С.П. Шарый

**Курс**

**ВЫЧИСЛИТЕЛЬНЫХ**

**МЕТОДОВ**

Курс

ВЫЧИСЛИТЕЛЬНЫХ

МЕТОДОВ

С. П. Шарый

Институт вычислительных технологий СО РАН

Новосибирск – 2012

**Книга является систематическим учебником по курсу вычислительных ме-тодов и написана на основе лекций, читаемых автором на механико-матема-тическом факультете Новосибирского государственного университета. Осо-бенностью книги является изложение методов интервального анализа и ре-зультатов конструктивной математики, связанных с традиционными разде-лами численного анализа.**

**c С.П. Шарый, 2011 г.**

Оглавление

|  |  |  |
| --- | --- | --- |
| Предисловие | | 7 |
| Глава 1. Введение | | 8 |
| **1.1** | **Погрешности вычислений . . . . . . . . . . . . . . . . . .** | **9** |
| **1.2** | **Обусловленность математических задач . . . . . . . . . .** | **9** |

1. **Интервальная арифметика . . . . . . . . . . . . . . . . . . 12**
2. **Интервальные расширения функций . . . . . . . . . . . . 15**

|  |  |  |
| --- | --- | --- |
| **1.5** | **Элементы конструктивной математики . . . . . . . . . .** | **18** |
| **1.6** | **Сложность задач и трудоёмкость алгоритмов . . . . . . .** | **20** |
| **Литература к Главе 1 . . . . . . . . . . . . . . . . . . . . . . . .** | | **21** |
| Глава 2. Численные методы анализа | | 23 |

1. **Интерполирование функций . . . . . . . . . . . . . . . . . 24**

|  |  |  |
| --- | --- | --- |
| **2.1а** | **Постановка задачи . . . . . . . . . . . . . . . . . .** | **24** |
| **2.1б** | **Интерполяционный полином Лагранжа . . . . . .** | **28** |
| **2.1в** | **Разделённые разности и их свойства . . . . . . . .** | **30** |
| **2.1г** | **Интерполяционный полином Ньютона . . . . . . .** | **35** |
| **2.1д** | **Погрешность алгебраической интерполяции . . . .** | **37** |
| **2.1е** | **Полиномы Чебышёва . . . . . . . . . . . . . . . . .** | **41** |
| **2.1ж** | **Интерполяция с кратными узлами . . . . . . . . .** | **46** |
| **2.1з** | **Общие факты алгебраической интерполяции . . .** | **48** |
| **2.2 Сплайны . . . . . . . . . . . . . . . . . . . . . . . . . . . .** | | **52** |
| **2.2а** | **Элементы теории . . . . . . . . . . . . . . . . . . .** | **52** |
| **2.2б** | **Интерполяционные кубические сплайны . . . . . .** | **55** |

1. **Нелинейные методы интерполяции . . . . . . . . . . . . . 59**
2. **Численное дифференцирование . . . . . . . . . . . . . . . 61 2.4а Интерполяционный подход . . . . . . . . . . . . . . 62**

**3**

|  |  |  |
| --- | --- | --- |
| **4** | Оглавление | |
| **2.4б** | **Оценка погрешности** |  |
|  | **численного дифференцирования . . . . . . . . . .** | **66** |
| **2.4в** | **Метод неопределённых коэффициентов . . . . . .** | **70** |
| **2.4г** | **Полная погрешность дифференцирования . . . . .** | **71** |

1. **Приближение в евклидовых пространствах . . . . . . . . 74 2.5а Обсуждение постановки задачи . . . . . . . . . . . 74 2.5б Метод наименьших квадратов . . . . . . . . . . . . 76 2.5в Полиномы Лежандра . . . . . . . . . . . . . . . . . 81**
2. **Численное интегрирование . . . . . . . . . . . . . . . . . . 87 2.6а Простейшие квадратурные формулы . . . . . . . . 89 2.6б Квадратурная формула Симпсона . . . . . . . . . 92 2.6в Дальнейшие формулы Ньютона-Котеса . . . . . . 96**

|  |  |  |
| --- | --- | --- |
| **2.6г** | **Интерполяционные квадратурные формулы . . .** | **97** |
| **2.6д** | **Сходимость квадратур . . . . . . . . . . . . . . . .** | **98** |
| **2.6е** | **Составные квадратурные формулы . . . . . . . . .** | **101** |
| **2.6ж** | **Квадратурные формулы Гаусса . . . . . . . . . . .** | **104** |
| **2.6з** | **Выбор узлов для квадратурных формул Гаусса .** | **109** |
| **2.7 Правило Рунге для оценки погрешности . . . . . . . . . .** | | **114** |
| **Литература к Главе 2 . . . . . . . . . . . . . . . . . . . . . . . .** | | **115** |
| Глава 3. Численные методы линейной алгебры | | 118 |

1. **Задачи вычислительной линейной алгебры . . . . . . . . 118**

|  |  |  |
| --- | --- | --- |
| **3.1а** | **Особенности постановок задач . . . . . . . . . . .** | **118** |
| **3.1б** | **Сингулярные числа и сингулярные векторы . . .** | **120** |
| **3.1в** | **Матрицы с диагональным преобладанием . . . . .** | **125** |

1. **Нормы векторов и матриц . . . . . . . . . . . . . . . . . . 128**

|  |  |  |
| --- | --- | --- |
| **3.2а** | **Векторные нормы . . . . . . . . . . . . . . . . . . .** | **128** |
| **3.2б** | **Матричные нормы . . . . . . . . . . . . . . . . . .** | **131** |
| **3.2в** | **Подчинённые матричные нормы . . . . . . . . . .** | **133** |
| **3.2г** | **Топология на множествах векторов и матриц . . .** | **137** |
| **3.2д** | **Спектральный радиус . . . . . . . . . . . . . . . .** | **141** |
| **3.2е** | **Матричный ряд Неймана . . . . . . . . . . . . . .** | **144** |
| **3.2ж** | **Число обусловленности матриц . . . . . . . . . . .** | **146** |
| **3.2з** | **Примеры хорошообусловленных** |  |
|  | **и плохообусловленных матриц . . . . . . . . . . . .** | **150** |
| **3.2и** | **Практическое применение числа обусловленности** | **152** |

1. **Прямые методы решения линейных систем . . . . . . . . 155 3.3а Решение треугольных линейных систем . . . . . . 156 3.3б Метод Гаусса для решения линейных систем . . . 157**

|  |  |
| --- | --- |
| Оглавление | **5** |

**3.3в Матричная интерпретация метода Гаусса . . . . . 160 3.3г Существование LU-разложения . . . . . . . . . . . 163 3.3д Метод Гаусса с выбором ведущего элемента . . . . 165 3.3е Обусловленность и матричные преобразования . . 168 3.3ж QR-разложение матриц . . . . . . . . . . . . . . . . 171 3.3з Ортогональные матрицы отражения . . . . . . . . 173 3.3и Метод Хаусхолдера . . . . . . . . . . . . . . . . . . 176 3.3к Матрицы вращения . . . . . . . . . . . . . . . . . . 179 3.3л Ортогонализация Грама-Шмидта . . . . . . . . . . 181 3.3м Метод Холесского . . . . . . . . . . . . . . . . . . . 181 3.3н Метод прогонки . . . . . . . . . . . . . . . . . . . . 185**

1. **Итерационные методы для систем линейных уравнений . 190 3.4а Условие сходимости стационарных методов . . . . 191 3.4б Подготовка системы к итерационному процессу . 196 3.4в Оптимизация скалярного предобуславливателя . . 199 3.4г Итерационный метод Якоби . . . . . . . . . . . . . 202 3.4д Итерационный метод Гаусса-Зейделя . . . . . . . . 205 3.4е Методы релаксации . . . . . . . . . . . . . . . . . . 210**

|  |  |  |
| --- | --- | --- |
| **3.4ж** | **Оценка погрешности итерационного метода . . . .** | **214** |
| **3.5 Нестационарные итерационные методы . . . . . . . . . .** | | **217** |
| **3.5а** | **Краткая теория . . . . . . . . . . . . . . . . . . . .** | **217** |
| **3.5б** | **Метод наискорейшего спуска . . . . . . . . . . . .** | **220** |
| **3.5в** | **Метод сопряжённых градиентов . . . . . . . . . .** | **227** |

1. **Вычисление обратной матрицы и определителя . . . . . . 227**
2. **Линейная задача о наименьших квадратах . . . . . . . . 229**
3. **Решение проблемы собственных значений . . . . . . . . . 230 3.8а Обсуждение постановки задачи . . . . . . . . . . . 230 3.8б Обусловленность проблемы собственных значений 232**

|  |  |  |
| --- | --- | --- |
| **3.8в** | **Коэффициенты перекоса матрицы . . . . . . . . .** | **236** |
| **3.8г** | **Круги Гершгорина . . . . . . . . . . . . . . . . . .** | **241** |
| **3.8д** | **Степенной метод . . . . . . . . . . . . . . . . . . .** | **243** |
| **3.8е** | **Обратные степенные итерации . . . . . . . . . . .** | **251** |
| **3.8ж** | **Сдвиги спектра . . . . . . . . . . . . . . . . . . . .** | **251** |
| **3.8з** | **Метод Якоби . . . . . . . . . . . . . . . . . . . . . .** | **254** |
| **3.8и** | **Базовый QR-алгоритм . . . . . . . . . . . . . . . .** | **259** |
| **3.8к** | **Модификации QR-алгоритма . . . . . . . . . . . .** | **262** |
| **Литература к Главе 3 . . . . . . . . . . . . . . . . . . . . . . . .** | | **265** |

|  |  |  |
| --- | --- | --- |
| **6** | Оглавление | |
| Глава 4. Решение нелинейных уравнений и их систем | | 270 |

1. **Вычислительно-корректные задачи . . . . . . . . . . . . . 272 4.1а Формальные определения . . . . . . . . . . . . . . 272 4.1б Задача решения уравнений**

**не является вычислительно-корректной . . . . . . 274 4.1в** ε**-решения уравнений . . . . . . . . . . . . . . . . . 275 4.1г Недостаточность** ε**-решений . . . . . . . . . . . . . 277**

1. **Теоретические основы численных методов . . . . . . . . . . . . . . . . . . . . . . 279**
2. **Классические методы решения уравнений . . . . . . . . . 283 4.3а Предварительная локализация решений . . . . . . 284 4.3б Метод дихотомии . . . . . . . . . . . . . . . . . . . 285 4.3в Метод простой итерации . . . . . . . . . . . . . . . 285**

|  |  |  |
| --- | --- | --- |
| **4.3г** | **Метод Ньютона и его модификации . . . . . . . .** | **287** |
| **4.3д** | **Методы Чебышёва . . . . . . . . . . . . . . . . . .** | **287** |
| **4.4 Классические методы решения систем уравнений . . . .** | | **287** |
| **4.4а Метод простой итерации . . . . . . . . . . . . . . .** | | **287** |
| **4.4б** | **Метод Ньютона и его модификации . . . . . . . .** | **287** |

1. **Интервальные линейные системы уравнений . . . . . . . 287**

|  |  |  |
| --- | --- | --- |
| **4.6** | **Интервальные методы решения уравнений . . . . . . . .** | **289** |
|  | **4.6а Одномерный интервальный метод Ньютона . . . .** | **292** |
|  | **4.6б Многомерный интервальный метод Ньютона . . .** | **295** |
|  | **4.6в Метод Кравчика . . . . . . . . . . . . . . . . . . . .** | **297** |
| **4.7** | **Глобальное решение уравнений . . . . . . . . . . . . . . .** | **299** |
| **Литература к Главе 4 . . . . . . . . . . . . . . . . . . . . . . . .** | | **302** |
| Обозначения | | 305 |
| Краткий биографический словарь | | 308 |
| Предметный указатель | | 312 |

Предисловие

**Представляемый вниманию читателей курс методов вычислений напи-сан на основе лекций, которые читаются автором на механико-матема-тическом факультете Новосибирского государственного университета. Его содержание в основной своей части традиционно и повторяет на современном уровне тематику, заданную ещё в знаменитых ¾Лекци-ях о приближённых вычислениях¿ акад. А.Н. Крылова, первый в мире систематический учебник вычислительных методов. Условно материал книги можно назвать ¾вычислительные методы-1¿, поскольку в стан-дарте университетского образования, существует вторая часть курса, посвящённая численному решению дифференциальных уравнений, как обыкновенных, так и в частных производных, интегральных уравнений и др.**

**Вместе с тем, книга имеет ряд особенностей. Во-первых, в ней широ-ко представлены элементы интервального анализа и современные ин-тервальные методы для решения традиционных задач вычислительной математики. Во-вторых, автор счёл уместным поместить в книгу крат-кий очерк идей конструктивной математики и теории сложности вы-числений, тесно связанных с предметом математики вычислительной.**

**7**

Глава 1

Введение

**Курс методов вычислений является частью более широкой математи-ческой дисциплины вычислительной математики, которую можно неформально определить ¾как математику вычислений¿ или ¾матема-тику, возникающую в связи с разнообразными процессами вычисле-ний¿.**

**Три типа задач, в основном, интересуют нас в связи с процессом вычислений:**

* **Как конструктивно найти (вычислить) тот или иной математиче-ский объект или его конструктивное приближение?**
* **Какова трудоёмкость нахождения тех или иных объектов? может ли она быть уменьшена и как именно?**
* **Если алгоритм для нахождения некоторого объекта уже известен, то как наилучшим образом организовать вычисления по этому алгоритму на том или ином конкретном вычислительном устрой-стве? Например, чтобы при этом уменьшить погрешность вычис-ления и/или сделать его менее трудоёмким?**

**Вопросы из последнего пункта сделались особенно актуальными в свя-зи с развитием различных архитектур электронных вычислительных машин, в частности, с связи с вхождением в нашу повседневную жизнь многопроцессорных и параллельных компьютеров.**

**Ясно, что все три отмеченных выше типа вопросов тесно связаны между собой. К примеру, если нам удаётся построить алгоритм для**

**8**

|  |  |
| --- | --- |
| 1.1. Погрешности вычислений | **9** |

**решения какой-либо задачи, то, оценив сложность его исполнения, мы тем самым предъявляем и верхнюю оценку трудоёмкости решения этой задачи.**

**Развитие вычислительной математики в различные исторические периоды имело свои особенности и акценты. Так, в период ¾холодной войны¿ (примерно вторую половину XX века) на первый план выдви-нулась разработка и применение конкретных практических алгорит-мов для решения сложных задач математического моделирования (в основном, вычислительной физики, механики и управления). Нужно было запускать и наводить ракеты, улучшать характеристики самолё-тов и других сложных технических устройств т. п. Ранее, в XVII–XIX веках, вычислительные методы гармонично входили в сферу научных интересов крупнейших математиков Ньютона, Эйлера, Лобачевско-го, Гаусса, Якоби и многих других, чьи имена остались в названиях некоторых численных методов.**

**Исторически сложилось, что исследования по второму пункту отно-сятся, главным образом, к различным теориям вычислительной слож-ности и к теории алгоритмов, которая в 30-е годы XX века вычле-нилась из абстрактной математической логики. Но традиционная вы-числительная математика, предметом которой считается построение и исследование конкретных численных методов, также немало способ-ствует прогрессу в этой области.**

1. Погрешности вычислений

**Для правильного учёта погрешностей реализации вычислительных ме-тодов на различных устройствах и для правильной организации этих методов нужно знать детали конкретного способа вычислений. В совре-менных электронных цифровых вычислительных машинах (ЭЦВМ), на которых выполняется подавляющая часть современных вычисле-ний, эти детали реализации регламентируются стандартом IEEE 754 и его дополнением и развитием, формализованным в виде стандарта IEEE 854 [33, 25].**

1. Обусловленность математических задач

**Вынесенный в заголовок этого параграфа термин обусловленность означает меру чувствительности решения задачи к изменениям (воз-**

**10** 1. Введение

**мущениям) её входных данных. Ясно, что любая информация подобно-го сорта чрезвычайно важна при практических вычислениях, так как позволяет оценивать достоверность результатов, полученных в усло-виях приближённого характера этих вычислений. С другой стороны, зная о высокой чувствительности решения мы можем предпринимать необходимые меры для компенсации этого явления повышать раз-рядность вычислений, наконец, модифицировать или вообще сменить выбранный вычислительный алгоритм и т. п.**

**Существует несколько уровней рассмотрения поставленного вопро-са. Во-первых, следует знать, является ли вообще непрерывной зависи-мость решения задачи от входных данных. Задачи, решение которых не зависит непрерывно от их данных, называют некорректными (выше на стр. 74 в качестве примера таких задач мы встречались с задачей диф-ференцирования). Во-вторых, в случае наличия этой непрерывности желательно иметь некоторую количественную меру чувствительности решения как функции от входных данных.**

**Переходя к формальным конструкциям, предположим, что в рас-сматриваемой задаче по значениям из множества** D **входных данных мы должны вычислить решение задачи из множества ответов** S**. Отоб-ражение** φ:D → S**, сопоставляющее всякому** x **из** D **решение задачи из** S**, мы будем называть разрешающим отображением (или разре-шающим оператором). Отображение** φ **может быть выписано явным образом, если ответ к задаче задаётся каким-либо выражением. Часто разрешающее отображение задаётся неявно, как, например, при реше-нии системы уравнений**

F (a, x) = 0

**с входными параметрами** a**. Даже при неявном задании нередко можно теоретически выписать вид разрешающего отображения, как, к при-меру,** x=A−***1***b **при решении системы линейных уравнений** Ax=b **с квадратной матрицей** A**. Но в любом случае удобно предполагать суще-ствование этого отображения и некоторые его свойства. Пусть также** D **и** S **являются линейными нормированными пространствами. Очевид-но, что самый первый вопрос, касающийся обусловленности задачи, требует, чтобы разрешающее отображение** φ **было непрерывным отно-сительно некоторого задания норм в** D **и** S**.**

**Что касается числовой меры обусловленности математических за-дач, то существуют два подхода к её введению. Одни из них условно может быть назван дифференциальным, а другой основан на оценива-нии константы Липшица разрешающего оператора.**

|  |  |
| --- | --- |
| 1.2. Обусловленность математических задач | **11** |

**Пусть разрешающее отображение дифференцируемо по крайней ме-ре в интересующей нас точке** a **из множества входных данных** D**. Тогда можно считать, что**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| φ(a + a) | ≈ | φ(a) + φ0 | (a) | · | a, |  |
|  |  |  |  |  |

**и потому мерой чувствительности решения может служить** kφ0(a)k**. Для более детального описания зависимости различных компонент ре-шения** φ(a) **от** a **часто привлекают отдельные частные производные**

**∂φ**i **, т. е. элементы матрицы Якоби** φ0(a) **разрешающего отображения**

**∂a**j

φ**, которые при этом называют** **коэффициентами чувствительности.** **Интересна также мера относительной чувствительности решения, ко-торую можно извлечь из соотношения**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| kφ(a)k | ≈ | kφ(a)k | | · k | k | kak | | |  |
| φ(a + a) − φ(a) |  |  | φ0(a) | a |  |  | a | . |  |
|  |  |  |  |  |  |  |

**Второй подход к определению обусловленности требует нахождения как можно более точных констант** C***1* и** C***2* в неравенствах**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| kφ(a + | a) − φ(a)k ≤ C***1***k | | | | ak | | **(1.1)** |  |
| **и** |  |  |  |  |  |  |  |  |
| kφ(a + | a) − φ(a)k | ≤ | C | k | ak | . | **(1.2)** |  |
| kφ(a)k | |  | ***2*** kak | | |  |  |

**Величины этих констант, зависящие от задачи, а иногда и конкретных входных данных, берутся за меру обусловленности решения задачи.**

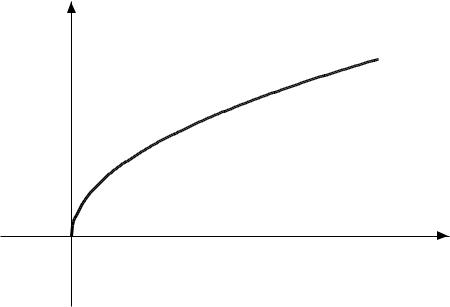
**В связи с неравенствами (1.1)–(1.2) напомним, что вещественная функция** f:R**n** ⊇D→R **называется** непрерывной по Липшицу **(или просто** липшицевой**), если существует такая константа** L**, что**

|f (x0) − f (x00)| ≤ L · dist (x0, x00)

**для любых** x0**,** x00∈D**. Величину** L **называют при этом** константой Лип-шица **функции** f **на** D**. Понятие непрерывности по Липшицу форма-лизует интуитивно понятное условие соразмерности изменения функ-ции изменению аргумента. Именно, приращение функции не должно превосходить приращение аргумента (по абсолютной величине или в некоторой заданной метрике) более чем в определённое фиксирован-ное число раз. При этом сама функция может быть и негладкой, как, например, модуль числа в окрестности нуля. Отметим, что понятие**

**12** 1. Введение

**непрерывности по Липшицу является более сильным свойством, чем просто непрерывность или даже равномерная непрерывность, так как влечёт за собой их обоих.**



**Рис. 1.1. Непрерывная функция** y ***=*** √x **имеет бесконечно большую скорость роста и не является липшицевой**

**Нетрудно видеть, что искомые константы** C***1* и** C***2* в неравенствах (1.1) и (1.2), характеризующие чувствительность решения задачи по отношению к возмущениям входных данных это не что иное, как константы Липшица для разрешающего отображения** φ **и произведение константы Липшица** L**ψ отображения** ψ:D → S**, действующего по правилу** a7→φ(a)/kφ(a)k **на норму** kak**. В последнем случае**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| kφ(a + a) − φ(a)k | / | L | **ψ**k | a | k ≤ | L | a | k · | k ak | . |  |
| kφ(a)k |  |  |  | **ψ**k | kak | |  |

1. Интервальная арифметика

**Исходной идеей создания интервальной арифметики является наблю-дение о том, что всё в нашем мире неточно и нам в реальности чаще всего приходится работать не с точными значениями величин, которые образуют основу классической ¾идеальной¿ математики, а с целыми диапазонами значений той или иной величины. Например, в цифровых ЭВМ множество вещественных чисел, которые точно представляются с помощью разрядной сетки, конечно, и из-за присутствия округления каждое из этих чисел, в действительности, является представителем целого интервала значений обычной вещественной оси** R**.**

**Нельзя ли организовать операции и отношения между диапазонами-интервалами так, как это сделано для обычных точных значений? С**

|  |  |
| --- | --- |
| 1.3. Интервальная арифметика | **13** |

**тем, чтобы можно было работать с ними, подобно обычным числам, опираясь на алгебраические преобразования, аналитические операции и т.п.? Ответ на эти вопросы положителен, хотя свойства получающей-ся ¾интервальной арифметики¿ оказываются во многом непохожими на привычные свойства операций с обычными числами.**

**Предположим, что нам даны переменные** x **и** y**, точные значения которых неизвестны, но мы знаем, что они могут находиться в ин-тервалах** [x, x] **и** [y, y] **соответственно. Что можно сказать о значении суммы** x+y**?**

**Складывая почленно неравенства**

x ≤ x ≤ x, y ≤ y ≤ y,

**получим**

x + y ≤ x + y ≤ x + y,

**так что** x+y∈x+y, x+y **.**

**На аналогичный вопрос, связанный с областью значений разности** x − y **можно ответить, складывая почленно неравенства**

x ≤ x ≤ x,

−y ≤ −y ≤ −y.

**Имеем** x−y∈x−y, x−y **. Для умножения**

x · y ∈ min{x y, x y, x y, x y}, max{x y, x y, x y, x y} .

**Рассмотрим множество всех вещественных интервалов** a:= [a,a] ={ x ∈ R | a ≤ x ≤ a }**, и бинарные операции сложение, вычитание,** **умножение и деление определим между ними ¾по представителям¿, т. е. в соответствии со следующим фундаментальным принципом:**

|  |  |
| --- | --- |
| a ? b := { a ? b | a ∈ a, b ∈ b } | **(1.3)** |

**для интервалов** a**,** b**, таких что выполнение точечной операции** a ? b**,** ? ∈ { +, −, · , / }**, имеет смысл для любых** a ∈ a **и** b ∈ b**. При этом ве-щественные числа** a **отождествляются с интервалами нулевой ширины** [a, a]**, называемыми также** **вырожденными интервалами. Кроме того,** **через** (−a) **условимся обозначать интервал** (−1)·a**.**

**14** 1. Введение

**Для интервальных арифметических операций развёрнутое опреде-ление, равносильное (1.3), задаётся следующими формулами:**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  | **(1.4)** |  |
| a + b = | a + b, |  | + b , | | |  |
| a |  |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| a − b = a − b, | | | |  | − b | | | | , |  |  |  |  |  |  |  |  |  |  |  |  | **(1.5)** |  |
| a |  |
| a · b = min{a b, a | | | | | |  | , |  | b, |  |  |  | } , max{a b, a |  | , |  | b, |  |  |  | } , | **(1.6)** |  |
| b | b | b | b |  |
| a | a | a | a |  |
| a/b = a · | 1/ |  | , 1/b | | | | | |  | **для** b630. | | | | | | | | | | | | **(1.7)** |  |
| b |  |  |

**Алгебраическая система** hIR,+,−,·, /i**, образованная множеством всех вещественных интервалов** a:= [a,a] ={x∈R|a≤x≤a} **с бинарными операциями сложения, вычитания, умножения и деления, которые определены формулами (1.4)–(1.7), называется** классическойинтервальной арифметикой**.**

**Алгебраические свойства классической интервальной арифметики существенно беднее, чем у поля вещественных чисел** R**. В частности, особенностью интервальной арифметики является отсутствие дистри-бутивности умножения относительно сложения: в общем случае**

(a + b)c =6 ac + bc.

**Например,**

[1, 2] · (1 − 1) = 0 =6 [−1, 1] = [1, 2] · 1 − [1, 2] · 1.

**Тем не менее, имеет место более слабое свойство**

|  |  |
| --- | --- |
| a(b + c) ⊆ ab + ac | **(1.8)** |

**называемое субдистрибутивностью умножения относительно сложе-ния. В ряде частных случаев дистрибутивность всё-таки выполняется:**

|  |  |  |
| --- | --- | --- |
| a(b + c) = ab + ac, | **если** a **вещественное число,** | **(1.9)** |
| a(b + c) = ab + ac, | **если** b,c≥0 **или** b,c≤0**.** | **(1.10)** |

**Сумма (разность) двух интервальных матриц одинакового размера есть интервальная матрица того же размера, образованная поэлемент-ными суммами (разностями) операндов.**

**Если** A= (a**ij** )∈IR**m**×**l и** B= (b**ij** )∈IR**l**×**n, то произведение матриц**

|  |  |
| --- | --- |
| 1.4. Интервальные расширения функций | **15** |

x

6***2***

-

x***1***

**Рис. 1.2. Интервальные векторы-брусы в** R***2*.**

A **и** B **есть матрица** C = ( c**ij** ) ∈ IR**m**×**n, такая что**

**l**

**X**

c**ij** := a**ik**b**kj** . **k*=1***

1. Интервальные расширения функций

**Задача об определении области значений функции на том или ином подмножестве области её определения, эквивалентная задаче оптими-зации, в интервальном анализе принимает специфическую форму за-дачи о вычислении так называемого интервального расширения функ-ции.**

Определение 1.4.1 **Пусть** D **непустое подмножество простран-ства** R**n. Интервальная функция** f:ID→IR**m называется интер-вальным продолжением вещественной функции** f:D→R**m, если** f (x) = f (x) **для всех** x ∈ D**.**

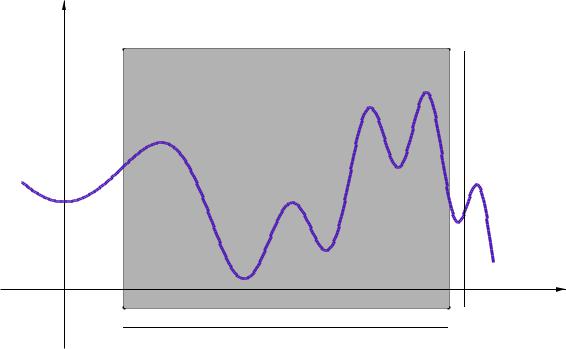
Определение 1.4.2 **Пусть** D **непустое подмножество простран-ства** R**n. Интервальная функция** f:ID→IR**m называется** интер-вальным расширением **вещественной функции** f : D → R**m, если**

1. f (x) **интервальное продолжение** f (x)**,**
2. f (x) **монотонна по включению, т. е.**

x0 ⊆ x00 ⇒ f (x0) ⊆ f (x00) **на** ID**.**

**Таким образом, если** f(x) **интервальное расширение функции** f (x)**, то для области значений** f **на брусе** X ⊂ D **мы получаем следу-**

**16** 1. Введение



|  |  |
| --- | --- |
|  | 6 |
|  | f (X) |
|  | - ? |
|  | X |
| **Рис. 1.3. Интервальное расширение функции** | |
| **даёт внешнюю оценку её области значений** | |

**ющую внешнюю (с помощью объемлющего множества) оценку:**

|  |  |  |
| --- | --- | --- |
|  | **[** | **[** |
|  | f (x) | x ∈ X = | f (x) =f (x) ⊆ f (X). |
|  | **x**∈X | **x**∈X |

**Эффективное построение интервальных расширений функций это важнейшая задача интервального анализа, поиски различных реше-ний которой продолжаются и в настоящее время. Уместно привести в рамках нашего беглого обзора некоторые общезначимые результаты в этом направлении. Первый из них часто называют ¾основной теоремой интервальной арифметики¿:**

Теорема 1.4.1 **Если для рациональной функции** f (x) = f (x***1*,** x***2*, . . . ,**

x**n**) **на брусе** x = (x***1***, x***2***, . . . , x**n**) ∈ IR**n** **определён результат** f \(x) **подстановки вместо её аргументов интервалов их изменения** x***1*,** x***2*,**

**. . . ,** x**n и выполнения всех действий над ними по правилам интер-вальной арифметики, то**

{ f (x) | x ∈ x } ⊆ f (x),

* **е.** f(x) **содержит множество значений функции** f(x) **на** x**.**

**Нетрудно понять, что по отношению к рациональной функции** f(x) **интервальная функция** f\(x)**, о которой идёт речь в Теореме 1.4.1, является интервальным расширением. Оно называется естественным интервальным расширением и вычисляется совершенно элементарно.**

|  |  |
| --- | --- |
| 1.4. Интервальные расширения функций | **17** |

Пример 1.4.1 **Для функции** f (x) = x/(x + 1) **на интервале** [1, 2] **об-ласть значений в соответствии с результатом Теоремы 1.4.1 можно оце-нить извне как**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| [1, 2] | = | [1, 2] | = [ ***1*** | | , 1]. |  |
|  |  |  |
| [1, 2] + 1 |  | [2, 3] | | ***3*** |  |  |
|  |  |  |  |

**Но если предварительно переписать выражение для функции в виде**

1

f (x) = 1 + 1/x ,

**разделив числитель и знаменатель дроби на** x=60**, то интервальное оценивание даст уже результат**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 |  | = | 1 |  |  | = [ ***2*** | , 1]. |  |
|  |  |  |  |  |  |
| 1 + 1/[1, 2] | | [1, | ***3*** |  |  |
|  | ]***3*** | |  |  |
|  |  |  |  | ***2*** |  |  |  |  |

**Он более узок, т. е. более точен, и совпадает к тому же с областью значений. Как видим качество интервального оценивания существенно зависит о вида выражения.**

**Использование естественного интервального расширения подчас да-ёт весьма грубые оценки областей значений функций, в связи с чем получили развитие более совершенные способы (формы) нахождения интервальных расширений. Одна из наиболее популярных так назы-ваемая центрированная форма:**

**n**

**X**

f**c**(x, x˜) = f (˜x) + g**i**(x, x˜)( x**i** − x˜**i**),

**i*=1***

**где** x˜ = ( x˜***1***, x˜***2***, . . . , x˜**n**) **некоторая фиксированная точка,**

**называемая ¾центром¿,**

g**i**(x, x˜) **интервальные расширения некоторых**

**функций** g**i**(x, x˜)**, которые строятся**

**по** f **и зависят в общем случае как**

**от** x˜**, так и от** x**.**

**В частности,** g**i**(x, x˜) **могут быть внешними интервальными оценками областей значений производных** ∂f(x)/∂x**i на** x**. За дальнейшей ин-формацией мы отсылаем заинтересованного читателя к книгам [1, 23, 28, 29], развёрнуто излагающим проблему построения интервальных расширений функций.**

**18** 1. Введение

**Для дальнейшего важно отметить, что точность интервального оце-нивания при использовании любой из форм интервального расширения критическим образом зависит от ширины бруса оценивания. Если обо-значить через** f(x) **точную область значений целевой функции на** x**, т. е.** f(x) ={f(x)|x∈x}**, то для естественного интервального расши-рения липшицевых функций имеет место неравенство**

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | **(1.11)** |  |
| dist f \(x), f (x) | ≤ C kwid xk |  |

**с некоторой константой** C**, и этот факт обычно выражают словами ¾естественное интервальное расширение имеет первый порядок точно-сти¿. Для центрированной формы верно соотношение**

|  |  |
| --- | --- |
| dist f**c**(x, x˜), f (x) ≤ 2 (wid g(x, x˜))>| x − x˜ |, | **(1.12)** |

**где** g(x, x˜) = (g***1***(x, x˜),g***2***(x, x˜), . . . ,g**n**(x, x˜) )**. В случае, когда интер-вальные оценки для функций** g**i**(x, x˜) **находятся с первым порядком**

**точности, общий порядок точности центрированной формы согласно (1.12) будет уже вторым. Вывод этих оценок заинтересованный чита-тель может найти, к примеру, в [23, 29].**

1. Элементы конструктивной математики

**Конструктивная математика это неформальное название той части современной математики, тех математических дисциплин, теории ал-горитмов, теории сложности вычислений, и ряда других в которых главным объектом изучения являются процессы построения тех или иных математических объектов,**

**В частности, теория алгоритмов и рекурсивных функций это ма-тематическая дисциплина, исследующая конструктивные свойства раз-личных математических объектов. Её основные понятия это алго-ритм, конструктивный объект, вычислимость, разрешимость и др.**

**Алгоритм это конечная последовательность инструкций, запи-санных на некотором языке и определяющих процесс переработки ис-ходных данных в искомые результаты (ответ решаемой задачи и т.п.). Алгоритм принципиально конечен и определяет собой конечный про-цесс. Далее, конструктивным объектом называется объект, который может быть построен с помощью некоторой конечной последователь-ности действий над каким-то конечным алфавитом. Таковы, например,**

|  |  |
| --- | --- |
| 1.5. Элементы конструктивной математики | **19** |

**рациональные числа. Строго говоря, конструктивные объекты и толь-ко они могут быть получены в качестве ответов при решении задачи на реальных цифровых ЭВМ с конечными быстродействием и объёмом памяти.**

**В частности, конечными машинами являются широко распростра-ненные ныне электронные цифровые вычислительные машины: они способны представлять, по сути дела, только конечные множества чи-сел. Таким образом, обречены на неудачу любые попытки использо-вать их для выполнения арифметических абсолютно точных операций над числовыми полями** R **и** C**, которые являются бесконечными (и да-же непрерывными) множествами, большинство элементов которых не представимы в цифровых ЭВМ.**

**Оказывается, что значительная часть объектов, с которыми работа-ют современная математика и её приложения, не являются конструк-тивными. В частности, неконструктивным является традиционное по-нятие вещественного числа, подразумевающее бесконечную процедуру определения всех знаков его десятичного разложения (которое в общем случае непериодично). Факт неконструктивности вещественных чисел может быть обоснован строго математически (см. [31]), и он указывает на принципиальные границы возможностей алгоритмического подхода и ЭВМ в деле решения задач математического анализа.**

**Тем не менее, и в этом океане неконструктивности имеет смысл вы-делить объекты, которые могут быть ¾достаточно хорошо¿ приближе-ны конструктивными объектами. На этом пути мы приходим к поня-тию вычислимого вещественного числа [31, 22]**1**: вещественное число** α

**называется вычислимым, если существует алгоритм, дающий по вся-кому натуральному числу** n **рациональное приближение к** α **с погреш-ностью n*1* . Множество всех вычислимых вещественных чисел образует**

**вычислимый континуум. Соответственно, вычислимая вещественная функция определяется как отображение из вычислимого континуума в вычислимый континуум, задаваемая алгоритмом преобразования про-граммы аргумента в программу значений.**

**Важно помнить, что и вычислимое вещественное число, и вычис-лимая функция это уже не конструктивные объекты. Но, как выяс-няется, даже ценой ослабления наших требований к конструктивности нельзя вполне преодолеть принципиальные алгоритмические трудно-сти, связанные с задачей решения уравнений. Для вычислимых веще-**

1 **Совершенно аналогичным является определение** конструктивного веществен-ного числа **у Б.А. Кушнера [30].**

**20** 1. Введение

**ственных чисел и функций ряд традиционных постановок задач оказы-вается алгоритмически неразрешимыми в том смысле, что построение общих алгоритмов их решения принципиально невозможно.**

**Например, алгоритмически неразрешимыми являются задачи**

1. **распознавания равно нулю или нет произвольное вычислимое ве-щественного число [30, 31, 32], распознавания равенства двух вы-числимых вещественных чисел [30, 31, 22, 24];**
2. **нахождения для каждой совместной системы линейных уравне-ний над полем конструктивных вещественных чисел какого-либо ее решения [30, 32];**
3. **нахождения нулей всякой непрерывной кусочно-линейной знако-переменной функции [32].**

**Приведённые выше результаты задают, как нам представляется, ту абсолютную и совершенно объективную мерку (в отличие от субъек-тивных пристрастий), с которой мы должны подходить к оценке тру-доёмкости тех или иных вычислительных методов. Получается, что необходимость переформулировки задачи решения уравнений и систем уравнений связана ещё и с тем, что в традиционной постановке эти задачи оказываются алгоритмически неразрешимыми! На фоне этого мрачного факта наличие даже экспоненциально трудного алгоритма с небольшим основанием ¾одноэтажной¿ экспоненты в оценке сложности (вроде** 2**n) можно рассматривать как вполне приемлемый вариант раз-решимости задачи, и именно это имеет место в ситуации с вычислением вращения векторного поля (степени отображения).**

**Вычислительная математика тесно примыкает к конструктивной**

1. Сложность задач и трудоёмкость алгоритмов

**Как правило, нас удовлетворит не всякий процесс решения интересу-ющей нас задачи, а лишь только тот, который выполним за практиче-ски приемлемое время. Соответственно, помимо алгоритмической раз-решимости задач огромную роль играет трудоёмкость тех или иных алгоритмов для их решения.**

**Получение оценок трудоёмкости задач является непростым делом. Если какой-то алгоритм решает поставленную задачу, то, очевидно,**

|  |  |
| --- | --- |
| Литература к Главе 1 | **21** |

**его трудоёмкость может служить верхней оценкой сложности решения этой задачи. Но вот получение нижних оценок сложности решения за-дач является чрезвычайно трудным. В явном виде таком нижние оцен-ки найдены лишь для очень небольшого круга задач, которые имеют, скорее, теоретическое значение. В этих условиях широкое распростра-нение получила альтернативная теория сложности, в основе которой лежат понятие сводимости задач друг к другу и вытекающее из него понятие эквивалентности задач по трудоёмкости.**

**NP-трудные задачи (универсальные переборные задачи) [10]**

Литература к Главе 1

Основная

1. **Алефельд Г., Херцбергер Ю. Введение в интервальные вычисления. –**

**Москва: Мир, 1987.**

1. **Барахнин В.Б., Шапеев В.П. Введение в численный анализ. – Санкт-Петербург–Москва– Краснодар: Лань, 2005.**
2. **Бабенко К.И. Основы численного анализа. – Москва: Наука, 1986.**
3. **Бауэр Ф.Л., Гооз Г. Информатика. В 2-х ч. – Москва: Мир, 1990.**
4. **Бахвалов Н.С., Жидков Н.П., Кобельков Г.М. Численные методы. –**

**Москва: Бином, 2003, а также другие издания этой книги.**

1. **Бахвалов Н.С., Корнев А.А., Чижонков Е.В. Численные методы. Реше-ния задач и упражнения. – Москва: Дрофа, 2008.**
2. **Березин И.С., Жидков Н.П. Методы вычислений. Т. 1–2. – Москва: Наука, 1966.**
3. **Вержбицкий В.М. Численные методы. Части 1–2. – Москва: ¾Оникс 21 век¿, 2005.**
4. **Волков Е.А. Численные методы. – Москва: Наука, 1987.**
5. **Гэри М., Джонсон Д. Вычислительные машины и труднорешаемые задачи.**

**– Москва: Мир, 1982.**

1. **Демидович Б.П., Марон А.А. Основы вычислительной математики. –**

**Москва: Наука, 1970.**

1. **Калиткин Н.Н. Численные методы. – Москва: Наука, 1978.**
2. **Крылов В.И., Бобков В.В., Монастырный П.И. Вычислительные методы. Т. 1–2. – Москва: Наука, 1976.**
3. **Кунц К.С. Численный анализ. – Киев: Техника, 1964.**
4. **Кунцман Ж. Численные методы. – Москва: Наука, 1979.**
5. **Мацокин А.М., Сорокин С.Б. Численные методы. Часть 1. Численный ана-лиз. Новосибирск: НГУ, 2006.**

**22** 1. Введение

1. **Меньшиков Г.Г. Локализующие вычисления. Конспект лекций. – Санкт-Петербург: СПбГУ, Факультет прикладной математики–процессов управле-ния, 2003.**
2. **Миньков С.Л., Миньков Л.Л. Основы численных методов. – Томск: Изда-тельство научно-технической литературы, 2005.**
3. **Райс Дж. Матричные вычисления и математическое обеспечение. – Москва: Мир, 1984.**
4. **Самарский А.А., Гулин А.В. Численные методы. – Москва: Наука, 1989.**
5. **Тыртышников Е.Е. Методы численного анализа. – Москва: Академия, 2007.**
6. **Успенский В.А., Семёнов А.Л. Теория алгоритмов: основные открытия и приложения. – Москва: Наука, 1987.**
7. **Шарый С.П. Конечномерный интервальный анализ. – Электронная книга, 2010 (см.** http://www.nsc.ru/interval/Library/InteBooks**)**
8. **Aberth O. Precise numerical methods using C++. – San Diego: Academic Press, 1998.**
9. **Goldberg D. What every computer scientist should know about floating point arithmetic // ACM Computing Surveys. – 1991. – Vol. 23, No. 1. – P. 5–48.**
10. **Hansen E., Walster G.B. Global optimization using interval analysis. – New York: Marcel Dekker, 2003.**
11. **Kreinovich V., Lakeyev A.V., Rohn J., Kahl P. Computational complexity and feasibility of data processing and interval computations. – Dordrecht: Kluwer, 1997.**
12. **Moore R.E., Kearfott R.B., Cloud M. Introduction to interval analysis. –**

**Philadelphia: SIAM, 2009.**

1. **Neumaier A. Interval methods for systems of equations. – Cambridge: Cambridge University Press, 1990.**

Дополнительная

1. **Кушнер Б.А. Лекции по конструктивному математическому анализу. –**

**Москва: Наука, 1973.**

1. **Мартин-Лёф П. Очерки по конструктивной математике. – Москва: Наука, 1975.**
2. **Математический Энциклопедический Словарь. – Москва: Большая Российская Энциклопедия, 1995.**
3. **IEEE Std 754-1985. IEEE Standard for Binary Floating-Point Arithmetic. – New York: IEEE, 1985.**

Глава 2

Численные методы анализа

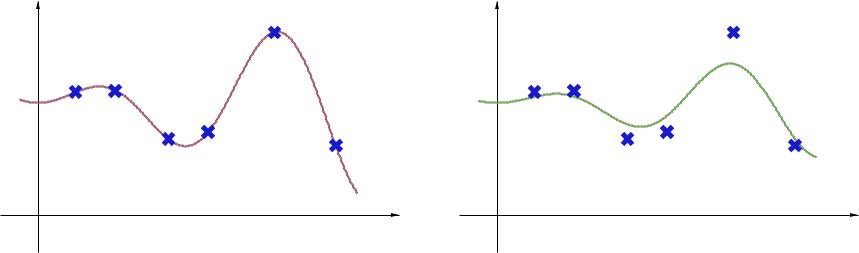
**Под численными методами анализа обычно понимаются методы реше-ния ряда задач, имеющих происхождение в классическом математи-ческом анализе. Традиционно сюда относят задачи интерполирования и приближения функций, задачу суммирования рядов, а также зада-чу численного нахождения определённых интегралов. Первыми в на-шем курсе будут рассмотрены задачи интерполирования и приближе-ния функций.**

**Задачи интерполирования и приближения функций (часто употреб-ляют также термин ¾аппроксимация функций¿) являются тесно свя-занными друг с другом задачами, которые укладываются в рамки сле-дующей единой неформальной схемы. Для заданных классов функций** F **и** G **найти способ, по которому любой заданной функции** f ∈ F **мо-жет быть сопоставлена функция** g∈G**, достаточно близкая (или даже наиболее близкая) к** f **в оговоренном заранее смысле. В зависимости от смысла, который вкладывается в понятие ¾близости¿ функций здесь могут получаться различные конкретные постановки задач. При этом часто считают, что** G⊂F**, и, кроме того, эти классы функций наделя-ются структурой линейного нормированного пространства и т. п.**

**Задача интерполирования получается из приведённой выше общей формулировки в случае, когда ¾близость¿ подразумевает совпадение функций** f **и** g **на некотором множестве точек из пересечения их об-ластей определения. Задача приближения (аппроксимации) функций**

**23**

**24** 2. Численные методы анализа



**Рис. 2.1. Различие задач интерполяции и приближения функций.**

**получается, когда ¾близость¿ понимается как малое отклонение** f **от** g

**относительно какой-то метрики (расстояния), заданной на пересечении семейств функций** F **и** G**. При этом равенство значений функции** g **за-данным значениям не требуется (см. Рис. 2.1). Разнообразные метрики, возникающие в практике математического моделирования, приводят к различным математическим задачам приближения.**

1. Интерполирование функций

**2.1а** **Постановка задачи**

**Задача интерполирования это задача восстановления (доопределе-ния) функции, которая задана на дискретном множестве точек** x**i,** i=0, 1, . . . , n**. Формальная её постановка такова.**

**Задан интервал** [a, b] **и конечное множество точек** x**i** ∈[a, b]**,** i=0, 1, . . . , n**, называемых** **узлами интерполяции. Даны значения** y**i,** i = 0, 1, . . . , n**. Требуется построить функцию** g(x) **от непрерывного аргу-мента** x**, называемую интерполирующей функцией (или интерполян-том), которая в узлах** x**i принимает значения** y**i,** i= 0,1, . . . , n**.**

**Практическая значимость задачи интерполяции станет более понят-ной, если упомянуть про её применение для вычисления различных функций, как элементарных, таких как** sin**,** cos**,** exp**,** log**, так и более сложных, которые принято называть специальными. С подобной за-дачей человеческая цивилизация столкнулась очень давно, столетия и даже тысячелетия назад, и типичным способом её решения в доком-пьютерную эпоху было составление для нужд практики таблиц та-буляция для значений интересующей нас функции при некоторых**

|  |  |
| --- | --- |
| 2.1. Интерполирование функций | **25** |

**специальных фиксированных значениях аргумента, покрывающих об-ласть её определения. Например,** sin **и** cos **для аргументов, кратных** 1◦**, одному угловому градусу. Но как, имея подобную таблицу, найти** **значение интересующей нас функции для аргумента, который не вы-ражается целым числом градусов, скажем, синус угла** 17◦230**?**

**Здесь на помощь приходит интерполяция восстановление значе-ния функции в промежуточных точках по ряду известных значений в некоторых фиксированных опорных. Собственно, сам термин ¾интер-полирование¿ (¾интерполяция¿) был впервые употреблён в 1656 го-ду Дж. Валлисом при составлении астрономических и математических таблиц. Он происходит от латинского слова interpolare, означающего ¾переделывать¿, ¾ремонтировать¿, ¾подновлять¿.**

**Интересно, что с появлением и развитием цифровых ЭВМ это при-менение интерполяции не кануло в лету. В начальный период развития ЭЦВМ преобладал алгоритмический подход к вычислению элементар-ных функций, когда основной упор делался на создании алгоритмов, способных ¾на голом месте¿ вычислить функцию, исходя из какого-нибудь её аналитического представления, например, в виде быстрос-ходящегося ряда и т. п. (см., к примеру, [23, 24]). Но затем, по мере удешевления памяти ЭВМ и повышения её быстродействия, постепен-но распространился подход, очень сильно напоминающий старый доб-рый табличный способ, но уже на новом уровне. Хранение сотен кило-байт или даже мегабайт цифровой информации никаких проблем сей-час не представляет, и потому в современных компьютерах програм-мы вычисления функций (элементарных и специальных), как правило, включают в себя библиотеки затабулированных значений функции для фиксированных аргументов, опираясь на которые строится значение в нужной нам точке.**

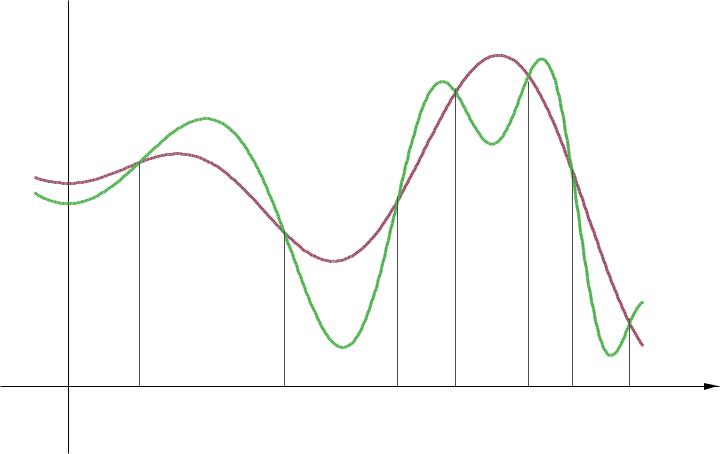
**Ещё один источник возникновения задачи интерполирования это желание иметь просто вычисляемое выражение для сложных функци-ональных зависимостей, которые могут быть заданы явно или неявно, но в исходной форме требуют очень большого труда для своего вычис-ления.**

**Если класс** G **интерполирующих функций достаточно широк, то яс-но, что решение поставленной задачи неединственно (см. Рис. 2.2). На-против, если** G **узок, то у задачи интерполяции может вовсе не быть решений. На практике выбор класса** G **обычно диктуется спецификой решаемой практической задачи.**

**Если, к примеру, заранее известно, что интерполируемая функция**

**26** 2. Численные методы анализа

 y



x***0*** x***1*** . . . . . . . x**n**

**Рис. 2.2. Задача интерполяции может иметь неединственное решение.**

**периодична, то в качестве интерполирующих функций естественно взять тоже периодические функции тригонометрические полиномы**

|  |  |  |
| --- | --- | --- |
| **m** |  |  |
| **X** |  |
|  | a**k** cos(kx) + b**k** sin(kx) , |  |
| **k*=0*** |  |  |

**(там, где требуется гладкость), либо пилообразные функции (в им-пульсных системах) и т. п.**

**Ниже мы подробно рассмотрим ситуацию, когда в качестве интер-**

|  |  |
| --- | --- |
| **полирующих функций берутся алгебраические полиномы** |  |
| a***0*** + a***1***x + a***2***x***2*** + · · · + a**m**x**m**, | **(2.1)** |

**которые просто вычислимы и являются хорошо изученным матема-тически объектом. При этом мы откладываем до §2.1з рассмотрение вопроса о том, насколько хороши полиномы для интерполирования. Вообще проблема наиболее адекватного выбора класса интерполиру-ющих функций** G **по заданному классу** F **является нетривиальной и требует, чтобы интерполирующие функции были ¾той же природы¿, что и интерполируемые функции из** F**. Если это не так, то задача ин-терполяции может решаться неудовлетворительно.**

Определение 2.1.1 **Интерполирование функций с помощью алгебра-ических полиномов называют** алгебраической интерполяцией**. Алгеб-раический полином** P**m**(x) =a***0*** +a***1***x+a***2***x***2*** +· · ·+a**m**x**m, решающий**

|  |  |
| --- | --- |
| 2.1. Интерполирование функций | **27** |

**задачу алгебраической интерполяции, называется** интерполяционнымполиномом**.**

**Как определить коэффициенты интерполяционного полинома? Подставим в него последовательно значения аргумента** x***0*,** x***1*, . . . ,**

x**n, получим соотношения**

a***0*** + a***1***x***0*** a***0*** + a***1***x***1***

**...** **...**

* a***2***x***20*** + · · · + a**m**x**m*0***
* a***2***x***21*** + · · · + a**m**x**m*1***

**.... .. ...**

* y***0***,
* y***0***,

|  |  |
| --- | --- |
| **.** | **(2.2)** |
| **.** |  |
| **.** |  |

a***0*** + a***1***x**n** + a***2***x***2*n** + · · · + a**m**x**mn** = y***0***,

**которые образуют систему линейных уравнений относительно неиз-вестных коэффициентов** a***0*,** a***1*,** a***2*, . . . ,** a**m искомого полинома. Решив её, можно построить и сам полином.**

**В самом общем случае, если мы не накладываем никаких ограни-чений на соотношение** m **и** n**, система (2.2) может не иметь решений, а если они существуют, то могут быть неединственными. Имеется, тем не менее, важный частный случай задачи алгебраической интервапо-ляции, в котором гарантируются разрешимость и единственность ре-шения.**

Теорема 2.1.1 **Если** m = n**, т. е. степень интерполяционного поли-нома на единицу меньше количества узлов, то решение задачи алгеб-раической интерполяции существует и единственно.**

Доказательство. **Матрица системы линейных уравнений (2.2) имеет** **вид**

1 x***0*** x***20*** . . . x**n*0***

1 x***1***

V (x***0***, x***1***, . . . , x**n**) = **.** **.**

**..** **..**

1 x**n**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ***2*** | .**. .**. **.**. | **n** |  |  |  |  |
| x**...*1*** |  | x**...*1*** |  | , | **(2.3)** |  |
| ***2*** | . . . | **n** |  |  |  |  |
| x**n** |  | x**n** |  |  |  |  |
|  |  |  |  |  |  |  |

**и является так называемой матрицей Вандермонда [17]. Её определи-тель равен, как известно, произведению**

**Y**

(x**j** − x**i**),

***1***≤**i<j**≤**n**

**28** 2. Численные методы анализа

**и он не зануляется, если узлы интерполяции** x**i попарно отличны друг от друга. Следовательно, система линейных уравнений (2.2) однознач-но разрешима тогда при любой правой части, т. е. при любых** y**i,** i=

0, 1, . . . , n**.**

**2.1б** **Интерполяционный полином Лагранжа**

**Теорема 2.1.1 и предшествующие ей рассуждения дают, в действитель-ности, конструктивный способ построения интерполяционного полино-ма через решение системы линейных алгебраических уравнений. Но в силу ряда причин этот способ не очень выгоден в вычислительном от-ношении. Решение систем линейных уравнений само по себе является не вполне тривиальной задачей. Кроме того, система (2.2) оказыва-ется очень чувствительной к возмущениям данных или, как принято говорить, плохо обусловлена (см. §1.2, а числовую оценку этой обу-словленности можно найти в §3.2з). Поэтому получаемый на этом пути интерполяционный полином может иметь большую погрешность.**

**На самом деле нам нечасто требуется знать коэффициенты канони-ческой формы (2.1) для интерполяционного полинома. Для большин-ства практических целей достаточно иметь некоторое конструктивное представление, позволяющее вычислять значения интерполяционного полинома в любой наперёд заданной точке.**

**Для отыскания такого представления заметим, что при фиксиро-ванных узлах** x***0*,** x***1*, . . . ,** x**n результат процесса интерполяции линей-ным образом зависит от данных** y***0*,** y***1*, . . . ,** y**n. Более точно, если поли-ном** P(x) **решает задачу интерполяции с данными** y= (y***0***, y***1***, . . . , y**n**)**, а полином** Q(x) **решает задачу интерполяции по тем же узлам с данными**

z = (z***0***, z***1***, . . . , z**n**)**, то для любых чисел** α**,** β ∈ R **полином** αP (x)+βQ(x) **решает задачу интерполяции для данных** αy+βz= (αy***0*** +βz***0***, αy***1*** +

βz***1***, . . . , αy**n** + βz**n**)**.**1

**Отмеченным свойством линейности можно воспользоваться для ре-шения задачи интерполяции ¾по частям¿, которые удовлетворяют от-дельным интерполяционным условиям, а затем собрать эти части во-**

1 **Сказанное можно выразить словами ¾оператор интерполирования линеен¿. В действительности, он даже является проектором, и эти наблюдения являются на-чалом большого и плодотворного направления теории приближения функций.**

(x − x***0***) · · · (x − x**i**−***1***)(x − x**i*+1***) · · · (x − x**n**)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 2.1. Интерполирование функций |  |  |  | **29** |  |
| **едино. Именно, будем искать интерполяционный полином в виде** | | | |  |  |
|  | **n** |  |  |  |  |
|  | **X** | |  | **(2.4)** |  |
| P**n**(x) = |  | y**i** φ**i**(x), | |  |
|  | **i*=0*** | |  |  |  |
| **где** φ**i**(x) **полином степени** n**, такой что** | | | |  |  |
| φ**i**(x**j** ) = δ**ij** = | **(** | 1, | **при** i=j, | **(2.5)** |  |
|  |  | 0, | **при** i6=j, |  |  |

i, j = 0, 1, . . . , n**, и посредством** δ**ij** **обозначен символ Кронекера. Поли-ном** y**i**φ**i**(x) **имеет степень** n **и решает задачу интерполяции по узлам** x***0*,** x***1*, . . . ,** x**n** **с данными** (0, . . . , 0, y**i**, 0, . . . , 0)**,** i = 0, 1, . . . , n**, и пото-му в силу представления (2.4) полином** P**n**(x) **в целом действительно удовлетворяет условиям задачи.**

**Коль скоро** φ**i**(x) **зануляется в точках** x***0*, . . . ,** x**i**−***1*,** x**i**−***1*, . . . ,** x**n, то ясно, что он должен иметь вид**

φ**i**(x) = Φ**i**(x − x***0***) · · · (x − x**i**−***1***)(x − x**i*+1***) · · · (x − x**n**) **(2.6)**

**с некоторым нормирующим множителем** Φ**i. Для определения** Φ**i под-ставим в выражение (2.6) значение аргумента** x=x**i и в силу (2.5) получим**

Φ**i**(x**i** − x***0***) · · · (x**i** − x**i**−***1***)(x**i** − x**i*+1***) · · · (x**i** − x**n**) = 1.

**Следовательно,**

1

Φ**i** = (x**i** − x***0***) · · · (x**i** − x**i**−***1***)(x**i** − x**i*+1***) · · · (x**i** − x**n**) ,

**и потому**

φ**i**(x) = (x**i** − x***0***) · · · (x**i** − x**i**−***1***)(x**i** − x**i*+1***) · · · (x**i** − x**n**) .

**Полиномы** φ**i**(x) **называют базисными полиномами Лагранжа, а иногда полиномами влияния** i**-го узла.**

**В целом, задачу алгебраической интерполяции решает полином**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **n** |  | **j**6***=*i**(x − x**j** ) | | | |  |
|  |  |  |  |
| P**n**(x) = | **i*=0*** | y**i** | **Qj*=*i**(x**i** | − | x**j** ) | . |  |
|  | **X** | | 6 |  |  |  |
|  |  |  | **Q** |  |  |  |  |

x**i*+2***

**30** 2. Численные методы анализа

**Его называют интерполяционным полиномом в форме Лагранжа или просто интерполяционным полиномом Лагранжа.**

**Далее нам потребуется его запись в несколько другом виде. Введём вспомогательную функцию**

ω**n**(x) = (x − x***0***) · · · (x − x**i**−***1***)(x − x**i**)(x − x**i*+1***) · · · (x − x**n**) **(2.7)**

**полином** (n+ 1)**-й степени, зануляющийся во всех узлах интерполя-**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **ции. Тогда** |  |  |  |  |  |  |  |  |  |  |  |  |
| φ**i**(x) = | |  |  |  | ω**n**(x) | |  |  | , |  | **(2.8)** |  |
| (x | − | | x**i**) ω0 | | (x**i**) | |  |  |
| **и в целом** |  |  |  | **n** |  |  |  |  |  |  |
| **n** | |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  | ω**n**(x) | | |  |  |  |  |
| P**n**(x) = |  | y**i** | |  |  |  | . |  |  |
|  | (x x**i**) ω**n**0(x**i**) | | | | | |  |  |
|  | **i*=0*** | |  |  | − |  |  |  |  |  |  |  |
|  | **X** | |  |  |  |  |  |  |  |  |  |

**2.1в** **Разделённые разности и их свойства**

**Пусть дана функция** f **и попарно различные точки** x***0*,** x***1*, . . . ,** x**n из её области определения, в которых функция принимает значения** f(x***0***)**,** f (x***1***)**, . . . ,** f (x**n**)**.** **Разделёнными разностями** **функции** f **, обозначаемы-ми** f∠(x**i**, x**i*+1***)**, называются отношения**

f ∠(x**i**, x**i*+1***) := f (x**i*+1***) − f (x**i**) , x**i*+1*** −x**i**

i = 0, 1, . . . , n − 1**. Разделённые разности второго порядка это вели-чины**

f ∠(x**i**, x**i*+1***, x**i*+2***) := f ∠(x**i*+1***, x**i*+2***) − f ∠(x**i**, x**i*+1***) ,

− x**i**

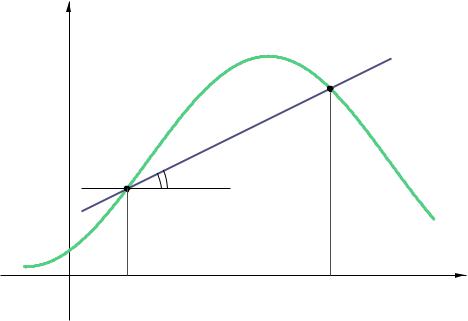
i = 0, 1, . . . , n − 2**, которые являются разделёнными разностями от раз-делённых разностей. Аналогичным образом вводятся разделённые раз-ности высших порядков: разделённая разность** k**-го порядка есть, по определению,**

f ∠(x**i**, x**i*+1***, . . . , x**i*+*k** ) := f ∠(x**i*+1***, . . . , x**i*+*k** ) − f ∠(x**i**, . . . , x**i*+*k**−***1***) ,

x**i*+*k** − x**i**

i = 0, 1, . . . , n − k**, т. е. она равна разделённой разности от разделённых** **разностей предыдущего** (k−1)**-го порядка. Для удобства и единооб-разия можно также считать, что сами значения функции являются разделёнными разностями нулевого порядка.**

|  |  |
| --- | --- |
| 2.1. Интерполирование функций | **31** |



x**i** x**i*+1***

**Рис. 2.3. Иллюстрация смысла разделённых разностей, как углового коэффициента секущей графика функции**

**Нетрудно увидеть геометрический смысл разделённой разности. Это угловой коэффициент (тангенс угла наклона к оси абсцисс) секущей графика функции** y=f(x)**, взятой между точками, имеющими аргу-менты** x**i и** x**i*+1*. В связи с этим уместно упомянуть, что разделённую разность иногда называют наклоном функции между заданными точ-ками (см. [16]). Разделённые разности-наклоны могут быть определены для функций многих переменных и даже для операторов, действующих из одного абстрактного пространства в другое. Интересно, что в нача-ле XX века для обозначения этой конструкции использовался также термин ¾подъём функции¿ [33].**

**Операция взятия разделённой разности является линейной: для лю-бых функций** f **,** g **и для любых скаляров** α**,** β **справедливо**

|  |  |
| --- | --- |
| (αf + βg)∠ = αf ∠ + βg∠. | **(2.9)** |

**Для разделённой разности первого порядка это очевидно следует из определения, а для разделённых разностей высших порядков показы-вается несложной индукцией по величине порядка.**

Предложение 2.1.1 **Имеет место представление**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **i*+*k** | | (x**j** |  | x**l**) | |  |  |
|  | **X** | |  |  |  |
| f ∠(x**i**, x**i*+1*** | , . . . , x**i*+*k** ) = |  | f (x**j** ) | |  | . | **(2.10)** |  |
| **i*+*k** |  |  |  |  |
|  |  | **Q** |  |  |  |  |  |  |
|  | **j*=*i** | |  | − |  |  |  |  |
|  |  | **l*=*i** |  |  |  |  |  |

**l*=***6**j**

**32** 2. Численные методы анализа

Доказательство. **Оно проводится индукцией по порядку** k **разделён-ной разности.**

**При** k= 1 **доказываемая формула, как нетрудно проверить, совпа-дает с определением разделённой разности первого порядка.**

**Пусть Предложение уже доказано при всех** k≤m **для некоторого положительного целого числа** m**. Тогда**

|  |  |  |  |
| --- | --- | --- | --- |
| f ∠(x**i**, x**i*+1***, . . . , x**i*+*m*+1***) | |  |  |
| = | f ∠(x**i*+1***, x**i*+1***, . . . , x**i*+*m*+1***) − f ∠(x**i**, x**i*+1***, . . . , x**i*+*m**) |  |  |
| x**i*+*m*+1*** − x**i** |  |  |
|  |  |
|  |  |  |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| = x**i*+*m*+1*** x**i** | | |  | **i*+*m*+1*** | **i*+*m*+1*** | **j** |  |  |  |
|  | 1 | − |  |  | (x**j** | x**l**) | |  |
| **j*=*i*+1*** |  |  |
|  | **X** | | f (x ) | | |  |  |
|  |  |  |  |  | − | |  |
|  |  |  |  |  | **l*=*i*+1*** |  |  |  |  |
|  |  |  |  |  | **lQ** |  |  |  |  |
|  |  |  |  |  | 6 |  |  |  |  |
|  |  |  |  |  | ***=*j** |  |  |  |  |

* f (x**i*+*m*+1***)

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  | **i*+*m** | |  |  |  |  |  |
| (x**i*+*m*+1*** − x**i**) | | | | | **l*=*Q** | |  |  |  |  |  |
|  | (x**i*+*m*+1*** − x**l**) | | | |  |  |
|  |  |  |  |  |  | **i*+1*** |  |  |  |  |  |
| 1 | |  |  |  |  | **i*+*m** | f (x ) | | |  |  |
| + |  |  |  |  |  |  |  | **j** |  |  |  |
| x**i*+*m*+1*** | − | x**i** | | **j*=*i*+1*** | **i*+*m*+1*** | (x**j** |  | x**l**) |  |
|  |  |  |  |  |  | − |  |
|  |  |  |  |  |  | **X** |  |  |  |  |
|  |  |  |  |  |  |  | **l*=*i*+1*** |  |  |  |  |
|  |  |  |  |  |  |  | **lQ** |  |  |  |  |
|  |  |  |  |  |  |  | 6 |  |  |  |  |
|  |  |  |  |  |  |  | ***=*j** |  |  |  |  |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| − **i*+*m** | f (x**j** ) | | |  |  |  |  |
| **i*+*m** | | | x**l**) | |  |
| **j*=*i** | | (x**j** | − |  |  |
| **X** | |  |  |  |  |  |
|  | **l*=*i** |  |  |  |  |  |  |
|  | **Q** |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |

**l*=***6**j**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **i*+*m** |  |  |  |  |  |
| − | f (x**j** ) | |  |  |  |
| **i*+*m** | | x**l**) |  |
| **j*=*i*+1*** | (x**j** | − |  |  |
| **X** | **l*=*i** |  |  |  |
|  |  |  |  |  |
|  | **Q** |  |  |  |  |
|  |  |  |  |  |  |

**l*=***6**j**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| − | f (x**i**) |  | . |  |
| **i*+*m** |  |  |
|  | **l*=*Q** | (x**i** − x**l**) | |  |
|  | (x**i*+*m*+1*** − x**i**) |  |
|  | **i*+1*** |  |  |  |

**В полученную сумму члены с** f(x**i**) **и** f(x**i*+*m*+1***) **входят по одному разу, причём их коэффициенты уже имеют вид, совпадающий с тем, что утверждает Предложение. Члены с остальными** f(x**j** ) **входят дважды, и после приведения подобных членов коэффициент при** f(x**j** ) **будет**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2.1. Интерполирование функций | | | | | | | | | | |  |  | **33** |  |
| **равен** | | |  |  |  |  |  |  |  |  |  |  |  |  |
| 1 | | |  |  | − | 1 |  |  |  |  |  |  |  |  |
|  |  | **i*+*m*+1*** |  |  | **i*+*m** |  |  |  |  |  |  |  |  |
|  |  | **Y** |  |  |  | **Yl** | − x**l**) | | | |  |  |  |  |
|  |  | (x**j** − x**l**) | | | | (x**j** |  |  |  |  |
|  |  | **l*=*i*+1*** |  |  |  | **l*=*i** |  |  |  |  |  |  |  |  |
|  |  | **l**6***=*j** |  |  |  | ***=*j** |  |  |  |  |  |  |  |  |
|  |  |  |  |  | 6 |  |  |  |  |  |  |  |  |
|  |  |  | x**i*+*m*+1*** − x**i** | | | |  |  |  |  |  |  |  |  |
| = | | |  | (x**j** − x**i**) − (x**j** − x**i*+*m*+1***) | | | | | |  | = | 1 | , |  |
|  |  | | | | | | **i*+*m*+1*** |  |
|  |  |  |  |  |  | **i*+*m*+1*** | | | | |  |  |  |
|  |  |  |  |  |  | **Yl** | | | | |  | **Y** |  |  |
|  |  |  |  | (x**i*+*m*+1*** − x**i**) | | | (x**j** − x**l**) | | | |  | (x**j** − x**l**) |  |  |
|  |  |  |  |  |  | **l*=*i** | | | | |  | **l*=*i** |  |  |
|  |  |  |  |  |  |  | ***=*j** | | | |  | **l*=*j** |  |  |
|  |  |  |  |  |  |  | 6 |  |  |  |  | 6 |  |  |
| **что и требовалось показать.** | | | | | | |  |  |  |  |  |  |  |  |

Следствие. **Разделённая разность симметричная функция своих** **аргументов, т. е. она не изменяется при любой их перестановке. Это непосредственно следует из вида выражения, стоящего в правой части представления (2.10).**

Пример 2.1.1 **Вычислим разделённые разности от полинома** g(x) =

x***3*** − 4x + 1**.**

**Прежде всего, воспользуемся свойством линейности разделённой разности (2.9), и будем искать по отдельности разделённые разности от мономов, образующих** g(x)**. Далее, вспомним известную из алгебры формулу**

|  |  |
| --- | --- |
| a**n** − b**n** = (a − b) a**n**−***1*** + a**n**−***2***b***2*** + · · · + ab**n**−***2*** + b**n**−***1*** . | **(2.11)** |

**Поэтому**

|  |  |  |
| --- | --- | --- |
| x***23*** − x***13*** | = x***22*** + x***2***x***1*** + x***12***. |  |
|  |
| x***2*** − x***1*** |  |
|  |  |

**Для линейного монома** (−4x) **разделённая разность находится триви-ально, а для константы она равна нулю. Следовательно, в целом**

g∠(x***1***, x***2***) = x***22*** + x***2***x***1*** + x***21*** − 4.

**34** 2. Численные методы анализа

**Вычислим вторую разделённую разность от** g(x)**:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| g∠(x***1***, x***2***, x***3***) = |  | g∠(x***2***, x***3***) − g∠(x***1***, x***2***) | | |  |  |
|  |  |  |
|  |  |  | x***3*** − x***1*** | |  |  |
| = |  | (x***32*** + x***3***x***2*** + x***22*** − 4) − (x***22*** + x***2***x***1*** + x***12*** − 4) | | | |  |
|  |  | | x***3*** − x***1*** | |  |
|  |  |  |  |  |
| = |  | x***32*** + (x***3*** − x***1***)x***2*** − x***12*** | | | = x***1*** + x***2*** + x***3***. |  |
|  |  |  | x***3*** − x***1*** | |  |  |
| **Третья разделённая разность** | | | | |  |  |
| g∠(x***1***, x***2***, x***3*** | , x***4***) = | |  | g∠(x***2***, x***3***, x***4***) − g∠(x***1***, x***2***, x***3***) | |  |
|  | x***4*** − x***1*** | |  |
|  |  |  |  |  |
|  | = | |  | (x***2*** + x***3*** + x***4***) − (x***1*** + x***2*** + x***3***) | |  |
|  |  | x***4*** − x***1*** | |  |
|  |  |  |  |  |

* + x***4*** − x***1*** = 1, x***4*** − x***1***
* **е. является постоянной. Четвертая и последующие разделённые раз-ности от** g(x) **будут, очевидно, тождественно нулевыми функциями.**

**Вообще, на основании формулы (2.11) нетрудно найти разделённую разность любого порядка от произвольного полинома. При этом взя-тие каждой разделённой разности уменьшает степень получающегося полинома на единицу, так что разделённые разности порядка более** n **от полинома степени** n **равны нулю. Именно это мы могли наблюдать в Примере 2.1в.**

**Сделанное наблюдение демонстрирует глубокую аналогию между разделёнными разностями и производными, применение которых так-же последовательно уменьшает степень полинома. В действительности, эта связь видна даже из определения разделённой разности, которую можно рассматривать как ¾недоделанную производную¿ и у которой отсутствует переход к пределу.**

Предложение 2.1.2 **(связь разделённых разностей с производными)**

**Пусть** f∈C**n**[a, b]**, т. е. функция** f **непрерывно дифференцируема** n **раз на интервале** [a, b]**, где расположены узлы** x***0*,** x***1*, . . . ,** x**n, и пусть** x = min{x***0***, x***1***, . . . , x**n**}**,** x = max{x***0***, x***1***, . . . , x**n**}**. Тогда**

f ∠(x***0***, x***1***, . . . , x**n**) = n1! f ***(*n*)***(ξ)

|  |  |
| --- | --- |
| 2.1. Интерполирование функций | **35** |

**для некоторой точки** ξ∈]x, x[**.**

**Для разделённых разностей первого порядка этот факт непосред-ственно следует из теоремы Лагранжа о среднем, а в общем случае его доказательство будет приведено несколько позже, в §2.1д.**

**2.1г** **Интерполяционный полином Ньютона**

**Выведем теперь другую форму интерполяционного полинома с учётом требования иметь такое выражение, которое в минимальной степени перестраивалось бы при смене набора узлов интерполяции.**

**Если представить**

|  |  |  |  |
| --- | --- | --- | --- |
| **n** |  |  |  |
| **X** | **(2.12)** |  |
| P**n**(x) = P***0***(x) + | P**i**(x) − P**i**−***1***(x) , |  |

**i*=1***

**где** P**i**(x) **интерполяционный полином, построенный по узлам** x***0*,** x***1*,**

**. . . ,** x**i, причём** P***0***(x) =f(x***0***)**, то при добавлении или удалении узлов перестройке должна подвергнуться лишь та часть выражения (2.12), которая вовлекает эти изменяемые узлы, а первые члены в (2.12) оста-нутся неизменными, коль скоро зависят только от первых узлов интер-поляции. Таким образом, стоящая перед нами задача окажется решён-ной, если будут найдены удобные и просто выписываемые выражения для разностей** P**i**(x)−P**i**−***1***(x)**.**

**Заметим, что разность** (P**i**(x)−P**i**−***1***(x)) **есть полином степени** i**, который обращается в нуль в узлах** x***0*,** x***1*, . . . ,** x**i**−***1*. Поэтому должно быть**

P**i**(x) − P**i**−***1***(x) = A**i**(x − x***0***)(x − x***1***) · · · (x − x**i**−***1***)

**с некоторой константой** A**i. Для её определения вспомним, что** P**i**(x**i**) =y**i** = f (x**i**) **по условию интерполяции. Следовательно,**

|  |  |  |  |
| --- | --- | --- | --- |
| A**i** = | f (x**i**) − P**i**−***1***(x**i**) | . |  |
|  |
|  | (x**i** − x***0***)(x**i** − x***1***) · · · (x**i** − x**i**−***1***) |  |  |

**Отсюда, пользуясь выражением для интерполяционного полинома в**

**36** 2. Численные методы анализа

**форме Лагранжа и Предложением 2.1.1, нетрудно вывести, что**

A**i** = **i**−***1***

**Q**

1

|  |  |  |  |
| --- | --- | --- | --- |
| · |  | **i**−***1*** |  |
| f (x**i**) − | f (x**j** ) |  |
|  |  | **X** |  |
|  |  |  |
|  |  |  |  |
|  |  |  |  |

|  |  |  |
| --- | --- | --- |
| **i**−***1*** |  |  |
| **k*=0***(x**i** − x**k**) |  |  |
| **k**6***=*j** |  |  |
| **Q** |  |  |
| **i**−***1*** |  |  |
|  |  |  |
|  |  |  |
| **Q** |  |  |
|  |  |  |

(x**i** − x**k**)

**k*=0***

= **i**−***1***f(x**i**)

**Q** (x**i** − x**k**)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | **j*=0*** | |  |  |  |
|  |  |  |  |  |  |
| **i**−***1*** |  | |  |  |  |  |
| − |  |  | 1 |  |  |  |
| **i**−***1*** | (x**i** |  | x**k**) |  |
| **j*=0*** |  | | − |  |
| **X** |  | |  |  |  |
|  |  | |  |  |  |  |
|  | **Q** | |  |  |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **k*=0***(x**j** − x**k**) | |  |  |  |  |
| **k**6***=*j** | |  |  |  |  |
|  | **i**−***1*** |  |  |  |  |
|  | **k*=0***(x**i** − x**k** ) | | |  |  |
|  | **k**6***=*j** |  |  |  |  |
| f (x**j** ) | **Q** |  |  |  |  |
|  | **i**−***1*** |  | x**k**) |  |  |
|  | (x**j** | − |  |  |  |
|  |  |  |  |  |  |
|  | **Q** |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

**k*=0***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | f (x**i**) | | **i**−***1*** |  |
| = |  | | **X** |  |
| (x**i** − x**k**) | |  |
| **i *1*** |  | − |  |
|  | − |  |  |  |
|  | **kQ** |  | **j*=0*** |  |
|  |  |  |  |
|  | ***=0*** |  |  |  |
|  | f (x**i**) | | **i**−***1*** |  |
| = |  |  | **X** |  |
| (x**i** | x**k**) |  |
| **i** | − | + |  |
|  | **Q** | **j*=0*** |  |
|  |  |  |  |

**k*=0*** **k*=0***

**k*=***6**j**

|  |  |  |
| --- | --- | --- |
| f (x**j** ) | |  |
|  | **i**−***1*** |  |
|  | **Q** |  |
| (x**i** − x**j** ) | **k*=0***(x**j** − x**k**) |  |
|  | **k**6***=*j** |  |
| f (x**j** ) | |  |
| (x**j** − x**i**) | **i**−***1*** |  |
|  | **Q** (x**j** − x**k**) |  |

**k*=0* k*=***6**i**

**i**

= **X** f (x**j** )

**i**

**j*=0*** **Q** (x**j** −x**k**)

**k*=0* k**6***=*j**

**k*=0* k*=***6**j**

= f ∠(x***0***, x***1***, . . . , x**i**).

**Окончательно**

P**n**(x) =

f (x***0***) + (x − x***0***) f ∠(x***0***, x***1***) + (x − x***0***)(x − x***1***) f ∠(x***0***, x***1***, x***2***) +

. . . + (x − x***0***)(x − x***1***) · · · (x − x**n**−***1***) f ∠(x***0***, x***1***, x***2***, . . . , x**n**).

**Это выражение называется интерполяционным полиномом в фор-ме Ньютона, или просто интерполяционным полином Ньютона. Оно**

|  |  |
| --- | --- |
| 2.1. Интерполирование функций | **37** |

**является равносильной формой записи интерполяционного полинома, широко применяемой во многих ситуациях, где использование формы Лагранжа по тем или иным причинам оказывается неудобным.**

**С учётом результата Предложения 2.1.2, т. е. равенства**

f ∠(x***0***, x***1***, . . . , x**n**) = n1! f ***(*n*)***(ξ)

**ясно видно, что интерполяционный полином Ньютона для гладкой фун-кции непрерывного аргумента является прямым аналогом формулы Тейлора (полинома Тейлора)**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| f (x ) + (x | − | x )f 0 | (x ) + (x | − | x )***2*** | f 00(x***0***) | + . . . + (x − x***0***) | **n** | f ***(*n*)***(x***0***) | . |  |
| 2! |  | n! |  |
| ***0*** | ***0*** | ***0*** | ***0*** |  |  |

**При этом аналогами степеней переменной** (x−x***0***)**k являются произ-ведения** (x−x***0***)(x−x***1***)· · ·(x−x**k** )**, которые в случае равномерного и упорядоченного расположения узлов** x***0*,** x***1*, . . . ,** x**k часто называют**

**обобщённой степенью [11].**

**2.1д** **Погрешность алгебраической интерполяции с простыми узлами**

**Задача интерполяции, успешно решённая в предшествующих парагра-фах, часто находится в более широком контексте, описанном во вве-дении к этой теме, на стр. 23. То есть, значения** y***0*,** y***1*, . . . ,** y**n при-нимаются в узлах** x***0*,** x***1*, . . . ,** x**n некоторой реальной функцией** f(x)**, свойства которой (хотя бы отчасти) известны. Насколько сильно будет отличаться от неё построенный нами интерполянт? Именно это отличие понимается под ¾погрешностью интерполяции¿.**

Определение 2.1.2 Остаточным членом **или** остатком **интерполиро-вания называется функция** R(f, x) =f(x)−g(x)**, являющаяся разно-стью рассматриваемой функции** f(x) **и интерполирующей её функции** g(x)**.**

Предложение 2.1.3 **Если точка** z **не совпадает ни с одним из узлов** **интерполирования, то в задаче алгебраической интерполяции оста-точный член** R**n**(f, z) :=f(x)−P**n**(x) **равен**

|  |  |
| --- | --- |
| R**n**(f, z) = f ∠(x***0***, x***1***, . . . , x**n**, z) · ω**n**(z), | **(2.13)** |

**38** 2. Численные методы анализа

**где функция** ω**n определяется посредством (2.7), т. е.**

**n**

**Y**

ω**n**(x) = (x − x**i**).

**i*=0***

Доказательство. **Выпишем для** f **интерполяционный полином Нью-тона по узлам** x***0*,** x***1*, . . . ,** x**n,** z**:**

P**n*+1***(x) = P**n**(x) + (x − x***0***)(x − x***1***) · · · (x − x**n**) f ∠(x***0***, x***1***, . . . , x**n**, z),

**где** P**n**(x) **полином Ньютона для узлов** x***0*,** x***1*, . . . ,** x**n. Подставляя в это соотношение значение** x=z**, получим**

P**n*+1***(z) = P**n**(z) + (z − x***0***)(z − x***1***) · · · (z − x**n**) f ∠(x***0***, x***1***, . . . , x**n**, z).

|  |  |  |
| --- | --- | --- |
| **Но** P**n*+1***(z) =f(z) **по построению полинома** P**n*+1*. Поэтому** | |  |
| R**n**(f, z) = | f (z) − P**n**(z) |  |
| = | (z − x***0***)(z − x***1***) · · · (z − x**n**) f ∠(x***0***, x***1***, . . . , x**n**, z), |  |
| **что и требовалось.** | |  |

Доказательство **Предложения 2.1.2 о связи разделённых разностей с** **производными.**

**Утверждение, сформулированное в этом Предложении, равносиль-но следующему: на** ]x***0***, x**n**[ **существует точка** ξ**, которая является нулём функции**

θ(x) = f ***(*n*)***(x) − n! f ∠(x***0***, x***1***, . . . , x**n**).

