

Исаков, В.Б.

Элементы численных методов: Учебное пособие для студентов, обучающихся по специальности Математика группы Педагогические специал. - М.: Академия, 2003.- 192 с. : ил.

ISBN 5-7695-0795-0, 20000 экз.

В пособии подробно излагается введение в теорию погрешностей и исследуется ряд несложных методов приближенного решения нелинейных уравнений, систем линейных и нелинейных уравнений, аналитического приближения табличных функций, численного интегрирования и дифференцирования, численного решения обыкновенных дифференциальных уравнений. Приводятся примеры и задачи качественного характера, задания для лабораторного практикума. Для студентов высших педагогических учебных заведений. Может быть полезно преподавателям и студентам средних профессиональных учебных заведений и учащимся средних школ с углубленным изучением математики.

Численные методы решения уравнений

ББК 22.193я73

УДК 519.6(075.8)

Оглавление

Основные обозначения - 6с.

Предисловие - 7-9с.

Глава 1 Введение в теорию погрешностей - 10-40с.

1.1 Основные источники погрешностей - 10-13с.

1.2 Расстояние - 13-17с.

1.2.1 Понятие о расстоянии - 13-14с.

1.2.2 Расстояние между числами - 14с.

1.2.3 Расстояние между векторами - 14-15с.

1.2.4 Расстояние между функциями и их значениями - 15-17с.

1.3 Абсолютная погрешность - 18-21с.

1.4 Десятичная запись приближенных чисел: основные понятия и правила - 21-28с.

1.4.1 Округление чисел - 21-22с.

1.4.2 Верные значащие цифры - 22-24с.

1.4.3 Правило записи приближенных чисел - 24-25с.

1.4.4 Правила записи знаков точного и приближенного равенств - 25-27с.

1.5 Относительная погрешность приближенных чисел - 28-29с.

1.6 Оценка точности приближенных векторов и функций - 29-32с.

1.6.1 Абсолютная погрешность приближенных векторов - 29с.

1.6.2 Абсолютная погрешность приближенных функций и их значений - 29-32с.

1.7 Оценка влияния погрешностей аргументов на значение функции - 32-38с.

1.7.1 Постановка задачи - 32-33с.

1.7.2 Оценка погрешности значения дифференцируемой функции - 33-36с.

1.8 Оценка погрешностей арифметических операций - 38-40с.

Глава 2 Приближенное решение уравнений - 41-75с.

2.1 Постановка задачи - 41-43с.

2.2 Отделение корней - 43-46с.

2.3 Метод половинного деления - 46-49с.

2.4 Метод хорд и метод касательных (общие вопросы) - 49-51с.

2.5 Метод хорд - 51-59с.

2.5.1 Рекуррентная формула - 51-53с.

- 2.5.2 Сходимость итерационной последовательности - 53-55с.
- 2.5.3 Оценка погрешностей приближений - 55-59с.
- 2.6 Метод касательных - 59-64с.
- 2.6.1 Рекуррентная формула - 59-61с.
- 2.6.2 Сходимость итерационной последовательности - 61-62с.
- 2.6.3 Оценка погрешностей приближений - 62-64с.
- 2.7 Комбинированный метод хорд и касательных - 64-66с.
- 2.8 Метод простой итерации - 66-75с.
- 2.8.1 Рекуррентная формула - 67-68с.
- 2.8.2 Сходимость итерационной последовательности - 68-71с.
- 2.8.3 Приведение уравнений к виду, пригодному для метода простой итерации - 71-73с.
- 2.8.4 Оценка погрешностей приближений - 74-75с.
- Глава 3 Аналитическое приближение табличных функций - 76-103с.**
- 3.1 Основные понятия - 76-79с.
- 3.1.1 Табличная функция - 76-77с.
- 3.1.2 Задача аналитического приближения табличных функций - 78-79с.
- 3.2 Интерполирование табличных функций - 79-81с.
- 3.2.1 Постановка задачи - 79-80с.
- 3.2.2 Полиномиальное интерполирование - 80-81с.
- 3.3 Оценка погрешности полиномиальной интерполяции - 82-84с.
- 3.4 Интерполяционный многочлен Лагранжа - 84-86с.
- 3.5 Интерполяционные многочлены Ньютона - 86-91с.
- 3.5.1 Конечные разности - 86-87с.
- 3.5.2 Первый интерполяционный многочлен Ньютона - 87-90с.
- 3.5.3 Второй интерполяционный многочлен Ньютона - 90-91с.
- 3.6 Линейное интерполирование - 91-97с.
- 3.6.1 Формула линейного интерполирования - 92-95с.
- 3.6.2 Обратное линейное интерполирование - 96-97с.
- 3.7 Приближение табличных функций по методу наименьших квадратов - 98-103с.
- 3.7.1 Постановка задачи - 98-101с.
- 3.7.2 Полиномиальное приближение по методу наименьших квадратов - 10-103с.

