



Н. Н. Калиткин

ЧИСЛЕННЫЕ МЕТОДЫ



Н. Н. Калиткин

Численные методы

2-е издание

Под редакцией А. А. Самарского

Рекомендовано Научно-методическим советом по математике
Министерства образования и науки Российской Федерации
в качестве учебного пособия для студентов университетов
и высших технических учебных заведений

Санкт-Петербург

«БХВ-Петербург»

2011

УДК 519.95(075.8)
ББК 22.193я73
К17

Калиткин Н. Н.

К17 Численные методы: учеб. пособие. —
2-е изд., исправленное. — СПб.: БХВ-Петербург, 2011. —
592 с.: ил. — (Учебная литература для вузов)

ISBN 978-5-9775-0500-0

Излагаются основные численные методы решения широкого круга математических задач, возникающих при исследовании физических и технических проблем. Книга начинается с простейших задач интерполирования, дифференцирования и интегрирования функций, решения уравнений и систем уравнений, а кончается методами решения дифференциальных и интегральных уравнений, описывающих процессы в сплошных средах. Для каждого метода даны практические рекомендации по применению. Для лучшего понимания алгоритмов приведены примеры численных расчетов.

*Для студентов, аспирантов и преподавателей университетов
и технических вузов, научных работников и инженеров-исследователей,
имеющих дело с численными расчетами*

УДК 519.95(075.8)
ББК 22.193я73

Рецензенты:

А. В. Гулин, доктор. физ.-мат. наук, проф., зам. заведующего кафедрой вычислительных методов факультета вычислительной математики и кибернетики МГУ им. М. В. Ломоносова.

Группа подготовки издания:

Главный редактор	<i>Екатерина Кондукова</i>
Зам. главного редактора	<i>Евгений Рыбаков</i>
Зав. редакцией	<i>Григорий Добин</i>
Корректор	<i>Виктория Пиотровская</i>
Дизайн серии	<i>Инны Тачиной</i>
Оформление обложки	<i>Елены Беляевой</i>
Фото	<i>Кирилла Сергеева</i>
Зав. производством	<i>Николай Тверских</i>

Лицензия ИД № 02429 от 24.07.00. Подписано в печать 28.12.10.

Формат 60×90¹/₁₆. Печать офсетная. Усл. печ. л. 37.

Тираж 1200 экз. Заказ №

"БХВ-Петербург", 190005, Санкт-Петербург, Измайловский пр., 29.

Санитарно-эпидемиологическое заключение на продукцию
№ 77.99.60.953.Д.005770.05.09 от 26.05.2009 г. выдано Федеральной службой
по надзору в сфере защиты прав потребителей и благополучия человека.

Отпечатано с готовых диапозитивов
в ГУП "Типография "Наука"
190034, Санкт-Петербург, 9 линия, 12

ISBN 978-5-9775-0500-0

© Калиткин Н. Н., 2011
© Оформление, издательство "БХВ-Петербург", 2011

ОГЛАВЛЕНИЕ

Предисловие редактора	9
Предисловие к первому изданию	11
Предисловие ко второму изданию	14
Г л а в а I	
Что такое численные методы?	
§ 1. Математические модели и численные методы	15
1. Решение задачи (15). 2. Численные методы (17). 3. История прикладной математики (19).	
§ 2. Приближенный анализ	20
1. Понятие близости (20). 2. Структура погрешности (26). 3. Корректность (28).	
Задачи	30
Г л а в а II	
Аппроксимация функций	
§ 1. Интерполирование	31
1. Приближенные формулы (31). 2. Линейная интерполя- ция (32). 3. Интерполяционный многочлен Ньютона (33). 4. По- грешность многочлена Ньютона (36). 5. Применения интерполя- ции (38). 6. Интерполяционный многочлен Эрмита (41). 7. Схо- димость интерполяции (44). 8. Нелинейная интерполяция (46). 9. Интерполяция сплайнами (50). 10. Монотонная интерполя- ция (53). 11. Многомерная интерполяция (54).	
§ 2. Среднеквадратичное приближение	58
1. Наилучшее приближение (58). 2. Линейная аппроксима- ция (60). 3. Суммирование рядов Фурье (64). 4. Метод наимень- ших квадратов (67). 5. Нелинейная аппроксимация (71).	
§ 3. Равномерное приближение	74
1. Наилучшее приближение (74). 2. Нахождение равномерного приближения (77).	
Задачи	78

Глава III

Численное дифференцирование

1. Полиномиальные формулы (80). 2. Простейшие формулы (82). 3. Метод Рунге—Ромберга (85). 4. Квазиравномерные сетки (89). 5. Быстропеременные функции (92). 6. Регуляризация дифференцирования (93).

Задачи 96

Глава IV

Численное интегрирование

§ 1. Полиномиальная аппроксимация 97

1. Постановка задачи (97). 2. Формула трапеций (98). 3. Формула Симпсона (100). 4. Формула средних (102). 5. Формула Эйлера (103). 6. Процесс Эйткена (105). 7. Формулы Гаусса—Кристоффеля (107). 8. Формулы Маркова (111). 9. Сходимость квадратурных формул (112).

§ 2. Нестандартные формулы 114

1. Разрывные функции (114). 2. Нелинейные формулы (115). 3. Метод Филона (117). 4. Переменный предел интегрирования (120). 5. Несобственные интегралы (120).

§ 3. Кратные интегралы 124

1. Метод ячеек (124). 2. Последовательное интегрирование (127).

§ 4. Метод статистических испытаний 130

1. Случайные величины (130). 2. Разыгрывание случайной величины (131). 3. Вычисление интеграла (134). 4. Уменьшение дисперсии (136). 5. Кратные интегралы (138). 6. Другие задачи (141).

Задачи 142

Глава V

Системы уравнений

§ 1. Линейные системы 143

1. Задачи линейной алгебры (143). 2. Метод исключения Гаусса (145). 3. Определитель и обратная матрица (148). 4. О других прямых методах (149). 5. Прогонка (150). 6. Метод квадратного корня (153). 7. Плохо обусловленные системы (155).

§ 2. Уравнение с одним неизвестным 157

1. Исследование уравнения (157). 2. Дихотомия (158). 3. Удаление корней (159). 4. Метод простых итераций (160). 5. Метод Ньютона (162). 6. Процессы высоких порядков (164). 7. Метод

секущих (164). 8. Метод парабол (166). 9. Метод квадрирования (168).	
§ 3. Системы нелинейных уравнений.....	170
1. Метод простых итераций (170). 2. Метод Ньютона (172). 3. Методы спуска (173). 4. Итерационные методы решения линейных систем (173).	
Задачи.....	175

Глава VI

Алгебраическая проблема собственных значений

§ 1. Проблема и простейшие методы.....	177
1. Элементы теории (177). 2. Устойчивость (181). 3. Метод интерполяции (184). 4. Трехдиагональные матрицы (186). 5. Почти треугольные матрицы (188). 6. Обратные итерации (188).	
§ 2. Эрмитовы матрицы.....	193
1. Метод отражения (193). 2. Прямой метод вращений (199). 3. Итерационный метод вращений (201).	
§ 3. Неэрмитовы матрицы.....	206
1. Метод элементарных преобразований (206). 2. Итерационные методы (212). 3. Некоторые частные случаи (214).	
§ 4. Частичная проблема собственных значений.....	215
1. Особенности проблемы (215). 2. Метод линеаризации (216). 3. Степенной метод (217). 4. Обратные итерации со сдвигом (218).	
Задачи.....	220

Глава VII

Поиск минимума

§ 1. Минимум функции одного переменного.....	221
1. Постановка задачи (221). 2. Золотое сечение (223). 3. Метод парабол (226). 4. Стохастические задачи (228).	
§ 2. Минимум функции многих переменных.....	229
1. Рельеф функции (229). 2. Спуск по координатам (231). 3. Наискорейший спуск (235). 4. Метод оврагов (238). 5. Сопряженные направления (238). 6. Случайный поиск (243).	
§ 3. Минимум в ограниченной области.....	245
1. Формулировка задачи (245). 2. Метод штрафных функций (245). 3. Линейное программирование (247). 4. Симплекс-метод (250). 5. Регуляризация линейного программирования (251).	
§ 4. Минимизация функционала.....	253

1. Задачи на минимум функционала (253). 2. Метод пробных функций (257). 3. Метод Рунге (261). 4. Сеточный метод (264).	
Задачи	268
Глава VIII	
Обыкновенные дифференциальные уравнения	
§ 1. Задача Коши	269
1. Постановка задачи (269). 2. Методы решения (271). 3. Метод Пикара (273). 4. Метод малого параметра (275). 5. Метод ломаных (276). 6. Метод Рунге—Кутты (279). 7. Метод Адамса (284). 8. Неявные схемы (286). 9. Специальные методы (288). 10. Особые точки (292). 11. Сгущение сетки (293).	
§ 2. Краевые задачи	296
1. Постановки задач (296). 2. Метод стрельбы (298). 3. Уравнения высокого порядка (302). 4. Разностный метод; линейные задачи (304). 5. Разностный метод; нелинейные задачи (308). 6. Метод Галеркина (314). 7. Разрывные коэффициенты (317).	
§ 3. Задачи на собственные значения	319
1. Постановки задач (319). 2. Метод стрельбы (320). 3. Фазовый метод (321). 4. Разностный метод (323). 5. Метод дополненного вектора (326). 6. Метод Галеркина (328).	
Задачи	329
Глава IX	
Уравнения в частных производных	
§ 1. Введение	331
1. О постановках задач (331). 2. Точные методы решения (334). 3. Автомодельность и подобие (336). 4. Численные методы (338).	
§ 2. Аппроксимация	341
1. Сетка и шаблон (341). 2. Явные и неявные схемы (343). 3. Невязка (344). 4. Методы составления схем (346). 5. Аппроксимация и ее порядок (351).	
§ 3. Устойчивость	355
1. Неустойчивость (355). 2. Основные понятия (356). 3. Принцип максимума (360). 4. Метод разделения переменных (363). 5. Метод энергетических неравенств (367). 6. Операторные неравенства (369).	
§ 4. Сходимость	371
1. Основная теорема (371). 2. Оценки точности (374). 3. Сравнение схем на тестах (378).	
Задачи	381

Глава X

Уравнение переноса

§ 1. Линейное уравнение	382
1. Задачи и решения (382). 2. Схемы бегущего счета (385). 3. Геометрическая интерпретация устойчивости (390). 4. Многомерное уравнение (394). 5. Перенос с поглощением (396). 6. Монотонность схем (398). 7. Диссипативные схемы (401).	
§ 2. Квазилинейное уравнение	404
1. Сильные и слабые разрывы (404). 2. Однородные схемы (409). 3. Псевдовязкость (410). 4. Ложная сходимость (413). 5. Консервативные схемы (415).	
Задачи	418

Глава XI

Параболические уравнения

§ 1. Одномерные уравнения	420
1. Постановки задач (420). 2. Семейство неявных схем (422). 3. Асимптотическая устойчивость неявной схемы (427). 4. Монотонность (429). 5. Явные схемы (431). 6. Наилучшая схема (433). 7. Криволинейные координаты (438). 8. Квазилинейное уравнение (440).	
§ 2. Многомерное уравнение	444
1. Экономичные схемы (444). 2. Продольно-поперечная схема (446). 3. Локально-одномерный метод (450). 4. Метод Монте-Карло (455).	
Задачи	456

Глава XII

Эллиптические уравнения

§ 1. Счет на установление	457
1. Стационарные решения эволюционных задач (457). 2. Оптимальный шаг (460). 3. Чебышёвский набор шагов (466).	
§ 2. Вариационные и вариационно-разностные методы	470
1. Метод Рунге (470). 2. Стационарные разностные схемы (472). 3. Прямые методы решения (473). 4. Итерационные методы (478).	
Задачи	482

Глава XIII

Гиперболические уравнения

§ 1. Волновое уравнение	483
-------------------------------	-----

1. Схема «крест» (483). 2. Неявная схема (487). 3. Двуслойная акустическая схема (489). 4. Инварианты (494). 5. Явная многомерная схема (495). 6. Факторизованные схемы (497).	
§ 2. Одномерные уравнения газодинамики.....	500
1. Лагранжева форма записи (500). 2. Псевдовязкость (503).	
3. Схема «крест» (506). 4. Неявная консервативная схема (509).	
5. О других схемах (513).	
Задачи.....	514
Глава XIV	
Интегральные уравнения	
§ 1. Корректно поставленные задачи.....	515
1. Постановки задач (515). 2. Разностный метод (518). 3. Метод последовательных приближений (523). 4. Замена ядра вырожденным (524). 5. Метод Галеркина (526).	
§ 2. Некорректные задачи.....	527
1. Регуляризация (527). 2. Вариационный метод регуляризации (530). 3. Уравнение Эйлера (534). 4. Некоторые приложения (540). 5. Разностные схемы (544).	
Задачи.....	546
Глава XV	
Статистическая обработка эксперимента	
1. Ошибки эксперимента (548). 2. Величина и доверительный интервал (550). 3. Сравнение величин (559). 4. Нахождение стохастической зависимости (564).	
Задачи.....	571
Приложение. Ортогональные многочлены.....	572
Список литературы.....	575
Предметный указатель.....	582

ПРЕДИСЛОВИЕ РЕДАКТОРА

Современное развитие физики и техники тесно связано с использованием электронных вычислительных машин (ЭВМ). В настоящее время ЭВМ стали обычным оборудованием многих институтов и конструкторских бюро. Это позволило от простейших расчетов и оценок различных конструкций или процессов перейти к новой стадии работы — детальному математическому моделированию (вычислительному эксперименту), которое существенно сокращает потребность в натурных экспериментах, а в ряде случаев может их заменить.