**Эта функция, в свою очередь, есть** n**-ая производная по** x **от функции погрешности аппроксимации** R**n**(f, x)**, т. е.**

f ***(*n*)***(x) − n! f ∠(x***0***, x***1***, . . . , x**n**) = R**n*(*n*)***(f, x),

**в чём можно убедиться непосредственным дифференцированием, бе-ря интерполяционный полином** P**n**(x) **в форме Ньютона. Поясним: в интерполяционном полиноме Ньютона только у разделённой разности** n**-го порядка** f ∠(x***0***, x***1***, . . . , x**n**) **коэффициент является полиномом** n**-ой** **степени от** x**, у остальных разделённых разностей эти коэффициенты суть полиномы меньших степеней, которые исчезнут при** n**-кратном**

|  |  |
| --- | --- |
| 2.1. Интерполирование функций | **39** |

**дифференцировании. А от полинома** n**-ой степени после дифференци-рования останется** n!**.**

**Итак, функция** R**n**(f, x) =f(x)−P**n**(x) **является** n**-кратно диффе-ренцируемой на** [a, b] **и обращается в нуль в** n+ 1 **различных точках** x***0*,** x***1*, . . . ,** x**n, которые без какого-либо ограничения общности можно** **считать упорядоченными, так что** x***0*** < x***1*** < . . . < x**n. В силу извест-ной теоремы Ролля производная** R**n**0(f, x) **обязана зануляться внутри** n

**интервалов** [x***0***, x***1***]**,** [x***1***, x***2***]**, . . . ,** [x**n**−***1***, x**n**]**, т. е. она имеет** n **корней. Далее, повторяя те же рассуждения в отношении второй производ-**

**ной** R**n**00(f, x)**, приходим к выводу, что она должна иметь на** ]x, x[ **не менее** n−1 **нулей, и т. д. вплоть до** R**n*(*n*)***(f, x)**, которая должна иметь на** ]x, x[ **хотя бы один нуль. Это и требовалось доказать.**

Теорема 2.1.2 **Если** f ∈ C**n*+1***[a, b]**, т. е. функция** f (x) **непрерывно** **дифференцируема** n+1 **раз на интервале** [a, b]**, то при её интерполиро-вании по попарно различным узлам** x***0*,** x***1*, . . . ,** x**n интерполяционным полиномом** P**n**(x) **остаточный член** R**n**(f, x) **может быть представ-лен в виде**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| R**n**(f, x) = | f ***(*n*+1)***(ξ(x)) | · ω**n**(x), | **(2.14)** |  |
| (n + 1)! |  |

**где** ξ(x) **некоторая точка, принадлежащая открытому интервалу** ] a, b [ **и зависящая от** x**, а функция** ω**n** **определена в** **(2.7).**

Доказательство. **Если** x = x**i** **для одного из узлов интерполирования,** **то** R**n**(f, x) = 0**, но в то же время и** ω**n**(x) = 0**, и потому в качестве** ξ **можно взять любую точку из открытого интервала** ]a, b[**.**

**Если же аргумент** x **остаточного члена не совпадает ни с одним из узлов интерполирования, то применяем результаты двух предшеству-ющих Предложений.**

**Полезны огрублённые оценки, удобные для применения при прак-тическом вычислении погрешности интерполирования:**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  | M**n*+1*** | |  | **(2.15)** |  |
| |R**n**(f, x)| | | | ≤ |  | · |ω**n**(x)|, | |  |
| (n + 1)! |  |
| **или даже совсем простая** | | |  |  |  |  |  |  |
| | | R | (f, x) | ≤ | M**n*+1***(b − a)**n*+1*** | | , | **(2.16)** |  |
| **n** | | | (n + 1)! | |  |  |  |

**40** 2. Численные методы анализа



x

ω**n**(x)

**Рис. 2.4. Типичное поведение полинома** ωn***(***x***)*: быстрый рост за пределами интервала узлов**

**где** M**n*+1*** = max **ξ**∈***[*a,b*]*** |f ***(*n*+1)***(ξ)|**.**

**Отметим, что полученные выше оценки (2.14) и её следствия (2.15) и (2.16) становятся неприменимыми, если гладкость функции** f **меньше, чем** n + 1**.**

**Поведение полинома** ω**n**(x) **при изменении** x **типично для полино-мов с вещественными корнями вообще. Если аргумент** x **находится на интервале** [x, x] **расположения его корней или ¾не слишком далёко¿ от него, то** ω**n**(x) **принимает относительно умеренные значения, так как формирующие его множители** (x−x**i**) **¾не слишком велики¿. Если же значения аргумента** x **существенно отличаются от его корней, то абсо-лютная величина** ω**n**(x)**, а вместе с ней и погрешность интерполяции, очень быстро растут. На Рис. 2.4 изображён график такого полинома нечётной (седьмой) степени. В связи с этим полезно на качественном уровне различать два случая. Если значения интерполируемой функ-ции ищутся в точках, далёких от интервала узлов интерполяции, ис-пользуют термин экстраполяция. Ей противопоставляется интерполя-ция в узком смысле, когда значения функции восстанавливаются на интервале, где расположены узлы, или недалеко от него. Как правило, экстраполяция сопровождается существенными ошибками и использу-ется ограниченным образом.**

|  |  |
| --- | --- |
| 2.1. Интерполирование функций | **41** |

**В рассмотренной выше постановке задачи интерполирования распо-ложение узлов считалось данным извне и фиксированным. Подобный подход соответствует тем практическим задачам, в которых измере-ния величины** y**i могут осуществляться, к примеру, лишь в какие-то фиксированные моменты времени** x**i, либо в определённых выделен-ных точках пространства и т. п., то есть заданы внешним образом и не могут быть изменены по нашему желанию.**

**Но существуют задачи, в которых мы можем управлять выбором уз-лов интерполирования. При этом естественно возникает вопрос о том, как сделать этот выбор наилучшим образом, чтобы погрешность ин-терполирования была как можно меньшей. В наиболее общей форму-лировке эта задача является весьма трудной, и её решение существенно завязано на свойства интерполируемой функции** f(x)**. Но имеет смысл рассмотреть и упрощённую постановку задачи, в которой на задан-ном интервале минимизируются значения полинома** ω**n**(x)**, а множи-тель** f ***(*n*+1)***(ξ(x))/(n+ 1)! **в выражении (2.14) для остаточного члена условно рассматривается как ¾приближённая константа¿.**

**Фактически, ответ на поставленный вопрос сводится к подбору уз-лов** x***0*,** x***1*, . . . ,** x**n в пределах заданного интервала** [a, b] **так, чтобы**

**полином** ω**n**(x) = (x−x***0***)(x−x***1***). . .(x−x**n**) **принимал ¾как можно мень-шие значения¿ на** [a, b]**. Конкретный смысл, который вкладывается в**

**эту фразу, может быть весьма различен, так как функция полином** ω**n**(x) **в нашем случае определяется своими значениями в бесконеч-ном множестве аргументов, и малость одних значений функции может иметь место наряду с очень большими значениями при других аргу-ментах. Ниже мы рассматриваем ¾отклонение от нуля¿ как максимум абсолютных значений функции на интервале, т.е.** max**x**∈***[*a,b*]*** ω**n**(x)**. Это условие является одним из наиболее часто встречающихся в приклад-ных задачах.**

**2.1е** **Полиномы Чебышёва**

**Полиномы Чебышёва это семейство полиномов, обозначаемых по традиции**2 **как** T**n**(x)**, и зависящее от неотрицательного целого пара-метра** n**. Они могут быть определены различными равносильными спо-собами, и наиболее просто и наглядно их тригонометрическое опреде-**

2 **Буква ¾T¿ является первой в немецком (Tschebyschev) и французском (Tchebychev) написаниях фамилии П.Л. Чебышёва, открывшего эти полиномы.**

|  |  |  |
| --- | --- | --- |
| **42** | 2. Численные методы анализа | |
| **ление:** |  |  |
| T**n**(x) = cos(n arccos x), | x ∈ [−1, 1], | **(2.17)** |

n = 0, 1, 2, . . . **. Полином степени** n **однозначно определяется своими** **значениями в** (n+ 1)**-точках, а формулой (2.17) мы фактически зада-ём значения функции в бесконечном множестве точек. Поэтому если посредством (2.17) в самом деле задаются полиномы, то они однознач-но определяются на всей вещественной оси, а не только для значений аргумента** x∈[−1,1]**.**

**Представление (2.17), в действительности, справедливо для любых** x ∈ R**, если под** arccos x **понимать комплексное значение и, соответ-ственно, рассматривать косинус от комплексного аргумента. Можно показать, что**

T**n**(x) = ch(n arch x),

**где ch гиперболический косинус, а arch обратная к нему функция, и это определение удобно применять при вещественных** x**, таких что**

|x| ≥ 1**.**

Предложение 2.1.4 T**n**(x) **полином степени** n**, и его старший ко-эффициент при** n≥1 **равен** 2**n**−***1*.**

Доказательство. **Мы проведём его индукцией по номеру** n **полинома** **Чебышёва. При** n= 0 **имеем** T***0***(x) = 1**, при** n= 1 **справедливо** T***1***(x) =x**, так что база индукции установлена.**

**Далее, из известной тригонометрической формулы**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 2 |  |  | 2 |  |  |
| cos α + cos β = 2 cos | α + β |  | cos | α − β |  |  |
|  |  |  |  |  |

|  |  |
| --- | --- |
| **следует, что** |  |
| cos (n + 1) arccos x + cos (n − 1) arccos x |  |
| = 2 cos(n arccos x) cos(arccos x) |  |
| = 2x cos(n arccos x). |  |
| **Следовательно,** |  |
| T**n*+1***(x) = 2x T**n**(x) − T**n**−***1***(x) | **(2.18)** |

**для любых** n= 1,2, . . . **.**

|  |  |
| --- | --- |
| 2.1. Интерполирование функций | **43** |

**Таким образом, если** T**n**(x) **и** T**n**−***1***(x) **являются полиномами, то** T**n*+1***(x) **также полином, степень которого на единицу выше степени** T**n**(x)**, а старший коэффициент в 2 раза больше.**

**Полученные в доказательстве рекуррентные формулы (2.18) позво-ляют последовательно выписывать явные алгебраические выражения для полиномов Чебышёва:**

T***0***(x) = 1,

T***1***(x) = x,

T***2***(x) = 2x***2*** − 1,

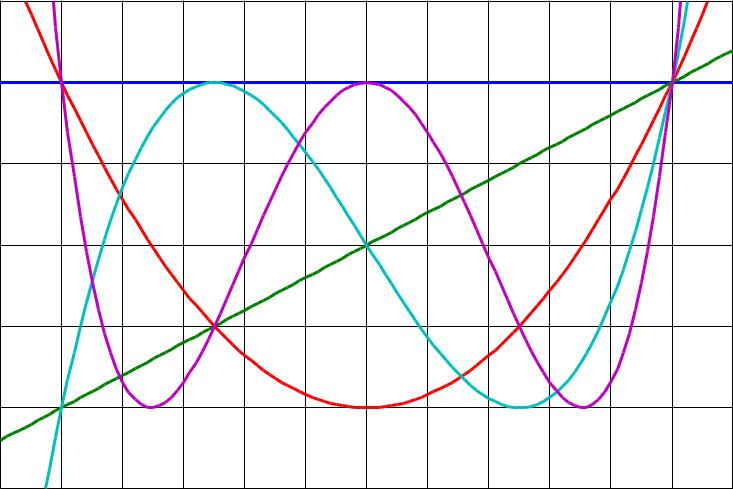
T***3***(x) = 4x***3*** − 3x,

T***4***(x) = 8x***4*** − 8x***2*** + 1,

T***5***(x) = 16x***5*** − 20x***3*** + 5x,

. . . .

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **По рекуррентным формулам (2.18) и следующим из них явным выра-** | | | | | | | | | | | |  |
| **жениям полиномы Чебышёва единообразно определяются для любых** | | | | | | | | | | | |  |
| **вещественных значений аргумента** x**.** | | | | | | |  |  |  |  |  |  |
| 1.5 |  |  |  |  |  |  |  |  |  |  |  |  |
| 1 |  |  |  |  |  |  |  |  | T***0*** |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
| 0.5 | T***4*** |  |  |  |  |  |  |  | T***1*** |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
| 0 |  |  |  |  |  |  |  |  |  |  |  |  |
| −0.5 |  |  |  |  |  |  |  |  |  |  |  |  |
| −1 |  |  |  |  |  | T***2*** |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  | T***3*** |  |  |  |  |  |  |  |  |  |  |  |
| −1.5 | −1 | −0.8 | −0.6 | −0.4 | −0.2 | 0 | 0.2 | 0.4 | 0.6 | 0.8 | 1 |  |
|  |  |
| **Рис. 2.5. Графики первых полиномов Чебышёва на интервале *[***−***1***.***2***, ***1***.***2]*.** | | | | | | | | | | | |  |



**Дальнейшие свойства полиномов Чебышёва:**

**44** 2. Численные методы анализа

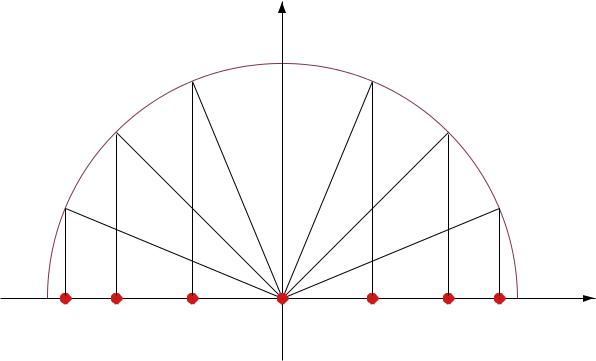
* 1. **При чётном (нечётном)** n **полином Чебышёва** T**n**(x) **есть чётная (нечётная) функция от** x**. Действительно, выражение для** T**n**(x) **при чётном** n **содержит только чётные степени** x**, а при нечётном** n **только нечётные степени** x**.**
  2. **Полином** T**n**(x) **имеет** n **вещественных корней, все они находятся**
* **интервале** [−1,1] **и выражаются формулой**

|  |  |  |  |
| --- | --- | --- | --- |
| ˚x**k** = cos | (2k − 1)π | ,k = 1, 2, . . . , n. |  |
| 2n |  |
|  |  |  |

**Геометрическая иллюстрация расположения корней полинома Чебы-шёва дана на Рис. 2.6, из неё видно что эти корни расположены суще-ственно неравномерно: они сгущаются к концам интервала** [−1,1]**, а в середине более разрежены.**

|  |  |  |  |
| --- | --- | --- | --- |
| **3)** max**x**∈***[***−***1*,*1]*** |T**n**(x)|= 1**, и этот максимум** | **достигается в точках** | |  |
| **s** |  |  |
| x**s** = cos(sπ/n)**,** s = 0, 1, . . . , n**, причём** T**n**(x**s**) = (−1) | | **,** s= 0,1, . . . , n**.** |  |

**Это непосредственно следует из тригонометрического определения по-линомов Чебышёва.**



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| −***1*** | ***0*** | ***1*** | x |  |
|  |  |

**Рис. 2.6. Иллюстрация расположения корней полиномов Чебышёва.**

* + ***1***−**n**

1. **Полином** T**n**(x) = 2T**n**(x)**,** n≥1**, среди всех полиномов** n**-ой степени со старшим коэффициентом 1 имеет на интервале** [−1,1] **наи-меньшее отклонение от нуля. Иными словами, если** Q**n**(x) **полином степени** n **со старшим коэффициентом 1, то**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **x** | max | | | | | Q | **n** | (x) | | ≥ | **x** | max | | | | | T˜ | (x) = 2***1***−**n**. | **(2.19)** |  |
|  | ***[*** | − | ***1*,*1]*** |  |  |  |  |  |  | ***[*** | − | ***1*,*1]*** |  | **n** | | |  |  |
|  | ∈ |  |  |  |  |  |  |  |  | ∈ |  |  |  |  |  |  |  |

|  |  |
| --- | --- |
| 2.1. Интерполирование функций | **45** |

Доказательство. **Предположим противное доказываемому, т. е. что** **для какого-то** Q**n**(x) **выполняется неравенство, противоположное по**

˜

**смыслу (2.19). Тогда полином** T**n**(x)−Q**n**(x) **имеет степень** n−1**, но в то же время в точках** x**s** = cos(sπ/n)**,** s= 0,1, . . . , n**, должно быть справедливо**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **n** | **s** | − | **n** | **s** |  | = sgn | (− | 1)**s** 2***1***−**n** | −= (**n** |  | 1)**s**. |
| sgn | T˜ | (x ) |  | Q | (x ) |  | = sgn | ( | 1)**s** 2***1***−**n** | Q | (x**s**) | |
|  |  |  |  |  |  |  |  | ˜ |  |  |  |  |
|  |  |  |  |  |  |  |  | − |  | − | |  |

**Итак, на интервале** [x**s**, x**s*+1***] **полином** T**n**(x)−Q**n**(x) **меняет знак, и потому обязан иметь корень. Коль скоро это происходит для** s= 0,1,

˜

**. . . ,** n−1**, т. е. всего** n **раз, то полином** T**n**(x)−Q**n**(x) **необходимо име-**

**ет** n **корней на** [−1,1]**, что невозможно из-за того, что степень этого полинома равна** n−1**.**

**Если мы осуществляем интерполяцию на интервале** [a, b]**, отличном от** [−1,1]**, то линейной заменой переменной**

y = ***12*** (b + a) + ***12*** (b − a) x

**интервал** [−1,1] **можно перевести в** [a, b]**. При этом**

x = 2y − (b + a) .

(b − a)

**Корням полинома Чебышёва на** [−1,1] **соответствуют тогда точки**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ˚y | = ***1*** | (b + a) + ***1*** | (b | − | a) cos | | | (2k − 1)π | | | | , | k = 1, 2, . . . , n. | |  |
| **k** | ***2*** | ***2*** |  |  |  |  |  |  | 2n | | |  |  |  |
| **из интервала** [a, b]**.** | | |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Итак, полином** | | |  | − |  | · |  |  |  |  | b − a | |  |  |  |
|  |  |  |  |  |  |  | **n** |  |  |
|  |  | 2***1***−***2*n**(b | |  | a)**n** |  | T | |  |  | 2x − (b + a) | | |  |  |
|  |  |  |  |  |  |  | | |  |

**наименее уклоняется от нуля на** [a, b] **среди всех полиномов** n**-ой сте-пени со старшим коэффициентом 1, и максимум его модуля на этом интервале равен** 2***1***−**n** (b−a)**n.**

**Возвращаясь к задаче интерполирования, можно сказать, что вы-бор узлов интерполирования по корням полиномов Чебышёва позволя-ет при прочих равных условиях существенно уменьшить погрешность. Помимо интерполирования полиномы Чебышёва имеют и другие важ-ные применения в различных задачах вычислительной математики.**

**46** 2. Численные методы анализа

**2.1ж** **Алгебраическая интерполяция с кратными узлами**

**Кратным узлом называют, по определению, узел, в котором информа-ция о функции задаётся более одного раза. Помимо значения функции это может быть какая-либо дополнительная информация о ней, к при-меру, значения производных и т.п. К задаче интерполяции с кратными узлами мы приходим, в частности, если количество узлов не больше, чем степень интерполяционного полинома, который нужно построить по этим узлам.**

**Рассмотрим следующую постановку задачи**

**Даны несовпадающие точки** x***0*,** x***1*, . . . ,** x**n узлы интерполирова-ния, в которых заданы значения** y**i*(*k*)*,** k= 0,1, . . . , N**i** −1**, их прини-мает интерполируемая функции** f **и её производные** f ***(*k*)***(x)**. При этом число** N**i называют кратностью узла** x**i. Требуется построить полином** H**m**(x) **степени** m**, такой что**

|  |  |
| --- | --- |
| H**m*(*k*)***(x**i**) = y**i*(*k*)***,i = 0, 1, . . . , n, k = 0, 1, . . . , N**i** − 1. | **(2.20)** |

Теорема 2.1.3 **Решение задачи алгебраической интерполяции с крат-ными узлами при** m=N***0***+N***1***+. . .+N**n**−1 **существует и единственно.**

Доказательство. **В канонической форме полином** H**m**(x) **имеет вид**

**m**

**X**

H**m**(x) = a**l**x**l**,

**l*=0***

**и для определения его коэффициентов** a***0*,** a***1*, . . . ,** a**m станем дифферен-цировать** H**m**(x) **и подставлять в него условия (2.20). Получим систему линейных алгебраических уравнений относительно** a***0*,** a***1*, . . . ,** a**m, в ко-**

**торой при** m=N***0***+N***1***+. . .+N**n**−1 **число уравнений** N***0***+N***1***+. . .+N**n**−1 **совпадает с числом неизвестных, равным** m+ 1**.**

**Итак, получена система линейных уравнений**

|  |  |
| --- | --- |
| Ga = y | **(2.21)** |

**с квадратной** (m+ 1)×(m+ 1)**-матрицей** G **и вектором неизвестных** a = (a***0***, a***1***, . . . , a**m**)> ∈ R**m*+1*, которую можно выписать даже в явном** **виде. Тем не менее, её прямое исследование весьма сложно, и потому мы пойдём окольным путём.**

|  |  |
| --- | --- |
| 2.1. Интерполирование функций | **47** |

**Для определения свойств матрицы** G **исследуем однородную систе-му, отвечающую нулевой правой части** y= 0**, т. е.**

Ga = 0.

**Вектор правой части** y **образован значениями интерполируемой функ-**

**ции и её производных** y**i*(*k*)* в узлах** x**i,** i= 0,1, . . . , n**. Однородная си-стема** Ga= 0 **отвечает случаю** y**i*(*k*)*** = 0 **для всех** i= 0,1, . . . , n**,** k=0, 1, . . . , N**i** − 1**. Каким является вектор решений** a **этой системы?**

**Если правая часть в (2.21) нулевая, то это означает, что полином** H**m**(x) **с учётом кратности имеет** N***0*** + N***1*** + . . .+ N**n** = m + 1 **корней, т. е.** **больше, чем его степень. Следовательно, он необходимо равен нулю,**

* **соответствующая однородная линейная система** Ga= 0 **имеет лишь нулевое решение. Как следствие, матрица** G **должна быть неособенной,**
* **неоднородная линейная система (2.21) однозначно разрешима при**

**любой правой части** y**. Это и требовалось доказать.**

**Задачу алгебраической интерполяции с кратными узлами в иссле-дуемой нами постановке часто называют также задачей эрмитовой ин-терполяции, а сам полином** H**m**(x) **решающий эту задачу, называют при этом интерполяционным полиномом Эрмита. Он может иметь, аналогично случаю интерполяции с простыми узлами, форму Лагран-жа либо форму Ньютона. Укажем сособ построения его лагранжевой формы.**

**Из выписанного при доказательстве Теоремы 2.1.3 представления вектора коэффициентов** a= (a***0***, a***1***, . . . , a**n**)> **интерполяционного поли-нома как решения системы (2.21) следует, что** a **линейно зависит от век-**

**тора данных** y**, т. е. от значений** y**i*(*k*)*,** k= 0,1, . . . , N**i** −1**,** i= 0,1, . . . , n**. Поэтому, как и в случае интерполирования с простыми узлами, можно**

**представить** H**m**(x) **в виде линейной комбинации**

|  |  |
| --- | --- |
|  | **n N**i−***1*** |
|  | **X X** |
| H**m**(x) = | y**i*(*k*)*** · φ**ik** (x), |
|  | **i*=0* k*=0*** |

**где** φ**ik** (x) **специальные ¾базисные¿ полиномы степени** m**, удовлетво-ряющие условиям**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| φ**ik** (x**j** ) = **(** | 1, | **при** | l = k | **и** i=j. | **(2.22)** |
| ***(*l*)*** | 0, | **при** | l 6= k | **или** i6=j, |  |

**48** 2. Численные методы анализа

**Полином** φ**ik** (x) **отвечает линейной системе (2.21) с вектор-столбцом правой части** y **вида** (0, . . . ,0,1,0, . . . ,0)>**, в котором все элементы ну-левые за исключением одного.**

**Каков конкретный вид этих базисных полиномов? Перепишем усло-вия (2.22) в виде**

|  |  |  |
| --- | --- | --- |
| φ**ik*(*l*)***(x**i**) = δ**kl**, | k = 0, 1, . . . , N**i** − 1, | **(2.23)** |
| φ**ik*(*l*)***(x**j** ) = 0, | j = 0, 1, . . . , i − 1, i + 1, . . . , n, | **(2.24)** |
|  | l = 0, 1, . . . , N**i** − 1. |  |

**Из второго условия следует, что**

φ**ik** (x) = (x − x***0***)**N*0*** . . . (x − x**i**−***1***)**N**i−***1*** (x − x**i*+1***)**N**i***+1*** . . . (x − x**n**)**N**n Q**ik**(x),

**где полином** Q**ik**(x) **должен определяться из первого условия (2.23). И так далее.**

**Мы не будем завершать этого построения, так как дальнейшие вы-кладки весьма громоздки, а алгоритм нахождения полинома из приве-дённых рассуждения вполне ясен . . .**

**Каков погрешность алгебраической интерполяции с кратными уз-лами?**

Теорема 2.1.4 **Если** f ∈ C**m*+1***[a, b]**, т. е. функция** f **непрерывно диф-ференцируема** m+ 1 **раз на интервале** [a, b]**, то погрешность её ин-терполирования по попарно различным узлам** x***0*,** x***1*, . . . ,** x**n** ∈[a, b] **полиномом** H**m**(x) **при условии** m=N***0*** +N***1*** +. . .+N**n** −1 **может быть представлена в виде**

|  |  |  |  |
| --- | --- | --- | --- |
|  | f ***(*m*+1)***(ξ(x)) | **n** |  |
| R**m**(f, x) = f (x) − H**m**(x) = | **iY** |  |
| (m + 1)! |  |
| · (x − x**i**)**N**i ,**(2.25)** |  |
|  |  | ***=0*** |  |

**где** ξ(x) **некоторая точка из** ]a, b[**, зависящая от** x**.**

Доказательство **этого результата вполне аналогично доказательству** **теоремы о погрешности интерполирования с простыми узлами и мы его опускаем.**

**2.1з** **Общие факты алгебраической интерполяции**

**Из математического анализа известна**

2.1. Интерполирование функций

Теорема Вейерштрасса. **Если** f : [a, b] → R **непрерывная** **то для всякого** >0 **существует полином** P**n**(x) **степени равномерно приближающий функцию** f **с погрешностью, не дящей , т. е. такой, что**

max |f (x) − P**n**(x)| ≤ .

**x**∈***[*a,b*]***

**49**

**функция,** n = n( )**,** **превосхо-**

**Этот результат служит теоретической основой приближения непре-рывных функций алгебраическими полиномами, обеспечивая существо-вание полинома, который сколь угодно близок к заданной непрерывной функции. Вместе с тем, теорема Вейерштрасса слишком обща и не да-ёт ответа на конкретные вопросы о решении задачи интерполирования, требующей совпадения значений функции и её интерполянта на данном множестве точек-узлов.**

**Как с теоретической, так и с практической точек зрения интересен вопрос о том, насколько малой может быть сделана погрешность ин-терполирования при возрастании числа узлов. Вообще, имеет ли место сходимость интерполяционных полиномов к интерполируемой функ-ции при неограниченном возрастании количества узлов?**

**Применение Теоремы 2.1.2 об оценке погрешности интерполирова-ния при возрастании числа узлов требует также увеличения гладко-сти интерполируемой функции** f **. Если эта гладкость ограничена, то сходимости интерполяционных полиномов к интерполируемой функ-ции может не быть. Один из первых примеров на этот счёт привёл С.Н. Бернштейн, рассмотрев на интервале** [−1,1] **алгебраическую ин-терполяцию функции** f(x) =|x| **по равноотстоящим узлам, включа-ющим и концы этого интервала. Можно показать (см. [9, 29]), что с возрастанием числа узлов соответствующий интерполяционный поли-ном не стремится к** |x| **ни в одной точке интервала** [−1,1]**, отличной от**

−1**,** 0 **и** 1**.**

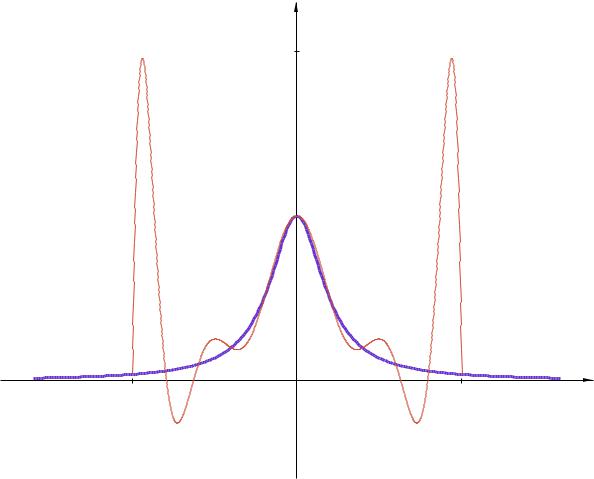
**Если интерполируемая функция** f **имеет бесконечную гладкость, т. е.** f∈C∞[a, b]**, и при этом её производные растут ¾не слишком быст-ро¿, так что**

|  |  |
| --- | --- |
| sup |f ***(*n*)***(x)| < M **n**, | **(2.26)** |

**x**∈***[*a,b*]***

**где** M **не зависит от** n**, то из Теоремы 2.1.2 следует, что погрешность**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **50** |  | 2. | Численные методы анализа | |  |
|  | ***2*** |  |  |  |  |
|  |  |  | P***10(***x***)*** |  |  |
| ψ***(***x***)*** |  |  |  |  |  |
| −***1*** | ***0*** |  | ***1*** | x |  |
|  |  |  |  |
| **Рис. 2.7. Интерполяция полиномом 10-й степени в примере Рунге** | | | | |  |



**итерполирования может быть оценена сверху как**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | (n + |  |  | **n*+1*** |  |  |
|  | M (b − a) | |  | , |  |
|  |  | |  |  |  |
|  |  | 1)! | |  |  |  |

**то есть очевидно сходится к нулю.**

**Но при алгебраическом интерполировании в самом общем случае погрешность всё-таки может не сходиться к нулю. По-видимому, наи-более известный из примеров такого рода привёл немецкий математик К. Рунге. В примере Рунге функция**

1 ψ(x) =1 + 25x***2***

**интерполируется алгебраическими полиномами на интервале** [−1,1] **с равномерным расположением узлов интерполяции** x**i** =−1 + 2i/n**,** i=0, 1, 2, . . . , n**. Оказывается, что тогда**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **n**lim | max | | | |ψ(x) − P**n**(x)| = ∞, |  |
| **x *[*** | − | ***1*,*1]*** |  |
| →∞ | ∈ |  |  |  |

P**n**(x) **интерполяционный полином** n**-ой степени. При этом с ростом** n **вблизи концов интервала интерполирования** [−1,1] **у полиномов** P**n**(x)

**возникают сильные осцилляции, размах которых стремится к бесконеч-ности (см. Рис. 2.7). Интересно, что на интервале** [−κ, κ]**, где** κ≈0.726**,**

|  |  |
| --- | --- |
| 2.1. Интерполирование функций | **51** |

**рассматриваемый интерполяционный процесс равномерно сходится к** ψ(x)**. Кроме того, полезно отметить, что функция** ψ(x) **имеет произ-водные всех порядков для любого вещественного аргумента** x**, но у кон-цов интервала интерполирования** [−1,1] **эти производные уже не удо-влетворяют условию (2.26). Таким образом, несмотря на простой вид, функция** ψ(x) **из примера Рунге своим поведением слишком непохожа на полиномы, производные от которых растут умеренно и, начиная с некоторого порядка, исчезают.**

**Чтобы строго сформулировать общие результаты о сходимости ал-гебраических интерполянтов, нам необходимо следующее**

Определение 2.1.3 **Говорят, что на интервале** [a, b] **задан** интерпо-ляционный процесс**, если задана бесконечная треугольная матрица уз-лов**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| x***(2)*** | | x***(2)*** | 0 | 0 |  | · · · | |  |  |
|  | x***0(1)*** | 0 | 0 | 0 |  |  |  | , |  |
| x***0(3)*** | | x***1(3)*** | x***2(3)*** | 0 |  | · · · | |  |
|  | ***0*** | ***1*** |  |  |  |  |  |
|  |  |  |  |  | · · · | |  |
|  | **...** | | | **. .** |  | **. .** |  |  |  |
|  | **...** | | | **.** | **.** |  |  |
|  | **...** | | |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |

**такая что в каждой её строке расположены различные точки ин-тервала** [a, b]**, т. е.** x***(*in*)*** ∈[a, b] **для всех** n **и всех** i= 0,1, . . . , n**, причём**

x***(*in*)*** 6= x***(*jn*)*** **для** i 6= j**.**

**Взяв элементы** n**-ой строки этой матрицы в качестве узлов интер-поляции, мы строим интерполяционный полином** P**n**(x)**.**

Определение 2.1.4 **Интерполяционный процесс для функции** f **на-зывается** сходящимся в точкеy∈[a, b]**, если** P**n**(y)→f(y) **при** n→ ∞**. Интерполяционный процесс для функции** f **называется** сходящимсяравномерно**, если** max**x**∈***[*a,b*]*** |f (x) − P**n**(x)| → 0 **при** n → ∞**.**

**Имеет место**

Теорема Фабера. **Не существует бесконечной треугольной матри-цы узлов из заданного интервала, такой что соответствующий ей интерполяционный процесс сходился бы равномерно для любой непре-рывной функции на этом интервале.**

**52** 2. Численные методы анализа

**В частности, даже для интерполяционного процесса по узлам поли-номов Чебышёва, рассмотренным в §2.1е и в некотором смысле наилуч-шим для интерполяции, существуют примеры непрерывных функций, для которых этот интерполяционный процесс всюду расходится. По-дробности можно найти в [29].**

**С другой стороны, справедлива**

Теорема Марцинкевича. **Для любой непрерывной на заданном ин-тервале функции** f **найдётся такая бесконечная треугольная матри-ца узлов из этого интервала, что соответствующий ей интерполя-ционный процесс для функции** f **сходится равномерно.**

**Доказательство этих теорем заинтересованный читатель может най-ти, к примеру, в [9, 10, 44].**

**Вывод из приведённых выше и, на первый взгляд кажущихся взаи-моисключающими, результатов состоит, прежде всего, в том, что мно-жество непрерывных функций является слишком широким, чтобы для него существовал один (или даже несколько) интерполяционных про-цессов, обеспечивающих сходимость для всех возможных ситуаций. С другой стороны, алгебраические полиномы, несмотря на определённые удобства работы с ними, оказываются слишком капризным инструмен-том интерполирования достаточно общих непрерывных функций, так что нам нужно озаботитьтся развитием более тонких методов интерпо-ляции. Этому и будут посвящёны следующие параграфы.**

1. Сплайны

**2.2а** **Элементы теории**

**Пусть заданный интервал** [a, b] **разбит на подынтервалы** [x**i**−***1***, x**i**]**,** i=1, 2, . . . , n**, так что** a = x***0*** **и** x**n** = b**.** **Сплайном** **на** [a, b] **называется** **функция, которая вместе со своими производными вплоть до некото-рого порядка непрерывна на всём интервале** [a, b]**, и на каждом подын-тервале** [x**i**−***1***, x**i**] **является некоторым полиномом. Максимальная на всём интервале** [a, b] **степень полиномов, задающих сплайн, называется степенью сплайна. Разность между степенью сплайна и наивысшим порядком его производной, которая непрерывна на** [a, b]**, называется дефектом сплайна. Наконец, точки** x**i,** i= 0,1, . . . , n**, называют узлами сплайна.**

|  |  |
| --- | --- |
| 2.2. Сплайны | **53** |

**Сплайны находят широкие применения в математике и могут ис-пользоваться для различных целей. Если сплайн применяется для ре-шения задачи интерполяции, то он называется интерполяционным. Другими словами, интерполяционный сплайн это сплайн, принима-ющий в заданных точках** x˜**i,** i= 0,1**, . . . ,** m**, узлах интерполяции требуемые значения** f**i, и эти узлы интерполяции, вообще говоря, могут не совпадать с узлами сплайна** x**i,** i= 0,1, . . . , n**.**

**Термин ¾сплайн¿ является удачным заимствованием из английско-го языка**3**, где слово ¾spline¿ означает гибкую (обычно стальную) ли-нейку, которую, изгибая, использовали чертёжники для проведения гладкой линии между данными фиксированными точками. В середине XX века этот термин вошёл в математику и далее перекочевал во мно-гие языки мира.**

**Таким образом, сплайны дефекта нуль это функции задаваемые на всём интервале** [a, b] **одной полиномиальной формулой, а термин ¾дефект¿ весьма точно выражает то, сколько сплайну ¾не хватает¿ до полноценного полинома. С другой стороны, именно наличие дефек-та обеспечивает сплайну б´ольшую гибкость в сравнении с полиномами и делает сплайны в некоторых ситуациях более удобным инструмен-том приближения и интерполирования функций. Чем больше дефект сплайна, тем больше он отличается от полинома и тем более специ-фичны его свойства. Но слишком большой дефект приводит к суще-ственному понижению общей гладкости сплайна. В значительном чис-ле приложений сплайнов вполне достаточным оказываются сплайны с минимально возможным дефектом 1, и только такие сплайны мы будем рассматривать далее в нашей книге.**

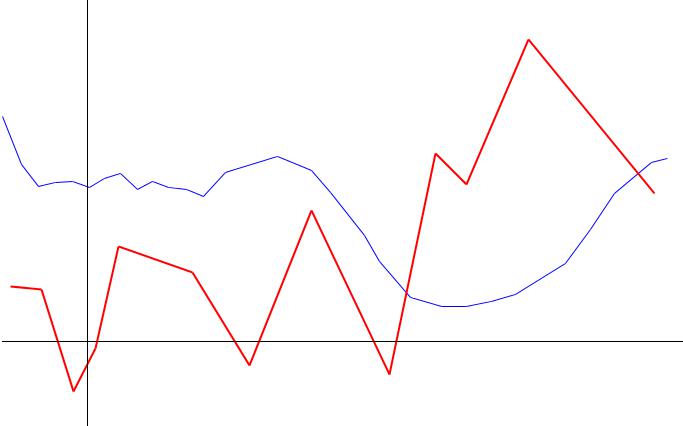
**Простейший ¾настоящий¿ сплайн имеет дефект 1 и степень 1, бу-дучи просто непрерывно склеенным в узлах** x**i,** i= 1,2, . . . , n−1**. Иными словами, это кусочно-линейная функция, имеющая несмотря на свою простоту богатые приложения в математике.**4 **Сплайны второй степени часто называют параболическими сплайнами.**

**Пусть имеется интерполяционный сплайн степени** d**, узлы которо-го** x***0*,** x***1*, . . . ,** x**n совпадают с узлами интерполирования. Этот сплайн полностью определяется** n(d+ 1) **значениями коэффициентов полино-**

3 **Пример неудачного заимствования термин ¾вейвлет¿, который в корне про-тиворечит фонетическому строю русского языка.**

4 **Вспомним, к примеру, ¾ломаные Эйлера¿, которые применяются для доказа-тельства существования решения задачи Коши для обыкновенных дифференци-альных уравнений.**

|  |  |  |
| --- | --- | --- |
| **54** | 2. | Численные методы анализа |
|  | 6 |  |
|  |  | - |
|  | **Рис. 2.8. Простейшие сплайны кусочно-линейные функции.** | |



**мов, задающих его на** n **подынтервалах** [x**i**−***1***, x**i**]**,** i= 1,2, . . . , n**. В то же время, в случае дефекта 1 имеется**

d(n − 1) **условий непрерывности в узлах** x***1*,** x***2*, . . . ,** x**n**−***1*** **самого** **сплайна и его производных вплоть до** (d−1)**-го порядка,**

(n + 1) **условие интерполяции в узлах** x***0*,** x***1*, . . . ,** x**n.**

**Всего** d(n−1) + (n+ 1) =n(d+ 1)−(d−1) **штук, и потому для опре-деления сплайна не хватает** d−1 **условий, которые обычно задают дополнительно на концах интервала** [a, b]**.**

**Сказанное имеет следущие важные следствия. Если решать зада-чу интерполяции с помощью сплайна чётной степени, требуя, чтобы на каждом подынтервале** [x**i**−***1***, x**i**] **сплайн являлся бы полиномом чёт-ной степени, то число** (d−1) **подлежащих доопределению параметров оказывается нечётным. Поэтому на одном из концов интервала** [a, b]

**приходится налагать больше условий, чем на другом. Это приводит, во-первых, к асимметрии задачи, и, во-вторых, может вызвать неустой-чивость при определении параметров сплайна. Наконец, интерполяци-онный сплайн чётной степени при некоторых естественных краевых условиях (периодических, к примеру) может просто не существовать.**

**Отмеченные недостатки могут решаться, к примеру, выбором узлов сплайна отличными от узлов интерполяции. Но мы далее рассмотрим сплайны нечётной степени 3.**

|  |  |
| --- | --- |
| 2.2. Сплайны | **55** |

**2.2б** **Интерполяционные кубические сплайны**

**Наиболее популярны в вычислительной математике сплайны третьей степени с дефектом 1, называемые также кубическими сплайнами.**

**Пусть задан набор узлов** x***0*,** x***1*, . . . ,** x**n** ∈[a, b]**, который мы будем**

**называть сеткой. Величину** h**i** =x**i** −x**i**−***1*,** i= 1,2, . . . , n**, мы будем называть шагом сетки. Кубический интерполяционный сплайн на ин-**

**тервале** [a, b] **с сеткой** a=x***0*** < x***1*** < . . . < x**n** =b **это функция** S(x)**, удовлетворяющая следующим условиям:**

1. S(x) **полином третьей степени на каждом** **из подинтервалов** [x**i**−***1***, x**i**]**,** i= 1,2, . . . , n**;**
2. S(x) ∈ C***2***[a, b]**;**
3. S(x**i**) = f**i,** i = 0, 1, 2, . . . , n**.**

**Для построения такого сплайна** S(x) **нужно определить** 4n **неизвестных величин по** 4 **коэффициента полинома третьей степени на каждом из** n **штук подинтервалов** [x**i**−***1***, x**i**]**,** i= 1,2, . . . , n**.**

* **нашем распоряжении имеются**

3(n − 1) **условий непрерывности самой функции** S(x)**, её первой**

* + **второй производных во внутренних узлах** x***1*,** x***2*, . . . ,** x**n**−***1*;**

(n + 1) **условие интерполяции** S(x**i**) = f**i,** i = 0, 1, 2, . . . , n**.**

**Таким образом, для определения** 4n **неизвестных величин мы имеем всего** (4n−2) **условий. Недостающие условия определяют различны-ми способами, среди которых могут быть задание на концах интерва-ла** [a, b] **первой или второй производной, либо условия периодичности:**

1. S0(a) = β***0*,** S0(b) = β**n;**
2. S00(a) = γ***0*,** S00(b) = γ**n;**
3. S***(*k*)***(a) = S***(*k*)***(b)**,** k = 0, 1, 2**.**

**Рассмотрим подробно второй случай задания краевых условий**

S00(a) = S00(x***0***) = γ***0***,

S00(b) = S00(x**n**) = γ**n**.

**56** 2. Численные методы анализа

**Будем искать кусочно-полиномиальное представление нашего кубиче-ского сплайна в виде**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| S(x) = α | + β | **i**−***1*** | (x | x | ) + γ | **i**−***1*** | (x − x**i**−***1***)***2*** | + ϑ | **i**−***1*** | (x − x**i**−***1***)***3*** |  |
|  |
| **i**−***1*** |  |  | − **i**−***1*** |  | 2 |  | 6 |  |
| **для** x∈[x**i**−***1***, x**i**]**,** i= 1,2, . . . , n**. При этом ясно, что** | | | | | | | | |  |  |  |
|  | S00(x**i**) = γ**i**, | | | | i = 0, 1, . . . , n − 1. | | | |  |  |  |

**Поскольку** S00(x) **является линейной функцией на** [x**i**−***1***, x**i**]**, то её вид полностью определяется двумя её крайними значениями** γ**i**−***1* и** γ**i на концах подинтервала** [x**i**−***1***, x**i**]**. Имеем поэтому**

S00(x) = γ**i**−***1*** x**i** − x + γ**i** x − x**i**−***1*** h**i** h**i**

**для** x∈[x**i**−***1***, x**i**]**,** i= 1,2, . . . , n**. Отметим, что в этих формулах мы уже задействовали известные значения** γ***0* и** γ**n второй производной** S00

**на левом и правом концах интервала** [a, b]**. Очевидно, что построенная таким образом функция** S00(x) **удовлетворяет условию ¾непрерывной склейки¿ в узлах** x***1*,** x***2*, . . . ,** x**n**−***1*, т. е.**

S00(x**i** − 0) = S00(x**i** + 0), i = 1, 2, . . . , n − 1.

**Беря дважды первообразную (неопределённый интеграл) от** S00(x)**, мы получим**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | S(x) = γ | | **i**−***1*** | | (x**i** − x)***3*** | | | + γ |  | (x − x**i**−***1***)***3*** | | | + C x + C | | |  | **(2.27)** |  |
|  |  |  |  |
|  |  |  |  |  | 6h**i** |  | **i** | 6h**i** | |  | ***1*** |  |  | ***2*** |  |  |
| **с какими-то константами** C***1* и** C***2*. Но нам будет удобно представить** | | | | | | | | | | | | | | | | | |  |
| **это выражение в виде** | | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| S(x) = γ | **i**−***1*** | (x**i** − x)***3*** | | + γ | |  | (x − x**i**−***1***)***3*** | | | + K | (x | x) + K | | (x | − | x | ), **(2.28)** |  |
|  | 6h**i** |  |  |  | **i** | 6h**i** | |  | ***1*** | **i** − |  | ***2*** |  | **i**−***1*** |  |  |

**где** K***1* и** K***2* также константы. Насколько законен переход к такой форме? Из сравнения (2.27) и (2.28) следует, что** C***1* и** C***2* должны быть связаны с** K***1* и** K***2* посредством формул**

C***1*** = −K***1*** + K***2***,

C***2*** = K***1***x**i** − K***2***x**i**−***1***.

|  |  |
| --- | --- |
| 2.2. Сплайны | **57** |

**Определитель выписанной системы линейных уравнений относительно** K***1*** **и** K***2*** **равен** x**i**−***1*** −x**i** = −h**i, он не зануляется, и потому переход от** C***1*** **и** C***2* к** K***1* и** K***2* это неособенная замена переменных. Следовательно, оба представления (2.27) и (2.28) совершенно равносильны друг другу.**

**Подставляя в выражение (2.28) значение** x=x**i**−***1* и используя усло-вие** S(x**i**−***1***) =f**i**−***1*, будем иметь**

γ**i**−***1*** (x**i** − x**i**−***1***)***3*** + K***1***(x**i** − x**i**−***1***) = f**i**−***1***,

6h**i**

**т. е.**

h***2***

γ**i**−***1*** 6**i** + K***1***h**i** = f**i**−***1***,

**откуда**

K***1*** = f**i**−***1*** − γ**i**−***1***h**i** . h**i** 6

**Совершенно аналогичным образом, подставляя в (2.28) значение** x=x**i и используя условие** S(x**i**) =f**i, найдём**

K***2*** = f**i** − γ**i**h**i** . h**i** 6

**Окончательно выражение сплайна на подинтервале** [x**i**−***1***, x**i**]**,** i= 1,2**,**

**. . . ,** n**, выглядит следующим образом**

S(x) = f**i**−***1*** x**i** − x + f**i** x − x**i**−***1*** h**i** h**i**

(x**i** − x)***3*** − h***2***(x**i** − x)

+ γ**i**−***1*** **i**

6h**i**

**(2.29)**

(x − x**i**−***1***)***3*** − h***2***(x − x**i**−***1***) + γ**i** **i** .

6h**i**

**Оно не содержит уже величин** α**i,** β**i и** ϑ**i, которые фигурировали ранее в представлении для** S(x)**, но неизвестными остались** γ***1*,** γ***2*, . . . ,** γ**n**−***1*.**

**Чтобы завершить определение вида сплайна, т. е. найти** γ***1*,** γ***2*, . . . ,** γ**n**−***1*, можно воспользоваться условием непрерывности первой произ-водной** S0(x) **в узлах** x***1*,** x***2*, . . . ,** x**n**−***1*:**

|  |  |
| --- | --- |
| S0(x**i** − 0) = S0(x**i** + 0),i = 1, 2, . . . , n − 1. | **(2.30)** |

**Продифференцировав формулу (2.29), будем иметь для** x∈[x**i**−***1***, x**i**]

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| S0(x) = | f**i** − f**i**−***1*** | − | γ | **i**−***1*** | 3(x**i** − x)***2*** − h**i*2*** | + γ | **i** | 3(x − x**i**−***1***)***2*** − h**i*2*** | . |  |
|  |
|  | h**i** |  | 6h**i** | | 6h**i** | |  |

**58** 2. Численные методы анализа

**Поэтому приравнивание согласно (2.30) производных, которые получе-ны в узлах интерполирования** x**i с соседних подинтервалов** [x**i**−***1***, x**i**] **и** [x**i**, x**i*+1***]**, приводит к системе уравнений**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 6 | |  | **i**−***1*** | 3 | |  |  | **i** | 6 | |  | **i*+1*** |  | h**i*+1*** | − | h**i** | |  |
|  |  | h**i** | γ |  | + | h**i** + h**i*+1*** | | γ | | + | h**i*+1*** | γ |  | = | f**i*+1*** − f**i** |  | f**i** − f**i**−***1*** | , |  |
|  |  |  |  | |  |  |  |  |  |
|  |  |  | i = 1, 2, . . . , n | | |  |  | 1, | |  |  |  |  |  |  | **(2.31)** | |  |
|  |  |  |  |  |  |  | − | |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | γ***0*** **и** γ**n** **заданы**. | | | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

**Это система линейных алгебраических уравнений относительно неиз-вестных переменных** γ***1*,** γ***2*, . . . ,** γ**n**−***1*, имеющая матрицу**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | ***1***h***2*** ***2*** |  | 2(h***2*** + h***3***) | h***3*** | | 0 | | |  |  |  |  |  |  |  |  |
|  | 2(h | + h | ) | h***2*** |  |  | **...** | | |  |  |  |  |  |  |  |  |
| 1 |  |  |  | h***3*** | 2(h***3*** + h***4***) | |  |  |  |  |  | , |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  |
| 6 |  |  |  | **.** | **. .** | **.** | | **..** |  | **.** | **..** | |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  | 0 |  |  | h | **n**−***1*** | | 2(h | **n**−***1*** | | + h | ) |  |  |  |
|  |  |  |  |  |  |  |  |  | **n** |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

**в которой ненулевыми являются лишь три диагонали главная и со-седние с ней сверху и снизу. Такие матрицы называются трёхдиаго-нальными. Кроме того, эта матрица обладает свойством диагонального преобладания (см. стр. 126), и потому в силу признака Адамара неосо-бенна. Следовательно, система линейных уравнений (2.31) однозначно разрешима при любой правой части, и искомый сплайн существует и единствен. Для нахождения решения системы (2.31) с трёхдиагональ-ной матрицей может быть с успехом применён метод прогонки, описы-ваемый ниже в §3.3н.**

**Интересен вопрос о погрешности интерполирования функций и их производных с помощью сплайнов, и ответ на него даёт следующая**

Теорема 2.2.1 **Пусть** f (x) ∈ C**p**[a, b]**,** p = 1, 2, 3**, а** S(x) **интерполя-ционный кубический сплайн с краевыми условиями (I), (II) или (III). Тогда**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **x** |  | f ***(*k*)***(x) | − |  | − | k = 0, . . . , p, |  |
| max | |  | S***(*k*)***(x) | = O(h**p k** ), |  |
|  |  |  |  |  |  |  |  |

**где** h= maxh**i.**

**i**

**Доказательство можно увидеть, к примеру, в [12, 15, 34].**

|  |  |
| --- | --- |
| 2.3. Нелинейные методы интерполяции | **59** |

**Отметим, что, в отличие от интерполяционных полиномов, последо-вательность интерполяционных кубических сплайнов на равномерной сетке узлов всегда сходится к интерполируемой непрерывной функции. Это относится, в частности, и к функции** f(x) = 1/(1+25x***2***) **из примера Рунге (см. §2.1з). Важно также, что с повышением гладкости интерпо-лируемой фукнции сходимость эта улучшается.**

**Интерполяционные кубические сплайны** S(x)**, удовлетворяющие на концах интервала** [a, b] **дополнительным условиям**

S00(a) = S00(b) = 0,

**называются естественными сплайнами. Их замечательное свойство состоит в том, что они минимизируют функционал**

E(ϕ) = **Z** **b** ϕ00(x) ***2*** dx,

**a**

**который в линейном приближении даёт выражение для энергии упру-гой деформации гибкой стальной линейки, принимающей форму, кото-рая описывается функцией** ϕ(x) **на заданном интервале** [a, b]**. Именно, если** S(x) **естественный сплайн, построенный по узлам** a=x***0*,** x***1*,**

**. . . ,** x**n** =b**, а** ϕ(x) **любая другая дважды гладкая функция, прини-мающая в этих узлах те же самые значения, то** E(ϕ)≥ E(S)**, причём неравенство строго для** ϕ6=S**. Сформулированное свойство называют**

**вариационным свойством естественных сплайнов.**

1. Нелинейные методы интерполяции

**Нелинейными называют методы интерполяции, в которых класс интер-полирующих функций** G **не является линейным векторным простран-ством. Важнейший частный случай нелинейных методов интерполяции это интерполяция с помощью рациональных функций вида**

|  |  |  |  |
| --- | --- | --- | --- |
| y = y(x) = | a***0*** + a***1***x + a***2***x***2*** + . . . | **(2.32)** |  |
| b***0*** + b***1***x + b***2***x***2*** + . . . |  |

**Итак, пусть в узлах** x***0*,** x***1*, . . . ,** x**n заданы значения функции** y***0*,** y***1*,**

**. . . ,** y**n. Нам нужно найти рациональную дробь вида (2.32), такую что**

y**i** = y(x**i**)**,** i = 0, 1, . . . , n**.**

**Поскольку дробь не меняется от умножения числителя и знамена-теля на одно и то же ненулевое число, то для какого-нибудь одного из**

**60** 2. Численные методы анализа

**коэффицентов** a**i или** b**i может быть выбрано произвольное наперёд за-данное значение. Кроме того, параметры** a**i и** b**i должны удовлетворять** n + 1 **условиям интерполяции в узлах, так что всего этих параметров** **потребуется** (n+ 2) **штук.**

|  |  |
| --- | --- |
| **Из формулы (2.32) следует, что** |  |
| a***0*** − b***0***y + a***1***x − b***1***xy + a***2***x***2*** − b***2***x***2***y + . . . = 0. | **(2.33)** |

**Коль скоро при** x=x**i должно быть** y=y**i,** i= 0,1, . . . , n**, то получаем** (n + 1) **равенств**

|  |  |
| --- | --- |
| a***0*** − b***0***y**i** + a***1***x**i** − b***1***x**i**y**i** + a***2***x**i*2*** − b***2***x**i*2***y**i** + . . . = 0, | **(2.34)** |

i = 0, 1, . . . , n**. Соотношения (2.33)–(2.34) можно трактовать, как усло-вие линейной зависимости с коэффициентами** a***0*,** −b***0*,** a***1*,** −b***1*, . . . для вектор-столбцов**

1

1

**...** ,

1

1

|  |  |  |
| --- | --- | --- |
| y***0*** |  |  |
| y**...*1*** | | , |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

y**n** y

x***0***

x***1***

**...** ,

x**n**

x

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| x***1***y***1*** | | |  | x***12*** | | |  | x***12***y***1*** | | |  |  |
|  | x***0***y***0*** |  |  |  | x***02*** |  |  |  | x***02***y***0*** |  |  |  |
|  | **...** |  | , |  | **...** |  | , |  | **...** |  | , . . . |  |
|  | x**n**y**n** |  |  |  | ***2*** |  |  |  | ***2*** |  |  |  |
|  |  |  |  | x**n** | |  |  |  | x**n**y**n** |  |  |  |
|  | xy |  |  |  | ***2*** |  |  |  | ***2*** |  |  |  |
|  |  |  |  |  | x |  |  |  | x y |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **размера** (n+ 2)**. Как следствие, определитель** | | | | | | | |  |  |  |  |
|  |  | 1 | y***1*** | x***1*** | x***1***y***1*** | x***12*** | x***12***y***1*** | . . . | |  |
|  |  | 1 | y***0*** | x***0*** | x***0***y***0*** | x***02*** | x***02***y***0*** | . . . | |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  | 1 | y***2*** | x***2*** | x***2***y***2*** | ***2*** | ***2*** | . . . | |  |  |
|  |  | x***2*** | x***2***y***2*** |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | **..** | **..** | **..** | **..** | **..** | **..** | **. .** | **.** |  |  |
|  |  | **. .** | | **.** | **.** | **.** | **.** |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | 1 | y**n** |  |  | ***2*** | ***2*** |  |  |  |  |
|  |  | x**n** x**n**y**n** x**n** | | | x**n**y**n** . . . | | |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | 1 | y | x | xy | ***2*** | ***2*** |  |  |  |  |
|  |  | x | x y . . . | | |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  | × | (n + 2) | |  |  |  |  |  |  |  |  |
| **размера** (n+ 2) |  | **должен быть равен нулю. Разлагая его по** | | | | | | |  |

**последней строке и разрешая полученное равенство нулю относитель-но** y**, мы действительно получим выражение для** y **в виде отношения двух многочленов от** x**. Реализация этого рецепта требует нахождения значений определителя числовых** (n+ 1)×(n+ 1)**-матриц, и далее в §3.6 мы рассмотрим соответствующие методы. Отметим, что в систе-мах компьютерной математики Scilab и** Matlab **для этого существует готовая встроенная функция** det**.**

|  |  |
| --- | --- |
| 2.4. Численное дифференцирование | **61** |

1. Численное дифференцирование

**Дифференцированием называется, как известно, процесс нахождения производной от заданной функции или же её численного значения в за-данной точке. Необходимость выполнения дифференцирования возни-кает весьма часто и вызвано огромным распространением этой опера-ции в современной математике и её приложениях. Производная бывает нужна и сама по себе, как мгновенная скорость тех или иных процессов,**

* **как вспомогательное средство для построения более сложных проце-дур, например, в методе Ньютона для численного решения уравнений**
* **систем уравнений (см. §§4.3г и 4.4б).**
  + **настоящее время наиболее распространены три следующих спо-соба вычисления производных:**
* **символьное (аналитическое) дифференцирование,**
* **численное дифференцирование,**
* **автоматическое (алгоритмическое) дифференцирование.**

**Символьным (аналитическим) дифференцированием называют про-цесс построения по функции, задаваемой каким-то выражением, про-изводной функции, основываясь на известных из элементарного ана-лиза правилах дифференцирования составных функций (суммы, раз-ности, произведения, частного, композиции, обратной функции и т. п.) и известных производных для простейших функций. На аналогичных принципах основывается автоматическое (алгоритмическое) дифферен-цирование, но при этом оперируют не выражениями для производных, а их численными значениями при данных значениях аргументов функ-ции. При этом символьное (аналитическое) дифференцирование и ав-томатическое (алгоритмическое) дифференцирование требуют знания выражения для функции или хотя бы компьютерной программы для её вычисления.**

**Численным дифференцированием называется процесс нахождения значения производной от функции, использующий значения этой функ-ции в некотором наборе точек её области определения. Таким образом, если функция задана таблично, т. е. лишь на конечном множестве зна-чений аргумента, либо процедура определения значений этой функ-ции является ¾чёрным ящиком¿ с неизвестной структурой, то альтер-натив численному дифференцированию нет. В частности, иногда мы**

**62** 2. Численные методы анализа

**вынуждены представлять в виде ¾чёрного ящика¿ вычисление значе-ний функции, аналитическое выражение для которой существует, но является слишком сложным для дифференцирования первыми двумя способами.**

**В основе численного дифференцирования лежат различные идеи. Самая первая состоит в том, чтобы доопределить (восстановить) таб-лично заданную функцию до функции непрерывного аргумента, к ко-торой уже применима обычная операция дифференцирования. Теория интерполирования, которой посвящены предшествующие параграфы, может оказаться в высшей степени полезной при реализации такого подхода. В частности, таблично заданную функцию можно заменить её интерполяционным полиномом, и его производные считать производ-ными рассматриваемой функции. Для этой процедуры годится также интерполяция сплайнами или какими-либо другими функциями.**

**2.4а** **Интерполяционный подход**

**Итак, пусть задан набор узлов** x***0*,** x***1*, . . . ,** x**n** ∈[a, b]**, т. е. сетка с шагом** h**i** = x**i** − x**i**−***1*,** i = 1, 2, . . . , n**. Кроме того, заданы значения функции** f***0*,** f***1*, . . . ,** f**n, такие что** f**i** = f (x**i**)**,** i = 0, 1, . . . , n**.**

**Рассмотрим простейший случай численного дифференцирования, когда применяется интерполяционный полином первой степени, кото-рый мы строим по двум соседним узлам сетки, т. е. по** x**i**−***1* и** x**i,** i=

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1, 2, . . . , n**:** |  |  |  |  |  |  |  |
| P***1*,i**(x) = | x − x**i** | f**i**−***1*** + | | x − x**i**−***1*** | f**i** | |  |
| x**i**−***1*** − x**i** |  |
|  |  |  | x**i** − x**i**−***1*** | | |  |
| = | f**i** − f**i**−***1*** | x + | f**i**−***1***x**i** − f**i**x**i**−***1*** | | | , |  |
|  | x**i** − x**i**−***1*** |  |  | x**i** − x**i**−***1*** | | |  |

**где у интерполяционного полинома добавлен дополнительный индекс ¾**i**¿, указывающий на ту пару узлов, по которым он построен. Поэтому производная равна**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| P***1***0**,i**(x) = | f**i** − f**i**−***1*** | = | f**i** − f**i**−***1*** | . |
|  | x**i** − x**i**−***1*** | | h**i** | |

**Это значение можно взять за приближение к производной от рассмат-риваемой функции на интервале** ]x**i**−***1***, x**i**[ **,** i= 1,2, . . . , n**.**

|  |  |
| --- | --- |
| 2.4. Численное дифференцирование | **63** |

**Во внутренних узлах сетки** x***1*,** x***2*, . . . ,** x**n**−***1*, где встречаются два подинтервала, производную можно брать по любой из возможных формул**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| f 0(x ) | ≈ | f |  |  | := | f**i** − f**i**−***1*** | | **разделённая разность назад,** |  |
| **x,i** | |  |
| **i** |  |  | x**i** − x**i**−***1*** | |  |  |
| f 0(x ) | ≈ | f |  |  | := | f**i*+1*** | − f**i** | **разделённая разность вперёд.** |  |
|  |  |  |  |
| **i** |  | **x,i** | |  | x**i*+1*** | − x**i** |  |  |

**Обе они примерно равнозначны и выбор конкретной из них может быть делом соглашения или удобства.**

**Построим теперь интерполяционные полиномы Лагранжа второй**

**степени по трём соседним точкам сетки** x**i**−***1*,** x**i,** x**i*+1*,** i= 1,2, . . . , n−1**. Имеем**

P***2*,i**(x) =

=

**Поэтому**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| (x − x**i**)(x − x**i*+1***) | | f | **i**−***1*** | + | (x − x**i**−***1***)(x − x**i*+1***) | | f | **i** |  |
|  |
| (x**i**−***1*** − x**i**)(x**i**−***1*** − x**i*+1***) | |  |  | (x**i** − x**i**−***1***)(x**i** − x**i*+1***) | |  |  |
| + | (x − x**i**−***1***)(x − x**i**) | | | | | f**i*+1*** |  |  |  |
|  | (x**i*+1*** − x**i**−***1***)(x**i*+1*** − x**i**−***1***) | | | | |  |  |  |  |

x***2*** − (x**i** + x**i*+1***)x + x**i**x**i*+1*** (x**i**−***1*** − x**i**)(x**i**−***1*** − x**i*+1***) f**i**−***1***

* x***2*** − (x**i**−***1*** + x**i*+1***)x + x**i**−***1***x**i*+1*** f**i**

(x**i** − x**i**−***1***)(x**i** − x**i*+1***)

* x***2*** − (x**i**−***1*** + x**i**)x + x**i**−***1***x**i** f**i*+1***.

(x**i*+1*** − x**i**−***1***)(x**i*+1*** − x**i**)

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| P 0 | (x) = | 2x − (x**i** + x**i*+1***) | | f | **i**−***1*** | + | 2x − (x**i**−***1*** + x**i*+1***) | | f | **i** |  |
|  |
| ***2*,i** |  | (x**i**−***1*** − x**i**)(x**i**−***1*** − x**i*+1***) | |  |  | (x**i** − x**i**−***1***)(x**i** − x**i*+1***) | |  |  |
|  |  | + | 2x − (x**i**−***1*** + x**i**) | | | | | f**i*+1***. |  |  |  |
|  |  |  | (x**i*+1*** − x**i**−***1***)(x**i*+1*** − x**i**) | | | | |  |  |  |  |

**Воспользуемся теперь обозначениями** h**i** :=x**i** −x**i**−***1*,** h**i*+1*** :=x**i*+1*** −x**i. Тогда** h**i** +h**i*+1*** =x**i*+1*** −x**i**−***1*, а результат предшествующих выкладок**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **64** |  |  |  |  |  |  |  |  |  | 2. Численные методы анализа | | | | |  |
| **может быть записан в виде** | | | | | | | |  |  |  |  |  |  |  |  |
| f 0(x) | ≈ | P | 0 |  | (x) = | 2x − x**i** − x**i*+1*** | | | | f | **i**−***1*** |  |  |  |  |
|  | ***2*,i** | | |  | h**i**(h**i** + h**i*+1***) | | | |  |  |  | **(2.35)** |  |
|  |  |  |  | 2x − x**i**−***1*** − x**i*+1*** | | |  |  |  | 2x − x**i**−***1*** − x**i** | |  |  |  |
|  |  | − |  | f | **i** | + | f | **i*+1*** | . |  |
|  |  |  |  | h**i**h**i*+1*** | | |  | h**i*+1***(h**i** + h**i*+1***) | |  |  |  |

**Формула (2.35) может применяться для вычисления производных в случае общей неравномерной сетки и для произвольного аргумен-та. Предположим теперь для простоты, что сетка равномерна, т. е.** h**i** = h = const**,** i = 1, 2, . . . , n**. Кроме того, для таблично заданной** **функции наиболее интересны производные в тех же точках, где задана сама функция, т. е. в узлах** x***0*,** x***1*, . . . ,** x**n. В точке** x=x**i из (2.35) получаем для первой производной формулу**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| f 0(x ) | ≈ | f | = | f**i*+1*** − f**i**−***1*** | , | **(2.36)** |  |
| **i** | ***˚*x,i** |  | 2h |  |  |  |

**называемую формулой центральной разности. Подставляя в (2.35) ар-гумент** x=x**i**−***1* и сдвигая в получающемся результате индекс на** +1**, получим**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| f 0(x ) | ≈ | −3f**i** + 4f**i*+1*** − f**i*+2*** | . |  |
| **i** | 2h | |  |

**Подставляя в (2.35) аргумент** x=x**i*+1* и сдвигая в получающемся ре-зультате индекс на** (−1)**, получим**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| f 0(x ) | ≈ | f**i**−***2*** − 4f**i**−***1*** + 3f**i** | . |  |
|  |
| **i** | 2h | |  |

**Займёмся теперь выводом формул для второй производной. Исполь-зуя интерполяционный полином второй степени, можно найти:**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2 | |  | 2 | | 2 | |  |  |
| P***2***00**,i**(x) = |  | f**i**−***1*** | − |  | f**i** + |  | f**i*+1***. |  |
| h**i**(h**i** + h**i*+1***) | h**i**h**i*+1*** | h**i*+1***(h**i** + h**i*+1***) |  |

**В частности, на равномерной сетке с** h**i** =h= const**,** i= 1,2, . . . , n

**имеем**

P 00 (x) = f**i**−***1*** − 2f**i** + f**i*+1*** .

***2*,i** h***2***

**Отметим, что полученные выражения для второй производной не за-висят от аргумента** x**.**

|  |  |
| --- | --- |
| 2.4. Численное дифференцирование | **65** |

**Несмотря на то, что проведённые выше рассуждения основывались на применении интерполяционного полинома Лагранжа, для взятия производных произвольных порядков на сетке общего вида удобнее ис-пользовать интерполяционный полином Ньютона, в котором члены яв-ляются полиномами возрастающих степеней.**

**Выпишем ещё без вывода формулы численного дифференцирова-ния, полученные по четырём точкам, т. е. с применением интерполяци-онного полинома третьей степени: для первой производной**

f 0(x**i**) ≈ 61h −11f**i** + 18f**i*+1*** − 9f**i*+2*** + 2f**i*+3*** , f 0(x**i**) ≈ 61h −2f**i**−***1*** − 3f**i** + 6f**i*+1*** − f**i*+2*** ,

f 0(x**i**) ≈ 61h f**i**−***2*** − 6f**i**−***1*** + 3f**i** + 2f**i*+1*** ,

f 0(x**i**) ≈ 61h −2f**i**−***3*** + 9f**i**−***2*** − 18f**i**−***1*** + 11f**i** ,

**для второй производной**

f 00(x**i**) ≈ h1***2*** 2f**i** − 5f**i*+1*** + 4f**i*+2*** − f**i*+3*** , f 00(x**i**) ≈ h1***2*** f**i**−***1*** − 2f**i** + f**i*+1*** .

* **последней формуле один из узлов никак не используется.**
  + **связи с численным дифференцированием, а также во многих дру-гих вопросах вычислительной математики чрезвычайно полезно поня-тие шаблона формулы, под которым мы будем понимать совокупность охватываемых этой формулой узлов сетки. Более точно, шаблон фор-мулы численного дифференцирования это множество узлов сетки, входящих в правую часть этой формулы. Например, шаблоном форму-лы для вычисления второй производной на равномерной сетке**

f 00(x**i**) ≈ f**i**−***1*** − 2f**i** + f**i*+1*** h***2***

**являются три точки** x**i**−***1*,** x**i,** x**i*+1*** **(см. Рис. 2.9).**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **66** |  |  |  | 2. Численные методы анализа | | |
|  | x**i**−***2*** x**i**−***1*** | | x**i** | x**i*+1*** x**i*+2*** | | |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |



**Рис. 2.9. Шаблон формулы второй разностной производной.**

**2.4б** **Оценка погрешности численного дифференцирования**

**Пусть для численного нахождения** k**-ой производной функций рассмат-ривается формула численного дифференцирования** Φ**, имеющая шаб-лон Ш и вовлекающая значения функции** f **в узлах этого шаблона. Если** f(x) **дифференцируемая необходимое число раз функция, та-кая что** f**i** =f(x**i**) **для всех узлов** x**i** ∈ **Ш, то какова может быть погрешность вычисления** f ***(*k*)***(x) **по формуле** Φ**? Вопрос этот можно адресовать как к целому интервалу значений аргумента, так и локаль-но, только к той точке** x**i, которая служит аргументом левой части формулы численного дифференцирования.**

**Ранее для погрешности интерполирования функций уже были по-лучены выражения (2.13) и (2.14), и потому заманчивой идеей является их прямое дифференцирование с целью нахождения желаемой погреш-ности формул, разработанных в предшествующем пункте. Этот путь оказывается очень непростым, так как применение, к примеру, выра-жения (2.14) требует достаточной гладкости функции** ξ(x)**, о которой мы можем сказать немногое. Даже если эта гладкость у** ξ(x) **имеется, полученные оценки будут содержать производные** ξ0(x) **и пр., о которых мы знаем ещё меньше. Наконец, шаблон некоторых формул численного дифференцирования содержит меньше точек, чем это необходимо для построения интерполяционных полиномов нужной степени. Такова, к примеру, формула ¾центральной разности¿ для первой производной.**

**Явные выражения для остаточного члена формул численного диф-ференцирования можно найти, например, в [20, 44], и оно получается методом, напоминающим вывод формулы для погрешности алгебраи-ческого интерполирования.**

**Рассмотрим ниже подробно более простой и достаточно универсаль-ный способ оценивания погрешностей, основанный на разложениях по формуле Тейлора. Суть этого способа заключается, во-первых, в вы-писывании по формуле Тейлора разложений для функций, входящих в**

|  |  |
| --- | --- |
| 2.4. Численное дифференцирование | **67** |

**правую часть формулы численного дифференцирования, и, во-вторых, в аккуратном учёте членов этих разложений с целью получения, по-возможности, наиболее точного выражения для ошибки.**

**Поясним эту методику на примере оценки погрешности для форму-лы ¾центральной разности¿ (2.36)**

f***˚*x,i** = f**i*+1*** − f**i**−***1*** .

2h

**Предположим, что имеется функция** f∈C***3***[x**i**−***1***, x**x*+1***]**. Подставляя её в (2.36) и разлагая относительно точки** x**i по формуле Тейлора с оста-точным членом в форме Лагранжа вплоть до членов второго порядка, получим**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| f***˚*x,i** | = 2h | | f (x**i**) + hf 0(x**i**) + | | | | |  | 2 f 00(x**i**) + | | | | |  | 6 f 000(ξ***+***) | | | |  |  |
|  |  | 1 |  |  |  |  |  | h***2*** | |  |  |  |  |  | h***3*** |  |  |  |  |  |
|  |  |  |  |  | − | f (x**i**) − hf 0 | | | | (x**i**) + 2 | | | | f 00 | | (x**i**) − 6 f 000(ξ−) | | | **!** |  |
|  |  |  |  |  |  | |  |  |  |  |  | h***2*** | |  |  |  | h***3*** | |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  | h***2*** | |  | h***2*** |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | = f 0(x**i**) + | | |  | f 000 | (ξ***+***) + |  |  | f 000(ξ−), | | | | |  |  |  |  |  |  |  |
|  | 12 | 12 | |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

**где** ξ***+* и** ξ− **некоторые точки из открытого интервала** ]x**i**−***1***, x**i*+1***[**. Поэтому**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  | h***2*** | |  |  |  |  |  |  | αh***2*** | |  |
|  | f***˚*x,i** − f 0 | | | (x**i**) = |  |  | f 000(ξ***+***) + f 000(ξ−) | | | | | = |  | , |  |
| **где** α= ***2*** | 12 | 6 |  |
| (f 000(ξ |  | ) + f 000(ξ−) **. В целом** | | | | |  | M |  |  |  |  |  |  |
| ***1*** |  | ***+*** |  |  |  |  |  | **справедлива оценка** | | | | | | |  |
|  |  |  |  | f***˚*x,i** − f 0 | | |  |  |
|  |  |  |  | (x**i**) ≤ | | 6 | ***3*** h***2***, | |  |  |  |  |
|  |  |  |  |  | | |  |  |  |  |  |  |  |  |  |

**в которой** M***3*** = max**ξ** |f000(ξ)| **для** ξ **из рассматриваемого множества значений аргумента. То есть, на функциях из** C***3* погрешность вычис-ления второй производной по формуле ¾центральной разности¿ равна** O(h***2***)**, если шаг сетки равен** h**.**

Определение 2.4.1 **Станем говорить, что приближённая формула** **(численного дифференцирования, интегрирования и т. п.) имеет** p**-ый**

порядок точности **(или** порядок аппроксимации**), если на равномерной** **сетке её погрешность является величиной** O(h**p**)**.**

**68** 2. Численные методы анализа

**Нередко понятие порядка точности распространяют и на неравно-мерные сетки, в которых шаг** h**i меняется от узла к узлу. Тогда роль величины** h **играет какой-нибудь ¾характерный размер¿, описывающий данную сетку, например,** h= max**i** h**i. Порядок точности важная ко-личественная мера погрешности формулы или метода, и при прочих равных условиях более предпочтительной является формула (метод) более высокого порядка точности. Но следует чётко осознавать, что по-рядок точности имеет асимптотический характер и отражает поведение погрешности при стремлении шага сетки к нулю. Если этого стремле-ния нет и шаг сетки ¾достаточно велик¿, то вполне возможны ситуации, когда метод меньшего порядка точности даёт лучшие результаты.**

**Из выкладок, проведённых для определения погрешности форму-лы ¾центральной разности¿, хорошо видна особенность метода разло-жений по формуле Тейлора: его локальный характер, вытекающий из свойств самой формулы Тейлора. Наши построения оказываются ¾при-вязанными¿ к определённому узлу (или узлам) сетки, относительно ко-торого и следует строить все разложения, чтобы обеспечить взаимные уничтожения их ненужных членов. Как следствие, в этом специальном узле (узлах) мы можем быстро оценить погрешность. Но за пределами этого узла (узлов), в частности, между узлами сетки всё гораздо слож-нее и не так красиво, поскольку взаимные уничтожения членов могут уже не происходить.**

**Какой порядок точности имеют другие формулы численного диф-ференцирования?**

**Методом разложений по формуле Тейлора для дважды гладкой функции** f **нетрудно получить оценки**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| f | **x,i** − | f 0(x ) | | ≤ | M***2*** | h, | f |  |  | f 0(x ) |  | M***2*** | h, |  |
| | | **i** | 2 |  | | | **x,i** − | | **i** | | ≤ | 2 |  |  |

**где** M***2*** = max**ξ** |f00(ξ)|**. Таким образом, разность вперёд и разность на-зад имеют всего лишь первый порядок точности.**

**Конспективно изложим другие результаты о точности формул чис-**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2.4. Численное дифференцирование | | | | | | | | | | | | | | | | | | | | | | |  |  |  |  |  |  |  |  | **69** |  |
| **ленного дифференцирования:** | | | | | | | | | | | | | | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  | f 0(x**i**) = | | | | 1 | | −3f**i** + 4f**i*+1*** | | | | | | | | | | − f**i*+2*** + O(h***2***), | | | | | | | | | | | |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  | 2h |  |  |
|  |  |  | f 0(x**i**) = | | | | 1 | | f**i**−***2*** − 4f**i**−***1*** + 3f**i** | | | | | | | | | | | | | | + O(h***2***), | | | | |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  | 2h |  |  |  |  |  |
|  |  |  | f 0(x**i**) = | | | | 1 | | −2f**i**−***1*** − 3f**i** | | | | | | | | | | + 6f**i*+1*** − f**i*+2*** + O(h***3***), | | | | | | | | | | | |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  | 6h |  |  |
|  |  |  | f 0(x**i**) = | | | | 1 | | f**i**−***2*** − 6f**i**−***1*** + 3f**i** | | | | | | | | | | | | | | + 2f**i*+1*** + O(h***3***). | | | | | | | |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  | 6h |  |  |
| **Оценим теперь погрешность формулы для второй производной** | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  |
|  |  |  |  |  |  | f 00(x ) | | | | ≈ | f | |  |  |  |  | = | f**i**−***1*** − 2f**i** + f**i*+1*** | | | | | | | | | | . |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  | **i** |  |  | **xx,i** | | | |  |  |  |  |  |  | h***2*** | | | | |  |  |  |  |  |
| **Для неё** | | | =h***2*** | | f**i** − hf**i**0 + | | | | | |  |  | 2 f**i**00 − | | | | | | 6 f**i**000 + 24 f IV | | | | | | | | | (ξ−) − 2f**i** | | |  |  |
| f**xx,i** | | |  |  |  |  |
|  |  |  |  | 1 |  |  |  |  |  |  |  | h***2*** | | |  | |  |  | h***3*** | |  | |  | h***4*** | |  | |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  | + | f**i** + hf**i**0 + | | | | | | | | |  | 2 f**i**00 + 6 f**i**000 | | | | | | | | + 24 f IV(ξ***+***) | | | **!** |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | h***2*** | |  |  |  | h***3*** | |  |  | h***4*** |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  | = f**i**00 | | + | h***2*** | f IV(ξ−) + f IV(ξ***+***) , | | | | | | | | | | | | | | | |  |  |  |  |  |  |  |  |  |  |
|  |  |  | 24 |  |  |  |  |  |  |  |  |  |  |

**где** ξ−**,** ξ***+* некоторые точки из открытого интервала** ]x**i**−***1***, x**i*+1***[ **, и если** f∈C***4*, то справедлива оценка**

|  |  |  |  |
| --- | --- | --- | --- |
| |f 00(x**i**) − f**xx,i**| ≤ | M***4*** | h***2***, |  |
| 12 |  |

**где** M***4*** = max**ξ** |fIV(ξ)|**. Таким образом, порядок точности этой форму-лы равен 2 на функциях из** C***4*.**

**Приведём ещё без вывода результат о погрешности формулы для вычисления второй производной вблизи края сетки (таблицы):**

|  |  |  |  |
| --- | --- | --- | --- |
| f 00(x**i**) = | 1 | 2f**i** − 5f**i*+1*** + 4f**i*+2*** − f**i*+3*** + O(h***2***), |  |
|  |  |  |  |
|  | h***2*** |  |  |
| f 00(x**i**) = | 1 | f**i**−***3*** − 4f**i**−***2*** + 5f**i**−***1*** − 2f**i** + O(h***2***). |  |
|  |  |  |  |
|  | h***2*** |  |  |

**70** 2. Численные методы анализа

**2.4в** **Метод неопределённых коэффициентов**

**Метод неопределённых коэффициентов это другой подход к полу-чению формул численного дифференцирования, особенно удобный в многомерном случае, когда построение интерполяционного полинома становится непростым.**

**Станем искать приближённое выражение для производной от функ-ции в виде**

**n**

**X**

|  |  |  |
| --- | --- | --- |
| f ***(*k*)***(x) ≈ | c**i**f (x**i**), | **(2.37)** |
|  | **i*=0*** |  |

**который мотивируется тем обстоятельством, что дифференцирование любого порядка является линейной операцией. При этом коэффициен-ты** c**i постараемся подобрать так, чтобы эта формула являлась точной формулой для всех полиномов степени не выше заданной. Учитывая линейность конструируемой формулы (2.37), можно брать** f(x) **равной** 1**,** x**,** x***2*, . . . ,** x**m. Каждое такое условие порождает какое-то линейное** **соотношение для коэффициентов** c**i, а в целом мы приходим к систе-ме линейных уравнений, для разрешимости которой естественно взять число неизвестных равным числу уравнений, т. е.** m=n**.**

**Получающаяся система линейных уравнений имеет вид**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| c***0***x***0*** | | | + c***1***x***1*** | | |  |
|  | c***0*** |  | + | c***1*** |  |  |
|  | **.** |  |  | **.** |  |  |
|  | **.** |  |  | **.** |  |  |
|  | **.** |  |  | **.** |  |  |
|  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  | **k** | ***1*** |  | **k** | ***1*** |  |
|  |  |  |
|  | c***0***x***0***− | | + c***1***x***1***− | | |  |
|  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  | **k** | + |  | **k** |  |
|  | c x | | c x | |  |
|  | ***0*** | ***0*** |  | ***1*** | ***1*** |  |
|  |  |  |  |  |  |  |
|  | c***0***x***0*k*+1*** | | + c***1***x***1*k*+1*** | | |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  | **.** |  |  | **.** |  |  |
|  | **.** |  |  | **.** |  |  |
|  | **.** |  |  | **.** |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  | c***0***x | **n** | + | c***1***x | **n** |  |
|  | ***0*** | ***1*** |  |
|  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| + . . . + | | | c**n** | = |  |  | 0, |  |  |  |  |
| + . . . + c**n**x**n** | | | | = |  |  | 0, |  |  |  |  |
| **.** | **.** |  | **.** |  |  |  | **.** |  |  |  |  |
|  |  | **.** | **.** |  |  |  | **.** |  |  |  |  |
|  |  |  | **.** |  |  |  | **.** |  |  |  |  |
| + . . . + c**n**x**nk**−***1*** = | | | | |  |  | 0, |  |  | **(2.38)** |  |
| + . . . + c**n**x**nk** | | | | = |  |  | k!, | |  |  |  |
|  | | | |  |  |  |  | |  |  |  |
| + . . . + c**n**x**nk*+1*** | | | | = | (k + 1)! x, | | | | |  |  |
| **.** | **.** |  | **.** |  |  |  | **.** |  |  |  |  |
|  |  | **.** | **.** |  |  |  | **.** |  |  |  |  |
|  |  |  | **.** |  |  |  | **.** |  |  |  |  |
| + . . . + c x**n** | | | | = n(n | − | 1) | · · · | (n | − | k + 1) x**n**−**k.** |  |
|  |  |  | **n n** |  |  |  |  |  |  |  |  |

**Матрица этой системы является матрицей Вандермонда вида (2.3) и неособенна для различных узлов** x***0*,** x***1*, . . . ,** x**n. При этом система ли-нейных уравнений однозначно разрешима относительно** c***0*,** c***1*, . . . ,** c**n при любой правой части, но содержательным является лишь случай** k ≤ n**. В противном случае, если** k > n**, правая часть системы (2.38)**

|  |  |
| --- | --- |
| 2.4. Численное дифференцирование | **71** |

**оказывается нулевой, и, как следствие, система также имеет только бес-содержательное нулевое решение. Этот факт имеет интуитивно ясное объяснение: нельзя построить формулу для вычисления производной** k**-го порядка от функции, используя значения этой функции не более** **чем в** k **точках.**

**Система (2.38) с матрицей Вандермонда, аналогичная системе (2.2), в общем случае является плохобусловленной. Но на практике её реше-ние вручную или на компьютере обычно не приводит к большим ошибкам, так как в отличие от случая интерполяции порядок системы (2.38), равный порядку производной, бывает, как правило, небольшим.**

**Интересен вопрос о взаимоотношении метода неопределённых ко-эффициентов и рассмотренного ранее в §2.4а интерполяционного под-хода к численному дифференцированию. К примеру, Ш.Е. Микеладзе в книге [42] утверждает, что любая формула численного дифференци-рования, полученная методом неопределённых коэффициентов, может быть выведена также с помощью интерполяционного подхода, отка-зывая последнему в оригинальности. Но нельзя отрицать также, что метод неопределённых коэффициентов конструктивно проще и техно-логичнее в применении, и уже только это обстоятельство оправдывает его существование.**

**2.4г** **Полная вычислительная погрешность численного дифференцирования**

**Рассмотрим поведение полной погрешности численного дифференци-рования при расчётах на реальных вычислительных устройствах. Под полной погрешностью мы понимаем суммарную ошибку численного нахождения производной, вызванную как приближённым характером самого метода, так и неточностями вычислений на современных циф-ровых ЭВМ из-за неизбежных ошибок округления и т. п.**

**Предположим, к примеру, что первая производная функции вычис-ляется по формуле ¾разность вперёд¿**

f 0(x**i**) ≈ f**x,i** = f**i*+1*** − f**i** . h

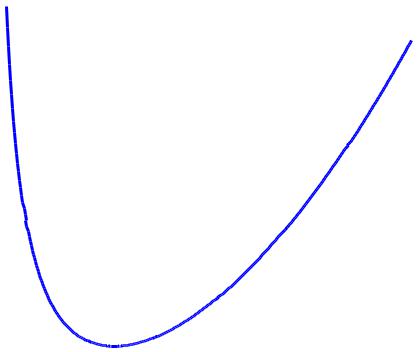
**Как мы уже знаем, её погрешность**

|f**x,i** − f 0(x**i**)| ≤ M***2***h ,

2

**72** 2. Численные методы анализа

6



|  |  |
| --- | --- |
|  | - |
| 0 | h |

**Рис. 2.10. Полная погрешность численного дифференцирования**

**где** M***2*** = max**ξ**∈***[*a,b*]*** |f00(ξ)|**. Если значения функции вычисляются с ошибками, то вместо точных** f**i и** f**i*+1* мы получаем их приближённые**

|  |  |  |  |
| --- | --- | --- | --- |
| ˜ ˜ |  |  |  |
| **значения** f**i и** f**i*+1*, такие что** |  |  |  |
| ˜ | **и** | ˜ |  |
| |f**i** − f**i**| ≤ δ | |f**i*+1*** − f**i*+1***| ≤ δ, |  |

**где через** δ **обозначена предельная абсолютная погрешность вычисле-ния значения функции. Тогда предельную полную вычислительную погрешность** E(h, δ) **нахождения первой производной функции можно оценить следующим образом**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| E(h, δ) = |  | | f˜**i*+1***h− f˜**i** − f 0(x**i**) | | | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| ≤ |  | | ˜ |  | ˜ | | |  |  |  | + | | | | f**i*+1***h− f**i** | | | |  | − f 0(x**i**) | | | |  |
|  | | f**i*+1***h− f**i** − f**i*+1***h− f**i** | | | | | | |  |  |
|  |  | |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  |  |  |  |  | |  |
|  |  | |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  |  |  |  |  | |  |
|  |  | | ˜ | − |  | h | |  | − | ˜ | |  |  | |  |  |  |  |  |  |  |  | |  |
| ≤ | | | (f**i*+1*** |  |  | f**i**) | | | + 2 | | | | | |  |  |  |  |  |  |
|  |  | |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  |  |
|  |  | |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  |  |
|  |  | |  |  | ˜ | | |  | ˜ | |  |  |  | |  |  |  |  |  | 2δ |  | M***2***h |  |  |
| ≤ |  | |f**i*+1*** − f**i*+1***| + |f**i** | | | | | | − f**i**| | | |  | + | | M***2***h | | | = | |  | + | . |  |
|  |  | | | | | |  | | |  |  | | |  | h |  |  |
|  |  |  |  |  | h | |  |  |  |  |  |  |  | 2 | |  |  |  | 2 | |  |  |

**Отметим, что эта оценка, во-первых, достижима при подходящем сочетании знаков фигурирующих в неравенствах величин, коль ско-**

|  |  |
| --- | --- |
| 2.4. Численное дифференцирование | **73** |

**ро достижимо само неравенство треугольника. Во-вторых, оценка не стремится к нулю при уменьшении шага** h**, так как первое слагаемое неограниченно увеличивается при** h→0**. В целом функция** E(h, δ) **при фиксированном** δ **имеет минимум, определяемый условием**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ∂h | = ∂h h + 2 | | | | | | = − h***2*** + | | | 2 | = 0. |
| ∂E(h, δ) |  | ∂ 2δ M***2***h | | | | |  | 2δ | | M***2*** |  |
|  |  |  |  |  |  |  |  |  |  |  |  |

**То есть, оптимальное значение шага численного дифференцирования, при котором достигается минимальная полная погрешность, равно**

**p**

h∗ = 2 δ/M***2***,

**и брать меньший шаг численного дифференцирования смысла нет. Зна-**

**чение достигающейся при этом полной погрешности есть** E(h∗, δ) =

√

2 δM***2*.**

**Совершенно аналогичная ситуация имеет место и при использова-нии других формул численного дифференцирования. Производная** k**-**

|  |  |  |  |
| --- | --- | --- | --- |
| **ого порядка определяется тогда формулой типа** | |  |  |
| f ***(*k*)***(x) = h−**k** | **X** | **(2.39)** |  |
| c**i**f (x**i**) + R**k**(f, x), |  |

**i**

**где** c**i** =O(1)**. Если эта формула имеет порядок точности** p**, то её остаточный член оценивается как** R**k**(f, x)≈c(x)h**p. Этот остаточный член определяет ¾идеальную¿ погрешность численного дифференци-рования в отсутствие ошибок вычисления функции, и он неограничен-но убывает при** h→0**.**

**Но если погрешность вычисления значений функции** f(x**i**) **равна** δ**, то в правой части (2.39) возникает ещё член, абсолютная величина** **которого оценивается сверху как**

**X**

δh−**k** |c**i**|.

**i**

**Она неограниченно возрастает при** h→0**. В целом график полной вы-числительной погрешности численного дифференцирования выглядит в этом случае примерно так, как на Рис. 2.10.**

**Практический вывод из вышесказанного состоит в том, что суще-ствует оптимальный шаг** h **численного дифференцирования, миними-зирующий полную вычислительную погрешность, и брать меньшее зна-чение шага нецелесообразно.**

**74** 2. Численные методы анализа

**Отмеченное нами обстоятельство потенциально сколь угодно боль-шого возрастания погрешности численного дифференцирования, в дей-ствительности, является отражением более глубокого факта некоррект-ности задачи дифференцирования, когда её решение не зависит непре-рывно от входных данных (см. §1.2). Если** f(x) **исходная функция, производную которой нам требуется найти, то возмущённая функция** f (x)+ **n*1*** sin(nx) **при** n → ∞ **будет очевидно сходиться к исходной, тогда** **как её производная**

f 0(x) + cos(nx)

**не сходится к производной** f0(x)**. При возмущении исходной функции слагаемым n*1*** sin(n***2***x) **производная вообще может сколь угодно сильно отличаться от производной исходной функции.**

1. Приближение в евклидовых пространствах

**2.5а** **Обсуждение постановки задачи**

* **этом параграфе мы займёмся задачей приближения функций и по-дробно рассмотрим случай, когда отклонение функций друг от друга измеряется в так называемой среднеквадратичной норме.**
  + **задаче приближения функций естественно приходят в ситуации, когда методы интерполирования по различным причинам не удовле-творяют практику. Эти причины могут носить чисто технический ха-рактер. К примеру, гладкость сплайна может оказаться недостаточной, либо его построение слишком сложным. Степень обычного интерпо-ляционного полинома может быть неприемлемо высокой для данного набора узлов интерполяции (а высокая степень это трудности при вычислении значений полинома). Но причины отказа от интерполяции**

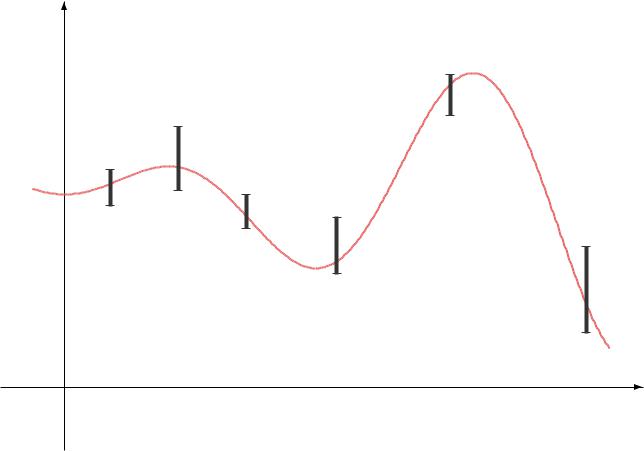
**могут иметь также принципиальный характер. В частности, это проис-ходит, если значения функции в узлах** x***0*,** x***1*, . . . ,** x**n известны неточно.**

* **этих условиях целесообразна коррекция самой постановки задачи.**

**Именно, имеет смысл отказаться от требования точного равенства** g(x**i**) **значениям функции** f**i** **в узлах** x***0*,** x***1*, . . . ,** x**n, допустив, к при-меру, принадлежности** g(x**i**)∈[f **i**, f **i**]**,** i= 0,1, . . . , n**,** f **i** ≤f **i для зна-**

**чений восстанавливаемой функции** g(x) **некоторым интервалам** [f **i**, f **i**]**. Наглядно-геометрически это означает построение функции** g(x) **из за-**

|  |  |
| --- | --- |
| 2.5. Приближение в евклидовых пространствах | **75** |
| **Рис. 2.11. Интерполяция функции, заданной с погрешностью** |  |



**данного класса** G**, которая в каждом узле сетки** x**i,** i= 0,1, . . . , n**, про-**

**ходит через некоторый ¾коридор¿** [f**i**, f**i**]**, см. Рис. 2.11.**

**Более общая постановка этой задачи предусматривает наличие неко-торой метрики (расстояния), которую мы будем обозначать через dist, и с помощью которой можно измерять отклонение вектора значений** (g(x***0***), g(x***1***), . . . , g(x**n**))> **функции** g(x) **в узлах сетки от вектора задан-ных значений** (f***0***, f***1***, . . . , f**n**)>**. Фактически, здесь dist это какое-то расстояние на пространстве** R**n*+1* всех** (n+ 1)**-мерных вещественных векторов. Соответствующая постановка задачи приближения (аппрок-симации) будет звучать тогда так:**

**Для заданных , набора узлов** x**i,** i= 0,1, . . . , n**, на интервале** [a, b] **и соответствующих им значений** f**i,** i = 0, 1, . . . , n**, найти** **функцию** g(x) **из класса функций** G**, такую что** dist (f, g)< **, где** f= (f***0***, f***1***, . . . , f**n**)> **и** g= (g(x***0***), g(x***1***), . . . , g(x**n**))>**.**

**При этом** g(x) **называют приближающей (аппроксимирующей) функ-цией, а сама эта новая задача является задачей приближения функции. Её важнейшей модификацией служит задача наилучшего приближе-ния, когда ищут приближающую (аппроксимирующую) функцию** g(x)**, которая доставляет минимум расстоянию** dist (f, g)**.**

**Наконец, выписанные выше формулировки являются дискретными вариантами общей задачи о приближении функции:**

dist (f, h)**.**

**76** 2. Численные методы анализа

**Для заданных** >0**, функции** f(x) **из** F **и метрики dist найти функцию** g(x) **из класса функций** G**, такую что** dist (f, g)< **.**

**Соответствующая общая формулировка задачи о наилучшем прибли-жении ставится так:**

**Для заданных функции** f(x) **из функцию** g(x) **из класса функций нижняя грань расстояний** inf**h**∈G

**ряющую** dist (f, g) = inf**h**∈G

F **и метрики dist найти** G**, на которой достигается** dist (f, h)**, т. е. удовлетво-**

**Решение** g **этой задачи обычно называется наилучшим приближени-ем для** f **. Отметим, что в каждом конкретном случае существование элемента наилучшего приближения требует отдельного исследования.**

**Формула Тейлора хороший и весьма важный в практическом от-ношении инструмент локального приближения функций. Но для боль-ших интервалов она не очень хороша.**

**До сих пор ничего не было сказано о выборе классов функций** F **и** G**, и в наших формулировках на их месте могут быть весьма произвольные множества. Но чаще всего** F **и** G **являются линейными пространствами, на которых определена некоторая норма** k·k**. Именно в ней измеряют отклонение функций (непрерывного или дискретного аргумента) друг от друга, так что**

dist (f, g) = kf − gk.