Глава 4 Численное интегрирование и дифференцирование - 104-129с.

4.1 Задача приближенного вычисления определенных интегралов - 104-108с.

4.2 Формулы прямоугольников - 108-114с.

4.2.1 Вывод формул - 108-110с.

4.2.2 Оценка погрешностей - 110-114с.

4.3 Формула трапеций - 114-18с.

4.3.1 Вывод формулы - 114-116с.

4.3.2 Оценка погрешностей - 116-118с.

4.4 Формула Симпсона - 118-120с.

4.5 Учет погрешностей квадратурных формул методом двойного пересчета - 120-123с.

4.6 Численное дифференцирование - 123-129с.

4.6.1 Постановка задачи - 123-126с.

4.6.2 Формулы численного дифференцирования на основе интерполяционного многочлена Ньютона - 126-129с.

Глава 5 Численное решение обыкновенных дифференциальных уравнений - 130-146с.

5.1 Необходимые сведения о дифференциальных уравнениях первого порядка - 130-133с.

5.2 Понятие численного решения задачи Коши - 133-134с.

5.3 Метод Эйлера - 135-137с.

5.4 Усовершенствования метода Эйлера - 137-141с.

5.4.1 Метод Эйлера-Коши - 138-140с.

5.4.2 Метод срединных точек - 140-141с.

5.5 Точность метода Эйлера и его модификаций - 141-144с.

5.6 Решение систем дифференциальных уравнений первого порядка - 144-145с.

5.7 Численное решение дифференциальных уравнений высших порядков - 145-146с.

Глава 6 Сжимающие отображения и метод итерации - 147-171с.

6.1 Основные определения - 147-152с.

6.2 Принцип сжимающих отображений - 152-156с.

6.3 Метод итерации для уравнений с одним неизвестным - 156-158с.

6.4 Метод итерации для систем линейных алгебраических уравнений - 158-165с.

6.4.1 Построение итерационной последовательности - 158-161с.

6.4.2 Достаточное условие сходимости итерационной последовательности - 161-165с.

6.5 Метод итерации для систем нелинейных уравнений - 165-171с.

Приложение. Лабораторные работы - 172-184с.

Литература - 185-186с.

Предметный указатель - 187-189с.

Основные обозначения

$\stackrel{def}{\equiv}$ — равно по определению.

\mathbf{N} — множество всех натуральных чисел.

$\mathbf{R} \stackrel{def}{=} (-\infty; +\infty)$ — множество всех действительных чисел.

\mathbf{R}^n — множество всех действительных n -мерных числовых векторов.

ρ — расстояние (метрика).

$\rho(x, y)$ — расстояние между элементами x и y некоторого метрического пространства X .

$\rho_1, \rho_2, \rho_\infty$ — метрики на \mathbf{R}^n .

$\mathbf{R}_1^n, \mathbf{R}_2^n, \mathbf{R}_\infty^n$ — метрические пространства, полученные введением на множестве \mathbf{R}^n метрик $\rho_1, \rho_2, \rho_\infty$ соответственно.

$A \approx a$ — A приближенно равняется a (a является приближением к A).

Δ_a — абсолютная погрешность приближенного значения a .

δ_a — относительная погрешность приближенного значения a .

○ — начало доказательств теорем.

● — конец доказательств и примеров.