В основе вычислительного эксперимента лежит решение уравнений математической модели численными методами. Изложению численных методов посвящено немало книг. Однако большинство этих книг ориентировано на студентов и научных работников математического профиля. Поэтому в настоящее время ощущается потребность в книге, рассчитанной на широкий круг читателей различных специальностей и сочетающей достаточную полноту изложения с разумной степенью строгости при умеренном объеме.

Предлагаемая книга отвечает этим требованиям. Она достаточно полно освещает тот круг вопросов, знание которого наиболее часто требуется в практике вычислений, и содержит ряд разделов, которые редко включают в учебные пособия. Умеренный объем достигнут за счет тщательного отбора материала и включения в книгу только наиболее эффективных и часто используемых на практике методов. Материал изложен четко и сжато, при этом большое внимание уделено рекомендациям по практическому применению алгоритмов; изложение пояснено рядом примеров. Для обоснования алгоритмов использован несложный математический аппарат, знакомый студентам физических и инженерных специальностей.

Книга рассчитана на читателя, который занимается не столько разработкой численных методов, сколько их применением к прикладным проблемам. Однако в процессе работы над книгой читатель знакомится с основными идеями построения вычислительных алгоритмов и с их обоснованием и приобретает знания, достаточные для разработки новых алгоритмов.

Эта книга является по существу введением в численные методы. Овладев ею, читатель затем может углубить свои знания, обратившись к руководствам по теории разностных схем и по методам численного решения отдельных классов задач.

Книга написана специалистом по теоретической и математической физике. Она возникла в результате работы автора над рядом актуальных проблем физики в Институте прикладной математики АН СССР и преподавания на физическом факультете МГУ.

Несомненно, книга окажется полезной широкому кругу читателей — студентам, аспирантам, научным сотрудникам и инженерам математических, физических и технических специальностей.

А. А. Самарский

ПРЕДИСЛОВИЕ К ПЕРВОМУ ИЗДАНИЮ

Сложные вычислительные задачи, возникающие при исследовании физических и технических проблем, можно разбить на ряд элементарных — таких как вычисление интеграла, решение дифференциального уравнения и т. п. Многие элементарные задачи являются несложными и хорошо изучены. Для этих задач уже разработаны методы численного решения, и нередко имеются стандартные программы решения их на ЭВМ. Есть и достаточно сложные элементарные задачи; методы решения таких задач сейчас интенсивно разрабатываются (например, решение уравнений бесстолкновительной плазмы).

Поэтому полная программа обучения численным методам должна состоять из ряда этапов. Во-первых, это освоение логарифмической линейки, клавишных вычислительных машин и программирования на ЭВМ. Во-вторых, основы численных методов, содержащие изложение классических элементарных задач (включая основные сведения о разностных схемах). В-третьих, курс теории разностных схем. И, в-четвертых, — ряд специальных курсов, которые сейчас нередко называют методами вычислительной физики: численное решение задач газодинамики, аэродинамики, переноса излучения, квантовой физики, квантовой химии и т. д.

Эта книга является введением в численные методы. Она начинается с простейших задач интерполирования функций и кончается недавно возникшим разделом вычислительной математики — методами решения некорректно поставленных задач. Книга написана на основе годового курса лекций, читавшихся автором сначала инженерам-конструкторам, а после переработки — студентам физического факультета МГУ.

Для каждой задачи существует множество методов решения. Например, хорошо обусловленную систему линейных уравнений можно решать методами Гаусса, Жордана, оптимального исключения, окаймления, отражений, ортогонализации и рядом других. Интерполяционный многочлен записывают в формах Лагранжа, Ньютона, Грегори — Ньютона, Бесселя, Стирлинга, Гаусса и Лапласа — Эверетта. Подобные методы обычно являются вариациями одного-

двух основных методов, и если даже в каких-то частных случаях имеют преимущества, то незначительные. Кроме того, многие методы создавались до появления ЭВМ, и ряд из них в качестве существенного элемента включает интуицию вычислителя. Появление ЭВМ потребовало переоценки старых методов, что до конца еще не сделано, и до сих пор по традиции большое количество неэффективных методов кочует из учебника в учебник. Отчасти это объясняется тем, что эффективность многих методов сильно зависит от мелких деталей алгоритма, почти не поддающихся теоретическому анализу; поэтому окончательный отбор лучших методов можно сделать только на основании большого опыта практических расчетов.

В этой книге сделана попытка такого отбора, опирающаяся на многолетний опыт решения большого числа разнообразных задач математической физики. Для большинства рассмотренных в книге задач изложены только наиболее эффективные методы с широкой областью применимости. Несколько методов для одной и той же задачи даны в том случае, если они имеют существенно разные области применимости, или если для данной задачи еще не разработано достаточно удовлетворительных методов.

Часто приходится слышать, что наступила эпоха ЭВМ, а «ручные» расчеты являются архаизмом. На самом деле это далеко не так. Прежде чем поручать ЭВМ большую задачу, надо сделать много оценочных расчетов и на их основе понять, какие методы окажутся эффективными для данной задачи. Конечно, даже в мелких расчетах ЭВМ с хорошим математическим обеспечением и набором периферийных устройств (телетайп, дисплей, графикопостроитель) оказывает большую пользу. Однако логарифмическая линейка и клавишные машины еще долго будут необходимы. Поэтому большинство методов, изложенных здесь, в равной мере пригодны для ЭВМ и «ручных» расчетов.

Основное внимание в книге уделено выработке практических навыков у читателя. Поэтому в первую очередь изложены алгоритмы, даны рекомендации по их применению и отмечены «маленькие хитрости» — те незначительные на первый взгляд практические приемы, которые сильно повышают эффективность алгоритма. Теоретическое обоснование методов приведено лишь в той мере, в какой оно необходимо для лучшего усвоения и практического применения.

В книгу включен ряд сведений, не относящихся к необходимому минимуму, но полезных читателю для лучшего понимания тонких

деталей вычислительных процессов. Чтобы не увеличивать объем книги и избежать сложных выкладок, эти сведения приведены, как правило, без доказательств, но со ссылками на дополнительную литературу. Некоторые сведения даны в форме задач в конце каждой главы.

Предполагается, что читатели знакомы с основами высшей математики, включая краткие сведения об уравнениях в частных производных. Необходимые дополнительные сведения, которые не содержатся в обязательных курсах университетов и вузов, сообщаются здесь в соответствующих разделах.

Книга разделена на главы, параграфы и пункты. В начале каждой главы кратко изложено ее содержание. Нумерация таблиц и рисунков — единая по всей книге, а нумерация формул — самостоятельная в каждой главе. Если ссылка не выходит за пределы данной главы, то указывается только номер формулы; если выходит — то номер главы и номер формулы. В конце книги дан список литературы. Приведенные в нем учебники и монографии рекомендуются для углубленного изучения отдельных разделов. Журнальные статьи даны для указания на оригинальные работы, их список не претендует на полноту; более полная библиография имеется в рекомендованных учебниках.

Общий подход к теории и практике вычислений, определивший стиль этой книги, сложился у меня под влиянием А. А. Самарского и В. Я. Гольдина за много лет совместной работы. Ряд актуальных тем был включен по инициативе А. Г. Свешникова и В. Б. Гласко. Много ценных замечаний сделали А. В. Гулин, Б. Л. Рождественский, И. М. Соболев, И. В. Фрязинов, Е. В. Шикин и сотрудники кафедры прикладной математической физики МИФИ. В оформлении рукописи мне помогли Л. В. Кузьмина и В. А. Красноярова. Я пользуюсь случаем искренне поблагодарить всех названных лиц, и в особенности Александра Андреевича Самарского.

Н. Н. Калиткин

ПРЕДИСЛОВИЕ КО ВТОРОМУ ИЗДАНИЮ

Первое издание вышло более 30-ти лет назад. Однако оно почти не устарело. В нем не хватает только методов решения жестких задач. Эти методы были первоначально разработаны для обыкновенных дифференциальных уравнений, но затем стали применяться к уравнениям в частных производных и к дифференциально-алгебраическим уравнениям. Чтобы их включить в книгу, пришлось бы вносить изменения в половину глав.

Поэтому в данном издании пришлось ограничиться лишь исправлением замеченных опечаток (их оказалось немного). Существенно изменен только список литературы. В нем выделены те немногие монографии и учебники, которые наиболее популярны у широкого круга вычислителей и рекомендуются в качестве первого чтения или компьютерного практикума. Список книг для дополнительного чтения значительно расширен. В частности, в него включена одна монография, специально посвященная жестким системам. В список журнальных статей добавлены наиболее существенные работы по жестким задачам, не включенные в указанную монографию.

Автор искренне благодарен Т. Г. Ермаковой и Л. В. Кузьминой за помощь в подготовке данного издания.

Н. Н. Калиткин

ГЛАВА I

ЧТО ТАКОЕ ЧИСЛЕННЫЕ МЕТОДЫ?

Глава I является вводной. В § 1 рассмотрены роль математики при решении физико-технических задач и место численных методов среди других математических методов и кратко изложена история численных методов. В § 2 разобраны основные понятия приближенного анализа: корректность постановки задач, определение близости точного и приближенного решений, структура погрешности.

§ 1. Математические модели и численные методы

1. Решение задачи. Физиков математика интересует не сама по себе, а как средство решения физических задач. Рассмотрим поэтому, как решается любая реальная задача — например, нахождение светового потока конструируемой лампы, производительности проектируемой химической установки или себестоимости продукции строящегося завода.

Одним из способов решения является эксперимент. Построим эту лампу, установку или завод и измерим интересующую нас характеристику. Если характеристика оказалась неудачной, то изменим проект и построим новый завод и т. д. Ясно, что мы получим достоверный ответ на вопрос, но слишком медленным и дорогим способом.

Другой способ — математический анализ конструкции или явления. Но такой анализ применяется не к реальным явлениям, а к некоторым математическим моделям этих явлений. Поэтому первая стадия работы — это *формулировка математической модели* (постановка задачи). Для физического процесса модель обычно состоит из уравнений, описывающих процесс; в эти уравнения в виде коэффициентов входят характеристики тел или веществ, участвующих в процессе. Например, скорость ракеты при вертикальном полете в вакууме определяется уравнением

$$\left(M - \int_0^t m(\tau) d\tau \right) \left(\frac{dv}{dt} + g \right) = cm(t), \quad (1)$$

где M — начальная масса ракеты, $m(t)$ — заданный расход горючего, g — ускорение поля тяготения, а c — скорость истечения газов, зависящая от калорийности топлива и среднего молекулярного веса продуктов сгорания.

Любое изучаемое явление бесконечно сложно. Оно связано с другими явлениями природы, возможно, не представляющими интереса для рассматриваемой задачи. Математическая модель должна охватывать важнейшие для данной задачи стороны явления. Наиболее сложная и ответственная работа при постановке задачи заключается в выборе связей и характеристик явления, существенных для данной задачи и подлежащих формализации и включению в математическую модель.

Если математическая модель выбрана недостаточно тщательно, то, какие бы методы мы ни применяли для расчета, все выводы будут недостаточно надежны, а в некоторых случаях могут оказаться совершенно неправильными. Так, уравнение (1) непригодно для запуска ракеты с поверхности Земли, ибо в нем не учтено сопротивление воздуха.

Вторая стадия работы — это *математическое исследование*. В зависимости от сложности модели применяются различные математические подходы. Для наиболее грубых и несложных моделей зачастую удается получить аналитические решения; это излюбленный путь многих физиков-теоретиков. Например, уравнение (1) легко интегрируется при $g = \text{const}$ и $m(t) = \text{const}$:

$$v = c \ln[M/(M - mt)] - gt.$$

Из-за грубости модели физическая точность этого подхода невелика; нередко такой подход позволяет оценить лишь порядки величин.