**Соответственно, в задаче наилучшего приближения функции** f **ищется такой элемент** g∈G**, на котором достигается** inf**h**∈Gkf−hk**.**

**Рассмотренные выше постановки задач дают начало большим и важным математическим теориям, в совокупности образующим тео-рию приближения функций (или теорию аппроксимации). Её ветвью является, в частности, теория равномерного приближения, когда от-**

**клонение функций оценивается в норме** kfk= max**x**∈***[*a,b*]*** |f(x)|**. Вы-бор различных норм (т. е. различных мер отклонения функций друг от**

**друга) и различных функциональных классов обуславливает огромное разнообразие задач теории приближения.**

**2.5б** **Метод наименьших квадратов**

**Рассмотрим подробно частный, но очень важный случай задачи о наи-лучшем приближении, в котором**

|  |  |
| --- | --- |
| 2.5. Приближение в евклидовых пространствах | **77** |

• **линейное нормированное пространство функций** F **таково, что на**

**нём задано скалярное произведение** h·,·i**, и норма в** F **определя-p**

**ется с его помощью как** kfk= hf, f i**,**

* **функциональный класс** G⊆F**, из которого выбирается искомый элемент наилучшего приближения, является конечномерным под-пространством в** F**.**

**Метод наименьших квадратов это общее название целого семей-ства идейно близких методов построения приближений, которые ос-нованы на минимизации суммы квадратов отклонений компонент ис-ходного вектора от приближающего. В случае, когда рассматривается приближение функций (или вообще элементов) каких-то абстрактных пространств естественным обобщением минимизации суммы квадратов является нахождение минимума нормы, порождённой скалярным про-изведением.**

**Напомним, что конечномерные линейные векторные пространства со скалярным произведением называются евклидовыми пространства-ми. Бесконечномерные линейные векторные пространства с выписан-ной нами нормой называются гильбертовыми пространствами при дополнительном условии полноты, т. е. существования в них предела всякой фундаментальной последовательности.**

**Приближение функций в норме, порождённой скалярным произ-ведением, часто называют среднеквадратичным приближением, так как в конечномерной ситуации, когда имеем** f= (f***0***, f***1***, . . . , f**n**)> **и** g = (g***0***, g***1***, . . . , g**n**)>**, обычно**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1 | | **n** |  |  |
| **X** | **(2.40)** |  |
|  |  |  |
| hf, gi = |  | **i*=0*** %**i**f**i**g**i**, |  |
| n + 1 |  |

**для некоторого положительного весового вектора** %= (%***0***, %***1***, . . . , %**n**)>**,** %**i** > 0**. То есть, в соответствующей норме отклонение одной функции** **от другой есть**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 1 | **n** | **!** | ***1*/*2*** |  |
|  |  |  |
| kf − gk = |  | , |  |
| n + 1 **i*=0*** %**i**(f**i** − g**i**)***2*** | |  |
|  |  | **X** |  |  |  |

**усреднение квадратов разностей компонент с какими-то весовыми множителями** %**i,** i= 0,1, . . . , n**. В частности, если известна информа-ция о точности задания отдельных значений функции** f**i, то веса** %**i**

**78** 2. Численные методы анализа

**можно назначать так, чтобы отразить величину этой точности, сопо-ставляя значениям** f**i с б´ольшей точностью б´ольший вес.**

**Нормирующий множитель n*+11* при сумме в (2.40) удобно взять для того, чтобы с ростом размерности** n **(при росте количества наблюде-ний, измельчении сетки и т. п.) ограничить рост величины скалярного произведения и обеспечить соизмеримость результатов при различных** n**. В общем случае, когда** f **и** g **функции непрерывного аргумента,** **обычно полагают**

**Z b**

hf, gi = %(x)f (x)g(x) dx, **(2.41)**

**a**

**для некоторой весовой функции** %(x)>0**, и при этом**

**Zb** **!*1*/*2***

***2***

kf − gk = %(x) f (x) − g(x) dx .

**a**

**Линейное пространство всех функций, квадрат которых интегрируем на заданном интервале** [a, b]**, со скалярным произведением, задаваемым посредством (2.41), называют пространством** L***2***[a, b]**. В курсах функ-ционального анализа показывается, что это гильбертово пространство. По этой причине оно очень популярно в самых различных математи-ческих дисциплинах, от теории уравнений в частных производных до статистики.**

**Итак, в ситуации, описанной в начале пункта, ищем приближение** g **для функции** f **в виде**

**m**

**X**

g(x) = c**j** ϕ**j** (x),

**j*=1***

**где** {ϕ**j** (x)}**mj*=1* базис** m**-мерного линейного подпространства** G⊆F**. Если через** Φ **обозначить квадрат нормы отклонения** f **от** g**, то имеем**

Φ = kf (x) − g(x)k***2*** = hf − g, f − gi

= hf, f i − 2hf, gi + hg, gi

|  |  |  |  |
| --- | --- | --- | --- |
| **m** | **m m** |  |  |
| **X** | **X X** | **(2.42)** |  |
| = hf, f i − 2c**j** hf, ϕ**j** i + | c**j** c**k**hϕ**j** , ϕ**k**i. |  |  |
| **j*=1*** | **j*=1* k*=1*** |  |  |

|  |  |
| --- | --- |
| 2.5. Приближение в евклидовых пространствах | **79** |

**Как видим,** Φ **есть квадратичная форма от аргументов** c***1*,** c***2*, . . . ,** c**m плюс ещё некоторые линейные члены относительно** c**j и постоянное слагаемое** hf, fi**. Так как для всех значений аргументов функция** Φ

**принимает только неотрицательные значения, то ясно, что она должна достигать своего минимума.**

**Действительно, после приведения к ¾главным осям¿ квадратичная форма в составе** Φ **обязательно должна получить вид суммы квад-ратов, так как иначе вся** Φ **была бы неограниченной снизу. Но сум-ма квадратов неограниченно возрастает при росте нормы аргумента** c = (c***0***, c***1***, . . . , c**m**) **и растёт быстрее линейных членов. Следовательно,** **при достаточно больших значениях** kck **значение** Φ **также сколь угодно велико. Более точно, для любого достаточно большого** K **существует такое** C**, что** Φ(x)> K **при** kxk> C**. Получается, что минимум непре-рывной функции** Φ **находится где-то в компактном шаре** kck ≤C**, т. е. действительно достигается функцией.**

**Для определения минимума функции** Φ **продифференцируем её по** c**j** **,** j = 1, 2, . . . , m**, и приравняем полученные производные к нулю:**

|  |  |  |
| --- | --- | --- |
| ∂Φ | **m** |  |
| **X** |  |
|  |  |
|  | = −2hf, ϕ**j** i + 2c**k**hϕ**j** , ϕ**k**i = 0. |  |
| ∂c**j** |  |
|  | **k*=1*** |  |

**Множитель** 2 **при сумме** c**k**hϕ**j** , ϕ**k**i **появляется оттого, что в двойной сумме из выражения (2.42) слагаемое с** c**j возникает дважды, один раз**

* **коэффициентом** hϕ**j** , ϕ**k**i**, а другой раз с коэффициентом** hϕ**k**, ϕ**j** i**.**
  + **целом, получаем систему линейных уравнений**

**m**

**X**

hϕ**j** , ϕ**k**i c**k** = hf, ϕ**j** i, j = 1, 2, . . . , m, **(2.43)**

**k*=1***

**с матрицей коэффициентов**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | hϕ***2*** | | , ϕ***1***i | | | hϕ***2*** | , ϕ***2***i | | | . . . | hϕ***2***, ϕ**n** | |  | i | | |  |
|  |  | ϕ***1*** | , ϕ***1*** |  |  | ϕ***1*** | , ϕ***2*** |  |  | . . . | ϕ***1*** | , ϕ**m** | |  |  |  |  |
| (ϕ***1***, ϕ***2***, . . . , ϕ**m**) = | h | | **...** | i | | h | **...** | i | | **. . .** | h | **...** | i | | | | , |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | ϕ**m**, ϕ***1*** | | |  | ϕ**m**, ϕ***2*** | | |  | . . . ϕ**m**, ϕ**m** | | | | |  |  |  |
|  |  | h |  |  | i | h |  |  | i |  | h |  |  |  | i |  |  |

**которая называется, как известно, матрицей Грама системы векторов** ϕ***1*,** ϕ***2*, . . . ,** ϕ**m. Из курса линейной алгебры и аналитической геометрии**

**80** 2. Численные методы анализа

**читателю должно быть известно, что матрица Грама это симметрич-ная матрица, неособенная тогда и только тогда, когда векторы** ϕ***1*,** ϕ***2*,**

**. . . ,** ϕ**m линейно независимы. При выполнении этого условия матри-ца Грама является ещё и положительно определённой. Таким образом, решение задачи наилучшего среднеквадратичного приближения суще-ствует и единственно в случае, когда** ϕ***1*,** ϕ***2*, . . . ,** ϕ**m образуют базис в подпространстве** G**.**

**Обратимся к практическим аспектам реализации метода наимень-ших квадратов. Каковы свойства системы уравнений (2.43) в смысле устойчивости её решения к возмущениям в данных и погрешностям вычислений на ЭЦВМ?**

**Наиболее простой вид матрица Грама имеет в случае, когда базис-ные функции** ϕ**j ортогональны друг другу, т. е. когда** hϕ**j** , ϕ**k**i= 0 **при** j 6= k**. При этом соответствующая система линейных уравнений (2.43)** **становится диагональной и решается тривиально**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| c**k** = | hf, ϕ**k** i | , | k = 1, 2, . . . , m. | **(2.44)** |  |
|  |
|  | hϕ**k** , ϕ**k**i | |  |  |  |
| **Соответствующее наилучшее приближение имеет вид суммы** | | | |  |  |
|  |  |  | **m** |  |  |
|  |  |  | **X** |  |  |
|  | g = | | c**k**ϕ**k** |  |  |

**k*=1***

**и называется (конечным) рядом Фурье для** f **по ортогональной системе векторов** {ϕ**j** (x)}**mj*=1*, а коэффициенты (2.44) называют при этом коэф-фициентами Фурье разложения** f **.**

**Кроме того, в случае ортогонального и близкому к ортогональному базиса** {ϕ**j** (x)}**mj*=1* решение системы (2.43) устойчиво к возмущениям в правой части и неизбежным погрешностям вычислений. Но в общем случае базис линейного подпространства** G **может быть неортогональ-ным, и тогда свойства системы уравнений (2.43) могут быть очень пло-хими в том смысле, что её решение будет чрезвычайно чувствительным к возмущениям и погрешностям.**

Пример 2.5.1 **Рассмотрим задачу о среднеквадратичном приближе-нии непрерывных функций на интервале** [0,1] **полиномами фиксиро-ванной степени** m**, беря нормой**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| kf k = | **Z *1*** | f (x) |  | ***2*** dx. |  |
|  |  |  |  |  |  |

***0***

|  |  |
| --- | --- |
| 2.5. Приближение в евклидовых пространствах | **81** |

**Если в качестве базиса в линейном пространстве полиномов мы возь-мём обычные степени**

1, x, x***2***, . . . , x**m**,

**то на месте** (i, j) **в соответствующей матрице Грама размера** (m+1)×(m+1) **будет стоять элемент**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Z*0*** | x**i**−***1***x**j**−***1*** dx = i + j − 1 | | ***0*** | = | i + j 1 ,i, j = 1, 2, . . . , m + 1 | | |  |
| ***1*** |  | x**i*+*j *1*** | ***1*** |  | 1 |  |  |  |
|  |  |  | − | |  |
|  |  | − |  |  |  |
|  |  |  |  |  |  |  |  |  |

**(сдвиг показателей степени на** (−1) **вызван тем, что строки и столб-цы матрицы нумеруются, начиная с единицы). Матрица** H= (h**ij** ) **с элементами** h**ij** = 1/(i+j−1) **называется матрицей Гильберта, и она является исключительно плохо обусловленной матрицей (см. §3.2з).**

**В то же время базис из тригонометрических полиномов вида**

|  |  |
| --- | --- |
| 1, cos(2πkx), sin(2πkx), | k = 1, 2, . . . , |

**является ортогональным на** [0,1]**, т. е. в вычислительном отношении очень хорошим для построения среднеквадратичных приближений.**

**Более детальный теоретический анализ и практический опыт пока-зывают, что в методе наименьших квадратов в качестве базиса** ϕ***1*,** ϕ***2*,**

**. . . ,** ϕ**n линейного подпространства** G⊂F **имеет смысл брать системы элементов, ортогональных, возможно, по отношению к другому ска-лярному произведению, так как это служит гарантией ¾разумной ма-лости¿ внедиагональных элементов матрицы Грама и, как следствие, её не слишком плохой обусловлености.**

**Отметим, что задачу приближения функций, значения которых за-даны приближённо, часто называют (особенно в практических прило-жениях) задачей сглаживания, поскольку получаемая приближающая функция, как правило, действительно ¾сглаживает¿ выбросы данных, вызванные случайными ошибками и т.п.**

**2.5в** **Полиномы Лежандра**

**Если базис, образованный последовательными степенями переменной** 1**,** x**,** x***2*, . . . ,** x**m** **является плохим для среднеквадратичного прибли-жения функций полиномами, то каков хороший ортогональный базис?**

**82** 2. Численные методы анализа

**Для его конструктивного построения можно воспользоваться, к приме-ру, известным из курса линейной алгебры процессом ортогонализации Грама-Шмидта (см. также §3.3л). Напомним, что по данной конечной линейно независимой системе векторов** v***1*,** v***2*,. . . ,** v**n этот процесс стро-ит ортогональный базис** q***1*,** q***2*, . . . ,** q**n линейной облочки векторов** v***1*,** v***2*,. . . ,** v**n** **по следующим расчётным формулам:**

**k**−***1***

q**k** ← v**k** − **X** hv**k** , q**i**i q**i**, k = 1, 2, . . . , n. **(2.45)**

**i*=1*** hq**i**, q**i**i

**Иногда получающийся ортогональный базис дополнительно нормиру-ют.**

**Конкретный вид ортогональных полиномов, которые получатся из** 1**,** x**,** x***2*, . . . , зависит от интервала** [a, b]**, на котором решается задача** **наилучшего среднеквадратичного приближения. Но мы можем суще-ственно облегчить свою задачу, если найдём семейство ортогональных полиномов для какого-нибудь одного, канонического, интервала** [α, β]**, а затем для любого другого интервала будем пользоваться формулой линейной замены переменной** y=kx+l**. Тогда для** a=kα+l**,** b=kβ+l

**в силу равенства**

**Z b** **Z β**

f (y) dy = k f (kx + l) dy,

**a** **α**

**вытекающего из формулы замены переменных в определённом инте-грале, получающиеся полиномы будут ортогональны на** [a, b]**. В каче-стве такого канонического интервала** [α, β] **обычно берётся** [−1,1]**, и тогда формула замены переменных принимает вид**

y = ***12*** (b − a) x + ***12*** (a + b),

**так что переменная** y **пробегает интервал** [a, b]**, если** x∈[−1,1]**. Обрат-ное преобразование даётся формулой**

1 x = b − a 2y − (a + b) ,

**которая позволяет заново построить ортогональные полиномы для лю-бого интервала.**

**Полиномами Лежандра называют полиномы, которые образуют ор-тогональную систему относительно скалярного произведения (2.41) с**

|  |  |
| --- | --- |
| 2.5. Приближение в евклидовых пространствах | **83** |

**простейшим весом** %(x) = 1 **на интервале** [−1,1]**. Они были введены в**

**широкий оборот французским математиком А. Лежандром ещё в 1785 году. Из общей теории скалярного произведения в линейных простран-ствах следует, что такие полиномы существуют и единственны. Если мы не требуем их нормированности, то они оказываются определённы-ми с точностью до постоянного множителя.**

**Применяя последовательно формулы (2.45), получим**

1, x, x***2*** − ***13*** , x***3*** − ***35*** x, . . .

**(два первых полинома ортогональны с самого начала).**

**Обычно вводят полиномы Лежандра с помощью формулы Родрига**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| L**n**(x) = | 1 d**n** | x***2*** −1**n**, | n = 0, 1, 2, . . . , | **(2.46)** |  |
| 2**n**n! dx**n** |  |

**которая даёт альтернативное представление для полиномов Лежанд-ра. Очевидно, что** L**n**(x)**, определяемый этой формулой, является ал-гебраическим полиномом** n**-ой степени со старшим коэффициентом, не равным нулю, так как при** n**-кратном дифференцировании полинома** (x***2*** − 1)**n** = x***2*n** − nx***2(*n**−***1)*** + . . . **степень понижается в точности на** n**.** **Коэффициент** 1/(2**n**n!) **перед производной в (2.46) взят затем, чтобы удовлетворить условию** L**n**(1) = 1**.**

**Из (2.46) с помощью применения** n**-кратного интегрирования по частям можно вывести и свойство ортогональности полиномов** L**n**(x)**.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Действительно, обозначая** | | |  |  |  |
|  |  | ψ(x) = (x***2*** − 1)**n**, | |  |  |
| **можно заметить, что** | | |  |  |  |
|  | d**k** | |  |  |  |
| ψ***(*k*)***(x) = |  | x***2*** −1**n**= 0 | **при** x=±1, | k = 0, 1, 2, . . . , n − 1. |  |
|  | dx**k** |  |  |  |  |

**Поэтому, если** Q(x) **является** n **раз непрерывно дифференцируемой**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **84** |  |  |  |  |  |  | 2. | Численные методы анализа | | |  |
| **функцией на** [−1,1]**, то получим** | | | | |  |  |  |  |  |  |  |
| **Z**−***1*** Q(x) L**n**(x) dx = 2**n**n! | | | | **Z**−***1*** Q(x) ψ***(*n*)***(x) dx | | | | | |  |  |
| ***1*** |  | 1 | | ***1*** |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  | ***1*** |  |  | **Z**−***1*** Q0(x) ψ***(*n**−***1)*** |  |  |
| =2**n**n! |  | Q(x) ψ***(*n**−***1)***(x) | | | ***1*** − | 2**n**n! | (x) dx |  |
|  |  |  |  |  |  |  |  | ***1*** |  |  |
| 1 |  |  |  | − | |  | 1 |  |  |  |

* − 1 **Z** ***1*** Q0(x) ψ***(*n**−***1)***(x) dx 2**n**n! −***1***
* · · ·
* (−1)**n** 1 **Z** ***1*** Q***(*n*)***(x) ψ(x) dx. 2**n**n! −***1***

**Если** Q(x) **любой полином степени меньше** n**, то его** n**-ая производная** Q***(*n*)***(x) **равна тождественному нулю, а потому из полученной формулы** **следует**

**Z *1***

Q(x) L**n**(x) dx = 0.

−***1***

**В частности, это верно и в случае, когда вместо** Q(x) **берётся полином Лежандра степени, меньшей** n**, что доказывает наше утверждение.**

**С другой стороны, если** Q(x) =L**n**(x)**, то**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Q***(*n*)***(x) = | 1 d***2*n** | | | x***2*** −1**n**= | | | | 1 | (2n)! |  |
|  |  |  |  |  | | | |  |  |  |
|  | 2**n**n! dx***2*n** | | |  | | | | 2**n**n! |  |  |
| **По этой причине** |  |  |  |  |  |  |  | **Z**−***1*** ψ(x) dx | |  |
| **Z**−***1*** L**n**(x) L**n**(x) dx = (−1)**n** (2**n**n!)***2*** | | | | | | | |  | |  |
| ***1*** |  |  |  |  |  | (2n)! | | ***1*** |  |  |
|  |  |  |  |  |  |  | |  |  |  |
|  |  |  |  | |  |  |  |  | |  |
|  |  |  | = 2(2**n**n!)***2*** | | | | **Z*0*** | (1 − x***2***)**n** dx. | |  |
|  |  |  |  |  | (2n)! | | ***1*** |  |  |  |
|  |  |  |  |  |  | |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |

|  |  |
| --- | --- |
| 2.5. Приближение в евклидовых пространствах | **85** |

**Для вычисления последнего интеграла воспользуемся заменой пере-менных** x= sinϕ**, получим**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Z** | ***0*** | (1 − x***2***)**n** dx = | **Z** | ***0*** | (cos ϕ)***2*n*+1***dϕ = (2n + 1)! . | | |
|  | ***1*** |  |  | **π/*2*** |  | (2**n**n!)***2*** |  |

**Следовательно, скалярное произведение** L**n**(x) **на себя равно**

2

.

2n + 1

**Выпишем первые полиномы Лежандра, как они даются формулой Родрига (2.46):**

L***0***(x) = 1,

L***1***(x) = x,

L***2***(x) = ***12*** (3x***2*** − 1),

L***3***(x) = ***12*** (5x***3*** − 3x),

L***4***(x) = ***18*** (35x***4*** − 30x***2*** + 3),

L***5***(x) = ***18*** (63x***5*** − 70x***3*** + 15x),

. . . .

**Важнейшие свойства полиномов Лежандра:**

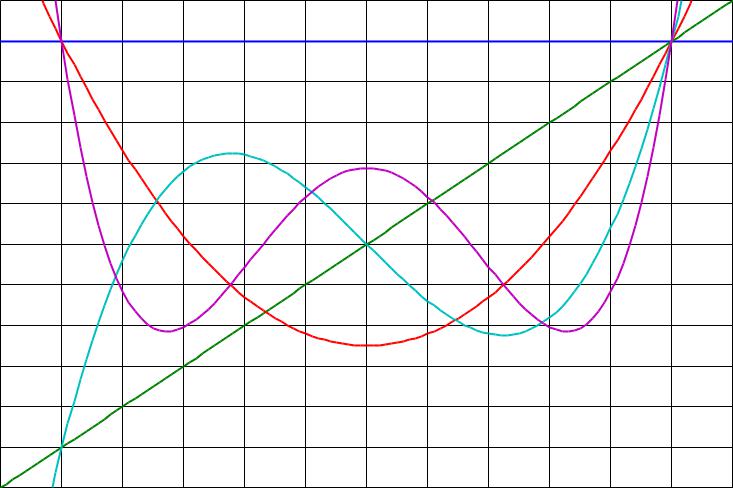
1. L**n**(1) = 1**,** L**n**(−1) = (−1)**n,** n = 0, 1, 2, . . .
2. **Полином Лежандра** L**n имеет** n **различных вещественных корней, которые расположены на интервале** [−1,1]**.**
3. **Для полиномов Лежандра справедливо рекуррентное представление**

(n + 1)L**n*+1***(x) = (2n + 1) xL**n**(x) − nL**n**−***1***(x).

1. **Полиномы Лежандра ортогональны друг другу на интервале** [−1,1]**:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ***1*** L**i**(x)L**j** (x) dx = |  | 2 | **если** | 6 |  |
| **Z**−***1*** |  | 0, |  | i = j, |  |
|  |  |  | , **если** | i = j. |  |
|  |  | 2i + 1 |  |  |  |
|  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **86** |  |  |  |  |  | 2. | Численные методы анализа | | | |
| 1 |  |  |  |  |  |  |  |  |  |  |
| 0.8 |  |  |  |  |  |  |  |  |  |  |
| 0.6 |  |  |  |  |  |  |  |  |  |  |
| 0.4 |  |  |  |  |  |  |  |  |  |  |
| 0.2 |  |  |  |  |  |  |  |  |  |  |
| 0 |  |  |  |  |  |  |  |  |  |  |
| −0.2 |  |  |  |  |  |  |  |  |  |  |
| −0.4 |  |  |  |  |  |  |  |  |  |  |
| −0.6 |  |  |  |  |  |  |  |  |  |  |
| −0.8 |  |  |  |  |  |  |  |  |  |  |
| −1 |  |  |  |  |  |  |  |  |  |  |
| −1 | −0.8 | −0.6 | −0.4 | −0.2 | 0 | 0.2 | 0.4 | 0.6 | 0.8 | 1 |
| **Рис. 2.12. Графики первых полиномов Лежандра на интервале *[***−***1***.***2***, ***1***.***2]*.** | | | | | | | | | | |



Предложение 2.5.1 **Все нули полиномов Лежандра** L**n**(x) **вещест-венны, различны и находятся на интервале** [−1,1]**.**

Доказательство. **Предположим, что среди корней полинома** L**n**(x)**,** **лежащих на** [−1,1]**, имеется** s **штук различных корней** η***1*,** η***2*, . . . ,** η**s нечётной кратности** α***1*,** α***2*, . . . ,** α**s, так что**

L**n**(x) = (x − η***1***)**α*1*** (x − η***2***)**α*2*** · · · (x − η**s**)**α**s γ(x),

**где в полиноме** γ(x) **присутствуют корни** L**n**(x)**, не лежащие на** [−1,1]**, а также те корни** L**n**(x) **из** [−1,1]**, которые имеют чётную кратность. Таким образом,** γ(x) **уже не меняет знака на интервале** [−1,1]**.**

**Ясно, что** s≤n**, и наша задача установить равенство** s=n**. Рассмотрим интеграл**

**Z b**

I = L**n**(x) (x − η***1***)(x − η***2***) · · · (x − η**s**) dx

**a**

* + **b**
* (x − η***1***)**α*1+1***(x − η***2***)**α*2*** ***+1*** · · · (x − η**s**)**α**s***+1*** γ(x) dx.

**a**

|  |  |
| --- | --- |
| 2.6. Численное интегрирование | **87** |

**Теперь** α***1*** + 1**,** α***2*** + 1**, . . . ,** α**s** + 1 **чётные числа, так что подынте-гральное выражение не меняет знак на** [−1,1]**. Это выражение не равно тождественному нулю, и потому определённо** I6= 0**.**

**С другой стороны, выражение для** I **есть скалярное произведение, в смысле** L***2***[−1,1]**, полинома** L**n**(x) **на полином** (x−η***1***)(x−η***2***)· · ·(x−η**s**) **степени не более** n−1**, если выполнено условие** s < n**. Следовательно, при этом должно быть** I= 0**.**

**Полученное противоречие снимается лишь при условии** s=n**, т. е. когда все корни полинома** L**n**(x) **различны и лежат на интервале** [−1,1]**.**

**Можно показать дополнительно, что нули полинома** L**n**(x) **переме-жаются с нулями полинома** L**n*+1***(x)**.**

**Помимо полиномов Лежандра существуют и другие семейства ор-тогональных полиномов, широко используемые в практических вычис-лениях. В частности, введённые в §2.1е полиномы Чебышёва образу-ют семейство полиномов, ортогональных на интервале** [−1,1] **с весом**

(1 − x***2***)−***1*/*2*.**

**Часто возникает необходимость воспользоваться ортогональными полиномами на бесконечных интервалах** [0,+∞] **или даже** [−∞,∞]**. Естественно, единичным весом** %(x) = 1 **тут не обойтись. Полиномы, ортогональные на этих бесконечных интервалах с весами** e−**x и** e−**x*2* называются полиномами Лагерра и полиномами Эрмита соответствен-но. Они также находят многообразные применения в задачах прибли-жения.**

1. Численное интегрирование

**Задачей численного интегрирования мы называем задачу нахождения определённого интеграла**

**Z b**

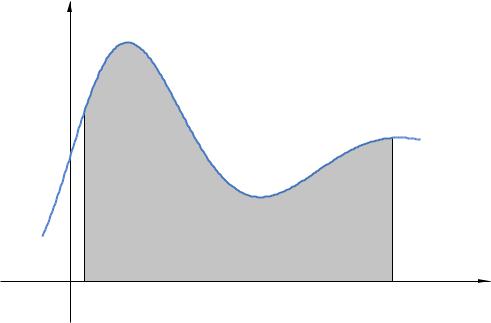
f (x) dx,

**a**

**на основе знания значений функции** f(x)**, без привлечения её первооб-разных и известной формулы Ньютона-Лейбница.**

**Подобная задача нередко возникает на практике, например, если подынтегральная функция** f(x) **задана таблично, т. е. набором своих значений в дискретном наборе точек, а не аналитической формулой. В**

**88** 2. Численные методы анализа



?

x

**Рис. 2.13. Вычисление определённого интеграла необходимо при нахождении площадей фигур с криволинейными границами**

**некоторых случаях численное нахождение интеграла приходится вы-полнять потому, что применение формулы Ньютона-Лейбница требует знания первообразной для интегрируемой функции, которая может не выражается через элементарные функции. Но даже если эта первооб-разная может быть найдена в конечном виде, её вычисление не всегда осуществляется просто (длинное и неустойчивое к ошибкам округления выражение и т. п.).**

**Наибольшее распространение в вычислительной практике получили**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **формулы вида** | **n** | c**k**f (x**k**), | **(2.47)** |  |
| **Z** **b** f (x) dx ≈ |  |
|  | **X** |  |  |  |

* **k*=0***

**где** c**k некоторые постоянные коэффициенты,** x**k точки из интерва-ла интегрирования** [a, b]**,** k= 0,1, . . . , n**. Подобные формулы называют**

**квадратурными формулами, коэффициенты** c**k это веса квадратур-ной формулы, а точки** x**k её узлы. В многомерном случае аналогич-ные приближённые равенства**

**Z** **n**

**X**

f (x) dx ≈ c**k**f (x**k** ),

* **k*=0***

**где** x, x**k** ∈R**m**, Ω **область в** R**m**, m≥2,

**называют кубатурными формулами. То, что квадратурные и кубатур-ные формулы являются линейными выражениями от значений инте-грируемой функции в узлах, объясняется линейной зависимостью са-**

|  |  |
| --- | --- |
| 2.6. Численное интегрирование | **89** |

**мого интеграла от подынтегральной функции. С другой стороны, квад-ратурные формулы можно рассматриваеть как обобщения интеграль-ных сумм Римана (через которые интеграл Римана и определяется), так как простейшие квадратурные формулы прямоугольников просто совпадают с этими интегральными суммами.**

**Как и ранее, совокупность узлов квадратурной (кубатурной) фор-мулы называют сеткой. Разность**

**Z b** **n**

**X**

Ψ = f (x) dx − c**k**f (x**k**)

* **k*=0***

**называется погрешностью квадратурной формулы или её остаточным членом.**

**Условие устойчивости вычислений по квадратурной формуле при значительном количестве узлов** n **вызывает требование положительно-сти весов** c**k (см. подробнее [32], §4.2.4). В частности, тогда при интегри-ровании функций, принимающих значения одного знака, мы избегаем потери точности при вычитании близких значений.**

**2.6а** **Формулы Ньютона-Котеса. Простейшие квадратурные формулы**

**Простейший приём построения квадратурных формул замена подын-тегральной функции** f(x) **на интервале интегрирования** [a, b] **на ¾более простую¿, легче интегрируемую функцию, которая интерполирует или приближает** f(x) **по узлам** x***0*,** x***1*, . . . ,** x**n. В случае, когда** f(x) **заменяет-ся интерполянтом, говорят о квадратурных формулах интерполяцион-ного типа, или, что равносильно, об интерполяционных квадратурных формулах.**

**Наконец, наиболее часто подынтегральную функцию интерполиру-ют полиномами, и подобные интерполяционные квадратурные форму-лы, полученные с помощью алгебраической интерполяции на равномер-ной сетке с простыми узлами, называют формулами Ньютона-Котеса. В зависимости от того, включаются ли концы интервала интегрирова-ния** [a, b] **в число узлов квадратурной формулы, различают формулы Ньютона-Котеса замкнутого типа и открытого типа.**

**Построим формулы Ньютона-Котеса для** n= 0,1,2**, причём для простоты будем строить формулы замкнутого типа за исключением случая** n= 0**, когда замкнутая формула просто невозможна.**

**90** 2. Численные методы анализа

**Если** n= 0**, то подынтегральная функция** f(x) **интерполируется полиномом нулевой степени, т. е. какой-то константой, равной значению**

f (x) **в узле** x***0*.**

**Наиболее естественно взять**

x***0*** = ***12*** (a + b),

**т. е. серединой интервала интегрирования** [a, b]**. При этом получаем квадратурную формулу**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Za** | f (x) dx ≈ (b − a) · f a | | 2 | , |  |
|  | **b** | | + b |  |  |
|  |  |  |  |  |
|  |  |  |  |  |  |

**называемую формулой прямоугольников, так как согласно ей интеграл берётся равным площади прямоугольника с основанием** (b−a) **и высо-той** f((a+b)/2)**. Эту формулу иногда называют также формулой сред-них прямоугольников, поскольку существуют формула ¾левых прямо-угольников¿ и формула ¾правых прямоугольников¿, соответствующие**

**случаям** x***0*** =a **и** x***0*** =b**.**

**Оценим погрешность формулы прямоугольников методом локаль-ных разложений, который ранее был использован при исследовании численного дифференцирования. Разлагая** f(x) **в окрестности точки** x***0*** = (a + b)/2 **по формуле Тейлора, получим**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| f (x) = f | 2 |  | + f 0 | | | |  |  |  | 2 |  |  | | · x − | |  |  | 2 | |  |  | + | | | f | 002 |  |  | · | x − a | | 2 |  | , |  |
|  | a + b | |  |  |  |  |  | a + b | | | |  |  |  |  |  | a + b | | | | |  |  |  |  | (ξ) | | |  |  |  | + b | ***2*** |  |  |
|  |  | |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **где** ξ∈[a, b]**, и далее** | | | | | **Za** | |  | f (x) dx − (b − a) · f a 2 | | | | | | | | | | | | | | | | | | |  | |  |  |  |  |  |  |  |
|  |  | Ψ = | | |  | |  |  | | | | | | | | | | | | | | | | | | |  | |  |  |  |  |  |  |  |
|  |  |  |  |  | **b** | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | + b |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  | **Za** | | f (x) − f a 2 | | | | | | | | | | |  |  |  | | | |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  | = |  | |  | | | | | | | | | | |  | dx | | | | | |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  | **b** | |  |  |  |  |  |  |  |  |  | + b | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  | **Za** | |  |  | f | 002 |  |  |  |  |  |  | |  |  |  |  |  |  | | |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  | = |  | |  |  |  |  |  |  |  | · x − a | | | | 2 |  |  |  |  | dx, | | |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  | **b** | |  |  |  | (ξ) | | | |  | |  |  | + b | | |  | | ***2*** |  |  |  |  |  |  |  |  |  |  |  |  |
| **поскольку** |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | **Za** | | x − a 2 | | | | | | | dx = **Z**−***(*a*+*b*)*/*2*** t dt = 0, | | | | | | | | | | | | | | | | | | |  |  |  |  |  |  |
|  |  |  | **b** |  |  |  |  |  | + b | |  |  |  |  |  |  |  | ***(*a*+*b*)*/*2*** | | | | | | |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

|  |  |
| --- | --- |
| 2.6. Численное интегрирование | **91** |

**интеграл от первого члена разложения зануляется. Следовательно, с учётом принятого нами обозначения**

M**p** = max |f ***(*p*)***(x)|

**x**∈***[*a,b*]***

**можно выписать оценку**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | |  | | ≤ | 2 | **Za** | **b** |  |  | − | 2 |  |  |  | 24 |  |  |
|  | Ψ |  | M***2*** |  |  | x |  | a + b | |  | ***2*** dx = | M***2***(b − a)***3*** | . |  |
|  |  |  |  |  |  |  |  |  |  |  |  |

**Отсюда, в частности, следует, что для полиномов степени не выше** 1

**формула (средних) прямоугольников даёт точное значение интеграла, коль скоро вторая производная подынтегральной функции тогда зану-**

**ляется и** M***2*** = 0**.**

**Полученная оценка точности неулучшаема, так как достигается на**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **функции** g(x) =x− ***21*** (a+b) ***2*. При этом** | | | | | | | | | | | a + b | |  |  |  |  |  |
|  |  |  | max | | | | g00(x) = 2, | | |  | g | = 0, | |  |  |
|  |  |  |  | 2 |  |  |  |
| M***2*** = **x**∈***[*a,b*]*** | | | | |  | | |  |  |  |  |  |  |  |  |
| **и потому** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **b** g(x) dx | − | (b | − | a) | · | g | a + b |  | = | (b − a)***3*** | |  | = |  | M***2***(b − a)***3*** | , |  |
|  |  | |  |  |
| **Za** |  |  |  | 2 |  | 12 | |  |  | 24 | |  |  |

**т. е. имеем точное равенство на погрешность.**

**Рассмотрим теперь квадратурную формулу Ньютона-Котеса, соот-ветствующую случаю** n= 1**, когда подынтегральная функция прибли-жается интерполяционным полиномом первой степени. Построим его по узлам** x***0*** =a **и** x***1*** =b**:**

1

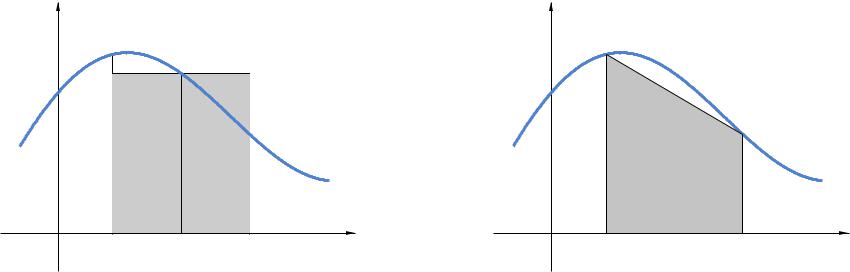
P***1***(x) = b − a (x − a) f (a) − (x − b) f (b) .

**Интегрируя это равенство, получим квадратурную формулу трапеций**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Za** | f (x) dx ≈ (b − a) · f (a) | | 2 | , |  |
|  | **b** | | + f (b) |  |  |
|  |  |  |  |  |
|  |  |  |  |  |  |

**название которой также навеяно геометрическим образом. Именно, точ-ное значение интеграла заменяется на значение площади трапеции (сто-ящей боком на оси абсцисс) с высотой** (b−a) **и основаниями, равными** f (a) **и** f (b) **(см. Рис. 2.14).**

**92** 2. Численные методы анализа



**Рис. 2.14. Иллюстрация квадратурных формул**

**средних прямоугольников и трапеций**

**Для оценивания погрешности формулы трапеций вспомним оценку (2.14) погрешности интерполяционного полинома, из которой следует, что**

f (x) − P***1***(x) = f 00(ξ(x)) · (x − a)(x − b) 2

**для некоторой точки** ξ(x)∈[a, b]**. Следовательно, для формулы трапе-ций**

Ψ = **Z** **b** f 00(ξ(x)) · (x − a)(x − b) dx,

**a** 2

**и окончательно**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | Ψ | | ≤ | 2 | **Za** | **b** | (x | − | a)(x | − |  | 12 | . |  |
|  |  | M***2*** |  |  |  | b) dx = M***2***(b − a)***3*** | |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |

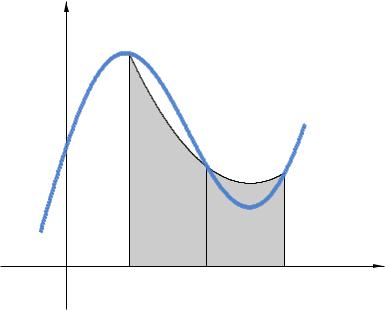
**Опять таки, эта оценка неулучшаема, поскольку достигается на функ-ции** g(x) = (x−a)***2*.**

**2.6б** **Квадратурная формула Симпсона**

**Построим квадратурную формулу типа Ньютона-Котеса для** n= 2**, т. е. для трёх узлов**

|  |  |  |  |
| --- | --- | --- | --- |
| x***0*** = a,x***1*** = | a + b | ,x***2*** = b, |  |
|  |  |  |  |
|  | 2 |  |  |

|  |  |
| --- | --- |
| 2.6. Численное интегрирование | **93** |



**Рис. 2.15. Иллюстрация квадратурной формулы Симпсона**

**по которым строится интерполяционный полином второй степени**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| P (x) = (b | | − | a)***2*** | | x − a | | | 2 |  | (x − b) f (a) | | | | |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  | + b | |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | 2 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  | | |  | | |  |  |  | + (x − a) x − a | | | | |  |  | f (b) . |  |
|  |  |  | + 2(x − a)(x − b) f | | | | | | | | 2 |  | 2 |  |  |
|  |  |  |  |  |  |  |  |  |  |  | a + b |  | |  |  |  | | + b | |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | | | | | | | | | | |  |  | | | | |  | |  |  |  |
| **Интегрируя его по** [a, b]**, получим приближённое равенство** | | | | | | | | | | | | | | | | | | | |  |  |
| **Za** | |  |  |  | ≈ 6 | | |  | · | |  |  |  | 2 |  |  | |  |  |  |  |
|  | **b** |  |  |  |  | b − a | | |  |  |  |  |  | a + b |  |  |  |  |  |  |  |
|  | f (x) dx | | | |  |  | f (a) + 4f | | |  |  | + f (b) , | |  |  | **(2.48)** |  |
|  |  |  | | |  |  |  |  |  |  |  |

**которое называется квадратурной формулой Симпсона или формулой парабол (см. Рис. 2.15), коль скоро она основана на приближении подын-тегральной функции подходящей параболой.**

**Покажем, что формула Симпсона является точной для любого по-линома третьей степени.**

**Отметим прежде всего, что для полиномов степени не выше второй этот факт следует прямо из того, что формула строилась нами как ин-терполяционная квадратурная формула, основанная на интерполяции подынтегральной функции полиномом второй степени. Поэтому доста-точно показать, что формула Симпсона точна для монома** x***3*. Имеем**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Za** |  | 4 |  |  |
| **b**x***3*** dx= | b***4*** | − a***4*** | , |  |
|  |  |
|  |  |  |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **94** | |  |  |  |  |  |  |  |  |  |  | 2. | Численные методы анализа | | | | | | | |  |
| **а с другой стороны** | | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 6 | |  |  | 2 | 6 | |  |  |  | 2 | |  |  |  |  |  |  |
|  | b − a |  | a***3*** + 4 | a + b | ***3*** | + b***3*** | = | b − a |  | a***3*** + | | | | a***3*** + 3a***2***b + 3ab***2*** + b***3*** | | | | | + b***3*** | |  |
|  |  |  |  |  |  |  |  | |  |  | |  |
|  |  |  |  |  |  |  | = b − a | | · |  | 3a***3*** | | + 3a***2***b + 3ab***2*** + 3b***3*** | | | | |  |  |  |  |
|  |  |  |  |  |  |  |  | 6 |  |  |  | 2 | |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  | = | b − a | a***3*** | | | + a***2***b + ab***2*** + b***3*** | | | = | b***4*** | | − a***4*** | | , |  |
|  |  |  |  |  |  |  |  | |  |
|  |  |  |  |  |  |  | 4 | |  |  |  |  |  |  |  | | | 4 | |  |  |

**что совпадает с результатом точного интегрирования. Следовательно, несмотря на то, что интерполяционный полином в**

**формуле Симпсона строился нами по трём точкам, его фактическая точность более высока, чем та, что обычно обеспечивается полиномом второй степени. Для правильной оценки погрешности формулы Симп-сона в этой ситуации нужно взять интерполяционный полином более высокой степени третьей. При наличии всего 3 узлов мы находимся в условиях задачи интерполяции с кратными узлами. Можно считать, к примеру, что кратным является узел** x***1*** = (a+b)/2**, формально решая задачу построения такого интерполяционного полинома** H***3***(x)**, что**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | H***3***(a) | | | = f | (a), | | , | H***3***0 | H***3***(b) | | | = f (b), | | , | **(2.49)** |  |
| H***3*** | a | | 2 |  | = f |  | 2 | 2 |  |  | = f 0 | 2 | **(2.50)** |  |
|  |  |  | + b |  |  |  | a + b |  |  | a + b | |  |  | a + b |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

**в которой значение производной в средней точке** (a+b)/2 **даже не используется.**

**В §2.1ж мы установили существование и единственность решения подобных задач, а также привели оценку (2.25) его погрешности:**

|  |  |  |  |
| --- | --- | --- | --- |
|  | f ***(*m*+1)***(ξ(x)) | **n** |  |
|  |  | **Y** |  |
| f (x) − H**m**(x) = |  | (x − x**i**)**N**i . |  |
| (m + 1)! |  |
|  |  | **i*=0*** |  |

**В случае решения задачи (2.49)–(2.50)**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| f (x) − H***3***(x) = | f ***(4)*** | (ξ(x)) | · (x − a) x − | a | + b |  | ***2*** |  |
|  |  |  | (x − b). |  |
|  | 24 |  | 2 |  |

**Далее, из того, что формула Симпсона точна для полиномов третьей**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2.6. Численное интегрирование | | | | |  |  |  |  |  |  |  |  |  | **95** |  |
| **степени, следует равенство** | | | | · |  |  |  | ***3*** |  |  |  |  |  |  |  |
| **Za** | ***3*** |  | 6 | ***3*** |  |  |  | 2 |  | ***3*** |  |
|  | **b** |  | b − a |  |  |  |  |  | a + b | | |  |  |  |  |
|  | H (x) dx = | |  | H (a) + 4H | | |  | + H (b) | |  |  |
|  |  | · |  |  |  | |  |  |
|  |  |  | 6 |  |  |  | 2 |  |  |  |  |  |  |  |
|  |  | = | b − a |  | f (a) + 4f | | a + b | | |  | + f (b) . | | |  |  |
|  |  |  |  |  |  |  |  |  |  |

**Отсюда уже нетрудно вывести выражение для погрешности квадра-турной формулы**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Za** | − 6 | | · |  | 2 |  |  |  |
| **b** |  | b − a |  |  | a + b |  |  |  |
| Ψ =f (x) dx |  |  | f (a) + 4f | + f (b) |  |  |
|  |  |  |  |  |  |
| **Z b** |  |  |  |  |  |  |  |  |

* f (x) − H***3***(x) dx

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **a** | | |  |  |  |  |  |  |  |  |  | · (x − a) x − | | | | | |  |  |  |  |  | (x − b) dx, | | |  |  |  |  |
| = | **Za** | | |  | f | | ***(4)*** | | | | 24 | | 2 | |  |  |  |  |  |  |  |
|  |  |  | **b** |  |  |  | (ξ(x)) | |  |  |  | |  |  | a + b ***2*** | | | | | |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  |  |  |  |
| **из которого следует оценка** | | | | | | | | | | | | |  | · (x − a) x − a | | | | | | |  |  |  |  |  |  |  |  |  |  |
| |Ψ| ≤ |  | | **a** | | |  |  |  |  |  |  | 24 |  | 2 |  |  |  | (x − b) dx | |  |  |  |
|  | **Z** | | |  | **b** | |  | f ***(4)***(ξ(x)) | | | | | |  |  |  |  | |  |  | + b | | | ***2*** |  |  |  |  |  |
|  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | |  | **b** | |  | | f | | ***(4)***(ξ(x)) | | | | · | (x − a) x − | | | | | | a + b | | | | ***2*** |  |  |  |  |  |
| ≤ **Za** | | | |  |  |  |  | 24 |  | 2 | |  |  | (x − b) | | dx | |  |  |
|  |  |  |  |  |  |  | |  |  |  |  |  |  |  |  | |  | |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  | |  |  |  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| ≤ |  | M | | | |  | |  |  | |  | **b** |  |  |  | | + b | | ***2*** | |  |  |  |  |  |  |  |  | **(2.51)** |  |
|  |  | 24***4*** · | | | | | |  | **a** (x − a) | | | x − a | | 2 |  |  | (x − b) dx | | | | | | , |  |  |  |
|  |  |  |  |  |  |  |  |  | **Z** | | |  |  |  | | |  |  |  | |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | ***2*** | (x − b) |  |
| **коль скоро подынтегральное выражение** (x−a)x−(a+b)/2 | | | | | | | | | | | | | | | | | | | | | | | | | | | |  |  |
|  |  |  | **на интервале интегрирования** [a, b]**. Здесь обозначено** | | | | | | | | | | | | | | | | | | | | | | | | | | |  |
| **не меняет знак** | | |  | ***(4)*** | | | | | (x)|**.** | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| M***4*** = max**x**∈***[*a,b*]*** |f | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

**Для вычисления фигурирующего в (2.51) интеграла сделаем замену переменных**

a + b

t = x −

2

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **96** |  |  |  |  |  |  |  |  |  |  |  |  | 2. Численные методы анализа | | | | | | | | |  |
| **тогда** |  | (x − a) x − a 2 | | |  |  |  | (x − b) dx | | | | | |  |  |  |  |  |  |  |  |  |
| **Za** | **b** |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  | + b | |  | ***2*** | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  | **Z**−***(*b**−**a*)2*** | | | | |  |  |  | 2 |  |  |  |  | − | | 2 |  | | |  |
|  |  | = |  | ***(*b**−**a*)*/*2*** | | | | t + | | b − a | | | t***2*** | t | |  |  | b − a | dt | | |  |
|  |  | **Z**−***(*b**−**a*)2*** | | | | |  |  |  |
|  |  |  | | |  |  |  |
|  |  |  |  |  |  | − |  | 4 |  |  |  |  | − 120 | | |  |
|  |  | = |  | ***(*b**−**a*)*/*2*** t***2*** | | | | | t***2*** | |  | (b − a)***2*** | | | | |  | dt = |  | (b − a)***5*** | . |  |
|  |  |  |  |  | | |  | |  |  |  |
| **Окончательно** | | |  |  |  |  |  |  | M***4*** (b − a)***5*** | | | | | |  |  |  |  |  |  |  |  |
|  |  |  | | | | Ψ | | | | ≤ | . | |  |  |  |  |  |  |
|  |  |  |  |  |  |  | 2880 | | |  |  |  |  |  |  |  |  |  |

**Как видим, более тонкие рассуждения позволили получить действи-тельно более точную оценку погрешности.**

**2.6в** **Дальнейшие формулы Ньютона-Котеса**

**Из других формул квадратурных формул Ньютона-Котеса отметим формулу ¾трёх восьмых¿**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Za** | ≈ 8 | | · |  | 3 |  |  | 3 |  |  |  |
| **b** |  | b − a |  |  | 2a + b |  |  | a + 2b |  |  |  |
| f (x) dx |  | f (a) + 3f |  |  | + 3f |  | + f (b) , **(2.52)** |  |
|  |  |  |  |  |  |  |  |

**погрешность которой оценивается как**

|Ψ| ≤ M***4*** (b − a)***5*** .

6480

**Можно показать, что формулы Ньютона-Котеса с нечётным числом узлов, один из узлов которых приходится на середину интервала ин-тегрирования, имеют (как формула Симпсона) повышенный порядок точности.**

**Формулы Ньютона-Котеса высоких порядков не очень употреби-тельны, так как они проигрывают по точности формулам Гаусса (изу-чаемым далее в §§2.6ж–2.6з) и другим квадратурным формулам. Кро-ме того, среди весов формул Ньютона-Котеса при** n= 8**,** 10 **и б´ольших значениях** n **встречаются отрицательные, что снижает ценность соот-ветствующих формул.**

|  |  |
| --- | --- |
| 2.6. Численное интегрирование | **97** |

**2.6г** **Общие интерполяционные квадратурные формулы**

**Квадратурные формулы интерполяционного типа были определены как формулы, получающиеся в результате замены подынтегральной функ-ции** f(x) **интерполяционным полиномом** P**n**(x)**. Выпишем его в форме Лагранжа**

**n**

**X**

P**n**(x) = f (x**i**) φ**i**(x),

**i*=0***

**где**

ω**n**(x)

φ**i**(x) = (x − x**i**) ω**n**0(x**i**) ,

**базисные полиномы Лагранжа (стр. 29), а функция** ω**n**(x) **определена в (2.7):**

ω**n**(x) = (x − x***0***)(x − x***1***) · · · (x − x**n**).

**Тогда коэффициенты квадратурной формулы (2.47) должны иметь вид**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| c**i** | = | **Za** | φ**i**(x) dx | = | **Za** | **b** | (x − x**i**) ω**n**0 | (x**i**) dx, | | i = 0, 1, . . . , n, **(2.53)** |  |
|  |  |  | **b** |  |  | ω**n**(x) |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |

**и эти значения весов** c**i, определяемых по узлам** x***0*,** x***1*, . . . ,** x**n, явля-ется характеристическим признаком интерполяционной квадратурной формулы.**

**Погрешность квадратурной формулы будет при этом равна**

* + **b**
* =R**n**(f, x) dx,

**a**

**где** R**n**(f, x) **остаточный член интерполяционной формулы. В §2.1д была получена оценка для** R**n**(f, x) **в виде (2.14)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| R**n**(f, x) = | | | f ***(*n*+1)***(ξ(x)) | · ω**n**(x), |  |
|  | | | (n + 1)! |  |  |
| **где** ξ(x)∈[a, b]**, и поэтому** | | |  |  |  |
| Ψ = (n + 1)! **Za** | | | f ***(*n*+1)***(ξ(x)) ω**n**(x) dx. | |  |
| 1 | |  | **b** |  |  |
|  | |  |  |  |  |
|  |  |  |  |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **98** |  |  | 2. Численные методы анализа |  |
| **В целом справедлива оценка** | | | |ω**n**(x)| dx, |  |
| |Ψ| ≤ (n + 1)! **Za** | | |  |
|  | M**n*+1*** |  | **b** |  |
|  |  |  |

**где** M**n*+1*** = max**x**∈***[*a,b*]*** |f ***(*n*+1)***(x)|**. Из неё можно ещё раз заключить, что квадратурная формула интерполяционного типа, построенная по**

(n + 1) **узлу, является точной для любого полинома степени не более** n**.** **Верно и обратное утверждение: если квадратурная формула (2.47), построенная по** (n+ 1) **узлу, является точной для любого полинома степени не выше** n**, то она является квадратурной формулой интерпо-ляционного типа, и потому её веса (коэффициенты) вычисляются по**

**формулам (2.53).**

**В самом деле, для базисных интерполяционных полиномов** φ**i**(x)

**выполнено свойство (2.5)**

**(**

0, **при** i=6j,

φ**i**(x**j** ) = δ**ij** =

1, **при** i=j,

**и они имеют степень** n**. Следовательно, получаем**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Zab** | **n** | **n** | c**k**δ**ik** = c**i**, |  |
| φ**i**(x) dx = **k*=0*** c**k** φ**i**(x**k** ) = | **l*=0*** |  |
|  | **X** | **X** | |  |

**т. е. имеет место равенство (2.53), что и требовалось доказать. Итак, нами обоснована**

Теорема 2.6.1 **Для того, чтобы квадратурная формула** **(2.47), по-строенная по** (n+ 1) **попарно различному узлу, была интерполяцион-ной, необходимо и достаточно, чтобы она являлась точной на поли-номах степени** n**.**

**Отметим, что при этом квадратурная формула может быть более точной также и для полиномов степени выше** n**, как, например, фор-мула Симпсона.**

**2.6д** **Сходимость квадратур**

**С теоретический точки зрения весьма важен вопрос о сходимости квад-ратур при неограниченном возрастании числа узлов. Похожий вопрос**

|  |  |
| --- | --- |
| 2.6. Численное интегрирование | **99** |

**вставал при исследовании о сходимости интерполяционного процесса, и мы обсуждали его в §2.1з. Но в случае квадратурных формул помимо бесконечной треугольной матрицы узлов**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| x***0(2)*** | | x***1(2)*** |  |  |
|  | x***0(1)*** |  |  |  |
| x***0*** | x***1*** | x***2*** |  |
|  |  |  |  |  |
| **...** | | | |  |
|  | ***(3)*** | ***(3)*** | ***(3)*** |  |

**..** **..** **..**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | · · · | |  |  |  |  |
| 0 · · · | | | |  | , | **(2.54)** |  |
|  |  |  |  |  |  |  |  |
|  |  | · · · | |  |  |  |  |
| **.** | **..** | **.** | **. .** |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |

**таких что** x***(*kn*)* лежат на интервале интегрирования** [a, b]**, необходимо задавать ещё и треугольную матрицу весовых коэффицинтов квадра-**

**турных формул**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| c***0(2)*** | | c***1(2)*** |  |  |
|  | c***0(1)*** |  |  |  |
| c***0*** | c***1*** | c***2*** |  |
|  |  |  |  |  |
| **...** | | | |  |
|  | ***(3)*** | ***(3)*** | ***(3)*** |  |

**..** **..** **..**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | · · · | |  |  |  |  |
| 0 · · · | | | |  | . | **(2.55)** |  |
|  |  |  |  |  |  |  |  |
|  |  | · · · | |  |  |  |  |
| **.** | **. .** | **.** | **..** |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |

Определение 2.6.1 **Будем говорить, что квадратурный процесс, за-даваемый зависящим от натурального параметра** n **семейством квад-ратурных формул**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Z** | **b** | **n** |  |  |
|  | f (x) dx ≈ **X** c**k*(*n*)***f x**k*(*n*)*** , | n = 0, 1, 2, . . . , |  |

* **k*=0***

**которые определяются матрицами узлов (2.54) и весов (2.55),** сходит-ся **для функции** f (x) **на интервале** [a, b]**, если**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **n** |  | f x**k** | |  | **b** |  |
| **n**→∞ **k*=0*** c**k** | | |  | **Za** |  |
|  | **X** | ***(*n*)*** |  | ***(*n*)*** |  |  |
| lim |  |  | = | f (x) dx, |  |
|  |  |  |  |  |

**т. е. если при неограниченном возрстании** n **предел результатов квад-ратурных формул равен точному интегралу.**

**Весьма общие достаточные условия для сходимости квадратур были сформулированы и обоснованы В.А. Стекловым [47], а впоследствии Д. Пойа [48] доказал также необходимость условий Стеклова.**

**100** 2. Численные методы анализа

Теорема 2.6.2 **(теорема Стеклова-Пойа)** **Для того, чтобы квадратур-ный процесс, порождаемый матрицами узлов (2.54) и весов (2.55) схо-дился для непрерывной на** [a, b] **функции, необходимо и достаточно, чтобы**

1. **этот процесс сходился для полиномов,**
2. **суммы абсолютных значений весов были равномерно по** n **огра-ничены, т. е. существала такая константа** C**, что**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **n** |  |  |  |  |  |
| **X** | c**k*(*n*)*** | ≤ C | **(2.56)** |  |
|  |  |  |
|  |  |  |  |

**k*=0***

**для всех** n= 0,1,2, . . . **.**

**Покажем достаточность условий теоремы Стеклова-Пойа. С этой целью, задавшись каким-то** >0**, найдём полином** P**N** (x)**, который равномерно с точностью приближает непрерывную подынтегральную функцию** f(x) **на рассматриваемом интервале** [a, b]**. Существование та-кого полинома обеспечивается теоремой Вейерштрасса (см. §2.1з). Да-лее преобразуем выражение для остаточного члена квадратурной фор-мулы:**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Z** | **b** |  | **n** |  |  |  |  |
| R**n**(f ) = |  |  | f (x) dx − **X** c**k*(*n*)***f x**k*(*n*)*** | | |  |  |  |
|  | **a** |  |  | **k*=0*** |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| = | **Za** | **b** |  | **b** | **n** | c**k*(*n*)***f x**k*(*n*)*** |  |  |
|  |  |  |  | f (x) − P**N** (x) dx + **Za** | P**N** (x) dx − **k*=0*** |  |  |  |
|  | **Z b** | |  |  | **X** |  |  |  |
| = |  | |  | f (x) − P**N** (x) dx |  |  |  |  |
|  |  |  |  |  |  |  |  |  |

**a**

**Z b**

+

**a**

|  |  |  |
| --- | --- | --- |
| **n** |  |  |
| P**N** (x) dx − **X** | |  |
| **k*=0*** | |  |
| **n** |  |  |
| + **k*=0*** c**k*(*n*)*** |  |
| **X** |  |  |

**!**

c***(*kn*)***P**N** x***(*kn*)***

P**N** x***(*kn*)*** − f x***(*kn*)***

|  |  |
| --- | --- |
| 2.6. Численное интегрирование | **101** |

**Отдельные слагаемые полученной суммы оцениваются следующим об-разом:**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **ab** |  | f (x) − P**N** (x) dx | |  |  | ≤ (b − a), | | | | |  |
| **Z** |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | **b** |  | **n** |  |  |  |  |  |  |  |  |  |
|  |  |  |  | ***(*** |  |  |  |  | ***(*n*)*** |  |  |  |
|  | **a** | P**N** (x) dx −c**k** | | | |  |  | P**N** | x**k** | ≤ , |  |
| **Z** |  |  | **k*=0*** |  |  |  |  |  |  |  |  |  |
|  |  |  | **X** |  | **n*)*** | | |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |

**так как квадратурный процесс сходится на полиномах,**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | c**k*(*n*)*** | P**N** x**k*(*n*)*** − f x**k*(*n*)*** |  | ≤ | **n** | c**k*(*n*)*** . |  |
| **n** |  |  |  |  |  |  |
| **k*=0*** |  |  | **k*=0*** |  |  |
| **X** |  |  | **X** |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |

**Поэтому в целом для достаточно больших значений** n **имеем**

|R**n**(x)| ≤ (b − a + C + 1),

**откуда и следует сходимость рассматриваемого квадратурного процес-са.**

**Доказательство необходимости условия теоремы Стеклова-Пойа мож-но найти помимо оригинальной статьи [48] также в книге [29].**

**В формулировке теоремы фигурирует величина**

**n**

**X** c**k** ,

**k*=0***

**сумма абсолютных значений весов, которая играет очень важную роль при оценке качества различных квадратурных формул, будучи коэффициентом увеличения погрешности в данных.**

**2.6е** **Составные квадратурные формулы**

**Рассмотренные выше квадратурные формулы дают приемлемую по-грешность при небольших длинах интервала интегрирования** [a, b] **в случае, когда подынтегральная функция не претерпервает сильных и резких изменений. Но если величина** (b−a) **относительно велика или интегрируемая функция имеет большие производные, то погрешность**

**102** 2. Численные методы анализа

**вычисления интеграла делается значительной или даже большой. То-гда для получения требуемой точности вычисления интеграла приме-няют составные квадратурные формулы, основанные на разбиении ин-тервала интегрирования на подынтервалы меньшей длины, по каждо-му из которых вычисляется значение ¾элементарной квадратуры¿, а затем искомый интеграл приближается их суммой.**5

**Будм считать, что погрешность** Ψ **используемой квадратурной фор-мулы имеет оценку**

|Ψ| ≤ C(b − a)**p**,

**где** C **константа, зависящая от типа квадратурной формулы и ин-тегрируемой функции, причём в любом случае** p >1**. К примеру, для формулы средних прямоугольников** C=M***2***/24 **и** p= 3**, а для фор-мулы Симпсона** C=M***4***/2880 **и** p= 5**. Если интервал интегрирования разбит точками** r***1*,** r***2*, . . . ,** r**N** −***1* на** N **равных частей** [a, r***1***]**,** [r***1***, r***2***]**, . . . ,** [r**N** −***1***, b] **длины** h = (b−a)/N **, то можно вычислить по рассматриваемой** **формуле интегралы**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Zr**i i***+1*** f (x) dx, | i = 0, 1, . . . , N − 1, | |  |  |
| **r** |  |  |  |  |
| **где** r***0*** =a**,** b=r**N , а затем положить** | |  |  |  |
| **Z** **b** f (x) dx ≈ | **N** −***1* Z r**i***+1*** f(x) dx. | | **(2.57)** |  |
| **a** | **X** | **r**i |  |  |
| **i*=0*** |  |  |
|  |  |  |  |

**Тогда на каждом подынтервале**

|Ψ| ≤C b − a **p** =Ch**p**,

N

**а полная погрешность интегрирования при использовании представле-**

˜

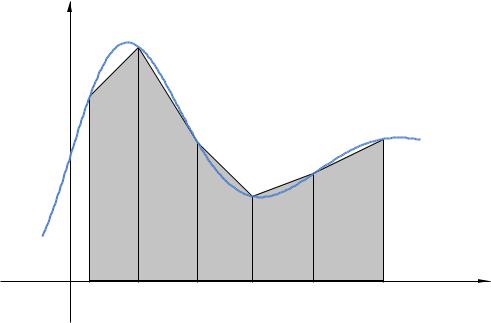
**ния (2.57) есть такая величина** Ψ**, что**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | |  | | ≤ |  | N | | **p** |  | N **p**−***1*** | | − |  |  |
|  | Ψ˜ |  | CN | b − a |  | = C | (b − a)**p** | = C(b |  | a)h**p**−***1***. |  |
|  |  |  |  |  |
|  |  |  |  |  |  |  |

**Как видим, эта погрешность уменьшилась в** N **p**−***1* раз, и потенциаль-но таким способом погрешность вычисления интеграла можно сделать сколь угодно малой.**

5 **Интересно, что при этом подынтегральная функция интерполируется на всём интервале интегрирования *[***a, b***]* при помощи сплайна.**

|  |  |
| --- | --- |
| 2.6. Численное интегрирование | **103** |



**Рис. 2.16. Составная квадратурная формула трапеций**

**Число** (p−1) **часто называют порядком точности (составной) квад-ратурной формулы. Ясно, что основная идея составных квадратурных формул работает и в случае неравномерного разбиения интервала ин-тегрирования на более мелкие части, но анализ погрешности проводить тогда труднее.**

**Для равномерного разбиения интервала интегрирования составные квадратурные формулы выглядят особенно просто. Выпишем их явный вид для рассмотренных выше простейших квадратур Ньютона-Котеса и равномерного разбиения интервала интегрирования** [a, b] **на** N **рав-ных частей длины** h **каждая.**

**Составная формула прямоугольников**

|  |  |  |
| --- | --- | --- |
| **Z** **b** f (x) dx ≈ h | **N** | f(r**i**−***1*/*2***). |
|  | **X** |  |

* **i*=1***

**где** r**i**−***1*/*2*** =r**i** −h/2**. Её полная погрешность**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| | | Ψ˜ | | ≤ | M | ***2*** | (b − a)h***2*** | , |  |
|  |  | 24 |  |  |

**т. е. она имеет второй порядок точности. Составная формула трапеций**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **b** | f (x) dx ≈ h | **N** −***1*** | ***21*** f (b) . |  |
| **Za** |  | ***21*** f (a) + **i*=1*** f (r**i**) + |  |  |
|  |  | **X** |  |  |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **104** |  |  |  |  |  |  |  | 2. Численные методы анализа | |  |
| **Её полная погрешность** | | |  |  |  |  |  |  |  |  |
|  | | | | Ψ˜ | | | ≤ | M | ***2*** | (b − a)h***2*** | , |  |
|  | 12 |  |
|  |  |  |  |  |  |
| **т. е. порядок точности тоже второй.** | | | | | | | | |  |  |
| **Составная формула Симпсона** | | | | | | |  |  |  |  |
| **b** | f (x) dx ≈ 6 | | |  | **N** | f (r**i**−***1***) + 4f (r**i**−***1*/*2***) + f (r**i**) , | | | |  |
| **Za** | **i*=1*** | |  |
|  |  | h | | **X** | | |  |  |  |  |
|  |  |  |  |  |  |  |
| **где** r**i**−***1*/*2*** =r**i** −h/2**. Её полная погрешность** | | | | | | | | | |  |
|  | | | | Ψ˜ | | | ≤ | M | ***4*** | (b − a)h***4*** | , |  |
|  | 2880 |  |
|  |  |  |  |  |  |

**т. е. формула имеет четвёртый порядок точности.**

**Как видим, при использовании составных квадратурных формул увеличение точности вычисления интеграла достигается ценой допол-нительных трудозатрат. В рассматриваемом нами одномерном случае эти трудозатраты растут всего лишь линейно, хотя и здесь необходи-мость вычисления сложной подынтегральной функции может иногда быть обременительной. Но при возрастании размерности интеграла, ко-гда необходимо прибегнуть к составным кубатурным формулам, рост трудозатрат делается уже значительным, имея тот же порядок, что и размерность пространства. Так же растёт и погрешность суммиро-вания результатов интегрирования по отдельным подобластям общей области интегрирования. Поэтому и эффект увеличения точности со-ставной формулы при возрастании размерности становится всё менее ощутимым.**

**2.6ж** **Квадратурные формулы Гаусса**

**Параметрами квадратурной формулы (2.47)**

**Z b** **n**

**X**

f (x) dx ≈ c**k**f (x**k**)

* **k*=0***

**являются узлы** x**k и веса (коэффициенты)** c**k. Однако, строя квадра-турные формулы Ньютона-Котеса, в частности, формулу трапеций и**

|  |  |
| --- | --- |
| 2.6. Численное интегрирование | **105** |

**формулу Симпсона, мы заранее задавали положение узлов и по ним на-ходили веса. Таким образом, возможности общей формулы (2.47) были использованы не полностью, так как для достижения наилучших ре-зультатов можно было бы управлять положением её узлов. Только в формуле прямоугольников положение единственного узла было выбра-но из соображений симметрии, и это привело к существеному улучше-нию точности.**

**Здесь, правда, возникает весьма важный методический вопрос: как измерять это ¾улучшение¿ квадратурной формулы? Что брать крите-рием её точности? В идеальном случае желательно было бы минимизи-ровать погрешность квадратурной формулы для тех или иных классов функций, но в такой общей постановке задача делается сложной (хотя и не неразрешимой).**

**Один из естественных ответов на этот вопрос состоит в том, что-бы в качестве меры точности квадратурной формулы брать степень полиномов, на которых она даёт точные результаты.**

Определение 2.6.2 Алгебраической степенью точности **квадратур-ной формулы называют наибольшую степень алгебраических полино-мов, для которых эта квадратурная формула является точной.**

**Как следствие, сформулированную в начале пункта задачу опти-мизации узлов можно переформулировать следующим образом: для заданного фиксированого числа узлов из интервала интегрирования нужно построить квадратурную формулу, которая имеет наивысшую алгебраическую степень точности, т. е. является точной на полиномах наиболее высокой степени. Удовлетворяющие поставленному требова-нию формулы действительно существуют и называются квадратурны-ми формулами Гаусса, поскольку впервые были рассмотрены в начале XIX века К.Ф. Гауссом.**

**Далее для удобства мы будем записывать квадратурные формулы**

|  |  |  |  |
| --- | --- | --- | --- |
| **Гаусса не в виде (2.47), а как** |  |  |  |
| **Z** **b** f (x) dx ≈ | **n** | c**k**f (x**k**), | **(2.58)** |
|  | **X** |  |  |

|  |  |  |
| --- | --- | --- |
| **a** | **k*=1*** |  |
|  |  |  |

**нумеруя узлы с** k= 1**, а не с нуля. Требование точного равенства для любого полинома степени** m **в этой формуле равносильно в силу её линейности тому, что формула является точной для одночленов** f(x) =

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **106** |  |  | 2. Численные методы анализа | |  |
| x**l,** l = 0, 1, 2, . . . , m**, т. е.** | |  |  |  |  |
| **Z** | **b** | **n** |  |  |  |
| x**l** dx = | **X** c**k**x**kl**, | l = 0, 1, 2, . . . , m. | **(2.59)** |  |

|  |  |  |
| --- | --- | --- |
| **a** | **k*=1*** |  |
|  |  |

**Это система из** (m+ 1) **нелинейных уравнений с** 2n **неизвестными вели-чинами** c***1*,** c***2*, . . . ,** c**n,** x***1*,** x***2*, . . . ,** x**n. Число уравнений совпадает с чис-лом неизвестных при** m+ 1 = 2n**, т. е.** m= 2n−1**, и это в общем случае есть максимальное возможное значение для** m**, так как при б´ольших значениях** m **получающаяся система уравнений переопределена и, ско-рее всего, окажется неразрешимой.**

**Итак, наивысшая алгебраическая степень точности квадратурной формулы, построенной по** n **узлам, в общем случае может быть рав-на** 2n−1**. Для двух узлов это** 3**, при трёх узлах имеем** 5**, и т. д. Для сравнения напомним, что алгебраическая точность формул трапеций и Симпсона, построенных по двум и трём узлам соответственно, равна всего** 1 **и** 3**. При возрастании числа узлов этот выигрыш в алгебраиче-ской степени точности формул Гаусса, достигнутый за счёт разумного расположения узлов, нарастает.**

**При небольших** n **система уравнений (2.59) может быть решена с помощью несложных аналитических преобразований.**

**Пусть** n= 1**, тогда система уравнений (2.59) принимает вид**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **(** c***1***x***1*** | | | = ***21*** (b***2*** − a***2***). |  |
|  | c***1*** | | = b − a, |  |
| **Отсюда** |  |  |  |  |
| c***1*** = b − a, | | |  |  |
| x***1*** = | 1 | (b***2*** − a***2***) = ***21*** (a + b). | |  |
|  |  |
| 2c***1*** |  |

**Как легко видеть, получающаяся квадратурная формула это фор-мула (средних) прямоугольников**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Za** | **b** | f (x) dx ≈ (b − a) · f a | | 2 | . |  |
|  |  |  | | + b |  |  |
|  |  |  | |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

**Действительно, нам известно, что она резко выделяется своей точно-стью среди родственных квадратурных формул.**

|  |  |
| --- | --- |
| 2.6. Численное интегрирование | **107** |

**Пусть** n= 2**, тогда алгебраическая степень точности** m= 2n−1 = 3**. Система уравнений (2.59) для узлов и весов принимает вид**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| c***1***x***1*** | | + c***2***x***2*** | = | ***21*** | (b***2*** | |  | a***2***), | |  |
|  |  | c***1*** + c***2*** | = b − a, | | | | |  |  |  |
| ***2*** | ***2*** |  | ***1*** |  | ***3*** | − | ***3*** |  |  |
|  |  |  | (b |  | ), |  |
| c***1***x + c***2***x = | | | | ***3*** |  |  | a |  |
|  | ***1*** | ***2*** |  |  |  | − |  |  |  |
|  |  | + c***2***x***23*** | = |  |  |  |  |  |  |
| c***1***x***13*** | | ***41*** (b***4*** − a***4***). | | | | | |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |

**Пользуясь симметричностью этой системы уравнений можно положить** c***1*** = c***2*, что даёт**

c***1*** = c***2*** = ***12*** (b − a),

|  |  |  |  |
| --- | --- | --- | --- |
| **а из первого и второго уравнений позволяет получить** | | |  |
|  | x***1*** + x***2*** = a + b. | | **(2.60)** |
| **В то же время из первого и третьего уравнений следует** | | |  |
| x***2*** | + x***2*** | = ***2*** (b***2*** + ab + b***2***). | **(2.61)** |
| ***1*** | ***2*** | ***3*** |  |

**Возводя (2.60) в квадрат и вычитая из результата уравнение (2.61), будем иметь**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| x***1***x***2*** | = ***1*** | (b***2*** | + 4ba + a***2***). | **(2.62)** |
|  | ***6*** |  |  |  |

**Соотношения (2.60) и (2.62) на основе теоремы Виета позволяют сде-лать вывод, что** x***1* и** x***2* являются корнями квадратного уравнения**

x***2*** − (a + b) x + ***16*** (b***2*** + 4ba + a***2***) = 0,

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **так что** |  |  | √ |  |  |  |  |  |
| x***1*,*2*** = | ***1*** | (a + b) ± | ***3*** | | (b − a). | **(2.63)** |  |
| ***2*** | ***6*** | |  |  |

**Удовлетворение полученными решениями четвёртого уравнения систе-мы проверяется прямой подстановкой. В целом мы получили квадра-турную формулу Гаусса**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Za** | 2 | · | ***1*** | ***2*** |  |  |  |
| **b** | b − a |  |  |  |  |  |  |
| f (x) dx = |  | f (x ) + f (x ) , | | | **(2.64)** |  |
|  |  |  |

**где узлы** x***1* и** x***2* определяются из формулы (2.63).**

**108** 2. Численные методы анализа

Пример 2.6.1 **Вычислим по выведенной выше формуле Гаусса с дву-мя узлами интеграл**

**Z π/*2***

cos x dx,

***0***

**точное значение которого равно** 1**. Согласно (2.63) и (2.64) имеем**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **π/*2*** | cos x dx ≈ |  | π/2 | · cos | π |  | √ |  |  | π | | + cos |  | π |  | √ |  |  | π | **!** |  |
|  |  |  |  |  |  |  |  |
| **Z** | ***0*** | 2 | | 4 | − 63 | | | 2 | |  | 4 | | + 63 | | | 2 | |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | = | 0.998473. | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

**Формула Ньютона-Котеса с двумя узлами** 0 **и** π/2 **формула трапеций даёт для этого интеграла значение**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Z*0*** | **π/*2*** | cos x dx ≈ 2 | | · | cos 0 + cos 2 | |  | = 0.785398, |
|  |  |  | π |  |  | π |  |  |

**точность которого весьма низка.**

**Чтобы получить с формулами Ньютона-Котеса точность вычисле-ния рассматриваемого интеграла, сравнимую с той, что даёт формула Гаусса, приходится брать больше узлов. Так, формула Симпсона (2.48), использующая три узла** 0**,** π/4 **и** π/2**, приводит к результату**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Z*0*** | cos x dx ≈ | 6 | |  |  | · cos 0 + 4 cos 4 | | | | | + cos 2 | |  |  |
|  | **π/*2*** | π/2 | | | |  |  |  |  | π |  | π |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  | √ |  |  |  |  |  |  |  |  |
|  |  | π | |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  | = |  |  |  | | 1 + 2 2 | | | = 1.00228, | | |  |  |  |
|  | 12 |  |  |  |  |

**погрешность которого по порядку величины примерно равна погреш-ности ответа по формуле Гаусса (2.64), но всё-таки превосходит её в полтора раза.**

**С ростом** n **сложность системы уравнений (2.59) для узлов и весов быстро нарастает, так что в общем случае остаётся неясным, будет ли разрешима эта система (2.59) при любом наперёд заданном** n**. Будут ли у неё вещественные решения? Будут ли эти решения принадлежать интервалу** [a, b]**, чтобы служить удобными узлами квадратурной фор-мулы?**

|  |  |
| --- | --- |
| 2.6. Численное интегрирование | **109** |

**Получение ответов на эти вопросы непосредственно из системы урав-нений (2.59) представляется громоздким и малоперспективным. Обыч-но следуют другим путём, расчленяя получившуюся задачу на отдель-ные подзадачи**

* 1. **исследования выбора узлов и**
  2. **построения весов по формулам (2.53), путём интегрирования коэффициентов интерполяционного полинома Лагранжа.**
* **последнем случае мы пользуемся тем фактом, что конструируемая квадратурная формула Гаусса оказывается квадратурной формулой интерполяционного типа. Это прямо следует из Теоремы 2.6.1.**

**2.6з** **Выбор узлов для квадратурных формул Гаусса**

Теорема 2.6.3 **Квадратурная формула** **(2.58)**

**Z b** **n**

**X**

f (x) dx ≈ c**k**f (x**k**)

* **k*=1***

**точна на полиномах степени** (2n−1) **тогда и только тогда, когда**

1. **формула (2.58) является формулой интерполяционного типа;**
2. **её узлы** x***1*,** x***2*, . . . ,** x**n являются корнями такого полинома**

ω(x) = (x − x***1***)(x − x***2***) · · · (x − x**n**),

**что**

**Z b**

ω(x) q(x) dx = 0

**a**

**для любого полинома** q(x) **степени не выше** (n−1)**.**

**Выражение**

**Z b**

ω(x) q(x) dx

**a**

**интеграл от произведения двух функций, уже встречались нам в §2.5б. Мы могли видеть в §2.5б, что на пространстве** L***2***[a, b] **всех инте-грируемых с квадратом функций оно задаёт скалярное произведение, т. е. симметричную билинейную форму. По этой причине утверждение**

**110** 2. Численные методы анализа

**Теоремы 2.6.3 часто формулируют так: для того, чтобы квадратурная формула**

**Z b** **n**

**X**

f (x) dx ≈ c**k**f (x**k**)

* **k*=1***

**была точна на полиномах степени** (2n−1)**, необходимо и достаточно, чтобы эта формула имела интерполяционный тип, а её узлы** x***1*,** x***2*, . . . ,** x**n** **были корнями полинома** ω(x)**, который ортогонален на** [a, b] **любому** **полиному** q(x) **степени не выше** (n−1)**.**

Доказательство. **Необходимость. Пусть рассматриваемая квадратур-ная формула точна на полиномах степени** (2n−1)**. Таковым является, в частности, полином** ω(x)q(x)**, имеющий степень** n+ (n−1)**. Тогда**

|  |  |  |
| --- | --- | --- |
| **Zab** ω(x) q(x) dx = | **n** |  |
| **k*=1*** c**k** ω(x**k**) q(x**k** ) = 0, |  |
|  | **X** |  |

**поскольку все** ω(x**k**) = 0**. И это соотношение верно для любого полино-ма** q(x) **степени не выше** n−1**.**

**Справедливость условия 1) следует из того доказанного в §2.6г фак-та, что если квадратурная формула (2.58) является точной для любого полинома степени не выше** n−1**, то она является квадратурной фор-мулой интерполяционного типа.**

**Достаточность. Пусть имеется полином** ω(x) **степени** n**, имеющий** n **различных корней на интервале** [a, b] **и удовлетворяющий условию** **ортогональности с любым полиномом** q(x) **степени не выше** (n−1)**. Покажем, что интерполяционная квадратурная формула, построенная по узлам** x***1*,** x***2*, . . . ,** x**n, которые являются корнями** ω(x)**, будет точна на полиномах степени** 2n−1**.**

**Пусть** f(x) **произвольный полином степени** 2n−1**. Тогда после деления его на** ω(x) **получим представление**

f (x) = ω(x) q(x) + r(x),

**где** q(x) **и** r(x) **соответственно частное и остаток от деления** f(x) **на** ω(x)**. При этом оба полинома** q(x) **и** r(x) **имеют степени меньше** n**. Отсюда**

**Z b** **Z b** **Z b** **Z b**

f (x) dx = ω(x) q(x) dx + r(x) dx = r(x) dx,

**a** **a** **a** **a**

|  |  |
| --- | --- |
| 2.6. Численное интегрирование | **111** |

**в силу сделанного нами предположения об ортогональности** ω(x) **всем полиномам степени не выше** n−1**.**

**Но рассматриваемая квадратурная формула имеет интерполяцион-ный тип и построена по** n **узлам, и потому она является точной на полиномах степени не выше** n−1**, в частности на полиноме** r(x)**. Сле-довательно,**

**Z b**

**a**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **n** | **n** |  | **n** |  |
| **X** | **X** |  | **X** |  |
| r(x) dx =c**k**r(x**k** ) = | c**k** f (x**k** ) − ω(x**k** ) q(x**k**) = | | c**k**f (x**k**), |  |
| **k*=1*** | **k*=1*** |  | **k*=1*** |  |

**коль скоро** ω(x**k**) = 0**. Итак,**

**Z b** **n**

**X**

f (x) dx = c**k**f (x**k**),

* + **k*=1***
* **е. исследуемая квадратурная формула действительно является точ-**

**ной на полиномах степени** 2n−1**.**

**Подведём промежуточные итоги. Процедура построения квадратур-ных формул Гаусса разделена нами на две отдельные задачи нахожде-ния узлов и вычисления весов. В свою очередь, узлы квадратурной формулы, как выясняется, можно взять корнями некоторых специаль-ных полиномов** ω(x) **из формулировки Теоремы 2.6.3. В этих полино-мах легко угадываются знакомые нам с §2.5в ортогональные полиномы, являющиеся полиномами Лежандра для** [a, b] = [−1,1]**, и соответствую-щим образом преобразованные для произвольного интервала интегри-рования** [a, b]**.**

**Обращаясь к построению квадратурных формул Гаусса, отметим, что отдельное нахождение их узлов и весов для каждого конкретно-го интервала интегрирования** [a, b] **является весьма трудозатратным, и если бы нам нужно было проделывать эту процедуру всякий раз при смене интервала** [a, b]**, то практическое применение формул Гаусса зна-чительно потеряло бы свою привлекательность. Естественная идея со-стоит в том, чтобы в этой ситуации воспользоваться формулой замены**

|  |  |  |
| --- | --- | --- |
| **переменных в определённом интеграле** | |  |
| **Z** **b** f (x) dx = **Z** **β** f (ϕ(y)) ϕ0(y) dy | | **(2.65)** |
| **a** | **α** |  |

**для** a=ϕ(α)**,** b=ϕ(β)**, и свести задачу вычисления определённого ин-теграла от** f(x) **по** [a, b] **к задаче вычисления интеграла по некоторому**

**112** 2. Численные методы анализа

**¾каноническому¿ интервалу. В качестве такового обычно берут** [−1,1]**, т. е. тот интервал, для которого строятся ортогональные полиномы Ле-жандра. Но при замене переменных не должно портиться свойство ор-тогональности полиномов, требуемое Теоремой 2.6.3. Нетрудно понять, что это условие будет соблюдено при простейшей линейной замене пе-ременных, когда** ϕ0= const**. Именно, если**

x = ***12*** (a + b) + ***12*** (b − a) y,

**то переменная** x **пробегает интервал** [a, b]**, если** y∈[−1,1]**. Обратное преобразование даётся формулой**

1 y = b − a 2x − (a + b) .

**В частности, если** y**k,** k= 0,1, . . . , n**, корни полинома Лежандра, лежащие на интервале** [−1,1]**, то узлы нашей квадратурной формулы на интервале интегрирования** [a, b] **суть** x**k** = ***12*** (a+b) + ***12*** (b−a)y**k.**

**Наконец, веса** c**k любой интерполяционной квадратурной формулы выражаются в виде интегралов (2.53), которые в случае формул Гаусса (когда узлы нумеруются с единицы), принимают вид**

**Z b**

c**k** = φ**k**(x) dx, k = 1, 2, . . . , n,

**a**

**где** φ**k**(x) **базисный полином Лагранжа (стр. 29). Тогда в силу фор-мулы замены переменных (2.65)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| c**k** | = **Za** | φ**k**(x) dx = | ***21*** (b − a) | **Z**−***1*** φ(y) dy, |
|  |  | **b** |  | ***1*** |

k = 0, 1, . . . , n**, коль скоро**

dx = d ***12*** (a + b) + ***12*** (b − a) y = ***12*** (b − a) dy.