Предисловие

Курс численных методов является важной составной частью математической подготовки студентов педагогических специальностей и направлений. Его значение в настоящее время определяется не только увеличивающимися возможностями применения методов вычислительной математики в вузовском учебном процессе, но и проникновением численных алгоритмов приближенного решения задач в среднее (прежде всего инновационное) образование, т.е. в сферу профессиональной деятельности учителя. При этом намечающийся переход средней школы к 12-летнему обучению с выделением физико-математического направления в старших классах даст возможность существенно расширить эту сферу.

Ввиду того что разумное применение и квалифицированное преподавание методов приближенного численного анализа затруднительны без основательной подготовки, будущему учителю математики, физики или информатики следует глубоко вникать в суть изучаемых методов приближений и оценок погрешностей, знать их обоснование и соответствующий математический инструментарий. Здесь уместно отметить, что освоение и использование таких распространенных компьютерных пакетов прикладной математики, как, например, MathCad и Maple, также требуют солидной теоретической базы.

Между тем теоретическая составляющая курса численных методов в педагогических вузах чаще всего невелика. Вследствие недостатка аудиторного времени на полноценное изучение обязательного учебного материала преподаватель вынужден ориентироваться в основном только на постановочные и рецептурные вопросы, что особенно характерно при обучении одновременно по двум специальностям и тем более на вечернем и заочном отделениях. Для решения проблемы фундаментализации курса в таких условиях нужна доступная и понятная для студентов литература, содержащая выводы всех методов приближений и формул оценок их погрешностей, анализ условий их применения, доказательства сходимости итерационных процессов и, что очень важно, упражнения, побуждающие к теоретическим микроисследованиям в этих областях.

В учебном пособии излагаются теория погрешностей и основные разделы программ курса численных методов, предусмотренных государственными образовательными стандартами высшего профессионального образования по педагогическим специальностям

«Математика», «Физика» и «Информатика», при этом акцент сделан именно на теоретическую сторону рассматриваемых вопросов. Одним из основных побудительных мотивов для написания работы явилось стремление показать значение математического анализа как фундамента вычислительной математики, взаимопроникновение и взаимосвязь численных методов и методов математического анализа. Вместе с тем чтобы эта цель не оказалась затененной громоздкими выкладками, содержание пособия ограничено несложными и традиционными для педвузовского курса методами приближений, которые либо применяются для решения наиболее часто встречающихся задач из математического анализа, либо опираются на средства этой дисциплины.

Изложение ведется достаточно строго, но без излишней формализации. Для удобства читателя используемые дополнительные факты и теоремы приведены в соответствующих местах со ссылкой на источники, откуда они взяты. Необходимый для понимания содержания книги уровень математической подготовки не выходит за пределы обозначенного в стандартах минимума знаний. Лишь при изучении гл. 6, посвященной применению в численных методах теоремы Банаха, потребуются начальные сведения о метрических пространствах и сжимающих отображениях, которые обычно не рассматриваются в курсе математики для специальностей «Физика» и «Информатика». В связи с данным обстоятельством глава начинается с подробного введения в эту тематику.

Содержание подкреплено значительным количеством примеров и упражнений, среди которых преобладают задачи качественного характера, нацеленные на активизацию самостоятельной работы студентов по изучению теории. Упражнения являются непосредственным продолжением основного материала и полезным (а часто и необходимым) дополнением к нему. Они могут пригодиться в системе практических занятий и курсовых работ. Часть задач отмечена звездочкой. Это означает, что при их решении надо проявить некоторую изобретательность. Поскольку теоретические задачи требуют самостоятельных исследований, а вычислительные задачи просты, ответы к ним не приводятся. В то же время к ряду упражнений даны краткие указания по их выполнению.

В приложении приведены задания для лабораторного практикума.

Пособие не претендует на полноту освещения всего изучаемого и используемого в педагогических вузах аппарата приближенных вычислений. В связи с этим полезно иметь в виду то, что некоторые не вошедшие в него методы достаточно просто и понятно изложены в руководствах [5, 8, 13, 15]. Для пополнения знаний по различным разделам вычислительной математики можно воспользоваться также книгами [1, 2, 6, 7, 10, 18, 20, 21] и другой учебной литературой для университетов и технических вузов. Желаящим иметь

готовые алгоритмы и компьютерные программы реализации численных методов рекомендуются работы [8, 19, 21, 24 – 26].

