут следующий вид:

$$\begin{aligned} b_1 + b_2 + b_3 &= \theta, \\ b_2 c_2 + b_3 c_3 &= \frac{\theta^2}{2}, \\ b_2 c_2^2 + b_3 c_3^2 &= \frac{\theta^3}{3}, \\ b_3 a_{32} c_2 &= \frac{\theta^3}{6}. \end{aligned} \tag{5.1}$$

Это следует из теоремы 2.1, если  $y(x_0 + h)$  заменить на  $y(x_0 + \theta h)$ . Простые вычисления показывают, что эти уравнения неразрешимы, если требовать независимости  $c_2$ ,  $c_3$  и  $a_{32}$  от  $\theta$ . Поэтому ослабим наше требование. Так как непрерывная формула используется только один раз для каждой выдаваемой величины, мы будем требовать, чтобы ее точность соответствовала *глобальной погрешности* формулы третьего порядка, а она является величиной  $O(h^3)$ . Итак, мы требуем *второго порядка* при всех  $\theta$  и *третьего* — при  $\theta = 1$ . Выбрав  $c_2 = 1/2$ ,  $c_3 = 1$  и решая пер-

вые три уравнения при произвольном  $\theta$ , а последнее — при  $\theta = 1$ , мы получим схему (табл. 5.1), которая является расширением одного из методов Кутты (1901).

Таблица 5.1. Непрерывный метод Рунге—Кутты

0			
1		$\frac{1}{2}$	
$\frac{1}{2}$			
1		-1	2
	$\theta \left(1 + \theta \left(-\frac{3}{2} + \frac{2}{3} \theta\right)\right)$	$\theta^2 \left(2 - \frac{2}{3} \theta\right)$	$\theta^2 \left(\frac{2}{3} - \frac{\theta}{2}\right)$

**Пример 5.2.** Однако существует простой и элегантный способ интерполяции третьего порядка, если добавить вычисление функции  $f_1 = f(x_0 + h, y_1)$ , которая все равно будет нужна на следующем шаге. Мы имеем тогда два значения функции  $y_0, y_1$  и два значения производной  $f_0 = k_1, f_1$  и можем, следовательно, сделать кубическую интерполяцию. Для этого случая схема с разделенными разностями примет следующий вид:

$x_0$	$y_0$			
$x_0$	$y_0$	$f_0$		
$x_1$	$y_1$	$d$	$(d - f_0)/h$	
$x_1$	$y_1$	$f_1$	$(f_1 - d)/h$	$(f_1 - 2d + f_0)/h^2$

(5.2)

где

$$d = (y_1 - y_0)/h = \sum_i b_i k_i$$

и интерполяционная формула Ньютона дает

$$\begin{aligned} y(x_0 + \theta h) = & y_0 + h \theta [f_0 + \theta (d - f_0) + \\ & + \theta (\theta - 1) (f_1 - 2d + f_0)] \end{aligned} \quad (5.3)$$

(см. также Абрамович и Стиган (1964), формула 25.2.26).

**Пример 5.3.** Предыдущая идея допускает красивое расширение, если выполнить ричардсоновскую экстраполацию на двух шагах, как описано в теореме 4.1. Для большей точности воспользуемся экстраполированным значением  $y_2 := \hat{y}_2$  и сделаем вычисление производной  $f_2 = f(x_0 + 2h, y_2)$ , которая используется на следующем шаге. Далее, поскольку постоянная  $C$

в (4.1) определяется из (4.3) и (4.4), мы можем найти экстраполяцией лучшее значение и для  $y_1$ :

$$y_1 := y_1 + \frac{y_2 - w}{(2^p - 1)2}. \quad (5.4)$$

Однако без дополнительного вычисления в нашем распоряжении нет производной при новом значении  $y_1$ . Существует две возможности для увеличения точности:

А) воспользоваться интерполяционным многочленом четвертой степени, опираясь на значения  $y_0, f_0, y_1, y_2, f_2$ ;

Б) использовать в качестве  $f_1$  производную при значении  $y_1$  до экстраполяции и интерполировать при помощи многочлена пятой степени, опирающегося на значения  $y_0, f_0, y_1, f_1, y_2, f_2$  (пробные вычисления показали, что дополнительное вычисление функции при новом значении  $y_1$  — это бесполезная трата сил).

Если обозначить  $d_1 = (y_1 - y_0)/h$  и  $d_2 = (y_2 - y_1)/h$ , то аналогичная (5.3) интерполяционная формула Ньютона для случая А) дает

$$\begin{aligned} y(x_0 + \theta h) = & y_0 + h\theta [f_0 + \theta(d_1 - f_0 + (\theta - 1)((d_2 - \\ & - 3d_1 + 2f_0)/4 + \\ & + (\theta - 2)(f_2 - 2d_2 + 2d_1 - f_0)/4))], \end{aligned} \quad (5.5)$$

а для случая Б):

$$\begin{aligned} y(x_0 + \theta h) = & y_0 + h\theta [f_0 + \theta(d_1 - f_0 + (\theta - 1)(f_1 - 2d_1 + f_0) + \\ & + (\theta - 1)((d_2 - 4f_1 + 5d_1 - 2f_0)/4 + \\ & + (\theta - 2)(2f_0 - 6d_2 + 8f_1 - 6d_1 + 2f_0)/8))]. \end{aligned} \quad (5.6)$$

**Численный пример.** Для уравнений (4.16) с начальными значениями  $y_1(3) = \exp(\sin(9))$ ,  $y_2(3) = \exp(\cos(9))$  мы произвели сравнение двух предыдущих методов. Мы рассчитали решение на 10 двойных шагах при  $h = 0.05$  с помощью классического

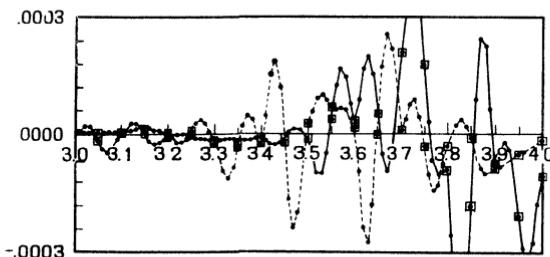


Рис. 5.1. Погрешность интерполяции формулы (5.5) на 10 двойных шагах РК-метода четвертого порядка для уравнения (4.16) (квадратиками обозначены вычисленные РК-методом точки).

метода Рунге—Кутты и затем для каждого двойного шага произвели интерполяцию 10 промежуточных значений. Полученные погрешности представлены графиками на рис. 5.1 и 5.2. Отчетливо видно, что формула (5.6) предпочтительнее. Формула (5.5) достаточна только для методов третьего порядка.

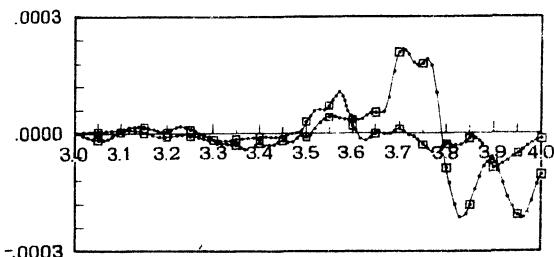


Рис. 5.2. Погрешность интерполяции формулы (5.6) в сочетании с классическим РК-методом четвертого порядка для примера (4.16).

### Непрерывные вложенные формулы

Непрерывные расширения методов Рунге—Кутты построили Хорн (1983) и Энрайт, Джексон, Нёрсетт и Томсен (1985). Хорн нашел, что формулу Фельберга (табл. 4.5) можно приспособить для получения глобального решения 4-го порядка точности, если добавить еще одно вычисление функции (правой части уравнения).

Для предпочтительного метода Дормана и Принса (табл. 4.6) непрерывное расширение 4-го порядка возможно даже без дополнительного вычисления функции. Это решение, при  $\theta = 1$  переходящее в решение  $y_1$  пятого порядка точности, определяется следующими формулами:

$$b_1(\theta) = \theta(1 + \theta(-1337/480 + \theta(1039/360 + \theta(-1163/1152)))),$$

$$b_2(\theta) = 0,$$

$$b_3(\theta) = 100\theta^2(1054/9275 + \theta(-4682/27825 + \theta(379/5565)))/3,$$

$$b_4(\theta) = -50^2(27/40 + \theta(-9/5 + \theta(83/96)))/2,$$

$$b_5(\theta) = 18225\theta^2(-3/250 + \theta(22/375 + \theta(-37/600)))/848,$$

$$b_6(\theta) = -22\theta^2(-3/10 + \theta(29/30 + \theta(-17/24)))/7,$$

$$y(x_0 + \theta h) \approx y_0 + h \sum_{j=1}^6 b_j(\theta) k_j. \quad (5.7)$$

Эти формулы вместе с коэффициентами из табл. 4.6 дают отличный инструмент для плотной выдачи результатов, графического представления решения, для решения уравнений с запаздывающим аргументом (см. разд. II.15), для «неявной» выдачи и т. п.

### «Неявная» выдача

Нередко значение  $x_{end}$ , для которого желательно вычислить и выдать решение, заранее не известно: оно зависит от вычисляемого решения неявным образом. Такие ситуации встречаются очень часто. Примером может служить отыскание периодических решений и предельных циклов, обсуждавшихся в разд. I.16; нам надо было узнать, когда решение достигнет сечения Пуанкаре в первый раз.

Такие задачи очень легко решаются методом Ньютона; производные решения по  $x$  известны из дифференциального уравнения. Пусть мы хотим определить  $x$ , которое удовлетворяет уравнению

$$F(x, y(x)) = 0. \quad (5.8)$$

**Алгоритм 5.4.** Вычисляйте решение дифференциального уравнения шаг за шагом, пока знак  $F(x_{i+1}, y_{i+1})$  не станет отличным от знака  $F(x_i, y_i)$  [это, однако, не вполне надежно, так как  $F$  может изменить знак дважды на интервале интегрирования; для повышения надежности можно использовать формулы (5.3), (5.5) или (5.6)]. Затем делайте итерации по Ньютону для отыскания решения уравнения (5.8):

$$x_i := x_i - \frac{F(x_i, y_i)}{F'(x_i, y_i)},$$

где

$$F'(x, y) = \frac{\partial F}{\partial x}(x, y) + \frac{\partial F}{\partial y^1}(x, y) \cdot f^1(x, y) + \cdots + \frac{\partial F}{\partial y^n}(x, y) \cdot f^n(x, y).$$

После этого новое значение  $y_1$  вычисляется с помощью одного шага методом РК с положительным или отрицательным значением  $h$ . Все численные примеры в разд. I.16 были рассчитаны этим способом.

Решение уравнения (5.8) можно рассчитать также с помощью непрерывного метода РК. В случае метода 5-го порядка, предложенного Дорманом и Принсом, это делается с помощью формул (5.7).

### Уравнения с разрывными производными

«Если вы пишете программу, не рассчитанную на любые ситуации, то рано или поздно найдется кто-нибудь, кто станет решать задачу с разрывами. Вам следует это учесть...»

(A. R. Кертис, 1986)

Функции, стоящие в правых частях дифференциальных уравнений, нередко бывают не аналитическими; они могут быть достаточно гладкими всюду, кроме некоторых поверхностей, на

которых терпят разрыв  $f_i(x, y)$  или их производные. Если разрывна сама  $f$ , задача Коши может вообще не иметь решения. Например, решение уравнения

$$y' = \begin{cases} -1 & (y \geq 0), \\ 1 & (y < 0) \end{cases} \quad (5.9)$$

перестает существовать после достижения оси  $x$ , так что начальная задача при  $y(0) = 0$  не имеет решения.

Однако обычно существует решение, которое удовлетворяет дифференциальному уравнению во всех точках непрерывности  $f$ . Возможны следующие подходы к решению таких задач:

**Способ 1.** На поверхности разрыва остановить вычисление решения (используя алгоритм 5.4, если необходимо) и возобновить его далее уже с новыми значениями  $f$ .

**Способ 2.** Игнорировать наличие разрыва и применять алгоритм с контролем локальной погрешности, автоматически управляющий длиной шага. Нет надобности говорить, что это может приводить к большим затратам времени счета и не дает особо надежных результатов.

**Численный пример.** Мы применили способ 2 с алгоритмом DOPR15, описанным в разд. II.4, к следующему примеру:

$$y' = \begin{cases} 1/(1+y), & \text{если } 0 \leq y \leq 2, \\ \begin{array}{l} \text{a)} 1 \\ \text{b)} 1/3 \\ \text{c)} (5-y)/9 \end{array} \end{cases}, \quad \text{если } 2 < y \quad (5.10)$$

( $y(0) = 1$ ,  $y(5) = ?$ ). Правильные ответы таковы:

a) 4.5; b)  $17/6$ ; c)  $5 - 3 \exp(-5/18)$ .

Для каждого случая мы провели расчет 81 раз, изменяя допустимую погрешность ( $10^{-8} \geq tol \geq 10^{-7}$ ), и результаты представили на рис. 5.3. По осям в логарифмических масштабах отложены погрешности вычисленных значений  $y(5)$  и число обращений к вычислению  $f$ .

Как видим, при изменении  $tol$  результаты ведут себя совершенно хаотично. Как правило, точность снижается на один — два десятичных знака, а число шагов очень велико. Другой пример подобного рода будет рассмотрен в разд. II.9 и II.10.

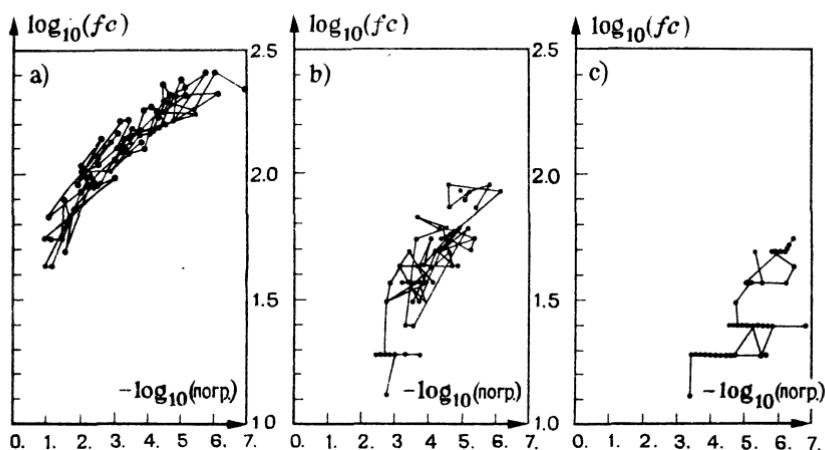


Рис. 5.3. Численные погрешности, полученные для (5.10) в случае игнорирования наличия разрыва.

### Длина начального шага

«Если все сделать с защитой от дураков, выведутся еще более замечательные дураки».

(Услышано от д-ра Пиркля в Бадене.)

В течение многих лет было принято задавать программе длину начального шага. Считалось, что пользователи имеют приблизительное представление о подходящей длине шага из математических соображений или из предшествующего опыта. Во всяком случае, плохой выбор начального  $h$  быстро исправляется механизмом управления длиной шага. Тем не менее, если делается очень плохой выбор и это происходит слишком часто, можно потратить впустую много времени счета. Поэтому ряд авторов (напр., Уоттс (1983), Хайндмарш (1980)) предложил идеи, как заставить компьютер делать этот выбор. Эти идеи в большинстве случаев опираются на гипотезу, что элементарные дифференциалы для «высоких» деревьев  $t_{21}, t_{32}, t_{44}, t_{59}$  и т. д. (см. табл. 2.1) имеют поведение типа  $\|f\|^q$ , как это имеет место для линейных дифференциальных уравнений с постоянными коэффициентами. Поэтому достаточно вычислить  $f(x, y)$  в заданной начальной точке, чтобы получить представление о порядке величины  $\|f\|$ . Несколько более тонкий подход воплощает следующий алгоритм.

а) Вычислите  $f(x, y)$  в начальной точке.

b) Положите

$$den = \left( \frac{1}{\max(|x_0|, |x_{end}|)} \right)^{p+1} + \|f\|^{p+1},$$

где  $p$  — порядок метода.

c) Положите

$$h = \left( \frac{tol}{den} \right)^{1/(p+1)}.$$

Очень часто, однако, начальные условия находятся в особом положении, где большинство компонент  $f_i$  может оказаться нулями. Поэтому добавим еще два шага к нашему алгоритму:

d) Сделайте один шаг методом Эйлера с длиной шага  $h$ , полученной в с).

e) Повторите шаги а) — с) с новым начальным значением и выберите меньшую из двух полученных длин шага.

Алгоритм типа приведенного или другой, подобный ему, обычно хорошо угадывает длину начального шага или, по крайней мере, предотвращает очень плохой выбор его. Лишь в некоторых случаях имеется дополнительная информация для выбора начального  $h$ , например из предыдущего опыта или решения аналогичных задач.

### Численное определение производных по начальным условиям и параметрам

Для эффективного решения краевых задач методом стрельбы, о котором говорилось в разд. I.15, нужны производные решений по (недостающим) начальным условиям. Аналогично, если мы хотим подобрать неизвестные параметры по имеющимся данным, скажем с помощью метода наименьших квадратов в нелинейном случае, нужны производные решений по параметрам, входящим в дифференциальное уравнение.

Однако результат численного определения этих производных ни в коем случае не будет гладкой функцией начальных условий или параметров из-за действия механизма управления погрешностью с его операторами условных переходов (IF ... THEN ...) и отбрасыванием забракованных шагов.

Пример. Для модели брюсселятора

$$\begin{aligned} y'_1 &= 1 + y_1^2 y_2 - (B + 1) y_1, & y_1(0) &= 1.3, \\ y'_2 &= B y_1 - y_1^2 y_2, & y_2(0) &= B \end{aligned} \tag{5.11}$$

мы рассчитали численное решение при  $x = 20$  как функцию от  $B$  при  $2.9 \leq B \leq 3.1$  с шагом  $\Delta B = 0.0001$ . Использовался метод Дормана и Принса с допустимой локальной погрешностью

$tol = 3 \times 10^{-4}$ . Графики разделенных разностей для  $\frac{\partial y_1}{\partial B}$  и  $\frac{\partial y_2}{\partial B}$  при  $x = 20$  представлены на рис. 5.4а и ясно показывают отсутствие гладкости.

Простой способ преодоления этой трудности состоит в замораживании всей последовательности длин шагов: выбранные про-

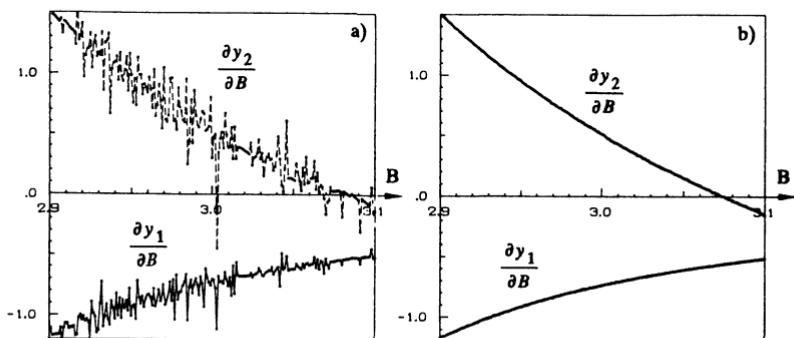


Рис. 5.4. Разделенные разности для численного решения системы (5.11): а) без замораживания длин шагов; б) с замораживанием длин шагов.

граммой в расчете с первым значением параметра  $B$  длины шагов заносятся в память и вновь используются во втором расчете для  $B + \Delta B$ . Разделенные разности результатов таких двух расчетов будут гораздо лучше аппроксимировать  $\frac{dy}{dB}$ . Эта процедура при очень малых  $\Delta B$  идентична тому, что называют «внутренним дифференцированием», но программируется она намного проще. Результаты, полученные для задачи (5.11) с замораживанием длин шагов, показаны на рис. 5.4б.

## Упражнения

1. (Бутчер.) Испытайте вашу любимую программу на следующем примере:

$$\begin{aligned} y'_1 &= f_1(y_1, y_2), \quad y_1(0) = 1, \\ y'_2 &= f_2(y_1, y_2), \quad y_2(0) = 0, \end{aligned}$$

где  $f$  определяется так:

```

if (|y1| < |y2|) then
    f1 = 0, f2 = sign (y1)
else
    f2 = 0, f1 = -sign (y2)
end if,

```

Вычислите  $y_1(8)$ ,  $y_2(8)$ . Покажите, что точное решение периодично.

2. Решите численно задачу  $y' = f(y)$ ,  $y(0) = 1$ ,  $y(3) = ?$ , где

$$f(y) = \begin{cases} y^2, & \text{если } 0 \leq y \leq 2, \\ \text{a)} \quad 1 \\ \text{b)} \quad 4 \\ \text{c)} \quad -4 + 4y \end{cases} \quad \text{если } 2 < y.$$

*Замечание.* Правильный ответ: а) 4.5; б) 12; в)  $\exp(10) + 1$ .

3. Рассмотрите  $s$ -стадийный метод Рунге—Кутты. Обозначьте  $s^*$  число различных значений  $c_i$ . Докажите, что порядок любого непрерывного расширения не превосходит  $s^*$ .

*Указание.* Пусть  $q(x)$  — многочлен степени  $s^*$ , удовлетворяющий условиям  $q(c_i) = 0$  при  $i = 1, \dots, s$ . Исследуйте выражение  $\sum_i b_i(\theta) q(c_i)$ .

## II.6. Явные методы Рунге — Кутты высших порядков

«Мы теперь конструируем формулу РК 10-го порядка...  
(Дж. Бутчер, 1969)

### Барьеры Бутчера

Для методов порядка 5 надо удовлетворить 17 условиям порядка (см. табл. 2.3). Выбрав  $s = 5$ , мы должны определить 15 параметров. Уже Кутта поднимал вопрос о том, что решение, возможно, все-таки существует, но он не надеялся на это и обратился прямо к случаю  $s = 6$  (см. разд. II.2, упр. 5). Вопрос, поставленный Куттой, оставался открытым в течение более 60 лет. Около 1963 года ответ на него был найден независимо в трех работах (Ческино, Кунцман (1963), Шанкс (1966), Бутчер (1964b, 1965b)). Работы Бутчера являются наиболее глубокими, и в нашем изложении мы будем следовать в основном его идеям.

**Теорема 6.1.** При  $p \geq 5$  не существует явных методов РК порядка  $p$  с числом стадий  $s = p$ .

*Доказательство.* Рассмотрим сначала случай  $s = p = 5$ . Введем (3, 3)-матрицы  $U$  и  $V$ :

$$U = \begin{bmatrix} \sum_i b_i a_{i2} & \sum_i b_i a_{i3} & \sum_i b_i a_{i4} \\ \sum_i b_i a_{i2} c_2 & \sum_i b_i a_{i3} c_3 & \sum_i b_i a_{i4} c_4 \\ g_2 & g_3 & g_4 \end{bmatrix},$$

$$V = \begin{bmatrix} c_2 & c_2^2 & \sum_l a_{2l} c_l - c_2^2/2 \\ c_3 & c_3^2 & \sum_l a_{3l} c_l - c_3^2/2 \\ c_4 & c_4^2 & \sum_l a_{4l} c_l - c_4^2/2 \end{bmatrix}, \quad (6.1)$$

где

$$g_h = \sum_{i,j} b_i a_{ij} a_{jh} - \frac{1}{2} \sum_i b_i a_{ih} (1 - c_h). \quad (6.2)$$

Из условий порядка для порядка 5 следует, что

$$UV = \begin{bmatrix} \frac{1}{6} & \frac{1}{12} & 0 \\ \frac{1}{12} & \frac{1}{20} & 0 \\ 0 & 0 & 0 \end{bmatrix}. \quad (6.3)$$

С помощью леммы 1.5 получаем отсюда, что  $g_4 = 0$  и, следовательно,  $c_4 = 1$ , как в лемме 1.4. Положим теперь в (6.1)

$$g_j = \left[ \sum_i b_i a_{ij} - b_j (1 - c_j) \right] (c_j - c_5). \quad (6.4)$$

Тривиальными вычислениями можно проверить, что матрица  $UV$  имеет снова вид (6.3). На этот раз отсюда следует, что  $c_4 = c_5$ , т. е.  $c_5 = 1$ . Следовательно, выражение

$$\sum_{i, j, k} b_i (1 - c_i) a_{ij} a_{jk} c_k \quad (6.5)$$

должно быть равно нулю (потому что  $2 \leq k < j < i$ ). Однако, выполнив умножение под знаком суммы, т. е. представив ее как разность двух сумм, и применяя к каждой соответствующее условие порядка (определенное из табл. 2.1), найдем, что выражение (6.5) должно быть равно  $1/120$ . Получилось противоречие.

Для случая  $p = s = 6$  надо рассмотреть все «одноногие» деревья, т. е. деревья, у которых корню инцидентно только одно ребро, на которое «привиты» деревья пятого порядка. Соответствующие условия порядка имеют вид

$$\sum_{i, j, \dots} b_i a_{ij} \text{ ( выражения для порядка 5) } = \frac{1}{\gamma(t)}.$$

Если положить  $b'_j = \sum_i b_i a_{ij}$ , эти условия будут иметь ту же структуру, что и в случае  $p = s = 5$ , и можно снова воспользоваться теми же идеями, что и выше. Но значения  $\gamma(t)$  будут здесь уже другими, и поэтому произведение  $UV$  вместо (6.3) будет равно

$$UV = \begin{bmatrix} \frac{1}{(s-2)!} & \frac{2!}{(s-1)!} & 0 \\ \frac{2!}{(s-1)!} & \frac{3!}{s!} & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (s=6) \quad (6.3')$$

Далее, для  $p = s = 7$  будем использовать деревья-«аисты», у которых условия порядка вида

$$\sum_{i, j, \dots} b_i a_{ij} a_{jk} \text{ ( выражения для порядка 5) } = \frac{1}{\gamma(t)}$$

и будем полагать  $b''_j = \sum_{i,j} b_i a_{ij} a_{jk}$ , и т. д. Мы надеемся, что общий случай  $p = s \geq 5$  теперь ясен. ■

### Шестистадийные процессы пятого порядка

Мы продемонстрируем во всех подробностях, как конструируются такие процессы, следуя при этом идеям, которые позволили Бутчеру (1964b) построить 7-стадийные формулы 6-го порядка. Такие же предположения, все более и более усложняясь, использовались потом во всех последующих построениях явных методов Рунге—Кутты (ЯРК) высоких порядков.

«Разыскивая такие процессы, мы, руководствуясь анализом, проведенным в предыдущем разделе, сделаем следующие предположения:»

$$\sum_{i=1}^6 b_i a_{ij} = b_j (1 - c_j), \quad j = 1, \dots, 6, \quad (6.6)$$

$$\sum_{i=1}^6 a_{ij} c_j = \frac{c_i^2}{2}, \quad i = 3, \dots, 6, \quad (6.7)$$

$$b_2 = 0. \quad (6.8)$$

С выгодами, которые дает условие (6.6), мы познакомились уже в разд. II.1 (см. лемму 1.3): оно позволяет пренебречь всеми «одноногими» деревьями, кроме  $t_{21}$ .

Использование условия (6.7) вместе с условием (6.8) дает аналогичный эффект. Например, равенства

$$\sum_{i, j, k} b_i a_{ij} c_j a_{ik} c_k = \frac{1}{5 \cdot 2 \cdot 2},$$

$$\sum_{i, l} b_i c_i^2 a_{il} c_l = \frac{1}{5 \cdot 2},$$

$$\sum_i b_i c_i^4 = \frac{1}{5}$$

становятся эквивалентными. Каждая отходящая от корня ветвь, состоящая из двух ребер и двух вершин, заменяется двумя ветвями, несущими по одной вершине каждая. Условие (6.7) не может выполняться для всех  $i$ , потому что при  $i = 2$  оно выглядит так:  $a_{21} c_1 = c_2^2 / 2$ , откуда следует, что  $c_2 = 0$ , а тогда и все остальные  $c_i$  будут равны нулю.

После описанного выше приведения останутся только деревья, соответствующие квадратурным условиям порядка

$$\sum_i b_i c_i^{q-1} = \frac{1}{q}, \quad q = 1, 2, 3, 4, 5 \quad (6.9)$$

(которые выражают условия того, что квадратурная формула с весами  $b_i$  и узлами  $c_i$ , аппроксимирующая интеграл по отрезку

[0, 1], будет точной для многочленов меньше пятой степени) и еще двум уравнениям:

$$\sum_{i, j, k} b_i c_i a_{ij} a_{jk} c_k = \frac{1}{5 \cdot 3 \cdot 2}, \quad (6.10)$$

$$\sum_{i, l} b_i c_i a_{il} c_l^2 = \frac{1}{5 \cdot 3}. \quad (6.11)$$

Вычитая последнее уравнение из предыдущего, получим

$$\sum_{i, l} b_i c_i a_{il} \left( \sum_k a_{jk} c_k - \frac{c_l^2}{l^2} \right) = 0.$$

В силу (6.7) выражение в скобках не обращается в нуль только при  $j = 2$ , откуда

$$\sum_i b_i c_i a_{i2} = 0. \quad (6.12)$$

Упрощение можно продолжить. Из (6.6) следует, что  $c_6 = 1$ . Рассмотрим теперь условия порядка, аналогичные (6.10) и (6.11), но без  $c_i$  в левой части, и проделаем с ними те же операции, которые мы выполнили над (6.10) и (6.11), а из результата вычтем равенство (6.12). Мы получим уравнение

$$\sum_i b_i (1 - c_i) a_{i2} = 0, \quad (6.12')$$

где отличны от нуля только три члена суммы. Сделаем последнее упрощение. Вычитая и складывая (6.11) с другими условиями порядка, мы придем к равенству

$$\sum_{i, l} b_i (1 - c_i) a_{il} c_j (c_j - c_3) = \frac{1}{60} - \frac{c_3}{24}. \quad (6.13)$$

(Здесь берется разность условий порядка, соответствующих деревьям  $t_{43}$  и  $t_{53}$  из табл. 2.1, и из нее вычитается умноженная на  $c_3$  разность условий, соответствующих деревьям  $t_{32}$  и  $t_{42}$ . — Прим. ред.) При этом опять число членов в сумме слева уменьшилось. Результатом проведенного рассмотрения является

**Теорема 6.2.** Следующий алгоритм дает 6-стадийные явные методы РК порядка 5 с пятым свободными параметрами:

1.  $c_2, c_3, c_4, c_5$  могут быть выбраны в качестве свободных параметров, подчиненных лишь некоторым очевидным ограничениям.

2.  $c_6 = 1$ , что следует из (6.6) при  $j = 6$ .

3.  $b_1, b_3, b_4, b_5, b_6$  определяются из линейной системы (6.9), а  $b_2 = 0$ .

4.  $a_{32}$  находим из (6.7) при  $i = 3$ ,  $a_{42} = \lambda$  выбираем произвольно,  $a_{43}$  определяем из (6.7) при  $i = 4$ .

5.  $a_{52}$  находим из (6.12'),  $a_{54}$  — из (6.13).
6.  $a_{53}$  находим из (6.7) при  $i = 5$ .
7.  $a_{63}, a_{64}, a_{65}$  определяем из линейной системы, образованной уравнениями (6.6) при  $j = 3, 4, 5$ .
8.  $a_{62}$  находим из (6.7) при  $i = 6$ .
9.  $a_{i1} (i = 2, \dots, 6)$  находятся из (1.9). ■

Примерами таких процессов являются метод Кутты—Ньюстрёма (разд. II.2, упр. 5), формула Фельберга (табл. 4.5) и формула пятого порядка Дормана—Принса (табл. 4.6). Возможен и иной выбор свободных параметров (см. упр. 1).

### Семистадийные процессы шестого порядка

Такие процессы вывел Бутчер (1964b), следуя описанным выше путем. На этот раз, кроме уравнений (6.6)–(6.11), надо дополнительно учесть еще шесть уравнений, так что вычисления становятся более сложными. Бутчер пришел к алгоритму, где  $c_2, c_3, c_5, c_6$  являются свободными параметрами. Среди множества возможных примеров мы выбрали метод, представленный в табл. 6.1, потому что он обладает положительными весами  $b_1, b_3, \dots, b_7$ .

Таблица 6.1. Метод Бутчера 6-го порядка

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	$-\frac{41}{260}$	$\frac{22}{13}$	$\frac{43}{156}$	$-\frac{118}{39}$	$\frac{32}{195}$	$\frac{80}{39}$	
	$-\frac{13}{200}$	0	$\frac{11}{40}$	$\frac{11}{40}$	$\frac{4}{25}$	$\frac{4}{25}$	$\frac{13}{200}$

### Дальнейшие барьеры Бутчера

Обратимся теперь к методам 7-го порядка. Оказывается, что существование таких методов с восемью стадиями невозможно — этому препятствует следующий барьер:

**Теорема 6.3.** [Бутчер (1965b)] При  $p \geq 7$  не существует явных методов РК порядка  $p$ , имеющих  $s = p + 1$  стадию. ■

Мы не воспроизводим здесь доказательства, поскольку оно намного сложнее, чем у теоремы 6.1.

Что касается методов порядка 8, то Кёртис (1970), а также Купер и Вернер (1972) построили такие процессы с  $s = 11$ . Долгое время оставался открытым вопрос, существуют ли такие методы с 10 стадиями. Мечте Джона Бутчера разрешить этот трудный вопрос до своего 50-летия не суждено было осуществиться. Но в конце концов ему удалось доказать, что такие методы не существуют, к 60-летию Далквиста.

**Теорема 6.4.** [Бутчер (1985b)]. При  $p \geq 8$  не существует явных методов РК порядка  $p$ , имеющих  $s = p + 2$  стадии. ■

Доказательство этой теоремы еще более сложно, и мы снова отошлем читателя к оригинальной статье Бутчера.

### Формула десятого порядка

Наивысший порядок, фактически достигнутый для явно построенных ЯРК-методов, равен *десяти* (Книга рекордов Гиннесса, с. 333). Кёртис (1975) сконструировал 18-стадийный метод порядка 10. Его построение было основано только на упрощающих предположениях типа (6.7), (6.8) и их расширениях. Позже Хайрер (1978) сконструировал 17-стадийный метод, используя полный арсенал идей упрощения. Коэффициенты не являются рациональными числами, поскольку они построены на базе квадратурной формулы Лобатто 10-го порядка. Как мы увидим далее, этот метод может быть полезен, если требуется точность выше чем примерно  $10^{-15}$ . Поэтому мы не воспроизводим здесь коэффициентов метода в виде десятичных дробей [они приведены с 21 десятичным знаком в работе Хайрера (1978)]. Вместо этого мы представим их здесь в виде решений последовательности линейных уравнений, откуда их можно получить с любой желаемой точностью (далее с. п. = свободный параметр):

$$c_1 = 0, \quad c_2 = c_{18} = \frac{1}{2} \text{ (с. п.)}, \quad c_{17} = 1, \\ c_6 = c_{13} = 0.766654 \text{ (с. п.)},$$

$$c_9 = \frac{1}{2} \left[ 1 - \left[ \frac{7 - 2\sqrt{7}}{21} \right]^{1/2} \right], \quad c_{12} = \frac{1}{2} \left[ 1 - \left[ \frac{7 + 2\sqrt{7}}{21} \right]^{1/2} \right],$$

$$c_{11} = 1 - c_9, \quad c_{10} = 1 - c_{12} \quad (\text{четыре узла квадратурной формулы Лобатто}),$$

$$(1 - c_{14}) = \frac{4(1 - c_{13}) - 3(1 - c_{12})}{6(1 - c_{13}) - 4(1 - c_{12})} (1 - c_{12}),$$

$$c_7 = c_{14}, \quad (1 - c_{15}) = \frac{2}{3} (1 - c_{14}), \quad c_8 = c_{15}, \quad c_4 = \frac{3}{2} c_3,$$

$$c_5 = \frac{4c_4 - 3c_6}{6c_4 - 4c_6} c_6, \quad c_8 = \frac{20c_6c_7 - 15c_9(c_6 + c_7) + 12c_9^2}{30c_6c_7 - 20c_9(c_6 + c_7) + 15c_9^2} c_9,$$

$$b_{13} = 0.13 \text{ (c. p.)}, \quad b_{14} = 0.18 \text{ (c. p.)}, \quad b_{15} = 0.12 \text{ (c. p.)},$$

$$b_{16} = 1/30 \text{ (c. n.)},$$

$$b_4 = 0, \ b_5 = 0, \ b_8 = 0, \ b_3 + b_{16} = 0,$$

$$b_3 + b_{15} = 0, \quad b_6 + b_{13} = 0, \quad b_7 + b_{14} = 0,$$

$$b_{10} = b_{12} = \frac{1}{60}(14 - \sqrt{7}), \quad b_9 = b_{11} = \frac{1}{60}(14 + \sqrt{7}),$$

$$b_1 = b_{17} = -\frac{1}{30}$$

(веса квадратурной формулы Лобатто).

Вычисление ненулевых коэффициентов  $a_{ij}$  производится затем в четыре этапа, как показано в табл. 6.2.

### Таблица 6.2.

*Этап 1.* Используются уравнения

$$\begin{aligned}
 & \sum_{j=1}^{i-1} a_{ij} = c_i, \quad i = 2, \dots, 17, \\
 & \sum_{j=1}^{i-1} a_{ij} c_j = \frac{c_i^2}{2}, \quad i = 3, \dots, 14, 17, \\
 & \sum_{j=1}^{i-1} a_{ij} c_j^2 = \frac{c_i^3}{3}, \quad i = 4, \dots, 14, 17, \\
 & \sum_{j=1}^{i-1} a_{ij} c_j^3 = \frac{c_i^4}{4}, \quad i = 6, \dots, 14, 17, \\
 & \sum_{j=1}^{i-1} a_{ij} c_j^4 = \frac{c_i^5}{5}, \quad i = 9, 10, 11, 12, 17, \\
 & \sum_{j=1}^{12} a_{13,j} c_j^4 = \sum_{l=1}^5 a_{6,l} c_l^4, \quad \sum_{j=1}^{13} a_{14,j} c_j^4 = \sum_{l=1}^6 a_{7,l} c_l^4
 \end{aligned} \tag{6.14}$$

для  $i \leq 10$ . При  $i = 2$  имеется единственное условие  $a_{2,1} = c_2$ . При  $i = 3$  имеем уже два условия:

$$\begin{aligned}
 a_{3,1} + a_{3,2} &= c_3, \\
 a_{3,1} c_1 + a_{3,2} c_2 &= \frac{c_3^2}{2}.
 \end{aligned}$$

Из этой линейной системы определяются  $a_{3,1}$  и  $a_{3,2}$ . Аналогичным образом все ненулевые  $a_{ij}$  при  $i \leq 10$  вычисляются путем решения линейных систем с матрицами типа Вандермонда.

*Этап 2.* Здесь используются условия

$$\begin{aligned}
 & \sum_{l=j+1}^{17} b_l a_{ij} = b_j (1 - c_j), \quad j = 11, 12, 13, 14, 15, 16, \\
 & \sum_{l=j+1}^{17} b_l c_l a_{il} = \frac{1}{2} b_l (1 - c_l^2), \quad j = 11, 12, 13, 14, 15, \\
 & \sum_{l=j+1}^{17} b_l c_l^2 a_{il} = \frac{1}{3} b_l (1 - c_l^3), \quad j = 11, 12, 13, 14, \\
 & \sum_{l=j+1}^{17} b_l c_l^3 a_{il} = \frac{1}{4} b_l (1 - c_l^4), \quad j = 11, 12.
 \end{aligned} \tag{6.15}$$

При  $j = 16$  имеется единственное условие, которое дает

$$a_{17, 16} = \frac{b_{16}}{b_{17}} (1 - c_{16}).$$

Два условия при  $j = 15$  дают коэффициенты  $a_{16, 15}$  и  $a_{17, 15}$ , и т. д. Здесь опять приходится решать линейные системы с матрицей типа Вандермонда.

*Этап 3.*

$$\begin{aligned} a_{15, 6} &= -a_{15, 18}, \quad a_{15, 7} = -a_{15, 14}, \quad a_{16, 8} = -a_{16, 15}, \quad a_{17, 2} = -a_{17, 18}, \\ a_{17, 8} &= -a_{17, 15}, \quad a_{18, 4} = a_{6, 4}, \quad a_{18, 5} = a_{6, 5}, \quad a_{14, 4} = a_{7, 4}, \\ a_{14, 5} &= a_{7, 5} \quad a_{15, 2} = a_{3, 2}. \end{aligned}$$

Значения  $a_{15, 1}$  и  $a_{16, 1}$  вычисляются из первого уравнения (6.14) при  $i = 15$  и  $i = 16$  соответственно.

*Этап 4.* Введем величины

$$w_i = \sum_{j=1}^{i-1} a_{ij} c_j^5, \quad i = 11, 12, 13, 14, 17, \quad (6.16)$$

и вычислим  $w_i$ , решая систему пяти уравнений

$$\sum_{i=11, 12, 13, 14, 17} b_i c_i^k w_i = \frac{1}{(7+k)6} - \sum_{i=6, 7, 9, 10} b_i c_i^k \sum_{j=1}^{i-1} a_{ij} c_j^5, \quad (6.17)$$

где  $k = 0, 1, 2, 3, 4$  ( $k = 4$  добавлено произвольно). Затем воспользуемся уравнениями (6.16) вместе с (6.14) при  $i = 11, 12, 13, 14, 17$ , чтобы для этих значений  $i$  вычислить  $a_{ij}$  с  $j = 6, \dots, 10$ . Наконец, для вычисления оставшихся  $a_{i1}$  воспользуемся первым уравнением (6.14).

Читателя, желающего глубже разобраться в проверке условий порядка, мы отшлем к оригинальной статье Хайрера (1978).

Дальнейшее повышение порядка методов теряет смысл, если нет особой нужды в них как в инструменте для вычислений с очень высокой точностью. Получение формул еще более высокого порядка превращается в сложную проблему, бросающую вызов математикам. Трудности растут быстрее, чем по экспоненте, а методы становятся громадными монстрами и делают к тому же управление длиной шага все более и более трудным.

### Вложенные формулы высоких порядков

Фельбергу — гению манипуляций с формулами — по существу первому удалось построить вложенные формулы высоких порядков. Его крупнейшим успехом была формула 7-го порядка с оценкой погрешности с помощью формулы 8-го порядка. Этот его метод часто используется во всех вычислениях, требующих высокой точности, например в астрономии. Мы с большим удовольствием воспроизводим в табл. 6.3 оригинал страницы из

Таблица 6.3. Метод Фельберга, порядок 7(8)

$k\backslash\lambda$	$\alpha_k$	$B_{k\lambda}$										$c_k$	$\hat{c}_k$
0	0	0	0	1	2	3	4	5	6	7	8	9	10
1	$\frac{2}{27}$	$\frac{2}{27}$	$\frac{1}{27}$	$\frac{1}{36}$	$\frac{1}{12}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{8}$	$\frac{25}{16}$	$\frac{16}{5}$	$\frac{1}{4}$	$\frac{1}{5}$	$\frac{34}{105}$
2	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{36}$	$\frac{1}{12}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{25}$	$\frac{25}{16}$	$\frac{16}{5}$	$\frac{1}{4}$	$\frac{1}{5}$	$\frac{34}{105}$	$\frac{9}{35}$
3	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{1}{4}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{34}{105}$	$\frac{9}{35}$
4	$\frac{5}{12}$	$\frac{5}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{1}{4}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{34}{105}$	$\frac{9}{35}$
5	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{34}{105}$	$\frac{9}{35}$
6	$\frac{5}{6}$	$\frac{5}{6}$	$\frac{108}{108}$	$\frac{108}{108}$	$\frac{0}{108}$	$\frac{0}{108}$	$\frac{0}{108}$	$\frac{0}{108}$	$\frac{27}{54}$	$\frac{27}{54}$	$\frac{27}{54}$	$\frac{27}{54}$	$\frac{9}{280}$
7	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{31}{300}$	$\frac{31}{300}$	$\frac{0}{300}$	$\frac{0}{300}$	$\frac{0}{300}$	$\frac{0}{300}$	$\frac{61}{225}$	$\frac{61}{225}$	$\frac{61}{225}$	$\frac{61}{225}$	$\frac{9}{280}$
8	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{0}{53}$	$\frac{0}{53}$	$\frac{0}{53}$	$\frac{0}{53}$	$\frac{704}{67}$	$\frac{704}{67}$	$\frac{704}{67}$	$\frac{704}{67}$	$\frac{9}{280}$
9	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{91}{108}$	$\frac{91}{108}$	$\frac{0}{108}$	$\frac{0}{108}$	$\frac{0}{108}$	$\frac{0}{108}$	$\frac{45}{9}$	$\frac{45}{9}$	$\frac{45}{9}$	$\frac{45}{9}$	$\frac{9}{280}$
10	1	$\frac{2383}{4100}$	0	0	$-\frac{341}{164}$	$-\frac{4496}{1025}$	$-\frac{301}{135}$	$-\frac{2133}{54}$	$-\frac{45}{60}$	$-\frac{17}{6}$	$-\frac{1}{12}$	$-\frac{1}{12}$	$\frac{41}{840}$
11	0	$\frac{3}{205}$	0	0	0	0	0	0	$-\frac{6}{41}$	$-\frac{3}{41}$	$-\frac{3}{41}$	$-\frac{3}{41}$	$-\frac{41}{840}$
12	1	$-\frac{1777}{4100}$	0	0	$-\frac{341}{164}$	$-\frac{4496}{1025}$	$-\frac{289}{82}$	$-\frac{2193}{82}$	$-\frac{51}{4100}$	$-\frac{33}{82}$	$-\frac{12}{82}$	$-\frac{1}{82}$	$-\frac{41}{840}$

Таблица 6.4. Метод Дормана—Принса 8(7) (DOPRI8)

$\epsilon_i$	$a_{ij}$	$b_i$	$b_i$
0			
$\frac{1}{18}$	$\frac{1}{18}$		
$\frac{1}{12}$	$\frac{1}{48}$	$\frac{1}{16}$	
$\frac{1}{8}$	$\frac{1}{32}$	$\frac{3}{32}$	
$\frac{5}{16}$	$\frac{5}{16}$	$\frac{75}{64}$	
$\frac{3}{8}$	$\frac{3}{80}$	$\frac{0}{16}$	
$\frac{59}{400}$	$-29443841$	$0$	$0$
$\frac{93}{200}$	$6144563906$	$692538347$	$11256000000$
$\frac{246}{130487787}$	$-16016141$	$61564180$	$227897113$
$\frac{12}{1299019798}$	$-1028468189$	$158732637$	$633445777$
$1$	$188892177$	$0$	$-3765042795$
$1$	$718116043$	$0$	$-3185094517$
$1$	$402863854$	$0$	$-506849293$
$491063109$	$5434740067$	$543043805$	$914296604$
		$\frac{14005451}{3155480064}$	$\frac{13451932}{455176623}$
		$0$	$0$
		$\frac{-59238491}{106877825}$	$\frac{-808719846}{976000145}$
		$0$	$0$
		$\frac{181666767}{75867731}$	$\frac{1757004468}{5445159321}$
		$0$	$0$
		$\frac{56129285}{797845732}$	$\frac{656045339}{265891186}$
		$0$	$0$
		$\frac{-104159430}{1371343529}$	$\frac{-3867574721}{1518517206}$
		$0$	$0$
		$\frac{760417239}{151163299}$	$\frac{465885868}{32736535}$
		$0$	$0$
		$\frac{11880643}{751138087}$	$\frac{5011238}{667516719}$
		$0$	$0$
		$\frac{123872331}{1001029789}$	$\frac{3962137247}{1879008318}$
		$0$	$0$
		$\frac{393006217}{1396673457}$	$\frac{306593473}{597172653}$
		$0$	$0$
		$\frac{103282649}{3398467676}$	$\frac{597172653}{3398467676}$
		$0$	$0$
		$\frac{523286602}{5731586787}$	$\frac{55666358}{3962137247}$
		$0$	$0$
		$\frac{-4093664535}{808688257}$	$\frac{-55666358}{879700083}$
		$0$	$0$
		$\frac{850066563}{102754527}$	$\frac{2220607170}{1805937418}$
		$0$	$0$
		$\frac{-13158990841}{11173962825}$	$\frac{-160528059}{3936647629}$
		$0$	$0$
		$\frac{6184727034}{1978049860}$	$\frac{1}{4}$
		$0$	$0$

отчета Фельберга (1968), тем более что более поздняя журнальная публикация содержит опечатку.

Методы Фельберга страдают двумя недостатками:

- 1) как уже упоминалось, они задуманы для продолжения интегрирования по формуле  $p$ -го порядка с минимизированными коэффициентами погрешности и поэтому имеют тенденцию недооценивать погрешность;
- 2) для задач вида  $y' = f(x)$ , сводящихся к квадратурам, они дают оценки погрешностей, тождественно равные нулю.

Первые вложенные формулы высших порядков, свободные от второго недостатка, построил Вернер (1978). Его метод порядков 8 (9) включен нами в число тестируемых в приведенном ниже численном примере.

Прекрасные численные результаты дает метод порядков 8 (7), который предложил Принс и Дорман (1981), приложившие много усилий, чтобы минимизировать коэффициенты погрешности для аппроксимации 8-го порядка. Коэффициенты их формулы воспроизведены в табл. 6.4 (с разрешения авторов).

Реализующая этот метод подпрограмма DOPRI8, написанная на Фортране, включена в приложение к этой книге. В ней осуществлена минимизация требований к объему памяти (см. упр. 5).

### Численный пример

Все упомянутые выше методы были запрограммированы с несколько более сложным управлением длиной шага. Их испытание проводилось на ЭВМ Univac 1100/60 (машиначная точность  $= 1.73 \times 10^{-18}$ ). Решался пример (4.16) при задании допустимой погрешности в интервале  $10^{-6} \geq tol \geq 10^{-17}$ . Метод Бутчера 6-го порядка, который на рис. 4.3 был методом самого высокого порядка, здесь сравнивался с методами еще более высоких порядков. Из рис. 6.1 опять ясно, что формула Фельберга недооценивает погрешность. Отчетливо видно превосходство метода Дормана—Принса при  $tol \geq 10^{-13}$ ; далее его превосходит метод 10-го порядка.

Коэффициенты в DOPRI8 не являются точными рациональными числами: это лишь аппроксимации непрерывными дробями с точностью до 18 десятичных знаков. Поэтому программа DOPRI8 не должна использоваться при слишком высоких требованиях к точности (при  $tol \leq 10^{-16}$ ). На рис. 6.2 показаны результаты, которые получены на дающей высокую точность машине CDC для задачи  $\dot{y}_1 = y_2/6$ ,  $\dot{y}_2 = -y_1 y_2^2/54$ ,  $y_1(4) = 2$ ,  $y_2(4) = 1.5$ , точное решение которой  $y_1(9) = 3$ ,  $y_2(9) = 1$ .

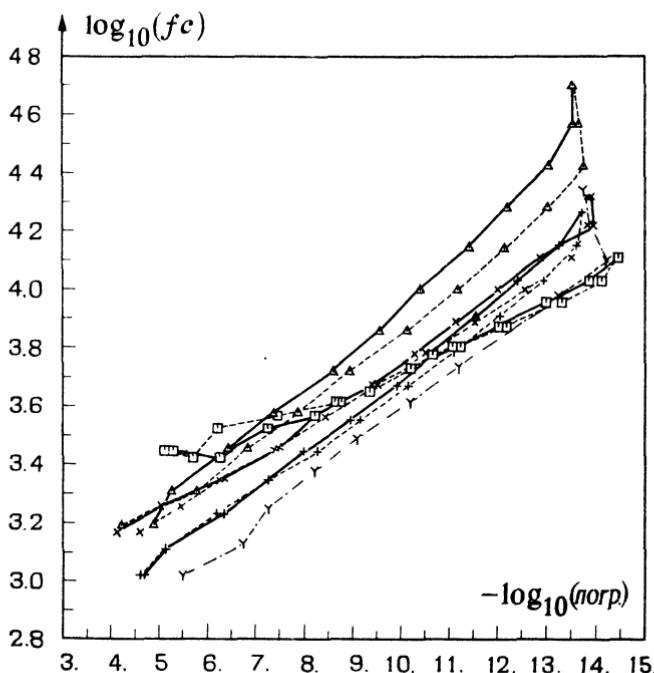


Рис. 6.1. Число обращений к подпрограмме вычисления функции ( $fc$ ) в зависимости от глобальной погрешности для значений допустимой погрешности от  $10^{-6}$  до  $10^{-17}$ . Сплошная линия — без локальной экстраполяции, штриховая — с локальной экстраполяцией; квадраты — метод Хайрера 10-го порядка, табл. 6.2; треугольники — метод Бутчера 6-го порядка, табл. 6.1; прямые крестики — метод Фельберга 7 (8), табл. 6.3; косые крестики — метод Вернера 8 (9), (1978); штрих-пунктирная линия — программа DOPRI8, табл. 6.4.

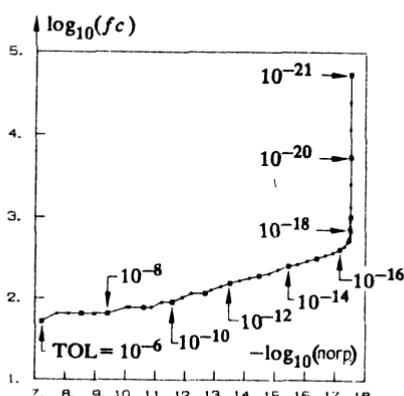


Рис. 6.2. Результаты для метода Дормана — Принса (7,8) при малых значениях допустимой погрешности.

### Упражнения

- Найдите условия, которым надо подчинить свободные коэффициенты, чтобы метод, описанный в теореме 6.2, мог быть снабжен вложенной формулой 4-го порядка для оценки погрешности. Как выбрать коэффициенты, чтобы все  $b_i$  были неотрицательны? Чтобы коэффициенты погрешности для 6-го порядка были минимальны?
- Постройте 6-стадийную формулу 5-го порядка, у которой  $c_3 = 1/3$ ,  $c_4 = 1/2$ ,  $c_5 = 2/3$ .
- (Бутчер) Покажите, что для любого метода РК порядка 5

$$\sum_i b_i \left( \sum_j a_{ij} c_j - \frac{c_i^2}{2} \right)^2 = 0$$

(следовательно, не существует методов ЯРК 5-го порядка, у которых все  $b_i > 0$ ).

**Указание.** Раскройте скобки и воспользуйтесь условиями порядка.

- Напишите (или возьмите) программу с методом РК высокого порядка и рассчитайте орбиту Аренсторфа для ограниченной задачи трех тел (0.1) (см. введение к этой главе) с начальными условиями

$$\begin{aligned} y_1(0) &= 0.994, \quad y'_1(0) = y_2(0) = 0, \\ y'_2(0) &= -2.0317326295573368357302057924, \\ \mu &= 0.012277471, \quad \mu' = 1 - \mu. \end{aligned}$$

Вычислите решение при

$$x_{\text{end}} = 11.124340337266085134999734047.$$

Начальные условия выбраны так, чтобы решение было периодическим (с такой точностью). Если начертить орбиту, соответствующую решению, на ней будет на одну петлю меньше, чем на той, что приведена во введении (рис. 0.1).

- (Шампайн, 1979.) Покажите, что метод Дормана—Принса 8 (7) (табл. 6.4) можно реализовать в программе таким образом, чтобы требуемая память зависела от размера  $n$  системы как  $9n$ .  
**Указание.** а) Воспользуйтесь расположением нулей в таблице коэффициентов. б) (Метод III): вычислив величины  $k_1, k_2, \dots, k_9$  [см. (1.8)], подсчитайте суммы

$$\sum_{j=1}^9 a_{ij} k_j \quad \text{для } i = 11, 12, 13, \quad \sum_{j=1}^9 b_j k_j, \quad \sum_{j=1}^9 \bar{b}_j k_j.$$

Тогда освободится место в памяти, занятое величинами  $k_2, k_3, \dots$  и т. д., которые больше не нужны. См. программу DOPR18 в приложении.

## II.7. Неявные методы Рунге — Кутты

«Было традицией рассматривать только явные процессы».  
(Дж. К. Бутчер, 1964)

«Быстро действующие вычислительные машины сделали возможным использование преимуществ сложных методов».

(П. К. Хаммер и Дж. У. Холлингсуорт, 1955)

### Введение

Первые *неявные* методы РК использовал Коши (1824) для ... (вы правильно угадали!) оценки погрешностей (*«Méthodes diverses qui peuvent être employées au Calcul numérique...»*<sup>1</sup>), см. упр. 6). Коши применил теорему о среднем значении к интегралу, рассмотренному в разд. I.8 и II.1,

$$y(x_1) = y(x_0) + \int_{x_0}^{x_1} f(x, y(x)) dx, \quad (7.1)$$

и получил формулу

$$y_1 = y_0 + hf(x_0 + \theta h, y_0 + \Theta(y_1 - y_0)), \quad (7.2)$$

где  $0 \leqslant \theta, \Theta \leqslant 1$  (так называемый « $\theta$ -метод»). Предельными случаями ее являются  $\theta = \Theta = 0$  (явный метод Эйлера) и  $\theta = \Theta = 1$ :

$$y_1 = y_0 + hf(x_1, y_1). \quad (7.3)$$

Эту формулу называют *неявным методом Эйлера*.

Для повышения эффективности вычислительного процесса применим, как это мы сделали в разд. II.1, «правило средней точки» ( $\theta = \Theta = 1/2$ ) и, положив  $k_1 = (y_1 - y_0)/h$ , получим из (7.2) формулу

$$\begin{aligned} k_1 &= f\left(x_0 + \frac{h}{2}, y_0 + \frac{h}{2}k_1\right), \\ y_1 &= y_0 + hk_1. \end{aligned} \quad (7.4)$$

Этот метод называют *неявным правилом средней точки*.

<sup>1</sup> «Различные методы, которые могут быть использованы в численных расчетах...» — Франц.

Есть еще и другая возможность: аппроксимируя интеграл (7.1) по правилу трапеций, получим

$$y_1 = y_0 + \frac{h}{2} [f(x_0, y_0) + f(x_1, y_1)]. \quad (7.5)$$

Рассмотрим еще квадратурную формулу Радо:

$$\begin{aligned} y(x_1) - y(x_0) &= \int_{x_0}^{x_0+h} f(x, y(x)) dx \approx \\ &\approx \frac{h}{4} \left( f(x_0, y_0) + 3f\left(x_0 + \frac{2}{3}h, y\left(x_0 + \frac{2}{3}h\right)\right) \right). \end{aligned}$$

Здесь надо как-то аппроксимировать значение  $y(x_0 + 2h/3)$ . Одна из возможностей состоит в использовании квадратичной интерполяции по значениям  $y_0$ ,  $y'_0$  и  $y(x_1)$ :

$$y\left(x_0 + \frac{2}{3}h\right) \approx \frac{5}{9}y_0 + \frac{4}{9}y(x_1) + \frac{2}{9}hf(x_0, y_0).$$

Получающийся при этом метод предложили Хаммер и Холлингсупорт (1955):

$$\begin{aligned} k_1 &= f(x_0, y_0), \\ k_2 &= f\left(x_0 + \frac{2}{3}h, y_0 + \frac{h}{3}(k_1 + k_2)\right), \\ y_1 &= y_0 + \frac{h}{4}(k_1 + 3k_2). \end{aligned} \quad (7.6)$$

Схемы (7.4) и (7.6) имеют такой же вид, как формулы (1.8), если суммирование в аргументах  $f$  распространить до « $s$ ».

**Определение 7.1.** Пусть  $b_i$ ,  $a_{ij}$  ( $i, j = 1, \dots, s$ ) — вещественные числа. Метод

$$k_i = f\left(x_0 + c_i h, y_0 + h \sum_{j=1}^s a_{ij} k_j\right), i = 1, \dots, s, \quad (7.7)$$

$$y_1 = y_0 + h \sum_{i=1}^s b_i k_i$$

называется  $s$ -стадийным методом Рунге—Кутты. Если  $a_{ij} = 0$  при  $i \leq j$ , имеем явный метод (ЯРК). Если  $a_{ij} = 0$  при  $i > j$  и хотя бы одно  $a_{ii} \neq 0$ , имеем диагонально неявный метод РК (ДНРК). Если, кроме того, все диагональные коэффициенты одинаковы,  $a_{ii} = \gamma$  ( $i = 1, \dots, s$ ), мы говорим об однократно диагонально неявном методе РК (ОДНРК-методе). Во всех остальных

случаях мы говорим о *неявных* методах Рунге—Кутты (НРК-методах)<sup>1)</sup>. ■

Использованную выше запись коэффициентов ЯРК-методов в виде таблицы легко распространить на все случаи, включая в таблицу все ненулевые  $a_{ij}$ , в том числе лежащие на диагонали и выше. Такая запись для методов (7.4) и (7.6) приведена в табл. 7.1.

Таблица 7.1. Неявные методы Рунге—Кутты

«Неявное правило средней точки»		Метод Хаммера—Холлингсуорта		
$\frac{1}{2}$	$\frac{1}{2}$	0 $\frac{2}{3}$	0 $\frac{1}{3}$	0 $\frac{1}{3}$
	1		$\frac{1}{4}$	$\frac{3}{4}$

Интерес к неявным методам РК возобновился и даже вырос в связи с проблемой *жестких* дифференциальных уравнений (см. том II).

### Существование численного решения

Для неявных методов значения  $k_i$  уже нельзя вычислять последовательно, так как (7.7) в этом случае представляет собой систему уравнений, определяющих величины  $k_i$  неявным образом. В случае методов ДНРК имеем последовательность систем размерности  $n$ : для  $k_1$ , затем для  $k_2$  и т. д. Для полностью неявных методов все  $s$  неизвестных ( $k_i$ ,  $i = 1, \dots, s$ , и каждый вектор  $k_i$  имеет размерность  $n$ ) должны определяться одновременно, что еще более увеличивает трудность задачи. В связи с этим естествен вопрос: а имеют ли вообще решение уравнения (7.7)? (На этом основании редакция журнала возвратила Бутчеру первую версию его статьи (1964а).)

**Теорема 7.2.** Пусть  $f$  — непрерывная и удовлетворяющая условию Липшица с постоянной  $L$  функция в некоторой окрестности начальных условий. Если выполнено неравенство

$$h < \frac{1}{L \max_i \sum_j |a_{ij}|}, \quad (7.8)$$

<sup>1)</sup> В иностранной литературе стали традиционными следующие обозначения: ERK, DIRK, SDIRK, IRK, которые соответствуют приведенным в определении ЯРК, ДНРК, ОДНРК, НРК. — Прим. ред.

то существует единственное решение уравнений (7.7), которое может быть получено итерированием. Если  $f(x, y)$  раз непрерывно дифференцируема, то все  $k_i$  как функции  $h$  тоже принадлежат множеству  $C^p$ .

**Доказательство.** Существование решения (7.7) будем доказывать с помощью итераций:

$$k_i^{(n+1)} = f \left( x_0 + c_i h, y_0 + h \sum_{j=1}^s a_{ij} k_j^{(n)} \right).$$

(«...оп la résoudra facilement par des approximations successives...» (Cauchy (1824)<sup>1</sup>). Определим  $K \in \mathbb{R}^{sn}$  как  $K = (k_1, \dots, k_s)_T$  и введем норму  $\|K\| = \max_i (\|k_i\|)$ . Тогда (7.7) можно записать в виде  $K = F(K)$ , где

$$F_i(K) = f \left( x_0 + c_i h, y_0 + h \sum_{j=1}^s a_{ij} k_j \right), \quad i = 1, \dots, s.$$

С помощью условия Липшица, повторно используя неравенство треугольника, легко показать, что

$$\|F(K_1) - F(K_2)\| \leq h L \max_i \sum_j |a_{ij}| \cdot \|K_1 - K_2\|.$$

Отсюда в силу (7.8) следует, что  $F(K)$  — сжимающее отображение. Принцип сжимающих отображений обеспечивает тогда существование и единственность решения, а также сходимость последовательных приближений к неподвижной точке.

Доказательство дифференцируемости опирается на теорему о неявных функциях из классического анализа. Запишем (7.7) в виде

$$\Phi(h, K) = K - F(K) = 0.$$

Матрица частных производных  $\partial\Phi/\partial K$  при  $h = 0$  является единичной матрицей. Поэтому решение уравнения  $\Phi(h, K) = 0$ , которое при  $h = 0$  переходит в  $k_i = f(x_0, y_0)$ , непрерывно дифференцируемо в окрестности точки  $h = 0$ . ■

Для жестких дифференциальных уравнений (у которых  $L$  очень велико) неравенство (7.8) дает весьма сильное ограничение длины шага, неприемлемое на практике. Мы вернемся к этому вопросу во втором томе.

Определение порядка остается таким же, как и для явных методов, и условия, определяющие порядок метода («условия порядка»), выводятся точно тем же путем, что и в разд. II.2.

<sup>1</sup>) «... легко разрешить с помощью последовательных приближений ...» (Коши (1824)). — Франц.

**Пример 7.3.** Рассмотрим неявные двухстадийные методы порядка 3. Условия порядка принимают следующий вид (см. теорему 2.1):

$$b_1 + b_2 = 1,$$

$$b_1 c_1 + b_2 c_2 = \frac{1}{2},$$

$$b_1 c_1^2 + b_2 c_2^2 = \frac{1}{3},$$

$$b_1 (a_{11}c_1 + a_{12}c_2) + b_2 (a_{21}c_1 + a_{22}c_2) = \frac{1}{6}. \quad (7.9)$$

Из первых трех уравнений вытекает следующее условие ортогональности (оно встречается в теории квадратурных формул Гаусса):

$$\int_0^1 (x - c_1)(x - c_2) dx = 0, \text{ откуда } c_2 = \frac{2 - 3c_1}{3 - 6c_1}, \quad (7.10)$$

а также

$$b_1 = \frac{c_2 - 1/2}{c_2 - c_1}, \quad b_2 = \frac{c_1 - 1/2}{c_1 - c_2} \quad (c_1 \neq 1/2, c_1 \neq c_2).$$

В четвертое уравнение подставим  $a_{21} = c_2 - a_{22}$ ,  $a_{11} = c_1 - a_{12}$  и будем рассматривать  $a_{12}$  и  $c_1$  как свободные параметры. Это дает

$$a_{22} = \frac{1/6 - b_1 (a_{12} (c_2 - c_1) + c_1^2)}{b_2 (c_2 - c_1)}. \quad (7.11)$$

Положив  $a_{12} = 0$ , мы получим однопараметрическое семейство методов ДНРК порядка 3. Метод ОДНРК получится, если потребовать еще  $a_{11} = a_{22}$  [Нёрсетт (1974), см. табл. 7.2].

Таблица 7.2. ОДНРК-метод порядка 3

$\gamma$	$\gamma$	0	$\gamma = \frac{3 \pm \sqrt{3}}{6}$
$1 - \gamma$	$1 - 2\gamma$	$\gamma$	
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	

Коэффициенты методов четвертого порядка должны удовлетворять еще четырем условиям, хотя осталось только два свободных параметра. Тем не менее в этом случае существует единственное решение, приведенное в табл. 7.3.

Таблица 7.3. Метод Хаммера—Холлингсурта порядка 4

$\frac{1}{2} - \frac{\sqrt{3}}{6}$	$\frac{1}{4}$	$\frac{1}{4} - \frac{\sqrt{3}}{6}$
$\frac{1}{2} + \frac{\sqrt{3}}{6}$	$\frac{1}{4} + \frac{\sqrt{3}}{6}$	$\frac{1}{4}$
	$\frac{1}{2}$	$\frac{1}{2}$

### Методы Кунцмана и Бутчера порядка 2s

Ясно, что формула (7.4) и метод, приведенный в табл. 7.3, являются обобщениями соответственно одноточечной и двух-

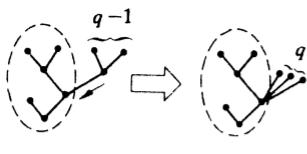


Рис. 7.1. Редукция с  $C(q)$ .

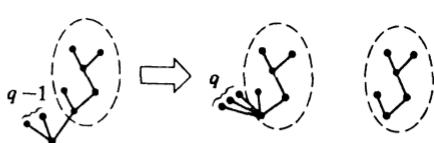


Рис. 7.2. Редукция с  $D(q)$ .

точечной квадратурных формул Гаусса. Кунцман (1961) [см. также Ческино и Кунцман (1963)] и Бутчер (1964а) открыли, что при всех  $s$  существуют НРК-методы порядка 2s. Главным инструментом доказательства этого являются следующие упрощающие предположения:

$$C(\eta): \sum_{l=1}^s a_{il} c_l^{q-1} = \frac{c_i^q}{q}, \quad i = 1, \dots, s, \quad q = 1, \dots, \eta, \quad (7.12)$$

(эти предположения аналогичны предположениям (6.7) и позволяют производить редукцию условий порядка, соответствующих тем или иным деревьям, как схематически показано на рис. 7.1),

а также

$$D(\xi): \sum_{i=1}^s b_i c_i^{q-1} a_{ij} = \frac{b_j}{q} (1 - c_j^q), \quad j = 1, \dots, s, \quad q = 1, \dots, \xi, \quad (7.13)$$

(ср. с предположениями (6.6) и леммой 1.3); эти предположения делают возможной окончательную редукцию всех деревьев, как показано на рис. 7.2 (попробуйте сделать это на каком-нибудь примере).

В результате, если  $\eta$  и  $\xi$  достаточно велики, все деревья приводятся к деревьям, соответствующим квадратурным условиям порядка (ср. (6.9))

$$B(p): \sum_{i=1}^s b_i c_i^{q-1} = \frac{1}{q}, \quad q = 1, \dots, p. \quad (7.14)$$

Итак, установлена

**Теорема 7.4** (Бутчер) *Если удовлетворяются равенства (7.12), (7.13) и (7.14), причем  $p \leq \xi + \eta + 1$ ,  $p \leq 2\eta + 2$ , то метод имеет порядок  $p$ . ■*

Для получения формул порядка  $2s$  Бутчер принял условия (7.14) с  $p = 2s$  (т. е. что  $c_i$  и  $b_i$  являются узлами и весами квадратурной формулы Гаусса) и (7.12) с  $\eta = s$ . Как можно показать, тогда (7.13) тоже выполняется с  $\xi = s$ . Тогда применима теорема 7.4 с  $p = 2s$ . (Ниже будет дано простое объяснение полученного для порядка результата с помощью идеи коллокации.) Методы, соответствующие  $s = 3$  и  $4$ , приведены в табл. 7.4 и 7.5. При  $s = 5$  коэффициенты метода могут быть еще выражены в радикалах. Они приведены в работе Бутчера (1964а). Первое сообщение о впечатляющих численных результатах, даваемых этими методами, появилось в диссертации Зоммера (1965) из области небесной механики.

Таблица 7.4. Метод Кунцмана—Бутчера порядка 6

$\frac{1}{2} - \frac{\sqrt{15}}{10}$	$\frac{5}{36}$	$\frac{2}{9} - \frac{\sqrt{15}}{15}$	$\frac{5}{36} - \frac{\sqrt{15}}{30}$
$\frac{1}{2}$	$\frac{5}{36} + \frac{\sqrt{15}}{24}$	$\frac{2}{9}$	$\frac{5}{36} - \frac{\sqrt{15}}{24}$
$\frac{1}{2} + \frac{\sqrt{15}}{10}$	$\frac{5}{36} + \frac{\sqrt{15}}{30}$	$\frac{2}{9} + \frac{\sqrt{15}}{15}$	$\frac{5}{36}$
	$\frac{5}{18}$	$\frac{4}{9}$	$\frac{5}{18}$

Таблица 7.5. Метод Кунцмана—Бутчера порядка 8

$\frac{1}{2} - \omega_8$	$\omega_1$	$\omega'_1 - \omega_3 + \omega'_4$	$\omega'_1 - \omega_3 - \omega'_4$	$\omega_1 - \omega_5$
$\frac{1}{2} - \omega'_2$	$\omega_1 - \omega'_3 + \omega_4$	$\omega'_1$	$\omega'_1 - \omega'_5$	$\omega_1 - \omega'_3 - \omega_4$
$\frac{1}{2} + \omega'_2$	$\omega_1 + \omega'_3 + \omega_4$	$\omega'_1 + \omega'_5$	$\omega'_1$	$\omega_1 + \omega'_3 - \omega_4$
$\frac{1}{2} + \omega_2$	$\omega_1 + \omega_5$	$\omega'_1 + \omega_3 + \omega'_4$	$\omega'_1 + \omega_3 - \omega'_4$	$\omega_1$
	$2\omega_1$	$2\omega'_1$	$2\omega'_1$	$2\omega_1$
$\omega_1 = \frac{1}{8} - \frac{\sqrt{30}}{144},$			$\omega'_1 = \frac{1}{8} + \frac{\sqrt{30}}{144},$	
$\omega_2 = \frac{1}{2} \left( \frac{15 + 2\sqrt{30}}{35} \right)^{1/2},$			$\omega'_2 = \frac{1}{2} \left( \frac{15 - 2\sqrt{30}}{35} \right)^{1/2},$	
$\omega_3 = \omega_2 \left( \frac{1}{6} + \frac{\sqrt{30}}{24} \right),$			$\omega'_3 = \omega'_2 \left( \frac{1}{6} - \frac{\sqrt{30}}{24} \right),$	
$\omega_4 = \omega_2 \left( \frac{1}{21} + \frac{5\sqrt{30}}{168} \right),$			$\omega'_4 = \omega'_2 \left( \frac{1}{21} - \frac{5\sqrt{30}}{168} \right),$	
$\omega_5 = \omega_2 - 2\omega_3,$			$\omega'_5 = \omega'_2 - 2\omega'_3.$	

### НРК-методы, основанные на квадратурной формуле Лобатто

Квадратурные правила Лобатто [Лобатто (1852), Радо (1880)] являются модификацией идеи квадратур Гаусса: они требуют, чтобы первый и последний узлы совпадали с концами интервала интегрирования, т. е.  $c_1 = 0$ ,  $c_s = 1$ . С этими узлами удобнее обращатьсяся, а в пошаговой процедуре их можно использовать дважды. Остальные  $c_i$  выбираются затем так, чтобы оптимизировать квадратурную формулу. Для интервала интегрирования  $[-1, 1]$  они являются нулями ортогональных многочленов Якоби  $P_{s-2}^{(1,1)}(x)$  или производных многочленов Лежандра  $P'_{s-1}(x)$  (см., напр., Абрамович и Стиган (1964), 25.4.32). В результате получаются квадратурные формулы порядка  $2s - 2$ . Бутчер (1964а, с) обнаружил, что квадратурные правила Лобатто можно распространить на НРК-методы, у которых первая строка и последний столбец матрицы коэффициентов состоят из нулей. При этом первая и последняя стадии становятся *явными*, число неявных стадий понижается до  $s - 2$ , тогда как порядок метода остается равным  $2s - 2$ . Эти методы характеризуются условиями  $B(2s - 2)$  и  $C(s - 1)$  [см. (7.12), (7.14)]. Коэффициенты методов с  $s = 3$  и  $4$  даны в табл. 7.6.

Таблица 7.6. Полученные Бутчером формулы Лобатто порядков 4 и 6

	0	0	0	0	0
0	0 0 0	$\frac{5 - \sqrt{5}}{10}$	$\frac{5 + \sqrt{5}}{60}$	$\frac{1}{6}$	$\frac{15 - 7\sqrt{5}}{60}$
$\frac{1}{2}$	$\frac{1}{4} \frac{1}{4} 0$	$\frac{5 + \sqrt{5}}{10}$	$\frac{5 - \sqrt{5}}{60}$	$\frac{15 + 7\sqrt{5}}{60}$	$\frac{1}{6}$
1	0 1 0				
	$\frac{1}{6} \frac{2}{3} \frac{1}{6}$	1	$\frac{1}{6}$	$\frac{5 - \sqrt{5}}{12}$	$\frac{5 + \sqrt{5}}{12}$
			$\frac{1}{12}$	$\frac{5}{12}$	$\frac{5}{12}$
				$\frac{1}{12}$	

Во втором томе будет показано, что эти методы, хотя и имеют преимущества с точки зрения соотношения между порядком и числом неявных стадий, но для жестких дифференциальных уравнений недостаточно устойчивы.

### НРК как коллокационные методы

«Es ist erstaunlich, dass die Methode trotz ihrer Primitivität und der geringen Rechenarbeit in vielen Fällen... sogar gute Ergebnisse liefert».

«Удивительно, что этот метод, несмотря на свою примитивность и малые вычислительные затраты, во многих случаях... дает даже хорошие результаты».

(Коллатц, 1951)

После 1970 года появилась серия работ [Райт (1970), Хулм (1972), де Боор и Шварц (1973), Нёрсерт и Ваннер (1979)], которые мало-помалу выяснили, что многие методы НРК некоторым образом эквивалентны коллокационным методам и что свойства их порядков совершенно естественны. Позже, в главе о жестких уравнениях, мы познакомимся с обобщением этой идеи (так называемая «возмущенная коллокация»), которое охватывает *все* методы Рунге—Кутты.

Понятие коллокации является старым и общеупотребительным в численном анализе [см., например, многие страницы у Коллатца (1960), или отчет Фрейзера, Джонса и Скана (1937)]. Для обыкновенных дифференциальных уравнений оно состоит в отыскании многочлена степени  $s$ , у которого в  $s$  заданных точках производные совпадают («коллоцируют») с векторным полем дифференциального уравнения.

**Определение 7.5.** Пусть  $s$  — положительное целое число и  $c_1, \dots, c_s$  — вещественные числа (как правило, выбираемые на

интервале  $[0, 1]$ ). Определим коллокационный многочлен  $u(x)$  степени  $s$  следующим образом:

$$u(x_0) = y_0 \text{ (начальное условие),} \quad (7.18a)$$

$$u'(x_0 + c_i h) = f(x_0 + c_i h, u(x_0 + c_i h)), i = 1, \dots, s. \quad (7.18b)$$

Численное решение дифференциального уравнения задаётся при этом выражением

$$y_1 = u(x_0 + h). \quad (7.18c)$$

Если некоторые из узлов  $c_i$  совпадают, условие коллокации (7.18b) будет содержать старшие производные. Тогда оно приводит к методам, использующим старшие производные (см. разд. II.12). Поэтому мы будем пока считать все узлы различными.

**Теорема 7.6.** Коллокационный метод (7.18) эквивалентен  $s$ -стадийному методу НРК, у которого

$$a_{ij} = \int_0^{c_j} l_j(t) dt, \quad b_j = \int_0^1 l_j(t) dt \quad (i, j = 1, \dots, s), \quad (7.19)$$

где  $l_j(s)$  являются многочленами Лагранжа

$$l_j(t) = \prod_{k \neq j} (t - c_k) / \prod_{k \neq j} (c_j - c_k). \quad (7.20)$$

**Доказательство.** Обозначим  $u'(x_0 + c_i h) = k_i$ , тогда

$$u'(x_0 + th) = \sum_{j=1}^s k_j \cdot l_j(t) \quad (\text{Лагранж}).$$

Теперь проинтегрируем это выражение:

$$u(x_0 + c_i h) = y_0 + b \int_0^{c_i} u'(x_0 + th) dt$$

и подставим в (7.18b), учитывая также (7.19). При этом получается НРК-метод (7.7). ■

**Теорема 7.7.** НРК-метод, у которого все  $c_i$  различны и порядок большие или равен  $s$ , является коллокационным методом тогда и только тогда, если предположения (7.12) выполнены для  $\eta = s$ .

**Доказательство.** Условия (7.12) однозначно определяют коэффициенты  $a_{ij}$ . Запишем их в виде равенства

$$\sum_{j=1}^s a_{ij} p(c_j) = \int_0^{c_i} p(s) ds, \quad (7.21)$$

справедливого для всех многочленов  $p$  степени, меньшей чем  $s$ . Даваемые формулой (7.19)  $a_{ij}$  удовлетворяют этому соотношению: подставив (7.19) в (7.21), мы получим интерполяционную формулу Лагранжа. ■

**Теорема 7.8.** Пусть  $M(x) = \prod_{i=1}^s (x - c_i)$  и  $M$  ортогональна многочленам степени  $r = 1$ , т. е.

$$\int_0^1 M(s) s^{q-1} ds = 0, \quad q = 1, \dots, r. \quad (7.22)$$

Тогда метод (7.18) имеет порядок  $p = s + r$ .

*Доказательство.* Предлагаемое доказательство использует формулу Грёбнера—Алексеева, которая позволяет хорошо понять подоплеку результата. Другое доказательство указано ниже, в упр. 7. Соотношение (7.22) означает, что квадратурная формула

$$\int_{x_0}^{x_0+h} g(s) ds = h \sum_{j=1}^s b_j g(x_0 + c_j h) + err(g) \quad (7.23)$$

имеет порядок  $s + r = p$ , и для ее погрешности справедлива оценка

$$|err(g)| \leq C h^{p+1} \cdot \max |g^{(p)}(x)|. \quad (7.24)$$

Далее основная идея доказательства состоит в следующем. Будем рассматривать равенство

$$u'(x) = f(x, u(x)) + (u'(x) - f(x, u(x)))$$

как возмущенное уравнение

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

и применим квадратурную формулу (7.23) для приближенного интегрирования формулы Грёбнера—Алексеева (I.14.18). Согласно (7.18b), результат тождественно равен нулю, так как в точках коллокаций возмущение обращается в нуль. Таким образом, из (7.23) и (7.24) следует оценка

$$\|y(x) - u(x)\| = \|err(g)\| \leq C \cdot h^{p+1} \cdot \max \|g^{(p)}(s)\|, \quad (7.25)$$

где

$$g(s) = D_3 y(x, s, u(s)) \cdot (u(s) - f(s, u(s)))$$

и  $D_3$  — символ частной производной по третьему аргументу  $y$ , т. е. по начальному условию  $u(s)$ . Итак, мы видим, что локальная погрешность ведет себя как  $O(h^{p+1})$ .

Остается, однако, маленькая техническая деталь: показать, что производные  $g(x)$  остаются ограниченными при  $h \rightarrow 0$ . Эти

производные выражаются через частные производные  $f(x, y)$  и производные  $u(x)$ . Эти последние остаются ограниченными при  $h \rightarrow 0$ , как показывает следующая теорема. ■

**Теорема 7.9.** [Хулм (1972)] Коллокационный многочлен  $u(x)$  порождает непрерывный НРК-метод порядка  $s$ , т. е. для всех  $x$ ,  $x_0 \leq x \leq x_0 + h$  имеем

$$\|y(x) - u(x)\| \leq Ch^{s+1}. \quad (7.26)$$

Более того, и для производных  $u(x)$  справедливы оценки

$$\|y^{(k)}(x) - u^{(k)}(x)\| \leq C \cdot h^{s+1-k}, \quad k = 0, \dots, s. \quad (7.27)$$

**Доказательство.** Точное решение  $y(x)$  удовлетворяет условиям коллокации всюду, следовательно, также и в точках  $x_0 + c_i h$ . Поэтому, точно так же, как это делалось при доказательстве теоремы 7.6, применим интерполяционную формулу Лагранжа к  $y'(x)$ :

$$y'(x) = \sum_{j=1}^s f(x_0 + c_j h, y(x_0 + c_j h)) l_j(t) + q(x), \quad x = x_0 + th,$$

где погрешность интерполяции  $q(x) = O(h^s)$ , если  $y(x)$  достаточно гладкая. Учитывая, что  $x = x_0 + th$ , и интегрируя эту формулу, получим

$$y(x) = y_0 + \sum_{j=1}^s f(x_0 + c_j h, y(x_0 + c_j h)) h \int_0^t l_j(s) ds + R(x),$$

где  $R(x) = O(h^{s+1})$ . Вычтем отсюда соответствующее выражение для  $u(x)$ :

$$\begin{aligned} y(x) - u(x) &= h \sum_{j=1}^s (f(x_0 + c_j h, y(x_0 + c_j h)) - \\ &\quad - f(x_0 + c_j h, u(x_0 + c_j h))) \int_0^t l_j(s) ds + R(x). \end{aligned} \quad (7.28)$$

Далее наша задача сводится к очевидным оценкам. Положим

$$B_0 = \max_{0 \leq t \leq 1} \sum_I \left| \int_0^t l_j(s) ds \right|,$$

$$B_h = \max_{0 \leq t \leq 1} \sum_I |l_I^{(k-1)}(t)| \quad \text{для } k = 1, 2, \dots$$

и воспользуемся условием Липшица для  $f$ . Для величины

$$M = \max_{x_0 \leq x \leq x_0 + h} \|y(x) - u(x)\|$$

получается неравенство

$$M \leq hLB_0M + \max \|R(x)\|.$$

Из него следует оценка (7.26). Дифференцируя (7.28)  $k$  раз и делая оценки с учетом того, что  $R^{(k)}(x) = O(h^{s+1-k})$ , получим

$$\max \|y^{(k)}(x) - u^{(k)}(x)\| \leq h^{1-k}LB_kM + O(h^{s+1-k}),$$

откуда следует (7.27), потому что  $M = O(h^{s+1})$ . ■

### Упражнения

- Постройте одноточечный коллокационный метод ( $s = 1$ ) с  $c_t = \theta$  и сравните его с формулой (7.2). Определите его порядок в зависимости от  $\theta$ .
- Рассчитайте все коллокационные методы второго порядка с  $s = 2$  при произвольных  $c_1$  и  $c_2$ .
- Рассмотрите методы из упр. 2, соответствующие выбору  $c_1 = 1/3$ ,  $c_2 = 1$ , а также  $c_1 = 0$ ,  $c_2 = 2/3$ . Определите порядки полученных методов и объясните результат.
- Интерпретируйте неявное правило средней точки (7.4) и явный метод Эйлера как коллокационные методы. Является ли метод (7.5) коллокационным? А метод (7.6)?
- Определите порядки методов из табл. 7.7.

Таблица 7.7. Методы Эле

Метод Эле II <sub>A</sub> , порядок 5				Метод Эле III <sub>A</sub> , порядок 4			
$\frac{4 - \sqrt{6}}{10}$	$\frac{88 - 7\sqrt{6}}{360}$	$\frac{296 - 169\sqrt{6}}{1800}$	$\frac{-2 + 3\sqrt{6}}{225}$	0	0	0	0
$\frac{4 + \sqrt{6}}{10}$	$\frac{296 + 169\sqrt{6}}{1800}$	$\frac{88 + 7\sqrt{6}}{360}$	$\frac{-2 - 3\sqrt{6}}{225}$	$\frac{1}{2}$	$\frac{5}{24}$	$\frac{1}{3}$	$-\frac{1}{24}$
1	$\frac{16 - \sqrt{6}}{36}$	$\frac{16 + \sqrt{6}}{36}$	$\frac{1}{9}$	1	$\frac{1}{6}$	$\frac{2}{3}$	$\frac{1}{6}$
	$\frac{16 - \sqrt{6}}{36}$	$\frac{16 + \sqrt{6}}{36}$	$\frac{1}{9}$		$\frac{1}{6}$	$\frac{2}{3}$	$\frac{1}{6}$

*Указание.* Воспользуйтесь теоремами 7.7 и 7.8.  
Обобщите на произвольные  $s$  [Эле (1968)].

- [Коши (1824)]. Из уравнения (7.2) найдите условия, которым следует подчинить функцию  $f(x, y)$ , чтобы для скалярного дифференциального уравнения выполнялись неравенства

$y_1(\text{ЯЭ}) \geq y(x_1) \geq y_1(\text{НЭ})$ , где  $y_1(\text{ЯЭ})$  ( $y_1(\text{НЭ})$ ) означает численное решение, полученное на одном шаге с помощью явного (неявного) метода Эйлера. Для задачи

$$y' = \cos \frac{x+y}{5}, \quad y(0) = 0,$$

рассчитайте обоими методами пять шагов с  $h = 0.2$ , чтобы оценить верхнюю и нижнюю границы решения  $y(1)$ .

Результат Коши:  $0.9659 \leq y(1) \leq 0.9810$ . Для одного шага с  $h = 1$  он получил  $0.926 \leq y(1) \leq 1$ .

Вычислите точное решение элементарным интегрированием уравнения.

7. [Бутчер (1964a)] Дайте алгебраическое доказательство теоремы 7.8.

Указание. Из теоремы 7.7 следует (7.12) для  $\eta = s$ . Далее, условия (7.14) с  $p = s + r$  (теория квадратурных формул Гаусса) имеют следствием (7.13) для  $\xi = r$ . Чтобы убедиться в этом, умножьте два вектора  $u_l = \sum_i b_i c_i^{q-1} a_{il}$  и  $v_l = b_l (1 - c_l^q)/q$  ( $j = 1, \dots, s$ ) на матрицу Вандермонда

$$V = \begin{pmatrix} 1 & 1 & \dots & 1 \\ c_1 & c_2 & \dots & c_s \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ c_1^{s-1} & c_2^{s-1} & \dots & c_s^{s-1} \end{pmatrix}.$$

Наконец, примените теорему 7.4.

## II.8. Асимптотическое разложение глобальной погрешности

«Mein Verzicht auf das Restglied war leichtsinnig...»

«Мой отказ от остаточного члена был легкомысленным ...»

(В. Ромберг, 1979)

Нашей следующей целью будет усовершенствование ричардсонского метода экстраполяции (см. разд. II.4) посредством выполнения *повторных* экстраполяций и исключения все большего числа членов  $Ch^{p+k}$  погрешности. Хорошую теоретическую основу для этой процедуры дает изучение асимптотического поведения глобальной погрешности. Для задач типа  $y' = f(x)$ , сводящихся к интегрированию функций, решение этой проблемы дает формула Эйлера—Маклорена, и дальнейшее развитие связано с работами Ромберга (1955) и его последователей. Первые строгие результаты для дифференциальных уравнений связаны с именами Хенрихи (1962) и Грэгга (1964) [см. также Штеттер (1973)]. Последовательное исключение членов погрешности мы изложим здесь, следуя работе Хайрера и Любиха (1984), которая допускает обобщение и на многошаговые методы.

### Локальная погрешность

Пусть имеется одношаговый метод, который мы запишем в виде

$$y_{n+1} = y_n + h\Phi(x_n, y_n, h). \quad (8.1)$$

Если метод имеет порядок  $p$ , то в каждой точке решения он обладает локальной погрешностью вида

$$\begin{aligned} y(x+h) - y(x) - h\Phi(x, y(x), h) &= \\ &= d_{p+1}(x)h^{p+1} + \dots + d_{N+1}(x)h^{N+1} + O(h^{N+2}), \end{aligned} \quad (8.2)$$

если правая часть дифференциального уравнения дифференцируема достаточное число раз. Для методов Рунге—Кутты первые члены погрешности были вычислены в разд. II.3 (см. теорему 3.2).

### Глобальная погрешность

Численное решение в точке  $x = x_0 + nh$  обозначим теперь  $y_n := y_h(x)$ . Из теоремы 3.6 известно, что глобальная погрешность ведет себя как  $h^p$ . Мы будем искать функцию  $e_p(x)$  такую, что

$$y_h(x) - y(x) = e_p(x)h^p + o(h^p). \quad (8.3)$$

Основная идея состоит в том, чтобы рассматривать величину

$$y_h(x) - e_p(x) h^p = : \hat{y}_h(x) \quad (8.4)$$

как численное решение, даваемое другим методом:

$$\hat{y}_{n+1} = \hat{y}_n + h \hat{\Phi}(x_n, \hat{y}_n, h).$$

Сравнивая с (8.1), мы видим, что функция приращения этого нового метода есть

$$\hat{\Phi}(x, \hat{y}, h) = \Phi(x, \hat{y} + e_p(x) h^p, h) - (e_p(x+h) - e_p(x)) h^{p-1}. \quad (8.5)$$

Наша задача состоит в нахождении такой функции  $e_p(x)$  с условием  $e_p(x_0) = 0$ , чтобы метод с функцией приращения  $\hat{\Phi}$  имел порядок  $p+1$ .

Разлагая локальную погрешность одношагового метода  $\hat{\Phi}$  по степеням  $h$ , мы получим

$$\begin{aligned} y(x+h) - y(x) - h \hat{\Phi}(x, y(x), h) &= \\ &= \left( d_{p+1}(x) - \frac{\partial f}{\partial y}(x, y(x)) e_p(x) + e'_p(x) \right) h^{p+1} + O(h^{p+2}). \end{aligned} \quad (8.6)$$

Здесь использовано равенство

$$\frac{\partial \Phi}{\partial y}(x, y, 0) = \frac{\partial f}{\partial y}(x, y). \quad (8.7)$$

Член, содержащий  $h^{p+1}$ , исчезнет, если определить  $e_p(x)$  как решение задачи Коши

$$e'_p(x) = \frac{\partial f}{\partial y}(x, y(x)) e_p(x) - d_{p+1}(x), \quad e_p(x_0) = 0. \quad (8.8)$$

Применяя теорему 3.6 к методу  $\hat{\Phi}$ , имеем теперь

$$y_h(x) - y(x) = e_p(x) h^p + O(h^{p+1}). \quad (8.9)$$

Первый член искомого асимптотического разложения определен.

Повторим теперь эту процедуру, отправляясь от метода с функцией приращения  $\hat{\Phi}$ . Этот метод имеет порядок  $p+1$ , и равенство (8.7) тоже удовлетворяется. В конечном результате на этом пути получается следующая

**Теорема 8.1.** [Грэгг (1964)] Пусть имеется метод с достаточно гладкой функцией приращения  $\Phi$ , удовлетворяющий условию согласованности  $\Phi(x, y, 0) = f(x, y)$  и обладающий разложением (8.2) для локальной погрешности. Тогда глобальная погрешность имеет асимптотическое разложение вида

$$y_h(x) - y(x) = e_p(x) h^p + \dots + e_N(x) h^N + E_h(x) h^{N+1}, \quad (8.10)$$

где  $e_j(x)$  — решения неоднородных дифференциальных уравнений вида (8.8) с начальными условиями  $e_j(x_0) = 0$ , а  $E_h(x)$  ограничена при  $x_0 \leq x \leq x_{\text{end}}$  и  $0 \leq h \leq h_0$ . ■

Свойства дифференцируемости функций  $e_j^0(x)$  зависят от гладкости  $f$  и  $\Phi$  (см. (8.8) и (8.2)). Разложение (8.10) будет служить теоретическим основанием во всех обсуждениях экстраполяционных методов.

### Примеры

1. Для уравнения  $y' = y$ ,  $y(0) = 1$  и метода Эйлера при  $h = 1/n$  и  $x = 1$  с помощью биномиальной теоремы получим:

$$\begin{aligned} y_h(1) &= \left(1 + \frac{1}{n}\right)^n = 1 + 1 + \left(1 - \frac{1}{n}\right) \frac{1}{2!} + \\ &\quad + \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \frac{1}{3!} + \dots \end{aligned}$$

Отсюда, раскрывая скобки, можно получить

$$y_h(1) - y(1) = \sum_{i=1}^{\infty} h^i \sum_{l=1}^{\infty} \frac{S_{i+l}^{(l)}}{(i+l)!} = -1.359h + 1.246h^3 \pm \dots,$$

где  $S_i^{(l)}$  — числа Стирлинга первого рода [1730 г.; см. Абрамович и Стиган (1964), разд. 24.1.3]. Мы получили, очевидно, ряд Тейлора для функции

$$\begin{aligned} (1+h)^{1/h} - e &= \exp\left(1 - \frac{h}{2} + \frac{h^2}{3} \pm \dots\right) - e = \\ &= e\left(-\frac{1}{2}h + \frac{11}{24}h^2 - \frac{7}{16}h^3 \pm \dots\right). \end{aligned}$$

Радиус сходимости этого ряда равен единице.

2. Для дифференциального уравнения  $y' = f(x)$  и правила трапеции (7.5) разложение (8.10) примет вид известной формулы Эйлера—Маклорена (1736):

$$y_h(1) - \int_0^1 f(x) dx = \sum_{k=1}^N \frac{h^{2k}}{(2k)!} B_{2k} (f^{(2k-1)}(1) - f^{(2k-1)}(0)) + Err.$$

При  $N \rightarrow \infty$  ряд обычно расходится из-за быстрого роста чисел Бернуlli при больших  $k$ . Однако формула может быть полезна при небольших значениях  $N$ . Формулы такого рода называют *асимптотическими разложениями* [Пуанкаре (1893)].

### Переменная длина шага

Теорема 8.1 справедлива не только при постоянной длине шага. Для случая переменной длины шага разумно предположить, что существует функция  $\tau(x) > 0$ , такая что длины шагов зависят от параметра  $h$  следующим образом:

$$x_{n+1} - x_n = \tau(x_n) h. \quad (8.11)$$

Тогда разложение локальной погрешности (8.2) примет вид

$$\begin{aligned} y(x + \tau(x) h) - y(x) - h\tau(x) \Phi(x, y(x), \tau(x) h) = \\ = d_{p+1}(x) \tau^{p+1}(x) h^{p+1} + \dots, \end{aligned}$$

а вместо (8.5) получим

$$\begin{aligned} \widehat{\Phi}(x, \widehat{y}, \tau(x) h) = \Phi(x, \widehat{y} + e_p(x) h^p, \tau(x) h) - \\ - \frac{h^p}{h\tau(x)} [e_p(x + \tau(x) h) - e_p(x)]. \end{aligned}$$

Поэтому вместо (8.6) разложение локальной погрешности нового метода будет иметь вид

$$\begin{aligned} y(x + \tau(x) h) - y(x) - h\tau(x) \widehat{\Phi}(x, y(x), \tau(x) h) = \\ = \tau(x) \left[ d_{p+1}(x) \tau^p(x) - \frac{\partial f}{\partial y}(x, y(x)) e_p(x) + e'_p(x) \right] h^{p+1} + O(h^{p+2}), \end{aligned}$$

и доказательство обобщенной теоремы 8.1 требует лишь небольших изменений.

### Отрицательные значения $h$

Наиболее важные экстраполяционные алгоритмы используют асимптотические разложения с четными степенями  $h$ . Чтобы дать теоретическое обоснование для этих методов, надо объяснить смысл  $y_h(x)$  при  $h$  отрицательных.

**Обоснование.** Запишем (8.1) в виде

$$y_h(x + h) = y_h(x) + h\Phi(x, y_h(x), h) \quad (8.1')$$

и заменим  $h$  на  $-h$ :

$$y_{-h}(x - h) = y_{-h}(x) - h\Phi(x, y_{-h}(x), -h).$$

Теперь заменим здесь  $x$  на  $x + h$ :

$$y_{-h}(x) = y_{-h}(x + h) - h\Phi(x + h, y_{-h}(x + h), -h). \quad (8.12)$$

Для определения  $y_{-h}(x + h)$  мы получили уравнение, которое при достаточно малых  $h$  имеет единственное решение (по теореме о неявных функциях). Запишем это решение в виде

$$y_{-h}(x + h) = y_{-h}(x) + h\Phi^*(x, y_{-h}(x), h). \quad (8.13)$$

Сравнивая (8.12) и (8.13) и обозначая  $A = y_{-h}(x + h)$ ,  $B = y_{-h}(x)$ , мы приходим к следующему определению.

**Определение.** Пусть  $\Phi(x, y, h)$  — функция приращения некоторого метода. Определим тогда функцию приращения  $\Phi^*(x, y, h)$  присоединенного метода следующей парой формул:

$$\begin{aligned} B &= A - h\Phi(x + h, A, -h), \\ A &= B + h\Phi^*(x, B, h). \end{aligned} \quad (8.14)$$

**Пример:** Для явного метода Эйлера присоединенным методом является неявный метод Эйлера.

**Теорема 8.2.** Пусть  $\Phi$  — метод Рунге—Кутты (7.7) с коэффициентами  $a_{ij}$ ,  $b_j$ ,  $c_i$  ( $i, j = 1, \dots, s$ ). Тогда присоединенный метод  $\Phi^*$  эквивалентен  $s$ -стадийному методу Рунге—Кутты с коэффициентами

$$\begin{aligned} c_i^* &= 1 - c_{s+1-i}, \\ a_{ij}^* &= b_{s+1-j} - a_{s+1-i, s+1-j}, \\ b_i^* &= b_{s+1-i}. \end{aligned}$$

**Доказательство.** Формулы (8.14) показывают, что для определения присоединенного метода надо в (7.7) поменять местами  $y_0$  и  $y_1$ ,  $h$  и  $-h$  и затем вместо  $x_0$  подставить  $x_0 + h$ . В результате получается

$$\begin{aligned} k_i &= f(x_0 + (1 - c_i)h, y_0 + h \sum_i (b_j - a_{ij}) k_j), \\ y_1 &= y_0 + h \sum_i b_i k_i. \end{aligned}$$

Чтобы сохранить упорядоченность  $c_1, \dots, c_s$  (обычно естественную), сделаем перестановку значений  $k_i$  в обратном порядке, для чего заменим все индексы  $i$  на  $s + 1 - i$ . ■

### Свойства присоединенного метода

**Теорема 8.3.**  $\Phi^{**} = \Phi$ .

**Доказательство.** Это свойство (давшее повод к названию «присоединенный») доказывается заменой  $h$  на  $-h$  и затем  $x$  на  $x + h$ ,  $B$  на  $A$  и  $A$  на  $B$  в формулах (8.14). ■

**Теорема 8.4.** Присоединенный метод имеет тот же порядок, что и исходный. Его главный член погрешности отличается от главного члена погрешности исходного метода множителем  $(-1)^p$ .

**Доказательство.** Заменим в (8.2)  $h$  на  $-h$ , затем подставим  $x + h$  вместо  $x$  и упорядочим заново члены по степеням  $h$ . При

этом получается

$$\begin{aligned} g(x) + d_{p+1}(x) h^{p+1} (-1)^p + O(h^{p+2}) = \\ = y(x+h) - h\Phi(x+h, y(x+h), -h). \end{aligned}$$

(Здесь учтено, что  $d_{p+1}(x+h) = d_{p+1}(x) + O(h)$ .) Теперь обозначим  $B$  левую часть этого тождества, положим  $A = y(x+h)$  и применим (8.14). Это приводит к равенству

$$y(x+h) = g(x) + d_{p+1}(x) h^{p+1} (-1)^p + h\Phi^*(x, g(x), h) + O(h^{p+2}),$$

выражающему утверждение теоремы. ■

**Теорема 8.5.** *Присоединенный метод имеет такое же асимптотическое разложение глобальной погрешности (8.10), что и исходный метод, но о заменой  $h$  на  $-h$ .*

**Доказательство.** Повторим с отрицательным  $h$  ту процедуру, которая привела к доказательству теоремы 8.1. Первый отделенный член, соответствующий (8.9), будет

$$g_{-h}(x) - g(x) = e_p(x) (-h)^p + O(h^{p+1}). \quad (8.9')$$

Справедливость этого выражения вытекает из того факта, что решение (8.8) при начальном условии  $e_p(x_0) = 0$  изменяет знак вместе с неоднородным членом  $d_{p+1}(x)$ . Итак, первый член разложения установлен. Чтобы иметь возможность продолжить процедуру определения дальнейших членов разложения, надо показать, что преобразование (8.4) перестановочно с операцией перехода к присоединенному методу, т. е. что имеет место равенство

$$(\widehat{\Phi})^* = (\Phi^*)^-. \quad (8.15)$$

Для доказательства (8.15) заметим, что из (8.4) и определения  $\widehat{\Phi}$  вытекает равенство

$$\begin{aligned} y_h(x+h) - e_p(x+h) h^p = \\ = y_h(x) - e_p(x) h^p + h\widehat{\Phi}(x, y_h(x) - e_p(x) h^p, h). \end{aligned}$$

Здесь опять подставим  $-h$  вместо  $h$ , а затем заменим  $x$  на  $x+h$ . Наконец, положив  $B = y_{-h}(x) - e_p(x) (-h)^p$  и  $A = y_{-h}(x+h) - e_p(x+h) (-h)^p$ , применим (8.14) и получим

$$\begin{aligned} y_{-h}(x+h) - e_p(x+h) (-h)^p = \\ = y_{-h}(x) - e_p(x) (-h)^p + h\widehat{\Phi}^*(x, y_{-h}(x) - e_p(x) (-h)^p, h). \quad (8.16) \end{aligned}$$

С другой стороны, если выполнить преобразование

$$\widehat{y}_{-h}(x) = y_{-h}(x) - e_p(x) (-h)^p \quad (8.4')$$

(см. теорему 8.4) и подставить в (8.13), мы получим опять равенство (8.16), но на этот раз с  $(\Phi^*)^\wedge$  вместо  $\widehat{\Phi}^*$ , что и доказывает формулу (8.15). ■

### Симметричные методы

**Определение 8.6.** Метод называют *симметричным*, если  $\Phi = \Phi^*$ .

**Пример.** Правило трапеции (7.5) и неявное правило средней точки (7.4) симметричны: перестановки  $y_1 \leftrightarrow y_0$ ,  $h \leftrightarrow -h$  и  $x_0 \leftrightarrow x_0 + h$  оставляют эти методы инвариантными.

Следующие две теоремы принадлежат Ваннеру (1973).

**Теорема 8.7.** *Если*

$$a_{s-i+1, s-j+1} + a_{ij} = b_{s-j+1} = b_i, \quad (8.17)$$

*то соответствующий метод РК симметричен.*

**Доказательство:** см. теорему 8.2. Условие  $c_i = 1 - c_{s+1-i}$  проверяется суммированием равенств (8.17) по  $j$  от 1 до  $s$ . ■

**Теорема 8.8.** *Коллокационный метод с симметрично распределенными точками коллокаций симметричен.*

**Доказательство.** Если  $c_i = 1 - c_{s+1-i}$ , то многочлены Лагранжа обладают свойством  $l_i(t) = l_{s+1-i}(1-t)$ . Условия (8.17) тогда легко получаются из (7.19). ■

Эти результаты позволяют теперь доказать следующее важное свойство симметричных методов, подмеченное уже давно:

**Теорема 8.9.** *Если метод симметричен и выполнены условия теоремы 8.1, то асимптотическое разложение (8.10) содержит только четные степени  $h$ :*

$$y_h(x) - y(x) = e_{2q}(x)h^{2q} + e_{2q+2}(x)h^{2q+2} + \dots \quad (8.18)$$

и  $e_{2j}(x_0) = 0$ .

**Доказательство.** Если  $\Phi^* = \Phi$ , то в силу (8.13)  $y_{-h}(x) = y_h(x)$  и утверждение следует из теоремы 8.5. ■

**Пример.** Самый старый пример такого разложения — формула Эйлера—Маклорена.

### Упражнения

1. Пусть одношаговый метод (8.1) имеет порядок  $p \geq 2$  и кроме условия  $\Phi(x, y, 0) = f(x, y)$  выполняется еще равенство

$$\frac{\partial \Phi}{\partial h}(x, y, 0) = \frac{1}{2} \left( \frac{\partial f}{\partial x}(x, y) + \frac{\partial f}{\partial y}(x, y) \cdot f(x, y) \right). \quad (8.19)$$

Покажите, что тогда главный член локальной погрешности метода  $\tilde{\Phi}$ , определяемого формулой (8.5), будет иметь вид

$$d_{p+2}(x) = d_{p+2}(x) - \frac{1}{2} \frac{\partial f}{\partial y}(x, y(x)) d_{p+1}(x) - \frac{1}{2} d'_{p+1}(x).$$

Проверьте, что (8.19) выполняется для всех методов РК с  $p \geq 2$ .

2. Рассмотрите применение метода второго порядка

$$\begin{array}{c|cc} 0 & & \\ \hline 1 & 1 \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array}$$

к задаче  $y' = y$ ,  $y(0) = 1$ . Покажите, что

$$d_3(x) = \frac{1}{6} e^x, \quad d_4(x) = \frac{1}{24} e^x, \quad e_2(x) = -\frac{1}{6} x e^x, \quad d_4(x) = -\frac{e^x}{8}.$$

3. Рассмотрите метод второго порядка

$$\begin{array}{c|ccc} 0 & & & \\ \hline \frac{1}{2} & & \frac{1}{2} & \\ \hline 1 & 0 & 1 \\ \hline & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{array}$$

Покажите, что для этого метода

$$d_3(x) = \frac{1}{24} \left( F(t_{32})(y(x)) - \frac{1}{2} F(t_{31})(y(x)) \right),$$

$$d_4(x) = \frac{1}{24} \left( F(t_{44})(y(x)) + \frac{1}{4} F(t_{43})(y(x)) - \frac{1}{4} F(t_{41})(y(x)) \right)$$

(обозначения соответствуют табл. 2.1). Покажите, что отсюда вытекают равенства

$$d_4(x) = 0 \text{ и } e_8(x) = 0,$$

так что один шаг экстраполяции по Ричардсону повышает порядок метода на 2. (Установите связь между этим методом и алгоритмом ГБШ из разд. II.9.)