**Получается, что веса квадратурной формулы Гаусса для произвольно-го интервала интегрирования** [a, b] **вычисляются простым умножением весов для канонического интервала** [−1,1] **на множитель *12*** (b−a)**.**

**Для интервала** [−1,1] **узлы квадратурных формул Гаусса (т. е. кор-ни полиномов Лежандра) и веса тщательно затабулированы для пер-вых натуральных чисел** n**, и их конкретные значения могут быть най-дены, к примеру, в подробных руководствах по численным методам**

|  |  |
| --- | --- |
| 2.6. Численное интегрирование | **113** |

**[3, 5, 11, 19, 44] или в специальных справочниках, например, в [21]. В частности, в [5] значения весов и узлов фомул Гаусса приведены для небольших** n **с 16 значащими цифрами, а в справочнике [21] с 15 значащими цифрами вплоть до** n= 48**. Таким образом, практическое применение квадратур Гаусса обычно не встречает затруднений.**

**Приведём для примера квадратурную формулу Гаусса для трёх уз-лов и интервала** [−1,1]**:**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Z**−***1*** f (x) dx | ≈ 9 f (x***1***) + 9 f (x***2***) + 9 f (x***3***), | | | | | | | |  |
| ***1*** | 5 | | 8 | |  | 5 | |  |  |
|  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
| x***1*,*3*** | = 0.774596669240, | | | | | x***2*** = 0. | | |  |

**Погрешность квадратурной формулы Гаусса, построенной по** n **уз-лам** x***1*,** x***2*, . . . ,** x**n** ∈[a, b] **может быть оценена по формуле Маркова**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Ψ(f ) = f | | ***(2*n*)***(ξ) | **b** | ω(x) | ***2*** dx, |  |
| (2n)! | **Za** |  |
|  |  |  |  |  |  |  |

**где** ξ∈]a, b[ **,** ω(x) = (x−x***1***)(x−x***2***)· · ·(x−x**n**)**. Её доказательство несложно и может быть найдено, например, в [44, 29]. Из формулы Маркова следует, что формулы Гаусса с большим числом узлов целе-сообразно применять лишь для достаточно гладких функций.**

**В заключение темы отметим, что на практике нередко требует-ся включение во множество узлов квадратурной формулы каких-либо фиксированных точек интервала интегрирования, например, его кон-цов (одного или обоих), либо каких-то выделенных внутренних точек. Основная идея формул Гаусса может быть применена и в этом случае, что приводит к квадратурам Маркова.**

**Построение квадратурных формул Гаусса основывалось на оптими-зации алгебраического порядка точности квадратур. Эта идея может быть модифицирована и приспособлена к другим ситуациям, когда ал-гебраические полиномы и их степень уже не являются наиболее адек-ватным мерилом точности квадратурной формулы. Например, мож-но развивать квадратуры наивысшего тригонометрического порядка точности, которые окажутся практичнее при вычислении интегралов от осциллирующих функций [44].**

**114** 2. Численные методы анализа

1. Правило Рунге для оценки погрешности

**Предположим, что нам необходимо численно найти интеграл или про-изводную функции, либо найти решение дифференциального или ин-тегрального уравнения, т. е. какой-либо задачи, где фигурирует сетка на некотором интервале вещественной оси. Пусть для решения этой задачи имеется численный метод порядка** p**, так то его погрешность равна** Ch**p, где** h **шаг рассматриваемой сетки, а** C **величина, на-прямую от** h **не зависящая. Как правило, значение** C **не известно точно**

* **его нахождение непосредственно из задачи является делом трудным**
* **малоперспективным. Мы могли видеть, к примеру, что для задачи численного интегрирования выражение для этой константы вовлекает оценки для производных высоких порядков от подынтегральной функ-ции. Во многих случаях их практическое вычисление не представляется возможным, так что оценки эти носят, главным образом, теоретический характер.**
  + **Рунге принадлежит идея использовать для определения значе-ния** C **результаты нескольких расчётов на различных сетках. Далее, после того как величина** C **будет определена, мы можем использовать её значение для практической оценки погрешности приближённых ре-шений нашей задачи, которые получаются с помощью численного ме-тода.**

**Предположим для простоты анализа, что численные решения рас-**

**сматриваемой задачи расчитаны на сетках с шагом** h **и** h/2 **и равны соответственно** I**h и** I**h/*2*, а точное решение есть** I**. Тогда**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  | I**h** − I = Ch**p**, | | | | | | | | | | | = C |  | 2**p** . | | |  |  |
|  |  | I**h/*2*** − I = C | | | | | 2 | | | | | | |  |  |  |
|  |  |  |  |  |  |  |  |  | h | | |  | | **p** | h**p** | | |  |  |  |
| **Вычитая второе равенство из первого, получим** | | | | | | | | | | | | | | | | | | |  |  |
| I | **h** − | I |  | = Ch**p** | | − |  | C | | h**p** | | |  | = Ch**p** | | | 2**p** − 1 | | , |  |
| **h/*2*** |  | 2**p** | | |  | |  |
|  |  |  |  |  |  |  |  |  |  |  |  | 2**p** |  |  |
| **так что** |  |  |  |  | 2**p** | |  |  |  |  | I**h** −I**h/*2*** | | | | |  |  |  |  |  |
|  |  |  | C = | |  | · | | | . |  |  |  |  |
|  |  |  | 2**p** − 1 | | |  | | | h**p** |  |  |  |  |  |

**Зная константу** C**, можно уже находить оценку погрешности расчитан-ных решений** I**h,** I**h/*2* или любых других**

|  |  |
| --- | --- |
| 2.7. Правило Рунге для оценки погрешности | **115** |

Литература к Главе 2

Основная

1. **Барахнин В.Б., Шапеев В.П. Введение в численный анализ. – Санкт-Петербург–Москва– Краснодар: Лань, 2005.**
2. **Бабенко К.И. Основы численного анализа. – Москва: Наука, 1986.**
3. **Бахвалов Н.С., Жидков Н.П., Кобельков Г.М. Численные методы. –**

**Москва: Бином, 2003, а также другие издания этой книги.**

1. **Бахвалов Н.С., Корнев А.А., Чижонков Е.В. Численные методы. Реше-ния задач и упражнения. – Москва: Дрофа, 2008.**
2. **Березин И.С., Жидков Н.П. Методы вычислений. Т. 1–2. – Москва: Наука, 1966.**
3. **Вержбицкий В.М. Численные методы. Части 1–2. – Москва: ¾Оникс 21 век¿, 2005.**
4. **Волков Е.А. Численные методы. – Москва: Наука, 1987.**
5. **Гельфонд А.О. Исчисление конечных разностей. – Москва: Наука, 1967.**
6. **Гончаров В.Л. Теория интерполирования и приближения функций. –**

**Москва: ГИТТЛ, 1954.**

1. **Даугавет И.К. Введение в теорию приближения функций. – Ленинград: Из-дательство Ленинградского университета, 1977.**
2. **Демидович Б.П., Марон А.А. Основы вычислительной математики. –**

**Москва: Наука, 1970.**

1. **Завьялов Ю.С., Квасов Б.И., Мирошниченко В.Л. Методы сплайн-функций. – Москва: Наука, 1980.**
2. **Калиткин Н.Н. Численные методы. – Москва: Наука, 1978.**
3. **Канторович Л.В., Акилов Г.П. Функциональный анализ. – Москва: Наука, 1984.**
4. **Кобков В.В., Шокин Ю.И. Сплайн-функции в численном анализе. – Ново-сибирск: Издательство НГУ, 1983.**
5. **Коллатц Л. Функциональный анализ и вычислительная математика. –**

**Москва: Мир, 1969.**

1. **Кострикин А.Н. Введение в алгебру. Часть 1. Основы алгебры. – Москва: Физматлит, 2001.**
2. **Кронрод А.С. Узлы и веса квадратурных формул. – Москва: Наука, 1964.**
3. **Крылов А.Н. Лекции о приближённых вычислениях. – Москва: ГИТТЛ, 1954, а также более ранние издания.**
4. **Крылов В.И., Бобков В.В., Монастырный П.И. Вычислительные методы. Т. 1–2. – Москва: Наука, 1976.**
5. **Крылов В.И., Шульгина Л.Т. Справочная книга по численному интегриро-ванию. – Москва: Наука, 1966.**

**116** 2. Численные методы анализа

1. **Кунц К.С. Численный анализ. – Киев: Техника, 1964.**
2. **Люстерник Л.А., Червоненкис О.А., Янпольский А.Р. Математический анализ. Вычисление элементарных функций. – Москва: ГИФМЛ, 1963.**
3. **Мак-Кракен Д., Дорн У. Численные методы и программирование на ФОР-ТРАНе. Москва: Мир, 1977.**
4. **Мацокин А.М. Численный анализ. Вычислительные методы линейной алгеб-ры. Конспекты лекций для преподавания в III семестре ММФ НГУ. Ново-сибирск: НГУ, 2009–2010.**
5. **Мацокин А.М., Сорокин С.Б. Численные методы. Часть 1. Численный ана-лиз. Новосибирск: НГУ, 2006.**
6. **Миньков С.Л., Миньков Л.Л. Основы численных методов. – Томск: Изда-тельство научно-технической литературы, 2005.**
7. **Мысовских И.П. Интерполяционные кубатурные формулы. – Москва: Нау-ка, 1981.**
8. **Натансон И.П. Конструктивная теория функций. – Москва–Ленинград: ГИТТЛ, 1949.**
9. **Никольский С.М. Квадратурные формулы. – Москва: Наука, 1988.**
10. **Ортега Дж., Пул У. Введение в численные методы решения дифференци-альных уравнений. – Москва: Наука, 1986.**
11. **Самарский А.А., Гулин А.В. Численные методы. – Москва: Наука, 1989.**
12. **Стефенсен И.Ф. Теория интерполяции. – Москва: Объединённое научно-техническое издательство НКТП СССР, 1935.**
13. **Стечкин С.Б., Субботин Ю.Н. Сплайны в вычислительной математике. –**

**Москва: Наука, 1976.**

1. **Тихонов А.Н., Арсенин В.Я. Методы решения некорректных задач. –**

**Москва: Наука, 1974.**

1. **Тыртышников Е.Е. Методы численного анализа. – Москва: Академия, 2007.**
2. **Фихтенгольц Г.М. Курс дифференциального и интегрального исчисления. Т. 1. Москва: Наука, 1966.**

Дополнительная

1. **Абрамовиц , Стиган И. Таблицы специальных функций. – Москва: Наука, .**
2. **Каханер Д., Моулер К., Нэш С. Численные методы и программное обеспе-чение. – Москва: Мир, 1998.**
3. **Квасов Б.И. Методы изогеометрической аппроксимации сплайнами. –**

**Москва: Физматлит, 2006.**

1. **Меньшиков Г.Г. Локализующие вычисления. Конспект лекций. – Санкт-Петербург: СПбГУ, Факультет прикладной математики–процессов управле-ния, 2003.**

2.7. Правило Рунге для оценки погрешности **117**

1. **Микеладзе Ш.Е. Численные методы математического анализа. – Москва: ГИТТЛ, 1953.**
2. **Милн В.Э. Численный анализ. – Москва: Издательство иностранной литера-туры, 1951.**
3. **Мысовских И.П. Лекции по методам вычислений. – Санкт-Петербург: Изда-тельство Санкт-Петербургского университета, 1998.**
4. **Сегё Г. Ортогонльные многочлены. – Москва: Физматлит, 1962.**
5. **Соболев С.Л. Введение в теорию кубатурных формул. – Москва: Наука, 1974.**
6. **Стеклов В.А. О приближённом вычислении определённых интегралов // Из-вестия Академии Наук. – 1916. – Т. 10, №6. – С. 169–186.**

**¨**

**[48] Polya G. Uber Konvergenz von Quadraturverfahren // Mathematische Zeitschrift.**

**– 1933. – Bd. 37. – S. 264–286.**

**[49]** **Aberth O. Precise numerical methods using C++. – San Diego: Academic Press, 1998.**

**[50] Moore R.E., Kearfott R.B., Cloud M. Introduction to interval analysis. –**

**Philadelphia: SIAM, 2009.**

Глава 3

Численные методы линейной алгебры

1. Задачи вычислительной линейной алгебры

**3.1а** **Особенности постановок задач**

**Численные методы линейной алгебры это один из классических раз-делов вычислительной математики, который вычленился даже в от-дельное научное направление в середине прошлого века в связи с бур-ным развитием математических вычислений на ЭВМ. Традиционный, исторически сложившийся список задач вычислительной линейной ал-гебры по состоянию на 50–60-е годы прошлого века можно найти в капитальной книге Д.К. Фаддеева и В.Н. Фаддеевой, и он включал**

* + **решение систем линейных алгебраических уравнений,**
  + **вычисление определителей матриц,**
  + **нахождение обратной матрицы**
  + **нахождение собственных значений и собственных векторов матриц,**
* **также многочисленные разновидности этих задач.**

**118**

|  |  |
| --- | --- |
| 3.1. Задачи вычислительной линейной алгебры | **119** |

**Но ¾всё течёт, всё меняется¿. По мере развития науки и технологий в фокусе развития вычислительной линейной алгебры оказались новые задачи. Вот как формулирует список важнейших задач в 2001 году американский специалист Дж. Деммель в книге [14]:**

* **решение систем линейных алгебраических уравнений;**
* **линейная задача о наименьших квадратах:**

**найти вектор** x**, минимизирующий** hAx−b, Ax−bi **для заданных** m×n**-матрицы** A **и** m**-вектора** b**;**

* **нахождение собственных значений и собственных векторов матриц;**
* **нахождение сингулярных чисел и сингулярных векторов матриц.**

**Постановку последней задачи мы будем обсуждать ниже в §3.1б. Вторая задача из этого списка линейная задача о наименьших квад-ратах возникает в связи с решением переопределённых СЛАУ, ко-торые, к примеру, получаются при обработке экспериментальных дан-ных.**

**С точки зрения классических разделов математики решение выпи-санных задач не встречает затруднений:**

**– решение квадратной СЛАУ получается покомпонентно по фор-муле Крамера, как частное двух определителей;**

**– для вычисления собственных значений матрицы нужно выписать характеристическое уравнение и найти его корни.**

* **так далее. Но практическая реализация этих теоретических рецептов наталкивается на почти непреодолимые трудности. К примеру, явная**

**формула для определителя** n×n**-матрицы выражает его как сумму** n! **слагаемых, каждое из которых есть произведение** n **элементов из** **разных строк и столбцов матрицы. Раскрытие определителя по этой**

**формуле требует** n!(n−1) + (n−1)≈n!(n−1) **арифметических опера-ций, и потому из-за взрывного роста факториала**1 **решение СЛАУ по правилу Крамера при** n≈ **20–30 делается уже невозможным даже на самых современных ЭВМ.**

1 **Напомним в этой связи известную в математическом анализе асимптотическую**

√

формулу Стирлинга n***!*** ≈ ***2***πn ***(***n/e***)***n**, где**e ***= 2***.***7182818*** . . .

**120** 3. Численные методы линейной алгебры

**Производительность современных ЭВМ принято выражать в так называемых флопах (сокращение от английской фразы floating point operation), и 1 флоп это одна усреднённая арифметическая операция с плавающей точкой в секунду. Для наиболее мощных на сегодняшний день ЭВМ скорость работы измеряется так называемым терафлопами,** 10***12*** **операций с плавающей точкой в секунду. Для круглого счёта мож-но даже взять производительность нашего гипотетического компьюте-ра равной 1 петафлоп =** 10***15* операций с плавающей точкой в секунду. Решение на такой вычислительной машине системы линейных алгеб-раических уравнений размера** 30×30 **по правилу Крамера потребует примерно**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 30 **компонент решения** | · |  |  | 29 · 30! **операций** | | |  |
|  | 10***15*** **флоп** · 3600 чассек · 24 | | час | · 365 **суток** |  |
|  |  |  |
|  |  | сутки |  |
| ***12*** |  | **Для сравнения, возраст Земли в настоящее** | | | | |  |
| **лет, т. е. около** 7.3·10 **лет.** | |  |
| ***9*** | | **лет.** | | |  |
| **время оценивается в** 4.5·10 | |  |  |  |

**Помимо неприемлемой трудоёмкости ещё одной причиной непри-годности для реальных вычислений некоторых широко известных ал-горитмов из ¾чистой математики¿ является сильное влияние на их ре-зультаты неизбежных погрешностей счёта (округлений и т.п.). Напри-мер, решение СЛАУ по правилу Крамера очень неустойчиво к ошиб-кам.**

**3.1б** **Сингулярные числа и сингулярные векторы матрицы**

**Для квадратной матрицы** A **собственные значения** λ **и собственные векторы** x **удовлетворяют матричному уравнению**

Ax = λ x

**Часто эти собственные векторы называют правыми собственными век-торами, поскольку иногда возникает необходимость рассмотрения ле-вых собственных векторов** y**, таких что**

y>A = µ y>.

**Транспонируя это соотношение, получим**

A>y = µ y,

|  |  |
| --- | --- |
| 3.1. Задачи вычислительной линейной алгебры | **121** |

**т. е. левые собственные векторы матрицы** A **совпадают с правыми соб-ственными векторами матрицы** A> **(что объясняет редкость самостоя-тельного использования понятия левого собственного вектора). Кроме того, транспонирование векового уравнения** det(A−λI) = 0 **приводит к соотношению**

det(A> − λI) = 0,

**которое показывает, что** λ=µ **собственные значения матриц** A **и** A> **совпадают. Таким образом, для определения собственных значе-ний матрицы и её левых и правых собственных векторов необходимо решить систему уравнений**

**(**

Ax = λ x,

**(3.1)**

A>y = λ y,

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **или, в матричной форме,** |  | y **!** |  | y **!** . |  |
| 0 A> **!** | | = λ |  |
| A | 0 | x |  | x |  |

**Система уравнений (3.1) является ¾распавшейся¿: в ней первые** n **урав-нений и последние** n **уравнений не зависят друг от друга. Поэтому ре-шать её также можно по частям, что обычно и делают на практике.**

**Изменим соотношения (3.1), чтобы они ¾завязались¿ друг на друга, поменяв в правых частях векторы** x **и** y**:**

**(**

Ax = σ y,

**(3.2)**

A>y = σ x,

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **или, в матричной форме,** | y **!** |  | y **!** . |  |
| A0> **!** | = σ |  |
| 0 A | x |  | x |  |

**Фигурально можно сказать, что при этом векторы** x **и** y **становятся ¾право-левыми¿ или ¾лево-правыми¿ собственными векторами матри-цы** A**. Аналоги собственных чисел матрицы** A**, которые мы переобозна-чили через** σ**, также получают новое содержание.**

**122** 3. Численные методы линейной алгебры

Определение 3.1.1 **Неотрицательные решения** σ **системы матрич-ных уравнений (3.2) называются** сингулярными числами **матрицы** A**. Удовлетворяющие системе (3.2) векторы** x **называются** правыми син-гулярными векторами **матрицы** A**, а векторы** y левыми сингуляр-ными векторами A**.**

**Отметим, что это определение имеет смысл уже для произвольных прямоугольных матриц, а не только для квадратных, как было в случае собственных значений и собственных векторов.**

Предложение 3.1.1 **Сингулярные числа матрицы** A **суть неотрица-тельные квадратные корни из собственных чисел матрицы** A>A **или матрицы** AA>**.**

**Формулировка этого утверждения требует разъяснений, так как в случае прямоугольной** m×n**-матрицы** A **размеры квадратных матриц** A>A **и** AA> **различны: первая из них это** n × n**-матрица, а вторая** m × m**-матрица. Соответственно, количество собственных чисел у них** **будет различным.**

**Из известного неравенства для ранга произведения матриц следу-ет, что если** m < n**, то** n×n**-матрица** A>A **имеет неполный ранг, не превосходящий** m**, а потому её собственные числа с** (m+ 1)**-го по** n**-ое заведомо нулевые. Аналогично, если** m > n**, то** m×m**-матрица** AA> **также имеет неполный ранг, который не превосходит** n**, и её собствен-ные числа с** (n+ 1)**-го по** m**-ое равны нулю. Таким образом, для** m×n**-матрицы содержательный смысл имеет рассмотрение лишь** min{m, n}

**штук сингулярных чисел, что устраняет вышеотмеченную кажущуюся неоднозначность.**

Доказательство. **Умножая обе части второго уравнения из (3.2) на** σ**,** **получим** A>(σy) =σ***2*** x**. Теперь подставим сюда значение** σy **из первого уравнения (3.2):** A>Ax=σ***2***x**.**

**С другой стороны, умножая на** σ **обе части первого уравнения (3.2), получим** A(σx) =σ***2***y**. Подставив сюда значение** σx **из второго уравне-ния (3.2), получим** AA>y=σ***2***y**.**

**Осталось показать, что собственные значения у матриц** A>A **и** AA>

**неотрицательны, чтобы иметь возможность извлекать из них квадрат-ные корни. Очевидно, это достаточно сделать лишь для одной из вы-писанных матриц.**

|  |  |
| --- | --- |
| 3.1. Задачи вычислительной линейной алгебры | **123** |

**Коль скоро матрица** A>A **симметрична, любое её собственное значе-ние** λ **вещественно. Кроме того, если** v **соответствующий собственный вектор, то** 0≤(Av)>(Av) =v>(A>Av) =v>λv=λv>v**, откуда в силу** v>v ≥ 0 **следует** λ ≥ 0**.**

**Итак, задаваемые Определением 3.1.1 сингулярные числа** m×n**-матрицы это набор из** min{m, n} **неотрицательных вещественных чисел, которые обычно нумеруют в порядке убывания:**

σ***1*** ≥ σ***2*** ≥ . . . ≥ σ***min***{**m,n**} ≥ 0.

**Из доказательства Предложения 3.1.1 следует также, что правыми син-гулярными векторами матрицы** A **являются правые собственные век-торы матрицы** A>A**, а левыми сингулярными векторами матрицы** A **левые собственные векторы для** A>A **или, что равносильно, правые собственные векторы матрицы** AA>**.**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Пример 3.1.1 **Для** 2 × 2**-матрицы** | | | | | | |  |  | |  |  |  |  |
|  |  |  |  | A = |  | 3 | 4 |  |  | **(3.3)** |  |
|  |  |  |  |  |  | 1 | 2 |  |  |  |  |  |  |
| **нетрудно выписать характеристическое уравнение** | | | | | | | | | | | | |  |
|  |  | 3 |  | 4 − λ | | |  |  | − |  | − |  |  |
| det |  | 1 − λ | | 2 |  | = λ***2*** | | |  | 4λ |  | 2 = 0, |  |
| ***1*** |  | √ |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
| **и найти его корни *2*** | (5± 33) **собственные значения матрицы, прибли-** | | | | | | | | | | | |  |

**зительно равные** −0.372 **и** 5.372**. Для определения сингулярных чисел**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **образуем** | 14 | 20 | | , |  |  |  |  |  |  |
| A>A = |  |  |  |  |  |  |
|  | 10 | 14 | |  |  |  |  |  |  |  |
| **и вычислим её собственные значения, равные** 15± | | | | | | √ |  |  |  |  |
|  | 221**. Получается,** | | |  |
| **что сингулярные числа матрицы** A **суть** | | |  | 15 ± | √ |  |  |  |  |  |
|  |  | 221**, т. е. примерно** | | | |  |
| 0.366 **и** 5.465 **(с точностью до трёх знаков** | | | **после запятой).** | | | | | | |  |
|  | **p** |  |  |  |  |  |  |
| **С другой стороны, для матрицы** |  | , |  |  |  |  |  |  |  |  |
| 13 | 4 |  |  |  | **(3.4)** | | | |  |
|  | 2 |  |  |  |  |  |  |  |  |  |

−

**124** 3. Численные методы линейной алгебры

**которая отличается от матрицы (3.3) лишь противоположным знаком**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **элемента на месте** (2,1)**, собственные значения это комплексно-сопря-** | | | | | | | |  |
| ***1*** | | | | √ |  |  |  |  |
|  |  |  |
| **жённая пара *2*** (5±i | | | | 15) ≈ 2.5 ± 1.936 i**, а сингулярные числа суть** | | | |  |
| **p** | 15 ± √ |  | **, т. е. приблизительно** 1.954 **и** 5.117**.** | | | |  |  |
| 125 |  |

**Можно заметить, что максимальные сингулярные числа рассмот-ренных матриц превосходят наибольшие из модулей собственных чи-сел. Мы увидим ниже (см. §3.2д), что это не случайно, и наибольшее сингулярное число всегда превосходит максимум модулей собственных чисел матрицы.**

**Важнейший результат, касающийся сингулярных чисел и сингуляр-ных векторов матриц, который служит одной из основ их широкого применения в разнообразных вопросах математического моделирова-ния это**

Теорема 3.1.1 **(теорема о сингулярном разложении матрицы)**

**Для любой матрицы** A∈R**m**×**n существуют ортогональные матрицы** U ∈ R**m**×**m** **и** V ∈ R**n**×**n, такие что** A = U ΣV > **с** m × n**-матрицей** Σ

**диагонального вида**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 0***1*** | | σ***2*** | 0 | · · · | |  | 0 |  |  |  |
|  |  | σ | 0 | 0 |  |  |  | 0 |  |  |  |
| Σ = |  | 0 0 σ***3*** | | | · · · | |  | 0 |  | , |  |
|  | **. . .** | | | | · · · **.** | | | |  |  |  |
|  | **. . .** | | | | **. .** | **.** | | |  |  |  |
|  | **. . .** | | | |  | **. .** | | |  |  |  |
|  |  | 0 | 0 | 0 |  |  | **.** | **. .** |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  | · · · | |  |  |  |  |  |

**где** σ***1*,** σ***2*, . . . ,** σ***min***{**m,n**} **сингулярные числа матрицы** A**.**

**Для квадратных матриц доказательство этого факта может быть легко выведено из известного полярного разложения матрицы, т. е. её представления в виде** A=QS**, где** Q **ортогональная матрица,** S **симметричная (см., к примеру, [9, 21]). Симметричную матрицу** S **мож-но ортогональными преобразованиями подобия привести к диагональ-ному виду:** S=T>DT **, где** T **ортогональна, а** D **диагональная. Поэтому** A= (QT>)DT **. Это уже почти требуемое представление для** A**, поскольку произведение ортогональных матриц** Q **и** T > **тоже орто-гонально. Нужно лишь убедиться в том, что по диагонали в** D **стоят сингулярные числа матрицы** A**.**

|  |  |
| --- | --- |
| 3.1. Задачи вычислительной линейной алгебры | **125** |

**Рассмотрим произведение** A>A**:**

A>A = (QT >)DT >(QT >)DT = T >D>(QT >)>(QT >)DT

= T >D>DT = T >D***2***T.

**Как видим, матрица** A>A **подобна диагональной матрице** D***2*, их соб-ственные числа потому совпадают и, следовательно, собственные числа** A>A **суть квадраты диагональных элементов** D**. Это и требовалось до-казать.**

**В общем случае доказательство Теоремы 3.1.1 не очень сложно и может быть найдено, к примеру, в книгах [12, 36, 38]. Фактически, этот результат показывает, как с помощью сингулярных чисел мат-рицы элегантно представляется действие соответствующего линейного оператора из одного векторного пространства в другое, возможно с от-личающимися друг от друга размерностями.**

**Задача нахождения сингулярных чисел и сингулярных векторов матриц, последняя из списка на стр. 119, по видимости является част-ным случаем третьей задачи, относящейся к нахождению собственных чисел и собственных векторов. Но, с одной стороны, она сделалась очень важной и в теории, и в приложениях вычислительной линей-ной алгебры, а, с другой, численные методы для её решения весьма специализированы, так что она рассматривается в списке задач уже ¾отдельной строкой¿. Рассказ об истории сингулярного разложения и его многочисленных применений можно найти, к примеру, в живо на-писанной статье [68].**

**Как видим, в современном списке задач на первые места выдвину-лась линейная задача о наименьших квадратах. А некоторые старые и уважаемые задачи как бы отошли на второй план.**

**3.1в** **Матрицы с диагональным преобладанием**

**В приложениях линейной алгебры и теории матриц часто возникают матрицы, в которых диагональные элементы в том или ином смыс-ле ¾преобладают¿ над остальной, недиагональной частью матрицы. Это обстоятельство может быть, к примеру, следствием особенностей рассматриваемой математической модели, в которой связи составля-ющих её частей с самими собой (они и выражаются диагональными элементами) сильнее, чем с остальными. Такие матрицы обладают ря-дом замечательных свойств, изложению одного из которых и посвящён**

**126** 3. Численные методы линейной алгебры

**этот пункт. Следует отметить, что сам смысл, вкладываемый в понятие ¾преобладания¿ может быть различен, и ниже мы рассмотрим простей-ший и наиболее популярный.**

Определение 3.1.2 **Квадратную** n × n**-матрицу** A = (a**ij** ) **называ-ют** матрицей с диагональным преобладанием**, если для любого** i=1, 2, . . . , n **имеет место**

**X**

|a**ii**| > |a**ij** |.

**j*=***6**i**

**В условиях этого определения некоторые авторы говорят о ¾стро-гом диагональном преобладании¿. Со своей стороны, мы будем гово-рить о нестрогом диагональном преобладании в случае выполнения неравенств**

|  |  |  |
| --- | --- | --- |
| **X** | **(3.5)** |  |
| |a**ii**| ≥|a**ij** | |  |

**j*=***6**i**

**для любого** i= 1,2, . . . , n**. Иногда при этом необходимо уточнять, что речь идёт о диагональном преобладании ¾по строкам¿, поскольку име-ет также смысл диагональное преобладание ¾по столбцам¿, которое определяется совершенно аналогичным образом.**

Теорема 3.1.2 **(признак Адамара)** **Матрица с диагональным преоб-ладанием неособенна.**

Доказательство. **Предположим, что, вопреки доказываемому, рас-сматриваемая матрица** A= (a**ij** ) **является особенной. Тогда для неко-торого ненулевого вектора** y∈R**n,** y= (y***1***, y***2***, . . . , y**n**)>**, выполняется равенство** Ay= 0**, т. е.**

**n**

**X**

a**ij** y**j** = 0, i = 1, 2, . . . , n. **(3.6)**

**j*=1***

**Выберем среди компонент вектора** y **ту, которая имеет наиболь-шее абсолютное значение. Пусть она имеет номер** l**, так что** |y**l**|=max***1***≤**j**≤**n** |x**j** |**, причём в силу сделанного выше предположения** |y**l**| > 0**.** **Следствием** l**-го из равенств (3.6) является соотношение**

**X**

−a**ll**y**l** = a**lj** y**j** , **j*=***6**l**

|  |  |
| --- | --- |
| 3.1. Задачи вычислительной линейной алгебры | **127** |

**что влечёт**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  | **X** | |  |  |
| |a**ll**| |y**l**| = | **X** | |  |  |  |  |
|  | a**lj** y**j** | |  | ≤ | |a**lj** | |y**j** | |  |  |
|  |  |  |  |  |  |  |
|  |  | **j**6***=*l** |  |  | **j**6***=*l** |  |  |  |
| ≤ |  | | **j** | | | |a**lj** | ≤ |y**l**| | **j**6***=*l** |a**lj** |. |  |
| ***1***≤**j**≤**n** | | **j**6***=*l** |  |
|  |  | max | y |  | **X** |  | **X** |  |
|  |  |  |  |  |  |  |

**Сокращая теперь обе части полученного неравенства на** |y**l**|>0**, будем**

**иметь**

**X**

|a**ii**| ≤ |a**ij** |,

**j*=***6**i**

**что противоречит наличию диагонального преобладания в матрице** A **по условию теоремы. Итак,** A **действительно должна быть неособенной матрицей.**

**Доказанный выше результат часто именуют также ¾теоремой Леви-Деспланка¿ (см., к примеру, [39]), но мы придерживаемся в этом пунк-те терминологии, принятой в [9, 32]. В книге М. Пароди [32] можно прочитать, в частности, некоторые сведения об истории вопроса.**

**Внимательное изучение доказательства признака Адамара показы-вает, что в нём нигде не использовался факт принадлежности элемен-тов матрицы и векторов какому-то конкретному числовому полю** R **или** C**. Таким образом, признак Адамара справедлив и для комплекс-ных матриц. Кроме того, он может быть обобщен на матрицы, удовле-творяющие нестрогому диагональному преобладанию (3.5).**

**Вещественная или комплексная** n×n**-матрица** A= (a**ij** ) **называ-ется разложимой , если существует разбиение множества** {1,2, . . . , n} **первых** n **натуральных чисел на два непересекающихся подмножества** I **и** J**, таких что** a**ij** = 0 **при** i ∈ I **и** j ∈ J**. Эквивалентное определе-ние: матрица** A∈R **n**×**n разложима, если путём перестановок строк и столбцов она может быть приведена к блочно-треугольному виду**

**!**

A***11*** A***12***

* + A***22***
* **квадратными блоками** A***11* и** A***22*. Матрицы, не являющиеся разложи-мыми, называются неразложимыми. Важнейший пример неразложи-мых матриц это матрицы, все элементы которых не равны нулю, в частности, неотрицательны.**

**128** 3. Численные методы линейной алгебры

**Обобщением признака Адамара является**

Теорема 3.1.3 **(теорема Таусски)** **Если для квадратной неразложи-мой матрицы** A **выполнены условия нестрогого диагонального преоб-ладания (3.5), причём хотя бы одно из этих неравенств выполнено строго, то матрица** A **неособенна.**

**Доказательство можно найти, к примеру, в [9].**

1. Нормы векторов и матриц

**3.2а** **Векторные нормы**

**Норму можно рассматривать как обобщение на многомерный и аб-страктный случаи понятия абсолютной величины числа. Вообще, и норма, и абсолютная величина являются понятиями, которые форма-лизуют интуитивно ясное свойство ¾размера¿ объекта, его ¾величины¿, т. е. того, насколько он мал или велик безотносительно к его располо-жению на прямой, плоскости или в пространстве. Такова, например, длина вектора как направленного отрезка в привычном нам трёхмер-ном евклидовом пространстве** R***3*.**

**Формальное определение даётся следующим образом.**

Определение 3.2.1 Нормой **в вещественном или комплексном ли-нейном пространстве** X **называется функция** k · k:X→R***+*, удо-влетворяющая следующим свойствам (аксиомам нормы):**

**(ВН1)** kak ≥0 **для любого** a∈X**, причём** kak= 0⇔a= 0

**неотрицательность,**

**(ВН2)** kα ak=|α| · kak **для любых** a∈X **и** α∈R **или** C

**абсолютная однородность,**

**(ВН3)** ka+bk ≤ kak+kbk **для любых** a, b, c∈X

**¾неравенство треугольника¿.**

**Само пространство** X **называется при этом** нормированным линей-ным пространством**.**

|  |  |
| --- | --- |
| 3.2. Нормы векторов и матриц | **129** |

**Далее в качестве линейного пространства** X **у нас всюду рассмат-риваются** R**n или** C**n.**

**Не все нормы, удовлетворяющие выписанным аксиомам одинаково практичны, и часто от нормы требуют выполнения ещё тех или иных дополнительных условий. К примеру, удобно иметь дело с абсолютной нормой, значение которой зависит лишь от абсолютных значений ком-понент векторов. В общем случае норма вектора этому условию может и не удовлетворять.**

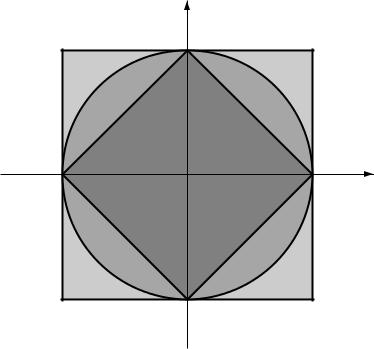
**Примеры наиболее часто используемых норм векторов в** R**n и** C**n:**

**n**

**X**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| kak***1*** = | | a**i**| **,** | |  |  |  |  |  |
| **i*=1*** | | |a**i**|***2*!*1*/*2*** |  |  |  |  |  |
| kak***2*** = | **n** | = |  | ha, ai **,** | |  |
|  | **X** |  |  | **p** |  |  |  |
|  | **i*=1*** |  |  |  |  |  |  |
| kak∞ = | max |  |  |  |  |  |  |
| ***1*** **i n** | a**i**| **.** | |  |  |  |  |  |
|  | ≤ ≤ |  |  |  |  |  |  |

**Вторая из этих норм часто называется евклидовой, а третья чебы-шёвской. Замечательность евклидовой нормы состоит в том, что она порождена скалярным произведением в** R**n. Иногда можно встретить и другие названия этих норм.**



**Рис. 3.1. Шары единичного радиуса в различных нормах.**

**Первая и вторая нормы** **это частные случаи более общей кон-**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **130** |  | 3. Численные методы линейной алгебры | |  |
| **струкции** p**-нормы** |  | |a**i**|**p!*1*/p** |  |  |
| kak**p** = | **n** | **для** p≥1. |  |
|  | **X** | |  |  |

**i*=1***

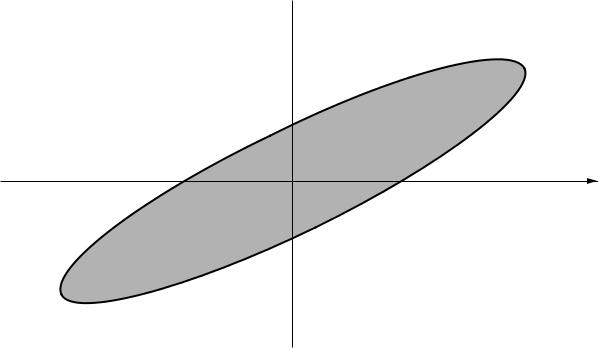
**Чебышёвская норма также может быть получена из** p**-нормы с помо-щью предельного перехода** p→ ∞**, что и объясняет индекс ¾**∞**¿ в её обозначении.**

**Ещё одной важной и популярной конструкцией нормы является так называемая энергетическая норма векторов, которая порождается какой-либо симметричной положительно-определённой матрицей (эр-митовой в комплексном случае). Если** A **такая матрица, то выра-жение** hAx, yi**, как нетрудно проверить, есть симметричная билиней-ная положительно-определённая форма, т. е. скалярное произведение**

**векторов** x **и** y**. Следовательно, можно определить норму вектора** x**, p**

**как** hAx, xi**, т. е. как корень из произведения** x **на себя в этом новом скалярном произведении. Это и есть энергетическая норма** x **относи-тельно матрицы** A**, обычно обозначаемая как** kxk**A** :=hAx, xi***1*/*2*. Её часто называют также** A**-нормой векторов, если в задаче имеется в виду какая-то конкретная симметричная положительно определённая матрица** A**.**

 x***2***



x***1***

**Рис. 3.2. Шар единичного радиуса в энергетической норме при большом разбросе спектра порождающей матрицы**

**Характерная особенность энергетической нормы, которая в ряде случаев оборачивается её недостатком, возможность существенного искажения обычного геометрического масштаба ¾величины¿ объектов.**

|  |  |
| --- | --- |
| 3.2. Нормы векторов и матриц | **131** |

**Она вызывается разбросом собственных значений порождающей мат-рицы** A **и приводит к тому, что векторы из** R**n, имеющие одинаковую энергетическую норму, существенно различны по обычной евклидовой длине (Рис. 3.2). С другой стороны, использование энергетической нор-мы, которая порождена матрицей, фигурирующей в постановке задачи (системе линейных алгебраических уравнений, задаче на собственные значения и т. п.) часто является удобным и оправданным, альтернати-вы ему очень ограничены. Примеры будут рассмотрены в §3.5б и §3.5в.**

**Нормы будут нужны нам как сами по себе, для оценивания ¾ве-личины¿ тех или иных объектов, так и для измерения ¾отклонения¿ одного вектора от другого. Как уже отмечалось ранее, на нормиро-ванном пространстве** X **расстояние (метрика) может быть естественно заданы как**

dist (a, b) = ka − bk.

**В этом смысле задание на некотором линейном пространстве** X **нормы автоматически определяет на нём и топологию, т. е. запас открытых и замкнутых множеств, структуру близости, с помощью которой можно будет, в частности, выполнять предельные переходы. Более точно, мы считаем, что переменная** a∈X **сходится к пределу** a∗ **относительно рассматриваемой нормы, если** ka−a∗k →0**.**

**3.2б** **Матричные нормы**

**Помимо векторов основным объектом вычислительной линейной алгеб-ре являются также матрицы, и потому нам будут нужны матричные нормы для того, чтобы оценивать ¾величину¿ той или иной матри-цы, а также для того, чтобы измерять ¾отклонение¿ одной матрицы от другой как**

|  |  |
| --- | --- |
| dist (A, B) := kA − Bk. | **(3.7)** |

**Множество матриц само является линейным (векторным) простран-ством, а матрица это составной многомерный объект, во многом аналогичный вектору. Поэтому вполне естественно прежде всего по-требовать от матричной нормы тех же свойств, что и для векторной нормы. Формально, матричной нормой мы будем называть функцию** k ·k : R**m**×**n** → R***+*** **или** k ·k : C**m**×**n** → R***+*, удовлетворяющую следующим** **аксиомам:**

**(МН1)** kAk ≥0 **для любой матрицы** A**, причём** kAk= 0⇔A= 0

**неотрицательность,**

**132** 3. Численные методы линейной алгебры

**(МН2)** kα Ak=|α| · kAk **для любых матрицы** A **и** α∈R **или** C

**абсолютная однородность,**

**(МН3)** kA+Bk ≤ kAk+kBk **для любых матриц** A, B, C

**¾неравенство треугольника¿.**

**Эти три условия выражают взгляд на матрицу, как на ¾вектор раз-мерности** m×n**¿, и должны быть признаны недостаточными, если мы хотим учесть специфику матриц как объектов, между которыми определена также операция умножения. В частности, множество всех квадратных матриц фиксированного размера наделено более богатой структурой, нежели линейное векторное пространство, и обычно в свя-зи с ним используют уже термин ¾кольцо¿, обозначающее множество с двумя взаимносогласованными бинарными операциями сложением и умножением.**

**Свойства матриц, как элементов кольца, обычно отражает четвёр-тая аксиома матричной нормы:**

**(МН4)** kABk ≤ kAk · kBk **для любых матриц** A, B

**¾субмультипликативность¿.**

**Она фиксирует связь нормы матриц с операцией их умножения и ино-гда формулируется расширительно для всех матриц, которые можно перемножать друг с другом.**

**Особый интерес и в теории и на практике представляют ситуации, когда нормы векторов пространства** R**n или** C**n и нормы матриц, дей-ствующих на этом пространстве, существуют не сами по себе, но в некотором смысле согласованы друг с другом. Инструментом такого согласования может как раз-таки выступать аксиома субмультиплика-тивности, понимаемая в расширенном смысле, т. е. для любых матриц** A **и** B **таких размеров, что произведение** AB **имеет смысл. В частности,** **для** n×1**-матриц** B**, являющихся векторами из** R**n.**

Определение 3.2.2 **Векторная и матричная нормы называются** со-гласованными**, если для любой матрицы** A ∈ R**m**×**n** **( или** C**m**×**n)**

|  |  |
| --- | --- |
| kAxk ≤ kAk · kxk | **(3.8)** |

**для всех** x∈R**n ( или** C**n).**

|  |  |
| --- | --- |
| 3.2. Нормы векторов и матриц | **133** |

**Рассмотрим примеры конкретных матричных норм. Фробениусова норма матрицы** A= (a**ij** ) **определяется как**

|  |  |  |  |
| --- | --- | --- | --- |
| kAk**F** = | |a|**ij*2*** | **!** . |  |
|  | **X** | ***1*/*2*** |  |
|  |  |  |
|  | **i,j** |  |  |

**Ясно, что она удовлетворяет первым трём аксиомам матричной нор-мы просто потому, что задаётся совершенно так же, как евклидова векторная норма. Свойство субмультипликативности устанавливается несложной проверкой. Кроме того, фробениусова норма согласована с евклидовой векторной нормой** k · k***2*. Матричная норма**

kAk***max*** = n max |a**ij** |,

**i,j**

**определённая на множестве квадратных** n×n**-матриц, согласована с евклидовой и чебышёвской нормами векторов.**

**Оказывается, что для любой нормы квадратных матриц можно по-добрать подходящую векторную норму, с которой она будет согласо-вана. Иными словами, среди матричных норм квадратных матриц нет таких, которые не были бы ни с чем согласованными. Действительно, если для данной матричной нормы** k · k **определить норму** kvk **вектора** v **как** k(v, v, . . . , v)k**, т. е. как норму матрицы** (v, v, . . . , v)**, составленной** **из** n **штук векторов** v **как из столбцов, то**

kAvk = k(Av, Av, . . . , Av)k = kA · (v, v, . . . , v)k ≤ kAk · k(v, v, . . . , v)k = kAk · kvk,

**так что требуемое согласование действительно будет достигнуто.**

**3.2в** **Подчинённые матричные нормы**

**В предшествующем пункте мы могли видеть, что с заданной вектор-ной нормой могут быть согласованы различные матричные нормы. И наоборот, для матричной нормы возможна согласованность со многи-ми векторными нормами. В этих условиях при проведении различных выкладок и выводе оценок наиболее выгодно оперировать как можно меньшими из согласованных матричных норм: огрубить оценку (3.8) всегда можно, а вот сделать её более точной гораздо труднее! Конкрет-ная величина нормы, к примеру, может оказать сильное влияние на**

**134** 3. Численные методы линейной алгебры

**количество итераций, которые мы вынуждены будем сделать в итера-ционных численных методах для достижения той или иной расчётной точности.**

**Ясно, что для фиксированной матрицы** A **и заданной векторной нормы** k · k **значения всех согласованных с ней матричных норм от** A

**ограничены снизу выражением**

sup kAxk ,

**x*=0***6 kxk

**поскольку из требования согласованности вытекает** kAk ≥ kAxk/kxk**.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Предложение 3.2.1 **Соотношением** | | | | |  |  |
| A | k | 0 | = sup | kAxk | **(3.9)** |  |
| k |  |  | x |  |  |
|  |  |  | **x**6***=0*** k k | |  |  |

**задаётся матричная норма.**

Доказательство. **Отметим, прежде всего, что вместо ¾**sup**¿ в выра-жении (3.9) можно брать ¾**max**¿, так как**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **x**6***=0*** kkxkk | | = **x**6***=0*** |  | |  |  | k**y**k***=1*** k |  | k |  |  |
| A kxk | |  | = | Ay | , |  |
| sup Ax | | sup |  | x |  | sup |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |

**а единичная сфера любой нормы компактна в конечномерном про-странстве** R**n [38, 48]. Непрерывная функция** kAyk **достигает на этом компактном множестве своего максимума.**

**Проверим теперь для нашей конструкции выполнение аксиом нор-мы. Если** A6= 0**, то найдётся ненулевой вектор** y**, такой что** Ay6= 0**. Ясно, что его можно считать нормированным, т. е.** kyk= 1**. Тогда** kAyk > 0**, и потому** maxk**y**k***=1*** kAyk > 0**, что доказывает первую ак-сиому нормы.**

**Абсолютная однородность доказывается тривиально, покажем для (3.9) справедливость неравенства треугольника. Очевидно,**

k(A + B)yk ≤ kAyk + kByk,

**и потому**

max k(A + B)yk ≤ max kAyk + max kByk.

k**y**k***=1*** k**y**k***=1*** k**y**k***=1***

|  |  |
| --- | --- |
| 3.2. Нормы векторов и матриц | **135** |

**Приступая к обоснованию субмультипликативности, отметим, что по самому построению** kAxk ≤ kAk0kxk **для любого вектора** x**. По этой причине**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| k | AB | 0 | = max | |  | (AB)y | | k | = | k | (AB)˜y | | |  | **для некоторого** y˜ **с** | | | | | | | y˜ = 1 |  |
| k |  | **y** | ***=1*** k | |  |  |  |  |  | k | |  |  |  |  |  |  |  | k k |  |
|  |  |  | k k |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  | A | 0 | · k | | By˜ |  | A | |  | max | | k | By | k | = | A 0 | k | B | 0. |  |  |
|  |  |  | ≤ k k | | k ≤ k k · | | | | | **y** | ***=1*** |  |  | k k | k |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  | k k |  |  |  |  |  |  |  |  |  |  |  |
| **Это завершает доказательство предложения.** | | | | | | | | | | | | | | | | | |  |  |  |  |  |  |

**Доказанный результат мотивирует**

Определение 3.2.3 **Матричная норма, определяемая для заданной** **векторной нормы** k · k **на линейном векторном пространстве** X **как**

kAk0 = sup kAxk = sup kAyk,

**x*=0***6 kxk k**y**k***=1***

**называется** подчинённой **к** k · k **матричной нормой (или** индуцирован-ной**, или** операторной **нормой).**

**Последний термин операторная норма, популярен потому, что конструкция этой нормы хорошо отражает взгляд на матрицу как на оператор, задающий отображения линейных векторных пространств**

R**n** → R**m** **или** C**n** → C**m.**

**Какими являются подчинённые матричные нормы для наших век-торных норм** k · k***1*,** k · k***2* и** k · k∞**? Вообще, существуют ли простые вы-ражения для этих подчинённых норм? Являются ли матричные нормы** kAk**F** **(фробениусова) и** kAk***max*** **подчинёнными для каких-либо вектор-ных норм?**

**Ответ на последний вопрос отрицателен. В самом деле, для единич-**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ной** n×n**-матрицы** I **имеем** | √ |  |  |  |
| kIk***max*** = n, |  |  |  |
| kIk**F** = n, | | |  |

**тогда как из определения подчинённой нормы следует, что должно быть**

kIk = sup kIyk = max kyk = 1.

k**y**k***=1*** k**y**k***=1***

**Ответом на первые вопросы является**

**136** 3. Численные методы линейной алгебры

Предложение 3.2.2 **Для векторной** 1**-нормы подчинённой матрич-ной нормой для** m×n**-матриц является**

|  |  |
| --- | --- |
| kAk***1*** | = ***1***≤**j**≤**n i*=1*** |a**ij** |**!** |
|  | **m** |
|  | **X** |
|  | max |

**максимальная сумма модулей элементов по столбцам.**

**Для чебышёвской векторной нормы (**∞**-нормы) подчинённой мат-ричной нормой для** m×n**-матриц является**

|  |  |  |
| --- | --- | --- |
| kAk∞ = ***1*** **i m** | | | **ij** |**!** |
|  | **n** |  |
|  | **X** |  |
| max | a |  |
| ≤ ≤ | **j*=1*** |  |

**максимальная сумма модулей элементов по строкам. Матричная норма, подчинённая евклидовой норме векторов** kxk***2*,**

**есть** kAk***2*** =σ***max***(A) **наибольшее сингулярное число матрицы** A**.**

Доказательство. **Первая часть предложения обосновывается следу-ющей выкладкой:**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| kAxk***1*** = | **i*=1*** |(Ax)**i**| = | | | **i*=1*** | **j*=1*** a**ij** x**j** | |  | ≤ | | **i*=1*** **j*=1*** | a**ij** x**j** | | | |  |  |  |  |
|  | **m** |  |  | **m** | **n** |  |  |  |  | **m** | **n** |  |  |  |  |  |
|  | **X** |  |  | **X X** | |  |  |  | **X X** | |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | **m n** |  |  |  | **n** | **m** |  |  |  |  | **n** | | x**j** | | **m** | | a**ij** | |  |  |
| = **i*=1*** **j*=1*** | a**ij** | | x**j** | = | | | | | **j*=1*** **i*=1*** | a**ij** | | x**j** | = **j*=1*** | | | | | | | **i*=1*** |  |
|  | **X X** |  |  | **ij** |**!** · | **X X** | |  |  |  |  | **X** |  | **X** | |  |  |
| ≤ | ***1* j n** | | |  | | | **j** | |  |  | k k***1*** k k***1*** | | |  |  |  |  |  |
|  |  | **m** |  |  | **n** |  |  |  |  |  |  |  |  |  |  |  |
|  | max | **X** | a |  | **X** | x | = | |  | A | x , |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  | ≤ ≤ | **i*=1*** |  |  | **j*=1*** |  |  |  |  |  |  |  |  |  |  |  |

**причём все неравенства в этой цепочке достижимы на векторе** x **в виде столбца единичной матрицы с тем номером** j**, на котором достигается**

**Pm**

max**j** **i*=1***|a**ij** |**. Аналогичным образом доказывается и вторая часть**

**предложения.**

**Для доказательства последней части предложения рассмотрим** n×n**-матрицу** A>A**. Она симметрична, её собственные числа веществен-ны и неотрицательны, будучи квадратами сингулярных чисел матрицы** A**. Кроме того, ортогональным преобразованием подобия** A>A **может** **быть приведена к диагональному виду:** A>A=Q>ΛQ**, где** Q **ортого-нальная** n×n**-матрица,** Λ **диагональная** n×n**-матрица, имеющая на диагонали собственные значения** A>A**, т. е. числа** σ**i*2***(A)**,** i= 1,2, . . . , n**.**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 3.2. Нормы векторов и матриц | | | | | | | | | |  |  |  |  |  |  |  |  |  |  |  |  |  | **137** | | | |  |
| **Далее имеем** | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  | √ |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| k k***2*** | **x**6***=0*** | k x ***2*** | | |  | **x**6***=0*** | | | |  | x ***2*** |  |  |  | **x**6***=0*** | | **p** | | x | ***2*** | | |  |  |  |  |  |
| A | = max | Axk***2*** | | | = max | | | | |  | x>A>Ax | |  | = max | | |  | x>Q>ΛQx | | | | | | | | |  |
|  |  | |  |  | |  | k | k |  |  |  |  |  |  |  |  |
|  |  | k | k |  |  |  |  |  |  |  | k k |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | **p** |  |  | |  |  |  |  |  |  | **p** | |  |  |  |  | | **s** | |  |  |  | |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | **P** | **i** y**i*2*** | | | |  |
|  | **x**6***=0*** |  | Qx ***2*** | | |  |  |  |  | **y**6***=0*** | y | ***2*** |  | **y**6***=0*** | |  |
|  |  |  | (Qx)>Λ(Qx) | | | | | | |  |  |  |  | y>Λy | | |  |  |  |  |  | **i** | σ**i*2***y**i*2*** | | | |  |
|  | = max |  |  |  |  |  |  |  |  | = max | |  |  |  |  | = max | | |  |  |  |  |  |  |  |  |  |
|  |  | k |  | k | |  |  |  | k k | | |  |  |  |  | **P** | | | | |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | ≤ **y**6***=0*** |  |  |  | **s** |  | | | **!** | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  | **Pi**y**i*2*** | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  | **~~i~~** | y***2*** | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | max | σ***max*** | | |  | **~~i~~** | | |  | = σ***max***, | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | **~~P~~** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

**где мы воспользовались в выкладках тем, что** kQxk***2*** =kxk***2* для любой ортогональной матрицы** Q**. С другой стороны, эта оценка достижима:**

|  |  |
| --- | --- |
| **n** |  |
| **достаточно взять в качестве вектора** y∈R***2*** | **столбец единичной мат-** |
| **рицы с номером, равным месту элемента** σ***max* на диагонали в** Λ**, а в** | |
| **самом начале выкладок положить** x=Q>y**.** |  |

**Норму матриц** k · k***2*, подчинённую евклидовой векторной норме, часто называют также спектральной нормой матриц. Отметим, что она не является абсолютной нормой (см. Пример 3.1.1), т. е. зависит не только от абсолютных значений элементов матрицы. В то же время,** k · k***1*** **и** k · k∞ **это абсолютные матричные нормы, что следует из вида** **их выражений.**

**3.2г** **Топология на множествах векторов и матриц**

**Термином топология для заданного множества обычно обозначают се-мейство всех открытых подмножеств данного множества, позволяю-щее определять окрестности и, как следствие, близость одно элемента множества к другому, сходимость и т.п. понятия.**2 **Топологию можно задавать различными способами, например, простым описанием того, какие именно множества считаются открытыми. Наиболее популярно задание топологии с помощью функции расстояния (метрики) или же с помощью различных норм.**

2 **Топологией называется также математическая дисциплина, изучающая, глав-ным образом, свойства объектов, инвариантные относительно непрерывных отоб-ражений.**

**138** 3. Численные методы линейной алгебры

**Говорят, что векторные нормы топологически эквивалентны (или просто эквивалентны), если эквивалентны порождаемые ими в про-странстве топологии, т. е. любое открытое (замкнутое) относительно одной нормы множество является открытым (замкнутым) и в другой норме, и наоборот. При этом, в частности, предельный переход в одной норме влечёт существование предела в другой, и обратно. Из матема-тического анализа известен простой критерий эквивалентности двух норм:**

Предложение 3.2.3 **Нормы** k · k0 **и** k · k00 **на пространстве** X **экви-валентны тогда и только тогда, когда существуют положительные константы** C***1* и** C***2*, такие что для любых** a∈X

|  |  |
| --- | --- |
| C***1***kak0 ≤ kak00 ≤ C***2***kak0. | **(3.10)** |

**Формулировка этого предложения имеет кажущуюся асимметрию, так как для значений одной из эквивалентных норм предъявляется дву-сторонняя ¾вилка¿ из значений другой нормы с подходящими множи-телями-константами. Но нетрудно видеть, что из (3.10) немедленно сле-дует**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1 | kak00 ≤ kak0 ≤ | 1 | kak00, |  |
|  |  |  |
| C***2*** | C***1*** |  |

**так что существование ¾вилки¿ для одной нормы автоматически под-разумевает существование аналогичной ¾вилки¿ и для другой.**

Предложение 3.2.4 **В линейных пространствах** R**n** **или** C**n**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| kak***2*** ≤ kak***1*** ≤ | √ |  |  |  |
| √ | n kak***2***, | |  |
| kak∞ ≤ kak***2*** ≤ |  |  |  |
|  | n kak∞, | |  |

1

n kak***1*** ≤ kak∞ ≤ kak***1***,

**т. е. все три рассмотренные выше векторные нормы эквивалентны друг другу.**

Доказательство. **Справедливость правого из первых неравенств сле-дует из известного неравенства Коши-Буняковского** |ha, bi| ≤ kak***2*** kbk***2*, применённого к случаю** b= (sgna***1***,sgna***2***, . . . ,sgna**n**)>**. Для обосно-вания левого из первых неравенств заметим, что в силу определений**

|  |  |
| --- | --- |
| 3.2. Нормы векторов и матриц | **139** |

**2-нормы и 1-нормы**

kak***22*** = |a***1***|***2*** + |a***2***|***2*** + . . . + |a**n**|***2***, kak***21*** = |a***1***|***2*** + |a***2***|***2*** + . . . + |a**n**|***2***

* + 2 |a***1***a***2***| + 2 |a***1***a***3***| + . . . + 2 |a**n**−***1***a**n**|,
* **все слагаемые** 2|a***1***a***2***|**,** 2|a***1***a***3***|**, . . . ,** 2|a**n**−***1***a**n**| **неотрицательны. В част-ности, равенство** kak***22*** =kak***12* и ему равносильное** kak***22*** =kak***12* возмож-ны лишь в случае, когда у вектора** a **все компоненты равны нулю за исключением одного.**

**Обоснование остальных неравенств даётся следующими несложны-ми выкладками:**

**p**

kak***2*** = |a***1***|***2*** + |a***2***|***2*** + . . . + |a**n**|***2*** **p**

* max**i** |a**i**|***2*** = max**i** |a**i**| = kak∞,

**p**

kak***2*** = |a***1***|***2*** + |a***2***|***2*** + . . . + |a**n**|***2***

≤ **p**n max**i** |a**i**|***2*** = √n max**i** |a**i**| = √n kak∞, kak∞ = max**i** |a**i**| ≤ |a***1***| + |a***2***| + . . . + |a**n**| = kak***1*** kak***1*** = |a***1***| + |a***2***| + . . . + |a**n**| ≤ n max**i** |a**i**| ≤ n kak∞

**Нетрудно видеть, что все эти неравенства достижимые (точные).**

**В действительности, этот факт является частным случаем общего результата анализа о том, что в конечномерном векторном простран-стве все нормы топологически эквивалентны друг другу.**

**Коль скоро векторы являются сложными объектами, составленны-ми из своих компонент, то помимо определённой выше сходимости по норме имеет смысл рассматривать покомпонентную сходимость, при которой один вектор сходится к другому тогда и только тогда, когда все компоненты первого вектора сходятся к соответствующим компо-нентам второго:**

a → a∗ ⇔ a**k** → a∗**k** **для всех индексов** k = 1, 2, . . . , n.

**Интересен вопрос о том, как соотносятся между собой сходимость по норме и сходимость всех компонент вектора. Нетрудно показать, что**

**140** 3. Численные методы линейной алгебры

**в случае конечномерных пространств эти понятия равносильны друг другу.**

**Совершенно аналогично случаю векторных норм можно рассмот-реть понятие эквивалентности норм матриц, и критерием эквивалент-ности матричных норм будет служить Предложение 3.2.3. В силу из-вестного факта из математического анализа в конечномерном линей-ном пространстве всех** m×n**-матриц все нормы эквивалентны. Тем не менее, конкретные константы эквивалентности играют огромную роль при выводе различных оценок, и их значения даёт следующее**

Предложение 3.2.5 **Во множествах матриц** R**n**×**n** **и** C**n**×**n**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | | |  |  |  |  | √ | |  |  |  |
| ~~√~~ | |  |  | kAk***2*** | ≤ kAk***1*** | ≤ |  | n kAk***2***, | | |  |
| ~~n~~ | |  |
| 1 | |  |  |  |  |  | √ |  | |  |  |
| ~~√~~ |  | | kAk∞ | | ≤ kAk***2*** | ≤ |  | n kAk∞, | | |  |
| ~~n~~ | |  |

1

n kAk***1*** ≤ kAk∞ ≤ n kAk***1***.

Доказательство. **Докажем первую пару неравенств:**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| kAk***2*** | = k**y**k***2=1*** kAyk***2*** | | | ≤ k**y**k***2*** | | ***=1*** |  |  |  | kAyk***2*** |  |
|  |  | n |  |
|  | max |  |  | max | | | √ | | | |  |
| **Кроме того,** |  |  |  |  |  |  |  |  |  |  |  |
|  | max | | |  | max | | | kAe**i**k***2*** | | |  |
| kAk***2*** = **y** | | k | ***2=1*** kAyk***2*** ≥ | |  | **i** |  |  |
|  | k |  |  |  |  |  |  |  |  |  |
| **где** e**i** i**-ый столбец единичной матрицы.** | | | | | | |  |  |  |  |  |

**Из эквивалентности матричных норм следует, в частности, суще-ствование для любой нормы** k · k **такой константы** C**, что**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **i,j** | **ij** | ≤ | ***1* j n** | | | **ij** |**!** | k | k***1*** ≤ | k k |
|  |  | **m** |  |  |  |  |
|  |  | **X** |  |  |  |  |
| max a | max | a |  | = | A | C A |
|  | ≤ ≤ | **i*=1*** |  |  |  |  |

**(вместо 1-нормы матриц в этой выкладке можно было бы взять, к при-меру,** ∞**-норму). Поэтому** |a**ij** | ≤CkAk**, так что сходимость последова-тельности матриц в любой норме равносильна сходимости последова-тельностей всех элементов этих матриц.**

|  |  |
| --- | --- |
| 3.2. Нормы векторов и матриц | **141** |

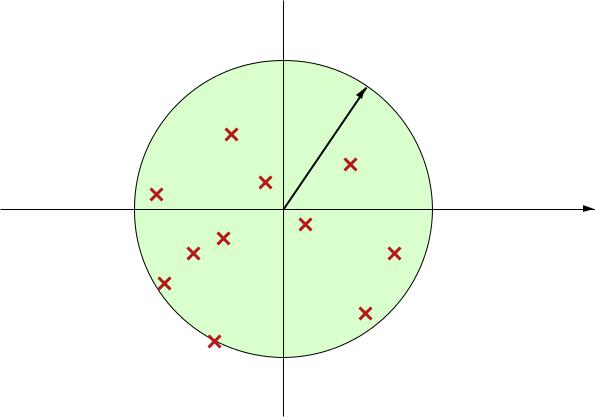
**В целом множество матриц с введённым на нём посредством (3.7) расстоянием для любой матричной нормы является полным метриче-ским пространством, т. е. любая фундаментальная (¾сходящаяся в се-бе¿) последовательность имеет в нём предел. Это следует из предше-ствующего рассуждения и из факта полноты вещественной оси** R **и комплексной плоскости** C**.**

**3.2д** **Спектральный радиус**

Определение 3.2.4 **Спектральным радиусом квадратной матрицы** **называется наибольший из модулей её собственных чисел.**

**Эквивалентное определение: спектральным радиусом матрицы на-зывается наименьший из радиусов кругов комплексной плоскости** C **с центром в начале координат, который содержит весь спектр матрицы. Эта трактовка хорошо объясняет и сам термин. Обычно спектральный радиус матрицы** A **обозначают** ρ(A)**.**

 Im



ρ

|  |  |
| --- | --- |
| 0 | Re |

**Рис. 3.3. Иллюстрация спектрального радиуса матрицы: крестиками обозначены точки спектра.**

Предложение 3.2.6 **Спектральный радиус матрицы не превосходит** **любой её нормы.**

Доказательство. **Рассмотрим сначала случай, когда основным линей-ным пространством является** C**n, и рассматриваемая** n×n**-матрица** A**, вообще говоря, также комплексная.**

**142** 3. Численные методы линейной алгебры

**Пусть** λ **собственное значение матрицы, а** v=60 **соответствую-щий собственный вектор** A**, так что** Av=λv**. Воспользуемся тем уста-новленным в §3.2б фактом, что любая матричная норма согласована с некоторой векторной нормой, и возьмём от обеих частей равенства** Av = λv **норму, согласованную с рассматриваемой** kAk**. Получим**

|  |  |
| --- | --- |
| kAk · kvk ≥ kAvk = kλvk = |λ| · kvk, | **(3.11)** |

**где** kvk ≥0**, и потому сокращение на эту величину обеих частей нера-венства (3.11) даёт** kAk ≥ |λ|**. Коль скоро наше рассуждение справед-ливо для любого собственного значения, то в самом деле** ρ(A)≤ kAk**.**

**Рассмотрим теперь случай вещественной матрицы** A**, действующей на** R**n. Если** λ **её вещественное собственное значение, то проведён-ные выше рассуждения остаются полностью справедливыми. Если** λ **комплексное собственное значение матрицы** A**, то комплексифици-руем рассматриваемое линейное пространство, т. е. перейдём от** R **к пространству** R⊕iR**, где** i **мнимая единица, а ¾**⊕**¿ означает прямую сумму линейных пространств (см. [10, 33]).**

**Элементами** R⊕iR **служат упорядоченные пары** (x, y)>**, где** x**,** y∈R**. Сложение и умножение на скаляр** (a+ ib)∈C **определяются для них следующим образом**

(x, y)> + (x0, y0)> = (x + x0, y + y0)>,

(a + i b) · (x, y)> = (ax − by, ay + bx)>.

**Введённые пары векторов** (x, y)> **обычно записывают в виде** x+iy**, при-чём** x **и** y **называются соответственно вещественной и мнимой частями вектора из** R⊕iR**. Оператор, действующий на** R⊕iR **и порождаемый матрицей** A**, может быть представлен в матричном виде как**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| A = | 0 | A | . | **(3.12)** |
|  | A | 0 |  |  |

**Его диагональность объясняется тем, что вещественная матрица** A

**независимо действует на вещественную и мнимую части векторов из**

R ⊕ iR**.**

**Без ущерба для общности можно считать, что рассматриваемая на-ми норма матрицы, т. е.** kAk**, является подчинённой (операторной) нор-мой, так как такие нормы являются наименьшими из всех согласован-ных матричных норм. Пусть** k · k **векторная норма в** R**, которой под-чинена наша матричная норма. Зададим в** R⊕iR **норму векторов как**

|  |  |
| --- | --- |
| 3.2. Нормы векторов и матриц | **143** |

k(x, y)>k = kxk + kyk**. Тогда ввиду (3.12) и с помощью рассуждений,** **аналогичных доказательству Предложения 3.2.2, нетрудно показать, что подчинённая матричная норма для** A **во множестве комплексных матриц есть** kAk= max{kAk,kAk}=kAk**. Кроме того, теперь для** A

**справедливы рассуждения, проведённые выше в случае комплексной матрицы.**

**Спектральный радиус в общем случае не является матричной нор-мой. Хотя для любого скаляра**

ρ(αA) = |α| ρ(A),

**т. е. спектральный радиус обладает абсолютной однородностью, акси-омы (МН1) и (МН3) матричной нормы для него не выполняются. Во-первых, для ненулевой матрицы**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 0 | 0 | 1 0 |  |  |  |
|  |  | 1 |  |  |  |  |
|  |  |  | **.. . .. .** |  | **(3.13)** |  |
|  |  |  | 0 1 |  |  |  |
|  |  | 0 |  |  |  |  |
|  |  | 0 |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

**жордановой клетки, отвечающей собственному значению** 0**, спек-тральный радиус равен нулю. Во-вторых, если** A **матрица вида (3.13), то** ρ(A>) =ρ(A) = 0**, но** ρ(A+A>)>0**, и потому неверно неравенство треугольника**

ρ( A + A>) ≤ ρ(A) + ρ(A>).

**Тем не менее, для симметричных матриц спектральный радиус есть норма, которая совпадает со спектральной нормой** k · k***2*. Это было до-казано в Предложении 3.2.2.**

**Менее очевидная связь между спектральным радиусом и нормами матриц выражается формулой Гельфанда**

**n** ***1*/n**

ρ(A) = lim A ,

**n**→∞

**которая справедлива для любой из матричных норм. Доказательство можно найти, к примеру, в [48]. Формула Гельфанда показывает, в част-ности, что с помощью спектрального радиуса адекватно описывается асимпототическое поведение норм степеней матрицы. Это подтвержда-ет и следующий результат.**

**144** 3. Численные методы линейной алгебры

Предложение 3.2.7 **Для сходимости степеней** A**k** → 0 **при** k → ∞ **необходимо, чтобы** ρ(A)<1**, т. е. чтобы спектральный радиус мат-рицы** A **был меньше** 1**.**

Доказательство. **Пусть** λ **собственное число матрицы** A **(возможно,** **комплексное), а** v=60 **соответствующий ему собственный вектор (который также может быть комплексным). Тогда** Av=λv**, и потому**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| A***2***v = A(Av) = A(λv) = λ(Av) = λ***2***v, | | |  |  |
| A***3***v = A(A***2***v) = A(λ***2***v) = λ***2***(Av) = λ***3***v, | | |  |  |
| . . . | . . . | , |  |  |
| **так что в целом** |  |  |  |  |
|  | A**k** | v = (λ**k**)v. | **(3.14)** |  |
| **Если последовательность** | **степеней** A**k,** k= 0,1,2, . . .**, сходится к нуле-** | | |  |
|  |  |  |  |

**вой матрице, то нулевой предел имеет и вся левая часть выписанного равенства. Как следствие, к нулевому вектору должна сходиться и пра-вая часть в (3.14). Это возможно лишь в случае** |λ|<1**.**

**Ниже в §3.4а мы увидим, что условие** ρ(A)<1 **является, в дей-ствительности, также достаточным для сходимости к нулю степеней матрицы** A**.**

**3.2е** **Матричный ряд Неймана**

**Как известно из математического анализа, операцию суммирования можно обобщить на суммы бесконечного числа слагаемых, которые на-зываются рядами. Суммой ряда называется при этом предел (если он существует) сумм конечного числа слагаемых, когда количество сла-гаемых неограниченно возрастает. Совершенно аналогичная конструк-ция применима также к суммированию векторов и матриц, а не только чисел. Именно, суммой матричного ряда**

∞

**X**

A***(*k*)***,

**k*=0***

**где** A***(*k*)*** **матрицы одного размера, мы будем называть предел частич-**

|  |  |
| --- | --- |
| **ных сумм** | **Pk*=0*** A***(*k*)*** **при** N → ∞**.** |
|  | **N** |

|  |  |
| --- | --- |
| 3.2. Нормы векторов и матриц | **145** |

Предложение 3.2.8 **Пусть** X n × n**-матрица и в некоторой мат-ричной норме** kXk<1**. Тогда матрица** (I−X) **неособенна, для обрат-ной матрицы справедливо представление**

∞

**X**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| (I | − | X)−***1*** = | |  |  | X**k**, | | |  |  |  | **(3.15)** |  |
|  |  |  | **k*=0*** | | |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
| **и имеет место оценка** |  |  |  |  |  |  |  |  |  |  |  |  |  |
| (I − X)−***1*** ≤ | | | |  |  |  | 1 |  |  | . |  | **(3.16)** |  |
|  |  | | | | |  |  |
|  | 1 − kXk | | | | |  |  |
| **Фигурирующий в (3.15)** | **аналог геометрической прогрессии для мат-** | | | | | | | | | | | |  |
| **риц называется матричным рядом Неймана.** | | | | | | | | |  |  |  |  |  |
| Доказательство. **Обозначим** S**N** = | | | |  | **N** |  | X**k** | | **частичную сумму мат-** | | | |  |
| **ричного ряда Неймана. Коль скоро** | | | **Pk*=0*** | | | |  | kX**k**k ≤ | | |  | kXk**k** |  |
| kS**N** ***+*p** − S**N** k =**N *+*p** | | X**k** | ≤ |  | **N *+*p** | | | **N *+*p** |  |
|  |  |  |  |  |  | **X** | |  |  |  | **X** |  |  |
| **X** | |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |

**k*=*N *+1*** **k*=*N *+1*** **k*=*N *+1***

=kXk**N *+1*** ·1− kXk**N *+*p** →01 − kXk

**при** N→ ∞ **и любых целых положительных** p**, то последовательность** S**N** **является фундаментальной (последовательностью Коши) в полном** **метрическом пространстве** n×n**-матриц с расстоянием, порождённым рассматриваемой нормой** k · k**. Следовательно, частичные суммы** S**N ряда Неймана имеют предел** S= lim**N** →∞S**N , причём**

(I − X)S**N** = (I − X)(I + X + X***2*** + . . . + X**N** ) = I − X**N** ***+1*** → I

**при** N→ ∞**, поскольку тогда** kX**N *+1***k ≤ kXk**N *+1*** →0**. Этот предел** S **удовлетворяет соотношению** (I−X)S=I**, и потому** S= (I−X)−***1*.**

**Наконец,**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| k(I − X)−***1***k = |  | ∞X**k** |  | ≤ | ∞kX**k**k ≤ | ∞kXk**k** = |  | 1 |  |  | , |  |
| 1 | − k | X |  |  |
|  |  | **k*=0*** |  |  | **k*=0*** | **k*=0*** |  |  | k | |  |
|  |  | **X** |  |  | **X** | **X** |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |

**где для бесконечных сумм неравенство треугольника может быть обос-новано предельным переходом по аналогичным неравенствам для ко-нечных сумм. Это завершает доказательство предложения.**

**146** 3. Численные методы линейной алгебры

**Матричный ряд Неймана является простейшим из матричных сте-пенных рядов, т. е. сумм вида**

∞

**X**

c**k**X**k**

**k*=0***

**для какого-то счётного набора коэффицентов** c**k,** k= 0,1,2, . . . **. C по-мощью матричных степенных рядов можно определять значения ана-литических функций от матриц, например, логарифм от матрицы. Мы далее не будем касаться этой важной и интересной темы, находящей многочисленные приложения: она развивается в рамках так называе-мой теории представлений линейных операторов.**

**3.2ж** **Число обусловленности матриц**

**Рассмотрим систему линейных алгебраических уравнений**

Ax = b

**с неособенной квадратной матрицей** A **и вектором правой части** b6= 0**, а также систему**

(A + A) x˜ = b + b,

**где** A∈R**n**×**n и** b∈R**n возмущения матрицы и вектора пра-вой части. Насколько сильно решение** x˜ **возмущённой системы может отличаться от решения** x **исходной системы уравнений?**

**Пусть** x˜ =x+ x**, так что**

(A + A)(x + x) = b + b.

**Вычитая из этого уравнения исходное невозмущённое уравнение, полу-**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **чим** |  |  |  |  |
| (ΔA) x + (A + | | A) x = b, | **(3.17)** |  |
| **или** |  |  |  |  |
| (ΔA)(x + x) + A x = b, | | |  |  |
| **так что** |  |  |  |  |
| x = A−***1*** −(ΔA) x˜ + b . | | |  |  |
| **Беря нормы от обеих частей** | **этого соотношения, получаем** | |  |  |
|  |  |  |  |

k xk ≤ kA−***1***k · k Ak kx˜k + k bk

|  |  |
| --- | --- |
| 3.2. Нормы векторов и матриц | **147** |

**при согласовании векторных и матричных норм. Поделив обе части на** kx˜k > 0**, придём к неравенству**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| kkx˜kk | ≤ kA−***1***k · k Ak + kkx˜kk | | | | | | | | |  |  |  |
| x |  |  |  |  |  |  | b |  |  |  |  |  |
|  |  | k |  | k k k · |  |  |  |  |  |  |  |  |
|  |  |  | kAk | | | kAk · kx˜k | |  |  |
|  | = |  | A−***1*** | A | k Ak | | + |  | k bk | . | **(3.18)** |  |
|  |  |  | |  |  |  |

**Это весьма практичная апостериорная**3 **оценка относительной ошибки решения в зависимости от относительных ошибок в матрице** A **и пра-вой части** b**, коль скоро** kAk · kx˜k ≥ kAx˜k ≈ kbk **и потому знаменатель второго слагаемого в скобках из правой части неравенства примерно равен** kbk**. Эту оценку удобно применять после того, как приближённое решение системы уже найдено. Фигурирующая в оценке (3.18) величи-на** kA− ***1***k kAk**, на которую суммарно умножаются ошибки в матрице и правой части, имеет своё собственное название, так как играет важней-шую роль в вычислительной линейной алгебре.**

Определение 3.2.5 **Для квадратной неособенной матрицы** A **вели-чина** kA−***1***k kAk **называется её** числом обусловленности **(относитель-но выбранной нормы матриц).**

**Мы будем обозначать число обусловленности матрицы** A **посред-ством** cond(A)**, иногда с индексом, указывающим выбор нормы. Если же матрица** A **особенна, то удобно положить** cond(A) = +∞**.**

**Выведем теперь априорную оценку погрешности, которая не будет опираться на знание вычисленного решения и годится для получения оценки до решения СЛАУ.**4

**После вычитания точного уравнения из приближённого мы получи-ли (3.17)**

(ΔA) x + (A + A) x = b.

3 **От латинского словосочетания ¾a posteriori¿, означающего знание, полученное из опыта.**

4 **От латинского словосочетания ¾a priori¿, означающего в философии знание, полученное до опыта и независимо от него.**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **148** |  |  |  | 3. | | | Численные методы линейной алгебры | | | | |  |
| **Отсюда** |  |  |  |  |  |  | − |  |  |  |  |  |
|  |  |  |  |  |  | ***1*** |  |  | ***1*** |  |
| x = (A + A)−***1*** | | | | | | |  | (ΔA) x + b | | | |  |
| = | A(I + |  | ***1*** |  | − |  |  | ***1*** | − | ***1*** − |  |  |
|  |  | A | | |  |  | A) | |  | (ΔA) x + b |  |
| = I + A− | | |  |  |  | A − A− −(ΔA) x + b . | | | | | |  |

**Беря нормы от обеих частей этого равенства и пользуясь субмульти-пликативностью нормы и неравенством треугольника, получим**

|  |  |  |  |
| --- | --- | --- | --- |
| k xk ≤ | I + A−***1*** | A −***1*** | · kA−***1***k · k Ak kxk + k bk , |
|  |  |  |  |

**откуда после деления обеих частей на** kxk>0**:**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | k | xk | ≤ |  | I + A−***1*** | A −***1*** |  | · kA−***1***k · | k Ak + | k | bk | . |  |
|  | k | x | k | x |  |
|  | k | |  |  |  |  | k | |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Если возмущение** A **матрицы** A **не слишком велико, так что** | | | | | | | | | | | | |  |
|  |  |  |  |  | kA−***1*** Ak ≤ kA−***1***k k Ak < 1, | | | | | |  |  |  |
| **то обратная матрица** (I+A−***1*** | | | | | | A)−***1*** **разлагается в матричный ряд Ней-** | | | | | | |  |

**мана (3.15), и мы можем воспользоваться вытекающей из этого оценкой (3.16). Тогда**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| kxk | ≤ | 1 − kA−***1***k k Ak | | | | · k | | | |  | k | | kkxkk | | | | | | |  |  |
| k xk |  |  | kA−***1***k |  |  |  |  |  |  | A + | | |  |  | b | |  | | |  |  |
|  |  |  |  |  |  |  |  |  |  | kAk kxk | | | | | |  |  |
|  |  | 1 − kA−***1***k k Ak | | | | · kAk | | | | | | |  |  |  |
|  | = | kA−***1***k · kAk | | | |  |  |  | k Ak | | | + |  |  | k bk | | |  |  |  |  |
|  |  | | |  |  | |  |  |  |  | |  |  |
|  | ≤ | 1 | cond(A) |  | k | Ak · | | | | kkAkk + kkbkk , | | | | | | | | | | **(3.19)** |  |
|  |  |  | cond(A) | | |  |  |  |  |  |  | A |  |  |  |  | b | | |  |  |
|  |  |  |  |  |  | |  |  | |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | − |  | · |  | |  | | |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  | kAk | |  |  |  |  |  |  |  |  |  |  |  |  |

**поскольку** kAk kxk ≥ kAxk=kbk**.**

**Оценка (3.19) важная априорная оценка относительной погреш-ности численного решения системы линейных алгебраических уравне-ний через оценки относительных погрешностей матриц и правой части.**

|  |  |
| --- | --- |
| 3.2. Нормы векторов и матриц | **149** |

**Если величина** kAk **достаточно мала, то множитель усиления отно-сительной ошибки в данных**

cond(A)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 1 | − | cond(A) | · | k Ak |  |
| kAk |  |
|  |  |  |

**близок к числу обусловленности матрицы** A**.**

**Понятие числа обусловленности матрицы и полученные с его помо-щью оценки имеют большое теоретическое значение, но их практиче-ская полезность напрямую зависит от наличия эффективных способов вычисления или хотя бы оценивания числа обусловленности матриц. Фактически, определение числа обусловленности требует знания неко-торых характеристик обратной матрицы, и в случае общих матричных норм удовлетворительного решения задачи оценивания** cond(A) **не су-ществует до сих пор.**

**Тем не менее, существует практически важный частный случай, ко-гда нахождение числа обусловленности матрицы может быть выполне-но достаточно эффективно. Это случай спектральной матричной нор-мы** k · k***2*, подчинённой евклидовой норме векторов.**

**Напомним, что если** λ **собственное значение квадратной неособен-ной матрицы, то** λ−***1* это собственное значение обратной матрицы, отвечающее тому же собственному вектору: если** C n×n**-матрица и** Cv = λv**, то** v = λC−***1***v ⇒ C−***1***v = λ−***1***v**. Применяя это соображение** **к матрице** A>A**, можем заключить, что если** λ***1*,** λ***2*, . . . ,** λ**n её соб-**

**ственные значения, то у обратной матрицы** (A>A)−***1*** =A−***1***(A>)−***1* соб-ственными значениями являются** λ−***11*,** λ−***21*, . . . ,** λ−**n*1*. Но** A−***1***(A>)−***1*** =

A−***1***(A− ***1***)>**, а потому в силу Предложения 3.1.1 выписанные числа** λ−***11*,** λ−***21*, . . . ,** λ−**n*1*** **образуют набор квадратов сингулярных чисел матрицы** A−***1*. Следовательно, сингулярные числа матрицы** A **суть обратные ве-личины для сингулярных чисел матрицы** A−***1*.**

**В частности, для любой неособенной квадратной матрицы** A **спра-ведливо равенство** σ***max***(A−***1***) =σ***min***−***1***(A)**, и поэтому относительно спек-тральной нормы число обусловленности матрицы есть**

σ***max***(A)

cond***2***(A) = σ***min***(A) .

**Этот результат помогает понять большую роль сингулярных чисел и важность алгоритмов для их нахождения в современной вычислитель-**

**150** 3. Численные методы линейной алгебры

**ной линейной алгебре. В совокупности с ясным геометрическим смыс-лом евклидовой векторной нормы (2-нормы) это вызывает преимуще-ственное использование этих норм для ряда задач, как в теории, так и на практике.**

**Наконец, если матрица** A **симметрична, то её сингулярные числа совпадают с модулями собственных значений, и тогда**

|  |  |  |  |
| --- | --- | --- | --- |
| cond***2***(A) = | max**i** |λ**i**(A)| | **(3.20)** |  |
|  |
| min**i** |λ**i**(A)| |  |
|  |  |  |

**спектральное число обусловленности равно отношению наибольшего и наименьшего по модулю собственных значений матрицы. Для сим-метричных положительно определённых матриц эта формула прини-мает совсем простой вид**

cond***2***(A) = λ***max***(A) .

λ***min***(A)

**3.2з** **Примеры хорошообусловленных и плохообусловленных матриц**

**Условимся называть матрицу хорошо обусловленной, если её число обу-словленности невелико. Напротив, если число обусловленности матри-цы велико, станем говорить, что матрица плохо обусловлена. Естествен-но, что эти определения имеют неформальный характер, так как зави-сят от нестрогих понятий ¾невелико¿ и ¾велико¿. Тем не менее, они весьма полезны в практическом отношени, в частности, потому, что позволяют сделать наш язык более выразительным.**

**Отметим, что для любой подчинённой матричной нормы**

cond(A) = kA−***1***k kAk ≥ kA−***1***Ak = kIk = 1,

**так что соответствующее число обусловленности матрицы всегда не меньше единицы.**

**Примером матриц, обладающих наилучшей возможной обусловлен-ностью относительно спектральной нормы, являются ортогональные матрицы, для которых** cond***2*** Q= 1**. Действительно, если** Q **ортогональ-на, то** kQxk***2*** =kxk***2* для любого вектора** x**. Следовательно,** kQk***2*** = 1**. Кроме того,** Q−***1*** =Q> **и тоже ортогональна, а потому** kQ−***1***k***2*** = 1**.**

**Самым популярным примером плохообусловленных матриц явля-ются, пожалуй, матрицы Гильберта** H**n** = (h**ij** )**, которые встретились**

|  |  |
| --- | --- |
| 3.2. Нормы векторов и матриц | **151** |

**нам в §2.5б при обсуждении среднеквадратичного приближения алгеб-раическими полиномами на интервале** [0,1]**. Это симметричные матри-цы, образованные элементами**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | |  | , | i, j = 1, 2, . . . , n, | | |  |
| h**ij** = |  |  |  |
| i + j − 1 |  |
| **так что, к примеру,** | | | ***2*** | ***3*** | ***4*** | . |  |
|  | H***3*** = | |  |
|  |  |  | 1 | ***1*** | ***1*** |  |  |
|  |  |  | ***2*** | ***3*** |  |  |
|  |  | | ***3*** |  |  |
|  | ***4*** | ***5*** |  |
|  |  | | ***1*** | ***1*** | ***1*** |  |  |
|  |  |  |  |  |
|  |  | | ***1*** | ***1*** | ***1*** |  |  |
|  |  |  |  |  |

**Число обусловленности матриц Гильберта исключительно быстро рас-тёт в зависимости от их размера** n**. Воспользовавшись какими-либо системами компьютерной математики (например, Scilab,** Matlab **и им подобными), нетрудно найти следующие числовые данные:**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| cond***2*** H***2*** | | = | | | | 19.3, | |  |  |  |
| cond***2*** H***3*** | | = | | | | 524, | |  |  |  |
| · · · | |  |  |  |  |  |  |  |  |  |
| cond***2*** H***10*** | | = | | | | 1.6 · 10***13***, | |  |  |  |
| · · · . | | | | | |  |  |  |  |  |
| **Существует общая формула [70, 69] :** | | | | | | |  |  |  |  |
| cond***2*** H**n** = O |  | √ | |  |  | ***4*n** | ≈O(34**n**/√ |  |  |  |
|  |  |  |  |  |  |
| (1 + 2) | | | | | |  | ), |  |
| n |  |
| √ | |  | |  | |  |
| n | |  |

**где** O **¾о большое¿, известный из математического анализа символ Э. Ландау. Интересно, что матрицы, обратные к матрицам Гильберта имеют целочисленные элементы [65], которые также очень быстро рас-тут с размерностью.**

**На этом фоне для матрицы Вандермонда (2.3) оценка числа обу-словленности (см. [50])**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| √ |  |  |  | √ |  |  |  |  |
|  | (1 + | | 2)**n**−***1*** | | |  |
|  |  |
| cond***2*** V (x***0***, x***1***, . . . , x**n**) ≥ 2 | |  | √ |  | | |  |  |
|  | n + 1 | | | |  |

**152** 3. Численные методы линейной алгебры

**представляется даже не слишком плохой.**5 **Но она и не хороша, а опыт реальных вычислений свидетельствует о том, что ошибки в линейных системах с матрицей Вандермонда могут весьма сильно влиять на от-вет.**

**3.2и** **Практическое применение числа обусловленности матриц**

**Оценки (3.18) и (3.19) на возмущения решений систем линейных ал-гебраических уравнений являются неулучшаемыми на всём множестве матриц, векторов правых частей и их возмущений. Но ¾плохая обу-словленность¿ матрицы не всегда означает высокую чувствительность решения конкретной системы по отношению к тем или иным конкрет-ным возмущениям. Если, к примеру, правая часть имеет нулевые ком-поненты в направлении сингулярных векторов, отвечающих наимень-шим сингулярным числам матрицы системы, то решение СЛАУ зави-сит от возмущений этой правой части гораздо слабее, чем показывает оценка (3.19) (см., к примеру, §3.7 книги [39]). И определение того, ка-кая конкретно у нас правая часть плохая или не очень не менее трудно, чем само решение данной системы линейных уравнений.**

**Из сказанного должна вытекать известная осторожность и осмотри-тельность по отношению к выводам, которые делаются о практической разрешимости и достоверности решений какой-либо системы линейных уравнений лишь на основании того, велико или мало число обусловлен-ности их матрицы.**

**Тривиальный пример: число обусловленности диагональной матри-цы может быть сколь угодно большим, но решение СЛАУ с такими матрицами почти никаких проблем не вызывает!**

**Наконец, число обусловленности малопригодно для оценки разбро-са решения СЛАУ при значительных и больших изменениях элементов матрицы и правой части (скажем, начиная с нескольких процентов от исходного значения). Получаемые при этом с помощью оценок (3.18) и (3.19) результаты типично завышены во много раз, и для решения упомянутой задачи более предпочтительны методы интервального ана-лиза (см., к примеру, [64]).**

5 **Аналогичные по смыслу, но более слабые экспонециальные оценки снизу для числа обусловленности матрицы Вандермонда выводятся также в книге [39].**

|  |  |
| --- | --- |
| 3.2. Нормы векторов и матриц | **153** |

Пример 3.2.1 **Рассмотрим** 2 × 2**-систему линейных уравнений**

* **!**

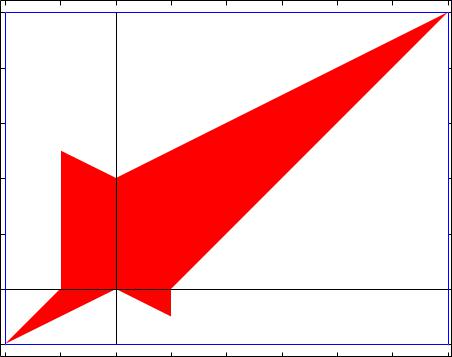
|  |  |  |
| --- | --- | --- |
| a***11*** | a***12*** | b***1*** |
|  |  | x = |
| a***21*** | a***22*** | b***2*** |

**в которой элементы матрицы и правой части заданы приближённо и могут принимать значения из интервалов**

|  |  |  |
| --- | --- | --- |
| a***11*** ∈ [2, 4], | a***12*** ∈ [−2, 0], | b***1*** ∈ [−1, 1] |
| a***12*** ∈ [−1, 1], | a***22*** ∈ [2, 4], | b***2*** ∈ [0, 2]. |

**При этом обычно говорят [64], что задана интервальная система ли-нейных алгебраических уравнений**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| [ | [2, 4] | | [−2, 0] | **!** x = | [−1, 1] | **!** | **(3.21)** |  |
| − | 1, 1] | [2, 4] | [0, 2] |  |
|  |  |  |  |  |  |  |  |



|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2.5 |  |  |  |  |  |  |  |  |
| 2 |  |  |  |  |  |  |  |  |
| 1.5 |  |  |  |  |  |  |  |  |
| 1 |  |  |  |  |  |  |  |  |
| 0.5 |  |  |  |  |  |  |  |  |
| 0 |  |  |  |  |  |  |  |  |
| −0.5 |  |  |  |  |  |  |  |  |
| −1 | −0.5 | 0 | 0.5 | 1 | 1.5 | 2 | 2.5 | 3 |

**Рис. 3.4. Множество решений интервальной линейной системы (3.21).**

**Множеством её решений называют множество, образованное всевоз-можными решениями систем линейных алгебраических уравнений той же структуры, коэффициенты матрицы и компонетны правой части которой принадлежат заданным интервалам. В частности, множество решений рассматриваемой нами системы (3.21) изображено на Рис. 3.4. Мы более подробно рассматриваем интервальные линейные системы уравнений в §4.5.**

**154** 3. Численные методы линейной алгебры

**Подсчитаем оценки возмущений решения системы (3.21), которые получаются на основе числа обусловленности. Её можно рассматрива-еть, как систему, получающуюся путём возмущения ¾средней системы¿**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | **!** |  | **!** |  |
|  | 3 | −1 | x = | 0 |  |
|  | 0 | 3 |  | 1 |  |
| **на величину** |  | a***22*!** , | |  |  |
| A = | a***21*** | k Ak∞ ≤ 2, |  |
|  | a***11*** | a***12*** |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **в матрице и величину** | b***2*!** |  |  |
| b = | ,k bk∞ ≤ 1, |  |
|  | b***1*** |  |  |

**в правой части.**

**Обусловленность средней матрицы относительно** ∞**-нормы равна** 1.778**,** ∞**-норма средней матрицы равна** 4**, а** ∞**-норма средней правой** **части это 1. Следовательно, по формуле (3.19) получаем**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | k | xk | / 24. |  |  |  |
|  |  | kxk | | |  |  |  |
| **Поскольку решение средней системы есть** x˜ = (1/3,1/9)>**, имеющее** | | | | | ∞ | **-** |  |
| **норму** | ***1*** |  |  |  |  |  |
| ***3*** | **, то оценкой разброса решений расматриваемой системы урав-** | | | | |  |
| **нений является** x˜±x**, где** k | | | xk∞ ≤ 8**, т. е. двумерный брус** | |  |  |  |

**!**

[−7.667, 8.333]

.