Выражаю искреннюю признательность профессору В. Н. Алексюку за поддержку и ценные советы по отдельным темам пособия, доценту В. А. Попову, который внимательно прочел рукопись книги и оказал существенную помощь в ее улучшении, а также рецензентам за их труд по оценке работы и ряд полезных замечаний.

Глава 1. ВВЕДЕНИЕ В ТЕОРИЮ ПОГРЕШНОСТЕЙ

1.1. Основные источники погрешностей

Численные методы в настоящее время относятся к основным методам решения задач математики и различных ее приложений. Они характеризуются тем, что сводят процесс решения математической задачи к некоторой конечной последовательности операций над числами и приводят к результатам, представленным в виде чисел, числовых векторов и матриц, числовых таблиц и т. п. Их значение возрастает параллельно с развитием вычислительной техники.

В то же время полученные численными методами результаты обычно содержат погрешности, являясь лишь приближениями к истинным ответам. Вызвано это рядом объективных причин, среди которых есть не связанные непосредственно с методами вычислений. Чтобы разобраться в них, проанализируем основные этапы математического решения прикладных задач, а именно:

1. Построение математической модели задачи.
2. Определение исходных данных.
3. Решение полученной математической задачи.

Погрешности появляются уже на первом этапе, ибо *математическая модель* задачи — это приближенное, идеализированное описание задачи на языке математики. При моделировании объекты и процессы задачи-оригинала, взаимосвязи между ее параметрами заменяются на математические понятия и соотношения. Ради того чтобы получаемая в итоге математическая задача оказалась доступной для дальнейших исследований, учитывают лишь наиболее важные параметры, условия и особенности исходной задачи. Понятно, что чем меньше факторов отбрасывается, тем точнее получается модель.

Несмотря на приближенность результатов математического моделирования, без него в приложениях математики не обойтись. Оно представляет собой обязательную ступень при переходе от нематематической задачи к математической. Более того, удовлетворительное исследование многих явлений реального мира оказывается возможным лишь тогда, когда удается построить их математические модели.

Следующей причиной появления погрешностей является то, что установить точные значения исходных параметров во многих случаях невозможно. Серьезные проблемы с этим возникают не только при исследовании, например, космических объектов или земной атмосферы, когда для определения исходных данных приходится прибегать к различным прикидкам и сложным измерительным процедурам, но и при решении достаточно простых бытовых задач.

Проиллюстрируем сказанное на примере.

Пример 1.1. Пусть требуется найти объем некоторого предмета, имеющего форму прямого кругового цилиндра (цистерна, бак и т.п.). Если не учитывать обычно имеющие место шероховатости и вмятины на поверхности, закругления на стыке между дном и стенкой, а также другие возможные отклонения такого предмета от идеального прямого кругового цилиндра, его объем можно вычислить по формуле $v = \pi r^2 h$. Ясно, что результат будет приближенным. Во-первых, здесь реальный предмет заменен его моделью — математическим цилиндром. Во-вторых, значения подставляемых в эту формулу числовых параметров могут быть только приближенными. При любом способе измерений радиуса основания r и высоты h предмета получим их значения с погрешностями, зависящими прежде всего от точности измерительных приборов. Иррациональное число $\pi = 3,141592653\dots$ обычно заменяют приближенным значением, погрешность которого зависит от количества сохраняемых при округлении цифр. •

После того как математическая модель построена и определены исходные данные, необходимо подобрать метод решения полученной математической задачи. Круг математических методов условно подразделяется на аналитические, численные и графические методы. Нас интересуют численные методы, которые, в свою очередь, делятся на точные и приближенные.

Численный метод называется *точным*, если он дает принципиальную возможность после выполнения конечного числа операций над точными числами получить точное решение задачи. К таким методам относится, например, алгоритм решения квадратного уравнения. Если все коэффициенты уравнения $x^2 + 2x - 3 = 0$ не содержат погрешностей, алгоритм приведет к точным корням $x_1 = -3$, $x_2 = 1$. Используемая в примере 1.1 формула вычислений также не порождает сама по себе погрешностей. Подставив в нее точные значения $r = 2$, $h = 4,5$, получим точное иррациональное число $v = 18\pi$.