4. Обсудите симметрию других классов методов НРК из разд. II.7. (*Ответ:* они не симметричны.)

## II.9. Экстраполяционные методы

«Следующий метод аппроксимации можно считать или не считать новым, но я считаю, что он важен для практики...»

(С. А. Кори, 1906)

«Открытие  $h^2$ -экстраполяции было навеяно теорией, подсказавшей арифметические эксперименты, которые дали желаемые результаты».

(Л. Ф. Ричардсон, 1927)

«Экстраполяция является мощным средством...» (Р. Булирш — И. Штёр, 1966)

«Экстраполяция, по-видимому, не является особенно эффективным способом ..., наши исследования поставили вопрос, есть ли вообще смысл заниматься ею как отдельным методом».

(Л. Ф. Шампайн и Л. С. Бака, 1984)

### Определение метода

Пусть имеется задача Коши  $y' = f(x, y)$ ,  $y(x_0) = y_0$ . Задавшись базовой длиной шага  $H > 0$ , выберем последовательность положительных целых чисел

$$n_1 < n_2 < n_3 < \dots \quad (9.1)$$

и определим соответствующие длины шагов  $h_1 > h_2 > h_3 > \dots$  как  $h_i = H/n_i$ . Возьмем теперь численный метод порядка  $p$  и, выполняя  $n_i$  шагов длины  $h_i$ , вычислим приближенное решение нашей начальной задачи

$$y_{h_i}(x_0 + H) = : T_{i, 1} : \quad (9.2)$$

(это установившееся обозначение; первоначально буква  $T$  символизировала правило трапеций). Выполнив вычисления для ряда последовательных значений  $i$ , исключим из асимптотического разложения (8.10) возможно большее число членов, вычисляя интерполирующий многочлен

$$p(h) = e_0 + e_p h^p + e_{p+1} h^{p+1} + \dots + e_{p+k-2} h^{p+k-2}, \quad (9.3)$$

для которого

$$p(h_i) = T_{i,1}, \quad i = j, j-1, \dots, j-k+1. \quad (9.4)$$

И в заключение сделаем «экстраполяцию к пределу»  $h \rightarrow 0$ , т. е. будем рассматривать значение

$$p(0) = e_0 =: T_{j,k} \quad (9.5)$$

как численное решение, даваемое методом экстраполяции.

Условия (9.4) дают систему  $k$  линейных уравнений для определения  $k$  неизвестных  $e_0, e_p, \dots, e_{p+k-2}$ .

**Пример.** Для  $k = 2, n_1 = 1, n_2 = 2$  приведенное выше определение совпадает с процедурой, описанной в начале разд. II.4.

**Теорема 9.1.** Точность аппроксимации решения задачи Коши величиной  $T_{j,k}$  соответствует численному методу порядка  $p + k - 1$ .

**Доказательство.** Сравним (9.3), (9.4) с асимптотическим разложением (8.10), которое перепишем в форме

$$\begin{aligned} T_{i,1} &= g(x_0 + H) + e_p(x_0 + H) h_i^p + \dots \\ &\quad \dots + e_{p+k-2}(x_0 + H) h_i^{p+k-2} + \Delta_i, \end{aligned} \quad (9.4')$$

где взято  $N = p + k - 1$  и

$$\Delta_i = e_{p+k-1}(x_0 + H) h_i^{p+k-1} + E_{h_i}(x_0 + H) h_i^{p+k} = O(H^{p+k})$$

(потому что  $e_{p+k-1}(x_0) = 0$  и  $h_i \leq H$ ). Формула (9.4') представляет линейную систему уравнений с неизвестными  $y(x_0 + H), H^p e_p(x_0 + H), \dots, H^{p+k-2} e_{p+k-2}(x_0 + H)$  и с матрицей

$$A = \left( \begin{array}{cccc} 1 & \frac{1}{n_i^p} & \cdots & \frac{1}{n_i^{p+k-2}} \\ \vdots & \vdots & & \vdots \\ 1 & \frac{1}{n_{i-k+1}^p} & \cdots & \frac{1}{n_{i-k+1}^{p+k-2}} \end{array} \right),$$

похожей на матрицу Вандермонда. Она отличается от (9.4) только возмущающим членом  $\Delta_i$  в правой части, имеющим порядок  $O(H^{p+k})$ . Матрица  $A$  обратима (см. упр. 6). Вычтя эти формулы друг из друга, легко получить, в частности, оценку

$$|y(x_0 + H) - e_0| \leq \|A^{-1}\|_\infty \cdot \max |\Delta_i| = O(H^{p+k}). \blacksquare$$

*Замечание.* Случай  $p = 1$  (а также  $p = 2$  для разложений по степеням  $h^2$ ) можно рассмотреть и иначе, интерпретируя разность  $y(x_0 + H) - e_0$  как погрешность интерполяции [см. (9.21)].

Большое достоинство рассматриваемого метода состоит в том, что он дает полную таблицу результатов вычислений

$$\begin{matrix} T_{11} \\ T_{21} & T_{22} \\ T_{31} & T_{32} & T_{33} \\ \dots & \dots & \dots \end{matrix}$$

которые образуют последовательность вложенных методов и позволяют легко оценить локальную погрешность и выбрать стратегии для методов переменного порядка.

Несколько числовых последовательностей используются в качестве (9.1):

«последовательность Ромберга» [Ромберг (1955)]:

$$1, 2, 4, 8, 16, 32, \dots \quad (9.6)$$

«последовательность Булирша» (также упомянутую Ромбергом (1955)):

$$1, 2, 3, 4, 6, 8, 12, 16, 24, 32, \dots \quad (9.7)$$

в которой чередуются степени 2 с  $1.5 \times 2^k$ . Для этой последовательности в высших порядках требуется меньшее число вычислений функции, чем для предыдущей, и она стала известной благодаря успеху «алгоритма Грэгга—Булирша—Штёра» [Булирш—Штёр (1966)].

Приведенные выше последовательности обладают тем свойством, что при интегрировании задачи  $y' = f(x)$  многие значения функции можно запомнить и заново использовать для меньших  $h_i$ . Кроме того,  $\liminf (n_{i+1}/n_i)$  для этих последовательностей удовлетворяет «условию Тёплица», т. е. он строго больше единицы, что позволяет доказывать сходимость при  $j = k \rightarrow \infty$  [Бауэр, Рутисхаузер, Штифель (1963)]. Однако при работе с дифференциальными уравнениями с фиксированным или ограниченным порядком, как показал Дойфлхард (1983), наиболее экономичной является «гармоническая последовательность»

$$1, 2, 3, 4, 5, 6, 7, \dots \quad (9.8)$$

### Алгоритм Эйткена—Невилла

При  $p = 1$  формулы (9.3) и (9.4) переходят в классическую задачу интерполяции, и можно экономно вычислять значения  $T_{j, k}$ , пользуясь классическими методами. Поскольку нам нужны значения интерполяционных многочленов только при  $h = 0$ , наиболее экономичным будет основанный на идеях Жордана (1928) «алгоритм Эйткена—Невилла» [Эйткен (1932), Невилл (1934)], который приводит к формуле

$$T_{j, k+1} = T_{j, k} + \frac{T_{j, k} - T_{j-1, k}}{(n_j/n_{j-k}) - 1}. \quad (9.9)$$

### Рациональная экстраполяция

Многие авторы в 60-х годах считали, что вместо полиномов в (9.3) лучше использовать рациональные функции. В этом случае формулу (9.9) нужно заменить другой:

$$T_{j, k+1} = T_{j, k} + \frac{T_{j, k} - T_{j-1, k}}{\frac{n_j}{n_{j-k}} \left[ 1 - \frac{T_{j, k} - T_{j-1, k}}{T_{j, k} - T_{j-1, k-1}} \right] - 1} \quad (9.10)$$

[Булирш и Штёр (1964)], где

$$T_{j, 0} = 0.$$

В случае систем дифференциальных уравнений деление векторных величин следует понимать покомпонентно.

Позже численные эксперименты [Дойфлхард (1983)] показали, что рациональная экстраполяция почти никогда не является более выгодной, чем полиномиальная.

### Вычислительный пример

Мы решали задачу

$$y' = (-y \sin x + 2 \operatorname{tg} x) y, \quad y\left(\frac{\pi}{6}\right) = \frac{2}{\sqrt{3}}, \quad (9.11)$$

с точным решением  $y(x) = 1/\cos x$ . Использовался метод Эйлера, был взят базовый шаг  $H = 0.2$ . На рис. 9.1 для трех последовательностей чисел шагов (9.6), (9.7) и (9.8) сопоставлены в двойном логарифмическом масштабе объем вычислительной работы  $\lg(1 + n_j - 1 + n_{j-1} - 1 + \dots + n_{j-k+1} - 1)$  и получаемая точность  $-\lg(|T_{j, k} - y(x_0 + H)|)$ .

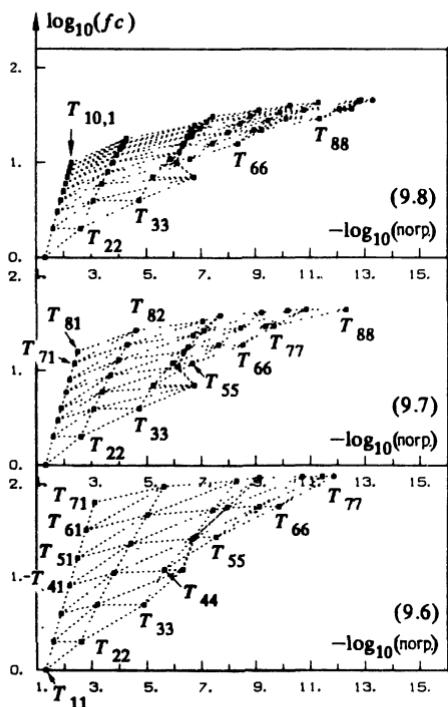


Рис. 9.1. Точность экстраполяции метода Эйлера для примера (9.11). Вверху — гармоническая последовательность, посередине — последовательность Булирша, внизу — последовательность Ромберга.

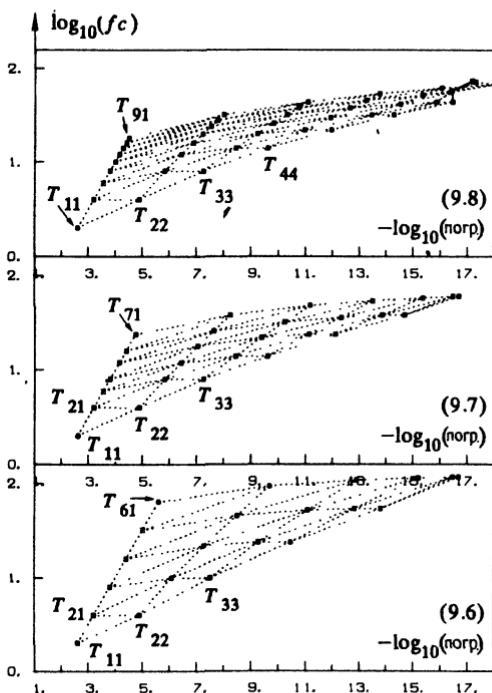
### Экстраполяция с помощью симметричных методов

Если опорный метод является *симметричным*, то, как мы знаем, он имеет асимптотическое разложение по степеням  $h^2$  (теорема 8.9), и каждая экстраполяция исключает две степени  $h$ . Поэтому мы можем просто заменить в (9.3)  $h$  на  $h^2$  и для  $p = 2$  [т. е.  $q = 1$  в формуле (8.18)] тоже воспользоваться алгоритмом Эйткена—Невилла с такой модификацией. Вместо формулы (9.9) мы получим

$$T_{j, h+1} = T_{j, h} + \frac{T_{j, h} - T_{j-1, h}}{(n_j/n_{j-h})^2 - 1}. \quad (9.12)$$

**Вычислительный пример.** На этот раз мы решали задачу (9.11) с помощью неявного правила средней точки (7.4). Хотя этот метод неявный, но благодаря простой структуре дифференциального уравнения его применение требует решения лишь квадратного уравнения. Мы использовали ту же длину базового шага  $H = 0.2$  и те же последовательности чисел шагов (9.6), (9.7) и (9.8), что

Рис. 9.2. Точность экстраполяции неявного правила средней точки. Вверху — гармоническая последовательность, посередине — последовательность Булирша, внизу — последовательность Ромберга.



и выше. Аналогичным образом представлены и погрешности на рис. 9.2.

### Метод Грэгга, или ГБШ

«Будучи полностью явным, алгоритм Грэгга столь идеально приспособлен в качестве опорного для экстраполяции по Ричардсону, что никакой другой симметричный двухшаговый алгоритм не может конкурировать с ним». (Х. И. Штеттер, 1970)

Здесь лучше всего привести цитату из работы Штеттера (1970): «Разложения по степеням  $h^2$  крайне важны для эффективного применения ричардсоновской экстраполяции. Поэтому, когда Грэгг доказал в 1963 г., что величина  $S_h(x)$ , получаемая с помощью алгоритма

$$y_1 = y_0 + hf(x_0, y_0), \quad (9.13a)$$

$$y_{i+1} = y_{i-1} + 2hf(x_i, y_i), \quad i = 1, 2, \dots, 2n, \quad (9.13b)$$

$$S_h(x) = \frac{1}{4} (y_{2n-1} + 2y_{2n} + y_{2n+1}), \quad (9.13c)$$

где  $x = x_0 + 2nh$ ,  $x_i = x_0 + ih$ , обладает асимптотическим разложением по четным степеням  $h$  и имеет удовлетворительные свойства устойчивости, это явилось большим достижением. Это

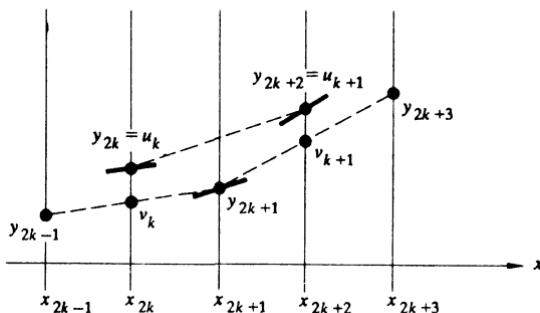


Рис. 9.3. Симметричность метода Грэгга.

привело к построению очень мощного экстраполяционного алгоритма ГБШ (Грэгга—Булирша—Штёра)<sup>1)</sup> ...»

Данное Грэггом доказательство этого свойства было очень длинным и сложным, и опять-таки «большим достижением» явилась элегантная идея Штеттера интерпретировать (9.13b) как одноступенчатый алгоритм, для чего переписать (9.13), введя нечетные и четные индексы. Введем для этого следующие величины:

$$\begin{aligned} h^* &= 2h, \quad x_k^* = x_0 + kh^*, \quad u_0 = v_0 = y_0, \\ u_h &= y_{2h}, \quad v_h = y_{2h+1} - hf(x_{2h}, y_{2h}) = \frac{1}{2}(y_{2h+1} + y_{2h-1}). \end{aligned} \quad (9.14)$$

Тогда метод (9.13) можно переписать в виде (см. рис. 9.3)

$$\begin{pmatrix} u_{h+1} \\ v_{h+1} \end{pmatrix} = \begin{pmatrix} u_h \\ v_h \end{pmatrix} + h^* \left( \begin{pmatrix} f(x_k^* + \frac{h^*}{2}, v_k + \frac{h^*}{2}f(x_k^*, u_k)) \\ \frac{1}{2}[f(x_k^* + h^*, u_{h+1}) + f(x_k^*, u_k)] \end{pmatrix} \right). \quad (9.15)$$

Как видно из рис. 9.3, этот метод, отображающий пару  $(u_h, v_h)$  в  $(u_{h+1}, v_{h+1})$ , является симметричным. Симметрию можно проверить

<sup>1)</sup> В иностранной литературе принята аббревиатура GBS (Gragg — Bulirsch — Stoer). — Прим. ред.

и аналитически (см. определение 8.6), сделав перестановки  $u_{k+1} \leftrightarrow u_k$ ,  $v_{k+1} \leftrightarrow v_k$ ,  $h^* \leftrightarrow -h^*$ ,  $x_k^* \leftrightarrow x_k^* + h^*$ . Простые вычисления показывают, что при этом формула (9.15) не изменяется. Метод (9.15) согласован с дифференциальным уравнением (системой!)

$$\begin{aligned} u' &= f(x, v), \quad u(x_0) = y_0, \\ v' &= f(x, u), \quad v(x_0) = y_0 \end{aligned} \quad (9.16)$$

(это видно, если в функции приращения устремить  $h^*$  к нулю). Точным решением этой системы является просто  $u(x) = v(x) = y(x)$ . Поэтому из теоремы 8.9 следует, что

$$y(x) - u_{h^*}(x) = \sum_{l=1}^t a_l(x) (h^*)^{2l} + (h^*)^{2l+2} A(x, h^*), \quad (9.17a)$$

$$y(x) - v_{h^*}(x) = \sum_{l=1}^t b_l(x) (h^*)^{2l} + (h^*)^{2l+2} B(x, h^*) \quad (9.17b)$$

и  $a_j(x_0) = b_j(x_0) = 0$ . Мы видим из (9.14), что  $y_h(x)$  обладает разложением по четным степеням  $h$  при условии, что число шагов — четное.

### Сглаживающий шаг

«Предупреждаем читателя о возможной неустойчивости, особенно при применении формул 25.5.2...» (Дж.П. Дэвис и И. Полански в «Справочнике» под ред. Абрамовича и Стиган)

«Явное правило средней точки» (9.13b), использованное в рассмотренной выше конструкции, имеет все же плохую репутацию «слабо неустойчивого». Мы обсудим подробно это свойство во введении к гл. IV (том II). Зигзагообразное поведение погрешности подсказало идею заменить  $y_{2n}$  взвешенным средним из трех величин

$$S_h(x_0 + 2nh) = \frac{1}{4} (y_{2n-1} + 2y_{2n} + y_{2n+1}) = \frac{1}{2} (u_n + v_n)$$

(см. (9.13c) и (9.14)). Теперь из (9.17) вытекает следующий результат, полученный Грэггом.

**Теорема 9.2.** Пусть  $f(x, y) \in C^{2t+2}$ . Тогда численное решение, определяемое формулами (9.13), обладает асимптотическим разложением вида

$$y(x) - S_h(x) = \sum_{j=1}^t e_j(x) h^{2j} + h^{2t+2} C(x, h), \quad (9.18)$$

где  $e_j(x_0) = 0$ , а  $C(x, h)$  ограничена при  $x_0 \leq x \leq 0$  и  $0 \leq h \leq h_0$ .

**Доказательство.** Разложение (9.18) получится, если сложить равенства (9.17a) и (9.17b) и подставить  $h^* = 2h$ , причем  $e_j(x) = (a_j(x) + b_j(x)) 2^{2j-1}$ . ■

### Вычислительный алгоритм и пример

Итак, этот метод может быть использован для ричардсоновской экстраполяции тем же способом, что и рассмотренные выше симметричные методы. Последовательности чисел шагов выбираются, однако, с условием, что  $n_j$  должны быть четными, т. е.

$$2, 4, 8, 16, 32, 64, \dots, \quad (9.6')$$

$$2, 4, 6, 8, 12, 16, \dots, \quad (9.7')$$

$$2, 4, 6, 8, 10, 12, \dots, . \quad (9.8')$$

Далее полагаем

$$T_{i,j} := S_{h_i}(x_0 + H)$$

и вычисляем основанные на разложении по степеням  $h^2$  экстраполированные значения  $T_{i,j}$  по формуле Эйткена—Невилла (9.12).

#### Пример вычислений.

Мы применили этот алгоритм опять к задаче (9.11), взяли длину базового шага  $H = 0.2$  и, вычислив *погрешности*, сопоставили их с *объемом вычислительной работы* (измеряемой количеством вычислений значений  $f$ , которое равно теперь  $1 + n_j + n_{j-1} + \dots + n_{j-k+1}$  из-за дополнительного вычисления функции для сглаживающего шага). Результаты показаны на рис. 9.4.

Эти результаты почти такие же, как для неявного правила сред-

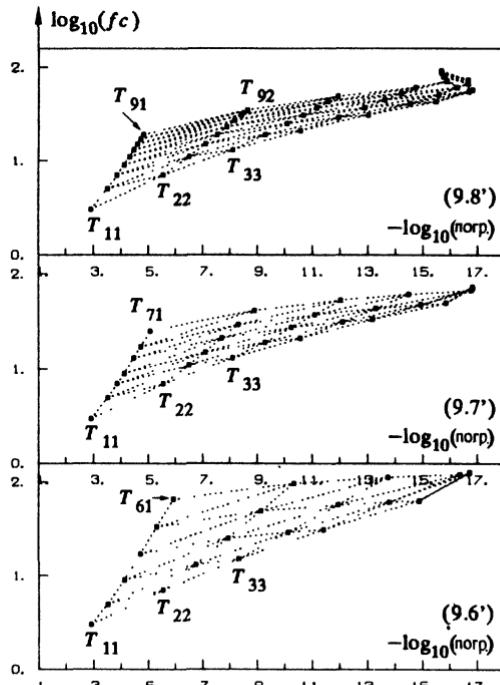


Рис. 9.4. Точность экстраполяции метода Грэгга для примера (9.11).

ней точки, но теперь опорный метод — явный, и это намного ценнее. Для всех трех последовательностей ясно видно превосходство над экстраполяцией, опирающейся на метод Эйлера, особенно в диагональных  $T_{kh}$  для больших значений  $k$ .

### Асимптотическое разложение для нечетных индексов

Для полноты мы хотим еще вывести из (9.17b) существование разложения по степеням  $h^2$  у  $y_{2k+1}$ , хотя это не имеет практического значения для описанного выше вычислительного алгоритма.

**Теорема 9.3.** (Грэгг) Для  $x = x_0 + (2k + 1)h$  имеет место разложение

$$y(x) - y_h(x) = \sum_{j=1}^l \hat{b}_j(x) h^{2j} + h^{2l+2} \hat{B}(x, h), \quad (9.19)$$

где коэффициенты  $\hat{b}_j$ , вообще говоря, отличны от соответствующих коэффициентов для четных индексов и  $\hat{b}_j(x_0) \neq 0$ .

**Доказательство.** Значение  $y_{2k+1}$  можно вычислить (см. рис. 9.3) либо делая шаг вперед от  $v_k$ , либо шаг назад от  $v_{k+1}$ . Для симметрии возьмем среднее из этих двух выражений и запишем

$$y_{2k+1} = \frac{1}{2} (v_k + v_{k+1}) + \frac{h}{2} (f(x_k^*, u_k) - f(x_{k+1}^*, u_{k+1})). \quad (9.20)$$

Вычитая отсюда точное решение, мы получим

$$\begin{aligned} 2(y_h(x) - y(x)) &= v_{2h}(x - h) - y(x - h) + \\ &+ v_{2h}(x + h) - y(x + h) + y(x - h) - 2y(x) + y(x + h) + \\ &+ h [f(x - h, u_{2h}(x - h)) - f(x + h, u_{2h}(x + h))]. \end{aligned}$$

Все полученное выражение симметрично по  $h$  благодаря симметрии  $u_{2h}(x)$  ( $u_{2h}(\xi) = u_{-2h}(\xi)$ ) и  $v_{2h}(x)$ . Поэтому асимптотическое разложение для  $y_{2k+1}$  не содержит нечетных степеней  $h$ . ■

Оба выражения, как для четных индексов, так и для нечетных, можно по-прежнему скомбинировать в одну формулу (см. упр. 2).

### Существование явных методов Рунге—Кутты произвольного порядка

Легко видеть, что каждое из выражений  $T_{j,k}$  представляет некоторый явный метод РК (см. упр. 1). Если применить к (9.18) известную формулу для погрешности полиномиальной интерполяции (см., например, Абрамович и Стиган (1964), формула 25.2.27), мы получим

$$y(x_0 + H) - T_{j,k} = \frac{(-1)^k}{n_j^2 \dots n_{l-k+1}^2} e_h(x_0 + H) H^{2k} + O(H^{2k+2}). \quad (9.21)$$

Так как  $e_k(x_0) = 0$ , то

$$y(x_0 + H) - T_{k,k} = \frac{(-1)^k}{n_1^2 \dots n_{j-k+1}^2} e_k''(x_0) H^{2k+1} + O(H^{2k+2}). \quad (9.22)$$

Это показывает, что  $T_{j,k}$  представляет некоторый метод ЯРК порядка  $2k$ .

В качестве приложения этого результата может быть указана

**Теорема 9.4.** (Грэгг.) Для четных  $p$  существуют явные методы РК порядка  $p$ , имеющие  $s = p^3/4 + 1$  стадию.

**Доказательство.** Этот результат получается путем подсчета числа вычислений функции  $f$ , необходимых в алгоритме ГБШ при использовании гармонической последовательности и без завершающего сглаживающего шага. ■

**Замечание.** Экстраполяция, опирающаяся на метод Эйлера, приводит к ЯРК-методам с  $s = p(p-1)/2 + 1$  стадиями. Это лишний раз показывает важность разложения по степеням  $h^2$ .

### Управление порядком и длиной шага

Экстраполяционные методы имеют то преимущество, что у них на каждом шаге можно менять не только длину шага, но и порядок метода (т. е. число столбцов в таблице  $T_{j,k}$ ). Благодаря этой двойной свободе их оптимальная практическая реализация в виде программы более сложна, чем для методов РК фиксированного порядка. Первые программы были разработаны Булиршем и Штёром (1966) и их студентами. Очень успешные экстраполяционные программы, разработанные П. Дойфлхардом и его сотрудниками, описаны в работе Дойфлхарда (1983).

Выбор длины шага может быть сделан точно таким же способом, как и для вложенных методов фиксированного порядка (см. разд. II.4). Если вычислены первые  $k$  строк экстраполяционной таблицы, то мы имеем  $T_{k,k}$  в качестве аппроксимации наивысшего порядка (согласно (9.22), равного  $2k$ ) и, кроме того,  $T_{k,k-1}$  порядка  $2k-2$ . Поэтому для управления длиной шага естественно использовать выражение

$$err_h = \|T_{k,k-1} - T_{k,k}\|. \quad (9.23)$$

Как в разд. II.4, для оптимальной длины шага получаем формулу

$$H_k = H \cdot 0.94 \cdot (0.65 tol / err_h)^{1/(2k-1)}. \quad (9.24)$$

Заметим, что здесь гарантийный фактор отчасти зависит от порядка.

Для выбора оптимального порядка нам необходима мера вычислительной работы, которая позволила бы сравнивать различ-

ные методы. Работа, затрачиваемая на вычисление  $T_{kk}$ , может измеряться числом  $A_k$  вычислений функции  $f$ . Для алгоритма ГБШ это число определяется рекуррентной формулой

$$\begin{aligned} A_1 &= n_1 + 1, \\ A_k &= A_{k-1} + n_k. \end{aligned} \quad (9.25)$$

Однако большое число вычислений функции может быть компенсировано большой длиной шага  $H_k$ , даваемой формулой (9.24). Поэтому рассмотрим в качестве меры вычислительной работы *работу на единичный шаг*:

$$W_k = \frac{A_k}{H_k}. \quad (9.26)$$

Теперь идея оптимального выбора порядка (т. е. индекса  $k$ ) состоит в том, чтобы минимизировать  $W_k$ .

Опишем теперь *комбинированное управление выбором порядка и длины шага* несколько подробнее. Допустим, что в некоторой точке интегрирования длина шага  $H$  и индекс  $k$  ( $k > 2$ ) уже выбраны. Тогда шаг реализуется следующим образом. Сначала вычисляется  $k - 1$  строка экстраполяционной таблицы, а также величины  $H_{k-2}$ ,  $W_{k-2}$ ,  $err_{k-1}$ ,  $H_{k-1}$ ,  $W_{k-1}$ .

а) *Сходимость в строке  $k - 1$ .* Если  $err_{k-1} \leq tol$ , то значение  $T_{k-1, k-1}$  принимается в качестве вычисленного результата, и интегрирование продолжается с новыми значениями

$$k_{new} = \begin{cases} k, & \text{если } W_{k-1} < 0.9 \cdot W_{k-2}, \\ k - 1 & \text{иначе;} \end{cases} \quad (9.27)$$

$$H_{new} = \begin{cases} H_{k_{new}}, & \text{если } k_{new} \leq k - 1, \\ H_{k-1} (A_k / A_{k-1}), & \text{если } k_{new} = k. \end{cases}$$

В (9.27) единственная нетривиальная формула — это выбор длины шага  $H_{new}$  в случае увеличения порядка ( $k_{new} = k$ ). В этом случае мы хотим избежать вычисления  $err_k$ , так что  $H_k$  и  $W_k$  остаются неизвестными. Но так как мы предположили, что значение  $k$  близко к оптимальному, то имеем  $W_k \approx W_{k-1}$ , что и ведет к предлагаемой формуле увеличения длины шага.

б) *Монитор сходимости.* Если  $err_{k-1} > tol$ , надо сначала решить, можно ли ожидать сходимости по крайней мере в строке  $k + 1$ . Из равенства (9.22) следует, что в асимптотическом пределе

$$\|T_{k, k-2} - T_{k, k-1}\| \approx \left( \frac{n_2}{n_k} \right)^2 err_{k-1}, \quad (9.28)$$

где  $err_{k-1}$  вычисляется по формуле (9.23). К сожалению,  $err_k$  нельзя сравнивать с (9.28), так как в асимптотической формуле оказывается влияние различных факторов, зависящих от решаемого дифференциального уравнения (ср. (9.22)). Если же тем не менее предположить, что  $err_k$  в  $(n_2/n_1)^2$  раз меньше чем (9.28), то мы получим, что  $err_k \approx (n_1/n_k)^2 err_{k-1}$ . Поэтому уже на данном этапе шаг отбрасывается, если

$$err_{k-1} > \left( \frac{n_{k+1} n_k}{n_1 n_k} \right)^2 \cdot tol \quad (9.29)$$

и пересчитывается заново с  $k_{new} \leq k - 1$  и  $H_{new}$  в соответствии с (9.27). Если же выполнено противоположное (9.29) неравенство, то рассчитывается следующая строка экстраполяционной таблицы, т. е.  $T_{k,k}$ ,  $err_k$ ,  $H_k$ ,  $W_k$ .

с) *Сходимость в строке k.* Если  $err_k \leq tol$ , то значение  $T_{kk}$  принимается в качестве вычисленного решения, и интегрирование продолжается с новыми значениями  $k$  и  $H$ :

$$k_{new} = \begin{cases} k - 1, & \text{если } W_{k-1} < 0.9 \cdot W_k, \\ k + 1, & \text{если } W_k < 0.9 \cdot W_{k-1}, \\ k & \text{во всех других случаях;} \end{cases} \quad (9.30)$$

$$H_{new} = \begin{cases} H_{k_{new}}, & \text{если } k_{new} \leq k, \\ H_k (A_{k+1}/A_k), & \text{если } k_{new} = k + 1. \end{cases}$$

д) *Второй монитор сходимости.* Если  $err_k > tol$ , то, как и в пункте б), проверяется соотношение

$$err_k > \left( \frac{n_{k+1}}{n_1} \right)^2 \cdot tol. \quad (9.31)$$

Если оно удовлетворено, то шаг отбрасывается и пересчитывается заново с  $k_{new} \leq k$  и  $H_{new}$  из (9.30). В противном случае продолжаем счет:

е) *В надежде на сходимость в строке  $k + 1$*  вычисляются  $T_{k+1,k+1}$ ,  $err_{k+1}$ ,  $H_{k+1}$ ,  $W_{k+1}$ . Если  $err_{k+1} \leq tol$ , то значение  $T_{k+1,k+1}$  принимается в качестве вычисленного решения, и интегрирование продолжается после нового выбора порядка:

$$k_{new} := k,$$

$$\text{if } (W_{k-1} < 0.9 \cdot W_k) \quad k_{new} := k - 1, \quad (9.32)$$

$$\text{if } (W_{k+1} < 0.9 \cdot W_{k_{new}}) \quad k_{new} := k + 1.$$

Если  $err_{k+1} > tol$ , то шаг отбрасывается и пересчитывается заново с  $k_{new} \leq k$  и  $H_{new}$ , выбранным по формуле (9.24).

Рекомендуются следующие небольшие изменения приведенного алгоритма:

1) Ограниченнность памяти машины приводит к ограничению максимального числа столбцов  $k_{max}$  экстраполяционной таблицы (например,  $k_{max} = 9$ ). Поэтому на вновь назначаемый индекс  $k_{new}$  надо наложить ограничения  $2 \leq k_{new} \leq k_{max} - 1$ . Это обеспечивает возможность активизации пункта e) на каждом шаге.

2) После отбрасывания шага ни длина шага, ни порядок не должны увеличиваться.

3) Норму в (9.23) следует понимать, как масштабированную норму (ср. (4.6)). Например, компоненты погрешности могут быть отнесены к большей из двух величин:  $10^{-6}$  (к примеру) и  $i$ -й компоненты численного решения в начале и конце рассматриваемого интервала.

### Численное исследование комбинированного управления длиной шага и порядком

В нижеследующих примерах мы покажем, как изменяются длина шага и порядок для приведенного выше алгоритма. Для этой цели нами была написана на Фортране подпрограмма ODEX, текст которой приведен в приложении.

В качестве первого примера снова возьмем модель брюссельатора (4.15). В нижней части рис. 9.5 показаны обе компоненты решения (ср. рис. 4.1 и 4.2 в разд. II.4). В средней и верхней части рисунка приведены графики изменения длин шагов ( $H_{max} = 1.7$ ) и порядков при трех значениях  $tol$ :  $10^{-8}$  (пунктир),  $10^{-6}$  (штриховая линия) и  $10^{-4}$  (сплошная линия). Легко заметить, что программа ODEX автоматически выбирает подходящий порядок (который существенно зависит от значения  $tol$ ). Вертикальные участки на графиках длин шагов указывают на отбраковку и пересчет соответствующего шага.

Рассмотрим теперь поведение механизма управления порядком метода вблизи разрывов. В следующем примере:

$$y' = -\text{sign}(x)|1 - |x|| \cdot y^2, \quad y(-2) = \frac{2}{3}, \quad -2 \leq x \leq 2, \quad (9.33)$$

имеется разрыв первой производной  $y'(x)$  при  $x = 0$  и два разрыва второй производной (при  $x = \pm 1$ ). Результаты расчетов показаны на рис. 9.6. При вычислениях с  $tol = 10^{-8}$  разрывы при  $x = \pm 1$  не были распознаны. В остальных случаях видно, как порядок

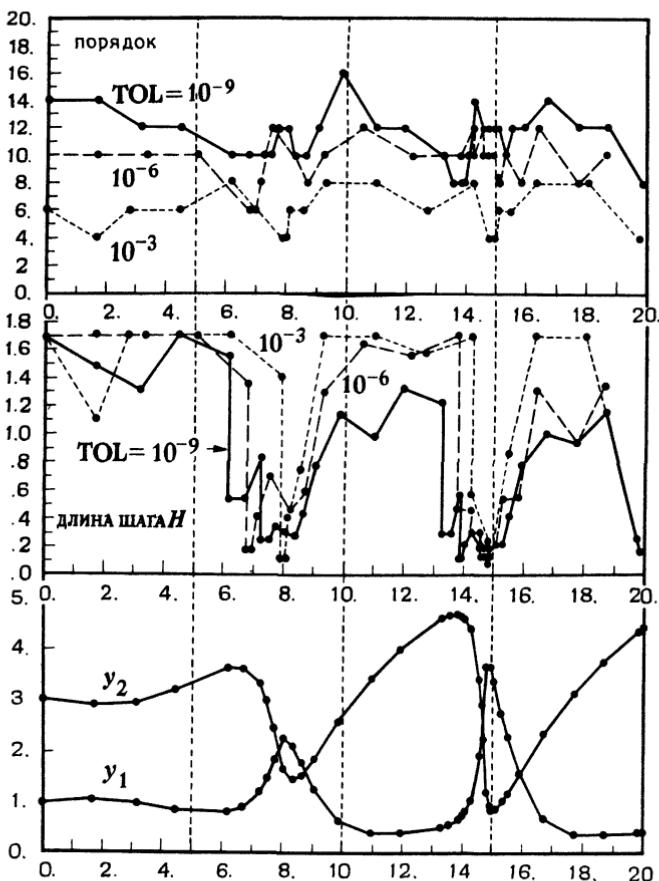


Рис. 9.5. Решение, изменения длины шага и порядка для подпрограммы ОDEX.

понижается до 4-го (наименьшего возможного) в окрестности точек разрыва, так что эти точки проходятся довольно эффективно.

### Упражнения

- Покажите, что экстраполированные методы Эйлера  $T_{3,1}$ ,  $T_{3,2}$ ,  $T_{3,3}$  [с последовательностью чисел шагов (9.8)] эквивалентны методам Рунге—Кутты из табл. 9.1. Найдите также схемы РК, соответствующие первым элементам алгоритма ГБШ.

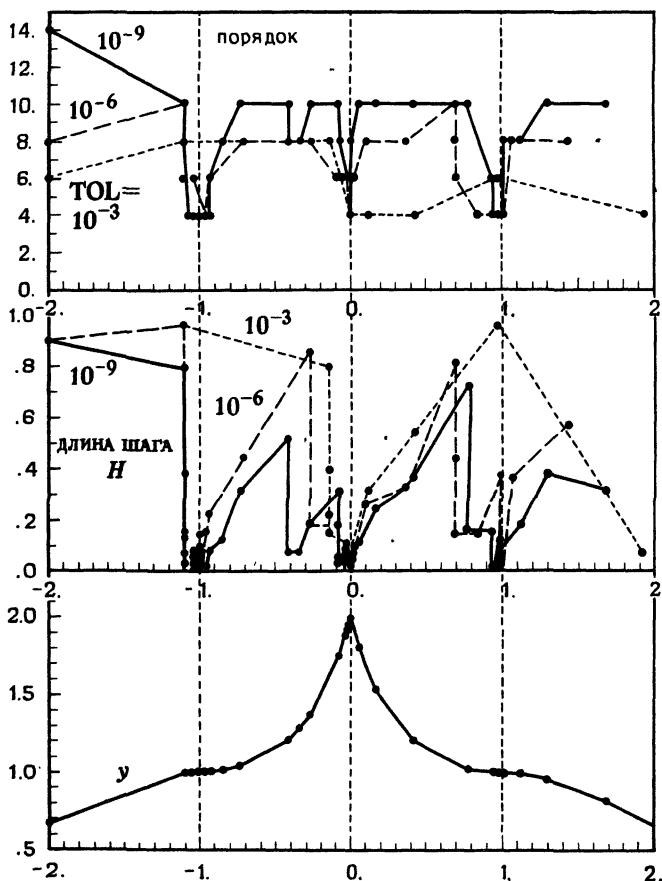


Рис. 9.6. Решение, изменения длины шага и порядка для подпрограммы ODEX, полученные на примере с разрывной функцией (9.33).

2. Комбинируя формулы (9.17a) и (9.19), получите формулу

$$y(x) - y_h = \sum_{l=1}^l [\alpha_j(x) + (-1)^k \beta_j(x)] h^{2l} + h^{2l+2} E_N(x, h)$$

для асимптотического разложения, соответствующего методу Грэгга в той его части, которая определяется формулами (9.13a и b).

Таблица 9.1. Экстраполяционные методы как методы Рунге—Кутты

$T_{8,1}$ порядка 1	$T_{8,2}$ порядка 2	$T_{8,3}$ порядка 3
0	0	0
$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{2}$
$\frac{2}{3}$	$\frac{1}{3}$ $\frac{1}{3}$	$\frac{1}{3}$ 0
	$\frac{2}{3}$	$\frac{2}{3}$ $\frac{1}{3}$ 0 $\frac{1}{3}$
	$\frac{1}{3}$ $\frac{1}{3}$ $\frac{1}{3}$	$\frac{1}{3}$ 0 $\frac{1}{3}$
	0 -1 1 1	0 -2 $\frac{3}{2}$ $\frac{3}{2}$

3. [Штеттер (1970)]. Докажите, что для любого вещественного  $b$  (обычно заключенного между 0 и 1) метод

$$\begin{aligned} y_1 &= y_0 + h [bf(x_0, y_0) + (1 - b)f(x_1, y_1)], \\ y_{i+1} &= y_{i-1} + h [(1 - b)f(x_{i-1}, y_{i-1}) + 2bf(x_i, y_i) + \\ &\quad + (1 - b)f(x_{i+1}, y_{i+1})], \quad i = 1, 2, \dots, \end{aligned}$$

обладает разложением по степеням  $h^2$ . Покажите, что это свойство сохраняется и для сглаживающего шага

$$S_h(x) = \frac{1}{2} [y_{2n} + y_{2n-1} + h(1 - b)f(x_{2n-1}, y_{2n-1}) + hb f(x_{2n}, y_{2n})].$$

4. [Штеттер (1970)]. Является ли шаг методом Эйлера (9.13a) существенным для существования разложения по степеням  $h^2$ ? Докажите, что если заменить (9.13a) стартовой процедурой

$$y_1 = y_0 + h\Phi(x_0, y_0, h)$$

первого порядка, то разложение по степеням  $h^2$  получится при условии, что величины  $y_{-1} = y_0 - h\Phi(x_0, y_0, -h)$ ,  $y_0$  и  $y_1$  удовлетворяют (9.13b) при  $i = 0$ .

*Указание.* Штеттер в своем доказательстве опирался на соображения о возмущении начальных условий. Более естественная возможность состоит в замене шагов по Эйлеру в (9.15) новой стартовой процедурой.

5. Получите численные решения системы дифференциальных уравнений

$$\begin{aligned} y' &= z, & y(0) &= 30, \\ z' &= 0.032 - 0.4z^2, & z(0) &= 0 \end{aligned}$$

при длинах шагов  $h = 0.5$  и  $h = 0.25$ , пользуясь методом из упр. 3 разд. II.8. Вычислите погрешности решения и экстраполированного результата. Проверьте вычислениями утверждение, содержащееся в указанном упражнении.

*Замечание.* Точное решение имеет вид:

$$y(x) = 30 + 2.5 \ln(\operatorname{ch}(\sqrt{0.0128}x)),$$

$$z(x) = \sqrt{0.08} \operatorname{th}(\sqrt{0.0128}x).$$

6. Покажите, что определитель матрицы

$$A = \begin{vmatrix} 1 & a_1^p & \dots & a_1^{p+n-2} \\ 1 & a_2^p & \dots & a_2^{p+n-2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & a_n^p & \dots & a_n^{p+n-2} \end{vmatrix}$$

при  $0 < a_1 < a_2 < \dots < a_n$  строго положителен.

*Указание.* Вычтите из столбца  $n$  столбец  $n - 1$ , умноженный на  $a_1$ , затем из столбца  $n - 1$  столбец  $n - 2$ , умноженный на  $a_1$ , и т. д. и, наконец, вычтите из столбца 2 столбец 1, умноженный на  $a_1^p$ . Затем вынесите положительные множители  $(a_2 - a_1)$ ,  $(a_3 - a_1)$ ,  $\dots$   $(a_n - a_1)$  и покажите, что определитель  $A$  можно представить в виде суммы определителей аналогичных матриц меньшей размерности, умноженных на положительные коэффициенты.

## II.10. Сравнение вычислительных качеств

«По-видимому, программа DIFEX1 является воплощением современного состояния экстраполяционных методов...»

(Л. Ф. Шампайн, Л. С. Бака, 1984)

Теперь уместно представить некоторые результаты сравнения вычислительных качеств описанных выше программ ODEX (метод экстраполяции) и DOPRI8 (метод Дормана и Принса порядка 8). Тексты обеих программ приведены в приложении. Мы сочли интересным включить в число сравниваемых еще и широко известную программу DIFEX1 [Дойфлхард (1983)]. Хотя эта программа воплощает точно тот же математический метод, что и ODEX, она отличается многими деталями стратегии выбора порядка, основанной на некоторой информационно-теоретической модели.

Из большого числа использованных нами для тестирования примеров дифференциальных уравнений мы отобрали здесь следующие шесть систем [первый пример приобрел известность благодаря статье Булирша и Штёра (1966), а следующие два взяты из знаменитого набора Detest set, Халл, Энрайт, Феллен и Седжвик (1972)].

JACB: дифференциальные уравнения для эллиптических функций Якоби [см. Абрамович и Стиган (1964), 16.16]:

$$\begin{aligned}y'_1 &= y_2 y_3, & y_1(0) &= 0, \\y'_2 &= -y_1 y_3, & y_2(0) &= 1, \\y'_3 &= -0.51 y_1 y_2, & y_3(0) &= 1,\end{aligned}$$

$$0 \leq x \leq 20.$$

TWOB: уравнения задачи двух тел:

$$\begin{aligned}y'_1 &= y_3, & y_1(0) &= 0.5, \\y'_2 &= y_4, & y_2(0) &= 0, \\y'_3 &= -\frac{y_1}{(y_1^2 + y_2^2)^{3/2}}, & y_3(0) &= 0, \\y'_4 &= -\frac{y_2}{(y_1^2 + y_2^2)^{3/2}}, & y_4(0) &= \sqrt{3}, \\0 &\leq x \leq 20.\end{aligned}$$

VDPL: уравнение Ван-дер-Поля (см. разд. I.16):

$$y'_1 = y_2, \quad y_1(0) = 2,$$

$$y'_2 = (1 - y_1^2)y_2 - y_1, \quad y_2(0) = 0,$$

$$0 \leq x \leq 20.$$

BRUS: «брюсселятор» (см. разд. I.16):

$$y'_1 = 2 + y_1^2 y_2 - 9.533 y_1, \quad y_1(0) = 1,$$

$$y'_2 = 8.533 y_1 - y_1^2 y_2, \quad y_2(0) = 4.2665,$$

$$0 \leq x \leq 20.$$

LAGR: уравнение Лагранжа для висящей струны (уравнение (6.2) из разд. I.6 при  $n = 10$ ,  $K = 1$ ). Начальные условия: все  $y_i(0) = 0$ , за исключением  $y_8(0) = 1$ . Интервал интегрирования:  $0 \leq x \leq 10$ .

PLEI: задача небесной механики, которую мы назвали «Плеяды»: семь звезд с координатами  $x_i$ ,  $y_i$  и массами  $m_i = i$  ( $i = 1, \dots, 7$ ) движутся в одной плоскости:

$$x''_i = \sum_{l \neq i} m_l (x_l - x_i) / r_{il},$$

$$y''_i = \sum_{l \neq i} m_l (y_l - y_i) / r_{il},$$

где

$$r_{ij} = ((x_i - x_j)^2 + (y_i - y_j)^2)^{3/2}, \quad i, j = 1, \dots, 7.$$

Начальные условия:

$$x_1(0) = 3, \quad x_2(0) = 3, \quad x_3(0) = -1, \quad x_4(0) = -3,$$

$$x_5(0) = 2, \quad x_6(0) = -2, \quad x_7(0) = 2,$$

$$y_1(0) = 3, \quad y_2(0) = -3, \quad y_3(0) = 2, \quad y_4(0) = 0,$$

$$y_5(0) = 0, \quad y_6(0) = -4, \quad y_7(0) = 4,$$

$x'_i(0) = y'_i(0) = 0$  для всех  $i$ , за исключением:

$$x'_6(0) = 1.75; \quad x'_7(0) = -1.5; \quad y'_4(0) = -1.25; \quad y'_5(0) = 1.$$

Интервал интегрирования  $[0, 3]$ .

Точные решения этих задач представлены графиками на рис. 10.1 как функции  $x$ . Исключение сделано для задачи PLEI, решение которой представлено в виде траекторий в плоскости  $x, y$ .

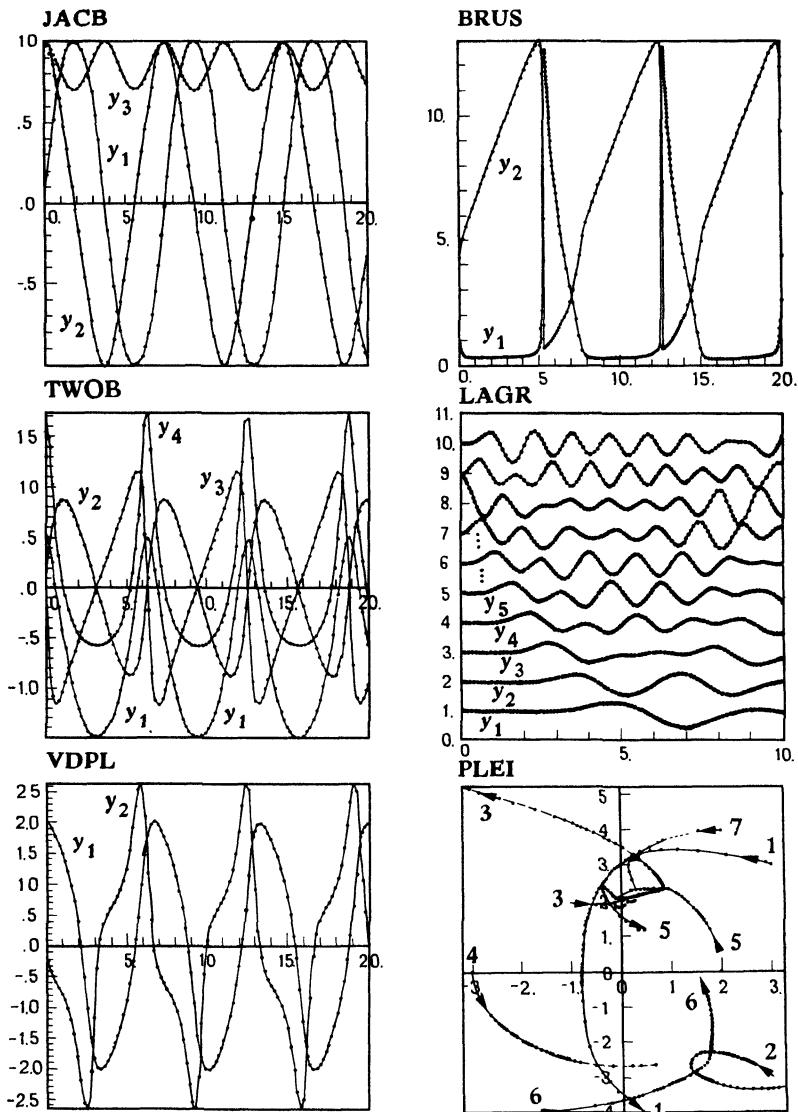


Рис. 10.1. Численные решения приведенных выше шести задач.

Чтобы дать лучшее представление о вычислительных трудностях, встретившихся в задаче PLEI, мы привели на рис. 10.2 графики решений  $x_i$ ,  $y_i$  и их производных  $x'_i$ ,  $y'_i$  как функции независимой переменной.

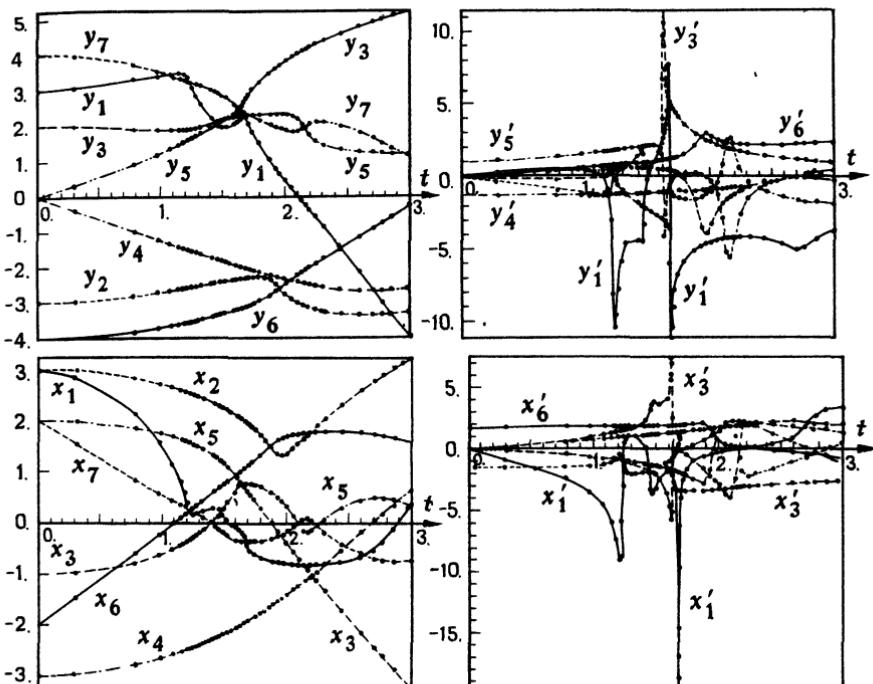


Рис. 10.2. Точные решения задачи PLEI.

### Результаты расчетов

Каждый из указанных выше примеров решался со следующими значениями допусков локальной погрешности:  $tol = 10^{-3}$ ,  $10^{-6}$ ,  $10^{-9}$ ,  $10^{-12}$  и  $10^{-15}$ . На конце интервала интегрирования вычисленное решение сравнивалось с точным решением (которое вычислялось на компьютере CDC с  $tol = 10^{-22}$ ). В качестве меры эффективности мы сопоставляем на графиках рис. 10.3 в двойном логарифмическом масштабе полученную точность —  $\lg(|err|)$  и число обращений к вычислению функций. Видно, что обе экстраполяционные программы (ODEX и DIFEX1) дают почти равнозначенные результаты. Программа DOPRI8 с ее фиксированным порядком точности явно превосходит их при  $10^{-3} \leq tol \leq 10^{-12}$ , и только при самых высоких требованиях к точности начинает сказываться высокий порядок точности экстраполяционных программ.

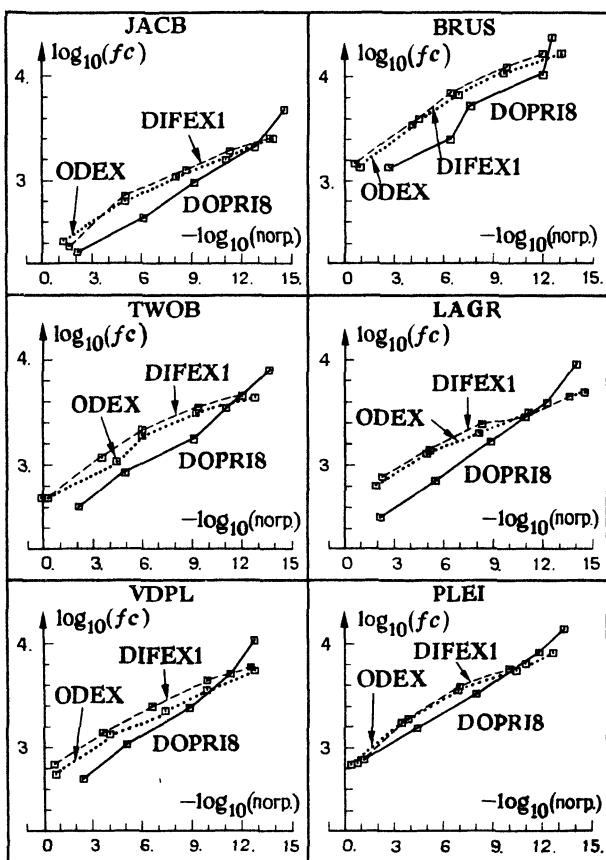


Рис. 10.3. Диаграммы «точность — объем вычислений» для приведенных выше примеров.

### Пример с негладким решением

Уравнение (9.33) содержит в интервале интегрирования три точки, в которых терпят разрыв первая или вторая производные. В этом случае получаемая точность — вопрос удачи: как расположатся границы шагов интегрирования относительно точек разрыва. Этот пример мы решили для очень большого числа значений допустимой погрешности ( $tol = 10^{-3-i/10}$ ,  $i = 0, 1, \dots, 90$ ). Полученные значения погрешностей и числа обращений к вычислению функций ведут себя совершенно хаотически. Для всех трех сравниваемых программ они представлены на рис. 10.4.

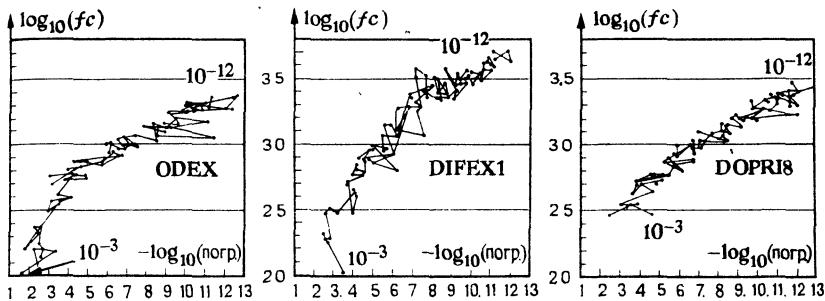


Рис. 10.4. Диаграммы «точность — объем вычислений» для уравнения (9.33).

Можно заметить, что DIFEX1 делает приблизительно вдвое больше вычислений функции, чем две другие программы, из которых ODEX слегка выигрывает благодаря понижению порядка в окрестности точек нарушения гладкости.

### Заключение

Для гладких задач и допустимой локальной погрешности от  $10^{-3}$  до  $10^{-12}$  предпочтительнее программа DOPRI8, имеющая фиксированный порядок точности. Экстраполяционные программы превосходят ее при очень малых значениях допустимой погрешности благодаря автоматическому *увеличению* порядка точности. Программа ODEX превосходит две другие при решении задач с негладкой функцией благодаря предусмотренному в ней автоматическому *понижению* порядка.

## II.11. Композиция В-рядов

Теперь мы установим центральную теорему о композиции объектов, которые мы назвали В-рядами (в честь Бутчера). Эта теорема будет иметь много приложений к более общим методам, например к методам со старшими производными (разд. II.12), методам Розенброка (том II) и др. Она заменит также теорию условий порядка, развитую в разд. II.2.

### Композиция методов РК

Пятистадийных методов ЯРК пятого порядка не существует (см. разд. II.6). Бутчеру (1969) принадлежит идея подыскать такие различные пятистадийные методы, определенная *композиция* которых давала бы результат пятого порядка точности («эффективный порядок»). Это послужило отправной точкой очень красивой алгебраической теории численных методов, хотя сама идея большого практического интереса не представляет, главным образом из-за проблемы изменения длины шага.

Пусть имеется два метода, например трехстадийных:

$$\begin{array}{c|ccc} 0 & 0 \\ \hat{c}_2 & \hat{a}_{21} & & \tilde{c}_2 & \tilde{a}_{21} \\ \hat{c}_3 & \hat{a}_{31} & \hat{a}_{32} & \tilde{c}_3 & \tilde{a}_{31} & \tilde{a}_{32} \\ \hline b_1 & b_2 & b_3 & b_1 & b_2 & b_3 \end{array} . \quad (11.1)$$

Отправляясь от начального условия  $y_0$ , применим их один за другим, сделав по одному шагу одинаковой длины  $h$  (ср. формулы (2.3)):

$$g_i = y_0 + h \sum_l \hat{a}_{il} f(g_j), \quad y_1 = y_0 + h \sum_l b_{jl} f(g_j), \quad (11.2)$$

$$l_i = y_1 + h \sum_l \tilde{a}_{il} f(l_j), \quad y_2 = y_1 + h \sum_l b_{jl} f(l_j). \quad (11.3)$$

Если подставить  $y_1$  из (11.2) в (11.3) и собрать все  $g_i$ ,  $l_i$  вместе, то легко видеть, что композицию можно интерпретировать как большой метод РК с коэффициентами

$$\begin{array}{c|cccccc} 0 & & & & & & \\ \hat{c}_2 & \hat{a}_{21} & & & & & \\ \hat{c}_3 & \hat{a}_{31} & \hat{a}_{32} & & & & \\ \sum \hat{b}_i & \hat{b}_1 & \hat{b}_2 & \hat{b}_3 & & & \\ \sum \hat{b}_i + \tilde{c}_2 & \hat{b}_1 & \hat{b}_2 & \hat{b}_3 & \tilde{a}_{21} & & \\ \sum \hat{b}_i + \tilde{c}_3 & \hat{b}_1 & \hat{b}_2 & \hat{b}_3 & \tilde{a}_{31} & \tilde{a}_{32} & \\ \hline \hat{b}_1 & \hat{b}_2 & \hat{b}_3 & \hat{b}_1 & \hat{b}_2 & \hat{b}_3 & \end{array} = \begin{array}{c|cccccc} 0 & & & & & & \\ c_2 & a_{21} & & & & & \\ c_3 & a_{31} & a_{32} & & & & \\ c_4 & a_{41} & a_{42} & a_{43} & & & \\ c_5 & a_{51} & a_{52} & a_{53} & a_{54} & & \\ c_6 & a_{61} & a_{62} & a_{63} & a_{64} & a_{65} & \\ \hline b_1 & b_2 & b_3 & b_4 & b_5 & b_6 & \end{array} \quad (11.4)$$

Теперь интересно посмотреть условия порядка нового метода. Для этого надо вычислить выражения

$$\sum b_i, \quad 2 \sum b_i c_i, \quad 3 \sum b_i c_i^2, \quad 6 \sum b_i a_{ij} c_i \text{ и т. д.}$$

(см. табл. 2.1). Если подставить сюда выражения из первой таблицы (11.4), то не слишком трудные для низких порядков вычисления показывают, что эти величины могут быть выражены через соответствующие величины для двух методов (11.1). Для первого метода мы обозначим эти выражения  $a(t)$ , для второго метода  $b(t)$ , а для композитного метода — через  $ab(t)$ :

$$a(\tau) = \sum b_i, \quad a(t_{21}) = 2 \sum b_i \hat{c}_i, \quad a(t_{31}) = 3 \cdot \sum b_i \hat{c}_i^2, \dots \quad (11.5a)$$

$$b(\tau) = \sum b_i, \quad b(t_{21}) = 2 \sum b_i \tilde{c}_i, \quad b(t_{31}) = 3 \cdot \sum b_i \tilde{c}_i^2, \dots \quad (11.5b)$$

$$ab(\tau) = \sum b_i, \quad ab(t_{21}) = 2 \sum b_i c_i, \quad ab(t_{31}) = 3 \cdot \sum b_i c_i^2, \dots \quad (11.5c)$$

Здесь  $\tau, t_{ij}$  — деревья из табл. 2.1. В этих обозначениях упомянутые формулы будут иметь такой вид:

$$\begin{aligned} ab(\tau) &= a(\tau) + b(\tau), \\ ab(t_{21}) &= a(t_{21}) + 2b(\tau)a(\tau) + b(t_{21}), \\ ab(t_{31}) &= a(t_{31}) + 3b(\tau)a(\tau)^2 + 3b(t_{21})a(\tau) + b(t_{31}), \\ ab(t_{32}) &= a(t_{32}) + 3b(\tau)a(t_{21}) + 3b(t_{21})a(\tau) + b(t_{32}) \end{aligned} \quad (11.6)$$

и т. д.

Интересно, конечно, понять теперь общую структуру этих формул для произвольных деревьев. Однако, пользуясь старыми средствами, это сделать нелегко («...громоздкие, утомительные вычисления показывают, что...»). Наряду с этим существует проблема идентификации различных методов, дающих одинаковые численные результаты (см. упр. 1 ниже). Кроме того, мы хотим, чтобы теория включала процессы более общие, чем методы Рунге—Кутты, например точное решение или методы со старшими производными.

## В-ряды

Разрешить все эти проблемы можно, непосредственно рассматривая композицию рядов Тейлора из разд. II.2. Определим множество всех деревьев

$$T = \{\emptyset\} \cup T_1 \cup T_2 \cup \dots,$$

в которое включено и «пустое» дерево  $\emptyset$  порядка 0, представляющее начальные условия; соответствующим ему элементарным дифференциалом является тождественное отображение:  $F(\emptyset)(y) = y$ . Аналогично определим множество всех помеченных деревьев:

$$LI = \{\emptyset\} \cup LT_1 \cup LT_2 \cup \dots$$

**Определение 11.1.** [Хайрер и Ваннер (1974)] Пусть  $a(\emptyset)$ ,  $a(t)$ ,  $a(t_{21})$ ,  $a(t_{31})$ , ... — последовательность вещественных коэффициентов, определенных для всех деревьев, т. е.  $a: T \rightarrow \mathbb{R}$ . Назовем В-рядом следующий ряд:

$$\begin{aligned} a(\emptyset) y^J + h a(t) f^J(y) + \frac{h^2}{2!} a(t_{21}) f_K^J(y) f^K(y) + \dots = \\ = \sum_{t \in LT} \frac{h^{\rho(t)}}{\rho(t)!} a(t) F^J(t)(y) =: B^J(a, y) \end{aligned} \quad (11.7)$$

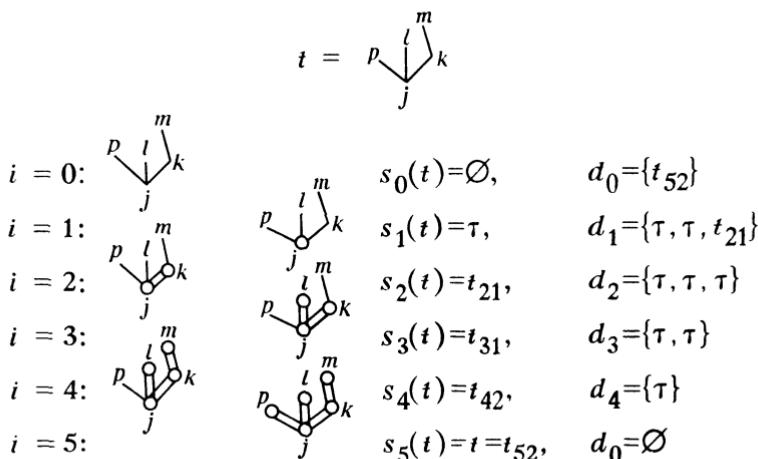
(см. определения 2.2, 2.3 и теорему 2.11).

Согласно теореме 2.11, приближенное (численное) решение, даваемое методом РК, является В-рядом, а в силу теоремы 2.6 это верно и для точного решения, у которого все коэффициенты В-ряда равны единице.

Обычно представляет интерес только конечное число членов таких рядов, не превосходящее порядков рассматриваемых методов или степени гладкости функции  $f$ . Поэтому все наши дальнейшие результаты будут верны с точностью до членов погрешности  $O(h^{k+1})$ .

**Определение 11.2.** Пусть  $t \in LT$  — какое-то помеченное дерево порядка  $q = \rho(t)$ , и фиксирано некоторое целое число  $i$  ( $0 \leq i \leq q$ ). Обозначим тогда  $s_i(t) = s$  субдерево, образованное первыми  $i$  индексами, и  $d_i(t)$  множество субдеревьев, образованных индексами  $i+1, \dots, q$  (разностное множество). На графах мы будем выделять субдерево  $s$ , обозначая его вершины кружками, а ребра — двойными линиями.

**Пример 11.3.**



«И был свет» (Гайдн).

**Определение 11.4.** Пусть  $a: T \rightarrow \mathbb{R}$  и  $b: T \rightarrow \mathbb{R}$  суть две последовательности коэффициентов, такие что  $a(\emptyset) = 1$ . Определим тогда для дерева  $t$  порядка  $q = \rho(t)$  композицию

$$ab(t) = \frac{1}{\alpha(t)} \sum \left[ \sum_{i=0}^q \binom{q}{i} b(s_i(t)) \prod_{z \in d_i(t)} a(z) \right], \quad (11.8)$$

где первая сумма берется по всем  $\alpha(t)$  различным возможным индексациям дерева  $t$  (см. определение 2.5).

**Пример 11.5.** Легко видеть, что формулы (11.6) являются частными случаями (11.8). Дерево  $t$  из примера 11.3 допускает шесть различных индексаций, что дает в результате

$$\begin{aligned} ab(t_{52}) &= b(\emptyset) a(t_{52}) + 5b(\tau) a(\tau)^2 a(t_{21}) + \\ &+ 10 \left[ \frac{1}{2} b(t_{21}) a(\tau) a(t_{21}) + \frac{1}{2} \cdot b(t_{21}) a(\tau)^3 \right] + \\ &+ 10 \left[ \frac{1}{6} b(t_{31}) a(t_{21}) + \frac{4}{6} b(t_{31}) a(\tau)^2 + \frac{1}{6} b(t_{32}) a(\tau)^2 \right] + \\ &+ 5 \left[ \frac{1}{2} b(t_{41}) a(\tau) + \frac{1}{2} b(t_{42}) a(\tau) \right] + b(t_{52}) \end{aligned} \quad (11.9)$$

(обозначения деревьев соответствуют табл. 2.1).

Докажем теперь основную теорему этого раздела.

**Теорема 11.6.** [Хайрер и Ваннер (1974)] Пусть, как и выше,  $a: T \rightarrow \mathbb{R}$  и  $b: T \rightarrow \mathbb{R}$  — две последовательности коэффициентов, такие что  $a(\emptyset) = 1$ . Тогда композиция двух соответствующих им В-рядов является тоже В-рядом:

$$B(b, B(a, y)) = B(ab, y), \quad (11.10)$$

где «произведение»  $ab: T \rightarrow \mathbb{R}$  приведено в определении 11.4.

**Доказательство.** Введем следующее обозначение для внутреннего ряда:

$$B(a, y) = g(h). \quad (11.11)$$

Тогда доказательство становится очень похожим на рассуждения, приведенные в разд. II.2 (см. рис. 2.2), с той лишь разницей, что вместо  $f^J(g)$  мы отправляемся теперь от

$$B(b, g) = \sum_{s \in LT} \frac{h^{\rho(s)}}{\rho(s)!} b(s) F(s)(g) \quad (11.12)$$

и должны вычислить производные этой функции. Выберем для примера член  $s = t_{32}$  из этого ряда:

$$\frac{h^3}{3!} b(t_{32}) f_L^K(g) f_M^L(g) f^M(g). \quad (11.13)$$

По формуле Лейбница  $q$ -я производная этого выражения при  $h = 0$  равна

$$\binom{q}{3} b(t_{32}) [f_L^K(g) f_M^L(g) f^M(g)]^{(q-3)}|_{h=0}. \quad (11.14)$$

Пользуясь классическими правилами дифференциального исчисления, вычислим теперь, как это делалось в лемме 2.8, производные выражения

$$f_L^K(g) f_M^L(g) f^M(g). \quad (11.15)$$

Для первой производной это дает

$$f_{LN}^K \cdot (g^N)' f_M^L f^M + f_L^K f_{MN}^L \cdot (g^N)' f^M + f_L^K f_M^L f_N^M \cdot (g^N)',$$

и т. д. Воспользуемся опять графическим представлением этого процесса дифференцирования (рис. 11.1).

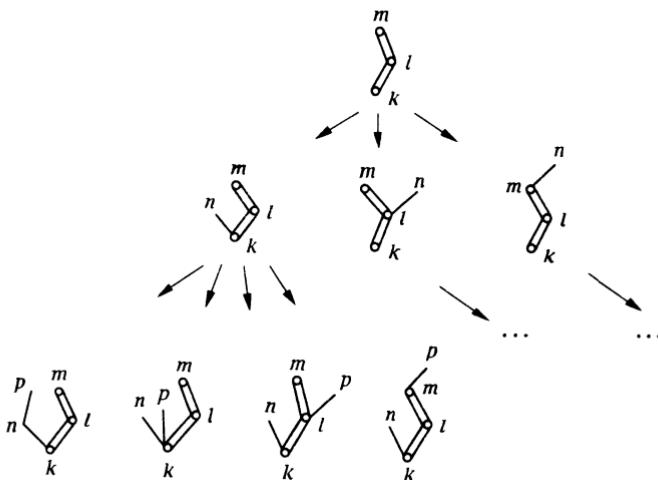


Рис. 11.1. Производные формулы (11.15).

Мы видим, что получаются деревья  $t$  порядка  $q$ , у которых  $s_3(u) = s$  (где  $3 = \rho(s)$ ) и элементы множеств  $d_3(u)$  не имеют разветвлений. Соответствующие выражения аналогичны выражениям (2.6;  $q - 1$ ) из леммы 2.8. Наконец, надо подставить выражения для производных  $g$  (см. (11.11)) и произвести перегруппировку членов. Тогда, как и на рис. 2.3, длинные ветви из  $d_3(u)$  заменятся деревьями  $z$  порядка  $\delta$ , умноженными на  $a(z)$ . Таким образом, для каждого данного дерева  $t$  получаются как раз коэффициенты, описываемые формулой (11.8).

Множитель  $1/\alpha(t)$  обязан своим происхождением тому обстоятельству, что в  $B(ab, y)$  член с  $ab(t)F(t)$  появляется  $\alpha(t)$  раз. ■

Поскольку  $hf(y) = B(b, y)$  является B-рядом частного вида, у которого  $b(t) = 1$ , а все остальные  $b(t) = 0$ , то из доказанной теоремы вытекает

**Следствие 11.7.** Если  $a: T \rightarrow \mathbb{R}$  — последовательность коэффициентов, такая, что  $a(\emptyset) = 1$ , то

$$hf(B(a, y)) = B(a', y),$$

причем

$$\begin{aligned} a'(\emptyset) &= 0, & a'(\tau) &= 1, \\ a'([t_1, \dots, t_m]) &= \rho(t) a(t_1) \dots a(t_m), \end{aligned} \quad (11.16)$$

где  $t = [t_1, \dots, t_m]$  означает, что  $d_1(t) = \{t_1, t_2, \dots, t_m\}$  (см. определение 2.12).

**Доказательство.** Формулы (11.16) получаются из (11.8) при  $i = 1$ ,  $q = \rho(t)$  с учетом того факта, что выражение в скобках не зависит от индексации дерева  $t$ . ■

### Условия порядка для методов Рунге—Кутты

В качестве приложения следствия 11.7 продемонстрируем вывод условий порядка. Запишем метод РК (2.3) в виде

$$g_i = y_0 + \sum_l a_{ij} k_j, \quad k_i = hf(g_i). \quad (11.17)$$

Допустим, что  $g_i$  и  $k_i$  являются B-рядами, коэффициенты которых обозначим  $g_{ij}$ ,  $k_{ij}$ :

$$g_i = B(g_i, y_0), \quad k_i = B(k_i, y_0).$$

Тогда следствие 11.7 позволяет нам непосредственно переписать формулы (11.17) в виде

$$\begin{aligned} g_i(\emptyset) &= 1, & g_i(t) &= \sum_l a_{ij} k_j(t), \\ k_i(\tau) &= 1, & k_i(t) &= \rho(t) g_i(t_1) \dots g_i(t_m), \end{aligned} \quad (11.17')$$

откуда легко получаются формулы (2.17), (2.19) и теорема 2.11.

Аналогично, если мы положим  $y(h) = B(y, y_0)$  для точного решения дифференциального уравнения и сравним (умноженную на  $h$ ) производную  $hy'(h)$  ряда (11.7) с выражением  $hf(y)$  из следствия 11.7, мы сразу же получим, что  $y(t) = 1$  для всех  $t$ , так что теорема 2.13 становится ненужной.

### «Эффективный порядок» Бутчера

Разобравшись в том, что такая композиция методов, будем теперь искать такие методы  $a$  и  $d$ , что  $da = ed$  вплоть до членов, соответствующих порядку 5 (здесь  $e$  — некоторый метод пятого порядка). Кроме того, построим еще методы  $b$  и  $c$ , такие, что

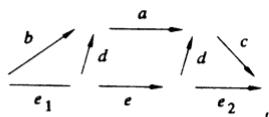


Рис. 11.2.

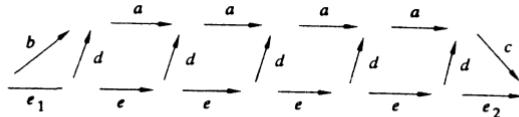


Рис. 11.3.

$e_1d = b$  и  $dc = e_2$ , где  $e_1$  и  $e_2$  — тоже методы пятого порядка (см. рис. 11.2). Тогда композиция методов  $ba^{k-2}c$  по своим вычислительным свойствам соответствует результату выполнения  $k$  шагов методом пятого порядка (рис. 11.3). Один из возможных наборов коэффициентов, вычисленный Бутчером (1969), приведен в табл. 11.1.

**Таблица 11.1** Метод Бутчера с эффективным порядком 5. Вверху слева — метод  $a$ , справа — метод  $b$ , внизу — метод  $c$ .

0		0		
$\frac{1}{5}$	$\frac{1}{5}$			
$\frac{2}{5}$	0	$\frac{2}{5}$		
$\frac{1}{2}$	$\frac{3}{16}$	0	$\frac{5}{16}$	
1	$\frac{1}{4}$	0	$-\frac{5}{4}$	2
	$\frac{1}{6}$	0	$\frac{2}{3}$	$\frac{1}{6}$
	$\frac{19}{144}$	0	$\frac{25}{48}$	$\frac{2}{9}$
			$\frac{1}{8}$	

0				
$\frac{1}{5}$	$\frac{1}{5}$			
$\frac{2}{5}$	0	$\frac{2}{5}$		
$\frac{3}{4}$	$\frac{161}{192}$	$-\frac{19}{12}$	$\frac{287}{192}$	
1	$-\frac{27}{28}$	$\frac{19}{7}$	$-\frac{291}{196}$	$\frac{36}{49}$
	$\frac{7}{48}$	0	$\frac{475}{1008}$	$\frac{2}{7}$
			$\frac{7}{72}$	

## Упражнения

1. Покажите, что пары методов, приведенные в табл. 11.2—11.4, дают идентичные численные результаты (по крайней мере для достаточно малых  $h$ ).

*Ответ.* Эквивалентность методов а) очевидна, если в одном из них переставить стадии; в случае б) достаточно отбросить лишние стадии [Далквист и Йелч (1979)]; в с) надо отождествить равные стадии [Штеттер (1973), Хундсдорфер и Спейкер (1981)]. См. также обзор Бутчера (1984) о «пространстве Рунге—Кутты».

2. Продолжите формулы (11.6), вычислив композицию  $ab(t)$  для всех деревьев порядков 4 и 5.
3. Проверьте, что методы, приведенные в табл. 11.1, действительно удовлетворяют свойствам точности, о которых говорилось выше.

Таблица 11.2. Эквивалентные методы а)

0	0	1	0	1
1	1	0	0	0
		$\frac{1}{4}$	$\frac{3}{4}$	
			$\frac{3}{4}$	$\frac{1}{4}$

Таблица 11.3 Эквивалентные методы б)

1	2	0	0	-1		1	2	-1
3	0	1	2	0		2	1	1
7	0	3	4	0				
2	1	0	0	1				
	$\frac{1}{2}$	0	0	$\frac{1}{2}$		$\frac{1}{2}$	$\frac{1}{2}$	

Таблица 11.4. Эквивалентные методы с)

1	1	1	1	-2		1	3	-2
1	2	2	-1	-2		-1	2	-3
1	-1	-1	5	-2				
-1	-1	2	1	-3				
	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$		$\frac{3}{4}$	$\frac{1}{4}$	

4. Докажите, что множество

$$G = \{a: T \rightarrow \mathbb{R}; \quad a(\emptyset) = 1\}$$

с законом композиции из определения 11.4 является группой (некоммутативной).

*Указание.* Воспользуйтесь теоремой 11.6.

## II.12. Методы, использующие старшие производные

В разд. I.8 уже рассматривалось вычисление старших производных решений дифференциального уравнения

$$(y^J)' = f^J(x, y^1, \dots, y^n), \quad J = 1, \dots, n. \quad (12.1)$$

«Цепное правило» (правило дифференцирования суперпозиции функций) дает

$$\begin{aligned} (y^J)'' &= \frac{\partial f^J}{\partial x}(x, y) + \frac{\partial f^J}{\partial y^1}(x, y) \cdot f^1(x, y) + \dots + \\ &+ \frac{\partial f^J}{\partial y^n}(x, y) \cdot f^n(x, y). \end{aligned} \quad (12.2)$$

Поэтому введем дифференциальный оператор  $D$ , действие которого на функцию  $\Psi(x, y)$  определяется равенством

$$\begin{aligned} (D\Psi)(x, y) &= \frac{\partial \Psi}{\partial x}(x, y) + \frac{\partial \Psi}{\partial y^1}(x, y) \cdot f^1(x, y) + \\ &+ \dots + \frac{\partial \Psi}{\partial y^n}(x, y) \cdot f^n(x, y). \end{aligned}$$

Поскольку  $Dy^J = f^J$ , формулы, обобщающие (12.2), будут иметь вид:

$$(y^J)^{(l)} = (D^l y^J)(x, y), \quad l = 0, 1, 2, \dots. \quad (12.3)$$

Эти обозначения удобны для определения нового класса методов, сочетающих в себе черты как методов Рунге—Кутты, так и методов, использующих ряды Тейлора.

**Определение 12.1.** Пусть заданы вещественные коэффициенты

$$a_{il}^{(r)}, b_l^{(r)} \quad (i, j = 1, \dots, s; \quad r = 1, \dots, q).$$

Метод, определяемый формулами

$$\begin{aligned} k_i^{(l)} &= \frac{h^l}{l!} (D^l y) \left( x_0 + c_l h, \quad y_0 + \sum_{r=1}^q \sum_{l=1}^s a_{il}^{(r)} k_l^{(r)} \right), \\ y_1 &= y_0 + \sum_{r=1}^q \sum_{l=1}^s b_l^{(r)} k_l^{(r)}, \end{aligned} \quad (12.4)$$

называют *s-стадийным методом РК с использованием производных q-го порядка, явным*, если все  $a_{il}^{(r)}$  равны нулю при  $i \leq j$ , и *неявным* — в противном случае.

Так как  $Dx = 1$ ,  $D^l x = 0$  ( $l \geq 2$ ), то естественным обобщением условий (1.9) здесь будет

$$c_l = \sum_{i=1}^s a_i^{(1)}. \quad (12.5)$$

Определение 12.1 взято из работы Кастионгера и Ваннера (1972), но конкретные методы этого типа публиковались и ранее. В частности, такую структуру имеют очень удачные методы Фельберга (1958, 1964).

### Коллокационные методы

Естественным способом получения  $s$ -стадийных методов с производными до  $q$ -го порядка является использование идеи коллокации с *кратными узлами*. Другими словами, соотношение (7.18b) надо заменить на

$$\begin{aligned} u^{(1)}(x_0 + c_i h) &= (D^l y)(x_0 + c_i h, u(x_0 + c_i h)), \quad i = 1, \dots, s, \\ l &= 1, \dots, q_i, \end{aligned} \quad (12.6)$$

где  $u(x)$  — это многочлен степени  $q_1 + q_2 + \dots + q_s$ , а заданные целые числа  $q_1, \dots, q_s$  — «кратности» узлов  $c_1, \dots, c_s$ . Например, при  $q_1 = m, q_2 = \dots = q_s = 1$  получаются методы типа Фельберга.

Для обобщения результатов и идей разд. II.7 надо заменить использованную в теореме 7.6 интерполяцию по Лагранжу *интерполированием с кратными узлами* [Эрмит (1878): «Je me suis proposé de trouver un polynôme...»<sup>1)</sup>]. Дело в том, что формулу (12.6) можно интерпретировать как обычное условие коллокации, где узлы собраны в  $s$  групп, причем  $i$ -я группа содержит  $q_i$  «бесконечно» близких друг к другу узлов (теорема Ролля). Запишем формулу для интерполяционного многочлена Эрмита  $p(t)$  степени  $\sum q_j - 1$  в таком виде:

$$p(t) = \sum_{i=1}^s \sum_{r=1}^{q_j} \frac{1}{r!} l_{jr}(t) p^{(r-1)}(c_j). \quad (12.7)$$

Здесь «базисные» многочлены  $l_{jr}(t)$  степени  $\sum q_j - 1$  должны удовлетворять условиям

$$l_{jr}^{(k)}(c_i) = \begin{cases} r!, & \text{если } i = j \text{ и } k = r - 1, \\ 0 & \text{в остальных случаях.} \end{cases} \quad (12.8)$$

Их лучше всего получать из интерполяционной формулы Ньютона (с кратными узлами)<sup>2)</sup>. Как и в разд. II.7, применим теперь

<sup>1)</sup> «Я поставил себе целью найти многочлен...». — Франц.

<sup>2)</sup> См., например, Бахвалов (1975), гл. II, § 6. — Прим. ред.

к многочлену  $p(t) = hu'(x_0 + th)$  нашу интерполяционную формулу:

$$hu'(x_0 + th) = \sum_{l=1}^s \sum_{r=1}^{q_j} l_{lr}(t) k_l^{(r)}, \quad (12.9)$$

где

$$k_l^{(r)} = \frac{h^r}{r!} u^{(r)}(x_0 + c_l h). \quad (12.10)$$

Подставляя в (12.6) выражение

$$u(x_0 + c_i h) = y_0 + \int_0^{c_i} hu'(x_0 + th) dt$$

с учетом (12.9), легко убедиться, что доказана

**Теорема 12.2.** Коллокационный метод (12.6) эквивалентен НРК методу (12.4) с  $s$  стадиями и использованием  $q$ -х производных, у которого

$$a_{ij}^{(r)} = \int_0^{c_i} l_{jr}(t) dt, \quad b_i^{(r)} = \int_0^1 l_{jr}(t) dt. \quad ■ \quad (12.11)$$

Теоремы 7.7, 7.8 и 7.9 непосредственно обобщаются на случай «конфлюентных» (т. е. с совпадающими узлами) квадратурных формул. Таким образом, метод РК с  $q$ -ми производными имеет такой же порядок, как и соответствующая ему квадратурная формула

$$\int_0^1 p(t) dt \approx \sum_{j=1}^s \sum_{r=1}^{q_j} b_j^{(r)} p^{(r-1)}(c_j).$$

«Алгебраическое» доказательство этого результата более сложно (ср. упр. 7 из разд. II.7). Для случая  $q_j = q$  оно дано у Каствонгера и Ваннера (1972б).

Предположениям (7.12) соответствуют теперь следующие формулы:

$$\sum_{j=1}^s \sum_{r=1}^{q_j} a_{ij}^{(r)} \binom{\rho}{r} c_j^{\rho-r} = c_i^\rho, \quad \rho = 1, 2, \dots, \sum q_i. \quad (12.12)$$

После того как выбраны  $c_i$ , эти уравнения однозначно определяют значения  $a_{ij}^{(r)}$ , так как они образуют линейную систему с «конфлюентной» матрицей Вандермонда [см., например, Гаучи (1962)]. Формула (12.12) получается, если в (12.7) положить  $p(t) = t^{\rho-1}$ , а затем проинтегрировать от 0 до  $c_i$ .

**Примеры методов.** Квадратурные формулы «гауссовского» типа с кратными узлами существуют при нечетных  $q$  [Струд и Станку (1965)]. Им соответствуют неявные методы Рунге—Кутты с использованием производных  $q$ -го порядка [Кастлюнгер и Ваннер (1972б)]. При  $s = 1$ , конечно,  $c_1 = 1/2$ , что дает

$$b_1^{(2k)} = 0, \quad b_1^{(2k+1)} = 2^{-2k}, \quad a_{11}^{(k)} = -(-1)^k 2^{-k}.$$

Приведем также коэффициенты для случая  $s = 2$  и  $q_1 = q_2 = 3$ . Узлы  $c_i$  и веса  $b_i^{(k)}$  нашли Струд и Станку. Метод имеет восьмой порядок:

$$c_1 = 0.185394435825045, \quad c_2 = 1 - c_1,$$

$$b_1^{(1)} = 0.5, \quad b_2^{(1)} = b_1^{(1)},$$

$$b_1^{(2)}/2! = 0.0240729420844974, \quad b_2^{(2)} = -b_1^{(2)},$$

$$b_1^{(3)}/3! = 0.00366264960671727, \quad b_2^{(3)} = b_1^{(3)}.$$

$$(a_{ik}^{(1)}) = \begin{pmatrix} 0.201854115831005 & -0.0164596800059598 \\ 0.516459680005959 & 0.298145884168994 \end{pmatrix},$$

$$(a_{ik}^{(2)}) = \begin{pmatrix} -0.0223466569080541 & 0.00868878773082417 \\ 0.0568346718998190 & -0.0704925410770490 \end{pmatrix},$$

$$(a_{ik}^{(3)}) = \begin{pmatrix} 0.0116739668400997 & -0.00215351251065784 \\ 0.0241294101509615 & 0.0103019308002039 \end{pmatrix}.$$

### Методы Обрешкова

Мы рассмотрим теперь частный случай коллокационных методов с  $s = 2$ ,  $c_1 = 0$ ,  $c_2 = 1$ . Эти методы называются *методами Обрешкова*. Их коэффициенты могут быть получены в замкнутой форме посредством повторного интегрирования по частям [Нёрпсетт (1974а)].

**Лемма 12.3.** Пусть задано целое положительное число  $m$  и многочлен  $F(t)$  степени  $m$ . Тогда равенство

$$\sum_{l=0}^m h^l (D^l y)(x_1, y_1) F^{(m-l)}(0) = \sum_{l=0}^m h^l (D^l y)(x_0, y_0) F^{(m-l)}(1) \quad (12.13)$$

определяет метод со старшими производными, имеющий порядок  $m$ .

**Доказательство.** Обозначив  $y(x)$  точное решение дифференциального уравнения, будем исходить из соотношения

$$h^{m+1} \int_0^1 y^{(m+1)}(x_0 + ht) F(1-t) dt = O(h^{m+1}).$$

Преобразуем этот интеграл, применяя повторно интегрирование по частям до тех пор, пока не будут использованы все производные многочлена  $F(1-t)$ . Мы получим

$$\sum_{j=0}^m h^j y^{(j)}(x_1) F^{(m-j)}(0) = \sum_{j=0}^m h^j y^{(j)}(x_0) F^{(m-j)}(1) + O(h^{m+1}).$$

Вычитая это равенство из (12.13), мы найдем, что разность левых частей — величина порядка  $O(h^{m+1})$ . В силу теоремы о неявной функции это показывает, что равенство (12.13) определяет значение  $y_1$  с таким порядком точности, если (постоянное) значение  $F^{(m)}$  отлично от нуля. ■

**Замечание.** Мы взяли здесь для  $F$  аргумент  $1-t$  (вместо более естественного  $t$ ), чтобы избежать чередования знаков.

При удачном выборе многочлена  $F(t)$  большинство его производных при  $t=0$  и  $t=1$  должно, конечно, обращаться в нуль. Тогда метод (12.13), сохраняя порядок  $m$ , будет наиболее экономичным. Взяв

$$F(t) = \frac{t^k (t-1)^l}{(k+l)!},$$

мы получим методы, имеющие порядок  $m = k+l$ , которые обычно называют *методами Обрешкова*:

$$\begin{aligned} y_1 - \frac{l}{k+h} \cdot \frac{h}{1!} (Dy)(x_1, y_1) + \frac{l(l-1)}{(k+l)(k+l-1)} \cdot \frac{h^2}{2!} (D^2y)(x_1, y_1) \pm \\ \pm \cdots = y_0 + \frac{k}{k+l} \cdot \frac{h}{1!} (Dy)(x_0, y_0) + \\ + \frac{k(k-1)}{(k+l)(k+l-1)} \cdot \frac{h^2}{2!} (D^2y)(x_0, y_0) + \dots \end{aligned} \quad (12.14)$$

Здесь после члена  $l$ -й степени в левой части и после члена  $k$ -й степени — в правой все коэффициенты автоматически обращаются в нуль. Частными случаями этих методов являются:

$k=1, l=0$  — явный метод Эйлера,

$k=0, l=1$  — неявный метод Эйлера,

$k=1, l=1$  — правило трапеций.

### Методы Фельберга

Другой класс методов с высшими производными предложил Фельберг (1958, 1964). Его идея состояла в том, чтобы из решения задачи  $y' = f(x, y)$ ,  $y(x_0) = y_0$  вычесть  $m$  членов ряда Тейлора (см. разд. I.8)

$$y(x) := y(x) - \sum_{i=0}^m Y_i(x - x_0)^i \quad (12.15)$$

и получившееся дифференциальное уравнение  $\dot{y}'(x) = \dot{f}(x, \dot{y}(x))$ , где

$$\dot{f}(x, \dot{y}(x)) = \dot{f}\left(x, \dot{y} + \sum_{i=0}^m Y_i(x-x_0)^i\right) - \sum_{i=1}^m Y_i i(x-x_0)^{i-1}, \quad (12.16)$$

решать методом Рунге—Кутты. Зная, что решение уравнения (12.16) и его первые  $m$  производных обращаются в нуль при начальных условиях, можно надеяться достичь более высоких порядков.

Постараемся пояснить эту идею. Для упрощения обозначений уберем шляпки над буквами и будем считать, что преобразование (12.15) выполнено заранее. Разложим в ряд Тейлора решение неавтономного уравнения, как это сделано в разд. II.1. Тогда из (1.6) мы получим, что

$$f = 0,$$

$$f_x + f_y f = 0,$$

$$f_{xx} + 2f_{xy}f + f_{yy}f^2 + f_y(f_x + f_yf) = 0 \quad \text{и т. д.}$$

(см. также упр. 3 из разд. II.2). Из этих формул следует, в свою очередь, что  $f = 0$ ,  $f_x = 0, \dots, \frac{\partial^{m-1}f}{\partial x^{m-1}} = 0$ . Поэтому *обращаются в нуль* все элементарные дифференциалы порядков  $\leq m$  и большинство из них для более высоких порядков. Соответствующие условия порядка могут быть опущены. Первыми ненулевыми членами будут

$$\frac{\partial^m f}{\partial x^m} \quad \text{для порядка } m+1,$$

$$\frac{\partial^{m+1} f}{\partial x^{m+1}} \quad \text{и} \quad \dot{f}_y \cdot \frac{\partial^m f}{\partial x^m} \quad \text{для порядка } m+2,$$

и т. д. Им соответствуют условия порядка:

$$\sum_{i=1}^s b_i c_i^m = \frac{1}{m+1}$$

— для порядка  $m+1$ ,

$$\sum_{i=1}^s b_i c_i^{m+1} = \frac{1}{m+2} \quad \text{и} \quad \sum_{i,l} b_i a_{il} c_i^m = \frac{1}{(m+1)(m+2)}$$

— для порядка  $m+2$ , и т. д.

Условие  $\sum a_{il} = c_i$ , которое обычно позволяет сгруппировать некоторые члены в (1.6), становится ненужным, потому что все эти члены равны нулю.

Для полного понимания вопроса надо рассмотреть его с точки зрения применения *разделяющегося метода к разделяющейся системе* уравнений  $y' = f(x, y)$ ,  $x' = 1$ . Эти понятия разъясняются в разд. II.14 (см. рис. 14.4).

**Пример 12.4.** Для 3-стадийного метода решение системы уравнений, образованной четырьмя условиями для порядка  $m+3$ , указал Фельберг (1964). Выбор  $c_1 = c_3 = 1$  максимально упрощает вычисление выражения (12.16); при этом остальные коэффициенты определяются однозначно (см. табл. 12.1).

Таблица 12.1. Метод Фельберга порядка  $m+3$

1		$\theta = \frac{m+1}{m+3}$
θ	$\frac{\theta^m}{m+3}$	
1	$-\frac{1}{m+1}$	$\frac{2}{(m+1)\theta^m}$
	0	$\frac{m+3}{(m+1)(m+2)2\theta^m}$
	0	$\frac{1}{2m+4}$

В этой же работе Фельберга (1964) был выведен вложенный метод с двумя добавочными стадиями порядков  $m+3$  ( $m+4$ ). Все эти методы широко применялись для проведения научных расчетов в 60-х годах.

### Общая теория условий порядка

Из тех же соображений, что и в разд. II.2, мы предположим, что система (12.1) является автономной. Общая форма условий порядка для метода (12.4) была найдена в диссертации Каствонгера [Каствонгер и Ваннер (1972)]. Позже она была получена как простое следствие теоремы о композиции В-рядов [Хайрер и Ваннер (1974)]. Суть здесь в том, что, согласно теореме 2.6, выражение

$$\frac{h^i}{i!} (D^i y)(y_0) = \sum_{t \in LT, \rho(t)=i} \frac{h^i}{i!} F(t)(y_0) = B(\delta_i, y_0) \quad (12.17)$$

является В-рядом с коэффициентами

$$\delta_i(t) = \begin{cases} 1, & \text{если } \rho(t) = i, \\ 0 & \text{в остальных случаях.} \end{cases} \quad (12.18)$$

Таким образом, мы получим обобщение следствия 11.7

$$\frac{h^i}{i!} (D^i y)(B(a, y_0)) = B(a^{(i)}, y_0), \quad (12.19)$$

где, согласно формуле (11.8) с  $q = \rho(t)$ ,

$$\mathbf{a}^{(l)}(t) = \frac{1}{\alpha(t)} \binom{q}{i} \sum_{z \in d_i(t)} \prod_{z \in d_i(t)} \mathbf{a}(z) \quad (12.20)$$

(здесь сумма берется по всем  $\alpha(t)$  различным индексациям дерева  $t$ ). Это дает нам рекуррентные формулы для вычисления коэффициентов В-рядов, появляющихся в (12.4).

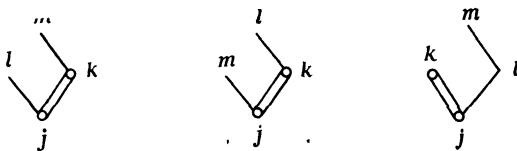


Рис. 12.1. Различные способы индексации дерева  $t_{42}$ .

**Пример 12.5.** Изображенное на рис. 12.1 дерево  $t_{42}$  допускает три различные индексации, из которых две порождают одно и то же разностное множество  $d_2(t)$ , так что формула (12.20) принимает вид

$$\mathbf{a}''(t_{42}) = 2(\mathbf{a}(\tau))^2 + \mathbf{a}(t_{21}). \quad (12.21)$$

Для всех остальных деревьев до четвертого порядка включительно  $\alpha(t) = 1$ . Итак, для вторых производных с помощью формулы (12.20) можно составить следующую таблицу:

$$\begin{aligned} \mathbf{a}''(\tau) &= 0, \\ \mathbf{a}''(t_{21}) &= 1, \\ \mathbf{a}''(t_{31}) &= 3\mathbf{a}(\tau), \\ \mathbf{a}''(t_{32}) &= 3\mathbf{a}(\tau), \\ \mathbf{a}''(t_{41}) &= 6(\mathbf{a}(\tau))^2, \\ \mathbf{a}''(t_{42}) &= 4(\mathbf{a}(\tau))^2 + 2\mathbf{a}(t_{21}), \\ \mathbf{a}''(t_{43}) &= 6(\mathbf{a}(\tau))^2, \\ \mathbf{a}''(t_{44}) &= 6\mathbf{a}(t_{21}) \end{aligned} \quad (12.22)$$

(обозначения деревьев см. в табл. 2.1).

Итак, пусть вид выражений  $a^{(l)}(t)$  установлен. Перепишем теперь формулы (12.4) в виде

$$k_t^{(l)} = \frac{h^l}{l!} (D^l y)(g_t),$$

$$g_t = y_0 + \sum_{r=1}^q \sum_{l=1}^s a_{rl}^{(r)} k_l^{(r)},$$

$$y_1 = y_0 + \sum_{r=1}^q \sum_{l=1}^s b_l^{(r)} k_l^{(r)}$$

и предположим, что выражения  $k_t^{(l)}$ ,  $g_t$ ,  $y_1$  являются В-рядами:

$$k_t^{(l)} = B(k_t^{(l)}, y_0), \quad g_t = B(g_t, y_0), \quad y_1 = B(y_1, y_0).$$

Тогда коэффициенты этих рядов определяются по рекуррентным формулам:

$$k_t^{(1)}(t) = \rho(t) g_t(t_1) \dots g_t(t_m), \quad k_t^{(1)}(\tau) = 1 \text{ (см. (11.17'))},$$

$$k_t^{(2)}(t) = g_t''(t) \text{ (см. (12.22))},$$

$$k_t^{(3)}(t) = g_t'''(t) \text{ (см. упр. 1 или упр. 2), и т. д.};$$

$$g_t(t) = \sum_{r=1}^q \sum_{l=1}^s a_{rl}^{(r)} k_l^{(r)}(t)$$

и, наконец,

$$y_1(t) = \sum_{r=1}^q \sum_{l=1}^s b_l^{(r)} k_l^{(r)}(t).$$

Метод (12.4), удовлетворяющий условиям (12.5), имеет порядок  $p$ , если, как обычно,

$$y_1(t) = 1 \text{ для всех } t, \text{ у которых } \rho(t) \leq p.$$

Дальнейшие подробности и конкретные методы приведены в статье Каствонгера и Ваннера (1972); см. также упр. 3.

### Упражнения

- Обобщите пример 12.5 и получите формулы для  $a^{(3)}(t)$  для всех деревьев до четвертого порядка включительно.
- (Каствонгер.) Докажите следующий вариант формулы (12.20), который позволяет обобщить соотношения (11.17') более не-

посредственным образом, а также может быть использован для получения формул из примера 12.5:

$$\mathbf{a}^{(l)}(t) = \frac{\rho(t)}{i} \sum_{\substack{\lambda_1 + \dots + \lambda_m = l-1 \\ \lambda_1, \dots, \lambda_m \geq 0}} a^{(\lambda_1)}(t_1) \dots a^{(\lambda_m)}(t_m),$$

если  $t = [t_1, \dots, t_m]$ .

**Указание:** см. статью Хайрера и Ваннера (1973), раздел 5.

3. Покажите, что следующие условия:

$$\sum_t b_t^{(1)} = 1,$$

$$2 \sum_t b_t^{(1)} c_t + \sum_t b_t^{(2)} = 1,$$

$$3 \sum_t b_t^{(1)} c_t^2 + 3 \sum_t b_t^{(2)} c_t = 1,$$

$$6 \sum_{t,l} b_t^{(1)} a_{tl}^{(1)} c_l + 3 \sum_t b_t^{(1)} e_t + 3 \sum_t b_t^{(2)} c_t = 1,$$

где

$$c_t = \sum_l a_{tl}^{(1)}, \quad e_t = \sum_l a_{tl}^{(2)},$$

обеспечивают третий порядок метода (12.4).

4. [Цурмюль (1952), Альбрехт (1955)]. Продифференцируйте систему дифференциальных уравнений первого порядка  $y' = f(x, y)$ . Вы придетете к задаче Коши

$$y'' = (D^2 y)(x, y), \quad y(x_0) = y_0, \quad y'(x_0) = f_0.$$

Примените к этой задаче специальный метод для систем дифференциальных уравнений высших порядков (см. следующий раздел). Вы получите метод, использующий высшие производные. Покажите, что нижеследующий метод имеет шестой порядок:

$$k_1 = h^2 g(x_0, y_0),$$

$$k_2 = h^2 g\left(x_0 + \frac{h}{4}, y_0 + \frac{h}{4} f_0 + \frac{1}{32} k_1\right),$$

$$k_3 = h^2 g\left(x_0 + \frac{h}{2}, y_0 + \frac{h}{2} f_0 + \frac{1}{24} (-k_1 + 4k_2)\right),$$

$$k_4 = h^2 g\left(x_0 + \frac{3h}{4}, y_0 + \frac{3h}{4} f_0 + \frac{1}{32} (3k_1 + 4k_2 + 2k_3)\right),$$

$$y_1 = y_0 + h f_0 + \frac{1}{90} (7k_1 + 24k_2 + 6k_3 + 8k_4).$$

Здесь

$$g(x, y) = (D^2 y)(x, y) = Df(x, y) = f_x(x, y) + f_y(x, y) f(x, y).$$

## II.13. Численные методы для дифференциальных уравнений второго порядка

«Mutationem motus proportionalem esse vi motrici impressae».  
 «Изменение движения пропорционально приложенной силе».  
 (Ньютона, 1687)

Многие встречающиеся на практике дифференциальные уравнения представляют собой системы уравнений второго порядка:

$$y'' = f(x, y, y'). \quad (13.1)$$

Главным образом это связано с тем, что силы пропорциональны ускорениям, т. е. вторым производным. Как уже упоминалось в разд. I.1, такая система может быть преобразована в систему дифференциальных уравнений первого порядка, но удвоенной размерности; для этого в качестве новой зависимой переменной рассматривают вектор  $(y, y')^T$ :

$$\begin{pmatrix} y \\ y' \end{pmatrix}' = \begin{pmatrix} y' \\ f(x, y, y') \end{pmatrix}, \quad \begin{aligned} y(x_0) &= y_0, \\ y'(x_0) &= y'_0. \end{aligned} \quad (13.2)$$

Чтобы получить численное решение уравнения (13.1), можно применить, например, метод Рунге—Кутты к системе (13.2). Это дает

$$\begin{aligned} k_i &= y'_0 + h \sum_l a_{il} k'_l, \\ k'_i &= f\left(x_0 + c_i h, y_0 + h \sum_l a_{il} k_l, y'_0 + h \sum_l a_{il} k'_l\right), \\ y_1 &= y_0 + h \sum_l b_i k_l, \\ y'_1 &= y'_0 + h \sum_l b_i k'_l. \end{aligned} \quad (13.3)$$

Подставив первую формулу (13.3) в остальные, мы получим

$$\begin{aligned} k'_i &= f\left(x_0 + c_i h, y_0 + c_i h y'_0 + h^2 \sum_l \bar{a}_{il} k'_l, y'_0 + h \sum_l a_{il} k'_l\right), \\ y_1 &= y_0 + h y'_0 + h^2 \sum_l \bar{b}_i k'_l, \\ y'_1 &= y'_0 + h \sum_l b_i k'_l, \end{aligned} \quad (13.4)$$

где

$$\bar{a}_{ij} = \sum_k a_{ik} a_{kj}, \quad \bar{b}_i = \sum_l b_j a_{ji}. \quad (13.5)$$

Для программирования представление (13.4) предпочтительнее, чем (13.3), потому что оно требует почти вдвое меньше машинной памяти. Это может оказаться важным преимуществом, в частности если система уравнений (13.1) имеет большую размерность.

### Методы Нюстрёма

Р. Х. Мерсон: «...Я не видел статьи Нюстрёма. Была она на английском?»

Дж. Беннетт: «Нет, нет, не на финском, а на немецком».

(1957)

Э. И. Нюстрём (1925) был первым, кто рассмотрел методы вида (13.4), в которых коэффициенты не обязательно удовлетворяют соотношениям (13.5) («Da bis jetzt die direkte Anwendung der Runge'schen Methode auf den wichtigen Fall von Differentialgleichungen zweiter Ordnung nicht behandelt war, ...» Nyström, 1925.)<sup>1)</sup> Прямые методы подобного рода получили название методов Нюстрёма.

**Определение 13.1.** Метод Нюстрёма (13.4) имеет порядок  $p$ , если для достаточно гладких задач вида (13.1)

$$y(x_0 + h) - y_1 = O(h^{p+1}), \quad (13.6)$$

$$y'(x_0 + h) - y'_1 = O(h^{p+1}). \quad (13.7)$$

Пример метода Нюстрёма, где условия (13.5) не выполняются, приведен в табл. 13.1. Нюстрём утверждал, что этот метод применять проще, чем «метод Рунге—Кутты», и что он уменьшает

Таблица 13.1. Метод Нюстрёма четвертого порядка

	0							
$c_i$	$\frac{1}{2}$	$\frac{1}{8}$				$\frac{1}{2}$		
	$\frac{1}{2}$	$\frac{1}{8}$	0	$\leftarrow \bar{a}_{ij}$		0	$\frac{1}{2}$	$\leftarrow a_{ij}$
	1	0	0	$\frac{1}{2}$		0	0	1
$b_i \rightarrow$		$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	0	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{2}{6}$
						$\frac{1}{6}$	$\frac{2}{6}$	$\frac{1}{6}$

<sup>1)</sup> «Так как до сих пор *прямое* применение метода Рунге к важному случаю дифференциальных уравнений второго порядка не рассматривалось ...». — Нюстрём, 1925. — *Nem.*

объем вычислений примерно на 25 %. Это, конечно, неверно, если метод Рунге—Кутты применяется в виде (13.4) (см. также упр. 2).