[−7.889, 8.111]

**По размерам он в более чем в 4 (четыре) раза превосходит оптимальные (точные) покоординатные оценки множества решений, равные**

**!**

[−1, 3]

.

[−0.5, 2.5]

**При использовании других норм результаты, даваемые формулой (3.19), совершенно аналогичны своей грубостью оценивания возмуще-ний решений.**

|  |  |
| --- | --- |
| 3.3. Прямые методы решения линейных систем | **155** |

**Отметим в заключение этой темы, что задача оценивания разбро-са решений СЛАУ при вариациях входных данных является в общем случае NP-трудной, т. е. требует для своего решения экспоненциально больших трудозатрат [66, 67].**

1. Прямые методы решения систем линейных алгебраических уравнений

**Решение систем линейных алгебраических уравнений вида**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| a***21***x***1*** | | + a***22***x***2*** | + . . . + a***2*n**x**n** = b***2*** | | | | , |  |
|  | a***11***x***1*** | + a***12***x***2*** | + . . . + a***1*n**x**n** = b***1*** | | | | , |  |
| **.** | **.** | **.** |  | **.** | **.** | **(3.22)** |  |
|  | **.** | **.** | **.** | **.** | **.** |  |  |
|  |  |  |  |
|  | **.** | **.** |  |  | **..** | **.** |  |  |
|  |  |  |  |  |  |  |  |  |
|  | a**n*1***x***1*** + a**n*2***x***2*** + . . . + a**mn**x**n** = b**m**, | | | | | | |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |

**с коэффициентами** a**ij и свободными членами** b**i, или, в краткой форме,**

|  |  |
| --- | --- |
| Ax = b | **(3.23)** |

**с** m×n**-матрицей** A= (a**ij** ) **и** m**-вектором правой части** b= (b**i**)**, являет-ся важной математической задачей. Она встречается как сама по себе, так и в качестве составного элемента в технологической цепочке реше-ния более сложных задач. Например, решение нелинейных уравнений или систем уравнений часто сводится к последовательности решений линейных систем (метод Ньютона).**

**По характеру вычислительного алгоритма методы решения уравне-ний и систем уравнений традиционно разделяют на прямые и итера-ционные. В прямых методах искомое решение получается в результате выполнения конечной последовательности действий, так что эти мето-ды нередко называют ещё конечными или даже точными. Напротив, в итерационных методах решение достигается как предел некоторой по-следовательности приближений, которая конструируется по решаемой системе.**

**Идея, лежащая в основе большинства прямых методов для решения систем линейных алгебраических уравнений, состоит в том, чтобы эк-вивалентными преобразованиями привести решаемую систему к наибо-лее простому виду, из которого решение находится уже непосредствен-но. В качестве простейших могут выступать системы с диагональны-ми, двухдиагональными, треугольными и т. п. матрицами. Чем меньше**

**156** 3. Численные методы линейной алгебры

**ненулевых элементов остаётся в матрице преобразованной системы, тем проще и устойчивее её решение, но, с другой стороны, тем сложнее и неустойчивее приведение к такому виду. На практике обычно стремят-ся к компромиссу между этими взаимно противоположными требова-ниями. Мы, в основном, рассмотрим методы, основанные на приведе-нии к треугольному виду.**

**Наконец, для простоты мы далее подробно разбираем случай систем уравнений, в которых число неизвестных** n **равно числу уравнений** m**, т. е. имеющих квадратную** n×n**-матрицу коэффициентов.**

**3.3а** **Решение треугольных линейных систем**

**Напомним, что треугольными матрицами называют матрицы, у ко-торых все элементы ниже главной диагонали либо все элементы вы-ше главной диагонали нулевые (так что ненулевые элементы образуют треугольник):**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | × × ·**.** ·**. .**· × | | | | × | |  |  |
| U = |  | × | **. ..** | ×**...** | ×**..** |  | , |  |
|  |  |  |  |  | **.** |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  | 0 |  | × | × |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  | × | |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | × | |  |  |  |  |
| L = | × | | × | **.. .** | |  |
|  | × | | × |  |  |  |
|  |  | **..** | **.. ..** | | **.** |  |
|  |  | **. .** | | |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  | × | | × | · · · | |  |

,

×

× ×

**где крестиками ¾**×**¿ обозначены ненулевые элементы. В первом слу-чае говорят о верхней (или правой) треугольной матрице, а во втором о нижней (или левой) треугольной матрице. Соответственно, тре-угольными называются системы линейных алгебраических уравнений, матрицы которых имеют треугольный вид верхний или нижний.**

**Рассмотрим для определённости линейную систему** Lx=b **с нижней треугольной матрицей** L= (l**ij** )**,** l**ij** = 0 **при** j > i**. Её первое уравнение содержит только одну неизвестную переменную** x***1*. Найдём её значение и подставим его во второе уравнение системы, в котором в результа-те останется также всего одна неизвестная переменная** x***2*. Вычислим** x***2*** **и затем подставим известные значения** x***1*** **и** x***2*. И так далее. Реше-ние линейной системы с нижней треугольной матрицей выполняется по следующему простому алгоритму**

|  |  |
| --- | --- |
| 3.3. Прямые методы решения линейных систем | **157** |

DO FOR i = 1 TO n

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | **X** |  |  |  |
| x**i** ← |  | b**i** − **j<i** l**ij** x**j** | **.** l**ii** | , | **(3.24)** |

END DO

**позволяющему последовательно друг за другом вычислить искомые значения неизвестных переменных, начиная с первой. Естественно на-звать этот процесс прямой подстановкой, коль скоро он выполняется по возрастанию индексов компонент вектора неизвестных** x**. Анало-гичный процесс для верхних треугольных систем называется обратной подстановкой он идёт в обратном направлении, т. е. от** x**n к** x***1*.**

**3.3б** **Метод Гаусса для решения линейных систем уравнений**

**Хорошо известно, что умножение какого-либо уравнения системы на ненулевое число, а также сложение уравнений системы приводят к рав-носильной системе уравнений. Воспользуемся этими свойствами для преобразования решаемой системы линейных алгебраических уравне-ний к более простому виду.**

**Предположим, что в системе линейных уравнений** Ax=b **коэф-фициент** a***11*** =60**. Умножим первое уравнение системы на** (−a***21***/a***11***) **и сложим со вторым уравнением. При этом коэффициент** a***21* во вто-ром уравнении занулится, а получившаяся система будет совершенно равносильна исходной.**

**Проделаем подобное преобразование с остальными 3-м, 4-м и т. д. до** n**-го уравнениями системы, т. е. будем умножать первое уравнение на** (−a**i*1***/a***11***) **и складывать с** i**-ым уравнением системы. В результа-те получим равносильную исходной систему линейных алгебраических уравнений, в которой неизвестная переменная** x***1* присутствует лишь в первом уравнении. Матрица получившейся СЛАУ станет выглядеть**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **158** |  |  | 3. | Численные методы линейной алгебры | | | |  |
| **следующим образом** | |  |  |  |  |  |  |  |
|  | | 0 |  | × · · · × | |  |  |  |
| × |  |  |
|  |  | × |  |  |  |  |  |  |
|  | | 0 | × × ·**.** ·**. .**· ×**...** | | |  | . |  |
|  | |  | × | × |  |  |  |  |
|  | | **.** | **. . .** | | **.** |  |  |  |
|  | | **..** | **..** | **..** | **.. ..** |  |  |  |
|  | |  |  |  |  |  |  |  |
|  | |  |  |  |  |  |  |  |
|  | | 0 | × | × · · · × | |  |  |  |
|  | |  |  |  |  |

**Рассмотрим теперь в преобразованной системе уравнения со 2-го по** n**-е. Они образуют квадратную** (n−1)×(n−1)**-систему линейных уравнений, к которой, если элемент на месте** (2,2) **не сделался равным нулю, можно заново применить вышеописанную процедуру исключе-ния. Её результатом будет обнуление поддиагональных элементов 2-го столбца матрицы СЛАУ. И так далее.**

**Выполнив** (n−1) **шагов подобного процесса c 1-м, 2-м, . . . ,** n−1**-м столбцами матрицы данной системы, мы получим, в конце концов, линейную систему с верхней треугольной матрицей, которая несложно решается с помощью обратной подстановки. Описанное преобразование системы линейных алгебраических уравнений к равносильному тре-угольному виду называется прямым ходом метода Гаусса, и его псев-докод выглядит следующим образом:**

|  |  |  |  |
| --- | --- | --- | --- |
| DO FOR j = 1 | TO n − 1 |  |  |
| DO FOR | i = j + 1 TO n |  |  |
| r**ij** ← a**ij** /a**jj** | |  |  |
| DO FOR k = j + 1 TO | | n |  |
|  | a**ik** ← a**ik** − r**ij** a**jk** |  | **(3.25)** |
| END DO | |  |  |
| b**i** ← b**i** − r**ij** b**j** | |  |  |
| END DO |  |  |  |
| END DO |  |  |  |
|  |  |  |  |

**Он выражает процесс последовательного обнуления поддиагональных элементов** j**-го столбца матрицы системы** A**,** j= 1,2, . . . , n−1**, и со-**

|  |  |
| --- | --- |
| 3.3. Прямые методы решения линейных систем | **159** |

**ответствующие преобразования вектора** b**. Матрица системы при этом приводится к верхнему треугольному виду. Далее следует обратный ход метода Гаусса, который является процессом обратной подстановки для решения полученной верхней треугольной системы:**

DO FOR i = n TO 1 STEP (−1)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | **X** |  |  |  |
| x**i** ← |  | b**i** − **j>i** a**ij** x**j** | **.** a**ii** | , | **(3.26)** |

END DO

**Он позволяет последовательно вычислить, в обратном порядке, иско-мые значения неизвестных, начиная с** n**-ой. Отметим, что в нашем псев-докоде прямого хода метода Гаусса зануление поддиагональных эле-ментов уже учтено нижними границами внутренних циклов.**

**Отметим, что существует много различных версий метода Гаусса. Весьма популярной является, к примеру, вычислительная схема един-ственного деления. При её выполнении сначала делят первое уравнение системы на** a***11*** 6= 0**, что даёт**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| x***1*** + | a***12*** | x***2*** + · · · + | a***1*n** | x**n** = | b***1*** | . | **(3.27)** |  |
| a***11*** | a***11*** | a***11*** |  |

**Умножая затем уравнение (3.27) на** a**i*1* и вычитая результат из** i**-го уравнения системы для** i= 2,3, . . . , n**, добиваются обнуления поддиа-гональных элементов первого столбца. Затем процедура повторяется в отношении 2-го уравнения и 2-го столбца получившейся СЛАУ, и так далее.**

**Схема единственного деления совершенно эквивалентна алгоритму (3.25) и отличается от него лишь тем, что для каждого столбца де-ление в ней выполняется действительно только один раз, тогда как все остальные операции это умножение и сложение. С другой сторо-ны, уравнения преобразуемой системы в схеме единственного деления дополнительно масштабируются диагональными коэффициентами при неизвестных, и иногда это бывает нежелательно.**

**160** 3. Численные методы линейной алгебры

**3.3в** **Матричная интерпретация метода Гаусса**

**Заметим, что умножение первого уравнения системы на** r**i*1*** =−a**i*1***/a***11* и сложение его с** i**-ым уравнением могут быть представлены в матрич-ном виде как умножение системы** Ax=b **слева на матрицу**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | 0 | 1 |  |  |  |  |  |
|  | 1 |  | 0 |  |  |  |  |
| **...** | **. ..** |  |  |  |
|  |  |  |  |  |  |  |  |
|  | r**i*1*** | 1 |  |  |  | , |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  | **.** |  | 1 |  |  |  |  |
|  | **..** |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  | 0 |  |  |  |  |  |
|  | 0 |  | 1 |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |

**которая отличается от единичной матрицы наличием одного дополни-тельного ненулевого элемента** r**i*1* на месте** (i,1)**. Исключение поддиа-гональных элементов первого столбца матрицы СЛАУ это последо-вательное домножение системы слева на матрицы**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| r***21*** | | 1 | 0 | | | |  | 0 | | 1 | 0 | |  |  |  |
|  | 1 |  |  |  |  |  |  |  | 1 |  |  |  |  |  |  |
|  |  |  | **. ..** |  |  |  | , |  | r***31*** | 0 **.. .** |  |  |  | , |  |
|  | 0 |  |  |  |  | |  |  |  |  |  |  |  |  |  |
|  | **.** |  |  |  |  | |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  |
|  | **.** |  |  | 1 |  | |  |  | **.** |  |  |  |  |  |  |
|  | **.** |  |  |  |  | |  |  | **.** |  |  | 1 |  |  |  |
|  |  |  |  |  |  | |  |  | **.** |  |  |  |  |  |  |
|  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  |
|  | 0 | 0 |  | 1 |  | |  |  | 0 | 0 |  | 1 |  |  |  |
|  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  | |  | 0 | 1 |  | 0 |  |  |  |  |  |
|  |  |  |  |  | |  | 1 |  |  |  |  |  |  |  |  |
|  |  | **и т. д. до** | |  | |  | **...** |  | **...** |  |  | . |  |  |  |
|  |  |  |  |  | |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  | |  | 0 |  |  | 1 |  |  |  |  |  |
|  |  |  |  |  | |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  | | r**n*1*** | | 0 |  | 1 |  |  |  |  |  |
|  |  |  |  |  | |  | |  |  |  |  |  |  |  |  |
|  |  |  |  |  | |  |  |  |  |  |  |  |  |  |  |

**Нетрудно убедиться, что умножение матриц выписанного выше ви-**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 3.3. Прямые методы решения линейных систем | | | | | | | | | | | | |  |  |  |  | **161** |  |
| **да выполняется по простому правилу:** | | | | | | | | | |  |  |  |  |  |  |  |  |  |
| 1 | | 1 |  | 0 | |  |  |  | | 1 |  | 1 | **...** | 0 | |  |  |  |
|  | r**i*1*** | **...** |  |  |  |  |  |  | |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  |
|  |  |  | 1 |  |  |  |  |  | · |  |  |  |  | 1 |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  | **..** | **.** |  |  |  | | r**k*1*** |  |  |  | **..** | **.** |  |  |  |
|  |  |  |  |  |  |  |  | |  |  |  |  |  |  |  |
|  |  | 0 |  |  |  |  |  |  | |  |  |  | 0 |  |  |  |  |  |
|  |  |  |  |  |  | 1 |  | |  |  |  |  |  | 1 |  |  |
|  |  |  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  |
|  |  |  | 1 | |  | 1 |  |  |  | 0 | |  |  |  |  |  |  |  |
|  |  | = |  | r**i*1*** |  |  |  | **.. .** |  |  |  |  |  | . |  |  |  |  |
|  |  |  |  |  |  |  |  | 1 |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  | r**k*1*** |  |  |  |  |  | **.** | **. .** |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  | 0 | |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  | 1 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

**Оно также остаётся верным в случае, когда у сомножителей в первом столбце на взаимнодополнительных местах присутствует более одно-го ненулевого элемента. Следовательно, обнуление поддиагональных элементов первого столбца и соответствующие преобразования правой части это не что иное, как умножение обеих частей СЛАУ слева на матрицу**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | r***21*** | | 1 |  | 0 | |  |  |  |  |  |
|  |  |  | 1 |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  | **..** | 0 | 1 | **..** | **.** |  |  |  |  |  |
| E***1*** | = |  | **.** |  |  |  | . | **(3.28)** |  |
|  | r***31*** |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  | r |  | 0 |  |  | 1 |  |  |  |  |
|  |  |  | **n*1*** |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |

**Аналогично, обнуление поддиагональных элементов** j**-го столбца матрицы СЛАУ и соответствующие преобразования правой части мож-**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **162** |  |  | 3. | Численные методы линейной алгебры | | | | | | |  |
| **но интерпретировать как умножение системы слева на матрицу** | | | | | | | | | |  |  |
|  |  | 1 | **...** |  | 0 | |  |  |  |  |  |
| E | = |  |  | 1 |  |  |  |  | , | **(3.29)** |  |
|  |  |  |  |  |  |  |  |
| **j** |  |  |  | r**j*+1*,j** | 1 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  | 0 | **.** | **.** | **..** |  |  |  |  |  |
|  |  |  | **..** |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  | r**nj** |  |  | 1 |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |

**а в целом метод Гаусса представляется как последовательность умно-жений решаемой СЛАУ слева на матрицы** E**j ,** j= 1,2, . . . , n−1**. При этом**

|  |  |
| --- | --- |
| E**n**−***1*** · · · E***2***E***1***A = U | **(3.30)** |

**верхней треугольной матрице.**

**Коль скоро все** E**j нижние треугольные матрицы, их произведе-ние также является нижним треугольным. Кроме того, все** E**j неособен-ны (нижние треугольные с единичной главной диагональю). Поэтому неособенно и их произведение** E**n**−***1*** · · ·E***2***E***1*. Обозначим**

* + = E**n**−***1*** · · · E***2***E***1*** −***1***,

**нижняя треугольная матрица, для которой в силу (3.30) справедливо**

A = LU

**т. е. исходная матрица СЛАУ оказалась представленной в виде произ-ведения нижней треугольной** L **и верхней треугольной** U **матриц. Это представление называют LU-разложением матрицы или треугольным разложением.**

**Соответственно, преобразования матрицы** A **в прямом ходе метода Гаусса можно трактовать как её разложение на нижний треугольный** L **и верхний треугольный** U **множители. В результате вместо решения** **исходной СЛАУ, равносильной**

(LU ) x = b,

**мы можем выполнить два несложных решения треугольных систем ли-нейных алгебраических уравнений**

**(**

Ly = b,

U x = y

**(3.31)**

|  |  |
| --- | --- |
| 3.3. Прямые методы решения линейных систем | **163** |

**с помощью прямой и обратной подстановок соответственно. Отметим, что при реализации метода Гаусса на компьютере можно**

**хранить треугольные сомножители** L **и** U **на месте** A**, так как у** L **по диагонали стоят единицы.**

**3.3г** **Существование LU-разложения**

**И в прямом, и в обратном ходе метода Гаусса встречаются операции деления, которые не выполнимы в случае, если делитель равен нулю. Тогда не может быть выполнен и метод Гаусса в целом. Естествен-но задаться вопросом о достаточных условиях реализуемости метода Гаусса, или, иначе, условиях существования LU-разложения матрицы. Следующий простой результат даёт частичный ответ на него.**

Теорема 3.3.1 **Пусть** A = (a**ij** ) **квадратная матрица и все её ве-дущие миноры отличны от нуля, т. е.**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| a***11*** 6= 0, | det | a***11*** | a***12*** | 6= 0, | . . . , | det A 6= 0. |  |
| a***21*** | a***22*** |  |
| **Тогда матрица** | A **представима в виде** | | | |  |  |  |

A = LU

**произведения нижней треугольной матрицы** L **и верхней треуголь-ной матрицы** U **. Это разложение единственно при условии, что диа-гональными элементами в** L **являются единицы.**

Доказательство **мы проведём индукцией по размеру** n **матрицы** A**.** **Если** n= 1**, то утверждение теоремы очевидно. Тогда матрицы** L=

(l**ij** ) **и** U = (u**ij** ) **являются просто числами, и достаточно взять** l***11*** = 1

**и** u***11*** =a***11*.**

**Пусть теорема верна для матриц размера** (n−1)×(n−1)**. Тогда представим** n×n**-матрицу** A **в блочном виде**

|  |  |  |  |
| --- | --- | --- | --- |
|  | a***11*** | |  |
|  | a***21*** | |  |
| A = | **...** | |  |
|  | a | **n*1*** |  |
|  |  |  |
|  |  |  |  |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| a***22*** | | . . . a***2*n** | |  |  | A**n** ***1*** | z |  |
| a***12*** | | . . . a***1*n** | |  |  |  | a**nn** **!** , |  |
|  | **...** | **.. . ...** | |  | = | v− |  |  |
| a | **n*2*** | . . . a | **nn** |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |

**164** 3. Численные методы линейной алгебры

**где** A**n**−***1*** (n − 1) × (n − 1)**-подматрица** A**,** z **вектор-столбец размера** n − 1**,**

v **вектор-строка размера** n − 1**,** **такие что**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | a***2*n** | | |  |  |  |
|  |  |  | a***1*n** |  |  |  |
| z = |  | **...** | , |  |
|  |  | a | **n**−***1*,n** |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |  |

v = a**n*1*** a**n*2***

. . . a**n,n**−***1*** .

**Требование разложения** A **на треугольные множители диктует равен-**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **ство** | v− |  | a**nn!** | = | x− |  | l**nn!** | · | 0− |  | u**nn!** | , |  |
| A = |  |  |  |  |
|  | A**n** | ***1*** | z |  | L**n** | ***1*** | 0 |  | U**n** | ***1*** | y |  |  |

**где** L**n**−***1***, U**n**−***1*** (n − 1) × (n − 1)**-матрицы,** x **вектор-строка размера** n − 1**,**

y **вектор-столбец размера** n−1**.**

**Следовательно, используя правила перемножения матриц по блокам,**

|  |  |
| --- | --- |
| A**n**−***1*** = L**n**−***1***U**n**−***1***, |  |
| z = L**n**−***1***y, | **(3.32)** |
| v = x U**n**−***1***, | **(3.33)** |
| a**nn** = xy + l**nn**u**nn**. | **(3.34)** |

**Первое из этих соотношений выполнено в силу индукционного пред-положения, причём оно должно однозначно определять** L**n**−***1* и** U**n**−***1*, если потребовать по диагонали в** L**n**−***1* единичные элементы. Далее, по условию теоремы** detA**n**−***1*** =60**, а потому матрицы** L**n**−***1* и** U**n**−***1* также должны быть неособенны. По этой причине системы линейных уравне-ний относительно** x **и** y

x U**n**−***1*** = v **и** L**n**−***1***y = z,

**которыми являются равенства (3.32)–(3.33), однозначно разрешимы. Найдя из них векторы** x **и** y**, мы сможем из соотношения (3.34) восста-новить** l**nn и** u**nn. Если дополнительно потребовать** l**nn** = 1**, то значение** u**nn** **находится однозначно и совпадает с** (a**nn** − xy)**.**

|  |  |
| --- | --- |
| 3.3. Прямые методы решения линейных систем | **165** |

* + **формулировке Теоремы 3.3.1 ничего не говорится о том, реализу-ем ли метод Гаусса для соответствующей системы линейных алгебра-ических уравнений. Но нетрудно понять, что в действительности тре-буемое Теоремой 3.3.1 условие отличия от нуля ведущих миноров в матрице СЛАУ является достаточным признаком выполнимости рас-смотренного в §3.3б простейшего варианта метода Гаусса.**
  + **самом деле, к началу** j**-го шага прямого хода, на котором предсто-ит обнулить поддиагональные элементы** j**-го столбца матрицы СЛАУ,**

**её ведущей (верхней угловой)** j×j**-подматрицей является треугольная матрица, которая получена из исходной ведущей подматрицы преобра-**

**зованиями метода Гаусса предыдущих** j−1 **шагов. Эти преобразова-ния линейное комбинирование строк не изменяют определителя,**

* **потому неравенство нулю всех ведущих миноров влечёт отличие от**

**нуля всех диагональных элементов ведущих подматриц преобразован-ной матрицы. В частности, при этом всегда** a**jj** 6= 0**, так что деление на этот элемент в (3.25) выполнимо.**

**3.3д** **Метод Гаусса с выбором ведущего элемента**

**Ведущим элементом в методе Гаусса называют элемент матрицы ре-шаемой системы, на который выполняется деление при исключении поддиагональных элементов очередного столбца.**6 **В алгоритме, опи-санном в §3.3б, ведущим всюду брался диагональный элемент** a**jj , но, вообще говоря, при решении конкретных СЛАУ более предпочтитель-ным может оказаться другой выбор ведущего элемента.**

**Условия Теоремы 3.3.1 заведомо выполнены, к примеру, для СЛАУ с положительно определёнными матрицами (в силу известного крите-рия Сильвестера), или если матрица СЛАУ имеет диагональное преоб-ладание. Но в общем случае проверка условий этой теоремы является весьма непростой, поскольку вычисление ведущих миноров матрицы требует немалых трудозатрат, и, по существу, ничуть не проще самого метода Гаусса.**

**С другой стороны, зануление какого-либо ведущего минора может произойти и в неособенной матрице, так что сформулированный Теоре-мой 3.3.1 признак это достаточное и довольно грубое условие. Вооб-ще, желательно уметь выполнять метод Гаусса для СЛАУ с произволь-**

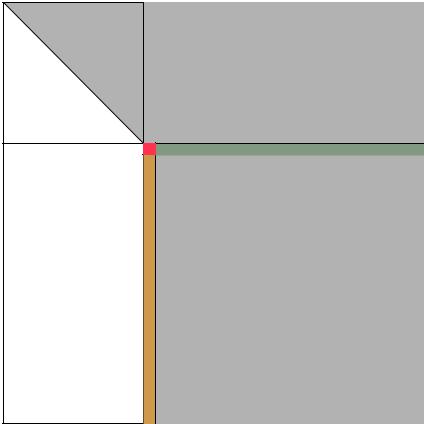
6 **Иногда в русской математической литературе его назыают** главным **элементом.**

**166** 3. Численные методы линейной алгебры

**ными неособенными матрицами, что равносильно его модификации та-ким образом, чтобы ведущий элемент всегда был отличен от нуля.**

**Отметим, что изменение порядка уравнений в системе приводит к равносильной системе уравнений. При этом в матрице СЛАУ перестав-ляются строки, так что она существенно меняется. Нельзя ли восполь-зоваться этим изменением для организации успешного выполнения ме-тода Гаусса?**

**Частичным выбором ведущего элемента на** j**-ом шаге прямого хо-да метода Гаусса называют процесс его выбора, как максимального по модулю элемента из всех элементов столбца с тем же** j**-ым номером, лещащих на диагонали и ниже, и сопровождаемый соответствующей перестановкой строк матрицы и компонент правой части (т. е. уравне-ний СЛАУ).**



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| j −1 | ( |  | 0 |  |
|  |  | ( | 0 |  |
|  |  | j −1 |  |
|  |  |  |  |
| **Рис. 3.5. Структура матрицы СЛАУ перед началом** | | | |  |
|  | j**-го шага прямого хода метода Гаусса** | | |  |

**Метод Гаусса с частичным выбором ведущего элемента всегда реа-лизуем для системы линейных алгебраических уравнений с неособен-ной матрицей. В самом деле, в преобразованиях прямого хода метода Гаусса не изменяется свойство определителя матрицы быть равным ну-лю или не быть равным нулю и, как следствие, правая нижняя подмат-рица, которой предстоит быть преобразованной на оставшихся шагах метода Гаусса, имеет ненулевой определитель, т.е. в первом её столбце обязан найтись ненулевой элемент. Его и можно сделать ведущим.**

|  |  |
| --- | --- |
| 3.3. Прямые методы решения линейных систем | **167** |

**Каково матричное представление метода Гаусса с выбором ведущего элемента?**

**Введём матрицы транспозиции**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1 | **. ..** | · · · | |  |  |  | ← |  |  |
|  |  | 1 |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  | 0 |  |  | 1 |  |  |  | i**-ая строка** |  |
| P = |  | **.** | **.** | **. .** | **.** |  |  |  |  |  |
|  | **..** |  | **..** |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  | 1 |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  | · · · | | **.** |  |  | ← |  |  |
|  |  |  | **. .** |  |  |  |  |
|  |  | 1 |  |  | 0 |  |  | j**-ая строка** |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  | 1 |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |

**элементарные матрицы перестановок, которые получаются из еди-ничной матрицы перестановкой двух её строк (или столбцов) с номе-рами** i **и** j**. Умножение на такую матрицу слева приводит к переста-новке** i**-ой и** j**-ой строк, а умножение справа к перестановке** i**-го и** j**-го столбцов. Тогда для метода Гаусса с частичным выбором ведущего** **элемента справедливо следующее матричное представление**

(E**n**−***1***P**n**−***1***) · · · (E***1***P***1***)A = U,

**где** E***1*,** E***2*, . . . ,** E**n**−***1* матрицы перестановок, при помощи которых выполняется перестановка строк на 1-м, 2-м, . . . ,** (n−1)**-м шагах метода Гаусса.**

**Несмотря на то, что метод Гаусса с частичным выбором ведущего элемента теоретически спасает положение, на практике для некоторых плохих СЛАУ он может работать не очень хорошо. По этой причине для обеспечения устойчивости вычислительного процесса по методу Гаусса иногда имеет смысл выбирать ведущий элемент более тщательно, чем это делается при частичном выборе**

**Ещё одним простым способом системы равносильного преобразова-ния системы уравнений является перенумерация переменных. Ей со-ответствует перестановка строк матрицы, а вектор правых частей при этом неизменен.**

**Полным выбором ведущего элемента называют его выбор, как мак-симального по модулю элемента из всей правой нижней подматрицы, которой предстоит быть преобразованной на оставшихся шагах метода**

**168** 3. Численные методы линейной алгебры

**Гаусса (а не только из её первого столбца, как было при частичном выборе), и сопровождаемый соответствующей перестановкой строк и столбцов матрицы и компонент правой части.**

**Напомним, что матрицей перестановки называется матрица, полу-чающаяся из единичной матрицы перестановкой произвольного числа её строк (столбцов).**

Теорема 3.3.2 **Для неособенной матрицы** A **существуют матрицы**

ˇ ˆ

**перестановок** P **и** P **, такие что**

ˇ ˆ

P AP = LU,

**где** L**,** U **нижняя и верхняя треугольные матрицы, причём диаго-нальными элементами в** L **являются единицы. В этом представлении**

ˇ ˆ

**можно ограничиться лишь одной из матриц** P **или** P **.**

**3.3е** **Число обусловленности и матричные преобразования**

**Пусть матрица** A **умножается на матрицу** B**. Как при этом изменяется число обусловленности получающейся матрицы?**

**Имеем**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | kABk ≤ kAk kBk, | | | | ≤ kA−***1***k kB−***1***k, | | |  |  |
|  | (AB)−***1*** | | = | B−***1***A−***1*** |  |  |
| **и поэтому** |  |  |  |  |  |  |  |  |  |
| cond(AB) = | | (AB)−***1*** kABk ≤ cond A***1***· cond B. | | | | | | **(3.35)** |  |
|  |  | **если** C=AB**, то** A=CB | | | | − | **, и в силу доказанного** | |  |
| **С другой стороны,** | | |  |  |  |  |  |  |

**неравенства**

cond(A) ≤ cond(C) · cond(B−***1***) = cond(AB) · cond(B),

**коль скоро** cond(B−***1***) = cond(B)**. Поэтому** cond(AB) ≥ cond(A)/cond(B).

**Аналогичным образом из** B=CA−***1* следует** cond(AB) ≥ cond(B)/cond(A).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 3.3. Прямые методы решения линейных систем | | | |  | **169** |  |
| **Объединяя полученные неравенства, в целом получаем оценку** | | | | |  |  |
| cond(AB) ≥ max | cond(A) | | cond(B) | . | **(3.36)** |  |
|  | , |  |  |
| cond(B) | cond(A) |  |

**Неравенства (3.35)–(3.36) кажутся грубыми, но они достижимы. В самом деле, пусть** A **неособенная симметричная матрица с собственн-ными значениями** λ***1*,** λ***2*, . . . и числом обусловленности (стр. 150)**

max**i** |λ**i**(A)| . min**i** |λ**i**(A)|

**Ясно, что у матрицы** A***2* собственные векторы будут теми же, а соб-ственные значения равны** λ***21*,** λ***22*, . . . , так что число обусловленности матрицы** A***2* равно**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| min**ii**(λ**ii**(A))***2*** | = min**ii** | | ||λ**ii**(A)||***2*** | = | min**ii** | | ||λ**ii**(A)|| | | , |
| max (λ (A))***2*** |  | max | λ (A) ***2*** |  |  | max | λ (A) |  | ***2*** |

**и в верхней оценке (3.35) получаем равенство. Нижняя оценка (3.36) достигается, к примеру, при** B=A−***1*.**

**Оценка (3.35) показывает, в частности, что при преобразованиях и разложениях матриц число обусловленности может существенно расти. Рассмотрим, к примеру, решение системы линейных алгебраических уравнений** Ax=b **методом Гаусса в его матричной интерпретации. Обнуление поддиагональных элементов первого столбца матрицы** A **это умножение исходной СЛАУ слева на матрицу** E***1*, имеющую вид (3.28), так что мы получаем систему**

|  |  |
| --- | --- |
| (E***1***A) x = E***1***b | **(3.37)** |

**с матрицей** E***1***A**, число обусловленности которой оценивается как**

cond(E***1***A) ≤ cond(E***1***) cond(A).

**Далее мы обнуляем поддиагональные элементы второго, третьего и т. д. столбцов матрицы системы (3.37), умножая её слева на матрицы** E***2*,** E***3*, . . . ,** E**n**−***1*** **вида (3.29). В результате получаем верхнюю треугольную** **систему линейных уравнений**

U x = y,

**170** 3. Численные методы линейной алгебры

**в которой** U=E**n**−***1*** . . . E***2***E***1***A**,** y=E**n**−***1*** . . . E***2***E***1***b**, и число обусловлен-ности матрицы** U **оценивается сверху как**

cond(U ) ≤ cond(E**n**−***1***) · . . . · cond(E***2***) · cond(E***1***) · cond(A). **(3.38)**

**Если** E**j отлична от единичной матрицы, то** cond(E**j** )>1**, причём несмотря на специальный вид матриц** E**j правая и левая части неравен-ства (3.38) могут отличаться не очень сильно (см. примеры ниже). Как следствие, обусловленность верхней треугольной матрицы** U **, а также матриц, в которые преобразуется матрица** A **исходной СЛАУ на проме-жуточных шагах прямого хода метода Гаусса, может быть существенно хуже, чем у исходной матрицы** A**.**

Пример 3.3.1 **Для** 2 × 2**-матрицы (3.3)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| A = |  | 3 | 4 |  |
|  |  | 1 | 2 |  |

**число обусловленности равно** cond***2***(A) = 14.93**. Выполнение для неё преобразований прямого хода метода Гаусса приводит к матрице**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| A˜ = |  | 0 | 22 | , |
|  |  | 1 |  |  |

−

˜

**число обусловленности которой** cond***2***(A) = 4.27**, т. е. уменьшается. С другой стороны, для матрицы (3.4)**

|  |  |  |  |
| --- | --- | --- | --- |
| B = | −3 | 4 | , |
|  | 1 | 2 |  |

**число обусловленности** cond***2***(B) = 2.62**. Преобразования метода Гаусса превращают её в матрицу**

|  |  |  |  |
| --- | --- | --- | --- |
| B˜ = | 0 | 10 | , |
|  | 1 | 2 |  |

˜

**для которой число обусловленности уже равно** cond***2***(B) = 10.4**, т. е. существенно возрастает.**

**Числовые данные этого примера читатель может проверить с по-мощью систем компьютерной математики, таких как Scilab,** Matlab **и им аналогичные.**

|  |  |
| --- | --- |
| 3.3. Прямые методы решения линейных систем | **171** |

**Фактически, ухудшение обусловленности и, как следствие, всё боль-шая чувствительность решения к погрешностям в данных это допол-нительная плата за приведение матрицы (и всей СЛАУ) к удобному для решения виду. Можно ли уменьшить эту плату? И если да, то как?**

**Хорошей идеей является привлечение для матричных преобразова-ний ортогональных матриц, которые имеют наименьшую возможную обусловленность в спектральной норме (и небольшие числа обуслов-леннсти в других нормах). Умножение на такие матрицы, по крайней мере, не будет ухудшать обусловленность получающихся систем линей-ных уравнений и устойчивость их решений к ошибкам вычислений.**

**3.3ж** **QR-разложение матриц**

Определение 3.3.1 **Для матрицы** A **представление** A = QR **в виде** **произведения ортогональной матрицы** Q **и правой треугольной мат-рицы** R **называется QR-**разложением**.**

**По поводу этого определения следует пояснить, что правая тре-угольная матрица это то же самое, что верхняя треугольная матри-ца, которую мы условились обозначать** U **. Другая терминология обу-словлена здесь историческими причинами, и частичное её оправдание состоит в том, что QR-разложение матрицы действительно ¾совсем другое¿, нежели LU-разложение. Впрочем, в математической литера-туре можно встретить тексты, где LU-разложение матрицы называется ¾LR-разложением¿ (от английских слов left-right), т. е. разложением на левую и правую треугольные матрицы.**

Теорема 3.3.3 **QR-разложение существует для любой квадратной** **матрицы.**

Доказательство. **Если** A **неособенная матрица, то, как было пока-зано при доказательстве Теоремы 3.3.4,** A>A **симметричная положи-тельно определённая матрица. Следовательно, существует её разложе-ние Холесского**

A>A = R>R,

**где** R **правая (верхняя) треугольная матрица. При этом** R **очевидно**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **172** |  |  |  |  | 3. | Численные методы линейной алгебры | | |  |
| **неособенна. Тогда матрица** Q:=AR−***1* ортогональна, поскольку** | | | | | | | | |  |
|  | > |  | ***1*** |  | >AR−***1*** | ***1*** | ***1*** | ***1*** |  |
| Q |  | Q = AR−***1*** | |  | = (R−***1***)>A>A R−***1*** | |  |  |

= (R− )> R>R R− = (R− )>R> RR− = I.

**Следовательно, в целом** A=QR**, где определённые выше сомножители** Q **и** R **удовлетворяют условиям теоремы.**

**Рассмотрим теперь случай особенной матрицы** A**. Известно, что любую особенную матрицу можно приблизить последовательностью неособенных. Например, это можно сделать с помощью матриц** A**k** =A + **k*1*** I**, начиная с достаточно больших натуральных номеров** k**. При** **этом собственные значения** A**k суть** λ(A**k** ) =λ(A) + **k*1* , и если величина k*1* меньше расстояния от нуля до ближайшего собственного значения** A**, то** A**k** **неособенна.**

**В силу уже доказанного для всех матриц из последовательности** {A**k**} **существуют QR-разложения:**

A**k** = Q**k**R**k**,

**где все** Q**k ортогональны, а** R**k правые треугольные матрицы. В ка-честве ортогонального разложения для** A **можно было бы взять преде-лы матриц** Q**k и** R**k, если таковые существуют. Но сходятся ли куда-нибудь последовательности этих матриц при** k→ ∞**, когда** A**k** →A**? Ответ на это вопрос может быть отрицательным, а потому приходится действовать более тонко, выделяя из** {A**k**} **подходящую подпоследова-тельность.**

**Множество ортогональных матриц компактно, поскольку является замкнутым (прообраз единичной матрицы** I **при непрерывном отоб-ражении** X7→X>X**) и ограничено (**kXk***2*** ≤1**). Поэтому из последо-вательности ортогональных матриц** {Q**k**} **можно выбрать сходящуюся подпоследовательность** {Q**k**l}∞**l*=1*. Ей соответствуют подпоследователь-ности** {A**k**l} **и** {R**k**l}**, причём первая из них также сходится.**

**Обозначим** Q:= lim**l**→∞Q**k**l **, и это также ортогональная матрица. Тогда**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **l**→∞ | Q**k**l **k**l | **l**→∞ **k**l | · **l**→∞ | **k**l | = Q>A = R | | |  |
| lim | > A | = lim Q> | lim A |  |  |
| **правой треугольной матрице, поскольку все** Q>A**k** | | | | | |  | **были правыми** |  |
|  |  |  |  |  | **k**l | l |  |  |

**треугольными матрицами** R**k**l **. Таким образом, в целом снова** A=QR **с ортогональной** Q **и правой треугольной** R**, как и требовалось.**

|  |  |
| --- | --- |
| 3.3. Прямые методы решения линейных систем | **173** |

**Если известно QR-разложение матрицы** A**, то решение исходной СЛАУ, равносильной**

(QR)x = b

**сводится к решению треугольной системы линейных алгебраических уравнений**

|  |  |
| --- | --- |
| Rx = Q>b. | **(3.39)** |

**Ниже в §3.8и мы встретимся и с другими важными применениями QR-разложения матриц при численном решении проблемы собственных значений.**

**Хотя для неособенных матриц доказательство Теоремы 3.3.3 носит конструктивный характер, оно существенно завязано на разложение Холесского, а потому находить с его помощью QR-разложение неудоб-но. На практике основным инструментом получения QR-разложения является техника, использующая так называемые матрицы отражения и матрицы вращения, описанию которых посвящены следующие пунк-ты.**

**3.3з** **Ортогональные матрицы отражения**

Определение 3.3.2 **Для вектора** u ∈ R**n** **с единичной евклидовой нор-мой,** kuk***2*** = 1**, матрица** H=H(u) =I−2uu> **называется** матрицейотражений **или** матрицей Хаусхолдера**. Вектор** u **называется** порож-дающим **для матрицы отражений** H(u) **или** вектором Хаусхолдера**.**

Предложение 3.3.1 **Матрицы отражений являются ортогональны-ми симметричными матрицами, причём**

H(u) · u = −u,

H(u) · v = v **для любого вектора** v∈R**n, ортогонального** u**.**

Доказательство **проводится непосредственной проверкой.** **Симметричность матрицы** H(u)**:**

H> = I − 2uu> > = I> − 2uu> >

= I − 2 u> >u> = I − 2uu> = H.

**174** 3. Численные методы линейной алгебры

**Ортогональность:**

H>H = I − 2uu> I − 2uu>

* I − 2uu> − 2uu> + 4uu>uu>
* I − 4uu> + 4u(u>u)u> = I.

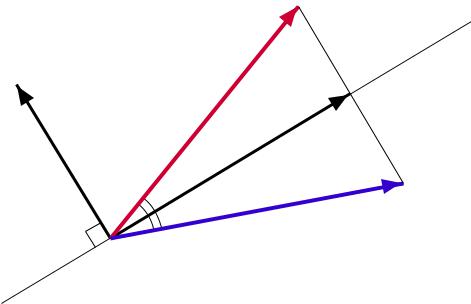
**Собственные векторы:**

H(u) · u = I − 2uu> u = u − 2u(u>u) = u − 2u = −u,

**т. е. порождающий вектор** u **является собственным вектором матрицы отражения, отвечающим собственному значению** −1**. Далее, если** u>v=0**, то**

H(u) · v = I − 2uu> v = v − 2u(u>v) = v,

**т. е. любой вектор** v**, ортогональный порождающему вектору** u**, являет-ся собственным вектором матрицы отражения, отвечающим собствен-ному значению** 1**.**



x

|  |  |  |
| --- | --- | --- |
| u | v |  |
|  |  |

Hx

**Рис. 3.6. Геометрическая интерепретация действия матрицы отражения.**

**Из последних двух свойств матриц отражения следует геометри-ческая интерпретация, которая мотивирует их название. Эти матрицы действительно осуществляют преобразование отражения вектора отно-сительно гиперплоскости, ортогональной порождающему вектору** u**.**

**Чтобы убедиться в этом, представим произвольный вектор** x **в виде** αu + v**, где** u **порождающий матрицу отражения вектор, а** v **ему** **ортогональный, т. е.** u>v= 0 **(см. Рис. 3.6). Тогда**

H(u) · x = H(u) · (αu + v) = −αu + v.

|  |  |
| --- | --- |
| 3.3. Прямые методы решения линейных систем | **175** |

Предложение 3.3.2 **Для любого ненулевого вектора** x ∈ R**n** **суще-ствует матрица отражений, переводящая его в вектор, коллинеар-ный заданному вектору** e **единичной длины,** kek***2*** = 1**.**

Доказательство. **Если** H **искомая матрица отражений, и** u **по-рождающий её вектор Хаусхолдера, то утверждение предложения тре-бует равенства**

Hx = x − 2 uu> x = γe **(3.40)**

**с некоторым коэффициентом** γ6= 0**. Отдельно рассмотрим два случая когда векторы** x **и** e **неколлинеарны, и когда они коллинеарны друг другу.**

**В первом случае можно переписать (3.40) в виде равенства**

|  |  |
| --- | --- |
| 2u u>x = x − γe, | **(3.41)** |

**правая часть которого заведомо не равна нулю. Тогда и числовой мно-житель** u>x **в левой части обязан быть ненулевым, и из соотношения (3.41) можно заключить, что**

1

u = 2u>x (x − γe),

**т. е. что вектор** u**, порождающий искомую матрицу отражения, должен быть коллинеарен вектору** (x−γe)**.**

**Коэффициент** γ **можно определить из того соображения, что ор-тогональная матрица** H **не изменяет длин векторов, а потому, взяв евклидову норму от обеих частей равенства (3.40), получим**

kHxk***2*** = kxk***2*** = |γ|, **т. е.** γ=±kxk***2***.

**Следовательно, вектор Хаусхолдера** u **коллинеарен вектору**

u˜ = x ± kxk***2*** e,

**и для окончательного определения** u **остаётся лишь применить норми-ровку:**

u = u˜ . ku˜k***2***

**Тогда** H=I−2uu> **искомая матрица отражений.**

**Обсудим теперь случай, когда** x **коллинеарен** e**. Тогда предшеству-ющая конструкция частично теряет смысл, так как вектор** u˜ =x−γe

**176** 3. Численные методы линейной алгебры

**может занулиться при подходящем выборе знака множителя** γ**. Но да-же если это произойдёт, существует ещё возможность выбора другого знака для** γ**, что спасает положение, поскольку при этом наверняка** x − γe 6= 0**. Более строго можно сказать, что тогда конкретный знак** **у множителя** γ=±kxk***2* выбирается из условия максимизации нормы вектора** (x−γe)**.**

**С другой стороны, в случае коллинеарных векторов** x **и** e **мы можем указать явную формулу для вектора Хаусхолдера:**

u = x . kxk***2***

**При этом**

u>x = x>x = kxk***2*** 6= 0, kxk***2***

**и для соответствующей матрицы отражений имеет место**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Hx = x − 2 uu> x = x − 2u u>x | = x − 2 kxk***2*** | x | | = −x. |  |
|  |  |  |
| kxk***2*** |  |
| **Итак, условие предложения удовлетворено и в этом случае.** | | | |  |  |

**В доказательстве предложения присутствовала неоднозначность в выборе знака в выражении** u˜ =x± kxk***2*** e**, если** x **и** e **неколлинеарны. В действительности, годится любой знак, и его конкретный выбор может определяться, как мы увидим, удобством реализации.**

**3.3и** **Метод Хаусхолдера**

**В основе метода Хаусхолдера для решения систем линейных алгебра-ических уравнений (называемого также методом отражений) лежит та же идея, что и в методе Гаусса: привести эквивалентными преобра-зованиями исходную систему к треугольному виду, а затем воспользо-ваться обратной подстановкой (3.26). Но теперь это приведение выпол-няется более глубокими, чем в методе Гаусса, преобразованиями мат-рицы, именно, путём последовательного умножения на специальным образом подобранные матрицы отражений.**

Предложение 3.3.3 **Для любой квадратной матрицы** A **существу-ет конечная последовательность** H***1*,** H***2*, . . . ,** H**n**−***1* из матриц отра-жения и, возможно, единичных матриц, таких что матрица**

H**n**−***1***H**n**−***2*** · · · H***2***H***1***A = R

|  |  |
| --- | --- |
| 3.3. Прямые методы решения линейных систем | **177** |

**является правой треугольной матрицей.**

**Для формального описания алгоритма очень удобно применять си-стему обозначений матрично-векторных объектов, укоренившуюся в языках программирования высокого уровня** Fortran**,** Matlab**, Scilab и др. В частности, посредством** A(p:q, r:s) **обозначается сечение мас-сива** A**, которое определяется как массив той же размерности с элемен-тами, стоящими на пересечении строк с номерами с** p **по** q **и столбцов с номерами с** r **по** s**. То есть, диапазоны изменения индексов элементов** A**, из которых образована новая матрица, указаны в соответствующих** **позициях.**

Доказательство **предложения конструктивно.**

**Действительно, используя результат Предложения 3.3.2, возьмём в качестве** H***1* матрицу отражений, которая переводит 1-й столбец** A **в вектор, коллинеарный** (1,0, . . . ,0)>**, если хотя бы один из элементов** a***21*,** a***31*, . . . ,** a**n*1*** **не равен нулю. Иначе полагаем** H***1*** = I**. Затем перехо-дим к следующему шагу.**

**В результате выполнения первого шага матрица СЛАУ приводится, как и в методе Гаусса, к виду**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | 0 | × | × · · · × | |  |
|  |  | × |  |
|  | | 0 | × × ·**.** ·**. .**· ×**...** | | |  |
|  | |  | × | × |  |  |
|  | | **.** | **. . .** | | **.** |  |
|  | | **..** | **..** | **..** | **.. ..** |  |
|  | |  |  |  |  |  |
|  | |  |  |  |  |  |
|  | | 0 |  |  |  |  |
|  | |  | × | × · · · × | |  |

.

**где крестиками ¾**×**¿ обозначены ненулевые элементы. Проделаем те-перь то же самое с матрицей** A(2 :n,2 :n)**, обнулив у неё поддиагональ-ные элементы первого столбца, который является вторым в исходной матрице. И т. д.**

**Далее определим** H**i,** i= 2,3, . . . , n−1**, как матрицу отражений, по-рождаемую вектором Хаусхолдера** u**i, который имеет нулевыми первые** i − 1 **компонент и подобран так, что** H**i** **аннулирует поддиагональные** **элементы** i**-го столбца в матрице** H**i**−***1*** · · ·H***2***H***1***A**.**

**Можно положить**

* 0 **!**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| H**i** = |  |  | , |  |
|  |  |  |  |  |
| 0 | | ˜ | |  |
|  | | H**i** | |  |

**178** 3. Численные методы линейной алгебры

**где в верхнем левом углу стоит единичная матрица размера** (j−1)×

˜

(j − 1)**, а** H**i** **матрица отражений размера** (n−j+ 1)×(n−j+ 1)**,**

**которая переводит вектор** A(j:n, j) **в** (n−j+ 1)**-вектор** (1,0, . . . ,0)>**, т. е. обнуляет поддиагональные элементы** j**-го столбца в** A**.**

**Таблица 3.1. Треугольное разложение матрицы с помощью отражений Хаусхолдера**

DO FOR j = 1 TO n − 1

**вычислить вектор Хаусхолдера** u**, отвечающий отражению, которое переводит вектор** A(j:n, j)

−

**в** (n j + 1)**-вектор** (1, 0, . . . , 0)>**;**

˜ >

H ← I − 2uu **;**

˜

A(i : n, i : n) ← H A(i : n, i : n)

END DO

**При реализации описанного алгоритма даже не нужно формиро-**

˜

**вать в явном виде матрицу отражений** H**, так как умножение на неё можно выполнить по экономичной формуле**

(I − 2uu>)A(i : n, i : n) = A(i : n, i : n) − 2u u>A(i : n, i : n) .

**Определённым недостатком метода Хаусхолдера и описываемого в следующем пункте метода вращений в сравнении с методом Гаусса яв-ляется привлечение неарифметической операции извлечения квадрат-ного корня, которая приводит к иррациональностям. Это не позволя-ет реализовать алгоритмы в поле рациональных чисел, к примеру, в программных системах так называемых ¾безошибочных вычислений¿, которые оперируют рациональными дробями с целочисленными чис-лителем и знаменателем.**

|  |  |
| --- | --- |
| 3.3. Прямые методы решения линейных систем | **179** |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **3.3к** | **Матрицы вращения** | | | | |  |  |  |  |  |
| **Матрицей вращения называется матрица** G(k, l, θ) **вида** | | | | | | | | | |  |
| 1 | **...** | · · · | | − |  |  |  | ← |  |  |
|  | 1 |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  | cos θ |  |  |  | sin θ |  |  |  | k**-ая строка** |  |
|  | **.** | **.** | **..** |  | **.** |  |  | , |  |  |
|  | **..** |  |  | **..** |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  | 1 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  | · · · | |  | **.** |  |  | ← |  |  |
|  |  |  | **..** |  |  |  |  |
|  | sin θ |  |  | cos θ | |  |  | l**-ая строка** |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  | 1 |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |

**где все не выписанные явно элементы вне главной диагонали равны ну-лю. Таким образом,** G(k, l, θ) **это матрица, которая отличается от еди-ничной матрицы лишь элементами, находящимися в позициях** (k, k)**,** (k, l)**,** (l, k) **и** (l, l) **для данных натуральных индексов** k**,** l**. Нетрудно** **проверить, что она ортогональна.**

**Матрица**

**!**

|  |  |
| --- | --- |
| cos θ | − sin θ |
| sin θ | cos θ |

**задаёт, как известно, вращение плоскости на угол** θ **вокруг начала ко-ординат. Матрица** G(k, l, θ) **также задаёт вращение пространства** R**n на угол** θ **вокруг оси, проходящей через начало координат и ортогональ-ной гиперплоскости** 0x**k**x**l. Матрицы вращения** G(k, l, θ) **называют так-же матрицами Гивенса, и мы будем обозначать их иногда посредством** G(k, l)**, если конкретная величина угла** θ **несущественна.**

**Если вектор** a= (a***1*** a***2***)> **ненулевой, то, взяв**

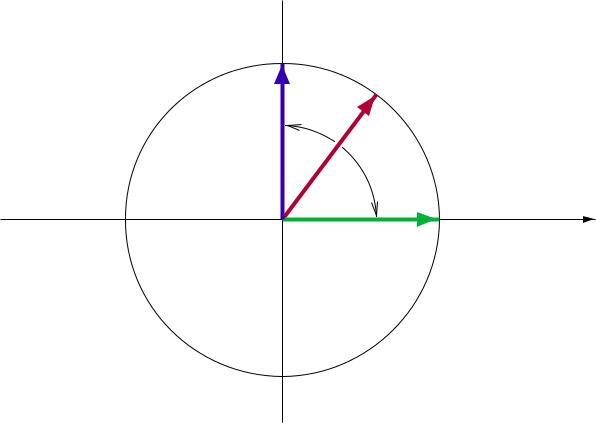
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | kak***2*** | |  | kak***2*** | |  | k | k***2*** |  | **q** |  |  |  |  |
|  |  |  |  | ***1*** | ***2*** |  |  |
| cos θ = | a***1*** | , | sin θ = | −a***2*** | , | **где** | a |  | = |  | a***2*** | + a***2*** | , |  |
|  |  |  |  |

**мы можем с помощью матрицы двумерного вращения занулить вторую компоненту вектора:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **!** | **!** |  | **!** |
| cos θ | − sin θ | a***1*** | = | kak***2*** . |
| sin θ | cos θ | a***2*** |  | 0 |

**180** 3. Численные методы линейной алгебры

 *x*2



|  |  |
| --- | --- |
| 0 | *x*1 |

**Рис. 3.7. Подходящим вращением можно занулить любую из компонент двумерного вектора.**

**Аналогично может быть занулена первая компонента** a**, путём домно-жения на матрицу вращения с**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| cos θ = | a***2*** | , | sin θ = | a***1*** | . |  |
|  |  |  |
|  | kak***2*** | |  | kak***2*** | |  |

**В общем случае умножение матрицы вращения** G(k, l, θ) **слева на любую матрицу** A **приводит к тому, что в произведении** G(k, l, θ)A **строки** k**-ая и** l**-ая становятся линейными комбинациями строк с этими же номерами из** A**. Как следствие рассуждений предшествующего аб-заца, при умножении матрицы** A **слева на матрицу вращения** G(k, l, θ) **со специально подобранным** θ **можно получить нуль в произвольной наперёд заданной позиции** k**-ой или** l**-ой строки. Поэтому любая дан-ная матрица** A **может быть приведена к правому треугольному виду** R **с помощью последовательности умножений на матрицы вращения.** **К примеру, мы можем один за другим занулить поддиагональные эле-мента первого столбца, потом второго, третьего и т.д., аналогично тому, как это делалось в методе Гаусса. При этом зануление поддиагональ-ных элементов второго и последующих столбцов никак не испортит по-лученные ранее нулевые элементы предшествующих столбцов, так как линейное комбинирование нулей даст снова нуль. В целом, существует набор матриц вращения** G(1,2)**,** G(1,3)**, . . . ,** G(1, n)**, . . . ,** G(n−1, n)**, таких что**

G(n − 1, n) · · · G(1, n) · · · G(1, 3) G(1, 2) A = R

|  |  |
| --- | --- |
| 3.3. Прямые методы решения линейных систем | **181** |

**правой треугольной матрице.**

**Это ещё один конструктивный способ получения QR-разложения, реализуемый даже более просто, чем метод отражений Хаусхолдера. Но для плотно заполненных матриц он в полтора раза более трудоё-мок, чем разложение с помощью матриц отражения, хотя для разре-женных матриц более предпочтителен в силу своей большей гибкости при занулении отдельных элементов.**

**3.3л** **Ортогонализация Грама-Шмидта**

**Ортогонализацией называют процесс построения по заданному базису линейного пространства некоторого ортогонального базиса, который имеет ту же самую линейную оболочку**

**по конечной линейно независимой системе векторов** v***1*,** v***2*,. . . ,** v**n процесс Грама-Шмидта строит ортогональный базис** q***1*,** q***2*, . . . ,** q**n ли-нейной облочки векторов** v***1*,** v***2*,. . . ,** v**n в соответствии со следующими расчётными формулами:**

**k**−***1***

q**k** ← v**k** − **X** hv**k** , q**i**i q**i**, k = 1, 2, . . . , n. **(3.42)**

**i*=1*** hq**i**, q**i**i

**3.3м** **Метод Холесского**

**Напомним, что** n×n**-матрица называется положительно определённой, если** hAx, xi>0 **для любых** n**-векторов** x**.**

Теорема 3.3.4 **Матрица** A **является симметричной положительно** **определённой тогда и только тогда, когда существует неособенная нижняя треугольная матрица** C**, такая что** A=CC>**. При этом матрица** C **из выписанного представления единственна.**

Определение 3.3.3 **Представление** A = CC> **называется** разложе-нием Холесского**, а нижняя треугольная матрица** C множителем Холесского **для** A**.**

Доказательство. **Если** A = CC>**, причём** C **неособенна, то**

hAx, xi = (Ax)> x = x>A>x = x> CC> >x

= x>CC>x = (C>x)>(C>x) = kC>xk***22*** > 0

**182** 3. Численные методы линейной алгебры

**для всех ненулевых** x∈R**n. Кроме того,** A **симмметрична по построе-нию. Таким образом, она является симметричной положительно опре-делённой матрицей.**7

**Обратно, пусть матрица** A **симметрична и положительно опреде-лена. В силу критерия Сильвестера все её ведущие миноры положи-тельны, а потому на основании Теоремы 3.3.1 о существовании LU-разложения мы можем заключить, что** A=LU **для некоторых неосо-бенных нижней треугольной матрицы** L= (l**ij** ) **и верхней треугольной матрицы** U **. Мы дополнительно потребуем, чтобы все диагональные элементы** l**ii в** L **были единицами, так что это разложение будет даже однозначно определённым.**

**Так как**

|  |  |  |  |
| --- | --- | --- | --- |
| **то** | LU = A = A> = LU > = U >L>, |  |  |
|  | U = L−***1***U >L>, |  |  |
| **и далее** | U L> −***1*** = L−***1***U >. |  |  |
|  | **(3.43)** |  |

**Слева в этом равенстве стоит произведение верхних треугольных мат-риц, а справа произведение нижних треугольных. Равенство, сле-довательно, возможно, лишь в случае, когда левая и правая его ча-сти это диагональная матрица, которую мы обозначим через** D=diag {d***1***, d***2***, . . . , d**n**}**. Тогда из (3.43) вытекает**

U = L−***1***U >L> = DL>,

**и потому**

|  |  |
| --- | --- |
| A = LU = LDL>. | **(3.44)** |

**Ясно, что в силу неособенности** L **и** U **матрица** D **также должна быть неособенна, так что по диагонали у неё стоят ненулевые элементы** d**i,** i = 1, 2, . . . , n**. Более того, все** d**i** **должны быть положительными.** **Если предположить противное, т. е. что** d**i** ≤0 **для некоторого** i**, то, беря** i**-ый столбец единичной матрицы в качестве вектора** x**, получим**

|  |  |
| --- | --- |
| **X** |  |
| hAx, xi = x>Ax = x>LDL>x = d**i** l**ij*2*** ≤ 0. | **(3.45)** |

**j**≤**i**

* + **Это рассуждение никак не использует факт треугольности** C **и обосновывает,**
* **действительности, более общее утверждение: произведение матрицы на её транс-понированную является положительно определённым.**

|  |  |
| --- | --- |
| 3.3. Прямые методы решения линейных систем | **183** |

**Неравенство (3.45) очевидно противоречит положительной определён-ности матрицы** A**.**

**Как следствие, из диагональных элементов матрицы** D **можем из-**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **влекать квадратные корни. Если обозначить получающуюся при этом** | | | | | | | |  |
| ***1*/*2*** | √ |  | √ |  | √ |  |  |  |
|  |  |  |  |
| **диагональную матрицу через** D***1*/*2*** | = diag { | d***1***, | | d***2***, . . . , d**n**}**, то окон-** | | | |  |
| **чательно можем взять** C=LD | **. Это представление для множителя** | | | | | | |  |
| **Холесского, в действительности, единственно, так как по** A **при сделан-** | | | | | | | |  |

**ных нами предположениях единственным образом определяется ниж-няя треугольная матрица** L**, а матричные преобразования, приведшие к формуле (3.44) и её следствиям, также дают однозначно определён-ный результат.**

**Доказанная теорема мотивирует следующий прямой метод решения систем линейных уравнений, аналогичный (3.31). Если разложение Хо-лесского уже найдено, то решение исходной СЛАУ** Ax=b**, равносиль-ной** CC>x=b**, сводится к решению двух систем линейных уравнений с треугольными матрицами:**

**(**

Cy = b,

**(3.46)**

C>x = y.

**Для решения первой системы применяем прямую подстановку, а для решения второй системы обратную.**

**Как конструктивно найти разложение Холесского?**

**Для определения расчётных формул нового метода рассмотрим ра-венство**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | A = CC>, | | | |  |  |  |  |
| **где** |  | c***21*** | c***22***0 | | |  |  |  |
|  |  |  |  |
|  |  | c***11*** |  | **..... .** | |  | . |  |
| A = (a**ij** ), | C = | **...** |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  | c | c | **n*2*** |  | c |  |  |
|  |  | **n*1*** |  | · · · |  | **nn** |  |
|  |  |  |  |  |  |  |  |
| **В силу правил умножения матриц имеем** | | | | |  |  |  |  |
|  | **j** |  |  |  |  |  |  |  |
| a**ij** = | **X** |  |  | **при** i≥j, | |  |  |  |
| c**ik**c**jk** | |  |  |  |  |

**k*=1***

**184** 3. Численные методы линейной алгебры

**или, в подробной записи,**

c***2*j*1*** + c***2*j*2*** + . . . + c***2*j,j**−***1*** + c***2*jj** c**i*1***c**j*1*** + c**i*2***c**j*2*** + . . . + c**ij** c**jj**

= a**jj** ,

= a**ij** , i = j + 1, . . . , n, **(3.47)** j = 1, 2 . . . , n.

**Из выписанной системы уравнений, в действительности, можно нахо-дить в матрице** C **элементы** c**ij последовательно друг за другом по столбцам. Именно,**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **при** j= 1 | **(** c**i*1*** | = a**i*1***/c***11***,i = 2, 3, . . . , n, | | |
|  | c***11*** | = √ | a***11*** | , |

* **p**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **при** j= 2 | | c***22*** = | a***22*** − c***212***, | |  |  |  |  |  |  |  |
|  |  | c**i*2*** = a**i*2*** − c**i*1***c***21*** /c***22***, | | | | | i = 3, 4, . . . , n, | | | |  |
| · · · |  |  | · · · |  |  |  |  |  |  |  |  |
| **Псевдокод этого процесса выглядит следующим образом** | | | | | | | | | | |  |
|  |  |  | |  |  |  |  |  |  |  |  |
|  |  | DO FOR j = 1 TO | | n |  |  |  |  |  |  |  |
|  |  | c**jj** ← | a**jj** − | c**jk*2*** |  | ***1*/*2*** |  |  |  |  |  |
|  |  | **!** | |  |  |  |  |  |
|  |  |  |  | **j**−***1*** |  |  |  |  |  |  |  |
|  |  |  |  | **X** |  |  |  |  |  |  |  |
|  |  |  |  | **k*=1*** |  |  |  |  |  |  |  |
|  |  | DO FOR | i = j + 1 TO | | | n |  |  | **(3.48)** | |  |
|  |  | c**ij** | ← a**ij** − **j**−***1*** | | | c**ik**c**jk** | **!** | c**jj** |  |
|  |  |  | |  |
|  |  |  |  | **X** | |  |  |  |  | |  |
|  |  |  |  | **k*=1*** | |  |  |  |  |  |  |
|  |  | END DO |  |  |  |  |  |  |  |  |  |
|  |  | END DO |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |

**Если** A **симметричная положительно определённая матрица, то в силу доказанных выше результатов система (3.47) обязана иметь ре-шение, и этот алгоритм успешно прорабатывает до конца, находя его. Если же матрица** A **не является положительно определённой, то алго-ритм (3.48) аварийно прекращает работу при попытке извлечь корень**

|  |  |
| --- | --- |
| 3.3. Прямые методы решения линейных систем | **185** |

**из отрицательного числа либо разделить на нуль. Это самый эконо-мичный способ проверки положительной определённости симметрич-ной матрицы.**

**Метод решения СЛАУ, основанный на разложении Холесского и использующий соотношения (3.46) и алгоритм (3.48), называют мето-дом Холесского. Он были предложен в 1910 году А.-Л. Холесским в неопубликованной рукописи, которая, тем не менее, сделалась широко известной во французской геодезической службе. Позднее метод неод-нократно переоткрывался, и потому иногда в связи с ним используют-ся также термины ¾метод квадратного корня¿ или ¾метод квадратных корней¿, данные другими его авторами.**

**Замечательным свойством метода Холесского является то, что обу-словленность множителей Холесского, вообще говоря, является луч-шей, чем у исходной СЛАУ, так как равна корню квадратному из обу-словленности исходной матрицы СЛАУ. То есть, в отличие от метода Гаусса, треугольные системы линейных уравнений из (3.46), к реше-нию которых сводится задача, менее чувствительны к ошибкам, чем исходная линейная система.**

**3.3н** **Метод прогонки**

**До сих пор не делалось никаких дополнительных предположений о структуре нулевых и ненулевых элементов в матрице системы. Но для большого числа систем линейных уравнений, встречающихся в прак-тике математического моделирования ненулевые элементы заполняют матрицу не полностью, образуя в ней те или иные правильные струк-туры ленты, блоки, их комбинации и т. п. Естественно попытаться использовать это обстоятельство при конструировании более эффек-тивных численных методов для решения СЛАУ с такими матрицами.**

**Метод прогонки, предложенный в 1952–53 годах И.М. Гельфандом и О.В. Локуциевским, предназначен для решения линейных систем урав-нений с трёхдиагональными матрицами. Это важный в приложениях случай СЛАУ, возникающий, к примеру, при решении многих краевых задач для дифференциальных уравнений. По определению, трёхдиа-гональными называются матрицы, все ненулевые элементы которых сосредоточены на трёх диагоналях главной и двух соседних с ней. Иными словами, для трёхдиагональной матрицы** A= (a**ij** ) **неравенство** a**ij** =6 0 **имеет место лишь при** i = j **и** i = j ± 1**.**

**Обычно трёхдиагональные системы линейных уравнений записыва-**

**186** 3. Численные методы линейной алгебры

**ют в некотором каноническом виде, даже без обращения к матрично-векторной форме:**

|  |  |
| --- | --- |
| a**i**x**i**−***1*** + b**i**x**i** + c**i**x**i*+1*** = d**i**, 1 ≤ i ≤ n, | **(3.49)** |

**причём полагают** a***1*** =c**n** = 0**. Подобный вид и обозначения оправ-дываются тем, что соответствующие СЛАУ получаются действитель-но ¾локально¿, как дискретизация дифференциальных уравнений, свя-зывающих значения искомых величин также локально, в окрестности какой-либо рассматриваемой точки.**

**Например, в §2.4 мы могли видеть, что**

u00(x) ≈ u**i**−***1*** − 2u**i** + u**i*+1*** , h***2***

**и потому решение конечно-разностными методами краевых задач для различных дифференциальных уравнений второго порядка приводит к трёхдиагональным матрицам и системам с такими матрицами. Соот-ношения вида (3.49)**

a**i**x**i**−***1*** + b**i**x**i** + c**i**x**i*+1*** = d**i**, i = 1, 2, . . . ,

**называют также трёхточечными разностными уравнениями или раз-ностными уравнениями второго порядка.**

**Для СЛАУ с трёхдиагональной матрицей прямой ход метода Гаусса без выбора ведущего элемента (т. е. без перестановок строк и столбцов матрицы) приводит к системе с двухдиагональной матрицей вида**

|  |  |  |  |
| --- | --- | --- | --- |
| × | × | **...** |  |
|  | × | **...** |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  | 0 |  |  |
|  |  |  |
|  |  |  |  |
|  |  |  |  |

* ,
* ×

×

**в которой ненулевые элементы присутствуют лишь на главной и первой верхней побочной диагоналях. Следовательно, формулы обратного хо-да метода Гаусса вместо (3.26) должны иметь следующий двучленный вид**

|  |  |
| --- | --- |
| x**i** = ξ**i*+1***x**i*+1*** + η**i*+1***,i = n, n − 1, . . . , 1, | **(3.50)** |

x**n*+1***

|  |  |
| --- | --- |
| 3.3. Прямые методы решения линейных систем | **187** |

**где для единообразия формул рассматривается вспомогательная фик-тивная компонента** x**n*+1* вектора неизвестных. Оказывается, что вели-чины** ξ**i и** η**i в соотношениях (3.50) можно несложным образом выра-зить через элементы исходной системы уравнений.**

**Уменьшим в (3.50) все индексы на единицу**

x**i**−***1*** = ξ**i**x**i** + η**i**

**и подставим полученное соотношение в** i**-ое уравнение системы, по-лучим**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |
| a**i** ξ**i**x**i** + η**i** + b**i**x**i** + c**i**x**i*+1*** = d**i**. | | | | | | |  |
| **Отсюда** |  | c**i** |  |  | d**i** − a**i**η**i** |  |  |
| x = |  | x | + | . |  |
| − a**i**ξ**i** + b**i** | |  |  |
| **i** | **i*+1*** |  | a**i**ξ**i** + b**i** | |  |

**Сравнивая эту формулу с двучленными расчётными формулами (3.50), можем заключить, что**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ξ**i*+1*** | = − | | c**i** | | , | **(3.51)** |  |
| a**i**ξ**i** + b**i** | |  |
| η**i*+1*** | = | d**i** − a**i**η**i** | | , |  | **(3.52)** |  |
|  |  | a**i**ξ**i** + b**i** | | |  |  |  |

**для** i= 1,2, . . . , n**. Это формулы прямого хода прогонки. Вместе с формулами обратного хода (3.50) они определяют метод прогонки для решения систем линейных алгебраических уравнений с трёхдиагональ-ной матрицей.**

**Для начала расчётов требуется знать величины** ξ***1* и** η***1* в прямом ходе и** x**n*+1* в обратном. Формально они неизвестны, но фактически полностью определяются условием** a***1*** =c**n** = 0**. Действительно, кон-кретные значения** ξ***1* и** η***1* не влияют на результаты решения, потому что в формулах прямого хода прогонки они встречаются с множителем** a***1*** = 0**. Кроме того, из формул прямого хода следует, что**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | c**n** | 0 | |  |  |
| ξ**n*+1*** = − |  | = − |  | = 0, |  |
| a**n**ξ**n** + b**n** | a**n**ξ**n** + b**n** |  |

**а это коэффициент при** x**n*+1* в обратном ходе прогонки. Поэтому и**

**может быть произвольным. Итак, для начала прогонки можно**

**положить**

|  |  |
| --- | --- |
| ξ***1*** = η***1*** = x**n*+1*** = 0. | **(3.53)** |

**188** 3. Численные методы линейной алгебры

**Дадим теперь достаточные условия осуществимости метода прогон-ки, т. е. того, что знаменатели в расчётных формулах прямого хода не обращаются в нуль.**

Предложение 3.3.4 **Если в системе линейных алгебраических урав-нений с трёхдиагональной матрицей (3.49) имеет место диагональ-ное преобладание, т. е.**

|b**i**| > |a**i**| + |c**i**|, i = 1, 2, . . . , n,

**то метод прогонки c выбором начальных значений в соответствии с (3.53) является реализуемым.**

Доказательство. **Покажем по индукции, что в рассматриваемой реа-лизации прогонки неравенство** |ξ**i**|<1 **справедливо для всех** i**. Прежде всего,** ξ***1*** = 0 **и потому** |ξ***1***|<1**. Далее, предположим, что для некото-рого индекса** i **уже установлена оценка** |ξ**i**|<1**. Если соответствующее** c**i** = 0**, то из (3.51) следует** ξ**i*+1*** = 0**, и индукционный переход доказан.** **Поэтому пусть** c**i** =60**. Тогда справедлива следующая цепочка соотно-шений**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ξ |  | = |  | | c**i** |  | = |  |  |  | c**i**| |  |  |  |
| | | − a**i**ξ**i** + b**i** | | |  |  | |a**i**ξ|**i** + b**i**| | | | |  |  |
| | **i*+1*** |  |  |  |  |  |
|  |  |  |  | |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  | | |c**i**| |  |  |  |  | **из оценки снизу для модуля суммы** | | | |  |
|  |  | ≤ |b**i**| − |a**i**| · |ξ**i**| | | | | | | | |  |
|  |  |  |  |  |  |  |
|  |  | < |  |  | |c**i**| | |  | | |  | **в силу диагонального преобладания** | | |  |
|  |  |  |  |  |  |
|  |  |  |  | |a**i**| + |c**i**| − |a**i**| · |ξ**i**| | | | | | | | | |  |  |
|  |  | = |  |  | |c**i**| |  |  |  |  |  | |c**i**| |  | = 1, |  |
|  |  |  | |a**i**|(1 − |ξ**i**|) + |c**i**| | | | | | | ≤ |c**i**| | | |  |
|  |  |  |  |  |  |

**где при переходе ко второй строке мы воспользовались известным нера-венством для модуля суммы двух чисел:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| |x + y| ≥ |  | |x| − |y| |  | . | **(3.54)** |
|  |  |  |  |  |  |

**Итак,** |ξ**i**|<1 **оказывается доказанным для всех прогоночных коэффи-циентов** ξ**i,** i= 1,2, . . . , n+ 1**.**

|  |  |
| --- | --- |
| 3.3. Прямые методы решения линейных систем | **189** |

**Как следствие, для знаменателей прогоночных коэффициентов** ξ**i и**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| η**i** **в формулах (3.51)–(3.52) имеем** | | | | | | | | |  |  |
| a**i**ξ**i** + b**i** |  | ≥ | |b**i**| − |a**i**ξ**i** | | | | | | **по неравенству (3.54)** |  |  |
|  |  | = | b**i** | | − | | a**i**ξ**i** | | |  | **из-за диагонального преобладания** |  |  |
|  |  |  | | |  |  |  |  |  |
|  |  | > | |a**i**| + |c**i**| − |a**i**| · |ξ**i**| **из-за диагонального преобладания** | | | | | | |  |
|  |  | = |a**i**|(1 − |ξ**i**|) + |c**i**| | | | | | | |  |  |
|  |  | ≥ | |c**i**| ≥ 0 | | |  | **в силу оценки** |ξ**i**|<1, | |  |  |
| **то есть строгое отделение от нуля. Это и требовалось доказать.** | | | | | | | | |  |  |

**Отметим, что существуют и другие условия реализуемости метода прогонки. Например, некоторые из них требуют ¾более мягкое¿ нестро-гое диагональное преобладание в матрице, но зато другие более жёст-кие условия на коэффициенты системы. Весьма полуярна, в частности, такая формулировка [16]:**

Предложение 3.3.5 **Если в трёхдиагональной системе линейных ал-гебраических уравнений (3.49) побочные диагонали не содержат нулей,**

**т. е.** a**i** 6= 0**,** i= 2,3, . . . , n**, и** c**i** 6= 0**,** i= 1,2, . . . , n−1**, имеет место нестрогое диагональное преобладание**

|b**i**| ≥ |a**i**| + |c**i**|, i = 1, 2, . . . , n,

**но хотя бы для одного** i **это неравенство является строгим, то метод прогонки реализуем.**

**Разработано немало различных модификаций метода прогонки, ко-торые хорошо приспособлены для решения тех или иных специальных систем уравнений, как трёхдиагональных, так и более общих, имеющих ленточные или даже блочно-ленточные матрицы [16]. В частности, су-ществует матричная прогонка.**

**Нетрудно убедиться, что реализация прогонки требует линейного в зависимости от размера системы количества арифметических операций (примерно** 8n**), т. е. весьма экономична.**

**190** 3. Численные методы линейной алгебры

1. Стационарные итерационные методы решения систем линейных уравнений

**В итерационных методах решения уравнений и систем уравнений ре-шение** x**? получается как предел некоторой последовательности при-ближений** {x***(*k*)***}∞**k*=0*, так что**

x**?** = lim x***(*k*)***.

**k**→∞

**Естественно, что на практике переход к пределу невозможен в силу ко-нечности наших вычислений, и потому при реализации итерационных методов вместо** x**? обычно довольствуются нахождением какого-то до-статочно хорошего приближения** x***(*k*)* к** x**?. Подробнее мы рассмотрим этот вопрос в §3.4ж.**

**Общая схема итерационных методов выглядит следующим образом: выбираем одно или несколько начальных приближений** x***(0)*,** x***(1)*, . . . ,** x***(*m*)*, а затем по их известным значениям последовательно вычисляем**

x***(*k*+1)*** ← T**k**(x***(0)***, x***(1)***, . . . , x***(*k*)***), k = m, m + 1, m + 2, . . . , **(3.55)**

**где** T**k отображение, называемое оператором перехода или операто-ром шага (**k**-го). Конечно, в реальных итерационных процессах каждое следующее приближение, как правило, зависит не от всех предшеству-ющих приближений, а лишь от какого-то их фиксированного конечного числа. Более точно, итерационный процесс (3.55) называют** p**-шаговым, если его последующее приближение** x***(*k*+1)* является функцией только от** x***(*k*)*,** x***(*k** −***1)*, . . . ,** x***(*k**−**p*+1)*. В частности, наиболее простыми в этом отношении являются одношаговые итерационные методы**

|  |  |
| --- | --- |
| x***(*k*+1)*** ← T**k**(x***(*k*)***), | k = 0, 1, 2, . . . , |

**в которых** x***(*k *+1)* зависит лишь от значения одной предшествующей итерации** x***(*k*)*. Для начала работы одношаговых итерационных процес-сов нужно знать лишь одно начальное приближение** x***(0)*.**

**Итерационный процесс называется стационарным, если оператор перехода** T**k не зависит от номера шага** k**, т. е.** T**k** =T **, и нестационар-ным в противном случае. Линейным итерационным процессом будут называться итерации, в которых оператор перехода имеет вид**

T**k**(x***(*k*)*** , x***(*k**−***1)***, . . . , x***(*k**−**p*+1)*** )

=C***(*k,k*)***x***(*k*)*** +C***(*k,k**−***1)***x***(*k**−***1)*** +. . .+C***(*k,k**−**p*+1)*** x***(*k**−**p*+1)*** +d

|  |  |
| --- | --- |
| 3.4. Итерационные методы для систем линейных уравнений | **191** |

**с какими-то матрицами** C***(*k,k*)*,** C***(*k,k**−***1)*, . . . ,** C***(*k,k**−**p*+1)* подходящих раз-меров.**

**Причины, по которым для решения систем линейных уравнений итерационные методы могут оказаться более предпочтительными, чем прямые, заключаются в следующем. Большинство итерационных мето-дов являются самоисправляющимися, т. е. такими, в которых погреш-ность, допущенная в вычислениях, при сходимости исправляется в ходе итерирования и не отражается на окончательном результате. Это сле-дует из конструкции оператора перехода, в котором обычно по самому его построению присутствует информация о решаемой системе, что мы увидим далее на примерах. При выполнении алгоритма эта информа-ция на каждом шаге вносится в итерационный процесс и оказывает влияние на его ход. Напротив, прямые методы решения СЛАУ этим свойством не обладают, так как, оттолкнувшись от исходной системы, мы далее уже не возвращаемся к ней, а оперируем с её следствиями, которые никакой обратной связи от исходной системы не получают.**

**Нередко итерационные процессы сравнительно несложно програм-мируются, так как представляют собой повторяющиеся единообразные процедуры, применённые к последовательным приближениям к реше-нию. При решении СЛАУ с разреженными матрицами в итерационных процессах можно легче, чем в прямых методах, учитывать структу-ру нулевых и ненулевых элементов матрицы и основывать на этом упрощённые формулы матрично-векторного умножения, которые су-щественно уменьшают общую трудоёмкость алгоритма. Наконец, быст-ро сходящиеся итерационные методы могут обеспечивать выигрыш по времени и для СЛАУ общего вида, если требуют число итераций, мень-шее размерности системы.**

**3.4а** **Условие сходимости стационарных одношаговых итерационных методов**

Теорема 3.4.1 **Стационарный одношаговый итерационный процесс**

|  |  |  |
| --- | --- | --- |
| x***(*k*+1)*** ←Cx***(*k*)*** +d, | k = 0, 1, 2, . . . , | **(3.56)** |

**сходится при любом начальном приближении** x***(0)* тогда и только то-гда, когда спектральный радиус матрицы** C **меньше единицы, т. е.**

ρ(C) < 1**.**

Доказательство **этого утверждения будет разбито на две части, ре-зультат каждой из которых представляет самостоятельный интерес.**

**192** 3. Численные методы линейной алгебры

Предложение 3.4.1 **Если** kCk < 1 **в какой-нибудь матричной норме,** **то стационарный одношаговый итерационный процесс**

x***(*k*+1)*** ← Cx***(*k*)*** + d, k = 0, 1, 2, . . . ,

**сходится при любом начальном приближении** x***(0)*.**

Доказательство. **Если** kC k < 1 **для какой-нибудь матричной нор-мы, то в силу Предложения 3.2.8 матрица** (I−C) **неособенна и имеет обратную. Следовательно, система уравнений** (I−C)x=d**, как и рав-носильная ей**

x = Cx + d,

**имеют единственное решение, которое мы обозначим** x**?:**

x**?** = Cx**?** + d.

**Вычтем это равенство из соотношения** x***(*k*)*** =Cx***(*k**−***1)*** +d**,** k= 0,1,2, . . . **, получим**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| x***(*k*)*** −x**?**=C x***(*k**−***1)*** −x**?**. | | | | | | | |  |  |  |  |
| **Взятие нормы от обеих частей** | **приводит к цепочке неравенств** | | | | | | | | | |  |
|  |  | |  |  |  |  |  |  | |  |
|  |  |  |  | ***2*** | ***(*k *2)*** |  |  |  |
| kx***(*k*)*** − x**?**k ≤ kCk · x***(*k**−***1)*** − x**?** | | | | | | | | |  |  |  |
| ≤ k | | C | k |  | x | − | − | x | | |  |
|  |  | · |  |  | **?** |  |  |
|  |  |  |  |  | . . . | |  |  |  |  |
| ≤ . . . | | | |  |  |  |  |  |
| ≤ kCk**k** ·x***(0)*** −x**?** , | | | | | | | | | | **(3.57)** |  |
| **при выводе которой мы пользуемся** | | | | **тем фактом, что всякая матрич-** | | | | | | |  |
|  |  |  |  |  |  |  |  |

**ная норма согласована с некоторой векторной нормой и именно в этой**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **норме оцениваем отклонение** x***(*k*)* от** x**?.** | | | | k |  |  |  | **си-** | |  |
| **Правая часть неравенства (3.57) сходится к нулю при** | | | | → ∞ | | **в** |  |
|  | ***(*k*)*** |  |  |
| **лу условия** k | C | k < 1**, поэтому** | **последовательность приближений** | | | k | x |  | k |  |
|  | **?** |  |  |  |  |  |
| **действительно сходится к пределу** x **.** | | | |  |  |  |  |  | |  |

**Побочным следствием доказательства Предложения 3.4.1 является демонстрация роли нормы матрицы перехода** kCk **как коэффициента подавления погрешности приближений к решению СЛАУ, что следует из неравенств (3.57): чем меньше** kCk**, тем быстрее убывает эта погреш-ность на каждом отдельном шаге итерационного процесса.**

|  |  |
| --- | --- |
| 3.4. Итерационные методы для систем линейных уравнений | **193** |

Предложение 3.4.2 **Для любой квадратной матрицы** A **и любого** > 0 **существует такая подчинённая матричная норма** k · k **, что**

ρ(A) ≤ kAk ≤ ρ(A) + .