Для большинства реальных задач точных методов решения вообще не существует, а если они и имеются, то бывают настолько трудоемкими, что не представляют практического интереса. Кроме того, они часто сопряжены с бесконечными вычислительными процессами. Например, поиск точного значения функции может свестись к нахождению суммы числового ряда, что в общем случае практически осуществить невозможно. Вследствие этого основным инструмен-

том вычислительной математики являются *приближенные* численные методы, приводящие обычно к приближенным результатам даже при точных исходных данных и точных вычислениях. Возникают так называемые *погрешности метода*.

Именно такие методы излагаются в последующих главах книги.

В процессе реализации численных методов приходится выполнять арифметические операции над приближенными числами, вычислять значения функций, а также округлять исходные данные, промежуточные и окончательные результаты. При этом оказывается, что погрешность результата любого арифметического действия, как правило, превышает погрешности исходных данных. Известно также, что значения корней $\sqrt[n]{x}$ и трансцендентных функций ($\sin x$, $\operatorname{tg} x$, e^x и т. п.) чаще всего удастся найти только приближенно с помощью математических таблиц, компьютерных средств вычислений или приближенных формул вида

$$\sin x \approx x - \frac{x^3}{3!} + \frac{x^5}{5!}.$$

Перечисленные погрешности называются *вычислительными*.

Выше было сказано, что точные численные методы должны давать «принципиальную возможность» получения точных решений. На самом деле эта возможность часто не реализуется из-за вычислительных погрешностей. Например, чтобы найти и записать в десятичной форме корни уравнения $x^2 + x - 1 = 0$, надо подставить в соответствующие формулы округленное значение числа $\sqrt{5}$. Понятно, что тогда получатся приближенные значения корней.

Наконец, определенное влияние на результаты оказывает степень точности используемых средств вычислений. Для представления чисел в них выделяется ограниченное количество десятичных разрядов, что влечет автоматическое округление числовых данных. Чем больше разрядная сетка, тем погрешность округлений меньше и, следовательно, с тем большей точностью можно производить вычисления.

Подытоживая все сказанное выше, отмечаем, что *основными источниками погрешностей* являются:

- замена реальной задачи математической моделью;
- затруднения в определении точных исходных данных;
- применение приближенных методов;
- арифметические действия над приближенными числами;
- вычисление значений функций;
- округление чисел;
- ограниченность разрядной сетки вычислительных устройств.

При решении конкретных задач влияние тех или иных источников различно. Иногда некоторые из них могут отсутствовать вообще или накладывать ничтожный отпечаток на результаты.

Поскольку *приближенные результаты решения задач бесполезны без информации о степени их точности*, в процессе вычислений обязательно следует вести учет погрешностей, а изучение необходимых для этого правил и методов должно занимать важное место в курсе вычислительной математики.

Погрешности математических моделей и исходных данных не зависят от вычислительных методов, поэтому способы их определения и оценки не входят в круг проблем данной книги. В то же время без сведений о точности исходных данных нельзя учесть погрешности результатов. Будем считать, что они известны вычислителю.

Далее в гл. 1 познакомимся с понятиями и правилами, необходимыми для оценки вычислительных погрешностей. Основную роль при этом будет играть понятие расстояния между точными и приближенными значениями величин. В последующих главах одновременно с приближенными численными методами приведем способы оценивания порождаемых ими погрешностей.

1.2. Расстояние

1.2.1. Понятие о расстоянии

При решении рассматриваемых нами математических задач искомые результаты будут представлять собой либо числовые величины¹⁾, либо конечномерные числовые векторы (векторные величины), либо действительные функции одной переменной (функциональные величины). Например, числами являются действительный корень уравнения с одним неизвестным и определенный интеграл, числовым вектором — решение системы линейных алгебраических уравнений, функцией одной переменной — решение обыкновенного дифференциального уравнения.

Возникает вопрос: как измерять или оценивать степень близости приближенных значений перечисленных величин к точным значениям?

В нашем сознании мера близости между двумя какими-то объектами неразрывно связана с понятием *расстояния*. Известно также, что *расстояние может определяться разными способами*. Когда мы говорим, что «Казань ближе к Москве, чем Иркутск», то обычно имеем в виду расстояние «по прямой», однако в повседневной жизни расстояние между городами часто измеряют «по железной дороге», «по шоссе» и т. д.