Действительное улучшение может быть получено в том случае, когда правая часть уравнения (13.1) не зависит от  $y'$ :

$$y'' = f(x, y). \quad (13.8)$$

Тогда метод Ньюстрёма принимает вид

$$k'_i = f\left(x_0 + c_i h, y_0 + c_i hy'_0 + h^2 \sum_l \bar{a}_{il} k'_l\right),$$

$$y_1 = y_0 + hy'_0 + h^2 \sum_i b_i k'_i, \quad (13.9)$$

$$y'_1 = y'_0 + h \sum_i b_i k'_i,$$

и коэффициенты  $a_{ij}$  больше не нужны. Примеры приведены в табл. 13.2. Обратите внимание, что метод пятого порядка здесь

Таблица 13.2. Методы для уравнения  $y'' = f(x, y)$

Ньюстрём, порядок 4			Ньюстрём, порядок 5		
$c_i$	0		$\bar{a}_{ij}$	0	
	$\frac{1}{2}$	$\frac{1}{8}$		$\frac{1}{5}$	$\frac{1}{50}$
	1	0 $\frac{1}{2}$		$\frac{2}{3}$	$\frac{-1}{27} \quad \frac{7}{27}$
				1	$\frac{3}{10} \quad \frac{-2}{35} \quad \frac{9}{35}$
$b_i$	$\frac{1}{6} \quad \frac{1}{3} \quad 0$		$b_i$	$\frac{14}{336} \quad \frac{100}{336} \quad \frac{54}{336} \quad 0$	
$b_i$	$\frac{1}{6} \quad \frac{4}{6} \quad \frac{1}{6}$		$b_i$	$\frac{14}{336} \quad \frac{125}{336} \quad \frac{162}{336} \quad \frac{35}{336}$	

требует только четырех вычислений  $f$ . Это — существенное улучшение по сравнению с методами Рунге—Кутты, которые требуют не менее шести вычислений  $f$  (теорема 6.1).

Нашей ближайшей целью будет вывод условий порядка для методов Ньюстрёма. Для этого мы обобщим теорию, развитую в разд. II.2, на дифференциальные уравнения второго порядка [Хайрер и Ваннер (1976)].

## Производные точного решения

Как и для уравнений первого порядка, мы можем ограничиться системами автономных дифференциальных уравнений

$$(y^J)'' = f^J(y^1, \dots, y^n, y^1, \dots, y^n) \quad (13.11)$$

(добавив, если надо, уравнение  $x'' = 0$ ). Верхний индекс  $J$  обозначает  $J$ -ю компоненту соответствующего вектора. Вычислим теперь производные точного решения уравнения (13.11). Вторую производную дает само уравнение (13.11):

$$(y^J)'' = f^J(y, y'). \quad (13.12; 2)$$

Повторное дифференцирование этого уравнения с использованием (13.11) дает

$$(y^J)^{(3)} = \sum_K \frac{\partial f^J}{\partial y^K}(y, y') \cdot y'^K + \sum_K \frac{\partial f^J}{\partial y'^K}(y, y') f^K(y, y'), \quad (13.12; 3)$$

$$(y^J)^{(4)} = \sum_{K, L} \frac{\partial^2 f^J}{\partial y^K \partial y^L}(y, y') \cdot y'^K \cdot y'^L + \quad (13.12; 4)$$

$$+ \sum_{K, L} \frac{\partial^2 f^J}{\partial y^K \partial y'^L}(y, y') \cdot y'^K \cdot f^L(y, y') +$$

$$+ \sum_K \frac{\partial f^J}{\partial y^K}(y, y') f^K(y, y') +$$

$$+ \sum_{K, L} \frac{\partial^2 f^J}{\partial y'^K \partial y^L}(y, y') f^K(y, y') \cdot y'^L +$$

$$+ \sum_{K, L} \frac{\partial^2 f^J}{\partial y'^K \partial y'^L}(y, y') f^K(y, y') f^L(y, y') +$$

$$+ \sum_{K, L} \frac{\partial f^J}{\partial y'^K}(y, y') \frac{\partial f^K}{\partial y^L}(y, y') \cdot y'^L +$$

$$+ \sum_{K, L} \frac{\partial f^J}{\partial y'^K}(y, y') \frac{\partial f^K}{\partial y'^L}(y, y') f^L(y, y').$$

Продолжение этого процесса становится еще более сложным, чем для дифференциальных уравнений первого порядка. Поэтому графическое представление приведенных выше формул будет очень полезным. Чтобы различать производные по  $y$  и по  $y'$ , нам понадобятся два рода вершин: «тонкие» и «толстые». На рис. 13.1 показаны графы, соответствующие приведенным выше формулам.

Дадим теперь точное определение этих графов.

**Определение 13.2.** Помеченным  $N$ -деревом порядка  $q$  назовем пару  $(t, t')$ , состоящую из помеченного дерева

$$t: A_q - \{j\} \rightarrow A_q$$

(см. определение 2.2) и отображения

$$t': A_q \rightarrow \{\text{«тонкая»}, \text{«толстая» (вершина)}\},$$

которая удовлетворяет следующим условиям:

а) корень  $t$  — всегда «толстая» вершина, т. е.

$$t'(j) = \text{«толстая»};$$

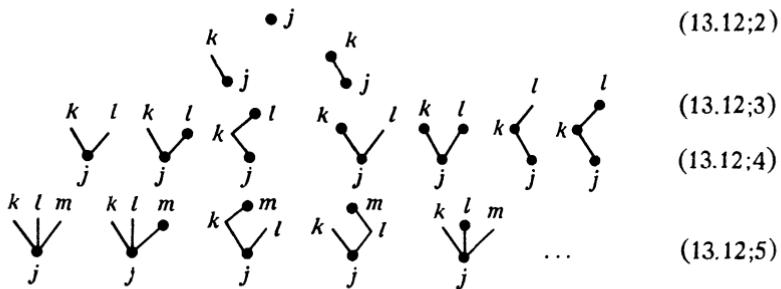


Рис. 13.1. Производные точного решения.

б) любая «тонкая» вершина может иметь не более одного «сына», который должен быть «толстым».

Множество всех помеченных  $N$ -деревьев порядка  $q$  будем обозначать  $LNT_q$ .

Условие б) в определении 13.2 мотивируется тем, что все производные функции  $g(y, y') = y'$  тождественно обращаются в нуль, за исключением первой производной по  $y'$ .

В дальнейшем мы будем называть *конечной вершиной* всякую вершину, не имеющую «сыновей», и там, где не может возникнуть недоразумения, будем обозначать помеченное  $N$ -дерево просто  $t$  вместо  $(t, t')$ .

**Определение 13.3.** Для помеченного  $N$ -дерева  $t$  будем обозначать через

$$F^J(t)(y, y')$$

выражение, которое представляет собой *сумму* по всем индексам «толстых» вершин  $t$ , кроме  $j$  (индекса корня), и по всем индексам «тонких» конечных вершин. *Общий член* этой суммы является произведением выражений

$$\frac{\partial^r f^K}{\partial y^L \dots \partial y'^M \dots}(y, y') \quad \text{и} \quad y'^K.$$

Множитель первого типа соответствует каждой толстой вершине  $k$ , толстые «сыновья» которой имеют индексы  $m, \dots$ , а индексы  $l, \dots$  принадлежат либо ее тонким «сыновьям» (если они являются конечными вершинами), либо (толстым) «сыновьям» этих тонких «сыновей» вершины  $k$  (и тогда индекс промежуточной тонкой вершины исчезает из формулы); случаю  $r = 0$  соответствуют, очевидно, конечные толстые вершины. Множитель  $y'^K$  соответствует тонкой конечной вершине с индексом  $k$ . Вектор  $F(t)(y, y')$  мы опять будем называть *элементарным дифференциалом*.

Примеры можно найти в табл. 13.3. Еще раз подчеркнем, что индексы тонких вершин, не являющихся конечными, не играют никакой роли в предыдущем определении. Далее, по аналогии с определением 2.4 дадим

Таблица 13.3. SN-деревья, элементарные дифференциалы и условия порядка

$t$	$t$	$\rho(t)$	$\alpha(t)$	$\gamma(t)$	$F^J(t)(y, y')$	$\Phi_j(t)$
$t_1$		1	1	1	$f^J$	1
$t_2$		2	1	2	$\sum_K f_K^J y'^K$	$c_j$
$t_3$		3	1	3	$\sum_{K,L} f_{KL}^J y'^K y'^L$	$c_j^2$
$t_4$		3	1	6	$\sum_L f_L^J f^L$	$\sum_l \bar{a}_{jl}$
$t_5$		4	1	4	$\sum_{K,L,M} f_{KLM}^J y'^K y'^L y'^M$	$c_j^3$
$t_6$		4	3	8	$\sum_{L,M} f_{LM}^J y'^L f^M$	$\sum_m c_j \bar{a}_{jm}$
$t_7$		4	1	24	$\sum_{L,M} f_L^J f_M^L y'^M$	$\sum_l \bar{a}_{jl} c_l$
$t_8$		5	1	5	$\sum_{K,L,M,P} f_{KLMP}^J y'^K y'^L y'^M y'^P$	$c_j^4$
$t_9$		5	6	10	$\sum_{L,M,P} f_{LMP}^J y'^L y'^M f^P$	$\sum_p c_j^2 \bar{a}_{jt}$
$t_{10}$		5	3	20	$\sum_{L,P} f_L^J f^L f^P$	$\sum_l \bar{a}_{ju} \bar{a}_{jp}$
$t_{11}$		5	4	30	$\sum_{L,P,M} f_{LP}^J f_M^L y'^M y'^P$	$\sum_l c_j \bar{a}_{ju} c_l$
$t_{12}$		5	1	60	$\sum_{L,P,M} f_L^J f_M^L f_P^M y'^M y'^P$	$\sum_l \bar{a}_{ju} c_l^2$
$t_{13}$		5	1	120	$\sum_{L,P} f_L^J f_P^L f^P$	$\sum_{l,p} \bar{a}_{ju} \bar{a}_{lp}$

**Определение 13.4.** Два помеченных N-дерева  $(t, t')$  и  $(u, u')$  **эквивалентны**, если они отличаются только перестановкой их индексов, т. е. если они имеют одинаковый порядок  $q$ , и существует взаимно однозначное отображение  $\sigma: A_q \rightarrow A_q$ , такое что  $\sigma(j) = j$ ,  $t\sigma = \sigma u$  на  $A_q - \{j\}$  и  $t'\sigma = u'$ .

Например, второе и четвертое помеченные N-деревья в формуле (13.12; 4) на рис. 13.1 эквивалентны; эквивалентны также второе и пятое деревья в формуле (13.12; 5).

**Определение 13.5.** Класс эквивалентности помеченных N-деревьев  $q$ -го порядка называется **N-деревом порядка  $q$** . Множество всех N-деревьев порядка  $q$  обозначается  $NT_q$ . Далее, будем обозначать через  $\alpha(t)$  число элементов в классе эквивалентности  $t$ , т. е. число различных возможных монотонных индексаций  $t$  (ср. определение 2.5).

Представители N-деревьев до пятого порядка включительно приведены в табл. 13.3. Теперь мы в состоянии записать в замкнутом виде формулу для производных точного решения системы (13.11).

**Теорема 13.6.** Для точного решения уравнения (13.11) справедлива формула

$$(y^J)^{(q)} = \sum_{t \in LNT_{q-1}} F(t)(y, y') = \sum_{t \in NT_{q-1}} \alpha(t) F(t)(y, y'). \quad (13.12; q)$$

*Доказательство.* Общая формула здесь получается, как и в разд. II.2, путем продолжения вычислений, которые привели к формулам (13.12; 2—4). ■

### Производные численного решения

Запишем формулы метода (13.4) несколько иначе:

$$\begin{aligned} g_i &= y_0 + c_i h y'_0 + \sum_l \bar{a}_{il} h^2 f(g_l, g'_l), \\ g'_i &= y'_0 + \sum_l a_{il} h f(g_l, g'_l), \\ g_l &= y_0 + h y'_0 + \sum_i b_{il} h^2 f(g_i, g'_i), \\ g'_l &= y'_0 + \sum_i b_{il} h f(g_i, g'_i). \end{aligned} \quad (13.13)$$

Теперь уравнения для промежуточных величин  $g_i$ ,  $g'_i$  имеют такой же вид, как и для  $y_1$ ,  $y'_1$ . В (13.13) входят выражения вида  $h^2 \varphi(h)$  и  $h \varphi(h)$ . Поэтому, кроме формулы (2.4), нам понадобится еще формула

$$(h^2 \varphi(h))^{(q)}|_{h=0} = q(q-1)(\varphi(h))^{(q-2)}|_{h=0}. \quad (13.14)$$

Будем теперь последовательным дифференцированием вычислять значения производных  $g_i^J$  и  $g'_i^J$  при  $h = 0$ :

$$(g_i^J)^{(1)}|_{h=0} = c_i y_0^J, \quad (13.15; 1)$$

$$(g'_i^J)^{(1)}|_{h=0} = \sum_l a_{il} f^J|_{y_0, y'_0}, \quad (13.16; 1)$$

$$(g_2^J)^{(2)}|_{h=0} = 2 \sum_l \bar{a}_{il} f^J|_{y_0, y'_0}. \quad (13.15; 2)$$

Теперь для дальнейшего дифференцирования нужна формула

$$(f^J(g_l, g'_l))^{(1)} = \sum_K \frac{\partial f^J}{\partial y^K} (g_l, g'_l) (g_l^K)^{(1)} + \sum_K \frac{\partial f^J}{\partial y'^K} (g_l, g'_l) (g_l^K)^{(1)}, \quad (13.17)$$

с помощью которой получаем

$$\begin{aligned} (g'_i^J)^{(2)}|_{h=0} &= 2 \sum_l a_{il} c_l \sum_K \frac{\partial f^J}{\partial y^K} \cdot y'^K|_{y_0, y'_0} + \\ &+ 2 \sum_{l, k} a_{il} a_{lk} \sum_K \frac{\partial f^J}{\partial y'^K} \cdot f^K|_{y_0, y'_0}, \end{aligned} \quad (13.16; 2)$$

$$\begin{aligned} (g_i^J)^{(3)}|_{h=0} &= 3 \cdot 2 \sum_l \bar{a}_{il} c_l \sum_K \frac{\partial f^J}{\partial y^K} \cdot y'^K|_{y_0, y'_0} + \\ &+ 3 \cdot 2 \sum_{l, k} \bar{a}_{il} a_{lk} \sum_K \frac{\partial f^J}{\partial y'^K} \cdot f^K|_{y_0, y'_0}. \end{aligned} \quad (13.15; 3)$$

Чтобы записать общую формулу, нам потребуется

**Определение 13.7.** Для помеченного N-дерева  $t$  обозначим через  $\Phi_j(t)$  выражение, являющееся суммой по индексам всех толстых вершин  $t$ , кроме  $j$  (индекса корня); общий член этой суммы представляет собой произведение множителей следующего вида:

$a_{kl}$ , если толстая вершина  $k$  имеет толстого «сына»  $l$ ;

$\bar{a}_{kl}$ , если толстая вершина  $k$  соединена с  $l$  через своего тонкого «сына», и

$c_k^m$ , если толстая вершина  $k$  соединена с  $m$  тонкими конечными вершинами.

**Теорема 13.8.** Для производных функций  $g_i, g'_i$  из формул метода (13.13) имеют место равенства

$$(g_i^J)^{(q+1)}|_{h=0} = (q+1) \sum_{t \in LNT_q} \gamma(t) \sum_l \bar{a}_{il} \Phi_l(t) F^J(t)(y_0, y'_0), \quad (13.15; q+1)$$

$$(g_i^J)^{(q)}|_{h=0} = \sum_{t \in LNT_q} \gamma(t) \sum_l \Phi_l(t) F^J(t)(y_0, \dot{y_0}), \quad (13.16; q)$$

где величина  $\gamma(t)$  вычисляется согласно определению 2.10.

**Доказательство.** Для небольших значений  $q$  эти формулы получены выше; в общем случае они доказываются, как в теореме 2.11. Система уравнений (13.2) является частным случаем так называемых *разделяющихся систем* (см. разд. II.14, где теорема 13.8 получает новое освещение). ■

Ввиду подобия формул для  $y_1, \dot{y}_1$  формулам для  $g_i, g'_i$  справедлива также следующая

**Теорема 13.9.** Производные определяемого методом (13.13) численного решения  $y_1, \dot{y}_1$  удовлетворяют следующим равенствам:

$$(y_1^J)^{(q)}|_{h=0} = q \sum_{t \in LNT_{q-1}} \gamma(t) \sum_l b_l \Phi_l(t) F^J(t)(y_0, \dot{y_0}), \quad (13.18; q)$$

$$(y_1^J)^{(q-1)}|_{h=0} = \sum_{t \in LNT_{q-1}} \gamma(t) \sum_l b_l \Phi_l(t) F^J(t)(y_0, \dot{y_0}). \quad (13.19; q-1)$$

### Условия порядка

Порядок метода Ньюстрёма (определение 13.1) исследуется путем сравнения разложений Тейлора для  $y_1, \dot{y}_1$  с такими же разложениями для точного решения  $y(x_0 + h), \dot{y}(x_0 + h)$ .

**Теорема 13.10.** Чтобы метод Ньюстрёма (13.4) имел порядок  $p$ , необходимы и достаточны следующие условия:

$$\sum_i b_i \Phi_i(t) = \frac{1}{(\rho(t) + 1) \cdot \gamma(t)} \text{ для } N\text{-деревьев } t \text{ с } \rho(t) \leq p-1, \quad (13.20)$$

$$\sum_i b_i \Phi_i(t) = \frac{1}{\gamma(t)} \text{ для } N\text{-деревьев } t \text{ с } \rho(t) \leq p. \quad (13.21)$$

Здесь  $\rho(t)$  обозначает порядок  $N$ -дерева  $t$ , величина  $\Phi_i(t)$  задана определением 13.7, а  $\gamma(t)$  — формулой (2.17).

**Доказательство.** Достаточность вытекает непосредственно из теорем 13.6 и 13.9. Необходимость может быть доказана тем же способом, что и для уравнений первого порядка (ср. упр. 4 из разд. II.2). ■

Обсудим вкратце вопрос, может ли дополнительная свобода выбора параметров метода (13.4) (полученная отказом от предположений (13.5)) привести к его существенному улучшению. Поскольку условия порядка для методов РК (теорема 2.13) являются подмножеством условий (13.21) (см. ниже упр. 3), то увеличить

порядок метода эта дополнительная свобода не позволяет. Можно лишь уменьшить некоторые (но не все) коэффициенты погрешности. Поэтому мы обратимся к методам Нюстрёма (13.9) для дифференциальных уравнений второго порядка специального вида (13.8).

Чтобы исследовать условия порядка для методов (13.9), запишем уравнение (13.8) в автономной форме:

$$y'' = f(y). \quad (13.22)$$

Специальный вид этого уравнения имеет следствием тождественное обращение в нуль тех элементарных дифференциалов, которые содержат производные по  $y'$ . Следовательно, надо рассмотреть лишь подмножество  $N$ -деревьев, имеющих следующий специальный вид:

**Определение 13.11.**  $N$ -дерево  $t$  называется *специальным  $N$ -деревом* ( $SN$ -деревом), если его толстые вершины имеют только тонких «сыновей».

**Теорема 13.12.** *Метод Нюстрёма (13.9) для дифференциального уравнения специального вида (13.8) имеет порядок  $p$  в том и только том случае, если*

$$\sum_i b_i \Phi_i(t) = \frac{1}{(\rho(t) + 1)\gamma(t)} \text{ для } SN\text{-деревьев } t \text{ с } \rho(t) \leq p-1, \quad (13.23)$$

$$\sum_i b_i \Phi_i(t) = \frac{1}{\gamma(t)} \quad \text{для } SN\text{-деревьев } t \text{ с } \rho(t) \leq p. \quad (13.24)$$

■

Все  $SN$ -деревья до пятого порядка включительно вместе с соответствующими элементарными дифференциалами, а также необходимыми для записи условий порядка выражениями  $\Phi_i$  и значениями  $\rho$ ,  $\alpha$  и  $\gamma$  представлены в табл. 13.3.

## О конструировании методов Нюстрёма

При конструировании методов Нюстрёма полезно использовать рассмотренные ниже упрощающие предположения.

**Лемма 13.13.** *Условие (13.23) становится следствием (13.24), если предположить, что выполнены равенства*

$$b_i = b_i(1 - c_i), \quad i = 1, \dots, s. \quad (13.25)$$

**Доказательство.** Пусть  $t$  является  $SN$ -деревом порядка не выше  $p-1$ . Обозначим  $u$   $SN$ -дерево порядка  $\rho(t)+1$ , которое

получается из  $t$  добавлением новой ветви с тонкой вершиной, исходящей от корня  $t$ . Согласно определению 13.7,  $\Phi_t(u) = c_i \Phi_i(t)$ , а из формулы (2.17) следует, что  $\gamma(u) = (\rho(t) + 1) \gamma(t)/\rho(t)$ . Заключение леммы следует из того, что

$$\sum_i b_i \Phi_i(t) = \sum_i b_i \Phi_i(t) - \sum_i b_i \Phi_i(u) = \frac{1}{\gamma(t)} - \frac{1}{\gamma(u)} = \frac{1}{(\rho(t) + 1) \gamma(t)},$$

где мы воспользовались предположением (13.25). ■

**Лемма 13.14.** Пусть  $t$  и  $u$  — два SN-дерева, изображенные на рис. 13.2 (предполагается, что очерченные кругами их части идентичны). Условия порядка для  $t$  и  $u$  одинаковы, если предположить выполнение равенств

$$\sum_i \bar{a}_{ij} = \frac{c_j^2}{2}, \quad i = 1, \dots, s. \quad (13.26)$$



Рис. 13.2. Деревья из леммы 13.14.

**Доказательство.** Из определения 13.7 и формулы (13.26) следует, что  $\Phi_t(t) = \Phi_t(u)/2$ , а из формулы (2.17) получаем, что  $\gamma(t) = 2\gamma(u)$ . Таким образом, условия порядка для  $t$  и  $u$  совпадают. ■

Условия (13.25) позволяют нам освободиться от уравнений (13.23), а роль условий (13.26) аналогична той роли, которую условия (1.9) играли для методов Рунге—Кутты. Условия (13.26) выражают то обстоятельство, что величины  $g_i$  в формулах (13.13) аппроксимируют  $y(x_0 + c_i h)$  с точностью до членов  $O(h^3)$ . Из леммы 13.14 следует, что могут быть опущены все SN-деревья, имеющие хотя бы одну толстую конечную вершину (т. е. деревья  $t_4, t_6, t_9, t_{10}, t_{13}$  в табл. 13.3).

С помощью упрощающих предположений (13.25) и (13.26) теперь легко построить явные методы Ньюстрема пятого порядка при  $s = 4$ . Условия порядка для деревьев  $t_1, t_2, t_3, t_5$  и  $t_8$  показывают, что квадратурная формула с узлами  $c_1 = 0, c_2, c_3, c_4$  и весами  $b_1, b_2, b_3, b_4$  имеет пятый порядок. Следовательно, узлы  $c_i$  должны удовлетворять соотношению ортогональности

$$\int_0^1 x(x - c_2)(x - c_3)(x - c_4) dx = 0,$$

и мы видим, что при выборе квадратурной формулы все еще остаются две степени свободы. После этого выбора коэффициенты  $\bar{a}_{ij}$  определяются уже однозначно, и их можно вычислить следующим образом. Значение  $\bar{a}_{21}$  определяется из (13.26) при  $i = 2$ . Условия порядка для  $t_7$  и  $t_{11}$  образуют систему из двух линейных уравнений с неизвестными

$$\sum_l \bar{a}_{8j} c_j \quad \text{и} \quad \sum_l \bar{a}_{4j} c_j.$$

Затем с помощью (13.26) при  $i = 3$  находим  $\bar{a}_{31}$  и  $\bar{a}_{32}$ . Наконец, условие порядка для  $t_{12}$  позволяет определить величину  $\sum_j \bar{a}_{4j} c_j^q$ , и оставшиеся коэффициенты  $\bar{a}_{41}$ ,  $\bar{a}_{42}$ ,  $\bar{a}_{43}$  могут быть вычислены из линейной системы уравнений с матрицей типа матрицы Вандермонда. Метод из табл. 13.2 был получен именно таким путем.

При конструировании методов еще более высоких порядков полезно использовать дальнейшие упрощающие предположения, например равенство

$$\sum_l \bar{a}_{ij} c_l^q = \frac{c_l^{q+2}}{(q+2)(q+1)}, \quad (13.27)$$

которое при  $q = 0$  сводится к (13.26), а также равенство

$$\sum_i b_i c_i^q \bar{a}_{ij} = b_j \frac{c_j^{q+2}}{(q+2)(q+1)} - \frac{c_j}{q+1} + \frac{1}{q+2}, \quad (13.28)$$

которое можно рассматривать как обобщение условия (1.12). Дальнейшие подробности читатель найдет в ряде работ, где представлены методы Ньюстрёма высших порядков [Хайрер и Ваннер (1976), а также Альбрехт (1955), Баттин (1976), Беентьев и Герритсен (1976), Хайрер (1977, 1982)].

### Глобальная сходимость

Вводя переменную  $z_n = (y_n, y'_n)^T$ , метод Ньюстрёма (13.4) можно записать в такой форме:

$$z_1 = z_0 + h\Phi(x_0, z_0, h), \quad (13.29)$$

где

$$\Phi(x_0, z_0, h) = \begin{pmatrix} y'_0 + h \sum_i b_i k'_i \\ \sum_i b_i k'_i \end{pmatrix}.$$

Формула (13.29) представляет специальный одношаговый метод для дифференциального уравнения (13.2). Для метода Ньюстрёма

$p$ -го порядка локальная погрешность  $(y(x_0 + h) - y_1, y'(x_0 + h) - y'_1)^T$  может быть оценена сверху величиной  $Ch^{p+1}$  (см. определение 13.1), что согласуется с формулой (3.24). Следовательно, теоремы о сходимости из разд. II.3, а также полученные в разд. II.8 результаты об асимптотических разложениях глобальной погрешности и здесь остаются в силе.

### Программная реализация методов Ньюстрёма

Для того, чтобы программа была эффективной, необходим механизм управления длиной шага. Его можно организовать так же, как и для методов РК (см. разд. II.4). Для оценки локальной погрешности можно применять экстраполяцию по Ричардсону или конструировать вложенные методы Ньюстрёма.

Ряд вложенных методов Ньюстрёма построил Фельберг (1972). В этих методах для управления длиной шага используется аппроксимация  $y(x_0 + h)$  с  $(p + 1)$ -м порядком точности. Аппроксимация  $y'(x_0 + h)$  с точностью до  $(p + 1)$ -го порядка не требуется, так как для пошагового продолжения решения применяется аппроксимация более низкого порядка.

Как и для дифференциальных уравнений первого порядка, здесь оказалась лучше другая стратегия — использовать аппроксимацию более высокого порядка для пошагового продолжения решения. Дорман и Принс (1978) сконструировали метод порядков 7 (6), коэффициенты которого приведены в табл. 13.4. Методы такого же типа, но более высоких порядков опубликовали недавно Филиппи и Грэф (1986).

### Экстраполяционный метод для уравнения $y'' = f(x, y)$ .

Les calculs originaux, comprenant environ 3.000 pages in-folio avec 358 grandes planches, et encore 3.800 pages de développements mathématiques correspondants, appartiennent maintenant à la collection de manuscrits de la Bibliothèque de l'Université, Christiania.

Оригинальные вычисления, содержащие около 3000 страниц ин-фолио с 358 большими таблицами, и еще 3800 страниц соответствующих математических выкладок принадлежат теперь коллекции рукописей библиотеки университета Христиании.

(Штёрмер, 1921)

Переписав дифференциальное уравнение (13.8) в виде системы уравнений первого порядка

$$\begin{pmatrix} y \\ y' \end{pmatrix}' = \begin{pmatrix} y' \\ f(x, y) \end{pmatrix}, \quad \begin{pmatrix} y \\ y' \end{pmatrix}(x_0) = \begin{pmatrix} y_0 \\ y'_0 \end{pmatrix}, \quad (13.30)$$

Таблица 13.4. Метод Дормана—Принса порядка 7 (6)

$c_i$	$a_{ii}$	$\hat{a}_i$	$b_i$	$\hat{b}_i$
0				
		$\frac{1}{20}$	$\frac{1}{20}$	$\frac{1}{20}$
$\frac{1}{10}$	$\frac{1}{200}$		0	0
$\frac{1}{5}$	$\frac{1}{150}$		0	0
$\frac{3}{8}$	$\frac{171}{8192}$	$\frac{315}{8192}$	0	0
$\frac{1}{2}$	$\frac{5}{288}$	$\frac{25}{528}$	$\frac{8}{45}$	$\frac{16}{45}$
$\frac{7-\sqrt{21}}{14}$	$\frac{1003-205\sqrt{21}}{90552}(751)$ $+113\sqrt{21})$	$\frac{25}{43218}-\frac{624}{237695}(361)$ $-137\sqrt{21})$	$\frac{3411-745\sqrt{21}}{24696}$ $-79\sqrt{21})$	$\frac{7}{360}(7+\sqrt{21})$ $\frac{7}{360}(7+\sqrt{21})$
$\frac{7+\sqrt{21}}{14}$	$\frac{793+187\sqrt{21}}{90552}(331)$ $+55\sqrt{21})$	$\frac{25}{43218}(1044$ $+247\sqrt{21})$	$\frac{-128}{9745659}(14885$ $+3779\sqrt{21})$	$\frac{7}{360}(7-\sqrt{21})$ $\frac{7}{360}(7-\sqrt{21})$
1	$\frac{-(157-3\sqrt{21})}{378}\frac{25(143-10\sqrt{21})}{2772}$	$\frac{-25}{3969}(876$ $+18\sqrt{21})$	$\frac{1280}{596673}(913$ $+26\sqrt{21})$	$\frac{7}{4228}(1777$ $+37\sqrt{21})$
1	$\frac{1}{20}$	0	$\frac{8}{45}$	$\frac{7(7+\sqrt{21})}{360}0$ $\frac{7(7-\sqrt{21})}{360}0$

можно применить алгоритм ГБШ (9.13) прямо к системе (13.30), что дает

$$y_1 = y_0 + hy'_0, \quad (13.31a)$$

$$y'_1 = y'_0 + hf(x_0, y_0),$$

$$y_{i+1} = y_{i-1} + 2hy'_i, \quad (13.31b)$$

$$y'_{i+1} = y'_{i-1} + 2hf(x_i, y_i), \quad i = 1, 2, \dots, 2n,$$

$$S_h(x) = (y_{2n-1} + 2y_{2n} + y_{2n+1})/4, \quad (13.31c)$$

$$S'_h(x) = (y'_{2n-1} + 2y'_{2n} + y'_{2n+1})/4.$$

Здесь  $S_h(x)$  и  $S'_h(x)$  аппроксимируют  $y(x)$  и  $y'(x)$  при  $x = x_0 + H$ , причем  $H = 2nh$  и  $x_i = x_0 + ih$ . Сделаем теперь следующее важное замечание: для вычисления значений  $y_0, y_2, y_4, \dots, y_{2n}$  (с четными индексами) и  $y'_1, y'_3, \dots, y'_{2n+1}$  (с нечетными индексами) нужно вычислить только значения функции  $f(x_0, y_0), f(x_2, y_2), \dots, f(x_{2n}, y_{2n})$ . Мы знаем, далее, что как  $y_{2n}$ , так и  $(y_{2n-1} + y_{2n+1})/2$  обладают асимптотическими разложениями погрешности по четным степеням  $h$  (см. формулы (9.17)). Поэтому понятно, что формулы (13.31c) следует заменить на

$$\begin{aligned} S_h(x) &= y_{2n}, \\ S'_h(x) &= (y'_{2n-1} + y'_{2n+1})/2 \end{aligned} \quad (13.31c')$$

[Грэгг (1965)]. При использовании этого конечного шага число вычислений функции уменьшается в два раза. Вычисленные приближенные значения можно теперь использовать для экстраполяции. Взяв гармоническую последовательность (9.8'), положим

$$T_{11} = S_h(x_0 + H), \quad T'_{11} = S'_h(x_0 + H)$$

и рассчитаем экстраполированные выражения  $T_{i,j}$  и  $T'_{i,j}$  с помощью формулы Эйткена—Невилла (9.12).

*Замечание.* Исключая в (13.31b) значения  $y'_j$ , мы получим эквивалентную формулу

$$y_{i+2} - 2y_i + y_{i-2} = (2h)^2 f(x_i, y_i), \quad (13.32)$$

часто называемую *правилом Штёрмера*. Но для программирования предпочтительнее формулировка (13.31b), потому что она более устойчива по отношению к погрешностям округления (см. разд. III.10).

### Численные эксперименты

Этот раздел мы завершим сравнением вычислительных свойств разных методов, для чего нами написаны следующие три программы:

ODEX2 — программа, основанная на описанном выше экстраполяционном методе. Она организована так же, как и ODEX (экстраполяционная программа для дифференциальных уравнений первого порядка). В частности, стратегия выбора порядка и длины шага описана в разд. II.9.

DOPRIN основана на вложенных методах Ньюстрёма фиксированного порядка с коэффициентами из табл. 13.4. Выбор длины шага делается как в программе DOPR18 и был описан в разд. II.4.

FILGR9 использует вложенные методы Ньюстрёма порядков 9(8), которые предложили Филиппи и Грэф (1986). Она организована так же, как и DOPRIN.

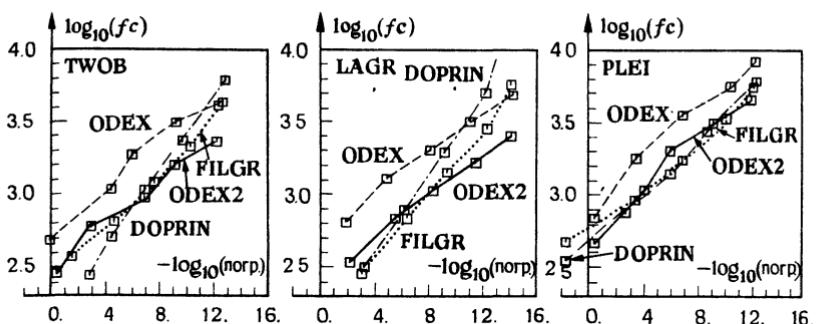


Рис. 13.3. Диаграммы «точность — объем вычислений» для приведенных выше примеров.

Листинги программ ODEX2 и DOPRIN приведены в приложении.

С помощью этих программ решались три системы дифференциальных уравнений вида  $y'' = f(x, y)$  из разд. II.10 (задачи TWOB, LAGR и PLEI). Каждая из задач решалась при следующих задаваемых значениях допустимой локальной погрешности  $tol = 10^{-3}, 10^{-6}, 10^{-9}, 10^{-12}, 10^{-15}$ . На рис. 13.3, как и ранее, по осям отложены в логарифмических масштабах точность (величина погрешности  $err$ ) и число  $fc$  вызовов подпрограммы вычисления функции. Чтобы продемонстрировать превосходство специальных методов для уравнения  $y'' = f(x, y)$ , мы включили также результаты расчетов по программе ODEX, предназначеннной для решения дифференциальных уравнений первого порядка (см. разд. II.9).

Видно, что ODEX2 действительно считает лучше, чем ODEX. Эта программа считает почти вдвое быстрее, в согласии с теоретическими соображениями. При больших требованиях к точности ( $tol < 10^{-10}$ ) она дает наилучшие результаты. Действительно, экстраполяционная программа позволяет доходить до очень высоких порядков точности. При требованиях к точности  $10^{-3} \geq 10^*$

$\gg tol \gg 10^{-8}$  более предпочтительны программы DOPRIN и FILGR9, использующие методы фиксированных порядков. Однако при решении этих задач не заметно их преимуществ перед программой DOPRI8, рассчитанной на системы уравнений первого порядка (ср. рис. 10.3).

Испытывалась также программа DIFEX2, написанная Дойфлхардом и Бауэром (см. Дойфлхард (1985)). Она показала результаты, аналогичные полученным для ODEX2.

### Системы высших порядков

Обобщение идей этого раздела на системы уравнений *более высоких порядков*

$$y^{(n)} = f(x, y, y', \dots, y^{(n-1)}) \quad (13.33)$$

представляется теперь более или менее непосредственным. И в этом случае действительное улучшение возможно только когда правая часть уравнения (13.33) зависит лишь от  $x$  и  $y$ . Этой теме посвятил свою замечательную работу Цурмюль (1948). Таблицы условий порядка и методов приведены у Хебзаккера (1982).

### Упражнения

1. Убедитесь, что методы из табл. 13.2 имеют, соответственно, четвертый и пятый порядки.
2. Коэффициенты погрешности метода Ньюстрёма  $p$ -го порядка определяются формулами

$$\begin{aligned} e(t) &= 1 - (\rho(t) + 1) \varphi(t) \sum_i b_i \Phi_i(t) \text{ для } \rho(t) = p, \\ e'(t) &= 1 - \varphi(t) \sum_i b_i \Phi_i(t) \quad \text{для } \rho(t) = p+1. \end{aligned} \quad (13.34)$$

- a) Покажите, что упрощающее предположение (13.25) имеет следствием соотношение

$$e(t) = -\rho(t) e'(u) \text{ при } \rho(t) = p,$$

где  $N$ -дерево  $u$  получается из  $t$  добавлением ветви с тонкой вершиной, исходящей из корня  $t$ .

b) Для метода Ньюстрёма из табл. 13.1 вычислите коэффициенты погрешности и сравните их с таковыми для классического метода Рунге—Кутты.

3. Покажите, что условия порядка для методов РК (теорема 2.13) составляют подмножество условий (13.21). Они соответствуют тем  $N$ -деревьям, у которых все вершины толстые.

4. Иногда порядок  $p$  метода Ньюстрёма (13.9) определяют смягченными требованиями

$$\begin{aligned} y(x_0 + h) - y_1 &= O(h^{p+1}), \\ y'(x_0 + h) - y'_1 &= O(h^p) \end{aligned} \quad (13.35)$$

(см. Ньюстрём (1925)). Покажите, что условия (13.35) не достаточны для получения глобальной сходимости порядка  $p$ . *Указание.* Исследуйте асимптотическое разложение глобальной погрешности с помощью теоремы 8.1 и формулы (8.8).

5. Покажите, что вычисляемые описанным в этом разделе экстраполяционным методом приближения  $T_{kk}$  и  $T'_{kk}$  эквивалентны приближенным решениям, которые дает некоторый метод Ньюстрёма порядка  $p = 2k$ , имеющий  $s = \frac{p^2}{8} + \frac{p}{4} + 1$  стадий.

## II.14. Р-ряды для разделяющихся обыкновенных дифференциальных уравнений

Divide ut regnes  
Разделяй и властвуй  
(Макиавелли)

В разд. II.2 и II.11 мы развили теорию В-рядов, позволяющую найти условия порядка для методов Рунге—Кутты, ориентированных на решение автономных дифференциальных уравнений. Это предполагает, что интегрирование  $x$  и  $y$  производится одинаковым образом, т. е. что выполнены условия (1.9). Мы уже встречались, однако, с некоторыми ситуациями, в которых эта теория оказывалась недостаточной.

1) Если  $x$  и  $y$  интегрируются различным образом, т. е. условия (1.9) не выполнены, то, как мы видели (см. упр. 6 в разд. II.1), возникают дополнительные условия порядка. Чтобы разобраться в их структуре, надо обобщить теорию В-рядов на разделяющееся обыкновенное дифференциальное уравнение (2.1).

2) Рассмотренный в разд. II.12 подход Фельберга требует изучения методов РК, не удовлетворяющих условию (1.9), причем в точке  $(x_0, y_0)$  производные  $f, f_x, f_{xx}, \dots$  равны нулю.

3) В разд. II.13 рассматривались прямые методы для дифференциального уравнения второго порядка  $y'' = f(y, y')$ . Чтобы найти общую структуру условий порядка для этих методов, мы обобщили теорию В-рядов на разделяющееся обыкновенное дифференциальное уравнение

$$\begin{pmatrix} y \\ y' \end{pmatrix}' = \begin{pmatrix} y' \\ f(y, y') \end{pmatrix}. \quad (14.1)$$

Для таких ситуаций желательно, конечно, иметь единую теорию. С этой целью Хайрер (1981) предложил обобщение теории В-рядов на общий случай *разделяющегося обыкновенного дифференциального уравнения*:

$$\begin{pmatrix} y_a \\ y_b \end{pmatrix}' = \begin{pmatrix} f_a(y_a, y_b) \\ f_b(y_a, y_b) \end{pmatrix}, \quad (14.2)$$

который включает все перечисленные выше ситуации, так что для них условия порядка получаются как частные случаи.

Другие важные приложения этой теории встречаются при изучении условий порядка для методов, в которых к разным

частям системы обыкновенных дифференциальных уравнений применяются различные вычислительные схемы. Такими частями могут быть, например, линейная и нелинейная части (на эту ситуацию ориентированы  $W$ -методы, см. том II), или «нежесткая» и «жесткая» компоненты дифференциального уравнения (соответствующие методы называют разделяющимися, или составными).

Для простоты обозначений мы будем рассматривать систему (14.2) как разделяющуюся на две компоненты. Обобщение на большее число компонент не представляет трудностей. Чтобы лучше понять, как должны быть в этом случае определены элементарные дифференциалы и соответствующие им деревья, мы рассмотрим сначала производные точного решения уравнения (14.2).

### Производные точного решения; Р-деревья

Чтобы не выписывать суммы и лишние индексы, предположим, что  $y_a$  и  $y_b$  в (14.2) являются скалярными величинами. Все дальнейшие формулы останутся справедливыми и для векторов, если производные интерпретировать как полилинейные отображения. Дифференцируя (14.2) и подставляя вместо производных снова (14.2), для первой компоненты, т. е.  $y_a$ , получим:

$$y_a^{(1)} = f_a, \quad (14.3.1)$$

$$y_a^{(2)} = \frac{\partial f_a}{\partial y_a} f_a + \frac{\partial f_a}{\partial y_b} f_b, \quad (14.3.2)$$

$$\begin{aligned} y_a^{(3)} = & \frac{\partial^2 f_a}{\partial y_a^2} (f_a, f_a) + \frac{\partial^2 f_a}{\partial y_b \partial y_a} (f_b, f_a) + \\ & + \frac{\partial f_a}{\partial y_a} \frac{\partial f_a}{\partial y_a} f_a + \frac{\partial f_a}{\partial y_a} \frac{\partial f_a}{\partial y_b} f_b + \frac{\partial^2 f_a}{\partial y_a \partial y_b} (f_a, f_b) + \quad (14.3.3) \\ & + \frac{\partial^2 f_a}{\partial y_b^2} (f_b, f_b) + \frac{\partial f_a}{\partial y_b} \frac{\partial f_b}{\partial y_a} f_a + \frac{\partial f_a}{\partial y_b} \frac{\partial f_b}{\partial y_b} f_b. \end{aligned}$$

Аналогичные формулы получаются и для производных  $y_b$ .

Для графического представления этих формул нам потребуются вершины двух сортов. Как и в разд. II.13, мы будем пользоваться «тонкими» и «толстыми» вершинами, отвечающими  $f_a$  и  $f_b$ , соответственно. Тогда формулы (14.3) можно графически представить так, как показано на рис. 14.1.

Дадим теперь формальное определение этих графов.

**Определение 14.1.** Помеченным Р-деревом порядка  $q$  назовем пару  $(t, t')$ , где  $t$  — помеченное дерево

$$t: A_q - \{j\} \rightarrow A_q$$

(см. определение 2.2) и  $t'$  — отображение

$$t': A_q \rightarrow \{\text{«тонкая», «толстая» (вершина)}\}.$$

Обозначим  $LTP_q^a$  множество всех помеченных Р-деревьев порядка  $q$  с «тонким» корнем (т. е.  $t'(j) = \text{«тонкая»}$ ). Аналогично,  $LTP_q^b$  — множество всех помеченных Р-деревьев  $q$ -го порядка с «толстым» корнем.

$$\begin{array}{c} j \\ \cdot \\ k \\ \backslash \\ j \end{array} \quad (14.3.1)$$

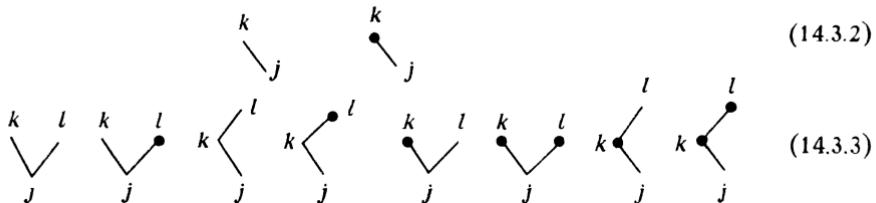
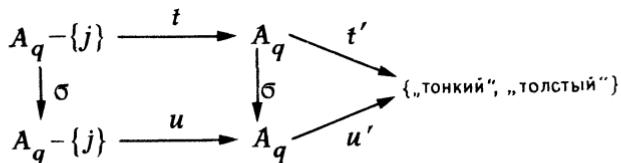


Рис. 14.1. Производные точного решения  $y_a$

В (14.3.3) второй и пятый члены равны. В связи с этим дадим

**Определение 14.2.** Два помеченных Р-дерева  $q$ -го порядка  $(t, t')$  и  $(u, u')$  эквивалентны, если существует такое взаимно однозначное отображение  $\sigma: A_q \rightarrow A_q$ , что  $\sigma(j) = j$  и выполняются перестановочные соотношения, представленные следующей диаграммой:



**Определение 14.3.** Класс эквивалентности помеченных Р-деревьев  $q$ -го порядка называется Р-деревом порядка  $q$ . Множество всех Р-деревьев порядка  $q$  с «тонким» корнем обозначается  $TP_q^a$ , а с «толстым» корнем —  $TP_q^b$ . Порядок Р-дерева  $t$  будем обозначать через  $\rho(t)$ , а число элементов в классе эквивалентности  $t$  — через  $\alpha(t)$ .

Ниже, в табл. 14.1, приведены примеры Р-деревьев и числа  $\rho(t)$  и  $\alpha(t)$  для них. Мы обсудим сначала рекурсивное представление Р-деревьев, играющее фундаментальную роль в излагаемой теории (оно обобщает индуктивное определение деревьев 2.12).

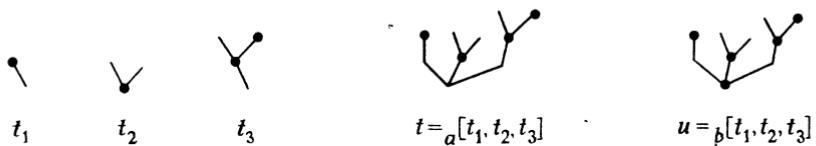


Рис. 14.2. Рекуррентное определение Р-деревьев.

**Определение 14.4.** Пусть  $t_1, \dots, t_m$  — Р-деревья. Будем обозначать

$$t = {}_a[t_1, \dots, t_m] \quad (14.4)$$

(единственное) Р-дерево с «тонким» корнем, такое что после удаления корня и инцидентных ему ребер остаются Р-деревья  $t_1, \dots, t_m$ . Совершенно аналогично обозначим через  ${}_b[t_1, \dots, t_m]$  Р-дерево, получаемое объединением  $t_1, \dots, t_m$  на одном новом «толстом» корне. Наконец, введем обозначения  $\tau_a$  и  $\tau_b$  для «тонкого» и «толстого» Р-деревьев порядка 1. Рис. 14.2 иллюстрирует это определение.

Займемся теперь уточнением связи между Р-деревьями и отдельными членами в формулах (14.3). Введем еще обозначение

$$w(t) = \begin{cases} a, & \text{если корень } t \text{ «тонкий»,} \\ b, & \text{если корень } t \text{ «толстый».} \end{cases} \quad (14.5)$$

**Определение 14.5.** Элементарные дифференциалы, соответствующие системе (14.2), определяются индуктивно следующим образом:

$$F(\tau_a)(y) = f_a(y),$$

$$F(\tau_b)(y) = f_b(y)$$

и

$$F(t)(y) = \frac{\partial^m f_{w(t)}(y)}{\partial y_w(t_1) \cdots \partial y_w(t_m)} \cdot (F(t_1)(y), \dots, F(t_m)(y)),$$

где  $y = (y_a, y_b)$  и  $t = {}_a[t_1, \dots, t_m]$  или  $t = {}_b[t_1, \dots, t_m]$ .

Элементарные дифференциалы для Р-деревьев до третьего порядка включительно выписаны в табл. 14.1.

Вернемся теперь к началу раздела и продолжим дифференцирование формул (14.3). Легко видеть, что в терминах помеченных Р-деревьев дифференцирование  $F(t)(y_a, y_b)$  можно интерпретировать как добавление новой ветви с «тонкой» или «толстой» вершиной и новым индексом суммирования поочередно к каждой

Таблица 14.1. Р-деревья и их элементарные дифференциалы

P-дерево	(14.4)	$p(t)$	$\alpha(t)$	элементарный дифференциал
.	$\tau_a$	1	1	$f_a$
\	$a[\tau_a]$	2	1	$\frac{\partial f_a}{\partial y_a} f_a$
\	$a[\tau_b]$	2	1	$\frac{\partial f_a}{\partial y_b} f_b$
\vee	$a[\tau_a, \tau_a]$	3	1	$\frac{\partial^2 f_a}{\partial y_a^2}(f_a, f_a)$
\vee	$a[\tau_a, \tau_b]$	3	2	$\frac{\partial^2 f_a}{\partial y_a \partial y_b}(f_a, f_b)$
\vee	$a[\tau_b, \tau_b]$	3	1	$\frac{\partial^2 f_a}{\partial y_b^2}(f_b, f_b)$
\langle	$a[a[\tau_a]]$	3	1	$\frac{\partial f_a}{\partial y_a} \frac{\partial f_a}{\partial y_a} f_a$
\langle	$a[a[\tau_b]]$	3	1	$\frac{\partial f_a}{\partial y_a} \frac{\partial f_a}{\partial y_b} f_b$
\langle	$a[b[\tau_a]]$	3	1	$\frac{\partial f_a}{\partial y_b} \frac{\partial f_b}{\partial y_a} f_a$
\langle	$a[b[\tau_b]]$	3	1	$\frac{\partial f_a}{\partial y_b} \frac{\partial f_b}{\partial y_b} f_b$
...	...	...	...	...
.	$\tau_b$	1	1	$f_b$
\	$b[\tau_a]$	2	1	$\frac{\partial f_b}{\partial y_a} f_a$
\	$b[\tau_b]$	2	1	$\frac{\partial f_b}{\partial y_b} f_b$
...	...	...	...	...

из вершин помеченного Р-дерева  $t$ . Тем же способом, каким мы доказывали теорему 2.6 для неразделяющихся дифференциальных уравнений, может быть доказана

**Теорема 14.6.** Производные точного решения уравнения (14.2) выражаются формулами

$$\begin{aligned} F(t) &= \sum_{t \in LTP_q^a} F(t)(y_a, y_b) = \sum_{t \in TP_q^a} \alpha(t) F(t)(y_a, y_b), \\ F(t) &= \sum_{t \in LTP_q^b} F(t)(y_a, y_b) = \sum_{t \in TP_q^b} \alpha(t) F(t)(y_a, y_b). \end{aligned} \quad (14.3.q)$$

## P-ряды

Мы уже видели в разд. II.11, какую важную роль играет ключевая лемма — следствие 11.7 — при выводе условий порядка для методов Рунге—Кутты. Поэтому обобщим данный результат и на разделяющиеся обыкновенные дифференциальные уравнения.

Прежде всего, удобно ввести два новых Р-дерева  $\mathcal{O}_a$  и  $\mathcal{O}_b$  порядка 0, которым соответствуют элементарные дифференциалы  $F(\mathcal{O}_a)(y) = y_a$  и  $F(\mathcal{O}_b)(y) = y_b$ . Далее, положим

$$\begin{aligned} TP^a &= \{\mathcal{O}_a\} \cup TP_1^a \cup TP_2^a \cup \dots, \\ TP^b &= \{\mathcal{O}_b\} \cup TP_1^b \cup TP_2^b \cup \dots, \\ LTP^a &= \{\mathcal{O}_a\} \cup LTP_1^a \cup LTP_2^a \cup \dots, \\ LTP^b &= \{\mathcal{O}_b\} \cup LTP_1^b \cup LTP_2^b \cup \dots. \end{aligned} \quad (14.6)$$

**Определение 14.7.** Пусть  $\mathbf{c}(\mathcal{O}_a), \mathbf{c}(\mathcal{O}_b), \mathbf{c}(\tau_a), \mathbf{c}(\tau_b), \dots$  — система вещественных коэффициентов, определенных для всех Р-деревьев:  $\mathbf{c}: TP^a \cup TP^b \rightarrow \mathbb{R}$ .

Тогда ряд

$$P(\mathbf{c}, y) = (P_a(\mathbf{c}, y), P_b(\mathbf{c}, y))^T,$$

где

$$\begin{aligned} P_a(\mathbf{c}, y) &= \sum_{t \in LTP^a} \frac{h^{\rho(t)}}{\rho(t)!} \mathbf{c}(t) F(t)(y), \\ P_b(\mathbf{c}, y) &= \sum_{t \in LTP^b} \frac{h^{\rho(t)}}{\rho(t)!} \mathbf{c}(t) F(t)(y), \end{aligned}$$

называется Р-рядом.

Теорема 14.6 просто утверждает, что точное решение уравнения (14.2) является Р-рядом

$$(y_a(x_0 + h), y_b(x_0 + h))^T = P(y, (y_a(x_0), y_b(x_0))),$$

у которого  $y(t) = 1$  для всех Р-деревьев  $t$ .

**Теорема 14.8.** Пусть  $\mathbf{c}: TP^a \cup TP^b \rightarrow \mathbb{R}$  — такая последовательность коэффициентов, у которой  $\mathbf{c}(\mathcal{O}_a) = \mathbf{c}(\mathcal{O}_b) = 1$ . Тогда

$$h \begin{pmatrix} f_a(P(\mathbf{c}, (y_a, y_b))) \\ f_b(P(\mathbf{c}, (y_a, y_b))) \end{pmatrix} = P(\mathbf{c}'(y_a, y_b)), \quad (14.7)$$

где

$$\begin{aligned} \mathbf{c}'(\mathcal{O}_a) &= \mathbf{c}'(\mathcal{O}_b) = 0, \\ \mathbf{c}'(\tau_a) &= \mathbf{c}'(\tau_b) = 1, \end{aligned} \quad (14.8)$$

$$\mathbf{c}'(t) = \rho(t) \mathbf{c}(t_1) \dots \mathbf{c}(t_m),$$

если  $t = {}_a[t_1, \dots, t_m]$  или  $t = {}_b[t_1, \dots, t_m]$ .

Доказательство этой теоремы родственно доказательству теоремы 11.6. Более подробно оно приведено в работе Хайрера (1981). ■

Применим теперь теорему 14.8 к случаям, указанным в начале этого раздела.

### Методы РК с нарушением условия (1.9)

Рассмотрим применение метода РК (1.8) к неавтономному дифференциальному уравнению  $y' = f(x, y)$ . Мы не предполагаем условие (1.9) выполненным, так что  $x$ - и  $y$ -компоненты интегрируются по-разному. Поэтому добавим к нашему дифференциальному уравнению еще уравнение  $x' = 1$  и получим систему

$$\begin{aligned} y' &= f(x, y), \\ x' &= 1. \end{aligned} \quad (14.9)$$

Она имеет вид (14.2), если положить  $y_a = y$ ,  $y_b = x$ ,  $f_a(y_a, y_b) = f(x, y)$  и  $f_b(y_a, y_b) = 1$ . Поскольку  $f_b$  — постоянная, все содержащие производные  $f_b$  элементарные дифференциалы тождественно обращаются в нуль. Следовательно, те Р-деревья, у которых хотя бы одна толстая вершина не является конечной, рассматривать не надо. Остается рассмотреть множество

$$T_a = \{t \in TP_a; \text{ все толстые вершины конечные}\}. \quad (14.10)$$

Для вывода условий порядка метода РК

$$\begin{aligned} k_i &= h f(x_0 + c_i h, g_i), \\ g_i &= g_0 + \sum_j a_{ij} k_j, \\ y_1 &= y_0 + \sum_i b_i k_i \end{aligned} \quad (14.11)$$

мы предположим, что  $k_i$  и  $g_i$  являются первыми компонентами Р-рядов:

$$k_i = P_a(k_i, (y_0, x_0)), \quad g_i = P_a(g_i, (y_0, x_0)).$$

Кроме того, положим

$$x_0 + c_i h = P_b(g_i, (y_0, x_0)),$$

так что  $g_i(\emptyset_b) = 1$ ,  $g_i(\tau_b) = c_i$  и  $g_i(t) = 0$  для остальных Р-деревьев с «толстым» корнем. Теперь из (14.11) на основе теоремы 14.8 получаем, что для всех  $t \in T_a$

$$\begin{aligned} k_i(t) &= g'_i(t), \\ g_i(t) &= \sum_j a_{ij} k_j(t). \end{aligned} \quad (14.12)$$

Мы получили рекуррентные соотношения для определения коэффициентов  $k_i(t)$ . Условия того, что метод будет иметь порядок  $p$ , принимают вид

$$\sum_i b_i k_i(t) = 1 \text{ при } t \in T_x, \rho(t) \leq p. \quad (14.13)$$

Мы видим, что обычные условия порядка (типа (1.11)) исчезают, за тем исключением, что для толстых вершин вместо  $\sum_j a_{ij}$  представляются  $c_i$ .

На рис. 14.3 представлены первые элементы множества  $T_x$ .



Рис. 14.3. Элементы множества  $T_x$ .

**Пример.** Условие (14.13) для пятого, восьмого и последнего деревьев рис. 14.3 принимает соответственно вид

$$3 \sum_{i,j} b_i c_i a_{ij} = 1,$$

$$3 \cdot 2 \sum_{i,l} b_i a_{il} c_l = 1,$$

$$6 \cdot 3 \cdot 2 \sum_{i,l,k,t} b_i a_{il} c_j a_{jk} a_{kl} c_l = 1.$$

### Методы Фельберга

Если для точного решения уравнения  $y' = f(x, y)$  значения  $y(x), y'(x), \dots, y^{(n)}(x)$  обращаются в нуль при  $x = x_0$ , то и  $f, df/dx, \dots, d^{n-1}f/dx^{n-1}$  равны нулю в точке  $(x_0, y_0)$  (см. разд. II.12),

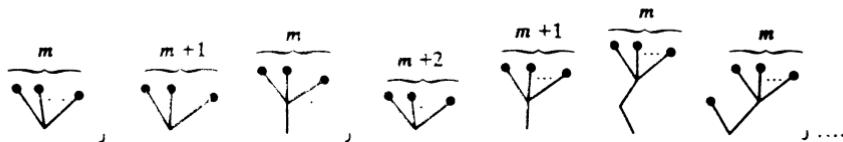


Рис. 14.4. Р-деревья, которые следует рассмотреть для метода Фельберга (разд. II.12).

методы Фельберга). Для этого случая на рис. 14.4 мы привели элементы множества  $T_x$ , для которых элементарные дифференциалы в общем случае не обращаются в нуль в точке  $(x_0, y_0)$ .

## Методы Нюстрёма

В качестве последнего приложения теоремы 14.8 представим новый вывод условий порядка для методов Нюстрёма (разд. II.13). Дифференциальное уравнение второго порядка  $y'' = f(y, y')$  может быть записано в форме разделяющейся системы:

$$\begin{pmatrix} y \\ y' \end{pmatrix}' = \begin{pmatrix} y' \\ f(y, y') \end{pmatrix}. \quad (14.14)$$

В обозначениях (14.2) мы имеем  $y_a = y$ ,  $y_b = y'$ ,  $f_a(y_a, y_b) = y_b$ ,  $f_b(y_a, y_b) = f(y_a, y_b)$ . Вследствие специального вида  $f_a$  должны быть рассмотрены только те Р-деревья, которые удовлетворяют условию

«тонкие вершины могут иметь не более одного сына, который должен быть толстым» (14.15)

(см. определение 13.2). Следовательно, существенными Р-деревьями являются

$$TN_q^a = \{t \in TP_q^a; \quad t \text{ удовлетворяет (14.15)}\},$$

$$TN_q^b = \{t \in TP_q^b; \quad t \text{ удовлетворяет (14.15)}\}.$$

Отсюда вытекает, что каждый элемент  $TN_{q+1}^a$  может быть записан в виде  $t = {}_a[u]$ , где  $u \in TN_q^b$ . Поэтому существует взаимно однозначное отображение  $TN_{q+1}^a$  на  $TN_q^b$ , которое не изменяет элементарные дифференциалы:

$$F({}_a[u])(y_a, y_b) = \frac{\partial y_b}{\partial y_b} \cdot F(u)(y_a, y_b) = F(u)(y_a, y_b).$$

Из этого свойства следует, что

$$hP_b(c, (y_a, y_b)) = P_a(c', (y_a, y_b)), \quad (14.16)$$

$$\text{где } c'(t) = \rho(t) c(u), \text{ если } t = {}_a[u]. \quad (14.17)$$

Эти обозначения согласуются с (14.8).

Условия порядка для метода (13.13) можно теперь вывести следующим образом. Предположим, что  $g_i$  и  $g'_i$  представляют Р-ряд:

$$g_i = P_a(c_i, (y_0, y'_0)), \quad g'_i = P_b(c_i, (y_0, y'_0)).$$

Тогда из теоремы 14.8 следует, что

$$hf(g_i, g'_i) = P_b(c'_i, (y_0, y'_0)). \quad (14.18)$$

Умножив это соотношение на  $h$ , с помощью (14.16) получим:

$$h^2f(g_i, g'_i) = P_a(c''_i, (y_0, y'_0)). \quad (14.19)$$

Здесь  $c_t'' = (c_t')'$ , т. е.

$$c_t''(t) = 0 \text{ для } t = \varnothing_a \text{ и } t = \tau_a,$$

$$c_t''(a[\tau_b]) = 1,$$

$$c_t''(t) = \rho(t)(\rho(t) - 1)c_t(t_1) \dots c_t(t_m),$$

если  $t = a[b[t_1, \dots, t_m]]$ .

Если подставить выражения (14.18) и (14.19) в (13.13), мы получим, как обычно, рекуррентные соотношения

$$c_t(\tau_a) = c_t,$$

$$c_t(t) = \sum_i \bar{a}_{ij} c_i''(t), \text{ если корень «тонкий»,}$$

$$c_t(t) = \sum_i a_{ij} c_i'(t), \text{ если корень «толстый»}.$$

Точно таким же способом получают и условия порядка (для  $p$ -го порядка)

$$\begin{aligned} \sum_i b_i c_i''(t) &= 1 \text{ для } t \in TN_q^a, q = 2, \dots, p; \\ \sum_i b_i c_i'(t) &= 1 \text{ для } t \in TN_q^b, q = 1, \dots, p. \end{aligned} \quad (14.20)$$

### Упражнения

1. Обозначим через  $\alpha_q$  число элементов множества  $TP_q^a$  (число Р-деревьев порядка  $q$  с «тонким» корнем, см. табл. 14.2). Докажите, что

$$\begin{aligned} \alpha_1 + \alpha_2 x + \alpha_3 x^2 + \dots &= \\ = (1-x)^{-2\alpha_1} (1-x^2)^{-2\alpha_2} \times & \\ \times (1-x^3)^{-2\alpha_3} \dots & \end{aligned}$$

Вычислите несколько первых чисел  $\alpha_q$  и сравните их с числами

$$\alpha_q = \text{card}(T_q) \text{ из табл. 2.2.}$$

2. Докажите, что не существует явного 4-стадийного метода РК порядка 4, который не удовлетворял бы условию (1.9).

*Указание.* Воспользуйтесь техникой доказательства леммы 1.4.

3. Покажите, что условия порядка (14.20) совпадают с теми, которые дает теорема 13.10.

Таблица 14.2. Число элементов множества  $TP_q^a$

$q$	$\alpha_q$	$q$	$\alpha_q$
1	1	6	458
2	2	7	2 058
3	7	8	9 498
4	26	9	44 947
5	107	10	216 598

## II.15. Дифференциальные уравнения с запаздывающим аргументом

«Детальное изучение реального мира вынуждает нас принимать в расчет, хотя и с неохотой, тот факт, что скорость изменения физических систем зависит не только от их состояния в настоящий момент, но также и от их истории в прошлом.»  
(Беллман и Кук, 1963)

Дифференциальные уравнения с запаздывающим аргументом — это такие уравнения, как

$$y'(x) = f(x, y(x), y(x - \tau)), \quad (15.1)$$

или

$$y'(x) = f(x, y(x), y(x - \tau_1), y(x - \tau_2)), \quad (15.2)$$

или даже еще более общего вида. Здесь производная решения зависит также от его значений в предыдущих точках.

Запаздывание встречается во многих моделях прикладной математики. Запаздывание (в более общем случае — отклонения аргумента) может быть также источником интересных математических явлений, таких как неустойчивости, предельные циклы, периодический характер решений.

### Существование

Для уравнений типа (15.1) или (15.2), если величина запаздывания  $\tau$  ограничена снизу положительной постоянной, вопрос о существовании решения решается легко. Предположим, что на отрезке  $x_0 - \tau \leq x \leq x_0$  решение известно:

$$y(x) = \varphi(x).$$

Тогда на отрезке  $x_0 \leq x \leq x_0 + \tau$  функция  $y(x - \tau)$  является известной функцией  $x$  и (15.1) становится обыкновенным дифференциальным уравнением, к которому применимы известные теоремы существования. Определив  $y(x)$  при  $x_0 \leq x \leq x_0 + \tau$ , мы можем затем перейти к вычислению решения при  $x_0 + \tau \leq x \leq x_0 + 2\tau$  и т. д. Этот «метод шагов», таким образом, для всех  $x$  позволяет получить результаты о существовании и единственности решения. Заинтересованному читателю мы рекомен-

дум книги Беллмана и Кука (1963) и Драйвера (1977, особенно гл. V)<sup>1)</sup>.

**Пример 1.** Рассмотрим уравнение

$$y'(x) = -y(x-1), \quad y(x) = 1 \quad \text{при } -1 \leq x \leq 0. \quad (15.3)$$

Применяя описанный выше «метод шагов», получим

$$y(x) = 1 - x \quad \text{при } 0 \leq x \leq 1,$$

$$y(x) = 1 - x + \frac{(x-1)^2}{2!} \quad \text{при } 1 \leq x \leq 2,$$

$$y(x) = 1 - x + \frac{(x-1)^2}{2!} - \frac{(x-2)^3}{3!} \quad \text{при } 2 \leq x \leq 3,$$

и т. д. Решение представлено на рис. 15.1. Заметим, что решение имеет разрывные производные, несмотря на то, что правая часть

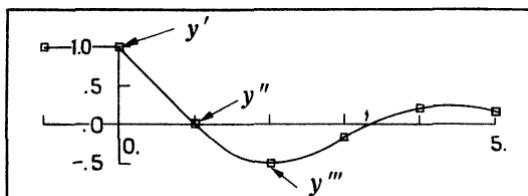


Рис. 15.1. Решение уравнения (15.3).

дифференциального уравнения и начальная функция принадлежат классу  $C^\infty$ . Это происходит оттого, что начальная функция не удовлетворяет дифференциальному уравнению. Однако с каждым шагом эти разрывы все более сглаживаются.

**Пример 2.** Этот пример ясно показывает, что решения уравнения с запаздывающим аргументом зависят не только от начального условия, но и от всей истории процесса от  $x_0 - \tau$  до  $x_0$ .

$$y'(x) = -1.4 \cdot y(x-1), \quad (15.4)$$

a)  $\varphi(x) = 0.8 \quad \text{при } -1 \leq x \leq 0,$

b)  $\varphi(x) = 0.8 + x \quad \text{при } -1 \leq x \leq 0,$

c)  $\varphi(x) = 0.8 + 2x \quad \text{при } -1 \leq x \leq 0.$

<sup>1)</sup> На русском языке можно рекомендовать книги Эльсгольца (1951) и Эльсгольца и Норкина (1963), указанные в списке дополнительной литературы в конце книги. — Прим. ред.

Решения показаны на рис. 15.2. Объяснение колебательного поведения решений будет дано ниже.

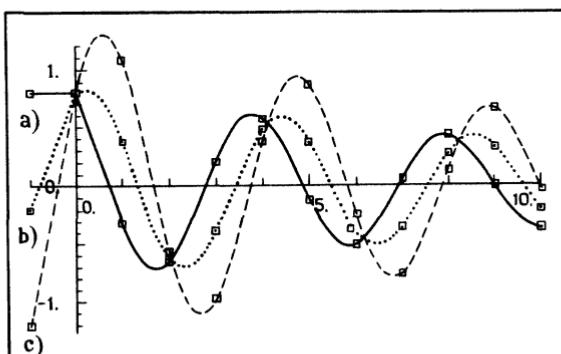


Рис. 15.2. Решения задач (15.4).

### Методы с постоянной длиной шага для постоянного запаздывания

Применяя метод Рунге—Кутты (1.8) к уравнению с запаздывающим аргументом (15.1), мы получим

$$g_i^{(n)} = y_n + h \sum_l a_{ij} f(x_n + c_j h, g_l^{(n)}, y(x_n + c_j h - \tau)),$$

$$y_{n+1} = y_n + h \sum_l b_j f(x_n + c_j h, g_l^{(n)}, y(x_n + c_j h - \tau)).$$

Сразу возникает вопрос, как задавать значения  $y(x_n + c_j h - \tau)$ . Но если запаздывание постоянное и длина шага выбрана так, что  $\tau = kh$ , где  $k$  — некоторое целое число, то естественно использовать уже вычисленные  $k$  шагов назад значения решения:

$$g_i^{(n)} = y_n + h \sum_l a_{ij} f(x_n + c_j h, g_l^{(n)}, \gamma_l^{(n)}), \quad (15.5a)$$

$$y_{n+1} = y_n + h \sum_l b_j f(x_n + c_j h, g_l^{(n)}, \gamma_l^{(n)}), \quad (15.5b)$$

где

$$\gamma_l^{(n)} = \begin{cases} \varphi(x_n + c_j h - \tau), & \text{если } n < k, \\ g_l^{(n-k)}, & \text{если } n \geq k. \end{cases} \quad (15.5c)$$

Это можно интерпретировать как последовательное решение уравнения

$$y'(x) = f(x, y(x), \varphi(x - \tau)) \quad (15.1a)$$

для интервала  $[x_0, x_0 + \tau]$ , затем уравнений

$$\begin{aligned} y'(x) &= f(x, y(x), z(x)), \\ z'(x) &= f(x - \tau, z(x), \varphi(x - 2\tau)) \end{aligned} \quad (15.1b)$$

для интервала  $[x_0 + \tau, x_0 + 2\tau]$ , затем

$$\begin{aligned} y'(x) &= f(x, y(x), z(x)), \\ z'(x) &= f(x - \tau, z(x), v(x)), \\ v'(x) &= f(x - 2\tau, v(x), \varphi(x - 3\tau)) \end{aligned} \quad (15.1c)$$

для интервала  $[x_0 + 2\tau, x_0 + 3\tau]$ , и т. д. Это точный численный аналог упомянутого выше «метода шагов».

**Теорема 15.1.** Если  $c_i$ ,  $a_{ij}$ ,  $b_j$  — коэффициенты метода Рунге—Кутты  $p$ -го порядка, то процесс (15.5) сходится с порядком  $p$ .

**Доказательство.** Последовательность (15.1a), (15.1b), ... — это последовательность систем обыкновенных дифференциальных уравнений, нормально решаемых методом РК  $p$ -го порядка. Поэтому результат непосредственно вытекает из теоремы 3.6. ■

**Замечание.** Для коллокационного метода, основанного на гауссовой квадратурной формуле, из теоремы 15.1 вытекает сверхсходимость — результат, который получил Беллен (1984).

**Численный пример.** Мы интегрировали уравнение (15.12) (см. ниже) на интервале  $0 \leq x \leq 10$  при  $a = 1.4$  с начальными условиями  $y(x) = 0$  при  $-1 \leq x < 0$  и  $y(0) = 0.1$ , взяв следующие длины шага:  $h = 1, 1/2, 1/4, 1/8, \dots$ . Использовался метод Бутчера шестого порядка (см. табл. 6.1). Результаты вычислений представлены в табл. 15.1. Хорошо виден 6-й порядок сходимости.

Таблица 15.1. Результаты расчетов для метода (15.5)

$h$	число шагов	число вычислений функции	максимальная погрешность
1	10	70	$1.69 \times 10^{-2}$
1/2	20	140	$4.76 \times 10^{-4}$
1/4	40	280	$4.76 \times 10^{-6}$
1/8	80	560	$7.80 \times 10^{-8}$
1/16	160	1120	$8.90 \times 10^{-10}$
1/32	320	2240	$2.14 \times 10^{-11}$
1/64	640	4480	$3.41 \times 10^{-13}$

### Методы с переменной длиной шага

К сожалению, метод (15.5) не позволяет произвольно менять длину шага, и применение его к уравнениям с переменной величиной запаздывания не столь просто. Если мы желаем достичь полной гибкости, нам необходима *глобальная аппроксимация решения*. Бесполезно иметь приближения только на некоторой последовательности точек. Поэтому подходящими методами для таких задач являются многошаговые методы типа Адамса или ФДН (см. разд. III.1), или же непрерывные методы Рунге—Кутты. Ввиду легкости программирования мы выбрали метод Дормана и Принса порядка 5 (4), присоединив к нему формулы (5.7) из разд. II.5. Превосходство этой процедуры над многими другими интерполяционными методами показал Готтвальд (1985), проведя обширное численное тестирование их. Для каждого интервала интегрирования решение в целом представляется многочленом 4-й степени. Поэтому после каждого успешного шага интегрирования подпрограмма RETARD (незначительная модификация подпрограммы DOPRI5) вызывает подпрограмму STORE, которая записывает в память (в общий блок COEF) коэффициенты этих многочленов вместе с соответствующими значениями  $x$ . Тогда при вызове вещественной функции YLAG (I, X) производится поиск нужного интервала для  $x$  и вычисляется соответствующий многочлен для I-го решения. Поэтому, чтобы подпрограмма RETARD решала уравнение с запаздывающим аргументом, скажем (15.4), надо, чтобы подпрограмма FCN включала такой оператор:

$$F(1) = -1.4 * YLAG(1, X - 1., PHI).$$

Тексты подпрограмм RETARD, STORE и YLAG вместе с примеромзывающей программы приведены в приложении. С их помощью были получены результаты, представленные на рис. 15.1 и 15.2.

Как мы видели, решение имеет разрывы производных в некоторых точках, например, для уравнения (15.1) в точках  $x_0 + \tau$ ,  $x_0 + 2\tau$ ,  $x_0 + 3\tau$ , ... и т. д. Поэтому точность улучшится, а время счета сократится, если эти точки выделить и возобновлять в них вызов подпрограммы RETARD.

При конструировании усложненных программ для решения дифференциальных уравнений с запаздывающим аргументом более общего чем (15.1) вида возникает ряд трудностей, в частности, нахождение всех точек разрыва производных до определенного порядка, а также экономное распоряжение памятью для хранения всех коэффициентов, представляющих запомненное решение, и своевременное освобождение от того, что больше не понадобится.

### Характеристические значения экспоненциальных решений

На рис. 15.1 и 15.2 можно заметить, что решения после начальной фазы как будто приближаются к функциям, похожим на  $e^{\alpha x} \cos \beta(x - \delta)$ . Попытаемся определить значения  $\alpha$  и  $\beta$ . Рассмотрим уравнение

$$y'(x) = \lambda y(x) + \mu y(x-1). \quad (15.6)$$

Выбор запаздывания  $\psi = 1$  не означает потери общности, так как любое запаздывание  $\psi \neq 1$  можно свести к  $\psi = 1$  заменой переменных.

Будем искать решение вида

$$y(x) = e^{\varphi x}, \text{ где } \varphi = \alpha + i\beta. \quad (15.7)$$

Подставляя (15.7) в уравнение (15.6), получим следующее «характеристическое уравнение» для  $\varphi$ :

$$\varphi - \lambda - \mu e^{-\varphi} = 0. \quad (15.8)$$

При  $\mu \neq 0$  оно имеет бесконечное множество решений. В самом деле, если  $|\varphi|$  становится большим, то из (15.8) видно, что  $\mu e^{-\varphi}$  тоже должно быть большим, поскольку  $\lambda$  фиксировано:

$$\varphi \approx \mu e^{-\varphi}. \quad (15.8')$$

Отсюда вытекает, что  $\varphi = \alpha + i\beta$  расположено вблизи мнимой оси. Следовательно,  $|\varphi| \approx |\beta|$ , а из (15.8') находим, что

$$|\beta| \approx |\mu| e^{-\alpha}.$$

Поэтому корни уравнения (15.8) асимптотически располагаются на кривых  $-\alpha = \ln |\beta| - \ln |\mu|$ . Из (15.8') видно также, что характеристическое уравнение имеет корень, когда аргумент выражения  $\mu e^{-i\beta}$  близок к  $\pi/2$  (при  $\beta > 0$ ), т. е. когда

$$\beta \approx \arg \mu - \frac{\pi}{2} + 2k\pi, \quad k = 1, 2, \dots$$

Таким образом, имеется две последовательности характеристических значений, уходящих в бесконечность по логарифмическим кривым слева от мнимой оси, и расстояние между двумя последовательными значениями асимптотически приближается к  $2\pi$ .

Итак, «общее решение» уравнения (15.6) — это похожая на ряд Фурье суперпозиция решений типа (15.7) [Райт (1946), см. также Беллман и Кук (1963), гл. 4]. Чем меньше величина  $\operatorname{Re} \varphi$ , тем быстрее эти решения затухают при  $x \rightarrow \infty$ . Поэтому домини-

ирующими решениями (при условии, что соответствующие коэффициенты не равны нулю) являются те, у которых наибольшие значения  $\operatorname{Re} \gamma$ , т. е. наиболее близкие к началу координат. Уравнению (15.3) соответствует характеристическое уравнение  $\gamma + \gamma e^{-\gamma} = 0$  с решением  $\gamma = -0.31813 \pm 1.33724i$ , а для (15.4) характеристическое уравнение  $\gamma + 1.4e^{-\gamma} = 0$  имеет решение  $\gamma = -0.08170 \pm 1.51699i$ . Это очень хорошо объясняет асимптотическое поведение решений на рис. 15.1 и 15.2.

Для случая матричных уравнений

$$\dot{y}(x) = Ay(x) + By(x-1),$$

где матрицы  $A$  и  $B$  не приводятся одновременно к диагональному виду, полагаем  $y(x) = ve^{\gamma x}$ , где  $v \neq 0$  — некоторый постоянный вектор, и, подставляя в уравнение, получаем

$$\gamma v = Av + Be^{-\gamma v}.$$

Это уравнение имеет нетривиальные решения только при условии, что  $\gamma$  удовлетворяет характеристическому уравнению для этого более общего случая:

$$\det(\gamma I - A - Be^{-\gamma}) = 0. \quad (15.8'')$$

Решения уравнения (15.8'') расположены в комплексной плоскости подобно решениям уравнения (15.8), только теперь в каждой полосе шириной  $2\pi$  вместо одной точки будут находиться  $r = \operatorname{rank}(B)$  точек.

## Устойчивость

Все решения уравнения (15.6) *устойчивы* при  $x \rightarrow \infty$ , если все характеристические корни уравнения (15.8) имеют отрицательную вещественную часть. Этот результат можно получить как из указанной выше теоремы о разложении, так и из теории преобразования Лапласа [см., например, Беллман и Кук (1963), гл. 1], между которыми, впрочем, имеется тесная связь.

Чтобы найти границу области устойчивости, будем искать такие значения  $(\lambda, \mu)$ , при которых первое решение  $\gamma$  переходит через мнимую ось, т. е.  $\gamma = i\theta$ , где  $\theta$  вещественно. Подставляя в (15.8), получим

$$\begin{aligned} \lambda &= -\mu && \text{при } \theta = 0 \text{ (вещественное } \gamma\text{),} \\ \lambda &= i\theta - \mu e^{-i\theta} && \text{при } \theta \neq 0. \end{aligned}$$

Во втором случае, разделяя вещественную и мнимую части, для вещественных  $\lambda$  и  $\mu$  имеем

$$\lambda = \frac{\theta \cos \theta}{\sin \theta}, \quad \mu = -\frac{\theta}{\sin \theta}.$$

Эти линии нанесены на рис. 15.3. Они разделяют плоскость  $(\lambda, \mu)$  на область устойчивости и область неустойчивости решений уравнения (15.6). Этот результат получил Хейес (1950).

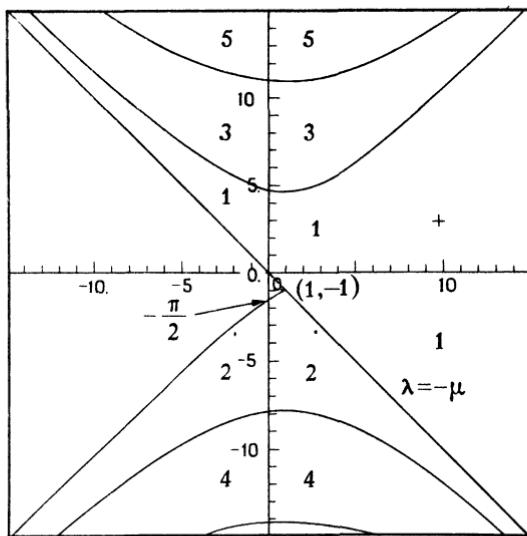


Рис. 15.3. Область устойчивости для уравнения  $y'(x) = \lambda y(x) + \mu y(x - 1)$ .

Положив  $\theta = \pi/2$ , мы найдем, что решения уравнения  $y'(x) = \mu y(x - 1)$  будут *устойчивыми* при

$$-\frac{\pi}{2} \leq \mu \leq 0 \quad (15.9a)$$

и *неустойчивыми* при

$$\mu < -\frac{\pi}{2} \text{ и при } \mu > 0. \quad (15.9b)$$

### Пример из динамики популяций

«Лорд Черуэлл привлек мое внимание к одному уравнению, эквивалентному уравнению (8) [здесь уравнение (15.12)] с  $a = \ln 2$ , которое он встретил, занимаясь приложением вероятностных методов к задаче о распределении простых чисел. Я признателен ему за то, что таким образом он познакомил меня с интересной проблемой.»  
 (Э. М. Райт, 1945)

Мы хотим теперь продемонстрировать обсуждавшиеся выше явления, а также возможности наших программ на нескольких примерах прикладных задач. Впечатляющий список различных

приложений дифференциальных уравнений с запаздывающим аргументом во всевозможных областях науки приведен в книге Драйвера (1977).

Пусть  $y(x)$  представляет численность популяции некоторого вида, и рассматривается ее изменение во времени. Простейшая модель неограниченного экспоненциального роста  $y' = \lambda y$  была уже давно заменена гипотезой, что темп роста уменьшается с увеличением количества особей  $y$  из-за болезней и недостатка пищи и места, что приводит к следующей модели:

$$y'(x) = k(a - y(x)) \cdot y(x) \quad (15.10)$$

[Верхюльст (1845), Пёрл (1922)]. «Nous donnerons le nom *logistique* à la courbe caractérisée par l'équation précédente»<sup>1)</sup> (Верхюльст). Это уравнение решается в элементарных функциях (упр. 1). Все решения с начальными условиями  $y_0 > 0$  при  $x \rightarrow \infty$  асимптотически стремятся к  $a$ . Если предположить, что скорость роста зависит от численности предыдущего поколения, уравнение (15.10) переходит в уравнение с запаздывающим аргументом:

$$y'(x) = k(a - y(x - \tau)) \cdot y(x) \quad (15.11)$$

[Каннингем (1954), Райт (1955), Какутани и Маркус (1958)]. Подставляя в (15.11) новую функцию  $z(x) = kty(tx)$  и после преобразований переобозначив снова  $z$  через  $y$  и  $\tau$  через  $a$ , получим уравнение

$$y'(x) = (a - y(x - 1)) \cdot y(x). \quad (15.12)$$

Оно имеет точку равновесия  $y(x) = a$ . Подстановкой  $y(x) = a + z(x)$  и последующей линеаризацией можно получить уравнение  $z'(x) = -az(x - 1)$ , и условие (15.9) показывает, что эта точка равновесия локально устойчива при  $0 < a \leq \pi/2$ . Следовательно, характеристическое уравнение (в данном случае  $\gamma + ae^{-\gamma} = 0$ ) имеет два вещественных решения тогда и только тогда, если  $a < 1/e = 0,368$ . При этом возможны монотонные решения уравнения (15.12); в противном случае они носят колебательный характер. При  $a > \pi/2$  равновесное решение неустойчиво: рождается периодический предельный цикл.

Представленные на рис. 15.4 решения рассчитаны тоже с помощью приведенных в приложении программ RETARD, STORE, YLAG с подпрограммой FCN вида

$$F(1) = (A - YLAG(1, X - 1., PHI)) * Y(1)$$

для значений  $A = 0.35, 0.5, 1., 1.4$  и  $1.6$ .

<sup>1)</sup>«Мы назвали *логистической* кривую, характеризуемую предыдущим уравнением». — Франц.

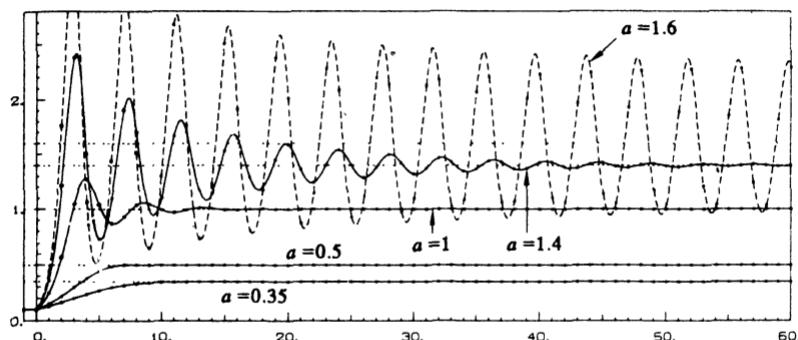


Рис. 15.4. Решение задачи популяционной динамики (15.12).

### Моделирование эпидемии

«De tous ceux qui ont traité cette matière, c'est sans contredit M. de la Condamine qui l'a fait avec plus de succès. Il est déjà venu à bout de persuader la meilleure partie du monde raisonnable de la grande utilité de l'inoculation: quant aux autres, il serait inutile de vouloir employer la raison avec eux: puisqu'ils n'agissent pas par principes. Il faut les conduire comme des enfants vers leur mieux...»

«Из всех, кто занимался этим предметом, наибольшего успеха, бесспорно, добился г-н де ля Кондамин. Он уже окончательно убедил большинство здравомыслящих людей в большой пользе прививок; что до остальных, то бесполезно надеяться повлиять на них доводами рассудка, потому что они в принципе непоколебимы. Их следует, как детей, вести к их благу...»  
(Даниил Бернулли, 1760)

Даниил Бернулли («Docteur en medicine, Professeur de Physique en l'Université de Bâle, Associé étranger de l'Academie des Sciences»<sup>1)</sup>) был первым, кто применил дифференциальное исчисление для моделирования развития инфекционных заболеваний в своей статье о вакцинации против оспы (1760). В начале нашего столетия математическое моделирование эпидемий вновь привлекло интерес ученых. В конечном счете это привело к появлению классической модели Кермака—Макендрика (1927). Пусть  $y_1(x)$  является мерой восприимчивой части населения,  $y_2(x)$  —

<sup>1)</sup> «Доктор медицины, профессор физики Базельского университета, иностранный член Академии наук — Франц.

зараженной, и  $y_3(x)$  — невосприимчивой (например, приобретшей иммунитет) части. Естественно предположить, что число вновь заражаемых в единицу времени людей пропорционально произведению  $y_1(x)y_2(x)$ , как в бимолекулярных химических реакциях (см. разд. I.16). Если, кроме того, предположить, что скорость роста числа невосприимчивых людей пропорциональна числу зараженных, мы придем к следующей модели:

$$\begin{aligned} y'_1 &= -y_1 y_2, \quad y'_2 = y_1 y_2 - y_2, \quad y'_3 = y_2 \end{aligned} \quad (15.13)$$

(для простоты мы выбрали все константы скоростей равными единице). Эту систему можно проинтегрировать элементарным

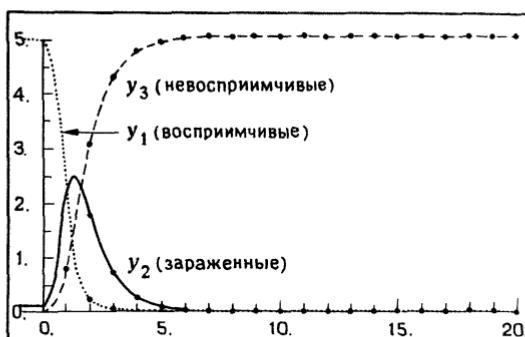


Рис. 15.5. Решение модели Кермака—Макендики (15.13).

методами (разделите второе уравнение на первое и решите  $dy_2/dy_1 = -1 + 1/y_1$ ). На рис. 15.5 показано численное решение с начальными условиями  $y_1(0) = 5$ ,  $y_2(0) = 0.1$ ,  $y_3(0) = 0$ . Видно, как вспыхивает и разражается эпидемия, но в конце концов каждый приобретает иммунитет, и далее ничего не происходит.

Но если допустить, что приобретенный иммунитет через определенное время  $\tau$  теряется, и люди становятся вновь восприимчивыми к заболеванию, мы получим периодические вспышки болезни. Если ввести еще инкубационный период  $\tau_2$ , то вместо (15.13) мы получим следующую модель:

$$\begin{aligned} y'_1(x) &= -y_1(x)y_2(x-1) + y_2(x-10), \\ y'_2(x) &= y_1(x)y_2(x-1) - y_2(x), \\ y'_3(x) &= y_2(x) - y_2(x-10) \end{aligned} \quad (15.14)$$

(здесь взято  $\tau = 10$ ,  $\tau_2 = 1$ ). Решение системы (15.14) для начальных значений  $y_1(x) = 5$ ,  $y_2(x) = 0.1$ ,  $y_3(x) = 1$  при  $x \leq 0$  показано на рис. 15.6 и описывает периодические вспышки болезни.

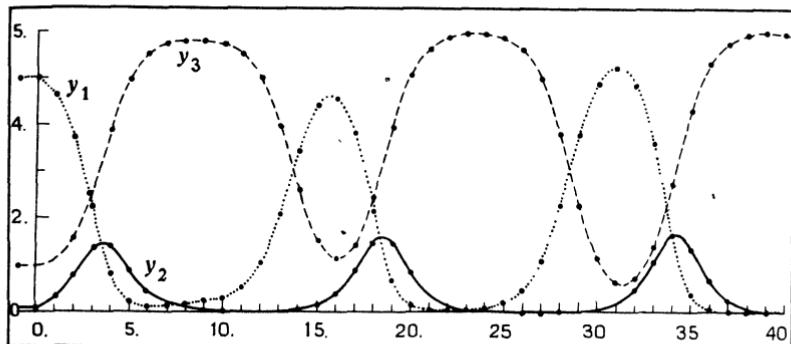
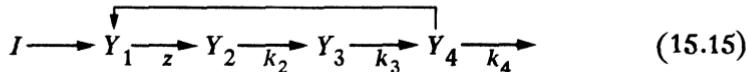


Рис. 15.6. Периодические вспышки заболевания, модель (15.14).

### Пример из кинетики ферментативных реакций

Следующий пример более сложен, чем предыдущие. Он описывает кинетику реакций, где катализатором служит фермент (энзим) [Окамото и Хаяси (1984)]. Рассмотрим следующую цепочку реакций:



Здесь  $I$  — внешний субстрат, запас которого поддерживается постоянным; совместное действие  $n$  молекул конечного продукта  $Y_4$  ингибирует стадию реакции  $Y_1 \rightarrow Y_2$ , так что ее константа скорости имеет вид

$$z = \frac{k_1}{1 + \alpha (y_4(x))^n}.$$

Обычно считают, что молекула ингибитора должна перемещаться в место нахождения регулирующего энзима посредством таких процессов, как диффузия или активный перенос. Рассматривая эти медленные процессы как причину запаздывания по времени, можно получить следующую модель:

$$\begin{aligned} y'_1(x) &= I - zy_1(x), \\ y'_2(x) &= zy_1(x) - y_2(x), \\ y'_3(x) &= y_2(x) - y_3(x), \\ y'_4(x) &= y_3(x) - 0.5y_4(x), \\ z &= \frac{1}{1 + 0.0005 (y_4(x-4))^8}. \end{aligned} \quad (15.16)$$

Эта система имеет точку равновесия  $zy_1 = y_2 = y_3 = 0.5y_4 = I$ . Линеаризация ее в окрестности равновесной точки дает

$$\begin{aligned} y'_1(x) &= -c_1y_1(x) + c_2y_4(x-4), \\ y'_2(x) &= c_1y_1(x) - y_2(x) - c_2y_4(x-4), \\ y'_3(x) &= y_2(x) - y_3(x), \\ y'_4(x) &= y_3(x) - 0.5y_4(x). \end{aligned} \quad (15.17)$$

Здесь введено обозначение  $c_1 := y_1 = I(1 + 0.004I^3)$  для равновесного значения  $y_1$  и положено  $c_2 = c_1 \cdot I^3 \cdot 0.006$ . Подставляя

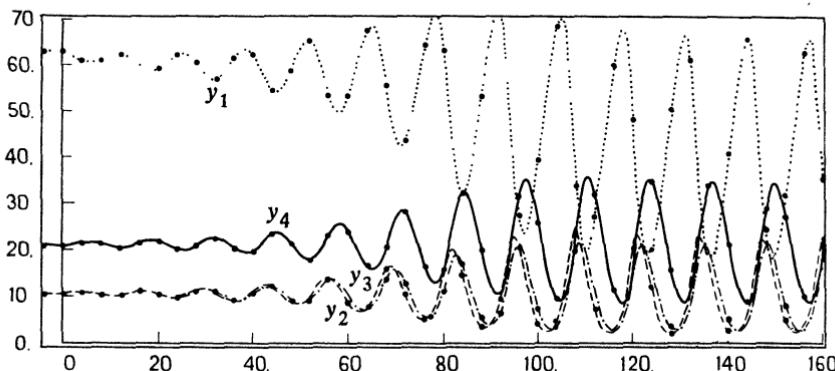


Рис. 15.7. Решения задачи ферментативной кинетики (15.16),  $I = 10.5$ . Начальные значения близки к равновесному состоянию.

сюда вектор  $\mathbf{y}(x) = v \cdot e^{vx}$ , мы придем к характеристическому уравнению (см. уравнение (15.8')), которое после некоторых преобразований примет вид

$$(c_1 + \gamma)(1 + \gamma)^2(0.5 + \gamma) + c_2\gamma e^{-4v} = 0. \quad (15.18)$$

Как и в статье Окамото—Хаяси, положим  $I = 10.5$ . Тогда уравнение (15.18) имеет одну пару комплексных корней, лежащих в  $C^+$ , а именно,

$$\gamma = 0.04246 \pm 0.47666i,$$

и равновесное решение неустойчиво (см. рис. 15.7). Период колебаний решения линеаризованного уравнения равен, таким образом,  $T = 2\pi/0.47666 = 13.18$ . Поэтому решения уравнения (15.16) стремятся к предельному циклу с приблизительно тем же периодом.

### Одна математическая модель в иммунологии

Нашу серию примеров мы завершим рассмотрением модели, которую предложил Марчук (1975) для описания борьбы вирусов  $V(t)$ , антител  $F(t)$  и плазматических клеток  $C(t)$  в человеческом организме, пораженном вирусным заболеванием. Уравнения модели таковы:

$$\begin{aligned}\frac{dV}{dt} &= (h_1 - h_2 F) V, \\ \frac{dC}{dt} &= \xi(m) h_3 F (t - \tau) V(t - \tau) - h_5 (C - 1), \\ \frac{dF}{dt} &= h_4 (C - F) - h_6 F V.\end{aligned}\quad (15.19)$$

Первое уравнение — такое же, как в модели Лотки—Вольтерры для системы хищник—жертва. Второе уравнение описывает рождение новых плазматических клеток с запаздыванием по времени, вызванное инфекцией, в отсутствие которой второй член порождает равновесие при  $C = 1$ . Третье уравнение моделирует синтез антител плазматическими клетками ( $h_4 C$ ) и их убыль в результате старения ( $-h_6 F$ ) и в результате связывания антигенов ( $-h_6 F V$ ). Наконец, множитель  $\xi(m)$  имеет вид

$$\xi(m) = \begin{cases} 1, & \text{если } m \leq 0.1, \\ (1-m) \frac{10}{9}, & \text{если } 0.1 \leq m \leq 1 \end{cases}$$

и выражает факт замедления процесса образования плазматических клеток, когда организм ослаблен вирусной инфекцией. Характеристика степени ослабления  $m(t)$  определяется четвертым уравнением

$$\frac{dm}{dt} = h_6 V - h_7 m,$$

где первый член выражает ослабление, а второй — восстановление.

Эта модель позволяет, меняя коэффициенты  $h_1, h_2, \dots, h_6$ , моделировать все состояния организма: устойчивое здоровье, неустойчивое здоровье, острые формы заболевания, хронические формы и т. д. (подробнее см. гл. 2 в книге Марчука (1983)). На рис. 15.8 показаны решения, рассчитанные нами по этой модели для следующих значений параметров:  $\tau = 0.5$ ,  $h_1 = 2$ ,  $h_2 = 0.8$ ,  $h_3 = 10^4$ ,  $h_4 = 0.17$ ,  $h_5 = 0.5$ ,  $h_6 = 0.12$ ,  $h_7 = 8$  и при начальных условиях  $V(t) = \max(0, t + 10^{-6})$ , если  $t \leq 0$ ,  $C(0) = 1$ ,  $F(t) = 1$ , если  $t \leq 0$ ,  $m(0) = 0$ . Получаемая картина зависит от

величины  $h_6$ . При  $h_6 = 10$  мы видим полное выздоровление (определенное условием  $V(t) < 10^{-16}$ ), а при  $h_6 = 300$  наблюдаются периодические вспышки болезни, связанные с ослаблением организма:  $m(t)$  достигает значений, близких к единице.

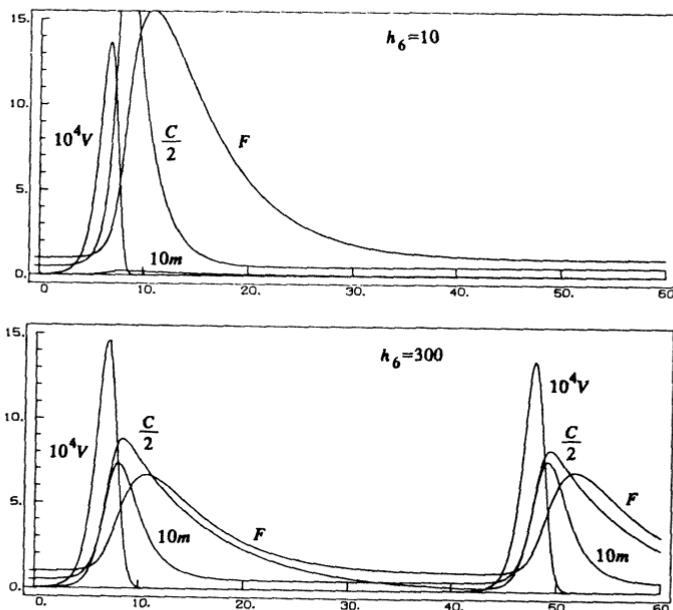


Рис. 15.8. Решения иммунологической модели Марчука.

### Интегро-дифференциальные уравнения

Нередко оказывается, что предполагать поведение системы в момент  $x$  зависящим от ее состояния в какой-то определенный предыдущий момент  $x - \tau$  не очень реалистично: правильней было бы считать, что система испытывает влияние предшествующих состояний, имевших место в течение некоторого периода времени. В таком случае уравнение (15.1) следует заменить, например, уравнением

$$y'(x) = f\left(x, y(x), \int_{x-\tau}^x K(x, \xi, y(\xi)) d\xi\right). \quad (15.20)$$

Численное решение таких задач обходится намного дороже. Если  $K(x, \xi, y)$  обращается в нуль в окрестности диагонали  $x = \xi$ , то можно, в конце концов, использовать программу RETARD, где при каждом обращении к вычислению функции придется вызывать подпрограмму, вычисляющую интеграл.

К счастью, многие интегро-дифференциальные уравнения удается свести к обыкновенным дифференциальным уравнениям или к дифференциальным уравнениям с запаздывающим аргументом, вводя вместо интеграла новые переменные.

**Пример.** [Вольтерра (1934)] Рассмотрим уравнение

$$y'(x) = \left( \varepsilon - \alpha y(x) - \int_0^x y(\xi) k(x - \xi) d\xi \right) \cdot y(x), \quad (15.21)$$

описывающее динамику популяции. Здесь интегральный член описывает уменьшение скорости размножения вследствие загрязнения среды. Пусть теперь, например,  $k(x) = c$ . Тогда, положив

$$\int_0^x y(\xi) d\xi = v(x), \quad y(x) = v'(x),$$

мы получим обыкновенное дифференциальное уравнение

$$v''(x) = [\varepsilon - \alpha v'(x) - cv(x)] v'(x).$$

Такой же метод можно применить и к уравнению (15.20) с «вырожденным ядром», т. е. когда

$$K(x, \xi, y) = \sum_{i=1}^m a_i(x) b_i(\xi, y). \quad (15.22)$$

Подставляя (15.22) в (15.20) и вводя

$$v_i(x) = \int_{x-\tau}^x b_i(\xi, y(\xi)) d\xi, \quad (15.23)$$

мы получим дифференциальное уравнение с запаздывающим аргументом:

$$\begin{aligned} y'(x) &= f(x, y(x), \sum_{i=1}^m a_i(x) v_i(x)), \\ v'_i(x) &= b_i(x, y(x)) - b_i(x - \tau, y(x - \tau)), \\ i &= 1, \dots, m. \end{aligned} \quad (15.20')$$

### Упражнения

- Найдите решение уравнения Верхюльста—Пёрла (15.10).
- Найдите точки равновесия для уравнения Марчука (15.19) и исследуйте их устойчивость.
- Пусть ядро  $k(x)$  в уравнении Вольтерры (15.21) имеет вид

$$k(x) = p(x) e^{-\beta x},$$

где  $p(x)$  — некоторый многочлен. Покажите, что эта задача может быть преобразована в обыкновенное дифференциальное уравнение.

4. Рассмотрите интегро-дифференциальное уравнение

$$y'(x) = f\left(x, y(x), \int_0^x K(x, \xi, y(\xi)) d\xi\right) \quad (15.24)$$

и покажите, что:

а) Для вырожденного ядра (15.22) задача (15.24) эквивалентна обыкновенному дифференциальному уравнению

$$\begin{aligned} y'(x) &= f\left(x, y(x), \sum_{j=1}^m a_j(x) v_j(x)\right), \\ v'_j(x) &= b_j(x, y(x)). \end{aligned} \quad (15.25)$$

б) Применение явного метода Рунге—Кутты  $p$ -го порядка к задаче (15.25) приводит к формулам

$$\begin{aligned} y_{n+1} &= y_n + h \sum_{i=1}^s b_i f(x_n + c_i h, g_i^{(n)}, u_i^{(n)}), \\ g_i^{(n)} &= y_n + h \sum_{l=1}^{i-1} a_{il} f(x_n + c_l h, g_l^{(n)}, u_l^{(n)}), \\ u_i^{(n)} &= F_n(x_n + c_i h) + h \sum_{l=1}^{i-1} a_{il} K(x_n + c_l h, x_n + c_i h, g_l^{(n)}), \end{aligned} \quad (15.26)$$

где

$$F_0(x) = 0,$$

$$F_{n+1}(x) = F_n(x) + h \sum_{i=1}^s b_i K(x, x_n + c_i h, g_i^{(n)})$$

(эти формулы получил Пузе (1963)).

с) При применении метода (15.26) к задаче (15.24) сходимость порядка  $p$  получается и в том случае, если ядро не удовлетворяет требованию (15.22).

*Указание.* Аппроксимируйте ядро вырожденным ядром.

5. [Дзеннаро (1986)]. Для дифференциального уравнения с запаздывающим аргументом (15.1) рассмотрите метод (15.5), в котором формула (15.5c) заменена на

$$\Psi_i^{(n)} = \begin{cases} \varphi(x_n + c_j h - \tau), & \text{если } n < k, \\ q_{n-k}(c_j), & \text{если } n \geq k, \end{cases} \quad (15.5c')$$

где  $q_n(\theta)$  — многочлен, даваемый непрерывным методом Рунге—Кутты (см. разд. II.5):

$$q_n(\theta) = y_n + h \sum_{l=1}^s b_l(\theta) f(x_n + c_l h, g_l^{(n)}, \gamma_l^{(n)}).$$

Докажите, что:

а) Из условий ортогональности

$$\int_0^1 \theta^{q-1} \left[ \gamma(t) \sum_{l=1}^s b_l(\theta) \Phi_l(t) - \theta^{\rho}(t) \right] d\theta = 0 \quad (15.27)$$

$$\text{при } q + \rho(t) \leq p$$

следует сходимость порядка  $p$ , если исходный метод Рунге—Кутты имеет порядок  $p$  для обыкновенных дифференциальных уравнений.

*Указание.* Воспользуйтесь теорией В-рядов и формулой Грёбнера—Алексеева (14.18) из разд. I.14.

б) Если для данного метода Рунге—Кутты многочлены  $b_j(\theta)$  степени не выше  $[(p+1)/2]$  таковы, что

$$\int_0^1 \theta^{q-1} b_j(\theta) d\theta = \frac{1}{q} b_j(1 - c_j^q) \quad (15.28)$$

$$\text{при } q = 1, \dots, [(p-1)/2] \text{ и } b_j(0) = 0, b_j(1) = b_j,$$

то условия (15.27) удовлетворяются. Кроме того, имеют место условия порядка

$$\sum_{l=1}^s b_l(\theta) \Phi_l(t) = \frac{\theta^{\rho}(t)}{\gamma(t)} \quad \text{при } \rho(t) \leq [(p+1)/2].$$

с) Условия (15.28) однозначно определяют вид многочленов  $b_j(\theta)$  степени  $[(p+1)/2]$ .

6. Решите уравнение Вольтерры (15.21) с  $k(x) = c$  и сравните решение с решением задачи «без загрязнения среды» (15.10). Какой популяции живется лучше: при наличии загрязнения или без него?

## Глава III

# Многошаговые методы и общие линейные методы

Эта глава посвящена изучению многошаговых и общих многозначных методов. После прослеживания их исторического развития (Адамс, Нюстрём, Милн, БДФ) мы изучаем в последующих разделах порядок, устойчивость и свойства сходимости этих методов. Сходимость наиболее удобно исследовать в форме одношаговых методов в пространствах высших размерностей. Разделы III.5 и III.6 посвящены методам с переменной длиной шага и методам Нордсика. Затем мы обсуждаем различные имеющиеся программы и сравниваем их характеристики на численных примерах из разд. II.10, а также на некоторых уравнениях высокой размерности. Прежде чем закончить главу разделом о специальных методах для уравнений второго порядка, мы обсудим две темы, для которых развита обширная теория: первая из них — общие линейные методы, включая методы Рунге—Кутты, многошаговые методы и многие их обобщения, а другая — асимптотическое разложение глобальной погрешности таких методов.

### III.1. Классические линейные многошаговые формулы

«... и если бы я располагал только своими собственными возможностями, то мне пришлось бы прервать начатое. Но в этот момент профессор Дж. К. Адамс предоставил мне вполне удовлетворительный метод решения с помощью квадратур дифференциального уравнения Лапласа, теоретически описывающего формы капель жидкости...»

(Ф. Башфорт, 1883)

Еще раньше, чем методы Рунге—Кутты, был рассмотрен другой способ усовершенствования метода Эйлера — методы Адамса. Джон Куч Адамс разработал их для решения задачи, с которой столкнулся Ф. Башфорт при исследовании капиллярных процессов. И задача, и схемы численного интегрирования опубликованы в работе Башфорта (1883). В действительности появление этих методов должно датироваться не позднее чем 1855 годом, так как в этом году Ф. Башфорт обратился в Королевское общество с просьбой о предоставлении государственной субсидии. Он писал: «...но я обязан г-ну Адамсу, предложившему метод решения дифференциального уравнения

$$\frac{\frac{ddz}{du^2}}{\left(1 + \frac{dz^2}{du^2}\right)^{3/2}} + \frac{1}{u} \frac{dz}{du} - 2\alpha z = \frac{2}{b},$$

представленного в виде

$$\frac{b}{p} + \frac{b}{x} \sin \varphi = 2 + 2ab^2 \frac{z}{b} = 2 + \beta \frac{z}{b},$$

поскольку это уравнение описывает теоретическую форму капли с точностью, превосходящей наиболее тонкие измерения».