Для более точных и сложных моделей аналитические решения удается получить сравнительно редко. Обычно теоретики пользуются приближенными математическими методами (например, разложением по малому параметру), позволяющими получить удовлетворительные качественные и количественные результаты. Наконец, для наиболее сложных и точных моделей основными методами решения являются численные; как правило, они требуют проведения расчетов на ЭВМ. Эти методы зачастую позволяют добиться хорошего количественного описания явления, не говоря уже о качественном.

Во всех случаях математическая точность решения должна быть несколько (в 2–4 раза) выше, чем ожидаемая физическая точ-

ность модели. Более высокой математической точности добиваться бессмысленно, ибо общую точность ответа это все равно не повысит. Но более низкая математическая точность недопустима (для облегчения решения задачи нередко в ходе работы делают дополнительные математические упрощения; это снижает ценность результатов).

Наконец, третья стадия работы — это *осмысливание математического решения* и сопоставление его с экспериментальными данными. Если расчеты хорошо согласуются с контрольными экспериментами, то это свидетельствует о правильном выборе модели; такую модель можно использовать для расчета процессов данного типа. Если же расчет и эксперимент не согласуются, то модель необходимо пересмотреть и уточнить.

2. Численные методы являются одним из мощных математических средств решения задачи. Простейшие численные методы мы используем всюду, например, извлекая квадратный корень на листке бумаги. Есть задачи, где без достаточно сложных численных методов не удалось бы получить ответа; классический пример — открытие Нептуна по аномалиям движения Урана.

В современной физике таких задач много. Более того, часто требуется выполнить огромное число действий за короткое время, иначе ответ будет не нужен. Например, суточный прогноз погоды должен быть вычислен за несколько часов; коррекцию траектории ракеты надо рассчитать за несколько минут (напомним, что для расчета орбиты Нептуна Леверье потребовалось полгода); режим работы прокатного стана должен исправляться за секунды. Это невысказимо без мощных ЭВМ, выполняющих тысячи или даже миллионы операций в секунду.

Современные численные методы и мощные ЭВМ дали возможность решать такие задачи, о которых полвека назад могли только мечтать. Но применять численные методы далеко не просто. Цифровые ЭВМ умеют выполнять только арифметические действия и логические операции. Поэтому помимо разработки математической модели, требуется еще разработка алгоритма, сводящего все вычисления к последовательности арифметических и логических действий. Выбирать модель и алгоритм надо с учетом скорости и объема памяти ЭВМ: чересчур сложная модель может оказаться машине не под силу, а слишком простая — не даст физической точности.

Сам алгоритм и программа для ЭВМ должны быть тщательно

проверены. Даже проверка программы нелегка, о чем свидетельствует популярное утверждение: «В любой сколь угодно малой программе есть по меньшей мере одна ошибка». Проверка алгоритма еще более трудна, ибо для сложных алгоритмов не часто удается доказать сходимостью классическими методами. Приходится использовать более или менее надежные «экспериментальные» проверки, проводя пробные расчеты на ЭВМ и анализируя их (смотри, например, главу IX, § 4, п. 3).

Строгое математическое обоснование алгоритма редко бывает исчерпывающим исследованием. Например, большинство доказательств сходимости итерационных процессов справедливо только при точном выполнении всех вычислений; практически же число сохраняемых десятичных знаков редко превосходит 5–6 при «ручных» вычислениях и 10–12 при вычислениях на ЭВМ. Плохо поддаются теоретическому исследованию «маленькие хитрости» — незначительные на первый взгляд детали алгоритма, сильно влияющие на его эффективность. Поэтому окончательную оценку метода можно дать только после опробования его в практических расчетах.

К чему приводит пренебрежение этими правилами — видно из принципа некомпетентности Питера: «ЭВМ многократно увеличивает некомпетентность вычислителя».

Для сложных задач разработка численных методов и составление программ для ЭВМ очень трудоемки и занимают от нескольких недель до нескольких лет. Стоимость комплекса отлаженных программ нередко сравнима со стоимостью экспериментальной физической установки. Зато проведение отдельного расчета по такому комплексу много быстрее и дешевле, чем проведение отдельного эксперимента. Такие комплексы позволяют подбирать оптимальные параметры исследуемых конструкций, что не под силу эксперименту.

Однако численные методы не всемогущи. Они не отменяют все остальные математические методы. Начиная исследовать проблему, целесообразно использовать простейшие модели, аналитические методы и прикидки. И только разобравшись в основных чертах явления, надо переходить к полной модели и сложным численным методам; даже в этом случае численные методы выгодно применять в комбинации с точными и приближенными аналитическими методами.

Современный физик или инженер-конструктор для успешной

работы должен одинаково хорошо владеть и «классическими» методами, и численными методами математики.

3. История прикладной математики. Раздел математики, имеющий дело с созданием и обоснованием численных алгоритмов для решения сложных задач различных областей науки, часто называют прикладной математикой; американцы применение численных методов к физическим задачам называют вычислительной физикой. Главная задача прикладной математики — фактическое нахождение решения с требуемой точностью; этим она отличается от классической математики, которая основное внимание уделяет исследованию условий существования и свойств решения.

В истории прикладной математики можно выделить три основных периода.

Первый начался 3–4 тысячи лет назад. Он был связан с ведением конторских книг, вычислением площадей и объемов, расчетами простейших механизмов; иными словами — с несложными задачами арифметики, алгебры и геометрии. Вычислительными средствами служили сначала собственные пальцы, а затем — счеты. Исходные данные содержали мало цифр, и большинство выкладок выполнялось точно, без округлений.

Второй период начался с Ньютона. В этот период решались задачи астрономии, геодезии и расчета механических конструкций, сводящиеся либо к обыкновенным дифференциальным уравнениям, либо к алгебраическим системам с большим числом неизвестных. Вычисления выполнялись с округлением; нередко от результата требовалась высокая точность, так что приходилось сохранять до 8 значащих цифр.

Вычислительные средства стали разнообразнее: таблицы элементарных функций, затем — арифмометр и логарифмическая линейка; к концу этого периода появились неплохие клавишные машины с электромотором. Но скорость всех этих средств была невелика, и вычисления занимали дни, недели и даже месяцы.

Третий период начался примерно с 1940 г. Военные задачи — например, наводка зенитных орудий на быстро движущийся самолет — требовали недоступных человеку скоростей и привели к разработке электронных систем. Появились электронные вычислительные машины (ЭВМ).

Скорость даже простейших ЭВМ настолько превосходила скорость механических средств, что стало возможным проводить вычисления огромного объема. Это позволило численно решать новые

классы задач; например, процессы в сплошных средах, описываемые уравнениями в частных производных.

Сначала для решения этих задач использовались численные методы, разработанные в «доэлектронный» период. Но применение ЭВМ быстро привело к переоценке методов. Многие старые методы оказались неподходящими для автоматизированных расчетов. Стали быстро разрабатываться новые методы, ориентированные прямо на ЭВМ (например, метод Монте-Карло).

Мощности ЭВМ быстро растут. Если в 50-е гг. в СССР вступила в строй первая «Стрела» со скоростью 2000 операций в секунду и памятью 1024 ячейки, то сейчас во многих вычислительных центрах страны работают БЭСМ-6 со скоростью в 300 раз больше и памятью в 30 раз больше. А наилучшие современные ЭВМ имеют скорость до 30 миллионов операций в секунду при практически неограниченной оперативной памяти с прямой адресацией. Становятся возможными расчеты все более сложных задач. Это служит стимулом для разработки новых численных методов.

§ 2. Приближенный анализ

1. Понятие близости. Если требуется определить некоторую величину y по известной величине x , то символически задачу можно записать в виде $y = A(x)$. Здесь и y , и x могут быть числами, совокупностью чисел, функцией одного или нескольких переменных, набором функций и т. д. Если оператор A настолько сложен, что решение не удается явно выписать или точно вычислить, то задачу решают приближенно.

Например, пусть надо вычислить $y = \int_a^b x(t)dt$. Можно приближенно заменить $x(t)$ многочленом $\bar{x}(t)$ или другой функцией, интеграл от которой легко вычислить. А можно заменить интеграл суммой $\sum_t x(t_i)\Delta t_i$, вычислить которую тоже несложно. Таким образом, приближенный метод заключается в замене исходных данных на близкие данные \bar{x} и (или) замене оператора на близкий оператор \bar{A} , так чтобы значение $\bar{y} = \bar{A}(\bar{x})$ легко вычислялось. При этом мы ожидаем, что значение \bar{y} будет близко к искомому решению.

Но что такое «близко»? Очевидно, для двух чисел x_1 и x_2 надо требовать малости $|x_1 - x_2|$; а близость двух функций можно

определить разными способами. Эти вопросы рассматриваются в функциональном анализе, некоторые понятия которого будут сейчас изложены.

Множество элементов x любой природы называется *метрическим пространством*, если в нем введено расстояние $\rho(x_1, x_2)$ между любой парой элементов (*метрика*), удовлетворяющее следующим аксиомам:

- а) $\rho(x_1, x_2)$ — вещественное неотрицательное число,
- б) $\rho(x_1, x_2) = 0$, только если $x_1 = x_2$,
- в) $\rho(x_1, x_2) = \rho(x_2, x_1)$,
- г) $\rho(x_1, x_3) \leq \rho(x_1, x_2) + \rho(x_2, x_3)$.

Последовательность элементов x_n метрического пространства называется *сходящейся* (по метрике) к элементу x , если $\rho(x_n, x) \rightarrow 0$ при $n \rightarrow \infty$. Последовательность x_n называется *фундаментальной*, если для любого $\varepsilon > 0$ найдется такое $k(\varepsilon)$, что $\rho(x_n, x_m) < \varepsilon$ при всех n и $m > k$.

Метрическое пространство называют *полным*, если любая фундаментальная последовательность его элементов сходится к элементу того же пространства. Примером неполного пространства является множество рациональных чисел $x = (n/m)$ с метрикой $\rho(x_1, x_2) = |x_1 - x_2|$. Последовательность $x_k = (1 + 1/k)^k$ ему принадлежит, является фундаментальной, а сходится к иррациональному числу e , т. е. не к элементу данного пространства. Если переменные y, x принадлежат неполным пространствам, то обосновать сходимость численных методов очень трудно: даже если удастся доказать, что при $x_n \rightarrow x$ последовательность y_n фундаментальная, то отсюда еще не следует, что она сходится к элементу данного пространства, т. е. к решению допустимого класса.

Элементами наших множеств будут числа, векторы, матрицы, функции и т. п. Сами множества обычно являются линейными нормированными пространствами, ибо в них определены операции сложения элементов и умножения их на число и введена *норма* каждого элемента $\|x\|$, причем выполнены следующие аксиомы:

$$x_1 + x_2 = x_2 + x_1, \quad (x_1 + x_2) + x_3 = x_1 + (x_2 + x_3);$$

существует единственный элемент θ такой, что $x + \theta = x$ для любого x (будем использовать для θ обозначение 0); для всякого x существует единственный элемент $-x$ такой, что $x + (-x) = \theta$;

$$a(x_1 + x_2) = ax_1 + ax_2; \quad (a + b)x = ax + bx; \quad (3)$$

$$\begin{aligned} a(bx) &= (ab)x; \quad 1 \cdot x = x; \quad 0 \cdot x = \theta \text{ единствен}; \\ \|x\| &\geq 0 \text{ — вещественное число}; \quad \|ax\| = |a| \cdot \|x\|; \\ \|x\| &= 0 \text{ только при } x = 0; \quad \|x_1 + x_2\| \leq \|x_1\| + \|x_2\|. \end{aligned}$$

Линейное нормированное пространство есть частный случай метрического пространства, а норма определяется метрикой. Полное линейное нормированное пространство называется *банаховым*. Практически всегда величины, с которыми мы будем оперировать, являются элементами банаховых пространств; это важно при доказательстве сходимости численных методов.

Рассмотрим некоторые примеры банаховых пространств, с которыми нам часто придётся встречаться. Выполнимость аксиом (3) и полноту читатели легко проверят сами.

а) Множество всех действительных чисел с нормой $\|x\| = |x|$.

б) Пространство C — множество функций $x(t)$, определенных и непрерывных при $0 \leq t \leq 1$, с чебышёвской нормой $\|x\|_c = \max |x(t)|$. Сходимость в этом пространстве называется *равномерной*. Условие $0 \leq t \leq 1$ здесь и в следующем примере принято для удобства; оно не является существенным, и можно определять функции на любом конечном отрезке.