Доказательство. **Левое из выписанных неравенств было обосновано** **ранее в Предложении 3.2.6, и потому содержанием сформулированного результата является, фактически, правое неравенство, дающее оценку сверху для нормы матрицы через спектральный радиус.**

**С помощью преобразования подобия приведём матрицу** A **к жорда-новой канонической форме**

S−***1***AS = J,

**где**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | λ***1*** | | 1 | **..** |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | | λ***1*** | **..** | **.** | 1 | 0 | |  |  |  | 0 |  |  |  |  |
|  |  | |  |  | **.** |  |  |  |  |  |  |  |  |  |  |
|  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | |  |  |  | λ***1*** |  |  |  |  |  |  |  |  |  |  |
|  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| J = |  | |  |  |  |  | λ***2*** 1 | |  |  |  |  |  |  | , |  |
|  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | 0 | |  |  | **..** | **.** | **..** | **.** |  | 0 |  |  |  |
|  |  | |  |  |  |  |  |  |  |  |  |
|  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | |  |  |  |  |  |  | λ***2*** | |  |  |  |  |  |  |
|  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  | **..** | **.** |  |  |  |
|  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | | 0 | |  |  | 0 | |  |  |  | **..** | **.** |  |  |  |
|  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

**а** S **некоторая неособенная матрица, осуществляющая преобразова-ние подобия.**

**Положим**

D := diag {1, , ***2***, . . . , **n**−***1***}

**диагональной матрице с числами** 1**, , *2*, . . . , n**−***1* по главной диа-**

**194** 3. Численные методы линейной алгебры

**гонали. Тогда нетрудно проверить, что**

(SD )−***1***A(SD ) = D−***1***(S−***1***AS)D = D−***1***J D

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | | ***1*** | λ | **...** |  |  |  |  |  |  |  |  |  |  |  |
|  |  | λ |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | |  | ***1*** | **...** |  |  |  |  |  |  |  |  |  |  |  |
|  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | |  |  |  | λ***1*** |  |  |  |  |  |  |  |  |  |  |
|  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| = |  | |  |  |  |  | λ***2*** |  |  |  |  |  |  |  | , |  |
|  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  | **. .** | **.** | **. .** | **.** |  |  |  |  |  |
|  |  | |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | |  |  |  |  |  |  | λ***2*** | |  |  |  |  |  |  |
|  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  | **.** | **..** |  |  |  |
|  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | |  |  |  |  |  |  |  |  |  | **.** | **..** |  |  |  |
|  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

**матрица в ¾модифицированной¿ жордановой форме, которая отли-чается от обычной жордановой формы присутствием вместо** 1 **на верх-ней побочной диагонали. Действительно, умножение на диагональную матрицу слева это умножение строк матрицы на соответствующие диагональные элементы, а умножение на диагональную матрицу спра-ва равносильно умножению столбцов на элементы диагонали. Два та-ких умножения на** D−***1* слева и на** D **справа компенсируют друг друга на главной диагонали матрицы. Но на верхней побочной диагона-ли от этих умножений остаётся множитель** = **i*+1*** −**i,** i= 0,1, . . . , n−1**.**

**Определим векторную норму**

kxk := (SD )−***1***x ∞.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 3.4. Итерационные методы для систем линейных уравнений | | | | | | | | | | | | | | | | | | | | | | | |  | **195** |  |
| **Тогда для подчинённой ей матричной нормы имеет место** | | | | | | | | | | | | | | | | | | | | | | | |  |  |  |
| k k |  | **x**6***=0*** |  | k | x |  | |  |  | **x**6***=0***(SD )−***1***x | | | | | | | | | | |  | | |  |  |  |
| A | = max | |  | Axk = max | | | | | | | | | |  | (SD )−***1***Ax | | | | | | | ∞ | |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  | | |  |  |  |  |  |  |  | | |  |  |  |
|  |  |  |  | k k | |  |  | ***1*** |  |  |  |  |  |  | |  |  |  |  |  | |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  |  | | ∞ | | ***1*** |  |  |
|  |  |  |  |  | (SD )− A(SD | | | | | | | | |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | = | 6 |  |  |  |  |  | k | k∞ | | | | | )y | |  |  |  |  | **после замены** y= (SD)−x | | | | | |  |
|  | **x*=0*** |  |  |  |  |  | y |  |  |  |  |  |  |  | ∞ | | |  |  |
|  |  | max |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  | ∞ | | |  |  |  |  |  | − |  |  |  |  |  |  |  |
|  | = **x**6***=0*** | |  |  |  | y |  |  |  |  |  |  |  |  |  | ∞ | | |  |  |  |
|  |  | max |  |  | (D−***1***J D )y | | | | | |  | | | = | |  | D | |  | ***1***J D |  |  | |  |  |  |
|  |  |  |  | k k∞ | | | |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |
|  | = **максимум сумм модулей элементов в** D−***1***J D **по строкам** | | | | | | | | | | | | | | | | | | | | | | | | |  |
|  | ≤ | max λ (A) + | | | | | | |  | = ρ(A) + , | | | | | | | | | |  |  |  |  |  |  |  |
|  | **i** | | | | **i** | | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **где** λ**i**(A)i**-ое собственное значение матрицы** A**.** | | | | | | | | | | | | | | | | | | | | | | | |  |  |  |

Доказательство **теоремы о сходимости одношагового стационарного** **итерационного процесса. Сначала покажем необходимость условия тео-ремы.**

**Пусть порождаемая в итерационном процессе последовательность** {x***(*k*)***} **сходится. Её пределом при этом может быть только решение** x**?** **системы** x=Cx+d**, т. е. должно быть** lim**k**→∞x***(*k*)*** =x**?, в чём можно убедиться, переходя в соотношении** x***(*k*+1)*** =Cx***(*k*)*** +d **к пределу по** k → ∞**. Тогда, вычитая из расчётной формулы итерационного процесса** **равенство для решения, получим**

x***(*k*)***

**откуда**

x***(*k*)***

* + x**?** = C(x***(*k**−***1)*** − x**?**),
* x**?** = C(x***(*k**−***1)*** − x**?**)

=C***2***(x***(*k**−***2)*** −x**?**)

= · · · · · ·

= C**k**(x***(0)*** − x**?**).

**Так как левая часть этих равенств при** k→ ∞ **сходится к нулю, то должна сходиться к нулю и правая, причём при любом векторе** x***(0)*. Это возможно лишь когда** C**k** →0**, и потому в силу Предложения 3.2.7**

**196** 3. Численные методы линейной алгебры

**можем заключить, что спектральный радиус** C **должен быть строго меньше** 1**.**

**Достаточность. Если** ρ(C)<1**, то, согласно Предложению 3.4.2, можно выбрать подчинённую (а, значит, и согласованную) норму** k · k **так, чтобы выполнялось неравенство** kCk<1**. Теперь мы находимся в условиях Предложения 3.4.1, которое утверждает сходимость итераци-онного процесса (3.56)**

|  |  |  |
| --- | --- | --- |
| x***(*k*+1)*** ←Cx***(*k*)*** +d, | k = 0, 1, 2, . . . . |  |
| **Это завершает доказательство Теоремы 3.4.1.** | |  |

**Доказанные результаты теорема и предложения проясняют роль спектрального радиуса среди различных характеристик матри-цы. Мы могли видеть в §3.2д, что спектральный радиус не является матричной нормой, но, как выясняется, его с любой степенью точ-ности можно приблизить некоторой подчинённой матричной нормой. Кроме того, понятие спектрального радиуса оказывается чрезвычайно полезным при исследовании итерационных процессов и вообще степе-ней матрицы. К примеру, ранее мы установили (Предложение 3.2.7), что из сходимости степеней матрицы** A**k при** k→ ∞ **к нулевой матрице вытекает** ρ(A)<1**. Теперь результат Предложения 3.4.2 позволяет ска-зать, что это условие на спектральный радиус является и достаточным. Кроме того, более точно можно переформулировать условия сходимо-сти матричного ряда Неймана: он сходится для матрицы** A **тогда и только тогда, когда** ρ(A)<1**.**

**3.4б** **Подготовка линейной системы**

* + **итерационному процессу**
* **этом параграфе мы исследуем различные способы приведения систе-мы линейных алгебраических уравнений**

|  |  |
| --- | --- |
| Ax = b | **(3.58)** |

**к системе в рекуррентном виде**

|  |  |
| --- | --- |
| x = Cx + d, | **(3.59)** |

**отталкиваясь от которого можно организовывать итерационный про-цесс. Фактически, это равносильно рассмотрению вопроса о том, как**

|  |  |
| --- | --- |
| 3.4. Итерационные методы для систем линейных уравнений | **197** |

**связан предел стационарного одношагового итерационного процесса (3.56) с интересующим нас решением системы линейных алгебраиче-ских уравнений** Ax=b**. При этом нас будет интересовать не всякое приведение системы (3.58) к виду (3.59), но лишь такое, которое удо-влетворяет условию сходимости стационарного одношагового итераци-онного процесса, выведенному в предшествующем пункте.**

**Простейший способ состоит в том, чтобы добавить к обеим частям исходной системы по вектору неизвестной переменной** x**, т. е.**

|  |  |
| --- | --- |
| x + Ax = x + b, | **(3.60)** |

**а затем член** Ax **перенести в правую часть:**

x = (I − A)x + b.

**Весьма часто он непригоден, так как спектральный радиус матрицы**

* = I − A **оказывается больше** 1**. К примеру, это происходит, если у** A **есть отрицательные собственные значения, или если у** A **имеются** **собственные значения большие по модулю, чем** 2**.**

**Итак, необходим активный способ управления свойствами матрицы**

* **в получающейся системе рекуррентного вида** x=Cx+d**. Одним из важнейших инструментов такого управления служит предобуслвлива-ние исходной системы.**

Определение 3.4.1 Предобуславливанием **системы линейных алге-браических уравнений** Ax=b **называется умножение слева обеих её частей на некоторую матрицу** Λ**. Сама эта матрица** Λ **называется**

предобуславливающей матрицей **или, коротко,** предобуславливателем**.**

**Цель предобуславливания изменение (вообще говоря, улучшение) свойств матрицы системы, так как теперь вместо**

Ax = b

**мы получаем**

(ΛA) x = Λb.

**Продуманный выбор предобуславливателя может, к примеру, изменить выгодным нам образом расположение спектра матрицы** A**, так необхо-димое для организации сходящихся итерационных процессов.**

**198** 3. Численные методы линейной алгебры

**Хорошая идея состоит в том, чтобы выполнить предобуславливание до перехода к системе (3.60), т. е. до добавления вектора неизвестных** x **к обеим частям исходной СЛАУ. Тогда вместо** Ax = b **будем иметь** (ΛA)x = Λb**, и далее получаем**

x = (I − ΛA)x + Λb.

**Теперь в полученном рекуррентном виде можно с помощью подходя-щего выбора** Λ **добиваться требуемых свойств матрицы** (I−ΛA)**.**

**Каким образом следует выбирать предобуславливатели? Совершен-но общего рецепта на этот счёт не существует, и теория разбивается здесь на набор рекомендаций для ряда более или менее конкретных важных случаев.**

**Например, если в качестве предобуславливающей матрицы взять** Λ = A−***1*** **или хотя бы приближённо равную обратной к** A**, то вместо** **системы** Ax=b **получим** (A−***1***A)x=A−***1***b**, т. е. систему уравнений**

Ix = A−***1***b

**(или близкую к ней), матрица которой обладает всеми возможными до-стоинствами (малой обусловленностью и т. п.). Ясно, что нахождение подобного предобуславливателя не менее трудно, чем решение исход-ной системы, но сама идея примера весьма плодотворна, и на практике в качестве предобуславливателей часто берут несложно вычисляемые обратные матрицы к той или иной части матрицы** A**. Например, к глав-ной диагонали матрицы или же к трём диагоналям главной и двум побочным.**

**Другой способ приведения СЛАУ к рекуррентному виду основан на расщеплении матрицы системы.**

Определение 3.4.2 Расщеплением **матрицы** A **называется её пред-ставление в виде** A=G+ (−H) =G−H**, где** G **неособенная мат-рица.**

**Если известно некоторое расщепление матрицы** A**,** A=G−H**, то вместо исходной системы** Ax=b **мы можем рассмотреть**

(G − H) x = b,

**равносильную**

Gx = Hx + b,

|  |  |
| --- | --- |
| 3.4. Итерационные методы для систем линейных уравнений | **199** |

**так что**

x = G−***1***Hx + G−***1***b.

**На основе полученного рекуррентного вида можно организовать ите-рации**

|  |  |
| --- | --- |
| x***(*k*+1)*** ←G−***1***Hx***(*k*)*** +G−***1***b, | **(3.61)** |

**задавшись каким-то начальным приближением** x***(0)*. Таким образом, всякое расщепление помогает конструированию итерационных процес-сов.**

**Но практическое значение имеют не все расщепления, а лишь те, в которых матрица** G **обращается ¾относительно просто¿, чтобы органи-зация итерационного процесса не сделалсь более сложной задачей, чем решение исходной СЛАУ. Другое требование к матрицам, образующим расщепление, состоит в том, чтобы норма обратной для** G**, т. е.** kG−***1***k**, была ¾достаточно маленькой¿, поскольку** kG−***1***Hk ≤ kG−***1***k kHk**. В аб-солютной норме** kHk ≤ kAk**, но вот если** G−***1* имеет большую норму, то может оказаться** ρ(G−***1***H)>1 **и для итерационного процесса (3.61) не будут выполнеными условия сходимости.**

**Несложно обращаемыми матрицами являются**

1. **диагональные матрицы,**
2. **треугольные матрицы,**
3. **трёхдиагональные матрицы,**
4. **. . .**

**Ниже мы подробно рассмотрим итерационные процессы, соответству-ющие первым двум пунктам этого списка.**

**3.4в** **Оптимизация скалярного предобуславливателя**

**Напомним, что скалярной матрицей называется матрица, кратная еди-ничной (из-за своего родства скалярам, т. е. числам из** R **или** C**). Сейчас мы исследуем подробно описанную в предшествующем параграфе воз-можность управления итерационным процессом на простейшем приме-ре предобуславливания с помощью скалярной матрицы, когда** Λ=τ I**,**

τ ∈ R**.**

|  |  |
| --- | --- |
| **Итак, рассматриваем итерационный процесс** |  |
| x***(*k*+1)*** ← (I − τ A) x***(*k*)*** + τ b, | **(3.62)** |

**200** 3. Численные методы линейной алгебры

τ = const**, который часто называют** **методом простой итерации. Если** λ**i,** i = 1, 2, . . . , n **собственные числа матрицы** A **(вообще говоря, они** **комплексны), то собственные числа матрицы** (I−τ A) **равны** (1−τ λ**i**)**. Ясно, что в случае, когда среди** λ**i имеются числа с разным знаком** Re λ**i, добиться их локализации в единичном круге никаким выбором** τ **невозможно.**

**Далее рассмотрим практически важный частный случай, когда** A **симметричная положительно определённая матрица, так что все** λ**i,** i = 1, 2, . . . , n**, вещественны и положительны. Они нечасто бывают из-вестны точно, но нередко более или менее точно известен интервал их расположения на вещественной оси** R**. Будем предполагать, что**

λ**i** ∈ [µ, M ]**,** i = 1, 2, . . . , n**.**

**Матрица** (I−τ A) **также симметрична, и потому её спектральный радиус совпадает с 2-нормой. Чтобы обеспечить сходимость итераци-онного процесса и добиться её наибольшей скорости, нам нужно мини-**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **мизировать величину** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| k | I | − τ Ak***2*** | = **λ**i | | max | | |  | | | 1 | − | τ λ | **i**| | , |  |
|  | ∈ | ***[*µ,M*]*** | | |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **т. е. найти** τ**, при котором достигается** | | | | | | |  |  |  |  |  |  |  |  |  |
|  |  | min | max | |  |  |  |  |  |  |  |  |  |  |  |
|  |  | **τ λ**i | ∈ | ***[*µ,M*]*** |1 − τλ**i**|. | | | | | | | | |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Обозначив** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | g(τ) = |  | max | | | | | 1 | − | | τ λ , | |  |  |  |
|  |  | **µ λ M** | | | |  |  | | |  |  |  |
|  |  |  |  | ≤ ≤ | | |  |  |  |  |  |  |  |  |  |

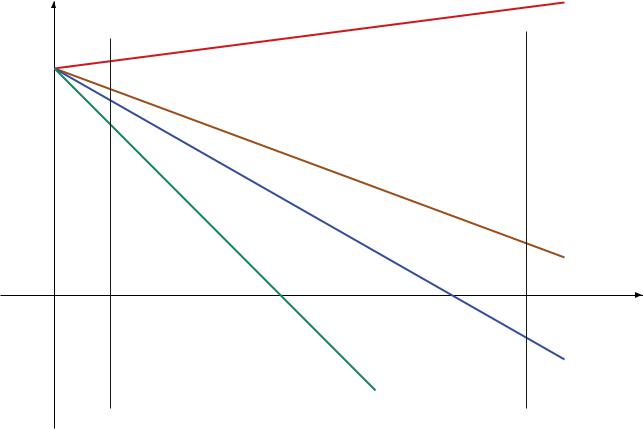
**обратимся для минимизации функции** g(τ) **к геометрической иллю-страции Рис. 3.8.**

**При** τ≤0 **функция** (1−τ λ) **не убывает по** λ**, и при положительных** λ**, очевидно, не меньше 1 по абсолютной величине. Следовательно, в** **нашем анализе имеет смысл ограничится значениями** τ >0**.**

**При** 0< τ≤M−***1* функция** 1−τ λ **на интервале** λ∈[µ, M] **неотри-цательна и монотонно убывает. Поэтому** g(τ) = max|1−τ λ|= 1−τ µ **и достигается на левом конце интервала** [µ, M]**.**

**При** τ > M−***1* велична** 1−τ M **отрицательна, так что график функ-ции** 1−τ λ **на интервале** λ∈[µ, M] **пересекает ось абсцисс. При этом на правом конце** |1−τ λ| **растёт с ростом** τ**, а на левом конце** |1−τ µ| **убывает с ростом** τ**.**

|  |  |  |  |
| --- | --- | --- | --- |
| 3.4. Итерационные методы для систем линейных уравнений | | | **201** |
| 1 |  |  |  |
|  |  | λ |  |
| 0 | µ | M |  |
| **Рис. 3.8. Графики функций *1*** −τ λ **для различных** τ | | |  |



**При некотором** τ=τопт **наступает момент, когда эти значения на концах интервала сравниваются друг с другом:**

1 − τ µ = −(1 − τ M ).

**Он и является моментом достижения оптимума, поскольку дальнейшее увеличение** τ **приводит к росту** −(1−τ M) **на правом конце интервала, а уменьшение** τ **ведёт к росту** 1−τ µ **на левом конце. Отсюда**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| τопт = | 2 | , | **(3.63)** |  |
| M + µ |  |

**а значение оптимума, равное коэффициенту подавления погрешности (как следствие из неравенств (3.57)), есть**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| k | I | − | τ | A | = min |  | max | | | | 1 | − | τ λ | = 1 | − | τ | µ = | M − µ | . **(3.64)** |  |
|  | опт | k***2*** | **τ** | **λ**i | ∈ | ***[*µ,M*]*** |  | **i**| |  | опт |  | M + µ |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

**Ясно, что оно меньше единицы, т. е. даже с помощью простейшего ска-лярного предобуславливателя мы добились сходимости итерационного процесса.**

**Полезно переписать полученное выражение, используя число обу-словленности матрицы** A**. Так как** µ≤λ***min***(A) **и** λ***max***(A)≤M **, то**

cond(A) = λ***max***(A) ≤ M . λ***min***(A) µ

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **202** |  |  |  |  |  | 3. Численные методы линейной алгебры | | | | | | |  |
| **Поэтому, принимая во внимание тот факт, что функция** | | | | | | | | | | | | |  |
|  |  |  | f (x) = | | |  | x − 1 | = 1 |  | 2 |  |  |  |
|  |  |  |  |  | − x + 1 | | | |  |
|  |  |  |  |  |  |  | x + 1 | |  |
| **возрастает при положительных** x**, можем заключить, что** | | | | | | | | | | | | |  |
| k | I | τ | A | k***2*** | = |  | M/µ − 1 | |  | cond(A) − 1 | | . |  |
|  | − | опт |  |  | M/µ + 1 | | ≥ cond(A) + 1 | | | |  |

**Получается, что чем больше** cond(A)**, т. е. чем хуже обусловленность матрицы** A **исходной системы, тем медленнее сходимость нашего ите-рационного процесса. Мы увидим далее, что это характерно для пове-дения большинства итерационных методов.**

**3.4г** **Итерационный метод Якоби**

**Пусть в системе линейных алгебраических уравнений** Ax=b **диаго-нальные элементы матрицы** A= (a**ij** ) **отличны от нуля, т. е.** a**ii** 6= 0**,** i = 1, 2, . . . , n**. Это условие не является обременительным, так как для** **неособенной матрицы** A **перестановкой строк (соответствующей пере-становке уравнений системы) можно всегда добиться ненулевых диаго-нальных элементов.**

**В развёрнутом виде рассматриваемая система имеет вид**

**n**

**X**

a**ij** x**j** = b**i**, i = 1, 2, . . . , n,

**j*=1***

**и, выражая из** i**-го уравнения** i**-ю компоненту вектора неизвестных,**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **получим** | | b**i** − |  | a**ij** x**j** ,i = 1, 2, . . . , n. |  |
| x**i** = a**ii** | |  |  |
| 1 | |  | **X** |  |  |
|  |  |  |
|  |  |  |
|  |  |  | **j**6***=*i** |  |  |

**Нетрудно понять, что полученные соотношения дают представление исходной СЛАУ в рекуррентном виде, необходимое для организации итераций:**

> x = T (x), T (x) = T***1***(x), T***2***(x), . . . , T**n**(x) ,

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **где** T**i**(x) =a**ii** | | b**i** − |  | a**ij** x**j** ,i = 1, 2, . . . , n. |  |
| 1 | |  | **X** |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  | **j**6***=*i** |  |  |

|  |  |
| --- | --- |
| 3.4. Итерационные методы для систем линейных уравнений | **203** |

**Псевдокод этого итерационного процесса выписан в Табл. 3.2, где вспо-могательная переменная** k **счётчик числа итераций.**

**Таблица 3.2. Итерационный метод Якоби для решения СЛАУ**

k ← 0 **;**

**выбираем начальное приближение** x***(0)* ;** DO WHILE **( метод не сошёлся )**

DO FOR i = 1 TO n

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| x**i** | ← a**ii** | | b**i** − |  | a**ij** x**j** |  |  |
| ***(*k*+1)*** | 1 | |  | **X** | ***(*k*)*** | **(3.65)** |  |
|  |  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  | **j**6***=*i** |  |  |  |
| END DO |  |  |  |  |  |  |  |
| k ← k + 1 **;** |  |  |  |  |  |  |  |
| END DO |  |  |  |  |  |  |  |

**Он был предложен ещё в середине XIX века К.Г.Якоби и часто (осо-бенно в старых книгах по численным методам) называется ¾методом одновременных смещений¿. Под ¾смещениями¿ здесь имеются в ви-ду коррекции компонент очередного приближения к решению, выпол-няемые на каждом шаге итерационного метода. Смещения-коррекции ¾одновременны¿ потому, что все компоненты следующего приближе-ния** x***(*k*+1)* насчитываются по единообразным формулам, основанным лишь на известном предыдущем приближении. В следующем парагра-фе будет рассмотрен итерационный процесс, устроенный несколько по-другому, в котором смещения-коррекции компонент приближения к ре-шению не одновременны в том смысле, что последовательно находятся не только из предыдущего приближения, но ещё и одна через другую.**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **204** |  |  |  |  |  | 3. | Численные методы линейной алгебры | | | | | |  |
|  |  |  |  | ˜ | ˜ |  |  |  |  |  |  |  |  |
| **Пусть** A=D+L+U **, где** | | | | | | |  |  |  |  |  |  |  |
|  |  | 0 | 0 | ·**.** ·**. .**· | a***2*,n**− | ***1*** |  | a***2*n** | |  |  |  |
|  |  | a***12*** | **.. .** | a***1*,n** | ***1*** |  | a***1*n** | |  |  |  |  |
| U˜ = |  |  | **...** − |  |  | **...** |  |  | **строго верхняя** |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  | 0 |  | a**n** | | ***1*,n** | |  |  |  |
|  |  |  | 0 |  |  |  |  | − |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  | **треугольная матрица,** |  |
|  |  |  |  |  |  |  |  | 0 | |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| D = diag {a***11***, a***22***, . . . , a**nn**} | | | | | |  |  |  | |  | **диагональ матрицы,** | |  |
|  | a***21*** | | 0 |  | 0 | |  |  |  |  |  |
|  |  | 0 |  |  |  |  |  |  | |  |  |  |  |
| L˜ = |  |  |  |  |  |  |  |  |  |  |
|  | **.** | **..** | | 0 |  |  |  | |  |  | **строго нижняя** |  |
|  | a***31*** | a***32*** | **...** |  |  |  |  | |  |  |
|  |  | **..** | **.. ..** | |  |  |  |  | |  |  |  |  |
|  |  |  |  |  |  | |  |  | **треугольная матрица.** |  |
|  |  | a**n,n**−***1*** | | 0 |  | |  |  |  |
|  |  | a**n*1*** | a**n*2*** |  |  | |  |  |  |  |
|  |  |  |  | · · · |  |  |  |  | |  |  |  |  |

**Тогда итерационный метод Якоби может быть представлен как метод, основанный на таком расщеплении матрицы системы** A=G−H**, что**

˜ ˜

G = D, H = −(L + U ).

**Соответственно, в матричном виде метод Якоби записывается как**

***(*k*+1)*** −***1*** ˜ ˜ ***(*k*)*** −***1***

x ← −D (L + U ) x + D b.

**Теперь нетрудно дать условия его сходимости, основываясь на общем результате о сходимости стационарных одношаговых итераций (Теоре-ма 3.4.1). Именно, метод Якоби сходится тогда и только тогда, когда**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***1*** |  |  |  |  | ρ D−***1***(L˜ + U˜ ) < 1. | | | | | |  |  |  |  |  |
| **Матрица** D− | (L˜ + U˜) **просто выписывается по исходной системе и** | | | | | | | | | | | | | |  |
| **имеет вид** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | |  | 0 |  |  | a***12***/a***11*** | | | | . . . a***1*n**/a***11*** | | |  |  |  |
| a***21***/a**..** | | | | | ***22*** |  |  | 0**..** |  | .**. .**. . a**n*2***/a**..** | | ***22*** | . | **(3.66)** |  |
| **.** | | | |  |  |  |  | **.** |  | **.** | **.** |  |  |  |  |
|  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| a | | ***1*n** | /a | **nn** | | a | **n*2*** | /a | **nn** | . . . | 0 |  |  |  |  |
|  | |  |  |  | |  |  |  |  |  |  |  |  |  |  |
|  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

|  |  |
| --- | --- |
| 3.4. Итерационные методы для систем линейных уравнений | **205** |

**Но нахождение её спектрального радиуса является задачей, сравнимой по сложности с выполнением самого итерационного процесса и пото-му непрактично. Для быстрой и грубой оценки спектрального радиуса можно воспользоваться какой-нибудь матричной нормой и результатом Предложения 3.2.6. Полезен также следующий достаточный признак сходимости:**

Предложение 3.4.3 **Если в системе линейных алгебраических урав-нений** Ax=b **матрица** A **имеет диагональное преобладание, то метод Якоби для этой системы сходится при любом начальном приближе-нии.**

Доказательство. **Диагональное преобладание в матрице** A = (a**ij** )

**означает, что**

**X**

|a**ii**| > |a**ij** |, i = 1, 2, . . . , n.

**j*=***6**i**

**Следовательно,**

* a**ij**

< 1,i = 1, 2, . . . , n,

a**ii**

**j*=***6**i**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **что равносильно** |  |  | a**ii** **!** | | | | < 1. |  |
| ***1***≤**i**≤**n** | **j*=*i** | |  |
| max | **X** | | | a**ij** | |  |  |  |
|  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  | 6 |  |  |  |  |  |  |
| **В выражении, стоящем в левой** | | **части неравенства, легко угадать под-** | | | | | |  |
|  |  |  |  |  |  |  |
| **чинённую чебышёвскую норму (**∞**-норму) матрицы** D−***1***(L˜+U˜)**, выпи-** | | | | | | | |  |
| **санной нами в (3.66). Таким образом,** | | | | | |  |  |  |
| D−***1*** | (L˜ + U˜) ∞ < 1, | | | | | | |  |
|  |  |  |  |  | |  |  |  |

**откуда, ввиду результата Предложения 3.4.1, следует доказываемое.**

**3.4д** **Итерационный метод Гаусса-Зейделя**

**В итерационном методе Якоби при организации вычислений по ин-струкции**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| x**i** | ← a**ii** | | b**i** − | **X** | a**ij** x**j** | **!**,i = 1, 2, . . . , n, | **(3.67)** |  |
| ***(*k*+1)*** |  | 1 |  | ***(*k*)*** |  |  |  |
|  |  |  |  |  |

**j*=***6**i**

**206** 3. Численные методы линейной алгебры

**Таблица 3.3. Итерационный метод Гаусса-Зейделя**

**для решения линейных систем уравнений**

k ← 0 **;**

**выбираем начальное приближение** x***(0)* ;** DO WHILE **( метод не сошёлся )**

DO FOR i = 1 TO n

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| x**i** | ← a**ii** | | b**i** − | **i**−***1*** | a**ij** x**j** | − | **n** | a**ij** x**j** |  |  |
| ***(*k*+1)*** | 1 | |  | ***(*k*+1)*** |  | ***(*k*)*** |  |  |
|  |  |  |  | **X** |  |  | **X** |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  | **j*=*i** |  |  | **j*=*i*+1*** |  |  |  |
| END DO |  |  |  |  |  |  |  |  |  |  |

k ← k + 1 **;**

END DO

**компоненты очередного приближения** x***(*k*+1)* находятся последователь-**

**но одна за другой, так что к моменту вычисления** i**-ой компоненты вектора** x***(*k*+1)* уже найдены** x***(1*k*+1)* ,** x***(2*k*+1)* , . . . ,** x***(*i**−**k*+1)1* . Но метод Якоби**

**никак не использует эти новые значения, и при вычислении любой ком-поненты следующего приближения всегда опирается только на вектор** x***(*k*)*** **предшествующего приближения. Если итерационный процесс схо-дится к решению, то естественно ожидать, что все компоненты** x***(*k*+1)* ближе к искомому решению, чем** x***(*k*)*, а посему немедленное вовлечение их в процесс вычисления будет способствовать ускорению сходимости.**

**На этой идее основан итерационный метод Гаусса-Зейделя**8**, псев-докод которого представлен в Табл. 3.3 (где, как и ранее,** k **счётчик итераций). В нём суммирование в формуле (3.67) для вычисления** i**-ой компоненты очередного приближения** x***(*k*+1)* к решению разбито на две части по индексам, предшествующим** i**, и по индексам, большим** i**.**

**Первая часть суммы использует новые вычисленные значения** x***(1*k*+1)* ,**

**. . . ,** x***(*i**−**k*+1)1*, тогда как вторая компоненты** x***(*i*+1*k*)*, . . . ,** x***(*nk*)* из старого приближения. Метод Гаусса-Зейделя иногда называют также итера-**

**ционным методом ¾последовательных смещений¿, а его основная идея**

8 **В отчествественной литературе по вычислительной математике нередко исполь-зуется также термин** метод Зейделя**.**

=x***(*k*)***

|  |  |
| --- | --- |
| 3.4. Итерационные методы для систем линейных уравнений | **207** |

**немедленного вовлечения уже полученной информации в вычислитель-ный процесс с успехом применима и для нелинейных итерационных схем.**

**Чтобы получить для метода Гаусса-Зейделя матричное представле-ние, перепишем его расчётные формулы в виде**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **i** |  |  | **n** |  |  |  |  |  |
| **X** |  |  | **X** | |  |  |  |  |
| a**ij** x**j*(*k*+1)*** | = − | | | a**ij** x**j*(*k*)*** + b**i**, | | i = 1, 2, . . . , n. |  |  |
| **j*=1*** |  |  | **j*=*i*+1*** | |  |  |  |  |
|  |  |  |  |  |  | ˜ ˜ | **, можем** |  |
| **Используя введённые в предыдущем пункте матрицы** D**,** L **и** U | | | | | | |  |
| **записать эти формулы в виде** | | | |  |  |  |  |  |
|  |  |  | ˜ | ***(*k*+1)*** | ˜ ***(*k*)*** | + b, |  |  |
|  | (D + L)x | | |  | = −U x |  |  |
| **т. е.** |  |  | (D + L˜)−***1***U˜ x***(*k*)*** + (D + L˜)−***1***b. | | | | **(3.68)** |  |
| x***(*k*+1)*** = | | − |  |
|  |  |  |  |  |  |  |  |

**Таким образом, необходимое и достаточное условие сходимости метода Гаусса-Зейделя из любого начального приближения это неравенство**

˜ −***1*** ˜

ρ (D + L) U < 1.

Предложение 3.4.4 **Если в системе линейных алгебраических урав-нений** Ax=b **матрица** A **имеет диагональное преобладание, то ме-тод Гаусса-Зейделя для этой системы сходится при любом начальном приближении.**

Доказательство. **Отметим, прежде всего, что в условиях диагональ-ного преобладания в** A **решение** x**? рассматриваемой линейной системы существует (вспомним признак неособенности Адамара). Пусть, как и ранее,** x***(*k*)* его приближение, полученное на** k**-ом шаге итерационного**

**процесса. Исследуем поведение погрешности решения** z***(*k*)*** − x**?**

**в зависимости от номера итерации** k**.**

**Чтобы получить формулу для** z***(* k*)*, предварительно перепишем со-отношения, которым удовлетворяет точное решение** x**?: вместо**

**n**

**X**

a**ij** x**?j** = b**i**, i = 1, 2, . . . , n.

**j*=1***

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **208** |  |  | 3. | Численные методы линейной алгебры | | |
| **можно придать им следующий эквивалентный вид** | | | | | | |
| x**i?** =a**ii** | | b**i** − | a**ij** x**j?** − | a**ij** x**j?!** | , | i = 1, 2, . . . , n. |
| 1 | | **i**−***1*** |  | **n** |  |  |
|  |  | **X** |  | **X** |  |  |
|  |  | **j*=1*** |  | **j*=*i*+1*** |  |  |

**Вычитая затем почленно эти равенства из расчётных формул метода Гаусса-Зейделя, т. е. из**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| x**i*(*k*+1)*** = |  | 1 | |  |  |
| a**ii** | | | |  |
|  |  |
| **получим** | |  |  |  |  |
| z**i*(*k*+1)*** = | | 1 | | |  |
|  | a**ii** | |  |
|  |  |  |  |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| b**i** − | a**ij** x**j*(*k*+1)*** − | | | a**ij**x**j*(*k*)*!** | | , | i = 1, 2, . . . , n, |
|  | **i**−***1*** |  |  | **n** |  |  |  |
|  | **X** |  |  | **X** |  |  |  |
|  | **j*=1*** |  |  | **j*=*i*+1*** |  |  |  |
| − | a**ij** z**j*(*k*+1)*** | − |  | a**ij**z**j*(*k*)*!** | , |  | i = 1, 2, . . . , n, |
| **i**−***1*** | |  |  | **n** |  |  |  |
| **X** | |  |  | **X** |  |  |  |
| **j*=1*** | |  | **j*=*i*+1*** | |  |  |  |

**Беря абсолютное значение от обеих частей этого равенства и пользу-ясь неравенством треугольника для абсолютных значений, будем иметь**

**для** i= 1,2, . . . , n

|  |  |  |  |
| --- | --- | --- | --- |
|  | z**i*(*k*+1)*** |  | ≤ |
|  |  |  |  |
|  |  |  | ≤ |

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **j*=1*** |  | a**ii** |  | · |  | z**j** | |  | | + | |  |
| **X** | | a**ij** |  | | |  | ***(*k*+1)*** |  |  |  |
| **i**−***1*** |  |  |  |  |  |  | |  |  |  |
| **i**−***1*** |  | a**ij** |  | · |  |  | z***(*k*+1)*** | |  |  | ∞ |  |
| **j*=1*** |  | a**ii** |  |  |  |  | |  |
| **X** | |  |  | | | |  |  |  |  |
|  |  |  |  | | | |  |  |  | |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **j*=*i*+1*** | a**ii** ·z**j** | | | |  |  |
| **n** | a**ij** |  |  |  |  |  |
| **X** |  | ***(*k*)*** |  |
|  |  |  |  |  |
| **n** |  |  |  | · z | |  |
| + **j*=*i*+1*** a**ii** | | | |  |
| **X** | | a**ij** |  | |  |  |
|  |  |  |
|  |  |  |  | |  |  |
|  |  |  |  |  |  |  |

***(*k*)*** ∞. **(3.69)**

**С другой стороны, условие диагонального преобладания в матрице** A**, т. е. то, что**

|  |  |  |
| --- | --- | --- |
| **X** |  |  |
| |a**ij** | < |a**ii**|, | i = 1, 2, . . . , n, |  |
| **j**6***=*i** |  |  |
| **означает существование константы** κ**,** 0≤κ<1**, такой что** | |  |
| **X** |  |  |
| |a**ij** | ≤ κ |a**ii**|, | i = 1, 2, . . . , n. | **(3.70)** |

**j*=***6**i**

**По этой причине**

* a**ij**

≤ κ,i = 1, 2, . . . , n,

a**ii**

**j**6***=*i**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 3.4. Итерационные методы для систем линейных уравнений | | | | | | | | | | | | | | | | | | | | | | | | | | | |  | **209** |  |
| **и, следовательно,** | | | | | | | |  |  |  |  |  |  |  | ≤ κ − κ **j*=1*** a**ii** | | | | | | | | | = κ 1 − **j*=1*** a**ii** **!**. | | | | | |  |
| **j*=*i*+1*** | | a**ii** | | |  | ≤ κ − **j*=1*** a**ii** | | | | | | | | |  |
|  | **n** |  | a**ij** | |  |  |  | **i**−***1*** | |  | a**ij** | |  | |  |  |  |  | **i**−***1*** | | a**ij** | |  |  |  | **i**−***1*** | | a**ij** |  |  |
| **X** | |  |  |  |  | **X** | | |  |  |  |  |  |  | **X** | |  |  |  |  | **X** | |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| i = 1, | |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | 2, . . . , n**. Подставляя полученную оценку в неравенства (3.69),** | | | | | | | | | | | | | | | | | | | | | | | | | | | |  |
| **приходим к соотношениям** | | | | | | | | | | | |  |  |  |  | ∞ + κ | | | 1 − **j*=1*** a**ii** **!** | | | | | | |  | ∞, |  |  |  |
|  | z**i** |  |  |  | ≤ | | **j*=1*** a**ii** · | | |  | z***(*k*+1)*** | | | | | z***(*k*)*** | **(3.71)** | |  |
|  | ***(*k*+1)*** | | |  |  |  | **i**−***1*** | a**ij** |  | |  |  |  |  |  |  |  |  |  |  | **i**−***1*** | | | a**ij** |  |  |  |  |  |  |
|  |  |  |  |  | **X** |  |  |  |  |  |  |  |  |  |  | **X** | | |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| i = 1, 2, . . . , n**.** | | | | | | |  |  |  |  |  |  |  | ***(*k*+1)*** | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  | z**i*(*k*)*** |  | ***(*k*)*** | |  |  |  |  |  |  |  |  |  |
|  | **Предположим, что** max***1***≤**i**≤**n** | | | | | | | | | | | | | | | | **достигается при** i=m**, так что** | | | | | | | | | | | |  |
|  |  |  |  |  |  |  |  |  |  |  |  | z | |  |  |  |  | = |  | z**m** . | | |  |  |  |  |  | **(3.72)** | |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | ∞ | |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |

**Рассмотрим теперь отдельно** m**-ое неравенство из (3.71). Привлекая равенство (3.72), можем написать**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | z***(*k*+1)*** | ∞ | ≤ | **m**−***1*** | | |  | a**mj** | · | |  | z***(*k*+1)*** | |  |  | ∞ |  |  |  |  | **m**−***1*** | |  |  | a**mj** | | |  | **!** | | z***(*k*)*** | ∞, |  |
|  |  |  |  | **X** | | | |  |  | | |  |  |  | |  |  |  |  |  | **X** | | |  |  |  |  |  | | | |  |  |
|  |  |  |  |  |  |  |  |  |  | | |  |  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **то есть** | |  |  |  | **!** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | **!** | |  |  |  |  |  |  |  |  |
|  | **m**−***1*** | | a**mj** | |  |  | z***(*k*+1)*** | |  | ∞ ≤ κ | | | | |  |  |  | **m**−***1*** |  | a**mj** | |  |  | z***(*k*)*** | | | ∞. **(3.73)** | | |  |
|  | 1 − **j*=1*** a**mm** | | | |  | |  |  | 1− **j*=1*** | | | a**mm** | | |  |  |
|  | **X** | |  |  |  |  |  |  |  |  |  |  |  |  |  | **X** | |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Коль скоро** | | |  |  |  |  |  |  | 1− **j*=1*** | | | | a**mm** | | | | | | > 0 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  | **m**−***1*** |  | | a**mj** | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  | **X** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

**в силу диагонального преобладания в матрице** A**, то мы можем со-кратить на эту величину обе части неравенства (3.73). Окончательно**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **получаем** | |  |  |  |  |  |  |  |  |
| **что при** | κ < 1 | z***(*k*+1)*** | z***(*k*)*** |  |
| | | | |  | ∞ ≤ κ |  | ∞, |  |  |
|  |  |  |  |  |  |
|  |  | **означает сходимость метода Гаусса-Зейделя.** | | | | | |  |  |

**Фактически, в доказательстве Предложения 3.4.4 мы получили да-же оценку уменьшения чебышёвской нормы погрешности через ¾меру диагонального преобладания¿, в качестве которой может выступать ве-личина** κ**, определённая посредством (3.70).**

**210** 3. Численные методы линейной алгебры

Теорема 3.4.2 **Если в системе линейных алгебраических уравнений** Ax = b **матрица** A **является симметричной положительно опреде-лённой, то метод Гаусса-Зейделя сходится к решению из любого на-чального приближения.**

Доказательство **опускается.**

**Метод Гаусса-Зейделя был сконструирован как модификация мето-да Якоби, и, казалось бы, должен работать лучше. Так оно и есть ¾в среднем¿, на случайно выбранных системах метод Гаусса-Зейделя работает несколько быстрее, что можно показать математически стро-го при определённых допущениях на систему. Но в целом ситуация не столь однозначна. Для СЛАУ размера** 3×3 **и более существуют при-меры, на которых метод Якоби расходится, но метод Гаусса-Зейделя сходится, так же как существуют и примеры другого свойства, когда метод Якоби сходится, а метод Гаусса-Зейделя расходится.**

**3.4е** **Методы релаксации**

**Одим из принципов, который кладётся в основу итерационных методов решения систем уравнений, является так называемый принцип релак-сации.**9 **Под эти понимается такая организация итераций, при которой на каждом шаге процесса уменьшается какая-либо величина, характе-ризующая погрешность решения системы.**

**Поскольку само решение** x**? нам неизвестно, то оценить напрямую погрешность** x***(*k*)*** −x**? не представляется возможным. По этой причине о степени близости** x***(*k *)* к** x**? судят на основании косвенных призна-ков, важнейшим среди которых является величина невязки решения. Невязка определяется как разность левой и правой частей уравнения после подстановки в него приближения к решению, и в нашем случае это** Ax***(*k*)*** −b**.**

**В этих условиях конкретное применение принципа релаксации мо-жет заключаться в том, что на каждом шаге итерационного процесса стремятся уменьшить абсолютные значения компонент вектора невяз-ки либо её норму. В этом смысле итерационные процессы Якоби и Гаусса-Зейделя можно также рассматривать как релаксационные, в ко-торых на каждом шаге компоненты очередного приближения вычис-ляются из условия зануления соответствующих компонент невязки на**

9 **От латинского слова ¾relaxatio¿** **уменьшение напряжения, ослабление.**

|  |  |
| --- | --- |
| 3.4. Итерационные методы для систем линейных уравнений | **211** |

**основе уже полученной информации о решении. Правда, это делается ¾локально¿, для отдельно взятой компоненты, и без учёта влияния ре-зультатов вычисления этой компоненты на другие компоненты невяз-ки.**

**Различают релаксацию полную и неполную, в зависимости от того, добиваемся ли мы на каждом отдельном шаге итерационного процесса (или его подшаге) наибольшего возможного улучшения рассматрива-емой функции от погрешности или нет. Локально полная релаксация может казаться наиболее выгодной, но глобально, с точки зрения сходи-мости процесса в целом, тщательно подобранная неполная релаксация нередко приводит к более эффективным методам.**

**Очень популярной реализацией высказанных выше общих идей яв-ляется метод решения систем линейных алгебраических уравнений, в котором для улучшения сходимости берётся ¾взвешенное среднее¿ зна-чений компонент предшествующей** x***(* k*)* и последующей** x***(*k*+1)* итераций метода Гаусса-Зейделя. Более точно, зададимся вещественным числом** ω**, которое будем называть** **параметром релаксации, и** i**-ую компоненту** **очередного** k+ 1**-го приближения положим равной**

ωx***(*ik*+1)*** + (1 − ω)x***(*ik*)***,

**где** x***(*ik*)*** **приближение, полученное в результате** k**-го шага алгоритма,**

**а** x***(*ik*+1)*** **приближение, которое было бы получено на основе** x***(1*k*+1)* ,**

**. . . ,** x***(*i**−**k*+1)1*,** x***(*i*+1*k*)*, . . . ,** x***(*nk*)* с помощью метода Гаусса-Зейделя. Псевдо-код метода релаксации для решения систем линейных алгебраических**

**уравнений представлен в Табл. 3.4.**

**Рсчётные формулы нового метода можно записать в виде**

|  |  |  |
| --- | --- | --- |
|  | **i**−***1*** | **n** |
|  | **X** | **X** |
| a**ii**x**i*(*k*+1)*** | + ωa**ij** x**j*(*k*+1)*** = (1 − ω) a**ii**x**i*(*k*)*** − ω | a**ij** x**j*(*k*)*** + ωb**i**, |
|  | **j*=1*** | **j*=*i*+1*** |

˜

**для** i= 1,2, . . . , n**. Применяя введённые выше в §3.4д матрицы** D**,** L**,**

˜

U **, можно придать этим соотношениям более компактный вид**

˜ ***(*k*+1)*** ˜ ***(*k*)***

D − ωL x = (1 − ω)D + ωU x + ωb,

**откуда**

***(*k*+1)*** ˜ −***1*** ˜ ***(*k*)*** ˜ −***1***

x = D − ωL (1 − ω)D + ωU x + D − ωL ωb,

**212** 3. Численные методы линейной алгебры

**Таблица 3.4. Псевдокод метода релаксации**

**для решения систем линейных уравнений**

k ← 0 **;**

**выбираем начальное приближение** x***(0)* ;** DO WHILE **( метод не сошёлся )**

DO FOR i = 1 TO n

x***(*ik*+1)*** ← (1 − ω) x***(*ik*)***+

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| a**ii** | | b**i** − | **i**−***1*** | a**ij** x**j** | − | **n** | a**ij** x**j** |  |  |
| ω | |  | ***(*k*+1)*** |  | ***(*k*)*** |  |  |
|  |  |  | **X** |  |  | **X** |  |  |  |
|  |  |  |  |  |  |
|  |  |  | **j*=*i** |  |  | **j*=*i*+1*** |  |  |  |

END DO

k ← k + 1 **;**

END DO

* = 0, 1, 2, . . . **.**
  + **зависимости от конкретного значения параметра релаксации при-нято различать три случая:**

**если** ω <1**, то говорят о ¾нижней релаксации¿,**

**если** ω= 1**, то имеем итерации Гаусса-Зейделя, если** ω >1**, то говорят о ¾верхней релаксации¿.**10

**Последний случай может показаться экзотичным, но во многих ситу-ациях он действительно обеспечивает улучшение сходимости итераций в сравнении с методом Гаусса-Зейделя. Несколько упрощённое объяс-нение этого явления может состоять в том, что если направление от** x***(*k*)*** **к** x***(*** **k*+1)*** **оказывается удачным в том смысле, что приближает к ис-комому решению, то имеет смысл пройти по нему и дальше, за** x***(*k*+1)* . Это соответствует случаю** ω >1**.**

**Важно отметить, что метод релаксации также укладывается в из-ложенную в §3.4б схему итерационных процессов, порождаемых рас-**

10 **В англоязычной литературе по вычислительной линейной алгебре этот ме-тод обычно обозначают аббревиатурой SOR(**ω**), которая происходит от термина ¾Successive OverRelaxation¿ последовательная верхняя релаксация.**

|  |  |
| --- | --- |
| 3.4. Итерационные методы для систем линейных уравнений | **213** |

**щеплением матрицы решаемой системы уравнений. Именно, мы берём** A = G**ω** − H**ω** **с матрицами**

|  |  |
| --- | --- |
| ˜ | ˜ |
| G**ω** = D − ωL, | H**ω** = (1 − ω)D + ωU . |

**Необходимое и достаточное условие сходимости метода релаксации при-нимает поэтому вид**

* G−**ω*1***H**ω** < 1.

**Для некоторых специфичных, но очень важных задач математической физики значение релаксационного параметра** ω**, при котором величи-**

**на** ρ G−**ω*1***H**ω достигает минимума, находится относительно просто. В более сложных задачах для оптимизации** ω **требуется весьма трудный анализ спектра матрицы перехода** G−**ω*1***H**ω из (3.61). Обзоры состония дел в этой области читатель может найти в [63, 43, 47].**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Предложение 3.4.5 **Пусть** C**ω** = | | ˜ −***1*** |  | ˜ | **мат-** | | | | |  |
| D −ωL | (1 −ω)D + ωU | |  |
|  | **релаксации. Тогда** ρ(C) | | | ≥ | ω | − | 1 | | | **, и** |  |
| **рица оператора перехода метода** |  |  | **ω** | | |  |  |  |

**потому неравенство** 0< ω <2 **на параметр релаксации необходимо для сходимости метода.**

Доказательство. **Преобразуем матрицу** C**ω** **для придания ей более** **удобного для дальнейших преобразований вида:**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **ω** = D − ωL | | | −***1*** | ***1*** ˜ −−***1*** | | | | |  |  |  |  |  |  | ˜ |  |
| C |  | ˜ |  | (1 |  |  |  |  |  |  |  | ˜ |  |  |  |
| − |  |  |  |  | ω)D + ωU | | | | | | |  | ˜ |  |
|  |  | ***1*** ˜ |  | −***1*** |  | ***1*** | | − | |  |  |  |  |
| = D(I |  | ωD− | | | L) | |  |  | (1 |  |  | ω)D + ωU | | | |  |
| = I − ωD−***1*** ˜ | | | | −***1*** | | D | − |  |  |  | − | |  | ***1*** | ˜ |  |
|  |  |  | L | |  |  |  | (1 | |  |  | ω)D + ωU | | |  |

= I − ωD− L (1 − ω)I + ωD− U .

**Желая исследовать расположение собственных чисел** λ**i**(C**ω** ) **матрицы** C**ω** **, рассмотрим её характеристический полином**

−***1*** ˜ −***1*** −***1***˜

φ(λ) = det(C**ω** − λI) = det I − ωD L (1 − ω)I + ωD U − λI

* + p**n**λ**n** + p**n**−***1***λ**n**−***1*** + . . . + p***1***λ + p***0***,
* **котором** p**n** = (−1)**n по построению. Свободный член** p***0* характери-**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **214** |  |  |  |  | 3. | | Численные методы линейной алгебры | | | | | | | | | | | | | | | | |  |
| **стического полинома может быть найден как** φ(0)**:** | | | | | | | | | | | | | | | | | | | |  | ***1*** |  |  |  |
|  | ***0*** |  |  | − | ***1*** | ˜ |  |  | ***1*** | |  |  | − | |  |  |  |  |  |  | ˜ |  |  |
| p |  | = det | I |  | ***1*** ˜ | |  | −***1*** | | (1 | | |  |  |  |  |  |  | ***1*** | ˜ |  |  |  |  |
|  |  | ωD− | L |  |  |  |  |  | ω)I + ωD− | | | | | U |  |  |  |  |
|  |  | = det (1 | |  | ω)I | ωD | | | ***1***U˜ | | |  |  | = (1 | | |  |  | ω)**n**, |  |  | U |  |
|  |  | = det (I | | − | ωD− | L)− | | |  | · | det (1 | | | | | − | | ω)I + ωD− | | | |  |  |
|  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  |
|  |  |  | − | +***1*** ˜ | | |  | − |  |  |  |  |  | − | |  |  |  |  |  |  |
| **коль скоро матрица** (I−ωD− | | | | | | L) | | | **нижняя треугольная и диагональ-** | | | | | | | | | | | | | | |  |
| **ными элементами имеет единицы, а** | | | | | | | | | | |  |  |  | (1 − ω)I + ωD−***1***U˜ | | | | | | | |  | **верхняя** |  |
|  |  |  |  |  |  |  |  |  |  |  |  | **главной диагонали.** | | | | | | | | | |  |  |
| **треугольная, с элементами** (1−ω) **по** | | | | | | | | | | | | | |  |  |  |  |  |  |  |  |  |  |

**С другой стороны, по теореме Виета свободный член характеристи-ческого полинома матрицы, делённый на старший коэффициент, равен произведению его корней, т. е. собственных чисел матрицы, умножен-ному на** (−1)**n, и поэтому**

**n**

**Y**

λ**i**(C**ω** ) = (1 − ω)**n**.

**i*=1***

**Отсюда необходимо следует**

max |λ**i**(C**ω** )| ≥ |ω − 1|,

***1***≤**i**≤**n**

**что и доказывает Предложение.**

Теорема 3.4.3 **Если в системе линейных алгебраических уравнений** Ax = b **матрица** A **является симметричной положительно опреде-лённой, то для всех значений параметра** ω**,** 0< ω <2**, метод релак-сации сходится к решению из любого начального приближения.**

Доказательство **опускается.**

**3.4ж** **Оценка погрешности стационарного итерационного метода**

**Пусть задан сходящийся стационарный одношаговый итерационный метод**

x***(*k*+1)*** ← Cx***(*k*)*** + d, k = 0, 1, 2, . . . ,

**в котором некоторая норма матрицы** C **меньше единицы, т. е.** kCk<1**. Ясно, что ввиду результатов §3.4а последнее допущение не ограничива-ет общности нашего рассмотрения. Как оценить отклонение по норме**

|  |  |
| --- | --- |
| 3.4. Итерационные методы для систем линейных уравнений | **215** |

**очередного приближения** x***(*k*)* от предела** x**?** := lim**k**→∞x***(* k*)*, не зная самого этого предела и наблюдая лишь за итерационной последова-тельностью** x***(0)*,** x***(1)*, . . . ,** x***(*k*)*, . . . ?**

**Как и прежде, имеем**

|  |  |
| --- | --- |
| x***(*k*)*** =Cx***(*k**−***1)*** +d, |  |
| x**?** = Cx**?** + d. |  |
| **Вычитая второе равенство из первого, получим** |  |
| x***(*k*)*** −x**?**=C x***(*k**−***1)*** −x**?**. | **(3.74)** |

**Перенесём** x***(*k *)* в правую часть этого соотношения, а затем добавим к обеим частям по** x***(*k**−***1)*:**

x***(*k**−***1)*** −x**?**=x***(*k**−***1)*** −x***(*k*)*** +C x***(*k**−***1)*** −x**?**.

**Беря нормы от обеих частей полученного равенства и применяя затем неравенство треугольника, приходим к оценке**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| x***(*k**−***1)*** | − x**?** |  | ≤ | x***(*k*)*** −x***(*k**−***1)*** |  | + kCk · | x***(*k**−***1)*** | − x**?** | , |
|  |  |  |  |  |  |  |  |  |  |

**где векторная норма согласована с используемой матричной нормой для** C**. Перенесение в левую часть второго слагаемого из правой части и последующее деление обеих частей неравенства на положительную величину** (1− kCk) **даёт**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | x***(*k**−***1)*** | − x**?** |  | ≤ | 1 | x***(*k*)*** −x***(*k**−***1)*** | . | **(3.75)** |  |
| 1 − kCk |  |
|  |  |  |  |  |  |  |  |  |  |

**Чтобы ещё улучшить вид этой оценки, вспомним, что из (3.74) сле-**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **дует** |  |  |  | − x**?** | |  | ≤ kCk · | |  | x***(*k**−***1)*** −x**?** | | | | |  |  |  |  |  |
|  |  | x***(*k*)*** | |  |  | . |  |  | **(3.76)** |  |
|  |  |  |  |  |  |  |  |  | |  | k | C | k |  |  |  |  |  |
| **Домножая обе** | **части неравенства (3.75) на** | | | | | | | | | | |  |  | **и учитывая оценку** | | | |  |
|  |  |  |  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  |
| **(3.76), получаем окончательно** | | | | | | | | |  |  |  |  |  |  |  |  |  |  |  |
|  |  | x***(*k*)*** | − | x**?** |  | ≤ |  | 1 − kCk | | x***(*k*)*** | − | x***(*k**−***1)*** | | | . | **(3.77)** |  |
|  |  |  |  |  |  | kCk |  |  |  |  |  |

**Выведенная оценка может быть использована на практике как для оценки погрешности какого-то приближения из итерационной последо-вательности, так и для определения момента окончания итераций, т. е.**

**216** 3. Численные методы линейной алгебры

**того, достигнута ли желаемая точность приближения к решению или нет.**

Пример 3.4.1 **Рассмотрим систему линейных алгебраических урав-**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **нений** |  | 3 | 4 | x = |  | 5 |  | , |
|  |  | 2 | 1 |  |  | 0 |  |  |

**точное решение которой равно** (−1,2)>**. Пусть для решения этой си-стемы организован итерационный метод Гаусса-Зейделя с начальным приближением** x***(0)*** = (0,0)>**. Через сколько итераций компоненты оче-редного приближения к решению станут отличаться от точного реше-ния не более, чем на** 10−***3*?**

**Исследуемый нами вопрос требует чебышёвской нормы** k · k∞ **для измерения отклонения векторов друг от друга, подчинённая матричная норма задаётся выражением из Предложения 3.2.2. Матрица перехода метода Гаусса-Зейделя согласно (3.68) равна**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| − | 2 | 0 | −***1*** |  | 0 | 1 | = |  | 0 | −0.5 | , |  |
| 3 | 4 |  | 0 | 0 |  | 0 | 0.375 |  |  |

**так что её** ∞**-норма равна** 0.5**. Следовательно, в оценке (3.77) имеем**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| kCk | = | 0.5 | = 1, |  |
| 1 − kCk | 1 − 0.5 |  |
|  |  |  |

**и потому должно быть справедливым неравенство**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| x***(*k*)*** −x**?** |  | ∞ | ≤ | x***(*k*)*** −x***(*k**−***1)*** | ∞. | **(3.78)** |
|  |  |  |  |  |  |  |

**Оно показывает, что компоненты очередного приближения отличаются от компонент точного решения не более, чем компоненты приближений друг от друга.**

**Запустив итерации Гаусса-Зейделя, мы можем видеть, что**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| x***(0)*** | = | (0, | 0)>, |  |
| x***(1)*** | = | (0, | 1.25)>, |  |
| x***(2)*** | = | ( | 0.625, 1.71875)>, |  |
| · · · |  | − | · · · |  |
|  |  |  |  |  |
| x***(8)*** | = (−0.998957, 1.999218)>, | | |  |
| x***(9)*** | = | ( | 0.999609, 1.999707)>, |  |
|  |  | − |  |  |

|  |  |
| --- | --- |
| 3.5. Нестационарные итерационные методы | **217** |

**т. е. 9-я итерация отличается от предыдущей 8-й меньше чем на** 10−***3*, и потому согласно оценке (3.78) получаем требуемую погрешность. То, что она действительно выполняется, видно из сравнения** x***(9)* с извест-**

**ным нам точным решением** (−1,2)>**.**

1. Нестационарные итерационные методы для линейных систем

**3.5а** **Краткая теория**

**В основу нестационарных итерационных методов для решения систем линейных алгебраических уравнений могут быть положены различные идеи.**

**В качестве первого примера рассмотрим метод простой итерации (3.62)**

x***(*k*+1)*** ← (I − τ A) x***(*k*)*** + τ b, k = 0, 1, 2, . . . ,

**исследованный нами в §3.4в. Если переписть его в виде**

|  |  |  |
| --- | --- | --- |
| x***(*k*+1)*** ←x***(*k*)*** −τ Ax***(*k*)*** −b , | k = 0, 1, 2, . . . , | **(3.79)** |

**то расчёт каждой последующей итерации** x***(*k*+1)* может трактоваться как вычитание из** x***(* k*)* поправки, пропорциональной вектору текущей невязки** (Ax***(*k*)*** −b)**. Но при таком взгляде на итерационный процесс можно попытаться изменять параметр** τ **в зависимости от шага, т. е. взять** τ=τ**k переменным, рассмотрев итерации**

|  |  |  |
| --- | --- | --- |
| x***(*k*+1)*** ←x***(*k*)*** −τ**k**Ax***(*k*)*** −b , | k = 0, 1, 2, . . . . | **(3.80)** |

**Эту нестационарную версию метода простой итерации часто связыва-ют с именем Л.Ф. Ричардсона, предложившего её идею ещё в 1910 го-ду. Он, к сожалению, не смог развить удовлетворительной теории вы-бора параметров** τ**k, и для решения этого вопроса потребовалось ещё несколько десятилетий развития вычислительной математики. Отме-тим, что задача об оптимальном выборе параметров** τ**k на группе из нескольких шагов приводит к так называемым чебышёвским цикличе-ским итерационным методам (см. [63, 35, 43]).**

**Можно пойти по намеченному выше пути дальше, рассмотрев неста-ционарное обобщение итерационного процесса**

x***(*k*+1)*** ← (I − ΛA) x***(*k*)*** + Λb, k = 0, 1, 2, . . . ,

**218** 3. Численные методы линейной алгебры

**который получен в результате матричного предобуславливания исход-ной системы линейных алгебраических уравнений. Переписав его вы-числительную схему в виде**

x***(*k*+1)*** ← x***(*k*)*** − Λ Ax***(*k*)*** − b , k = 0, 1, 2, . . . ,

**нетрудно увидеть возможность изменения предобуславливающей мат-рицы** Λ **в зависимости от номера шага. Таким образом, приходим к весьма общей схеме нестационарных линейных итерационных процес-сов**

x***(*k*+1)*** ← x***(*k*)*** − Λ**k** Ax***(*k*)*** − b , k = 0, 1, 2, . . . ,

**где** {Λ**k**}∞**k*=0* некоторая последовательность матриц, выбор которой зависит, вообще говоря, от начального приближения** x***(0)*.**

**Другой популярный путь построения нестационарных итерацион-ных методов для решения уравнений использование вариационных принципов.**

**Интуитивно понятный термин вариация был введён Ж.-Л. Лагран-жем для обозначения малого изменения (¾шевеления¿) независимо-го переменного или функционала. Соответственно, метод вариаций это метод исследования экстремальных задач, основанный на иссле-довании зависимости функционалов от вариаций аргументов. Но со временем ¾вариационными¿ стали называть методы решения различ-ных уравнений, сводящие исходную постановку задачи к определённым задачам на нахождение экстремума. Соответственно, вариационными принципами называют переформулировки интересующих нас задач в виде каких-либо задач на нахождение экстремумов. Тогда итерацион-ные методы решения СЛАУ могут конструироваться как итерационные процессы для отыскания экстремумов тех или иных функционалов.**

**Вариационные принципы получаются весьма различными способа-ми. Некоторые из них вытекают из содержательного (физического, ме-ханического и пр.) смысла решаемой задачи. Например, в классической механике хорошо известны ¾припцип наименьшего действия Лагран-жа¿, в оптике существует ¾принцип Ферма¿. В последнее столетие име-ется тенденция всё меньше связывать вариационные принципы с кон-кретным физическим содержанием, они становятся абстрактным ма-тематическим инструментом решения разнообразных задач.**

**Строго говоря, в вычислительном отношении получающаяся в ре-зультате сведения оптимизационная задача может быть не вполне эк-вивалентна исходной, так как задача нахождения устойчивого реше-ния уравнения может превратиться в неустойчивую задачу о проверке**

|  |  |
| --- | --- |
| 3.5. Нестационарные итерационные методы | **219** |

**точного равенства экстремума какому-то определённому значению (см. §4.1б). Но если существование решение уравнения известно априори, до начала процедуры численного исследования (например, на основе каких-либо теорем существования), то вариационные методы становят-ся важным подспорьем практических вычислений.**

**Как именно можно переформулировать задачу решения СЛАУ в виде экстремальной задачи? По-видимому, простейший способ может основываться на том факте, что точное решение** x**? зануляет норму невязки** kAx−bk**, доставляя ей, таким образом, наименьшее возмож-ное значение. Желая приобрести гладкость получаемого функционала, обычно берут евклидову норму невязки, или даже её квадрат, т. е. ска-лярное произведение** hAx−b, Ax−bi**, чтобы не привлекать взятия кор-ня. Получающаяся задача называется линейной задачей о наименьших квадратах, и мы рассмотрим её подробнее в §3.7.**

**Ещё одним фактом, который может служить теоретической основой для вариационных методов решения систем линейных алгебраических уравнений является**

Предложение 3.5.1 **Вектор** x**?** ∈ R**n** **является решением системы** **линейных алгебраических уравнений** Ax=b **с симметричной поло-жительно определённой матрицей** A **тогда и только тогда, когда он доставляет минимум функционалу** Φ(x) = ***12*** hAx, xi − hb, xi**.**

Доказательство. **Если** A **симметричная положительно-определён-ная матрица, то, как мы видели в §3.2а, выражением *12*** hAx, xi **задаётся так называемая энергетическая норма векторов на** R**n.**

**Далее, пусть** x**? решение рассматриваемой системы линейных ал-гебраических уравнений, так что** Ax**?** =b**. В силу единственности** x**? некоторый вектор** x∈R**n является решением системы тогда и только тогда, когда** x−x**?** = 0**, или, иными словами,** kx−x**?**k***2*A** = 0**.**

**С другой стороны, учитывая симметричность матрицы** A **и равен-ство** Ax**?** =b**, получаем**

kx − x**?**k***2*A** = ***12*** hA(x − x**?**), x − x**?**i

* ***12*** hAx, xi − ***12*** hAx**?**, xi − ***12*** hAx, x**?**i + ***12*** hAx**?**, x**?**i
* ***12*** hAx, xi − hb, xi + ***12*** hAx**?**, x**?**i

|  |  |
| --- | --- |
| = Φ(x) + ***21*** hAx**?**, x**?**i, | **(3.81)** |

**220** 3. Численные методы линейной алгебры

**т. е. энергетическая норма погрешности отличается от функционала** Φ(x) **лишь постоянным слагаемым** ***12*** hAx**?**, x**?**i **(которое заранее неиз-вестно из-за незнания нами** x**?). Следовательно,** Φ(x) **действительно достигает своего единственного минимума одновременно с** kx−x∗k***2*A, т. е. на решении рассматриваемой линейной системы.**

**Функционал** Φ(x) = ***12*** hAx, xi−hb, xi **является квадратичной формой от вектора переменных** x**, и его часто называют функционалом энер-гии из-за его сходства с выражениями для различных видов энергии в физических системах.**11

**3.5б** **Метод наискорейшего спуска**

**В предшествующем пункте были предложены две вариационные пе-реформулировки задачи решения системы линейных алгебраических уравнений. Как находить минимум соответствующих функционалов? Прежде, чем строить конкретные численные алгоритмы, рассмотрим общую схему.**

**Пусть** f:R**n** →R **некоторая функция, ограниченная снизу на всём пространстве** R**n и принимающая своё наименьшее значение в** x**?, так что**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **?** |  | f (x) | **для любых** x∈R | **n** | . |  |
| f (x) ≥ f (x ) = **x**∈Rn | |  |  |
|  | min |  |  |  |  |  |

**Нам нужно найти точку** x**?, в которой достигается наименьшее зна-чение функции** f **. Типичным методом решения этой задачи являет-ся построение последовательности значений аргумента** {x***(*k*)***}**, которая ¾минимизирует¿ функцию** f **в том смысле, что**

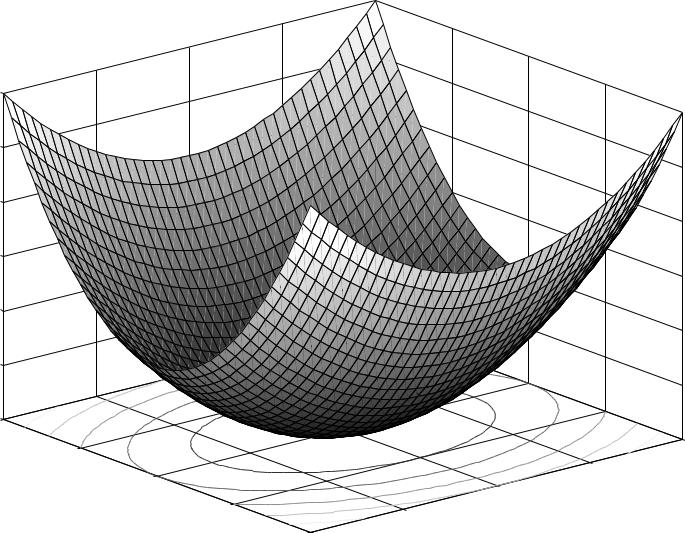
lim f (x***(*k*)***) = min f (x).

**k**→∞ **x**∈Rn

**Саму функцию** f **, для которой ищется экстремум, в теории оптимиза-ции называют обычно целевой функцией. Если построенная последова-тельность** {x***(*k*)***} **сходится к некоторому пределу** x**?, то он и является решением задачи в случае непрерывности функции** f **.**

11 **К примеру, кинетическая энергия тела массы** m**, движущегося со скоростью** v**, равна** mv***2*** /***2*. Энергия упругой деформации пружины с жёсткостью** k**, растянутой или сжатой на величину** x**, равна** kx***2***/***2*. И т. д.**

|  |  |
| --- | --- |
| 3.5. Нестационарные итерационные методы | **221** |



**Рис. 3.9. График функционала энергии и его линии уровня.**

**Метод градиентного спуска, является способом построения после-довательности, которая является минимизирующей для определённо-го класса дифференцируемых функций, и заключается в следующем. Пусть уже найдено какое-то приближение** x***(*k*)*,** k= 0,1,2, . . . **, к точ-ке минимума функции** f(x)**. Естественная идея состоит в том, что-бы из** x***(*k*)* сдвинуться по направлению наибольшего убывания целевой функции, которое противоположно направлению градиента** f0(x***(*k*)***)**, т. е. взять**

|  |  |
| --- | --- |
| x***(*k*+1)*** ←x***(*k*)*** −**k**f0(x***(*k*)*** ), | **(3.82)** |

**где k величина шага, которая выбирается из условия убывания целевой функции на рассматриваемой итерации. Далее мы можем по-вторить этот шаг ещё раз и ещё . . . столько, сколько требуется для достижения требуемого приближения к минимуму.**

**В интересующем нас случае минимизации функционала энергии** Φ(x)**, порождаемого системой линейных уравнений с симметричной по-ложительно определённой матрицей, имеем**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ∂x**l** | = ∂x**l** | |  | 2 **i*=1*** **j*=1*** a**ij** x**i**x**j** − **i*=1*** b**i**x**i** | | | = **j*=1*** a**lj** x**j** − b**l**, | |  |
| ∂Φ(x) |  | ∂ |  | 1 | **n n** | **n** |  | **n** |  |
|  |  |  |  |  | **X X** | **X** |  | **X** |  |
|  |  |  |  |  |  |  |  |  |  |

**k**(Ax***(*k*)***

**k в методе наиско-**

**222** 3. Численные методы линейной алгебры

l = 1, 2, . . . , n**, так что**

Φ0(x) = Ax − b,

**т. е. градиент функционала** Φ **равен невязке решаемой системы линей-ных уравнений в рассматриваемой точке. Важнейшим выводом из это-го факта является тот, что метод простой итерации (3.62)–(3.79) явля-ется ни чем иным, как методом градиентного спуска (3.82) для миними-зации функционала энергии** Φ**, в котором шаг k выбран постоянным и равным** τ**.**

**Вообще, выбор величины шага k является очень ответственным делом, так как от него зависит и наличие сходимости, и её скорость. Спуск по направлению антиградиента обеспечивает убывание целевой функции лишь при достаточно малых шагах, и потому при неудачно большой величине шага мы можем попасть в точку, где значение функ-ционала не меньше, чем в текущей точке. С другой стороны, слиш-ком малый шаг приведёт к медленному движению в сторону решения. Для градиентного метода с постоянным шагом его трактовка как мето-да простой итерации позволяет, опираясь на результат §3.4в, выбрать шаг k** = const**, который наверняка обеспечивает сходимость процес-са. Именно, если положительные числа** µ **и** M **это нижняя и верхняя границы спектра положительно определённой матрицы** A **решаемой системы, то в соответствии с (3.63) следует взять**

2

**k** = = M + µ .

**Другой способ выбора шага состоит в том, чтобы потребовать k наибольшим возможным, обеспечивающим убывание функционала** Φ

**вдоль выбранного направления спуска по антиградиенту. При этом получается разновидность градиентного спуска, называемая методом наискорейшего спуска, которая была развита в конце 40-х годов XX**

**века Л.В. Канторовичем.**

**Для определения конкретной величины шага**

**рейшего спуска нужно подставить выражение** x***(*k*)*** − **k** Φ0(x***(*k*)***) =x***(*k*)*** −

− b) **в аргумент функционала энергии и продифференциро-**

**вать получающееся отображение по** **k. Для удобства выкладок обо-**

|  |  |
| --- | --- |
| 3.5. Нестационарные итерационные методы | **223** |

**значим невязку**

Φ x***(*k*)*** − τ**k**r***(*k*)***

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| r***(*k*)*** := Ax***(*k*)*** − b **и шаг** τ**k** := | | | | **k. Имеем** | | | |  |  |  |
|  | ***1*** | ***(*k*) (*k*)*** | ***(*k*)*** | ***(*k*)*** |  |  | ***1 2*** | ***(*k*)*** | ***(*k*)*** |  |
| = | ***21*** | A(x***(*k*)*** − τ**k**r***(*k*)***), x***(*k*)*** − τ**k**r***(*k*)*** | | | |  | − hb, x***(*k*)*** − τ**k**r***(*k*)***i | | | |
| = ***2*** Ax , x | | | − τ**k** Ax , r |  | + ***2*** τ**k** Ar , r | | | |  |  |

− h b, x***(*k*)***i + τ**k**h b, r***(*k*)***i.

**При дифференцировании выписанного выражения по** τ**k не зависящие от него члены исчезнут, мы получим**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| d | Φ x | − τ**k**r |  |  | = − | Ax | ***(*k*)***, r | ***(*k*)*** | + τ**k**h ***(*k*)*** | , r |  | ***(***i**k*)*** | h | i |  |
| dτ**k** |  |  |  |
|  | ***(*k*)*** |  | ***(*k*)*** |  |  | ***(*k*) (*k*)*** | | | Ar***(*k*)*** |  | ***(*k*)*** +b, r***(*k*)*** | | |  |  |
|  |  |  |  |  | = τ**k** Ar | |  |  |  |  |  |
|  |  |  |  |  | , r |  | − Ax − b, r | | |  |  |  |  |

* τ**k** Ar***(*k*)*** , r***(*k*)*** − hr***(*k*)*** , r***(*k*)***i.

**Таким образом, в точке экстремума по** τ**k из условия**

dΦ x***(*k*)*** −τ**k**r***(*k*)*** = 0dτ**k**

**необходимо следует**

hr***(*k*)*** , r***(*k*)*** i

τ**k**= Ar***(*k*)***, r***(*k*)*** .

**Легко видеть, что при найденном значении** τ**k функционалом энер-гии действительно достигается минимум по выбранному направлению спуска, так как тогда его вторая производная, равная** hAr ***(*k*)*** , r***(*k*)***i**, по-ложительна. В целом, псевдокод метода наискорейшего градиентно-го спуска для решения системы линейных алгебраических уравнений** Ax = b **представлен в Табл. 3.5.**

Теорема 3.5.1 **Если** A **симметричная положительно определён-ная матрица, то последовательность** {x***(*k*)***}**, порождаемая методом наискорейшего спуска, сходится к решению** x**? системы уравнений** Ax = b **из любого начального приближения** x***(0)*, и быстрота этой** **сходимости оценивается неравенством**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| k |  | − |  | k**A** | ≤ | M + µ | | | **k** | k |  | − |  | k**A** |  |  |  |
|  | x***(*k*)*** |  | x**?** |  |  |  | M − µ |  |  | x***(0)*** |  | x**?** |  | , | **(3.83)** |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |

**где** µ**,** M **нижняя и верхняя границы спектра матрицы** A**.**

**224** 3. Численные методы линейной алгебры

**Таблица 3.5. Псевдокод метода наискорейшего спуска**

**для решения систем линейных уравнений.**

**выбираем начальное приближение** x***(0)* ;**

k ← 0 **;**

DO WHILE **( метод не сошёлся )**

r***(*k*)*** ← Ax***(*k*)*** − b **;**

hr***(*k*)*** , r***(*k*)***i

τ**k**←hAr***(*k*)*** , r***(*k*)***i**;**

x***(*k*+1)*** ←x***(*k*)*** −τ**k**r***(*k*)* ;**

k ← k + 1 **;**

END DO

Доказательство **оценки (3.83) и теоремы в целом будет получено пу-тём сравнения метода наискорейшего спука с методом градиентного спуска с постоянным оптимальным шагом. В самом деле, из равенства (3.81)**

kx − x**?**k***2*A** = Φ(x) + hAx**?**, x**?**i,

**следует, что метод, обеспечивающий лучшее убывание значения функ-ционала энергии одновременно обеспечивает лучшее приближение к решению в энергетической норме.**

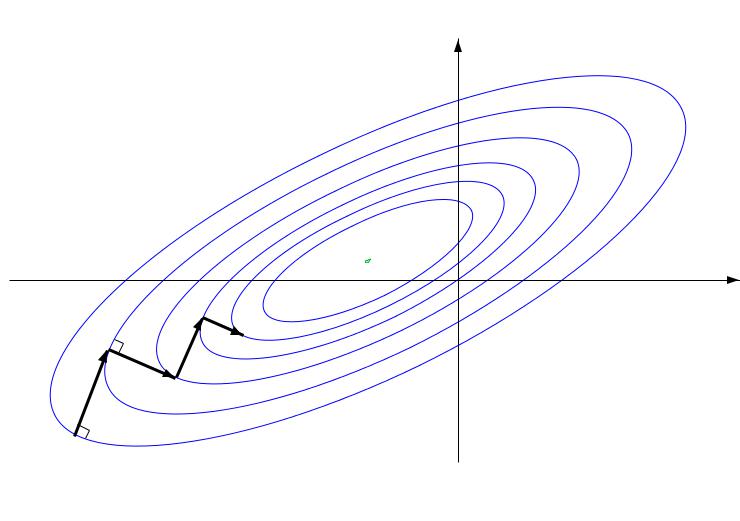
* + **методе градиентного спуска с постоянным шагом совпадающем**
* **методом простой итерации (3.62) или (3.79) имеем**

x***(*k*+1)*** − x**?** = (I − τ A)(x***(*k*)*** − x**?**).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 3.5. Нестационарные итерационные методы | | | | | | | | | | |  |  |  | **225** |  |
| **По этой причине** |  | A(I τA)(x | | | | | | |  |  |  |  |  |  |  |
| kx***(*k*+1)*** − x**?**k**A*2*** | = |  | x ), (I |  |  | x ) |  |
| = | A(x***(*k*+1)*** | | | | | − x**?**), x***(*k*+1)*** − x**?** | | | | |  |  | x**?**) |  |
|  | = | (IτA)A(x***(*k*)*** | | | | | | | − x**?**), (I − τA)(x***(*k*)*** | | | |  |  |
|  |  |  |  |  | − |  |  | ***(*k*)*** | − | **?** | − | τ A)(x***(*k*)*** | − | **?** |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | I | −τ A | | ***22*** | | A(x***(*k*)*** − x**?**), (x***(*k*)*** − x**?**) | | | | | − |  |
|  | ≤ k | | I | − | τ Ak | ***2*** |  | x***(*k*)*** | − | x**?** ***2*** . |  |  | |  |  |
|  | ≤ k | |  | − | k***2*** | | k | | k**A** |  |  |  |  |  |

**При этом у метода наискорейшего спуска оценка заведомо не хуже этой оценки, в которой взято значение параметра шага** τ= 2/(M+µ)**, оптимальное для спуска с постоянным шагом. Тогда в соответствии с оценкой (3.64) для метода простой итерации получаем**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| k |  | − |  | k**A** | ≤ | M + µ k | |  | − |  | k**A** |  |  |
|  | x***(*k*+1)*** |  | x**?** |  |  | M − µ |  | x***(*k*)*** |  | x**?** |  | , |  |
|  |  |  |  |  |  |  |  |
| **откуда следует доказываемая оценка (3.83).** | | | | | | | |  |  |  |  |  |  |



x**?**

x***(0)***

**Рис. 3.10. Иллюстрация работы метода наискорейшего спуска.**

**Интересно и поучительно рассмотреть геометрическую иллюстра-цию работы метода наискорейшего спуска.**

**Коль скоро** A **симметричная матрица, она может быть приведена к диагональной форме** D **ортогональным преобразованием подобия:**

A = Q>DQ,

**226** 3. Численные методы линейной алгебры

**причём в силу положительной определённости матрицы** A **по диаго-нали в** D **стоят положительные элементы собственные значения** A**. Подставляя это представление в выражение для функционала энергии** Φ(x)**, получим**

Φ(x) = Q>DQx, x − 2hb, xi

* D(Qx), Qx − 2hQb, Qxi
* hDy, yi − 2hQb, yi,

**где обозначено** y=Qx**. Видим, что в изменённой системе координат, которая получается с помощью ортогонального линейного преобразо-вания переменных, выражение для функционала энергии есть сумма квадратов с коэффициентами, равными собственным значениям мат-рицы член** hDy, yi**, минус линейный член** 2hQb, yi**. Таким образом, график функционала энергии это эллиптический параболоид, воз-можно, сдвинутый относительно начала координат и ещё повёрнутый, а его поверхности уровня (линии уровня в двумерном случае) эллипсо-иды (эллипсы), в центре которых находится искомое решение системы уравнений. При этом форма эллипсоидов уровня находится в зависи-мости от разброса коэффициентов при квадратах переменных, т. е. от числа обусловленности матрицы** A**. Чем больше эта обусловленность, тем сильнее сплющены эллипсоиды уровня.**

**Градиент функционала энергии нормален к поверхностям уровня, и именно по этим направлениям осуществляется ¾спуск¿ движение в сторону решения. Шаг в методе наискорейшего спуска идёт на макси-мально возможную величину до пересечения с касательным эллип-соидом. Поэтому траектория метода наискорейшего спуска является ломаной, звенья которой перпендикулярны друг другу (см. Рис. 3.10).**

**Хотя доказательство Теоремы 3.5.1 основано на мажоризации наи-скорейшего спуска методом простой итерации и может показаться до-вольно грубым, в действительности оценка (3.83) весьма точно пере-даёт особенности поведения метода, а именно, замедление сходимости при** M µ**. Тот факт, что в случае плохой обусловленности матрицы системы движение к решению в методе наискорейшего спуска весьма далеко от оптимального, подтверждается вычислительной практикой и может быть понято на основе геометрической интерпретации. Искомое решение находится при этом на дне глубокого и вытянутого оврага, а метод ¾рыскает¿ от одного склона оврага к другому вместо того, чтобы идти напрямую к глубочайшей точке решению.**

|  |  |
| --- | --- |
| 3.6. Вычисление обратной матрицы и определителя | **227** |

**Другой популярный подход к выбору итерационных параметров** τ**k в нестационарном итерационном процессе (3.80), названный методом минимальных невязок, был предложен С.Г. Крейном и М.А. Красно-сельским в работе [22].**

**Если сходимость методов наискорейшего спуска и минимальных невязок принципиально не лучше сходимости метода простой итера-ции, то имеют ли они какое-либо практическое значение? Вспомним, что наша оптимизация метода простой итерации основывалась на зна-нии границ спектра симметричной положительно определённой матри-цы СЛАУ. Для работы методов наискорейшего спуска и минимальных невязок этой информации не требуется.**

**3.5в** **Метод сопряжённых градиентов**

**В практике вычислений для решения систем линейных алгебраических уравнений большое распространение получили методы сопряжённых градиентов, которые являются двухшаговыми нестационарными ите-рационными алгоритмами.**

1. Вычисление обратной матрицы и определителя матрицы

**Нечасто, но всё-таки иногда приходится вычислять обратную к данной матрице. Например, это необходимо при нахождении дифференциала операции обращения матрицы** A7→A−***1*, равного**

d(A−***1***) = −A−***1***(dA) A−***1***.

**Матрица** A−***1*, обратная к данной матрице** A**, является решением матричного уравнения**

AX = I.

**Но это уравнение распадается на** n **уравнений относительно вектор-ных неизвестных, соответствующих отдельным столбцам неизвестной матрицы** X**, и мы можем решать получающиеся уравнения порознь.**

**Из сказанного следует способ нахождения обратной матрицы: нуж-но решить** n **штук систем линейных уравнений**

Ax = e***(*j*)***, j = 1, 2, . . . , n, **(3.84)**

**228** 3. Численные методы линейной алгебры

**где** e***(*j*)*** j**-ый столбец единичной матрицы** I**. Это можно сделать, к примеру, методом Гаусса или любым другим из рассмотренных выше методов, которые здесь особенно удобны в своей матричной трактовке. В самом деле, сначала мы можем выполнить один раз LU-разложение исходной матрицы** A**, а затем хранить его и использовать посредством схемы (3.31) для различных правых частей уравнений (3.84).**

**Другой подход конструирование чисто матричных процедур, не опирающихся на методы решения систем линейных уравнений с век-торными неизвестными. Известен итерационный метод Шульца для обращения матриц: задавшись специальным начальным приближени-ем** X***(0)*, выполняют итерации**

X***(*k*+1)*** ← X***(*k*)*** 2I − AX***(*k*)*** , k = 0, 1, 2, . . . . **(3.85)**

**Метод Шульца это не что иное как метод Ньютона для решения системы уравнений, применённый к** X−***1*** −A= 0**. Его можно также рассматривать как матричную версию известной процедуры для вы-числения обратной величины (см. [13], глава 3).**

**Отметим, что гораздо чаще встречается необходимость вычисле-ния произведения обратной матрицы** A−***1* на какой-то вектор** b**, и это произведение всегда следует находить как решение системы уравнений** Ax = b **какими-либо из методов для решения СЛАУ. Такой способ за-ведомо лучше, чем вычисление** A−***1***b **через нахождение обратной** A−***1*, как по точности, так и по трудоёмкости.**

**Рассмотрим теперь вычисление определителя матрицы.**

**Мы уже отмечали раньше, что выполняемые в методе Гаусса преоб-разования линейное комбинирование строк не изменяют величины определителя матрицы. Следовательно,** detA **равен определителю по-лучающейся в итоге верхней треугольной матрицы** U **, т. е.** detA **есть произведение диагональных элементов** U **.**

**Другая возможная трактовка этого результата состоит в том, что если** A=LU **, то, как известно из линейной алгебры,**

det A = det L · det U.