В отличие от одношаговых методов, в которых численное решение получают только из дифференциального уравнения и начального условия, алгоритм Адамса состоит из двух частей: первая из них — это *стартовая процедура* для определения  $y_1, \dots, y_{k-1}$  (приближенных значений точного решения в точках  $x_0 + h, \dots, x_0 + (k-1)h$ ), а вторая — *многошаговая формула* для получения приближенного значения точного решения  $y(x_0 + kh)$ . Затем эта формула применяется рекурсивно для того, чтобы по численному решению на  $k$  последовательных шагах вычислить  $y(x_0 + (k+1)h)$  и т. д.

Стартовые значения можно получить несколькими способами. Дж. К. Адамс вычислял их с помощью разложения точного решения в ряд Тейлора (как описано в разд. I.8, см. также упр. 2). Другой способ состоит в использовании какого-либо одношагового метода, например метода Рунге—Кутты (см. гл. II). Стартовые значения часто также вычисляют методами Адамса низкого порядка с очень малым шагом.

### Явные методы Адамса

Теперь тем же способом, что и Адамс, выведем первую группу явных многошаговых формул. Для точек сетки введем обозначение  $x_i = x_0 + ih$  и предположим, что нам известны численные приближенные значения  $y_n, y_{n-1}, \dots, y_{n-k+1}$  точного решения  $y(x_n), \dots, y(x_{n-k+1})$  дифференциального уравнения

$$y' = f(x, y), \quad y(x_0) = y_0. \quad (1.1)$$

Адамс рассматривает (1.1) в интегральной форме:

$$y(x_{n+1}) = y(x_n) + \int_{x_n}^{x_{n+1}} f(t, y(t)) dt. \quad (1.2)$$

В правую часть (1.2) входит искомое решение  $y(x)$ . Но поскольку нам известны его приближенные значения  $y_{n-k+1}, \dots, y_n$ , то мы располагаем также и величинами

$$f_i = f(x_i, y_i) \text{ при } i = n - k + 1, \dots, n, \quad (1.3)$$

а поэтому естественно заменить функцию  $f(t, y(t))$  в (1.2) интерполяционным многочленом, проходящим через точки  $\{(x_i, f_i) \mid i = n - k + 1, \dots, n\}$ . Его можно выразить через разности назад вида

$$\nabla^0 f_n = f_n, \quad \nabla^{j+1} f_n = \nabla^j f_n - \nabla^j f_{n-1}$$

следующим образом:

$$p(t) = p(x_n + sh) = \sum_{j=0}^{k-1} (-1)^j \binom{-s}{j} \nabla^j f_n \quad (1.4)$$

[интерполяционная формула Ньютона; см., например, работу Хенричи (1962)]. Тогда численный аналог (1.2) задается формулой

$$y_{n+1} = y_n + \int_{x_n}^{x_{n+1}} p(t) dt,$$

или после подстановки (1.4)

$$y_{n+1} = y_n + h \sum_{j=0}^{k-1} \gamma_j \nabla^j f_n, \quad (1.5)$$

где коэффициенты  $\gamma_j$  удовлетворяют равенству

$$\gamma_j = (-1)^j \int_0^1 \binom{-s}{j} ds \quad (1.6)$$

(их числовые значения см. в табл. 1.1). Далее для этих коэффициентов будет выведено простое рекуррентное соотношение (формула (1.11)).

Таблица 1.1. Коэффициенты явных методов Адамса

$j$	0	1	2	3	4	5	6	7	8
$\gamma_j$	1	$\frac{1}{2}$	$\frac{5}{12}$	$\frac{3}{8}$	$\frac{251}{720}$	$\frac{95}{288}$	$\frac{19\,087}{60\,480}$	$\frac{5257}{17\,280}$	$\frac{1\,070\,017}{3\,628\,800}$

Частные случаи (1.5). Для  $k = 1, 2, 3, 4$ , выразив разности назад через  $f_{n-j}$ , получим следующие формулы:

$$k = 1: y_{n+1} = y_n + hf_n,$$

$$k = 2: y_{n+1} = y_n + h \left[ \frac{3}{2} f_n - \frac{1}{2} f_{n-1} \right], \quad (1.5')$$

$$k = 3: y_{n+1} = y_n + h \left[ \frac{23}{12} f_n - \frac{16}{12} f_{n-1} + \frac{5}{12} f_{n-2} \right],$$

$$k = 4: y_{n+1} = y_n + h \left[ \frac{55}{24} f_n - \frac{59}{24} f_{n-1} + \frac{37}{24} f_{n-2} - \frac{9}{24} f_{n-3} \right].$$

Замечание. Для  $k = 1$  мы имеем явный метод Эйлера.

### Неявные методы Адамса

Формулы (1.5) получены при интегрировании интерполяционного многочлена (1.4) от  $x_n$  до  $x_{n+1}$ , т. е. вне интервала интерполяции  $(x_{n-k+1}, x_n)$ . Хорошо известно, что вне этого интервала интерполяционный многочлен обычно дает довольно плохое приближение. Таким образом, явные методы Адамса не очень точны. Поэтому Адамс исследовал также методы, где (1.4) заменен на интерполяционный многочлен, дополнительно использующий точку  $(x_{n+1}, f_{n+1})$ , т. е. на

$$p^*(t) = p^*(x_n + sh) = \sum_{j=0}^k (-1)^j \binom{-s+1}{j} \nabla^j f_{n+1}. \quad (1.7)$$

Подставляя этот многочлен в (1.2), получим следующий неявный метод:

$$y_{n+1} = y_n + h \sum_{l=0}^k \gamma_l^* \nabla^l f_{n+l}, \quad (1.8)$$

где коэффициенты  $\gamma_l^*$  удовлетворяют равенству

$$\gamma_l^* = (-1)^l \int_0^1 \binom{-s+1}{l} ds, \quad (1.9)$$

а для  $l \leq 8$  приведены в табл. 1.2. Для этих коэффициентов также можно вывести простое рекуррентное соотношение (упр. 3).

Таблица 1.2. Коэффициенты неявных методов Адамса

$k$	0	1	2	3	4	5	6	7	8
$\gamma_l^*$	$1 - \frac{1}{2}$	$-\frac{1}{12}$	$-\frac{1}{24}$	$-\frac{19}{720}$	$-\frac{3}{160}$	$-\frac{863}{60480}$	$-\frac{275}{24192}$	$-\frac{33953}{3628800}$	

Полученные таким образом формулы имеют общий вид:

$$y_{n+1} = y_n + h [\beta_k f_{n+1} + \dots + \beta_0 f_{n-k+1}]. \quad (1.8')$$

Первые из них суть

$$k = 0: y_{n+1} = y_n + h f_{n+1},$$

$$k = 1: y_{n+1} = y_n + h \left[ \frac{1}{2} f_{n+1} + \frac{1}{2} f_n \right], \quad (1.8'')$$

$$k = 2: y_{n+1} = y_n + h \left[ \frac{5}{12} f_{n+1} + \frac{8}{12} f_n - \frac{1}{12} f_{n-1} \right],$$

$$k = 3: y_{n+1} = y_n + h \left[ \frac{9}{24} f_{n+1} + \frac{19}{24} f_n - \frac{5}{24} f_{n-1} + \frac{1}{24} f_{n-2} \right].$$

В частных случаях  $k = 0$  и  $k = 1$  мы имеем неявный метод Эйлера и правило трапеций соответственно. В действительности это одношаговые методы, и они уже рассмотрены в гл. II.7.

Вся группа методов (1.8) более точно, чем (1.5), аппроксимирует решение дифференциального уравнения. Мы подробно остановимся на этом вопросе после введения понятий порядка и константы погрешности (разд. III.2). Повышение точности получено за счет того, что формула (1.8) определяет  $y_{n+1}$  только неявно. Таким образом, в общем случае на каждом шаге требуется решать нелинейное уравнение. Это можно сделать с помощью ме-

тогда простой итерации. Пусть  $y_{n+1}^{(0)}$  аппроксимирует  $y_{n+1}$ ; определим

$$y_{n+1}^{(l)} = y_n + h [\beta_k f(x_{n+1}, y_{n+1}^{(l-1)}) + \beta_{k-1} f_n + \dots + \beta_0 f_{n-k+1}]. \quad (1.10)$$

Тогда, если шаг  $h$  достаточно мал, то при  $l \rightarrow \infty$   $y_{n+1}^{(l)}$  сходится к численному решению (1.8)  $y_{n+1}$ . На практике в качестве  $y_{n+1}^{(0)}$  используют значение, полученное по явной формуле (1.5) («предиктор»), а затем выполняют одну итерацию формулы (1.10) («корректор»). Этот метод использован в работах Мултона (1926) и Милна (1926). Дж. К. Адамс решал неявное уравнение (1.8) методом Ньютона, как это принято теперь в случае жестких уравнений (см. т. II).

*Замечание.* Формулу (1.5) часто приписывают Адамсу и Башфорту, а (1.8) — Адамсу и Мултону [Мултон (1926)]. В действительности обе формулы принадлежат Адамсу.

### Рекуррентные соотношения для $\gamma_i$

Используя эйлеров метод производящих функций [см., например, работу Хенричи (1962)], можно вывести простые рекуррентные соотношения для  $\gamma_i$ . Обозначим через  $G(t)$  ряд

$$G(t) = \sum_{j=0}^{\infty} \gamma_j t^j.$$

Из определения  $\gamma_j$  и биномиальной теоремы получим

$$\begin{aligned} G(t) &= \sum_{j=0}^{\infty} (-t)^j \int_0^1 \binom{-s}{j} ds = \int_0^1 \sum_{j=0}^{\infty} (-t)^j \binom{-s}{j} ds = \\ &= \int_0^1 (1-t)^{-s} ds = -\frac{t}{(1-t) \log(1-t)}. \end{aligned}$$

Это равенство можно записать в виде

$$-\frac{\log(1-t)}{t} G(t) = \frac{1}{1-t},$$

или

$$\left(1 + \frac{1}{2}t + \frac{1}{3}t^2 + \dots\right)(\gamma_0 + \gamma_1 t + \gamma_2 t^2 + \dots) = 1 + t + t^2 + \dots.$$

Сравнивая коэффициенты при  $t^m$ , получим требуемое рекуррентное соотношение

$$\gamma_m + \frac{1}{2}\gamma_{m-1} + \frac{1}{3}\gamma_{m-2} + \dots + \frac{1}{m+1}\gamma_0 = 1. \quad (1.11)$$

Аналогичная формула для неявных методов Адамса рассматривается в упр. 3.

## Явные методы Ньюстрёма

«Приближенное интегрирование нашло, особенно в последнее время, обширную область применения в точных науках и в технике».

(Е. Дж. Ньюстрём, 1925)

В своей обзорной статье по численному интегрированию дифференциальных уравнений (на которую мы уже ссылались в разд. II.13) Ньюстрём (1925) также представил новый класс многошаговых методов. Вместо (1.2) он рассматривает интегральное уравнение

$$y(x_{n+1}) = y(x_{n-1}) + \int_{x_{n-1}}^{x_{n+1}} f(t, y(t)) dt. \quad (1.12)$$

Точно так же, как Адамс, он заменяет неизвестную функцию  $f(t, y(t))$  многочленом  $p(t)$ , заданным с помощью (1.4), и получает формулу

$$y_{n+1} = y_{n-1} + h \sum_{j=0}^{k-1} \alpha_j \nabla^j f_n \quad (1.13)$$

с коэффициентами

$$\alpha_j = (-1)^j \int_{-1}^1 \left( \frac{-s}{j} \right) ds. \quad (1.14)$$

Первые шесть из этих коэффициентов приведены в табл. 1.3. Е. Дж. Ньюстрём рекомендовал формулы (1.13) по той причине, что коэффициенты  $\alpha_j$  больше подходили ему для расчетов, чем  $\gamma_j$  из (1.5). Этот, несомненно, разумный при ручном счете совет теряет смысл в случае вычислений на ЭВМ.

Таблица 1.3. Коэффициенты явных методов Ньюстрёма

$j$	0	1	2	3	4	5	6
$\alpha_j$	2	0	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{29}{90}$	$\frac{14}{45}$	$\frac{1139}{3780}$

*Частные случаи.* При  $k = 1$  имеем формулу

$$y_{n+1} = y_{n-1} + 2hf_n. \quad (1.13')$$

Этот простейший двухшаговый метод называется *правилом средней точки*. Его симметричность оказалась крайне полезным свой-

ством для экстраполяционных схем из разд. II.9. Случай  $k = 2$  не дает ничего нового, поскольку  $\kappa_1 = 0$ . Для  $k = 3$  получим

$$y_{n+1} = y_{n-1} + h \left[ \frac{7}{3} f_n - \frac{2}{3} f_{n-1} + \frac{1}{3} f_{n-2} \right]. \quad (1.13'')$$

### Методы Милна—Симпсона

Снова рассмотрим интегральное уравнение (1.12). Но теперь заменим подынтегральное выражение на многочлен  $p^*(t)$  из (1.7), в котором дополнительно к  $f_n, \dots, f_{n-k+1}$  при интерполяции используется также значение  $f_{n+1}$ . Действуя, как обычно, получим на этот раз неявные формулы

$$y_{n+1} = y_{n-1} + h \sum_{l=0}^k \kappa_l^* \nabla^l f_{n+1}. \quad (1.15)$$

Коэффициенты  $\kappa_l^*$  определяются формулой

$$\kappa_l^* = (-1)^l \int_{-1}^1 \binom{-s+1}{l} ds, \quad (1.16)$$

а числовые значения для первых из них приведены в табл. 1.4.

Таблица 1.4. Коэффициенты методов Милна—Симпсона

$l$	0	1	2	3	4	5	6
$\kappa_l^*$	2	-2	$-\frac{1}{3}$	0	$-\frac{1}{90}$	$-\frac{1}{90}$	$-\frac{37}{3780}$

Запишем методы для нескольких значений  $k$ , выразив разности назад в (1.15) через  $f_{n-j}$ :

$$k = 0: y_{n+1} = y_{n-1} + 2hf_{n+1},$$

$$k = 1: y_{n+1} = y_{n-1} + 2hf_n,$$

$$k = 2: y_{n+1} = y_{n-1} + h \left[ \frac{1}{3} f_{n+1} + \frac{4}{3} f_n + \frac{1}{3} f_{n-1} \right],$$

$$k = 4: y_{n+1} = y_{n-1} + h \left[ \frac{29}{90} f_{n+1} + \frac{44}{90} f_n + \frac{4}{90} f_{n-1} + \frac{4}{90} f_{n-2} - \frac{1}{90} f_{n-3} \right]. \quad (1.15')$$

В частном случае  $k = 0$  мы имеем в точности неявный метод Эйлера с шагом  $2h$ . Для  $k = 1$  получим выведенное ранее правило средней точки. Частный случай  $k = 2$  дает интересный метод,

известный как *метод Милна* [см. работы Милна (1926), (1970)]. Это прямое обобщение правила Симпсона.

Было исследовано и много других подобных методов. Все они основаны на интегральном уравнении вида

$$y(x_{n+1}) = y(x_{n-l}) + \int_{x_{n-l}}^{x_{n+1}} f(t, y(t)) dt, \quad (1.17)$$

в котором  $f(t, y(t))$  заменяется интерполяционным многочленом  $p(t)$  [формула (1.4)] или  $p^*(t)$  [формула (1.7)]. Например, для  $l = 3$  получим

$$y_{n+1} = y_{n-3} + h \left[ \frac{8}{3} \hat{f}_n - \frac{4}{3} \hat{f}_{n-1} + \frac{8}{3} \hat{f}_{n-2} \right]. \quad (1.18)$$

Данную формулу Милн (1926) использовал в качестве «предиктора» для своего метода: для решения неявного уравнения (1.15') он применяет одну или две итерации метода простой итерации с начальным приближением, вычисленным по формуле (1.18).

### Методы, основанные на дифференцировании

«Моя фамилия Gear. — Как? — Gear, dshii, ii, ay, аге. — Jiea?»  
(В парижской гостинице.)

Все рассмотренные до сих пор многошаговые формулы основаны на численном интегрировании, т. е. интеграл в (1.17) аппроксимируется с помощью какой-либо квадратурной формулы. Теперь перейдем к многошаговым методам, воплотившим совсем другую идею — численного дифференцирования искомой функции.

Предположим, что известны приближенные значения  $y_{n-k+1}, \dots, y_n$  решения дифференциального уравнения (1.1). Чтобы вывести формулу для  $y_{n+1}$ , рассмотрим интерполяционный многочлен  $q(x)$ , проходящий через точки  $\{(x_i, y_i) | i = n - k + 1, \dots, n + 1\}$ . Как и многочлен (1.7), его можно выразить через разности назад, а именно

$$q(x) = q(x_n + sh) = \sum_{j=0}^k (-1)^j \binom{-s+1}{j} \nabla^j y_{n+1}. \quad (1.19)$$

Определим теперь неизвестное значение  $y_{n+1}$  таким образом, чтобы многочлен  $q(x)$  удовлетворял дифференциальному уравнению по крайней мере в одном узле сетки, т. е.

$$q'(x_{n+1-r}) = f(x_{n+1-r}, y_{n+1-r}). \quad (1.20)$$

Для  $r = 1$  получим явные формулы. При  $k = 1$  и  $k = 2$  они эквивалентны явному методу Эйлера и правилу средней точки соответственно. В случае  $k = 3$  формула имеет вид

$$\frac{1}{3} y_{n+1} + \frac{1}{2} y_n - y_{n-1} + \frac{1}{6} y_{n-2} = h\dot{f}_n. \quad (1.21)$$

Однако она неустойчива, как и все остальные формулы при  $k > 3$  (см. разд. III.3), и поэтому непригодна для расчетов.

Значительно интереснее формулы, получаемые из (1.20) при  $r = 0$ . Это *неявные* формулы

$$\sum_{j=0}^k \delta_j^* \nabla^j y_{n+1} = h\dot{f}_{n+1} \quad (1.22)$$

с коэффициентами

$$\delta_j^* = (-1)^j \left. \frac{d}{ds} \binom{-s+1}{j} \right|_{s=1},$$

которые после дифференцирования биномиального коэффициента

$$(-1)^j \binom{-s+1}{j} = \frac{1}{j!} (s-1) s (s+1) \dots (s+j-2)$$

приводятся к виду

$$\delta_0^* = 0, \quad \delta_j^* = \frac{1}{j!} \quad \text{при } j \geq 1. \quad (1.23)$$

Поэтому (1.22) сводится к формуле

$$\sum_{j=1}^k \frac{1}{j!} \nabla^j y_{n+1} = h\dot{f}_{n+1}. \quad (1.22')$$

Эти многошаговые методы, известные как *формулы дифференцирования назад* (ФДН- или BDF-методы), широко используются для интегрирования жестких дифференциальных уравнений (см. т. II). Они введены в работе Кёртиса и Хиршфельдера (1952); Митчелл и Крэггс (1953) называют их «стандартными пошаговыми методами».

Для полноты изложения приведем эти формулы, выразив в них разности назад через  $y_{n-j}$ :

$$k = 1: y_{n+1} - y_n = h\dot{f}_{n+1},$$

$$k = 2: \frac{3}{2} y_{n+1} - 2y_n + \frac{1}{2} y_{n-1} = h\dot{f}_{n+1},$$

$$k = 3: \frac{11}{6} y_{n+1} - 3y_n + \frac{3}{2} y_{n-1} - \frac{1}{3} y_{n-2} = h\dot{f}_{n+1},$$

$$k = 4: \frac{25}{12} y_{n+1} - 4y_n + 3y_{n-1} - \frac{4}{3} y_{n-2} + \frac{1}{4} y_{n-3} = h\dot{f}_{n+1},$$

$$k = 5: \frac{137}{60} y_{n+1} - 5y_n + 5y_{n-1} - \frac{10}{3} y_{n-2} + \frac{5}{4} y_{n-3} - \frac{1}{5} y_{n-4} = hf_{n+1}, \quad (1.22'')$$

$$k = 6: \frac{147}{60} y_{n+1} - 6y_n + \frac{15}{2} y_{n-1} - \frac{20}{3} y_{n-2} + \frac{15}{4} y_{n-3} - \frac{6}{5} y_{n-4} + \frac{1}{6} y_{n-5} = hf_{n+1}.$$

При  $k > 6$  BDF-методы неустойчивы (см. разд. III.3).

### Упражнения

- Пусть заданы дифференциальное уравнение  $y' = y^2$ ,  $y(0) = 1$  и точные стартовые значения  $y_i = 1/(1 - x_i)$  при  $i = 0, 1, \dots$

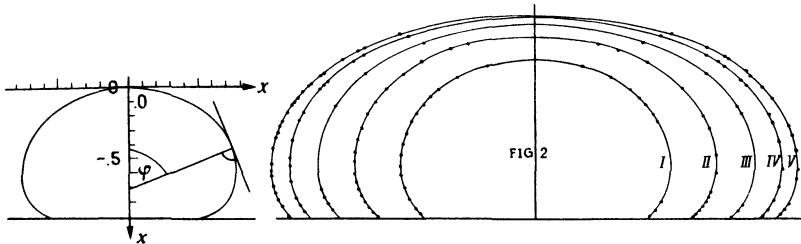


Рис. 1.1. Решение дифференциального уравнения (1.24) и иллюстрация из книги Башфорта.

- ...,  $k = 1$ . Применить методы Адамса и изучить выражение  $y(x_k) — y_k$  при малых шагах.
- Рассмотрим дифференциальное уравнение из начала данного раздела. Оно описывает форму капли и может быть задано в виде [Ф. Башфорт (1883); та же задача, что и в упр. 2 из разд. II.1, но в другой системе координат]

$$\frac{dx}{d\varphi} = \rho \cos \varphi, \quad \frac{dz}{d\varphi} = \rho \sin \varphi, \quad (1.24)$$

где

$$\frac{1}{\rho} + \frac{\sin \varphi}{x} = 2 + \beta z. \quad (1.25)$$

Можно считать  $\rho$  функцией координат  $x$  и  $z$  и понимать под ней радиус кривизны. Через  $\varphi$  обозначен угол между нормалью к кривой и осью  $z$  (рис. 1.1). Заданы следующие начальные ус-

ловия:  $x(0) = 0$ ,  $z(0) = 0$ ,  $\rho(0) = 1$ . Решить это уравнение с помощью предложенного Адамсом способа:

а) Предположим, что  $\rho$  можно представить в виде

$$\rho = 1 + b_2 \varphi^2 + b_4 \varphi^4 + \dots$$

Подставив это выражение в (1.24), в результате интегрирования получим первые члены разложения  $x(\varphi)$  и  $z(\varphi)$  в ряд Тейлора с коэффициентами, зависящими от  $b_2$ ,  $b_4$ , ..., и т. д. Эти параметры можно вычислить из (1.25), приравняв коэффициенты при  $\varphi^m$ . Таким образом, получают решение при малых значениях  $\varphi$  (стартовые значения).

б) с помощью одной из предложенных многошаговых формул вычислить решение для заданного  $\beta$  (например,  $\beta = 3$ ) на интервале  $[0, \pi]$ .

3. Доказать, что определенные в (1.9) коэффициенты  $\gamma_i^*$  удовлетворяют соотношениям

$$\gamma_0^* = 1,$$

$$\gamma_m^* + \frac{1}{2} \gamma_{m-1}^* + \frac{1}{3} \gamma_{m-2}^* + \dots + \frac{1}{m+1} \gamma_0^* = 0 \quad \text{при } m \geq 1.$$

4. Пусть коэффициенты  $\kappa_j$ ,  $\kappa_j^*$ ,  $\gamma_j$ ,  $\gamma_j^*$  определены в (1.14), (1.16), (1.6), (1.9) соответственно. Показать, что выполняются равенства

$$\kappa_j = 2\gamma_j - \gamma_{j-1}, \quad \kappa_j^* = 2\gamma_j^* - \gamma_{j-1}^* \quad \text{при } j \geq 0$$

(где  $\gamma_{-1} = \gamma_{-1}^* = 0$ ).

*Указание:* разбив интеграл в (1.14) на два слагаемых, можно установить равенство  $\kappa_j = \gamma_j + \gamma_j^*$ . Соотношение  $\gamma_j^* = \gamma_j - \gamma_{j-1}$  получается с помощью известного тождества для биномиальных коэффициентов.

## III.2. Локальная погрешность и условия порядка

«Вы знаете, я ведь многошаговый человек... и никому не говорите, что моей первой программой для первой шведской ЭВМ была реализация метода Рунге—Кутты...»

(Г. Далквист, 1982, после нескольких бокалов вина; печатается с его разрешения)

Исследования Далквиста (1956, 1959) положили начало общей теории многошаговых методов, а широкую известность она приобрела благодаря книге Хенричи (1962). Для всех многошаговых формул из предыдущего раздела общим является линейное вхождение в них значений численного решения  $y_i$ , а также  $f_i$ . Поэтому рассмотрим разностное уравнение общего вида

$$\alpha_h y_{n+h} + \alpha_{h-1} y_{n+h-1} + \dots + \alpha_0 y_n = h (\beta_h f_{n+h} + \dots + \beta_0 f_n), \quad (2.1)$$

включающее все эти формулы как частные случаи. В (2.1) параметры  $\alpha_i$ ,  $\beta_i$  — вещественные числа,  $h$  обозначает величину шага, а

$$f_i = f(x_i, y_i), \quad x_i = x_0 + ih.$$

В этой главе мы будем считать, что выполнены следующие условия:

$$\alpha_h \neq 0, \quad |\alpha_0| + |\beta_0| > 0. \quad (2.2)$$

Первое предположение означает, что неявное уравнение (2.1) разрешимо относительно  $y_{n+h}$  по крайней мере при достаточно малых  $h$ . Второе соотношение из (2.2) всегда можно получить, уменьшив при необходимости индекс  $k$ .

Формула (2.1) называется *линейным многошаговым*, а точнее, *линейным k-шаговым методом*. Будем различать явные ( $\beta_k = 0$ ) и неявные ( $\beta_k \neq 0$ ) многошаговые методы.

### Локальная погрешность многошагового метода

Поскольку численное решение, полученное многошаговым методом, зависит не только от задачи Коши (1.1), но и от выбора стартовых значений, определение локальной погрешности в этом случае будет не таким простым, как для одношаговых методов (сравнить с разд. II.2 и II.3).

**Определение 2.1.** Локальной погрешностью многошагового метода (2.1) называется величина

$$y(x_k) - y_k,$$

где  $y(x)$  — точное решение задачи  $y' = f(x, y)$ ,  $y(x_0) = y_0$ , а  $y_k$  — численное решение, полученное по формуле (2.1) при точных стартовых значениях  $y_l = y(x_l)$ ,  $l = 0, 1, \dots, k-1$ .

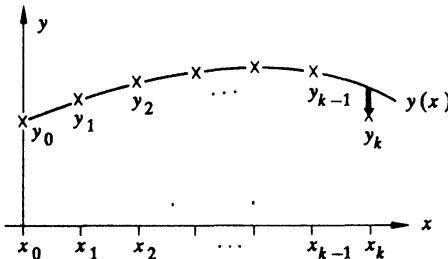


Рис. 2.1. Иллюстрация локальной погрешности.

В случае  $k = 1$  это определение совпадает с определением локальной погрешности для одношаговых методов. Чтобы показать связь с другими ее возможными определениями, сопоставим формуле (2.1) линейный разностный оператор

$$L(y, x, h) = \sum_{i=0}^k [\alpha_i y(x + ih) - h\beta_i y'(x + ih)]. \quad (2.3)$$

Здесь  $y(x)$  — какая-либо дифференцируемая функция, которая задана на интервале, включающем значения  $x + ih$ ,  $i = 0, 1, \dots, k$ .

**Лемма 2.2.** Рассмотрим дифференциальное уравнение (1.1) с непрерывно дифференцируемой функцией  $f(x, y)$  и решением  $y(x)$ . Тогда для локальной погрешности выполняется равенство

$$y(x_k) - y_k = \left( \alpha_k I - h\beta_k \frac{\partial f}{\partial y}(x_k, \eta) \right)^{-1} L(y, x_0, h).$$

Здесь  $\eta$  в скалярном случае является некоторым промежуточным значением между  $y(x_k)$  и  $y_k$ . Если  $f$  — векторная функция, то  $\frac{\partial f}{\partial y}(x_k, \eta)$  — матрица Якоби, строки которой вычислены, возможно, при разных значениях, принадлежащих отрезку между  $y(x_k)$  и  $y_k$ .

**Доказательство.** По определению 2.1,  $y_k$  задается в неявном виде уравнением

$$\sum_{i=0}^{k-1} [\alpha_i y(x_i) - h\beta_i f(x_i, y(x_i))] + \alpha_k y_k - h\beta_k f(x_k, y_k) = 0.$$

Подставляя в него (2.3), получим

$$L(y, x_0, h) = \alpha_k [y(x_k) - y_k] - h \beta_k [f(x_k, y(x_k)) - f(x_k, y_k)],$$

откуда по теореме о среднем следует утверждение леммы. ■

Эта лемма показывает равенство локальной погрешности и главного члена выражения  $\alpha_k^{-1} L(y, x_0, h)$ , которое иногда также называют локальной погрешностью (Далквист 1956, 1959). Для явных методов оба определения совпадают.

### Порядок многошагового метода

Определив локальную погрешность многошагового метода, для него таким же образом, как для одношаговых методов, можно ввести понятие порядка.

**Определение 2.3.** Говорят, что многошаговый метод (2.1) имеет порядок  $p$ , если выполняется одно из следующих двух условий:

i)  $L(y, x, h) = O(h^{p+1})$  для всех достаточно гладких функций  $y(x)$ ;

ii) локальная погрешность формулы (2.1) равна  $O(h^{p+1})$  для всех достаточно гладких дифференциальных уравнений (1.1).

Отметим, что по лемме 2.2 условия i) и ii) эквивалентны. Следующая наша задача — выразить порядок многошагового метода через его свободные параметры  $\alpha_i$  и  $\beta_i$ . Далквист (1956) первым указал на фундаментальную роль многочленов

$$\begin{aligned}\rho(\zeta) &= \alpha_k \zeta^k + \alpha_{k-1} \zeta^{k-1} + \dots + \alpha_0, \\ \sigma(\zeta) &= \beta_k \zeta^k + \beta_{k-1} \zeta^{k-1} + \dots + \beta_0.\end{aligned}\tag{2.4}$$

Назовем их производящими многочленами многошагового метода (2.1).

**Теорема 2.4.** Многошаговый метод (2.1) имеет порядок  $p$  тогда и только тогда, когда выполняется одно из следующих эквивалентных условий:

i)  $\sum_{i=1}^k \alpha_i = 0$  и  $\sum_{i=0}^k \alpha_i i^q = q \sum_{i=0}^k \beta_i i^{q-1}$  при  $q = 1, \dots, p$ ;

ii)  $\rho(e^h) - h\sigma(e^h) = O(h^{p+1})$  при  $h \rightarrow 0$ ;

iii)  $\frac{\rho(\zeta)}{\log \zeta} - \sigma(\zeta) = O((\zeta - 1)^p)$  при  $\zeta \rightarrow 1$ .

**Доказательство.** Подставляя в (2.3) разложение в ряд Тейлора (при необходимости усеченное) для  $y(x + ih)$  и  $y'(x + ih)$ , по-

лучим

$$\begin{aligned} L(y, x, h) &= \sum_{i=0}^k \left[ \alpha_i \sum_{q>0} \frac{i^q}{q!} h^q y^{(q)}(x) - h \beta_i \sum_{r>0} \frac{i^r}{r!} h^r y^{(r+1)}(x) \right] = \\ &= y(x) \sum_{i=0}^k \alpha_i + \sum_{q>1} \frac{h^q}{q!} y^{(q)}(x) \left[ \sum_{i=0}^k \alpha_i i^q - q \sum_{i=0}^k \beta_i i^{q-1} \right], \end{aligned} \quad (2.5)$$

откуда следует эквивалентность условия i) равенству  $L(y, x, h) = O(h^{p+1})$  для всех достаточно гладких функций  $y(x)$ .

Остается доказать эквивалентность всех трех условий теоремы 2.4. Эквивалентность условий i) и ii) следует из тождества

$$L(\exp, 0, h) = \rho(e^h) - h\sigma(e^h),$$

в котором через  $\exp$  обозначена экспоненциальная функция, и полученного из (2.5) равенства

$$L(\exp, 0, h) = \sum_{i=0}^k \alpha_i + \sum_{q>1} \frac{h^q}{q!} \left[ \sum_{i=0}^k \alpha_i i^q - q \sum_{i=0}^k \beta_i i^{q-1} \right].$$

С помощью преобразования  $\zeta = e^h$  (или  $h = \log \zeta$ ) условие ii) можно записать в виде

$$\rho(\zeta) - \log \zeta \sigma(\zeta) = O((\log \zeta)^{p+1}) \text{ при } \zeta \rightarrow 1,$$

что эквивалентно условию iii), поскольку

$$\log \zeta = (\zeta - 1) + O((\zeta - 1)^2) \text{ при } \zeta \rightarrow 1. \blacksquare$$

*Замечание.* Для многошаговых методов условия первого порядка, называемые обычно условиями *согласованности*, можно также записать в виде

$$\rho(1) = 0, \quad \rho'(1) = \sigma(1). \quad (2.6)$$

Если читателю понятно доказательство приведенных выше условий порядка, то и более общий случай неравномерных сеток не представит для него трудностей [см. разд. III.5 и книгу Штеттера (1973)].

*Пример 2.5. Порядок явных методов Адамса.* Сначала определим, для каких дифференциальных уравнений явные методы Адамса теоретически дают точное решение. Это происходит в том случае, когда многочлен  $p(t)$  в (1.4) равен  $f(t, y(t))$ . Предположим теперь, что  $f(t, y) = f(t)$  не зависит от  $y$  и является многочленом, степень которого меньше  $k$ . Тогда с помощью явных ме-

тодов Адамса получается точное решение дифференциальных уравнений

$$y' = qx^{q-1} \text{ при } q = 0, 1, \dots, k.$$

Это означает, что локальная погрешность равна нулю, а, следовательно, по лемме 2.2

$$0 = L(x^q, 0, h) = h^q \left[ \sum_{i=0}^k \alpha_i i^q - q \sum_{i=0}^k \beta_i i^{q-1} \right] \text{ при } q = 0, \dots, k.$$

Эти равенства совпадают с условием i) из теоремы 2.4 при  $p = k$ , а поэтому явные методы Адамса имеют порядок не ниже  $k$ . Далее показано, что он и не превосходит  $k$  (пример 2.7).

**Пример 2.6.** В неявных методах Адамса вместо  $p(i)$  используется многочлен  $p^*(i)$ , заданный в (1.7). Поскольку степень  $p^*(i)$  на единицу выше, чем у  $p(i)$ , такие же, как в примере 2.5, рассуждения показывают, что неявные методы Адамса имеют порядок не ниже  $k + 1$ .

Аналогично можно рассмотреть все остальные методы из разд. III.1 (см. упр. 2 и табл. 2.1).

### Константа погрешности многошаговых методов

Порядок многошагового метода показывает, насколько быстро погрешность стремится к нулю при  $h \rightarrow 0$ . Однако методы одного порядка могут иметь разные погрешности; они различаются *константой погрешности*. Из формулы (2.5) видно, что разностный оператор  $L$ , соответствующий многошаговому методу порядка  $p$ , для всех достаточно гладких функций  $y(x)$  удовлетворяет равенству

$$L(y, x, h) = C_{p+1} h^{p+1} y^{(p+1)}(x) + O(h^{p+2}), \quad (2.7)$$

где константа  $C_{p+1}$  вычисляется по формуле

$$C_{p+1} = \frac{1}{(p+1)!} \left[ \sum_{i=0}^k \alpha_i i^{p+1} - (p+1) \sum_{i=0}^k \beta_i i^p \right]. \quad (2.8)$$

Эта константа не подходит в качестве меры точности, поскольку для нее можно получить любое числовое значение, умножив формулу (2.1) на соответствующую постоянную, что не изменит численного решения  $\{y_n\}$ . Для этого больше подошла бы константа  $\alpha_k^{-1} C_{p+1}$ , так как локальная погрешность многошагового метода удовлетворяет соотношению (см. лемму 2.2 и формулу (2.7))

$$y(x_h) - y_h = \alpha_k^{-1} C_{p+1} h^{p+1} y^{(p+1)}(x_0) + O(h^{p+2}). \quad (2.9)$$

Но такое определение тоже не может нас удовлетворить по некоторым причинам, которые станут понятны из следующего объяснения. Обозначим через  $e_n$  глобальную погрешность, масштабированную на  $h^p$ :

$$e_n = \frac{y_n - y(x_n)}{h^p}.$$

Вычитая (2.1) из (2.3) и учитывая (2.7), получим

$$\sum_{i=0}^k \alpha_i e_{n+i} = h^{1-p} \sum_{i=0}^k \beta_i (f(x_{n+i}, y_{n+i}) - f(x_{n+i}, y(x_{n+i}))) - C_{p+1} h^{(p+1)}(x_n) + O(h^2). \quad (2.10)$$

Теперь, чтобы в (2.10) относящийся к погрешности член внести под знак суммы с коэффициентами  $\beta_i$ , используем равенство

$$y^{(p+1)}(x_n) = \frac{1}{\sigma(1)} \sum_{i=0}^k \beta_i y^{(p+1)}(x_{n+i}) + O(h). \quad (2.11)$$

Линеаризуем выражение

$$f(x_{n+i}, y_{n+i}) - f(x_{n+i}, y(x_{n+i})) = \frac{\partial f}{\partial y}(x_{n+i}, y(x_{n+i})) h^p e_{n+i} + O(h^{2p})$$

и подставим его, а также (2.11), в (2.10). Пренебрегая членами  $O(h^2)$  и  $O(h^{2p})$ , мы можем рассматривать полученную формулу как многошаговый метод, примененный к задаче

$$e'(x) = \frac{\partial f}{\partial y}(x, y(x)) e(x) - Cy^{(p+1)}(x), \quad e(x_0) = 0, \quad (2.12)$$

где

$$C = \frac{C_{p+1}}{\sigma(1)} \quad (2.13)$$

является естественной мерой глобальной погрешности и поэтому называется *константой погрешности*.

Другой вывод определения (2.13) будет указан в разделе, посвящённом глобальной сходимости (см. упр. 2 из разд. III.4). Кроме того, решение задачи (2.12) дает первый член асимптотического разложения глобальной погрешности (см. разд. III.9).

**Пример 2.7. Константа погрешности явных методов Адамса.** Рассмотрим дифференциальное уравнение  $y' = f(x)$  с функцией  $f(x) = (k+1)x^k$  и точным решением  $y(x) = x^{k+1}$ . Поскольку это уравнение интегрируется точно явным методом Адамса (см. пример 2.5), имеет место равенство

$$y(x_k) - y(x_{k-1}) = h \sum_{j=0}^k \gamma_j \nabla^j f_{k-1}.$$

Поэтому локальную погрешность  $k$ -шагового явного метода Адамса (1.5) можно записать в виде

$$y(x_k) - y_k = h\gamma_k \nabla^k f_{k-1} = h^{k+1} \gamma_k f^{(k)}(x_0) = h^{k+1} \gamma_k y^{(k+1)}(x_0).$$

В силу неравенства  $\gamma_k \neq 0$  из этой формулы следует, что порядок  $k$ -шагового метода не превосходит  $k$  (сравните с примером 2.5). Кроме того, учитывая, что  $\alpha_k = 1$ , при сопоставлении последнего соотношения с формулой (2.9) придем к равенству  $C_{k+1} = \gamma_k$ . И наконец, для методов Адамса  $\rho(\zeta) = \zeta^k - \zeta^{k-1}$  и  $\rho'(1) = 1$ , а поэтому с помощью (2.6) получим значение константы погрешности  $C = \gamma_k$ .

В табл. 2.1 приведены константы погрешности всех рассмотренных выше многошаговых методов (отметим, что для явного метода Ньюстрёма и метода Милна—Симпсона  $\sigma(1) = 2$ ).

Таблица 2.1. Порядок и константа погрешности многошаговых методов

Метод	Номер формулы	Порядок	Константа погрешности
явный Адамса	(1.5)	$k$	$\gamma_k$
неявный Адамса	(1.8)	$k + 1$	$\gamma_{k+1}^*$
правило средней точки	(1.13')	2	$1/6$
Ньюстрёма, $k > 2$	(1.13)	$k$	$x_k/2$
Милна, $k = 2$	(1.15')	4	$-1/180$
Милна—Симпсона, $k > 3$	(1.15)	$k + 1$	$\gamma_{k+1}^*/2$
ФДН	(1.22')	$k$	$-1/(k + 1)$

### Неприводимые методы

Пусть производящие многочлены  $\rho(\zeta)$  и  $\sigma(\zeta)$  метода (2.1) заданы формулами (2.4), и предположим, что они имеют общий множитель  $\varphi(\zeta)$ . Тогда определим многочлены

$$\rho^*(\zeta) = \frac{\rho(\zeta)}{\varphi(\zeta)}, \quad \sigma^*(\zeta) = \frac{\sigma(\zeta)}{\varphi(\zeta)},$$

которые являются производящими многочленами нового, более простого многошагового метода. С помощью оператора сдвига  $E$ , заданного соотношением

$$Ey_n = y_{n+1} \text{ или } E y(x) = y(x + h),$$

этот многошаговый метод можно записать в компактной форме

$$\rho^*(E) y_n = h \sigma^*(E) f_n.$$

Умножив обе части формулы на  $\varphi(E)$ , мы увидим, что любое ее решение  $\{y_n\}$  будет также и решением исходного метода

$\rho(E) y_n = h\sigma(E) f_n$ . Таким образом, оба метода в сущности совпадают. Обозначим через  $L^*$  разностный оператор, сопоставленный новому приведенному методу, а через  $C_{p+1}^*$  — его константу из (2.7). Из равенства

$$\begin{aligned} L(y, x, h) &= \varphi(E) L^*(y, x, h) = C_{p+1}^* h^{p+1} \varphi(E) y^{(p+1)}(x) + O(h^{p+2}) \\ &= C_{p+1}^* \varphi(1) h^{p+1} y^{(p+1)}(x) + O(h^{p+2}) \end{aligned}$$

сразу получим  $C_{p+1} = \varphi(1) C_{p+1}^*$ , и поэтому также выполняется соотношение  $C_{p+1}/\sigma(1) = C_{p+1}^*/\sigma^*(1)$ . Итак, оба метода имеют одинаковые константы погрешности.

Проведенный анализ показывает, что многошаговые методы, у которых производящие многочлены имеют общий множитель, не представляют интереса. Поэтому обычно мы будем предполагать, что

$$\rho(\zeta) \text{ и } \sigma(\zeta) \text{ не имеют общих множителей.} \quad (2.13)$$

Многошаговые методы, удовлетворяющие этому свойству, называются *неприводимыми*.

### Ядро Пеано многошаговых методов

Порядок и константа погрешности не дают полного описания погрешности, так как в ее разложении последующие члены ряда могут значительно превосходить  $C_{p+1}$ . Поэтому было предпринято несколько попыток — первоначально для квадратурных формул — получить полное описание погрешности. Следующие результаты опираются на идеи Пеано (1913):

**Теорема 2.8.** *Пусть многошаговый метод (2.1) имеет порядок  $p$ , а  $q$  — некоторое целое число ( $1 \leq q \leq p$ ). Тогда для любой функции  $y(x)$ , непрерывно дифференцируемой  $q+1$  раз, выполняется равенство*

$$L(y, x, h) = h^{q+1} \int_0^k K_q(s) y^{(q+1)}(x + sh) ds, \quad (2.14)$$

где

$$K_q(s) = \frac{1}{q!} \sum_{i=0}^k \alpha_i (i-s)_+^q - \frac{1}{(q-1)!} \sum_{i=0}^k \beta_i (i-s)_+^{q-1}$$

называется  $q$ -м ядром Пеано многошагового метода (2.1).

**Замечание.** В определении ядра Пеано использовано обозначение

$$(i-s)_+^r = \begin{cases} (i-s)^r & \text{при } i-s > 0, \\ 0 & \text{при } i-s \leq 0. \end{cases}$$

Поэтому  $K_q(s)$  является кусочно-полиномиальной функцией и удовлетворяет равенству

$$K_q(s) = \frac{1}{q!} \sum_{i=1}^k \alpha_i (i-s)^q - \frac{1}{(q-1)!} \sum_{i=1}^k \beta_i (i-s)^{q-1} \text{ при } s \in [j-1, j]. \quad (2.15)$$

*Доказательство.* По теореме Тейлора с интегральным представлением остаточного члена

$$\begin{aligned} y(x+ih) &= \sum_{r=0}^q \frac{i^r}{r!} h^r y^{(r)}(x) + h^{q+1} \int_0^i \frac{(i-s)^q}{q!} y^{(q+1)}(x+sh) ds, \\ hy'(x+ih) &= \sum_{r=0}^q \frac{i^{r-1}}{(r-1)!} h^r y^{(r)}(x) + h^{q+1} \int_0^i \frac{(i-s)^{q-1}}{(q-1)!} y^{(q+1)}(x+sh) ds. \end{aligned}$$

Подставив эти два выражения в (2.3), с помощью таких же рассуждений, как в доказательстве теоремы 2.4, получим, что при  $q \leq p$  стоящие перед интегралом многочлены сокращаются. Тогда из равенства

$$\int_0^i \frac{(i-s)^q}{q!} y^{(q+1)}(x+sh) ds = \int_0^k \frac{(i-s)_+^q}{q!} y^{(q+1)}(x+sh) ds$$

следует утверждение теоремы. ■

Кроме представления (2.15), ядро Пеано  $K_q(s)$  обладает следующими свойствами:

$$\begin{aligned} K_q(s) &= 0 \quad \text{при } s \in (-\infty, 0) \cup [k, \infty) \\ \text{и } q &= 1, \dots, p; \end{aligned} \quad (2.16)$$

$K_q(s)$  — непрерывно дифференцируемая  $q-2$  раза функция и  $K'_q(s) = -K_{q-1}(s)$  при  $q = 2, \dots, p$  (при  $q = 2$  в обобщенном смысле);

$K_1(s)$  — кусочно-линейная функция с разрывами в точках  $0, 1, \dots, k$ . В точке  $j$  она имеет скачок  $\beta_j$ , а на интервале  $(j-1, j)$  ее угловой коэффициент равен  $-(\alpha_j + \alpha_{j+1} + \dots + \alpha_k)$ ;

константа  $C_{p+1}$  из (2.8) вычисляется по формуле

$$C_{p+1} = \int_0^k K_p(s) ds. \quad (2.19)$$

Приведем доказательство утверждений (2.16) и (2.19). Равенство  $K_q(s) = 0$  при  $s \geq k$  и  $q \leq p$  вытекает прямо из определения

ния ядра Пеано. Чтобы доказать это равенство и для  $s < 0$ , рассмотрим многочлен  $y(x) = (x - s)^q$ , где  $s$  является параметром. Тогда из теоремы 2.8 следует, что

$$L(y, 0, 1) = \sum_{i=0}^k \alpha_i (i - s)^q - q \sum_{i=0}^k \beta_i (i - s)^{q-1} = 0 \quad \text{при } q \leq p,$$

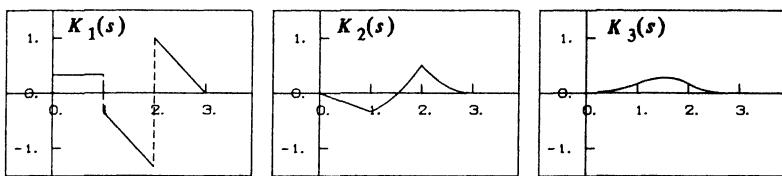


Рис. 2.2. Ядра Пеано трехшагового метода Ньюстрёма.

а поэтому  $K_q(s) = 0$  при  $s < 0$ , что и завершает доказательство свойства (2.16). Соотношение (2.17) устанавливается с помощью интегрирования по частям выражения (2.14). В качестве примера на рис. 2.2 изображены ядра Пеано для 3-шагового метода Ньюстрёма (1.13').

### Упражнения

#### 1. Построение многошаговых методов.

а) Для любого многочлена  $\rho(\zeta)$  степени  $k$ , удовлетворяющего условию  $\rho(1) = 0$ , существует единственный многочлен  $\sigma(\zeta)$ , степень которого не превосходит  $k$ , такой что соответствующий многошаговый метод имеет порядок не ниже  $k + 1$ ;

б) При условии а) существует единственный многочлен  $\sigma(\zeta)$  степени  $< k$ , такой что соответствующий многошаговый метод (который в этом случае будет явным) имеет порядок не ниже  $k$ .

*Указание.* Используйте условие iii) из теоремы 2.4.

#### 2. Найдите многошаговый метод вида

$$y_{n+2} + \alpha_1 y_{n+1} + \alpha_0 y_n = h (\beta_1 f_{n+1} + \beta_0 f_n),$$

имеющий максимально возможный порядок. Изучите полученную формулу на примере  $y' = y$ ,  $y(0) = 1$ ,  $h = 0.1$ .

3. Проверьте приведенные в табл. 2.1 значения порядка и константы погрешности для формул дифференцирования назад.
4. Покажите, что в явных и неявных методах Адамса, а также в формулах дифференцирования назад ядро Пеано  $K_p(s)$

не меняет знака. Выведите из этого свойства соотношение

$$L(y, x, h) = h^{p+1} C_{p+1} y^{(p+1)}(\zeta),$$

где  $\zeta \in (x, x + kh)$ , а  $C_{p+1}$  — константа из (2.8).

5. Пусть  $y(x)$  — точное решение уравнения  $y' = f(x, y)$ , а стартовые значения равны  $y_i = y(x_i)$ ,  $i = 0, 1, \dots, k - 1$ . Предположим, что  $f$  непрерывна и удовлетворяет условию Липшица по  $y$  (необязательно, чтобы  $f$  была дифференцируемой). Докажите, что для локальной погрешности согласованного многошагового метода (для которого верны равенства (2.6)) выполняется соотношение

$$\|y(x_h) - y_h\| \leq h\omega(h),$$

где  $\omega(h) \rightarrow 0$  при  $h \rightarrow 0$ .

### III.3. Устойчивость и первый барьер Далквиста

«... с тех пор автору не раз приходилось встречаться с такими методами численного интегрирования дифференциальных уравнений, которые обладали подкапающе малой погрешностью аппроксимации, но тем не менее таили в себе большую опасность вычислительной неустойчивости».

(Х. Рутисхаузер, 1952)

В своей знаменитой статье Рутисхаузер отметил, что высокого порядка и малой локальной погрешности еще недостаточно для того, чтобы многошаговый метод был пригодным для практических расчетов. Численное решение может быть «неустойчивым» даже при очень малом шаге  $h$ . На этот же факт указал Тодд (1950), применив ряд разностных методов к дифференциальным уравнениям второго порядка. При изложении мы будем опираться в основном на работу Далквиста (1956), в которой этот эффект изучен систематически. История развития понятий численной устойчивости интересно представлена в статье Далквиста (1985) «33 года численной неустойчивости, часть I».

Начнем с примера из работы Далквиста (1956) (см. упр. 2 из разд. III.2). Среди всех явных 2-шаговых методов выберем формулу максимально возможного порядка. С помощью теоремы 2.4 легко определить, что это метод 3-го порядка, имеющий вид

$$y_{n+2} + 4y_{n+1} - 5y_n = h(4f_{n+1} + 2f_n). \quad (3.1)$$

Его применение к дифференциальной задаче

$$y' = y, \quad y(0) = 1 \quad (3.2)$$

дает линейное разностное уравнение

$$y_{n+2} + 4(1-h)y_{n+1} - (5+2h)y_n = 0. \quad (3.3)$$

В качестве стартовых возьмем значения точного решения  $y_0 = 1$  и  $y_1 = \exp(h)$ . На рис. 3.1 изображены точное решение  $\exp(x)$  и численное решение для шагов  $h = 1/10$ ,  $h = 1/20$ ,  $h = 1/40$  и т. д. Несмотря на малую локальную погрешность, результаты являются неудовлетворительными. Из рис. 3.1 видно, что с уменьшением шага они все ухудшаются.

Этот факт легко объяснить. Подставим в (3.3)  $y_j = \zeta^j$ , как это обычно делается для линейных разностных уравнений [Ла-

гранж (1775)]. В результате получим характеристическое уравнение

$$\zeta^2 + 4(1-h)\zeta - (5+2h) = 0. \quad (3.4)$$

Общее решение (3.3) вычисляется по формуле

$$y_n = A\zeta_1^n(h) + B\zeta_2^n(h), \quad (3.5)$$

где

$$\zeta_1(h) = 1 + h + O(h^2), \quad \zeta_2(h) = -5 + O(h)$$

— корни уравнения (3.4), а коэффициенты  $A$  и  $B$  определяются по стартовым значениям  $y_0$  и  $y_1$ . Поскольку  $\zeta_1(h)$  аппроксимирует

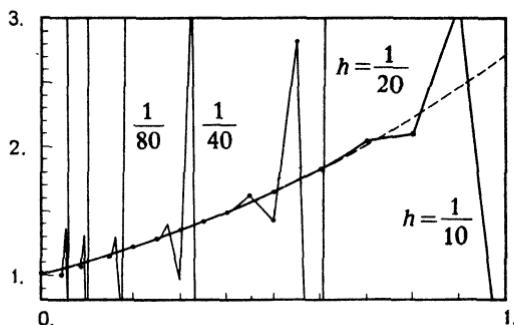


Рис. 3.1. Численное решение, полученное неустойчивым методом (3.1).

$\exp(h)$ , первый член в (3.5) аппроксимирует значение точного решения  $\exp(x)$  в точке  $x = nh$ . Возмущение в наш метод вносит второй член из (3.5), называемый часто *паразитным решением*: так как при  $h \rightarrow 0$   $\zeta_2(h)$  по абсолютному значению превосходит единицу, это паразитное решение с ростом  $n$  становится очень большим и начинает преобладать в решении  $y_n$ .

Теперь обратимся к вопросу об устойчивости метода общего вида (2.1). Важную роль играет поведение решения при  $n \rightarrow \infty$  (или  $h \rightarrow 0$ ) и фиксированном  $nh$ . Очевидно, что (2.1) при  $h \rightarrow 0$  сводится к формуле

$$\alpha_k y_{n+k} + \alpha_{k-1} y_{n+k-1} + \cdots + \alpha_0 y_n = 0. \quad (3.6)$$

Ее можно рассматривать как численный метод (2.1), примененный к решению дифференциального уравнения

$$y' = 0. \quad (3.7)$$

Подставив в формулу (3.6)  $y_j = \zeta^j$  и разделив ее на  $\zeta^n$ , получим, что  $\zeta$  должно быть корнем уравнения

$$\rho(\zeta) = \alpha_k \zeta^k + \alpha_{k-1} \zeta^{k-1} + \cdots + \alpha_0 = 0. \quad (3.8)$$

Как и в разд. I.5, здесь снова возникают трудности, когда (3.8) имеет корень *кратности*  $m > 1$ . В этом случае [Лагранж (1792), см. далее упр. 1] решениями (3.6) являются величины  $y_n = n^{j-1} \zeta^n$  ( $j = 1, \dots, m$ ), и суперпозиция приводит к следующему результату.

**Лемма 3.1.** Пусть многочлен  $\rho(\zeta)$  имеет корни  $\zeta_1, \dots, \zeta_l$  кратностей  $m_1, \dots, m_l$  соответственно. Тогда общее решение (3.6) задается формулой

$$y_n = p_1(n) \zeta_1^n + \cdots + p_l(n) \zeta_l^n, \quad (3.9)$$

где  $p_j(n)$  — многочлены степеней  $m_j - 1$ . ■

Из формулы (3.9) видно, что для ограниченности  $y_n$  при  $n \rightarrow \infty$  требуется, чтобы корни (3.8) лежали в круге единичного радиуса, а корни, принадлежащие единичной окружности, были простыми.

**Определение 3.2.** Многошаговый метод (2.1) называется *устойчивым*, если его производящий многочлен  $\rho(\zeta)$  [формула (3.8)] удовлетворяет *корневому условию*, т. е.

- i) корни  $\rho(\zeta)$  лежат на (или внутри) единичной окружности;
- ii) корни, принадлежащие единичной окружности, являются простыми.

**Замечание.** Чтобы отличать это понятие устойчивости от остальных, его иногда называют *нуль-устойчивостью* или *D-устойчивостью* в честь Далквиста.

**Примеры.** Для явного и неявного методов Адамса  $\rho(\zeta) = \zeta^k - \zeta^{k-1}$ . Кроме простого корня, равного 1, этот многочлен имеет нулевой корень кратности  $k - 1$ . Поэтому методы Адамса устойчивы.

Это утверждение верно также для явного метода Ньюстрема и методов Милна—Симпсона, для которых  $\rho(\zeta) = \zeta^k - \zeta^{k-2}$ . Заметим, что в этом случае имеется еще простой корень — 1, который может представлять опасность при решении некоторых дифференциальных уравнений (см. т. II).

### Устойчивость формул дифференцирования назад

Для ФДН-формул исследовать устойчивость сложнее. Поскольку характеристический многочлен выражения  $\nabla^l y_{k+n} = 0$  равен  $\zeta^{k-l} (\zeta - 1)^l = 0$ , из представления (1.22') следует, что эти

методы имеют производящий многочлен  $\rho(\zeta)$  вида

$$\rho(\zeta) = \sum_{j=1}^k \frac{1}{j} \zeta^{k-j} (\zeta - 1)^j. \quad (3.10)$$

При изучении его нулей удобнее рассматривать многочлен

$$p(z) = (1-z)^k \rho\left(\frac{1}{1-z}\right) = \sum_{j=1}^k \frac{z^j}{j}, \quad (3.11)$$

полученный с помощью преобразования  $\zeta = 1/(1-z)$ . Он равен  $k$ -й частичной сумме ряда Тейлора для функции  $-\log(1-z)$ .

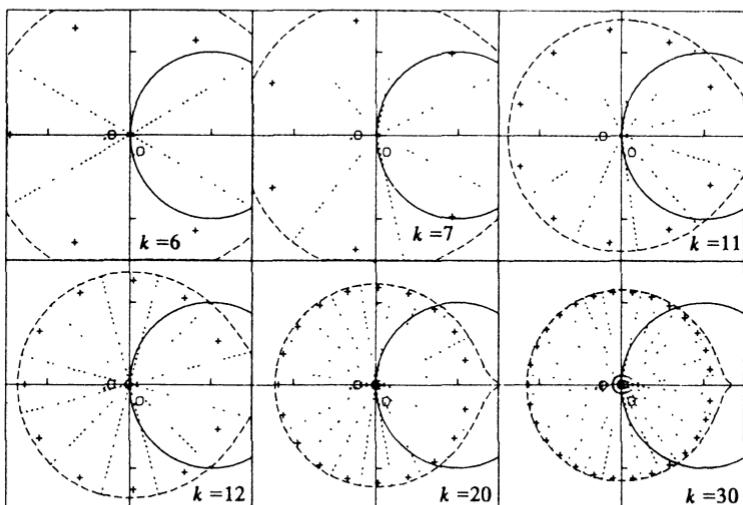


Рис. 3.2. Корни многочлена  $p(z)$ , определенного формулой (3.11).

Поскольку корни  $p(z)$  и  $\rho(\zeta)$  связаны указанным преобразованием, выполняется следующая лемма.

**Лемма 3.3.**  *$k$ -шаговая ФДН-формула (1.22') устойчива тогда и только тогда, когда все корни многочлена (3.11) лежат вне круга  $\{z; |z-1| \leq 1\}$ , а корни на его границе являются простыми.* ■

Корни многочлена (3.11) для различных значений  $k$  изображены на рис. 3.2.

**Теорема 3.4.**  *$k$ -шаговая ФДН-формула (1.22') устойчива при  $k \leq 6$  и неустойчива при  $k \geq 7$ .*

**Доказательство.** Первое утверждение легко проверить простым перебором конечного числа случаев [см. рис. 3.2]. Впервые это отметили Митчелл и Крэггс (1953)]. Однако вторую часть

теоремы доказать сложнее, так как она относится к бесконечному числу случаев. Ее первое полное доказательство содержится в техническом отчете Крайера (1971), а сокращенный вариант опубликован в его же статье (1972). Другое доказательство, основанное на критерии Шура—Кона, дано в работе Кридона и Миллера (1975) [см. также книгу Григорьева (1977)]. Схема этого доказательства приведена ниже в упр. 4. Мы изложим доказательство из статьи Хайрера и Ваннера (1983), где используется представление

$$p(z) = \int_0^{\sigma} \sum_{l=1}^k \zeta^{l-1} d\zeta = \int_0^{\sigma} \frac{1 - \zeta^k}{1 - \zeta} d\zeta = \int_0^{\sigma} (1 - e^{ik\theta} s^k) \varphi(s) ds, \quad (3.12)$$

в котором

$$\zeta = se^{i\theta}, \quad z = re^{i\theta}, \quad \varphi(s) = \frac{e^{i\theta}}{1 - se^{i\theta}}.$$

Разобъем комплексную плоскость на  $k$  секторов:

$$S_j = \left\{ z; \frac{2\pi}{k} \left( j - \frac{1}{2} \right) < \arg(z) < \frac{2\pi}{k} \left( j + \frac{1}{2} \right) \right\},$$

$$j = 0, 1, \dots, k-1.$$

На лучах, ограничивающих  $S_j$ , выполняется равенство  $e^{ik\theta} = -1$ , а поэтому (3.12) сводится к интегралу

$$p(z) = \int_0^{\sigma} (1 + s^k) \varphi(s) ds$$

с положительной весовой функцией. Следовательно, аргумент  $p(z)$  всегда принадлежит углу, который покрывает функция  $\varphi(s)$ , т. е. углу между  $e^{i\theta}$  и  $e^{i\pi} = -1$ . Таким образом, на этих лучах  $\arg(p(z))$  не может совершить полный оборот, а поскольку  $\arg(z^k)$  делает один полный оборот на бесконечности между  $\theta = 2\pi(j - 1/2)/k$  и  $\theta = 2\pi(j + 1/2)/k$ , то в силу принципа аргумента [см., например, книгу

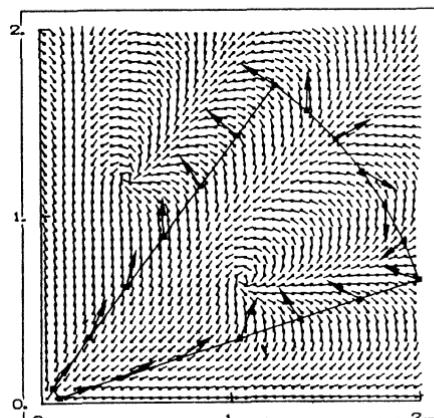


Рис. 3.3. Аргумент многочлена  $p(z)$ , определенного формулой (3.11).

Хенричи (1974) [каждому сектору  $S_j$  ( $j = 1, \dots, k - 1$ , случай  $j = 0$  исключается) принадлежит ровно один корень  $p(z)$  (см. рис. 3.3)].

Для завершения доказательства нам остается только ограничить область, где расположены нули  $p(z)$ . Сделаем это следующим образом. Заметим, что в (3.12) член  $s^k$  становится большим при  $s > 1$ . Поэтому разобьем (3.12) на два интеграла:  $p(z) = I_1 - I_2$ , где

$$I_1 = \int_0^r \varphi(s) ds - \int_0^1 e^{ik\theta} s^k \varphi(s) ds, \quad I_2 = e^{ik\theta} \int_1^r s^k \varphi(s) ds.$$

Из неравенства  $|\varphi(s)| \leq B(\theta)$ , в котором

$$B(\theta) = \begin{cases} \frac{1}{\sin \theta} & \text{при } 0 < \theta \leq \frac{\pi}{2}, \\ 1 & \text{при } \frac{\pi}{2} \leq \theta \leq \pi, \end{cases}$$

получим оценку

$$|I_1| \leq \left(r + \frac{1}{k+1}\right) B(\theta) < r B(\theta) \frac{k+2}{k+1}, \quad (r > 1). \quad (3.13)$$

Затем из положительности  $s^k$  выведем соотношение

$$I_2 = e^{ik\theta} \Phi \int_1^r s^k ds,$$

где  $\Phi$  принадлежит выпуклой оболочке множества  $\{\varphi(s); 1 \leq s \leq r\}$ . Любой ее элемент можно представить в виде

$$\Phi = \alpha \varphi(s_1) + (1 - \alpha) \varphi(s_2) = \frac{\varphi(s_1) \varphi(s_2)}{\varphi(s)},$$

где  $\hat{s} = \alpha s_2 + (1 - \alpha) s_1$ ,  $0 \leq \alpha \leq 1$ ,  $1 \leq s_1, s_2 \leq r$ . Поскольку  $|\varphi(s)|$  монотонно убывает при  $s \geq 1$ , выполняется неравенство  $|\Phi| \geq |\varphi(r)|$ . Элементарные геометрические соображения приводят к соотношению  $|\Phi| \geq 1/2r$ , и в результате получим оценку

$$|I_2| \geq \frac{r^{k+1} - 1}{2r(k+1)} > \frac{r(r^{k-1} - 1)}{2k+2}, \quad (r > 1). \quad (3.14)$$

Из (3.13) и (3.14) следует, что при

$$r \geq R(\theta) = ((2k+4)B(\theta) + 1)^{1/(k-1)} \quad (3.15)$$

верно соотношение  $|I_2| > |I_1|$ , а поэтому  $p(z)$  не может обращаться в нуль. Кривая  $R(\theta)$ , также нанесенная на рис. 3.2, отделяет от секторов  $S_j$  область, которую мы назовем «пирогом с сыром мадам Имхоф». Каждый его кусок (при  $j \neq 0$ ) должен содержать корень многочлена  $p(z)$ . Как показывает простой анализ, принадлежащая  $S_1$  часть «пирога» при  $k = 12$  настолько мала, что обеспечивает наличие нулей  $p(z)$  внутри круга  $\{z;$

$|z - 1| \leq 1$ . Поскольку  $R(\theta)$  при фиксированном  $\theta$  монотонно убывает с ростом  $k$ , как и  $R(\pi/k)$ , то же утверждение верно для всех  $k \geq 12$ . При  $6 < k < 12$  на неустойчивость методов указывают и численные расчеты (см. рис. 3.2 или упр. 4). ■

### Наивысший достижимый порядок устойчивых многошаговых методов

Вполне естественно возникает вопрос об устойчивости многошаговых методов максимально возможного порядка. Далквист (1956) провел его исследование, результатом которого стал знаменитый «первый барьер Далквиста».

Как следует из подсчета условий порядка (теорема 2.4), для того чтобы линейный многошаговый метод имел порядок  $p$ , его параметры должны удовлетворять  $p + 1$  линейным уравнением. Поскольку метод содержит  $2k + 1$  свободных параметров (без ограничения общности можно считать  $a_k = 1$ ), его наивысший достижимый порядок равен  $2k$ . Это утверждение можно действительно проверить (см. упр. 5). Однако, как показывает следующая теорема, такие методы не имеют практического значения.

**Теорема 3.5.** (Первый барьер Далквиста) Порядок  $p$  устойчивого линейного  $k$ -шагового метода подчиняется следующим ограничениям

$$p \leq k + 2 \text{ при четных } k,$$

$$p \leq k + 1 \text{ при нечетных } k,$$

$$p \leq k \text{ при } \beta_k/a_k \leq 0 \text{ (в частности, для явных методов).}$$

Отложим доказательство теоремы, чтобы дать ряд необходимых для него обозначений и лемм. Сначала введем «греко-римское преобразование»

$$\xi = \frac{z + 1}{z - 1} \quad \text{или} \quad z = \frac{\xi + 1}{\xi - 1}. \quad (3.16)$$

Оно отображает круг  $|\xi| < 1$  в полуплоскость  $\operatorname{Re} z < 0$ , верхнюю полуплоскость  $\operatorname{Im} z > 0$  — в нижнюю, окружность  $|\xi| = 1$  — в мнимую ось, точку  $\xi = 1$  в  $z = \infty$ , а  $\xi = -1$  в  $z = 0$ . Теперь рассмотрим многочлены

$$R(z) = \left( \frac{z - 1}{2} \right)^k \rho(\xi) = \sum_{l=0}^k a_l z^l, \quad (3.17)$$

$$S(z) = \left( \frac{z - 1}{2} \right)^k \sigma(\xi) = \sum_{l=0}^k b_l z^l.$$

Поскольку нули  $R(z)$  и  $\rho(\zeta)$  связаны преобразованием (3.16), условие устойчивости многошагового метода можно выразить в терминах  $R(z)$  следующим образом: все нули  $R(z)$  лежат в левой полуплоскости  $\operatorname{Re} z \leq 0$ , и на мнимой оси нет кратных нулей.

**Лемма 3.6.** *Пусть многошаговый метод является устойчивым и имеет по крайней мере нулевой порядок. Тогда*

$$\text{i) } a_k = 0 \text{ и } a_{k-1} = 2^{1-k}\rho'(1) \neq 0;$$

*ii) все ненулевые коэффициенты  $R(z)$  имеют один и тот же знак.*

*Доказательство.* Разделив формулу (3.17) на  $z^k$  и подставив в нее  $z = \infty$ , найдем значение коэффициента  $a_k = 2^{-k}\rho(1)$ . Это выражение должно быть равным нулю вследствие условия нулевого порядка. Тем же способом получим коэффициент  $a_{k-1} = -2^{1-k}\rho'(1)$ , который отличен от нуля, так как в силу устойчивости корень 1 многочлена  $\rho(\zeta)$  не может быть кратным. Второе утверждение леммы доказывается с помощью разложения на множители

$$R(z) = a_{k-1} \prod (z + x_j) \prod ((z + u_j)^2 + v_j^2),$$

где  $-x_j$  — вещественные корни, а  $-u_j \pm iv_j$  — пары комплексно сопряженных корней. По условию устойчивости  $x_j \geq 0$  и  $u_j \geq 0$ , откуда следует, что все коэффициенты  $R(z)$  имеют один и тот же знак. ■

Далее выразим условия порядка из теоремы 2.4 в терминах многочленов  $R(z)$  и  $S(z)$ .

**Лемма 3.7.** *Многошаговый метод имеет порядок  $p$  тогда и только тогда, когда выполняется равенство*

$$R(z) \left( \log \frac{z+1}{z-1} \right)^{-1} - S(z) = C_{p+1} \left( \frac{2}{z} \right)^{p-k} + O \left( \left( \frac{2}{z} \right)^{p-k+1} \right)$$

при  $z \rightarrow \infty$ . (3.18)

*Доказательство.* Сначала заметим, что член  $O((\zeta - 1)^p)$  в условии iii) теоремы 2.4 по формуле (2.7) равен  $C_{p+1}(\zeta - 1)^p + O((\zeta - 1)^{p+1})$ . Применяя преобразование (3.16), получим (3.18), так как  $\zeta - 1 = 2/(z - 1) = 2/z + O((2/z)^2)$  при  $z \rightarrow \infty$ . ■

**Лемма 3.8.** *Для коэффициентов разложения в ряд Лорана*

$$\left( \log \frac{z+1}{z-1} \right)^{-1} = \frac{z}{2} - \mu_1 z^{-1} - \mu_3 z^{-3} - \mu_5 z^{-5} - \dots \quad (3.19)$$

*при всех  $j \geq 0$  выполняется неравенство  $\mu_{2j+1} > 0$ .*

*Доказательство.* Рассмотрим ту ветвь функции  $\log \zeta$ , которая является аналитической на комплексной плоскости  $\zeta$  с разрезом вдоль отрицательной части вещественной оси и удовлетворяет

равенству  $\log 1 = 0$ . Преобразование (3.16) отображает этот разрез в отрезок вещественной оси  $[-1, +1]$ . Поэтому функция  $\log((z+1)/(z-1))$  — аналитическая на комплексной плоскости с разрезом вдоль этого отрезка (см. рис. 3.4). Формула

$$\log \frac{z+1}{z-1} = \frac{2}{z} \left( 1 + \frac{z^{-2}}{3} + \frac{z^{-4}}{5} + \frac{z^{-6}}{7} + \dots \right) \quad (3.20)$$

показывает существование разложения (3.19). Чтобы доказать

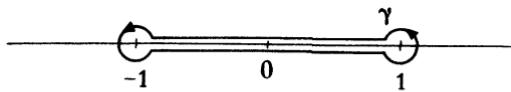


Рис. 3.4. Разрез на плоскости  $z$  с кривой  $\gamma$ .

положительность его коэффициентов, используем формулу Коши для коэффициентов функции  $f(z) = \sum_{n \in \mathbb{Z}} a_n (z - z_0)^n$ :

$$a_n = \frac{1}{2\pi i} \int_{\gamma} \frac{f(z)}{(z - z_0)^{n+1}} dz;$$

они в данном случае имеют вид

$$\mu_{2j+1} = -\frac{1}{2\pi i} \int_{\gamma} z^{2j} \left( \log \frac{z+1}{z-1} \right)^{-1} dz$$

[Коши (1831), см. также работу Бенке—Зоммера (1962)]. Здесь  $\gamma$  — произвольная кривая вокруг отрезка  $[-1, 1]$ , например, такая, как на рис. 3.4.

Заметим, что  $\log((z+1)/(z-1)) = \log((1+x)/(1-x)) = -i\pi$ , если  $z$  стремится к вещественному значению  $x \in (-1, 1)$  сверху, и  $\log((z+1)/(z-1)) = \log((1+x)/(1-x)) + i\pi$ , если  $z$  стремится к  $x$  снизу. Отсюда следует, что

$$\begin{aligned} \mu_{2j+1} &= -\frac{1}{2\pi i} \int_{-1}^1 x^{2j} \left[ \left( \log \frac{1+x}{1-x} + i\pi \right)^{-1} - \left( \log \frac{1+x}{1-x} - i\pi \right)^{-1} \right] dx \\ &= \int_{-1}^1 x^{2j} \left[ \left( \log \frac{1+x}{1-x} \right)^2 + \pi^2 \right]^{-1} dx > 0. \quad ■ \end{aligned}$$

Другое доказательство этой леммы, без использования теории функций комплексного переменного, описано в упр. 10.

*Доказательство теоремы 3.5.* Подставив в (3.18) ряд (3.19), получим

$$R(z) \left( \log \frac{z+1}{z-1} \right)^{-1} - S(z) = \\ = \text{многочлен } (z) + d_1 z^{-1} + d_2 z^{-2} + O(z^{-3}), \quad (3.21)$$

где

$$\begin{aligned} d_1 &= -\mu_1 a_0 - \mu_3 a_2 - \mu_5 a_4 - \dots \\ d_2 &= -\mu_3 a_1 - \mu_5 a_3 - \mu_7 a_5 - \dots \end{aligned} \quad (3.22)$$

Из леммы 3.6 и положительности  $\mu_j$  (лемма 3.8) следует, что в формулах для  $d_1$  и  $d_2$  все слагаемые имеют один и тот же знак. Поскольку  $a_{k-1} \neq 0$ , для четных  $k$  выполняется неравенство  $d_2 \neq 0$ , а для  $k$  нечетных  $d_1 \neq 0$ . Тогда первые два ограничения теоремы 3.5 — это прямое следствие формулы (3.18).

И наконец, при  $\beta_k/\alpha_k < 0$  докажем неравенство  $p \leq k$ : предположим противное, т. е. что порядок больше  $k$ . Тогда по формуле (3.18) многочлен  $S(z)$  равен главному члену разложения  $R(z) \times \times (\log((z+1)/(z-1)))^{-1}$ , что можно записать (подставив  $\mu_j = 0$  при четных  $j$ ) в виде соотношения

$$S(z) = R(z) \left( \frac{z}{2} - \sum_{j=1}^{k-1} \mu_j z^{-j} \right) + \sum_{j=1}^{k-1} \left( \sum_{s=j}^{k-1} \mu_s a_{s-j} \right) z^{-j}.$$

При  $z = 1$  получим

$$\frac{S(1)}{R(1)} = \left( \frac{1}{2} - \sum_{j=1}^{k-1} \mu_j \right) + \sum_{j=1}^{k-1} \left( \sum_{s=j}^{k-1} \mu_s a_{s-j} \right) \frac{1}{R(1)}. \quad (3.23)$$

В силу формулы (3.17) и равенств  $S(1) = \beta_k$ ,  $R(1) = \alpha_k$  достаточно доказать, что  $S(1)/R(1) > 0$ . Формула (3.19) при  $z \rightarrow 1$  дает соотношение

$$\sum_{j=1}^{\infty} \mu_j = \frac{1}{2},$$

а поэтому первое слагаемое в (3.23) строго положительно. Непротиворечивость второго слагаемого следует из лемм 3.6 и 3.8. ■

Устойчивые многошаговые методы, достигающие максимально возможного порядка  $k+2$ , имеют особую структуру.

**Теорема 3.9.** Устойчивые многошаговые ( $k$ -шаговые) методы порядка  $k+2$  являются симметричными, т. е.

$$\alpha_j = -\alpha_{k-j}, \quad \beta_j = \beta_{k-j} \quad \text{при всех } j. \quad (3.24)$$

**Замечание.** Для симметричных многошаговых методов из определения следует равенство  $\rho(\zeta) = -\zeta^k \rho(1/\zeta)$ . Поскольку ве-

личина  $1/\zeta_i$  — нуль многочлена  $\rho(\zeta)$ , как и  $\zeta_i$ , все корни устойчивых симметричных многошаговых методов лежат на единичной окружности и являются простыми.

*Доказательство.* При сравнении формул (3.18) и (3.21) видно, что  $d_1 = 0$  — необходимое условие порядка  $k + 2$ . При предложении об устойчивости метода из леммы 3.6 следует, что все четные коэффициенты  $R(z)$  равны нулю. Поэтому  $k$  четно, а  $R(z)$  удовлетворяет соотношению  $R(z) = -R(-z)$ . По определению  $R(z)$  оно эквивалентно равенству  $\rho(\zeta) = -\zeta^k \rho(1/\zeta)$ , которое и дает первое условие из (3.24). С помощью указанного соотношения для  $R(z)$  из формулы (3.18) получим равенство  $S(z) = -S(-z) = O((2/z)^2)$ , которое означает, что  $S(z) = S(-z)$ . Преобразовав последнее соотношение к эквивалентному виду в терминах  $\sigma(\zeta)$ , придем ко второму условию из (3.24). ■

### Упражнения

1. Рассмотрим линейное разностное уравнение (3.6) с характеристическим многочленом

$$\rho(\zeta) = \alpha_k \zeta^k + \alpha_{k-1} \zeta^{k-1} + \dots + \alpha_0.$$

Пусть  $\rho(\zeta)$  имеет различные корни  $\zeta_1, \dots, \zeta_l$ , а  $m_j \geq 1$  — кратность корня  $\zeta_j$ . Покажите, что при  $1 \leq j \leq l$ ,  $1 \leq i \leq m_j - 1$  последовательности

$$\left\{ \binom{n}{i} \zeta_j^{n-i} \right\}_{n \geq 0}$$

образуют систему  $k$  линейно независимых решений уравнения (3.6).

2. Покажите, что все корни многочлена  $p(z)$ , заданного формулой (3.11), за исключением простого нулевого корня, принадлежат кольцу

$$\frac{k}{k-1} \leq |z| \leq 2.$$

*Указание.* Используйте лемму, которую можно найти в работе Мардена (1966): если все коэффициенты многочлена  $a_k z^k + a_{k-1} z^{k-1} + \dots + a_0$  вещественные и положительные, то его корни лежат в кольце  $\rho_1 \leq |z| \leq \rho_2$ , где  $\rho_1 = \min(a_j/a_{j+1})$ , а  $\rho_2 = \max(a_j/a_{j+1})$ .

3. Примените лемму из предыдущего примера к многочлену  $\rho(\zeta)/(\zeta - 1)$  и покажите, что ФДН-формулы устойчивы при  $k = 1, 2, 3, 4$ .

4. Дайте отличное от приведенного выше доказательство теоремы 3.4, применив к многочлену

$$f(z) = z^k \rho\left(\frac{1}{z}\right) = \sum_{l=1}^k \frac{1}{l} (1-z)^l \quad (3.25)$$

критерий Шура—Кона [см., например, книгу Мардена (1966), гл. X]. Для заданного многочлена с вещественными коэффициентами

$$f(z) = a_0 + a_1 z + \dots + a_k z^k$$

рассмотрим коэффициенты  $a_i^{(j)}$ , определяемые формулами

$$a_i^{(0)} = a_i, \quad i = 0, 1, \dots, k,$$

$$a_i^{(j+1)} = a_0^{(j)} a_i^{(j)} - a_{k-i}^{(j)} a_{k-j-1}^{(j)} \quad \text{при } i = 0, \dots, k-j-1, \quad (3.26)$$

а также произведения

$$P_1 = a_0^{(1)}, \quad P_{j+1} = P_j a_0^{(j+1)} \quad \text{при } j = 1, \dots, k-1; \quad (3.27)$$

Затем обозначим через  $n$  число отрицательных, а через  $p$  — положительных элементов среди  $P_1, \dots, P_k$ . Тогда  $f(z)$  имеет не менее  $n$  нулей внутри круга единичного радиуса и не менее  $p$  нулей вне этого круга.

а) Докажите следующие формулы для коэффициентов многочлена (3.25):

$$a_0 = \sum_{l=1}^k \frac{1}{l}, \quad a_1 = -k, \quad a_2 = \frac{k(k-1)}{4},$$

$$a_{k-2} = (-1)^k \frac{k(k-1)}{2(k-2)}, \quad a_{k-1} = (-1)^{k-1} \frac{k}{k-1},$$

$$a_k = (-1)^k \frac{1}{k}. \quad (3.28)$$

б) Проверьте, что коэффициенты  $a_0^{(1)}$  из (3.26) имеют такие знаки, как в табл. 3.1. Для  $k < 13$  эти громоздкие расчеты можно вы-

Таблица 3.1. Знаки  $a_0^{(j)}$

полнить на ЭВМ, используя при этом язык символьных преобразований, чтобы избежать погрешностей округления. Для всех  $k > 2$  легко проверить неравенства  $a_0^{(1)} > 0$  и  $a_0^{(2)} > 0$ . Чтобы установить соотношение  $a_0^{(3)} = (a_0^{(2)})^2 - (a_{k-2}^{(2)})^2 < 0$  для  $k \geq 13$  рассмотрите выражение

$$a_0^{(2)} - (-1)^k a_{k-2}^{(2)} = a_0^{(1)} [a_0^2 - a_k^2 - a_0 |a_{k-2}| + a_2 |a_k|] - |a_{k-1}^{(1)}| (a_0 + |a_1|) (|a_{k-1}| + a_1), \quad (3.29)$$

которое можно записать в виде  $(a_0 + |a_k|) \varphi(k)$ , где

$$\begin{aligned} \varphi(k) &= (a_0 - |a_k|) (a_0^2 - a_k^2 - a_0 |a_{k-2}| + a_2 |a_k|) - \\ &\quad - |a_{k-1}^{(1)}| (a_1 + |a_{k-1}|) = \\ &= a_0^3 - a_0^2 \left( \frac{k}{2} + \frac{1}{2} + \frac{1}{k-2} + \frac{1}{k} \right) + \\ &\quad + a_0 \left( \frac{5k}{4} + \frac{1}{4} + \frac{1}{2k-4} - \frac{1}{k-1} - \frac{1}{(k-1)^2} - \frac{1}{k^2} \right) - \\ &\quad - \left( k - \frac{3}{4} - \frac{1}{k-1} - \frac{1}{4k} - \frac{1}{k^3} \right). \end{aligned}$$

Покажите, что  $\varphi(13) < 0$ , а с ростом  $k$  ( $k \geq 13$ )  $\varphi$  монотонно убывает (заметим, что  $a_0 = a_0(k)$  действительно зависит от  $k$  и  $a_0(k+1) = a_0(k) + 1/(k+1)$ ). И наконец, из отрицательности (3.29) выведите неравенство  $a_0^{(3)} < 0$  при  $k \geq 13$ .

с) Используйте табл. 3.1 и критерий Шура—Кона для проверки теоремы 3.4.

5. (Многошаговые методы максимального порядка.) Проверьте следующие утверждения:
  - а) не существует  $k$ -шаговых методов порядка  $2k+1$ ;
  - б) существует единственный (неявный)  $k$ -шаговый метод порядка  $2k$ ;
  - с) существует единственный явный  $k$ -шаговый метод порядка  $2k-1$ .
6. Докажите, что симметричные многошаговые методы всегда имеют четный порядок. Уточним формулировку: если симметричный многошаговый метод имеет порядок  $2s-1$ , то в действительности это также метод порядка  $2s$ .
7. Покажите, что следующие многочлены описывают все устойчивые 4-шаговые методы 6-го порядка:

$$\rho(\zeta) = (\zeta^2 - 1)(\zeta^2 + 2\mu\zeta + 1), \quad |\mu| < 1,$$

$$\begin{aligned} \sigma(\zeta) &= \frac{1}{45} (14 - \mu) (\zeta^4 + 1) + \frac{1}{45} (64 + 34\mu) \zeta (\zeta^2 + 1) + \\ &\quad + \frac{1}{15} (8 + 38\mu) \zeta^3. \end{aligned}$$

Подсчитайте константу погрешности и убедитесь, что ее нельзя сделать произвольно малой.

*Ответ.*  $C = -(16 - 5\mu)/(7560(1 + \mu))$ .

8. Докажите следующие оценки для константы погрешности:  
а) для устойчивых методов порядка  $k + 2$

$$C \leq -2^{-l-k}\mu_{k+1};$$

- б) для устойчивых методов порядка  $k + 1$  при нечетных  $k$

$$C \leq -2^{-k}\mu_k;$$

- с) для устойчивых явных методов порядка  $k$  ( $\mu_j = 0$  при четных  $j$ )

$$C \geq 2^{l-k} \left( \frac{1}{2} - \sum_{j=1}^{k-1} \mu_j \right).$$

Покажите, что все эти оценки оптимальны.

*Указание.* Сравните формулы (3.18) и (3.21) и используйте соотношение  $\sigma(1) = 2^{k-l}a_{k-1}$  из леммы 3.6.

9. Коэффициенты  $\mu_j$  из формулы (3.19) удовлетворяют рекуррентному соотношению

$$\mu_{2j+1} + \frac{1}{3}\mu_{2j-1} + \dots + \frac{1}{2j+1}\mu_1 = \frac{1}{4j+6}. \quad (3.30)$$

Первые из этих коэффициентов имеют следующие числовые значения:

$$\mu_1 = \frac{1}{6}, \quad \mu_3 = \frac{2}{45}, \quad \mu_5 = \frac{22}{945}, \quad \mu_7 = \frac{214}{14175}.$$

10. Другой способ доказательства леммы 3.8. Умножив равенство (3.30) на  $(2j+3)$  и вычитая из него такую же формулу, но с заменой  $j$  на  $j-1$ , получим

$$(2j+3)\mu_{2j+1} + \sum_{i=0}^{j-1} \mu_{2i+1} \left( \frac{2j+3}{2j-2i+1} - \frac{2j+1}{2j-2i-1} \right) = 0.$$

Покажите, что выражение в скобках является отрицательным, и по индукции докажите утверждение леммы 3.8.

### III.4. Сходимость многошаговых методов

«...Метод Адамса значительно превосходит любой другой метод. И если применение его не стало в достаточной мере всеобщим (особенно в Германии, где более популярны методы, которые развили Рунге, Хайн и Кутта), то причина этого, возможно, в том, что до сих пор отсутствовало дальнее исследование точности интегрирования по Адамсу. Этот пробел мы хотим здесь восполнить...»

(Р. фон Мизес, 1930)

Сходимость методов Адамса исследована фон Мизесом (1930) в его основополагающей статье, за которой последовал целый поток публикаций, где улучшались оценки погрешности или изложенные в этой статье идеи применялись к другим отдельным многошаговым методам, как, например, в работах Толлмина (1938), Фрике (1949), Вайсингера (1950), Виториса (1953). Однако в общем случае сходимость методов (2.1) первым доказал Далквист (1956), сформулировав ее необходимые и достаточные условия. Подлинное изящество внесли в доказательство идеи Бутчера (1966), основанные на представлении многошаговых методов в форме одношаговых, но в пространстве большей размерности. Кроме того, такое представление можно легко распространить на более общий класс методов численного интегрирования (см. разд. III.8).

Мы не можем надеяться на разумные результаты по сходимости численных методов, если задача

$$y' = f(x, y), \quad y(x_0) = y_0 \quad (4.1)$$

не будет иметь единственное решение. Поэтому сделаем следующие предположения, естественные в нашем случае, как это видно из разд. 1.7 и 1.9:

функция  $f$  непрерывна на множестве  
 $D = \{(x, y); x \in [x_0, \hat{x}], \|y(x) - y\| \leq b\}$ ,  
где через  $y(x)$  обозначено точное решение задачи (4.1),  
а  $b$  — некоторое положительное число. (4.2a)

Кроме того,  $f$  удовлетворяет условию Липшица  $\|f(x, y) - f(x, z)\| \leq L \|y - z\|$  при  $(x, y), (x, z) \in D$ . (4.2b)

Если мы применим к задаче (4.1) многошаговый метод (2.1) с шагом  $h$ , то получим последовательность  $\{y_i\}$ . Для заданных  $x$

и  $h$ , таких что  $(x - x_0)/h = n$  — целое число, введем следующее обозначение численного решения:

$$y_h(x) = y_n, \text{ если } x - x_0 = nh. \quad (4.3)$$

Предполагается, что многошаговый метод — «хороший» в смысле сходимости численного решения  $y_h(x)$  к точному решению  $y(x)$  при  $h \rightarrow 0$ . Кроме того, мы надеемся на быструю сходимость в случае, когда функция  $f$  является более гладкой, чем предусмотрено условиями (4.2).

#### Определение 4.1. (Сходимость)

i) Линейный многошаговый метод (2.1) называется *сходящимся*, если для всех задач Коши (4.1), удовлетворяющих требованиям (4.2),

$$y(x) - y_h(x) \rightarrow 0 \quad \text{при } h \rightarrow 0, x \in [x_0, \chi],$$

при условии, что для стартовых значений выполняется соотношение

$$y(x_0 + ih) - y_h(x_0 + ih) \rightarrow 0 \quad \text{при } h \rightarrow 0, i = 0, 1, \dots, k-1.$$

ii) Метод (2.1) сходится с порядком  $p$ , если для любой задачи (4.1) с достаточно гладкой функцией  $f$  существует положительное  $h_0$ , такое, что

$$\|y(x) - y_h(x)\| \leq Ch^p \quad \text{при } h \leq h_0$$

и стартовых значениях, отвечающих неравенству

$$\|y(x_0 + ih) - y_h(x_0 + ih)\| \leq C_0 h^p$$

$$\text{при } h \leq h_0, i = 0, 1, \dots, k-1.$$

Очевидно, что в этом определении мы предполагаем существование решения задачи (4.1) на  $[x_0, \chi]$ .

В данном разделе мы ставим цель доказать, что устойчивость совместно с согласованностью являются необходимым и достаточным условием сходимости многошагового метода. Это выражается знаменитой формулой

$$\text{сходимость} = \text{устойчивость} + \text{согласованность}$$

[сравните также с работой Лакса—Рихтмайера (1956)]. Начнем с изучения необходимых условий сходимости.

**Теорема 4.2.** *Если многошаговый метод (2.1) сходится, то он обязательно является:*

- i) *устойчивым и*
- ii) *согласованным (т. е. имеет первый порядок):*

$$\rho(1) = 0, \rho'(1) = \sigma(1)).$$

*Доказательство.* Применение многошагового метода (2.1) к дифференциальному уравнению  $y' = 0$ ,  $y(0) = 0$  дает разностное уравнение (3.6). Предположим противное, т. е. что  $\rho(\zeta)$  имеет корень  $\zeta_1$ , модуль которого  $|\zeta_1| > 1$ , или кратный корень  $\zeta_2$  на единичной окружности. Тогда  $\zeta_1^n$  и  $n\zeta_2^n$  будут расходящимися решениями (3.6). Умножив стартовые значения на  $\sqrt{h}$ , мы добьемся их сходимости к  $y_0 = 0$  при  $h \rightarrow 0$ . Решения  $y_h(x) = \sqrt{h}\zeta_1^{x/h}$  и  $y_h(x) = (x/\sqrt{h})\zeta_2^{x/h}$  останутся при этом расходящимися при любом фиксированном  $x$ , что противоречит предложению о сходимости. Поэтому метод (2.1) должен быть устойчивым.

Теперь рассмотрим задачу Коши  $y' = 0$ ,  $y(0) = 1$  с точным решением  $y(x) = 1$ . Ей также соответствует разностное уравнение (3.6), которое в новых обозначениях имеет вид

$$\alpha_k y_h(x + kh) + \alpha_{k-1} y_h(x + (k-1)h) + \dots + \alpha_0 y_h(x) = 0.$$

При подстановке  $h \rightarrow 0$  из сходимости сразу следует, что  $\rho(1) = 0$ .

И наконец, применим метод (2.1) к задаче  $y' = 1$ ,  $y(0) = 0$ . Ее точное решение  $y(x) = x$ . Поскольку мы уже установили равенство  $\rho(1) = 0$ , легко проверить, что частное решение разностного уравнения имеет вид  $y_n = nhK$ , или  $y_h(x) = xK$ , где  $K = \sigma(1)/\rho'(1)$ . Для сходимости необходимо, чтобы выполнялось равенство  $K = 1$ . ■

Примечательно, что, хотя теорема 4.2 доказывается с помощью почти тривиальных дифференциальных уравнений, условия i) и ii) окажутся не только необходимыми, но и достаточными для сходимости.

### Представление в виде одношагового метода

Для дальнейшего изложения полезно переписать многошаговый метод как одношаговый, но в пространстве большей размерности [см. работы Бутчера (1966), Скила (1976)]. Для этого введем величину  $\psi = \psi(x_i, y_i, \dots, y_{i+k-1}, h)$ , неявно определяемую уравнением

$$\psi = \sum_{l=0}^{k-1} \beta_l f(x_i + lh, y_{i+l}) + \beta_k f\left(x_i + kh, h\psi - \sum_{l=0}^{k-1} \alpha_l y_{i+l}\right), \quad (4.4)$$

где  $\alpha'_j = \alpha_j/\alpha_k$ ,  $\beta'_j = \beta_j/\alpha_k$ . Тогда многошаговую формулу (2.1) можно представить в виде

$$y_{i+k} = - \sum_{l=0}^{k-1} \alpha'_l y_{i+l} + h\psi. \quad (4.5)$$

Введем векторы размерности  $mk$  ( $m$  — размерность системы дифференциальных уравнений)

$$Y_i = (y_{i+k-1}, y_{i+k-2}, \dots, y_i)^T, \quad i \geq 0, \quad (4.6)$$

и

$$A = \begin{bmatrix} -\alpha'_{k-1} & -\alpha'_{k-2} & \dots & -\alpha'_0 \\ 1 & 0 & \dots & 0 \\ & 1 & \ddots & 0 \\ & & \vdots & \vdots \\ & & 1 & 0 \end{bmatrix}, \quad e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (4.7)$$

и, добавив ряд тривиальных тождеств, запишем многошаговый метод (4.5) в компактной форме

$$Y_{i+1} = (A \otimes I) Y_i + h \Phi(x_i, Y_i, h), \quad i \geq 0, \quad (4.8)$$

где

$$\Phi(x_i, Y_i, h) = (e_1 \otimes I) \psi(x_i, Y_i, h). \quad (4.8a)$$

Здесь через  $A \otimes I$  обозначено тензорное произведение Кронекера, т. е. блочная матрица размерности  $mk$ , с  $(m, m)$ -блоками  $a_{ij}I$ . Читатели, не знакомые с обозначением и свойствами этого произведения, могут для простоты считать (4.1) скалярным уравнением ( $m = 1$ ) и  $A \otimes I = A$ .

Следующие леммы показывают, как в этих новых обозначениях выражаются понятия порядка и устойчивости.

**Лемма 4.3.** Пусть  $y(x)$  — точное решение задачи (4.1). Для  $i = 0, 1, \dots$  определим вектор  $\hat{Y}_{i+1}$  численного решения после одного шага

$$\hat{Y}_{i+1} = (A \otimes I) Y(x_i) + h \Phi(x_i, Y(x_i), h)$$

с точными стартовыми значениями

$$Y(x_i) = (y(x_{i+k-1}), y(x_{i+k-2}), \dots, y(x_i))^T.$$

i) Если многошаговый метод (2.1) имеет первый порядок, а  $f$  удовлетворяет условиям (4.2), то существует  $h_0 > 0$ , такое что для  $h \leq h_0$

$$\|Y(x_{i+1}) - \hat{Y}_{i+1}\| \leq h \omega(h), \quad 0 \leq i \leq k/h - k,$$

где  $\omega(h) \rightarrow 0$  при  $h \rightarrow 0$ .

ii) Если многошаговый метод (2.1) имеет порядок  $p$ , а  $f$  дифференцируема достаточночное число раз, то существует постоянная  $M$ , такая что для всех достаточно малых  $h$

$$\|Y(x_{i+1}) - \hat{Y}_{i+1}\| \leq M h^{p+1}, \quad 0 \leq i \leq k/h - k.$$

*Доказательство.* По определению 2.1 первая компонента вектора  $Y(x_{t+1}) - \hat{Y}_{t+1}$  является локальной погрешностью. Поскольку остальные компоненты равны нулю, из упражнения 5 в разд. III.2 и определения 2.3 следует утверждение теоремы. ■

**Лемма 4.4.** *Предположим, что многошаговый метод (2.1) устойчив. Тогда существует векторная норма (в  $\mathbb{R}^{mk}$ ), такая что матрица  $A$  из (4.7) удовлетворяет соотношению*

$$\|A \otimes I\| \leq 1$$

*в подчиненной матричной норме.*

*Доказательство.* Если  $\lambda$  — корень многочлена  $\rho(\zeta)$ , то  $(\lambda^{k-1}, \lambda^{k-2}, \dots, 1)$  является собственным вектором матрицы  $A$ , соответствующим собственному числу  $\lambda$ . Поэтому собственные числа матрицы  $A$  (т. е. корни  $\rho(\zeta)$ ) удовлетворяют корневому условию из определения 3.2. Следовательно, ее каноническая жорданова форма имеет вид (см. разд. 1.12):

$$T^{-1}AT = J = \begin{array}{c|c} \lambda_1 & \\ \cdots & \\ \hline \lambda_l & 0 \\ \hline & \lambda_{l+1} \ v_{l+1} \\ 0 & \cdots \ v_{k-1} \\ & \lambda_k \end{array} \quad (4.9)$$

где  $\lambda_1, \dots, \lambda_l$  — собственные числа, которые равны по модулю единице и должны быть простыми. Теперь, умножив столбцы матрицы  $T$  на соответствующие числа, установим, что  $|v_j| < 1 - |\lambda_j|$  при  $j = l + 1, \dots, k - 1$ . Тогда в силу (9.11') из гл. I выполняется неравенство  $\|J \otimes I\|_\infty \leq 1$ . С помощью преобразования  $T$  из (4.9) определим норму  $\|x\| := \|(T^{-1} \otimes I)x\|_\infty$ . Итак, получим

$$\begin{aligned} \|(A \otimes I)x\| &= \|(T^{-1} \otimes I)(A \otimes I)x\|_\infty = \|(J \otimes I)(T^{-1} \otimes I)x\|_\infty \\ &\leq \|(T^{-1} \otimes I)x\|_\infty = \|x\|, \end{aligned}$$

а следовательно,  $\|A \otimes I\| \leq 1$ . ■

### Доказательство сходимости

Теперь можно доказать теорему о сходимости многошаговых методов.

**Теорема 4.5.** *Если многошаговый метод (2.1) устойчив и имеет первый порядок, то он сходится. Если метод (2.1) устойчив и имеет порядок  $p$ , то он сходится с порядком  $p$ .*

*Доказательство.* Как и в теореме о сходимости одношаговых методов (разд. II.3), мы можем считать без ограничения общности, что функция  $f(x, y)$  определена для всех  $y \in \mathbb{R}^m$ ,  $x \in [x_0, \hat{x}]$  и удовлетворяет на этом множестве (глобальному) условию Липшица. Отсюда следует, что при достаточно малых  $h$  для функций  $\psi(x_i, Y_i, h)$  и  $\Phi(x_i, Y_i, h)$  выполняется условие Липшица по второму аргументу (с константой Липшица  $L^*$ ). Поэтому для  $G$  —

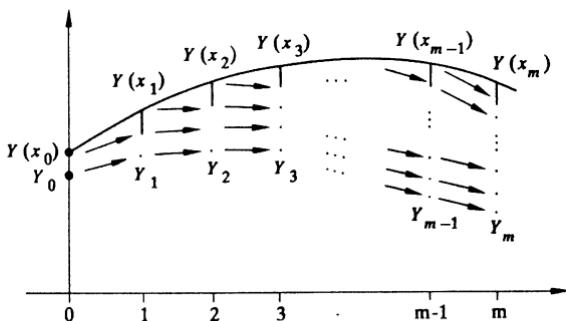


Рис. 4.1. Веер леди Уиндермир для многошаговых методов.

функции перехода от вектора  $Y_i$  к  $Y_{i+1}$ , заданной формулой (4.8), из леммы 4.4 получим

$$\|G(Y_i) - G(Z_i)\| \leq (1 + hL^*) \|Y_i - Z_i\|. \quad (4.10)$$

Далее доказательство проводится так же, как для одношаговых методов, а его иллюстрация представлена на рис. 4.1. Здесь стрелки указывают на действие  $G$ . Из леммы 4.3 известно, что  $\|Y(x_{i+1}) - G(Y(x_i))\| \leq h\omega(h)$ . Это соотношение совместно с (4.10) показывает, что локальная погрешность  $Y(x_{i+1}) - G(Y(x_i))$  этапа  $i + 1$  на этапе  $m$  дает погрешность, не превосходящую  $h\omega(h) \times (1 + hL^*)^{m-i+1}$ . Таким образом,

$$\begin{aligned} \|Y(x_m) - Y_m\| &\leq \|Y(x_0) - Y_0\|(1 - hL^*)^m + \\ &+ h\omega(h)[(1 + hL^*)^{m-1} + (1 + hL^*)^{m-2} + \dots + 1] \leq \\ &\leq \|Y(x_0) - Y_0\| \exp(mhL^*) + \frac{\omega(h)}{L^*} (\exp(mhL^*) - 1). \end{aligned} \quad (4.11)$$

Сходимость метода (2.1) является прямым следствием формулы (4.11). Если многошаговый метод имеет порядок  $p$ , то сходимость с порядком  $p$  доказывается точно так же, но с заменой  $\omega(h)$  на  $Mh^p$ . ■

## Упражнения

1. Рассмотрим функцию (при  $x \geq 0$ )

$$f(x, y) = \begin{cases} 2x, & y \leq 0, \\ 2x - \frac{4y}{x}, & 0 < y < x^2, \\ -2x, & y \geq x^2. \end{cases}$$

а) Покажите, что  $y(x) = x^3/3$  — единственное решение задачи  $y' = f(x, y)$ ,  $y(0) = 0$ , хотя  $f$  и не удовлетворяет условию Липшица в окрестности начала координат.

б) Примените к этой задаче правило средней точки (2.2') со стартовыми значениями  $y_0 = 0$ ,  $y_1 = -h^2$  и проверьте, что численное решение в точке  $x = nh$  имеет вид  $y_h(x) = (-1)^n x^2$  (см. книгу Григорьева (1977)).

2. Другое обоснование выбора константы погрешности: предположим, что 1 — единственное собственное значение матрицы  $A$  из (4.7), равное по модулю единице. Покажите, что ему соответствуют правый собственный вектор  $(1, 1, \dots, 1)^T$  и левый собственный вектор  $(1, 1 + \alpha'_{k-1}, 1 + \alpha'_{k-1} + \alpha'_{k-2}, \dots)$ . Тогда глобальный вклад локальной погрешности после большого числа шагов равен

$$A^\infty = \begin{pmatrix} C_{p+1} \\ 0 \\ \vdots \\ 0 \end{pmatrix} = C \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}. \quad (4.12)$$

Умножив это уравнение слева на левый собственный вектор, покажите, что с учетом (2.6)  $C$  будет константой погрешности, определенной с помощью (2.13).

*Замечание.* Для многошаговых методов с несколькими собственными значениями, равными по модулю 1, формула (4.12) также верна, но с заменой  $A^\infty$  на  $E$  (см. разд. III.8).

### III.5. Многошаговые методы с переменным шагом

«Это был крепкий орешек ...»  
(Тирольское присловье)

Из разд. II.4 становится понятно, что в эффективной программе численного интегрирования должно быть предусмотрено изменение величины шага. Однако в случае многошаговых методов эта процедура вызывает затруднения, так как формулы из предыдущих разделов требуют значений численного решения в равнотстоящих точках. Принципиально существуют две возможности преодоления этого затруднения:

- i) воспроизведение стартовых значений на новой (равномерной) сетке с помощью интерполяции многочленами;
- ii) построение методов, приведенных в соответствие узлам неравномерной сетки.

Данный раздел посвящен второму подходу. Исследуем для этого согласованность, устойчивость и сходимость. Его современная реализация (стратегия выбора порядка и шага) будет рассмотрена в разд. III.7.

#### Методы Адамса с переменным шагом

Вероятно, Ф. Ческино (1961) первым предложил «гладкий» переход от шага  $h$  к новому шагу  $\omega h$ . Его идеи развили Форрингтон (1961), а позднее Крох (1969): рассмотрим произвольную сетку  $(x_n)$  и обозначим величины шагов через  $h_n = x_{n+1} - x_n$ . Будем считать величины  $y_j$ , аппроксимирующие  $y(x_j)$ , известными при  $j = n - k + 1, \dots, n$  и введем обозначение  $f_j = f(x_j, y_j)$ . Точно так же, как в разд. III.1, обозначим через  $p(t)$  интерполяционный многочлен, проходящий через точки  $(x_j, f_j)$ ,  $j = n - k + 1, \dots, n$ . Используя интерполяционную формулу Ньютона, запишем

$$p(t) = \sum_{j=0}^{k-1} \prod_{i=0}^{j-1} (t - x_{n-i}) \tilde{f}[x_n, x_{n-1}, \dots, x_{n-j}], \quad (5.1)$$

где разделенные разности  $\tilde{f}[x_n, \dots, x_{n-j}]$  определяются рекур-сивно:

$$\begin{aligned} \tilde{f}[x_n] &= f_n, \\ \tilde{f}[x_n, \dots, x_{n-j}] &= \frac{\tilde{f}[x_n, \dots, x_{n-j+1}] - \tilde{f}[x_{n-1}, \dots, x_{n-j}]}{x_n - x_{n-j}}. \end{aligned} \quad (5.2)$$

Для практических расчетов формулу (5.1) удобно переписать в виде [см. статью Кроха (1969)]

$$p(t) = \sum_{l=0}^{k-1} \prod_{t=0}^{l-1} \frac{t - x_{n-t}}{x_{n+1} - x_{n-t}} \Phi_l^*(n), \quad (5.1')$$

где

$$\Phi_l^*(n) = \prod_{t=0}^{l-1} (x_{n+1} - x_{n-t}) \cdot f[x_n, \dots, x_{n-l}]. \quad (5.3)$$

Теперь определим значение, приближающее  $y(x_{n+1})$ :

$$y_{n+1} = y_n + \int_{x_n}^{x_{n+1}} p(t) dt. \quad (5.4)$$

Подставляя в (5.4) формулу (5.1'), получим

$$y_{n+1} = y_n + h_n \sum_{l=0}^{k-1} g_l(n) \Phi_l^*(n), \quad (5.5)$$

где

$$g_l(n) = \frac{1}{h_n} \int_{x_n}^{x_{n+1}} \prod_{t=0}^{l-1} \frac{t - x_{n-t}}{x_{n+1} - x_{n-t}} dt. \quad (5.6)$$

Формула (5.5) — это обобщение явного метода Адамса (1.5) на случай переменного шага. Отметим, что при постоянном шаге приведенные выше выражения сводятся к следующим (упр. 1):

$$g_l(n) = \Psi_l, \quad \Phi_l^*(n) = \nabla^l f_n.$$

Подобным образом можно вывести *неявные* методы Адамса с переменным шагом. По аналогии с разд. III.1 обозначим через  $p^*(t)$  интерполяционный многочлен степени  $k$ , проходящий через точки  $(x_j, f_j)$ ,  $j = n - k + 1, \dots, n, n + 1$ . (Значение  $f_{n+1} = f(x_{n+1}, y_{n+1})$  содержит неизвестное решение  $y_{n+1}$ .) Снова используя интерполяционную формулу Ньютона, получим:

$$p^*(t) = p(t) + \prod_{l=0}^{k-1} (t - x_{n-l}) f[x_{n+1}, x_n, \dots, x_{n-k+1}].$$

Численное решение, определяемое формулой

$$y_{n+1} = y_n + \int_{x_n}^{x_{n+1}} p^*(t) dt,$$

теперь можно представить в виде

$$y_{n+1} = p_{n+1} + h_n g_k(n) \Phi_k(n + 1). \quad (5.7)$$

Здесь  $p_{n+1}$  — приближенное значение решения, полученное с помощью явного метода Адамса

$$p_{n+1} = y_n + h_n \sum_{l=0}^{k-1} g_l(n) \Phi_l^*(n),$$

а

$$\Phi_k(n+1) = \sum_{t=0}^{k-1} (x_{n+1} - x_{n-t}) f[x_{n+1}, x_n, \dots, x_{n-k+1}]. \quad (5.8)$$

### Рекуррентные соотношения для $g_l(n)$ , $\Phi_l(n)$ и $\Phi_l^*(n)$

«Если допустить произвольное изменение шага, то наибольшим недостатком будут затраты на подсчет коэффициентов метода интегрирования».