Расстояние, или, как еще его называют, *метрика*, является одним из основных понятий математического анализа. Оно определяется

¹⁾ Под числом далее всегда понимаем *действительное число*.

с помощью аксиом, согласующихся с обыденными представлениями о расстоянии.

О п р е д е л е н и е 1.1. Пусть X — некоторое непустое множество. Функция ρ , ставящая в соответствие любым двум элементам $x, y \in X$ число $\rho(x, y) \geq 0$ с выполнением условий

1. $\rho(x, y) = 0 \Leftrightarrow x = y$ (аксиома тождества);

2. $\rho(x, y) = \rho(y, x)$ (аксиома симметрии);

3. $\rho(x, y) \leq \rho(x, z) + \rho(z, y)$ (аксиома треугольника),

называется *расстоянием*, или *метрикой*, на множестве X . При этом число $\rho(x, y)$ называется *расстоянием между x и y* .

Множество X с введенной на нем метрикой ρ называется *метрическим пространством*. На одном и том же множестве понятие расстояния можно ввести разными способами. При этом получатся разные метрические пространства.

1.2.2. Расстояние между числами

Пусть $X = \mathbf{R}$, где \mathbf{R} — множество всех действительных чисел. Расстояние между $x, y \in \mathbf{R}$ определим как *модуль разности* между этими числами:

$$\rho(x, y) = |y - x|. \quad (1.1)$$

Геометрически это расстояние равно длине отрезка, соединяющего точки x и y числовой прямой.

1.2.3. Расстояние между векторами

Пусть $X = \mathbf{R}^n$ — множество всех n -мерных числовых векторов и $x = (x_1, x_2, \dots, x_n)$, $y = (y_1, y_2, \dots, y_n)$ — два таких вектора. Расстояние между ними можно ввести разными способами. Наиболее часто применяются следующие определения:

$$\rho_1(x, y) = \sum_{i=1}^n |y_i - x_i|; \quad (1.2)$$

$$\rho_2(x, y) = \sqrt{\sum_{i=1}^n (y_i - x_i)^2}; \quad (1.3)$$

$$\rho_\infty(x, y) = \max_{i=1, \dots, n} |y_i - x_i|. \quad (1.4)$$

Каждое из этих расстояний порождает свое метрическое пространство. В частности, множество n -мерных числовых векторов с расстоянием ρ_2 называется *n -мерным евклидовым пространством*, а метрика ρ_2 называется *евклидовой*.

Пример 1.2. Даны двумерные векторы $x = (1, 1)$ и $y = (2, -1)$. Вычислим расстояния между ними всеми тремя способами, а также расстояния между их соответствующими координатами. На основании формул (1.2) – (1.4) имеем: $\rho_1(x, y) = |2 - 1| + |-1 - 1| = 3$, $\rho_2(x, y) = \sqrt{1^2 + 2^2} = \sqrt{5}$, $\rho_\infty(x, y) = \max\{1, 2\} = 2$. Расстояние между координатами вычисляется по формуле (1.1): $\rho(x_1, y_1) = |2 - 1| = 1$, $\rho(x_2, y_2) = |-1 - 1| = 2$. •

Нетрудно убедиться, что расстояния между соответствующими координатами векторов не превосходят расстояния между векторами при любом способе его определения:

$$|y_i - x_i| \leq \rho(x, y) \quad (i = 1, \dots, n). \quad (1.5)$$

При оценке погрешностей векторов обычно применяют расстояние (1.4). Естественно связывать близость двух векторов с близостью их соответствующих координат, а ρ_∞ делает это лучше, чем другие метрики. Как видно из формулы (1.4), если дано $\epsilon > 0$, то

$$\rho_\infty(x, y) \leq \epsilon \Leftrightarrow |y_i - x_i| \leq \epsilon \quad \text{при всех } i = 1, 2, \dots, n. \quad (1.6)$$

Следовательно, два вектора удалены не больше чем на ϵ в смысле метрики ρ_∞ тогда и только тогда, когда все их соответствующие координаты отличаются друг от друга не более чем на ϵ .