Класс непрерывных функций часто еще сужают, накладывая на функции дополнительные требования: липшиц-непрерывности, однократной или многократной дифференцируемости и т. д. Напомним некоторые определения.

Функция $x(t)$ называется *равномерно-непрерывной* на отрезке, если для сколь угодно малого $\omega > 0$ найдется такое δ , что $|x(t_1) - x(t_2)| \leq \omega$ для любой пары точек отрезка, удовлетворяющих условию $|t_1 - t_2| \leq \delta$. Таким образом, устанавливается функциональная связь между ω и δ . Величина $\omega(\delta)$ называется *модулем непрерывности* функции. Функция, непрерывная во всех точках замкнутого отрезка $a \leq t \leq b$, является на этом отрезке ограниченной и равномерно-непрерывной (теорема Кантора); следовательно, пространство C — множество ограниченных и равномерно-непрерывных функций. Если $\omega(\delta) \leq K\delta$, где K — некоторая константа, то функцию называют *липшиц-непрерывной*. Нетрудно видеть, что если функция имеет ограниченную производную, то она липшиц-непрерывна, причем $K = \sup |x'(t)|$.

в) Пространство L_p — множество функций $x(t)$, определенных при $0 \leq t \leq 1$ и интегрируемых по модулю с p -й степенью, если норма определена

$$\|x\|_{L_p} = \left[\int_0^1 |x(t)|^p dt \right]^{1/p}.$$

Сходимость в такой норме называют сходимостью *в среднем*. Пространство L_2 называют *гильбертовым*, а сходимость в нем — *среднеквадратичной*.

Разницу между равномерной близостью и близостью в среднем иллюстрирует рис. 1. Функция x_2 равномерно близка к функции x_1 , а функция x_3 близка в среднем, т. е. мало отличается от x_1 на большей части отрезка, но может сильно отличаться от нее на небольших участках.

Выбирая метрические пространства, т. е. выбирая множества X , Y и определяя в них метрики, мы тем самым уславливаемся, в каких классах функций можно брать начальные данные и искать решение. Поэтому в конкретной задаче выбор пространств должен в первую очередь определяться физическим смыслом задачи, и лишь во вторую — чисто математическими соображениями (такими, например, как возможность доказать сходимость). Например, при расчете прочности самолета нужна равномерная близость приближенного решения к точному, а близости в среднем недостаточно: перенапряжение в маленьком участке может разрушить конструкцию. А в задаче о нагреве тела потоком тепла даже норма L_1 удовлетворительна, ибо температура тела определяется интегралом от потока по времени.

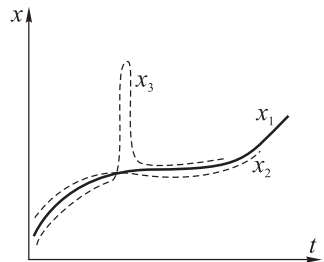


Рис. 1

Нетрудно показать, что между разными нормами (если они существуют) выполняются определенные соотношения. Если функции $x(t)$ определены при $0 \leq t \leq 1$, тогда

$$\|x(t)\|_{L_1} \leq \|x(t)\|_{L_2} \leq \dots \leq \|x(t)\|_C. \quad (4)$$

В самом деле, например:

$$\|x(t)\|_{L^p}^p = \int_0^1 |x(t)|^p dt \leq \int_0^1 \max |x(t)|^p dt = \max |x(t)|^p = \|x(t)\|_C^p.$$

Следовательно, из равномерной сходимости вытекает сходимость в среднем, в частности — среднеквадратичная. Поэтому чебышёвскую норму называют *более сильной*, чем гильбертову.

г) Координатные бесконечно мерные пространства, элементами которых являются счетные множества чисел $x = \{x_1, x_2, \dots\}$. По аналогии с пространствами функций, в них обычно вводят норму $\|x\|_c = \sup |x_i|$ или

$$\|x\|_{l_p} = \lim_{n \rightarrow \infty} \left(\frac{1}{n} \sum_{i=1}^n |x_i|^p \right)^{1/p},$$

а само пространство называют соответственно c или l_p .

д) Конечномерные пространства $c^{(n)}$, $l_p^{(n)}$, элементами которых являются группы из n чисел $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$; их можно считать координатами векторов в n -мерном пространстве, $l_2^{(n)}$ называют евклидовым. Нормы векторов вводят по аналогии со случаем (г), например,

$$\|\mathbf{x}\|_p = \left(\frac{1}{n} \sum_{i=1}^n |x_i|^p \right)^{1/p}.$$

Для конечномерных векторов между разными нормами существуют соотношения

$$\|\mathbf{x}\|_1 \leq \|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_c \leq \sqrt{n} \|\mathbf{x}\|_2 \leq n \|\mathbf{x}\|_1, \quad (5)$$

которые легко проверить. Поэтому из сходимости в одной из этих норм следует сходимость во всех остальных нормах. Нормы, обладающие таким свойством, называют *эквивалентными*.

Отметим, что если последовательность векторов \mathbf{x}_m не сходится, но $\mathbf{x}_m / \|\mathbf{x}_m\|$ сходится, то говорят о сходимости векторов *по направлению*.

е) В пространстве квадратных матриц порядка n наиболее употребительны следующие нормы:

$$\begin{aligned} \|A\|_c &= \max_i \left(\sum_{j=1}^n |a_{ij}| \right), & \|A\|_1 &= \max_j \left(\sum_{i=1}^n |a_{ij}| \right), \\ \|A\|_M &= n \cdot \max_{i,j} |a_{ij}|, & \|A\|_E &= \left(\sum_{i,j=1}^n |a_{ij}|^2 \right)^{1/2}, \\ \|A\|_2 &= \sqrt{\max \mu_i}, \end{aligned} \quad (6)$$

где μ_i — собственные значения эрмитовой матрицы $A^H A$ (здесь A^H — матрица, эрмитово сопряженная по отношению к A). Первые две нормы не имеют специальных названий, третья называется максимальной, четвертая — сферической или евклидовой и пятая — спектральной. Между ними выполняются некоторые соотношения, аналогичные (5).

Интересна связь между нормами матриц и векторов, на которые матрицы действуют. Норма матрицы называется *согласованной* с нормой вектора, если $\|A\mathbf{x}\| \leq \|A\| \cdot \|\mathbf{x}\|$. Наименьшая из норм матрицы, согласованных с данной нормой вектора: $\|A\| = \sup(\|A\mathbf{x}\|/\|\mathbf{x}\|)$, называется нормой матрицы, *подчиненной* данной норме вектора.

Приведем пример подчиненной нормы. Из цепочки неравенств

$$\begin{aligned} \|A\mathbf{x}\|_c &= \max_i \left| \sum_{j=1}^n a_{ij} \mathbf{x}_j \right| \leq \max_i \left[\left(\max_j |\mathbf{x}_j| \right) \sum_{k=1}^n |a_{ik}| \right] = \\ &= \|\mathbf{x}\|_c \cdot \max_i \left(\sum_{k=1}^n |a_{ik}| \right) = \|A\|_c \cdot \|\mathbf{x}\|_c \end{aligned} \quad (7)$$

следует, что $\|A\|_c$ согласована с $\|\mathbf{x}\|_c$. Кроме того, для любой матрицы A существует такой вектор \mathbf{x} , что неравенство (7) обращается в равенство. Для его нахождения положим $x_j = \pm 1$; знаки выберем так, чтобы они совпадали со знаками элементов a_{ij} той строки матрицы i , в которой $\sum_{j=1}^n |a_{ij}|$ максимальна.

Тогда именно сумма по этой строке будет максимальна в левой части (7), и неравенство превратится в равенство. Это означает, что $\|A\|_c$ есть наименьшая из норм, согласованных с $\|\mathbf{x}\|_c$: если мы возьмем еще меньшую $\|A\|$, то при этом векторе \mathbf{x} для нее знак неравенства (7) будет обратным, т. е. она не будет согласованной. Следовательно, $\|A\|_c$ подчинена $\|\mathbf{x}\|_c$.

Без доказательства укажем, что $\|A\|_c$ подчинена $\|\mathbf{x}\|_1$, и спектральная норма подчинена $\|\mathbf{x}\|_2$. Сферическая норма согласована с

$\|\mathbf{x}\|_2$, а максимальная норма согласована со всеми рассмотренными выше векторными нормами.

2. Структура погрешности. Есть четыре источника погрешности результата: математическая модель, исходные данные, приближенный метод и округления при вычислениях. Погрешность математической модели связана с физическими допущениями и здесь рассматриваться не будет.

Исходные данные зачастую неточны; например, это могут быть экспериментально измеренные величины. В прецизионных физических измерениях точность доходит до 10^{-12} , но уже характерная астрономическая и геодезическая точность равна 10^{-6} , а во многих физических и технических задачах погрешность измерения бывает 1–10%. Погрешность исходных данных δx приводит к так называемой *неустранимой* (она не зависит от математика) погрешности решения $\delta y = A(x + \delta x) - A(x)$. В следующем пункте будут рассмотрены случаи, когда неустраняемая погрешность может становиться недопустимо большой.

Погрешность метода связана с тем, что точные оператор и исходные данные заменяются приближенными. Например, заменяют интеграл суммой, производную — разностью, функцию — многочленом или строят бесконечный итерационный процесс и обрывают его после конечного числа итераций. Методы строятся обычно так, что в них входит некоторый параметр; при стремлении параметра к определенному пределу погрешность метода стремится к нулю, так что эту погрешность можно регулировать. Погрешность метода мы будем исследовать при рассмотрении конкретных методов.

Погрешность метода целесообразно выбирать так, чтобы она была в 2–5 раз меньше неустранимой погрешности. Большая погрешность метода снижает точность ответа, а заметно меньшая — невыгодна, ибо это обычно требует значительного увеличения объема вычислений.

Вычисления как на бумаге, так и на ЭВМ выполняют с определенным числом значащих цифр. Это вносит в ответ *погрешность округления*, которая накапливается в ходе вычислений.

Рассмотрим накопление погрешности при простейших вычислениях. Пусть исходные данные x_i известны с относительной погрешностью $\Delta_i > 0$, т. е. заключены между $x_i(1 - \Delta_i)$ и $x_i(1 + \Delta_i)$; их абсолютные погрешности равны $\Delta_i|x_i|$. Тогда при сложении или вычитании двух чисел результат равен $x_1 \pm x_2$ с абсолютной по-

грешностью не более $\Delta_1|x_1| + \Delta_2|x_2|$, т. е. при этих операциях абсолютные погрешности складываются. При умножении (делении) результат равен $x_1x_2(x_1/x_2)$ с относительной погрешностью не более $\Delta_1 + \Delta_2$, т. е. складываются относительные погрешности. На современных ЭВМ числа записываются с 10–12 десятичными знаками, поэтому в расчете на них погрешность единичного округления $\Delta = 10^{-10} \div 10^{-12}$ обычно пренебрежимо мала по сравнению с погрешностью метода и неустранимой погрешностью.

При решении больших задач выполняются миллиарды действий. Казалось бы, начальные ошибки возрастут в 10^9 раз и погрешность ответа будет огромной. Однако при отдельных действиях фактические погрешности чисел могут иметь разные знаки и компенсировать друг друга. Согласно статистике при N одинаковых действиях среднее значение суммарной ошибки превышает единичную примерно в \sqrt{N} раз, а вероятность заметного отклонения суммарной ошибки от среднего значения очень мала. Значит, если нет систематических причин, то случайное накопление ошибок не слишком существенно.

Систематические причины возникают, например, если алгоритм таков, что в нем есть вычитание близких по величине чисел: хотя абсолютная ошибка при этом невелика, относительная ошибка $\Delta = (\Delta_1|x_1| + \Delta_2|x_2|)/(x_1 - x_2)$ может стать большой. Например, при нахождении корней квадратного уравнения по обычной формуле

$$x^2 + px - q = 0, \quad x_{1,2} = -0,5p \pm \sqrt{0,25p^2 + q}$$

в случае, когда $0 < q \ll p$, относительная ошибка округления для положительного корня x_1 велика. Это надо заранее предусмотреть и преобразовать формулу так, чтобы избавиться от подобных вычитаний:

$$x_1 = q/(0,5p + \sqrt{0,25p^2 + q}).$$

Этот пример очень прост. Существуют гораздо более сложные алгоритмы, где ошибки округления очень опасны: например, нахождение корней многочлена очень высокой степени (глава V, § 2, п. 8) или итерационное решение разностных схем для эллиптических уравнений при помощи чебышёвского набора параметров (глава XII, § 1). В этих случаях только после серьезного исследования удалось так видоизменить алгоритм, чтобы довести ошибки округления до безопасного уровня.