**Определитель нижней треугольной матрицы** L **равен** 1**, коль скоро на её диагонали стоят все единицы. Следовательно, как и ранее,** detA=det U **, а точнее произведению всех диагональных элементов в верхней** **треугольной матрице** U **.**

|  |  |
| --- | --- |
| 3.7. Линейная задача о наименьших квадратах | **229** |

**Совершенно аналогичные выводы можно сделать и при использо-вании других матричных разложений. Например, если нам удалось по-лучить** A=QR **разложение исходной матрицы в произведение ор-тогональной и правой треугольной, то, коль скоро** detQ= 1**, искомый определитель** detA= detR **и вычисляется по** R **как произведение её диагональных элементов.**

1. Линейная задача о наименьших квадратах

**Для заданных** m×n**-матрицы** A **и** m**-вектора** b **линейной задачей о наименьших квадратах называют задачу отыскания такого вектора** x**, который доставляет минимум квадратичной форме** hAx−b, Ax−bi**, или, что равносильно, квадрату евклидовой нормы невязки** kAx−bk***22*. Ясно, что для матриц** A **полного ранга в случае** m≤n**, когда чис-ло строк матрицы не превосходит числа столбцов, искомый минимум, как правило, равен нулю. Для квадратной матрицы** A **линейная задача о наименьших квадратах, фактически, равносильна решению системы линейных алгебраических уравнений** Ax=b **и несёт особую специфи-ку лишь когда** A **имеет неполный ранг, т. е. особенна. Теоретически и практически наиболее важный случай линейной задачи о наименьших квадратах соответствует** m > n**. Он находит многочисленные и разно-образные применения при обработке данных**

**Коль скоро**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| hAx − b, Ax − b i = hAx, Axi − hb, Axi − hAx, bi + hb, bi | | | | |  |  |
|  |  | = hAx, Axi − 2hAx, bi + hb, bi, | | |  |  |
| **то** | | |  |  |  |  |
|  | ∂ | | ∂ | |  |  |
|  |  | hAx − b, Ax − b i = |  | hAx, Axi − 2hAx, bi |  |  |
|  | ∂x**i** | ∂x**i** |  |  |
| **Система линейных алгебраических уравнений** | | | | |  |  |
|  |  | A>A x = A>b | | | **(3.86)** |  |

**называется нормальной системой уравнений для линейной задачи о наименьших квадратах с матрицей** A **и вектором** b**. Её решение и до-ставляет искомый минимум выражению** kAx−bk***22***

**230** 3. Численные методы линейной алгебры

1. Численное решение проблемы собственных значений

**3.8а** **Обсуждение постановки задачи**

**Ненулевой вектор называется собственным вектором квадратной мат-рицы, если в результате умножения на эту матрицу он переходит в коллинеарный ему, т. е. отличающийся от исходного только скалярным множителем. Сам скаляр, который является коэффициентом пропор-циональности исходного вектора и его образа при действии матрицы, называют собственным значением матрицы. Проблемой собственных значений называют задачу определения собственных значений и соб-ственных векторов матриц: для заданной** n×n**-матрицы** A **найти числа** λ **и** n**-векторы** v =6 0**, удовлетворяющие условию**

|  |  |
| --- | --- |
| Av = λv. | **(3.87)** |

**Как собственные числа** λ**, так и собственные векторы** v **являются, вооб-ще говоря, комплексными. Из курса линейной алгебры читателю долж-но быть известно, что задача нахождения собственных значений мат-рицы** A **эквивалентна задаче нахождения корней характеристического (векового) уравнения матрицы** det(A−λI) = 0**.**

**Иногда при упоминании этой задачи подчёркивают ¾алгебраиче-ская проблема собственных значений¿, чтобы уточнить, что речь идёт о матрицах конечных размеров, конечномерной ситуации и т. п. в отли-чие, скажем, от задачи нахождения собственных значений операторов в бесконечномерных пространствах функций. Слово ¾проблема¿ также уместно в этом контексте, поскольку рассматриваемая задача сложна и имеет много аспектов.**

**Собственные значения матриц нужно знать во многих приложени-ях. Например, задача определения частот собственных колебаний меха-нических систем (весьма важная при проектировании и механических конструкций и не только) сводится к нахождению собственных значе-ний так называемых матриц жёсткости этих систем.**

Пример 3.8.1 **Линейные динамические системы с дискретным вре-менем вида**

|  |  |  |
| --- | --- | --- |
| x***(*k*+1)*** =Ax***(*k*)*** +b***(*k*)***, | k = 0, 1, 2 . . . , | **(3.88)** |

**служат моделями разнообразных процессов окружающего нас мира, от биологии до экономики.**

Ax***(*k*)***

|  |  |
| --- | --- |
| 3.8. Решение проблемы собственных значений | **231** |

**Общее решение такой системы есть сумма частного решения исход-ной системы (3.88) и общего решения однородной системы** x***(*k*+1)*** =

**без свободного члена. Если искать нетривиальные решения од-**

**нородной системы в виде** x***(*k*)*** =λ**k** h**, где** λ **ненулевой скаляр и** h∈Rn**-вектор, то нетрудно убедиться, что** λ **должно быть собственным** **значением** A**, а** h **собственным вектором матрицы** A**.**

**Различают полную проблему собственных значений и частичную проблему собственных значений. В полной проблеме требуется нахож-дение всех собственных чисел и собственных векторов. Частичная про-блема собственных значений это задача нахождения некоторых соб-ственных чисел матрицы и/или некоторых собственных векторов. К примеру, наибольшего по модулю собственного значения, или несколь-ких наибольших по модулю собственных значений и соответствующих им собственных векторов.**

**Ясно, что собственные векторы матрицы определяются неоднознач-но, с точностью до скалярного множителя. В связи с этим часто говорят**

* **нахождении одномерных инвариантных подпространств матрицы. Инвариантные подпространства матрицы могут иметь и б´ольшую раз-мерность, и в любом случае их знание доставляет важную информацию**
* **рассматриваемом линейном операторе, позволяя упростить его пред-ставление. В последние десятилетия задача определения для матрицы тех или иных инвариантных подпространств, не обязательно одномер-ных, также включается в ¾проблему собственных значений¿.**

**Особенности проблемы собственных значений, осложняющие её ре-шение:**

* + **необходимость выхода в комплексную плоскость** C**, даже для вещественных матриц;**
  + **нелинейный характер задачи, несмотря на традицию отнесения её к вычислительной линейной алгебре.**

**Последнее обстоятельство нетрудно осознть из рассмотрения основного соотношения (3.87)**

Av = λv,

**которое является системой уравнений относительно** λ **и** v**, причём в его правой части суммарная степень неизвестных переменных равна двум: 2 = (1 при** λ**) + (1 при** v**).**

**Система уравнений (3.87) кажется недоопределёной, так как содер-жит** n+ 1 **неизвестных, которые нужно найти из** n **уравнений. Но на**

**232** 3. Численные методы линейной алгебры

**самом деле можно замкнуть её, к примеру, каким-нибудь условием нор-мировки собственных векторов (**kvk= 1 **в какой-то норме) или требова-нием, чтобы какая-нибудь компонента** v **принимала бы заданное зна-чение. Последнее условие иногда даже более предпочтительно ввиду своей линейности.**

**Если собственное значение** λ**i матрицы** A **уже найдено, то, как из-вестно, определение соответствующих собственных векторов сводится к решению системы линейных алгебраических уравнений**

(A − λ**i**I)x = 0

**с особенной матрицей. Но на практике часто предпочитают пользо-ваться для нахождения собственных векторов специализированными вычислительными процедурами. Многие из них позволяют вычислять собственные векторы одновременно с собственными значениями мат-риц.**

**В заключение нашего обсуждения коснёмся алгоритмического ас-пекта проблемы собственных значений. Напомним известную в алгеб-ре теорему Абеля-Руффини: для алгебраических полиномов степени выше 4 не существует прямых методов нахождения корней. Как след-ствие, мы не вправе ожидать существования прямых методов решения проблемы собственных значений для произвольных матриц размера более** 4×4**, и потому рассматриваемые ниже методы существенно итерационные.**

**3.8б** **Обусловленность проблемы собственных значений**

**Спектр матрицы, как множество точек комплексной плоскости** C**, непре-рывно зависит от элементов матрицы. Соответствующий результат ча-сто называют теоремой Островского (читатель может увидеть деталь-ное изложение этой теории в книгах [18, 24, 31, 39, 48]. Но собственные векторы (инвариантные подпространства) матрицы могут изменять-ся в зависимости от матрицы разрывным образом даже в совершенно обычных ситуациях.**

Пример 3.8.2 **[48] Рассмотрим матрицу**

|  |  |  |  |
| --- | --- | --- | --- |
| A = 1 | 0 | 1 | . |
|  | + ε | δ |  |

|  |  |
| --- | --- |
| 3.8. Решение проблемы собственных значений | **233** |

**Её собственные значения суть числа** 1 **и** 1 +ε**, и при** εδ=60 **соответ-ствующими нормированными собственными векторами являются**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| √ε***2*** +δ***2*** | | | ε |  |  | 0 |  |  |
| 1 | |  | −δ |  | **и** | 1 | . |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

**Выбирая подходящим образом отношение** ε/δ**, можно придать первому собственном вектору любое направление, сколь бы малыми ни явля-лись значения** ε **и** δ**.**

**Если положить** ε= 0**, то**

|  |  |  |  |
| --- | --- | --- | --- |
| A = | 0 | 1 | . |
|  | 1 | δ |  |

**При** δ=60 **у матрицы** A **будет всего один собственный вектор, хотя при надлежащем** δ **её можно сделать сколь угодно близкой к единичной матрице, имеющей два линейно независимых собственных вектора.**

**При более пристальном изучении проблемы собственных значений выясняется, что, несмотря на непрерывную зависимость собственных значений от элементов матрицы, скорость их изменения может быть сколь угодно большой (даже для матриц фиксированного размера), если они соответствуют так называемым нелинейным элементарным делителям матрицы жордановым клеткам размера 2 и более.**

Пример 3.8.3 **Собственные значения матрицы**

λ 1

A =

ε λ

**возмущённой жордановой** 2×2**-клетки равны** λ±√ε**, так что мгновенная скорость их изменения, равная** √ε/ε**, при** ε= 0 **бесконечна. Это же явление имеет место и для произвольной жордановой клетки, размером более двух.**

**Всюду далее в этом параграфе большую роль будут играть матри-цы, которые преобразованием подобия можно привести к диагонально-му виду так называемые матрицы простой структуры или диаго-нализуемые матрицы.**12 **Собственные числа таких матриц зависят от возмущений гораздо более ¾плавным образом¿, чем в общем случае.**

12 **Такие матрицы называют также** недефектными**.**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **234** |  |  |  |  | 3. |  | Численные методы линейной алгебры | | | |  |
| Теорема Бауэра-Файка. **Если** A **квадратная матрица простой** | | | | | | | | | | |  |
| **структуры,** λ**i**(A) | **её собственные числа,** V | | | | | | | | **матрица из соб-** | |  |
| **ственных векторов** A**, а** | | | | | ˜ |  |  |  |  |  |  |
| λ **собственное число возмущённой мат-** | | | | | |  |
| **рицы** A+A**, то** |  |  |  | − |  |  | ≤ | ***2***k | k***2*** |  |  |
| **i** |  | ˜ | **i** | **(3.89)** |  |
| min | |  |  | λ (A) |  |  | cond (V ) | A . |  |
|  |  |  |  |  |  |  |  |  |  |  |  |

˜

Доказательство. **Если** λ **совпадает с каким-то из собственных значе-ний исходной матрицы** A**, то левая часть доказываемого неравенства**

**зануляется, и оно, очевидно, справедливо. Будем поэтому предпола-**

˜

**гать, что** λ **не совпадает ни с одним из** λ**i**(A)**,** i= 1,2, . . . , n**. Следова-тельно, если, согласно условию теоремы**

V −***1***AV = D,

**где** D= diag{λ***1***, λ***2***, . . . , λ**n**} **диагональная матрица с собственными**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **числами матрицы** | | | | A **по диагонали, то матрица** | | | | | | ˜ |  |
| D − λI **неособенна.** |  |
| **С другой стороны, матрица** A+ | | | | | | | | | ˜ | **является особенной по** |  |
| A − λI |  |
| **построению, так что особенна и матрица** V−***1*** A+A−λI˜V **. Но** | | | | | | | | | | |  |
|  | − |  |  | − |  |  | = (D − λI˜ | ) I + (D λI˜)−***1***V −***1***(ΔA)V , | | |  |
| V |  | ***1*** | A + A |  | λI˜ |  | V = (D λI˜ | ) + V −***1***(ΔA)V | | |  |
|  |  |  |  |  |  |  | − |  | − |  |  |

**откуда можно заключить о том, что особенной должна быть матрица**

˜ −***1*** −***1***

I + (D − λI) V (ΔA)V **. Как следствие, матрица**

˜ −***1*** −***1***

(D − λI) V (ΔA)V

**имеет собственное значение** −1**, и потому любая норма этой матрицы должна быть не меньшей** 1**. В частности, это верно для спектральной нормы:**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Отсюда** |  |  | (D − λI˜)−***1***V −***1***(ΔA)V ***2*** ≥ 1. | | | | | | | | | | | | | | | | | | |  |  |  |
|  | ***1***≤**i**≤**n** | |  | | − | ˜ | | − | | ***1*** |  | · k | |  |  | ***1*** |  |  |  |  | k***2*** | ≥ |  |  |
|  |  | |  | **i** |  | λ) | |  | |  |  |  | | V | − | | k***2*** k k***2*** k | | | V |  |  | 1. |  |
|  | max |  | (λ | |  |  | |  |  |  |  |  |  |  |  |  |  |  | A |  |  |  |  |  |
| **Последнее** | **неравенство равносильно** | | | | | | | | | | | | | | |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  | **i** − | |  |  |  |  |  |  | ≤ k | | |  |  |  | k***2*** k k***2*** k k***2*** | | | | |  |
|  | ***1***≤**i**≤**n** | | |  |  | | ˜ | | − | |  |  |  | | |  |  | − |  | | | | |  |
|  | min | |  | (λ |  |  |  | |  |  | ***1*** |  |  |  |  | V | | ***1*** | A |  | V |  | , |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 3.8. Решение проблемы собственных значений | | | | | | | | |  | **235** |  |
| **или** |  |  | − |  |  | ≤ |  | k | k***2*** |  |  |
| **i** | ˜ | **i** | ***2*** |  |  |
| min |  |  | λ (A) |  |  | cond (V ) |  | A , | |  |
| **как и требовалось.** |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |

**Теорема Бауэра-Файка показывает, что, каково бы ни было возму-щение** A **матрицы простой структуры** A**, для любого собственного**

˜

**значения** λ **возмущённой матрицы** A+ A **найдётся собственное значе-**

˜

**ние** λ**i матрицы** A**, отличающееся от** λ **не более чем на величину спек-тральной нормы возмущения** kAk***2*, умноженную на число обуслов-ленности матрицы собственных векторов. Таким образом, обусловлен-ность матрицы собственных векторов матрицы может служить мерой обусловленности проблемы собственных значений. Практическую цен-ность неравенства (3.89) и теоремы Бауэра-Файка в целом снижает то обстоятельство, что собственные векторы матрицы определены с точ-ностью до скалярного множителя, и потому** cond***2***(V) **есть величина, заданная не вполне однозначно. Наилучшим выбором для** cond***2***(V) **в неравенстве был бы, очевидно, минимум чисел обусловленности мат-риц из собственных векторов, но его нахождение является сложной задачей. Тем не менее, прикидочные оценки и качественные выводы на основе теоремы Бауэра-Файка делать можно.**

Предложение 3.8.1 **Матрицы простой структуры образуют отк-рытое всюду плотное подмножество во множестве всех квадратных матриц.**

**Как следствие, матрицы с нелинейными элементарными делителя-ми, которые соответствуют жордановым клеткам размера 2 и более, составляют множество первой бэровской категории во множестве всех матриц. Подобные множества, называемые также тощими, являются в топологическом смысле наиболее разреженными и бедными множе-ствами (см. [17, 46]). Но на долю таких матриц приходятся главные трудности, с которыми сталкиваются при решении проблемы собствен-ных значений. В этом отношении задача нахождения сингулярных чи-сел и сингулярных векторов является принципиально другой, так как симметричная матрица** A>A **(эрмитова матрица** A∗A **в комплексном случае) всегда имеет простую структуру, т. е. диагонализуема.**

**236** 3. Численные методы линейной алгебры

**3.8в** **Коэффициенты перекоса матрицы**

**Целью этого пункта является детальное исследование устойчивости ре-шения проблемы собственных значений в упрощённой ситуации, когда все собственные значения матрицы** A **различны. Именно в этом случае, как было отмечено в §3.8б, собственные векторы непрерывно зависят от элементов матрицы и, более того, существуют их конечные диффе-ренциалы.**

**Нам будет необходим вспомогательный результат, касающийся так называемой сопряжённой задачи на собственные значения. Так назы-вают задачу нахождения собственных чисел и собственных векторов для эрмитово сопряжённой матрицы** A∗**:**

A∗y = µy,

**где** µ∈C **собственное значение матрицы** A∗ **и** y∈C**n соответ-ствующий собственный вектор. Как связаны между собой собственные значения и собственные векторы исходной** A **и сопряжённой** A∗ **мат-риц?**

**Два набора векторов** {r***1***, r***2***, . . . , r**m** } **и** {s***1***, s***2***, . . . , s**m**} **в евклидовом пространстве, состоящие из одного и того же их числа, называются биортогональными, если** hr**i**, s**i**i= 0 **при** i=6j**. Приставка ¾би¿ означает отнесение этого свойства к двум наборам векторов.**

**Ясно, что выполнение свойства биортогональности существенно за-висит от порядка нумерации векторов в пределах каждого из наборов, так что в определении биортогональности неявно предполагается, что необходимые нумерации существуют и рассматриваемые наборы упо-рядочены в соответствии с ними. Нетрудно также понять, что если какой-либо набор векторов биортогонален сам себе, то он ортогонален в обычном смысле.**

Предложение 3.8.2 **Собственные значения эрмитово-сопряжённых** **матриц попарно комплексно сопряжены друг другу. Собственные век-торы сопряжённых задач биортогональны.**

Доказательство. **Определитель матрицы не меняется при её транс-понировании,** detA>= detA**. Комплексное сопряжение элементов мат-рицы влечёт комплексное сопряжение её определителя,** detA= detA**. Следовательно,**



det(A − λI) = det(A> − λI) = det A> − λI = det(A∗ − λI).

|  |  |
| --- | --- |
| 3.8. Решение проблемы собственных значений | **237** |

**Отсюда мы можем заключить, что комплексное число** λ **является кор-нем характеристического полинома** det(A−λI) = 0 **матрицы** A **тогда и только тогда, когда ему сопряжённое** λ **является корнем полинома** det(A∗ − λI)**, который является характеристическим для матрицы** A∗**.** **Это доказывает первое утверждение.**

**Пусть** x **и** y **собственные векторы матриц** A **и** A∗ **соответственно, а** λ **и** µ **отвечающие им собственные числа этих матриц. Для дока-зательства второго утверждения выпишем следующую цепочку преоб-разований:**

hx, yi = λ1 hλx, yi = λ1 hAx, yi = λ1 hx, A∗yi = λ1 hx, µyi = µλ hx, yi.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Поэтому** |  |  |  |  |  |  |
| hx, yi − | µ | hx, yi = 0, | | | |  |
| λ |  |
| **то есть** |  |  |  |  | = 0. |  |
| hx, yi 1 − λ | | | | |  |
|  |  | | µ |  |  |  |

**Если** x **и** y **являются собственными векторами матриц** A **и** A∗**, отвечаю-щими собственным значениям** λ **и** µ**, которые не сопряжены комплекс-но друг другу, то второй сомножитель** (1−µ/λ)6= 0**. По этой причине необходимо** hx, yi= 0**, что и требовалось доказать.**

**Обратимся теперь к анализу возмущений собственных чисел и соб-ственных векторов матриц. Пусть** A **данная матрица и** dA **диф-ференциал (главная линейная часть) её возмущения, так что** A+dA **это близкая к** A **возмущённая матрица. Как изменятся собственные значения и собственные векторы матрицы** A+dA **в сравнении с соб-ственными значениями и собственными векторами** A**?**

**Имеем**

Ax**i** = λ**i**x**i**,

(A + dA)(x**i** + dx**i**) = (λ**i** + dλ**i**)(x**i** + dx**i**),

**где через** λ**i обозначены собственные значения** A**,** x**i собственные век-торы,** i= 1,2, . . . , n**, причём последние образуют базис в** R**n, коль скоро по предположению** A **является матрицей простой структуры. Пренебре-гая членами второго порядка малости, можем выписать равенство**

|  |  |
| --- | --- |
| (dA) x**i** + A (dx**i**) = λ**i**(dx**i**) + (dλ**i**) x**i**. | **(3.90)** |

**238** 3. Численные методы линейной алгебры

**Пусть** y***1*,** y***2*, . . . ,** y**n собственные векторы эрмитово-сопряжённой матрицы** A∗**, соответствующие её собственным значениям** λ***1*,** λ***2*, . . . ,** λ**n. Умножая скалярно равенство (3.90) на** y**j** **, получим**

|  |  |
| --- | --- |
| h(dA) x**i**, y**j** i + hA (dx**i**), y**j** i = λ**i**hdx**i**, y**j** i + (dλ**i**)hx**i**, y**j** i. | **(3.91)** |

**В частности, при** j=i **имеем**

h(dA) x**i**, y**i**i + hA (dx**i**), y**i**i = λ**i**hdx**i**, y**i**i + (dλ**i**)hx**i**, y**i**i,

**где соседние со знаком равенства члены можно взаимно уничтожить: они оказываются одинаковыми, коль скоро**

hA(dx**i**), y**i**i = hdx**i**, A∗y**i**i = hdx**i**, λ**i**y**i**i = λ**i**hdx**i**, y**i**i.

**Следовательно,**

h(dA) x**i**, y**i**i = (dλ**i**)hx**i**, y**i**i,

**и потому**

dλ**i** = h(dA) x**i** , y**i**i . hx**i**, y**i**i

**Пусть теперь** j6=i**. Тогда** hx**i**, y**j** i= 0 **в силу Предложения 3.8.2, и потому**

hA(dx**i**), y**j** i = h dx**i**, A∗y**j** i = h dx**i**, λ**j** y**j** i = λ**j** h dx**i**, y**j** i.

**Подставляя этот результат в (3.91), будем иметь**

h(dA) x**i**, y**j** i + λ**j** h dx**i**, y**j** i = λ**i**h dx**i**, y**j** i.

**Поэтому**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| h | dx , y | **j** i | = | h(dA) x**i**, y**j** i | . |  |
|  |
| **i** |  | λ**i** − λ**j** | |  |

**Разложим** dx**i по базису из собственных векторов невозмущённой матрицы** A**:**

**n**

**X**

dx**i** = α**ij** x**j** .

**j*=1***

**Так как собственные векторы задаются с точностью до множителя, то в этом разложении коэффициенты** α**ii содержательного смысла не**

h(dA) x**i**, y**j** i

|  |  |
| --- | --- |
| 3.8. Решение проблемы собственных значений | **239** |

**имеют, и можно считать, что** α**ii** = 0 **(напомним, что мы, в действитель-ности, ищем возмущение одномерного инвариантного подпространства матрицы). Для остальных коэффициентов имеем** hdx**i**, y**j** i=α**ij** hx**j** , y**j** i**, опять таки в силу Предложения 3.8.2. Следовательно, для** i6=j

α**ij** = (λ**i** − λ**j** )hx**j** , y**j** i .

**Перейдём теперь к оцениванию возмущений собственных значений и собственных векторов. Из формулы для дифференциала** dλ**i и из неравенства Коши-Буняковского следует**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | | dλ | **i**| ≤ | |  | kdAk***2*** kxk***2*** kyk***2*** | | | = ν | dA | , |  |
|  |  |
|  |  | hx**i**, y**i**i | |  |  | **i**k k***2*** |  |  |
| **где посредством** |  |  |  |  |  |  |  |  |  |  |  |
|  | ν**i** | = | kx**i**k***2*** ky**i**k***2*** | | | , | i = 1, 2, . . . , n, | | |  |  |
|  |  |  |  | hx**i**, y**i**i | | |  |  |  |  |  |

**обозначены величины, называемые коэффициентами перекоса матри-цы** A**, отвечающие собственным значениям** λ**i,** i= 1,2, . . . , n**.**

**Ясно, что** ν**i** ≥1**, и можно интерпретировать коэффициенты пере-**

**коса как**

1

ν**i** =cosϕ**i** ,

**где** ϕ**i угол между собственными векторами** x**i и** y**i исходной и сопря-жённой матриц. Коэффиценты перекоса характеризуют, таким обра-зом, обусловленность проблемы собственных значений в смысле второ-го подхода §1.2.**

**Для симметричной (или, более общо, эрмитовой) матрицы коэффи-циенты перекоса равны** 1**. В самом деле, сопряжённая к ней задача на собственные значения совпадает с ней самой, и потому в наших обо-**

**значениях** x**i** =y**i,** i= 1,2, . . . , n**. Следовательно,** hx**i**, y**i**i=hx**i**, x**i**i=kx**i**k***2*** ky**i**k***2*, откуда и следует** ν**i** = 1**.**

**Это наименьшее возможное значение коэффициентов перекоса, так что численное нахождение собственных значений симметричных (эр-митовых в комплексном случае) матриц является наиболее устойчи-вым.**

**Что касается возмущений собственных векторов, то коэффициенты** α**ij** **их разложения оцениваются сверху как**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | α | k(dA) x**i**k***2*** ky**j** k***2*** | | ≤ | kdAk***2*** ν | | , |  |
| | | **ij** | ≤ |  |  |  |  | **j** |  |  |
|  |  | |λ**i** − λ**j** | · |hx**j** , y**j** i| | |  | |λ**i** − λ**j** | |  |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **240** | 3. Численные методы линейной алгебры | | | |  |
| **и потому имеет место общая оценка** | |  |  |  |  |
|  | **X** | ν**j** | | **(3.92)** |  |
|  |  |  |  |
| kdx**i**k***2*** | ≤ kdAk***2*** · kxk***2*** · **j*=*i** |  | . |  |
| |λ**i** − λ**j** | |  |
|  | 6 |  |  |  |  |

**Отметим значительную разницу в поведении возмущений собствен-ных значений и собственных векторов матриц. Из оценки (3.92) сле-дует, что на чувствительность отдельного собственного вектора вли-яют коэффициенты перекоса всех собственных значений матрицы, а не только того, которое отвечает этому вектору. Кроме того, в зна-менателях слагаемых из правой части (3.92) присутствуют разности** λ**i** − λ**j** **, которые могут быть малыми при близких собственных значе-ниях матрицы. Как следствие, собственные векторы при этом очень чувствительны к возмущениям в элементах матрицы. Это мы могли наблюдать в Примере 3.8б. В частности, даже для симметричных (эр-митовых) матриц задача отыскания собственных векторов может ока-заться плохообусловленной.**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Пример 3.8.4 **Вещественная** 20 × 20**-матрица** | | | | | |  |  |  |
| 20 | | 19 20 |  | 0 | |  |  |
|  |  | 20 |  |  |  |  |  |  |
|  |  | 18 20 | |  |  |  |  |  |
|  |  | **. .** |  | **..** |  |  | , |  |
|  |  | **.** | **.** |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  | 2 20 | |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  | ε |  |  |  | 1 |  |  |  |
|  |  |  |  |  |  |  |  |  |

**в которой ненулевыми являются две диагонали главная и верхняя побочная, а также элемент в позиции** (20,1)**, называется матрицей Уилкинсона (см. [40]). При** ε= 0 **эта матрица имеет, очевидно, раз-личные собственные значения 1, 2, . . . , 18, 19, 20. Но в общем случае характеристическое уравнение матрицы Уилкинсона есть**

(20 − λ)(19 − λ) . . . (1 − λ) − 20***19***ε = 0,

**и его свободный член, который равен** 20!−20−***19***ε**, зануляется при** ε=20−***19*** · 20! ≈ 4.64 · 10 −***7*. Как следствие, матрица будет иметь при этом** **нулевое собственное значение, т. е. сделается особенной.**

|  |  |
| --- | --- |
| 3.8. Решение проблемы собственных значений | **241** |

**Величина возмущения, изменившего в рассмотренном примере наи-меньшее собственное значение с 1 до 0, примерно равна одинарной точ-ности представления на ЭЦВМ машинных чисел в районе единицы. Как видим, несмотря на то, что все собственные числа матрицы раз-личны и, следовательно, являются гладкими функциями от элементов матрицы, скорость их изменения настолько велика, что практически мы как будто бы имеем дело с разрывными функциями.**

**3.8г** **Круги Гершгорина**

**Пусть** A= (a**ij** ) **квадратная матрица из** R**n**×**n или** C**n**×**n. Если** λ∈C

**её собственное значение, то**

|  |  |
| --- | --- |
| Av = λv | **(3.93)** |

**для некоторого собственного вектора** v∈C**n. Предположим, что в** v **наибольшее абсолютное значение имеет компонента с номером** l**, так что** |v**l**|= max***1***≤**j**≤**n** |v**j** |**. Коль скоро собственные векторы матрицы определены с точностью до скалярного множителя, можно нормиро-вать** v **так, чтобы**

v = (x***1***, . . . , x**l**−***1***, 1, x**l*+1***, . . . , x**n**)>,

**причём** |v**j** | ≤1 **для** j6=l**.**

**Рассмотрим** l**-ую компоненту векторного равенства (3.93):**

**n**

**X**

a**lj** v**j** = λv**l** = λ.

**j*=1***

**Из неё следует при сделанных относительно вектора** v **предположени-ях, что**

**n**

**X**

λ − a**ll** = a**lj** v**j** .

**j*=***6**l**

**По этой причине**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |λ − a**ll**| = |  | **n** | a**lj** v**j** |  | ≤ | **n** | |a**lj** v**j** | = | **n** | |a**lj** | |v**j** | ≤ | **n** | |a**lj** |, |  |
|  |  |  |  |  |  | **X** |  | **X** |  | **X** |  |  |
|  | **X** | |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | **j**6***=*l** |  |  |  | **j**6***=*l** |  | **j**6***=*l** |  | **j**6***=*l** |  |  |

**Не зная собственного вектора** v**, мы не располагаем и номером** l **его наибольшей по модулю компоненты. Но можно действовать наверня-ка, рассмотрев дизъюнкцию соотношений выписанного выше вида для**

**242** 3. Численные методы линейной алгебры

**всех** l= 1,2. . . , n**, так как хотя бы для одного из них непременно спра-ведливы выполненные нами рассуждения. Потому в целом, если** λ **какое-либо собственное значение рассматриваемой матрицы** A**, должно выполняться хотя бы одно из неравенств**

**X**

|λ − a**ii**| ≤ |a**ij** |, i = 1, 2, . . . , n.

**j*=***6**i**

**Каждое из этих соотношений на** λ **определяет на комплексной плоско-сти** C **круг с центром в точке** a**ii и радиусом, равным Pj*=***6**i** |a**ij** |**. Мы приходим, таким образом, к результату, который был установлен в 1931 году С.А. Гершгориным:**

Теорема 3.8.1 **Все собственные значения** λ(A) **любой вещественной** **или комплексной** n×n**-матрицы** A= (a**ij** ) **расположены в объединении**

**P**

**кругов комплексной плоскости с центрами** a**ii и радиусами j*=***6**i** |a**ij** |**,** i = 1, 2, . . . , n**, т. е.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| λ(A) ∈ | **n** | **n** z ∈ C |z − a**ii**| ≤ | **j**6***=*i** |a**ij** | **o**. |  |
| **i*=1*** |  |
|  | **[** |  | **P** |  |
|  |  |  |
|  |  |  |  |  |

**Фигурирующие в условиях теоремы круги комплексной плоскости**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **n** z ∈ C |z − a**ii**| ≤ **Pj**6***=*i** |a**ij** | **o**, | | | | | i = 1, 2, . . . , n, |  |
| **называются** | **кругами Гершгорина.** | |  |  |  |  |
|  |  |  |  |  |  |
| Пример 3.8.5 **Для** 2 × 2**-матрицы (3.3)** | | | | |  |  |
|  |  | 3 | 4 | , |  |  |
|  |  | 1 | 2 |  |  |  |

**рассмотренной в Примере 3.1.1 (стр. 123), собственные значения суть**

√

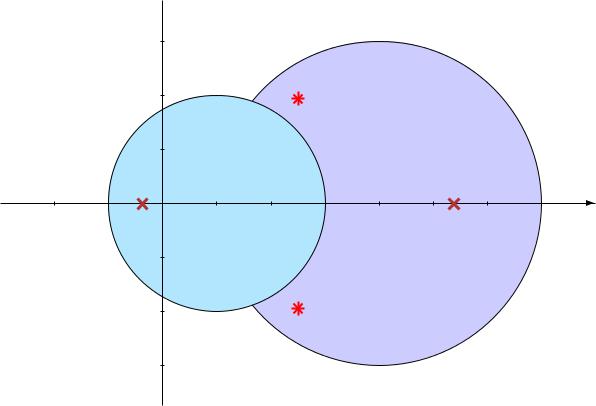
***12*** (5 ± 33)**, они приблизительно равны** −0.372 **и** 5.372**. На Рис. 3.11,** **показывающем соответствующие матрице круги Гершгорина, эти соб-ственные значения выделены крестиками.**

**Для матрицы (3.4)**

|  |  |  |
| --- | --- | --- |
| −3 | 4 | , |
| 1 | 2 |  |

|  |  |
| --- | --- |
| 3.8. Решение проблемы собственных значений | **243** |

 Im



2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| -2 | 0 | 2 | 4 | 6 |

Re

-2

**Рис. 3.11. Круги Гершгорина для матриц (3.3) и (3.4).**

**которая отличается от матрицы (3.3) лишь противоположным знаком**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **элемента на месте** (2,1)**, круги Гершгорина те же. Но собственные зна-** | | | | |  |
| ***1*** | √ |  |  |  |  |
|  |  |  |
| **чения у неё комплексные, равные *2*** (5±i | 15)**, т. е. приблизительно** | | | |  |
| 2.5 ± 1.936 i **. Они выделены на Рис. 3.11 звёздочками.** | | | |  |  |

**Бросается в глаза ¾избыточность¿ кругов Гершгорина, которые в качестве области локализации собственных значений очерчивают очень большую область комплексной плоскости. Это характерно для матриц с существенной внедиагональной частью. Но если недиагональные эле-менты матрицы малы сравнительно с диагональными, то информация, даваемая кругами Гершгорина, становится весьма точной.**

**3.8д** **Степенной метод**

**Если для собственных значений** λ**i,** i= 1,2, . . . , n**, некоторой матрицы справедливо неравенство** |λ***1***|>|λ***2***| ≥ |λ***3***| ≥. . .|λ**n**|**, то** λ***1* называют**

**доминирующим собственным значением, а соответствующий ему соб-ственный вектор доминирующим собственным вектором. Степен-ной метод, описанию которого посвящён этот пункт, предназначен для решения частичной проблемы собственных значений нахождения до-минирующих собственного значения и собственного вектора матрицы.**

**Лежащая в его основе идея чрезвычайно проста и состоит в том, что если у матрицы** A **имеется собственное значение** λ***1*, превосходящее**

**244** 3. Численные методы линейной алгебры

**по модулю все остальные собственные значения, то при действии этой матрицей на произвольный вектор** x∈C**n направление** v***1*, отвечаю-щее этому собственному значению** λ***1*, скорее всего, будет растягивать-ся сильнее остальных (при** λ***1*** ≥1**) или сжиматься меньше остальных (при** λ***1*** ≤1**). При повторном умножении** A **на результат** Ax **пред-шествующего умножения эта компонента ещё более удлинится в срав-нении с остальными. Повторив рассмотренную процедуру умножения достаточное количество раз, мы получим вектор, в котором полностью преобладает направление** v***1*, т. е. практически будет получен прибли-жённый собственный вектор.**

**В качестве приближённого собственного значения матрицы** A **мож-но при этом взять ¾отношение¿ двух последовательных векторов, по-рождённых нашим процессом** x***(*k*+1)*** =A**k*+1***x***(0)* и** x***(*k *)*** =A**k**x***(0)*,** k = 0, 1, 2, . . . **. Слово ¾отношение¿ взято здесь в кавычки потому, что** **употреблено не вполне строго: ясно, что векторы** x***(*k*+1)* и** x***(*k*)* могут оказаться неколлинеарными, и тогда их ¾отношение¿ смысла иметь не будет. Возможны следующие пути решения этого вопроса:**

**1) рассматривать отношение каких-нибудь фиксированных компо-нент векторов** x***(*k*+1)* и** x***(*k*)*, т. е.** x***(*ik*+1)***/x***(*ik*)* для некоторого** i∈

{1, 2, . . . , n}**;**

1. **рассматривать отношение проекций последовательных прибли-жений** x***(*k*+1)* и** x***(*k*)* на какое-нибудь направление** l***(*k*)*, т. е.**

|  |  |  |  |
| --- | --- | --- | --- |
| hx***(*k*+1)*** , l***(*k*)***i | . | **(3.94)** |  |
|  |
| hx***(*k*)*** , l***(*k*)***i | |  |  |

**Во втором случае мы обозначили направление проектирования через** l***(*k*)*, подчёркивая его возможную зависимость от номера шага** k**. Ясно** **также, что это направление** l***(*k*)* не должно быть ортогональным векто-ру** x***(*k*)* , чтобы не занулился знаменатель в (3.94)).**

**Последний способ кажется более предпочтительным в вычислитель-ном отношении, поскольку позволяет избегать капризного поведения в одной отдельно взятой компоненте вектора** x***(* k*)*, когда она может сде-латься очень малой по абсолютной величине или совсем занулиться, хотя в целом вектор** x***(*k*)* будет иметь значительную длину. Наконец, в качестве вектора, задающего направление проектирования во втором варианте, естественно взять сам** x***(*k*)*, вычисляя на каждом шаге отно-**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 3.8. Решение проблемы собственных значений | | | **245** |  |
| **шение** | |  |  |  |
|  | hx***(*k*+1)***, x***(*k*)***i | , | **(3.95)** |  |
|  | hx***(*k*)***, x***(*k*)***i |  |
|  |  |  |  |

**где** x***(*k*)*** =A**k**x***(0)*.**

**Для организации вычислительного алгоритма степенного метода требуется разрешить ещё два тонких момента, связанных с реализа-цией на ЭВМ.**

**Во-первых, это возможное неограниченное увеличение (при** λ***1*** >1**) или неограниченное уменьшение (при** λ***1*** <1**) норм векторов** x***(*k*)* и** x***(*k*+1)*, участвующих в нашем процессе. Разрядная сетка современных** **цифровых ЭВМ, как известно, конечна и позволяет представлять чис-ла из ограниченного диапазона. Чтобы избежать проблем, вызванных выходом за этот диапазон (¾переполнение¿ и ¾исчезновение порядка¿), имеет смысл нормировать** x***(*k*)*. При этом наиболее удобна нормировка в евклидовой норме** k · k***2*, так как тогда знаменатель отношения (3.95) сделается равным единице.**

**Во-вторых, при выводе степенного метода мы неявно предполагали, что начальный вектор** x***(0)* выбран так, что он имеет ненулевую проек-цию на направление доминирующего собственного вектора** v***1* матрицы** A**. В противном случае произведения любых степеней матрицы** A **на** x***(0)*** **будут также иметь нулевых проекцию, и никакой дифференци-ации длины компонент** A**k** x***(0)*, на которой собственно и основан сте-пенной метод, не произойдёт. Это затруднение может быть преодолено на основе какой-нибудь априорной информации о доминирующем соб-ственном векторе матрицы. Кроме того, при практической реализации степенного метода на цифровых ЭВМ неизбежные ошибки округления, как правило, приводят к появлению ненулевых компонент в направле-нии** v***1*, которые затем в процессе итерирования растянутся на нужную величину. Но, строго говоря, это может не происходить в некоторых исключительных случаях, и потому при ответственных вычислениях рекомендуется многократный запуск степенного метода с различными начальными векторами (так называемый мультистарт).**

˜

**В псевдокоде, представленном в Табл. 3.6,** λ **это приближённое доминирующее собственное значение матрицы** A**, а** x***(*k*)* текущее при-ближение к его нормированному собственному вектору.**

Теорема 3.8.2 **Пусть** n×n**-матрица** A **диагонализуема и у неё имеет-ся доминирующее собственное значение, которое является простым,**

**246** 3. Численные методы линейной алгебры

**Таблица 3.6. Степенной метод для нахождения доминирующего собственного значения матрицы**

k ← 0**;**

**выбираем вектор** x***(0)*** 6= 0**; нормируем** x***(0)*** ←x***(0)***/kx***(0)***k**;** DO WHILE **( метод не сошёлся )**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| y***(*k*+1)*** ←Ax***(*k*)* ;** | | | | | | | | |  |  |  | ***2*;** |  |
| x***(*k** |  |  |  | y |  |  |  | / y | |  |  |  |
| ˜ |  |  | ***(*k*+1)*** | | | , x | ***(*k*)*** | | **;** |  |  |  |  |
| λ ← y | | |  |  |  |  |  |  |  |  |  |
|  | ***+1)*** | ← | |  | ***(*k*+1)*** | | |  | k | ***(*k*+1)*** | k |  |  |
|  |  |  |  |  |  |  |  |  |  |

k ← k + 1**;**

END DO

**т. е. ему соответствует один линейный элементарный делитель. Ес-ли начальный вектор** x***(0)* не ортогонален доминирующему собствен-ному вектору матрицы** A**, то степенной метод сходится.**

Доказательство. **При сделанных нами предположениях о матрице** A

**она может быть представлена в виде**

A = V DV −***1***,

**где** D= diag{λ***1***, λ***2***, . . . , λ**n**} **диагональная матрица с собственными значениями** λ***1*,** λ***2*, . . . ,** λ**n по диагонали, а** V **матрица, осуществля-ющая преобразование подобия. Матрица** V **составлена из собственных векторов** v**i матрицы** A **как из столбцов:**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  | (v***1***)***2*** | | (v***2***)***2*** | . . . | (v**n**)***2*** | |  |  |
|  | · · · v**n** |  |  | (v***1***)***1*** | (v***2***)***1*** | . . . | (v**n**)***1*** |  |  |  |
| V = v***1*** v***2*** | = | **...** | **...** | **. . .** | **...** | , |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  | (v***1***)**n** | (v***2***)**n** | . . . | (v**n**)**n** |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |

**где через** (v**i**)**j обозначена** j**-ая компонента** i**-го собственного вектора**

|  |  |
| --- | --- |
| 3.8. Решение проблемы собственных значений | **247** |

**матрицы** A**. При этом можно считать, что** kv**i**k***2*** = 1**. Следовательно,**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  | **k**x***(0)*** |  |  |  | **k** раз | | |  | · · · |  |  |  |
| A**k** x***(0)*** | = | V DV −***1*** |  | = | V DV −***1*** |  | V DV −***1*** | | | |  | V DV −***1*** | x***(0)*** |  |
|  |  |  |  |  | **|** |  |  |  | **{z** |  |  |  |  | **}** |  |

* V D V −***1***V D V −***1***V · · · V −***1***V DV −***1***x***(0)***
* V D**k**V −***1***x***(0)*** = V D**k**z

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | λ***2*k** z***2*** | | |  | **k** |  | (λ***2*** | | /λ***1***)**k** | (z***2***/z***1***) | |  |  |
|  |  | λ***1*k** z***1*** |  |  |  |  |  |  | 1 |  |  |  |  |
| = V | **...** | = λ***1*** z***1*** | | V |  | **...** |  | , |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | **k** |  |  |  |  |  | (λ**n** | **k** |  |  |  |  |
|  |  | λ**n**z**n** |  |  |  |  |  | /λ***1***) (z**n**/z***1***) | |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |

**где обозначено** z=V−***1***x***(0)*. Необходимое условие проведения этой вы-кладки** z***1*** =60 **выполнено потому, что в условиях теоремы** x***(0)* имеет ненулевую компоненту в направлении** v***1*.**

**Коль скоро у матрицы** A **имеется доминирующее собственное зна-чение, т. е.**

|λ***1***| > |λ***2***| ≥ . . . ≥ |λ**n**|,

**то** λ***2***/λ***1*,** λ***3***/λ***1*, . . . ,** λ**n**/λ***1* все по модулю меньше единицы, и потому при** k→ ∞ **вектор**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| (λ***2*** | /λ***1***)**k**(z***2***/z***1***) |  |  |  |
|  | 1 |  |  |  |
| **...** | **(3.96)** |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

(λ**n**/λ***1***)**k**(z**n**/z***1***)

**сходится к вектору** (1,0,0, . . . ,0)>**. Соответственно, произведение**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | (λ***2*** | | /λ***1***)**k** | (z***2***/z***1***) |  |  |
|  |  |  | 1 |  |  |  |
| V |  | **...** |  |  |
|  |  |  |  |  |  |  |
|  |  | (λ**n** | **k** |  |  |  |
|  |  | /λ***1***) (z**n**/z***1***) | |  |  |
|  |  |  |  |  |  |  |

**сходится к первому столбцу матрицы** V **, т. е. к собственному векто-ру, отвечающему** λ***1*. Вектор** x***(*k*)*, который отличается от** A**k**x***(0)* лишь**

|  |  |  |  |
| --- | --- | --- | --- |
| **248** | 3. Численные методы линейной алгебры | |  |
| **нормировкой, сходится к собственному вектору** v***1*, а величина** | | ˜ |  |
| λ = |  |
| hy***(*k*+1)***, x***(*k*)***i **сходится к** hAv***1***, v***1***i = hλ***1***v***1***, v***1***i = λ***1*.** | |  |  |

**Из проведённых выше выкладок следует, что быстрота сходимости степенного метода определяется отношениями** |λ**i**/λ***1***|**,** i= 2,3, . . . , n**, знаменателями геометрических прогрессий, стоящих в качестве эле-ментов вектора (3.96). Фактически, решающее значение имеет наиболь-шее из этих отношений, т. е.** |λ***2***/λ***1*** |**, зависящее от того, насколько мо-дуль доминирующего собственного значения отделён от модуля осталь-ной части спектра. Чем больше эта отделённость, тем быстрее сходи-мость.**

Пример 3.8.6 **Для матрицы (3.3)**

|  |  |  |  |
| --- | --- | --- | --- |
|  | 3 | 4 | , |
|  | 1 | 2 |  |

**при вычислениях с двойной точностью степеной метод с начальным**

**вектором** x***(0)*** = (1,1)> **за 7 итераций даёт семь верных знаков доми-**

√

**нирующего собственного значения *12*** (5 + 33)≈5.3722813**. Детальная картина сходимости показана в следующей табличке:**

|  |  |
| --- | --- |
| **Номер** | **Приближение** |
| **итерации** | **к собственному значению** |
|  |  |
| **1** | **5.0** |

* **5.3448276**
* **5.3739445**
* **5.3721649**
* **5.3722894**
* **5.3722808**
* **5.3722814**

**Быстрая сходимость объясняется малостью величины** |λ***2***/λ***1***|**, ко-торая, как мы могли видеть в Примере 3.1.1, для рассматриваемой матрицы равна всего лишь** 0.069**.**

**Для матрицы (3.4)**

|  |  |  |
| --- | --- | --- |
| 13 4 |  | , |
| 2 |  |  |

−

|  |  |
| --- | --- |
| 3.8. Решение проблемы собственных значений | **249** |

**при тех же исходных условиях степенной метод порождает последо-**

˜

**вательность значений** λ**, которая случайно колеблется от примерно** 0.9 **до** 4 **и очевидным образом не имеет предела. Причина нали-чие у матрицы двух одинаковых по абсолютной величине комплексно-сопряжённых собственных значений** 2.5±1.936i **(см. Пример 3.1.1).**

**Наконец, необходимое замечание о сходимости степенного метода в комплексном случае. Так как комплексные числа описываются парами вещественных чисел, то комплексные одномерные инвариантные про-странства матрицы имеют вещественную размерность 2. Даже будучи нормированными, векторы из такого подпространства могут отличать-ся на скалярный множитель** e***i*ϕ, так что если не принять специальных мер, то в степенном методе видимой стабилизации координатных пред-ставлений комплексных собственных векторов может не наблюдаться. Тем не менее, о факте сходимости или расходимости можно при этом судить по стабилизации приближения к собственному значению. Либо кроме нормировки собственных векторов следует предусмотреть ещё приведение их к такой форме, в которой их координатные представле-ния будут определяться более ¾жёстко¿, например, требованием, чтобы первая компонента вектора была бы чисто вещественной.**

Пример 3.8.7 **Рассмотрим работу степенного метода в применении к** **матрице**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 3 | 4i |  | , |
|  | 1 | 2i |  |  |

**имеющей собственные значения**

λ***1*** = −0.4308405 − 0.1485958 i,

λ***2*** = 1.4308405 + 4.1485958 i.

**Доминирующим собственным значением здесь является** λ***2*.**

**Начав итерирование с вектора** x***(0)*** = (1,1)>**, уже через 7 итераций мы получим 6 правильных десятичных знаков в вещественной и мни-мой частях собственного значения. Но вот в порождаемых алгоритмом нормированных векторах** x***(*k*)***

x***(9)*** = −0.0113198 − 0.4322252 i , −0.1165906 − 0.8941252 i

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **250** |  | 3. Численные методы линейной алгебры | | |
| x***(10)*** | = | 0.4049144 | − 0.1516281 i | , |
|  |  | 0.8072493 | − 0.4017486 i |  |
|  |  | 0.2753642 | + 0.3333485 i |  |
| x***(11)*** | = 0.6429975 | | + 0.6321451 i , | |
|  |  | 0.2253495 | + 0.3690045 i |  |

x***(12)*** = −0.3879509 + 0.8139702 i ,

**и так далее нелегко ¾невооружённым глазом¿ узнать один и тот же собственный вектор, который ¾крутится¿ в одномерном комплексном инвариантном подпространстве. Но если поделить все получающиеся векторы на их первую компоненту, то получим один и тот же результат**

|  |  |  |  |
| --- | --- | --- | --- |
| 2.0742979 | 1.0.2154202 i | , |  |
|  | − |  |  |
| **и теперь уже налицо факт сходимости собственных векторов.** | | |  |

**Как ведёт себя степенной метод в случае, когда матрица** A **не явля-ется диагонализуемой? Полный анализ ситуации можно найти, напри-мер, в книгах [40, 42]. Наиболее неблагоприятен при этом случай, когда доминирующее собственное значение находится в жордановой клетке размера два и более. Теоретически степенной метод всё таки сходится к этому собственному значению, но уже медленнее любой геометриче-ской прогрессии.**

Пример 3.8.8 **Рассмотрим работу степенного метода в применении к** **матрице**

|  |  |  |  |
| --- | --- | --- | --- |
|  | 0 | 1 | , |
|  | 1 | 1 |  |

**т. е. к жордановой** 2×2**-клетке с собственным значением** 1**.**

**Запустив степенной метод из начального вектора** x***(0)*** = (1,1)>**, бу-дем иметь следующее**

***(*k*+1)***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 3.8. Решение проблемы собственных значений | | | | **251** |  |
|  | **Номер** |  | **Приближение** |  |  |
|  |  |  |  |
|  | **итерации** |  | **к собственному значению** |  |  |
|  |  |  |  |  |  |
| **1** | |  | **1.5** |  |  |
| **3** | |  | **1.3** |  |  |
| **10** | |  | **1.0990099** |  |  |
| **30** | |  | **1.0332963** |  |  |
| **100** | |  | **1.009999** |  |  |
| **300** | |  | **1.0033333** |  |  |
| **1000** | |  | **1.001** |  |  |
|  |  |  |  |  |  |

**То есть, для получения** n **верных десятичных знаков собственного зна-чения приходится делать примерно** 10**n итераций, что, конечно же, непомерно много. При увеличении размера жордановой клетки сходи-мость степенного метода делается ещё более медленной.**

**3.8е** **Обратные степенные итерации**

**Обратными степенными итерациями для матрицы** A **называют опи-санный в прошлом параграфе степенной метод, применённый к об-ратной матрице** A−***1*. Явное нахождение обратной матрицы при этом не требуется, так как в степенном методе используется лишь резуль-тат** y***(*k*+1)* её произведения на вектор** x***(*k*)* очередного приближения, а это, как мы знаем, эквивалентно решению системы линейных уравне-**

**ний** Ay = x***(*k*)*. Псевдокод получающегося метода представлен в**

**Табл. 3.7.**

**Практическая реализация решения системы линейных уравнений (5-я строка псевдокода) может быть сделана достаточно эффективной, если предварительно выполнить LU- или QR-разложение матрицы** A**, а затем на каждом шаге метода использовать формулы (3.31) или (3.39).**

**Так как собственные значения матриц** A **и** A− ***1* взаимно обратны, то обратные степенные итерации будут сходится к наименьшему по абсолютной величине собственному значению** A **и соответствующему собственному вектору.**

**3.8ж** **Сдвиги спектра**

**Сдвигом матрицы называют прибавление к ней скалярной матрицы, т. е. матрицы, пропорциональной единичной матрице, так что вместо матрицы** A **мы получаем матрицу** A+ϑI**. Если** λ**i**(A) **собственные**

**252** 3. Численные методы линейной алгебры

**Таблица 3.7. Обратные степенные итерации для нахождения**

**наименьшего по модулю собственного значения матрицы** A

k ← 0**;**

**выбираем вектор** x***(0)*** =60**; нормируем** x***(0)*** ←x***(0)***/kx***(0)***k**;**

DO WHILE **( метод не сошёлся )**

**найти** y***(*k*+1)* из системы** Ay***(*k*+1)*** =x***(*k*)*;**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| x***(*k** |  |  |  | y |  |  |  | / y | |  |  | ***2*;** |  |
| ˜ |  |  | ***(*k*+1)*** | | | , x | ***(*k*)*** | | **;** |  |  |  |  |
| λ ← y | | |  |  |  |  |  |  |  |  |  |
|  | ***+1)*** | ← | |  | ***(*k*+1)*** | | |  | k | ***(*k*+1)*** | k |  |  |
|  |  |  |  |  |  |  |  |  |  |

k ← k + 1**;**

END DO

**значения матрицы** A**, то для любого комплексного числа** ϑ **собствен-ными значениями матрицы** A+ϑI **являются числа** λ**i**(A) +ϑ**, тогда как собственные векторы остаются неизменными. Цель сдвига преобра-зование спектра матрицы для того, чтобы улучшить работу тех или иных алгоритмов решения проблемы собственных значений.**

**Если, к примеру, у матрицы** A **наибольшими по абсолютной вели-чине были два собственных значения** −2 **и** 2**, то прямое применение к ней степенного метода не приведёт к успеху. Но у матрицы** A+I **эти собственные значения перейдут в** −1 **и** 3**, второе собственное число станет наибольшим по модулю, и теперь уже единственным. Соответ-ственно, степенной метод сделается применимым к новой матрице.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Пример 3.8.9 **Для матрицы (3.4)** |  |  |  |  |
| −3 | 4 | , |  |
| 1 | 2 |  |  |  |

**как было отмечено в Примере 3.8д, простейший степенной метод расхо-дится из-за существования двух наибольших по абсолютной величине собственных значений.**

|  |  |
| --- | --- |
| 3.8. Решение проблемы собственных значений | **253** |

**Но если сдвинуть эту матрицу на** 2i**, то её спектр (см. Рис. 3.11) поднимется ¾вверх¿, абсолютные величины собственных значений пе-рестанут совпадать, и степенной метод окажется применимым.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Степенные итерации для ¾сдвинутой¿ матрицы** | | | |  |  |
| 1 +32i | 2 | 2i |  | **(3.97)** |  |
| 4 + |  |
| − |  |  |  |  |  |

**довольно быстро сходятся к наибольшему по модулю собственному зна-**

√

**чению *52*** + (2 + ***12*** 15) i≈2.5 + 3.9364917 i **. Детальная картина сходи-мости при вычислениях с двойной точностью и начальным вектором** x***(0)*** = (1, 1)> **показана в следующей табличке:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Номер** | **Приближение** | |  |
|  | **итерации** | **к собственному значению** | |  |
|  |  |  |  |  |
| **1** | | **2.0 + 2.0 i** | |  |
| **3** | | **2.0413223 + 4.3140496 i** | |  |
| **5** | | **2.7022202 + 3.9372711 i** | |  |
| **10** | | **2.5004558 + 3.945456 i** | |  |
| **20** | | **2.4999928 + 3.9364755 i** | |  |
| **В данном случае для матрицы (3.97) имеем** |λ***2***/λ***1***| ≈0.536**.** | | | |  |

**С помощью сдвигов матрицы можно любое её собственное значение, которое является крайней точкой выпуклой оболочки спектра, сделать наибольшим по модулю, обеспечив, таким образом, сходимость к нему итераций степенного метода. Но как добиться сходимости к другим соб-ственным значениям, которые лежат ¾внутри¿ спектра, а не ¾с краю¿? Здесь на помощь приходят обратные степенные итерации.**

**Другое важное следствие сдвигов изменение отношения** |λ***2***/λ***1***|**, величина которого влияет на скорость сходимости степенного метода, и отношения** λ**n**/λ**n**−***1*, которое определяет скорость сходимости обрат-ных степенных итераций. Подбирая нужным образом** ϑ**, часто можно добиться того, чтобы**

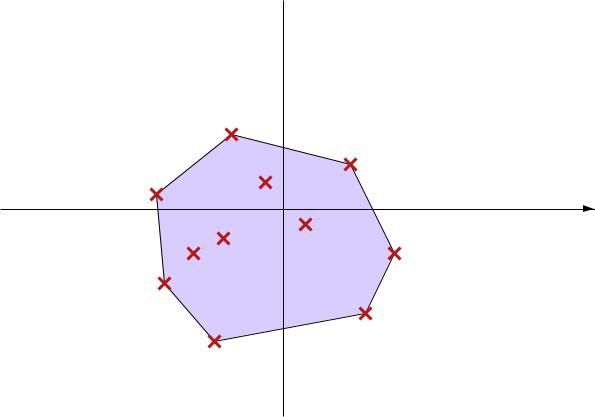
λ***2*** + ϑ λ***1*** + ϑ

**было меньшим, чем** |λ***2***/λ***1***|**, ускорив тем самым степенные итерации Обратные степенные итерации особенно эффективны в случае, ко-**

**гда имеется хорошее приближение к собственному значению и требу-ется найти соответствующий собственный вектор.**

**254** 3. Численные методы линейной алгебры

 Im



|  |  |
| --- | --- |
| 0 | Re |

**Рис. 3.12. С помощью сдвигов любую крайнюю точку спектра можно сделать наибольшей по модулю.**

**3.8з** **Метод Якоби для решения симметричной проблемы собственных значений**

**В этом параграфе мы рассмотрим численный метод для решения сим-метричной проблемы собственных значений, т. е. для вычисления соб-ственных чисел и собственных векторов симметричных матриц, кото-рый был впервые применён К.Г. Якоби в 1846 году к конкретной** 7×7**-матрице. Затем он был забыт на целое столетие, и вновь переоткрыт лишь после Второй мировой войны, когда началось бурное развитие вычислительной математики.**

**Идея метода Якоби состоит в том, чтобы подходящими преобразо-ваниями подобия от шага к шагу уменьшать величину**

**!*1*/*2***

**X**

a***2*ij**

**j**6***=*i**

**фробениусову норму внедиагональной части матрицы. Получающа-яся при этом последовательность матриц** A***(*k*)* будет стремиться к диа-гональной матрице с собственными значениями на главной диагонали. Инструментом реализации этого плана выступают элементарные орто-гональные матрицы вращений, рассмотренные в §3.3к. Почему именно ортогональные матрицы и почему вращений? Ответ на эти вопросы станет ясен позднее при анализе работы алгоритма.**

|  |  |
| --- | --- |
| 3.8. Решение проблемы собственных значений | **255** |

**Итак, положим** A***(0)*** :=A**. Если матрица** A***(*k*)*,** k= 0,1,2, . . . **, уже вычислена, то подберём матрицу** G(p, q, θ) **таким образом, чтобы сде-лать нулями пару внедиагональных элементов в позициях** (p, q) **и** (q, p) **в матрице** A***(*k*+1)*** :=G(p, q, θ)>A***(*k*)***G(p, q, θ)**. Желая сделать это, мы должны добиться выполнения равенства**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| a**qp*(*k*+1)*** | a**qq*(*k*+1)*** | **!** | sin θ |  | cos θ |  | a**qp*(*k*)*** |
| a**pp*(*k*+1)*** | a**pq*(*k*+1)*** | = | cos θ | − sin θ | | > | a**pp*(*k*)*** |
|  |  | = | 0 | **!** | , |  |  |
|  |  |  | × | 0 |  |  |  |

×

|  |  |  |  |
| --- | --- | --- | --- |
| a**pq*(*k*)*** | **!** sin θ | cos θ |  |
| a**qq*(*k*)*** |  |  |  |
|  | cos θ | − sin θ |  |

**где, как обычно, посредством ¾**×**¿ обозначены какие-то элементы, кон-кретное значение которых несущественно. Строго говоря, в результате рассматриваемого преобразования подобия в матрице** A***(*k*)* изменятся и другие элементы, находящиеся в строках и столбцах с номерами** p **и** q**. Этот эффект будет проанализирован ниже в Предложении 3.8.4.**

**Опуская индексы, обозначающие номер итерации и приняв сокра-щённые обзначения** c= cosθ**,** s= sinθ**, получим**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| × | **!** |  |  |  |  |  |  |  |
| 0 |  |  |  |  |  |  |  |
| 0 | = |  |  |  |  |  |  |  |
| × |  |  |  |  |  | **!** |  |
|  | sc(a**qq** − a**pp**) + a**pq** (c***2*** − s***2***) | | | a**pp**s***2*** + a**qq** c***2*** − 2sca**pq** | | |  |
|  | a**pp**c***2*** | + a**qq** s***2*** + 2sca**pq** | | sc(a**qq** − a**pp**) + a**pq** (c***2*** | | − s***2***) |  |  |
| **Приравнивание внедиагональных элементов нулю даёт** | | | | | |  |  |  |
|  |  | a**pp** − a**qq** | = | c***2*** − s***2*** | . |  |  |  |
|  |  | a**pq** | | sc | |  |  |  |

**Поделив обе части этой пропорции пополам, воспользуемся тригоно-метрическими формулами двойных углов**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| a**pp** − a**qq** | = | c***2*** − s***2*** | = | cos(2θ) | = | 1 | =: τ. |  |
| 2a**pq** | 2sc | sin(2θ) | tg(2θ) |  |
|  |  |  |  |  |

**Положим** t:= sinθ/cosθ= tgθ**. Вспоминая далее тригонометриче-скую формулу для тангенса двойного угла**

2 tg θ tg(2θ) =1−tg***2*** θ,

**256** 3. Численные методы линейной алгебры

**мы можем прийти к выводу, что** t **является корнем квадратного урав-нения**

t***2*** + 2τ t − 1 = 0

**с положительным дискриминантом** (4τ***2*** + 4)**, которое, следовательно, всегда имеет вещественные корни. Отсюда находится сначала** t**, а затем** c **и** s**, подробные формулы для которых мы не будем выписывать, чтобы** **не загромождать изложения.**

**Займёмся теперь обоснованием сходимости метода Якоби для реше-ния симметричной проблемы собственных значений.**

Предложение 3.8.3 **Фробениусова норма матрицы** A**, т. е.**

|  |  |  |  |
| --- | --- | --- | --- |
| kAk**F** = | **i,j*=1*** a**ij*2*** | **!** | , |
|  | **n** |  | ***1*/*2*** |
|  | **X** |  |  |

**не изменяется при умножениях на ортогональные матрицы слева или справа.**

Доказательство. **Напомним, что** **следом матрицы** A = (a**ij** )**, обозна-чаемым** trA**, называется сумма всех её диагональных элементов:**

**n**

**X**

tr A = a**ii**.

**i*=1***

**Привлечение понятия следа позволяет переписать определение фробе-ниусовой нормы матрицы таким образом**

kAk**F** = tr (A>A) ***1*/*2***.

**Следовательно, для любой ортогональной матрицы** Q **справедливо**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| kQAk**F** = | tr | (QA)>(QA) | | ***1*/*2*** |  |  |  |  |  |
| = |  | A>Q>QA | ***1*/*2*** | |  |  | ***1*/*2*** | = kAk**F** . |  |
| tr |  | = tr (A>A) | |  |  |
|  |  |  |  |  |  |  |  |  |

**Для доказательства аналогичного соотношения с умножением на орто-гональную матрицу справа заметим, что фробениусова норма не меня-ется при транспонировании матрицы. Следовательно,**

|  |  |  |  |
| --- | --- | --- | --- |
| kAQk**F** | = |  | Q>A> > **F** = kQ>A>k**F** = kA>k**F** = kAk**F** , |
|  |  |  |  |

|  |  |
| --- | --- |
| 3.8. Решение проблемы собственных значений | **257** |
| **что завершает доказательство Предложения.** |  |

Следствие. **Фробениусова норма матрицы не меняется при ортого-нальных преобразованиях подобия.**

**Для более точного описания меры близости матриц** A***(*k*)*, которые порождаются конструируемым нами методом, к диагональной матрице введём величину**

**!*1*/*2***

**X**

ND(A) = a***2*ij**

**j*=***6**i**

**фробениусову норму внедиагональной части матрицы. Ясно, что матрица** A **диагональна тогда и только тогда, когда** ND(A) = 0**.**

Предложение 3.8.4 **Пусть преобразование подобия матрицы** A **с по-мощью матрицы вращений** J **таково, что в матрице** B=J>AJ **за-нуляются элементы в позициях** (p, q) **и** (q, p)**. Тогда**

|  |  |
| --- | --- |
| ND***2***(B) = ND***2***(A) − 2a**pq*2***. | **(3.98)** |

**Итак, в сравнении с матрицей** A **в матрице** B **изменились элементы строк и столбцов с номерами** p **и** q**, но фробениусова норма недиаго-нальной части изменилась при этом так, как будто кроме зануления элементов** a**pq и** a**qp ничего не произошло.**

Доказательство. **Для** 2 × 2**-подматрицы**

**!**

a**pp** a**pq**

a**qp** a**qq**

**из матрицы** A **и соответствующей ей** 2×2**-подматрицы**

**!**

b**pp** 0

* + b**qq**
* **матрице** B **справедливо соотношение**

a***2*pp** + a***2*qq** + 2a***2*pq** = b***2*pp** + b***2*qq** ,

**258** 3. Численные методы линейной алгебры

**так как ортогональным преобразованием подобия фробениусову норма матрицы не изменяется. Но, кроме того,** kAk***2*F** =kBk***2*F , и потому**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **n** |  |  |  |  |
| **ND*2***(B) =kBk**F*2*** − | **X** |  |  |  |  |
| b**ii*2*** |  |  |  |  |
| = kAk**F*2*** − | **i*=1*** | − a**pp*2*** + a**qq*2*** | + b**pp*2*** + b**qq*2*** | **!** |  |
| a**ii*2*** |  |
|  | **n** |  |  |  |  |
|  | **X** |  |
|  | **i*=1*** |  |  |  |  |

|  |  |
| --- | --- |
|  | = ND***2***(A) − 2a**pq*2***, |
| **поскольку на диагонали у матрицы** A **изменились только два элемента** | |
| a**pp** **и** a**qq** **.** |  |

**Таблица 3.8. Метод Якоби для вычисления собственных**

**значений симметричной матрицы**

Вход

**Симметричная матрица** A**.**

**Допуск** **на норму внедиагональных элементов.**

Выход

**Матрица, на диагонали которой стоят приближёния, к собственным значениям** A**.**

Алгоритм

DO WHILE ND(A) > )

**выбрать ненулевой внедиагональный элемент** a**pq в** A **;**

**обнулить** a**pq и** a**qp преобразованием подобия с матрицей вращения** G(p, q, θ) **;**

END DO

**Теперь можно ответить на вопрос о том, почему в методе Якоби для преобразований подобия применяются именно ортогональные мат-**

|  |  |
| --- | --- |
| 3.8. Решение проблемы собственных значений | **259** |

**рицы. Как следует из результатов Предложений 3.8.3 и 3.8.4, ортого-нальные матрицы обладают замечательным свойством сохранения об-щей нормы всех элементов матрицы и, как следствие, ¾перекачивания¿ величины с внедиагональных элементов на диагональ. При других пре-образованиях подобия добиться этого было бы едва ли возможно.**

**Итак, всё готово для организации итерационного процесса приведе-ния симметричной матрицы к диагональному виду, при котором вне-диагональные элементы последовательно подавляются. Как уже от-мечалось, занулённые на каком-то шаге алгоритма элементы могут впоследствии вновь сделаться ненулевыми. Но результат Предложе-ния 3.8.4 показывает, что норма внедиагональной части матрицы при этом всё равно монотонно уменьшается.**

**Различные способы выбора ненулевых внедиагональных элементов, подлежащих обнулению, приводят к различным практическим версиям метода Якоби.**

**Выбор наибольшего по модулю внедиагонального элемента наи-лучшее для отдельно взятого шага алгоритма решение. Но поиск этого элемента имеет трудоёмкость** n(n−1)/2**, что может оказаться весьма дорогостоящим, особенно для матриц больших размеров. Преобразова-ние подобия с матрицей вращений обходится всего в** O(n) **операций!**

**Чаще применяют циклический обход столбцов (или строк) матри-цы, и наибольший по модулю элемент берут в пределах рассматривае-мого столбца (строки).**

**Наконец, ещё одна популярная версия это так называемый ¾ба-рьерный метод Якоби¿, в котором назначают величину ¾барьера¿ на значение модуля внедиагональных элементов матрицы, и алгоритм об-нуляет все элементы, модуль которых превосходит этот барьер. Затем барьер понижается, процесс обнуления повторяется заново, и так до тех пор, пока не будет достигнута требуемая точность.**

**3.8и** **Базовый QR-алгоритм**

**QR-алгоритм, изложению которого посвящён этот параграф, являет-ся одним из наиболее эффективных численных методов для решения полной проблемы собственных значений. Он был изобретён независимо В.Н. Кублановской (1960 год) и Дж. Фрэнсисом (1961 год). Публика-ция В.Н. Кублановской появилась раньше**13**, а Дж. Фрэнсис более пол-**

13 **Упоминая о вкладе В.Н. Кублановской в изобретение QR-алгоритма, обычно ссылаются на её статью 1961 года в ¾Журнале вычислительной математики и мате-**

**260** 3. Численные методы линейной алгебры

**но развил практичную версию QR-алгоритма.**

**Вспомним теорему о QR-разложении (Теорема 3.3.3, стр. 171): вся-кая квадратня матрица представима в виде произведения ортогональ-ной и правой (верхней) треугольной матриц. Ранее в нашем курсе мы уже обсуждали конструктивные способы выполнения этого разложе-ния с помощью матриц отражения Хаусхолдера, а также с помо-щью матриц вращений. Следовательно, далее можно считать, что QR-разложение выполнимо и основывать свои построения на этом факте.**

**Вычислительная схема базового QR-алгоритма для решения про-блемы собственных значений представлена в Табл. 3.9: мы разлагаем матрицу** A***(*k*)*, полученную на** k**-м шаге алгоритма,** k= 0,1,2, . . . **, на ортогональный** Q***(*k*)* и правый треугольный** R***(* k*)* сомножители и далее, поменяв их местами, умножаем друг на друга, образуя следующее при-ближение** A***(*k*+1)*.**

**Таблица 3.9. QR-алгоритм для нахождения**

**собственных значений матрицы** A

k ← 0**;**

A***(0)*** ← A**;**

DO WHILE **( метод не сошёлся )**

**вычислить QR-разложение** A***(*k*)*** =Q***(*k*)***R***(*k*)*;**

A***(*k*+1)*** ←R***(*k*)***Q***(*k*)*;**

k ← k + 1**;**

END DO

**Прежде всего отметим, что поскольку**

A***(*k*+1)*** =R***(*k*)***Q***(*k*)*** = Q***(*k*)*** > Q***(*k*)***R***(*k*)*** Q***(*k*)*** = Q***(*k*)*** >A***(*k*)***Q***(*k*)***,

**то все матрицы** A***(*k*)*,** k= 0,1,2, . . . **, ортогонально подобны друг другу и исходной матрице** A**. Результат о сходимости QR-алгоритма нефор-мальным образом может быть резюмирован в следующем виде: если** A

**матической физики¿ [62]. Но первое сообщение о QR-алгоритме было опубликовано ею раньше в Дополнении к изданию 1960 года книги [42].**

|  |  |
| --- | --- |
| 3.8. Решение проблемы собственных значений | **261** |

**неособенная вещественная матрица, то последовательность порож-даемых QR-алгоритмом матриц** A***(*k*)* сходится ¾по форме¿ к верхней блочно-треугольной матрице.**

**Это означает, что предельная матрица, к которой сходится QR-алгоритм, является верхней треугольной либо верхней блочно-треу-гольной, причём размеры диагональных блоков зависят, во-первых, от типа собственных значений матрицы (кратности и принадлежности ве-щественной оси** R**), и, во-вторых, от того, в вещественной или комплекс-ной арифметике выполняется QR-алгоритм.**

**Если алгоритм выполняется в вещественной (комплексной) арифме-тике и все собственные значения матрицы вещественны (комплексны) и различны по модулю, то предельная матрица верхняя треуголь-ная. Если алгоритм выполняется в вещественной (комплексной) ариф-метике и некоторое собственное значение матрицы вещественно (ком-плексно) и имеет кратность** p**, то в предельной матрице ему соответ-ствует диагональный блок размера** p×p**. Если алгоритм выполняется для вещественной матрицы в вещественной арифметике, то простым комплексно-сопряжённым собственным значениям (они имеют равные модули) отвечают диагональные** 2×2**-блоки в предельной матрице. На-конец, если некоторое комплексное собственное значение вещественной матрицы имеет кратность** p**, так что ему соответствует ещё такое же комплексно-сопряжённое собственое значение кратности** p**, то при вы-полнении QR-алгоритма в вещественной арифметике предельная мат-рица получит диагональный блок размера** 2p×2p**.**

Пример 3.8.10 **Проиллюстрируем работу QR-алгоритма на примере** **матрицы**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1 | −2 | − | 3 | , | **(3.99)** |  |
| −7 | | 8 | 9 |  |
|  | 4 | 5 |  | 6 |  |  |  |
|  |  |  |  |  |  |

**имеющей собственные значения**

2.7584148

6.1207926 ± 8.0478897 i

**Читатель может провести на компьютере этот увлекательный экспе-римент самостоятельно, воспользовавшись системами Scilab,** Matlab

**262** 3. Численные методы линейной алгебры

**или им подобными: все они имеют встроенную процедуру для QR-разложения матриц.**14

Пример 3.8.11 **Для ортогональной матрицы**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 1 | 0 | , | **(3.100)** |
|  | 0 | 1 |  |  |

**QR-разложением является произведение её самой на единичную мат-рицу. Поэтому в результате одного шага QR-алгоритма мы снова по-лучим исходную матрицу, которая, следовательно, и будет пределом итераций. В то же время, матрица (3.100) имеет собственные значе-ния, равные** ±1**, так что в данном случае QR-алгоритм не работает.**

**3.8к** **Модификации QR-алгоритма**

**Представленная в Табл. 3.9 версия QR-алгоритма на практике обычно снабжается рядом модификаций, которые существенно повышают её эффективность. Главными из этих модификаций являются**

1. **сдвиги матрицы, рассмотренные нами в §3.8ж, и**
2. **предварительное приведение матрицы к специальной верхней почти треугольной форме.**

**Можно показать (см. теорию в книгах [14, 39]), что, аналогично сте-пенному методу, сдвиги также помогают ускорению QR-алгоритма. Но в QR-алгоритме их традиционно организуют способом, представлен-ным в Табл. 3.10.**

**Особенность организации сдвигов в этом псевдокоде присутствие обратных сдвигов (в строке 6 алгоритма) сразу же вслед за прямыми (в 5-й строке). Из-за этого в получающемся алгоритме последовательно вычисляемые матрицы** A***(*k*)* и** A***(*k*+1)* ортогонально подобны, совершен-но так же, как и в исходной версии QR-алгоритма:**

A***(*k*+1)*** =R***(*k*)***Q***(*k*)*** +ϑ**k**I=Q***(*k*)*** >Q***(*k*)***R***(*k*)***Q***(*k*)*** +ϑ**k**Q***(*k*)*** >Q***(*k*)*** =Q***(*k*)*** >Q***(*k*)***R***(*k*)*** +ϑ**k**I Q***(*k*)*** =Q***(*k*)*** >A***(*k*)***Q***(*k*)***.

14 **В Scilab’е и Matlab’е она так и называется** qr**.**

|  |  |
| --- | --- |
| 3.8. Решение проблемы собственных значений | **263** |

**Таблица 3.10. QR-алгоритм со сдвигами для нахождения**

**собственных значений матрицы** A

k ← 0**;**

A***(0)*** ← A**;**

DO WHILE **( метод не сошёлся )**

**выбрать сдвиг** ϑ**k вблизи собственного значения** A **; вычислить QR-разложение** A***(*k*)*** −ϑ**k**I=Q***(*k*)***R***(*k*)*;**

A***(*k*+1)*** ←R***(*k*)***Q***(*k*)*** +ϑ**k**I**;**

k ← k + 1**;**

END DO

**То есть, представленная организация сдвигов позволила сделать их в одно и то же время локальными и динамическими по характеру.**

Пример 3.8.12 **Проиллюстрируем работу QR-алгоритма со сдвигами** **на знакомой нам матрице (3.99)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 1 | −2 | 3 |  |  |
|  |  |  |
| −7 | | 8 | −9 |  |
|  | 4 | 5 | 6 |  |  |
|  |  |  |  |

**из предыдущего примера**

Определение 3.8.1 **Матрица** H = (h**ij** ) **называется** верхней почти треугольной **или** хессенберговой матрицей (в форме Хессенберга)**, если** h**ij** = 0 **при** i > j + 1**.**

**264** 3. Численные методы линейной алгебры

**Наглядный ¾портрет¿ хессенберговой матрицы выглядит следую-**

**щим образом:**

× × × ×

H = ×

0

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| · · · | | × | × | |  |
| · · · | | × | × | . |  |
| **.. . ...** | | | **...** |  |  |
| **.** | **. .** |  |  |  |  |
|  |  | × |  |  |  |
|  |  |  | × | |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  | × | × |  |  |
|  |  |  |  |  |  |

Предложение 3.8.5 **Любую** n × n**-матрицу** A **можно привести к ор-тогонально подобной хессенберговой матрице** H=QAQ>**, где** Q **произведение конечного числа отражений или вращений.**

Доказательство. **Рассмотрим для определённости преобразования с** **помощью матриц отражения.**

**Возьмём матрицу отражения** Q***1*** =I−2uu> **так, чтобы первая ком-понента вектора Хаусхолдера** u **была нулевой и при этом**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | a***11*** | |  |  |  | a***11*** |  |  |  |
|  |  | a***21*** | |  |  |  | a***21***0 |  |  |  |
| Q***1*** | a***31*** | | = | 0 | , |  |
|  |  |  | **.** |  |  |  | **.** |  |  |  |
|  |  |  | **..** |  |  |  | **..** |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  | a | **n*1*** |  |  |  | 0 |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |

**т. е. занулялись бы элементы** a***31*, . . . ,** a**n*1* в первом столбце. Нетрудно видеть, что** Q***1* выглядит следующим образом**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  | 0 | × | · · · | | × | × | | |  |
|  |  |  | 1 | 0 | · · · | | 0 | 0 |  |  |  |
| Q***1*** | = |  | 0 | × **. . . ...** | | | | **...** | . | |  |
|  |  | **.** | | **. .** | | **..** |  |  |  | |  |
|  |  | **..** | | **..** | | × | × |  | |  |
|  |  |  |  |  |  |  |  | |  |
|  |  |  |  |  |  |  |  |  |  | |  |
|  |  |  |  |  |  |  |  |  |  | |  |
|  |  |  | 0 | × | · · · | | × | × |  | |  |
|  |  |  |  |  |  |

**Далее, когда** A **умножается на такую матрицу** Q***1* слева, то в ней не изменяются элементы первой строки. Когда матрица** Q***1***A **умножается на** Q>***1*** =Q***1* справа, то в ней не изменяются элементы первого столбца. Поэтому в матрице** Q***1***AQ>***1*, как и в** Q***1***A**, первый столбец имеет нули в позициях с 3-й по** n**-ую.**

|  |  |
| --- | --- |
| 3.8. Решение проблемы собственных значений | **265** |

**Далее выбираем матрицы отражения** Q***2*,** Q***3*, . . . ,** Q**n**−***2* так, чтобы умножение слева на** Q**i давало нули в позициях с** (i+ 2)**-ой по** n**-ую в** i**-ом столбце. При этом последующее умножения справа на** Q**i** **также не** **портит возникающую почти треугольную структуру результирующей матрицы. Получающаяся в итоге матрица** QAQ> **с** Q=Q**n**−***2*** . . . Q***1* действительно является верхней почти треугольной.**

Предложение 3.8.6 **Хессенбергова форма матрицы сохраняется при** **выполнении с ней QR-алгоритма.**

Доказательство. **При QR-разложении хессенберговой матрицы в ка-честве ортогонального сомножителя** Q **для матрицы** A***(*k*)*** −ϑI **полу-чается также хессенбергова матрица, так как** j**-ый столбец в** Q **есть линейная комбинация первых** j **столбцов матрицы** A***(*k*)*** −ϑI**. В свою очередь, матрица** RQ **произведение после перестановки сомножи-телей также получается хессенберговой. Добавление диагонального слагаемого** ϑI **не изменяет верхней почти треугольной формы матри-цы.**

**Смысл предварительного приведения к хессенберговой форме за-ключается в следующем. Хотя это приведение матрицы требует** O(n***3***)

**операций, дальнейшее выполнение одной итерации QR-алгоритма с хес-сенберговой формой будет теперь стоить всего** O(n***2***) **операций, так что общая трудоёмкость алгоритма составит** O(n***3***)**.**

Литература к Главе 3

Основная

1. **Бахвалов Н.С., Жидков Н.П., Кобельков Г.М. Численные методы. –**

**Москва: Бином, 2003, а также другие издания этой книги.**

1. **Бахвалов Н.С., Корнев А.А., Чижонков Е.В. Численные методы. Реше-ния задач и упражнения. – Москва: Дрофа, 2008.**
2. **Березин И.С., Жидков Н.П. Методы вычислений. Т. 1–2. – Москва: Наука, 1966.**
3. **Вержбицкий В.М. Численные методы. Части 1–2. – Москва: ¾Оникс 21 век¿, 2005.**
4. **Воеводин В.В. Вычислительные основы линейной алгебры. – Москва: Наука, 1977.**

**266** 3. Численные методы линейной алгебры

1. **Воеводин В.В. Линейная алгебра. – Москва: Наука, 1980.**
2. **Воеводин В.В., Кузнецов Ю.А. Матрицы и вычисления. – Москва: Наука, 1984.**
3. **Волков Е.А. Численные методы. – Москва: Наука, 1987.**
4. **Гантмахер Ф.Р. Теория матриц. – Москва: Наука, 1988.**
5. **Глазман И.М., Любич Ю.И. Конечномерный линейный анализ. – Москва: Наука, 1969.**
6. **Годунов С.К. Современные аспекты линейной алгебры. – Новосибирск: На-учная книга, 1997.**
7. **Голуб Дж., ван Лоун Ч. Матричные вычисления. – Москва: Мир, 1999.**
8. **Демидович Б.П., Марон А.А. Основы вычислительной математики. –**

**Москва: Наука, 1970.**

1. **Деммель Дж. Вычислительная линейная алгебра. – Москва: Мир, 2001.**
2. **Икрамов Х.Д. Численные методы линейной алгебры. – Москва: Знание, 1987.**
3. **Ильин В.П., Кузнецов Ю.И. Трёхдиагональные матрицы и их приложения.**

**– Москва: Наука, 1985.**

1. **Канторович Л.В., Акилов Г.П. Функциональный анализ. – Москва: Наука, 1984.**
2. **Като Т. Теория возмущений линейных операторов. – Москва: Мир, 1972.**
3. **Коллатц Л. Функциональный анализ и вычислительная математика. –**

**Москва: Мир, 1969.**

1. **Коновалов А.Н. Введение в вычислительные методы линейной алгебры. –**

**Новосибирск: Наука, 1993.**

1. **Кострикин А.Н. Введение в алгебру. Часть 2. Линейная алгебра. – Москва: Физматлит, 2001.**
2. **Красносельский М.А., Крейн С.Г. Итеративный процесс с минимальными невязками // Математический Сборник. – 1952. – Т. 31 (73), №2. – С. 315–334.**
3. **Крылов В.И., Бобков В.В., Монастырный П.И. Вычислительные методы. Т. 1–2. – Москва: Наука, 1976.**
4. **Ланкастер П. Теория матриц. – Москва: Наука, 1978.**
5. **Лоусон Ч., Хенсон Р. Численное решение задач методом наименьших квад-ратов. – Москва: Наука, 1986.**
6. **Марчук Г.И., Кузнецов Ю.А. Итерационные методы и квадратичные функ-ционалы. – Новосибирск: Наука, 1972.**
7. **Матрицы и квадратичные формы. Основные понятия. Терминология / Акаде-мия Наук СССР. Комитет научно-технической терминологии. – Москва: Нау-ка, 1990. – (Сборники научно-нормативной терминологии; Вып. 112).**
8. **Мацокин А.М. Численный анализ. Вычислительные методы линейной алгеб-ры. Конспекты лекций для преподавания в III семестре ММФ НГУ. Ново-сибирск: НГУ, 2009–2010.**

3.8. Решение проблемы собственных значений **267**

1. **Миньков С.Л., Миньков Л.Л. Основы численных методов. – Томск: Изда-тельство научно-технической литературы, 2005.**
2. **Мысовских И.П. Лекции по методам вычислений. – Санкт-Петербург: Изда-тельство Санкт-Петербургского университета, 1998.**
3. **Островский А.М. Решение уравнений и систем уравнений.**

**– Москва: Издательство иностранной литературы, 1963.**

1. **Пароди М. Локализация характеристических чисел матриц и её применения.**

**– Москва: Издательство иностранной литературы, 1960.**

1. **Прасолов В.В. Задачи и теоремы линейной алгебры. – Москва: Наука-Физматлит, 1996.**
2. **Райс Дж. Матричные вычисления и математическое обеспечение. – Москва: Мир, 1984.**
3. **Самарский А.А., Гулин А.В. Численные методы. – Москва: Наука, 1989.**
4. **Стренг Г. Линейная алгебра и её применения. – Москва: Мир, 1980.**
5. **Тихонов А.Н., Арсенин В.Я. Методы решения некорректных задач. –**

**Москва: Наука, 1974.**

1. **Тыртышников Е.Е. Матричный анализ и линейная алгебра. – Москва: Физ-матлит, 2007.**
2. **Тыртышников Е.Е. Методы численного анализа. – Москва: Академия, 2007.**
3. **Уилкинсон Дж. Алгебраическая проблема собственных значений. – Москва: Наука, 1970.**
4. **Уоткинс Д. Основы матричных вычислений. – Москва: ¾Бином. Лаборатория знаний¿, 2009.**
5. **Фаддеев Д.К., Фаддеева В.Н. Вычислительные методы линейной алгебры.**

**– Москва–Ленинград: Физматлит, 1960 (первое издание) и 1963 (второе изда-ние).**

1. **Федоренко Р.П. Итерационные методы решения разностных эллиптических уравнений // Успехи Математических Наук. – 1973. – Т. 28, вып. 2 (170). – С. 121–182.**
2. **Форсайт Дж.Э. Что представляют собой релаксационные методы? // Совре-менная математика для инженеров под ред Э.Ф.Беккенбаха. – Москва: Изда-тельство иностранной литературы, 1958. – С. 418–440.**
3. **Форсайт Дж., Молер К. Численное решение систем линейных алгебраиче-ских уравнений. – Москва: Мир, 1969.**
4. **Хаусдорф Ф. Теория множеств. – Москва: УРСС Эдиториал, 2007.**
5. **Хейгеман Л., Янг Д. Прикладные итерационные методы. – Москва: Мир, 1986.**
6. **Хорн Р., Джонсон Ч. Матричный анализ. – Москва: Мир, 1989.**
7. **Aberth O. Precise numerical methods using C++. – San Diego: Academic Press, 1998.**

**268** 3. Численные методы линейной алгебры

1. **Beckermann B. The condition number of real Vandermonde, Krylov and positive definite Hankel matrices // Numerische Mathematik. – 2000. – Vol. 85, No. 4. – P. 553–577.**
2. **Kelley C.T. Iterative methods for linear and nonlinear equations. – Philadelphia: SIAM, 1995.**
3. **Saad Y. Iterative methods for sparse linear systems. – 2000.**
4. **Scilab The Free Platform for Numerical Computation.** http://www.scilab.org
5. **Temple G. The general theory of relaxation methods applied to linear systems //**

**Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences. – 1939. – Vol. 169, No. 939. – P. 476–500.**

1. **Trefethen L.N., Bau D. III Numerical linear algebra. – Philadelphia: SIAM, 1997.**

Дополнительная

1. **Алексеев Е.Р., Чеснокова О.В., Рудченко Е.А. Scilab. Решение инженер-ных и математических задач. – Москва: Alt Linux – Бином, 2008.**
2. **Алефельд Г., Херцбергер Ю. Введение в интервальные вычисления. –**

**Москва: Мир, 1987.**

1. **Джордж А., Лю Дж. Численное решение больших разреженных систем уравнений. – Москва: Мир, 1984.**
2. **Дробышевич В.И., Дымников В.П., Ривин Г.С. Задачи по вычислитель-ной математике. – Москва: Наука, 1980.**
3. **Калиткин Н.Н. Численные методы. – Москва: Наука, 1978.**
4. **Крылов А.Н. Лекции о приближённых вычислениях. – Москва: ГИТТЛ, 1954, а также более ранние издания.**
5. **Кублановская В.Н. О некоторых алгорифмах для решения полной пробле-мы собственных значений // Журнал вычисл. матем. и мат. физики. – 1961. – Т. 1, №4. – С. 555–570.**
6. **Марчук Г.И. Методы вычислительной математики. – Москва: Наука, 1989.**
7. **Шарый С.П. Конечномерный интервальный анализ. – Электронная книга, 2010 (см.** http://www.nsc.ru/interval/Library/InteBooks**)**
8. **Gregory R.T., Karney D.L. A collection of matrices for testing computational algorithms. – Hantington, New York: Robert E. Krieger Publishing Company, 1978.**
9. **Kreinovich V., Lakeyev A.V, Noskov S.I. Approximate linear algebra is intractable // Linear Algebra and its Applications. – 1996. – Vol. 232. – P. 45– 54.**
10. **Kreinovich V., Lakeyev A.V., Rohn J., Kahl P. Computational complexity and feasibility of data processing and interval computations. – Dordrecht: Kluwer, 1997.**
11. **Moler C. Professor SVD // The MathWorks News & Notes. – October 2006. – P. 26–29.**

3.8. Решение проблемы собственных значений **269**

1. **Wilf H.S. Finite sections of some classical inequalities. – Heidelberg: Springer, 1970.**
2. **Todd J. The condition number of the finite segment of the Hilbert matrix //**

**National Bureau of Standards, Applied Mathematics Series. – 1954. – Vol. 39. – P. 109–116.**

Глава 4

Решение нелинейных уравнений и их систем

**Предметом рассмотрения этой главы книги является задача решения системы уравнений**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | F***2*** | | | ( x***1*** | , x***2*** | | , . . . , x**n**) = 0, | | | |  |
|  |  | F***1*** | | ( x***1*** | , x***2*** | | , . . . , x**n**) = 0, | | | |  |
|  |  |  |  | **.** |  |  | **.** |  |  | **.** |  |
|  |  |  |  | **.** |  |  | **.** |  | **.** |  |
|  |  |  |  |  |  |  | **.** |  |
|  |  |  |  | **.** |  |  |  |  | **.** |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | F**n**( x***1***, x***2***, . . . , x**n**) = 0, | | | | | | | | |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
| **над полем** | **вещественных** | | **чисел** | | | R**, или, кратко,** | | | | |  |
|  |  |  |  |  |  |

F (x) = 0,

**(4.1)**

**(4.2)**

**где** x= (x***1***, x***2***, . . . , x**n**)∈R**n** **вектор неизвестных переменных,**

F**i**(x)**,** i = 1, 2, . . . , n **некоторые вещественнозначные функции,** F (x) = ( F***1***(x), F***2***(x), . . . , F**n**(x) )> **вектор-столбец функций** F**i.**

**Нужно найти набор значений** x˜***1*,** x˜***2*, . . . ,** x˜**n переменных** x***1*,** x***2*, . . . ,** x**n, который и будет называться решением, обращающий в равенства все уравнения системы (4.1)–(4.2). В некоторых случаях желательно найти все решения системы, иногда достаточно одного. В случае отсутствия решений у системы (4.1)–(4.2) нередко требуется предоставить обос-нованный вывод этого факта в виде протокола работы программы и т. п.**

**270**

**271**

**Наряду с задачами, рассмотренными в Главе 2, интерполяцией и приближениями функций, вычислением интегралов, задача решения уравнений и систем уравнений является одной из классических задач численного анализа.**

**Всюду далее мы предполагаем, что функции** F**i**(x) **по меньшей мере непрерывны, а количество уравнений в системе (4.1)–(4.2) совпадает с количеством неизвестных переменных. Помимо записи систем уравне-ний в каноническом виде (4.1)–(4.2) часто встречаются и другие формы их представления, например,**

|  |  |
| --- | --- |
| G(x) = H(x) | **(4.3)** |

**с какими-то функциями** G**,** H**. Чрезвычайно важным частным случа-ем этой формы является известный нам рекуррентный вид системы уравнений (уравнения),**

|  |  |
| --- | --- |
| x = G(x). | **(4.4)** |

**в котором неизвестная переменная выражена через саму себя. В этом случае решение системы уравнений (или уравнения) есть неподвиж-ная точка отображения** G**, т. е. такой элемент области определения** G**, который переводится этим отображением сам в себя.**

**Как правило, системы уравнений различного вида могут быть при-ведены друг к другу равносильными преобразованиями. В частности, несложно установить связь решений уравнений и систем уравнений ви-да (4.1)–(4.2) с неподвижными точками отображений, т. е. с решениями уравнений в рекуррентом виде (4.4). Ясно, что**

F (x) = 0 ⇐⇒ x = x − ΛF (x),

**где** Λ **ненулевой скаляр в одномерном случае или же неособенная матрица в случае вектор-функции** F **. Поэтому решение уравнения**

F (x) = 0

**является неподвижной точкой отображения**

G(x) = x − ΛF (x).