(Ф. Т. Крох, 1973)

Значения  $\Phi_j^*(n)$  ( $j = 0, \dots, k-1$ ) и  $\Phi_k(n+1)$  можно эффективно вычислять с помощью рекуррентных соотношений

$$\begin{aligned} \Phi_0(n) &= \Phi_0^*(n) = f_n, \\ \Phi_{j+1}(n) &= \Phi_j(n) - \Phi_j^*(n-1), \\ \Phi_j^*(n) &= \beta_j(n) \Phi_j(n), \end{aligned} \quad (5.9)$$

которые сразу следуют из определений (5.3) и (5.8). Коэффициент

$$\beta_j(n) = \prod_{t=0}^{j-1} \frac{x_{n+1} - x_{n-t}}{x_n - x_{n-t-1}}$$

можно вычислить по формулам

$$\begin{aligned} \beta_0(n) &= 1, \\ \beta_j(n) &= \beta_{j-1}(n) \frac{x_{n+1} - x_{n-j+1}}{x_n - x_{n-j}}. \end{aligned}$$

Сложнее подсчитать коэффициенты  $g_j(n)$  [Ф. Т. Крох (1974)]. Введем  $q$ -кратный интеграл

$$c_{jq}(x) = \frac{(q-1)!}{h_n^q} \int_{x_n}^{x_{n+1}} \int_{x_n}^{\xi_{q-1}} \cdots \int_{x_n}^{\xi_1} \prod_{t=0}^{j-1} \frac{\xi_t - x_{n-t}}{x_{n+1} - x_{n-t}} d\xi_0 \cdots d\xi_{q-1} \quad (5.10)$$

и заметим, что

$$g_j(n) = c_{j1}(x_{n+1}).$$

**Лемма 5.1.** Верны следующие равенства:

$$c_{0q}(x_{n+1}) = \frac{1}{q}, \quad c_{1q}(x_{n+1}) = \frac{1}{q \cdot (q+1)},$$

$$c_{jq}(x_{n+1}) = c_{j-1,q}(x_{n+1}) - c_{j-1,q+1}(x_{n+1}) \frac{h_n}{x_{n+1} - x_{n-j+1}}.$$

**Доказательство.** Первые два соотношения следуют непосредственно из (5.10). Для доказательства рекуррентного соотношения введем разность

$$d(x) = c_{jq}(x) - c_{j-1,q}(x) \frac{x - x_{n-j+1}}{x_{n+1} - x_{n-j+1}} + c_{j-1,q+1}(x) \frac{h_n}{x_{n+1} - x_{n-j+1}}.$$

Ясно, что  $d^{(i)}(x_n) = 0$  при  $i = 0, 1, \dots, q-1$ . Кроме того, равна нулю  $q$ -я производная  $d(x)$ , так как по правилу Лейбница

$$\begin{aligned} \frac{d^q}{dx^q} \left[ c_{j-1,q}(x) \cdot \frac{x - x_{n-j+1}}{x_{n+1} - x_{n-j+1}} \right] &= \\ &= c_{j-1,q}^{(q)}(x) \frac{x - x_{n-j+1}}{x_{n+1} - x_{n-j+1}} + qc_{j-1,q}^{(q-1)}(x) \frac{1}{x_{n+1} - x_{n-j+1}} = \\ &= c_{j,q}^{(q)}(x) + c_{j-1,q+1}^{(q)}(x) \frac{h_n}{x_{n+1} - x_{n-j+1}}. \end{aligned}$$

Итак, выполняется тождество  $d(x) \equiv 0$ , и, подставив в него  $x = x_{n+1}$ , мы придем к утверждению леммы. ■

С помощью последнего рекуррентного соотношения можно последовательно вычислять  $c_{2q}(x_{n+1})$  при  $q = 1, \dots, k-1$ ;  $c_{3q}(x_{n+1})$  при  $q = 1, \dots, k-2; \dots$ ;  $c_{kq}(x_{n+1})$  при  $q = 1$ . Эта процедура позволяет эффективно определять коэффициенты  $g_j(n) = c_{j1}(x_{n+1})$  методов Адамса.

### Формулы дифференцирования назад с переменным шагом

ФДН-формулы (1.22) также можно обобщить естественным образом на случай переменного шага. Обозначим через  $q(t)$  интерполяционный многочлен степени  $k$ , проходящий через точки  $(x_i, y_i)$ ,  $i = n+1, n, \dots, n-k+1$ . Его можно выразить через разделенные разности

$$q(t) = \sum_{i=0}^k \prod_{l=0}^{i-1} (t - x_{n+1-l}) y[x_{n+1}, x_n, \dots, x_{n-j+1}]. \quad (5.11)$$

### Требование

$$q'(x_{n+1}) = f(x_{n+1}, y_{n+1})$$

сразу приводит к ФДН-формуле с переменным шагом

$$\sum_{l=1}^k h_n \prod_{i=1}^{l-1} (x_{n+1} - x_{n+1-i}) y[x_{n+1}, \dots, x_{n-l+1}] = h_n f(x_{n+1}, y_{n+1}). \quad (5.12)$$

Вычисление ее коэффициентов намного проще, чем в методах Адамса.

### Многошаговые методы общего вида с переменным шагом и их порядок согласованности

В теоретических исследованиях методов удобно использовать их форму записи с линейным вхождением  $y_j$  и  $f_j$ . Например, неявный метод Адамса (5.7) принимает вид

$$y_{n+1} = y_n + \frac{h_n}{6(1+\omega_n)} ((3+2\omega_n)f_{n+1} + (3+\omega_n)(1+\omega_n)f_n - \omega_n^2 f_{n-1}), \quad (5.13)$$

где для отношения шагов введено обозначение  $\omega_n = h_n/h_{n-1}$ . Или 2-шаговую ФДН-формулу (5.12) можно записать в форме

$$y_{n+1} - \frac{(1+\omega_n)^2}{1+2\omega_n} y_n + \frac{\omega_n^2}{1+2\omega_n} y_{n-1} = h_n \frac{1+\omega_n}{1+2\omega_n} f_{n+1}. \quad (5.14)$$

Чтобы дать единую теорию для всех этих многошаговых методов с переменным шагом, рассмотрим формулы следующего вида:

$$y_{n+k} + \sum_{j=0}^{k-1} \alpha_{jn} y_{n+j} = h_{n+k-1} \sum_{j=0}^k \beta_{jn} f_{n+j}. \quad (5.15)$$

Коэффициенты  $\alpha_{jn}$  и  $\beta_{jn}$  в действительности зависят от отношений  $\omega_i = h_i/h_{i-1}$ ,  $i = n+1, \dots, n+k-1$ . По аналогии со случаем постоянного шага дадим следующее определение.

**Определение 5.2.** Метод (5.15) согласован с порядком  $p$ , если для всех многочленов  $q(x)$ , степень которых не превосходит  $p$ , и всех сеток  $(x_j)$  выполняется равенство

$$q(x_{n+k}) + \sum_{j=0}^{k-1} \alpha_{jn} q(x_{n+j}) = h_{n+k-1} \sum_{j=0}^k \beta_{jn} q'(x_{n+j}).$$

По определению явный метод Адамса (5.5) имеет порядок  $k$ , неявный метод Адамса (5.7) — порядок  $k+1$ , а ФДН-формула (5.12) — порядок  $k$ .

Понятие согласованности надо, конечно, связать с локальной погрешностью. Действительно, если метод имеет порядок  $p$ , от-

ношения  $h_j/h_n$  ограничены при  $j = n + 1, \dots, n + k - 1$  и коэффициенты

$$\alpha_{jn}, \beta_{jn} \text{ ограничены,} \quad (5.16)$$

то для достаточно гладких функций  $y(x)$  разложение в ряд Тейлора приводит к соотношению

$$y(x_{n+k}) + \sum_{j=0}^{k-1} \alpha_{jn} y(x_{n+j}) - h_{n+k-1} \sum_{j=0}^k \beta_{jn} y'(x_{n+j}) = O(h_n^{p+1}). \quad (5.17)$$

Считая  $y(x)$  решением дифференциального уравнения, из тривиального обобщения леммы 2.2 на переменный шаг получим, что локальная погрешность в точке  $x_{n+k}$  (см. определение 2.1) также равна  $O(h_n^{p+1})$ .

По этой причине необходимо изучить условие (5.16). Очевидно, что методы (5.13) и (5.14) удовлетворяют ему всегда, когда отношение шагов  $h_n/h_{n-1}$  ограничено сверху. В общем случае мы имеем следующий результат.

**Лемма 5.3.** Для явных и неявных методов Адамса, а также для ФДН-формул коэффициенты  $\alpha_{jn}$  и  $\beta_{jn}$  ограничены, если для некоторого  $\Omega$  выполняется соотношение

$$h_n/h_{n-1} \leq \Omega.$$

**Доказательство.** Приведем доказательство только для явных методов Адамса. Для остальных методов оно мало отличается от данного, и поэтому опустим его. Из формулы (5.5) видно, что коэффициенты  $\alpha_{jn}$  не зависят от  $n$ , а поэтому ограничены. Значения  $\beta_{jn}$  состоят из произведений  $g_j(n)$  на коэффициенты выражения  $\Phi_j^*(n)$ , записанного в виде линейной комбинации величин  $f_n, \dots, f_{n-j}$ . Из формулы (5.6) следует соотношение  $|g_j(n)| \leq 1$ . В силу неравенства  $x_{n+1} - x_{n-j+1} \leq \max(1, \Omega^j)(x_n - x_{n-j})$  по индукции получим, что коэффициенты выражения  $\Phi_j^*(n)$  также ограничены. Поэтому и  $\beta_{jn}$  ограничены, что и доказывает лемму. ■

Условие  $h_n/h_{n-1} \leq \Omega$  — это вполне разумное требование, которое легко предусмотреть в программе.

## Устойчивость

«Это так просто...»

(Р. Д. Григорьев, семинар в Халле, 1983)

Для методов с переменным шагом изучение устойчивости было начато в статьях Гира—Ту (1974) и Гира—Ватанабе (1974). Дальнейшие исследования проведены Григорьевым (1983) и Крузе — Лисбоной (1984).

В разд. III. 3 мы увидели, что на равномерных сетках устойчивость эквивалентна ограниченности численного решения, полученного для скалярного дифференциального уравнения  $y' = 0$ . Сделаем и теперь то же самое, но для более общего случая. Применение метода (5.15) к задаче  $y' = 0$  дает разностное уравнение с переменными коэффициентами

$$\theta_{n+h} + \sum_{l=0}^{k-1} \alpha_{jn} y_{n+l} = 0.$$

Вводя вектор  $Y_n = (y_{n+k-1}, \dots, y_n)^T$ , перейдем к эквивалентному уравнению

$$Y_{n+1} = A_n Y_n$$

с сопровождающей матрицей

$$A_n = \begin{bmatrix} -\alpha_{k-1,n} & \dots & \dots & -\alpha_{1,n} & -\alpha_{0,n} \\ 1 & 0 & \dots & 0 & 0 \\ & & & \vdots & \vdots \\ & & 1 & 0 & 0 \\ & & & 1 & 0 \end{bmatrix}. \quad (5.18)$$

**Определение 5.4.** Метод (5.15) называется *устойчивым*, если для всех  $n$  и  $l \geq 0$  выполняется неравенство

$$\|A_{n+1} A_{n+l-1} \dots A_{n+1} A_n\| \leq M. \quad (5.19)$$

Отметим, что в общем случае  $A_n$  зависит от отношений шагов  $\omega_{n+1}, \dots, \omega_{n+k-1}$ . Поэтому условие (5.19) обычно будет приводить к ограничениям на эти величины.

Поскольку у методов Адамса (5.5) и (5.7) коэффициенты  $\alpha_{jn}$  не зависят от  $n$ , они устойчивы при любой последовательности шагов.

В следующих трех теоремах мы представим результаты об устойчивости методов общего вида с переменным шагом. Первая теорема взята из работы Крузе—Лисбоны (1984) и основана на теории возмущений: метод с переменным шагом считается возмущением сильно устойчивого метода с фиксированным шагом.

**Теорема 5.5.** Пусть метод (5.15) обладает следующими свойствами:

- a) порядок  $p \geq 0$ , т. е.  $1 + \sum_{l=0}^{k-1} \alpha_{jn} = 0$ ;
- b) коэффициенты  $\alpha_{jn} = \alpha_j(\omega_{n+1}, \dots, \omega_{n+k-1})$  — непрерывные в окрестности точки  $(1, \dots, 1)$  функции;

с) соответствующая формула с постоянным шагом сильно устойчива, т. е. все корни многочлена

$$\zeta^k + \sum_{j=0}^{k-1} \alpha_j(1, \dots, 1) \zeta^j = 0$$

лежат в открытом круге единичного радиуса  $|\zeta| < 1$ ; за исключением корня  $\zeta_1 = 1$ .

Тогда существуют вещественные числа  $\omega, \Omega$  ( $\omega < 1 < \Omega$ ), такие что при выполнении условия

$$\omega \leq h_n/h_{n-1} \leq \Omega \text{ при всех } n \quad (5.20)$$

метод устойчив.

*Доказательство.* Пусть  $A$  — сопровождающая матрица формулы с постоянным шагом. Как и в доказательстве леммы 4.4, приведем к канонической жордановой форме

$$T^{-1}AT = \left( \begin{array}{c|c} \widehat{A} & \begin{matrix} 0 \\ \vdots \\ 0 \\ 1 \end{matrix} \end{array} \right),$$

где по предположению с)  $\|\widehat{A}\|_1 < 1$ . Отметим, что стоящий в последнем столбце матрицы  $T$  собственный вектор матрицы  $\widehat{A}$ , соответствующий ее собственному числу 1, равен  $t_k = (1, \dots, 1)^T$ . Из предположения а) следует, что  $t_k$  является также собственным вектором и для всех матриц  $A_n$ . Поэтому верно равенство

$$T^{-1}A_n T = \left( \begin{array}{c|c} \widehat{A}_n & \begin{matrix} 0 \\ \vdots \\ 0 \\ 1 \end{matrix} \end{array} \right),$$

и в силу непрерывности  $\|\widehat{A}_n\|_1 \leq 1$  при достаточно близких к 1 значениях  $\omega_{n+1}, \dots, \omega_{n+k-1}$ . Теперь устойчивость следует из соотношения

$$\|T^{-1}A_n T\|_1 = \max(\|\widehat{A}_n\|_1, 1) = 1,$$

которое приводит к неравенству

$$\|A_{n+1} \dots A_{n+k-1} A_n\| \leq \|T\| \cdot \|T^{-1}\|. \blacksquare$$

Следующий результат [Григорьев (1983)] основан на сокращении размера матриц  $A_n$  на единицу. Идея заключается в исполь-

зовании преобразования

$$T = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ & 1 & 1 & \dots & 1 \\ & & 1 & \dots & 1 \\ 0 & \ddots & & \vdots & \\ & & & & 1 \end{pmatrix}, \quad T^{-1} = \begin{pmatrix} 1 & -1 & & & \\ & 1 & -1 & & \\ & & 1 & \dots & \\ 0 & & & \dots & -1 \\ & & & & 1 \end{pmatrix}.$$

Отметим, что в последнем столбце матрицы  $T$  стоит вектор  $t_k$  из предыдущего доказательства. С помощью простых вычислений получим

$$T^{-1}A_n T = \left( \begin{array}{c|c} A_n^* & 0 \\ \hline e_{k-1}^T & 1 \end{array} \right).$$

Здесь  $e_{k-1}^T = (0, \dots, 0, 1)$ , а матрица  $A_n^*$  имеет вид

$$A_n^* = \begin{pmatrix} -\alpha_{k-2,n}^* & -\alpha_{k-3,n}^* & \dots & -\alpha_{1,n}^* & -\alpha_{0,n}^* \\ 1 & 0 & \dots & . & 0 \\ & 1 & \dots & . & 0 \\ & & \ddots & : & : \\ & & & 1 & 0 \end{pmatrix}, \quad (5.21)$$

где

$$\alpha_{k-2,n}^* = 1 + \alpha_{k-1,n}, \quad \alpha_{0,n}^* = -\alpha_{0,n}$$

$$\alpha_{k-l-1,n}^* - \alpha_{k-l,n}^* = \alpha_{k-l,n} \quad \text{при } j = 2, \dots, k-1.$$

Заметим, что  $\alpha_{j,n}^*$  совпадают с коэффициентами многочлена, определяемого следующей формулой:

$$\begin{aligned} (\zeta^k + \alpha_{k-1,n}\zeta^{k-1} + \dots + \alpha_{1,n}\zeta + \alpha_{0,n}) &= \\ &= (\zeta - 1)(\zeta^{k-1} + \alpha_{k-2,n}\zeta^{k-2} + \dots + \alpha_{1,n}\zeta + \alpha_{0,n}). \end{aligned}$$

**Теорема 5.6.** Пусть метод (5.15) имеет порядок  $p \geq 0$ . В этом случае метод устойчив тогда и только тогда, когда для всех  $n$  и  $l \geq 0$  выполняются условия

$$a) \|A_{n+1}^* \dots A_{n+l}^* A_n^*\| \leq M_1,$$

$$b) \left\| e_{k-1}^T \sum_{l=n}^{n+l} \prod_{i=n}^{j-1} A_i^* \right\| \leq M_2.$$

*Доказательство.* С помощью индукции легко показать, что

$$T^{-1}A_{n+1} \dots A_n T = \left( \begin{array}{c|c} A_{n+1}^* \dots A_n^* & 0 \\ \hline b_{n+1}^T & 1 \end{array} \right),$$

где

$$b_{n+1}^T = e_{k-1}^T \sum_{l=n}^{n+l} \prod_{i=n}^{j-1} A_i^*. \quad \blacksquare$$

Поскольку в данной теореме размер рассматриваемых матриц уменьшается на единицу, она особенно полезна для исследования устойчивости двухшаговых методов.

*Пример.* Рассмотрим двухшаговый ФДН-метод (5.14). Здесь

$$\alpha_{0n} = \frac{\omega_{n+1}^2}{1 + 2\omega_{n+1}}, \quad \alpha_{1n} = -1 - \alpha_{0n}.$$

Матрица (5.21) в этом случае имеет вид

$$A_n^* = (-\alpha_{0n}^*), \quad -\alpha_{0n}^* = \frac{\omega_{n+1}^2}{1 + 2\omega_{n+1}}.$$

При  $|\alpha_{0n}^*| \leq q < 1$  выполняются условия теоремы 5.6, из которых следует устойчивость. Данное соотношение верно, если имеет место неравенство

$$0 < h_{n+1}/h_n < \Omega < 1 + \sqrt{2}.$$

Интересным следствием сформулированной теоремы является неустойчивость двухшаговых ФДН-формул в случае, когда величина шага растет не медленнее, чем определено соотношением  $h_{n+1}/h_n \geq 1 + \sqrt{2}$ .

Исследовать устойчивость  $k$ -шаговых ( $k \geq 3$ ) методов значительно сложнее, так как в них входит несколько отношений шагов  $\omega_{n+1}, \omega_{n+2}, \dots$

Для ФДН-методов высокого порядка Григорьев (1983) нашел следующие оценки вида (5.20), обеспечивающие устойчивость:

Таблица 5.1. Оценки (5.20) для  $k$ -шаговых ФДН-формул

$k$	2	3	4	5
$\omega$	0	0.836	0.979	0.997
$\Omega$	2.414	1.127	1.019	1.003

Эти ограничения, конечно, нереалистичны, так как они включают все патологические изменения шага.

Менее пессимистичные оценки можно получить, предположив, что шаг изменяется более гладко [Гир—Ту (1974)]: известно, что локальная погрешность имеет вид  $d(x_n) h_n^{p+1} + O(h_n^{p+2})$ , где  $d(x)$  — ее главный член. С помощью управления величиной шага эта погрешность сохраняется равной значению TOL. Поэтому, если  $d(x)$  находится вне некоторой окрестности нуля, то выполняется равенство

$$h_n = |\text{TOL}/d(x_n)|^{1/(p+1)} + O(h_n),$$

из которого следует (при условии  $h_{n+1}/h_n \leq \Omega$ ), что

$$h_{n+1}/h_n = |d(x_n)/d(x_{n+1})|^{1/(p+1)} + O(h_n).$$

Если  $d(x)$  дифференцируема, то получим

$$|h_{n+1}/h_n - 1| \leq Ch_n. \quad (5.22)$$

На этом предположении основано несколько результатов Гира—Ту об устойчивости. («Следовательно, мы можем рассчитывать, что метод будет устойчивым, если устойчив метод с фиксированным шагом ...»). Суммируя неравенства (5.22), получим условие

$$\sum_{j=n}^{n+l} |h_{j+1}/h_j - 1| \leq C(\lambda - x_0),$$

содержащее только отношения шагов. Имеем следующую теорему.

**Теорема 5.7.** Пусть коэффициенты  $\alpha_{jn}$  метода (5.15) являются непрерывно дифференцируемыми функциями от  $\omega_{n+1}, \dots, \omega_{n+k-1}$  в окрестности  $\{(\omega_{n+1}, \dots, \omega_{n+k-1}); \omega \leq \omega_j \leq \Omega\}$ , и предположим, что метод устойчив при постоянных шагах (т. е. при  $\omega_j = 1$ ). Тогда из условия

$$\sum_{j=n}^{n+l} |h_{j+1}/h_j - 1| \leq C \quad \text{при всех } n \text{ и } l \geq 0 \quad (5.23)$$

совместно с  $\omega \leq h_{j+1}/h_j \leq \Omega$  следует условие устойчивости (5.19).

**Доказательство.** Как и в доказательстве теоремы 5.5, обозначим через  $A$  сопровождающую матрицу формулы с постоянным шагом, а через  $T$  соответствующее преобразование, такое что  $\|T^{-1}AT\| = 1$ . Применяя теорему о среднем к выражению  $\alpha_j(\omega_{n+1}, \dots, \omega_{n+k-1}) - \alpha_j(1, \dots, 1)$ , придем к соотношению

$$\|T^{-1}A_n T - T^{-1}AT\| \leq K \sum_{j=n+1}^{n+k-1} |\omega_j - 1|.$$

Следовательно,

$$\|T^{-1}A_n T\| \leq 1 + K \sum_{j=n+1}^{n+k-1} |\omega_j - 1| \leq \exp \left( K \sum_{j=n+1}^{n+k-1} |\omega_j - 1| \right).$$

Из этого неравенства выведем, что

$$\|A_{n+1} \dots A_{n+k-1} A_n\| \leq \|T\| \cdot \|T^{-1}\| \cdot \exp(K \cdot (k-1)C). \quad \blacksquare$$

### Сходимость

Для методов Адамса с переменным шагом сходимость впервые изучил Пиотровски (1969). Для доказательства сходимости в общем виде введем вектор  $Y_n = (y_{n+k-1}, \dots, y_{n+1}, y_n)^T$ . По аналогии с (4.8) метод (5.15) приводится к эквивалентному виду:

$$Y_{n+1} = (A_n \otimes I) Y_n + h_{n+k-1} \Phi_n(x_n, Y_n, h_n), \quad (5.24)$$

где матрица  $A_n$  задана в (5.18), а

$$\Phi_n(x_n, Y_n, h_n) = (e_1 \otimes I) \Psi_n(x_n, Y_n, h_n).$$

Значение  $\Psi = \Psi_n(x_n, Y_n, h_n)$  неявно задается соотношением

$$\Psi = \sum_{l=0}^{k-1} \beta_{jn} f(x_{n+l}, y_{n+l}) + \beta_{kn} f\left(x_{n+k}, h\Psi - \sum_{l=0}^{k-1} \alpha_{jn} y_{n+l}\right).$$

Далее обозначим через

$$Y(x_n) = (y(x_{n+k-1}), \dots, y(x_{n+1}), y(x_n))^T$$

точные значения, аппроксимируемые вектором  $Y_n$ . Теперь теорему о сходимости можно сформулировать следующим образом.

**Теорема 5.8.** Предположим, что

- a) метод (5.15) устойчив, имеет порядок  $p$ , а его коэффициенты  $\alpha_{jn}$ ,  $\beta_{jn}$  ограничены;
- b) для стартовых значений выполняется условие

$$\|Y(x_0) - Y_0\| = O(h_0^p);$$

- c) отношения шагов ограничены ( $h_n/h_{n-1} \leq \Omega$ ).

Тогда метод сходится с порядком  $p$ , т. е. для любого дифференциального уравнения  $y' = f(x, y)$ ,  $y(x_0) = y_0$  с достаточно гладкой функцией  $f$  глобальная погрешность удовлетворяет соотношению

$$\|y(x_n) - y_n\| \leq Ch^p \quad \text{при } x_n \leq \hat{x},$$

где  $h = \max h_j$ .

**Доказательство.** Поскольку метод имеет порядок  $p$ , а коэффициенты и отношения шагов ограничены, из формулы (5.17) видно, что для локальной погрешности

$$\delta_{n+1} = Y(x_{n+1}) - (A_n \otimes I) Y(x_n) - h_{n+k-1} \Phi_n(x_n, Y(x_n), h_n) \quad (5.25)$$

верно соотношение

$$\delta_{n+1} = O(h_n^{p+1}). \quad (5.26)$$

Вычитая (5.24) из (5.25), придем к равенству

$$Y(x_{n+1}) - Y_n = (A_n \otimes I)(Y(x_n) - Y_n) +$$

$$+ h_{n+k-1} \{\Phi_n(x_n, Y(x_n), h_n) - \Phi_n(x_n, Y_n, h_n)\} + \delta_{n+1},$$

откуда по индукции следует

$$\begin{aligned} Y(x_{n+1}) - Y_{n+1} &= ((A_n \dots A_0) \otimes I)(Y(x_0) - Y_0) + \\ &+ \sum_{j=0}^n h_{j+k-1} ((A_n \dots A_{j+1}) \otimes I) \{\Phi_j(x_j, Y(x_j), h_j) - \\ &- \Phi_j(x_j, Y_j, h_j)\} + \sum_{j=0}^n ((A_n \dots A_{j+1}) \otimes I) \delta_{j+1}. \end{aligned}$$

Как и в доказательстве теоремы 4.5, получим, что  $\Phi_n$  удовлетворяет равномерному условию Липшица по  $Y_n$ . С учетом устойчивости и равенства (5.26) выведем отсюда оценку

$$\|Y(x_{n+1}) - Y_{n+1}\| \leq \sum_{j=0}^n h_{j+k-1} L \|Y(x_j) - Y_j\| + C_1 h^p.$$

Для решения этого неравенства введем последовательность  $\{\varepsilon_n\}$ , определенную следующим образом:

$$\begin{aligned} \varepsilon_0 &= \|Y(x_0) - Y_0\|, \\ \varepsilon_{n+1} &= \sum_{j=0}^n h_{j+k-1} L \varepsilon_j + C_1 h^p. \end{aligned} \quad (5.27)$$

По индукции легко показать, что

$$\|Y(x_n) - Y_n\| \leq \varepsilon_n. \quad (5.28)$$

Из (5.27) при  $n \geq 1$  получим соотношение

$$\varepsilon_{n+1} = \varepsilon_n + h_{n+k-1} L \varepsilon_n \leq \exp(h_{n+k-1} L) \varepsilon_n,$$

а поэтому и неравенство

$$\begin{aligned} \varepsilon_n &\leq \exp((\lambda - x_0) L) \varepsilon_1 = \\ &= \exp((\lambda - x_0) L) \cdot \{h_{k-1} L \|Y(x_0) - Y_0\| + C_1 h^p\}, \end{aligned}$$

которое совместно с (5.28) и завершает доказательство теоремы 5.8. ■

### Упражнения

- Докажите, что при постоянном шаге выражения  $g_I(n)$  и  $\Phi_I^*(n)$  [формулы (5.3) и (5.6)] сводятся к виду
 
$$g_I(n) = \gamma_I, \quad \Phi_I^*(n) = \nabla^I f_n,$$
 где  $\gamma_I$  определены в (1.6).
- [Григорьев (1983)]. Для  $k$ -шаговых ФДН-методов рассмотрим сетки с постоянным отношением шагов  $\omega$ , т. е.  $h_n = \omega h_{n-1}$  при всех  $n$ . В этом случае элементы матрицы  $A_n^*$  (см. (5.21)) не зависят от  $n$ . Покажите с помощью численных расчетов, что все собственные значения  $A_n^*$  по модулю меньше единицы при  $0 < \omega < R_k$ , где  $R_k$  имеют следующие значения:

$k$	2	3	4	5	6
$R_k$	2.414	1.618	1.280	1.127	1.044

### III.6. Методы Нордсика

Важные результаты представлены в статье Нордсика (1962), где для обыкновенных дифференциальных уравнений рассматривается класс методов, позволяющих изменять шаг удобным способом (см. разд. III.7). Уже сам автор указал на то, что его методы эквивалентны в некотором смысле неявным методам Адамса. Сначала мы изложим его вывод методов, а затем изучим их связь с линейными многошаговыми методами.

Нордсик (1962) отметил, что «...все методы численного интегрирования эквивалентны нахождению многочлена, аппроксимирующего  $y(x)$ ...». Его идея состоит в том, чтобы представить этот многочлен через производные от нулевого до  $k$ -го порядка включительно, т. е. с помощью вектора («вектора Нордсика»)

$$z_n = \left( y_n, hy'_n, \frac{h^2}{2!} y''_n, \dots, \frac{h^k}{k!} y^{(k)}_n \right)^T. \quad (6.1)$$

Величины  $y_n^{(l)}$  имеют смысл приближенных значений для  $y^{(l)}(x_n)$ , где  $y(x)$  — точное решение дифференциального уравнения

$$y' = f(x, y). \quad (6.2)$$

Чтобы определить процедуру интегрирования, необходимо задать правило нахождения  $z_{n+1}$  по известным  $z_n$  и дифференциальному уравнению (6.2). При использовании разложения в ряд Тейлора (например, при  $k = 3$ ) такое правило имеет вид

$$\begin{aligned} y_{n+1} &= y_n + hy'_n + \frac{h^2}{2!} y''_n + \frac{h^3}{3!} y'''_n + \frac{h^4}{4!} e, \\ hy'_{n+1} &= hy'_n + 2 \frac{h^2}{2!} y''_n + 3 \frac{h^3}{3!} y'''_n + 4 \frac{h^4}{4!} e, \\ \frac{h^2}{2!} y''_{n+1} &= \frac{h^2}{2!} y''_n + 3 \frac{h^3}{3!} y'''_n + 6 \frac{h^4}{4!} e, \\ \frac{h^3}{3!} y'''_{n+1} &= \frac{h^3}{3!} y'''_n + 4 \frac{h^4}{4!} e, \end{aligned} \quad (6.3)$$

где значение  $e$  выбирается таким образом, чтобы выполнялось равенство

$$y'_{n+1} = f(x_{n+1}, y_{n+1}). \quad (6.4)$$

Подставляя (6.4) во второе соотношение из (6.3), получим

$$4 \frac{h^4}{4!} e = h(f(x_{n+1}, y_{n+1}) - f_n^p), \quad (6.5)$$

где

$$hf_n^p = hy_n' + 2 \frac{h^2}{2!} y_n'' + 3 \frac{h^3}{3!} y_n'''.$$

С учетом этого выражения для  $e$  метод принимает вид

$$\begin{aligned} y_{n+1} &= y_n + hy_n' + \frac{h^2}{2!} y_n'' + \frac{h^3}{3!} y_n''' + \frac{1}{4} (f(x_{n+1}, y_{n+1}) - f_n^p), \\ hy_{n+1}' &= \quad hy_n' + 2 \frac{h^2}{2!} y_n'' + 3 \frac{h^3}{3!} y_n''' + h (f(x_{n+1}, y_{n+1}) - f_n^p), \\ \frac{h^2}{2!} y_n'' &= \quad \frac{h^2}{2!} y_n'' + 3 \frac{h^3}{3!} y_n''' + \frac{3}{2} h (f(x_{n+1}, y_{n+1}) - f_n^p), \\ \frac{h^3}{3!} y_n''' &= \quad \frac{h^3}{3!} y_n''' + h (f(x_{n+1}, y_{n+1}) - f_n^p). \end{aligned} \tag{6.6}$$

Первое уравнение — неявное относительно  $y_{n+1}$ , а остальные — явные. Отметим, что если величины  $y_n^{(l)}$  достаточно точно аппроксимируют  $y^{(l)}(x_n)$ , то значение  $e$  [формула (6.5)] аппроксимирует  $y^{(4)}(x_n)$ . С точки зрения точности это свойство выглядит привлекательно. Но, к сожалению, метод (6.6) является неустойчивым. Чтобы показать это, подставим в (6.6)  $f(x, y) = 0$ . В этом случае метод обращается в линейное преобразование

$$z_{n+1} = Mz_n \tag{6.7}$$

с матрицей

$$M = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 1/4 \\ 1 \\ 3/2 \\ 1 \end{pmatrix} (0 \ 1 \ 2 \ 3).$$

Видно, что собственные значения  $M$  равны  $1, 0, -(2 + \sqrt{3})$  и  $-1/(2 + \sqrt{3})$ , откуда следует неустойчивость метода (6.6), а поэтому и его практическая непригодность. С часто встречающимся явлением неустойчивости методов высокой точности мы уже столкнулись в разд. III.3.

Чтобы преодолеть указанную трудность, Нордсик предложил заменить постоянные  $1/4, 1, 3/2, 1$ , стоящие в (6.6) перед скобками, на произвольные значения  $(l_0, l_1, l_2, l_3)$  и использовать эту дополнительную степень свободы для достижения устойчивости. Такую модификацию можно записать в компактной форме:

$$z_{n+1} = (P \otimes I) z_n + (l \otimes I) (hf(x_{n+1}, y_{n+1}) - (e_1^T P \otimes I) z_n). \tag{6.8}$$

Здесь  $z_n$  задано в (6.1),  $P$  — треугольная матрица Паскаля, опре-

делаемая формулой

$$P_{ij} = \begin{cases} \binom{j}{i} & \text{при } 0 \leq i \leq j \leq k, \\ 0 & \text{в противном случае,} \end{cases}$$

$\mathbf{l} = (l_0, l_1, \dots, l_k)^T$ , а  $e_1 = (0, 1, 0, \dots, 0)^T$ . Отметим, что индексы векторов и матриц начинаются с нуля.

Для простоты обозначений в следующих теоремах будем далее рассматривать только скалярные дифференциальные уравнения, для которых метод (6.8) принимает вид

$$\mathbf{z}_{n+1} = P\mathbf{z}_n + l [h f_{n+1} - e_1^T P \mathbf{z}_n]. \quad (6.8')$$

Конечно, все результаты остаются верными и для систем уравнений. Условие (6.4), связывающее этот метод с дифференциальным уравнением, фиксирует значение

$$l_1 = 1. \quad (6.9)$$

Анализируя описанным выше способом устойчивость метода общего вида (6.8), мы придем к разностному уравнению (6.7) с матрицей

$$M = P - le_1^T P. \quad (6.10)$$

Например, при  $k = 3$  эта матрица имеет вид

$$M = \begin{pmatrix} 1 & 1 - l_0 & 1 - 2l_0 & 1 - 3l_0 \\ 0 & 0 & 0 & 0 \\ 0 & -l_2 & 1 - 2l_2 & 3 - 3l_2 \\ 0 & -l_3 & -2l_3 & 1 - 3l_3 \end{pmatrix}.$$

Заметим, что 1 и 0 являются собственными значениями матрицы  $M$ , а ее характеристический многочлен не зависит от  $l_0$ . Нордсик определил  $l_2, \dots, l_k$  таким образом, чтобы остальные собственные значения  $M$  были равны нулю. В случае  $k = 3$  это выполняется при  $l_2 = 3/4, l_3 = 1/6$ . Коэффициент  $l_0$  можно выбрать из условия обращения в нуль константы погрешности метода (см. далее теорему 6.2). В нашем случае получается  $l_0 = 3/8$ , а весь метод задается вектором

$$\mathbf{l} = \left( \frac{3}{8}, 1, \frac{3}{4}, \frac{1}{6} \right)^T.$$

Интересно отметить, что этот метод эквивалентен неявному 3-шаговому методу Адамса. Действительно, исключая с помощью формулы (6.8) с уменьшенными индексами члены  $(h^3/3!)\ddot{y}_n$  и  $(h^2/2!)\ddot{y}_n$ , придем к методу [см. формулу (1.8'')]

$$y_{n+1} = y_n + \frac{h}{24} \cdot [9y'_{n+1} + 19y'_n - 5y'_{n-1} + y'_{n-2}]. \quad (6.11)$$

### Эквивалентность многошаговым методам

Работы Деклу (1963), Осборна (1966) и Скила (1978) позволяют глубже понять связь между методами Нордсика и многошаговыми методами. Следующие две теоремы показывают, что любой метод Нордсика эквивалентен некоторой многошаговой формуле и имеет порядок не ниже  $k$ .

**Теорема 6.1.** Рассмотрим метод Нордсика (6.8) с коэффициентом  $l_1 = 1$ . Тогда первые две компоненты вектора  $z_n$  удовлетворяют линейной многошаговой формуле (при  $n \geq 0$ )

$$\sum_{i=0}^k \alpha_i y_{n+i} = h \sum_{i=0}^k \beta_i f_{h+i}, \quad (6.12)$$

где производящие многочлены имеют вид

$$\begin{aligned} \rho(\zeta) &= \det(\zeta I - P) e_1^T (\zeta I - P)^{-1} l, \\ \sigma(\zeta) &= \det(\zeta I - P) e_0^T (\zeta I - P)^{-1} l. \end{aligned} \quad (6.13)$$

*Доказательство.* Приведенное в самих указанных статьях доказательство значительно упрощается при использовании производящих функций (дискретное преобразование Лапласа):

$$Z(\zeta) = \sum_{n \geq 0} z_n \zeta^n, \quad Y(\zeta) = \sum_{n \geq 0} y_n \zeta^n, \quad F(\zeta) = \sum_{n \geq 0} f_n \zeta^n, \dots$$

Умножив формулу (6.8') на  $\zeta^{n+1}$  и проведя суммирование, получим

$$Z(\zeta) = \zeta P Z(\zeta) + l(hF(\zeta) - e_1^T P \zeta Z(\zeta)) + (z_0 - lh f_0). \quad (6.14)$$

Подобным образом линейный многошаговый метод можно переписать в виде

$$\hat{\rho}(\zeta) Y(\zeta) = h \hat{\sigma}(\zeta) F(\zeta) + p_{k-1}(\zeta), \quad (6.15)$$

где

$$\hat{\rho}(\zeta) = \zeta^k \rho(1/\zeta), \quad \hat{\sigma}(\zeta) = \zeta^k \sigma(1/\zeta), \quad (6.16)$$

а  $p_{k-1}$  — многочлен степени  $k-1$ , зависящий от стартовых значений. Для доказательства теоремы надо показать, что первые две компоненты  $Z(\zeta)$  удовлетворяют соотношению вида (6.15). Сначала перепишем уравнение (6.14):

$$\begin{aligned} Z(\zeta) &= (I - \zeta P)^{-1} l(hF(\zeta) - e_1^T P \zeta Z(\zeta)) + \\ &\quad + (I - \zeta P)^{-1} (z_0 - lh f_0). \end{aligned}$$

При этом первые две компоненты  $Z(\zeta)$  примут вид

$$\begin{aligned} Y(\zeta) &= e_0^T (I - \zeta P)^{-1} l(hF(\zeta) - e_1^T P \zeta Z(\zeta)) + e_0^T (I - \zeta P)^{-1} (z_0 - lh f_0), \\ hF(\zeta) &= e_1^T (I - \zeta P)^{-1} l(hF(\zeta) - e_1^T P \zeta Z(\zeta)) + e_1^T (I - \zeta P)^{-1} (z_0 - lh f_0). \end{aligned}$$

Теперь, сокращая стоящие в скобках члены и умножая равенство на  $\det(I - \zeta P)$ , мы придем к формуле (6.15) с многочленами

$$\begin{aligned}\hat{\rho}(\zeta) &= \det(I - \zeta P) e_1^T (I - \zeta P)^{-1} l, \\ \hat{\sigma}(\zeta) &= \det(I - \zeta P) e_0^T (I - \zeta P)^{-1} l, \\ p_{k-1}(\zeta) &= \det(I - \zeta P) (e_1^T (I - \zeta P)^{-1} l e_0^T (I - \zeta P)^{-1} - \\ &\quad - e_0^T (I - \zeta P)^{-1} l e_1^T (I - \zeta P)^{-1}) z_0.\end{aligned}\quad (6.17)$$

С помощью (6.16) мы сразу получим формулы (6.13). Итак, остается показать, что многочлен  $p_{k-1}$  в (6.17) имеет степень  $k-1$ . Поскольку  $P$  — матрица размера  $(k+1) \times (k+1)$ , многочлен  $p_{k-1}$  при  $|\zeta| \rightarrow \infty$  ведет себя как  $\zeta^{k-1}$ . И наконец, из соотношения (6.15) следует, что ряд Лорана для  $p_{k-1}$  не может содержать членов с отрицательными степенями. ■

Подставляя в (6.13)  $(\zeta I - P)^{-1} l = u$  и применяя к линейной системе  $(\zeta I - P) u = l$  правило Крамера, из (6.13) получим изящное выражение

$$\rho(\zeta) = \det \begin{pmatrix} \zeta - 1 & l_0 & -1 & \dots & -1 \\ 0 & l_1 & -2 & \dots & -k \\ 0 & l_2 & \zeta - 1 & \dots & . \\ . & . & . & & . \\ . & . & . & & . \\ . & . & . & & . \\ 0 & l_k & 0 & \dots & \zeta - 1 \end{pmatrix}, \quad (6.13a)$$

$$\sigma(\zeta) = \det \begin{pmatrix} l_0 & -1 & -1 & \dots & -1 \\ l_1 & \zeta - 1 & -2 & \dots & -k \\ l_2 & 0 & \zeta - 1 & \dots & . \\ . & . & . & & . \\ . & . & . & & . \\ . & . & . & & . \\ l_k & 0 & 0 & & \zeta - 1 \end{pmatrix}. \quad (6.13b)$$

Заметим, что  $\rho(\zeta)$  не зависит от  $l_0$ . Кроме того,  $\zeta_0 = 1$  является простым корнем  $\rho(\zeta)$  тогда и только тогда, когда  $l_k \neq 0$ . Выполняется соотношение

$$\rho'(1) = \sigma(1) = k! l_k. \quad (6.18)$$

Условие (6.9) эквивалентно равенству  $\alpha_k = 1$ .

**Теорема 6.2.** Пусть  $l_k \neq 0$ . Определенный в (6.13) многошаговый метод имеет порядок не ниже  $k$ , а его константа погрешности

(см. (2.13)) вычисляется по формуле

$$C = -\frac{b^T l}{k! l_k}.$$

Здесь компоненты вектора

$$b^T = (B_0, B_1, \dots, B_k) = \left( 1, -\frac{1}{2}, \frac{1}{6}, 0, -\frac{1}{30}, 0, \frac{1}{42}, \dots \right)$$

являются числами Бернули.

*Доказательство.* По теореме 2.4 порядок равен  $k$  тогда и только тогда, когда

$$\rho(\zeta) - \log \zeta \cdot \sigma(\zeta) = C_{k+1}(\zeta - 1)^{+1} + O((\zeta - 1)^{k+2}).$$

С учетом равенства  $\det(\zeta I - P) = (\zeta - 1)^{k+1}$  это условие эквивалентно условию

$$e_1^T (\zeta I - P)^{-1} l - \log \zeta \cdot e_0^T (\zeta I - P)^{-1} l = C_{k+1} + O((\zeta - 1)),$$

а в силу (6.18) достаточно показать, что

$$(\log \zeta \cdot e_0^T - e_1^T) (\zeta I - P)^{-1} = b^T + O((\zeta - 1)). \quad (6.19)$$

Обозначив левую часть равенства (6.19) через  $b^T(\zeta)$ , получим

$$(\zeta I - P)^T b(\zeta) = (\log \zeta \cdot e_0 - e_1). \quad (6.20)$$

В этой системе  $q$ -е уравнение ( $q \geq 2$ )

$$\zeta b_q(\zeta) - \sum_{j=0}^q \binom{q}{j} b_j(\zeta) = 0$$

эквивалентно уравнению

$$\frac{\zeta b_q(\zeta)}{q!} - \sum_{j=0}^q \frac{b_j(\zeta)}{j!} \frac{1}{(q-j)!} = 0,$$

которое является произведением Коши. Поэтому формулу (6.20) можно переписать в виде

$$\zeta \sum_{q \geq 0} \frac{t^q}{q!} b_q(\zeta) - e^t \sum_{q \geq 0} \frac{t^q}{q!} b_q(\zeta) = \log \zeta - t,$$

откуда следует, что

$$\sum_{q \geq 0} \frac{t^q}{q!} b_q(\zeta) = \frac{t - \log \zeta}{e^t - \zeta}.$$

Если в эту формулу подставить  $\zeta = 1$ , то получим равенство

$$\sum_{q \geq 0} \frac{t^q}{q!} b_q(1) = \frac{t}{e^t - 1},$$

которое означает, что  $b_q(1) = B_q$ , где  $B_q$  —  $q$ -е число Бернулли (см. книгу Абрамовича и Стиган, гл. 23).

Таким образом, мы показали, что любому методу Нордсика (6.8) соответствует линейный многошаговый метод, имеющий порядок не ниже  $k$ . Следующая наша задача — установить соответствие в обратном направлении.

**Теорема 6.3.** Пусть  $(\rho, \sigma)$  являются производящими многочленами  $k$ -шагового метода (6.12), имеющего порядок не ниже  $k$ , и предположим, что  $\alpha_k = 1$ . Тогда:

- существует единственный вектор  $l$ , такой что  $\rho$  и  $\sigma$  определяются формулами (6.13);
- если, кроме того, многошаговый метод неприводим, то существует невырожденное преобразование  $T$ , при котором решения систем (6.8') и (6.12) связаны соотношением

$$z_n = T^{-1} u_n, \quad (6.21)$$

где  $j$ -я компонента вектора  $u_n$  вычисляется по формулам

$$u_l^{(n)} = \begin{cases} \sum_{i=0}^j \{\alpha_{k-j+i} y_{n+i} - h \beta_{k-j+i} f_{n+i}\} & \text{при } 0 \leq j \leq k-1, \\ h f_n & \text{при } j = k. \end{cases} \quad (6.22)$$

**Доказательство.** а) Для любого многошагового метода порядка  $k$  многочлен  $\rho(\zeta)$  однозначно определяется по  $\sigma(\zeta)$  (см. теорему 2.4). Разлагая определитель из (6.13b) по первому столбцу, получим

$$\sigma(\zeta) = l_0(\zeta - 1)^k + l_1(\zeta - 1)^{k-1} r_1(\zeta) + \cdots + l_k r_k(\zeta),$$

где  $r_j(\zeta)$  — многочлен степени  $j$ , отвечающий условию  $r_j(1) \neq 0$ . Следовательно,  $l$  можно найти из  $\sigma(\zeta)$ .

б) Пусть  $y_0, \dots, y_{k-1}$  и  $f_0, \dots, f_{k-1}$  заданы. Тогда многочлен  $p_{k-1}(\zeta)$  в (6.15) удовлетворяет равенству

$$p_{k-1}(\zeta) = u_0^{(0)} + u_1^{(0)} \zeta + \cdots + u_{k-1}^{(0)} \zeta^{k-1}.$$

С другой стороны, если для метода Нордсика известен стартовый вектор  $z_0$ , вычисленный при значении  $l$  из утверждения а), то  $p_{k-1}(\zeta)$  определяется формулой (6.17). Приравнивая оба выражения, получим

$$\sum_{j=0}^{k-1} u_j^{(0)} \zeta^j = (\hat{\rho}(\zeta) e_0^T - \hat{\sigma}(\zeta) e_1^T) (I - \zeta P)^{-1} z_0. \quad (6.23)$$

Теперь обозначим через  $t_j^T$  ( $j = 0, \dots, k-1$ ) коэффициенты векторного многочлена

$$(\hat{\rho}(\zeta) e_0^T - \hat{\sigma}(\zeta) e_1^T) (I - \zeta P)^{-1} = \sum_{j=0}^{k-1} t_j^T \zeta^j \quad (6.24)$$

и возьмем  $t_k^T = e_1^T$ . Затем выберем в качестве  $T$  квадратную матрицу, у которой  $j$ -й столбец равен  $t_j^T$ , а поэтому  $u_0 = Tz_0$  вследствие (6.23) и  $hf_n = hy_n'$ . Проводя для  $y_n, \dots, y_{n+k-1}$  и  $f_n, \dots, f_{n+k-1}$  те же рассуждения, что и для  $y_0, \dots, y_{k-1}$ , приедем к равенству  $u_n = Tz_n$  для всех  $n$ .

Для завершения доказательства остается лишь проверить невырожденность  $T$ . Пусть вектор  $v = (v_0, v_1, \dots, v_k)^T$  отличен от нулевого и удовлетворяет равенству  $Tv = 0$ . По определению  $t_k^T$  имеет место равенство  $v_1 = 0$ , и из (6.24) следует (с учетом преобразования (6.16)), что

$$\rho(\zeta) \Phi_0(\zeta) = \sigma(\zeta) \varphi_1(\zeta), \quad (6.25)$$

где  $\tau_i(\zeta) = \det(\zeta I - P) e_i^T (\zeta I - P)^{-1} v$  — многочлены, степень которых не превосходит  $k$ . Более того, из правила Крамера следует, что степень  $\tau_1(\zeta)$  не превосходит  $k - 1$ ,  $v_1 = 0$ . Поэтому в силу (6.25) по крайней мере один из корней  $\rho(\zeta)$  должен быть корнем  $\sigma(\zeta)$ , что противоречит предположению о неприводимости метода.

В таблицах 6.1 и 6.2 даны векторы  $l$ , соответствующие неявным методам Адамса и ФДН-методам. Для этих двух классов методов мы более подробно исследуем их эквивалентность.

Таблица 6.1. Коэффициенты  $l_j$   $k$ -шаговых неявных методов Адамса

	$l_0$	$l_1$	$l_2$	$l_3$	$l_4$	$l_5$	$l_6$
$k = 1$	1/2	1					
$k = 2$	5/12	1	1/2				
$k = 3$	3/8	1	3/4	1/6			
$k = 4$	251/720	1	11/12	1/3	1/24		
$k = 5$	95/288	1	25/24	35/72	5/48	1/120	
$k = 6$	19 087/60 480	1	137/120	5/8	17/96	1/40	1/720

Таблица 6.2. Коэффициенты  $l_j$   $k$ -шаговых ФДН-методов

	$l_0$	$l_1$	$l_2$	$l_3$	$l_4$	$l_5$	$l_6$
$k = 1$	1	1					
$k = 2$	2/3	1	1/3				
$k = 3$	6/11	1	6/11	1/11			
$k = 4$	12/25	1	7/10	1/5	1/50		
$k = 5$	60/137	1	225/274	85/274	15/274	1/274	
$k = 6$	20/49	1	58/63	5/12	25/252	1/84	1/1764

## Неявные методы Адамса

Следующие результаты принадлежат Бирну и Хайндмаршу (1975). Поскольку их «эффективный пакет» EPISODE основан на представлении Нордсика для методов с переменным шагом, мы включили в рассмотрение и этот случай. Методы Адамса естественным образом определяют многочлен, аппроксимирующий неизвестное решение уравнения (6.2), а именно: при заданных  $y_n$  и  $f_n, \dots, f_{n-k+1}$   $k$ -шаговый метод Адамса эквивалентен построению многочлена  $p_{n+1}(x)$  степени  $k+1$ , удовлетворяющего следующим условиям:

$$\begin{aligned} p_{n+1}(x_n) &= y_n, \quad p_{n+1}(x_{n+1}) = y_{n+1}, \\ p'_{n+1}(x_j) &= f_j \text{ при } j = n - k + 1, \dots, n + 1. \end{aligned} \quad (6.26)$$

Условие (6.26) определяет  $y_{n+1}$  неявно. Заметим, что разность между двумя последовательными многочленами  $p_{n+1}(x) - p_n(x)$  обращается в нуль в точке  $x_n$ , а ее производная — в точках  $x_{n-k+1}, \dots, x_n$ . Поэтому, вводя обозначение  $e_{n+1} = y_{n+1} - p_n(x_{n+1})$ , мы можем записать эту разность в виде

$$p_{n+1}(x) - p_n(x) = \Lambda \left( \frac{x - x_{n+1}}{x_{n+1} - x_n} \right) e_{n+1}, \quad (6.27)$$

где  $\Lambda$  — единственный многочлен степени  $k+1$ , определяемый следующими условиями:

$$\begin{aligned} \Lambda(0) &= 1, \quad \Lambda(-1) = 0, \\ \Lambda' \left( \frac{x_j - x_{n+1}}{x_{n+1} - x_n} \right) &= 0 \text{ при } j = n - k + 1, \dots, n. \end{aligned} \quad (6.28)$$

Подставив в производную выражения (6.27) значение  $x = x_{n+1}$ , мы увидим, что при  $h_n = x_{n+1} - x_n$  выполняется соотношение

$$h_n f_{n+1} - h_n p'_n(x_{n+1}) = \Lambda'(0) e_{n+1}.$$

Если ввести вектор Нордсика

$$\tilde{z}_n = \left( p_n(x_n), h_n p'_n(x_n), \dots, \frac{h_n^{k+1}}{(k+1)!} p_n^{(k+1)}(x_n) \right)^T$$

и коэффициенты  $I_j$ ,

$$\Lambda(t) = \sum_{j=0}^{k+1} I_j t^j, \quad (6.29)$$

то (6.27) становится эквивалентным формуле

$$\tilde{z}_{n+1} = P \tilde{z}_n + I I_1^{-1} [h f_{n+1} - e_1^T P \tilde{z}_n], \quad (6.30)$$

в которой  $I = (I_0, I_1, \dots, I_{k+1})^T$ . Это метод вида (6.8'). Однако он имеет размерность  $k+2$ , а не  $k+1$ , как следовало ожидать

на основании теоремы 6.3. Причина этого заключается в следующем: пусть многошаговый метод, соответствующий методу (6.30), имеет производящие многочлены  $\tilde{\rho}(\zeta)$  и  $\tilde{\sigma}(\zeta)$ . Тогда из условий  $\Lambda(-1) = 0$  и  $\Lambda'(-1) = 0$  следуют равенства  $\tilde{\sigma}(0) = \tilde{\rho}(0) = 0$ , означающие, что метод приводим. Тем не менее метод (6.30) применим на практике, так как последнюю компоненту вектора  $\tilde{z}_n$  можно использовать для контроля величины шага.

*Замечание.* При  $k \geq 2$  определенные в (6.29) коэффициенты  $I_j$  зависят от отношений шагов  $h_j/h_{j-1}$ ,  $j = n - k + 2, \dots, n$ . Их можно вычислить из формулы

$$\Lambda(t) = \frac{\int\limits_{-1}^t \prod\limits_{j=1}^k (s - t_j) ds}{\int\limits_{-1}^0 \prod\limits_{j=1}^k (s - t_j) ds}, \quad (6.31)$$

в которой  $t_j = (x_{n-j+1} - x_{n+1})/(x_{n+1} - x_n)$  (см. также упр. 1).

### ФДН-методы

Один шаг  $k$ -шагового ФДН-метода заключается в построении многочлена  $q_{n+1}(x)$  степени  $k$ , удовлетворяющего условиям

$$\begin{aligned} q_{n+1}(x_j) &= y_j \text{ при } j = n - k + 1, \dots, n + 1, \\ q_{n+1}'(x_{n+1}) &= f_{n+1}, \end{aligned} \quad (6.32)$$

и в нахождении значения  $y_{n+1}$ , при котором это возможно. Как и для методов Адамса, в данном случае верно равенство

$$q_{n+1}(x) - q_n(x) = \Lambda\left(\frac{x - x_{n+1}}{x_{n+1} - x_n}\right) \cdot (y_{n+1} - q_n(x_{n+1})), \quad (6.33)$$

где  $\Lambda(t)$  — многочлен степени  $k$ , определяемый условиями

$$\begin{aligned} \Lambda\left(\frac{x_j - x_{n+1}}{x_{n+1} - x_n}\right) &= 0 \text{ при } j = n - k + 1, \dots, n, \\ \Lambda(0) &= 1. \end{aligned}$$

Вводя вектор

$$\tilde{z}_n = \left( q_n(x_n), h_n q'_n(x_n), \dots, \frac{h_n^k}{k!} q_n^{(k)}(x_n) \right)^T$$

и коэффициенты  $I_j$ , задаваемые формулой

$$\Lambda(t) = \sum_{l=0}^k I_l t^l,$$

преобразуем (6.33) к виду

$$\tilde{z}_{n+1} = P\tilde{z}_n + T_1^{-1} [h f_{n+1} - e_1^T P \tilde{z}_n]. \quad (6.34)$$

Вектор  $\boldsymbol{l} = (l_0, l_1, \dots, l_k)^T$  можно вычислить из формулы

$$\Lambda(t) = \prod_{j=1}^k \left(1 + \frac{t}{t_j}\right),$$

в которой  $t_j = (x_{n-j+1} - x_{n+1})/(x_{n+1} - x_n)$ . При постоянном шаге формула (6.34) соответствует формуле из теоремы 6.3, а коэффициенты  $l_j = \tilde{l}_j/l_1$  совпадают с приведенными в табл. 6.2.

### Упражнения

1. Пусть  $l_j^{(k)}$  ( $j = 0, \dots, k$ ) являются коэффициентами Нордсика для  $k$ -шаговых неявных методов Адамса (они определены с помощью теоремы 6.3, а их числовые значения даны в табл. 6.1). Обозначим через  $\tilde{l}_j^{(k)}$  ( $j = 0, \dots, k+1$ ) коэффициенты, введенные с помощью (6.29) и (6.31), в случае постоянного шага. Покажите, что выполняются соотношения

$$\frac{\tilde{l}_j^{(k)}}{l_1^{(k)}} = \begin{cases} l_j^{(k)} & \text{при } j = 0, \\ l_j^{(k+1)} & \text{при } j = 1, \dots, k+1. \end{cases}$$

Используйте их для проверки табл. 6.1.

2. а) Найдите матрицу  $T$  из теоремы 6.3 для 3-шагового неявного метода Адамса.

*Ответ:*

$$T^{-1} = \begin{pmatrix} 1 & 0 & 0 & 3/8 \\ 0 & 0 & 0 & 1 \\ 0 & 6 & 6 & 3/4 \\ 0 & 4 & 12 & 1/6 \end{pmatrix}.$$

Покажите, что вектор Нордсика  $z_n$  имеет вид

$$z_n = (y_n, h f_n, (3h f_n - 4h f_{n-1} + h f_{n-2})/4, (h f_n - 2h f_{n-1} + h f_{n-2})/6)^T.$$

б) Вектор  $\tilde{z}_n$  для двухшагового неявного метода Адамса (6.30) (при постоянном шаге) также задается формулой

$$\tilde{z}_n = (y_n, h f_n, (3h f_n - 4h f_{n-1} + h f_{n-2})/4, (h f_n - 2h f_{n-1} + h f_{n-2})/6)^T,$$

но теперь  $y_n$  менее точно аппроксимирует  $y(x_n)$ .

### III.7. Реализация и численное сравнение

При реализации многошаговых методов существует большая свобода выбора (даже если ограничиться рассмотрением методов Адамса). Можно непосредственно использовать *методы с переменным шагом* из разд. III.5 либо применять метод с постоянным шагом, определяя при этом с помощью интерполяции значения вне узлов сетки, необходимые при изменении шага. Кроме того, возможен выбор между формулировкой в *разделенных разностях* (5.7) и представлением Нордсика (6.30).

Исторически первый подход состоял в использовании формулы (1.8), дополненной интерполяцией (Дж. С. Адамс (1883): «Конечно, мы всегда можем изменить значение  $\omega$  (величину шага), если большая или меньшая скорость убывания последовательных разностей указывает на целесообразность увеличения или уменьшения интервала. Только при этом необходимо, выбрав одно из уже вычисленных значений или применив интерполяцию через них, найти координаты нескольких новых значений  $\varphi$ , разделенных новым интервалом»). С точки зрения теории метод (5.7) с переменным шагом выглядит более подходящим и красивым. Для обоих подходов изменение шага является достаточно дорогой процедурой, тогда как изменить порядок очень просто — надо только прибавить следующий член разложения (1.8). Если же реализуется представление Нордсика (6.30), то мы имеем обратную ситуацию. Теперь изменение порядка не является таким же естественным, как прежде, зато можно легко изменить шаг, умножив вектор Нордсика (6.1) на диагональную матрицу  $\text{diag}(1, \omega, \omega^2, \dots)$ , где  $\omega = h_{\text{new}}/h_{\text{old}}$  — отношение шагов. Это и послужило основной причиной введения такого представления.

#### Выбор шага и порядка

В литературе 60-х годов многое было сделано по вопросу о начале многошаговых вычислений и необходимости для этого методов Рунге—Кутты [см., например, работу Ралстона (1962)]. В настоящее время основанные на многошаговых методах программы в начале счета просто используют методы первого порядка и очень малый шаг и поэтому являются самостотельно стартующими.

Теперь предположим, что до точки  $x_n$  численное интегрирование протекало успешно, а для нахождения  $y_{n+1}$ , аппроксимирующего  $y(x_{n+1})$ , мы выбрали величину шага  $h_n$  и порядок  $k + 1$ . Чтобы решить, подходит ли  $y_{n+1}$ , мы должны иметь оценку локаль-

ной погрешности обрывания ряда. Такую оценку дает, например, формула

$$le_{k+1}(n+1) = y_{n+1} - y_{n+1}^*,$$

где значение  $y_{n+1}^*$  вычислено по неявной формуле Адамса порядка  $k+2$ . Вычитая формулу (5.7) из нее же, но с заменой  $k$  на  $k+1$ , получим

$$le_{k+1}(n+1) = h_n(g_{k+1}(n) - g_k(n)) \Phi_{k+1}(n+1). \quad (7.1)$$

Главный член этого выражения не изменится, если вместо  $\Phi_{k+1}(n+1)$  мы поставим

$$\Phi_{k+1}^p(n+1) = \prod_{i=0}^k (x_{n+1} - x_{n-i}) f^p[x_{n+1}, x_n, \dots, x_k]. \quad (7.2)$$

Верхний индекс  $p$  у  $f$  обозначает замену  $f_{n+1} = f(x_{n+1}, y_{n+1})$  на  $f(x_{n+1}, p_{n+1})$  при образовании разделенных разностей. Если решать неявное уравнение (5.7) итерационным методом с начальным приближением  $p_{n+1}$ , то  $\Phi_{k+1}^p(n+1)$  придется вычислять в любом случае. Поэтому единственной затратой при подсчете оценки

$$LE_{k+1}(n+1) = h_n(g_{k+1}(n) - g_k(n)) \Phi_{k+1}^p(n+1) \quad (7.3)$$

будет нахождение  $g_{k+1}(n)$ . После того как вычислено выражение (7.3), мы требуем выполнения (в некоторой подходящей норме) неравенства

$$\|LE_{k+1}(n+1)\| \leq TOL, \quad (7.4)$$

означающего, что данный шаг был успешным.

Если вместо (5.7) рассматривать представление Нордсика (6.30) то оценка локальной погрешности будет уже не такой простой, так как векторы  $l$  в (6.30) полностью различаются при разных порядках. Одну из возможных оценок погрешности можно найти в статье Бирна—Хиндмарша (1975).

Предположим теперь, что значение  $y_{n+1}$  принято. Далее мы должны выбрать новые шаг и порядок. Идея выбора шага состоит в нахождении наибольшего  $h_{n+1}$ , при котором предсказанная локальная погрешность является допустимой, т. е. выполняется соотношение

$$h_{n+1} \cdot |g_{k+1}(n+1) - g_k(n+1)| \cdot \|\Phi_{k+1}^p(n+2)\| \leq TOL.$$

Однако эта процедура на практике не используется из-за сложной зависимости величин  $g_j(n+1)$  и  $\Phi_{k+1}^p(n+2)$  от шага  $h_{n+1}$ . Кроме того, слишком дорого обходится подсчет коэффициентов  $g_{k+1}(n+1)$  и  $g_k(n+1)$ . Чтобы преодолеть эту трудность, пред-

положим, что сетка равномерная (это необоснованное предположение, но из него следует простая формула для нового шага). Тогда локальная погрешность (для метода порядка  $k + 1$ ) имеет вид  $C(x_{n+2}) h^{k+2} + O(h^{k+3})$ , где  $C$  — гладкая функция от  $x$ . Поэтому локальную погрешность в точке  $x_{n+2}$  можно приблизить ее значением в  $x_{n+1}$ , и так же, как в случае одношаговых методов (см. формулу (4.7) из разд. II.4), мы получим в качестве оптимальной величину шага

$$h(k+1) = \left( \frac{\text{TOL}}{LE_{k+1}(n+1)} \right)^{1/(k+2)}. \quad (7.5)$$

Локальная погрешность  $LE_{k+1}(n+1)$  вычисляется по формуле (7.3) или, снова при предположении о равномерности сетки, по формуле

$$LE_{k+1}(n+1) = h\gamma_{k+1}^* \Phi_{k+1}^p(n+1), \quad (7.6)$$

где значения  $\gamma_{k+1}^*$  приведены в таблице 1.2 (см. упр. 1 из разд. III.5 и упр. 4 из разд. III.1).

Теперь опишем, как можно определить оптимальный порядок. Поскольку при всех порядках необходимо одинаковое количество вычислений функции, для выбора нового порядка в основном применяются две стратегии. Можно выбирать порядок  $k + 1$  из условия либо минимальности оценки локальной погрешности, либо максимальности нового оптимального шага. Эти две стратегии не всегда эквивалентны, так как в формулу (7.5) входит показатель степени  $1/(k+2)$ . Более подробно этот вопрос будет рассмотрен ниже при описании программы DEABM. Необходимо отметить, что в каждой из реализаций методов Адамса — а их существует множество — описанный алгоритм усовершенствован и дополнен еще рядом процедур. Например, шаг сохраняется постоянным, если отношение  $h_{\text{new}}/h_{\text{old}}$  близко к 1, что упрощает вычисления коэффициентов  $g_j(n)$ .

### Некоторые распространенные программы

Для демонстрации стратегий выбора порядка и шага в многошаговых методах мы выбрали три программы: DEABM, EPISODE и LSODE.

DEABM — это модификация программы DE, описанной в книге Шампайна—Гордона (1975). Она датируется 1980 годом, а ее авторы — Шампайн и Уаттс. При численном тестировании мы применяли переработанную версию программы, относящуюся к февралю 1984 г. Европейские пользователи могут ее получить по адресу «Rechenzentrum der RWTH Aachen, Seffenter Weg 23, D — 5100 Aachen, West-Germany».

В этой программе реализовано представление (5.7) в разделенных разностях для формул Адамса с переменным шагом. При решении уравнения (5.7), нелинейного относительно  $y_{n+1}$ , в качестве предиктора ( $P$ ) принимают значение  $p_{n+1}$ , затем вычисляется ( $E$ ),  $f_{n+1}^p = f(x_{n+1}, p_{n+1})$  и выполняется одна итерация корректора ( $C$ ), которая и определяет  $y_{n+1}$ . В конце, если шаг оказался успешным, то вычисляется значение  $f_{n+1} = f(x_{n+1}, y_{n+1})$  для следующего шага. В этой реализации PECE на каждом успешном шаге требуется дважды вычислять значение функции. Теперь кратко опишем стратегию выбора порядка из этой программы: после выполнения шага методом  $k + 1$ -го порядка с помощью нескольких модифицированной формулы (7.6) вычисляются  $LE_{k-1}(n+1)$ ,  $LE_k(n+1)$  и  $LE_{k+1}(n+1)$ . Затем, если верно неравенство

$$\max(\|LE_{k-1}(n+1)\|, \|LE_k(n+1)\|) \leq \|LE_{k+1}(n+1)\|, \quad (7.7)$$

то порядок понижается на единицу. Повышение порядка рассматривается только в том случае, когда принят результат последнего шага, не нарушено соотношение (7.7) и примененный метод соответствовал сетке с постоянным шагом. Тогда вычисляется оценка

$$LE_{k+2}(n+1) = h\gamma_{k+2}^* \Phi_{k+2}(n+1),$$

в которой используется новое значение  $f_{n+1} = f(x_{n+1}, y_{n+1})$ , и, если имеет место неравенство

$$\|LE_{k+2}(n+1)\| < \|LE_{k+1}(n+1)\|,$$

порядок увеличивается на единицу. На рис. 7.1 показано изменение шага и порядка для примера из разд. II.4 (см. рис. 4.1, а также рис. 9.5 из разд. II.9). На графике нанесены решение, шаг и порядок для уровней точности  $10^{-3}$ ,  $10^{-6}$  и  $10^{-9}$ . По поведению выбираемой величины порядка видно сильное отличие от основанной на экстраполяции программы ODEX (разд. II.9). На участках наиболее быстрого изменения решения значительно меняется только шаг, но не порядок. Отметим также, что на больших интервалах шаг остается постоянным, а порядок меняется довольно часто (особенно при  $TOL = 10^{-9}$ ). Это согласуется с замечанием из книги Шампайна—Гордона (1975): «... слабое уменьшение оцениваемой погрешности может вызвать колебания порядка, что в свою очередь помогает программе продолжать счет с постоянным шагом».

EPISODE при значении параметра  $MF = 10$  реализует представление Нордика (6.30) переменного порядка для методов Адамса. Авторами этой программы являются Бирн и Хайндмарш (1975), а для ее получения следует обратиться по следую-

шим адресам: «National Energy Software Center, Build. 221, Argonne Nat. Lab., 9700 South Cass Avenue, Argonne IL 60439, USA» или «NEA Data Bank, B. P. No. 9 (Bat. 45), F91190 GIF-sur-YVETTE, France».

Мы использовали версию от 5 апреля 1977 г. У этой программы есть несколько отличий от DEABM. Нелинейное уравнение, первое в системе (6.30), решается методом простой итерации до сходимости. Пересчет  $f$  в конце не производится. Поэтому такой метод можно рассматривать как  $P$  (EC) $^M$ -метод, где число итераций  $M$

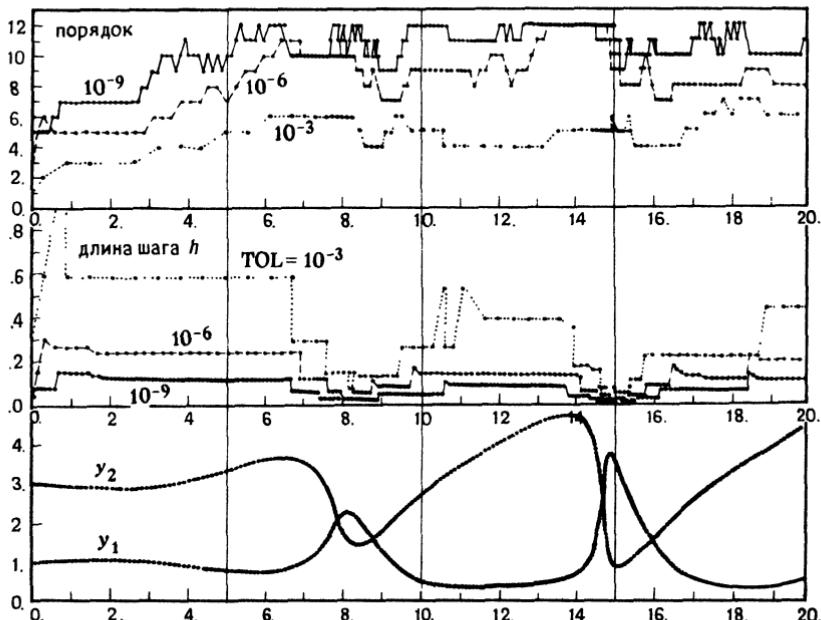


Рис. 7.1. Размер шага и изменение порядка для программы DEABM.

может меняться от шага к шагу. Например, в teste на рис. 7.2 ( $TOL = 10^{-6}$ ) при 340 шагах требуется только 410 вычислений функции. Отсюда видно, что для большинства шагов достаточно одной итерации. Выбор порядка в программе EPISODE основан на поиске максимального шага среди  $h_k(n+1)$ ,  $h_{k+1}(n+1)$  и  $h_{k+2}(n+1)$ . На рис. 7.2 представлено изменение шага и порядка при использовании этой программы для того же примера, что и раньше: заметим, что здесь выбирается намного более низкий порядок, чем в DEABM. Кроме того, порядок остается почти постоянным почти на всем интервале. Такое поведение оправданно, поскольку для представления Нордсика смена порядка не является естественной процедурой.

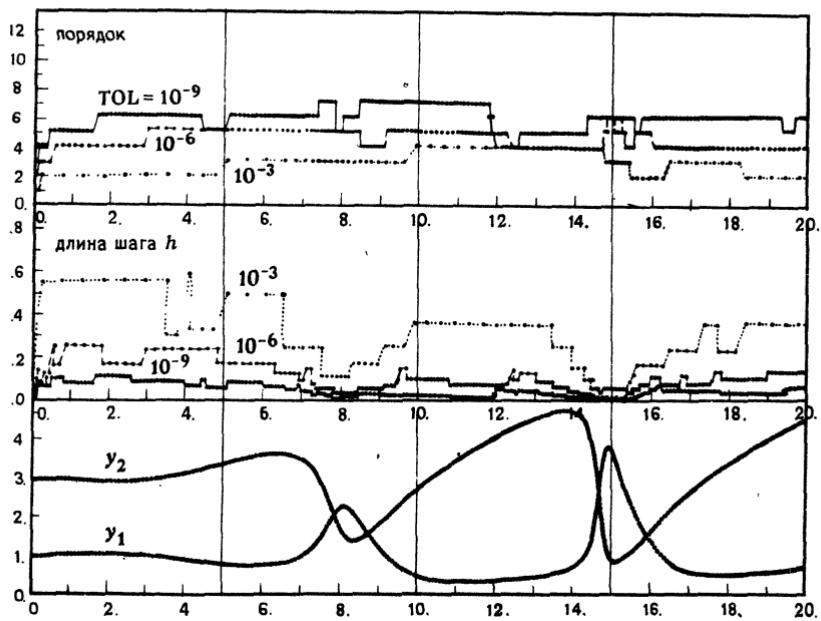


Рис. 7.2. Размер шага и изменение порядка для программы EPISODE.

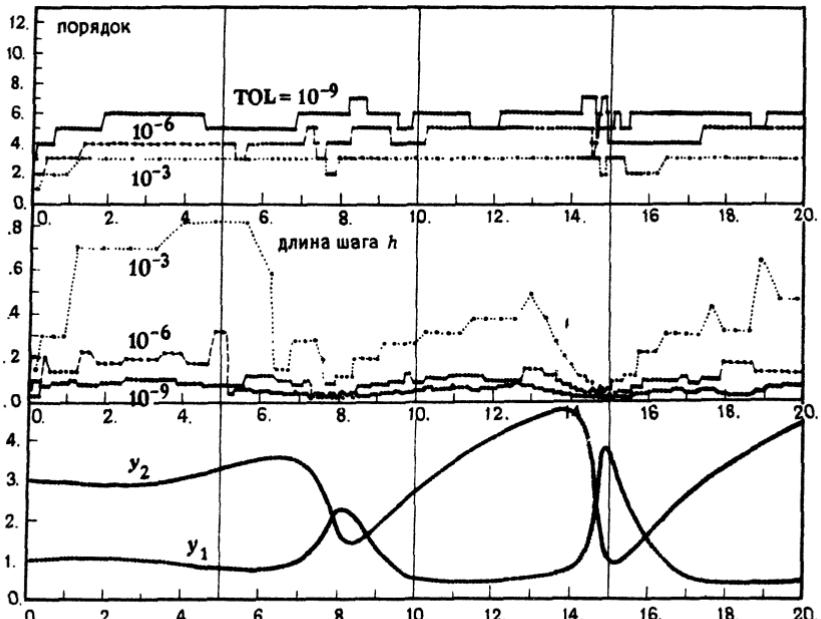


Рис. 7.3. Размер шага и изменение порядка для программы LSODE.

LSODE (с параметром  $MF = 10$ ) — это еще одна реализация методов Адамса. Ее предшественником является программа GEAR (Хайндмарш, 1972), которая в свою очередь представляет модернизированный и улучшенный вариант на основе программы DIFSUB (Гир, 1971). Мы использовали версию от 17 июня 1980 г. Ее можно получить из тех же источников, что и EPISODE. LSODE основана на представлении Нордсика для методов Адамса с постоянным шагом. По алгоритму изменения шага и порядка она близка к программе EPISODE (см. рис. 7.3).

### Сравнение численных результатов

«...Среди трех семейств методов метод Рунге—Кутты постоянного порядка наиболее прост, понятен в ряде аспектов и наименее эффективен».

(Шампайн и Гордон, 1975)

Интересной проблемой является, конечно, сравнение разных реализаций одного и того же численного метода. Помимо программ DEABM, EPISODE и LSODE мы рассмотрим подпрограмму DO2CAF из библиотеки NAG (библиотека NAG, Mark 10), которая представляет собой версию программы Хайндмарша GEAR. Еще мы включили в сравнение результаты счета по программе DOPRI8, основанной на методе Рунге—Кутты фиксированного порядка (разд. II.6), чтобы показать различие в поведении одношаговых и многошаговых методов.

Таблица 7.1. Используемые при сравнении программы

Название программы	Линия на рисунке	Объем памяти при больших $N$
DEABM	----	$22 \cdot N$
EPISODE	- - - - -	$18 \cdot N$
LSODE	- - - - -	$17 \cdot N$
DO2CAF	- - - - -	$19 \cdot N$
DOPRI8	.....	$9 \cdot N$

Все эти методы мы применили к численному решению шести задач из разд. II.10: JACB, TWOB, VDPL, BRUS, LAGR, PLEI (при пяти уровнях точности  $10^{-3}, 10^{-6}, 10^{-9}, 10^{-12}, 10^{-15}$ ). На рис. 7.4 нанесена в двойной логарифмической шкале зависимость числа вызовов функции от достигнутой точности. DEABM кажется лучше других программ (по крайней мере при  $TOL \leqslant 10^{-6}$ ). DOPRI8 и EPISODE требуют наибольшего количества вычислений функции. Авторы программы EPISODE рекомендуют применять ее «для задач, у которых сильно изменяется характер поведения решения». Основанная на методе Рунге—Кутты программа DOPRI8 требует намного меньших дополнительных затрат (вре-

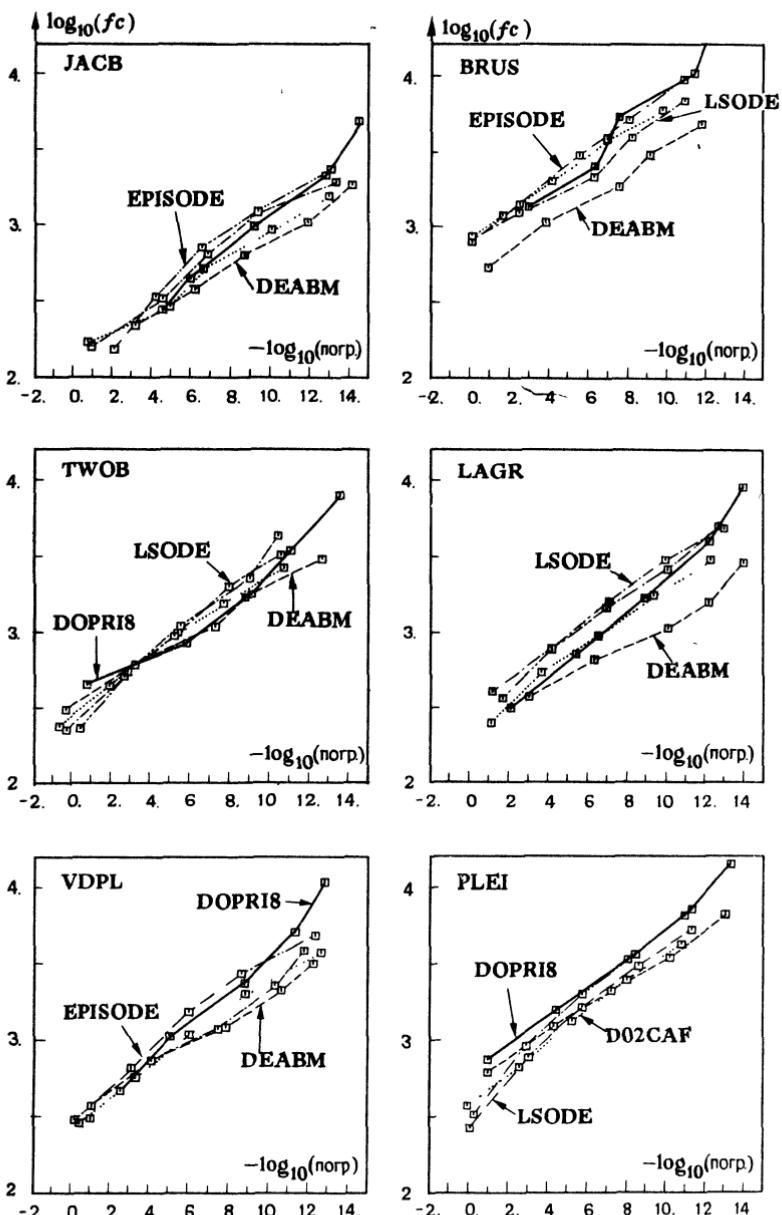


Рис. 7.4. Число обращений к подпрограммам вычисления функций ( $fc$ ) для различных программ при решении для шести задач из раздела 11.10.

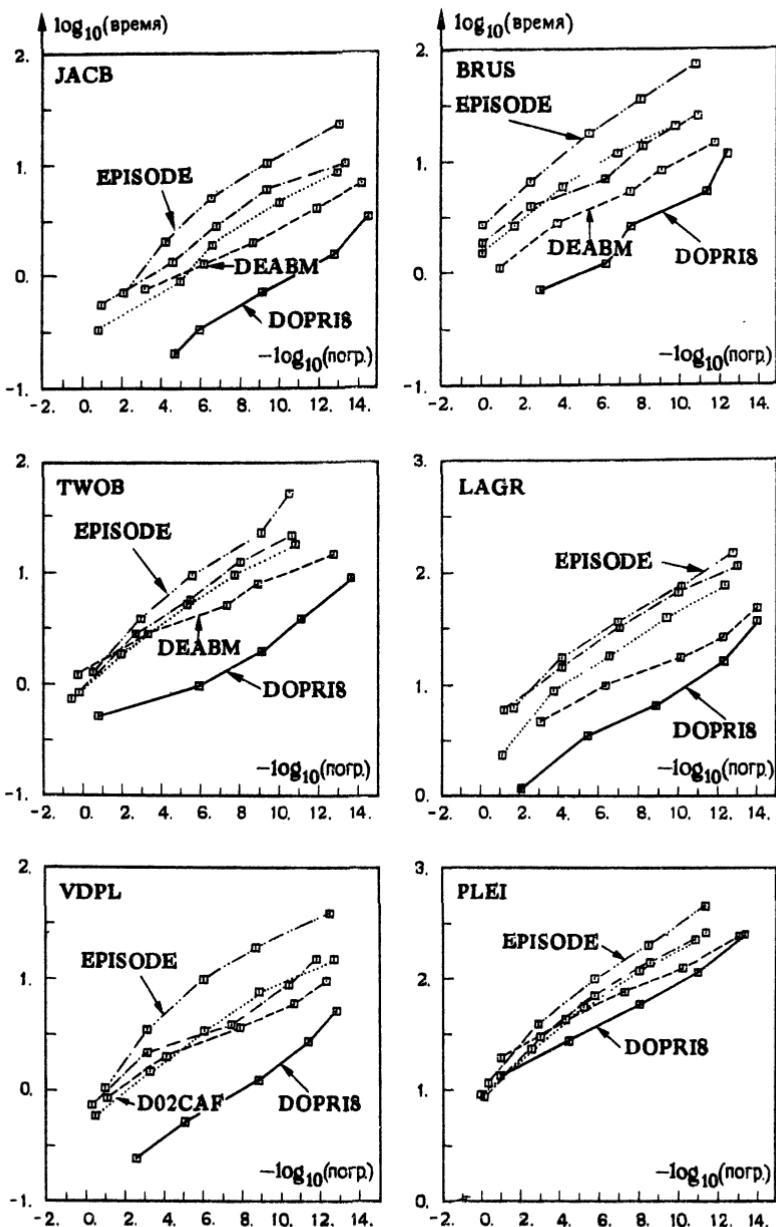


Рис. 7.5. Время работы центрального процессора (в секундах) для различных программ при решении шести задач из раздела 11.10.

мени счета на операции, не входящие в подпрограмму вычисления функции) по сравнению с многошаговыми методами. Это видно из рис. 7.5, на котором для тех же шести задач указана зависимость времени счета (на ЭВМ UNIVAC 1100/60) от достигнутой точности (в двойной логарифмической шкале).

Отметим, что в задачах, где вычисление функции обходится дешево (как в первых пяти примерах), программа с методом Рунге—Кутты затрачивает намного меньше времени, чем основанная на многошаговых методах, хотя в общем она должна повторить эти вычисления большее число раз. Для задачи PLEI, в которой вычисление правой части достаточно дорого, эта разница уже не столь велика.

Теперь обратимся к изучению случая, когда задача имеет очень большую размерность, а отдельные ее компоненты не очень сложны.

### Уравнения в частных производных.

Рассмотрим уравнение реакции с диффузией (брюсселятор с диффузией)

$$\begin{aligned}\frac{\partial u}{\partial t} &= 1 + u^2 v - 4.4u + \alpha \left( \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right), \\ \frac{\partial v}{\partial t} &= 3.4u - u^2 v + \alpha \left( \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} \right)\end{aligned}\quad (7.8)$$

при  $0 \leq x \leq 1$ ,  $0 \leq y \leq 1$ ,  $t \geq 0$ ,  $\alpha = 2 \cdot 10^{-3}$  совместно с граничными условиями Неймана

$$\frac{\partial u}{\partial n} = 0, \quad \frac{\partial v}{\partial n} = 0 \quad (7.9)$$

и начальными условиями

$$\begin{aligned}u(x, y, 0) &= 0.5 + y, \\ v(x, y, 0) &= 1 + 5x.\end{aligned}\quad (7.10)$$

С помощью метода прямых (см. разд. I.6) эту задачу можно привести к системе обыкновенных дифференциальных уравнений. Введем обозначения

$$x_i = \frac{i-1}{N-1}, \quad i = 1, \dots, N, \quad y_j = \frac{j-1}{N-1}, \quad j = 1, \dots, N,$$

и определим функции

$$\begin{aligned}U_{ij}(t) &= u(x_i, y_j, t), \\ V_{ij}(t) &= v(x_i, y_j, t).\end{aligned}\quad (7.11)$$

Дискретизация в (7.8) производных по пространственным переменным дает систему ОДУ размерности  $2N^2$ , состоящую из уравнений

$$\begin{aligned}U'_{ij} &= 1 + U_{ij}^2 V_{ij} - 4.4U_{ij} + \alpha(N-1)^2 (U_{i+1,j} + U_{i-1,j} + \\ &\quad + U_{i,j+1} + U_{i,j-1} - 4U_{ij}),\end{aligned}\quad (7.12)$$

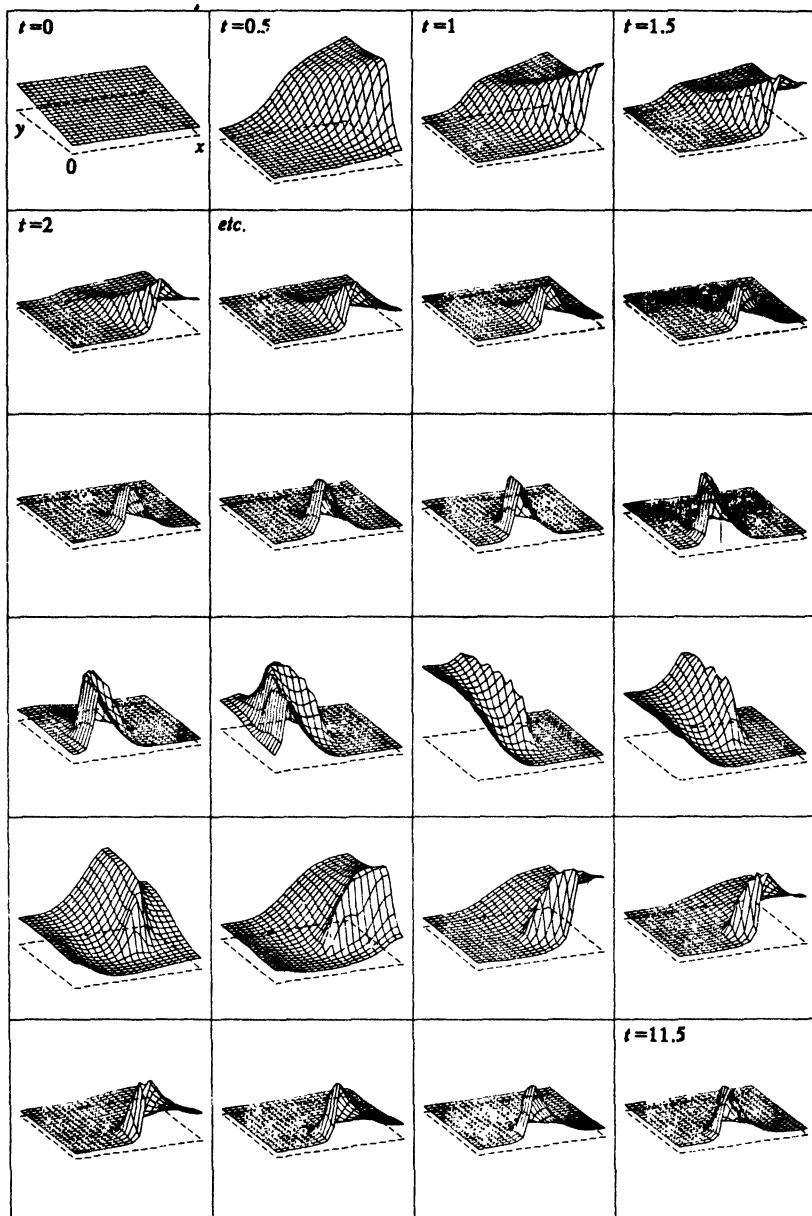
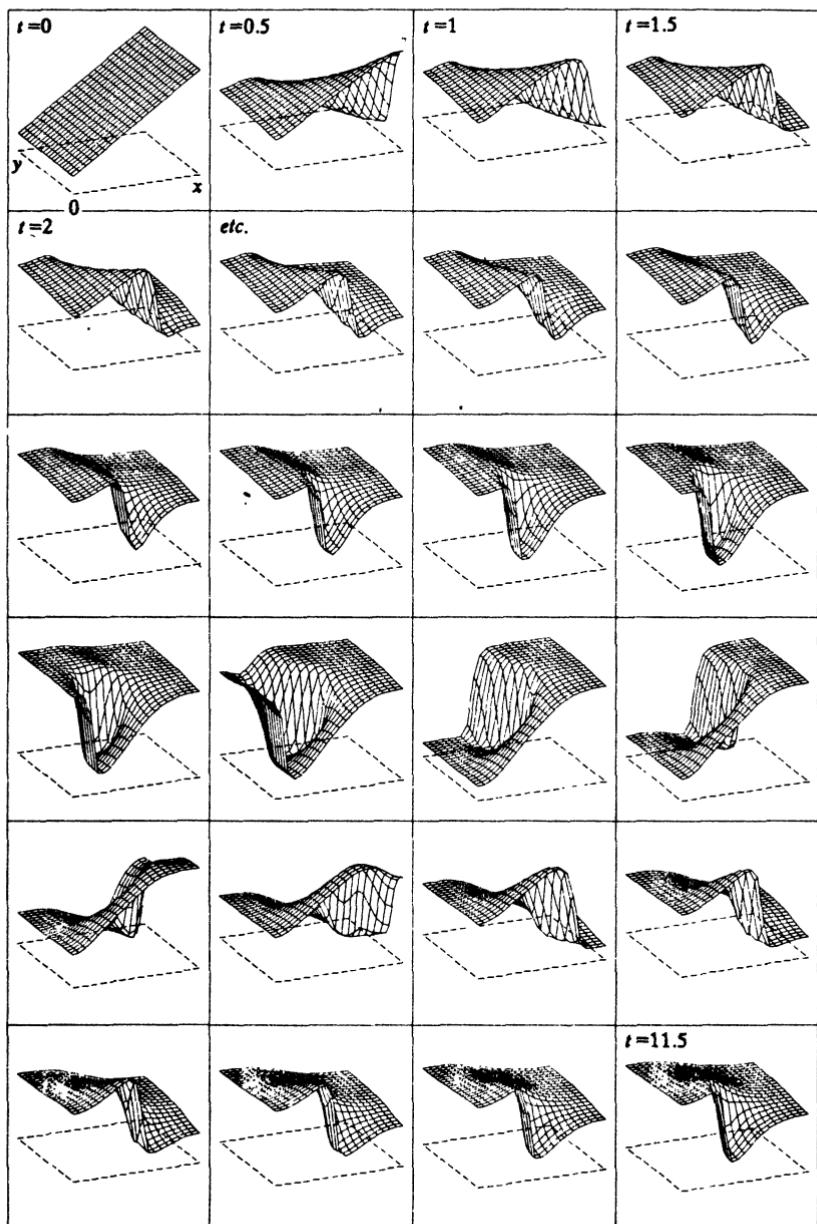


Рис. 7.6. Решение  $u(x, y, t)$  для  $t = 0, 0.5, \dots, 11.5$ .

Рис. 7.7. Решение  $v(x, y, t)$  для  $t = 0, 0.5, \dots, 11.5$ .

$$V_{ij}^* = 3.4U_{ij} - U_{ij}^2 V_{ij} + \alpha(N-1)^2(V_{i+1,j} + V_{i-1,j} + V_{i,j+1} + V_{i,j-1} - 4V_{ij}),$$

где  $i, j = 1, \dots, N$ . В (7.12) еще не определены значения  $U_{0j}$ ,  $U_{N+1,j}$ ,  $U_{i0}$ , ... . В силу граничных условий (7.9) их надо выбирать следующим образом:

$$U_{0j} = U_{2j}, U_{N+1,j} = U_{N-1,j}, U_{i0} = U_{i2}, U_{i,N+1} = U_{i,N-1}.$$

То же относится и к величинам  $V_{ij}$ .

Решение системы (7.12) как функция от  $(x_i, y_j)$  (при  $N = 21$ ) представлено на рис. 7.6 и 7.7 через равные промежутки времени:  $t = 0, 0.5, 1, 1.5, \dots, 11.5$ .

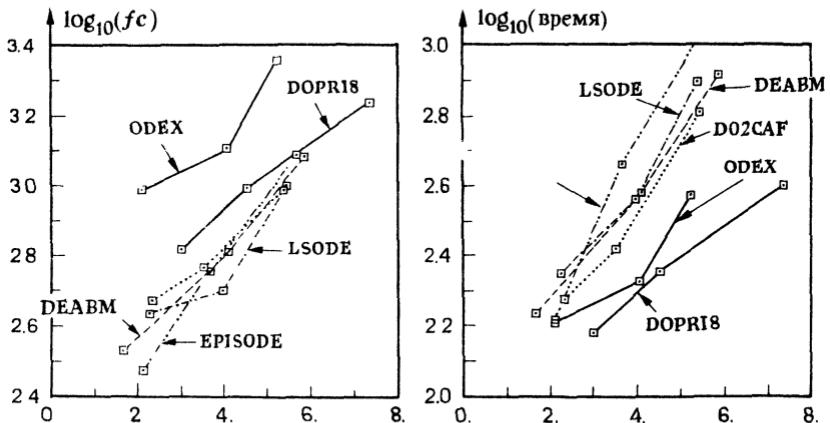


Рис. 7.8. Число вызовов функций  $fc$  и время работы центрального процессора для задачи (7.12) при  $N = 21$  в зависимости от достигнутого уровня точности. По оси абсцисс отложен взятый с обратным знаком десятичный логарифм допустимой погрешности.

Применим теперь различные программы к задаче (7.12) при  $N = 21$ . Ее размерность уже равна 882. Как и при сравнении результатов счета предыдущих задач, на рис. 7.8 показана зависимость количества обращений к вычислению правой части и времени счета от достигнутой точности. Мы использовали уровни точности  $10^{-2}, 10^{-4}, 10^{-6}$  для всех методов, за исключением DEABM, для которого они равны  $10^0, 10^{-2}, 10^{-4}$ . Это необходимо из-за применения в программах разных норм. Из сопоставления двух графиков на рис. 7.8 видно, что основанная на экстраполяционных методах программа ODEX (разд. II.9) вычисляет функцию наибольшее число раз, но при этом требует наименьших дополнительных затрат и сравнима по времени счета с программой DOPR18.

### III.8. Общие линейные методы

«...методы, достаточно общие для того, чтобы включить как частные случаи линейные многошаговые методы и методы Рунге — Кутты...»

(К. Барридж и Дж. С. Бутчер, 1980)

В примечательно короткое время (1964—1966) независимо друг от друга появилось множество работ, где предприняты попытки обобщить либо методы Рунге—Кутты в направлении многошаговых, либо, наоборот, последние в направлении первых. Это вызвано стремлением сделать преимущества многошаговых методов доступными для методов Рунге—Кутты либо «разрушить барьер Далквиста», модифицируя многошаговые формулы. Грэгг и Штеттер (1964) ввели «обобщенные многошаговые методы», Бутчер (1965а) — «модифицированные многошаговые методы», и в том же году появилась работа Гира (1965) по «гибридным методам». Годом позже Бирн и Ламберт (1966) опубликовали работу по «псевдометодам Рунге—Кутты». Все эти методы входят в класс «общих линейных методов», которому и посвящен данный раздел.

Приведем пример такого метода [Бутчер (1965а), порядок равен пяти]:

$$\begin{aligned}y_{n+1/2} &= y_{n-1} + \frac{h}{8} (9f_n + 3f_{n-1}), \\y_{n+1} &= \frac{1}{5} (28y_n - 23y_{n-1}) + \frac{h}{5} (32f_{n+1/2} - 60f_n - 26f_{n-1}), \\y_{n+1} &= \frac{1}{31} (32y_n - y_{n-1}) + \frac{h}{93} (64f_{n+1/2} + 15f_{n+1} + 12f_n - f_{n-1}).\end{aligned}\quad (8.1)$$

Теперь можно сделать выбор, для каких методов развивать теорию — для «обобщенных» многошаговых или «обобщенных» методов Рунге—Кутты. На основании разд. III.4, где показано, что теория сходимости становится красивее, если многошаговые методы рассматривать как одношаговые в пространстве большей размерности, мы выберем вторую возможность: поскольку в формуле (8.1) предыдущая информация представлена значениями  $y_n$ ,  $y_{n-1}$ , введем вектор  $u_n = (y_n, y_{n-1})^T$  и преобразуем последнюю строку в (8.1) к виду

$$\begin{pmatrix} y_{n+1} \\ y_n \end{pmatrix} = \begin{pmatrix} \frac{32}{31} & -\frac{1}{31} \\ 1 & 0 \end{pmatrix} \begin{pmatrix} y_n \\ y_{n-1} \end{pmatrix} + \begin{pmatrix} \frac{64}{93} & \frac{15}{93} & \frac{12}{93} & -\frac{1}{93} \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} hf(y_{n+1/2}) \\ hf(y_{n+1}) \\ f(y_n) \\ f(y_{n-1}) \end{pmatrix},$$

которая теперь будет иметь форму

$$u_{n+1} = Su_n + h\Phi(x_n, u_n, h). \quad (8.2)$$

Свойства таких общих методов исследованы в работах Бутчера (1966), Хайрера—Ваннера (1973), Скила (1976), Купера (1978) и других авторов. Очевидно, что ничто нам не мешает считать  $S$  и  $\Phi$  произвольными или по-другому интерпретировать  $u_n$ .

### Общая процедура интегрирования

Рассмотрим систему

$$y' = f(x, y), \quad y(x_0) = y_0, \quad (8.3)$$

где  $f$  удовлетворяет условию регулярности (4.2). Пусть система дифференциальных уравнений (8.3) имеет размерность  $m$ , а разностное уравнение (8.2) — размерность  $q \geq m$ , и  $x_n = x_0 + nh$  являются точками разбиения равномерной сетки. Рассматриваемый метод состоит из трех частей:

(I) *процедура продвижения на шаг*, т. е. формула (8.2), где квадратная матрица  $S$  не зависит от задачи (8.3);

(II) *функция точных значений*  $z(x, h)$ , которая служит для интерпретации величины  $u_n$ ;  $u_n$  должно быть приближенным значением  $z_n = z(x_n, h)$ , и тогда глобальная погрешность будет вычисляться как разность  $u_n - z_n$ . Предполагается, что по  $z(x, h)$  можно восстановить точное решение  $y(x)$  задачи (8.3);

(III) *стартовая процедура*  $\varphi(h)$ , которая определяет стартовое значение  $u_0 = \varphi(h)$ .  $\varphi(h)$  приближает величину  $z_0 = z(x_0, h)$ .

Таким образом, соответствующая (8.3) дискретная задача описывается формулами

$$u_0 = \varphi(h), \quad (8.4a)$$

$$u_{n+1} = Su_n + h\Phi(x_n, u_n, h), \quad n = 0, 1, 2, \dots, \quad (8.4b)$$

которые дают численное решение  $u_0, u_1, u_2, \dots$ . Отметим, что функция приращения  $\Phi(x, u, h)$ , стартовая процедура  $\varphi(h)$  и функция точных значений  $z(x, h)$  зависят от дифференциального уравнения (8.3), хотя это явно и не указано.

### Примеры метода (8.4)

**Пример 8.1.** Наиболее простой случай представляют *одношаговые методы*. Их характерная особенность состоит в том, что размерности дифференциальной и разностной задач совпадают (т. е.  $m = q$ ), а  $S$  — единичная матрица. Кроме того,  $\varphi(h) = y_0$  и  $z(x, h) = y(x)$ . Эти методы уже изучены в гл. II.

**Пример 8.2.** В разд. III.4 показано, что линейные *многошаговые* методы также относятся к классу (8.4). Для  $k$ -шаговых

методов размерность разностной задачи равна  $q = km$ , а процедура продвижения на шаг задается формулой (4.8). Стартовая процедура определяет вектор  $\varphi(h) = (y_{k-1}, \dots, y_1, y_0)^T$ , и, наконец, функция точных значений имеет вид  $z(x, h) = (y(x + +(k-1)h), \dots, y(x + h), y(x))^T$ .

Наиболее распространенный способ реализации неявного многошагового метода — это процесс *предиктор — корректор* [сравнить с (1.10) и разд. III.7]:  $y_{n+k}^{(0)}$  — приближенное значение для  $y_{n+k}$  — «предсказывают» с помощью явного многошагового метода, скажем,

$$\alpha_k^p y_{n+k}^{(0)} + \alpha_{k-1}^p y_{n+k-1} + \dots + \alpha_0^p y_n = h (\beta_{k-1}^p f_{n+k-1} + \dots + \beta_0^p f_n), \quad (8.5; P)$$

а затем «исправляют» (обычно один или два раза):

$$f_{n+k} := f(x_{n+k}, y_{n+k}^{(l-1)}), \quad (8.5; E)$$

$$\begin{aligned} \alpha_k y_{n+k}^{(l)} + \alpha_{k-1} y_{n+k-1} + \dots + \alpha_0 y_n = \\ = h (\beta_k f_{n+k} + \beta_{k-1} f_{n+k-1} + \dots + \beta_0 f_n). \end{aligned} \quad (8.5; C)$$

Если итерации (8.5) выполняются до сходимости, то этот процесс совпадает с примером 8.2. Однако на практике проводится лишь фиксированное число итераций (пусть их число равно  $M$ ), и теоретически метод уже не является «чисто» многошаговым. Мы различаем два вида методов предиктор—корректор (РС) в зависимости от того, оканчивается ли шаг коррекцией (8.5; С). Первый алгоритм символически обозначается  $P(EC)^M$ , а второй, когда  $f_{n+k}$  дополнительно вычисляется еще один раз для дальнейшего использования на последующих шагах, как  $P(EC)^{ME}$ . Теперь рассмотрим, как эти две процедуры можно выразить в форме методов (8.4).

**Пример 8.2а.**  $P(EC)^M E$ -методы. Стартовая процедура и функция точных значений остаются теми же, что и для многошаговых методов, и так же  $q = km$ . Кроме того,  $S = A \otimes I$ , где матрица  $A$  задана в (4.7), а  $I$  — единичная матрица размера  $m \times m$ . Заметим, что  $S$  зависит только от формулы корректора и не зависит от формулы предиктора. Функция приращения имеет вид

$$\Phi(x, u, h) = (e_1 \otimes I) \psi(x, u, h),$$

где  $e_1 = (1, 0, \dots, 0)^T$ . Для вектора  $u = (u^1, \dots, u^k)^T$ , у которого компоненты  $u^l \in \mathbb{R}^m$ , функция  $\psi(x, u, h)$  определяется следующим образом:

$$\begin{aligned} \psi(x, u, h) = \alpha_k^{-1} (\beta_k f(x + kh, y^{(M)}) + \\ + \beta_{k-1} f(x + (k-1)h, u^1) + \dots + \beta_0 f(x, u^k)), \end{aligned}$$

где значение  $y^{(M)}$  вычисляется из формул

$$\begin{aligned} \alpha_k^p y^{(0)} + \alpha_{k-1}^p u^1 + \cdots + \alpha_0^p u^k &= \\ &= h (\beta_{k-1}^p f(x + (k-1)h, u^1) + \cdots + \beta_0^p f(x, u^k)), \\ \alpha_k y^{(l)} + \alpha_{k-1} u^1 + \cdots + \alpha_0 u^k &= h (\beta_k f(x + kh, y^{(l-1)}) + \\ &\quad + \beta_{k-1} f(x + (k-1)h, u^1) + \cdots + \beta_0 f(x, u^k)) \end{aligned}$$

(при  $l = 1, \dots, M$ ).