Отметим, что в большинстве подобных задач неприятностей можно избежать, проводя расчет с двойной или тройной точностью. Такая возможность реализована в хороших математических обеспечениях ЭВМ; это в несколько раз увеличивает время расчета, зато позволяет пользоваться уже известными алгоритмами, а не разрабатывать новые.

При любых расчетах справедливо правило: надо удерживать столько значащих цифр, чтобы погрешность округления была существенно меньше всех остальных погрешностей.

3. Корректность. Задача $y = A(x)$ называется *корректно поставленной*, если для любых входных данных x из некоторого класса решение y существует, единственно и устойчиво по входным данным. Рассмотрим это определение подробнее.

Чтобы численно решать задачу $y = A(x)$, надо быть уверенным в том, что искомое решение существует. Естественно также требовать единственности решения точной задачи: численный алгоритм — однозначная последовательность действий, и она может привести к одному решению. Но этого мало.

Нас интересует решение y , соответствующее исходным данным x . Но реально мы имеем входные данные с погрешностью $x + \delta x$ и находим $y + \delta y = A(x + \delta x)$. Следовательно, неустранимая погрешность решения равна $\delta y = A(x + \delta x) - A(x)$. Если решение непрерывно зависит от входных данных, т. е. всегда $\|\delta y\| \rightarrow 0$ при $\|\delta x\| \rightarrow 0$, то задача называется *устойчивой* по входным данным; в противном случае задача неустойчива по входным данным.

Рассмотрим классический пример неустойчивости — задачу Коши для эллиптического уравнения в полуплоскости $y \geq 0$:

$$u_{xx} + u_{yy} = 0, \quad u(x, 0) = 0, \quad u_y(x, 0) = \varphi(x). \quad (8)$$

Входными данными является $\varphi(x)$. Если $\bar{\varphi}(x) = 0$, то задача имеет только тривиальное решение $\bar{u}(x, y) = 0$. Если же $\varphi_n(x) = \frac{1}{n} \cos nx$, то решением будет

$$u_n(x, y) = \frac{1}{n^2} \cos nx \cdot \operatorname{sh} ny.$$

Очевидно, $\varphi_n(x)$ равномерно сходятся к $\bar{\varphi}(x)$ при $n \rightarrow \infty$; но при этом если $y \neq 0$, то $u_n(x, y)$ не ограничено и никак не может сходиться к $\bar{u}(x, y)$. Этот пример связан с физической задачей о тяжелой жидкости, налитой поверх легкой; при этом действительно возникает так называемая релей-тейлоровская неустойчивость.

Отсутствие устойчивости обычно означает, что даже сравнительно небольшой погрешности δx соответствует весьма большое δy , т. е. получаемое в расчете решение будет далеко от искомого. Непосредственно к такой задаче численные методы применять бессмысленно, ибо погрешности, неизбежно появляющиеся при численном расчете, будут катастрофически нарастать в ходе вычислений.

Правда, сейчас развиты методы решения многих некорректных задач. Но они основаны на решении не исходной задачи, а близкой к ней вспомогательной корректно поставленной задачи, содержащей параметр α ; при $\alpha \rightarrow 0$ решение вспомогательной задачи должно стремиться к решению исходной задачи. Примеры таких методов (называемых регуляризацией) даны в следующих двух главах, а их строгое обоснование приведено в главе XIV, § 2.

На практике даже не всякую устойчивую задачу легко решить. Пусть $\|\delta y\| \leq C\|\delta x\|$, причем константа C очень велика. Задача формально устойчива, но фактическая неустранимая ошибка может быть большой. Этот случай называют *слабой* устойчивостью (или плохой обусловленностью). Примером является такая задача:

$$y''(x) = y(x), \quad (9a)$$

$$y(0) = 1, \quad y'(0) = -1. \quad (9б)$$

Общее решение дифференциального уравнения (9a) есть:

$$y(x) = 0,5[y(0) + y'(0)]e^x + 0,5[y(0) - y'(0)]e^{-x}.$$

Начальным условиям (9б) соответствует точное решение $y(x) = e^{-x}$; но небольшая погрешность начальных данных может привести к тому, что в решении добавится член вида ϵe^x , который при больших аргументах много больше искомого решения.

Очевидно, для хорошей практической устойчивости расчета константа C должна быть не слишком велика. Так, если начальные данные известны точно, т. е. могут быть заданы с точностью до ошибок округления $\Delta \sim 10^{-12}$, то необходимо, чтобы $C \ll 10^{12}$. Если же начальные данные найдены из эксперимента с точностью $\delta x \sim 0,001$, а требуемая точность решения $\delta y \sim 0,1$, то допустимо $C \leq 100$.

Даже если задача устойчива, то численный алгоритм может быть неустойчивым. Например, если производные заменяются разностями, то приходится вычитать близкие числа и сильно теряется

точность. Эти неточные промежуточные результаты используются в дальнейших вычислениях, и ошибки могут сильно нарастать.

По аналогии можно говорить о корректности алгоритма $\bar{y} = \bar{A}(\bar{x})$, подразумевая существование и единственность приближенного решения для любых входных данных \bar{x} некоторого класса, и устойчивость относительно всех ошибок в исходных данных и промежуточных выкладках. Однако в общем случае этим определением трудно пользоваться; только в теории разностных схем (глава IX) оно применяется успешно.

ЗАДАЧИ

1. Доказать выполнимость всех соотношений (4). Рассмотреть, как меняется форма записи этих соотношений при задании функции на произвольном конечном отрезке $a \leq t \leq b$.

2. Доказать утверждения о согласованности и подчиненности норм матриц, приведенные в конце п. 1 § 2.

ГЛАВА II

АППРОКСИМАЦИЯ ФУНКЦИЙ

В главе II рассмотрены способы построения приближенных формул для заданной функции. В § 1 изложен способ интерполяции; он несложен и обеспечивает хорошую точность на небольших отрезках. В § 2 рассмотрена среднеквадратичная аппроксимация, частным случаем которой является метод наименьших квадратов; она позволяет строить приближенные формулы, пригодные на больших отрезках. В § 3 кратко изложены основные сведения о равномерной аппроксимации.

§ 1. Интерполирование

1. Приближенные формулы. Если задана функция $y(x)$, то это означает, что любому допустимому значению x сопоставлено значение y . Но нередко оказывается, что нахождение этого значения очень трудоемко. Например, $y(x)$ может быть определено как решение сложной задачи, в которой x играет роль параметра, или $y(x)$ измеряется в дорогостоящем эксперименте. При этом можно вычислить небольшую таблицу значений функции, но прямое нахождение функции при большом числе значений аргумента будет практически невозможно.

Функция $y(x)$ может участвовать в каких-либо физико-технических или чисто математических расчетах, где ее приходится многократно вычислять. В этом случае выгодно заменить функцию $y(x)$ приближенной формулой, т. е. подобрать некоторую функцию $\varphi(x)$, которая близка в некотором смысле к $y(x)$ и просто вычисляется. Затем при всех значениях аргумента полагают $y(x) \approx \varphi(x)$. Близость получают введением в аппроксимирующую функцию свободных параметров $\mathbf{a} = \{a_1, a_2, \dots, a_n\}$ и соответствующим их выбором.

Подбор удачного вида функциональной зависимости $\varphi(x; \mathbf{a})$ — искусство; некоторые советы по этому поводу будут даны в § 1, п. 8. А определение наилучших (в требуемом смысле) параметров формулы делается стандартными методами, которые и будут рассмотрены в этой главе.

2. Линейная интерполяция. Пусть функция $y(x)$ известна только в узлах некоторой сетки x_i , т. е. задана таблицей. Если потребовать, чтобы $\varphi(x; \mathbf{a})$ совпадала с табличными значениями в n выбранных узлах сетки, то получим систему

$$\varphi(x_i; a_1, a_2, \dots, a_n) = y(x_i) \equiv y_i, \quad 1 \leq i \leq n, \quad (1)$$

из которой можно определить параметры a_k . Этот способ подбора параметров называется *интерполяцией* (точнее, *лагранжевой интерполяцией*). По числу используемых узлов сетки будем называть интерполяцию *одноточечной*, *двухточечной* и т. д.

Если $\varphi(x; \mathbf{a})$ нелинейно зависит от параметров, то интерполяцию назовем *нелинейной*; в этом случае нахождение параметров из системы (1) может быть трудной задачей. Сейчас мы рассмотрим *линейную* интерполяцию, когда $\varphi(x; \mathbf{a})$ линейно зависит от параметров, т. е. представима в виде так называемого *обобщенного многочлена*

$$\varphi(x; a_1, a_2, \dots, a_n) = \sum_{k=1}^n a_k \varphi_k(x). \quad (2)$$

Очевидно, функции $\varphi_k(x)$ можно считать линейно-независимыми, иначе число членов в сумме и параметров можно было бы уменьшить. На систему функций $\varphi_k(x)$ надо наложить еще одно ограничение. Подставляя (2) в (1), получим для определения параметров a_k следующую систему линейных уравнений:

$$\sum_{k=1}^n a_k \varphi_k(x_i) = y_i, \quad 1 \leq i \leq n. \quad (3)$$

Чтобы задача интерполяции всегда имела единственное решение, надо, чтобы при любом расположении узлов (лишь бы среди них не было совпадающих) определитель системы (3) был бы отличен от нуля:

$$\Delta \equiv \text{Det}\{\varphi_k(x_i)\} = \begin{vmatrix} \varphi_1(x_1) & \varphi_2(x_1) & \dots & \varphi_n(x_1) \\ \varphi_1(x_2) & \varphi_2(x_2) & \dots & \varphi_n(x_2) \\ \dots & \dots & \dots & \dots \\ \varphi_1(x_n) & \varphi_2(x_n) & \dots & \varphi_n(x_n) \end{vmatrix} \neq 0 \text{ при } x_i \neq x_j. \quad (4)$$

Система функций, удовлетворяющих требованию (4), называется *чебышёвской*. Таким образом, при линейной интерполяции надо

строить обобщенный многочлен по какой-нибудь чебышёвской системе функций.

Для линейной интерполяции наиболее удобны обычные многочлены, ибо они легко вычисляются и на клавишной машине и на ЭВМ. Другие системы функций сейчас почти не употребляются, хотя в теории подробно рассматривают интерполяцию тригонометрическими многочленами и экспонентами. Поэтому мы не приводим выражения обобщённого многочлена (2) через табулированные значения функции y_i ; вывести это выражение несложно.

3. Интерполяционный многочлен Ньютона. Рассмотрим систему $\varphi_k(x) = x^k$, $0 \leq k \leq n$; для удобства узлы интерполяции также перенумеруем с нулевого по n -й. Легко заметить, что определитель (4) в этом случае есть определитель Вандермонда

$$\Delta = \begin{vmatrix} 1 & x_0 & x_0^2 & \dots & x_0^n \\ 1 & x_1 & x_1^2 & \dots & x_1^n \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_n & x_n^2 & \dots & x_n^n \end{vmatrix} = \prod_{n \geq k > m \geq 0} (x_k - x_m). \quad (5)$$

Следовательно, алгебраический интерполяционный многочлен $\mathcal{P}_n(x)$ всегда существует и единствен (с точностью до формы записи). Применим для его вывода следующий прием.

Определим *разделенные разности* табулированной функции $y(x)$ при помощи соотношений

$$\begin{aligned} y(x_i, x_j) &= [y(x_i) - y(x_j)] / (x_i - x_j), \\ y(x_i, x_j, x_k) &= [y(x_i, x_j) - y(x_j, x_k)] / (x_i - x_k) \end{aligned} \quad (6)$$

и т. д. Разделенные разности первого, второго и более высоких порядков имеют размерности производных соответствующих порядков; в главе III показано, что они дают приближенные значения производных. Разделенные разности любого порядка можно выразить непосредственно через узловые значения функции, но вычислять их удобнее по рекуррентному соотношению (6).

Пусть $\mathcal{P}(x)$ есть многочлен степени n . Рассмотрим, что представляют собой его разделенные разности. Вычитая из него константу $\mathcal{P}(x_0)$, получим многочлен $\mathcal{P}(x) - \mathcal{P}(x_0)$, который обращается в нуль при $x = x_0$ и поэтому делится нацело на $x - x_0$. Следовательно, первая разделенная разность многочлена n -й степени $\mathcal{P}(x, x_0) = [\mathcal{P}(x) - \mathcal{P}(x_0)] / (x - x_0)$ есть многочлен степени $n-1$ относительно x и в силу симметрии выражения — относительно

x_0 . Аналогично, вторая разность $\mathcal{P}(x, x_0, x_1)$ есть многочлен степени $n - 2$; в самом деле, из (6) видно, что числитель этой разности обращается в нуль при $x = x_1$, и значит, нацело делится на $x - x_1$, а степень при этом понижается на единицу. Продолжая эти рассуждения, можно показать, что разность $\mathcal{P}(x, x_0, x_1, \dots, x_{n-1})$ есть многочлен нулевой степени, т. е. константа, а более высокие разделенные разности тождественно равны нулю.