**Обращаясь к решению нелинейных уравнений и их систем, мы обна-руживаем себя в гораздо более сложных условиях, нежели при решении систем линейных алгебраических уравнений. Стройная и весьма пол-ная теория разрешимости систем линейных уравнений, базирующаяся**

**272** 4. Решение нелинейных уравнений и их систем

**на классических результатах линейной алгебры, обеспечивала в слу-чае СЛАУ уверенность в существовании решения и его единственности. Для нелинейных уравнений столь общей и простой теории не существу-ет. Напротив, будучи объединёнными лишь общим признаком ¾отри-цания линейности¿, нелинейные уравнения отличаются огромным раз-нообразием**

1. Вычислительно-корректные задачи

**4.1а** **Формальные определения**

**Напомним общеизвестный факт: на вычислительных машинах (как электронных, так и механических, как цифровых, так и аналоговых) в условиях приближённого представления входных числовых данных и приближённого характера вычислений над полем вещественных чи-сел** R **мы в принципе можем решать лишь те постановки задач, ответы которых непрерывно зависят от входных данных, т. е. устойчивы по от-ношению к возмущениям в этих начальных данных. Дело в том, что ре-шение задачи на любой вычислительной машине сопровождается неиз-бежными ошибками и погрешностями, вызванными конечным харак-тером представления чисел, конечностью исполнительных устройств и т.п. Потенциально эти погрешности могут быть сделаны сколь угодно малыми, но в принципе избавиться от них не представляется возмож-ным. Получается, что реально**

* **мы решаем на вычислительной машине не исходную математиче-скую задачу, а более или менее близкую к ней,**
* **сам процесс решения на ЭВМ отличается от своего идеального математического прообраза.**

**Возникновение и бурное развитие компьютерной алгебры с её ¾без-ошибочными¿ вычислениями едва ли опровергает высказанный выше тезис, так как исходные постановки задач для систем символьных пре-образований требуют точную представимость входных данных, кото-рые поэтому подразумеваются целыми или, на худой конец, рациональ-ными с произвольной длиной числителя и знаменателя (см. [2]), а все преобразования над ними не выводят за пределы поля рациональных чисел.**

|  |  |
| --- | --- |
| 4.1. Вычислительно-корректные задачи | **273** |

**Под массовой задачей [13] будем понимать некоторый общий вопрос, формулировка которого содержит несколько свободных переменных параметров могущих принимать значения в пределах предписанных им множеств. В целом массовая задача** Π **определяется**

1. **общим списком всех её параметров с областями их определения,**
2. **формулировкой тех свойств, которым должен удовлетворять ответ, т. е. решение задачи.**

**Индивидуальная задача** I **получается из массовой задачи** Π **путём при-сваивания всем параметрам задачи** Π **каких-то конкретных значений. Наконец, разрешающим отображением задачи** Π **назовём отображе-ние, сопоставляющее каждому набору входных данных-параметров от-вет соответствующей индивидуальной задачи.**

Определение 4.1.1 **. Станем говорить, что массовая математи-ческая задача является** вычислительно корректной**, если её разреша-ющее отображение** P → A **из множества входных данных** P **во мно-жество** A **ответов задачи непрерывно относительно некоторых то-пологий на** P **и** A**, определяемых содержательным смыслом задачи.**

**Те задачи, ответы на которые неустойчивы по отношению к воз-мущениям входных данных, могут решаться на ЭВМ с конечной раз-рядной сеткой лишь опосредованно, после проведения мероприятий, необходимых для защиты от этой неустойчивости или её нейтрализа-ции.**

**Например, задача решения систем линейных уравнений** Ax=b **с невырожденной квадратной матрицей** A **является вычислительно-корректной. Если топология на пространстве** R**n её решений задается обычным евклидовым расстоянием и подобным же традиционным об-разом задаётся расстояние между векторами правой части и матрица-ми, то существуют хорошо известные неравенства (см. §1.2), оцениваю-щие сверху границы изменения решений** x **через изменения элементов матрицы** A**, правой части** b **и число обусловленности матрицы** A**.**

**Для систем нелинейных уравнений, могущих иметь неединствен-ное решение, топологию на множестве ответов** A **нужно задавать уже каким-либо расстоянием между множествами, например, с помощью так назыаемой хаусдорфовой метрики [8]. Напомним её определение.**

**274** 4. Решение нелинейных уравнений и их систем

**Если задано метрическое пространство с метрикой** %**, то расстоя-нием точки** a **до множества** X **называется величина** %(a, X)**, определя-емая как** inf**x**∈**X** %(a, x)**. Хаусдорфовым расстоянием между компакт-ными множествами** X **и** Y **называют величину**

|  |  |  |  |
| --- | --- | --- | --- |
| % (X, Y ) = max | **n x**∈**X** | **y**∈**Y** | **o** |
|  | max % (x, Y ), max % (y, X) . | | |

**При этом** %(X, Y) = +∞**, если** X=∅ **или** Y=∅**. Введённая таким об-разом величина действительно обладает всеми свойствами расстояния и может быть использована для задания топологии на пространствах решений тех задач, ответы к которым неединствены.**

**4.1б** **Задача решения уравнений не является вычислительно-корректной**

**Уже простейшие примеры показывают, что задача решения уравнений и систем уравнений не является вычислительно-корректной. Например, квадратное уравнение**

|  |  |
| --- | --- |
| x***2*** + px + q = 0 | **(4.5)** |
| **для** |  |
| p***2*** = 4q | **(4.6)** |

**имеет лишь одно решение** x=−p/2**. Но при любых сколь угодно ма-лых возмущениях коэффициента** p **и свободного члена** q**, нарушающих равенство (4.6), уравнение (4.5) теряет это единственное решение или же приобретает ещё одно (см. Рис. 4.1).**

**Аналогичным образом ведёт себя решение двумерной системы урав-нений, эквивалентной (4.5),**

**(**

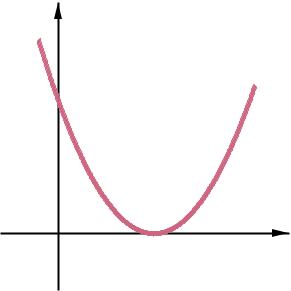
x + y = r,

xy = s

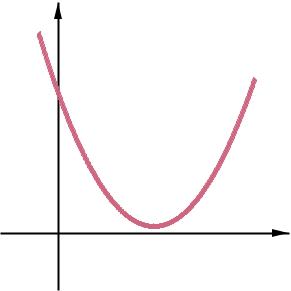
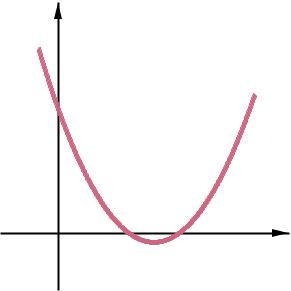
**при** s=r***2***/4**. При этом раздвоение решения не является большим грехом, коль скоро мы можем рассматривать хаусдорфово расстояние между целостными множествами решений. Но вот исчезновение един-ственного решения это чрезвычайное событие, однозначно указыва-ющее на разрывность разрешающего отображения.**

**Как видим, математическую постановку задачи нахождения реше-ний уравнений нужно ¾исправить¿, заменив какой-нибудь вычисли-тельно-корректной постановкой задачи. Приступая к поиску ответа на**

|  |  |
| --- | --- |
| 4.1. Вычислительно-корректные задачи | **275** |



x



x

x

**Рис. 4.1. Неустойчивая зависимость решений уравнения (4.5)–(4.6)**

**от сколь угодно малых шевелений его коэффициентов.**

**этот математический вопрос, отметим, прежде всего, что с точки зре-ния практических приложений задачи, которые мы обычно формули-руем в виде решения уравнений или систем уравнений, традиционно выписывая**

|  |  |
| --- | --- |
| F (x) = 0, | (**4.2**) |

**имеют весьма различную природу. Это и будет отправной точкой нашей ревизии постановки задачи решения уравнений и систем уравнений.**

**4.1в** ε**-решения уравнений**

**В ряде практических задач пользователю требуется не точное равен-ство некоторого выражения нулю, а лишь его ¾исчезающая малость¿ в сравнении с каким-то a priori установленным порогом. Аналогичным образом часто имеет смысл рассматривать соотношения, выражающие равенство двух каких-то выражений.**

**276** 4. Решение нелинейных уравнений и их систем

**Таковы, например, уравнения материального баланса в большин-стве физических, химических и других естественнонаучных расчётов. Точное равенство нулю здесь неявным образом и не требуется, так как масса молекулы, размеры атома, заряд элементарной частицы, дли-на световой волны и т.п. все это величины вполне конечные (хотя и весьма малые), обуславливающие точность тех или иных уравнений баланса и пр.**

**Например, не имеет смысла требовать, чтобы закон сохранения за-ряда выполнялся с погрешностью, меньшей чем величина элементарно-го электрического заряда (т. е. заряда электрона, равного** 1.6·10−***19* Кл). Также бессмысленно требовать, чтобы погрешность изготовления или подгонки деталей оптических систем была существенно меньшей дли-ны световой волны (от** 4·10−***7* м до** 7.6·10−***7* м в зависимости от цвета). А что касается температуры, то при обычных земных условиях определе-ние её с точностью превосходящей** 0.001 **градуса вообще проблематично в силу принципиальных соображений.**

**Наконец, ограниченная точность, с которой известны абсолютно все физические константы**1**, также воздвигает границы для требований ра-венства в физических соотношениях.**

**Во всех вышеприведенных примерах под решением уравнения по-нимается значение переменной, обращающее значение функции в пре-небрежимо малую величину:**

**Для заданных отображения** F:R**n** →R**n и** ε >0 **найти значения неизвестной переменной** x**, такие что** F(x)≈0 **с абсолютной погрешностью** ε **, т. е.** kF(x)k< ε **.**

**Решением этой задачи является целое множество точек, которые мы будем называть** ε**-решениями или почти решениями, если порог этой пренебрежимой несуществен или не оговорен явно.**

**Нетрудно понять, что** ε**-решения устойчивы к малым возмущениям исходных данных, а задача ¾о нахождении почти решений¿ является вычислительно-корректной, если отображение** F **непрерывно.**

1 **В лучшем случае относительная погрешность известных на сегодняшний день значений физических констант равна *10***−***10* , см. [51].**

|  |  |
| --- | --- |
| 4.1. Вычислительно-корректные задачи | **277** |

**Как уже отмечалось выше, в некоторых задачах система уравнений более естественно записывается не как (4.2), а в виде (4.3)**

G(x) = H(x),

**и требуется обеспечить с относительной погрешностью равенство её левой и правой частей:**

**Для заданных отображений** G**,** H:R**n** →R**n и** >0 **найти значения неизвестной переменной** x**, такие что** G(x) ≈ H(x) **с относительной погрешностью , т. е.**

kG(x) − H(x)k

< **.**

max{kG(x)k, kH(x)k}

**Решения этой задачи мы также будем называть -решениями системы уравнений вида (4.3).**

**Математические понятия, определения которых привлекают малый допуск , не являются новинкой. Таковы, к примеру,** ε**-энтропия мно-жеств в метрических пространствах,** ε**-субдифференциал функции,** ε**-оптимальные решения задач оптимизации и т.п. Наиболее близким, по-видимости, к нашим -решениям является концепция -спектра мат-рицы, которой интенсивно оперирует ряд авторов [11, 52, 47]. Говорят, что точка** z **на комплексной плоскости принадлежит -спектру матри-цы** A**, если существует комплексный вектор** v **единичной длины, такой что** k(A−zI)vk ≤ **, где** k · k **какая-то индуцированная матричная норма.**

**4.1г** **Недостаточность** ε**-решений**

**Но есть и принципиально другой тип задач, которые образно могут быть названы задачами ¾об определении перехода через нуль¿ и не сводятся к нахождению** ε**-решений. Таковы задачи, в которых требует-ся гарантированно отследить переход функции к значениям противо-положного знака (или, более общо, переход через некоторое критиче-ское значение). При этом, в частности, в любой окрестности решения**

**278** 4. Решение нелинейных уравнений и их систем

**должны присутствовать как положительные значения функции, так и её отрицательные значения, тогда как в задачах нахождения ¾почти решений¿ это условие может и не выполняться.**

**Фазовый переход в физической (или химической системе) типич-ная задача такого сорта, так как в процессе фазового перехода темпера-тура системы не меняется. Если мы хотим узнать, прошёл ли фазовый переход полностью, то нужно зафиксировать момент достижения мно-жества состояний, лежащего по другую сторону от границы раздела!**

**Другой пример. Рассмотрим систему линейных дифференциальных уравнений с постоянными коэффициентами**

|  |  |  |  |
| --- | --- | --- | --- |
| dx | = Ax, | **(4.7)** |  |
| dt |  |
|  |  |  |

**матрица которой** A=A(θ) **зависит от параметра** θ **(возможно, вектор-ного). Пусть при некотором начальном значении** θ=θ***0* собственные значения матрицы** A **имеют отрицательные вещественные части, так что все решения системы (4.7) устойчивы по Ляпунову (и даже асимп-тотически устойчивы). При каких значениях параметра** θ **рассматри-ваемая система сделается неустойчивой?**

**Традиционный ответ на этот вопрос: ¾Cрыв устойчивости в системе (4.7) произойдет при Re** λ(A(θ)) = 0**¿. Но он неправилен, так как для потери устойчивости необходимо не точное равенство нулю действи-тельных частей некоторых собственных чисел матрицы, а переход их через нуль в область положительного знака. Без этого перехода через мнимую ось и ¾ещё чуть-чуть дальше¿ система останется устойчивой, сколь бы близко мы не придвинули собственные значения к мнимой оси или даже достигли бы её. Здесь важен именно переход ¾через и за¿ критическое значение, в отсутствие которого качественное изменение в поведении системы не совершится, и этот феномен совершенно не ухва-тывается понятиями -решения из §4.1в или -спектра из [11, 52, 47].**

**Рассмотренная ситуация, в действительности, весьма типична для динамических систем, где условием совершения многих типов бифур-каций является переход некоторого параметра через так называемое бифуркационное значение. К примеру, при переходе через мнимую ось пары комплексных собственных чисел матрицы линеаризованной си-стемы происходит бифуркация Андронова-Хопфа (называемая также бифуркацией рождения цикла, см. [35]). И здесь принципиален имен-но переход через некоторый порог, а не близость к нему, на которую делается упор в понятиях -решения и -спектра.**

4.2. Теоретические основы численных методов **279**

**Нетрудно понять, что такое ¾переход через нуль¿ для непрерыв-ной функции одного переменного** f:R→R**. Но в многомерной ситуа-ции мы сталкиваемся с методическими трудностями, возникающими из необходимости иметь для нестрогого понятия ¾прохождение функции через нуль¿ чисто математическое определение. Из требования вычис-лительной корректности следует, что в любой окрестности такого ре-шения каждая из компонент** F**i**(x) **вектор-функции** F(x) **должна при-нимать как положительные, так и отрицательные значения. Но как именно? Какими должны (или могут) быть значения компонент** F**j** (x)**,**

* 6= i**, если** F**i**(x) > 0 **или** F**i**(x) < 0**?**
  + **разрешении этого затруднения нам на помощь приходят нелиней-**

**ный анализ и алгебраическая топология.**

1. Теоретические основы численных методов

**Этот параграф мы посвятим напоминанию некоторых важных фактов из математического анализа, которые существенно используются при численном решении уравнений и систем уравнений.**

**В математическом анализе хорошо известна**

Теорема Больцано-Коши. **Если функция** f : R → R **непрерывна** **на интервале** X⊂R **и на его концах принимает значения разных знаков, то внутри интервала** X **существует нуль функции** f **, т. е. точка** x˜∈X**, в которой** f(˜x) = 0**.**

**Часто её называют также ¾теоремой Больцано¿ (см., к примеру, [37]), так как именно Б. Больцано первым обнаружил это замечатель-ное свойство непрерывных функций. Многомерное обобщение этого ре-зультата было опубликовано более чем столетием позже в заметке [44]:**

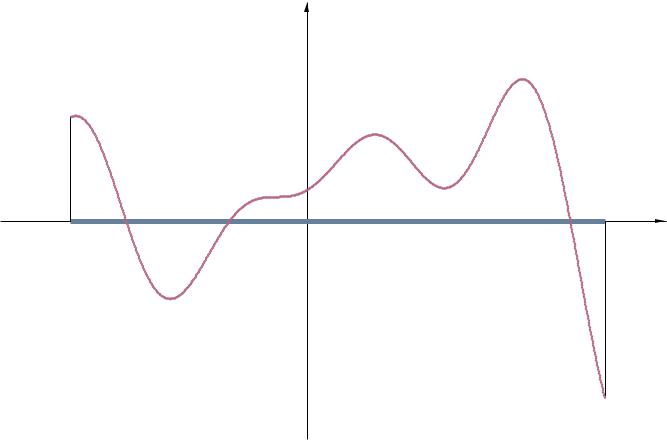
Теорема Миранды. **Пусть** f : R**n** → R**n,** f (x) = f***1***(x)**,** f***2***(x)**, . . . ,**

f**n**(x) > **функция, непрерывная на брусе** X ⊂ R**n** **со сторонами, па-раллельными координатным осям, и для любого** i= 1,2, . . . , n **имеет место либо**

f**i** X***1***, . . . , X**i**−***1***, X**i**, X**i*+1***, . . . , X**n** ≤ 0

**и** f**i** X***1***, . . . , X**i**−***1***, X**i**, X**i*+1***, . . . , X**n** ≥ 0,

|  |  |
| --- | --- |
| **280** | 4. Решение нелинейных уравнений и их систем |
|  | **Рис. 4.2. Иллюстрация теоремы Больцано-Коши** |



**либо**

f**i** X***1***, . . . , X**i**−***1***, X**i**, X**i*+1***, . . . , X**n** ≥ 0

* + f**i** X***1***, . . . , X**i**−***1***, X**i**, X**i*+1***, . . . , X**n** ≤ 0,
* **е. области значений каждой компоненты функции** f(x) **на соот-ветствующих противоположных гранях бруса** X **имеют разные зна-**

**ки. Тогда на брусе** X **существует нуль функции** f **, т. е. точка** x˜∈X**, в которой** f(˜x) = 0**.**

**Характерной особенностью теоремы Миранды является специаль-ная форма множества, на котором утверждается существование нуля функции: оно должно быть брусом со сторонами, параллельными ко-ординатным осям, т. е. интервальным вектором. Кроме того, для пол-ноценного применения теоремы Миранды нужно уметь находить или как-то оценивать области значений функций на подобных множествах.**

**Удобное средство для решения этой задачи предоставляют мето-ды интервального анализа. Задача об определении области значений функции на брусах из области её определения эквивалентна задаче оп-тимизации, но в интервальном анализе она принимает специфическую форму задачи о вычислении так называемого интервального расшире-ния функции (см. §1.4).**

**Аналогичным образом интервальные методы позволяют придать**

4.2. Теоретические основы численных методов **281**

**конструктивный характер следующему известному результату матема-тического анализа:**

Теорема Брауэра о неподвижной точке. **Пусть** D **выпук-лое компактное множество в** R**n. Если непрерывное отображение** g : R**n** → R**n** **переводит** D **в себя,** g(D) ⊆ D**, то оно имеет на** D **неподвижную точку** x∗**, т. е. такую что** x∗=g(x∗)**.**

**Если вместо произвольных выпуклых компактов ограничиться ин-тервальными векторами-брусами в** R**n, а для оценивания области зна-чений применять его внешнюю оценку в виде интервального расшире-ния, то условия теоремы Брауэра могут быть конструктивно проверены на компьютере.**

**С учётом сказанного выше во введении к главе (стр. 271) о рав-носильности рекуррентного вида систем уравнений (4.4) канонической форме (4.1)–(4.2) чрезвычайно полезными для вычислительной мате-матики оказываются результаты анализа, утверждающие существова-ние неподвижных точек у отображений, и именно такой является теоре-ма Брауэра. Наиболее часто существование неподвижных точек можно гарантировать у отображений, которые удовлетворяют тем или иным дополнительным условиям, и самыми популярными из них являются так называемые условия сжимаемости (сжатия) образа.**

**Напомним, что отображение** g:X→X **метрического пространства** X **с расстоянием** dist : X → R***+*** **называется** **сжимающим** **(или просто** **сжатием), если существует такая положительная постоянная** α <1**, что для любой пары элементов** x, y∈X **имеет место неравенство**

dist g(x), g(y) ≤ α · dist (x, y).

Теорема 4.2.1 **(теорема Банаха о неподвижной точке).** **Сжимающее** **отображение** g:X→X **полного метрического пространства** X **в себя имеет единственную неподвижную точку. Она может быть найдена методом последовательных приближений**

|  |  |
| --- | --- |
| x***(*k*+1)*** ←g(x***(*k*)***), | k = 0, 1, 2, . . . , |

**при любом начальном приближении** x***(0)*** ∈X**.**

**Доказательство этого результата можно найти, к примеру, в [17].**

**282** 4. Решение нелинейных уравнений и их систем

**Но иногда бывает полезно работать с векторнозначным расстоянием мультиметрикой, которая вводится на** R**n как**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Dist (x, y) := |  | dist (x**..** | ***1*** | , y***1***) | |
|  |  | dist (x**.** |  | , y | ) |
|  |  | **n** | | **n** |  |
|  |  |  |  |  |  |

∈ R**n** . **(4.8)**

***+***

**Для мультиметрических пространств аналогом теоремы Банаха о неподвижной точке для сжимающих отображений является приводи-мая ниже теорема Шрёдера о неподвижной точке. Перед тем, как дать её точную формулировку, введём**

Определение 4.2.1 **Отображение** g : X → X **мультиметрическо-го пространства** X **с мультиметрикой** Dist :X→R**n*+* называется** P -сжимающим **(или просто** P -сжатием**), если существует неотрица-тельная** n×n**-матрица** P **со спектральным радиусом** ρ(P)<1**, такая что для всех** x**,** y∈X **имеет место**

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | **(4.9)** |  |
| Dist g(x), g(y) | ≤ P · Dist (x, y). |  |

**Следует отметить, что математики, к сожалению, не придержива-ются здесь единой терминологии. Ряд авторов (см. [46]) за матрицей** P

**из (4.9) закрепляют отдельное понятие ¾оператора Липшица (матри-цы Липшица) отображения** g**¿, и в условиях Определения 4.2.1 говорят, что ¾оператор Липшица для** g **сжимающий¿.**

Теорема 4.2.2 **(теорема Шрёдера о неподвижной точке)** **Пусть отоб-ражение** g:R**n** ⊇X→R**n является** P **-сжимающим на замкнутом подмножестве** X **пространства** R**n с мультиметрикой** Dist **. Тогда для любого** x***(0)* последовательность итераций**

|  |  |
| --- | --- |
| x***(*k*+1)*** =g(x***(*k*)***), | k = 0, 1, 2, . . . , |

**сходится к единственной неподвижной точке** x∗ **отображения** g **в** X

**и имеет место оценка**

Dist ( x***(*k*)***, x∗) ≤ (I − P )−***1***P · Dist ( x***(*k*)***, x***(*k**−***1)***).

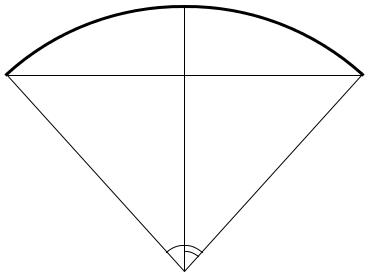
**Доказательство можно найти, например, в книгах [1, 18, 27, 46]**

|  |  |
| --- | --- |
| 4.3. Классические методы решения уравнений | **283** |

1. Классические методы решения уравнений

Пример 4.3.1 **Рабочие имеют кусок кровельного материала шири-ной** l= 3.3 **метра и хотят покрыть им пролёт шириной** h= 3 **метра, сделав крышу круглой, в форме дуги окружности. Для того, чтобы придать правильную форму балке, поддерживающей кровлю, нужно знать, какой именно радиус закругления крыши при этом получится (см. Рис. 4.3).**

*l*



*h*

*R*

**Рис. 4.3. Проектирование круглой крыши.**

**Обозначим искомый радиус закругления крыши через** R**. Если** 2α

**величина дуги (в радианах), соответствующей крыше, то**

l

= R.

2α

**С другой стороны, из рассмотрения прямоугольного треугольника с катетом** h/2 **и гипотенузой** R **получаем**

R sin α = h/2.

**Исключая из этих двух соотношений** R**, получим уравнение относи-тельно одной неизвестной** α**:**

|  |  |
| --- | --- |
| l sin α = αh. | **(4.10)** |

**Его решение не может быть выражено в явном виде, и потому далее мы обсудим возможности его численного решения.**

**284** 4. Решение нелинейных уравнений и их систем

**Уравнение (4.10) является простейшим нелинейным трансцендент-ным уравнением.**

**4.3а** **Предварительная локализация решений**

**Обычно первым этапом численного решения уравнений и систем урав-нений является предварительная локализация, т. е. уточнение местона-хождения, искомых решений. Это вызвано тем, что большинство чис-ленных методов для поиска решений имеют локальный характер, т. е. сходятся к этим решениям лишь из достаточно близких начальных при-ближений.**

**Для локализации решений могут применяться как численные, так и аналитические методы, а также их смесь гибридные методы, которые (следуя Д. Кнуту) можно назвать получисленными или полуаналити-ческими.**

**Особенно много аналитических результатов существует о локализа-ции решений алгебраических уравнений (корней полиномов), что, ко-нечно, имеет причину в очень специальном виде этих уравнений, до-пускающем исследование с помощью выкладок и т. п.**

Теорема 4.3.1 **Пусть для алгебраического уравнения вида**

a**n**x**n** + a**n**−***1***x**n**−***1*** + . . . + a***1***x + a***0*** = 0

**обозначено**

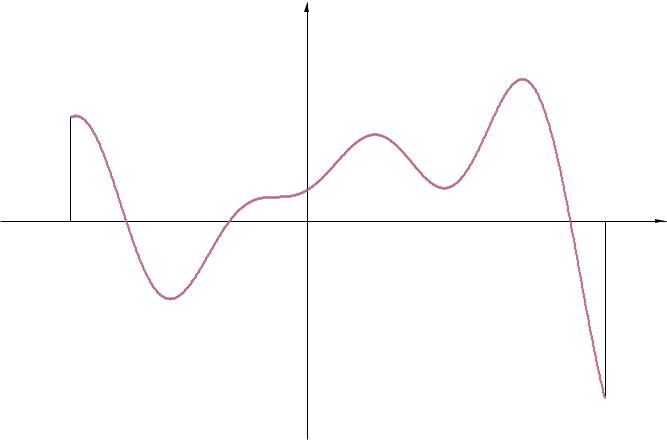
α = max{a***0***, . . . , a**n**−***1***}, β = max{a***1***, . . . , a**n**}.

**Тогда все решения этого уравнения принадлежат кольцу в комплекс-ной плоскости, определяемому условием**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1 |  | α | |  |
|  | ≤ |x| ≤ 1 + |  | . |  |
| 1 + β/|a***0***| | |a**n**| |  |

**Полезно правило знаков Декарта, утверждающее, что число поло-жительных корней полинома с вещественными коэффициентами равно числу перемен знаков в ряду его коэффициентов или на чётное число меньше этого числа. При этом корни считаются с учётом кратности, а нулевые коэффициенты при подсчёте числа перемен знаков не учи-тываются. Если, к примеру, заранее известно, что все корни данного полинома вещественны, то правило знаков Декарта даёт точное число**

|  |  |  |
| --- | --- | --- |
| 4.3. Классические методы решения уравнений | | **285** |
| **корней. Рассматривая полином с переменной** (−x) **можно с помощью** | | |
| **этого же результата найти число отрицательных корней исходного по-** | | |
| **линома.** | |  |
| **4.3б** | **Метод дихотомии** |  |
| **Этот метод часто называют также методом бисекции или методом по-** | | |
| **ловинного деления и т.п. Он заключается в последовательном делении** | | |
| **пополам интервала локализации корня уравнения, на концах которого** | | |
| **функция принимает значения разных знаков.** | |  |
|  | **Рис. 4.4. Иллюстрация метода дихотомии (деления пополам)** |  |



**Недостаток этого простейшего варианта метода дихотомии воз-можность потери решений, как это изображено на рисунке. Этого мож-но избежать, если отслеживать постоянный знак производной функ-ции.**

**4.3в** **Метод простой итерации**

**Методом простой итерации обычно называют стационарный одноша-говый итерационный процесс, который организуется после того, как исходное уравнение каким-либо способом приведено к равносильному рекуррентному виду** x= Φ(x)**. Далее, после выбора некоторого началь-ного приближения** x***(0)*, запускается итерационный процесс**

|  |  |
| --- | --- |
| x***(*k*+1)*** ←Φ(x***(*k*)***), | k = 0, 1, 2, . . . |

**286** 4. Решение нелинейных уравнений и их систем

**При благоприятных обстоятельствах последовательность** {x***(*k*)***} **схо-дится, и её пределом является решение исходного уравнения. Но в об-щем случае и характер сходимости, и вообще её наличие существенно зависят как от отображения** Φ**, так и от начального приближения к решению.**

Пример 4.3.2 **Уравнение (4.10) из Примера 4.3 нетрудно привести к** **рекуррентному виду**

l

α = h sin α,

**и, взяв в качестве начального приближения, например,** α***(0)*** = 1**, через 50 итераций**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| α***(*k*+1)*** | l | sin α***(*k*)***, | k = 0, 1, 2, . . . , | **(4.11)** |  |
| ← h |  |

**получить пять верных знаков точного решения** α∗= 0.748986642697. . .

**(читатель легко может самостоятельно проверить все числовые данные этого примера с помощью любой системы компьютерной математики).**

**Итерационный процесс (4.11) сходится к решению** α∗ **не из любо-го начального приближения. Если** α***(0)*** =πl**,** l∈Z**, то в результате итераций (4.11) получаем** α***(*k*)*** = 0**,** k= 1,2, . . .**. Если же** α***(0)* таково, что синус от него отрицателен, то итерации (4.11) сходятся к решению** (−α∗) **уравнения (4.10). И нулевое, и отрицательное решения очевидно** **не имеют содержательного смысла.**

**С другой стороны, переписывание исходного уравнения (4.11) в дру-гом рекуррентном виде**

* + = 1l arcsin(αh)

**приводит к тому, что характер сходимости метода простой итерации совершенно меняется. Из любого начального приближения, меньшего по модулю чем примерно** 0.226965**, итерации**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| α***(*k*+1)*** | 1 | arcsin α***(*k*)***h , | k = 0, 1, 2, . . . , |  |
| ← l |  |

**сходятся лишь к нулевому решению. Б´ольшие по модулю начальные приближения быстро выводят за границы области определения веще-ственного арксинуса, переводя итерации в комплексную плоскость, где они снова сходятся к нулевому решению. Таким образом, искомого ре-шения** α∗ **мы при этом никак не получаем.**

|  |  |
| --- | --- |
| 4.4. Классические методы решения систем уравнений | **287** |

**4.3г** **Метод Ньютона и его модификации**

**Для определения погрешности приближённого решения** x˜ **и контроля вычислений можно применять формулу**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| |x˜ − x∗| ≤ | f (˜x) | , | **(4.12)** |  |
| min**ξ**∈***[*a,b*]*** |f 0(ξ)| |  |

**которая следует из теоремы Лагранжа о среднем (формулы конечных приращений):**

f (˜x) − f (x∗) = f 0(ξ)(˜x − x∗),

**для некоторой точки** ξ∈{x,˜x∗}**, т. е. интервалу с концами** x˜ **и** x∗**. Ясно, что тогда**

|f (˜x) − f (x∗)| ≥ min |f 0(ξ)| · |x˜ − x∗|,

**ξ**

**и при** min**ξ**∈***[*a,b*]*** |f0(ξ)| 6= 0 **получаем оценку (4.12).**

**4.3д** **Методы Чебышёва**

1. Классические методы решения систем уравнений

**4.4а** **Метод простой итерации**

**4.4б** **Метод Ньютона и его модификации**

1. Интервальные линейные системы уравнений

**Предметом рассмотрения настоящего пункта являются интервальные системы линейных алгебраических уравнений (ИСЛАУ) вида**

|  |  |
| --- | --- |
| Ax = b, | **(4.13)** |

**где** A= (a**ij** ) **это интервальная** m×n**-матрица и** b= (b**i**) **интер-вальный** m**-вектор. Строго говоря, для интервальных уравнений ре-шения и множества решений могут быть определены разнообразными**

**288** 4. Решение нелинейных уравнений и их систем

**способами (см. [36]), но ниже мы ограничимся так называемым объеди-нённым множеством решений для (4.13), которое образовано всевоз-можными решениями** x **точечных систем** Ax=b**, когда матрица** A **и вектор** b **независимо пробегают** A **и** b **соответственно. Объединённое множество решений определяется строго как**

|  |  |
| --- | --- |
| Ξ(A, b) = { x ∈ R**n** | (∃ A ∈ A)(∃ b ∈ b)(Ax = b)}, | **(4.14)** |

**и ниже мы будем называть его просто множеством решений интер-вальной линейной системы (4.13), так как другие множества реше-ний нами не исследуются. Точное описание множества решений может расти экспоненциально с размерностью вектора неизвестных** n**, а по-тому является практически невозможным уже при** n**, превосходящем несколько десятков. С другой стороны, в большинстве реальных поста-новок задач точное описание на самом деле и не нужно. На практике бывает вполне достаточно нахождения оценки для множества реше-ний, т.е. приближенного описания, удовлетворяющего содержательно-му смыслу рассматриваемой задачи.**

**Приведём полезный технический результат, который часто исполь-зуется в связи с исследованием и оцениванием множества решений ин-тервальных систем линейных алгебраических уравнений.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Теорема 4.5.1 **(характеризация Бека)** | | |  | **Если** A∈IR**m**×**n,** b∈IR**m,** | |  |
| **то** | ∈ | **n** | |  |  | b |  |
|  |  | A |  |
| Ξ(A, b) = x |  | R**n** | Ax ∩ b 6= | | ∅ |  |
| = x ∈ R | 0 ∈ x − | | | | | . |  |
|  |  |  | ˜ | ˜ | ˜ |  |
| Доказательство. **Если** x˜ ∈ Ξ(A, b)**, то** Ax˜ = b **для некоторых** A ∈ A**,** | | | | | |  |
| ˜ |  | ˜ |  |  |  |  |
| b ∈ b**. Следовательно, по крайней мере** b ∈ Ax˜ ∩b**, так что действитель-** | | | | | |  |
| **но** Ax˜∩b6=∅**.** |  |  |  |  |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Наоборот,mесли** Ax˜∩b6=∅**, то это пересечение** Ax˜∩b **содержит** | | | |  |
| ˜ | ˜ | ˜ | **с** |  |
| **вектор** b∈R | **, для которого должно иметь место равенство** b=Ax˜ | |  |

˜

**некоторой** A∈A**. Итак,** x˜∈Ξ(A,b)**.**

**Второе равенство следует из того, что** Ax˜∩b6=∅ **тогда и только тогда, когда** 0∈Ax˜−b**.**

Теорема 4.5.2 **(характеризация Оеттли-Прагера)** **Для объединённого** **множества решений ИСЛАУ имеет место**

|  |  |  |  |
| --- | --- | --- | --- |
| x ∈ Ξ(A, b) ⇔ | (mid A) x − mid b |  | ≤ rad A · |x| + rad b, **(4.15)** |
|  |  |  |  |

|  |  |
| --- | --- |
| 4.6. Интервальные методы решения уравнений | **289** |

**где неравенство между векторами понимается покомпонентным об-разом.**

Доказательство. **Включение** p ⊆ q **для интервальных векторов-брусов** p **и** q **равносильно покомпонентному неравенству**

| mid q − mid p | ≤ rad q − rad p.

**Следовательно, характеризация Бека может быть переписана в следу-ющем виде:**

1. Интервальные методы решения уравнений и систем уравнений

**Задача решения уравнений и систем уравнений является одной из клас-сических задач вычислительной математики, для решения которой раз-вито немало эффективных подходов метод простой итерации, ме-тод Ньютона, их модификации и т.п. Преимущества и недостатки этих классических методов мы обсудили выше в §§4.3–4.4 (см. также [5, 27, 31, 42]). Для дальнейшего нам важны два факта:**

* **Для уравнений, в которых фигурируют функции, не обладаю-щие ¾хорошими¿ глобальными свойствами, все традиционные ме-тоды имеют локальный характер, т. е. обеспечивают отыскание решения, находящегося в некоторой (иногда достаточно малой) окрестности начального приближения. Задача нахождения всех решений уравнения или системы уравнений, как правило, рас-сматривается лишь в специальных руководствах и методы её ре-шения оказываются очень сложными.**
* **Гарантированные оценки погрешности найденного приближения к решению в традиционных методах дать весьма непросто.**

**Указание приближённого значения величины и его максимальной погрешности равносильно тому, что мы знаем левую и правую границы возможных значений этой величины, и поэтому можно переформули-ровать нашу задачу в следующем усиленном виде**

**290** 4. Решение нелинейных уравнений и их систем

**Для каждого решения системы уравнений**

F (x) = 0

, **(4.16)**

**на данном множестве** D⊆R **n найти гарантированные двусторонние границы**

**который будем называть задачей доказательного глобального реше-ния системы уравнений. Эпитет ¾доказательный¿ означает здесь, что получаемый нами ответ к задаче границы решений и т.п. имеет статус математически строго доказанного утверждения о расположе-нии решений (при условии, что ЭВМ работает корректно)**2

**Задача (4.16) оказывается чрезвычайно сложной, а в классическом численном анализе почти полностью отсутствуют развитые методы для её решения. Из часто используемых подходов, имеющих ограниченный успех, следует упомянуть аналитическое исследование, мультистарт, методы продолжения [27].**

**Итак, пусть к решению предъявлена система уравнений (4.2)**

F (x) = 0

**на брусе** X⊂R**n. Отметим, что эту задачу можно переписать в виде**

ran (F, X) 3 0,

**и потому техника интервального оценивания множеств значений функ-ций оказывается весьма полезной при решении вопроса о существова-нии или несуществовании решений системы (4.2). В частности, если нуль содержится во внутренней интервальной оценке множества зна-чений** ran (F,X) **отображения** F **, то на брусе** X **гарантированно на-ходится решение системы (4.2). С другой стороны, если в нашем рас-поряжении имеется интервальное расширение** F **функции** F **на** X**, то** F (X) ⊇ ran (F, X)**. Поэтому** 0 6 F (X) **влечёт вывод о том, что на** X

**нет решений рассматриваемой системы уравнений.**

2 **Термин ¾доказательные вычисления на ЭВМ¿ впервые систематически исполь-зовал К.И. Бабенко [49]. Этот термин является хорошим русским эквивалентом та-ких распространённых английских оборотов как verified computation, verification numerics и др.**

|  |  |
| --- | --- |
| 4.6. Интервальные методы решения уравнений | **291** |

**Далее, перепишем исходную систему (4.2) в равносильной рекур-**

|  |  |
| --- | --- |
| **рентной форме** |  |
| x = T (x) | **(4.17)** |

**с некоторым отображением** T:R**n** →R**n. Оно может быть взято, к примеру, в виде**

T (x) = x − F (x)

**либо**

T (x) = x − ΛF (x),

**с неособенной** n×n**-матрицей** Λ**, либо как-нибудь ещё. Пусть также** T : IR**n** → IR**n** **интервальное расширение отображения** T **. Ясно, что** **решения системы (4.17) могут лежать лишь в пересечении** X∩T(X)**. Поэтому если**

X ∩ T (X) = ∅,

**то в** X **нет решений системы уравнений (4.17). Коль скоро искомое ре-шение содержится и в** T(X)**, то для дальнейшего уточнения бруса, в котором может присутствовать решение, мы можем организовать ите-рации с пересечением**

|  |  |  |
| --- | --- | --- |
| X***(0)*** | ← X, | **(4.18)** |
| X***(*k*+1)*** | ← T (X***(*k*)***) ∩ X***(*k*)***,k = 0, 1, 2, . . . . | **(4.19)** |

**Следует особо отметить, что в получающихся при этом брусах нали-чие решения, вообще говоря, не гарантируется. Они являются лишь ¾подозрительными¿ на существование решения.**

**Но вот если для бруса** X **выполнено**

T (X) ⊆ X,

**то по теореме Брауэра о неподвижной точке (стр. 281) в** X **гарантиро-ванно находится решение системы (4.17). Для уточнения этого бруса мы снова можем воспользоваться итерациями (4.18)–(4.19). Таким об-разом, наихудшим, с точки зрения уточнения информации о решении системы, является случай**

T (X) % X. **(4.20)**

**Приведённую выше последовательность действий по обнаружению решения системы уравнений и уточнению его границ мы будем назы-вать далее кратко тестом существования. Условимся также считать,**

**292** 4. Решение нелинейных уравнений и их систем

**что его результатом является брус пересечения** (X∩T(X)) **либо пре-дел последовательности (4.18)–(4.19). Если этот брус непуст, то он либо наверняка содержит решение системы уравнений, либо является подо-зрительным на наличие в нём решения. Если же результат теста суще-ствования пуст, то в исходном брусе решений системы уравнений нет.**

**В действительности, каждый из изложенных выше приёмов уточ-нения решения допускает далеко идущие модификации и улучшения. Например, это относится к итерациям вида (4.18)–(4.19), которые мо-гут быть последовательно применены не к целым брусам** X***(*k*)*, а к от-дельным их компонентам в комбинации с различными способами при-ведения исходной системы к рекуррентному виду (4.17). На этом пути мы приходим к чрезвычайно эффективным алгоритмам, которые по-лучили наименование методов распространения ограничений (см., к примеру, [30]).**

**Как простейший тест существования, так и его более продвинутые варианты без особых проблем реализуются на ЭВМ и работают тем лучше, чем более качественно вычисляются интервальные расшире-ния функций** F **в (4.2) и** T **в (4.17) и чем меньше ширина бруса** X**. Последнее связано с тем, что погрешность оценивания области значе-ний функции посредством любого интервального расширения убывает с уменьшением размеров бруса, на котором производится это оценива-ние. (см. §1.4).**

**4.6а** **Одномерный интервальный метод Ньютона**

**В этом параграфе мы рассмотрим простейший случай одного уравне-ния с одним неизвестным.**

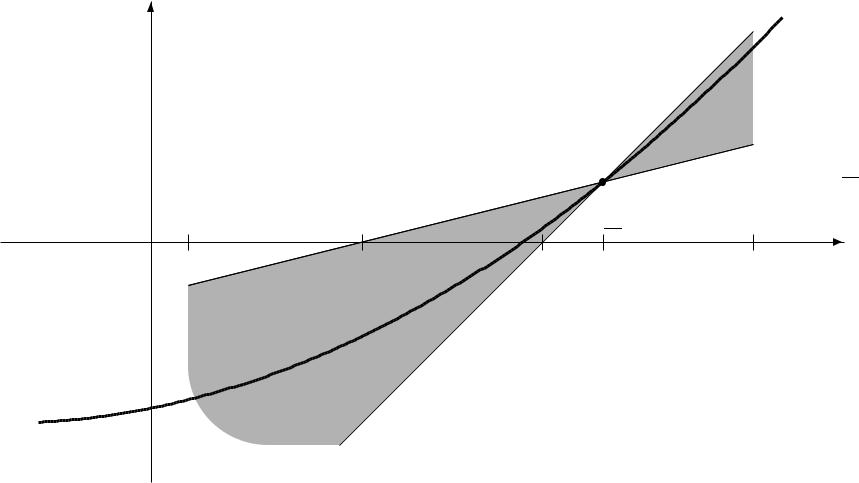
**Предположим, что** f:R⊇x→R **непрерывно дифференцируе-мая функция, имеющая нуль** x**? на интервале** x**, т. е.** f(x**?**) = 0**. Тогда для любой точки** x˜∈x **из этого же интервала в силу теоремы Лагран-жа о среднем значении**

f (˜x) − f (x**?**) = (˜x − x**?**) · f 0(ξ),

**где** ξ **некоторая точка между** x˜ **и** x**?. Но так как** f(x**?**) = 0**, то отсюда следует**

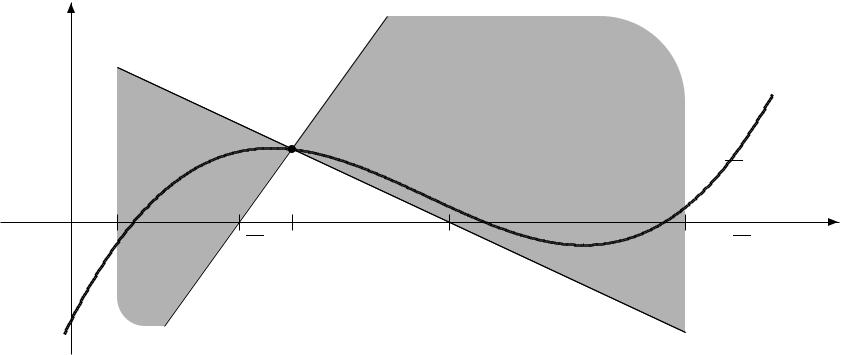
x**?** = x˜ − ff0(˜(xξ)) .

|  |  |  |
| --- | --- | --- |
| 4.6. Интервальные методы решения уравнений |  | **293** |
| **Если** f0(x) **является каким-либо интервальным расширением произ-** | | |
| **водной функции** f(x) **на** x**, то** f0(ξ)∈f0(x) **и** |  |  |
| f (˜x) |  |  |
| x**?** ∈ x˜ − f 0(x) . |  |  |
| **Интервальное выражение, фигурирующее в правой части этого вклю-** | | |
| **чения, будет играть в дальнейшем важную роль и потому достойно** | | |
| **выделения самостоятельным понятием.** |  |  |
| Определение 4.6.1 **Для заданной функции** f : R → R **отображение** | | |
| N : IR × R → IR, |  |  |
| **действующее по правилу** |  |  |
| f (˜x) |  |  |
| N(x, x˜) := x˜ − f 0(x) |  |  |
| **называется (одномерным)** интервальным оператором Ньютона**.** | |  |
| x | x˜ | x |
| x0 | x0 |  |
| y ***=*** f ***(***x***)*** |  |  |
| **Рис. 4.5. Иллюстрация работы одномерного** | |  |
| **интервального метода Ньютона. Ситуация 1.** | |  |



**Допустим на время, что** 06 f0(x)**, так что** N(x, x˜) **является вполне определённым конечным интервалом. Так как любой нуль функции**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **294** | 4. | Решение нелинейных уравнений и их систем | |  |
| f (x) **на** x **лежит также и в** N(x, x˜)**, то разумно взять в качестве следу-** | | | |  |
| **ющего более точного приближения к решению пересечение** | | | |  |
|  |  | x ∩ N(x, x˜), |  |  |
| **которое окажется, по крайней мере, не хуже** x**.** | | |  |  |
|  |  |  | y ***=*** f ***(***x***)*** |  |
| x | x˜ |  | x |  |
|  |  |  |  |
| x0 | x0 | x00 | x00 |  |
|  | **Рис. 4.6. Иллюстрация работы одномерного** | | |  |
|  | **интервального метода Ньютона. Ситуация 2.** | | |  |



**Далее, если** 0∈f0(x)**, мы можем придать смысл оператору Ньюто-на, воспользовавшись интервальной арифметикой Кахана. В действи-тельности, эта модификация даже усилит интервальный метод Ньюто-на, так как мы получим возможность отделять решения друг от дру-га: в результате выполнения шага интервального метода Ньютона при** 0 ∈ int f 0(x) **получаются, как правило, два непересекающися интерва-ла.**

**Свойства одномерного интервального метода Ньютона:**

1. **Всякий нуль функции** f **на исходном интервале** x **корректно вы-деляется методом.**
2. **Если на исходном интервале** x **нет нулей функции** f **, то этот факт будет установлен методом за конечное число итераций.**
3. **Если** 06 f0(x) **для некоторого** x**, то на следующем шаге метода будет исключена по крайней мере половина** x
4. **Если** 06 f0(x)**, то асимптотический порядок сходимости метода к нулю функции** f **на интервале** x **является квадратичным.**

|  |  |
| --- | --- |
| 4.6. Интервальные методы решения уравнений | **295** |

**4.6б** **Многомерный интервальный метод Ньютона**

**Переходя к решению систем нелинейных уравнений, следует отметить, что многомерные версии интервального метода Ньютона гораздо более многочисленны, чем одномерные, и отличаются очень большим раз-нообразием. В многомерном случае мы можем варьировать не только выбор точки** x˜**, вокруг которой осуществляется разложение, форму ин-тервального расширения производных или наклонов функции, как это было в одномерном случае, но также и способ внешнего оценивания множества решений интервальной линейной системы, к которой приво-дится оценивание бруса решения. В оставшейся части этого параграфа мы рассмотрим простейшую форму многомерного интервального мето-да Ньютона, а его более специальным версиям, которые связываются с именами Кравчика и Хансена-Сенгупты, будут посвящены отдельные параграфы.**

Определение 4.6.2 **[46]** **Для отображения** F:R**n** ⊇D***0*** →R**m**

**матрица** A∈IR**m**×**n называется** интервальной матрицей Липшица

**на** D⊆D***0*, если для любых** x, y∈D **равенство**

F (y) − F (x) = A(y − x)

**имеет место с некоторой вещественной** m×n**-матрицей** A∈A**.**

**Предположим, что на брусе** x **к решению предъявлена система нели-нейных уравнений**

|  |  |
| --- | --- |
| F (x) = 0. | **(4.21)** |

**Если** L **интервальная матрица Липшица отображения** F **на** x**, то для любых точек** x, x˜∈x **справедливо представление**

F (x) ∈ F (˜x) + L(x − x˜).

**В частности, если** x **решение системы уравнений (4.21), т. е.** F(x) = 0**,**

|  |  |
| --- | --- |
| **то** |  |
| 0 ∈ F (˜x) + L(x − x˜). | **(4.22)** |

**Вспомним характеризацию Бека для объединённого множества реше-ний ИСЛАУ (Теорема 4.5.1): получается, что точка** x **удовлетворяет**

**296** 4. Решение нелинейных уравнений и их систем

**включению (4.22) тогда и только тогда, когда она принадлежит объ-единённому множеству решений интервальной линейной системы**

|  |  |
| --- | --- |
| L(x − x˜) = −F (˜x). | **(4.23)** |

**Далее, если** Encl **процедура внешнего оценивания множества ре-шений ИСЛАУ, то справедливо включение**

x − x˜ ∈ Encl (L, −F (˜x)),

**так что**

x ∈ x˜ + Encl (L, −F (˜x)).

Определение 4.6.3 **Пусть для внешнего оценивания множеств ре-шений ИСЛАУ зафиксирована процедура** Encl**, а для отображения** F:R**n** ⊇ D → R**n** **известна интервальная матрица Липшица** L ∈ IR**n**×**n.** **Отображение**

N : ID × R**n** → IR**n**,

**задаваемое правилом**

N(x, x˜) = x˜ + Encl (L, −F (˜x)),

**называется** интервальным оператором Ньютона **на** ID **относительно точки** x˜**.**

**Как лучше выбирать центр разложения** x˜**? Имеет смысл делать это так, чтобы величина** kF(˜x)k **была, по-возможности, меньшей. Чем меньше будет норма вектор-функции** F(˜x)**, тем меньшим будет нор-ма векторов, образующих множество решений интервальной линейной системы**

L(x − x˜) = −F (˜x),

**которое мы должны пересекать с исходным брусом. Может быть, мы получим при этом более узкую внешнюю оценку множества решений исходной нелинейной системы и более точно определим статус исследу-емого бруса. Численные эксперименты как будто подтверждают этот вывод.**

**Процедуру для уточнения центра разложения можно организовать как метод типа Ньютона, коль скоро нам известна интервальная мат-рица Липшица.**

|  |  |
| --- | --- |
| 4.6. Интервальные методы решения уравнений | **297** |

**Наиболее неблагоприятной ситуацией при работе интервального ме-тода Ньютона является, конечно, появление включения**

N(x, x˜) ⊇ x.

**Тогда все последующие шаги зацикливаются на брусе** x **и не дают ника-кой дополнительной информации об искомых решениях системы. Как поступать в этом случае? Ответ на этот вопрос рассматривается в сле-дующем параграфе.**

**4.6в** **Метод Кравчика**

**Пусть на брусе** x∈IR**n задана система** n **нелинейных уравнений c** n

**неизвестными**

F (x) = 0,

**для которой требуется уточнить двусторонние границы решений. Возь-мём какую-нибудь точку** x˜∈x **и организуем относительно неё разло-жение функции** F **:**

F (x) ∈ F (˜x) + L(x − x˜),

**где** L∈R**n**×**n интервальная матрица Липшица отображения** F **на брусе** x**. Если** x **это точка решения системы, то**

|  |  |
| --- | --- |
| 0 ∈ F (˜x) + L(x − x˜), | **(4.22)** |

**но далее, в отличие от интервального метода Ньютона, мы не будем переходить к рассмотрению интервальной линейной системы (4.23), а домножим обе части этого включения слева на точечную** n×n**-матрицу, которую нам будет удобно обозначить как** (−Λ)**:**

0 ∈ −ΛF (˜x) − ΛL(x − x˜).

**Добавление к обеим частям получившегося соотношения по** (x−x˜)

**приводит к**

x − x˜ ∈ −ΛF (˜x) − ΛL(x − x˜) + (x − x˜),

**что равносильно**

x ∈ x˜ − ΛF (˜x) + (I − ΛL)(x − x˜),

**так как для неинтервального общего множителя** (x−x˜) **можно вос-пользоваться дистрибутивным соотношением (1.9). Наконец, если ре-шение** x **системы уравнений предполагается принадлежащим брусу** x**,**

**298** 4. Решение нелинейных уравнений и их систем

**мы можем взять интервальное расширение по** x∈x **правой части по-лученного включения, придя к соотношению**

x ∈ x˜ − ΛF (˜x) + (I − ΛL)(x − x˜),

Определение 4.6.4 **Пусть определены некоторые правила, сопостав-ляющие всякому брусу** x∈IR**n точку** x˜∈x **и вещественную** n×n**-матрицу** Λ **и пусть также** L∈IR**n**×**n интервальная матрица Липшица отображения** F:R**n** ⊇D→R**n на** D**. Отображение**

K : ID × R → IR**n**,

**задаваемое выражением**

K(x, x˜) := x˜ − ΛF (˜x) + (I − ΛL)(x − x˜),

**называется** оператором Кравчика **на** ID **относительно точки** x˜**.**

Теорема 4.6.1 **Пусть** F : R**n** ⊇ D → R**n** **непрерывное по Липшицу** **отображение,** L **его интервальная матрица Липшица и** x˜∈x⊆ID**. Тогда**

1. **каждое решение системы** F(x) = 0 **на брусе** x **лежит также в**

K(x, x˜)**;**

1. **если** x∩K(x, x˜) =∅**, то в** x **нет решений системы** F(x) = 0**;**
2. **если** K(x, x˜)⊆x**, то в** x **находится хотя бы одно решение си-стемы** F(x) = 0**;**
3. **если** x˜∈intx **и** ∅6=K(x, x˜)⊆intx**, то матрица** L **сильно неособенна и в** K(x, x˜) **содержится в точности одно решение системы** F(x) = 0**.**

**Оператор Кравчика это не что иное, как центрированная форма интервального расширения отображения** Φ(x) =x−ΛF(x)**, возникаю-щего в правой части системы уравнений после её приведения к рекур-рентному виду**

x = Φ(x).

|  |  |
| --- | --- |
| 4.7. Глобальное решение уравнений | **299** |

1. Глобальное решение уравнений и систем уравнений

**Если ширина бруса** X **велика, то на нём описанные в предшествующем параграфе методики уточнения решения могут оказаться малоуспеш-ными в том смысле, что мы получим включение (4.20), из которого нельзя вывести никакого определённого заключения ни о существова-нии решения на брусе** X**, ни о его отсутствии. Кроме того, сам этот брус, как область потенциально содержащая решение, нисколько не будет уточнён (уменьшен).**

**Тогда практикуют принудительное дробление** X **на более мелкие подбрусы. Наиболее популярна при этом бисекция разбиение бру-са** X **на две (равные или неравные) части вдоль какой-нибудь грани, например, на половинки**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| X0 | = |  | X***1***, . . . , [ X**ι**, mid X**ι** ], . . . , X**n** | | | , |  |
| X00 |  |  |  |  |  |  |  |
| = | X***1***, . . . , [ mid X**ι**, X**ι** ], . . . , X**n** | | |  |

**для некоторого номера** ι∈{1,2, . . . , n}**. При этом подбрусы** X0 **и** X00 **называются потомками бруса** X**. Далее эти потомки можно разбить ещё раз, и ещё . . . столько, сколько необходимо для достижения желае-мой малости их размеров, при которой мы сможем успешно выполнять на этих брусах рассмотренные выше тесты существования решений.**

**Если мы не хотим упустить при этом ни одного решения системы, то должны хранить все возникающие в процессе такого дробления подб-русы, относительно которых тестом существования не доказано строго, что они не содержат решений. Организуем поэтому рабочий список** L **из всех потомков начального бруса** X**, подозрительных на содержание решений.**3 **В целом же алгоритм глобального доказательного решения системы уравнений организуем в виде повторяющейся последователь-ности следующих действий:**

* **извлечение некоторого бруса из списка** L**,**
* **дробление этого бруса на потомки,**
* **проверка существования решений в каждом из подбрусов-потомков, по результатам которой мы**

3 **Хотя мы называем эту структуру данных ¾списком¿, в смысле программной реализации это может быть не обязательно список, но и** стек **(магазин), и** куча **[4].**

**300** 4. Решение нелинейных уравнений и их систем

* **либо выдаём этот подбрус в качестве ответа к решаемой задаче,**
* **либо заносим его в рабочий список** L

**для последующей обработки,**

* **либо исключаем из дальнейшего рассмотрения, как не содержащий решений рассматриваемой системы.**

**Кроме того, чтобы обеспечить ограниченность времени работы алго-ритма, на практике имеет смысл задаться некоторым порогом мелкости (малости размеров) брусов** δ**, при достижении которого дальше дро-бить брус уже не имеет смысла. Псевдокод получающегося алгоритма приведён в Табл. 4.1.**

**Отметим, что неизбежные ограничения на вычислительные ресур-сы ЭВМ могут воспрепятствовать решению этим алгоритмом задачи (4.16) ¾до конца¿, поскольку могут возникнуть ситуации, когда**

* 1. **размеры обрабатываемого бруса уже меньше** δ**, но нам ещё не удаётся ни доказать существование в нём решений, ни показать их отсутствие;**
  2. **размеры обрабатываемого бруса ещё больше** δ**, но вычислительные ресурсы уже не позволяют производить его обработку дальше: исчерпались выделенное время, память и т.п.**
* **реальных вычислениях остановка алгоритма Табл. 4.1 может проис-**

**ходить поэтому не только при достижении пустого рабочего списка** L

**(когда исчерпана вся область поиска решений), но и, к примеру, при достижении определённого числа шагов или времени счёта и т.п. Тогда**

**все брусы, оставшиеся в рабочем списке** L**, оказываются не до конца обработанными, и мы условимся так и называть их ¾недообрабо-танные¿. Итак, в общем случае результатом работы нашего алгоритма должны быть три списка брусов:**

**список** НавернякаРешения**, состоящий из брусов шириной меньше** δ**, которые гарантированно содержат решения,**

**список** ВозможноРешения**, состоящий из брусов шириной меньше** δ**, подозрительных на содержание решения, и**

4.7. Глобальное решение уравнений **301**

**Таблица 4.1. Простейший интервальный алгоритм глобального доказательного решения уравнений**

Вход

**Система уравнений** F(x) = 0**. Брус** X∈IR**n. Интервальное расширение** F:IX→IR **функции** F **. Заданная точность** δ >0 **локализации решений системы.**

Выход

**Список** НавернякаРешения **из брусов размера менее** δ**, которые гарантированно содержат решения системы уравнений в** X**.**

**Список** ВозможноРешения **из брусов размера менее** δ**, которые могут содержать решения системы уравнений в** X**.**

**Список** Недообработанные **из брусов размера более** δ**, которые могут содержать решения системы уравнений в** X**.**

Алгоритм

**инициализируем рабочий список** L **исходным брусом** X **;**

DO WHILE ( L =6 ∅ ) **и ( не исчерпаны ресурсы ЭВМ )**

**извлекаем из рабочего списка** L **брус** Y **;**

**применяем к** Y **тест существования решения, его результат обозначаем также через** Y **;**

IF **( в** Y **доказано отсутствие решений )** THEN

**удаляем брус** Y **из рассмотрения**

ELSE

IF **( (размер бруса** Y **) <** δ **)** THEN

**заносим** Y **в соответствующий из списков**

НавернякаРешения **или** ВозможноРешения

ELSE

**рассекаем** Y **на потомки** Y0 **и** Y00 **и заносим их в рабочий список** L

END IF

END IF

END DO

**все брусы из** L **перемещаем в список** Недообработанные**;**

**302** 4. Решение нелинейных уравнений и их систем

**список** Недообработанные**, состоящий брусов, которые алгоритму не удалось обработать ¾до конца¿ и которые имеют ширину не меньше** δ**,**

**такие что все решения рассматриваемой системы уравнений, не при-надлежащие брусам из списка** НавернякаРешения**, содержатся в брусах из списков** ВозможноРешения **и** Недообработанные**.**

**Алгоритмы описанного выше типа, дополненные различными усо-вершенствованиями, получили большое развитие в интервальном ана-лизе в последние десятилетия (см., например, книги [40, 41, 45, 46]), а реализованные на их основе программные комплексы существенно про-двинули практику численного решения уравнений и систем уравнений.**

Литература к Главе 4

Основная

1. **Алефельд Г., Херцбергер Ю. Введение в интервальные вычисления. –**

**Москва: Мир, 1987.**

1. **Акритас А. Основы компьютерной алгебры с приложениями. – Москва: Мир, 1994.**
2. **Барахнин В.Б., Шапеев В.П. Введение в численный анализ. – Санкт-Петербург–Москва– Краснодар: Лань, 2005.**
3. **Бауэр Ф.Л., Гооз Г. Информатика. В 2-х ч. – Москва: Мир, 1990.**
4. **Бахвалов Н.С., Жидков Н.П., Кобельков Г.М. Численные методы. –**

**Москва: Бином, 2003, а также другие издания этой книги.**

1. **Бахвалов Н.С., Корнев А.А., Чижонков Е.В. Численные методы. Реше-ния задач и упражнения. – Москва: Дрофа, 2008.**
2. **Березин И.С., Жидков Н.П. Методы вычислений. Т. 1–2. – Москва: Наука, 1966.**
3. **Берже М. Геометрия. Т. 1, 2. – Москва: Наука, 1984.**
4. **Вержбицкий В.М. Численные методы. Части 1–2. – Москва: ¾Оникс 21 век¿, 2005.**
5. **Волков Е.А. Численные методы. – Москва: Наука, 1987.**
6. **Годунов С.К. Современные аспекты линейной алгебры. – Новосибирск: На-учная книга, 1997.**
7. **Годунов С.К., Антонов А.Г., Кириллюк О.П., Костин В.И. Гарантиро-ванная точность решения систем линейных уравнений в евклидовых простран-ствах. – Новосибирск: Наука, 1992.**
8. **Гэри М., Джонсон Д. Вычислительные машины и труднорешаемые задачи.**

**– Москва: Мир, 1982.**

4.7. Глобальное решение уравнений **303**

1. **Демидович Б.П., Марон А.А. Основы вычислительной математики. –**

**Москва: Наука, 1970.**

1. **Дэннис Дж., мл., Шнабель Р. Численные методы безусловной оптимизации и решения нелинейных уравнений. – Москва: Мир, 1988.**
2. **Калиткин Н.Н. Численные методы. – Москва: Наука, 1978.**
3. **Канторович Л.В., Акилов Г.П. Функциональный анализ. – Москва: Наука, 1984.**
4. **Коллатц Л. Функциональный анализ и вычислительная математика. –**

**Москва: Мир, 1969.**

1. **Крылов А.Н. Лекции о приближённых вычислениях. – Москва: ГИТТЛ, 1954, а также более ранние издания.**
2. **Крылов В.И., Бобков В.В., Монастырный П.И. Вычислительные методы. Т. 1–2. – Москва: Наука, 1976.**
3. **Кунц К.С. Численный анализ. – Киев: Техника, 1964.**
4. **Мацокин А.М. Численный анализ. Вычислительные методы линейной алгеб-ры. Конспекты лекций для преподавания в III семестре ММФ НГУ. Ново-сибирск: НГУ, 2009–2010.**
5. **Мацокин А.М., Сорокин С.Б. Численные методы. Часть 1. Численный ана-лиз. Новосибирск: НГУ, 2006.**
6. **Меньшиков Г.Г. Локализующие вычисления. Конспект лекций. – Санкт-Петербург: СПбГУ, Факультет прикладной математики–процессов управле-ния, 2003.**
7. **Миньков С.Л., Миньков Л.Л. Основы численных методов. – Томск: Изда-тельство научно-технической литературы, 2005.**
8. **Мысовских И.П. Лекции по методам вычислений. – Санкт-Петербург: Изда-тельство Санкт-Петербургского университета, 1998.**
9. **Ортега Дж., Рейнболдт В. Итерационные методы решения нелинейных си-стем уравнений со многими неизвестными. – Москва: Мир, 1975.**
10. **Островский А.М. Решение уравнений и систем уравнений.**

**– Москва: Издательство иностранной литературы, 1963.**

1. **Самарский А.А., Гулин А.В. Численные методы. – Москва: Наука, 1989.**
2. **Семёнов А.Л., Важев И.В., Кашеварова Т.П. и др. Интервальные мето-ды распространения ограничений и их приложения // Системная информати-ка. – Новосибирск: Издательство СО РАН, 2004. – Вып. 9. С. 245–358.**
3. **Трауб Дж. Итерационные методы решения уравнений. – Москва: Мир, 1985.**
4. **Тыртышников Е.Е. Методы численного анализа. – Москва: Академия, 2007.**
5. **Успенский В.А., Семёнов А.Л. Теория алгоритмов: основные открытия и приложения. – Москва: Наука, 1987.**
6. **Фихтенгольц Г.М. Курс дифференциального и интегрального исчисления. Т. 1. Москва: Наука, 1966.**

**304** 4. Решение нелинейных уравнений и их систем

1. **Холодниок М., Клич А., Кубичек М., Марек М. Методы анализа нели-нейных динамических моделей. – Москва: Мир, 1991.**
2. **Шарый С.П. Конечномерный интервальный анализ. – Электронная книга, 2010 (см.** http://www.nsc.ru/interval/Library/InteBooks**)**
3. **Шилов Г.Е. Математический анализ. Функции одного переменного. Ч. 1–2. –**

**Москва: Наука, 1969.**

1. **Aberth O. Precise numerical methods using C++. – San Diego: Academic Press, 1998.**
2. **Akyildiz Y., Al-Suwaiyel M.I. No pathologies for interval Newton’s method //**

**Interval Computations. – 1993. – No. 1. – P. 60–72.**

1. **Hansen E., Walster G.B. Global optimization using interval analysis. – New York: Marcel Dekker, 2003.**
2. **Kearfott R.B. Rigorous global search: Continuous problems. – Dordrecht: Kluwer, 1996.**
3. **Kelley C.T. Iterative methods for linear and nonlinear equations. – Philadelphia: SIAM, 1995.**
4. **Kreinovich V., Lakeyev A.V., Rohn J., Kahl P. Computational complexity and feasibility of data processing and interval computations. – Dordrecht: Kluwer, 1997.**
5. **Miranda C. Un’ osservatione su un teorema di Brouwer // Bollet. Unione Mat. Ital. Serie II. – 1940. – Т. 3. – С. 5–7.**
6. **Moore R.E., Kearfott R.B., Cloud M. Introduction to interval analysis. –**

**Philadelphia: SIAM, 2009.**

1. **Neumaier A. Interval methods for systems of equations. – Cambridge: Cambridge University Press, 1990.**
2. **Trefethen L.N. Pseudospectra of linear operators // SIAM Review. 1997. – Vol. 39, No. 3. – P. 383–406.**
3. **Trefethen L.N., Bau D. III Numerical linear algebra. – Philadelphia: SIAM, 1997.**

Дополнительная

1. **Бабенко К.И. Основы численного анализа. – Москва: Наука, 1986.**
2. **Загускин В.Л. Справочник по численным методам решения алгебраических и трансцендентных уравнений. – Москва: Физматгиз, 1960.**
3. http://physics.nist.gov/cuu/Constants/index.html
4. http://web.comlab.ox.ac.uk/projects/pseudospectra/
5. **Scilab The Free Platform for Numerical Computation.** http://www.scilab.org

Обозначения

⇒**логическая импликация**

**логическая равносильность**

* **логическая конъюнкция, связка ¾и¿**
* **отображение множеств**

7→ **правило сопоставления элементов при отображении**

* **оператор присваивания в алгоритмах**
* **знак композиции отображений**

∅**пустое множество**

x ∈ X **элемент** x **принадлежит множеству** X

x 6 X **элемент** x **не принадлежит множеству** X

X ∪ Y **объединение множеств** X **и** Y

X ∩ Y **пересечение множеств** X **и** Y

X \ Y **разность множеств** X **и** Y

X ⊆ Y **множество** X **есть подмножество множества** Y

X × Y **прямое декартово произведение множеств** X **и** Y

* **множество натуральных чисел**
* **множество вещественных (действительных) чисел**

R***+*** **множество неотрицательных вещественных чисел**

* **множество комплексных чисел**

IR **классическая интервальная арифметика**

R**n** **множество вещественных** n**-мерных векторов**

C**n** **множество комплексных** n**-векторов**

**305**

|  |  |
| --- | --- |
| **306** | Обозначения |
| IR**n** | **множество** n**-мерных интервальных векторов** |
| R**m**×**n** | **множество вещественных** m×n**-матриц** |
| IR**m**×**n** | **множество интервальных** m×n**-матриц** |

* **комплексно сопряжённое к числу** z∈C

|  |  |  |  |
| --- | --- | --- | --- |
| sgn a | | **знак числа** a∈R |  |
| a**,** inf a | | **левый конец интервала** a |  |
|  | **,** supa | **правый конец интервала** a |  |
| a |  |
| mid a | | **середина интервала** a |  |
| wid a | | **ширина интервала** a |  |
| dist | | **метрика (расстояние)** |  |
| Dist | | **мультиметрика (векторнозначное расстие)** |  |
| dom f | | **область определения функции** f |  |
| ran (f, X) | | **область значений функции** f **на** X |  |
| int X | | **топологическая внутренность множества** X |  |
| cl X | | **топологическое замыкание множества** X |  |
| f ∠( · ) | | **разделённая разность от функции** f |  |
| δ**ij** | | **символ Кронекера, 1 при** i=j **и 0 иначе** |  |
| i | | **мнимая единица** |  |
| I | | **единичная матрица соответствующих размеров** |  |
| k · k | | **векторная или матричная норма** |  |
| A> | | **матрица, транспонированная к матрице** A |  |
| A−***1*** | | **матрица, обратная к матрице** A |  |
| ρ(A) | | **спектральный радиус матрицы** A |  |
| λ(A)**,** λ**i**(A) | | **собственные числа матрицы** A |  |
| σ(A)**,** σ**i**(A) | | **сингулярные числа матрицы** A |  |
| cond(A) | | **число обусловленности марицы** A |  |

diag {z***1***, . . . , z**n**} **диагональная** n × n**-матрица**

**с элементами** z***1***, . . . , z**n по главной диагонали**

**Если** x **вектор, то его подвектор, состоящий из компонент** x**k с ин-дексами** k **из некоторого индексного подмножества** K **обозначается че-рез** x**K , а дополнительный к нему вектор через** x6 **K или** x***=***6**k, если**

|  |  |
| --- | --- |
| Обозначения | **307** |

K = {k}**. Аналогичных соглашений будем придерживаться и в отноше-нии матриц, так что, к примеру, если** A **является** m×n**-матрицей, то** A***:*,**6***=*k** **это матрица размера** m × (n − 1)**, полученная из** A **удалением** k**-го столбца.**

**Интервалы и другие интервальные величины (векторы, матрицы и др.) всюду в тексте обозначаются жирным математическим шриф-том, например,** A**,** B**,** C**, . . . ,** x**,** y**,** z**, тогда как неинтервальные (то-чечные) величины никак специально не выделяются. Арифметические операции с интервальными величинами это операции классической интервальной арифметики** IR **(см. §1.3). Наконец, если не оговорено противное, под векторами (точечными или интервальными) всюду по-нимаются вектор-столбцы.**

**Конец доказательства теоремы или предложения и конец примера выделяются в тексте стандартным знаком ¾ ¿.**

**Значительная часть описываемых в книге алгоритмов снабжает-ся псевдокодами на неформальном алгоритмическом языке, в котором операторные скобки**

DO FOR **. . .** END DO **означают оператор цикла со счётчиком,** **который задаётся после** FOR**,**

DO WHILE **. . .** END DO **означают оператор цикла с предусловием,** **стоящим после** WHILE**,**

IF **. . .** THEN **. . .** ELSE **. . .** END **или** IF **. . .** THEN **. . .** END **означают** **условные операторы с условием, стоящим после** IF**.**

Краткий биографический словарь

**Абель, Нильс Хенрик (Niels Henrik Abel, 1802–1829) норвежский математик.**

**Адамар, Жак Саломон (Jacques Salomon Hadamard, 1865–1963) французский математик.**

**Андронов, Александр Александрович (1901–1952) советский физик и механик.**

**Бабенко, Константин Иванович (1919–1987) советский математик и механик.**

**Банах, Стефан (Stefan Bahach, 1892–1945) польский математик.**

**Бернштейн, Сергей Натанович (1880–1968) российский и советский математик.**

**Больцано, Бернард (Bernard Bolzano, 1781–1848) чешский теолог, философ и математик.**

**Брадис, Владимир Модестович (1890–1975)**

**русский и советский математик и педагог.**

**Брауэр, Лёйтзен Эгберт Ян (Luitzen Egbertus Jan Brouwer, 1881–1966) голландский математик.**

**Валлис, Джон (John Wallis, 1616–1703) английский математик.**

**Вандермонд, Александр Теофиль (Alexandre Theophill Vandermonde, 1735–1796) французский музыкант и математик.**

**308**

Обозначения **309**

**Вейерштрасс, Карл Теодор (Karl Theodor Weierstrass, 1815–1897) немецкий математик.**

**Виет, Франсуа (Fran¸cois Vi`ete, 1540–1603) французский математик.**

**Гаусс, Карл Фридрих (Carl Friedrich Gauss, 1777–1855)**

**немецкий математик, внёсший также фундаментальный вклад в численные методы, астрономию и геодезию.**

**Гельфанд, Израиль Моисеевич (1913–2009)**

**советский математик. С 1989 года жил и работал в США.**

**Гершгорин, Семён Аронович (1901–1933)**

**советский математик, живший и работавший в Ленинграде.**

**Гивенс, Джеймс Уоллес (James Wallace Givens, 1910–1993) американский математик.**

**Гильберт, Давид (David Hilbert, 1862–1943) немецкий математик.**

**Грам, Йорген Педерсен (Jorgen Pedersen Gram, 1850–1916) датский математик.**

**Евклид, или Эвклид (др.-греч.** Eυκλειδης**, ок. 300 г. до н. э.) древнегреческий математик.**

**Зейдель, Филипп Людвиг (Philipp Ludwig Seidel, 1821–1896) немецкий астроном и математик.**

**Канторович, Леонид Витальевич (1912–1986)**

**советский математик и экономист, один из создателей теории линейного программирования.**

**Котес, Роджер (Roger Cotes, 1682–1716) английский математик.**

**Коши, Огюстен Луи (Augustin Louis Cauchy, 1789–1857) французский математик.**

**Кравчик, Рудольф (Krawczyk, Rudolf) немецкий математик.**

**Крамер, Габриэль (Gabriel Cramer, 1704–1752) швейцарский математик.**

**Красносельский, Марк Александрович (1920–1997) советский и российский математик.**

**310** Обозначения

**Крейн, Селим Григорьевич (1917–1999)**

**советский и российский математик.**

**Кронекер, Леопольд (Leopold Kronecker, 1823–1891) немецкий математик.**

**Крылов, Алексей Николаевич (1863–1945)**

**русский и советский математик, механик и кораблестроитель.**

**Кублановская, Вера Николаевна (род. 1920) советский и российский математик.**

**Кузьмин, Родион Осиевич (1891–1949) русский и советский математик.**

**Лагранж, Жозеф Луи (Joseph Louis Lagrange, 1736–1813) французский математик и механик.**

**Лежандр, Адриен Мари (Adrien Marie Legendre, 1752–1833) французский математик и механик.**

**Лейбниц, Готфрид Вильгельм (Gottfried Wilhelm Leibnitz, 1646–1716) немецкий философ, математик и физик, один из создателей дифференциального и интегрального исчисления.**

**Липшиц, Рудольф (Rudolf Lipschitz, 1832–1903) немецкий математик.**

**Лобачевский, Николай Иванович (1792–1856)**

**русский математик, создатель неевклидовой геометрии.**

**Локуциевский, Олег Вячеславович (1922–1990) советский математик.**

**Ляпунов, Александр Михайлович (1857–1918)**

**русский математик и механик, основоположник математической теории устойчивости.**

**Марков, Андрей Андреевич (1856–1922) русский математик.**

**Марцинкевич, Юзеф (J´ozef Marcinkiewicz, 1910–1941) польский математик.**

**Миранда, Карло (Carlo Miranda, 1912–1982) итальянский математик.**

**Нейман, Карл Готфрид (Karl Gottfried Neumann, 1832–1925)**

|  |  |
| --- | --- |
| Обозначения | **311** |

**немецкий математик.**4

**Ньютон, Исаак (Isaac Newton, 1643–1727)**

**английский физик и математик, заложивший основы дифференциального и интегрального исчисления и механики.**

**Пойа (Полиа), Дъёрдь (иногда Джордж) (Gy¨orgy Polya, 1887–1985) венгерский и американский математик.**

**Ричардсон, Льюис Фрай (Lewis Fry Richardson, 1881–1953) английский математик, физик и метеоролог.**

**Родриг, Бенжамен Оленд (Benjamin Olinde Rodrigues, 1795–1851) французский математик и банкир.**

**Рунге, Карл Давид (Karl David Runge, 1856–1927) немецкий физик и математик.**

**Руффини, Паоло (Paolo Ruﬃni, 1765–1822) итальянский математик.**

**Симпсон, Томас (Thomas Simpson, 1710–1761) английский математик.**

**Стеклов, Владимир Андреевич (1863–1926) русский математик и механик.**

**Стирлинг, Джеймс (James Stirling, 1692–1770) шотландский математик.**

**Таусски, Ольга (Olga Tausski, 1906–1995) американский математик.**

**Тейлор, Брук (Brook Taylor, 1685–1731) английский математик.**

**Фабер, Георг (Georg Faber, 1877–1966) немецкий математик.**

**Фаддеев, Дмитрий Константинович (1907–1989) советский математик.**

**Фаддеева, Вера Николаевна (1906–1983) советский математик.**

4 **Не следует путать его с Джоном фон Нейманом (John von Neumann, 1903– 1957), американским математиком венгерского происхождения, разработчиком пер-вых ЭЦВМ, именем которого назван также спектральный признак устойчивости разностных схем.**

**312** Обозначения

**Фробениус, Фердинанд Георг (Ferdinand Georg Frobenius, 1849–1917) немецкий математик.**

**Фрэнсис, Джон (John G.F. Francis, род. 1934) английский математик и программист.**

**Хаусдорф, Феликс (Felix Hausdorﬀ, 1868–1942) немецкий математик.**

**Хаусхолдер, Элстон (Alston Scott Householder, 1904–1993) американский математик.**

**Хессенберг, Карл Адольф (Karl Adolf Hessenberg, 1904–1959) немецкий математик и инженер.**

**Холесский, Андре-Луи (Andr´e-Louis Cholesky, 1875–1918) французский геодезист и математик.**5

**Хопф, Хайнц (Heinz Hopf, 1896–1971)**

**немецкий и швейцарский математик.**

**Чебышёв, Пафнутий Львович (1821–1894)**

**русский математик и механик, внёсший основополагающий вклад в теорию приближений и теорию вероятностей.**

**Шрёдер, Иоганн (Johann Schr¨oder, 1925–2007) немецкий математик.**

**Эйлер, Леонард (Leonhard Euler, 1707–1783)**

**швейцарский и российский математик, механик и физик. Работе в России посвятил 31 год жизни, умер и похоронен в Петербурге.**

**Эрмит, Шарль (Charles Hermite, 1822–1901) французский математик.**

**Як´оби, Карл Густав (Carl Gustav Jacobi, 1804–1851) немецкий математик.**

5 **В русской научной литературе его фамилия нередко транслитерируется как ¾Холецкий¿ или даже ¾Халецкий¿.**

Предметный указатель

ε**-решения, 275** **формулы численного**

**алгебраическая степень точности,** **дифференцирования,**

**104** **62, 65**

**биортогональность, 235** **характеризация Бека, 287**

**целевая функция, 219** **характеризация Оеттли-Прагера,**

**число обусловленности, 146** **288**

**дефект сплайна, 52** **хессенбергова форма, 262**

**диагональное преобладание, 125** **интерполирование, 24**

**дифференцирование** **интерполяция эрмитова, 47**

**автоматическое, 61** **интервальная арифметика, 13**

**дифференцирование численное,** **интервальная арифметика**

**61** **классическая, 14**

**дифференцирование символьное,** **интервальное расширение, 15**

**61** **итерационные методы, 154**

**доминирующее собственное** **коэффициент чувствительности,**

**значение, 242** **11**

**доминирующий собственный** **коэффициенты Фурье, 80**

**вектор, 242** **коэффициенты перекоса, 238**

**экстраполяция, 40** **комплексификация, 141**

**эквивалентные нормы, 137** **конечные методы, 154**

**энергетическая норма, 129** **круги Гершгорина, 241**

**эрмитова интерполяция, 47** **линейная задача о наименьших**

**формула Маркова, 112** **квадратах, 228**

**формула Родрига, 82** **матрица Гильберта, 81, 150**

**формула Симпсона, 92** **матрица Грама, 79**

**формула кубатурная, 88** **матрица Липшица интервальная,**

**формула квадратурная, 88** **294**

**формула парабол, 92** **матрица Уилкинсона, 239**

**формула прямоугольников, 89** **матрица Вандермонда, 27, 150**

**формула трапеций, 91** **матрица неразложимая, 126**

**формулы Гаусса, 104** **матрица отражения, 172**

**формулы Ньютона-Котеса, 89** **матрица перестановки, 167**

**313**

**314** Предметный указатель

**матрица почти треугольная, 262** **251**

**матрица предобуславливающая,** **оператор Кравчика, 297**

**196** **оператор Ньютона интервальный,**

**матрица разложимая, 126** **292**

**матрица скалярная, 198** **ортогонализация Грама-Шмидта,**

**матрица строго нижняя** **81, 180**

**треугольная, 203** **основная теорема интервальной**

**матрица строго верхняя** **арифметики, 16**

**треугольная, 203** **почти решения, 275**

**матрица транспозиции, 166** **полином интерполяционный, 27**

**матрица трёхдиагональная, 184** **полином интерполяционный**

**матрица вращения, 178** **Лагранжа, 30**

**матричная норма, 131** **полином интерполяционный**

**матричный ряд Неймана, 144** **Ньютона, 37**

**мера диагонального** **полиномы Чебышёва, 41**

**преобладания, 208** **полиномы Лежандра, 82**

**метод Гаусса, 157** **порядок точности формулы, 67,**

**метод Гаусса-Зейделя, 205** **102**

**метод Хаусхолдера, 175** **правило Рунге, 113**

**метод Холесского, 184** **предобуславливание, 196**

**метод Шульца, 227** **предобуславливатель, 196**

**метод Якоби, 202** **пример Рунге, 50**

**метод градиентного спуска, 219** **принцип релаксации, 209**

**метод квадратного корня, 184** **принцип вариационный, 217**

**метод минимальных невязок, 226** **признак Адамара, 125**

**метод наискорейшего спуска, 223** **прямые методы, 154**

**метод отражений, 175** **расщепление матрицы, 197**

**метод прогонки, 186** **разделённая разность, 30**

**метод простой итерации, 198** **разложение Холесского, 180**

**метод релаксации, 211** **разложение сингулярное, 123**

**множитель Холесского, 180** **разностные уравнения**

**мультиметрика, 281** **трёхточечные, 185**

**непрерывность по Липшицу, 12** **рекуррентный вид системы, 195,**

**нестационарный итерационный** **270**

**процесс, 190** **рекуррентный вид уравнения, 270**

**невязка, 209** **ряд Фурье, 80**

**норма, 127** **сдвиг спектра, 250**

**норма энергетическая, 129** **схема единственного деления, 158**

**нормальная система уравнений,** **сингулярные числа, 121**

**228** **сингулярные векторы, 121**

**обобщённая степень, 37** **след матрицы, 255**

**обратные степенные итерации,** **собственный вектор, 229**

Предметный указатель

**собственное значение, 229 спектральная норма, 136 спектральный радиус, 140 сплайн, 52 стационарный итерационный**

**процесс, 190 степенной метод, 245 степень сплайна, 52 субдистрибутивность, 14 сжатие, 280 сжимающее отображение, 280 шаблон, 65**

**теорема Абеля-Руффини, 231 теорема Банаха о неподвижной точке, 280**

**теорема Бауэра-Файка, 233 теорема Больцано-Коши, 278 теорема Брауэра о неподвидной**

**точке, 280 теорема Фабера, 51 теорема Марцинкевича, 52 теорема Миранды, 278 теорема Стеклова-Пойа, 99**

**теорема Шрёдера о неподвижной точке, 281**

**теорема Таусски, 127 теорема Вейерштрасса, 49 теорема о сингулярном**

**разложении, 123 тригонометрические полиномы,**

**26 трёхдиагональная матрица, 58 узлы сплайна, 52 ведущий элемент, 164 векторная норма, 127 вырожденный интервал, 13 задача некорректная, 10, 73**

**задача о наименьших квадратах линейная, 228**

**задача приближения функции, 75 задача сглаживания, 81 задача вычислительно**

**315**

**корректная, 272** P**-сжатие, 281**

**LU-разложение, 161**

**QR-алгоритм, 259, 262**

**QR-разложение, 170**