**Пример 8.2б.** Для  $P$  ( $EC$ )<sup>M</sup>-методов формулировка в виде метода типа (8.4) усложняется, так как информация, которую следует перенести на следующий шаг, определяется не только значениями  $y_{n-k+1}, \dots, y_n$ , но зависит еще и от  $hf_{n+k-1}, \dots, hf_n$ , где  $hf_{n+j} = hf(x_{n+j}, y_{n+j}^{(M-1)})$ . Таким образом, размерность разностной задачи становится равной  $q = 2km$ . Обычная стартовая процедура (как для многошаговых методов) дает

$$\varphi(h) = (y_{k-1}, \dots, y_0, hf(x_{k-1}, y_{k-1}), \dots, hf(x_0, y_0))^T.$$

Если определить функцию точных значений как

$$\begin{aligned} z(x, h) &= (y(x + (k-1)h), \dots, y(x), \\ hy'(x + (k-1)h), \dots, hy'(x))^T, \end{aligned}$$

то процедура продвижения на шаг будет задана следующим образом:

$$S = \left( \begin{array}{c|ccccc} A & & B \\ \hline 0 & 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ 0 & 0 & 1 & 0 \end{array} \right), \quad \Phi(x, u, h) = \begin{pmatrix} \beta'_k \\ 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \end{pmatrix} \psi(x, u, h).$$

Здесь  $A$  — матрица из (4.7), а  $B$  — матрица размера  $k \times k$ , ранг которой равен единице:

$$(1, 0, \dots, 0) (\beta'_{k-1}, \dots, \beta'_0)^T, \quad \beta'_j = \beta_j / \alpha_k.$$

Для  $u = (u^1, \dots, u^k, hv^1, \dots, hv^k)^T$  функция  $\psi(x, u, h) \in \mathbb{R}^q$  имеет вид

$$\psi(x, u, h) = f(x + kh, y^{(M-1)}),$$

где  $y^{(M-1)}$  определяется из соотношений

$$\begin{aligned} \alpha_k^p y^{(0)} + \alpha_{k-1}^p u^1 + \cdots + \alpha_0^p u^k &= h (\beta_{k-1}^p v^1 + \cdots + \beta_0^p v^k), \\ \alpha_k y^{(l)} + \alpha_{k-1} u^1 + \cdots + \alpha_0 u^k &= h (\beta_k f(x + kh, y^{(l-1)}) + \\ &\quad + \beta_{k-1} v^1 + \cdots + \beta_0 v^k). \end{aligned}$$

Еще раз отметим, что  $S$  зависит только от формулы корректора.

**Пример 8.3.** *Методы Нордсика* также относятся к типу (8.4). Это следует непосредственно из представления (6.8). В этом случае функция точных значений

$$z(x, h) = \left( y(x), hy'(x), \frac{h^2}{2!} y''(x), \dots, \frac{h^k}{k!} y^{(k)}(x) \right)^T$$

состоит не только из значений точного решения, но содержит также и его производные.

**Пример 8.4.** *Циклические многошаговые методы.* Донельсон и Хансен (1971) исследовали возможность построения схемы дискретизации на основе нескольких отличных друг от друга  $k$ -шаговых методов, используемых циклически. Пусть  $S_j$  и  $\Phi_j$  определяют процедуру продвижения на шаг для  $j$ -го многошагового метода; тогда численное решение  $u_0, u_1, \dots$  вычисляется по следующим формулам:

$$u_0 = \varphi(h),$$

$$u_{n+1} = S_j u_n + h \Phi_j(x_n, u_n, h), \text{ если } n \equiv (j-1) \bmod m.$$

Чтобы получить метод (8.4) с не зависящей от номера шага матрицей  $S$ , рассмотрим один цикл как один шаг нового метода

$$\begin{aligned} u_0^* &= \varphi\left(\frac{h^*}{m}\right), \\ u_{n+1}^* &= S u_n^* + h^* \Phi(x_n^*, u_n^*, h^*) \end{aligned} \tag{8.6}$$

с шагом  $h^* = mh$ . Здесь  $x_n^* = x_0 + nh^*$ ,  $S = S_m \dots S_2 S_1$ , а  $\Phi$  выбирается соответственно. Например, в случае  $m = 2$

$$\begin{aligned} \Phi(x^*, u^*, h^*) &= \frac{1}{2} S_2 \Phi_1\left(x^*, u^*, \frac{h^*}{2}\right) + \\ &+ \frac{1}{2} \Phi_2\left(x^* + \frac{h^*}{2}, S_1 u^* + \frac{h^*}{2} \Phi_1\left(x^*, u^*, \frac{h^*}{2}\right), \frac{h^*}{2}\right). \end{aligned}$$

Интересно отметить, что на основе циклически используемых  $k$ -шаговых методов можно получить сходящиеся методы порядка  $2k - 1$  (или  $2k$  при четных  $k$ ). Таким способом можно разрушить «первый барьер Далквиста» (теорема 3.5). Для более подробного ознакомления см. книгу Штеттера (1973) и упр. 2.

**Пример 8.5. Общие линейные методы.**

«Следуя совету Аристотеля...» (оригинал на греческом языке можно найти в работе Бутчера) «...мы ищем наилучшее как среднее между крайностями».

(Дж. С. Бутчер, 1985а)

Эти методы введены в работе Барриджа—Бутчера (1980) и являются настолько общими, что включают все предыдущие примеры как частные случаи, но в то же время в них функция приращения выражается явно через дифференциальное уравнение и несколько свободных параметров. Методы задаются формулами

$$u_i^{(n+1)} = \sum_{j=1}^k a_{ij} u_j^{(n)} + h \sum_{l=1}^s b_{il} f(x_n + c_l h, v_l^{(n)}), \quad i = 1, \dots, k, \quad (8.7a)$$

$$v_l^{(n)} = \sum_{j=1}^k \tilde{a}_{ij} u_j^{(n)} + h \sum_{l=1}^s \tilde{b}_{il} f(x_n + c_l h, v_l^{(n)}), \quad i = 1, \dots, s. \quad (8.7b)$$

Этапы ( $u_i^{(n)}$ ,  $i = 1, \dots, k$ ) называются *внешними этапами*, так как они содержат всю необходимую информацию с предыдущего шага, которая используется при выполнении текущего шага. Этапы ( $v_l^{(n)}$ ,  $i = 1, \dots, s$ ) являются *внутренними* и остаются в «черном ящике» текущего шага. Очень часто некоторые внутренние этапы совпадают с внешними, как, например, в методе (8.1), где

$$v_n = (\dot{y}_{n+1/2}, \dot{y}_{n+1}, y_n, y_{n-1}).$$

Одношаговые методы Рунге—Кутты характеризуются условием  $k = 1$ . В конце данного раздела мы обсудим алгебраические условия, при которых линейные методы имеют порядок  $p$ .

**Пример 8.6.** Покажем, что проведенный в данном разделе анализ применим не только к численным методам дискретизации дифференциальных уравнений первого порядка. Для этого рассмотрим задачу Коши второго порядка

$$y'' = g(x, y), \quad y(x_0) = y_0, \quad y'(x_0) = y'_0, \quad (8.8)$$

которую, очевидно, можно также записать и как систему первого порядка (8.3). Заменив  $y''(x)$  центральной разностной производной, получим метод

$$y_{n+1} - 2y_n + y_{n-1} = h^2 g(x_n, y_n),$$

который при введении дополнительной переменной

$$hy'_n = y_{n+1} - y_n$$

можно представить в виде

$$\begin{pmatrix} y_{n+1} \\ y'_{n+1} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y_n \\ y'_n \end{pmatrix} + h \begin{pmatrix} y'_n \\ g(x_{n+1}, y_n + hy'_n) \end{pmatrix}.$$

Теперь он имеет форму (8.4) в функцией точных значений  $z(x, h) = (y(x), (y(x+h) - y(x))/h)^T$ . Здесь  $y(x)$  обозначает точное решение задачи (8.8).

Все методы Ньюстрёма (разд. II.13), конечно, попадают в этот класс, как и многошаговые методы для дифференциальных уравнений второго порядка. Более подробно они будут изучены в разд. III.10.

**Пример 8.7.** Многошаговые многоэтапные методы со старшими производными представляют, видимо, наиболее широкий класс задаваемых в явном виде линейных методов и являются обобщением методов из разд. II.12. В обозначениях данного раздела их можно записать следующим образом:

$$u_i^{(n+1)} = \sum_{l=1}^k a_{il} u_l^{(n)} + \sum_{r=1}^q \frac{h^r}{r!} \sum_{l=1}^s b_{il}^{(r)} D^r g(x_n + c_j h, v_l^{(n)}), \quad i = 1, \dots, k,$$

$$v_i^{(n)} = \sum_{l=1}^k \tilde{a}_{il} u_l^{(n)} + \sum_{r=1}^q \frac{h^r}{r!} \sum_{l=1}^s \tilde{b}_{il}^{(r)} D^r g(x_n + c_j h, v_l^{(n)}), \quad i = 1, \dots, s.$$

Такие методы изучены в работе Хайрера—Ваннера (1973).

### Устойчивость и порядок

Далее при изучении устойчивости, порядка и сходимости мы в основном будем опираться на работу Скила (1976). Для устойчивости численного метода требуется, чтобы полученное с его помощью решение оставалось ограниченным при  $h \rightarrow 0$ . Этим обосновано следующее определение.

**Определение 8.8.** Метод (8.4) называется *устойчивым*, если матрица  $S^n$  равномерно ограничена при всех  $n \geq 0$ .

Локальная погрешность метода (8.4) определяется точно так же, как для одношаговых (разд. II.3) и многошаговых (разд. III.2) методов.

**Определение 8.9.** Пусть  $z(x, h)$  является функцией точных значений метода (8.4) и  $z_n = z(x_n, h)$ . Тогда *локальная погрешность* имеет вид (см. рис. 8.1)

$$d_0 = z_0 - \varphi(h),$$

$$d_{n+1} = z_{n+1} - S z_n - h \Phi(x_n, z_n, h), \quad n = 0, 1, \dots \quad (8.9)$$

Порядок нельзя определить столь же просто. Требование, чтобы локальная погрешность имела величину  $O(h^{p+1})$  (как

для одношаговых и многошаговых методов), оказывается достаточным, но в общем случае не является необходимым для сходимости порядка  $p$ . Для формулировки соответствующего определения нам потребуется спектральное разложение матрицы  $S$ .

Во-первых, отметим, что если локальная погрешность (8.9) стремится к нулю при  $h \rightarrow 0$  (и фиксированном  $nh = x - x_0$ ), то выполняется равенство

$$0 = z(x, 0) - Sz(x, 0), \quad (8.10)$$

и поэтому 1 является собственным значением матрицы  $S$ , а  $z(x, 0)$  — соответствующим собственным вектором. Кроме того,

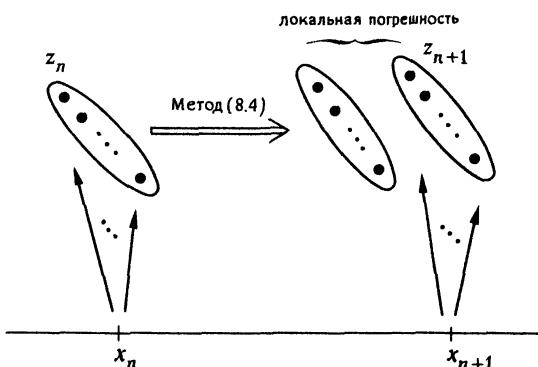


Рис. 8.1. Определение локальной погрешности.

вследствие устойчивости  $S$  не может иметь собственных значений вне единичной окружности, а собственные значения, равные по модулю единице, не могут порождать жордановы цепочки. Обозначим собственные значения, равные по модулю единице, через  $\zeta_1 (=1)$ ,  $\zeta_2, \dots, \zeta_l$ . Тогда жорданова каноническая форма  $S$  (см. (I.12.14)) будет блочно-диагональной матрицей:

$$S = T \operatorname{diag} \left\{ \begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{pmatrix}, \begin{pmatrix} \zeta_2 & & \\ & \ddots & \\ & & \zeta_2 \end{pmatrix}, \dots, \begin{pmatrix} \zeta_l & & \\ & \ddots & \\ & & \zeta_l \end{pmatrix}, \tilde{J} \right\} T^{-1}.$$

Если мы разложим ее на члены, соответствующие единичным собственным значениям, то получим

$$S = E + \zeta_2 E_2 + \dots + \zeta_1 E_1 + \tilde{E}, \quad (8.11)$$

где

$$\begin{aligned} E &= T \operatorname{diag} \{I, 0, 0, \dots\} T^{-1}, \\ E_2 &= T \operatorname{diag} \{0, I, 0, \dots\} T^{-1}, \\ &\dots \\ \tilde{E} &= T \operatorname{diag} \{0, 0, 0, \dots, \tilde{J}\} T^{-1}. \end{aligned} \quad (8.12)$$

Теперь все готово для следующего определения.

**Определение 8.10.** Метод (8.4) имеет *порядок*  $p$  (согласован с порядком  $p$ ), если для всех задач (8.3) с функцией  $f$ , непрерывно дифференцируемой  $p$  раз, локальная погрешность удовлетворяет соотношениям

$$\begin{aligned} d_0 &= O(h^p), \\ E(d_0 + d_1 + \dots + d_n) + d_{n+1} &= O(h^p) \\ \text{при } 0 < nh < \text{const.} \end{aligned} \quad (8.13)$$

*Замечание.* В работе Скила (1976) это свойство названо *квазисогласованностью порядка*  $p$ .

Если правая часть дифференциального уравнения (8.3)  $p$  раз непрерывно дифференцируема, то в общем случае  $\varphi(h)$ ,  $\Phi(x, u, h)$  и  $z(x, h)$  также являются гладкими, а поэтому локальную погрешность (8.9) можно разложить в ряд Тейлора по степеням  $h$ :

$$\begin{aligned} d_0 &= \gamma_0 + \gamma_1 h + \dots + \gamma_{p-1} h^{p-1} + O(h^p), \\ d_{n+1} &= d_0(x_n) + d_1(x_n)h + \dots + d_p(x_n)h^p + O(h^{p+1}). \end{aligned} \quad (8.14)$$

Тогда функции  $d_j(x)$  непрерывно дифференцируемы  $p - j + 1$  раз. Следующая лемма дает более практическое описание порядка методов (8.4).

**Лемма 8.11.** Пусть локальная погрешность метода (8.4) удовлетворяет соотношениям (8.14) с непрерывными функциями  $d_j(x)$ . В этом случае метод имеет порядок  $p$  тогда и только тогда, когда

$$d_n = O(h^p) \text{ при } 0 < nh < \text{const} \text{ и } Ed_p(x) = 0. \quad (8.15)$$

*Доказательство.* Условие (8.15) эквивалентно равенствам

$$d_n = O(h^p), \quad Ed_{n+1} = O(h^{p+1}) \text{ при } 0 < nh < \text{const}, \quad (8.16)$$

которые, очевидно, являются достаточными для порядка  $p$ . Покажем теперь, что они также и необходимы. Поскольку  $E^2 = E$

[см. (8.12)], из условий порядка  $p$  следует, что

$$d_n = O(h^p), \quad E(d_1 + \dots + d_n) = O(h^p) \text{ при } 0 \leq nh \leq \text{const}. \quad (8.17)$$

Лучше всего это видно при умножении (8.13) на  $E$ . Рассмотрим теперь пары  $(n, h)$ , для которых  $nh = x - x_0$  при некотором фиксированном  $x$ . Подставим (8.14) (заметим, что  $d_n = O(h^p)$ ) в выражение  $E(d_1 + \dots + d_n)$  и аппроксимируем полученную сумму с помощью соответствующего интеграла Римана:

$$\begin{aligned} E(d_1 + \dots + d_n) &= h^p E \sum_{l=1}^n d_p(x_{j-1}) + O(h^p) = \\ &= h^{p-1} E \int_{x_0}^x d_p(s) ds + O(h^p). \end{aligned}$$

Из (8.17) следует равенство  $E \int_{x_0}^x d_p(s) ds = 0$ , при дифференцировании которого придем к соотношению  $E d_p(x) = 0$ . ■

### Сходимость

В дополнение к численному решению, заданному формулами (8.4), рассмотрим возмущенное численное решение  $(\hat{u}_n)$ , т. е.

$$\begin{aligned} \hat{u}_0 &= \varphi(h) + r_0, \\ \hat{u}_{n+1} &= S\hat{u}_n + h\Phi(x_n, \hat{u}_n, h) + r_{n+1}, \quad n = 0, 1, \dots, N-1 \end{aligned} \quad (8.18)$$

при некотором возмущении  $R = (r_0, r_1, \dots, r_N)$ . Например, «точное» решение  $z_n = z(x_n, h)$  можно считать возмущенным при возмущении, равном локальной погрешности. Следующая лемма дает наилучшую из возможных качественных оценок разности  $u_n - \hat{u}_n$ , выраженную через возмущение  $R$ . Нам надо предположить, что функция приращения  $\Phi(x, u, h)$  удовлетворяет условию Липшица по  $u$  (в компактной окрестности решения). Это выполняется для всех разумных методов.

**Лемма 8.12.** *Предположим, что метод (8.4) устойчив, а последовательности  $(u_n)$  и  $(\hat{u}_n)$  заданы формулами (8.4) и (8.18) соответственно. Тогда существуют положительные постоянные  $c$  и  $C$ , такие что при любом возмущении  $R$  и  $hN \leq \text{const}$*

$$c \|R\|_s \leq \max_{0 \leq n \leq N} \|u_n - \hat{u}_n\| \leq C \|R\|_s,$$

где

$$\|R\|_s = \max_{0 \leq n \leq N} \left\| \sum_{j=0}^n S^{n-j} r_j \right\|.$$

*Замечание.* Норма  $\|R\|_S$  задана в  $\mathbb{R}^{(N+1)q}$ . Ее положительность объясняется следующим образом: если  $\|R\|_S = 0$ , то при  $n = 0, 1, 2, \dots$  мы получим последовательно  $r_0 = 0, r_1 = 0, \dots$

*Доказательство.* Введем обозначение  $\delta_n = \hat{u}_n - u_n$ . Тогда

$$\delta_{n+1} = S\delta_n + h\tilde{\delta}_n + r_{n+1}, \quad (8.19)$$

где

$$\tilde{\delta}_n = \Phi(x_n, \hat{u}_n, h) - \Phi(x_n, u_n, h).$$

По предположению существует константа  $L$ , такая что  $\|\tilde{\delta}_n\| \leq L\|\delta_n\|$ . Решая разностное уравнение (8.19), придем к равенствам  $\delta_0 = r_0$  и

$$\delta_{n+1} = \sum_{j=0}^n S^{n-j}h\tilde{\delta}_j + \sum_{j=0}^{n+1} S^{n+1-j}r_j. \quad (8.20)$$

В силу устойчивости существует постоянная  $B$ , для которой

$$\|S^n\|L \leq B \quad \text{при всех } n \geq 0. \quad (8.21)$$

Таким образом, из (8.20) следует соотношение

$$\|\delta_{n+1}\| \leq hB \sum_{j=0}^n \|\delta_j\| + \|R\|_S.$$

С помощью индукции по  $n$  получим

$$\|\delta_n\| \leq (1 + hB)^n \|R\|_S \leq \exp(\text{const} \cdot B) \|R\|_S,$$

что и доказывает второе неравенство леммы. Из (8.20) и (8.21)

$$\left\| \sum_{j=0}^n S^{n-j}r_j \right\| \leq (1 + nhB) \max_{0 \leq n \leq N} \|\delta_n\|,$$

а поэтому при  $Nh \leq \text{const}$  выполняется соотношение  $\|R\|_S \leq (1 + \text{const} \cdot B) \max_{0 \leq n \leq N} \|\hat{u}_n - u_n\|$ . ■

*Замечание.* Норму, которая, подобно  $\|\cdot\|_S$ , позволяет оценить сверху и снизу через возмущение величину  $\max \|u_n - \hat{u}_n\|$ , называют *минимальным функционалом устойчивости* для метода. Эти идеи более подробно изложены в работах Спейкера (1971) и Штеттера (1973).

С помощью данной леммы можно доказать следующую теорему.

**Теорема 8.13.** *Рассмотрим устойчивый метод (8.4) и предположим, что его локальная погрешность удовлетворяет соотношению (8.14) с непрерывно дифференцируемой функцией  $d_p(x)$ .*

В этом случае метод сходится с порядком  $p$ , т. е. для его глобальной погрешности  $u_n - z_n$  выполняется равенство

$$u_n - z_n = O(h^p) \text{ при } 0 \leq nh \leq \text{const}$$

тогда и только тогда, когда он согласован с порядком  $p$ .

*Доказательство.* В силу равенства  $ES = E$  [см. (8.11) и (8.12)] выполняется тождество

$$E(d_0 + \dots + d_n) + d_{n+1} = \sum_{j=0}^{n+1} S^{n+1-j} d_j - (S - E) \sum_{j=0}^n S^{n-j} d_j,$$

из которого следует, что при  $n \leq N - 1$  и  $D = (d_0, \dots, d_N)$

$$\|E(d_0 + \dots + d_n) + d_{n+1}\| \leq (1 + \|S - E\|) \|D\|_s. \quad (8.22)$$

Оценка снизу из леммы (8.12), но с заменой  $r_n$  и  $\hat{d}_n$  на  $d_n$  и  $z_n$  соответственно доказывает вторую часть утверждения, т. е. «только тогда».

Для доказательства первой части утверждения, т. е. «тогда», используем оценку сверху из этой же леммы. Нам надо показать, что из согласованности порядка  $p$  следует соотношение

$$\max_{0 \leq n \leq N} \left\| \sum_{j=0}^n S^{n-j} d_j \right\| = O(h^p). \quad (8.23)$$

В силу (8.11) и (8.12) выполняется тождество

$$S^{n-l} = E + \zeta_2^{n-l} E_2 + \dots + \zeta_l^{n-l} E_l + \tilde{E}^{n-l},$$

которое вместе с леммой (8.11) приводит к равенству

$$\begin{aligned} \sum_{j=0}^n S^{n-l} d_j &= h^p E_2 \sum_{j=1}^n \zeta_2^{n-l} d_p(x_{j-1}) + \dots \\ &\quad + h^p E_l \sum_{j=1}^n \zeta_l^{n-l} d_p(x_{j-1}) + \sum_{j=0}^n \tilde{E}^{n-l} d_j + O(h^p). \end{aligned}$$

Последний член этого выражения равен  $O(h^p)$ , так как в соответствующей норме  $\|\tilde{E}\| < 1$ , а поэтому

$$\left\| \sum_{j=0}^n \tilde{E}^{n-l} d_j \right\| \leq \sum_{j=0}^n \|\tilde{E}\|^{n-l} \|d_j\| \leq \frac{1}{1 - \|\tilde{E}\|} \cdot \max_{0 \leq n \leq N} \|d_n\|.$$

Остальное следует из того, что сумма  $\sum_{j=1}^n \zeta^{n-l} d(x_{j-1})$  равномерно ограничена по  $n$ , если  $\|\zeta\| = 1$ ,  $\zeta \neq 1$  и  $d$  имеет ограниченную вариацию. Проще всего это увидеть при суммировании по частям [Абель (1826)]:

$$\sum_{j=1}^n \zeta^{n-l} d(x_{j-1}) = \frac{1 - \zeta^n}{1 - \zeta} d(x_0) + \sum_{j=1}^n \frac{1 - \zeta^{n-l}}{1 - \zeta} (d(x_j) - d(x_{j-1})). \quad \blacksquare$$

## Условия порядка для общих линейных методов

Условия (8.15) не очень удобны для построения общего линейного метода (8.7) порядка  $p$ . Вместо них нам хотелось бы иметь алгебраические условия на свободные параметры, как в случае методов Рунге—Кутты. Мы покажем, как этого можно добиться с помощью теории  $B$ -рядов из разд. II.11 (см. также работу Барриджа и Мосса, 1980). Чтобы избежать тензорных произведений, в дальнейшем будем считать рассматриваемое дифференциальное уравнение скалярным. При этом все результаты верны и для систем. Далее предположим, что дифференциальное уравнение является автономным, а поэтому можно прямо применить теорию из разд. II.11. Это будет обосновано ниже в замечании 8.17.

Теперь предположим, что компоненты функции точных значений  $\mathbf{z}(x, h) = (z_1(x, h), \dots, z_k(x, h))^T$  можно разложить в  $B$ -ряд

$$z_i(x, h) = B(z_i, y(x)),$$

и тогда, введя обозначение  $\mathbf{z}(t) = (z_1(t), \dots, z_k(t))^T$ , запишем

$$\mathbf{z}(x, h) = \mathbf{z}(\emptyset) y(x) + h \mathbf{z}'(t) f(y(x)) + \dots \quad (8.24)$$

Прежде чем вывести условия порядка, отметим, что соотношение (8.7b) имеет смысл, только если  $v_j^{(n)} \rightarrow y(x_n)$  при  $h \rightarrow 0$ . В противном случае нельзя определить  $f(v_j^{(n)})$ . Поскольку  $u_j^{(n)}$  аппроксимирует  $z_j(x_n, h)$ , мы приходим к условию  $\sum \tilde{a}_{ij} \mathbf{z}_j(\emptyset) = 1$ . Совместно с (8.10) оно дает так называемые *условия предсогласованности*

$$A\mathbf{z}(\emptyset) = \mathbf{z}(\emptyset), \quad \tilde{A}\mathbf{z}(\emptyset) = e. \quad (8.25)$$

Здесь  $A$  и  $\tilde{A}$  — матрицы с элементами  $a_{ij}$  и  $\tilde{a}_{ij}$  соответственно, а  $e = (1, \dots, 1)^T$ . Напомним, что локальная погрешность (8.9) для общего линейного метода (8.7) задается формулой

$$d_i^{(n+1)} = z_i(x_n + h, h) - \sum_{j=1}^k a_{ij} z_j(x_n, h) - \sum_{j=1}^s b_{ij} h f(v_j), \quad (8.26a)$$

где

$$v_i = \sum_{j=1}^k \tilde{a}_{ij} z_j(x_n, h) + \sum_{j=1}^s b_{ij} h f(v_j). \quad (8.26b)$$

Чтобы вывести условия порядка, запишем  $v_i$  и  $d_i^{(n+1)}$  в виде  $B$ -рядов:

$$v_i = B(v_i, y(x_n)), \quad d_i^{(n+1)} = B(d_i, y(x_n)).$$

По теореме о композиции  $B$ -рядов и по формуле (11.10) из разд. II.11 получим

$$\begin{aligned} z_i(x_n + h, h) &= B(z_i, y(x_n + h)) = B(z_i, B(p, y(x_n))) = \\ &= B(pz_i, y(x_n)). \end{aligned}$$

Подставив все эти ряды в (8.26) и приравняв коэффициенты, придем к соотношениям

$$\begin{aligned} d_i(t) &= (pz_i)(t) - \sum_{l=1}^k a_{il} z_l(t) - \sum_{l=1}^s b_{il} v'_l(t) \\ v_i(t) &= \sum_{j=1}^k \tilde{a}_{ij} z_j(t) + \sum_{j=1}^s \tilde{b}_{ij} v'_j(t). \end{aligned} \quad (8.27)$$

Применение леммы (8.11) дает следующий результат.

**Теорема 8.14.** Рассмотрим вектор  $d(t) = (d_1(t), \dots, d_r(t))^T$  с компонентами  $d_i(t)$ , заданными формулой (8.27). Общий линейный метод (8.7) имеет порядок  $p$  тогда и только тогда, когда

$$\begin{aligned} d(t) &= 0 \quad \text{при } t \in T, \rho(t) \leq p-1, \\ Ed(t) &= 0 \quad \text{при } t \in T, \rho(t) = p. \end{aligned} \quad (8.28)$$

[Матрица  $E$  определена в (8.12)]. ■

### Следствие 8.15. Условия

$$d(t) = 0 \quad \text{при } t \in T, \rho(t) \leq p \quad (8.29)$$

являются достаточными для того, чтобы общий линейный метод имел порядок  $p$ . ■

**Замечание 8.16.** Выражение  $(pz_i)(t)$  в (8.27) можно найти с помощью формулы (11.8) из разд. II.11. Поскольку  $p(t) = 1$  для всех деревьев  $t$ , имеет место равенство

$$(pz_i)(t) = \sum_{j=1}^{\rho(t)} \binom{\rho(t)}{j} \frac{1}{\alpha(t)} \sum z_i(s_j(t)), \quad (8.30)$$

где вторая сумма взята по всем возможным индексациям помеченного дерева  $t$ . Эта довольно сложная формула значительно упрощается при предположении, что все коэффициенты  $z_i(t)$  функции точных значений зависят только от порядка  $t$ , т. е.

$$z_i(t) = z_i(u), \quad \text{если } \rho(t) = \rho(u). \quad (8.31)$$

В этом случае формула (8.30) принимает вид

$$(pz_i)(t) = \sum_{j=1}^{\rho(t)} \binom{\rho(t)}{j} z_i(\tau^j). \quad (8.32)$$

Здесь  $\tau^l$  представляет некоторое дерево порядка  $l$ , например,

$$\tau^l = [\underbrace{\tau, \dots, \tau}_{l-1}], \quad \tau^1 = \tau, \quad \tau^0 = \emptyset. \quad (8.33)$$

Обычно компоненты  $z(x, h)$  состоят из  $y(x)$ ,  $y(x + jh)$ ,  $hy'(x)$ ,  $h^2y''(x)$ , ..., а в этом случае предположение (8.31) выполняется.

**Замечание 8.17.** Покажем, как надо определить в (8.7) коэффициенты  $c_i$ , чтобы результаты, полученные выше для автономных систем, были верны и в общем случае. Для дифференциального уравнения  $x' = 1$  формула (8.7б) принимает вид

$$v_n = \tilde{A}u_n + h\tilde{B}e.$$

Предполагая, что уравнение  $x' = 1$  интегрируется точно, т. е.  $u_n = z(\emptyset)x_n + hz(\tau)$ , мы получим  $v_n = x_ne + hc$ , где вектор  $c = (c_1, \dots, c_s)^T$  задается формулой

$$c = \tilde{A}z(\tau) + \tilde{B}e. \quad (8.34)$$

При таком определении  $c_i$  численные результаты для задачи  $y' = f(x, y)$  и для расширенной автономной системы дифференциальных уравнений совпадают.

В заключение приведем в табл. 8.1 условия порядка до третьего включительно [при предположении (8.31)]. Заметим, что условия (8.25) всегда считаются выполненными, а  $c$  задано в (8.34). Через  $c^l$  обозначен вектор  $(c_1^l, \dots, c_s^l)^T$ .

### Построение общих линейных методов

Покажем на примере, как можно построить методы низкого порядка: возьмем  $k = s = 2$  и фиксируем функцию точных значений

$$z(x, h) = (y(x), y(x - h))^T.$$

Этот выбор удовлетворяет условиям (8.24) и (8.31) при

$$z(\emptyset) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad z(\tau) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \quad z(\tau^2) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \dots.$$

Поскольку вторая компонента в  $z(x + h, h)$  равна первой в  $z(x, h)$ , вполне естественно искать методы с матрицами

$$A = \begin{pmatrix} a_{11} & a_{12} \\ 1 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} b_{11} & b_{12} \\ 0 & 0 \end{pmatrix}.$$

Кроме того, наложим требование, чтобы матрица  $\tilde{B}$  имела вид

$$\tilde{B} = \begin{pmatrix} 0 & 0 \\ b_{21} & 0 \end{pmatrix}$$

и, следовательно, метод был явным.

Условие предсогласованности (8.25), формула (8.34) и условия порядка из табл. 8.1 дают следующие уравнения:

$$a_{11} + a_{12} = 1, \quad (8.35a)$$

$$\tilde{a}_{11} + \tilde{a}_{12} = 1, \quad \tilde{a}_{21} + \tilde{a}_{22} = 1, \quad (8.35b)$$

$$c_1 = -\tilde{a}_{12}, \quad c_2 = b_{21} - \tilde{a}_{22}, \quad (8.35c)$$

$$-a_{12} + b_{11} + b_{12} = 1, \quad (8.35d)$$

$$a_{12} + 2(b_{11}c_1 + b_{12}c_2) = 1, \quad (8.35e)$$

$$-a_{12} + 3(b_{11}c_1^2 + b_{12}c_2^2) = 1, \quad (8.35f)$$

$$-a_{12} + 3(b_{11}\tilde{a}_{12} + b_{12}(\tilde{a}_{22} + 2b_{21}c_1)) = 1. \quad (8.35g)$$

Таблица 8.1. Условия порядка для общих линейных методов

$t$	$\rho(t)$	Условие порядка
$\tau$	1	$Az(\tau) + Be = z(\tau) + z(\emptyset)$
$\tau^2$	2	$Az(\tau^2) + 2Bc = z(\tau^2) + 2z(\tau) + z(\emptyset)$
$\tau^3$	3	$Az(\tau^3) + 3Bc^2 = z(\tau^3) + 3z(\tau^2) + 3z(\tau) + z(\emptyset)$
$[\tau^2]$	3	$Az(\tau^3) + 3Bv(\tau^2) = z(\tau^3) + 3z(\tau^2) + 3z(\tau) + z(\emptyset),$ где $v(\tau^2) = \tilde{A}z(\tau^2) + 2\tilde{B}e$

Система содержит 9 уравнений и 11 неизвестных. Выбрав в качестве свободных параметров  $c_1$  и  $c_2$ , мы получим решение следующим образом: найдем  $a_{12}$ ,  $b_{11}$  и  $b_{12}$  из линейной системы (8.35d, e, f), затем  $\tilde{a}_{12}$ ,  $\tilde{a}_{22}$  и  $b_{21}$  из (8.35c, g), и, наконец,  $a_{11}$ ,  $\tilde{a}_{11}$  и  $\tilde{a}_{21}$  из (8.35a, b).

Частное решение при  $c_1 = 1/2$ ,  $c_2 = -2/5$  имеет вид

$$A = \begin{pmatrix} 16/11 & -5/11 \\ 1 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 104/99 & -50/99 \\ 0 & 0 \end{pmatrix}, \quad (8.36)$$

$$\tilde{A} = \begin{pmatrix} 3/2 & -1/2 \\ 3/2 & -1/2 \end{pmatrix}, \quad \tilde{B} = \begin{pmatrix} 0 & 0 \\ -9/10 & 0 \end{pmatrix}.$$

Этот устойчивый явный 2-шаговый, 2-этапный метод 3-го порядка принадлежит Бутчера (1984).

Построение методов более высокого порядка быстро становится очень сложным, и в этом случае большую помощь окажет использование упрощающих предположений.

**Теорема 8.18.** Пусть для функции точных значений выполняется условие (8.31). Тогда из упрощающих предположений

$$\tilde{A}z(\tau^j) + j\tilde{B}c^{j-1} = c^j, \quad j = 1, \dots, p-1 \quad (8.37)$$

вместе с соотношениями предсогласованности (8.25) и условиями порядка для «разветвленных деревьев»

$$d(\tau^j) = 0, \quad j = 1, \dots, p,$$

следует, что метод (8.7) имеет порядок  $p$ .

*Доказательство.* Используя формулы (8.27), по индукции получим:

$$v(t) = v(\tau^j) \quad \text{при } \rho(t) = j, \quad j = 1, \dots, p - 1$$

и, следовательно,

$$d(t) = d(\tau^j) \quad \text{при } \rho(t) = j, \quad j = 1, \dots, p. \blacksquare$$

Упрощающие предположения (8.37) допускают интересную интерпретацию: они эквивалентны тому, что внутренние этапы  $v_i^{(n)}$  совпадают до порядка  $p - 1$  включительно с точным решением в точках  $x_n + c_i h$ , т. е.

$$v_i^{(n)} - g(x_n + c_i h) = O(h^p).$$

В случае методов Рунге—Кутты (8.37) сводится к условиям (7.12) из разд. II.7.

Другие примеры общих линейных методов, удовлетворяющих соотношениям (8.37), можно найти в работах Барриджа—Мосса (1980) и Бутчера (1981). Советуем посмотреть также более поздние статьи Барриджа (1985) и Бутчера (1985а).

## Упражнения

1. Рассмотрите композиции (см. пример 8.5):

- а) явного и неявного методов Эйлера;
- б) неявного и явного методов Эйлера.

Каким методам они эквивалентны? Какой порядок имеют эти составные методы?

2. а) Предположим, что каждый из  $m$  многошаговых методов  $(\rho_i, \sigma_i)$ ,  $i = 1, \dots, m$ , имеет порядок  $p$ . Докажите, что порядок соответствующего циклического метода не ниже  $p$ .

б) Постройте устойчивый 2-циклический, 3-шаговый линейный многошаговый метод 5-го порядка: сначала найдите однопараметрическое семейство линейных 3-шаговых методов 5-го порядка (которые обязательно будут неустойчивыми).

*Ответ.*

$$\rho_c(\zeta) = c\zeta^3 + \left(\frac{19}{30} - c\right)\zeta^2 - \left(\frac{8}{30} + c\right)\zeta + \left(c - \frac{11}{30}\right),$$

$$\sigma_c(\zeta) = \left(\frac{1}{9} - \frac{c}{3}\right)\zeta^3 + \left(c + \frac{8}{30}\right)\zeta^2 + \left(\frac{19}{30} - c\right)\zeta + \left(\frac{c}{3} - \frac{1}{90}\right).$$

Затем определите  $c_1$  и  $c_2$ , при которых собственные значения матрицы  $S$  составного метода будут равны 1, 0, 0.

3. Докажите, что композиция двух разных общих линейных методов (с одной и той же функцией точных значений) снова даст общий линейный метод. Вследствие этого циклические методы из примера 8.4 относятся к общим линейным.
4. Предположим, что все собственные значения  $S$  (за исключением  $\zeta_1 = 1$ ) лежат внутри единичной окружности. Тогда минимальный функционал устойчивости равен

$$\|R\|_E = \max_{0 \leq n \leq N} \left\| r_n + E \sum_{j=0}^{n-1} r_j \right\|.$$

5. Проверьте для линейных многошаговых методов, что условия согласованности (2.6) эквивалентны согласованности первого порядка в смысле леммы 8.11.
6. Запишите метод (8.1) в виде общего линейного метода (8.7) и определите его порядок (ответ:  $p = 5$ ).

### III.9. Асимптотическое разложение глобальной погрешности

Асимптотическое разложение глобальной погрешности многошаговых методов изучено в знаменитой диссертации Грэгга (1964). Там приведено очень громоздкое доказательство, которое в измененном варианте можно найти также в книге Штеттера (1973). Скил (1976) высказал предположение о существовании асимптотических разложений для общих линейных методов. Ниже мы дадим доказательство (Хайрер и Любич (1984)), основанное на идеях из разд. II.8.

#### Поучительный пример

Чтобы понять, каким может быть ожидаемое асимптотическое разложение, начнем с примера. Рассмотрим простое дифференциальное уравнение

$$y' = -y, \quad y(0) = 1,$$

выберем постоянный шаг  $h$  и применим 3-шаговую формулу дифференцирования назад (1.22) с одной из следующих трех стартовых процедур:

$$y_0 = 1, \quad y_1 = \exp(-h), \quad y_2 = \exp(-2h) \quad (\text{точные значения}); \quad (9.1a)$$

$$y_0 = 1, \quad y_1 = 1 - h + \frac{h^2}{2} - \frac{h^3}{6}, \quad y_2 = 1 - 2h + 2h^2 - \frac{8h^3}{6}; \quad (9.1b)$$

$$y_0 = 1, \quad y_1 = 1 - h + \frac{h^2}{2}, \quad y_2 = 1 - 2h + 2h^2. \quad (9.1c)$$

Три левых графика на рис. 9.1 (расположенные в том же порядке, что и соответствующие стартовые процедуры) показывают глобальную погрешность, деленную на  $h^3$ , при пяти величинах шага:  $h = 1/5, 1/10, 1/20, 1/40, 1/80$ .

Для первых двух стартовых процедур можно отметить равномерную сходимость к функции  $e_3(x) = xe^{-x}/4$  [см. формулу (2.12)], а поэтому соотношение

$$y_n - y(x_n) = e_3(x_n)h^3 + O(h^4) \quad (9.2)$$

выполняется равномерно при  $0 \leq nh \leq \text{const}$ . В третьем случае имеет место сходимость к  $e_3(x) = (9+x)e^{-x}/4$  (упр. 2), но она уже не является равномерной. Поэтому (9.2) выполняется только для значений  $x_n$ , расположенных на некотором расстоянии от нуля, т. е. при  $0 < \alpha \leq nh \leq \text{const}$ . На трех правых графиках рис. 9.1 нанесены функции

$$(y_n - y(x_n) - e_3(x_n)h^3)/h^4. \quad (9.3)$$

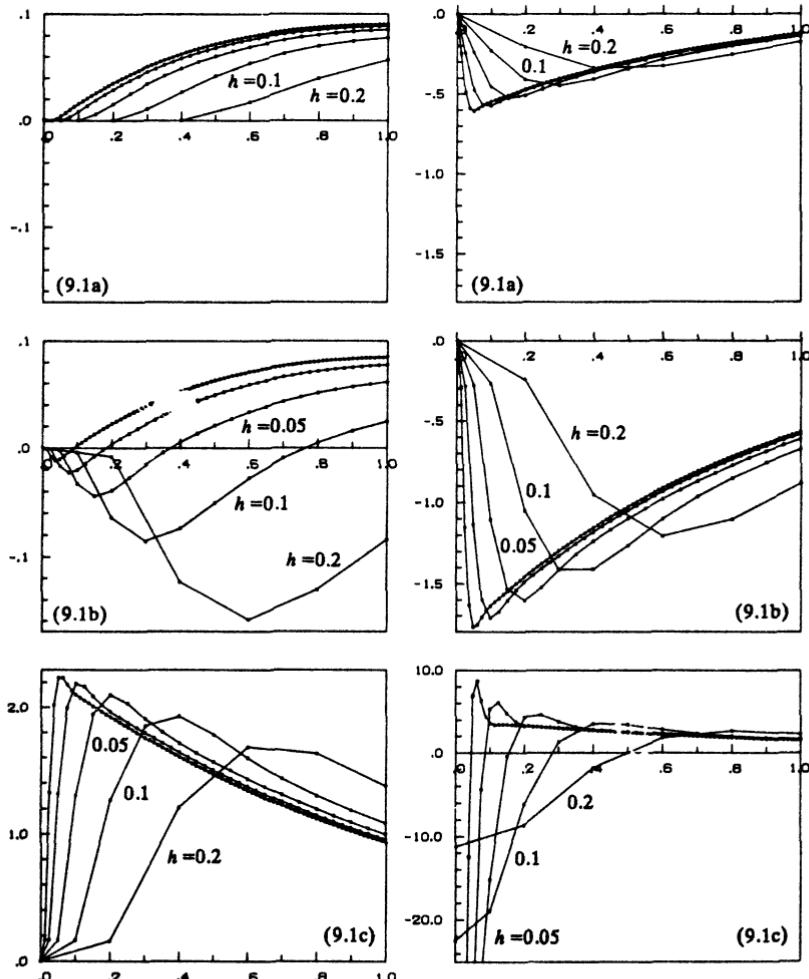


Рис. 9.1. Значения величин  $(y_n - y(x_n))/h^3$  (слева) и  $(y_n - y(x_n) - e_3(x_n))h^3/h^6$  (справа) для трехшагового ФДН-метода и для трех различных стартовых процедур.

Сходимость к функции  $e_4(x)$  можно наблюдать во всех случаях. Поскольку для третьей стартовой процедуры (9.1c)  $e_3(x_0) \neq 0$ , последовательность (9.3) в этом случае, очевидно, расходится как  $O(1/h)$ .

Из этого примера следует вывод о том, что для линейных многошаговых методов в общем случае не существует асимптотического разложения вида

$$y_n = y(x_n) = e_p(x_n)h^p + e_{p+1}(x_n)h^{p+1} + \dots,$$

которое было бы равномерным при  $0 < nh \leq \text{const}$ . В него необходимо еще добавить отвечающие возмущению члены

$$y_n - y(x_n) = (e_p(x_n) + \varepsilon_n^p)h^p + (e_{p+1}(x_n) + \varepsilon_n^{p+1})h^{p+1} + \dots, \quad (9.4)$$

которые будут компенсировать нерегулярность вблизи  $x_0$ . Если возмущения  $\varepsilon_n^l$  затухают экспоненциально (при  $n \rightarrow \infty$ ), то на некотором расстоянии от  $x_0$  они не влияют на асимптотическое разложение в  $x_n$ .

### Асимптотическое разложение для сильно устойчивых методов (8.4)

Для распространения методики из разд. II.8 на многошаговые методы полезно записать их как «одношаговый» метод в пространстве большей размерности [см. (4.8) и пример (8.2)]. Это означает, что мы изучаем асимптотическое разложение сразу для общего метода (8.4). Поскольку в (9.4) входят члены  $\varepsilon_n^l h^l$ , при итерационном доказательстве приведенной далее теоремы 9.1 мы придем к функциям приращения, которые также зависят от  $n$  и имеют вид

$$\Phi_n(x, u, h) = \Phi(x, u + h\alpha_n(h), h) + \beta_n(h). \quad (9.5)$$

Поэтому рассмотрим на равномерной сетке  $(x_n)$  следующую численную процедуру:

$$\begin{aligned} u_0 &= \varphi(h), \\ u_{n+1} &= S u_n + h \Phi_n(x_n, u_n, h), \end{aligned} \quad (9.6)$$

где  $\Phi_n$  задана формулой (9.5), а функция точных значений снова обозначена через  $z(x, h)$ . Изучение асимптотического разложения упрощается при дополнительных предположениях:

A1) Метод (9.6) *сильно устойчив*, т. е. он устойчив (определение 8.8) и 1 является единственным собственным значением  $S$ , равным по модулю единице. В этом случае спектральный радиус матрицы  $S-E$  (см. формулу (8.11)) меньше единицы;

A2)  $\alpha_n(h)$  и  $\beta_n(h)$  — многочлены, коэффициенты которых при  $n \rightarrow \infty$  экспоненциально убывают как  $O(\rho_0^*)$ . Здесь  $\rho_0$  обоз-

значает некоторое число между спектральным радиусом  $S-E$  и единицей, т. е.  $\rho(S-E) < \rho_0 < 1$ ;

А3) функции  $\varphi$ ,  $z$  и  $\Phi$  дифференцируемы достаточное число раз.

Предположение А3 позволяет нам разложить локальную погрешность, заданную формулами (8.9), в ряд Тейлора:

$$\begin{aligned} d_{n+1} &= z(x_n + h, h) - Sz(x_n, h) - h\Phi(x_n, z(x_n, h)) + \\ &\quad + h\alpha_n(h, h) - h\beta_n(h) = \\ &= d_0(x_n) + d_1(x_n)h + \dots + d_{N+1}(x_n)h^{N+1} = \\ &= h^2 \frac{\partial \Phi}{\partial u}(x_n, z(x_n, 0), 0)\alpha_n(h) - \dots - h\beta_n(h) + O(h^{N+1}). \end{aligned}$$

Выражения, содержащие  $\alpha_n(h)$ , можно еще больше упростить. Действительно, для гладкой функции  $G(x)$  выполняется равенство

$$G(x_n)\alpha_n(h) = G(x_0)\alpha_n(h) + hG'(x_0)n\alpha_n(h) + \dots + h^{N+1}R(n, h).$$

Отметим, что  $n!\alpha_n(h)$  снова является многочленом по  $h$ , а его коэффициенты затухают как  $O(\rho^n)$ , где  $\rho$  удовлетворяет условию  $\rho_0 < \rho < 1$ . С помощью тех же рассуждений можно показать ограниченность остаточного члена  $R(n, h)$  при  $0 \leq nh \leq \text{const}$ . Вследствие этого запишем локальную погрешность в виде

$$\begin{aligned} d_0 &= \gamma_0 + \gamma_1 h + \dots + \gamma_N h^N + O(h^{N+1}), \\ d_{N+1} &= (d_0(x_n) + \delta_n^0) + \dots + (d_{N+1}(x_n) + \delta_n^{N+1})h^{N+1} + O(h^{N+2}) \\ &\quad \text{при } 0 \leq nh \leq \text{const}. \end{aligned} \tag{9.7}$$

Функции  $d_j(x)$  являются гладкими, а возмущения  $\delta_n^j$  отвечают условию  $\delta_n^j = O(\rho^n)$ . Разложение (9.7) единствено, так как  $\delta_n^j \rightarrow 0$  при  $n \rightarrow \infty$ .

Метод (9.6) называется согласованным с порядком  $p$ , если локальная погрешность (9.7) удовлетворяет соотношениям (лемма 8.11)

$$d_n = O(h^p) \quad \text{при } 0 \leq nh \leq \text{const} \quad \text{и} \quad Ed_p(x) = 0. \tag{9.8}$$

Отметим, что в силу этого определения возмущения  $\delta_n^j$  должны равняться нулю при  $j = 0, \dots, p-1$ , но на  $\delta_n^p$  не наложено никаких условий. Из экспоненциального убывания этих членов следует, что по-прежнему в согласии с определением 8.10,

$$d_{n+1} + E(d_n + \dots + d_0) = O(h^p) \quad \text{при } 0 \leq nh \leq \text{const}.$$

Теперь можно легко проверить, что для метода (8.6) остаются справедливыми лемма (8.12) ( $\Phi_n$  удовлетворяет условию Липшица с той же константой, что и  $\Phi$ ) и теорема сходимости (8.13).

В следующей теореме мы используем то же обозначение, что и для одношаговых методов, а именно  $u_h(x) = u_n$  при  $x = x_n$ .

**Теорема 9.1.** (Хайрер, Любич, 1984) Пусть для метода (9.6) выполняются предположения A1—A3 и он согласован с порядком  $p$  ( $p \geq 1$ ). Тогда глобальная погрешность имеет асимптотическое разложение вида

$$h_n(x) - z(x, h) = e_p(x) h^p + \dots + e_N(x) h^N + E(x, h) h^{N+1}, \quad (9.9)$$

где величины  $e_j(x)$  даны в доказательстве (см. формулу (9.18)), а  $E(x, h)$  равномерно ограничена по  $h \in [0, h_0]$  при значениях  $x$  из компактных отрезков, не содержащих  $x_0$ . Имеет место более точный, чем (9.9), результат:

$$u_n - z_n = (e_p(x_n) + \varepsilon_n^p) h^p + \dots + (e_N(x_n) + \varepsilon_n^N) h^N + \tilde{E}(n, h) h^{N+1}, \quad (9.10)$$

где  $\varepsilon_n^j = O(\rho^n)$ ,  $\rho(S-E) < \rho < 1$  и  $\tilde{E}(n, h)$  ограничена при  $0 \leq nh \leq \text{const}$ .

**Замечание.** Из (9.10) и (9.9) получается формула

$$E(x_n, h) = \tilde{E}(n, h) + h^{-1} \varepsilon_n^N + h^{-2} \varepsilon_n^{N-1} + \dots + h^{p-N-1} \varepsilon_n^p,$$

а поэтому остаточный член  $E(x, h)$  в общем случае не является равномерно ограниченным по  $h$  при значениях  $x$  из отрезка  $[x_0, \bar{x}]$ . Однако если  $x$  находится на некотором расстоянии от  $x_0$ , например  $x \geq x_0 + \delta$  ( $\delta > 0$  фиксировано), то последовательность  $\varepsilon_n^j$  стремится к нулю быстрее, чем любая степень  $\delta/n \leq h$ .

**Доказательство.** а) Как и в случае одношаговых методов (см. доказательство теоремы 8.1 из гл. II), построим новый метод, который для заданных гладкой функции  $e(x)$  и последовательности  $\varepsilon_n$ , отвечающей условию  $\varepsilon_n = O(\rho^n)$ , имеет численное решение

$$u_n = u_n - (e(x_n) + \varepsilon_n) h^p. \quad (9.11)$$

Такой метод определяется формулами

$$\begin{aligned} \widehat{u}_0 &= \widehat{\varphi}(h), \\ \widehat{u}_{n+1} &= S\widehat{u}_n + h\widehat{\Phi}_n(x_n, \widehat{u}_n, h), \end{aligned} \quad (9.12)$$

где  $\widehat{\varphi}(h) = \varphi(h) - (e(x_0) + \varepsilon_0) h^p$  и

$$\begin{aligned} \widehat{\Phi}_n(x, u, h) &= \Phi_n(x, u + (e(x) + \varepsilon_n) h^p, h) - \\ &- (e(x+h) - Se(x)) h^{p-1} - (\varepsilon_{n+1} - Se_n) h^{p-1}. \end{aligned}$$

Поскольку  $\Phi_n$ , а также  $\widehat{\Phi}_n$  имеют вид (9.5), локальную погрешность можно разложить в ряд (9.7). Определим теперь  $e(x)$  и

$\varepsilon_n$  таким образом, чтобы метод (9.12) был согласован с порядком  $p + 1$ .

б) Локальную погрешность  $d_n$  метода (9.12) можно разложить следующим образом:

$$\begin{aligned} d_0 &= z_0 - \hat{u}_0 = (\gamma_p + e(x_0) + \varepsilon_0) h^p + O(h^{p+1}), \\ \hat{d}_{n+1} &= z_{n+1} - Sz_n - h\hat{\Phi}_n(x_n, z_n, h) = \\ &= d_{n+1} + ((I - S)e(x_n) + (\varepsilon_{n+1} - Se_n)) h^p + \\ &\quad + (-G(x_n)(e(x_n) + \varepsilon_n) + e'(x_n)) h^{p+1} + O(h^{p+2}). \end{aligned}$$

Здесь функция

$$G(x) = \frac{\partial \Phi_n}{\partial u}(x, z(x, 0), 0)$$

в силу (9.5) не зависит от  $n$ . Метод (9.12) согласован с порядком  $p$ , если [см. (9.8)]

- i)  $\varepsilon_0 = -\gamma_0 - e(x_0)$ ,
- ii)  $d_p(x) + (I - S)e(x) + \delta_n^p + \varepsilon_{n+1} - Se_n = 0$  при  $x = x_n$ ,
- iii)  $Ee'(x) = EG(x)e(x) - Ed_{p+1}(x)$ .

В настоящий момент предположим, что систему (i) — (iii) можно решить относительно  $e(x)$  и  $\varepsilon_n$ . В части г) доказательства будет показано, что это действительно так. По теореме сходимости 8.13 метод (9.12) сходится с порядком  $p + 1$ . Следовательно, равенство

$$\hat{u}_n - z_n = O(h^{p+1}) \text{ выполняется равномерно}$$

при  $0 \leq nh \leq \text{const}$ ,

откуда следует утверждение (9.10) при  $N = p$ .

в) Метод (9.12) удовлетворяет предположениям теоремы с заменой  $p$  на  $p + 1$  и  $\rho_0$  на  $\rho$ . Как и в теореме 8.1 (разд. II.8), с помощью индукции мы придем к утверждению теоремы.

г) Остается найти решение системы (i) — (iii). Условие (ii) выполняется, если для некоторой константы  $c$

$$\text{iia)} \quad d_p(x) = (S - I)(e(x) + c),$$

$$\text{iib)} \quad \varepsilon_{n+1} - c = S(\varepsilon_n - c) - \delta_n^p.$$

С помощью соотношения  $(I - S + E)^{-1}(I - S) = I - E$ , которое следует из равенств  $SE = E^2 = E$  [см. (8.11)], формула (iia) приводится к эквивалентному виду

$$(I - S + E)^{-1}d_p(x) = -(I - E)(e(x) + c). \quad (9.13)$$

Из (i) мы получим  $\varepsilon_0 - c = -\gamma_p - (e(x_0) + c)$ , а поэтому в силу (9.13)

$$(I - E)(\varepsilon_0 - c) = -(I - E)\gamma_p + (I - S + E)^{-1}d_p(x_0).$$

Поскольку  $Ed_p(x_0) = 0$ , это соотношение выполняется, в частности при условии

$$\varepsilon_0 - c = -(I - E)\gamma_p + (I - S + E)^{-1}d_p(x_0). \quad (9.14)$$

Числа  $\varepsilon_n - c$  определяются теперь рекуррентно соотношением (iib):

$$\begin{aligned} \varepsilon_n - c &= S^n(\varepsilon_0 - c) - \sum_{l=1}^n S^{n-l}\delta_{l-1}^p = \\ &= E(\varepsilon_0 - c) + (S - E)^n(\varepsilon_0 - c) - E \sum_{l=0}^{\infty} \delta_l^p + E \sum_{l=n}^{\infty} \delta_l^p - \\ &\quad - \sum_{l=1}^n (S - E)^{n-l}\delta_{l-1}^p, \end{aligned}$$

при выводе которого использовано равенство  $S^n = E + (S - E)^n$ . Если выбрать

$$c = E \sum_{l=0}^{\infty} \delta_l^p, \quad (9.15)$$

то указанная выше последовательность  $\{\varepsilon_n\}$  будет отвечать условию  $\varepsilon_n = O(\rho^n)$ , так как в силу (9.14)  $E(\varepsilon_0 - c) = 0$  и  $\delta_n^p = O(\rho^n)$ .

Для нахождения  $e(x)$  определим

$$v(x) = Ee(x).$$

С помощью формул (9.15) и (9.13) мы можем по  $v(x)$  восстановить  $e(x)$ :

$$e(x) = v(x) - (I - S + E)^{-1}d_p(x). \quad (9.16)$$

Теперь перепишем уравнение (iii) как дифференциальное

$$v'(x) = EG(x)(v(x) - (I - S + E)^{-1}d_p(x)) - Ed_{p+1}(x), \quad (9.17)$$

а условие (i) дает для него начальное значение  $v(x_0) = -E(\gamma_p + \varepsilon_0)$ . Такую задачу Коши можно решить относительно  $v(x)$ , а затем с помощью (9.16) получить из него  $e(x)$ . Эта функция и определенное выше  $\varepsilon_n$  являются решением системы (i) — (iii). ■

*Замечания.* а) Из (9.15) — (9.17) следует, что для главного члена погрешности выполняются соотношения

$$e'_p(x) = EG(x)e_p(x) - Ed_{p+1}(x) - (I - S + E)^{-1}d'_p(x) \quad (9.18)$$

$$e_p(x_0) = -E\gamma_p - E \sum_{l=0}^{\infty} \delta_l^p - (I - S + E)^{-1}d_p(x_0).$$

б) Поскольку функция  $e_{p+1}(x)$  — это и есть главный член погрешности для метода (9.12), она удовлетворяет дифференциаль-

ному уравнению (9.18) с заменой  $d_j$  на  $d_{j+1}$ . Поэтому по индукции для  $j \geq p$  мы получаем

$$e'_j(x) = EG(x)e_j(x) + \text{неоднородная функция } (x).$$

### Слабо устойчивые методы

Теперь изучим асимптотическое разложение для устойчивых методов, которые не обладают сильной устойчивостью. К этому типу относится, например, явное правило средней точки (I.13'), к которому мы обратились в связи с GBS-алгоритмом (разд. II.9). Как и в начале данного раздела, применим правило средней точки к задаче  $y' = -y$ ,  $y(0) = 1$  и рассмотрим следующие три стартовые процедуры:

$$y_0 = 1, \quad y_1 = \exp(-h), \quad (9.19a)$$

$$y_0 = 1, \quad y_1 = 1 - h + \frac{h^2}{2}, \quad (9.19b)$$

$$y_0 = 1, \quad y_1 = 1 - h. \quad (9.19c)$$

На рис. 9.2 на трех левых графиках представлена глобальная погрешность, деленная на  $h^2$ . Для первых двух стартовых процедур мы имеем сходимость к функции  $xe^{-x}/6$ , а для (9.19c) отношение погрешности к  $h^2$ , т. е.  $(y_n - y(x_n))/h^2$ , сходится к

$$e^{-x} \left( \frac{2x-3}{12} - \frac{e^x}{4} \right) \quad \text{при четных } n,$$

$$e^{-x} \left( \frac{2x-3}{12} + \frac{e^x}{4} \right) \quad \text{при нечетных } n.$$

Теперь вычтем из глобальной погрешности члены порядка  $h^2$  и разделим остаток на  $h^3$  в случае (9.19a) и на  $h^4$  для b) и c). Результат показан в правой части рис. 9.2.

Приведенный пример прекрасно иллюстрирует тот факт, что асимптотического разложения вида (9.9) или (9.10) больше нет, но оно существует, если рассматривать  $x_n$  только с четными  $n$ , а при нечетных  $n$  оно тоже существует, но будет уже другим (см. также упр. 2 из разд. II.9). Далее мы получим подобный результат для более общих методов.

Будем говорить, что метод вида (8.4) *слабо устойчив*, если он устойчив, но матрица  $S$  кроме  $\zeta_1 = 1$  имеет и другие собственные значения, равные по модулю 1; обозначим их через  $\zeta_2, \dots, \zeta_l$ . Тогда  $S$  можно представить в виде [см. (8.11)]

$$S = \zeta_1 E_1 + \zeta_2 E_2 + \dots + \zeta_l E_l + R, \quad (9.19)$$

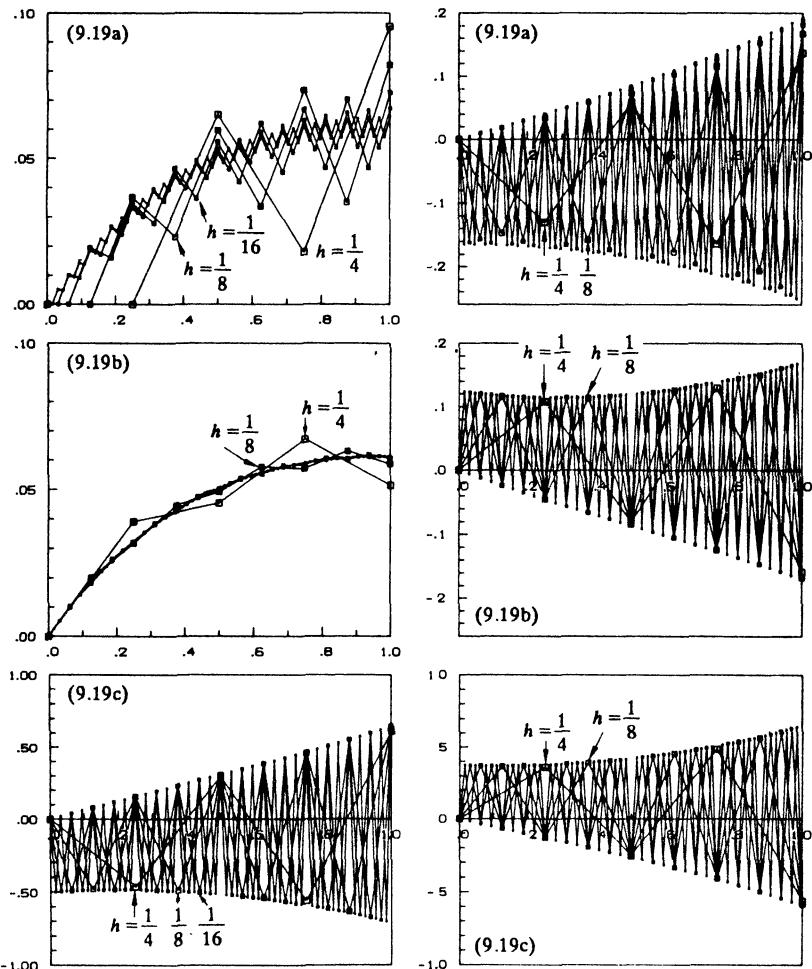


Рис. 9.2. Асимптотическое разложение правила средней (точки три различные стартовые процедуры).

где  $E_j$  — проекторы (соответствующие собственным значениям  $\zeta_j$ ), а спектральный радиус  $R$  удовлетворяет неравенству  $\rho(R) < 1$ .

При последующем изложении мы ограничимся случаем, когда все  $\zeta_j$  ( $j = 1, \dots, l$ ) являются корнями из единицы. Это позволяет просто доказать существование асимптотического разложения и в то же время представляет наиболее важный частный случай. Общая ситуация рассмотрена в работе Хайрера и Любича (1984).

**Теорема 9.2.** Пусть метод (9.6) с не зависящей от  $n$  функцией  $\Phi_n$  является устойчивым, согласованным с порядком  $p$  и удовлетворяет условию А3. Если для всех собственных значений (матрицы  $S$ ), равных по модулю единице, при некотором положительном целом  $q$  выполняется равенство  $\zeta_j^q = 1$  ( $j = 1, \dots, l$ ), то существует асимптотическое разложение вида ( $\omega = e^{2\pi i/q}$ )

$$u_n - z_n = \sum_{s=0}^{q-1} \omega^{ns} (e_{ps}(x_n) h^p + \dots + e_{Ns}(x_n) h^N) + E(n, h) h^{N+1}, \quad (9.20)$$

где  $e_{js}(x)$  — гладкие функции, а  $E(n, h)$  равномерно ограничена при  $0 < \delta \leq nh \leq \text{const}$ .

**Доказательство.** Основная идея доказательства состоит в том, чтобы рассматривать  $q$  последовательных шагов метода (9.6) как один метод с большим шагом. Введем обозначения  $\tilde{u}_n = u_{nq+i}$  ( $0 \leq i \leq q-1$ ,  $i$  фиксировано),  $\tilde{h} = qh$ ,  $\tilde{x}_n = x_i + nh$ . Тогда метод примет вид

$$\tilde{u}_{n+1} = S^q \tilde{u}_n + \tilde{h} \tilde{\Phi}(\tilde{x}_n, \tilde{u}_n, \tilde{h}), \quad (9.21)$$

где функция  $\tilde{\Phi}$  подобрана соответствующим образом. Например, при  $q = 2$

$$\begin{aligned} \tilde{\Phi}(\tilde{x}, \tilde{u}, \tilde{h}) &= \frac{1}{2} S \Phi\left(\tilde{x}, \tilde{u}, \frac{\tilde{h}}{2}\right) + \frac{1}{2} \Phi\left(\tilde{x} + \frac{\tilde{h}}{2}, S\tilde{u} + \right. \\ &\quad \left. + \frac{\tilde{h}}{2} \Phi\left(\tilde{x}, \tilde{u}, \frac{\tilde{h}}{2}\right), \frac{\tilde{h}}{2}\right). \end{aligned}$$

Из предположения о собственных значениях следует равенство

$$S^q = E_1 + \dots + E_l + R^q,$$

из которого видно, что метод (9.21) сильно устойчив. Прямой подсчет дает для локальной погрешности метода (9.21) формулы

$$d_0 = O(h^p),$$

$$d_{n+1} = (I + S + \dots + S^{q-1}) d_p(\tilde{x}_n) h^p + O(h^{p+1}).$$

С помощью подстановки (9.19) и равенства  $\zeta_j^q = 1$  при  $\tilde{E} = E_1 + \dots + E_l$  получим выражение

$$\begin{aligned}\tilde{E}(I + S + \dots + S^{q-1})d_p(x) &= \\ &= \tilde{E} \left( I - \tilde{E} + qE_1 + \sum_{j=2}^k \frac{1 - \zeta_j^q}{1 - \zeta_j} E_j + \sum_{l=1}^{q-1} R^l \right) d_p(x) = qE_1 d_p(x),\end{aligned}$$

которое в силу (8.15) равно нулю. Следовательно, метод (9.21) также согласован с порядком  $p$ . Таким образом, мы проверили, что метод (9.21) отвечает всем предположениям теоремы 9.1. Поэтому получим

$$u_{nq+l} - z_{nq+l} = \tilde{e}_{p,l}(x_{nq+l})h^p + \dots + \tilde{e}_{N,l}(x_{nq+l})h^N + E_l(n, h)h^{N+1},$$

где  $E_l(n, h)$  обладает требуемыми свойствами ограниченности. Если мы определим  $e_{js}(x)$  как решение системы Вандермонда

$$\sum_{s=0}^{q-1} \omega^{ls} e_{js}(x) = \tilde{e}_{j,l}(x),$$

то придем к (9.20). ■

### Сопряженный метод

Для метода (8.4) функция точных значений  $z(x, h)$ , стартовая процедура  $\varphi(h)$  и функция приращения  $\Phi(x, u, h)$  обычно определяются и при отрицательных  $h$  (см. примеры из разд. III.8). Как и в случае одношаговых методов (разд. II.8), мы укажем точный смысл численного решения  $u_h(x)$  при отрицательных  $h$ . Это естественным образом приводит нас к изучению асимптотических разложений по четным степеням  $h$ .

Введем обозначение  $u_h(x) = u_n$  при  $x = x_0 + nh$  ( $h > 0$ ). Тогда метод (8.4) примет вид

$$\begin{aligned}u_h(x_0) &= \varphi(h), \\ u_h(x+h) &= Su_h(x) + h\Phi(x, u_h(x), h) \quad \text{при } x = x_0 + nh.\end{aligned}\tag{9.22}$$

Сначала в (9.22) заменим  $h$  на  $-h$ :

$$\begin{aligned}u_{-h}(x_0) &= \varphi(-h), \\ u_{-h}(x-h) &= Su_{-h}(x) - h\Phi(x, u_{-h}(x), -h),\end{aligned}$$

а затем  $x$  на  $x+h$ , что даст нам формулы

$$\begin{aligned}u_{-h}(x_0) &= \varphi(-h), \\ u_{-h}(x) &= Su_{-h}(x+h) - h\Phi(x+h, u_{-h}(x+h), -h).\end{aligned}$$

При достаточно малых  $h$  это уравнение можно решить относительно  $u_{-h}(x+h)$  (теорема о неявной функции). Тогда мы получим

$$\begin{aligned} u_{-h}(x_0) &= \varphi(-h), \\ u_{-h}(x+h) &= S^{-1}u_{-h}(x) + h\Phi^*(x, u_{-h}(x), h). \end{aligned} \quad (9.23)$$

Метод (9.23), который снова относится к типу (8.4), называется *сопряженным методом* для (9.22). Его функция точных значений равна  $z^*(x, h) = z(x, -h)$ . Отметим, что при заданных  $S$  и  $\Phi$  новая функция приращения  $\Phi^*$  определяется парой формул

$$\begin{aligned} v &= Su - h\Phi(x+h, u, -h), \\ u &= S^{-1}v + h\Phi^*(x, v, h). \end{aligned} \quad (9.24)$$

**Пример 9.3.** Рассмотрим линейный многошаговый метод с производящими функциями

$$\rho(\zeta) = \sum_{j=0}^k \alpha_j \zeta^j, \quad \sigma(\zeta) = \sum_{j=0}^k \beta_j \zeta^j.$$

Тогда

$$S = \begin{pmatrix} -\alpha_{k-1}/\alpha_k & -\alpha_{k-2}/\alpha_k & \dots & -\alpha_0/\alpha_k \\ 1 & 0 & \dots & 0 \\ & 1 & \ddots & 0 \\ & & \ddots & \ddots \\ & & & 1 & 0 \end{pmatrix},$$

$$\Phi(x, u, h) = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \psi(x, u, h),$$

где  $\psi = \psi(x, u, h)$  — решение уравнения ( $u = (u_{k-1}, \dots, u_0)^T$ ):

$$\alpha_k \psi = \sum_{j=0}^{k-1} \beta_j f(x+jh, u_j) + \beta_k f\left(x+kh, h\psi - \sum_{j=0}^{k-1} \frac{\alpha_j}{\alpha_k} u_j\right).$$

Прямо из формулы (9.24) получается, что

$$S^{-1} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \\ \vdots & \vdots \\ \vdots & \ddots & \dots & 1 \\ -\alpha_k/\alpha_0 & -\alpha_{k-1}/\alpha_0 & \dots & -\alpha_1/\alpha_0 \end{pmatrix},$$

$$\Phi^*(x, v, h) = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \psi^*(x, v, h),$$

где  $\psi^* = \psi^*(x, v, h)$  (при  $v = (v_0, \dots, v_{k-1})^T$ ) задается следующим образом:

$$\begin{aligned} -\alpha_0 \psi^* &= \sum_{j=0}^{k-1} \beta_{k-j} f(x + (j - k + 1)h, v_j) + \\ &+ \beta_0 f\left(x + h, h\psi^* - \sum_{j=0}^{k-1} \frac{\alpha_{k-j}}{\alpha_0} v_j\right). \end{aligned}$$

Отсюда видно, что сопряженный метод снова является линейным многошаговым. Его производящие многочлены равны

$$\rho^*(\zeta) = -\zeta^k \rho\left(\frac{1}{\zeta}\right), \quad \sigma^*(\zeta) = \zeta^k \sigma\left(\frac{1}{\zeta}\right). \quad (9.25)$$

Нашей следующей целью будет доказательство того факта, что сопряженный метод имеет точно такое же асимптотическое разложение, что и исходный метод, но с заменой  $h$  на  $-h$ . Для этого необходимо, чтобы матрица  $S^{-1}$  также была устойчивой. Поэтому все собственные значения  $S$  должны лежать на единичной окружности.

**Теорема 9.4.** Пусть метод (9.22) является устойчивым, согласованным с порядком  $p$ , и предположим, что для всех собственных значений  $S$  при некотором целом положительном  $q$  выполняется равенство  $\zeta_j^q = 1$ . Тогда глобальная погрешность имеет асимптотическое разложение вида ( $\omega = e^{2\pi i/q}$ )

$$\begin{aligned} u_h(x_n) - z(x_n, h) &= \sum_{s=0}^{q-1} \omega^{ns} (e_{ps}(x_n) h^p + \dots + e_{Ns}(x_n) h^N) + \\ &+ E(x_n, h) h^{N+1}, \end{aligned} \quad (9.26)$$

которое справедливо и для положительных, и для отрицательных  $h$ . Остаточный член  $E(x, h)$  равномерно ограничен при  $|h| \leq h_0$  и  $x_0 \leq x \leq \hat{x}$ .

**Доказательство.** Как и в доказательстве теоремы 9.2, рассмотрим  $q$  последовательных шагов метода (9.22) в качестве одного нового метода. Из предположения о собственных значениях следует, что  $S^q = I$  — единичной матрице. Поэтому новый метод по существу является одношаговым. Единственное отличие состоит в том, что его стартовая процедура и функция точных значений могут зависеть от  $h$ . Простое распространение на этот случай теоремы 8.5 из гл. II (упр. 3) дает существование разложения

$$\begin{aligned} u_h(x_{nq+l}) - z(x_{nq+l}, h) &= \tilde{e}_{pl}(x_{nq+l}) h^p + \dots \\ &+ \tilde{e}_{Nl}(x_{nq+l}) h^N + E_l(x_{nq+l}, h) h^{N+1}. \end{aligned}$$

Оно справедливо и для положительных, и для отрицательных  $h$ , а остаточный член  $E_i(x, h)$  ограничен при  $|h| \leq h_0$  и  $x_0 \leq x \leq \hat{x}$ . Те же рассуждения, что и при доказательстве теоремы 9.2, приводят нас к требуемому разложению. ■

### Симметричные методы

Для общих линейных методов определить симметрию не так просто, как в случае одношаговых методов. В примере 9.3 мы увидели, что компоненты численного решения сопряженного метода имеют обратный порядок. Поэтому требования  $\Phi(h) = \Phi(-h)$ ,  $S = S^{-1}$  и  $\Phi = \Phi^*$  являются слишком ограничительными.

Однако во многих методах, представляющих практический интерес, функция точных значений удовлетворяет *отношению симметрии* вида

$$z(x, h) = Qz(x + qh, -h) \quad (9.27)$$

где  $Q$  — квадратная матрица, а  $q$  — целое число. Оно выполняется, например, в случае линейных многошаговых методов, у которых функция точных значений равна

$$z(x, h) = (y(x + (k-1)h), \dots, y(x))^T;$$

такие методы отвечают условию (9.27) при

$$Q = \begin{pmatrix} & & 1 \\ & \ddots & \\ 1 & \ddots & \end{pmatrix} \quad \text{и} \quad q = k-1. \quad (9.28)$$

**Определение 9.5.** Пусть функция точных значений удовлетворяет условию (9.27). Метод (9.22) называется *симметричным* [относительно (9.27)], если для численного решения выполняется условие, аналогичное (9.27):

$$u_h(x) = Qu_{-h}(x + qh). \quad (9.29)$$

**Пример 9.6.** Рассмотрим линейный многошаговый метод и предположим, что для производящих многочленов сопряженного метода (9.25) справедливо равенство

$$\rho^*(\zeta) = \rho(\zeta), \quad \sigma^*(\zeta) = \sigma(\zeta). \quad (9.30)$$

Это эквивалентно требованию [см. (3.24)]

$$\alpha_{k-j} = -\alpha_j, \quad \beta_{k-j} = \beta_j.$$

Прямой подсчет (с использованием формул из примера 9.3) показывает, что если отношение симметрии (9.29) выполняется при

$x = x_0$ , то оно справедливо и для всех  $x = x_0 + nh$ . Это налагает дополнительное условие на стартовую процедуру  $\varphi(h)$ .

В завершение продемонстрируем, как можно использовать теорему 9.4 для доказательства асимптотических разложений по четным степеням  $h$ . Обозначим через  $u_h^j(x)$   $j$ -ю компоненту  $u_h(x)$ . Из отношения симметрии (9.29) для многошаговых методов следует

$$u_{-h}^k(x) = u_h^1(x - (k - 1)h).$$

Кроме того, для любого многошагового метода

$$u_h^k(x) = u_h^1(x - (k - 1)h),$$

а поэтому для симметричных методов выполняется равенство

$$u_h^k(x) = u_{-h}^k(x).$$

Если многошаговый метод симметричен в смысле определения 9.5, то, как следует из теоремы 9.4, в асимптотическое разложение глобальной погрешности входят только четные степени  $h$ .

## Упражнения

1. Рассмотрим сильно устойчивый линейный многошаговый метод порядка  $p$ , записанный в форме (9.6) (см. пример 9.3), и введем обозначение

$$G(x) = \frac{\partial \Phi}{\partial u}(x, z(x, 0), 0).$$

- a) Докажите равенство

$$EG(x) \mathbf{1} = \mathbf{1} \frac{\partial f}{\partial y}(x, y(x)),$$

где матрица  $E$  взята из формулы (8.11), а  $\mathbf{1} = (1, \dots, 1)^T$ .

b) Покажите, что в разложении (9.9) функция  $e_p(x) = \mathbf{1} \hat{e}_p(x)$ , где

$$\hat{e}'_p(x) = \frac{\partial f}{\partial y}(x, y(x)) \hat{e}_p(x) - Cy^{(p+1)}(x),$$

а  $C$  — константа погрешности [см. (2.13)]. Определите также  $\hat{e}_p(x_0)$ .

2. Найдите функцию  $e_3(x)$  и возмущения  $(\varepsilon_n^3)_{n>0}$  в разложении (9.4) при использовании 3-шагового ФДН-метода для решения задачи  $y' = -y$ ,  $y(0) = 1$  со стартовой процедурой (9.1 с). Сравните ваш результат с рис. 9.1.

3. Рассмотрим метод

$$\begin{aligned} u_0 &= \varphi(h), \\ u_{n+1} &= u_n + h\Phi(x_n, z_n, h) \end{aligned} \tag{9.31}$$

с функцией точных значений  $z(x, h)$ .

a) Докажите, что глобальная погрешность имеет асимптотическое разложение вида

$$u_n - z_n = e_p(x_n)h^p + \dots + e_N(x_n)h^N + E(x_n, h)h^{N+1},$$

где  $E(x, h)$  равномерно ограничена при  $0 \leq h \leq h_0$  и  $x_0 \leq x \leq \hat{x}$ .

b) Покажите, что для метода (9.31) остается справедливой теорема 8.5 из гл. II.

4. Пусть многочлены  $\rho^*(\xi)$  и  $\sigma^*(\xi)$  заданы с помощью (9.25). При предположении

$$\rho^*(\xi) = C\rho(\xi), \quad \sigma^*(\xi) = C\sigma(\xi)$$

отношение симметрии (9.29) выполняется для всех  $x = x_0 + nh$ , если оно верно при  $x = x_0$ . По какой причине в (9.30)  $C = 1$ ?

### III.10. Многошаговые методы для дифференциальных уравнений второго порядка

«В 1904 г. мне потребовался подобный метод для вычисления траекторий заряженных частиц в магнитном поле; перепробовав различные уже известные методы и не найдя среди них удобного для моей задачи, я сам разработал достаточно простой метод, которым сразу и воспользовался».  
 (С. Штёрмер, 1921)

Дифференциальные уравнения второго порядка заслуживают дополнительного внимания вследствие своей значимости. В разд. II.13 мы уже видели, что для отдельных дифференциальных уравнений второго порядка прямое применение некоторых одношаговых методов более эффективно, чем использование классических методов Рунге—Кутты. Исследуем теперь, сохраняется ли эта ситуация и в случае многошаговых методов.

Рассмотрим дифференциальное уравнение второго порядка

$$y'' = f(x, y, y'), \quad (10.1)$$

где  $y$  может быть вектором. Обычным способом перепишем (10.1) в виде системы первого порядка и применим многошаговый метод

$$\begin{aligned} \sum_{i=0}^k \alpha_i y_{n+i} &= h \sum_{i=0}^k \beta_i y'_{n+i}, \\ \sum_{i=0}^k \alpha_i y'_{n+i} &= h \sum_{i=0}^k \beta_i f(x_{n+i}, y_{n+i}, y'_{n+i}). \end{aligned} \quad (10.2)$$

Если правая часть дифференциального уравнения не зависит от  $y'$ ,

$$y'' = f(x, y), \quad (10.3)$$

то естественно обратиться к численным методам, не содержащим первой производной  $y'$ . Исключая  $\{y'_n\}$  из уравнений (10.2), получим

$$\sum_{i=0}^{2k} \hat{\alpha}_i y_{n+i} = h^2 \sum_{i=0}^{2k} \hat{\beta}_i f(x_{n+i}, y_{n+i}), \quad (10.4)$$

где новые коэффициенты  $\hat{\alpha}_i$ ,  $\hat{\beta}_i$  определяются из формул

$$\sum_{i=0}^{2k} \hat{\alpha}_i \xi^i = \left( \sum_{i=0}^k \alpha_i \xi^i \right)^2, \quad \sum_{i=0}^{2k} \hat{\beta}_i \xi^i = \left( \sum_{i=0}^k \beta_i \xi^i \right)^2. \quad (10.5)$$

Далее мы исследуем методы (10.4), коэффициенты которых не обязательно удовлетворяют (10.5). При этом мы надеемся добиться того же порядка при меньшем числе шагов.

### Первые методы

«Мы сосчитали более 120 различных траекторий; эта огромная работа потребовала более 4500 часов...»  
(С. Штёрмер, 1907)

«Кроме того, его доклад довольно сух и скучен...»  
(Мнение Б. Римана об Энке, 1847)

Поскольку большинство дифференциальных уравнений небесной механики имеет вид (10.3), неудивительно, что первые попытки построения специальных методов для таких уравнений были сделаны астрономами.

В своем трактате о северном сиянии С. Штёрмер (1907) вывел дифференциальное уравнение второго порядка, описывающее движение заряженных частиц в магнитном поле Земли. Для обширных численных расчетов этих траекторий (см. приведенную выше цитату!) он построил точный и простой метод: сложив ряды Тейлора для  $y(x_n + h)$  и  $y(x_n - h)$ , мы получим

$$y(x_n + h) - 2y(x_n) + y(x_n - h) = h^2 y''(x_n) + \frac{h^4}{12} y^{(4)}(x_n) + \\ + \frac{h^6}{360} y^{(6)}(x_n) + \dots$$

Подставив сюда  $y''(x_n)$  из дифференциального уравнения (10.3) и пренебрегая членами более высокого порядка, придем к формуле

$$y_{n+1} - 2y_n + y_{n-1} = h^2 f_n,$$

которая дает первый простой и точный метод, называемый иногда методом Штёрмера или Энке. Чтобы увеличить точность, заменим производные  $y$  более высокого порядка на центральные разности  $f$ :

$$h^2 y^{(4)}(x_n) = \Delta^2 f_{n-1} - \frac{1}{12} \Delta^4 f_{n-2} + \dots,$$

$$h^4 y^{(6)}(x_n) = \Delta^4 f_{n-2} + \dots;$$

получим формулу

$$y_{n+1} - 2y_n + y_{n-1} = h^2 \left[ f_n + \frac{1}{12} \Delta^2 f_{n-1} - \frac{1}{240} \Delta^4 f_{n-2} + \dots \right]. \quad (10.6)$$

Она пока не очень практична, поскольку разделенные разности в правой части содержат неизвестные значения  $f_{n+1}$  и  $f_{n+2}$ . Пренебрегая разностями пятого порядка (т. е. полагая  $\Delta^4 f_{n-2} =$

$= \Delta^4 f_{n-4}$  и  $\Delta^2 f_{n-1} = \Delta^2 f_{n-2} + \Delta^3 f_{n-3} + \Delta^4 f_{n-4}$ ), придем к формуле «...формула, которая лежит в основе нашего метода...», Штёрмер, 1907)

$$\begin{aligned} y_{n+1} - 2y_n + y_{n-1} &= \\ = h^2 f_n + \frac{h^2}{12} \left[ \Delta^2 f_{n-2} + \Delta^3 f_{n-3} + \Delta^4 f_{n-4} - \frac{1}{20} \Delta^6 f_{n-8} \right]. \end{aligned} \quad (10.7)$$

«Если бы Очерки Астрономического общества были полностью свободны от грубых численных ошибок...»  
(П. Х. Кауэлл и А. С. Д. Кроммелин, 1910)

Через несколько лет Кауэлл и Кроммелин (1910) использовали те же идеи для исследования движения кометы Галлея. Они рассмотрели один дополнительный член в разложении (10.6), а именно:

$$\frac{31}{60\,480} \Delta^6 f_{n-8} \approx \frac{1}{1951} \Delta^6 f_{n-8}.$$

Прямое использование формулы (10.6) в ее варианте четвертого порядка

$$y_{n+1} - 2y_n + y_{n-1} = h^2 \left[ f_n + \frac{1}{12} \Delta^2 f_{n-1} \right] \quad (10.8)$$

обосновал Б. Нумеров (1924, 1927): вместо того, чтобы заменить  $\Delta^2 f_{n-1}$  на какое-то явное выражение, он попытался решить прямо это неявное нелинейное уравнение относительно  $y_{n+1}$ .

### Задача Штёрмера

«При достаточном навыке вычисляют около трех точек ( $R, z$ ) в час». (С. Штёрмер, 1907)

Предположим, что элементарный магнит помещен в начало координат, а его ось соответствует оси  $z$ . Тогда траектория  $(x(s), y(s), z(s))$  заряженной частицы в этом магнитном поле является решением системы (Штёрмер, 1907)

$$\begin{aligned} x'' &= \frac{1}{r^5} (3yzz' - (3z^2 - r^2)y'), \\ y'' &= \frac{1}{r^5} ((3z^2 - r^2)x' - 3xzz'), \\ z'' &= \frac{1}{r^5} (3xzy' - 3yzx'), \end{aligned} \quad (10.9)$$

где  $r^2 = x^2 + y^2 + z^2$ . При введении полярных координат  
 $x = R \cos \varphi, \quad y = R \sin \varphi \quad (10.10)$

система (10.9) принимает вид

$$R'' = \left( \frac{2\gamma}{R} + \frac{R}{r^8} \right) \left( \frac{2\gamma}{R^2} + \frac{3R^2}{r^6} - \frac{1}{r^8} \right),$$

$$z'' = \left( \frac{2\gamma}{R} + \frac{R}{r^8} \right) \frac{3Rz}{r^6}, \quad (10.11a)$$

$$\varphi' = \left( \frac{2\gamma}{R} + \frac{R}{r^8} \right) \frac{1}{R}, \quad (10.11b)$$

где  $\gamma$  — некоторая константа, появляющаяся при интегрировании  $\varphi''$ . Два уравнения (10.11a) образуют дифференциальное

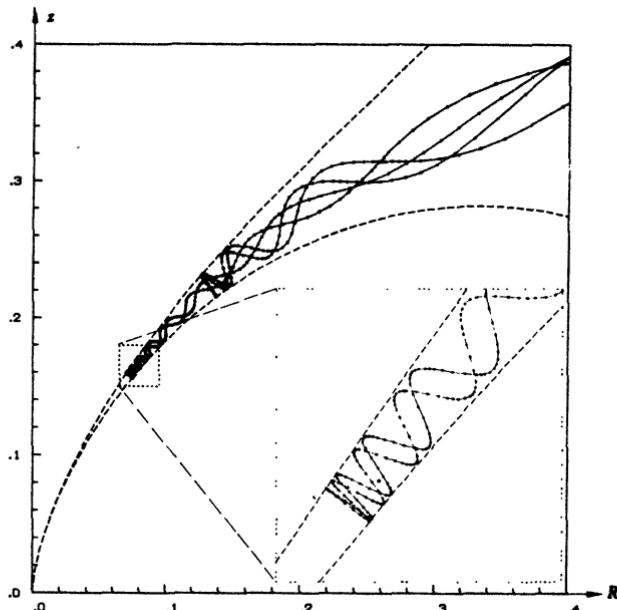


Рис. 10.1. Решения системы (10.11a) в плоскости  $(R, z)$ .

уравнение второго порядка типа (10.3) (отметим, что  $r^2 = R^2 + z^2$ ), которое можно решить численно с помощью методов из данного раздела. Тогда  $\varphi$  получается простым интегрированием равенства (10.11b). На рис. 10.1 показано два решения подсистемы (10.11a) с начальными условиями, поставленными самим Штёрмером:  $R_0 = 0.257453, \quad z_0 = 0.314687, \quad \varphi = -0.5, \quad r_0 =$

$= \sqrt{R_0^2 + z_0^2}$ ,  $Q_0 = 1 - (2\gamma/R_0 + R_0/r_0^3)^2$ ,  $R_0' = \sqrt{Q_0} \cos u$ ,  $z_0' = \sqrt{Q_0} \sin u$  при  $u \in \{0, \pi/4, \pi, 5\pi/4\}$ . На рис. 10.2 представлено трехмерное изображение в пространстве  $(x, y, z)$  для решения с начальными условиями  $u = \pi/4$  и  $5\pi/4$ .

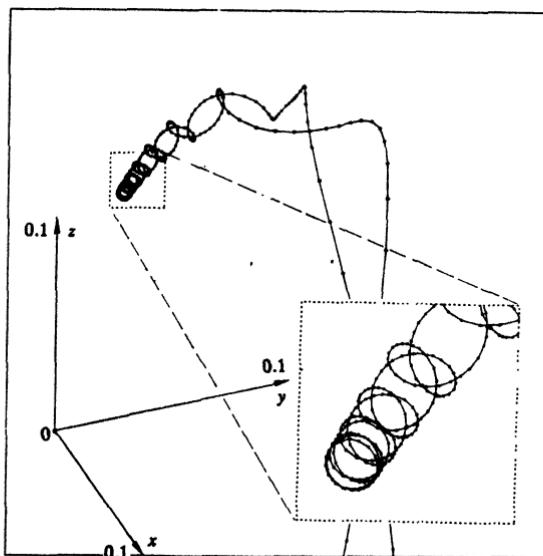


Рис. 10.2. Трехмерное изображение решения.

### Методы более высокого порядка

Формулу Штёрмера можно обобщить следующим образом: дважды проинтегрировав уравнение (10.3), мы получим

$$y(x+h) = y(x) + hy'(x) + h^2 \int_0^1 (1-s)[f(x+sh, y(x+sh))] ds. \quad (10.12)$$

Чтобы исключить первую производную  $y'(x)$ , запишем ту же формулу, заменив в ней  $h$  на  $-h$  и сложив два выражения:

$$\begin{aligned} y(x+h) - 2y(x) + y(x-h) &= \\ &= h^2 \int_0^1 (1-s)[f(x+sh, y(x+sh)) + f(x-sh, y(x-sh))] ds. \end{aligned} \quad (10.13)$$

Так же как при выводе формул Адамса (разд. III.1), заменим неизвестную функцию  $f(t, y(t))$  интерполяционным многочленом  $p(t)$ , заданным формулой (1.4). Это даст нам явный метод

$$y_{n+1} - 2y_n + y_{n-1} = h^2 \sum_{i=0}^{k-1} \sigma_i \nabla^i f_n \quad (10.14)$$

с коэффициентами

$$\sigma_j = (-1)^j \int_0^1 (1-s) \left[ \binom{-s}{j} + \binom{s}{j} \right] ds. \quad (10.15)$$

Их числовые значения даны в табл. 10.1, а способ вычисления описан в упр. 2.

Таблица 10.1. Коэффициенты метода (10.14)

$f$	0	1	2	3	4	5	6	7
$\sigma_j$	1	0	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{19}{240}$	$\frac{3}{40}$	$\frac{863}{12\,096}$	$\frac{275}{4032}$

Приведем частные случаи формулы (10.14):

$$k=2: y_{n+1} - 2y_n + y_{n-1} = h^2 f_n,$$

$$k=3: y_{n+1} - 2y_n + y_{n-1} = h^2 \left[ \frac{13}{12} f_n - \frac{1}{6} f_{n-1} + \frac{1}{12} f_{n-2} \right], \quad (10.14')$$

$$k=4: y_{n+1} - 2y_n + y_{n-1} = h^2 \left[ \frac{7}{6} f_n - \frac{5}{12} f_{n-1} + \frac{1}{3} f_{n-2} - \frac{1}{12} f_{n-3} \right].$$

При  $k=5$  метод (10.14) эквивалентен формуле (10.7), которая задает метод, использованный Штёрмером (1907, 1921), а при  $k=6$  получается метод из работы Кауэлла и Кроммелина (1910). Простейший из этих методов ( $k=1$ ) был успешно применен в качестве основы при построении экстраполационного метода (разд. II.13, формула (13.32)).

Для повышения точности можно также использовать интерполяционный многочлен  $p^*(t)$  из (1.7), который проходит через дополнительную точку  $(x_{n+1}, f_{n+1})$ . Заменив в (10.13)  $f(t, y(t))$  на  $p^*(t)$ , мы придем к неявному методу:

$$y_{n+1} - 2y_n + y_{n-1} = h^2 \sum_{i=0}^k \sigma_i^* \nabla^i f_{n+1}. \quad (10.16)$$

Коэффициенты  $\sigma_j^*$  определяются формулой

$$\sigma_j^* = (-1)^j \int_0^1 (1-s) \left[ \binom{-s+1}{j} + \binom{s+1}{j} \right] ds, \quad (10.17)$$

а их числовые значения для  $j \leq 6$  приведены в табл. 10.2.

Таблица 10.2. Коэффициенты неявного метода (10.16)

$t$	0	1	2	3	4	5	6	7
$\sigma_t^*$	1	-1	$\frac{1}{12}$	0	$-\frac{1}{240}$	$-\frac{1}{240}$	$-\frac{221}{60480}$	$-\frac{19}{6048}$

При  $k = 2$  (а также  $k = 3$ ) получается интересный метод

$$y_{n+1} - 2y_n + y_{n-1} = h^2 \left[ \frac{1}{12} f_{n+1} + \frac{10}{12} f_n + \frac{1}{12} f_{n-1} \right], \quad (10.16')$$

который обычно приписывают Нумерову.

Используя идеи Ньюстрёма и Милна для уравнений первого порядка, можно вывести и другие методы. При заменах  $h \rightarrow 2h$ ,  $2s \rightarrow s$  и  $x \rightarrow x - h$  формула (10.13) примет вид

$$\begin{aligned} y(x+h) - 2y(x-h) + y(x-3h) &= h^2 \int_0^2 (2-s), \\ &\text{if } (x+(s-1)h, y(x+(s-1)h)) + \\ &\quad + f(x-(s+1)h, y(x-(s+1)h)) ds. \end{aligned} \quad (10.18)$$

Замена  $f(t, y(t))$  на многочлен  $p(t)$  (или  $p^*(t)$ ) даст нам новые классы явных и неявных методов.

### Общая формулировка

Следующая наша цель — изучить устойчивость, согласованность и сходимость общих линейных многошаговых методов для задачи (10.3). Запишем их в виде

$$\sum_{i=0}^k \alpha_i y_{n+i} = h^2 \sum_{i=0}^k \beta_i f(x_{n+i}, y_{n+i}). \quad (10.19)$$

Производящие многочлены с коэффициентами  $\alpha_i$  и  $\beta_i$  снова обозначим через

$$\rho(\zeta) = \sum_{i=0}^k \alpha_i \zeta^i, \quad \sigma(\zeta) = \sum_{i=0}^k \beta_i \zeta^i. \quad (10.20)$$

Если метод (10.19) применить к задаче Коши

$$y'' = f(x, y), \quad y(x_0) = y_0, \quad y'(x_0) = y'_0, \quad (10.21)$$

то будет естественно потребовать согласованности стартовых значений с обоими начальными условиями, т. е.

$$\frac{y_t - y_0 - ihy'_0}{h} \rightarrow 0 \quad \text{при } h \rightarrow 0, \quad t = 0, 1, \dots, k-1. \quad (10.22)$$

### Условие устойчивости

Чтобы получить представление о том, какое условие устойчивости необходимо для сходимости, рассмотрим простую задачу:

$$y'' = 0, \quad y_0 = 0, \quad y'_0 = 0.$$

Ее численное решение удовлетворяет линейному разностному уравнению с характеристическим многочленом  $\rho(\zeta)$ . Те же рассуждения, что и при доказательстве теоремы 4.2, показывают, что необходимым для сходимости является следующее условие.

**Определение 10.1.** Метод (10.19) называется *устойчивым*, если для производящего многочлена  $\rho(\zeta)$  выполняются условия:

- i) Корни  $\rho(\zeta)$  лежат на (или внутри) единичной окружности;
- ii) Кратность корней на единичной окружности не превосходит двух.

### Одношаговое представление метода (10.19)

Для рассматриваемых методов можно построить теорию по аналогии с разд. III.2—III.4. Такой подход разработан Далквистом (1959) и описан также в книге Хенричи (1962). Мы предпочтаем переписать (10.19) в форме (8.4) и прямо применить результаты из разд. III.8 и III.9 (см. пример 8.7). Для этого мы могли бы ввести вектор  $u_n = (y_{n+k-1}, \dots, y_n)^T$ , что кажется вполне естественным. Но тогда соответствующая матрица  $S$  не будет удовлетворять условию устойчивости из определения 8.8 из-за двукратных корней, равных по модулю единице. Чтобы преодолеть это препятствие, мы отделим такие корни. Представим характеристический многочлен в виде произведения

$$\rho(\zeta) = \rho_1(\zeta) \cdot \rho_2(\zeta) \quad (10.23)$$

так, чтобы каждый из многочленов

$$\rho_1(\zeta) = \sum_{i=0}^l \gamma_i \zeta^i, \quad \rho_2(\zeta) = \sum_{i=0}^m \kappa_i \zeta^i \quad (l + m = k) \quad (10.24)$$

имел только простые корни с модулем 1. Без ограничения общности в дальнейшем будем считать, что  $m \geq l$  и  $\alpha_k = \gamma_l = \kappa_m = 1$ . С помощью оператора сдвига  $E$ , определенного в (2.10), метод (10.19) можно записать в виде

$$\rho(E) y_n = h^2 \sigma(E) f_n.$$

Основная идея заключается в том, чтобы ввести новую переменную  $\rho_2(E) y_n$  (обозначим ее через  $h v_n$ ). Тогда многошаговая формула будет эквивалентна системе

$$\rho_1(E) v_n = h \sigma(E) f_n, \quad \rho_2(E) y_n = h v_n. \quad (10.25)$$

Если ввести вектор

$$u_n = (v_{n+l-1}, \dots, v_n, y_{n+m-1}, \dots, y_n)^T,$$

то формулу (10.25) можно записать в виде

$$u_{n+1} = Su_n + h\Phi(x_n, u_n, h), \quad (10.26a)$$

где

$$S = \begin{pmatrix} G & 0 \\ 0 & K \end{pmatrix}, \quad \Phi(x_n, u_n, h) = \begin{pmatrix} \psi(x_n, u_n, h) \\ 0 \\ \vdots \\ 0 \\ v_n \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \quad (10.27)$$

Матрицы  $G$  и  $K$  являются сопровождающими:

$$G = \begin{pmatrix} -\gamma_{l-1} & -\gamma_{l-2} & \dots & -\gamma_0 \\ 1 & 0 & \dots & 0 \\ & 1 & . & 0 \\ & & \ddots & \\ & & & 1 \\ & & & 1 & 0 \end{pmatrix}, \quad K = \begin{pmatrix} -\kappa_{m-1} & -\kappa_{m-2} & \dots & -\kappa_0 \\ 1 & 0 & \dots & 0 \\ & 1 & . & 0 \\ & & \ddots & \\ & & & 1 \\ & & & 1 & 0 \end{pmatrix},$$

а  $\psi = \psi(x_n, u_n, h)$  задается неявно формулой

$$\psi = \sum_{j=0}^{k-1} \beta_j f(x_n + jh, y_{n+j}) + \beta_k f\left(x_n + kh, h^2 \psi - \sum_{j=0}^{k-1} \alpha_j y_{n+j}\right). \quad (10.28)$$

В ней  $\psi$  записана как функция от  $x_n$ ,  $(y_{n+k-1}, \dots, y_n)$  и  $h$ . Но из второго соотношения в (10.25) видно, что каждое из значений  $y_{n+k-1}, \dots, y_{n+m}$  можно выразить в виде линейной комбинации элементов  $u_n$ . Поэтому в действительности  $\psi$  является функцией переменных  $(x_n, u_n, h)$ .

Формула (10.26) задает процедуру продвижения на шаг. Соответствующая стартовая процедура имеет вид

$$\varphi(h) = (v_{l-1}, \dots, v_0, y_{m-1}, \dots, y_0)^T \quad (10.26b)$$

и в силу (10.25) определяется по значениям  $(y_{k-1}, \dots, y_0)^T$  единственным образом. Функцией точных значений является

$$z(x, h) = \left( \frac{1}{h} \rho_2(E) y(x + (l-1)h), \dots, \frac{1}{h} \rho_2(E) y(x), \right.$$

$$\left. y(x + (m-1)h), \dots, y(x) \right)^T. \quad (10.26c)$$

При нашем выборе многочленов  $\rho_1(\zeta)$  и  $\rho_2(\zeta)$  (оба имеют только простые корни с модулем 1) матрицы  $G$  и  $K$  ограничены при возведении в любую степень. Следовательно, это справедливо также и для  $S$ , и метод (10.26) *устойчив* в смысле определения 8.8.

### Согласованность и сходимость

Найдем условия на параметры  $\alpha_i$  и  $\beta_i$ , при которых метод (10.26) согласован с порядком  $p$  (в смысле леммы 8.11). Начнем с погрешности в начальных условиях:

$$d_0 = z(x_0, h) - \Phi(h).$$

Первые  $l$  компонент этого вектора имеют вид

$$\frac{1}{h} \rho_2(E) y(x_j) - v_j = \frac{1}{h} \sum_{i=0}^m \alpha_i (y(x_{i+j}) - y_{i+j}), \quad j = 0, \dots, l-1,$$

а последние  $m$  компонент равны

$$y(x_j) - y_j, \quad j = 0, \dots, m-1.$$

Мы видим, что условие, достаточное для выполнения соотношения  $d_0 = O(h^p)$ , состоит в следующем:

$$y(x_j) - y_j = O(h^{p+1}) \quad \text{при } j = 0, 1, \dots, k-1. \quad (10.29)$$

Затем рассмотрим локальную погрешность в точке  $x_n$ ,

$$d_{n+1} = z(x_n + h, h) - Sz(x_n, h) - h\Phi(x_n, z(x_n, h), h).$$

Она имеет все нулевые компоненты, за исключением первой, которая равна

$$d_{n+1}^{(1)} = \frac{1}{h} \rho(E) y(x_n) - h\psi(x_n, z(x_n, h), h).$$

Так же, как в разд. III.2, введем линейный разностный оператор

$$\begin{aligned} L(y, x, h) &= \rho(E) y(x) - h^2 \sigma(E) y''(x) = \\ &= \sum_{i=0}^k (\alpha_i y(x + ih) - h^2 \beta_i y''(x + ih)). \end{aligned}$$

Используя формулу (10.28), по теореме о среднем получим

$$d_{n+1}^{(1)} = \frac{1}{h} L(y, x_n, h) + h^2 \beta_k f'(x_{n+k}, \eta) \cdot d_{n+1}^{(1)}, \quad (10.30)$$

где  $\eta$  имеет то же значение, что и в лемме 2.2. Отсюда следует, что

$$d_{n+1} = O(h^{p+1}) \Leftrightarrow L(y, x_n, h) = O(h^{p+2}).$$

Применяя теперь теорему сходимости 8.13, придем к следующему результату.

**Теорема 10.2.** Предположим, что метод (10.19) устойчив, стартовые значения отвечают условию (10.29) и для всех достаточно гладких функций  $y(x)$  выполняется равенство

$$L(y, x, h) = O(h^{p+2}). \quad (10.31)$$

Тогда метод (10.19) сходится с порядком  $p$ , т. е.

$$\|y(x_n) - y_n\| \leq Ch^p \quad \text{при } 0 \leq nh \leq \text{const}. \quad \blacksquare$$

Следующая теорема дает полезный критерий для проверки соотношения (10.31).

**Теорема 10.3.** Многошаговый метод (10.19) удовлетворяет равенству (10.31) тогда и только тогда, когда выполняются эквивалентные условия:

$$\text{i)} \sum_{i=0}^k \alpha_i = 0, \quad \sum_{i=0}^k i\alpha_i = 0$$

$$\text{и } \sum_{i=0}^k \alpha_i i^q = q(q-1) \sum_{i=0}^k \beta_i i^{q-2} \quad \text{при } q = 2, \dots, p,$$

$$\text{ii)} \rho(e^h) - h^2\sigma(e^h) = O(h^{p+2}) \quad \text{при } h \rightarrow 0,$$

$$\text{iii)} \frac{\rho(\zeta)}{(\log \zeta)^2} - \sigma(\zeta) = O((\zeta-1)^p) \quad \text{при } \zeta \rightarrow 1.$$

*Доказательство* опускается, так как оно повторяет доказательство теоремы 2.4. ■

**Замечания.** 1. Многочлен  $\rho(\zeta)$  обязательно имеет двукратный корень  $\zeta_1 = 1$ . Поэтому многочлены  $\rho_1, \rho_2$  из (10.24) отвечают равенствам  $\rho_1(1) = 0$  и  $\rho_2(1) = 0$ . Таким образом, функция точных значений  $z(x, h)$  имеет предел при  $h \rightarrow 0$ .

2. Как и в случае методов Адамса, легко проверить, что метод (10.14) имеет порядок  $k$ , а (10.16) — порядок  $k+1$ .

### Асимптотическая формула для глобальной погрешности

Предположим, что метод (10.19) устойчив и согласован с порядком  $p$ . Тогда локальная погрешность для (10.26) равна

$$d_{n+1} = e_1 h^{p+1} C_{p+2} y^{(p+2)}(x_n) + O(h^{p+2}), \quad (10.32)$$

где

$$C_{p+2} = \frac{1}{(p+2)!} \sum_{i=0}^k (\alpha_i i^{p+2} - (p+2)(p+1)\beta_i i^p).$$

Формулу (10.32) можно проверить, разложив в (10.30)  $L(y, x_n, h)$  в ряд Тейлора. Применяя теорему 9.1 (если среди корней  $\rho(\zeta)$  только один имеет модуль 1) или теорему 9.2, получим, что глобальная погрешность метода (10.26) имеет вид

$$u_h(x) - z(x, h) = e(x)h^p + O(h^{p+1}),$$

где  $e(x)$  — решение уравнения

$$e^*(x) = E \frac{\partial \Phi}{\partial u}(x, z(x, 0), 0) e(x) - E e_1 \cdot C_{p+2} y^{(p+2)}(x). \quad (10.33)$$

Здесь  $E$  — это матрица, определенная в (8.12). Поскольку в локальной погрешности (10.32) нет членов порядка  $h^p$ , из (9.16) следует равенство  $e(x) = Ee(x)$ . Тогда (см. упр. 4а) эту функцию можно записать в виде

$$e(x) = \begin{pmatrix} \Psi(x) & 1 \\ \kappa(x) & 1 \end{pmatrix}.$$

Вычислив  $\frac{\partial \Phi}{\partial u}(x, z(x, 0), 0)$  и  $Ee_1$  (подробное описание см. в упр. 4), мы увидим, что (10.33) преобразуется в эквивалентную систему

$$\Psi'(x) = \frac{\sigma(1)}{\rho'_1(1)} \frac{\partial f}{\partial y}(x, y(x)) \kappa(x) - \frac{C_{p+2}}{\rho'_1(1)} y^{(p+2)}(x), \quad (10.34a)$$

$$\kappa'(x) = \frac{1}{\rho'_2(1)} \Psi(x). \quad (10.34b)$$

Дифференцируя (10.34b) и подставляя в полученное уравнение  $\Psi'(x)$  из (10.34a), мы, наконец, придем к уравнению

$$\kappa''(x) = \frac{\partial f}{\partial y}(x, y(x)) \kappa(x) - Cy^{(p+2)}(x), \quad (10.35)$$

где

$$C = \frac{C_{p+2}}{\sigma(1)}. \quad (10.36)$$

Здесь мы использовали соотношение  $\sigma(1) = \rho'_1(1) \cdot \rho'_2(1)$ , которое является прямым следствием равенства (10.23) и предположения о том, что порядок метода не ниже первого. Постоянная  $C$  в (10.35) называется константой погрешности метода (10.19). Она играет ту же роль, что и (2.13) для уравнений первого порядка.

Поскольку последняя компонента вектора  $u_n$  равна  $y_n$ , мы приходим к требуемому результату

$$y_n - y(x_n) = \kappa(x_n)h^p + O(h^{p+1}),$$

где  $\kappa(x)$  отвечает (10.35). Применяя результаты из разд. III.9, можно получить и последующие члены асимптотического разложения глобальной погрешности.