Перепишем соотношения (6) для случая, когда функция есть многочлен и первый аргумент равен x :

$$\begin{aligned}\mathcal{P}(x) &= \mathcal{P}(x_0) + (x - x_0)\mathcal{P}(x, x_0), \\ \mathcal{P}(x, x_0) &= \mathcal{P}(x_0, x_1) + (x - x_1)\mathcal{P}(x, x_0, x_1)\end{aligned}$$

и т. д. Эта цепочка соотношений конечна, ибо $(n + 1)$ -я разделенная разность многочлена тождественно равна нулю. Последовательно подставляя эти соотношения друг в друга, получим формулу

$$\begin{aligned}\mathcal{P}(x) &= \mathcal{P}(x_0) + (x - x_0)\mathcal{P}(x_0, x_1) + \\ &\quad + (x - x_0)(x - x_1)\mathcal{P}(x_0, x_1, x_2) + \dots \\ &\quad \dots + (x - x_0)(x - x_1) \dots (x - x_{n-1})\mathcal{P}(x_0, x_1, \dots, x_n),\end{aligned}\quad (7)$$

по которой многочлен n -й степени выражается при помощи разделенных разностей через свои значения в узлах x_0, \dots, x_n . Но значения интерполяционного многочлена в этих узлах по определению совпадают со значениями искомой функции, и поэтому разделенные разности $y(x)$ и $\mathcal{P}(x)$ тоже совпадают. Подставляя в (7) разделенные разности искомой функции и заменяя точное равенство на приближенное, получим интерполяционную формулу Ньютона

$$y(x) \approx y(x_0) + \sum_{k=1}^n (x - x_0)(x - x_1) \dots (x - x_{k-1})y(x_0, x_1, \dots, x_k). \quad (8)$$

Формула Ньютона удобна для вычислений и на ЭВМ, и на клавишной машине. Легко составить следующую таблицу 1 разделенных разностей для табулированной функции $y(x)$ и произвести вычисления по формуле (8).

З а м е ч а н и е 1. За точностью расчета удобно следить, визуально оценивая скорость убывания членов суммы (8). Если они убывают медленно, то на хорошую точность рассчитывать, вообще говоря, нельзя (подробнее см. пп. 6, 7). Если убывание быстрое, то

Таблица 1

x_0	$y(x_0)$			
x_1	$y(x_1)$	$y(x_0, x_1)$		
x_2	$y(x_2)$	$y(x_1, x_2)$	$y(x_0, x_1, x_2)$	
x_3	$y(x_3)$	$y(x_2, x_3)$	$y(x_1, x_2, x_3)$	$y(x_0, x_1, x_2, x_3)$

оставляют только те члены, которые больше допустимой погрешности; тем самым определяют, сколько узлов требуется подключить в расчет.

Пример. Покажем, как вычислять синус в первом квадранте, используя четыре известных значения. Составим таблицу 2 с четырьмя узлами, причем для удобства вычисления положим $y(x) = \sin(30^\circ \cdot x)$. Для проверки точности, используя разности верхней косой строки, вычислим

$$y(1,5) \approx 0 + 0,750 - 0,050 + 0,006 = 0,706.$$

Это приближенное значение мало отличается от точного значения $y(1,5) = \sin 45^\circ \approx 0,707$. Таким образом, достаточно помнить только верхнюю косую строку таблицы 2, чтобы вычислять синус с точностью около 0,001.

Таблица 2

x_i	$y(x_i)$	$y(x_i, x_{i+1})$	$y(x_i, x_{i+1}, x_{i+2})$	$y(x_i, \dots, x_{i+3})$
0	0,000			
1	0,500	0,500		
2	0,866	0,366	-0,0607	
3	1,000	0,134	-0,116	-0,016

З а м е ч а н и е 2. При заданном числе узлов многочлен Ньютона удобнее вычислять по схеме Горнера, записывая его в виде

$$y(x) = y(x_0) + (x - x_0)[y(x_0, x_1) + (x - x_1)[y(x_0, x_1, x_2) + \dots]].$$

Но если надо контролировать точность расчета и определять нужное число узлов, то удобнее форма (8).

Замечание 3. Для расчетов по формуле Ньютона безразличен порядок, в котором переenumerованы узлы интерполяции; это полезно помнить при подключении новых узлов в расчет.

Мы ограничились здесь общими формулами, пригодными для таблиц с переменным шагом. Во многих учебниках для таблиц с постоянным шагом вводят *конечные разности* $\Delta^n y$, связанные с разделенными разностями соотношением $\Delta^n y = n!y(x_0, x_1, \dots, x_n)$. Но это дань историческим традициям, ибо разделенные разности не менее удобны в расчетах, чем конечные.

Есть много разных форм записи интерполяционного многочлена общего вида: Ньютона, Лагранжа, Гаусса, Грегори — Ньютона, Лапласа — Эверетта и др. Наиболее удобной для вычислений с контролем точности и на ЭВМ и вручную является форма Ньютона (8). Большинство остальных форм рассчитано на определенные частные случаи расположения узлов интерполяции, но те выгоды, которые при этом получаются, обычно несущественны при расчетах на ЭВМ.

4. Погрешность многочлена Ньютона. Выше мы рассмотрели эмпирическое правило определения погрешности интерполяции по убыванию членов суммы (8). Проведем теперь строгое исследование погрешности метода, проистекающей от замены искомой функции интерполяционным многочленом Ньютона.

Погрешность удобно представить в следующем виде:

$$y(x) - \mathcal{P}_n(x) = \omega_n(x)r(x), \quad \omega_n(x) = \prod_{i=0}^n (x - x_i), \quad (9)$$

ибо эта погрешность заведомо равна нулю во всех узлах интерполяции. Введем вспомогательную функцию $q(\xi) = y(\xi) - \mathcal{P}_n(\xi) - \omega_n(\xi)r(x)$, где x играет роль параметра и принимает любое фиксированное значение. Очевидно, $q(\xi) = 0$ при $\xi = x_0, x_1, \dots, x_n$ и при $\xi = x$, т. е. обращается в нуль в $n + 2$ точках.

Предположим, что $y(x)$ имеет $n + 1$ непрерывную производную; тогда то же справедливо для $q(\xi)$. Между двумя нулями гладкой функции лежит нуль ее производной. Последовательно применяя это правило, получим, что между крайними из $n + 2$ нулей функции лежит нуль $n + 1$ -й производной. Но $q^{(n+1)}(\xi) = y^{(n+1)}(\xi) - (n+1)!r(x)$, и если в какой-то точке ξ^* , лежащей между указанными нулями, она обращается в нуль, то $r(x) = y^{(n+1)}(\xi^*) / (n + 1)!$ Заменяя погрешность (9) максимально возможной, получаем

оценку погрешности:

$$|y(x) - \mathcal{P}_n(x)| \leq \frac{M_{n+1}}{(n+1)!} |\omega_n(x)|, \quad M_{n+1} = \max |y^{(n+1)}(\xi)|, \quad (10)$$

где максимум производной берется по отрезку между наименьшим и наибольшим из значений x, x_0, x_1, \dots, x_n .

Оценить $\omega_n(x)$ при произвольном расположении узлов интерполяции сложно. Однако таблицы чаще всего имеют постоянный шаг $h = x_{i+1} - x_i$, а узлы интерполяции берутся из таблицы подряд. Тогда $\omega_n(x)$ имеет примерно такой вид, как показано на рис. 2 для $n = 5$: вблизи центрального узла интерполяции экстремумы невелики, вблизи крайних узлов — несколько больше, а если x выходит за крайние узлы интерполяции, то $\omega_n(x)$ быстро возрастает.

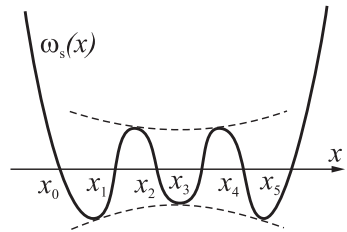


Рис. 2

Можно подобрать узлы интерполяции так, чтобы на заданном отрезке $\max |\omega_n(x)|$ был меньше, чем у любого другого многочлена той же степени. Для этого $\omega_n(x)$ должен быть многочленом Чебышёва первого рода (см. Приложение). Узлы этого многочлена расположены сравнительно редко в середине рассматриваемого отрезка и сгущаются у его концов. Но вне выбранного отрезка многочлен $\omega_n(x)$ по-прежнему будет быстро возрастать. Этот способ интерполяции довольно громоздок, а выигрыш в точности невелик; поэтому его используют только для специальных целей — например, при построении аппроксимирующих формул.

Термин *интерполяция* в узком смысле употребляют, если x заключен между крайними узлами интерполяции; если он выходит из этих пределов, то говорят об *экстраполяции*. Очевидно, что при экстраполяции далеко за крайний узел ошибка может быть велика, поэтому экстраполяция мало надежна. На практике рекомендуется пользоваться преимущественно интерполяцией.

При интерполяции на равномерной сетке выгодно выбирать из таблицы узлы так, чтобы искомая точка x попадала ближе к центру этой конфигурации узлов — это обеспечит более высокую точность. Для упрощения вычислений рассмотрим случай нечетного $n = 2k + 1$. Из симметрии полинома $\omega_n(x)$ очевидно, что в центральном интервале экстремум достигается точно в середине (см.

рис. 2). Этот экстремум равен

$$\left[\frac{h}{2} \cdot \frac{3h}{2} \cdot \frac{5h}{2} \cdots \frac{(2k+1)h}{2} \right]^2 = \left[\frac{(2k+1)!h^{k+1}}{k!2^{2k+1}} \right]^2.$$

Подставим эту величину в оценку (10). После несложных преобразований с использованием формулы Стирлинга $p! \approx \sqrt{2\pi p}(p/e)^p$ получим оценку ошибки в центральном интервале

$$|y(x) - \mathcal{P}_n(x)| < \sqrt{2/\pi n} M_{n+1} (h/2)^{n+1}. \quad (11)$$

Если величины производных $y(x)$ можно оценить, то отсюда легко определить число узлов, достаточное для получения заданной точности.

Из оценки (11) видно, что если перейти от таблиц с крупным шагом к таблицам с более мелким шагом, то погрешность метода будет убывать, как h^{n+1} . Поэтому говорят, что многочлен Ньютона $\mathcal{P}_n(x)$ имеет погрешность $\mathcal{O}(h^{n+1})$ и обеспечивает $n+1$ -й порядок точности интерполяции.

В главе III мы увидим, что между разделенными разностями и производными соответствующих порядков существует соотношение $y^{(n)}(x) \approx n!y(x_0, x_1, \dots, x_n)$. Если учесть это при определении величины членов суммы (8), то нетрудно заметить, что эмпирическая оценка погрешности по первому отброшенному члену близка к оценке (10), хотя является менее строгой. Оценки (10) и (11) можно провести до вычисления интерполяционного многочлена, т.е. это *априорные* оценки точности. Оценка же (по первому отброшенному члену делается после выполнения вычислений, т.е. является *апостериорной*). Поскольку обычно величины производных искомой функции заранее неизвестны, а в ходе вычисления многочлена Ньютона они фактически определяются, то на практике удобнее пользоваться апостериорной оценкой.

Далее мы не раз сможем убедиться, что строгие априорные оценки используются в основном при теоретическом исследовании методов. При практическом контроле точности расчетов обычно употребляют менее строгие (хотя тоже имеющие теоретическое обоснование), но более удобные апостериорные оценки.

5. Применения интерполяции. Интерполяция применяется во многих задачах, а не только для вычисления табулированной функции при любых значениях аргумента.

При помощи разделенных разностей контролируется точность таблиц. Для этого составляют таблицы разделенных разностей различных порядков для соседних узлов и анализируют их поведение.

Например, в таблице 3 приведена зависимость коэффициента теплопроводности высокотемпературной фазы циркония от температуры. Там же вычислены первая и вторая конечные разности. Видно, что вторая разность меняется беспорядочно, так что интерполировать более чем по двум точкам бессмысленно. По величине второй разности можно сказать, что случайная погрешность λ составляет около единицы третьего знака в большинстве точек, но в двух первых она может доходить до единицы второго знака (для систематической погрешности измерений эти соображения неприемлемы).

Таблица 3
Теплопроводность циркония

τ°, K	$\lambda \cdot 10^4, \text{ кал/см}\cdot\text{г}\cdot\text{сек}$	$\Delta_1 \lambda \cdot 10^6$	$\Delta_2 \lambda \cdot 10^8$
1200	561		
		79	
1300	640		-24
		55	
1400	695		-31
		21	
1500	716		-2
		19	
1600	735		-2
		17	
1700	752		+2
		19	
1800	771		-2
		17	
1900	788		-3
		14	
2000	802		+5
		19	
2100	821		

Подобный контроль полезен при анализе результатов измерений в физике и технике.

Интерполяцию применяют для *субтабулирования*— сгущения таблиц. Алгоритмы непосредственного вычисления многих функций очень сложны. Поэтому при математическом табулировании обычно функцию непосредственно вычисляют в небольшом числе узлов, т. е. таблицы имеют крупный шаг. Затем при помощи ин-

терполяции высокого порядка точности сетку узлов сгущают и составляют подробную таблицу. Шаг этой таблицы выбирают таким, чтобы простейшая интерполяция (двухточечная) обеспечивала требуемую точность.

В связи с этим отметим, что при ручных расчетах выгодны подробные таблицы, ибо они допускают применение простейших способов интерполяции, легко выполняемых на бумаге или клавишных машинах, а время поиска нужных узлов интерполяции невелико по сравнению со временем выполнения алгебраических действий. Наоборот, при расчетах на ЭВМ задание подробных таблиц невыгодно, поскольку они занимают много места в оперативной памяти, а время поиска становится много больше времени выполнения алгебраических действий; выгоднее таблицы с большим шагом, хотя при этом требуются более сложные и точные способы интерполяции.

Задачей *обратного интерполирования* называют нахождение x для произвольного y , если задана таблица $y_i = y(x_i)$. Для монотонных функций между прямым и обратным интерполированием нет разницы: можно читать таблицу наоборот, как задание $x_i = x(y_i)$. Единственное отличие будет в том, что «обратная» таблица $x(y_i)$ будет иметь переменный шаг, даже если «прямая» таблица имела постоянный. Но все наши формулы рассчитаны на переменный шаг. Отметим, что для достижения заданной точности прямая и обратная интерполяции требуют, вообще говоря, разного числа узлов.

Важный пример обратного интерполирования — решение уравнения $y(x) = 0$. Вычислим несколько значений функции $y(x_i)$, т. е. составим небольшую таблицу. Запишем ее в виде $x_i = x(y_i)$ и при помощи интерполяции найдем приближенное значение $x(0)$. Этот способ дает хорошие результаты, если функция достаточно гладкая, а корень лежит между рассчитанными узлами. Если корень расположен далеко от узлов, то способ ненадежен, ибо применяется экстраполяция.

Пример. Решим уравнение

$$y(x) \equiv (1+x)e^{0,5x} - 2,5 = 0. \quad (12)$$

Составим таблицу 4 значений функции; первым запишем столбец значений y , ибо в дальнейших вычислениях эта величина будет аргументом. Найдем разделенные разности и произведем вычисления по верхней косой строке:

$$x(0) \approx x_0 + (0 - y_0)x(y_0, y_1) + (0 - y_0)(0 - y_1)x(y_0, y_1, y_2) = 0,744.$$

Точное решение есть $x(0) = 0,732$, так что ошибка получилась небольшой. Для повышения точности в этом способе целесообразно взять новые узлы, близко расположенные к грубо найденному корню, а не увеличивать число узлов.

Таблица 4

y_i	x_i	$x(y_i, y_{i+1})$	$x(y_0, y_1, y_2)$
-1500	0		
		0,540	
-0,574	0,5		-0,076
		0,365	
0,797	1,0		

В этом курсе будут рассмотрены и другие примеры применения интерполирования.

6. Интерполяционный многочлен Эрмита. Пусть табулирована не только функция, но и ее производные вплоть до некоторого порядка. Тогда можно потребовать, чтобы в узлах интерполяции совпадали не только значения искомой функции $y(x)$ и интерполяционной функции $\varphi(x)$, но и значения их производных вплоть до некоторого порядка. Такую интерполяцию будем называть *эрмитовой*; если $\varphi(x)$ — алгебраический многочлен n -й степени, то он называется интерполяционным многочленом Эрмита и обозначается $\mathcal{H}_n(x)$.

Покажем, как построить этот многочлен. По $n+1$ узлу построим интерполяционный многочлен Ньютона $\mathcal{P}_n(x; x_0, x_1, \dots, x_n)$. Поскольку значения функции $y(x)$ и многочлена в узлах совпадают, то их средние наклоны на участках между узлами равны. Мысленно будем приближать узел x_n к узлу x_{n-1} ; при этом средний наклон будет стремиться к производной. Значит, после совпадения узлов получим многочлен, который в узле x_{n-1} правильно передает не только значение функции, но и значение первой производной. Символически обозначим его как $\mathcal{P}_n(x; x_0, x_1, \dots, x_{n-1}, x_{n-1})^*$.

Слияние трех узлов в один обеспечивает передачу не только наклона, но и кривизны, т. е. первой и второй производных и т. д.

*) Чтобы отличать его обозначение от разделенной разности, мы отделяем аргумент от узлов, по которым составлен многочлен, точкой с запятой.

Таким образом, многочлен

$$\mathcal{H}_n(x) = \mathcal{P}_n(x; \underbrace{x_0, x_0, \dots, x_0}_{m_0}, \underbrace{x_1, x_1, \dots, x_1}_{m_1}, \underbrace{x_p, x_p, \dots, x_p}_{m_p}), \quad (13)$$

$$\sum_{k=0}^p m_k = n + 1,$$

в узле x_k правильно передает значение функции и ее производных вплоть до порядка $m_k - 1$ и имеет минимально необходимую для этого степень. Оценка погрешности метода (10) в этом случае принимает следующий вид:

$$|y(x) - \mathcal{H}_n(x)| \leq \frac{M_{n+1}}{(n+1)!} |\Omega_n(x)|, \quad \Omega_n(x) = \prod_{k=0}^p (x - x_k)^{m_k}. \quad (14)$$

Очевидно, если сетка имеет шаг h , а точка x лежит между крайними узлами интерполяции, то $\Omega_n(x) = O(h^{n+1})$; следовательно, порядок точности эрмитовой интерполяции равен $n + 1$, т. е. числу коэффициентов интерполяционного многочлена.

Заметим, что обычный многочлен Ньютона с таким же числом коэффициентов (т. е. той же степени) также имеет погрешность $O(h^{n+1})$. Однако на одной и той же сетке численная величина погрешности многочлена Ньютона будет больше, чем у многочлена Эрмита: его вспомогательный многочлен $\omega_n(x)$ содержит больше узлов, чем $\Omega_n(x)$, и поэтому в него входят бóльшие сомножители. Очевидно также, что чем более высокие производные используются при построении интерполяционного многочлена Эрмита заданной степени, тем меньше требуемое число узлов, и тем меньше будет численная величина его погрешности (хотя порядок точности остается одним и тем же).

Выражением (13) нельзя пользоваться буквально. Если формально подставить в формулу Ньютона (8) совпадающие узлы, то потребуются вычислить разделенные разности, у которых некоторые узлы являются кратными. Выражения (6) для таких разностей содержат неопределенность типа $0/0$. Если кратность каждого узла не больше чем двойная, то эту неопределенность можно раскрыть

с помощью предельного перехода, например,

$$\begin{aligned} y(x_0, x_0) &= \lim_{x_1 \rightarrow x_0} \frac{y(x_0) - y(x_1)}{x_0 - x_1} = y'(x_0), \\ y(x_0, x_0, x_1) &= \frac{1}{x_0 - x_1} [y'(x_0) - y(x_0, x_1)], \\ y(x_0, x_0, x_1, x_1) &= \frac{1}{(x_0 - x_1)^2} [y'(x_0) - 2y(x_0, x_1) + y'(x_1)]. \end{aligned} \quad (15)$$

Если узлы имеют более высокую кратность, то удобнее дифференцировать формулу Ньютона (8). Например, если ее продифференцировать $m - 1$ раз, то обратятся в нуль все члены, содержащие разделенные разности порядка меньше $m - 1$. Затем положим $x = x_0 = x_1 = \dots$; тогда обратятся в нуль множители перед разделенными разностями порядка больше $m - 1$, и мы получим

$$y(\underbrace{x_0, x_0, \dots, x_0}_m) = \frac{1}{(m-1)!} y^{(m-1)}(x_0). \quad (16)$$

Но узлы более чем двойной кратности почти не встречаются в практике вычислений, ибо вторые и более высокие производные искомой функции редко табулируются.

Рассмотрим наиболее употребительные частные случаи интерполяционного многочлена Эрмита.

Первый случай — многочлен, который в одном узле x_0 совпадает с функцией и всеми ее заданными производными:

$$\begin{aligned} \mathcal{P}_n(x; x_0, x_0, \dots) &= \\ &= y(x_0) + (x - x_0)y'(x_0) + \frac{1}{2}(x - x_0)^2 y''(x_0) + \dots \end{aligned} \quad (17)$$

Очевидно, это отрезок ряда Тейлора; в этом случае $\Omega_n(x) = (x - x_0)^{n+1}$, и оценка (11) переходит в известную оценку точности ряда Тейлора.

Второй случай — многочлен, передающий в двух узлах значения функции и ее первой производной:

$$\begin{aligned} \mathcal{P}_n(x; x_0, x_0, x_1, x_1) &= y(x_0) + (x - x_0)\{y'(x_0) + \\ &+ (x - x_0)[y(x_0, x_0, x_1) + (x - x_1)y(x_0, x_0, x_1, x_1)]\}; \end{aligned} \quad (18)$$

разделенные разности сюда надо подставить из соотношения (15). Функция $\Omega_n(x) = (x-x_0)^2(x-x_1)^2$ внутри интервала интерполирования не превышает $(h/2)^4$, так что погрешность формулы (18) не более $0,026M_4h^4$; эта формула имеет четвертый порядок точности.

Для сравнения приведем без вывода общее выражение интерполяционного многочлена Эрмита

$$\mathcal{H}_n(x) = \sum_{k=0}^p \sum_{m=0}^{\alpha_k-1} \sum_{q=0}^{\alpha_k-1-m} \frac{y^{(m)}(x_k)}{m!q!} \left\{ (x-x_k)^{m+q} \times \right. \\ \left. \times \prod_{i \neq k} (x-x_i)^{\alpha_i} \right\} \left\{ \frac{d^q}{dx^q} \prod_{j \neq k} (x-x_j)^{-\alpha_j} \right\}_{x=x_k}.$$

Оно настолько громоздко, что пользоваться им для вычислений практически невозможно. Если все $\alpha_i = 1$, то обе внутренние суммы превращаются в одно слагаемое с $m = q = 0$, и многочлен Эрмита переходит в многочлен Ньютона в форме Лагранжа. Если все $\alpha_i = 2$, то получим

$$\mathcal{H}_n(x) = \sum_{k=0}^p \left\{ (x-x_k)y'_k + \left(1 - 2 \sum_{\substack{i=0 \\ i \neq k}}^p \frac{x-x_k}{x_k-x_i} \right) y_k \right\} \prod_{\substack{j=0 \\ j \neq k}}^p \left(\frac{x-x_j}{x_k-x_j} \right)^2;$$

можно проверить, что в случае двух узлов последнее выражение совпадает с (18) с точностью до формы записи. Но даже и это выражение оказывается очень громоздким.

Такая ситуация довольно часто встречается в прикладной математике. Общие формулы, рассчитанные на все случаи жизни, нередко оказываются настолько сложными, что их не применяют ни в одном конкретном случае. К тому же, в практических расчетах, как мы увидим далее, нецелесообразно использовать многочлены высоких степеней, поэтому в общих формулах нет серьезной необходимости. Трудоемкость же вычислений часто оказывается существенно меньшей при применении рекуррентных процедур типа формулы разделенных разностей (6).

7. Сходимость интерполяции. При каких условиях погрешность метода стремится к нулю, т. е. когда и как интерполяционный многочлен сходится к $y(x)$? На практике мы имеем два способа перехода к пределу. Первый состоит в том, чтобы, сохраняя степень интерполяционного многочлена, уменьшить шаг сетки, т. е. воспользоваться более подробными таблицами. Второй — сохраняя шаг сетки, увеличивать число используемых узлов, т. е. увеличивать степень многочлена.