### Порядковый барьер для устойчивых методов (10.19)

Вполне естественно встает вопрос о поиске среди устойчивых многошаговых методов (10.19) метода максимально возможного порядка.

**Теорема 10.4.** Порядок  $p$  устойчивого линейного многошагового метода (10.19) удовлетворяет ограничениям:

$$p \leq k + 2 \text{ при четных } k,$$

$$p \leq k + 1 \text{ при нечетных } k.$$

**Теорема 10.5.** Устойчивый многошаговый метод (10.19) порядка  $k + 2$  является симметричным, т. е.

$$\alpha_i = \alpha_{k-j}, \quad \beta_j = \beta_{k-j} \text{ при всех } j.$$

Эти две теоремы доказываются подобно теореме 3.5 и 3.9. Здесь мы опускаем доказательства, которые читатель может найти в оригинальной статье Далквиста (1959) или в книге Хенричи (1962).

### Погрешности округления

Обсудим кратко еще одно преимущество представления (10.26). Среди множества способов реализации многошагового метода остановимся на следующих двух:

i) хранение значений  $(y_{n+k-1}, \dots, y_n)$  и вычисление  $y_{n+k}$  прямо по формуле (10.19);

ii) хранение  $u_n$  и использование формулы (10.26). Назовем это стабилизацией (10.19).

Оба процесса математически эквивалентны. Однако с вычислительной точки зрения они различаются. При малых  $h$  рекуррентное соотношение (10.19) неустойчиво (из-за двукратных корней на единичной окружности), а рекуррентная формула (10.26) устойчива (матрица  $S$  ограничена при возведении в любую степень). Чтобы проиллюстрировать эту разницу, применим формулу Штёрмера (10.14'), отвечающую значению  $k = 4$ , к задаче  $y'' = -y$  с точным решением  $y(x) = \sin x$ . В качестве стартовых возьмем значения точного решения и выберем шаг  $h = 0.01$ , чтобы погреш-

ность обрывания ряда сохранялась меньшей, чем погрешности округления (одинарная точность при EPSMACH = 1.5E — 8). На рис. 10.3 показано абсолютное значение глобальной погреш-

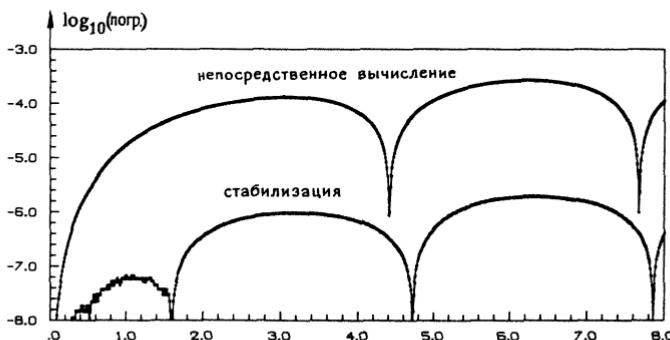


Рис. 10.3. Погрешности округления для двух реализаций формулы Штёрмера (10.14') при  $k = 4$ .

ности для обеих реализаций. Видно, что погрешность стабилизированного варианта меньше в 100 раз. Направленные вниз пики соответствуют смене знака глобальной погрешности.

### Упражнения

- Найдите решение задачи Штёрмера (10.11a) с помощью одного из методов данного раздела.
- а) Покажите, что производящие функции коэффициентов  $\sigma_l$  и  $\sigma_l^*$  [определенных в (10.15) и (10.17)]

$$S(t) = \sum_{l=0}^{\infty} \sigma_l t^l, \quad S^*(t) = \sum_{l=0}^{\infty} \sigma_l^* t^l$$

удовлетворяют формулам

$$S(t) = \left( \frac{t}{\log(1-t)} \right)^2 \frac{1}{1-t}, \quad S^*(t) = \left( \frac{t}{\log(1-t)} \right)^2.$$

- Подсчитайте коэффициенты  $d_j$  ряда

$$\sum_{l=0}^{\infty} d_j t^l = \left( \frac{\log(1-t)}{t} \right)^2 = \left( 1 + \frac{t}{2} + \frac{t^2}{3} + \frac{t^3}{4} + \dots \right)^2$$

и выведите рекуррентное соотношение для  $\sigma_l$  и  $\sigma_l^*$ .

- Докажите равенство

$$\sigma_l^* = \sigma_l - \sigma_{l-1}.$$

3. Пусть  $\rho(\zeta)$  является многочленом степени  $k$  с двукратным корнем 1. Тогда существует единственный многочлен  $\sigma(\zeta)$ , такой что соответствующий метод имеет порядок  $k+1$ .
4. Рассмотрим метод (10.26) и для простоты предположим, что дифференциальное уравнение является скалярным.
- а) Для любого вектора  $w \in \mathbb{R}^k$  его образ  $Ew$  с матрицей  $E$  из (8.12) удовлетворяет равенству

$$Ew = \begin{pmatrix} \gamma & 1 \\ \kappa & 1 \end{pmatrix},$$

где  $\gamma, \kappa$  — вещественные числа, а  $\mathbf{1}$  — вектор, у которого все элементы равны 1. Векторы  $\gamma\mathbf{1}$  и  $\kappa\mathbf{1}$  имеют размерности  $l$  и  $m$  соответственно.

б) Проверьте, что при  $e_1 = (1, 0, \dots, 0)^T$  справедливо соотношение

$$E \begin{pmatrix} \alpha e_1 \\ \beta e_1 \end{pmatrix} = \begin{pmatrix} (\alpha/\rho'_1(1)) \mathbf{1} \\ (\beta/\rho'_2(1)) \mathbf{1} \end{pmatrix}.$$

с) Покажите, что

$$E \frac{\partial \Phi}{\partial u}(x, z(x, 0), 0) \begin{pmatrix} \gamma & \mathbf{1} \\ \kappa & \mathbf{1} \end{pmatrix} = \begin{pmatrix} \left(\frac{\sigma(1)}{\rho'_1(1)}\right) \kappa \mathbf{1} \\ \left(\frac{1}{\rho'_2(1)}\right) \gamma \mathbf{1} \end{pmatrix}.$$

*Указание.* Если ввести  $Y_n = (y_{n+k-1}, \dots, y_n)^T$ , то по формуле (10.28)  $\psi$  можно представить функцией от  $(x_n, Y_n, h)$ . Векторы  $Y_n$  и  $u_n$  связаны соотношением

$$KY_n = Lu_n + O(h),$$

где

$$K = \begin{pmatrix} \kappa_m & \dots & \dots & \kappa_0 \\ \vdots & & & \vdots \\ \kappa_m & \dots & \dots & \kappa_0 \\ 1 & & & \vdots \\ & & & 1 \end{pmatrix}, \quad L = \begin{pmatrix} 0 & & 0 \\ & 1 & \\ 0 & & \ddots & 1 \end{pmatrix}.$$

Для вычисления  $\frac{\partial \psi}{\partial u}(x, z(x, 0), 0)$  используйте целое правило (дифференцирования сложной функции). См. также упр. 2 из разд. III.4 и упр. 1 из разд. III.9.

5. Определите константу погрешности (10.36) для методов (10.14) и (10.16).

*Ответ.*  $\sigma_k$  и  $\sigma_{k+1}^*$  соответственно.

# Приложение. Программы на Фортране

... но программные продукты находятся на разных стадиях разработки, от экспериментальных (эвфемизм для плохо написанных) до того, что мы можем назвать ...  
(Ч. Гир, 1985)

Следующие программы на Фортране были разработаны для наших численных расчетов. Они могут быть получены от авторов (Section de Mathématiques, Case Postale 240, CH—1211 Genéve 24, Switzerland) на диске фирмы ИБМ. Пожалуйста, пришлите 15 швейцарских франков.

## 1. Dopr5

Это программа явного метода Рунге—Кутты, основанная на формулах Дормана и Принса (см. табл. 4.6 в разд. II.4) с автоматическим управлением длиной шага. Наилучший метод, если допустима невысокая точность (от  $10^{-4}$  до  $10^{-7}$ ).

```
SUBROUTINE DOPRIS(N,FCN,X,Y,XEND,EPS,HMAX,H)
C -----
C      NUMERICAL SOLUTION OF A SYSTEM OF FIRST ORDER
C      ORDINARY DIFFERENTIAL EQUATIONS Y'=F(X,Y).
C      THIS IS AN EMBEDDED RUNGE-KUTTA METHOD OF ORDER (4)5
C      DUE TO DORMAND & PRINCE (WITH STEPSIZE CONTROL).
C      C.F. SECTION II.4
C
C      INPUT PARAMETERS
C -----
C      N          DIMENSION OF THE SYSTEM (N.LE.51)
C      FCN        NAME (EXTERNAL) OF SUBROUTINE COMPUTING THE
C                  FIRST DERIVATIVE F(X,Y):
C                  SUBROUTINE FCN(N,X,Y,F)
C                  REAL*4 X,Y(N),F(N)
C                  F(1)=... ETC.
C      X          INITIAL X-VALUE
C      XEND       FINAL X-VALUE (XEND-X POSITIVE OR NEGATIVE)
C      Y(N)       INITIAL VALUES FOR Y
C      EPS        LOCAL TOLERANCE
C      HMAX       MAXIMAL STEPSIZE
C      H          INITIAL STEPSIZE GUESS
C
C      OUTPUT PARAMETERS
C -----
C      Y(N)       SOLUTION AT XEND
C
C      EXTERNAL SUBROUTINE (TO BE SUPPLIED BY THE USER)
C -----
C      SOLOUS      THIS SUBROUTINE IS CALLED AFTER EVERY
C                  SUCCESSFULLY COMPUTED STEP (AND THE
C                  INITIAL VALUE):
C                  SUBROUTINE SOLOUS (NR,X,Y,N)
C                  REAL*4 X,Y(N)
C                  FURNISHES THE SOLUTION Y AT THE NR-TH
C                  GRID-POINT X (THE INITIAL VALUE IS CON-
C                  SIDERED AS THE FIRST GRID-POINT).
C                  SUPPLY A DUMMY SUBROUTINE, IF THE SOLUTION
C                  IS NOT DESIRED AT THE INTERMEDIATE POINTS.
C -----
```

```

REAL*4 K1(51),K2(51),K3(51),K4(51),K5(51),Y1(51),Y(N)
LOGICAL REJECT
COMMON/STAT/NFCN,NSTEP,NACCPT,NREJECT
C --- COMMON STAT CAN BE USED FOR STATISTICS
C --- NFCN      NUMBER OF FUNCTION EVALUATIONS
C --- NSTEP     NUMBER OF COMPUTED STEPS
C --- NACCPT    NUMBER OF ACCEPTED STEPS
C --- NREJECT   NUMBER OF REJECTED STEPS
C --- DATA NMMAX/3000./,UROUND/5.E-8/
C --- NMMAX     MAXIMAL NUMBER OF STEPS
C --- UROUND    SMALLEST NUMBER SATISFYING 1.+UROUND>1.
C ---           (TO BE ADAPTED BY THE USER)
C --- POSNEG=SIGN(1.,XEND-X)
C --- INITIAL PREPARATIONS
HMAX=ABS(HMAX)
H=AMIN1(AMAX1(1.E-4,ABS(H)),HMAX)
H=SIGN(H,POSNEG)
EPS=AMAX1(EPS,7.*UROUND)
REJECT=.FALSE.
NACCPT=0
NREJECT=0
NFCN=1
NSTEP=0
CALL SOLOUS(NACCPT+1,X,Y,N)
CALL FCN(N,X,Y,K1)
C --- BASIC INTEGRATION STEP
1 IF(NSTEP.GT.NMAX.OR.X+.1*H.EQ.X)GOTO 79
IF((X-XEND)*POSNEG+UROUND.GT.0.) RETURN
IF((X+H-XEND)*POSNEG.GT.0.) H=XEND-X
NSTEP=NSTEP+1
C --- THE FIRST 6 STAGES
DO 22 I=1,N
22 Y1(I)=Y(I)+H*.2*K1(I)
CALL FCN(N,X+.2*H,Y1,K2)
DO 23 I=1,N
23 Y1(I)=Y(I)+H*((3./40.)*K1(I)+(9./40.)*K2(I))
CALL FCN(N,X+.3*H,Y1,K3)
DO 24 I=1,N
24 Y1(I)=Y(I)+H*((44./45.)*K1(I)-(56./15.)*K2(I)+(32./9.)*K3(I))
CALL FCN(N,X+.8*H,Y1,K4)
DO 25 I=1,N
25 Y1(I)=Y(I)+H*((19372./6561.)*K1(I)-(25360./2187.)*K2(I)
& +(64448./6561.)*K3(I)-(212./729.)*K4(I))
CALL FCN(N,X+(8./9.)*H,Y1,K5)
DO 26 I=1,N
26 Y1(I)=Y(I)+H*((9017./3168.)*K1(I)-(355./33.)*K2(I)
& +(46732./5247.)*K3(I)+(49./176.)*K4(I)-(5103./18656.)*K5(I))
XPH=X+H
CALL FCN(N,XPH,Y1,K2)
DO 27 I=1,N
27 Y1(I)=Y(I)+H*((35./384.)*K1(I)+(500./1113.)*K3(I)
& +(125./192.)*K4(I)-(2187./6784.)*K5(I)+(11./84.)*K2(I))
C --- COMPUTE INTERMEDIATE SUM TO SAVE MEMORY
DO 61 I=1,N
61 K2(I)=(71./57600.)*K1(I)-(71./16695.)*K3(I)
& +(71./1920.)*K4(I)-(17253./339200.)*K5(I)+(22./525.)*K2(I)
C --- THE LAST STAGE
CALL FCN(N,XPH,Y1,K3)
DO 28 I=1,N
28 K4(I)=(K2(I)-(1./40.)*K3(I))*H
NFCN=NFCN+6
C --- ERROR ESTIMATION
ERR=0.
DO 41 I=1,N
DENOM=AMAX1(1.E-5,ABS(Y1(I)),ABS(Y(I)),2.*UROUND/EPS)
41 ERR=ERR+(K4(I)/DENOM)**2
ERR=SQRT(ERR/FLOAT(N))

```

```

C --- COMPUTATION OF HNEW
C --- WE REQUIRE .2<=HNEW/H<=10.
FAC=AMAX1(.1,AMIN1(5.,(ERR/EPS)**(1./5.)/.9))
HNEW=H/FAC
IF(ERR.LE.EPS)THEN
C --- STEP IS ACCEPTED
NACCPT=NACCPT+1
DO 44 I=1,N
K1(I)=K3(I)
44 Y(I)=Y1(I)
X=XPH
CALL SOLOUS(NACCPT+1,X,Y,N)
IF(ABS(HNEW).GT.HMAX)HNEW=POSNEG*HMAX
IF(REJECT)HNEW=POSNEG*AMIN1(ABS(HNEW),ABS(H))
REJECT=.FALSE.
ELSE
C --- STEP IS REJECTED
REJECT=.TRUE.
IF(NACCPT.GE.1)NREJCT=NREJCT+1
END IF
H=HNEW
GOTO 1
C --- FAIL EXIT
79 WRITE(6,979)X
979 FORMAT(' EXIT OF DOPRI5 AT X=',E11.4)
RETURN
END
C
SUBROUTINE SOLOUS(NRPNTS,X,Y,N)
REAL*4 Y(N)
RETURN
END

```

## 2. Dopri8

Это программа явного метода Рунге—Кутты, основанная на формулах Дорманда и Принса с автоматическим управлением длиной шага (табл. 4.6 из раздела II.6). Она написана для расчетов с двойной точностью и предпочтительна, если допустимая погрешность находится в пределах примерно от  $10^{-7}$  до  $10^{-18}$ . Не применяйте ее, если нужна точность выше  $10^{-16}$ .

```

SUBROUTINE DOPRI8(N,FCN,X,Y,XEND,EPS,HMAX,H)
C -----
C NUMERICAL SOLUTION OF A SYSTEM OF FIRST ORDER
C ORDINARY DIFFERENTIAL EQUATIONS Y=F(X,Y).
C THIS IS AN EMBEDDED RUNGE-KUTTA METHOD OF ORDER (7)8
C DUE TO DORMAND & PRINCE (WITH STEPSIZE CONTROL).
C F. SECTION II.6
C
C INPUT PARAMETERS
C -----
C N           DIMENSION OF THE SYSTEM (N.LE.51)
C FCN         NAME (EXTERNAL) OF SUBROUTINE COMPUTING THE
C             FIRST DERIVATIVE F(X,Y):
C             SUBROUTINE FCN(N,X,Y,F)
C             REAL*8 X,Y(N),F(N)
C             F(1)=... ETC.
C X           INITIAL X-VALUE
C XEND        FINAL X-VALUE (XEND-X POSITIVE OR NEGATIVE)
C Y(N)        INITIAL VALUES FOR Y
C EPS         LOCAL TOLERANCE
C HMAX        MAXIMAL STEPSIZE
C H           INITIAL STEPSIZE GUESS

```

```

C      OUTPUT PARAMETERS
C -----
C      Y(N)      SOLUTION AT XEND
C
C      EXTERNAL SUBROUTINE (TO BE SUPPLIED BY THE USER)
C -----
C      SOLOUT      THIS SUBROUTINE IS CALLED AFTER EVERY
C                     SUCCESSFULLY COMPUTED STEP (AND THE
C                     INITIAL VALUE):
C                     SUBROUTINE SOLOUT (NR,X,Y,N)
C                     REAL*8 X,Y(N)
C FURNISHES THE SOLUTION Y AT THE NR-TH
C GRID-POINT X (THE INITIAL VALUE IS CON-
C SIDERED AS THE FIRST GRID-POINT).
C SUPPLY A DUMMY SUBROUTINE, IF THE SOLUTION
C IS NOT DESIRED AT THE INTERMEDIATE POINTS.
C
C -----
IMPLICIT REAL*8 (A-H,O-Z)
REAL*8 K1(51),K2(51),K3(51),K4(51),K5(51),K6(51),K7(51)
*,Y(N),Y1(51)
LOGICAL REJECT
COMMON/STAT/NFCN,NSTEP,NACCPT,NREJCT
C --- COMMON STAT CAN BE USED FOR STATISTICS
C --- NFCN      NUMBER OF FUNCTION EVALUATIONS
C --- NSTEP     NUMBER OF COMPUTED STEPS
C --- NACCPT   NUMBER OF ACCEPTED STEPS
C --- NREJCT   NUMBER OF REJECTED STEPS
COMMON/CORF/C2,C3,C4,C5,C6,C7,C8,C9,C10,C11,C12,C13,
&A21,A31,A32,A41,A43,A51,A53,A64,A61,A64,A65,A71,A74,A75,A76,
&A81,A84,A85,A86,A87,A91,A94,A95,A96,A97,A98,A101,A104,A105,A106,
&A107,A108,A109,A111,A114,A115,A116,A117,A118,A119,A1110,A121,
&A124,A125,A126,A127,A128,A129,A1210,A1211,A131,A134,A135,A136,
&A137,A138,A139,A1310,A1311,B1,B6,B7,B8,B9,B10,B11,B12,B13,
&BH1,BH6,BH7,BH8,BH9,BH10,BH11,BH12
DATA NMAX/2000/,UROUND/1.73D-18/
C --- NMAX      MAXIMAL NUMBER OF STEPS
C --- UROUND    SMALLEST NUMBER SATISFYING 1.D0+UROUND>1.D0
C ---           (TO BE ADAPTED BY THE USER)
CALL CORFST
POSNEG=DSIGN(1.D0,XEND-X)
C --- INITIAL PREPARATIONS
HMAX=DABS(HMAX)
H=DMIN1(DMAX1(1.D-10,DABS(H)),HMAX)
H=DSIGN(H,POSNEG)
EPS=DMAX1(EPS,13.D0*UROUND)
REJECT=.FALSE.
NACCPT=0
NREJCT=0
NFCN=0
NSTEP=0
CALL SOLOUT(NACCPT+1,X,Y,N)
C --- BASIC INTEGRATION STEP
1  IF(NSTEP.GT.NMAX.OR.X+.03D0*H.EQ.X)GOTO 79
IF((X-XEND)*POSNEG+UROUND.GT.0.D0) RETURN
IF((X+H-XEND)*POSNEG.GT.0.D0)H=XEND-X
CALL FCN(N,X,Y,K1)
2  CONTINUE
NSTEP=NSTEP+1
C --- THE FIRST 9 STAGES
DO 22 I=1,N
22  Y1(I)=Y(I)+H*A21*K1(I)
CALL FCN(N,X+C2*H,Y1,K2)
DO 23 I=1,N
23  Y1(I)=Y(I)+H*(A31*K1(I)+A32*K2(I))
CALL FCN(N,X+C3*H,Y1,K3)
DO 24 I=1,N
24  Y1(I)=Y(I)+H*(A41*K1(I)+A43*K3(I))

```

```

      CALL FCN(N,X+C4*N,Y1,K4)
      DO 25 I=1,N
25   Y1(I)=Y(I)+H*(A51*K1(I)+A53*K3(I)+A54*K4(I))
      CALL FCN(N,X+C5*N,Y1,K5)
      DO 26 I=1,N
26   Y1(I)=Y(I)+H*(A61*K1(I)+A64*K4(I)+A65*K5(I))
      CALL FCN(N,X+C6*N,Y1,K6)
      DO 27 I=1,N
27   Y1(I)=Y(I)+H*(A71*K1(I)+A74*K4(I)+A75*K5(I)+A76*K6(I))
      CALL FCN(N,X+C7*N,Y1,K7)
      DO 28 I=1,N
28   Y1(I)=Y(I)+H*(A81*K1(I)+A84*K4(I)+A85*K5(I)+A86*K6(I)+A87*K7(I))
      CALL FCN(N,X+C8*N,Y1,K2)
      DO 29 I=1,N
29   Y1(I)=Y(I)+H*(A91*K1(I)+A94*K4(I)+A95*K5(I)+A96*K6(I)+A97*K7(I)
     & +A98*K2(I))
      CALL FCN(N,X+C9*N,Y1,K3)
      DO 30 I=1,N
30   Y1(I)=Y(I)+H*(A101*K1(I)+A104*K4(I)+A105*K5(I)+A106*K6(I)
     & +A107*K7(I)+A108*K2(I)+A109*K3(I))
C --- COMPUTE INTERMEDIATE SUMS TO SAVE MEMORY
      DO 61 I=1,N
        Y11S=A111*K1(I)+A114*K4(I)+A115*K5(I)+A116*K6(I)+A117*K7(I)
     & +A118*K2(I)+A119*K3(I)
        Y12S=A121*K1(I)+A124*K4(I)+A125*K5(I)+A126*K6(I)+A127*K7(I)
     & +A128*K2(I)+A129*K3(I)
        K4(I)=A131*K1(I)+A134*K4(I)+A135*K5(I)+A136*K6(I)+A137*K7(I)
     & +A138*K2(I)+A139*K3(I)
        K5(I)=B1*K1(I)+B6*K6(I)+B7*K7(I)+B8*K2(I)+B9*K3(I)
        K6(I)=BH1*K1(I)+BH6*K6(I)+BH7*K7(I)+BH8*K2(I)+BH9*K3(I)
        K2(I)=Y11S
61   K3(I)=Y12S
C --- THE LAST 4 STAGES
      CALL FCN(N,X+C10*N,Y1,K7)
      DO 31 I=1,N
31   Y1(I)=Y(I)+H*(K2(I)+A1110*K7(I))
      CALL FCN(N,X+C11*N,Y1,K2)
      XPH=X+H
      DO 32 I=1,N
32   Y1(I)=Y(I)+H*(K3(I)+A1210*K7(I)+A1211*K2(I))
      CALL FCN(N,XPH,Y1,K3)
      DO 33 I=1,N
33   Y1(I)=Y(I)+H*(K4(I)+A1310*K7(I)+A1311*K2(I))
      CALL FCN(N,XPH,Y1,K4)
      NFCM=NFCM+13
      DO 35 I=1,N
35   K5(I)=Y(I)+H*(K5(I)+B10*K7(I)+B11*K2(I)+B12*K3(I)+B13*K4(I))
      K6(I)=Y(I)+H*(K6(I)+BH10*K7(I)+BH11*K2(I)+BH12*K3(I))
C --- ERROR ESTIMATION
      ERR=0.D0
      DO 41 I=1,N
        DENOM=DMAX1(1.D-6,DABS(K5(I)),DABS(Y(I)),2.D0*UROUND/EPS)
41   ERR=ERR+((K5(I)-K6(I))/DENOM)**2
      ERR=DSQRT(ERR/DFLOAT(N))
C --- COMPUTATION OF HNEW
C --- WE REQUIRE .333<=HNEW/W<=6.
      FAC=DMAX1((1.D0/6.D0),DMIN1(3.D0,(ERR/EPS)**(1.D0/8.D0)/.9D0))
      HNEW=H/FAC
      IF(ERR.GT.EPS)GOTO 51
C --- STEP IS ACCEPTED
      NACCPT=NACCPT+1
      DO 44 I=1,N
44   Y(I)=K5(I)
      X=XPH
      CALL SOLOUT(NACCPT+1,X,Y,N)
      IF(DABS(HNEW).GT.HMAX)HNEW=POSNEG*HMAX
      IF(REJECT)HNEW=POSNEG*DMIN1(DABS(HNEW),DABS(H))

```

```

REJECT=.FALSE.
H=HNEW
GOTO 1
C --- STEP IS REJECTED
51  REJECT=.TRUE.
H=HNEW
IF(NACCPT.GE.1)NREJCT=NREJCT+1
NFCN=NFCN-1
GOTO 2
C --- FAIL EXIT
79  WRITE(6,979)X
979 FORMAT(' EXIT OF DOPRI8 AT X=',D16.7)
RETURN
END
C
SUBROUTINE COEFST
C --- THIS SUBROUTINE SETS THE COEFFICIENTS FOR THE DORMAND-PRINCE
C --- METHOD OF ORDER 8 WITH ERROR ESTIMATOR OF ORDER 7 AND 13 STAGES
IMPLICIT REAL*8 (A-H,O-Z)
COMMON/COEF/C2,C3,C4,C5,C6,C7,C8,C9,C10,C11,C12,C13,
& A21,A31,A32,A41,A43,A51,A53,A54,A61,A64,A65,A71,A74,A75,A76,
& A81,A84,A85,A86,A87,A91,A94,A95,A96,A97,A98,A101,A104,A105,A106,
& A107,A108,A109,A111,A114,A115,A116,A117,A118,A119,A110,A121,
& A124,A125,A126,A127,A128,A129,A1210,A1211,A131,A134,A135,A136,
& A137,A138,A139,A1310,A1311,B1,B6,B7,B8,B9,B10,B11,B12,B13,
& BH1,BH6,BH7,BH8,BH9,BH10,BH11,BH12
C2=1.D0/18.D0
C3=1.D0/12.D0
C4=1.D0/8.D0
C5=5.D0/16.D0
C6=3.D0/8.D0
C7=59.D0/400.D0
C8=93.D0/200.D0
C9=5490023248.D0/9719169821.D0
C10=13.D0/20.D0
C11=1201146811.D0/1299019798.D0
C12=1.D0
C13=1.D0
A21=C2
A31=1.D0/48.D0
A32=1.D0/16.D0
A41=1.D0/32.D0
A43=3.D0/32.D0
A51=5.D0/16.D0
A53=-75.D0/64.D0
A54=-A53
A61=3.D0/80.D0
A64=3.D0/16.D0
A65=3.D0/20.D0
A71=29443841.D0/614563906.D0
A74=77736538.D0/692538347.D0
A75=-28693883.D0/1125.D6
A76=23124283.D0/18.D8
A81=16016141.D0/946692911.D0
A84=61564180.D0/158732637.D0
A85=22789713.D0/633445777.D0
A86=545815736.D0/2771057229.D0
A87=-180193667.D0/1043307555.D0
A91=39632708.D0/573591083.D0
A94=-433636366.D0/683701615.D0
A95=-421739975.D0/2616292301.D0
A96=100302831.D0/723423059.D0
A97=790204164.D0/839813087.D0
A98=800635310.D0/3783071287.D0
A101=246121993.D0/1340847787.D0
A104=-37696042795.D0/15268766246.D0
A105=-309121744.D0/1061227803.D0

```

```

A106=-12992083.D0/490766935.D0
A107=6005943493.D0/2108947869.D0
A108=393006217.D0/1396673457.D0
A109=123872331.D0/1001029789.D0
A111=-1028468189.D0/846180014.D0
A114=8478235783.D0/508512852.D0
A115=1311729495.D0/1432422823.D0
A116=-10304129995.D0/1701304382.D0
A117=-48777925059.D0/3047939560.D0
A118=15336726248.D0/1032824649.D0
A119=-45442868181.D0/3398467696.D0
A110=3065993473.D0/597172653.D0
A121=185892177.D0/718116043.D0
A124=-3185094517.D0/667107341.D0
A125=-477755414.D0/1098053517.D0
A126=-703635378.D0/230739211.D0
A127=5731566787.D0/1027545527.D0
A128=5232866602.D0/850066563.D0
A129=-4093664535.D0/808688257.D0
A1210=3962137247.D0/1805957418.D0
A1211=65686358.D0/487910083.D0
A131=403863854.D0/491063109.D0
A134=-5068492393.D0/434740067.D0
A135=-411421997.D0/543043805.D0
A136=652783627.D0/914296604.D0
A137=11173962825.D0/925320556.D0
A138=-13158990841.D0/6184727034.D0
A139=3936647629.D0/1978049680.D0
A1310=-160528059.D0/685178525.D0
A1311=248638103.D0/1413531060.D0
B1=14005451.D0/335480064.D0
B6=-59238493.D0/1068277825.D0
B7=181606767.D0/758867731.D0
B8=561292985.D0/797845732.D0
B9=-1041891430.D0/1371343529.D0
B10=760417239.D0/1151165299.D0
B11=118820643.D0/751138087.D0
B12=-528747749.D0/2220607170.D0
B13=1.D0/4.D0
BH1=13451932.D0/455176623.D0
BH6=-808719846.D0/976000145.D0
BH7=1757004468.D0/5645159321.D0
BH8=656045339.D0/265891186.D0
BH9=-3867574721.D0/1518517206.D0
BH10=465885868.D0/322736535.D0
BH11=53011238.D0/667516719.D0
BH12=2.D0/45.D0
RETURN
END
C
SUBROUTINE SOLOUT(NSTEP,X,Y,N)
IMPLICIT DOUBLE PRECISION (A-H,O-Z)
DIMENSION Y(N)
RETURN
END

```

### 3. Одекс

Программа основана на методе экстраполяции (см. разд. II.9) с переменным порядком и переменной длиной шага. Хороша при всех уровнях точности, превосходна для очень высокой точности (например,  $10^{-20}$  или  $10^{-30}$ ).

```

SUBROUTINE ODEX (N,FCN,X,Y,XEND,EPS,HMAX,H)
C -----
C      NUMERICAL SOLUTION OF A SYSTEM OF FIRST ORDER
C      ORDINARY DIFFERENTIAL EQUATIONS Y'=F(X,Y).
C      THIS IS AN EXTRAPOLATION-ALGORITHM, BASED ON THE
C      EXPLICIT MIDPOINT RULE (WITH STEPSIZE CONTROL
C      AND ORDER SELECTION).
C      C.F. SECTION II.9
C
C      INPUT PARAMETERS
C -----
C      N           DIMENSION OF THE SYSTEM (N.LE.51)
C      FCN         NAME (EXTERNAL) OF SUBROUTINE COMPUTING THE
C                  FIRST DERIVATIVE F(X,Y):
C                  SUBROUTINE FCN(N,X,Y,F)
C                  REAL*8 X,Y(N),F(N)
C                  F(1)=... ETC.
C      X           INITIAL X-VALUE
C      XEND        FINAL X-VALUE (XEND.GT.X)
C      Y(N)        INITIAL VALUES FOR Y
C      EPS         LOCAL TOLERANCE
C      HMAX        MAXIMAL STEPSIZE
C      H           INITIAL STEPSIZE GUESS      PS → C ???
C
C      OUTPUT PARAMETERS
C -----
C      Y(N)        SOLUTION AT XEND          ER → ER LOC
C
C      EXTERNAL SUBROUTINE (TO BE SUPPLIED BY THE USER)
C -----
C      SOLOUT       THIS SUBROUTINE IS CALLED AFTER EVERY
C                  SUCCESSFULLY COMPUTED STEP (AND THE
C                  INITIAL VALUE):
C                  SUBROUTINE SOLOUT (NR,X,Y,N)
C                  REAL*8 X,Y(N)
C
C                  FURNISHES THE SOLUTION Y AT THE NR-TH
C                  GRID-POINT, X (THE INITIAL VALUE IS CON-
C                  SIDERED AS THE FIRST GRID-POINT).
C
C                  SUPPLY A DUMMY SUBROUTINE, IF THE SOLUTION
C                  IS NOT DESIRED AT THE INTERMEDIATE POINTS.
C
C      IMPLICIT REAL*8 (A-H,O-Z)
C      LOGICAL REJECT, LAST
C      REAL*8 Y(N)
C      EXTERNAL FCN
C      COMMON /STAT/NFCN,NSTEP,NACCP, NREJCT
C      --- COMMON STAT CAN BE USED FOR STATISTICS
C      --- NFCN      NUMBER OF FUNCTION EVALUATIONS
C      --- NSTEP     NUMBER OF COMPUTED STEPS
C      --- NACCP     NUMBER OF ACCEPTED STEPS
C      --- NREJCT    NUMBER OF REJECTED STEPS
C      COMMON /EXTABL/ DZ(51),T(9,51),NJ(9),HH(9),W(9),ERR,FAC,
1      A(9),EPSD4,UROUND,FAC1,FAC2,SAFE2
      DATA NJ/2,4,6,8,10,12,14,16,18/
      DATA A/3.D0,7.D0,13.D0,21.D0,31.D0,43.D0,57.D0,73.D0,91.D0/
      DATA NMAX/800/,KM/9/.UROUND/1.73D-18/
C      --- NMAX      MAXIMAL NUMBER OF STEPS
C      --- UROUND   SMALLEST NUMBER SATISFYING 1.D0+UROUND>1.D0

```

```

C ---          (TO BE ADAPTED BY THE USER)
      DATA FAC1/2.D-2/, FAC2/4.D0/, FAC3/.9D0/, FAC4/.8D0/
      DATA SAFR1/.65D0/, SAFR2/.94D0/
C --- INITIAL PREPARATIONS
      EPSD4=EPS*SAFR1
      NSTEP=0
      NREJECT=0
      NACCPt=0
      NFCN=0
      K=MAX0(3,MIN0(8,INT(-DLOG10(EPS)*.6D0+1.5D0)))
      H=DMIN1(H,HMAX,(XEND-X)/2.D0)
      CALL SOLOUT (NACCPt+1,X,Y,N)
      ERR=0.D0
      W(1)=0.D0
      REJECT=.FALSE.
      LAST=.FALSE.

C --- IS XEND REACHED IN THE NEXT STEP?
      10   H1=XEND-X
           IF (H1.LE.UROUND) GO TO 110
           H=DMIN1(H,H1,HMAX)
           IF (H.GE.H1-UROUND) LAST=.TRUE.
           CALL FCN(N,X,Y,DZ)
           NFCN=NFCN+1

C --- THE FIRST AND LAST STEP
      IF (NSTEP.EQ.0.OR.LAST) THEN
          NSTEP=NSTEP+1
          DO 20 J=1,K
              KC=J
              CALL MIDEX(J,X,Y,H,HMAX,N,FCN)
              IF (J.GT.1.AND.ERR.LE.EPS) GO TO 60
              GO TO 55
          END IF

C --- BASIC INTEGRATION STEP
      30   CONTINUE
          NSTEP=NSTEP+1
          IF (NSTEP.GE.NMAX) GO TO 120
          KC=K-1
          DO 40 J=1,KC
              CALL MIDEX(J,X,Y,H,HMAX,N,FCN)

C --- CONVERGENCE MONITOR
          IF (K.EQ.2.OR.REJECT) GO TO 50
          IF (ERR.LE.EPS) GO TO 60
          IF (ERR/EPS.GT.(DFLOAT(NJ(K+1)*NJ(K))/4.D0)**2) GO TO 100
      50   CALL MIDEX(K,X,Y,H,HMAX,N,FCN)
          KC=K
          IF (ERR.LE.EPS) GO TO 60

C --- HOPE FOR CONVERGENCE IN LINE K+1
      55   IF (ERR/EPS.GT.(DFLOAT(NJ(K+1))/2.D0)**2) GO TO 100
          KC=K+1
          CALL MIDEX(KC,X,Y,H,HMAX,N,FCN)
          IF (ERR.GT.EPS) GO TO 100

C --- STEP IS ACCEPTED
      60   X=X+H
          DO 70 I=1,N
              Y(I)=T(I,I)
              NACCPt=NACCPt+1
              CALL SOLOUT (NACCPt+1,X,Y,N)

C --- COMPUTE OPTIMAL ORDER
          IF (KC.EQ.2) THEN
              KOPT=3
              IF (REJECT) KOPT=2
              GO TO 80
          END IF
          IF (KC.LE.K) THEN
              KOPT=KC
              IF (W(KC-1).LT.W(KC)*FAC3) KOPT=KC-1
              IF (W(KC).LT.W(KC-1)*FAC3) KOPT=MIN0(KC+1,KM-1)
          END IF
      
```

```

ELSE
  KOPT=KC-1
  IF (KC.GT.3.AND.W(KC-2).LT.W(KC-1)*FAC3) KOPT=KC-2
  IF (W(KC).LT.W(KOPT)*FAC3) KOPT=MIN0(KC,KM-1)
END IF
C --- AFTER A REJECTED STEP
80  IF (REJECT) THEN
    K=MIN0(KOPT,KC)
    H=DMIN1(H,HH(K))
    REJECT=.FALSE.
    GO TO 10
  END IF
C --- COMPUTE STEPSIZE FOR NEXT STEP
  IF (KOPT.LE.KC) THEN
    H=HH(KOPT)
  ELSE
    IF (KC.LT.K.AND.W(KC).LT.W(KC-1)*FAC4) THEN
      H=HH(KC)*A(KOPT+1)/A(KC)
    ELSE
      H=HH(KC)*A(KOPT)/A(KC)
    END IF
  END IF
  K=KOPT
  GO TO 10
C --- STEP IS REJECTED
100 K=MIN0(K,KC)
  IF (K.GT.2.AND.W(K-1).LT.W(K)*FAC3) K=K-1
  NREJCT=NREJCT+1
  H=HH(K)
  REJECT=.TRUE.
  GO TO 30
C --- SOLUTION EXIT
110 CONTINUE
  RETURN
C --- FAIL EXIT
120 WRITE (6,*)
     ' MORE THAN ',NMAX,' STEPS '
  RETURN
END
C
  SUBROUTINE MIDEX(J,X,Y,H,HMAX,N,FCN)
C --- THIS SUBROUTINE COMPUTES THE J-TH LINE OF THE
C --- EXTRAPOLATION TABLE AND PROVIDES AN ESTIMATION
C --- OF THE OPTIMAL STEPSIZE
  IMPLICIT REAL*8 (A-H,O-Z)
  EXTERNAL FCN
  REAL*8 Y(N),DY(51),YH1(51),YH2(51)
  COMMON /STAT/NFCN,NSTEP,NACCP,NREJCT
  COMMON /EXTABL/ DZ(51),T(9,51),NJ(9),HH(9),W(9),ERR,FAC,
  1   A(9),EPSD4,UROUND,FAC1,FAC2,SAFE2
  HJ=H/DFLOAT(NJ(J))
C --- EULER STARTING STEP
  DO 30 I=1,N
    YH1(I)=Y(I)
  30  YH2(I)=Y(I)+HJ*DZ(I)
C --- EXPLICIT MIDPOINT RULE
  M=NJ(J)-1
  DO 35 MM=1,M
    CALL FCN(N,X+HJ*DFLOAT(MM),YH2,DY)
  35  DO 36 I=1,N
    YS=YH1(I)
    YH1(I)=YH2(I)
    YH2(I)=YS+2.D0*HJ*DY(I)
  36  CALL FCN(N,X+H,YH2,DY)
  DO 40 I=1,N
    T(J,I)=(YH1(I)+YH2(I)+HJ*DY(I))/2.D0
  40  NFCN=NFCN+NJ(J)

```

```

C --- POLYNOMIAL EXTRAPOLATION
IF (J.EQ.1) RETURN
DO 60 L=J,2,-1
  FAC=(DFLOAT(NJ(J))/DFLOAT(NJ(L-1)))**2-1.D0
DO 60 I=1,N
  T(L-1,I)=T(L,I)+(T(L,I)-T(L-1,I))/FAC
60  CONTINUE
ERR=0.D0
DO 65 I=1,N
C --- SCALING
  SCAL=DMAX1(DABS(Y(I)),DABS(T(1,I)),1.D-6,UROUND/EPSD4)
65  ERR=ERR+((T(1,I)-T(2,I))/SCAL)**2
  ERR=DSQRT(ERR/DFLOAT(N))
C --- COMPUTE OPTIMAL STEPSIZES
  EXP0=1.D0/DFLOAT(2*j-1)
  FACMIN=FAC1**EXP0
  FAC=DMIN1(FAC2/FACMIN,DMAX1(FACMIN,(ERR/EPSD4)**EXP0/SAFE2))
  FAC=1.D0/FAC
  HH(J)=DMIN1(H*FAC,HMAX)
  W(J)=A(J)/HH(J)
  RETURN
END
C
SUBROUTINE SOLOUT (NRPNTS,X,Y,N)
IMPLICIT REAL*8 (A-H,O-Z)
REAL*8 Y(N)
RETURN
END

```

#### 4. Odex2

Программа основана на методе экстраполяции (см. разд. II.13) с переменным порядком и переменной длиной шага. Предназначена для интегрирования дифференциальных систем второго порядка вида  $y'' = f(x, y)$ . Хороша для всех уровней точности.

```

SUBROUTINE ODEX2 (N,FCN,X,Y,YP,XEND,EPS,HMAX,H)
C -----
C      NUMERICAL SOLUTION OF A SYSTEM OF SECOND ORDER
C      ORDINARY DIFFERENTIAL EQUATIONS  Y''=F(X,Y).
C      THIS IS AN EXTRAPOLATION-ALGORITHM, BASED ON THE
C      EXPLICIT MIDPOINT RULE (WITH STEPSIZE CONTROL
C      AND ORDER SELECTION).
C      C.F. SECTION II.13
C
C      INPUT PARAMETERS
C -----
C      N          DIMENSION OF THE SYSTEM (N.LE.51)
C      FCN        NAME (EXTERNAL) OF SUBROUTINE COMPUTING THE
C                  SECOND DERIVATIVE F(X,Y):
C                  SUBROUTINE FCN(N,X,Y,F)
C                  REAL*8 X,Y(N),F(N)
C                  F(1)=... ETC.
C      X          INITIAL X-VALUR
C      XEND       FINAL X-VALUE (XEND.GT.X)
C      Y(N)       INITIAL VALUES FOR Y
C      YP(N)      INITIAL VALUES FOR Y'
C      EPS        LOCAL TOLERANCE
C      HMAX       MAXIMAL STEPSIZE
C      H          INITIAL STEPSIZE GUESS
C
C      OUTPUT PARAMETERS
C -----
C      Y(N)       SOLUTION AT XEND
C      YP(N)      DERIVATIVE OF SOLUTION AT XEND
C

```

```

C      EXTERNAL SUBROUTINE (TO BE SUPPLIED BY THE USER)
C -----
C      SOLUT2      THIS SUBROUTINE IS CALLED AFTER EVERY
C                  SUCCESSFULLY COMPUTED STEP (AND THE
C                  INITIAL VALUE):
C                  SUBROUTINE SOLUT2 (NR,X,Y,YP,N)
C                  REAL*8 X,Y(N),YP(N)
C                  FURNISHES THE SOLUTION (Y,YP) AT THE NR-TH
C                  GRID-POINT X (THE INITIAL VALUE IS CON-
C                  SIDERED AS THE FIRST GRID-POINT).
C                  SUPPLY A DUMMY SUBROUTINE, IF THE SOLUTION
C                  IS NOT DESIRED AT THE INTERMEDIATE POINTS.
C -----
C      IMPLICIT REAL*8 (A-H,O-Z)
C      LOGICAL REJECT,LAST
C      REAL*8 Y(N),YP(N)
C      EXTERNAL FCN
C      COMMON /STAT/NFCN,NSTEP,NACCPT,NREJCT
C      --- COMMON STAT CAN BE USED FOR STATISTICS
C      --- NFCN      NUMBER OF FUNCTION EVALUATIONS
C      --- NSTEP     NUMBER OF COMPUTED STEPS
C      --- NACCPT   NUMBER OF ACCEPTED STEPS
C      --- NREJCT    NUMBER OF REJECTED STEPS
C      COMMON /EXTABL/DZ(51),T(9,51),TP(9,51),NJ(9),HH(9),W(9),ERR,FAC,
1       A(9),EPSD4,UROUND,FAC1,FAC2,SAFE2
        DATA NJ/2,4,6,8,10,12,14,16,18/
        DATA A/2.D0,4.D0,7.D0,11.D0,16.D0,22.D0,29.D0,37.D0,46.D0/
        DATA NMAX/800/,KM/9/,UROUND/1.73D-18/
C      --- NMAX      MAXIMAL NUMBER OF STEPS
C      --- UROUND   SMALLEST NUMBER SATISFYING 1.D0+UROUND>1.D0
C      ---          (TO BE ADAPTED BY THE USER)
        DATA FAC1/2.D-2/,FAC2/4.D0/,FAC3/.9D0/,FAC4/.8D0/
        DATA SAFE1/.65D0/,SAFE2/.94D0/
C      --- INITIAL PREPARATIONS
        EPSD4=EPS*SAFE1
        NSTEP=0
        NREJCT=0
        NACCPT=0
        NFCN=0
        K=MAX0(3,MIN0(8,INT(-DLOG10(EPS)*.6D0+1.5D0)))
        H=DMIN1(H,HMAX,(XEND-X)/2.D0)
        CALL SOLUT2 (NACCPT+1,X,Y,YP,N)
        ERR=0.D0
        W(1)=0.D0
        REJECT=.FALSE.
        LAST=.FALSE.
C      --- IS XEND REACHED IN THE NEXT STEP?
10      H1=XEND-X
        IF (H1.LE.UROUND) GO TO 110
        H=DMIN1(H,H1,HMAX)
        IF (H.GE.H1-UROUND) LAST=.TRUE.
        CALL FCN(N,X,Y,DZ)
        NFCN=NFCN+1
C      --- THE FIRST AND LAST STEP
        IF (NSTEP.EQ.0.OR.LAST) THEN
          NSTEP=NSTEP+1
          DO 20 J=1,K
            KC=J
            CALL STORM(J,X,Y,YP,H,HMAX,N,FCN)
20          IF (J.GT.1.AND.ERR.LE.EPS) GO TO 60
            GO TO 55
        END IF
C      --- BASIC INTEGRATION STEP
30      CONTINUE
        NSTEP=NSTEP+1
        IF (NSTEP.GE.NMAX) GO TO 120
        KC=K-1
        DO 40 J=1,KC

```

```

40    CALL STOERM(J,X,Y,YP,H,HMAX,N,FCN)
C --- CONVERGENCE MONITOR
      IF (K.EQ.2.OR.REJECT) GO TO 50
      IF (ERR.LE.EPS) GO TO 60
      IF (ERR/EPS.GT.(DFLOAT(NJ(K+1)*NJ(K))/4.D0)**2).GO TO 100
\ 50   CALL STOERM(K,X,Y,YP,H,HMAX,N,FCN)
      KC=K
      IF (ERR.LE.EPS) GO TO 60
C --- HOPE FOR CONVERGENCE IN LINE K+1
      55   IF (ERR/EPS.GT.(DFLOAT(NJ(K+1))/2.D0)**2) GO TO 100
      KC=K+1
      CALL STOERM(KC,X,Y,YP,H,HMAX,N,FCN)
      IF (ERR.GT.EPS) GO TO 100
C --- STEP IS ACCEPTED
      60   X=X+H
          DO 70 I=1,N
          YP(I)=TP(1,I)
      70   Y(I)=T(1,I)
          NACOPT=NACOPT+1
          CALL SOLUT2 (NACOPT+1,X,Y,YP,N)
C --- COMPUTE OPTIMAL ORDER
      IF (KC.EQ.2) THEN
          KOPT=3
          IF (REJECT) KOPT=2
          GO TO 80
      END IF
      IF (KC.LE.K) THEN
          KOPT=KC
          IF (W(KC-1).LT.W(KC)*FAC3) KOPT=KC-1
          IF (W(KC).LT.W(KC-1)*FAC3) KOPT=MIN0(KC+1,KM-1)
      ELSE
          KOPT=KC-1
          IF (KC.GT.3.AND.W(KC-2).LT.W(KC-1)*FAC3) KOPT=KC-2
          IF (W(KC).LT.W(KOPT)*FAC3) KOPT=MIN0(KC,KM-1)
      END IF
C --- AFTER A REJECTED STEP
      80   IF (REJECT) THEN
          K=MIN0(KOPT,KC)
          H=DMIN1(H,HH(K))
          REJECT=.FALSE.
          GO TO 10
      END IF
C --- COMPUTE STEPSIZE FOR NEXT STEP
      IF (KOPT.LE.KC) THEN
          H=HH(KOPT)
      ELSE
          IF (KC.LT.K.AND.W(KC).LT.W(KC-1)*FAC4) THEN
              H=HH(KC)*A(KOPT+1)/A(KC)
          ELSE
              H=HH(KC)*A(KOPT)/A(KC)
          END IF
      END IF
      K=KOPT
      GO TO 10
C --- STEP IS REJECTED
      100  K=MIN0(K,KC)
          IF (K.GT.2.AND.W(K-1).LT.W(K)*FAC3) K=K-1
          NREJCT=NREJCT+1
          H=HH(K)
          REJECT=.TRUE.
          GO TO 30
C --- SOLUTION EXIT
      110  CONTINUE
          RETURN
C --- FAIL EXIT
      120  WRITE (6,*), ' MORE THAN ',NMAX,' STEPS '
          RETURN
      END
C

```

```

SUBROUTINE STOERM(J,X,Y,YP,H,HMAX,N,FCN)
C --- THIS SUBROUTINE COMPUTES THE J-TH LINE OF THE
C --- EXTRAPOLATION TABLE AND PROVIDES AN ESTIMATION
C --- OF THE OPTIMAL STEPSIZE
      IMPLICIT REAL*8 (A-H,O-Z)
      EXTERNAL FCN
      REAL*8 Y(N),YP(N),DY(51),YH1(51),ZH1(51)
      COMMON /STAT/NFCN,NSTEP,NACCPT,NREJCT
      COMMON /EXTABL/ DZ(51),T(9,51),TP(9,51),NJ(9),HH(9),W(9),ERR,FAC,
      1   A(9),EPSD4,UROUND,FAC1,FAC2,SAFE2
      HJ=H/DFLOAT(NJ(J))
      HJ2=HJ*2.D0
C --- EULER STARTING STEP
      DO 30 I=1,N
         YH1(I)=Y(I)
      30  ZH1(I)=YP(I)+HJ*DZ(I)
C --- EXPLICIT MIDPOINT (STOERMR) RULE
      M=NJ(J)/2-1
      IF (J.EQ.1) GO TO 37
      DO 35 MM=1,M
      DO 33 I=1,N
      33  YH1(I)=YH1(I)+HJ2*ZH1(I)
      CALL FCN(N,X+HJ2*DFLOAT(MM),YH1,DY)
      DO 35 I=1,N
      35  ZH1(I)=ZH1(I)+HJ2*DY(I)
C --- FINAL STEP
      37  CONTINUE
      DO 40 I=1,N
      40  YH1(I)=YH1(I)+HJ2*ZH1(I)
      CALL FCN(N,X+H,YH1,DY)
      DO 43 I=1,N
         T(J,I)=YH1(I)
      43  TP(J,I)=ZH1(I)+HJ*DY(I)
      NFCN=NFCN+M+1
C --- POLYNOMIAL EXTRAPOLATION
      IF (J.EQ.1) RETURN
      DO 60 L=J,2,-1
         FAC=(DFLOAT(NJ(J))/DFLOAT(NJ(L-1)))**2-1.D0
      DO 60 I=1,N
         T(L-1,I)=T(L,I)+(T(L,I)-T(L-1,I))/FAC
         TP(L-1,I)=TP(L,I)+(TP(L,I)-TP(L-1,I))/FAC
      60  CONTINUE
      ERR=0.D0
      DO 65 I=1,N
      65  SCAL=DMAX1(DABS(Y(I)),DABS(T(1,I))),1.D-6,UROUND/EPSD4)
      SCALP=DMAX1(DABS(YP(I)),DABS(TP(1,I))),1.D-6,UROUND/EPSD4)
      65  ERR=ERR+((T(1,I)-T(2,I))/SCAL)**2+((TP(1,I)-TP(2,I))/SCALP)**2
      ERR=DSQRT(ERR/DFLOAT(N**2))
C --- COMPUTE OPTIMAL STEPSIZES
      EXP0=1.D0/DFLOAT(2*J-1)
      FACMIN=FAC1**RXP0
      FAC=DMIN1(FAC2/FACMIN,DMAX1(FACMIN,(ERR/EPSD4)**EXP0/SAFE2))
      FAC=1.D0/FAC
      HH(J)=DMIN1(H*FAC,HMAX)
      W(J)=A(J)/HH(J)
      RETURN
      END
C
      SUBROUTINE SOLUT2 (NRPNTS,X,Y,YP,N)
      IMPLICIT REAL*8 (A-H,O-Z)
      REAL*8 Y(N),YP(N)
      RETURN
      END

```

## 5. Doprin

Программа явного метода Рунге—Кутты—Нюстрёма, основанная на формулах Дормана и Принса с автоматическим управлением длиной шага. Предназначена для интегрирования (см. табл. 13.4 из разд. 2.13) дифференциальных систем вида  $y'' = f(x, y)$ .

```

SUBROUTINE DOPRIN(N,FCN,X,Y,YP,XEND,EPS,HMAX,H)
C -----
C NUMERICAL SOLUTION OF A SYSTEM OF SECOND ORDER
C ORDINARY DIFFERENTIAL EQUATIONS Y''=F(X,Y).
C THIS IS AN EMBEDDED NYSTROEM METHOD OF ORDER (6)7
C DUE TO DORMAND & PRINCE (WITH STEPSIZE CONTROL)
C C.F. SECTION II.13
C
C INPUT PARAMETERS
C -----
C N           DIMENSION OF THE SYSTEM (N.LE.51)
C FCN         NAME (EXTERNAL) OF SUBROUTINE COMPUTING THE
C             SECOND DERIVATIVE F(X,Y):
C             SUBROUTINE FCN(N,X,Y,F)
C             REAL*8 X,Y(N),F(N)
C             F(1)=... ETC.
C X           INITIAL X-VALUE
C XEND        FINAL X-VALUE (XEND.GT.X)
C Y(N)        INITIAL VALUES FOR Y
C YP(N)       INITIAL VALUES FOR Y'
C EPS         LOCAL TOLERANCE
C HMAX        MAXIMAL STEPSIZE
C H           INITIAL STEPSIZE GUESS
C
C OUTPUT PARAMETERS
C -----
C Y(N)        SOLUTION AT XEND
C YP(N)       DERIVATIVE OF SOLUTION AT XEND
C
C EXTERNAL SUBROUTINE (TO BE SUPPLIED BY THE USER)
C -----
C SOLUT2      THIS SUBROUTINE IS CALLED AFTER EVERY
C             SUCCESSFULLY COMPUTED STEP (AND THE
C             INITIAL VALUE):
C             SUBROUTINE SOLUT2 (NR,X,Y,YP,N)
C             REAL*8 X,Y(N),YP(N)
C             FURNISHES THE SOLUTION (Y,YP) AT THE NR-TH
C             GRID-POINT X (THE INITIAL VALUE IS CON-
C             SIDERED AS THE FIRST GRID-POINT).
C             SUPPLY A DUMMY SUBROUTINE, IF THE SOLUTION
C             IS NOT DESIRED AT THE INTERMEDIATE POINTS.
C
C -----
C IMPLICIT REAL*8 (A-H,O-Z)
C REAL*8 K0(51),K1(51),K2(51),K3(51),K4(51),Y1(51),Y(N),YP(N)
C LOGICAL REJECT
C COMMON/STAT/NFCN,NSTEP,NACCPt,NREJCT
C --- COMMON STAT CAN BE USED FOR STATISTICS
C --- NFCN        NUMBER OF FUNCTION EVALUATIONS
C --- NSTEP       NUMBER OF COMPUTED STEPS
C --- NACCPt     NUMBER OF ACCEPTED STEPS
C --- NREJCT     NUMBER OF REJECTED STEPS
C DATA NMAX/2000/, UROUND/1.73D-18/
C --- NMAX        MAXIMAL NUMBER OF STEPS
C --- UROUND     SMALLEST NUMBER SATISFYING 1.D0+UROUND>1.D0
C --- (TO BE ADAPTED BY THE USER)
C     ALPH1=.1D0
C     ALPH2=.2D0

```

```

ALPH3=3.D0/8.D0      "
ALPH4=.5D0
SQ21=DSQRT(21.D0)
ALPH5=(7.D0-SQ21)/14.D0
ALPH6=(7.D0+SQ21)/14.D0
C0=.05D0
C4=16.D0/45.D0
C5=49.D0/180.D0
G10=1.D0/200.D0
G20=1.D0/150.D0
G21=1.D0/75.D0
G30=171.D0/8192.D0
G31=45.D0/4096.D0
G32=315.D0/8192.D0
G40=5.D0/288.D0
G41=25.D0/528.D0
G42=25.D0/672.D0
G43=16.D0/693.D0
G50=(1003.D0-205.D0*SQ21)/12348.D0
G51=(-18775.D0+4325.D0*SQ21)/90552.D0
G52=(15600.D0-3425.D0*SQ21)/43218.D0
G53=(-46208.D0+10112.D0*SQ21)/237699.D0
G54=(3411.D0-745.D0*SQ21)/24696.D0
G60=(793.D0-187.D0*SQ21)/12348.D0
G61=(-8275.D0-2825.D0*SQ21)/90552.D0
G62=(26100.D0+6175.D0*SQ21)/43218.D0
G63=-(1905280.D0+483712.D0*SQ21)/9745659.D0
G64=(3327.D0+797.D0*SQ21)/24696.D0
G65=-(581.D0+127.D0*SQ21)/1722.D0
G70=(-157.D0+3.D0*SQ21)/378.D0
G71=(3575.D0-250.D0*SQ21)/2772.D0
G72=-(21900.D0+1375.D0*SQ21)/3969.D0
G73=(1168640.D0+23040.D0*SQ21)/596673.D0
G74=(-1353.D0+26.D0*SQ21)/2268.D0
G75=(12439.D0+2639.D0*SQ21)/4428.D0
G76=(36.D0-7.D0*SQ21)/36.D0
G80=.05D0
G84=8.D0/45.D0
G85=(49.D0+7.D0*SQ21)/360.D0
G86=(49.D0-7.D0*SQ21)/360.D0
POSNEG=DSIGN(1.D0,XEND-X)
C --- INITIAL PREPARATIONS
HMAX=DABS(HMAX)
H=DMAX1(DMAX1(1.D-8,DABS(H)),HMAX)
H=DSIGN(H,POSNEG)
EPS=DMAX1(EPS,9.D0*UROUND)
REJECT=.FALSE.
NACCPT=0
NRJCT=0
NFCN=1
NSTEP=0
CALL SOLUT2(NACCPT+1,X,Y,YP,N)
CALL FCN(N,X,Y,K0)
C ----- BASIC INTEGRATION STEP -----
1 IF(NSTEP.GT.NMAX.OR.X+.05D0*H.EQ.X)GOTO 79
IF((X-XEND)*POSNEG+UROUND.GT.0.D0) RETURN
IF((X+H-XEND)*POSNEG.GT.0.D0)H=XEND-X
HP2=H*H
NSTEP=NSTEP+1
C --- THE FIRST 5 STAGES
DO 21 I=1,N
21 Y1(I)=Y(I)+ALPH1*H*YP(I)+HP2*G10*K0(I)
CALL FCN(N,X+ALPH1*H,Y1,K1)
DO 22 I=1,N
22 Y1(I)=Y(I)+ALPH2*H*YP(I)+HP2*(G20*K0(I)+G21*K1(I))
CALL FCN(N,X+ALPH2*H,Y1,K2)
DO 23 I=1,N

```

```

23   Y1(I)=Y(I)+ALPH3*H*YP(I)+HP2*(G30*K0(I)+G31*K1(I)+G32*K2(I))
      CALL FCN(N,X+ALPH3*H,Y1,K3)
      DO 24 I=1,N
24   Y1(I)=Y(I)+ALPH4*H*YP(I)+HP2*(G40*K0(I)+G41*K1(I)+G42*K2(I)-
     & G43*K3(I))
      CALL FCN(N,X+ALPH4*H,Y1,K4)
      DO 25 I=1,N
25   Y1(I)=Y(I)+ALPH5*H*YP(I)+HP2*(G50*K0(I)+G51*K1(I)+G52*K2(I)-
     & G53*K3(I)+G54*K4(I))
C --- COMPUTE INTERMEDIATE SUM TO SAVE MEMORY
      DO 61 I=1,N
         Y1S=G60*K0(I)+G61*K1(I)+G62*K2(I)+G63*K3(I)+G64*K4(I)
         K3(I)=G70*K0(I)+G71*K1(I)+G72*K2(I)+G73*K3(I)+G74*K4(I)
61   K2(I)=Y1S
C --- THE LAST 4 STAGES
      CALL FCN(N,X+ALPH5*H,Y1,K1)
      DO 26 I=1,N
26   Y1(I)=Y(I)+ALPH6*H*YP(I)+HP2*(K2(I)+G65*K1(I))
      CALL FCN(N,X+ALPH6*H,Y1,K2)
      DO 27 I=1,N
27   Y1(I)=Y(I)+H*YP(I)+HP2*(K3(I)+G75*K1(I)+G76*K2(I))
      XPH=X+H
      CALL FCN(N,XPH,Y1,K3)
      DO 28 I=1,N
28   Y1(I)=Y(I)+H*YP(I)+HP2*(G80*K0(I)+G84*K4(I)+G85*K1(I)+G86*K2(I))
      DO 30 I=1,N
30   K4(I)=YP(I)+H*(C0*(K0(I)+K3(I))+C4*K4(I)+C5*(K1(I)+K2(I)))
      CALL FCN(N,XPH,Y1,K1)
      DO 33 I=1,N
33   K2(I)=HP2*(K1(I)-K3(I))/20.D0
      NFCN=NFCN+8
C --- ERROR ESTIMATION
      ERR=0.D0
      DO 41 I=1,N
        DENOM=DMAX1(1.D-6,DABS(Y1(I)),DABS(Y(I)),2.D0*UROUND/EPS)
41   ERR=ERR+(K2(I)/DENOM)**2
      ERR=DSQRT(ERR/DFLOAT(N))
C --- COMPUTATION OF HNEW
C --- WE REQUIRE .2<=HNEW/H<=10.
      FAC=DMAX1(.1D0,DMIN1(5.D0,(ERR/EPS)**(1.D0/7.D0)/.9D0))
      HNEW=H/FAC
      IF(ERR.LE.EPS)THEN
C --- STEP IS ACCEPTED
        NACCPt=NACCPt+1
        DO 44 I=1,N
          YP(I)=K4(I)
          K0(I)=K1(I)
44       Y(I)=Y1(I)
          X=XPH
          CALL SOLUT2(NACCPt+1,X,Y,YP,N)
          IF(DABS(HNEW).GT.HMAX)HNEW=POSNEG*HMAX
          IF(REJECT)HNEW=POSNEG*DMIN1(DABS(HNEW),DABS(H))
          REJECT=.FALSE.
        ELSE
C --- STEP IS REJECTED
          REJECT=.TRUE.
          IF(NACCPt.GE.1)NREJCT=NREJCT+1
        END IF
        H=HNEW
        GOTO 1
C --- FAIL EXIT
79   WRITE(6,979)X
979  FORMAT(' EXIT OF DOPRIN AT X=',D16.7)
      RETURN
    END

```

```

C
      SUBROUTINE SOLUT2 (NRPNTS,X,Y,YP,N)
      IMPLICIT REAL*8 (A-H,O-Z)
      REAL*8 Y(N),YP(N)
      RETURN
      END

```

## 6. Retard

Модификация программы DOPRI5 для дифференциальных уравнений с запаздывающим аргументом (см. разд. II.15). Включен пример головной программы, вызывающей RETARD. Подпрограммы STORE и YLAG также полезны для «плотного» вывода и построения графиков.

```

C ----- SAMPLE CALLING PROGRAM FOR RETARD -----
C      SOLVING PROBLEM (15.12) WITH SAME INITIAL CONDITIONS
C      AS FOR TABLE 15.1.
      DIMENSION Y(1)
      COMMON/POSITS/IFIRST,LAST,X0,XLAST,IPOS,DISC
      COMMON/STAT/NFCN,NSTEP,NACCPY,NREJCT
      EXTERNAL FCN
      X0=0.
      LAST=0
      H=0.5
      Y(1)=0.1
      DO 1 I=1,10
         X=FLOAT(I-1)
         XEND=FLOAT(I)
         EPS=1.E-6
         HMAX=1.
         CALL RETARD(1,FCN,X,Y,XEND,EPS,HMAX,H)
         WRITE(6,*)X,Y(1)
         WRITE(6,*)" COMMON STAT: ",NFCN,NSTEP,NACCPY,NREJCT
    1    CONTINUE
      STOP
      END

      SUBROUTINE FCN(N,X,Y,F)
      DIMENSION Y(N),F(N)
      EXTERNAL PHI
      A=1.4
      F(1)=(A-YLAG(1,X-1.,PHI))*Y(1)
      RETURN
      END

      REAL FUNCTION PHI(I,X)
      IF(I.EQ.1)PHI=0.
      RETURN
      END

      SUBROUTINE RETARD(N,FCN,X,Y,XEND,EPS,HMAX,H)
C -----
C      NUMERICAL SOLUTION OF A SYSTEM OF FIRST ORDER
C      RETARDED DIFFERENTIAL EQUATIONS Y'=F(X,Y(X),Y(X-TAU),...).
C      THIS IS BASED ON AN EMBEDDED RUNGE-KUTTA METHOD OF ORDER (4)5
C      DUE TO DORMAND & PRINCE (WITH STEPSIZE CONTROL).
C      C.F. SECTIONS II.5 AND II.15
C
C      INPUT PARAMETERS
C -----
C      N          DIMENSION OF THE SYSTEM (N.LE.51)
C      FCN        NAME (EXTERNAL) OF SUBROUTINE COMPUTING THE
C                  FIRST DERIVATIVE F(X,Y):
C                  SUBROUTINE FCN(N,X,Y,F)
C                  REAL*4 X,Y(N),F(N)
C                  EXTERNAL PHI

```

```

C           F(1)=-YLAG(1,X-1.,PHI)+...
C           F(2)=...      ETC.
C           WHERE "PHI" IS THE (EXTERNAL) NAME OF
C           A REAL FUNCTION COMPUTING THE
C           I-TH COMPONENT OF THE INITIAL FUNCTION PHI(X)
C           REAL FUNCTION PHI(I,X)
C           IF (I.EQ.1) PHI=...      ETC.
C           X           INITIAL X-VALUE
C           XEND        FINAL X-VALUE (XEND>X)
C           Y(N)        INITIAL VALUES FOR Y
C           EPS         LOCAL TOLERANCE
C           HMAX        MAXIMAL STEP SIZE
C           H           INITIAL STEP SIZE GUESS
C
C           OUTPUT PARAMETERS
C   -----
C           Y(N)        SOLUTION AT XEND
C   -----
C           REAL*4 K1(51),K2(51),K3(51),K4(51),K5(51),K6(51),K7(51)
C           REAL*4 Y(51),Y(N)
C           LOGICAL REJECT,DISC
C           COMMON/POSITS/IFIRST,LAST,X0,XLAST,IPOS,DISC
C   -----
C   ----- MEANING OF THESE VARIABLES:
C   ---    IFIRST     LOWEST STEP NUMBER STILL IN MEMORY COEF;
C   ---    LAST       ADDRESS OF LAST DATA WRITTEN BY STORE ON COMMON BLOCK COEF
C   ---          MUST BE SET TO 0 IN THE CALLING PROGRAM
C   ---          BEFORE THE FIRST CALL.
C   ---    X0         INITIAL POINT, MUST BE SET IN THE CALLING PROGRAM
C   ---          BEFORE THE FIRST CALL.
C   ---    XLAST      =X+H OF LAST WRITTEN STEP;
C   ---    IPOS        POSITION OF LAST SUCCESSFUL SEARCH IN FUNCTION YLAG;
C   ---    DISC        LOGICAL VARIABLE, NECESSARY FOR THE DISTINCTION
C   ---          OF K7 AND K1 OF THE FOLLOWING STEP IN THE CASE
C   ---          WHEN Y(X0) IS DIFFERENT FROM PHI(X0).
C           COMMON/STAT/NFCN,NSTEP,NACCPt,NREJCT
C   ----- COMMON STAT CONTAINS STATISTICAL INFORMATION:
C   ---    NFCN        NUMBER OF FUNCTION EVALUATIONS
C   ---    NSTEP       NUMBER OF COMPUTED STEPS
C   ---    NACCPt     NUMBER OF ACCEPTED STEPS
C   ---    NREJCT     NUMBER OF REJECTED STEPS
C           COMMON/UROUND/UROUND
C           UROUND     SMALLEST NUMBER SATISFYING 1.+UROUND>1.
C   ---          (TO BE ADAPTED BY THE USER)
C           DATA UROUND/6.E-8/
C           DATA NMAX/3000/
C   ---    NMAX        MAXIMAL NUMBER OF STEPS
C   ---    INITIAL PREPARATIONS
C           HMAX=ABS(HMAX)
C           H=AMIN1(AMAX1(1.E-4,ABS(H)),HMAX)
C           H=SIGN(H,1.)
C           EPS=AMAX1(EPS,7.*UROUND)
C           REJECT=.FALSE.
C           NACCPt=0
C           NREJCT=0
C           NFCN=1
C           NSTEP=0
C           DISC=.TRUE.
C           CALL FCN(N,X,Y,K1)
C           IF (.NOT.DISC) CALL FCN(N,X,Y,K1)
C   ---    BASIC INTEGRATION STEP
C   1    DISC=.TRUE.
C           IF(NSTEP.GT.NMAX.OR.X+.1*H.EQ.X)GOTO 79
C           IF((X-XEND)+UROUND.GT.0.) RETURN
C           IF((X+H-XEND).GT.0.)H=XEND-X
C           NSTEP=NSTEP+1
C   ---    THE 7 RUNGE-KUTTA STAGES
C           DO 22 I=1,N
C   22   Y1(I)=Y(I)+H*.2*K1(I)
C           CALL FCN(N,X+.2*H,Y1,K2)
C           DO 23 I=1,N
C   23   Y1(I)=Y(I)+H*((3./40.)*K1(I)+(9./40.)*K2(I))
C           CALL FCN(N,X+.3*H,Y1,K3)

```

```

DO 24 I=1,N
24 Y1(I)=Y(I)+H*((44./45.)*K1(I)-(56./15.)*K2(I)+(32./9.)*K3(I))
CALL FCN(N,X+.8*H,Y1,K4)
DO 25 I=1,N
25 Y1(I)=Y(I)+H*((19372./6561.)*K1(I)-(25360./2187.)*K2(I)
& +(64448./6561.)*K3(I)-(212./729.)*K4(I))
CALL FCN(N,X+(8./9.)*H,Y1,K6)
DO 26 I=1,N
26 Y1(I)=Y(I)+H*((9017./3168.)*K1(I)-(355./33.)*K2(I)
& +(46732./5247.)*K3(I)+(49./176.)*K4(I)-(5103./18656.)*K5(I))
XPH=X+H
CALL FCN(N,XPH,Y1,K6)
DO 27 I=1,N
27 Y1(I)=Y(I)+H*((35./384.)*K1(I)+(500./1113.)*K3(I)
& +(125./192.)*K4(I)-(2187./6784.)*K5(I)+(1./84.)*K6(I))
DISC=.TRUE.
CALL FCN(N,XPH,Y1,K7)
DO 28 I=1,N
28 K2(I)=((71./57600.)*K1(I)-(71./16695.)*K3(I)+(71./1920.)*K4(I)
& -(17253./339200.)*K5(I)+(22./526.)*K6(I)-(1./40.)*K7(I))*H
NFCN=NFCN+6
C --- ERROR ESTIMATION
ERR=0.
DO 41 I=1,N
DENOM=AMAX1(1.E-5,ABS(Y1(I)),ABS(Y(I)),2.*UROUND/EPS)
41 ERR=ERR+(K2(I)/DENOM)**2
ERR=SQRT(ERR/FLOAT(N))
C --- COMPUTATION OF HNEW
C --- WE REQUIRE .2<=HNEW/H<=10.
FAC=AMAX1(.1,AMIN1(5.,(ERR/EPS)**(1./5.)/.9))
HNEW=H/FAC
IF(ERR.LE.EPS)THEN
C --- STEP IS ACCEPTED
NACOPT=NACOPT+1
CALL STORE(X,XPH,Y,N,K1,K3,K4,K5,K6)
DO 44 I=1,N
K1(I)=K7(I)
44 Y(I)=Y1(I)
C --- RECOMPUTE K1 IN THE CASE OF DISCONTINUOUS INITIAL PHASE
IF(.NOT.DISC) CALL FCN(N,XPH,Y,K1)
X=XPH
IF(ABS(HNEW).GT.HMAX)HNEW=HMAX
IF(REJECT)HNEW=AMIN1(ABS(HNEW),ABS(H))
REJECT=.FALSE.
ELSE
C --- STEP IS REJECTED
REJECT=.TRUE.
IF(NACOPT.GE.1)NREJCT=NREJCT+1
ENDIF
H=HNEW
GOTO 1
C --- FAIL EXIT
79 WRITR(6,979)X
979 FORMAT(' EXIT OF RETARD AT X=',E11.4)
RETURN
END
C
SUBROUTINE STORE(X,XPH,Y,N,FG1,FG3,FG4,FG5,FG6)
PARAMETER (NN=4,MXST=800)
DIMENSION Y(N),FG1(N),FG3(N),FG4(N),FG5(N),FG6(N)
COMMON/CORF/XSTOR(MXST),YSTOR(NN,MXST),
& C1(NN,MXST),C2(NN,MXST),C3(NN,MXST),C4(NN,MXST)
C --- COEFFICIENTS FOR GLOBAL SOLUTION ARE STORED IN COEF
COMMON/POSITS/IFIRST,LAST,X0,XLAST,IPOS,DISC
LAST=LAST+1
IFIRST=MAX0(1,LAST-MXST+1)
IADR=MOD(LAST-1,MXST)+1
XLAST=XPH
XSTOR(IADR)=X
DO 2 I=1,NN
YSTOR(I,IADR)=Y(I)
C1(I,IADR)=FG1(I)

```

```

C2(I,IADR)=-(1337./480.)*FG1(I)+(105400./27825.)*FG3(I)-
& (135./80.)*FG4(I)-(54675./212000.)*FG5(I)+(66./70.)*FG6(I)
C3(I,IADR)=(1039./360.)*FG1(I)-(468200./83475.)*FG3(I)+
& (9./2.)*FG4(I)+(400950./318000.)*FG5(I)-(638./210.)*FG6(I)
C4(I,IADR)=-(1163./1152.)*FG1(I)+(37900./16695.)*FG3(I)-
& (415./192.)*FG4(I)-(674325./508800.)*FG5(I)+(374./168.)*FG6(I)
2 CONTINUE
RETURN
END

C
REAL FUNCTION YLAG(I,X,PHI)
PARAMETER (NN=4,MXST=800)
LOGICAL DISC
COMMON/COEF/XSTOR(MXST),YSTOR(NN,MXST),
& C1(NN,MXST),C2(NN,MXST),C3(NN,MXST),C4(NN,MXST)
COMMON/POSITS/IFIRST,LAST,X0,XLAST,IPOS,DISC
COMMON/UROUND/UROUND
C --- INITIAL PHASE
IF (DISC) THEN
  IF(ABS(X-X0).LE.(3.*UROUND*ABS(X)))DISC=.FALSE.
  IF(X.LE.X0)THEN
    YLAG=PHI(I,X)
    RETURN
  END IF
END IF
C --- COMPUTE THE POSITION OF X
IF (X.LT.XSTOR(IFIRST)) THEN
  WRITE (6,*) ' MEMORY FULL, MXST = ',MXST
  STOP
END IF
IPOS=MAX0(IFIRST,MIN0(LAST,IPOS))
1 IADR=MOD(IPOS-1,MXST)+1
IF (X.LT.XSTOR(IADR).AND.IPOS.GT.IFIRST) THEN
  IPOS=IPOS-1
  GOTO 1
END IF
2 IADR=MOD(IPOS,MXST)+1
IF (IPOS.LT.LAST) THEN
  IF (X.GT.XSTOR(IADR)) THEN
    IPOS=IPOS+1
    GOTO 2
  END IF
END IF
C --- COMPUTE THE DESIRED APPROXIMATION
IADR=MOD(IPOS-1,MXST)+1
IF (IPOS.EQ.LAST) THEN
  H=XLAST-XSTOR(IADR)
ELSE
  H=XSTOR(MOD(IPOS,MXST)+1)-XSTOR(IADR)
END IF
S=(X-XSTOR(IADR))/H
YLAG=YSTOR(I,IPOS)+H*S*(C1(I,IPOS)+S*(C2(I,IPOS)-
& +S*(C3(I,IPOS)+S*C4(I,IPOS))))
RETURN
END

```

## Литература

Абель (N. H. Abel)

(1826): Untersuchungen über die Reihe:

$$1 + \frac{m}{1} x + \frac{m(m-1)}{1 \cdot 2} x^2 + \frac{m(m-1)(m-2)}{1 \cdot 2 \cdot 3} x^3 + \dots \text{ и. с. в.}$$

Crelle J. f. d. r. u. angew. Math. (in zwanglosen Heften), vol. 1, p.311—339.

(1827): Ueber einige bestimmte Integrale. Crelle J. f. d. r. u. angew. Math., vol. 2, p. 22—30.

Абрамович, Стиган (M. Abramowitz and I. A. Stegun)

(1964): Handbook of mathematical functions, Dover, 1000 pages. [Имеется перевод: Справочник по специальным функциям с формулами, графиками и математическими таблицами, под ред. М. А. Абрамовича и И. Стигана. — М.: Наука, 1979. — 832 с.]

Адамс (J. C. Adams)

(1883): см. Башфорт (1883).

Айнс (E. L. Ince)

(1944): Ordinary differential equations, Dover Publications, New York, 558 pp. [Имеется перевод предыдущего издания: Айнс Э. Л. Обыкновенные дифференциальные уравнения. — Харьков: Гостехиздат Украины, 1939. — 717 с.]

Алексеев В. М. (1961): Об одной оценке возмущений решений обыкновенных дифференциальных уравнений. — Вестник МГУ, серия I (математика, механика), № 2, с. 28—36.

Альбрехт (J. Albrecht)

(1955): Beiträge zum Runge-Kutta-Verfahren, ZAMM, vol. 35, p. 100—110.

Аренсторф (R. F. Arenstorf)

(1963): Periodic solutions of the restricted three body problem representing analytic continuations of Keplerian elliptic motions, Amer. J. Math., vol. LXXXV, p. 27—35.

Арцела (C. Arzelá)

(1895): Sulle funzioni di linee, Memorie dell. R. Accad. delle Sc. di Bologna, 5e serie, vol. V, p. 225—244, see also: vol. V, p. 257—270, vol. VI, (1896), p. 131—140.

Барридж (K. Burrage)

(1985): Order and stability properties of explicit multivalue methods. Appl. Numer. Anal., vol. 1; pp. 363—379.

Барридж, Бутчер (K. Burrage and J. C. Butcher)

(1980): Non-linear stability of a general class of differential equation methods, BIT, vol. 20, pp. 185—203.

Барридж, Мосс (K. Burrage and P. Moss)

(1980): Simplifying assumptions for the order of partitioned multivalue methods. BIT, vol. 20, pp. 452—465.

Баттин (R. H. Battin)

(1976): Resolution of Runge-Kutta-Nyström condition equations through eighth order, AIAA J., vol. 14, pp. 1012—1021.

- Бауэр, Рутисхаузер, Штифель (F. L. Bauer, H. Rutishauser and E. Stiefel) (1963): New aspects in numerical quadrature. Proc. of Symposia in Appl. Math. vol. 15, p. 199–218, Am. Math. Soc.
- Бахвалов Н. С. Численные методы. — М.: Наука, 1973 и 1975.
- Башфорд (F. Bashforth) (1883): An attempt to test the theories of capillary action by comparing the theoretical and measured forms of drops of fluid. With an explanation of the method of integration employed in constructing the tables which give the theoretical form of such drops, by J. C. Adams, Cambridge Univ. Press.
- Беентьев, Герритсен (P. A. Beentjes and W. J. Gerritsen) (1976): Higher order Runge-Kutta methods for the numerical solution of second order differential equations without first derivatives, Report NW 34/76, Math. Centrum, Amsterdam.
- Беллен (A. Bellen) (1984): One-step collocation for delay differential equations, J. Comput. Appl. Math. Vol. 10, p. 275–283.
- Беллман, Куки (R. Bellman and K. L. Cooke) (1963): Differential-Difference equations, Academic Press, 482 p. [Имеется перевод: Беллман Р., Куки К. Л. Дифференциально-разностные уравнения. — М.: Мир, 1967.]
- Бендикссон (I. Bendixson) (1893): Sur le calcul des intégrales d'un système d'équations différentielles par des approximations successives, Stock. Akad. Forh., Vol. 51, p. 599–612. (1901): Sur les courbes définies par des équations différentielles, Acta Mathematica, vol. 24, p. 1–88.
- Бенкке, Зоммер (H. Behnke and F. Sommer) (1962): Theorie der analytischen Funktionen einer komplexen Veränderlichen. Zweite Auflage. Springer Verlag, Berlin-Göttingen-Heidelberg.
- Бернулли Даниил (D. Bernoulli) (1760): Essai d'une nouvelle analyse de la mortalité causée par la petite vérole, et des avantages de l'inoculation pour la prévenir, Hist. et Mém. de l'Acad. Roy. Sciences, Paris, 1760, p. 1–45; Werke Band 2, p. 235–267.
- Бернулли Иоганн (Joh. Bernoulli) (1696): Problema novum Mathematicis propositorum. Acta Erud. Lips. MDCXCVI, p. 269, Opera Omnia vol. 1, p. 161 and 165, Lausanne & Genevae 1742. (1697): De Conoidibus et Sphaeroidibus quaedam. Solutio analytica Acquationis in Actis A. 1695, pag. 553 propositae, Acta Erud. Lips., MDCXCVII, p. 113–118. (1697b): Solutioque Problematis a se in Actis 1696, p. 269, proposit, de invenienda Linea Brachystochrona, Acta Erud. Lips. MDCXCVII, p. 206, Opera Omnia vol. 1, p. 187–193.
- Бернулли Якоб (Jac. Bernoulli) (1695): Explications, Annotationes & Additiones ad ea, quae in Actis super anni de Curva Elastica, Isochrona Paracentrica, & Velaria, hinc inde memorata, & partim controversa legundur: ubi de Linea mediarii directionum, aliisque novis, Acta Erudit. Lipsiae, Anno MDCXCV, p. 537–553. (1697): Solutio Problematum Fraternorum, Peculiaris Programmate Cal. Jan. 1697 Groningae, nec non Actorum Lips. mense Jun. & Dec. 1696, & Febr. 1697 propositorum: una cum Propositione reciproca aliorum. Acta Erud. Lips. MDCXCVII, p. 211–217.
- Бибербах (L. Bieberbach) (1923): Theorie der Differentialgleichungen, Grundlehren Bd. VI, Springer-Verlag. (1951): On the remainder of the Runge-Kutta formula in the theory of ordinary differential equations, Zeitschr. angew. Math. Phys. (ZAMP), vol. 2, p. 233–248.

- Биркгофф, Барна (G. Birkhoff and R. S. Varga)**  
 (1965): Discretization errors for well-set Cauchy problems I, *Journal of Math. and Physics*, vol. XLIV, p. 1—23.
- Бирн, Ламберт (G. D. Byrne and R. J. Lambert)**  
 (1966): Pseudo-Runge-Kutta methods involving two points, *J. Assoc. Comput. Mach.*, vol. 13, p. 114—123.
- Бирн, Хайндмарш (G. D. Byrne and A. C. Hindmarsh)**  
 (1975): A polyalgorithm for the numerical solution of ordinary differential equations, *ACM Trans. on Math. Software*, vol. 1, No. 1, p. 71—96.
- де Боор, Шварц (C. de Boor and B. Swartz)**  
 (1973): Collocation at Gaussian points, *SIAM J. Numer. Anal.* vol. 10, p. 582—606.
- Булирш, Штёр (R. Bulirsch and J. Stoer)**  
 (1964): Fehlerabschätzungen und Extrapolation mit rationalen Funktionen bei Verfahren vom Richardson-Typus, *Num. Math.* Vol. 6, p. 413—427.  
 (1966): Numerical treatment of ordinary differential equations by extrapolation methods, *Num. Math.* vol. 8, p. 1—13.
- Бутчэр (J. C. Butcher)**  
 (1963): Coefficients for the study of Runge-Kutta integration processes, *J. Austral. Math. Soc.*, vol. 3, p. 185—201.  
 (1963a): On the integration process of A. Huta, *J. Austral. Math. Soc.* vol. 3, p. 202—206.  
 (1964a): Implicit Runge-Kutta Processes. *Math. Comput.* vol. 18, p. 50—64.  
 (1964b): On Runge-Kutta processes of high order, *J. Austral. Math. Soc.* vol. IV, Part 2, p. 179—194.  
 (1964c): Integration processes based on Radau quadrature formulas, *Math. Comput.*, Vol. 18, p. 233—244.  
 (1965a): A modified multistep method for the numerical integration of ordinary differential equations. *J. ACM*, vol. 12, p. 124—135.  
 (1965b): On the attainable order of Runge-Kutta methods, *Math. of Comp.* vol. 19, p. 408—417.  
 (1966): On the convergence of numerical solutions to ordinary differential equations, *Math. Comput.*, vol. 20, p. 1—10.  
 (1967): A multistep generalization of Runge-Kutta methods with four or five stages. *J. ACM*, vol. 14, p. 84—99.  
 (1969): The effective order of Runge-Kutta methods, in: Conference on the numerical solution of differential equations, Lecture notes in Math., Vol. 109, p. 133—139.  
 (1972): An algebraic theory of integration methods, *Math. Comput.*, Vol. 26, p. 79—106.  
 (1975): An order bound for Runge-Kutta methods, *SIAM J. Num. Anal.* vol. 12, p. 304—315.  
 (1981): A generalization of singly-implicit methods, *BIT*, vol. 21, p. 175—189.  
 (1984): An application of the Runge-Kutta space, *BIT*, vol. 24, p. 425—440.  
 (1985a): General linear method: a survey, *Appl. Num. Math.*, vol. 1, p. 273—284.  
 (1985b): The non-existence of ten stage eighth order explicit Runge-Kutta methods, *BIT*, vol. 25, p. 521—540.
- Вайсингер (J. Weissinger)**  
 (1950): Eine verschärzte Fehlerabschätzung zum Extrapolationsverfahren von Adams, *ZAMP*, vol. 30, p. 356—363.
- Ван-дер-Полъ (B. Van der Pol)**  
 (1926): On «Relaxation Oscillations», *Phil. Mag.*, vol. 2, p. 978—992; reproduced in: B. van der Pol, Selected Scientific Papers, vol. 1, North-Holland Publ. Comp. Amsterdam (1960).
- Ваннер (G. Wanner)**

(1969): Integration gewöhnlicher Differentialgleichungen, Lie Reihen, Runge-Kutta-Methoden, BI Mannheim Htb. 831/831a, 182 p.

(1973): Runge-Kutta methods with expansions in even powers of  $h$ , Computing vol. 11, p. 81—85.

(1983): On Shi's counter example for the 16th Hilbert problem, Internal Rep. Sect. de Math., Univ. Genève. 1982; in german in: Jahrbuch Ueberblicke Mathematik 1983, ed. Chatterji, Fenyö, Kulisch, Laugwitz, Liedl, BI Mannheim, p. 9—24.

#### Бейерштрасс (K. Weierstrass)

(1858): Ueber ein die homogenen Functionen zweiten Grades betreffendes Theorem, nebst Anwendung desselben auf die Theorie der kleinen Schwingungen, Monatsber. der Königl. Akad. der Wiss., 4. März 1858, Werke Bd. 1, p. 233—246.

#### Вернер (J. H. Verner)

(1978): Explicit Runge-Kutta methods with estimates of the local truncation error, SIAM J. Numer. Anal., vol. 15, p. 772—790.

#### Верхольст (P. F. Verhulst)

(1845): Recherches mathématiques sur la loi d'accroissement de la population, Nuov. Mem. Acad. Roy. Bruxelles, vol. 18, p. 3—38.

#### Вольтерра (V. Volterra)

(1934): Remarques sur la Note de M. Régnier et Mlle Lambin. C. R. Acad. Sc. t. CXCIX, p. 1682. See also: V. Volterra — U. d'Ancona, Les associations biologiques au point de vue mathématique, Paris 1935.

#### Вронский (H. Wronski)

(1810): Premier principe des méthodes algorithmiques comme base de la technique algorithmique, publication refused by the Acad. de Paris (for more details see: S. Dickstein, Int. Math. Congress 1904, p. 517).

#### Гаусс (C. F. Gauss)

(1812): Disquisitiones generales circa seriem infinitam

$$1 + \frac{\alpha\beta}{1 \cdot \gamma} x + \frac{\alpha(\alpha+1)\beta(\beta+1)}{1 \cdot 2 \cdot \gamma(\gamma+1)} xx + \\ + \frac{\alpha(\alpha+1)(\alpha+2)\beta(\beta+1)(\beta+2)}{1 \cdot 2 \cdot 3 \cdot \gamma(\gamma+1)(\gamma+2)} x^3 + \text{etc.}$$

Werke vol. 3, p. 123—162.

#### Гаутчи (W. Gautschi)

(1962): On inverses of Vandermonde and confluent Vandermonde matrices, Numer. Math. vol. 4, p. 117—123.

#### Гиббонс (A. Gibbons)

(1960): A program for the automatic integration of differential equations using the method of Taylor series, Computer J. vol. 3, p. 108—111.

#### Гилл (S. Gill)

(1951): A process for the step-by-step integration of differential equations in an automatic digital computing machine, Proc. Cambridge Philos. Soc., vol. 47, p. 95—108.

#### Гир (C. W. Gear)

(1965): Hybrid methods for initial value problems in ordinary differential equations, SIAM J. Numer. Anal. ser. B, Vol. 2, p. 69—86.

(1971): Numerical initial value problems in ordinary differential equations, Prentice-Hall, 253 p.

#### Гир, Ватанабе (C. W. Gear and D. S. Watanabe)

(1974): Stability and convergence of variable order multistep methods, SIAM J. Num. Anal. 11, p. 1044—1058.

#### Гир, Ту (C. W. Gear and K. W. Tu)

(1974): The effects of variable mesh size on the stability of multistep methods, SIAM J. Num. Anal. 11, p. 1025—1043.

**Гобсон (E. W. Hobson)**

(1921): *The theory of functions of a real variable*, vol. 1, Cambridge, 670 p.

**Голдстайн (H. H. Goldstine)**

(1980): *A History of the Calculus of Variations*, Springer-Verlag, 410 p.

**Голуб, Уилкинсон (G. H. Golub and J. H. Wilkinson)**

(1976): *Ill-conditioned eigensystems and the computation of the Jordan canonical form*, SIAM Review, vol. 18, p. 578—619.

**Готтвальд (B. A. Gottwald)**

(1985): *Zur Modellierung zeitverzögter biologischer Prozesse*, Informatik Fachberichte, vol. 109, p. 350—354, Springer-Verlag.

**Грёбнер (W. Gröbner)**

(1960): *Die Lierien und ihre Anwendungen*, D. Verl. d. Wiss. Berlin, 2nd ed. 1967.

**Григориев (R. D. Grigorieff)**

(1977): *Numerik gewöhnlicher Differentialgleichungen* 2. Teubner Studienbücher, Stuttgart.

(1983): *Stability of multistep-methods on variable grids*, Numer. Math., 42, p. 359—377.

**Гронуолл (T. H. Gronwall)**

(1919): *Note on the derivatives with respect to a parameter of the solutions of a system of differential equations*, Ann. Math. Vol. 20, p. 292—296.

**Грэг (W. B. Gragg)**

(1964): *Repeated extrapolation to the limit in the numerical solution of ordinary differential equations*, Thesis, Univ. of California; see also SIAM J. Numer. Anal. vol. 2, p. 384—403 (1965)

(1965): *On extrapolation algorithms for ordinary initial value problems*, SIAM J. Num. Anal. ser. B, vol. 2, p. 384—403.

**Грэг, Штеттер (W. B. Gragg and H. J. Stetter)**

(1964): *Generalized multistep predictor-corrector methods*. J. ACM, vol. 11, p. 188—209.

**Гурвиц (A. Hurwitz)**

(1895): *Ueber die Bedingungen, unter welchen eine Gleichung nur Wurzeln mit negativen reelen Theilen besitzt*. Math. Ann., vol. 46, p. 273—284; Werke vol. 2, p. 533ff.

**Даламбер (J. le Rond d'Alembert)**

(1748): *Suite des recherches sur le calcul intégral, quatrième partie: Méthodes pour intégrer quelques équations différentielles*, Hist. Acad. Berlin, Tom IV, p. 275—291.

**Далквист (G. Dahlquist)**

(1956): *Convergence and stability in the numerical integration of ordinary differential equations*, Math. Scand., vol. 4, p. 33—53.

(1959): *Stability and error bounds in the numerical integration of ordinary differential equations*. Trans. of the Royal Inst. of Technn., Stockholm, Sweden, Nr. 130, 87 p.

(1985): *33 years of numerical instability*, part 1, BIT, vol. 25, p. 188—204.

**Далквист, Йелч (G. Dahlquist and R. Jeltsch)**

(1979): *Generalized disks of contractivity for explicit and implicit Runge-Kutta methods*. Report TRITA-NA-7906, NADA, Roy. Inst. Techn. Stockholm.

**Деклу (J. Descloux)**

(1963): *A note on a paper by A. Nordsieck*, Report No. 131, Dept. of Comp. Sci., Univ. of Illinois at Urbana-Champaign.

**Джованнини, Вейсс-Пармегиани, Ульрих (B. Giovannini, L. Weiss-Parmegiani and B. T. Ulrich)**

(1978): *Phase locking in coupled Josephson weak links*, Helvet. Physica Acta, vol. 51, p. 69—74.

**Дзеннаро (M. Zennaro)**

(1986): Natural continuous extensions of Runge—Kutta methods. *Math. Comput.* Vol. 46, p. 119—133.

**Дойфлхард (P. Deuflhard)**

(1980): Recent advances in multiple shooting techniques. In: Computational techniques for ordinary differential equations (Gladwell-Sayers, ed.), Section 10, p. 217—272, Academic Press.

(1983): Order and stepsize control in extrapolation methods, *Num. Math.* vol. 41, p. 399—422.

(1985): Recent progress in extrapolation methods for ordinary differential equations, *SIAM Rev.*, vol. 27, p. 505—535.

**Дойфлхард, Бадер (F. Deuflhard and G. Bader)**

(1982): Multiple shooting techniques revisited, Preprint Nr. 163, Univ. Heidelberg, Inst. Angew. Math., D-6900 Heidelberg 1, Summary published in: P. Deuflhard, E. Hairer, (1983) p. 74—94.

**Дойфлхард, Хайрер (P. Deuflhard and E. Hairer ed.)**

(1983): Numerical treatment of inverse problems in differential and integral equations, Birkhäuser, Boston, Basel, 357 p.

**Донельсон, Хансен (J. Donelson and E. Hansen)**

(1971): Cyclic composite multistep predictor-corrector methods, *SIAM J. Numer. Anal.*, vol. 8, p. 137—157.

**Дорман, Принс (J. R. Dormand, and P. J. Prince)**

(1978): New Runge-Kutta algorithms for numerical simulation in dynamical astronomy, *Celestial Mechanics*, vol. 18, p. 223—232.

(1980): A family of embedded Runge-Kutta formulae, *J. Comp. Appl. Math.* vol. 6, p. 19—26.

**Драйвер (R. D. Driver)**

(1977): Ordinary and delay differential equations, *Applied Math. Sciences* 20, Springer Verlag, 501 p.

**Жордан (C. Jordan)**

(1870): *Traité des Substitutions et des équations algébriques*, Paris, 667 p.

(1928): Sur une formule d'interpolation, *Atti Congresso Bologna*, vol. 6, p. 157—177.

**Зоммер (D. Sommer)**

(1965): Numerische Anwendung impliziter Runge-Kutta-Formeln, *ZAMP*, vol. 45, Sonderheft, p. T77—T79.

**Зонневельд (J. A. Zonneveld)**

(1963): Automatic integration of ordinary differential equations, Report R743, Mathematisch Centrum, Postbus 4079, 1009AB Amsterdam. Appeared in book form 1964.

**Ингланд (R. England)**

(1969): Error estimates for Runge-Kutta typesolutions to systems of ordinary differential equations: *The Computer J.* vol. 12, p. 166—170.

**Йосс, Джозеф (G. Iooss and D. D. Joseph)**

(1980): Elementary stability and bifurcation theory, Springer-Verlag, 286 p. (see chapter VII).

[Имеется перевод: Йосс Ж., Джозеф Д. Элементарная теория устойчивости и бифуркаций. —М.: Мир, 1983.]

**Какутани, Маркус (S. Kakutani and L. Marcus)**

(1958): On the non-linear difference-differential equation  $y'(t) = [A - B \times \times y(t - \tau)] y(t)$ , in: Contributions to the theory of nonlinear oscillations. Ed. by S. Lefschetz, Princeton, vol. IV, p. 1—18.

**Камке (E. Kamke)**

(1930): Ueber die eindeutige Bestimmtheit der Integrale von Differentialgleichungen II, *Sitz. Ber. Heidelberg Akad. Wiss. Math. Naturw. Kl.*, 17. Abhandl., see also *Math. Zeitschr.*, vol. 32, p. 101—107.

- (1942): Differentialgleichungen, Lösungsmethoden und Lösungen, Becker & Erler, Leipzig, 642 p.  
 [Имеется перевод: Камке Э. Справочник по обыкновенным дифференциальным уравнениям. — М.: ИЛ. 1951. — 828 с., а также перевод 6-го немецкого издания (1959): изд. 2-е. — М.: Физматгиз, 1961; изд. 3-е, 4-е, 5-е. — М.: Наука, 1965, 1971, 1976.]
- Каннингэм (W. J. Cunningham)  
 (1954): A nonlinear differential-difference equation of growth, Proc. Math. Acad. Sci., USA, vol. 40, p. 708—713.
- Кастлюнгер, Ваннер (K. H. Kastlunger and G. Wanner)  
 (1972): Runge-Kutta processes with multiple nodes, Computing vol. 9, p. 9—24  
 (1972b): On Turan type implicit Runge-Kutta methods, Computing vol. 9, p. 317—325.
- Кауэлл, Кроммелин (P. H. Cowell and A. C. D. Crommelin)  
 (1910): Investigation of the motion of Halley's comet from 1759 to 1910. Appendix to Greenwich Observations for 1909, Edinburgh, p. 1—84.
- Кейли (A. Cayley)  
 (1858): A memoir on the theory of matrices, Phil. Trans. of Royal Soc. of London, vol. CXLVIII, p. 17—37, Mathematical Papers vol. 2, Nr. 152, p. 475.
- Келлер (H. B. Keller)  
 (1968): Numerical methods for two-point boundary-value problems, Blaisdell Publ. Comp., 184 p. (see section 2.4).
- Кермак, Макендрик (W. O. Kermack and A. G. McKendrick)  
 (1927): Contributions to the mathematical theory of epidemics (Part 1), Proc. Roy. Soc., A., vol. 115, p. 700—721.
- Кертис (A. R. Curtis)  
 (1970): An eighth order Runge-Kutta process with eleven function evaluations per step, Numer. Math. Vol. 16, p. 268—277.  
 (1975): High-order explicit Runge-Kutta formulae, their uses, and limitations, J. Inst. Maths. Applies., vol. 16, p. 35—55.
- Кертис, Хиршфельдер (C. F. Curtiss and J. O. Hirschfelder)  
 (1952): Integration of stiff equations, Proc. of the National Academy of Sciences of U. S., vol. 38, p. 235—243.
- Кинг (R. King)  
 (1966): Runge-Kutta methods with constrained minimum error bounds, Math. Comput., vol. 20, p. 386—391.
- Клеро (A. C. Clairaut)  
 (1734): Solution de plusieurs problèmes où il s'agit de trouver des courbes dont la propriété consiste dans une certaine relation entre leurs branches, exprimée par une Equation donnée. Mémoires de Math. et de Phys. de l'Acad. Royale des Sciences, Paris, Année MDCCXXXIV, p. 196—215.
- Кнапп, Ваннер (H. Knapp and G. Wanner)  
 (1969): LIESE II, A program for ordinary differential equations using Lie-series, MRC Report No. 1008, Math. Research Center, Univ. Wisconsin. Madison, Wisc. 53706.
- Ковалевский (G. Kowalewski)  
 (1908): см. Лейбниц (1693).
- Когстрем, Рүэ (B. Kaogström and A. Ruhe)  
 (1980): An algorithm for numerical computation of the Jordan normal form of a complex matrix: ACM Trans. Math. Software, vol. 6, p. 398—419 (Received May 1975, revised Aug. 1977, accepted May 1979).
- Коллатц (L. Collatz)  
 (1951): Numerische Behandlung von Differentialgleichungen, Grundlehren Band IX, Springer Verlag, 458 p., second edition, 1955; third edition and english translation 1960.  
 [Имеется перевод: Коллатц Л. Численные методы решения дифференциальных уравнений. — М.: ИЛ, 1953.]

**Коллет, Экманн (P. Collet and J. P. Eckmann)**

(1980): Iterated maps on the interval as dynamical systems. Birkhäuser, 248 p.

**Кори (S. A. Corey)**

(1906): A method of approximation, Amer. Math. Monthly Vol. 15, p. 137—140.

**Коши (A. L. Cauchy)**

(1824): Résumé des Lecons données à l'Ecole Royale Polytechnique. Suite du Calcul Infinitésimal; published: Equations différentielles ordinaires, ed. Chr. Gilain, Johnson 1981.

(1831): Sur la mecanique celeste et sur un nouveau calcul appelé calcul des limites, lu à l'acad. de Turin le 11 oct. 1831; also: exerc. d'anal. et de physique math., 2, Paris 1841; Oeuvres (2), 12.

(1839—42): Several articles in Comptes Rendus de l'Acad. des Sciences de Paris. (Aug. 5, Nov. 21, 1839. June 29, Oct. 26, 1840, etc).

**Крайтер (C. W. Cryer)**

(1971): A proof of the instability of backward-difference multistep methods for the numerical integration of ordinary differential equations. Tech. Rep. No. 117. Comp. Sci. Dept., Univ. of Wisconsin, p. 1—52.

(1972): On the instability of high order backward-difference multistep methods, BIT, vol. 12, p. 17—25.

**Кридон, Миллер (D. M. Creedon and J. J. H. Miller)**

(1975): The stability properties of q-step backward-difference schemes, BIT, vol. 15, p. 244—249.

**Крох (F. T. Krogh)**

(1969): A variable step variable order multistep method for the numerical solution of ordinary differential equations, Information Processing 68, North-Holland, Amsterdam, p. 194—199.

(1973): Algorithms for changing the step size. SIAM J. Num. Anal., 10, p. 949—965.

(1974): Changing step size in the integration of differential equations using modified divided differences. Proceedings of the Conference on the Num. Sol. of ODE, Lecture Notes in Math. No. 362, Springer-Verlag, New York, p. 22—71.

**Крузе (M. Crouzeix)**

(1975): Sur l'approximation des équations différentielles opérationnelles linéaires par des méthodes de Runge-Kutta, Thèse d'état, Univ. Paris 6, 192 p.

**Крузе, Лисbonа (M. Crouzeix and F. J. Lisbona)**

(1984): The convergence of variable-stepsize, variable formula, multistep methods, SIAM J. Num. Anal. 21, p. 512—534.

**Крылов Н. М., Боголюбов Н. Н. Введение в нелинейную механику. — Киев: Изд-во АН УССР, 1937.**

**Куммер (E. E. Kummer)**

(1839): Note sur l'intégration de l'équation  $d^n y/dx^n = x^m y$  par des intégrales définies. Crelle J. f. d. r. u. angew. Math. vol. 19, p. 286—288.

**Кунцманн (J. Kuntzmann)**

(1959): Deux formules optimales du type de Runge-Kutta, Chiffres, vol. 2, p. 21—26.

(1961): Neuere Entwickelungen der Methode von Runge-Kutta, ZAMM, vol. 41, p. 28—31.

**Купер (G. J. Cooper)**

(1978): The order of convergence of general linear methods for ordinary differential equations. SIAM J. Numer. Anal., vol. 15, p. 643—661.

**Купер, Вернер (G. J. Cooper and J. H. Verner)**

(1972): Some explicit Runge-Kutta methods of high order, SIAM J. Numer. Anal., vol. 9, p. 389—405.

**Кутта (W. Kutta)**

(1901): Beitrag zur näherungsweisen Integration totaler Differentialgleichungen, Zeitschr. für Math. u. Phys., vol. 46, p. 435—453.

**Лагранж (J. L. de Lagrange)**

- (1759): *Recherches sur la nature et la propagation du son*, *Miscell. Taurinensis* t. I, *Oeuvres* t. 1, p. 39—148.
- (1762): *Solution de différents problèmes de Calcul Intégral*, *Miscell. Taurinensis*, t. III, *Oeuvres* t. 1, p. 471—668.
- (1774): *Sur les Intégrales particulières des Equations différentielles*, *Oeuvres*, tom. 4, p. 5—108.
- (1775): *Recherche sur les Suites Récurrentes*, *Nouveaux Mém. de l'Acad. royale des Sciences et Belles-Lettres*, Berlin, *Oeuvres* vol. 4, p. 159.
- (1788): *Mécanique analytique*, Paris, *Oeuvres* t. 11, et 12  
[Имеется перевод: Лагранж Ж. Аналитическая механика. — М.—Л.: Гос-техиздат, 1950, т. 1 и 2.]
- (1792): *Mémoire sur l'expression du terme général des séries récurrentes, lorsque l'équation génératrice a des racines égales*. *Nouv. Mém. de l'Acad. royale des Sciences de Berlin*, *Oeuvres* t. 5, p. 627—641.
- (1797): *Théorie des fonctions analytiques, contenant les principes du calcul différentiel, dégagés de toute considération d'infiniment petits, d'évanouissants, de limites et de fluxions, et réduits à l'analyse algébrique des quantités finies*, Paris, 1797, nouv. ed. 1813, *Oeuvres Tome 9*.

**Лакс, Рихтмайер (P. D. Lax and R. D. Richtmyer)**

- (1956): *Survey of the stability of linear limite difference equations*. *Comm. Pure Appl. Math.*, vol. 9, p. 267—293.

**Лаплас (P. S. Laplace)**

- (An XIII = 1805): *Supplément au dixième livre du Traité de mécanique céleste sur l'action capillaire*, Paris chez Courcier, 65 + 78 p.

**Лейбниц (G. W. Leibniz)**

- (1691): *Methodus, qua innumerarum linearum construction ex data proprietate tangentium seu aequatio inter abscissam et ordinatam ex dato volore subtangentialis, exhibetur*. Letter to Huygens, in: C. I. Gerhardt, *Leibnizens math. Schriften*, 1850, Band II, p. 116—121.
- (1693): (*Gothofredi Guilirimi Leibnitii*): *Supplementum Geometriae Dimensionariae seugeneralissima omnium tetra gonismorum effectio per motum: Similiterque multiplex constructio linea ex data tangentium conditione*, *Acta Bruditorum; Lipsiae*, p. 385—392; german translation: G. Kowalewski, *Leibniz über die Analysis des Unendlichen*, Ostwalds Klassiker, Nr. 162 (1908), p. 24—34.

**Лефевер, Николис (R. Lefever and G. Nicolis)**

- (1971): *Chemical Instabilities and sustained oscillations*, *J. theor. Biol.*, vol. 30, p. 267—284.

**Линделёф (E. Lindelöf)**

- (1894): *Sur l'application des méthode d'approximation successives a l'étude des intégrales réelles des équations différentielles ordinaires*, *J. de Math.*, 4e série vol. 10, p. 117—128.

**Липшиц (R. Lipschitz)**

- (1876): *Sur la possibilité d'intégrer complètement un système donné d'équations différentielles*, *Bulletin des Sciences Math. et Astr.*, Paris, vol. 10, p. 149—159.

**Лиувилль (J. Liouville)**

- (1836): *Sur le développement des fonctions ou parties de fonctions en séries dont les divers termes sont assujétis à satisfaire à une même équation différentielle du second ordre, contenant un paramètre variable*, *Journ. de Math. pures et appl.*, vol. 1, p. 253—265.

- (1838): *Sur la Théorie de la variation des constantes arbitraires*, *Liouville J. de Math.* vol. 3, p. 342—349.

- (1841): *Remarques nouvelles sur l'équation de Riccati*, *J. des Math. pures et appl.*, vol. 6, p. 1—13.

- Лобатто (R. Lobatto)**  
 (1852): *Lessen over Differential- en Integraal-Rekening*, 2 vol., La Haye 1851—1852.
- Лоренц (E. N. Lorenz)**  
 (1979): On the prevalence of aperiodicity in simple systems, *Global Analysis*, Calgary 1978, ed. by M. Grmela and J. E. Marsden, Lecture Notes in Mathematics, vol. 755, p. 53—75.
- Лоткин (M. Lotkin)**  
 (1951): On the accuracy of Runge-Kutta methods, *MTAC* vol. 5, p. 128—132.
- Лиенар (A. M. Liénard)**  
 (1928): Etude des oscillations entretenues, *Revue générale de l'Électricité*, tome XXIII, p. 901—912, et 946—954.
- Ляпунов (А. М. Liapunov)**  
 (1892): Problème général de la stabilité du mouvement, russ. trad. en français 1907 (*Annales de la Faculté des Sciences de Toulouse*), reprinted 1947 Princeton Univ. Press, 474 p.
- Ляпунов А. М.**  
 (1892): Общая проблема устойчивости движения. — В кн.: Собрание сочинений, т. II, с. 7—263. — М.—Л.: Изд-во АН СССР.
- Марден (M. Marden)**  
 (1966): *Geometry of polynomials*, American Mathematical Society, Providence, Rhode Island, 2nd edition
- Марсден, Мак-Кракен (J. E. Marsden and M. McCracken)**  
 (1976): *The Hopf bifurcation and its applications*, Springer, New York, p. 408.  
 [Имеется перевод: Марсден Дж., Мак-Кракен М. Бифуркация рождения цикла и ее приложения. — М.: Мир, 1980.]
- Марчук Г. И.**  
 (1975): Простейшая математическая модель вирусного заболевания. — Новосибирск, ВЦ СО АН СССР, препринт.  
 (1980): Математические модели в иммунологии. — М.: Наука.
- Мерсон (R. H. Merson)**  
 (1957): An operational method for the study of integration processes, Proc. Symp. Data Processing, Weapons Research Establishment, Salisbury, Australia, p. 110—1 to 110—25.
- Мизес, фон (R. von Mises)**  
 (1930): Zur numerischen Integration von Differentialgleichungen, *ZAMM*, vol. 10, p. 81—92.
- Милн (W. E. Milne)**  
 (1926): Numerical integration of ordinary differential equation, *Amer. Math. Monthly*, vol. 33, p. 455—460.  
 (1970): *Numerical solution of differential equations*. Dover Publications, Inc. New York, second edition. [Имеется перевод первого издания: Милн В. Э. Численное решение дифференциальных уравнений. — М.: ИЛ, 1955.]
- Митчелл, Крэггс (A. R. Mitchell and J. W. Craggs)**  
 (1953): Stability of difference relations in the solution of ordinary differential equations, *Math. Tables Aids Comput.*, vol. 7, p. 127—129.
- Молер, ван Лоан (C. Moler and C. Van Loan)**  
 (1978): Nineteen dubious ways to compute the exponential of a matrix: *SIAM Review*, Vol. 20, p. 801—836.
- Мултон (F. R. Moulton)**  
 (1926): *New methods in exterior ballistics*. Univ. Chicago Press
- Мур (R. E. Moore)**  
 (1966): *Interval Analysis*, Prentice-Hall, Inc., 145 p.  
 (1979): *Methods and applications of interval analysis*, *SIAM studies in Appl. Math.*, p. 190.

- Мюллер (M. Müller)**  
 (1926): Über das Fundamentaltheorem in der Theorie der gewöhnlichen Differentialgleichungen, *Math. Zeitschr.*, vol. 26, p. 619—645. (Kap. III).
- Мэй (R. M. May)**  
 (1976): Simple mathematical models with very complicated dynamics, *Nature*, vol. 261, p. 459—467.
- Невилл (E. H. Neville)**  
 (1934): Iterative interpolation, *Ind. Math. Soc. J. Vol. 20*, p. 87—120.
- Норсетт (S. P. Norsett)**  
 (1974a): One-step methods of Hermite type for numerical integration of stiff systems, *BIT*, vol. 14, p. 63—77.  
 (1974b): Semi explicit Runge-Kutta methods, Report No. 6/74, ISBN 82-7151-009-6, Dept. Math. Univ. Trondheim, Norway, 68+7pp.
- Норсетт, Ваннер (S. P. Norsett and G. Wanner)**  
 (1979): The real-pole sandwich for rational approximations and oscillation equations, *BTT* vol. 19, p. 79—94.
- Нордсик (A. Nordsieck)**  
 (1962): On numerical integration of ordinary differential equations, *Math. Comp.*, vol. 16, p. 22—49.
- Нумеров (B. Numerov (B. V. Noumerow))**  
 (1924): A method of extrapolation of perturbations. *Monthly notices of the Royal Astronomical Society* of  $d^2x/dt^2 = f(x, t)$  *Astron. Nachrichten*, vol. 230, p. 359—364.
- Ньютона (I. Newton)**  
 (1671): *Methodus Fluxionum et Serierum Infinitarum*, edita Londini 1736, Opuscula mathematica vol. 1. Traduit en français par M. de Buffon, Paris MDCCXL. [Имеется перевод: Ньютона И. Математические работы. — М.—Л.: Объединенное научно-техническое издательство НКТП СССР, Главная редакция технико-теоретической литературы, 1937, с. 25—166; Метод флюксий и бесконечных рядов.]  
 (1687): *Philosophiae naturalis principia mathematica*, Imprimatur S. Pepys. Reg. Soc. Praeses, julii 5, 1686, Londini anno MDCLXXXVII. [Имеется перевод: Иса. Ньютона, Математические начала натуральной философии/Перевод с латинского с примечаниями и пояснениями А. Н. Крылова. — В кн.: Собрание трудов академика А. Н. Крылова, т. VII. — М.—Л.: Изд-во АН СССР, 1936.]
- Нюстрём (H. J. Nyström)**  
 (1925): Ueber die numerische Integration von Differentialgleichungen, *Acta Soc. Sci. Fenn.*, vol. 50, No. 13, p. 1—54.
- Окамото, Хаяси (M. Okamoto and K. Hayashi)**  
 (1984): Frequency conversion mechanism in enzymatic feedback systems, *J. Theor. Biol.*, vol. 108, p. 529—537.
- Оливер (J. Oliver)**  
 (1975): A curiosity of low-order explicit Runge-Kutta methods, *Math. Comp.*, vol. 29, p. 1032—1036.
- Оборна (M. R. Osborne)**  
 (1966): On Nordsieck's method for the numerical solution of ordinary differential equations, *BTT*, vol. 6, p. 51—57.
- Парлетт (B. N. Parlett)**  
 (1976): A recurrence among the elements of functions of triangular matrices, *Linear Algebra Appl.*, vol. 14, p. 117—121.
- Пeano (G. Peano)**  
 (1888): Intégration par séries des équations différentielles linéaires, *Math. Annalen*, vol. 32, p. 450—456.  
 (1890): Démonstration de l'intégrabilité des équations différentielles ordinaires, *Math. Annalen* vol. 37, p. 182—228; see also the german translation and commentary: G. Mie, *Math. Annalen* vol. 43 (1893), p. 553—568.

- (1913): Resto nelle formule di quadratura, espresso con un integrale definito  
 Atti Della Reale Accad. Dei Lincei, Rendiconti, vol. 22, № 9, p. 562—569,  
 Roma.
- Перл, Рид (R. Pearl and L. J. Reed)  
 (1922): A further note on the mathematical theory of population growth, Proceedings of the National Acad. of Sciences, vol. 8, No. 12, p. 365—368.
- Перко (L. M. Perko)  
 (1984): Limit cycles of quadratic systems in the plane, Rocky Mountain J. of Math., vol. 14, p. 619—645.
- Перрон (O. Perron)  
 (1915): Ein neuer Existenzbeweis für die Integrale der Differentialgleichung  $y' = f(x, y)$ , Math. Annalen, vol. 76, p. 471—484.  
 (1918): zur Zeit in Felde): Ein neuer Existenzbeweis für die Integrale eines Systems gewöhnlicher Differentialgleichungen, Math. Annalen, vol. 78, p. 378.  
 (1929): Über Stabilität und asymptotisches Verhalten der Integrale von Differentialgleichungssystemen, Math. Z., vol. 29, p. 129—160 (see par. 3).
- Пикар (E. Picard)  
 (1890): Mémoire sur la théorie des équations aux dérivées partielles et la méthode des approximations successives, J. de Math. pures et appl., 4e série, vol. 6, p. 145—210.  
 (1891—1896): Traité d'Analyse, 3 vols, Paris.
- Пиотровски (P. Piotrowsky)  
 (1969): Stability, consistency and convergence of variable  $k$ -step methods for numerical integration of large systems of ordinary differential equations, Lecture Notes in Math. 109, Dundee 1969, p. 221—227.
- Пригожин, Лефевер и др. (I. Prigogine, R. Lefever, A. Goldbeter, M. Herschkowitz)  
 (1969): Nature, vol. 223, p. 913.
- Принс, Дорман (P. J. Prince and J. R. Dormand)  
 (1981): High order embedded Runge-Kutta formulae, J. Comp. Appl. Math., vol. 7, p. 67—75.
- Прюфер (H. Prüfer)  
 (1926): Neue Herleitung der Sturm-Liouvillschen Reihenentwicklung stetiger Funktionen. Math. Annalen vol. 95, p. 489—518.
- Пуанкаре (H. Poincaré)  
 (1881, 82, 85): Sur les courbes définies par les équations différentielles, J. de Math., 3e série, t. 7, p. 375—422, t. 8, p. 251—296, 4e série, t. 1, p. 167—244. [Имеется перевод: Пуанкаре А. О. кривых, определяемых дифференциальными уравнениями. — М.—Л.: Гостехиздат, 1947.]  
 (1893): Les méthodes nouvelles de la mécanique céleste, Tome II, 480 р., Gauthier-Villars Paris. [Имеется перевод: Пуанкаре А. Избранные труды, т. I, II. — М.: Наука, 1971.]
- Пуассон (S. D. Poisson)  
 (1835): Théorie mathématique de la chaleur, Paris, Buchelier, 532 p. Supplément 1837, p. 72.
- Пузэ (F. Pouzet)  
 (1963): Etude en vue de leur traitement numérique des équations intégrales de type Volterra. Rev. Francaise Traitement Information (Chiffres), vol. 6, p. 79—112.
- Радо (R. Radau)  
 (1880): Etude sur les formes d'approximation qui servent à calculer la valeur numérique d'une intégrale definite, Liouville J. de Mathém. pures et appl., Zeser., tome VI, p. 283—336. (Voir p. 307.)
- Райт Э. (E. M. Wright)  
 (1945): On a sequence defined by a non-linear recurrence formula, J. of London Math. Soc., vol. 20, p. 68—73.

- (1946): The non-linear difference-differential equation, Quart. J. of Math., vol. 17, p. 245—252.
- (1955): A non-linear difference-differential equation, J. f. d. r. u. angew. Math., vol. 194, p. 66—87.
- Райт К. (K. Wright)**
- (1970): Some relationships between implicit Runge-Kutta collocation and Lanczos  $\tau$  methods, and their stability properties, BTT vol. 10, p. 217—227.
- Ральстон (A. Ralston)**
- (1962): Runge-Kutta methods with minimum error bounds, Math. Comput. vol. 16, p. 431—437, corr. vol. 17, p. 488.
- Раус (H. J. Routh)**
- (1877): A Treatise on the stability of a given state of motions, being the essay to which the Adams prize was adjudged in 1877, in the University of Cambridge, London, p. 108.
- (1884): A Treatise on the dynamics of a system of rigid bodies, part A and II, 4th edition (1st ed. 1860, 6th ed. 1897, german translation with remarks of F. Klein 1898). [Имеется перевод: Раус Э. Дж. Динамика системы твердых тел. — М.: Наука, 1983. — т. I, II.]
- Ред (W. T. Red)**
- (1980): Sturmian theory for ordinary differential equations, Springer-Verlag, Appl. Math., Serie 31, p. 559.
- Риккати (J. Riccati)**
- (1712): Soluzione generale del Problema inverso intorno à raggi oscillatori ... determinar la curva, a cui consengs una tal'expressione, Giornale de'Letterati d'Italia, vol. 11, p. 204—220.
- (1723): Animadversions in aequationes differentiales secundi gradus, Acta Erud. Lips, anno MDCCXXIII, p. 502—510.
- Риман (B. Riemann)**
- (1854): Ueber die Darstellbarkeit einer Function durch eine trigonometrische Reihe, von dem Verfasser behufs seiner Habilitation an der Universität zu Göttingen der philosophischen Facultät eingereicht; collected works, p. 227—265. [Имеется перевод: Риман Б. Сочинения. — М.—Л.: Гостехиздат, 1948, с. 225—261: О возможности представления функций посредством тригонометрического ряда.]
- Ричардсон (L. F. Richardson)**
- (1910): The approximate arithmetical solution by finite differences of physical problems including differential equations, with an application to the stresses in a masonry dam, Phil. Trans., A, vol. 210, p. 307—357.
- (1927): The deferred approach to the limit, Phil. Trans., A, vol. 226, p. 299—349.
- Ромберг (W. Romberg)**
- (1955): Vereinfachte numerische Integration, Nomke Vid. Selsk Forhdl, vol. 28, p. 30—36.
- Роте (E. Rothe)**
- (1930): Zweidimensionale parabolische Randwertaufgaben als Grenzfall eindimensionaler Randvertaufgaben, Math. Annalen, vol. 102; p. 650—670.
- Рунге (C. Runge)**
- (1895): Ueber die numerische Auflösung von Differentialgleichungen, Math. Ann., vol. 46, p. 167—178.
- (1905): Ueber die numerische Auflösung totaler Differentialgleichungen, Göttinger Nachr., p. 252—257.
- Рунге, Кёниг (C. Runge and H. König)**
- (1924): Vorlesungen über numerisches Rechnen, Grundlehren XI, Springer-Verlag, p. 372.
- Рутисхаузер (H. Rutishauser)**
- (1952): Ueber die Instabilität von Methoden zur Integration gewöhnlicher Differentialgleichungen, ZAMP, vol. 3, p. 65—74.

- Руш, Абетс, Лалуа (N. Rouche, P. Habets and M. Laloy)  
 (1977): Stability theory by Liapunov's direct method, Appl. Math. Sci., 22,  
 Springer-Verlag, p. 396. [Имеется перевод: Руш Н., Абетс П., Лалуа М.  
 Прямой метод Ляпунова в теории устойчивости. — М.: Мир, 1980.]
- Рэлей (Lord Rayleigh)  
 (1883): On maintained vibrations, Phil. Mag. Ser. 5, vol. 15, p. 229—235.
- Рюэль, Такенс (D. Ruelle and F. Takens)  
 (1971): On the nature of turbulence, Commun. Math. Physics, vol. 20, p. 167—192.
- Саррафян (D. Sarafyan)  
 (1966): Error estimation for Runge-Kutta methods through pseudo-iterative  
 formulas; Techn. Rep. No. 14, Louisiana State Univ., New Orleans, May 1966.
- Симmons (D. F. Simmons)  
 (1972): Differential equations with applications and historical notes, McGraw-  
 Hill, p. 456.
- Скил (R. Skeel)  
 (1976): Analysis of fixed-stepsize methods, SIAM J. Numer. Anal., vol. 13,  
 p. 664—685.  
 (1978): Equivalent forms of multistep formulas, Report R-78-940, Dept.  
 of Comp. Sci., Univ. of Illinois at Urbana-Champaign
- Спайкер (M. N. Spijker)  
 (1971): On the structure of error estimates for finite difference methods; Numer.  
 Math., vol. 18, p. 73—100.
- Стеффенсен (J. F. Steffensen)  
 (1956): On the restricted problem of three bodies, K. danske Vidensk. Selsk.,  
 Mat-fys. Medd. 30 Nr. 18.
- Струд, Станку (A. H. Stroud and D. D. Stancu)  
 (1965): Quadrature formulas with multiple Gaussian nodes, SIAM J. Numer.  
 Anal., ser. B, vol. 2, p. 129—143.
- Тодд (J. Todd)  
 (1950): Notes on modern numerical analysis, I, Math. Tables Aids Comput.,  
 vol. 4, p. 39—44.
- Толлмин (W. Tollmien)  
 (1938): Über die Fehlerabschätzung beim Adamsschen Verfahren zur Integration  
 gewöhnlicher Differentialgleichungen, ZAMM, vol. 18, p. 83—90.
- Томсон, Тэйт (W. Tomson (Lord Kelvin) and P. G. Tait)  
 (1879): Treatise on natural philosophy (Vol. 1, Part I), Cambridge; New edition  
 1890, p. 508.
- Уайльд (O. Wilde)  
 (1892): Lady Windermere's Fan, Comedy in four acts [Имеется перевод: Оскар  
 Уайльд, Избранные произведения в двух томах. — М.: ГИХЛ, 1960.—  
 т. 2, с. 7—68; Веер леди Уиндермир.]
- Уилкинсон (J. H. Wilkinson)  
 (1965): The eigenvalue problem, Monographs on numerical analysis,  
 Oxford, p. 662. [Имеется перевод: Уилкинсон Дж. Х. Алгебраическая про-  
 блема собственных значений. — М.: Наука, 1970.]
- Уилкинсон, Райнш (J. H. Wilkinson and C. Reinsch)  
 (1970): Linear Algebra, Grundlehren Band 186, Springer-Verlag, p. 439.  
 [Имеется перевод: Уилкинсон, Райнш. Справочник алгоритмов на языке  
 АЛГОЛ. Линейная алгебра. — М.: Машиностроение, 1976.]
- Уинтнер, Мурнаган (A. Wintner and F. D. Murnaghan)  
 (1931): A canonical form for real matrices under orthogonal transformations,  
 Proc. Nat. Acad. Sci. U.S.A., vol. 17, p. 417—420.
- Ульманн (W. Uhlmann)  
 (1957): Fehlerabschätzungen bei Anfangswertaufgaben gewöhnlicher Differen-  
 tialgleichungssysteme I. Ordnung, ZAMM (Zeitschr. Angew. Math. Mech.)  
 vol. 37, p. 88—99.

**Уолтер (W. Walter)**

(1970): Differential and integral inequalities, Springer-Verlag p. 352, german edition 1964.

(1971): There is an elementary proof of Peano's theorem, Amer. Math. Monthly, vol. 78, p. 170—173.

**Уорр (H. A. Watts)**

(1983): Starting stepsize for an ODE solver, J. Comp. Appl. Math. vol. 9, p. 177—191.

**Фейгенбаум (M. Feigenbaum)**

(1978): Quantitative universality for a class of nonlinear transformations, J. Stat. Phys. vol. 19, p. 25—52, vol. 21 (1979), p. 669—706.

**Фельберг (E. Fehlberg)**

(1958): Eine Methode zur Fehlerverkleinerung beim Runge-Kutta Verfahren, ZAMM, vol. 38, p. 421—426.

(1964): New high-order Runge-Kutta formulas with step size control for systems of first and second order differential equations, ZAMM, vol. 44, Sonderheft T17-T19.

(1968): Classical fifth-, sixth-, seventh-, and eighth order Runge-Kutta formulas with step size control; NASA Technical Report 287 (1968); extract published in Computing vol. 4., p. 93—106 (1969).

(1969): Low-order classical Runge-Kutta formulas with step size control and their application to some heat transfer problems, NASA Technical Report 315 (1969), extract published in Computing vol. 6, p. 61—71 (1970).

(1972): Classical eighth-and lower-order Runge-Kutta-Nyström formulas with stepsize control for special second-order differential equations, NASA Technical Report R-381.

**Филд, Нойес (J. R. Field and R. M. Noyes)**

(1974): Oscillations in chemical systems. IV, Limit cycle behavior in a model of a real chemical reaction, J. Chem. Physics, vol. 60, p. 1877—1884.

**Филиппи, Грэф (S. Filippi and J. Gräf)**

(1986): New Runge-Kutta-Nyström formula-pairs of order 8 (7), 9 (8), 10 (9) and 11 (10) for differential equations of the form  $y' = f(x, y)$ , J. Comput. and Applied Math., vol. 14, p. 361—370.

**Фиторис (L. Vietoris)**

(1953): Der Richtungsfehler einer durch das Adamssche Interpolationsverfahren gewonnenen Näherungslösung einer Gleichung  $y' = f(x, y)$ . Oesterr. Akad. Wiss., Math.-naturw. Kl., Abt. IIa, vol. 162, p. 157—167 and p. 293—299.

**Флетчер, Соренсен (R. Fletcher and D. C. Sorensen)**

(1983): An algorithmic derivation of the Jordan canonical form, Amer. Math. Monthly, vol. 90, No. 1, p. 12—16.

**Форрингтон (C. V. D. Forrington)**

(1961—1962): Extensions of the predictor-corrector method for the solution of systems of ordinary differential equations, Comput. J. 4, p. 80—84.

**Франческини (V. Franceschini)**

(1980): Feigenbaum sequence of bifurcations in the Lorenz model, J. Stat. Phys., vol. 22, p. 397—406.

**Фрейзер, Джонс, Скан (R. A. Frazer, W. P. Jones and S. W. Skan)**

(1937): Approximations to functions and to the solutions of differential equations, Reports and Memoranda Nr. 1799 (2913): Aeronautical Research Committee, p. 33.

**Фрике (A. Fricke)**

(1949): Über die Fehlerabschätzung des Adamsschen Verfahrens zur Integration gewöhnlicher Differentialgleichungen erster Ordnung, ZAMM, vol. 29, p. 165—178.

**Фроммер (M. Frommer)**

(1934): Über das Auftreten von Wirbeln und Strudeln (geschlossener und spiralförmiger Integralkurven) in der Umgebung rationaler Unbestimmtheitsstellen,

Math. Ann., vol. 109, p. 395—424.

**Фукс (L. Fuchs)**

(1866): Zur Theorie der linearen Differentialgleichungen mit veränderlichen Coefficienten, Crelle J. f. d. r. u. angew. Math., vol. 66., p. 121—160. (published in «Programm der städtischen Gewerbeschule zu Berlin, Ostern 1865»)

**Функ (P. Funk)**

(1962): Variationsrechnung und ihre Anwendung in Physik und Technik, Springer-Verlag, Grundlehren Bd. 94, 676 p.

**Фурье (J. B. J. Fourier)**

(1807): Sur la propagation de la chaleur, unpublished manuscript; published: La théorie analytique de la chaleur, Paris 1822.

**Хайдмарш (A. C. Hindmarsh)**

(1972): GEAR: ordinary differential equation system solver UCID-30001, Rev. 2, LLL, Livermore, Calif.

(1980): LSODE and LSODI, two new initial value ordinary differential equation solvers, ACM Signum Newsletter 15, 4.

**Хайрер (E. Hairer)**

(1977): Méthodes de Nyström pour l'équation différentielle  $y'' = f(x, y)$ , Numer. Math., vol. 27, p. 283—300.

(1978): A Runge-Kutta method of order 10, J. Inst. Maths. Applies., vol. 21, p. 47—59.

(1981): Order conditions for numerical methods for partitioned ordinary differential equations, Numer. Math., vol. 36, p. 431—445.

(1982): A one-step method of order 10 for  $y'' = f(x, y)$ , IMA J. Num. Anal., vol. 2, p. 83—94.

**Хайрер, Ваннер (H. Hairer and G. Wanner)**

(1973): Multistep-multistage-multiderivative methods for ordinary differential equations, Computing, Vol. 11, p. 287—303.

(1974): On the Butcher group and General multi-value methods, Computing Vol. 13, p. 1—15.

(1976): A theory for Nyström methods, Numer. Math. vol. 25, p. 383—400.

(1983): On the instability of the BDF formulas, SIAM J. Numer. Anal., vol. 20, No. 6, p. 1206—1209.

**Хайрер, Любих (E. Hairer and Ch. Lubich)**

(1984): Asymptotic expansions of the global error of fixed stepsize methods, Numer. Math., vol. 45, p. 345—360.

**Халл (T. N. Hull)**

(1967): A search for optimum methods for the numerical integration of ordinary differential equations, SIAM Rev., vol. 9., p. 647—654.

**Халл, Джонстон (T. N. Hull and R. L. Johnston)**

(1964): Optimum Runge-Kutta methods, Math. Comput., vol. 18, p. 306—310.

**Халл, Энрайт и др. (T. E. Hull, W. H. Enright, B. M. Fellen and A. E. Sedgwick)**

(1972): Computing numerical methods for ordinary differential equations, SIAM, J. Numer. Anal., vol. 9, p. 603—637.

**Хаммер, Холлингсворт (P. C. Hammer and J. W. Hollingsworth)**

(1955): Trapezoidal methods of approximating solutions of differential equations, MTAC-vol. 9, p. 92—96.

**Хан В. (W. Hahn)**

(1967): Stability of motion, Springer, Verlag, p. 446.

**Хан Г. (H. Hahn)**

(1921): Theorie der reellen Funktionen, Springer-Verlag, Berlin, p. 600.

**Хаувен, ван дер (P. J. van der Houwen)**

(1977): Construction of integration formulas for initial value problems, North-Holland Amsterdam, p. 269.

- Хебзаккер (H. M. Hebsacker)**  
 (1982): Conditions for the coefficients of Runge-Kutta method for systems of n-th order differential equations, *J. Comput. Appl. Math.*, vol. 8, p. 3—14.
- Хейес (N. D. Hayes)**  
 (1950): Roots of the transzental equation associated with a certain difference differential equation, *J. of London Math. Soc.*, vol. 25, p. 226—232.
- Хенричи (P. Henrici)**  
 (1962): Discrete variable methods in ordinary differential equations, John Wiley & Sons, Inc., New York-London-Sydney.  
 (1974): Applied and computational complex analysis, Volume I, John Wiley & Sons, New York, p. 682.
- Хирш, Смайл (M. W. Hirsch and S. Smale)**  
 (1974): Differential equations, dynamical systems and linear algebra, Acad. Press, p. 358.
- Хобсон (E. W. Hobson)**  
 (1921): The theory of functions of a real variable, vol. I, Cambridge, 670 pp.
- Хайн (K. Heun)**  
 (1900): Neue Methode zur approximativem Integration der Differentialgleichungen einer unabhängigen Veränderlichen, *Zeitschr. für Math. u. Phys.*, vol. 45, p. 23—38.
- Хуфт (G. 't Hooft)**  
 (1974): Magnetic monopoles in unified gauge theories, *Nucl. Phys.* vol. B79, p. 276—284.
- Хопф (B. Hopf)**  
 (1942): Abzweigung einer periodischen Lösung von einer stationären Lösung eines Differentialsystems, *Ber. math. physik. Kl. Akad. a. Wiss. Leipzig, Bd. XCIV*, p. 3—22.
- Хорн (M. K. Horn)**  
 (1983): Fourth and fifth-order scaled Runge-Kutta algorithms for treating dense output, *SIAM J. Numer. Anal.* vol. 20, p. 558—568.
- Хулм (B. L. Hulme)**  
 (1972): One-step piecewise polynomial Galerkin methods for initial value problems, *Math. of Comput.* vol. 26, p. 415—426.
- Хундсдорфер, Спайкер (W. H. Hundsdorfer and M. N. Spijker)**  
 (1981): A note on B-stability of Runge-Kutta methods, *Num. Math.* Vol. 36, p. 319—331.
- Цурмюль (R. Zurmühl)**  
 (1948): Runge-Kutta Verfahren zur numerischen Integration von Differentialgleichungen n-ter Ordnung, *ZAMP*, vol. 28, p. 173—182.  
 (1952): Runge-Kutta Verfahren unter Verwendung höherer Ableitungen, *Z. angew. Math. Mech.*, vol. 32, p. 153—154.
- Ческино (F. Ceschino)**  
 (1961): Modification de la longueur du pas dans l'intégration numérique par les méthodes à pas fixes, *Chiffres* 2, p. 101—106.  
 (1962): Evaluation de l'erreur par pas dans les problèmes différentiels, *Chiffres* vol. 5, p. 223—229.
- Ческино, Кунцман (F. Ceschino and J. Kuntzmann)**  
 (1963): Problèmes différentiels de conditions initiales (méthodes numériques), Dunod Paris, p. 372.  
 (1963): Problèmes différentiels de conditions initiales, Dunod Paris; english translation: Numerical solutions of initial value problems, Prentice Hall, 1966.
- Шампайн (L. F. Shampine)**  
 (1979): Storage reduction for Runge-Kutta codes, *ACM Trans. Math. Software*, Vol. 5, p. 245—250.

- Шампайн, Бака (L. F. Shampine and L. S. Baca)**  
 (1984): Fixed versus variable order Runge-Kutta, Rept. SAND84-1410, Sandia Nat'l Labs., Albuquerque, New Mexico, p. 57. see also ACM Trans. Math. Softw., vol. 12, pp. 1—23, 1986.
- Шампайн, Гордон (L. F. Shampine and M. K. Gordon)**  
 (1975): Computer Solution of Ordinary Differential Equations, The Initial Value Problem, Freeman and Company, San Francisco, p. 318.
- Шампайн, Уоттс (L. F. Shampine and H. A. Watts)**  
 (1979): The art of writing a Runge-Kutta code, II, Appl. Math. Comput., vol. 5, p. 93—121.
- Шампайн, Уоттс, Давенпорт (L. F. Shampine, H. A. Watts and S. M. Davenport)**  
 (1976): Solving nonstiff ordinary differential equations — The state of the art, SIAM Rev., vol. 18, p. 376—410.
- Шанкс (H. B. Shanks)**  
 (1966): Solutions of Differential equations by evaluations of functions, Math. of Comp., vol. 20, p. 21—38.
- Шеффер (L. Scheffer)**  
 (1884): Zur Theorie der stetigen Funktionen einer reellen Veränderlichen, Acta Mathematica, vol. 5, p. 183—194.
- Ши Сонлин (Shi Songling)**  
 (1980): A concrete example of the existence of four limit cycles for plane quadratic systems, Sci Sinica, vol. 23, p. 153—158.
- Штёр, Булирш (J. Stoer and R. Bulirsch)**  
 (1973): Einführung in die numerische Mathematik, Springer Verlag, 1 st ed. 1973, English translation 1980 (see section 7.3.5.).  
 (1980): Introduction to Numerical Analysis, Springer-Verlag, 609 p., german edition 1973.
- Штёрмер (C. Störmer (C. Stoermer))**  
 (1907): Sur les trajectoires des corpuscules électrisés. Arch. sci. phys. nat., Genève, vol. 24, p. 5—18, 113—156, 221—247.  
 (1921): Méthodes d'intégration numérique des équations différentielles ordinaires. C. R. congr. intern. math., Strasbourg, p. 243—257.
- Штеттер (H. J. Stetter)**  
 (1970): Symmetric two-step algorithms for ordinary differential equations, Computing vol. 5, p. 267—280.  
 (1973): Analysis of discretization methods for ordinary differential equations, Springer-Verlag, Berlin—Heidelberg—New York. [Имеется перевод: Штеттер Х. Анализ методов дискретизации для обыкновенных дифференциальных уравнений. — М.: Мир., 1978. — 451 с.]
- Штурм (Ch. Sturm)**  
 (1829): Bulletin des Sciences de Féruccac, tome XI, p. 419, see also: Algèbre de Choquet et Mayer (1832).  
 (1836): Sur les équations différentielles linéaires du second ordre, Journal de Math., pures et appl. (Liouville), vol. 1, p. 106—186. (see also p. 253, p. 269, p. 373 of this volume).
- Шур (L. Schur)**  
 (1909): Über die charakteristischen Wurzeln einer linearen Substitution mit einer Anwendung auf die Theorie der Integralgleichungen, Math. Ann., vol. 66, p. 488—510.
- Эйлер (L. Euler)**  
 (1728): Nova methodus innumerabiles aequationes differentiales secundi gradus reducendi ad aequationes differentiales primi gradus, Comm. acad. scient. Petrop. vol. 3, p. 124—137.; Opera Omnia vol. XXII, p. 1—14.  
 (1743): De integratione aequationum differentialium altiorum graduum, Miscellanea Berolinensis vol. 7, p. 193—242; Opera Omnia vol. XXII, p. 108—149. See also: Letter from Euler to Joh. Bernoulli, 15, Sept. 1739.

(1744): *Methodus inveniendi lineas curvas maximi minimive proprietate gaudentes...*, Lausanne & Genevae, Opera Omnia (intr. by Caratheodory) vol. XXIV, p. 1—308. [Имеется перевод: Эйлер Л. Метод нахождения кривых линий, обладающих свойствами максимума либо минимума. — М.—Л.: Гостехиздат, 1934, 600 с.]

(1750): *Methodus aequationes differentiales altiorum gradium integrandi ulterius promota, Novi Comment. acad. scient. Petrop.* vol. 3, p. 3—35; *Opera Omnia* vol. XXII, p. 181—213.

(1756): *Elementa calculi variationum, presented September 16, 1756 at the Acad. of Science, Berlin; printed 1766, Opera Omnia vol. XXV.* p. 141—176.

(1768): *Institutionum Calculi Integralis, Volumen Primum, Opera Omnia* vol. XI. [Имеется перевод: Эйлер Л. Интегральное исчисление, т. I. — М.: Гостехиздат, 1956, 415 с.]

(1769): *Institutionum Calculi Integralis, Volumen Secundum, Opera Omnia* vol. XII. [Имеется перевод: Эйлер Л. Интегральное исчисление, т. II. — М.: Гостехиздат, 1957, 368 с.]

(1778): *Specimen transformationis singularis seriem, Nova acta. acad. Petrop.*, vol. 12 (1794): p. 58—70, *Opera Omnia* vol. XVI, p. 41—55.

**Эйспак (N. F. Eispack)**

(1974): B. T. Smith, J. M. Boyle, B. S. Garbow, Y. Jkebe, V. C. Klema, C. B. Moler: *Matrix Eigenesystem Routines, (Fortrantranslations of algorithms published in Reinsch and Wilkinson)*, Lecture Notes in Computer Science, vol. 6, Springer-Verlag.

**Эйткен (A. C. Aitken)**

(1932): On interpolation by iteration of proportional parts, without the use of differences; *Proc. Edinburgh Math. Soc. Second ser.* Vol. 3, p. 56—76.

**Эле (B. L. Ehle)**

(1968): High order A-stable methods for the numerical solution of systems of D. E.'s *BIT* vol. 8, p. 276—278.

**Эльтерманн (H. Eltermann)**

(1955): Fehlerabschätzung bei näherungsweiser Lösung von Systemen von Differentialgleichungen erster Ordnung *Math. Zeitschr.* vol. 62, p. 469—501.

**Энрайт, Халл, Линдберг (W. H. Enright, T. E. Hull and B. Lindberg)**

(1975): Computing numerical methods for stiff systems of O. S. O. D. E. s, *BIT*, vol. 15, p. 10—48.

**Энрайт и др. (W. H. Enright, K. R. Jackson, S. P. Norsett and P. G. Thomsen)**

(1985): Interpolants for Runge-Kutta formulas, Technical Report 180/85, Dept. Comput. Science, Toronto, Canada M5S 1A5, p. 36.

**Эрмит (Ch. Hermite)**

(1878): Extrait d'une lettre de M. Ch. Hermite à M. Borchardt sur la formule d'interpolation de Lagrange, *J. de Crelle*, vol. 84, p. 70; *Ouvres*, tome III, p. 432—443.

**Якоби (C. G. J. Jacobi)**

(1845): *Theoria novi multiplicatoris systemati aequationum differentialium vulgarium applicandi*, *Crelle J. f. d. r. u. angew. Math.*, vol. 29, p. 213—279, 333—376.

## **Дополнительная литература<sup>1)</sup>**

**Бахвалов Н. С.**

(1955): Некоторые замечания к вопросу о численном интегрировании дифференциальных уравнений методом конечных разностей. Доклады АН СССР, т. 104, № 6, с. 805—808.

**Бахвалов Н. С., Жидков Н. П., Кобельков Г. М.**

(1987): Численные методы. — М.: Наука. — 600 с.

**Бордовицна Т. В.**

(1984): Современные численные методы в задачах небесной механики. — М.: Наука. — 136 с.

**Бутчер (J. Ch. Butcher)**

(1987): The numerical analysis of ordinary differential equations: Runge—Kutta and general linear methods, Wiley, Chichester, 512 pp.

**Горбунов А. Д.**

(1967): Разностные уравнения и разностные методы решения задач Коши для систем обыкновенных дифференциальных уравнений. — М.: Изд. ВЦ МГУ.

**Крылов В. И., Бобков В. В., Монастырный П. И.**

(1982): Начала теории вычислительных методов. Дифференциальные уравнения. — Минск: Наука и техника. — 286 с.

**Лаврентьев М. А., Шабат Б. В.**

(1973): Методы теории функций комплексного переменного. — М.: Наука. — 736 с.

**Ламберт (J. D. Lambert)**

(1973): Computational methods in ordinary differential equations, Wiley, London—New York—Sidney—Toronto, 278 pp.

**Лапидус, Зайнфельд (L. Lapidus, J. H. Seinfeld)**

(1971): Numerical solution of ordinary differential equations, Academic Press, New York—London, 299 pp.

**Максвелл Д. К., Вышнеградский И. А., Стодола А.**

(1949): Теория автоматического регулирования. — М.: Гостехиздат.

**Самарский А. А., Гулин А. В.**

(1989): численные методы. — М.: Наука. — 432 с.

**Холл, Уатт (G. Hall and J. M. Watt, editors)**

(1976): Modern numerical methods for ordinary differential equations, Clarendon Press, Oxford. [Имеется перевод: Современные численные методы решения обыкновенных дифференциальных уравнений, редакторы Дж. Холл и Дж. Уатт. — М.: Мир, 1979. — 312 с.]

**Эльсгольц Л. Э., Норкин С. Б.**

(1971): Введение в теорию дифференциальных уравнений с отклоняющимся аргументом. — М.: Наука. — 296 с.

<sup>1)</sup> Литература, добавленная редактором перевода.

## Указатель обозначений

$A \otimes I$	тензорное произведение 362
$a(t)$	коэффициенты $B$ -ряда 259, 260
$B^J(a, y)$	$B$ -ряд 260
$B(p)$	квадратурные условия порядка 218
$b_J(\theta)$	непрерывный метод 191
$c_i, a_{ij}, b_j$	коэффициенты метода РК 139, 213
$C$	константа погрешности 339, 383, 448
$C_{p+1}$	константа локальной погрешности 338
$C(\eta)$	упрощающие предположения 217
$D$	дифференциальный оператор 266
$D \pm m(x)$	производные Дини 60
$D(\xi)$	упрощающие предположения 218
$d_i(t)$	разностное множество 260
$E$	главная часть $S$ 411
$e_p(x)$	коэффициенты разложения глобальной погрешности 226, 227
$f[x_n, \dots, x_{n-1}]$	разделенные разности 336
$F^I(t)(y)$	элементарный дифференциал 154, 156, 280, 281, 297
$g_j(n)$	коэффициенты методов Адамса с переменным шагом 367
$h$	длина шага 135, 136
$K_q(s)$	ядро Пеано 341
$l = (l_0, l_1, \dots)$	коэффициенты Нордсика 380, 386
$L$	постоянная Липшица 39, 57
$L(y, x, h)$	линейный разностный оператор 335, 446
$LNT_q$	помеченные $N$ -деревья 280
$LS_q$	специальные помеченные деревья 159
$LT_q, LT$	помеченные деревья 154, 259
$LTP_q^a, LTP^a$	помеченные $P$ -деревья 296, 299
$NT_q$	$N$ -деревья 282
$P(c, y)$	$P$ -ряд 299
$P(EC)^M$	предиктор—корректор 394, 406
$P(EC)^M E$	предиктор—корректор 393, 405
$\ Q\ $	матричная норма 58
$R(x, x_0)$	резольвента 70
$S$	матрица общего линейного метода 404
$s_t(t)$	субдерево 260

- $t_{21}, t_{31}, \dots$   
 $T_q, T$   
 $T_{i, 1}, T_{j, k}$   
 $TP_q^a, TP_a^k$   
 $V(y_1, \dots, y)$   
 $W(x)$   
 $\|y\|$   
 $y_h(x)$   
 $y(x, x_0, y_0)$   
 $z_n = z(x_n, h)$   
 $\alpha_j, \beta_j$   
 $\alpha(t)$   
 $\beta_j(n)$   
 $\varphi(t)$   
 $\Psi_j, \Psi_i$   
 $\nabla^I f_n$   
 $\mu(Q)$   
 $\phi(h)$   
 $\Phi_j(t)$   
 $\Phi_j(n)$   
 $\Phi_j^*(n)$   
 $\Phi(x_0, y_0, h)$   
 $\Phi^*(x, y, h)$   
 $\rho(t)$   
 $\rho(\zeta), \sigma(\zeta)$   
 $\Psi, \Psi_a$
- деревья 156  
корневые деревья 155, 259  
экстраполяционная таблица 234, 235, 290  
Р-деревья 296, 299  
функция Ляпунова 95  
матрица Вронского 70  
норма вектора 56  
ломаная Эйлера 38  
решение задачи Коши 105  
функция точных значений 404  
коэффициенты многошагового метода 334, 370  
число элементов в классе эквивалентности 155, 156, 281, 282, 298  
коэффициенты методов Адамса с переменным шагом 368  
произведения порядков деревьев 156, 160, 281  
коэффициенты явных, неявных методов Адамса 325, 326  
разности назад 324  
логарифмическая норма 65  
стартовая процедура 404  
весовые коэффициенты 156, 159, 281, 283  
разделенные разности 368  
разделенные разности 367  
функция приращения 169, 362, 404  
функция приращения сопряженного метода 230  
порядок дерева 154, 155, 281, 296, 298  
производящие многочлены 336  
деревья порядка 1 155, 156, 297

## Предметный указатель

- Абелл—Лиувилля—Якоби—Остроградского тождество** (Abel—Liouville—Jacobi—Ostrogradskii identity) 71  
**Автоматическое управление длиной шага** (automatic step size control) 177  
**Автономные системы** (autonomous systems) 75, 150  
**Адамса методы** (Adams methods) 323  
— — неявные (implicit ~ ~) 325  
— — с переменным шагом (variable step size ~ ~) 366  
— — в представлении Нордсика (~ ~ as Nordsieck methods) 379, 381, 386, 387  
— — рекуррентные соотношения для коэффициентов (~ ~, recurrence relations for the  $\gamma_i$ ) 327, 333  
— — явные (explicit ~ ~) 324  
**Алгоритм ГБШ** (GBS algorithm) 240, 290  
— — Евклида (Euclid's ~) 90, 92  
— — Рауса (Routh ~) 94, 98  
— — Эйткена—Невилла (Aitken—Neville ~) 237  
**Аренсторфа орбиты** (Arenstorf orbits) 135, 211  
**Аргумента запаздывание** (time lags, retarded arguments) 304  
— принцип (principle of the argument) 89, 349  
**Арцела—Асколи теорема** (theorem of Arzela—Ascoli) 44  
**Асимптотическая устойчивость** (asymptotic stability) 96  
**Асимптотические решения при малых  $\varepsilon$**  (~ solutions for small parameters) 118  
**Асимптотическое разложение глобальной погрешности** (~ expansion of the global error) 226, 288, 421, 447  
— — — симметричного метода (~ ~ ~ ~, symmetric method) 232  
— — — сопряженного метода (~ ~ ~ ~, adjoint ~) 231  
— — — по четным степеням  $h$  (~ ~ ~ ~ in  $h^2$ ) 232, 238, 435  
— — локальной погрешности (~ ~ of the local error) 226  
**Барьеры порядковые Бутчера** (order barriers of Butcher) 198, 202  
— — **Далквиста** (~ ~ of Dahlquist) 351, 449  
**Бернуlli уравнение** (Bernoulli equation) 19  
— — числа (~ numbers) 228, 384  
**Бесселя уравнения** (Bessel equation) 29  
**Бифуркация рождения цикла** (Hopf bifurcation) 121  
**Брахистохрона** (brachystochrone) 13, 19, 28  
**Брюсселатор** (Brusselator) 120, 132, 183, 253  
— — с диффузией (~ with diffusion) 399  
— — «полный» (full ~) 122  
**Булирша последовательность** (Bulirsch sequence) 236  
**Бутчера барьеры** (Butcher barriers) 198, 202  
— метод 6-го порядка (~'s 6-th order method) 202  
— методы порядка 2s (~'s methods of order 2s) 217  
— ряды ( $B$ -series) 258  
— формулы Лобатто (Butcher's Lobatto formulas) 220  
**Ван-дер-Поля уравнение** (Van der Pol's equation) 115, 131, 132, 253  
**Вариационная задача** (variational problem) 14, 36  
**Вариационное исчисление** (~ calculus) 13  
— — уравнение (~ equation) 106  
**Вариация постоянных** (variation of constants) 23, 72  
— — нелинейная (nonlinear ~ ~) 106  
**Веер леди Уиндермир** (Lady Windermere's fan) 41, 107, 170, 364  
**Вековые члены** («secular» terms) 32, 119  
**Вектор Нордсика** (Nordsieck vector) 379, 387, 389  
**Векторное поле** (~ field) 11  
**Векторные обозначения** (~ notation) 10, 56  
**Верхольста—Перла уравнение** (Verhulst—Pearl equation) 312  
**Вершина конечная** (end-vertex) 280  
— — толстая, тонкая («fat», «meagre» ~) 279  
**Висящая струна** (hanging string) 32  
**Вложенные формулы** (embedded formulas) 178, 206  
**Вольтерры уравнение** (Volterra's equation) 319, 321  
**Вронского матрица** (вронскиан) (Wronskian) 70  
— — определитель (~) 24

- Выбор порядка (order selection) 390  
 Выдача неявная (implicit output) 192  
 — плотная (dense  $\sim$ ) 188, 191, 469  
 Вырожденное ядро (degenerate kernel) 319, 320  
 Вычислительные примеры (numerical examples)  
 —, методы РК 4-го порядка ( $\sim \sim$ , 4th order RK methods) 145, 185  
 — — высших порядков ( $\sim \sim$ , high order methods) 209  
 — —, многошаговые методы ( $\sim \sim$ , multistep methods) 396  
 — —, сравнение программ ( $\sim \sim$ , comparisons of codes) 252, 396  
 — —, уравнения 2-го порядка ( $\sim \sim$ , 2nd order equations) 290  
 — —, экстраполяционные методы ( $\sim \sim$ , extrapolation methods) 237
- Гармоническая последовательность (harmonic sequence) 237  
 Гаусса квадратурные формулы (Gaussian quadrature formulas) 137, 217  
 ГБШ-алгоритм (GBS algorithm) 240, 290  
 Гершгорина теорема (Gershgorin's theorem) 99  
 Гилла метод (Gill's method) 145  
 Гильберта 16-я проблема (Hilbert's 16th problem) 131  
 Гипергеометрические функции (hypergeometric functions) 27  
 Главный член погрешности (principal error term) 168  
 Глобальная погрешность (global error) 169, 226, 421  
 Границные условия (boundary conditions) 110  
 Графы (graphs) 153  
 Грёбнера—Алексеева теорема (формула) (Gröbner—Alekseev theorem (formula)) 106, 222  
 Греко-латинское преобразование (Greek-Roman transformation) 351  
 Грина формула (Green's formula) 133  
 Гронуолла лемма (Gronwall lemma) 67  
 Грггга метод (Gragg's method) 239
- Даламбера уравнение (d'Alembert's equation) 31  
 Далквиста барьеры (Dahlquist barriers) 351, 449  
 Дерево (tree) 155  
 — помеченное (labelled  $\sim$ ) 154  
 — специальное (special  $\sim \sim$ ) 159  
 Дефект (см. тж. норма невязки) (defect) 61, 64
- Джозефсона переходы (Josephson junctions) 124, 125  
 Диаграммы «точность—объем вычислений» (work—precision diagrams) 146, 186, 194, 210, 238, 239, 242, 256, 257, 291, 397, 398, 402  
 Динамика популяций (population dynamics) 311, 319  
 Диши производные (Dini derivatives) 60  
 Дискретное преобразование Лапласа (discrete Laplace transformation) 362  
 Дифференциальные неравенства (differential inequalities) 60, 96  
 — для систем ( $\sim \sim$  for systems) 67  
 — уравнения (differential equations)  
 — — второго порядка (second order  $\sim \sim$ ) 9, 17, 276, 437  
 — — жесткие (stiff  $\sim \sim$ ) 124, 214, 331  
 — — с запаздывающим аргументом (delay  $\sim \sim$ ) 304  
 — — линейные (linear  $\sim \sim$ ) 16, 17, 21, 26, 69, 75  
 — — неявные (implicit  $\sim \sim$ ) 15  
 — — первого порядка (first order  $\sim \sim$ ) 9  
 — — в частных производных (partial  $\sim \sim$ ) 10, 399  
 ДНРК (DIRK) 213  
 Дормана—Принса метод (Dormand and Prince method) 182  
 — —, непрерывное расширение ( $\sim \sim \sim$ , continuous extension) 191  
 — — для уравнения 2-го порядка ( $\sim \sim \sim$  for 2nd order equations) 289
- Евклида алгоритм (Euclid's algorithm) 90, 92
- Жесткие дифференциальные уравнения (stiff differential equations) 124, 214, 331  
 Жорданова каноническая форма (Jordan canonical form) 80, 87, 95  
 — клетка ( $\sim$  block) 89  
 — цепочка векторов ( $\sim$  chain) 95
- Задача о брахистохроне (brachystochrone problem) 13, 19, 28  
 — двух тел (two body  $\sim$ ) 252  
 — о касательной обратная (inverse tangent  $\sim$ ) 12  
 — Коши (initial value  $\sim$ ) 10, 109  
 — краевая (boundary value  $\sim$ ) 10, 109  
 — начальная (с начальными условиями) (initial value  $\sim$ ) 10, 109  
 — на собственные значения (eigenvalues  $\sim$ ) 32, 93, 111

- трех тел ограниченная (restricted three body  $\sim$ ) 134
- Штурма—Лиувилля (Sturm—Liouville eigenvalue  $\sim$ ) 111
- Замораживание длин шагов (step size freeze) 196
- Запаздывающий аргумент (time lag, retarded argument) 304
- Зонневельда метод (Zonneveld's method) 180
- Иммунология (immunology)** 317
- Интегрирующий множитель (factor  $M(x, y)$ ) 17
- Интегро-дифференциальные уравнения (integro-differential equations) 318, 320
- Интервальный анализ (interval analysis) 53
- Интерполяционная формула Лагранжа (Lagrange interpolation formula) 222
- Ньютона (Newton's  $\sim\sim$ ) 324, 366
  - с кратными узлами ( $\sim\sim$  with multiple nodes) 267
- Интерполяционный многочлен Эрмита (Hermite's  $\sim$  polynomial) 267
- Каноническая форма Жордана (Jordan canonical form) 80, 87, 95
- Капли (drops) 147, 323, 332
- Кауэлла—Кроммелина метод (Cowell and Crommelin's method) 439, 442
- Квадратурные формулы Гаусса (Gaussian quadrature formulas) 137, 217
  - Лобатто (Lobatto  $\sim\sim$ ) 219
  - Радо (Radau  $\sim\sim$ ) 213
- Квазимонотонность (quasimonotone) 46, 47, 109
- Кинетика ферментативных реакций (enzyme kinetics) 315
  - химическая (chemical  $\sim$ ) 120
- Классический метод Рунге—Кутты («The» Runge—Kutta method) 144
- Клеро уравнение (Clairaut differential equation) 15, 19
- Коллокация (collocation) 220
  - с кратными узлами ( $\sim$  with multiple nodes) 267
- Комбинированное управление выбором порядка и длины шага (combined order- and step size control) 245, 247
- Композиция методов (composition of RK methods) 258
  - В-рядов ( $\sim$  of B-series) 261
- Константа погрешности методов Нордсиека (error constant of Nordsieck methods) 383
- многошаговых методов ( $\sim\sim$  of multistep methods) 338, 339
  - для уравнений 2-го порядка ( $\sim\sim$  for 2nd order equations) 448
- Корень дерева (root of a tree) 154
- Корневое условие ( $\sim$  condition) 347
- Корни характеристического уравнения (roots of characteristic equation) 22
- кратные (multiple  $\sim\sim\sim$ ) 25
- Корректор (corrector) 327, 393, 405
- Коши—Буняковского—Шварца неравенство (Cauchy—Schwarz inequality) 55
- Коши задача (initial value problem) 10, 109
  - произведение (Cauchy product) 51, 384
- Коэффициенты погрешности (error coefficients) 168
- Краевая задача (boundary value problem) 10, 109
- Кратные характеристические корни (multiple characteristic values) 25
- узлы интерполяционной формулы ( $\sim$  nodes of interpolation formula) 267
- Критерий Райса—Гурвица (Routh—Hurwitz criterion) 89
- Шура—Кона (Schur—Cohn  $\sim$ ) 356
- Критические точки (critical points) 83, 85, 87, 117
- Кронекера тензорное произведение (Kronecker tensor product) 362
- Кунцмана методы порядка 2s (Kuntzmann's methods of order 2s) 217, 218, 219
- Кутты методы (Kutta's  $\sim$ ) 144
- Лагранжа интерполяционная формула (Lagrange interpolation formula)** 222
  - метод (вариации постоянных) (method of  $\sim$  (variation of constants)) 23, 25
  - многочлены ( $\sim$  polynomials) 221
- Лапласа дискретное преобразование (discrete Laplace transformation)** 382
  - уравнение (формы капель) (differential equation of  $\sim$  (forms of drops)) 147, 323
- Лейбница формула (Leibniz' formula)** 151
- Лемма Гронуолла (Gronwall lemma) 67
  - фундаментальная (fundamental  $\sim$ ) 62, 108
- Линеаризация (linearization)** 75
- Линейные дифференциальные уравнения (linear differential equations)** 16, 17, 21

- — — неоднородные (inhomogeneous  $\sim \sim \sim$ ) 16, 22, 72
- — — однородные (homogeneous  $\sim \sim \sim$ ) 16, 21
- — — с постоянными коэффициентами ( $\sim \sim \sim$  with constant coefficients) 21, 75
- — — со слабыми особенностями ( $\sim \sim \sim$  with weak singularities) 26
- многошаговые ( $k$ -шаговые) методы ( $\sim$  multistep ( $k$ -step) methods) 334
- формулы ( $\sim \sim$  formulas) 323
- — — сходимость порядка  $p$  ( $\sim \sim \sim$ , convergence of order  $p$ ) 360
- Липшица условие* (Lipschitz condition) 15, 39, 57, 359
- — — одностороннее (one-sided  $\sim \sim$ ) 64
- Лобатто квадратурные формулы* (Lobatto quadrature formulas) 219
- Логарифмическая норма (logarithmic norm) 65, 68, 95
- Логистическая кривая (la courbe logistique) 312
- Локальная погрешность (local error) 170, 226, 334, 409
- экстраполяция ( $\sim$  extrapolation) 210
- Лоренца модель (Lorenz model) 125
- Ляпунова функции (Liapunov functions) 94, 99
  
- Марчук модель* (Marchuk's model) 317, 318
- Матрица Вронского (вронскиан) (Wronskian) 70
- комплексная (complex matrix) 58, 77
- нормальная (normal  $\sim$ ) 86
- ортогональная (orthogonal  $\sim$ ) 77
- симметричная (symmetric  $\sim$ ) 77
- сопряженная (adjoint  $\sim$ ) 77
- унитарная (unitary  $\sim$ ) 77
- Якоби (Jacobian) 105
- Маятник математический (pendulum) 132
- сферический (spherical  $\sim$ ) 122
- Мерсона метод* (Merson's method) 180
- Метод (method)
- вариации постоянных ( $\sim$  of variation of constants) 23, 25
- Гилла (Gill's  $\sim$ ) 145
- Грэгга (Gragg's  $\sim$ ) 239
- дискретизации по времени ( $\sim$  of discretization in time) 10
- Дормана—Принса (Dormand and Prince  $\sim$ ) 182
- — —, непрерывное расширение
- — —,  $\sim \sim \sim$ , continuous extension) 191
- — — для уравнений 2-го порядка ( $\sim \sim \sim$  for 2nd order equations) 289
- Зонневельда (Zonneveld's  $\sim$ ) 180
- Кауэлла—Кроммелина (Cowell and Crommelin's  $\sim$ ) 439, 442
- композитный (composite  $\sim$ ) 259
- Лагранжа ( $\sim$  of Lagrange) 23, 25
- мажорант (majorant  $\sim$ ) 50
- малого параметра (Newcomb—Lindstedt—Poincaré  $\sim$ ) 118
- Мерсона (Merson's  $\sim$ ) 180
- Милне (Milne  $\sim$ ) 330
- Нумерова (Numerov's  $\sim$ ) 439, 443
- Ньютона (Newton's  $\sim$ ) 192
- опорный (basic  $\sim$ ) 238
- параллельной стрельбы (multiple shooting technique) 111
- последовательных приближений Пикара (Picard iteration  $\sim$ ) 48, 58
- Принса—Дормана (Prince and Dormand's  $\sim$ ) 208, 210
- прямых ( $\sim$  of lines) 10, 32, 399
- разделяющийся (partitioned  $\sim$ ) 272
- рядов Тейлора (Taylor series  $\sim$ ) 49
- сечений Пуанкаре ( $\sim$  of Poincaré sections) 116
- симметричный (symmetric  $\sim$ ) 232, 484
- сопряженный (adjoint  $\sim$ ) 230, 431
- стрельбы (пристрелки) (shooting  $\sim$ ) 110
- Хойна (Heun's  $\sim$ ) 140
- частичной дискретизации ( $\sim$  of lines, or  $\sim$  of discretization in space) 10
- Ческино (Ceschino's method) 179
- Эйлера (Euler's  $\sim$ ) 13, 38, 56, 135, 137, 325
- — неявный (implicit (backward)  $\sim \sim$ ) 212, 326
- — —, оценка погрешности (error estimate for  $\sim \sim$ ) 42, 43
- — —, сходимость (convergence of  $\sim \sim$ ) 38
- Энке (Encke's  $\sim$ ) 438
- Методы (methods)
- Адамса (Adams  $\sim$ ) 323
- — неявные (implicit  $\sim \sim$ ) 325
- — — с переменным шагом (variable step size  $\sim \sim$ ) 366
- — в представлении Нордсиека ( $\sim \sim$  as Nordsieck methods) 379, 381, 386, 387
- — явные (explicit  $\sim \sim$ ) 324
- Бутчера (Butcher's  $\sim$ ) 202, 217, 220
- гибридные (hybrid  $\sim$ ) 403

- для уравнений 2-го порядка ( $\sim$  for 2nd order equations) 276, 277, 288, 437
- коллокационные (collocation  $\sim$ ) 220, 267
- Кунцмана (Kuntzmann's  $\sim$ ) 217, 218, 219
- Кутты (Kutta's  $\sim$ ) 144
- линейные многошаговые ( $k$ -шаговые) (Linear multistep ( $k$ -step)  $\sim$ ) 334
- Милна—Симпсона (Milne—Simpson  $\sim$ ) 329
- многошаговые (multistep  $\sim$  (formulas)) 134, 322
- как общие линейные методы ( $\sim\sim$  as general linear methods) 404
- многоэтапные со старшими производными (multi-step multi-stage multi-derivative  $\sim$ ) 409
- модифицированные (modified  $\sim\sim$ ) 403
- неприводимые (irreducible  $\sim\sim$ ) 340
- обобщенные (generalized  $\sim\sim$ ) 403
- , паразитные решения (parasitic solutions of  $\sim\sim$ ) 346
- симметричные (symmetric  $\sim\sim$ ) 354
- с переменным шагом (variable step size  $\sim\sim$ ) 366
- , устойчивость (stability of  $\sim\sim$ ) 345, 347
- , характеристическое уравнение (characteristic equation of a  $\sim\sim$ ) 346
- циклические (cyclic  $\sim\sim$ ) 407
- , ядро Пеано (Peano kernel of a  $\sim\sim$ ) 341
- непрерывные (continuous  $\sim$ ) 188, 308, 321
- неприводимые (irreducible  $\sim$ ) 340
- Нордсиека (Nordsieck  $\sim$ ) 379
- как общие линейные методы ( $\sim\sim$  as general linear methods) 407
- , эквивалентность многошаговым методам ( $\sim\sim$ , equivalence with multistep methods) 382
- Ньюстрёма (Nyström  $\sim$ ) 277, 328
- как общие линейные методы ( $\sim\sim$  as general linear methods) 409
- , конструирование (construction of  $\sim\sim$ ) 285
- многошаговые явные (explicit multistep  $\sim\sim$ ) 328
- , сходимость (convergence of  $\sim\sim$ ) 287
- —, условия порядка (order conditions for  $\sim\sim$ ) 284, 302
- Обрешкова (Obreschkoff  $\sim$ ) 269
- общие линейные (general linear  $\sim$ ) 403, 408
- одношаговые (one-step  $\sim$ ) 134, 226
- предиктор—корректор (predictor—corrector process) 405
- РК (см. методы Рунге—Кутты) (RK)
- Рунге—Кутты (Runge—Kutta methods) 134, 139, 213
- — вложенные (embedded  $\sim\sim$ ) 178, 206
- — — диагонально неявные (diagonal implicit  $\sim\sim$ ) 213
- — — для уравнений с запаздывающим аргументом ( $\sim\sim$  for delay équation) 306
- — — непрерывные (continuous  $\sim\sim$ ) 188, 308, 321
- — — неявные (implicit  $\sim\sim$ ) 212, 214
- — — однократно диагонально неявные (singly diagonal implicit  $\sim\sim$ ) 213
- — — с нарушением условия (1.9) ( $\sim\sim$  violating (1.9)) 300
- — — со старшими производными (higher derivative  $\sim\sim$ ) 266
- — — явные (explicit  $\sim\sim$ ) 139, 151, 213
- — — 4-го порядка (4th order  $\sim\sim$ ) 140
- сильно устойчивые (strictly stable  $\sim$ ) 423
- симметричные (symmetric  $\sim$ ) 232, 354, 434
- слабо устойчивые (weakly stable  $\sim$ ) 428
- ФДН (BDF-  $\sim$ ) 308, 331
- — в представлении Нордсиека ( $\sim\sim$  as Nordsieck methods) 386, 388
- — с переменным шагом (variable step size  $\sim\sim$ ) 369, 375, 378
- —, устойчивость (stability of  $\sim\sim$ ) 347
- Фельберга (Fehlberg's  $\sim$ ) 181, 207
- — с высшими производными (multiderivative  $\sim\sim$ ) 270, 301
- Хаммера—Холлингсворта (Hammer and Hollingsworth's  $\sim$ ) 213, 214, 217
- Штёрмера (Stoermer's  $\sim$ ) 438, 439, 442
- экстраполяционные (extrapolation  $\sim$ ), 234, 288
- — как ЯРК-методы ( $\sim\sim$  as ERK)

- methods) 243, 250
- Эле (Ehle's ~) 224
- Милна* метод (Milne method) 330
- Милна—Симпсона* методы (Milne—Simpson methods) 329
- Многочлен интерполяционный *Эрмита* (Hermite interpolation polynomial) 267
  - коллокационный (collocation ~) 221
  - характеристический (characteristic ~) 22, 32, 76, 91, 93, 122
- Многочлены *Лагранжа* (Lagrange polynomials) 221
- производящие (generating ~) 336
- Моделирование эпидемии (infectious disease modelling) 313
- Модель *Лоренца* (Lorenz model) 125
- Монитор сходимости (convergence monitor) 245
  
- Начального шага длина (starting step size)** 194
- Начальное условие (значение) (initial value)** 9, 109
- Невязка (defect) 61, 63
- Нелинейная формула вариации постоянных (nonlinear variation-of-constants formula) 106
- Неоднородное линейное уравнение (inhomogeneous linear equation) 16, 22, 72
- Непрерывные вложенные формулы (continuous embedded formulas) 191
  - методы РК (~ RK methods) 188, 308, 321
- Неприводимые методы (irreducible methods) 340
- Неравенства дифференциальные (differential inequalities) 60, 96
  - для систем (~ ~ for systems) 67
- Неустойчивость численная (numerical instability) 345
- Невявные методы *Адамса* (implicit Adams methods) 325
  - — — в представлении *Нордсика* (~ ~ ~ as Nordsieck methods) 379, 381, 386, 387
  - — — *Рунге—Кутты* (~ RK ~) 212
  - — — как коллокационные методы (~ ~ ~ as collocation ~) 220
  - — — основанные на квадратурной формуле *Гаусса* (~ ~ ~ based on Gaussian formulas) 217
  - — — — — *Лобатто* (~ ~ ~ ~ Lobatto quadrature) 219
  - — — — — существование решения (existence of solution of ~ ~ ~) 214
- Нордсика* вектор (Nordsieck vector) 379, 387, 389
  - методы (~ methods) 379, 382, 407
- Норма вектора (vector norm) 56, 57
  - — евклидова (Euclidean ~ ~) 56, 65
- Матричная (matrix ~) 58
- логарифмическая (logarithmic ~ ~) 65, 68, 95
- подчиненная (subordinate ~ ~) 58
- невязки (defect) 61
- Шура (Schur norm) 58
- Нормальные матрицы (normal matrices) 86
- НРК (IRK) 214
- Нуль-устойчивость (zero-stability) 347
- Нумерова метод (Numerov's method) 439, 443
- Ньютона интерполяционная формула (Newton's interpolation formula) 324, 366
  - — — с кратными узлами (~ ~ ~ with multiple nodes) 267
  - метод (~ method) 192
- Нюстрём методы (Nyström methods) 277, 328
  - как общие линейные методы (~ ~ as general linear methods) 409
  - — —, конструирование (construction of ~ ~) 285
  - — — многошаговые явные (explicit multistep ~ ~) 328
  - — —, сходимость (convergence of ~ ~) 287
  - — —, условия порядка (order conditions for ~ ~) 284, 302
  
- Обратная задача о касательной (inverse tangent problem) 12
- Обрешкова методы (Obreschkoff methods) 269
- Общие линейные методы (general linear ~) 403, 408
- Ограничennaя задача трех тел (restricted three body problem) 134
- Одностороннее условие *Липшица* (one-sided Lipschitz condition) 67
- ОДНРК (SDIRK) 213
- Оптимальная длина шага (optimal step size) 177
- Оптимальные формулы (optimal formulas) 144
- Оптимальный порядок многошагового метода (optimal order (multistep)) 392
  - — — экстраполационного метода (~ ~ ~ (extrapolation)) 244
- Орегонатор (Oregonator) 123, 124

- Ортогональная матрица (orthogonal matrix) 77  
 Ортогональное преобразование ( $\sim$  transformation) 77  
 Особая точка (critical point) 122  
 — регулярная (слабая) (regular (weak) singular point) 27  
 Особенность уравнения (singularity) 25  
 «Отец» (*«father»*) 154  
 Отношение симметрии (symmetry relation) 434  
 — шагов (step size ratio) 371, 390  
 Отображение *Пуанкаре* (Poincaré map) 117, 121, 129, 131  
 — «сыновей» на «отцов» (son-father mapping) 154  
 Оценка погрешности (error estimate) 60  
 — для ломаной Эйлера ( $\sim \sim$  for Euler polygon) 42, 43  
 — надежная (reliable error bounds) 53  
 — строгая (rigorous  $\sim \sim$ ) 166
- Паде* аппроксимация экспоненты (Padé approximation to  $\exp(z)$ ) 94, 99  
 Паразитные решения (parasitic solutions) 346  
 Пеано теорема существования (existence theorem of Peano) 43, 46, 56  
 — ядро (Peano kernel) 341  
 Периметр эллипса (perimeter of the ellipse) 29  
 Периодические решения (periodic solutions) 118, 133, 135  
*Пикара* метод последовательных приближений (Picard iteration method) 48, 53, 58  
 — — — для систем ( $\sim \sim \sim$  for systems) 58  
 «Пирог с сыром мадам Имхоф» (Madam Imhof's cheese pie) 350  
 «Плеяды» (*«the Pleiades»*) 253  
 Плотная выдача (dense output) 136, 188, 191, 469  
 Погрешность глобальная (накопленная) (global error) 169, 226, 421  
 — локальная (local  $\sim$ ) 170, 226, 334, 409  
 — — численная оценка (numerical estimation of  $\sim \sim$ ) 390, 391  
 — округления (rounding  $\sim$ ) 449  
 Погрешности главный член (principal  $\sim$  term) 168  
 — константа ( $\sim$  constant) 338, 339, 383, 448  
 — коэффициенты ( $\sim$  coefficients) 168  
 — оценка ( $\sim$  estimate) 63, 166
- — для метода Эйлера ( $\sim \sim$  of Euler's method) 42, 43  
 — — практическая (practical  $\sim \sim$ ) 175, 390  
 Помеченное дерево (labelled tree) 153, 154  
 — — специальное (special  $\sim \sim$ ) 159  
 — — *N*-дерево ( $\sim N\sim$ ) 280  
 — — *P*-дерево ( $\sim P\sim$ ) 295  
 Порядок дерева (order of a tree) 155  
 — — помеченного ( $\sim$  of a labelled  $\sim$ ) 154  
 — — методов Адамса ( $\sim$  of Adams methods) 337, 338  
 — — многошаговых ( $\sim$  of multistep  $\sim$ ) 336  
 — — — с переменным шагом ( $\sim$  of variable step size  $\sim \sim$ ) 370  
 — — общих линейных ( $\sim$  of general linear  $\sim$ ) 411  
 — — Рунге–Кутты ( $\sim$  of RK  $\sim$ ) 139  
 — — экстраполяционных ( $\sim$  of extrapolation  $\sim$ ) 235  
 — — оптимальный (optimal  $\sim$ ) 244, 392  
 — — эффективный (effective  $\sim$ ) 264  
 Последовательность Булирша (Bulirsch sequence) 236  
 — гармоническая (harmonic  $\sim$ ) 236  
 — Ромберга (Romberg  $\sim$ ) 236  
 Правило  $«3/8»$  ( $3/8$ -rule) 144  
 — Симпсона (Simpson  $\sim$ ) 143  
 — средней точки (midpoint  $\sim$ ) 137, 328  
 — — неявное (implicit  $\sim \sim$ ) 212, 214  
 — трапецидальный (trapezoidal  $\sim$ ) 213, 326  
 — цепное (chain  $\sim$ ) 17  
 — Штермера (Stoermer's  $\sim$ ) 290  
 Предельный цикл (limit cycle) 115, 131  
 Предиктор (predictor) 327, 393, 405  
 Предсогласованности условия (preconsistency conditions) 415  
 Преобразование «греко-латинское» (Greek-Roman transformation) 351  
 — Лапласа дискретное (discrete Laplace  $\sim$ ) 362  
 — ортогональное (orthogonal  $\sim$ ) 77  
 Приведение матрицы к диагональному виду (diagonalization) 75  
 Примеры вычислений (numerical examples)  
 — —, методы РК 4-го порядка ( $\sim \sim$ , 4th order RK methods) 145, 185  
 — — — высших порядков ( $\sim, \sim, \sim$ , high  $\sim \sim \sim$ ) 209  
 — —, многошаговые методы ( $\sim \sim$ , multistep methods) 396  
 — —, сравнение программ ( $\sim \sim$ , com-

- parisons of codes) 252, 396  
 —, уравнения 2-го порядка ( $\sim\sim$ , 2nd order equations) 290  
 —, экстраполяционные методы ( $\sim\sim$ , extrapolation methods) 237  
*Принса—Дормана* метод (Prince and Dormand's method) 208, 210  
 Принцип аргумента (principle of the argument) 89, 349  
 Произведение Коши (Cauchy product) 51, 384  
 — Кронекера (тензорное) (Kronecker tensor product) 362  
 Производные Дини (Dini derivatives) 60  
 — по начальным значениям ( $\sim$  with respect to initial values) 105  
 — по параметру ( $\sim\sim\sim$  параметры) 103  
 — — —, численное определение (numerical computation of  $\sim\sim\sim$ ) 195  
 Производящие многочлены многошагового метода (generating polynomials of a multistep method) 336  
 — функции коэффициентов методов Адамса ( $\sim$  functions for the  $y_i$ ) 327  
 Процедура продвижения на шаг (forward step procedure) 404  
 — стартовая (starting  $\sim$ ) 323, 404  
 Процесс предиктор—корректор (predictor—corrector process) 405  
 Псевдометоды Рунге—Кутты (pseudo Runge—Kutta methods) 403  
 Пуанкаре—Бендиクссона теорема (theorem of Poincaré—Bendixson) 117, 121  
 — отображение ( $\sim$  map) 117, 121, 129, 131  
 — метод малого параметра (method of  $\sim$  for a small) 118, 132  
 — — сечений ( $\sim\sim\sim$  sections) 116, 129, 192
- Равностепенная непрерывность** (equi-continuity) 44, 45
- Радо* квадратурная формула (Radau scheme) 213
- Разделенные разности (divided differences) 366, 390, 393
- Разделяющаяся система уравнений (partitioned system) 272, 294
- Разделяющийся метод ( $\sim$  method) 272
- Разложение Тейлора точного решения (Taylor expansion of exact solution) 138, 157  
 — — решения методом РК ( $\sim\sim\sim$  RK  $\sim$ ) 138
- Разности назад (backward differences) 324
- разделенные (divided  $\sim$ ) 366, 390, 393
- Разностное уравнение (difference equation) 33
- Распространение звука (propagation of sound) 31
- Рауса* алгоритм (Routh algorithm) 94, 98  
 — таблица ( $\sim$  tableau) 93
- Рауса—Гурвица* критерий ( $\sim$ —Hurwitz criterion) 89
- Рациональная экстраполяция (rational extrapolation) 237
- Реализация многошаговых методов (implementation of multistep methods) 390
- Регулярная особая точка (regular singular point) 27
- Регулятор паровой машины (steam engine governor) 100, 101
- Резольвента (resolvent) 70, 72, 75, 79, 105
- Решение с помощью разложения в ряд (solution by series expansion) 11  
 — паразитное (parasitic  $\sim$ ) 346
- Риккати* уравнение (Riccati equation) 17, 47
- Ричардсоновская* экстраполяция (Richardson extrapolation) 175, 189, 233
- РК (RK) — см. *Рунге—Кутты* методы
- Ромберга последовательность (Romberg sequence) 236
- Рунге* методы (Runge's methods) 140  
 — правило 177
- Рунге—Кутты* методы (Runge—Kutta methods) 134, 139, 213  
 — — вложенные (embedded  $\sim\sim\sim$ ) 178, 206  
 — — — диагонально неявные (diagonal implicit  $\sim\sim\sim$ ) 213  
 — — — для уравнений с запаздывающим аргументом ( $\sim\sim\sim$  for delay equations) 306  
 — — — непрерывные (continuous  $\sim\sim\sim$ ) 188, 191, 308, 321  
 — — — неявные (implicit  $\sim\sim\sim$ ) 212, 214  
 — — — однократно диагонально неявные (singly diagonal  $\sim\sim\sim$ ) 213  
 — — — с нарушением условия (1.9) ( $\sim\sim\sim$  violating (1.9)) 300  
 — — — со старшими производными (higher derivatives  $\sim\sim\sim$ ) 266  
 — — — эквивалентные (equivalent  $\sim$ )

- ~) 265
- явные (explicit ~ ~ ~) 139, 151, 198, 213, 228, 243
- — — 4-го порядка (4th order ~ ~ ~ ~) 140
- псевдометоды (pseudo ~ ~ ~) 403
  
- Сечения Пуанкаре** (Poincaré sections) 116, 129
- Сильно устойчивые методы (strictly stable methods) 423
- Симметричные методы (symmetric ~) 232, 354, 434
- Симпсона правило (Simpson rule) 143
- Система автономная (autonomous system) 75
  - дифференциальных уравнений 1-го порядка (first order ~ of differential equations) 9, 34
  - — 2-го порядка (second ~ ~ ~ ~) 10, 276
  - — квазимонотонная (quasimonotone ~ ~ ~ ~) 46, 67
  - — линейных 1-го порядка (1st order linear ~ ~ ~ ~) 34, 69, 75
  - — —, геометрическое представление (~ ~ ~ ~ ~ ~ ~ ~, geometric representation) 83
  - — — 2-го порядка (2nd ~ ~ ~ ~ ~ ~ ~) 32
  - разделяющаяся (partitioned ~) 272, 294
  - многочленов Штурма (Sturm sequence) 90
- Слабая особая точка (weak singular point) 27
- Слабые особенности (~ singularities) 25
  - в системах (~ ~ for systems) 73
- Слабо устойчивые методы (weakly stable methods) 428
- След матрицы (trace of a matrix) 71
- Собственная функция (eigenfunction) 113
- Собственное значение (число) (eigenvalue) 32, 75, 86, 99
  - устойчивое (stable ~) 122
- Собственный вектор (eigenvector) 32, 75, 95
- Согласованность (consistency) 337, 411, 446
- Соотношения ортогональности (orthogonality relations) 34
- Специальное помеченное дерево (special labelled tree) 159
- Сравнение вычислительных свойств методов и программ (numerical comparison of methods and codes) 145, 185, 209, 237, 252, 290, 396
- Стабилизация (stabilization) 449
- Стартовая процедура (starting procedure) 323, 404
- Стационарная точка (stationary point) 75
- Стационарные приближения при больших  $\varepsilon$  (steady-state approximations for  $\varepsilon$  large) 117
- Странные аттракторы (strange attractors) 125
- Струна (string) 32
- Существование решения дифференциального уравнения (existence of a solution of a differential equation) 37, 47, 55
  - численного (~ of a numerical—) 214
- Сферический маятник (spherical pendulum) 122
- Сходимость метода Эйлера (convergence of Euler's method) 38
  - в строке (экстраполяционного метода (~ in line k) 245
  - методов многошаговых (~ of multistep methods) 359
  - — для уравнений 2-го порядка (~ ~ ~ ~ for 2nd order equations) 446
    - — с переменным шагом (~ ~ variable step size ~ ~) 376
    - — Ньюстрёма (~ ~ Nyström ~) 287
    - — общих линейных (~ ~ general linear) 412
    - — Рунге—Кутты (~ ~ Runge—Kutta ~) 166
    - «Сын» (son) 154
- Тейлора ряд (Taylor series) 49, 138, 151
  - —, доказательство сходимости (~ ~ convergence proof) 50
  - —, рекуррентное вычисление коэффициентов (recursive computation of Taylor coefficients) 51
- Тензорное произведение Кронекера (Kronecker tensor product) 362
- Теорема Ариела—Асколи (theorem of Arzelà—Ascoli) 44
  - Гершгорина (Gershgorin's ~) 99
  - Грёбнера—Алексеева (Gröbner—Alekseev ~) 106
  - о среднем («mean value ~») 57
  - Пуанкаре—Бендиксона (~ of Poincaré—Bendixson) 117, 121
  - сравнения Штурма (Sturm's comparison of

- rison  $\sim$ ) 112
  - существования общая (general existence  $\sim$ ) 37
  - Пеано ( $\sim \sim$  of Peano) 43
  - для систем уравнений ( $\sim \sim$  for systems of equations) 55
  - , доказательство методом Пикара ( $\sim \sim$  proof using Picard iteration) 48
  - , — рядов Тейлора ( $\sim \sim \sim$  Taylor series) 49
  - Штурма (Sturm's  $\sim$ ) 90
- Точка критическая (critical point) 83, 85, 117, 121
- особая (singular  $\sim$ , singularity) 26, 85
  - регулярная (слабая) (regular (weak)  $\sim \sim$ ) 27
- равновесия (equilibrium  $\sim$ ) 83
- стационарная (stationary  $\sim$ ) 75
  
- Уатта* регулятор (Watt's steam engine governor) 100
- Унитарная матрица (unitary matrix) 77
- Управление длиной шага для методов многошаговых (step size control for multistep methods) 390
  - Рунге—Кутты ( $\sim \sim \sim$  Runge—Kutta  $\sim$ ) 177, 183
  - — — экстраполяционных ( $\sim \sim \sim$  extrapolation  $\sim$ ) 244, 245, 247
  - порядком методов многошаговых (order control for multistep methods) 390
    - — — экстраполяционных ( $\sim \sim \sim$  extrapolation  $\sim$ ) 244, 247, 248
- Упрощающие предположения (simplifying assumptions) 141, 217, 418
- Уравнение Бернулли (Bernoulli equation) 19
  - Бесселя (Bessel  $\sim$ ) 29
  - Ван-дер-Поля (Van der Pol's  $\sim$ ) 115, 131, 132, 253
  - в вариациях (вариационное) (variational  $\sim$ ) 106
  - Вольтерры (Volterra's  $\sim$ ) 319, 321
  - гипергеометрическое (hypergeometric  $\sim$ ) 27, 29
  - Даламбера (d'Alembert's  $\sim$ ) 31
  - индексное (index  $\sim$ ) 26
  - Клеро (Clairaut  $\sim$ ) 15, 19
  - колеблющейся струны ( $\sim$  for the vibrating string) 32
  - Лапласа (формы капель) (differential  $\sim$  of Laplace (describing the form of drops)) 147, 323
  - математического маятника (pendulum  $\sim$ ) 132
    - неоднородное линейное (inhomogeneous linear  $\sim$ ) 16, 22, 72
    - разностное (difference  $\sim$ ) 33
    - Риккати (Riccati  $\sim$ ) 17, 47
    - теплопроводности (heat  $\sim$ ) 35, 111
    - характеристическое (characteristic  $\sim$ ) 22, 33, 89, 309, 346
    - Эйлера для вариационной задачи (Euler  $\sim$  for a variational problem) 14, 36
- Уравнения автономные (autonomous equations) 75
  - в полных дифференциалах (total differential  $\sim$ ) 16
  - в частных производных (partial  $\sim \sim$ ) 10, 399
  - второго порядка (second order  $\sim$ ) 9, 10, 17
    - — — не зависящие от  $x$  ( $\sim \sim \sim$  independent of  $x$ ) 17
  - дифференциальные, см. Дифференциальные уравнения (differential equations)
  - с запаздывающим аргументом (delay  $\sim \sim$ ) 304
  - жесткие (stiff  $\sim \sim$ ) 124, 214, 331
  - интегро-дифференциальные (integro- $\sim \sim$ ) 318, 320
  - линейные, см. Линейные дифференциальные уравнения (linear  $\sim \sim$ )
  - перманентные (permanant  $\sim$ ) 75
  - с особенностями ( $\sim$  with singularities) 25
  - с разделяющимися переменными ( $\sim \sim$  separable variables) 16
  - с разрывными производными (discontinuous  $\sim$ ) 192, 197, 247
  - Эйлера (Euler's  $\sim$ ) 29
- Условие корневое (root condition) 347
  - Липшица (Lipschitz  $\sim$ ) 15, 39, 57, 359
  - одностороннее (one-sided  $\sim \sim$ ) 64
  - ортогональности (orthogonality relation) 114
- Условия граничные (boundary conditions) 110
  - начальные (initial value) 9, 109
  - предсогласованности (preconsistency conditions) 415
  - согласованности (consistency  $\sim$ ) 337
    - для уравнений 2-го порядка ( $\sim \sim$  for 2nd order equations) 446
    - общих линейных методов ( $\sim \sim$  for general linear methods) 411
    - порядка методом многошаговых (of-

- der  $\sim$  for multistep methods) 334, 336
- — — Ньюстрёма ( $\sim \sim \sim$  Nyström  $\sim$ ) 281, 284
- — — общих линейных ( $\sim \sim \sim$  general linear  $\sim$ ) 415
- — — Рунге—Кутты ( $\sim \sim \sim$  Runge—Kutta  $\sim$ ) 150, 152, 162, 163, 263
- — — со старшими производными ( $\sim \sim \sim$  multiderivative  $\sim$ ) 272
- Устойчивость асимптотическая (asymptotic stability) 96
- дифференциального уравнения (см. — решений) (stability of a differential equation)
- линеаризованного уравнения ( $\sim$  of a linearized  $\sim$ ) 88
- многошаговых методов ( $\sim \sim$  multistep formulas) 347
- — для уравнений 2-го порядка ( $\sim \sim \sim$  for 2nd order equations) 444
- — — с переменным шагом ( $\sim \sim$  variable step size  $\sim \sim$ ) 371
- неавтономных систем ( $\sim \sim$  non-autonomous systems) 97
- нелинейных систем ( $\sim \sim$  nonlinear  $\sim$ ) 96
- общих линейных методов ( $\sim \sim$  general linear methods) 409
- по Ляпунову ( $\sim$  in the sense of Liapunov) 88
- сильная (strict  $\sim$ ) 423
- слабая (weak  $\sim$ ) 428
- решений дифференциального уравнения ( $\sim$  of a differential equation) 88, 105
- — — с запаздывающим аргументом ( $\sim \sim$  delay  $\sim \sim$ ) 310
- формул дифференцирования назад ( $\sim \sim$  the BDF-formulas) 347
- численная (numerical  $\sim$ ) 93, 345
  
- Фаа ди Бруно** формула (Faa di Bruno's formula) 159
- Фазовое пространство (phase-space) 83
- Фазовый поток (flow of the differential equation) 126
- ФДН-методы (формулы) (BDF-methods (formulas)) 308, 331
- в представлении Нордсиека ( $\sim \sim$  as Nordsieck methods) 386, 388
- — с переменным шагом ( $\sim$  variable step size  $\sim$ ) 369, 375, 378
- —, устойчивость (stability of BDF-formulas) 347
- Фейгенбаума каскады (Feigenbaum cascades) 129
- число ( $\sim$  number) 131
- Фельберга методы (Fehlberg's methods) 181, 207
- — с высшими производными ( $\sim$  multiderivative  $\sim$ ) 270, 301
- Форма капель (form of drops) 147, 323, 332
- Формула вариации постоянных (variation of constants formula) 72, 105, 108
- — — нелинейная (nonlinear  $\sim \sim$   $\sim \sim$ ) 106
- Гробнера—Алексеева (Gröbner—Alekseev  $\sim$ ) 106, 222
- Грина (Green's  $\sim$ ) 133
- десятого порядка (10-th order  $\sim$ ) 203
- Дормана—Принса (Dormand and Prince  $\sim$ ) 182
- интерполяционная Лагранжа (Lagrange interpolation  $\sim$ ) 222
- Ньютона (Newton's  $\sim \sim$ ) 324, 366
- — — с кратными узлами ( $\sim \sim \sim$  with multiple nodes) 267
- Лейбница (Leibniz'  $\sim$ ) 151
- Эйлера (Euler's  $\sim$ ) 33
- Эйлера—Маклорена (Euler—MacLaurin  $\sim$ ) 228, 232
- Формулы вложенные высоких порядков (embedded formulas of high order) 206
- непрерывные (continuous  $\sim \sim$ ) 191
- Рунге—Кутты ( $\sim$  RK  $\sim$ ) 178
- дифференцирования назад, см. ФДН-методы (backward differentiation  $\sim$ )
- квадратурные Гаусса (Gaussian quadrature  $\sim$ ) 137, 217
- Лобатто (Lobatto  $\sim \sim$ ) 219
- Радау (Radau  $\sim \sim$ ) 213
- линейные многошаговые (linear multistep  $\sim$ ) 323
- оптимальные (optimal  $\sim$ ) 144
- Пузе (Pouzet  $\sim$ ) 320
- Фельберга (Fehlberg  $\sim$ ) 180, 206
- Фундаментальная лемма (fundamental lemma) 62, 108
- система решений ( $\sim$  system of solutions) 70
- Функции гипергеометрические (hypergeometric functions) 27
- Ляпунова (Liapunov  $\sim$ ) 94, 99
- собственные (eigenfunctions) 113
- Якоби эллиптические (Jacobi el-

- liptic functions) 252  
**Функция приращения** (increment function) 169  
 — общих линейных методов ( $\sim \sim$  for general linear methods) 404  
 — точных значений (correct value  $\sim$ ) 404  
**Фурье ряды** (Fourier series) 34, 35
- Хаммера—Холлингсвортса метод** (Hammer and Hollingsworth's method) 213, 214, 217  
**Характеристические значения** (characteristic values) 309  
**Характеристический многочлен** ( $\sim$  polynomial) 22, 32, 76, 91, 93, 122  
**Характеристическое уравнение** ( $\sim$  equation) 22, 33, 89, 309, 346  
**Химическая кинетика** (chemical kinetics) 120  
**Хеун метод** (Heun's method) 140  
**Хопфа бифуркация** (Hopf bifurcation) 121
- Цепное правило** (chain rule) 17, 158  
**Цикл предельный** (limit cycle) 115, 131  
**Циклические многошаговые методы** (cyclic multistep methods) 407
- Ческино метод** (Ceschino's method) 179  
**Числа Бернуlli** (Bernoulli numbers) 228, 384  
 — Стирлинга (Stirling  $\sim$ ) 228  
**Численные примеры**, см. Вычислительные примеры (numerical examples)  
**Число деревьев** (number of trees) 156  
 — условий порядка ( $\sim \sim$  order conditions) 163  
 — Фейгенбаума (Feigenbaum  $\sim$ ) 131  
**Чувствительность решений** (sensitivity of the solutions) 105
- Штёрмера методы** (Stoermer's methods) 438, 439, 442  
 — правило ( $\sim$  rule) 290
- Штурма—Лиувилля задача** (Sturm—Liouville eigenvalue problem) 111  
 — система многочленов (Sturm sequence) 90  
 — теорема (Sturm's theorem) 90  
 — сравнения ( $\sim$  comparison  $\sim$ ) 112
- Шура—Кона критерий** (Schur—Cohn criterion) 356  
 — норма (Schur norm) 58  
 — разложение ( $\sim$  decomposition) 76, 87  
 — форма матрицы ( $\sim$  form) 79
- Эйлера ломаная** (Euler polygon) 13, 38  
 — метод ( $\sim$ 's method) 13, 38, 56, 135, 137, 325  
 — неявный (implicit (backward)  $\sim$ ) 212, 326  
 — оценка погрешности (error estimate of  $\sim \sim$ ) 42, 43  
 — сходимость (convergence of  $\sim \sim$ ) 38  
 — уравнение для вариационной задачи ( $\sim$  equation for a variational problem) 14, 36  
 — уравнения ( $\sim$ 's equations) 29  
 — формула ( $\sim$ 's formula) 33
- Эйлера—Маклорена формула** (Euler—MacLaurin  $\sim$ ) 228, 232
- Эйткена—Невилла алгоритм** (Aitken—Neville algorithm) 237
- Эквивалентность методов РК** (equivalence of RK methods) 265  
 — помеченных деревьев ( $\sim \sim$  labelled trees) 155  
 —  $N$ -деревьев ( $\sim \sim \sim N$ -trees) 282  
 —  $P$ -деревьев ( $\sim \sim \sim P$ -trees) 296
- Эксперименты численные** 290 (см. тж. Вычислительные примеры)
- Экстраполяционные методы** (extrapolation methods) 234  
 — для уравнений 2-го порядка ( $\sim \sim$  for 2nd order equations) 288  
**Экстраполяция к пределу**  $h \rightarrow 0$  ( $\sim$  to the limit  $h \rightarrow 0$ ) 234  
 — локальная (local  $\sim$ ) 210  
 — по Ричардсону (Richardson  $\sim$ ) 175, 189, 233  
 — рациональная (rational  $\sim$ ) 237
- Эле методы** (Ehle's methods) 224
- Элементарный дифференциал** (elementary differential) 154, 156, 281, 297, 298
- Эмдена—Фаулера уравнение** (Emden—Fowler equation) 30
- Энке метод** (Encke's method) 438
- Эпидемии моделирование** (infectious disease modelling) 313
- Эрмита интерполяционный многочлен** (Hermite's interpolation polynomial) 267
- Этапы внешние, внутренние** (external, internal stages) 408
- Ядро вырожденное** (degenerate kernel) 319, 320  
 — Пеано (Peano  $\sim$ ) 341
- Якоби матрица** (Jacobian) 105  
 — эллиптические функции ( $\sim$  elliptic functions) 252
- ЯРК (ERK)** 139, 213

# Оглавление

Предисловие к русскому изданию . . . . .	5
Предисловие . . . . .	6
<b>Глава I. Классическая математическая теория . . . . .</b>	<b>8</b>
I.1. Терминология . . . . .	9
I.2. Наиболее ранние дифференциальные уравнения . . . . .	11
Ньютона . . . . .	11
Лейбница . . . . .	12
Вариационное исчисление . . . . .	13
Клеро . . . . .	15
I.3. Уравнения, разрешимые в квадратурах и элементарных функциях . . . . .	16
Уравнение с разделяющимися переменными . . . . .	16
Неоднородное линейное уравнение . . . . .	16
Уравнения в полных дифференциалах . . . . .	16
Уравнения второго порядка . . . . .	17
Упражнения . . . . .	18
I.4. Линейные дифференциальные уравнения . . . . .	21
Уравнения с постоянными коэффициентами . . . . .	21
Вариация постоянных . . . . .	23
Упражнения . . . . .	24
I.5. Уравнения со слабыми особенностями . . . . .	25
Линейные уравнения . . . . .	26
Нелинейные уравнения . . . . .	28
Упражнения . . . . .	29
I.6. Системы уравнений . . . . .	31
Лагранж . . . . .	31
Фурье . . . . .	34
Упражнения . . . . .	35
I.7. Общая теорема существования . . . . .	37
Сходимость метода Эйлера . . . . .	38
Теорема существования Пеано . . . . .	43
Упражнения . . . . .	46
I.8. Теория существования решения, основанная на итерационных методах и рядах Тейлора . . . . .	47
Метод последовательных приближений Пикара . . . . .	48
Метод рядов Тейлора . . . . .	49
Доказательство сходимости . . . . .	50
Рекурсивное вычисление коэффициентов ряда Тейлора . . . . .	51
Упражнения . . . . .	53
I.9. Теория существования решения для систем уравнений . . . . .	55
Векторные обозначения . . . . .	56
Подчиненные матричные нормы . . . . .	58
Последовательные приближения Пикара для систем . . . . .	58
Упражнения . . . . .	59
I.10. Дифференциальные неравенства . . . . .	60
Введение . . . . .	60
Фундаментальные теоремы . . . . .	61
Оценки с использованием односторонних условий Липшица . . . . .	64
Упражнения . . . . .	67

I.11. Системы линейных дифференциальных уравнений . . . . .	69
Матрица Вронского (вронсиан) . . . . .	70
Тождество Абеля—Лиувилля—Якоби—Остроградского . . . . .	71
Неоднородные линейные уравнения . . . . .	72
Упражнения . . . . .	72
I.12. Системы с постоянными коэффициентами . . . . .	75
Линеаризация . . . . .	75
Приведение к диагональному виду . . . . .	76
Разложение Шура . . . . .	76
Численные расчеты . . . . .	78
Каноническая форма Жордана . . . . .	80
Геометрическое представление . . . . .	83
Упражнения . . . . .	85
I.13. Устойчивость . . . . .	88
Введение . . . . .	88
Критерий Рауса—Гурвица . . . . .	89
Вопросы численной реализации . . . . .	93
Функция Ляпунова . . . . .	94
Устойчивость нелинейных систем . . . . .	96
Устойчивость неавтономных систем . . . . .	97
Упражнения . . . . .	98
I.14. Производные по параметрам и начальным значениям . . . . .	102
Производная по параметру . . . . .	103
Производные по начальным значениям . . . . .	105
Нелинейная формула вариации постоянных . . . . .	106
Упражнения . . . . .	108
I.15. Краевые задачи и задачи на собственные значения . . . . .	109
Краевые задачи . . . . .	109
Задачи Штурма—Лиувилля на собственные значения . . . . .	111
Упражнения . . . . .	114
I.16. Периодические решения, предельные циклы, странные аттракторы . . . . .	115
Доказательство существования . . . . .	116
Стационарные приближения при $\varepsilon$ . . . . .	117
Асимптотические решения при малых $\varepsilon$ . . . . .	118
Химические реакции . . . . .	120
Предельные циклы в системах больших размерностей, бифуркация . . . . .	121
Хопфа . . . . .	121
Странные аттракторы . . . . .	125
Каскады Фейгенбаума . . . . .	129
Упражнения . . . . .	131
 Глава II. Методы Рунге—Кутты и экстраполяционные методы . . . . .	134
II.1. Первые методы Рунге—Кутты . . . . .	137
Метод Эйлера для решения начальной задачи . . . . .	137
Общая формулировка методов Рунге—Кутты . . . . .	139
Обсуждение методов порядка 4 . . . . .	140
«Оптимальные» формулы . . . . .	144
Численный пример . . . . .	145
Упражнения . . . . .	147
II.2. Условия порядка для методов Рунге—Кутты . . . . .	150
Производные точного решения . . . . .	152
Условия для порядка 3 . . . . .	152
Деревья и элементарные дифференциалы . . . . .	153
Разложение Тейлора для точного решения . . . . .	157
Формула Фаа ди Бруно . . . . .	158

Производные численного решения . . . . .	159
Условия порядка . . . . .	162
Упражнения . . . . .	163
<b>II.3.</b> Оценка погрешности и сходимость методов Рунге—Кутты . . . . .	166
Строгие оценки погрешности . . . . .	166
Главный член погрешности . . . . .	168
Оценка глобальной погрешности . . . . .	169
Упражнения . . . . .	173
<b>II.4.</b> Практическая оценка погрешности и выбор длины шага . . . . .	175
Экстраполяция по Ричардсону . . . . .	175
Автоматическое управление длиной шага . . . . .	177
Вложение формулы Рунге—Кутты . . . . .	178
Формула Дормана и Принса . . . . .	182
Численное исследование механизма управления длиной шага . . . . .	183
Численное сравнение методов 4-го порядка . . . . .	185
Упражнения . . . . .	187
<b>II.5.</b> Дальнейшие вопросы практических вычислений . . . . .	188
Плотная выдача . . . . .	188
Непрерывные вложенные формулы . . . . .	191
«Неявшая» выдача . . . . .	192
Уравнения с разрывными производными . . . . .	192
Длина начального шага . . . . .	194
Численное определение производных по начальным условиям и параметрам . . . . .	195
Упражнения . . . . .	196
<b>II.6.</b> Явные методы Рунге—Кутты высших порядков . . . . .	198
Барьеры Бутчера . . . . .	198
Шестистадийные процессы пятого порядка . . . . .	200
Семистадийные процессы шестого порядка . . . . .	202
Дальнейшие барьеры Бутчера . . . . .	202
Формула десятого порядка . . . . .	203
Вложенные формулы высоких порядков . . . . .	206
Численный пример . . . . .	209
Упражнения . . . . .	211
<b>II.7.</b> Неявные методы Рунге—Кутты . . . . .	212
Введение . . . . .	212
Существование численного решения . . . . .	214
Методы Кунцмана и Бутчера порядка $2s$ . . . . .	217
НРК-методы, основанные на квадратурной формуле Лобатто . . . . .	219
НРК как коллокационные методы . . . . .	220
Упражнения . . . . .	224
<b>II.8.</b> Асимптотическое разложение глобальной погрешности . . . . .	226
Локальная погрешность . . . . .	226
Глобальная погрешность . . . . .	226
Примеры . . . . .	228
Переменная длина шага . . . . .	229
Отрицательные значения $h$ . . . . .	229
Свойства присоединенного метода . . . . .	230
Симметричные методы . . . . .	232
Упражнения . . . . .	232
<b>II.9.</b> Экстраполяционные методы . . . . .	234
Определение методы . . . . .	234
Алгоритм Эйткена—Невилла . . . . .	237
Рациональная экстраполяция . . . . .	237
Вычислительный пример . . . . .	237

Экстраполяция с помощью симметричных методов . . . . .	238
Метод Грэгга, или ГБШ . . . . .	239
Сглаживающий шаг . . . . .	241
Вычислительный алгоритм и пример . . . . .	242
Асимптотическое разложение для нечетных индексов . . . . .	243
Существование явных методов Рунге—Кутты произвольного порядка . . . . .	243
Управление порядком и длиной шага . . . . .	244
Численное исследование комбинированного управления длиной шага и порядком . . . . .	247
Упражнения . . . . .	248
<b>II.10. Сравнение вычислительных качеств . . . . .</b>	<b>252</b>
Результаты расчётов . . . . .	255
Пример с негладким решением . . . . .	256
Заключение . . . . .	257
<b>II.11. Композиция В-рядов . . . . .</b>	<b>258</b>
Композиция методов Рунге—Кутты . . . . .	258
В-ряды . . . . .	259
Условия порядка для методов Рунге—Кутты . . . . .	263
«Эффектный порядок» Бутчера . . . . .	264
Упражнения . . . . .	265
<b>II.12. Методы, использующие старшие производные . . . . .</b>	<b>266</b>
Коллокационные методы . . . . .	267
Методы Фельберга . . . . .	270
Общая теория условий порядка . . . . .	272
Упражнения . . . . .	274
<b>II.13. Численные методы для дифференциальных уравнений второго порядка . . . . .</b>	<b>276</b>
Методы Ньюстрема . . . . .	277
Производные точного решения . . . . .	279
Производные численного решения . . . . .	282
Условия порядка . . . . .	284
О конструировании методов Ньюстрема . . . . .	285
Глобальная сходимость . . . . .	287
Программная реализация методов Ньюстрема . . . . .	288
Численные эксперименты . . . . .	290
Система высших порядков . . . . .	292
Упражнения . . . . .	292
<b>II.14. Р-ряды для разделяющихся обыкновенных дифференциальных уравнений . . . . .</b>	<b>294</b>
Производные точного решения; Р-деревья . . . . .	295
Р-ряды . . . . .	299
Методы Рунге—Кутты с нарушением условия (1.9) . . . . .	300
Методы Фельберга . . . . .	301
Методы Ньюстрема . . . . .	302
Упражнения . . . . .	303
<b>II.15. Дифференциальные уравнения с запаздывающим аргументом . . . . .</b>	<b>304</b>
Существование . . . . .	304
Методы с постоянной длиной шага для постоянного запаздывания . . . . .	306
Методы с переменной длиной шага . . . . .	308
Характеристические значения экспоненциальных решений . . . . .	309
Устойчивость . . . . .	310
Пример из динамики популяций . . . . .	311
Моделирование эпидемии . . . . .	313
Пример из кинетики ферментативных реакций . . . . .	315
Одна математическая модель в иммунологии . . . . .	317

Интегро-дифференциальные уравнения . . . . .	318
Упражнения . . . . .	319
<b>Глава III. Многошаговые методы и общие линейные методы . . . . .</b>	<b>322</b>
III.1. Классические линейные многошаговые формулы . . . . .	323
Явные методы Адамса . . . . .	324
Неявные методы Адамса . . . . .	325
Рекуррентные соотношения для $\Psi$ . . . . .	327
Явные методы Ньютрема . . . . .	328
Методы Милна—Симпсона . . . . .	329
Методы, основанные на дифференцировании . . . . .	330
Упражнения . . . . .	332
III.2. Локальная погрешность и условия порядка . . . . .	334
Локальная погрешность многошагового метода . . . . .	334
Порядок многошагового метода . . . . .	336
Константа погрешности многошаговых методов . . . . .	338
Неприводимые методы . . . . .	340
Ядро Пеано многошаговых методов . . . . .	341
Упражнения . . . . .	343
III.3. Устойчивость и первый барьер Далквиста . . . . .	345
Устойчивость формул дифференцирования назад . . . . .	347
Наивысший достижимый порядок устойчивых многошаговых методов . . . . .	351
Упражнения . . . . .	355
III.4. Сходимость многошаговых методов . . . . .	359
Представление в виде одношагового метода . . . . .	361
Доказательство сходимости . . . . .	363
Упражнения . . . . .	365
III.5. Многошаговые методы с переменным шагом . . . . .	366
Методы Адамса с переменным шагом . . . . .	366
Рекуррентные соотношения для $g_i(n)$ , $\Phi_j(n)$ и $\Phi_j^*(n)$ . . . . .	368
Формулы дифференцирования назад с переменным шагом . . . . .	369
Многошаговые методы общего вида с переменным шагом и их порядок согласованности . . . . .	370
Устойчивость . . . . .	371
Сходимость . . . . .	376
Упражнения . . . . .	378
III.6. Методы Нордика . . . . .	379
Эквивалентность многошаговым методом . . . . .	382
Неявные методы Адамса . . . . .	387
ФДН-методы . . . . .	388
Упражнения . . . . .	389
III.7. Реализация и численное сравнение . . . . .	390
Выбор шага и порядка . . . . .	390
Некоторые распространенные программы . . . . .	392
Сравнение численных результатов . . . . .	396
Уравнения в частных производных . . . . .	399
III.8. Общие линейные методы . . . . .	403
Общая процедура интегрирования . . . . .	404
Примеры метода (8.4) . . . . .	404
Устойчивость и порядок . . . . .	409
Сходимость . . . . .	412
Условия порядка для общих линейных методов . . . . .	415
Построение общих линейных методов . . . . .	417
Упражнения . . . . .	419
III.9. Асимптотическое разложение глобальной погрешности . . . . .	421
Поучительный пример . . . . .	421

Асимптотическое разложение для сильно устойчивых методов (8.4) . . . . .	423
Слабо устойчивые методы . . . . .	428
Сопряженный метод . . . . .	431
Симметричные методы . . . . .	434
Упражнения . . . . .	435
<b>III.10. Многошаговые методы для дифференциальных уравнений второго порядка . . . . .</b>	<b>437</b>
Первые методы . . . . .	438
Задача Штермера . . . . .	439
Методы более высокого порядка . . . . .	441
Общая формулировка . . . . .	443
Условия устойчивости . . . . .	444
Одношаговое представление метода (10.19) . . . . .	444
Согласованность и сходимость . . . . .	446
Асимптотическая формула для глобальной погрешности . . . . .	447
Порядковый барьер для устойчивых методов (10.19) . . . . .	449
Погрешности округления . . . . .	449
Упражнения . . . . .	450
<b>Приложение. Программы на Фортране . . . . .</b>	<b>452</b>
<b>Литература . . . . .</b>	<b>473</b>
<b>Дополнительная литература . . . . .</b>	<b>492</b>
<b>Указатель обозначений . . . . .</b>	<b>493</b>
<b>Предметный указатель . . . . .</b>	<b>495</b>

## Научное издание

Эрнст Хайрер, Сиверт Пауль Нёрсетт, Герхард Ваннер

**РЕШЕНИЕ ОБЫКНОВЕННЫХ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ**

Заведующий редакцией чл.-корр. АН СССР В. И. Арнольд

Зам. зав. редакцией А. С. Попов

Научн. редактор С. В. Чудов

Мл. научн. редактор Т. Ю. Дехтярева

Художник Г. М. Чеховский

Художественный редактор В. И. Шаповалов

Технический редактор А. Л. Гулина

Корректор Т. И. Стифеева

ИБ № 7025

Сдано в набор 20.02.89. Подписано к печати 16.11.89.

Формат 60×90<sup>1/16</sup>. Бумага офсетная № 2. Печать офсетная. Гарнитура литературная. Объем бум. л. 16,00.  
Усл. печ. л. 32,00. Усл. кр.-отт. 32,00. Уч.-изд. л. 27,38.  
Изд. № 1/6445. Тираж 14 000 экз. Зак. 736. Цена 2 р. 20 к.Издательство «Мир»  
В/О «Совэкспорткнига» Государственного комитета СССР  
по печати,  
129820, ГСП, Москва, 1-й Рижский пер., 2.Типография № 6 ордена Трудового Красного Знамени издательства «Машиностроение»  
при Государственном комитете СССР по печати,  
193144, г. Ленинград, ул. Монсенько, 10.