Уменьшение шага. Если $y(x)$ имеет непрерывные производные вплоть до $n+1$ -й, то при интерполяции многочленом $\mathcal{P}_m(x)$

степени $m \leq n$ погрешность метода есть $O(h^{m+1})$. В этом случае при фиксированной степени многочлена и уменьшении шага сетки погрешность $|y(x) - \mathcal{P}_m(x)|$ неограниченно убывает. Если ограничена производная, входящая в оценку ошибки, то интерполяционный многочлен равномерно сходится к $y(x)$ на ограниченном отрезке $a \leq x \leq b$.

Строго говоря, для каждого значения x выбирают свои узлы интерполяции, ближайшие (на данной сетке) к точке x , т.е. составляют свой многочлен $\mathcal{P}_m(x)$. При этом точка x заведомо лежит между крайними узлами интерполяции, используемыми в данном многочлене. Поэтому входящий в оценку погрешности (10) полином $\omega_m(x)$ ограничен равномерно по x : $|\omega_m(x)| < \max_i |x - x_i|^{m+1} \leq (mh)^{m+1}$, где h — шаг сетки (для неравномерных сеток — максимальный шаг). Для заданной точности ε определим шаг сетки из условия $M_{m+1}(mh)^{m+1} \leq \varepsilon \cdot (m+1)!$, где $M_{m+1} = \max_{[a,b]} |y^{(m+1)}(x)|$.

Тогда для всех сеток с данным и более мелким шагом и любой точки отрезка $a \leq x \leq b$ погрешность интерполяционного многочлена $\mathcal{P}_m(x)$, узлы которого выбраны указанным выше образом, будет не более ε .

Аналогичные утверждения справедливы для интерполяционного многочлена Эрмита.

Увеличение числа узлов. Увеличивать степень интерполяционного многочлена далеко не всегда целесообразно. Во-первых, неизвестно, как быстро растет максимум производной M_m с увеличением ее порядка. Во-вторых, у функции может быть лишь конечное число производных. Рассмотрим интерполяцию на отрезке $a \leq x \leq b$, когда число узлов, используемых для построения интерполяционного многочлена, неограниченно возрастает.

Известно, что если $y(x)$ — целая функция, то при произвольном расположении узлов на $[a, b]$ многочлен $\mathcal{P}_n(x)$ равномерно сходится к $y(x)$ при $n \rightarrow \infty$. Но целая функция — это функция, разложимая в степенной ряд с бесконечным радиусом сходимости. Гораздо чаще приходится импонировать не целые функции, так что практическая ценность этого утверждения невелика.

Если же на $[a, b]$ функция имеет непрерывные производные сколь угодно высоких порядков, то это не гарантирует сходимости при произвольном расположении узлов. Например, возьмем функцию

$$y(x) = 0 \quad \text{при} \quad -1 \leq x \leq 0, \quad y(x) = e^{-1/x} \quad \text{при} \quad 0 < x \leq 1.$$

Ее график приведен на рис. 3. Все производные этой функции на $[-1, +1]$ непрерывны. Но если разместить все узлы интерполяции левее точки $x = 0$, то, очевидно, $\mathcal{P}_n(x) \equiv 0$, и никакой сходимости быть не может.

Правда, в этом примере расположение узлов было грубо неравномерным. Но равномерное расположение не всегда спасает. С. Н. Бернштейн в 1916 г. доказал, что для функции $y(x) = |x|$ на отрезке $[-1, +1]$, покрытом равномерной сеткой узлов, значения $\mathcal{P}_n(x)$ между узлами интерполяции неограниченно возрастают при $n \rightarrow \infty$. Это иллюстрируется рис. 4, где даны графики функции и двух многочленов разных степеней.

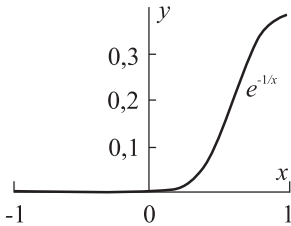


Рис. 3

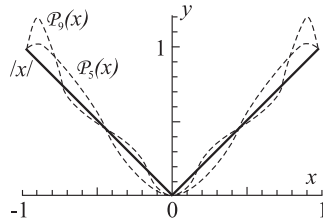


Рис. 4

Более того, для любой наперед заданной системы узлов можно найти такую непрерывную функцию, что построенные по этим узлам и функции многочлены Ньютона не будут равномерно сходиться.

Но сходимости в среднем для многочленов Ньютона всегда можно добиться следующим несложным выбором узлов. Пусть $\Phi_n(x)$ — система многочленов, ортогональных с весом $\rho(x)$ на отрезке $[a, b]$, и $x_m^{(n)}$ — нули этих многочленов. Используем эти точки в качестве узлов интерполяций; тогда

$$\int_a^b [\mathcal{P}_n(x) - y(x)]^2 \rho(x) dx \rightarrow 0$$

при $n \rightarrow \infty$ для любой непрерывной функции.

Для многочленов Эрмита получены более сильные результаты. Пусть функция $y(x)$ непрерывна на $[-1, +1]$; возьмем в качестве узлов нули многочленов Чебышёва первого рода $T_n(x)$ (см. Приложение); фиксируем в этих узлах значения функции, а вместо ее производной возьмем любые числа c_{in} , удовлетворяющие условию $\lim_{n \rightarrow \infty} \max_i |c_{in} \ln n/n| = 0$. Построенный по всем этим значениям многочлен $\mathcal{H}(2n-1)(x)$ равномерно сходится к $y(x)$ при $n \rightarrow \infty$. Очевидно, если $y(x)$ имеет ограниченную производную, то в качестве c_{in} можно брать значение производной в узлах.

Но и для многочленов Эрмита неудачный выбор узлов может испортить сходимость. Например, ряд Тейлора (17) расходится, если $|x - x_0|$ больше расстояния от x_0 до ближайшей особой точки в комплексной плоскости.

Выводы. На практике интерполировать многочленом высокой степени нежелательно. Если 3–5 узлов (точнее, свободных параметров) не обеспечивают требуемой точности, то обычно надо не увеличивать число узлов, а уменьшать шаг таблицы.

8. Нелинейная интерполяция. Полиномиальная интерполяция по оценке (11) имеет погрешность $\sim M_{n+1}(h/2)^{n+1}$, и при по-

вышении порядка точности формулы на единицу погрешность меняется примерно в $hM_{n+2}/2M_{n+1}$ раз. Если шаг достаточно мал, то погрешность при этом уменьшается. Но если шаг велик, или производные быстро растут с увеличением порядка, то погрешность может увеличиваться при увеличении порядка точности формулы. С этим часто приходится сталкиваться при работе с быстро меняющимися функциями.

Таблица 5

x_i	$y(x_i)$	$y(x_i, x_{i+1})$		
0	1			
		10		
1	11		50	
		110		170
2	121		550	
		1230		
3	1351			

Пример 1. Пусть требуется найти значение $y(0, 5)$, если функция задана таблицей 5 (в ней выписаны не только значения функции, но и разделенные разности). Используя интерполяционный многочлен Ньютона и ведя вычисления по верхней строке таблицы 5, запишем последовательно члены все более высоких порядков:

$$\varphi(0, 5) = 1 + 5 - 12, 5 + 63, 75 - \dots$$

Этот ряд содержит быстро возрастающие члены и совсем не похож на сходящийся; поэтому вычислить функцию с его помощью не удастся. Функция слишком быстро меняется или, что то же самое, шаг сетки слишком велик для данной функции (рис. 5, а).

Как интерполировать такие функции, если более подробных таблиц нет? Универсального рецепта, пригодного для любой функции, не существует. Однако для конкретной функции нередко удается найти свой способ интерполяции, дающей разумную точность. Такая интерполяция обычно нелинейна.

Для этого нужно располагать дополнительной информацией о

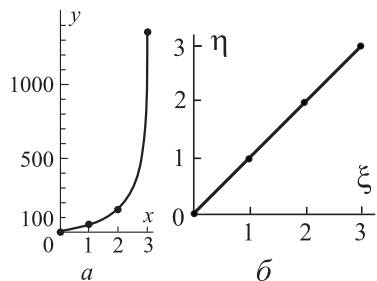


Рис. 5

качественном поведении функции. Часто ее можно получить, зная физический смысл $y(x)$. Например, проходящий через поглощающую среду свет ослабляется примерно по экспоненциальному закону; сопротивление движению в газе зависит от скорости примерно как v^m , где $m \approx 1$ для ламинарного движения, $m \approx 2$ для турбулентного и $m > 2$ вблизи звукового барьера. Нередко помогают формальные математические соображения — изучение графика функции и сравнение его с графиками хорошо изученных функций (в первую очередь элементарных).

Выяснив качественное поведение функции, стараются подобрать такое преобразование переменных $\eta = \eta(y)$, $\xi = \xi(x)$, чтобы в новых переменных график $\eta(\xi)$ мало отличался от прямой на протяжении нескольких шагов таблицы. Тогда в переменных $\eta(\xi)$ интерполяция многочленом невысокой степени будет давать хорошую точность. Вычисления заключаются в составлении таблицы для новых переменных $\eta_i = \eta(\xi_i)$, интерполяции по ней и нахождении $y = y(\eta)$ обратным преобразованием. Этот способ называют *методом выравнивания*.

Пример 2. Проиллюстрируем метод выравнивания на примере функции, заданной таблицей 5. Нетрудно заметить, что зависимость близка к показательной, $y(x) \approx 10^x$; значит, в переменных $\xi = x$, $\eta = \lg y$ график будет почти прямым (рис. 5, б). Составим новую таблицу 6 и проведем интерполяцию по формуле Ньютона

$$\eta^* = \eta(0,5) = 0 + 0,5207 + 0 - 0,0004 \approx 0,5203.$$

Теперь члены ряда быстро убывают, обеспечивая хорошую точность; считая, что точность η^* примерно равна последующему члену ряда, обратным преобразованием получим, что $y(\eta^*) \approx 3,314 \pm 0,1\%$. Очевидно, что удачно выбранное выравнивание позволило получить высокую точность интерполяции.

Замечание 1. Для каждой конкретной функции подбирают свой вид нелинейной интерполяции. Для других функций этот вид, как правило, будет давать плохую точность.

Замечание 2. Оценка погрешности такой интерполяции содержит старшие производные $\eta(\xi)$. Их трудно найти, поэтому на практике удобнее оценивать точность по скорости убывания членов в формуле Ньютона, как было сделано выше. Употребителен также следующий прием: для одного из узлов x_i вычисляют $y(x_i)$ интерполяцией по соседним узлам и сравнивают с табличным значением y_i .

Таблица 6

ξ_i	η_i	$\eta(\xi_i, \xi_{i+1})$		
0	0,0000			
1	1,0414	1,0414	0,0000	0,0011
2	2,0828	1,0414	0,0032	
3	3,1306	1,0478		

Пример 3. Отбросим в таблице 6 узел $\xi = 1$ и связанные с ним разделенные разности. По оставшимся трем узлам приближенно вычислим отброшенное значение $\eta(1) \approx 1,0414$ или $y(1) \approx 10,92$. Последняя величина отличается от табличного значения на 0,8%. Это вычисление велось фактически с шагом $h_0 = 2$ многочленом второй степени, имеющим погрешность $O(h^3)$. Значит, при вычислениях с шагом $h = 1$ погрешность должна уменьшиться в $(h_0/h)^3 = 8$ раз и составить 0,1%. Это хорошо согласуется с оценкой по последнему члену ряда, сделанной выше.

Замечание 3. Оба прямых преобразования $\eta(y)$, $\xi(x)$ и обратное преобразование $y(\eta)$ должны выражаться несложными формулами, иначе метод выравнивания будет малопригодным на практике. Удобны преобразования типа логарифмирования, вычисления экспонент, тригонометрических функций и другие, имеющиеся в библиотеках стандартных программ современных ЭВМ (или легко выполнимые на логарифмической линейке).

Замечание 4. В исходных переменных интерполяция нелинейна относительно параметров; в данном примере она имела вид $\varphi(x) = \exp\left(\sum_{k=0}^n a_k x^k\right)$. Однако в переменных η , ξ она линейна по параметрам. Такая нелинейность мало осложняет работу, поэтому интерполяцию подобного вида будем называть *квазилинейной*.

Встречаются случаи, когда метод выравнивания неприменим. Например, если $y(x) \approx a(x+b)^c$, то не удастся найти такие координаты, которые превращали бы график в прямую и не содержали бы явно параметров a , b , c . Тогда зависимость от параметров не сводится к линейной и отыскать параметры и выполнить интерполяцию нелегко. Такую интерполяцию будем называть *существенно нелинейной*; на практике она используется крайне редко.

Замечание 5. Если выравнивающие преобразования переменных просты, то иногда удастся явно выразить $\varphi(x)$ через табличные