Для других метрик соотношение (1.6) верно только в одну сторону: слева направо (убедитесь в этом, используя неравенство (1.5) и два вектора из упр. 1.3, приняв $\epsilon = 2$).

З а м е ч а н и е. Индексы в обозначениях приведенных метрик имеют следующее смысловое значение. Пусть $t > 0$ и

$$P_{y-x}(t) = \left(|y_1 - x_1|^t + \dots + |y_n - x_n|^t \right)^{\frac{1}{t}}.$$

Ясно, что $\rho_1(x, y) = P_{y-x}(1)$, $\rho_2(x, y) = P_{y-x}(2)$. Можно показать, что $\rho_\infty(x, y) = \lim_{t \rightarrow +\infty} P_{y-x}(t)$.

1.2.4. Расстояние между функциями и их значениями

Пусть f и g – действительные функции n действительных переменных ($n \geq 1$) с областями определения D_f и D_g , причем $D_f \cap D_g \neq \emptyset$. Значения f и g в точке $x = (x_1, \dots, x_n) \in D_f \cap D_g$ представляют собой действительные числа, поэтому *расстояние между значениями* определяется по формуле (1.1):

$$\rho(f(x), g(x)) = |f(x) - g(x)|.$$

Точно так же находится расстояние между двумя значениями одной и той же функции.

Расстояние между функциями будем рассматривать в случае, когда они являются функциями одной переменной. Если функции f, g определены и непрерывны на некотором отрезке $[a; b] \subset \subset D_f \cap D_g$, то за *расстояние между f и g на этом отрезке* примем число

$$\rho(f, g) = \max_{[a, b]} |f(x) - g(x)|.^{1)} \quad (1.7)$$

Оно существует в силу следующей теоремы:

Теорема 1.1 (вторая теорема Вейерштрасса). *Если функция непрерывна на отрезке $[a; b]$, то среди всех ее значений при $x \in [a; b]$ есть наибольшее и наименьшее [3, гл. IV, § 5].*

Пример 1.3. Узнаем, на каком расстоянии находятся друг от друга функции $f: f(x) = x$ и $g: g(x) = \sin x$ на отрезке $[0; 2]$ в смысле определения (1.7). Для этого найдем $\rho(f, g) = \max_{[0, 2]} |x - \sin x|$.

Введем на отрезке $[0; 2]$ функцию $\varphi: \varphi(x) = |x - \sin x| = x - \sin x$. Производная $\varphi'(x) = 1 - \cos x$ положительна на полуинтервале $(0; 2]$, поэтому φ возрастает на отрезке и принимает свое наибольшее значение в точке $x = 2$. Следовательно, $\rho(f, g) = 2 - \sin 2 \approx 1,09$. •

Обратим внимание на то, что формула (1.7) определяет расстояние между функциями как наибольшее из всех расстояний между значениями функций в точках $x \in [a; b]$. Отсюда, в частности, при любом $\varepsilon > 0$ вытекает справедливость соотношения:

$$\rho(f, g) \leq \varepsilon \Leftrightarrow |f(x) - g(x)| \leq \varepsilon \text{ при всех } x \in [a; b]. \quad (1.8)$$

Нам потребуется измерять степень близости между заданными на отрезке $[a; b]$ функциями, когда для одной из них, скажем f , известно аналитическое выражение, а для другой, g , известны лишь значения для аргументов $x_0, x_1, \dots, x_n \in [a; b]$. В этом случае за «расстояние» $\rho(f, g)$ на $[a; b]$ примем расстояние между числовыми векторами $(f(x_0), \dots, f(x_n))$ и $(g(x_0), \dots, g(x_n))$, координатами которых являются значения функций для этих аргументов. Его можно вычислить по формулам (1.2) – (1.4).

Замстим, что здесь величина $\rho(f, g)$ названа расстоянием в кавычках не случайно. Дело в том, что введенная таким образом с помощью конечного набора точек отрезка характеристика близости между двумя функциями не удовлетворяет аксиомам метрики. Например, легко проверить нарушение аксиомы тождества, взяв при фиксированной функции g различные функции f с одними и теми же значениями в точках x_i .

¹⁾ В математике широко используются и другие определения расстояния между непрерывными на отрезке функциями. Например, $\rho(f, g) = \left(\int_a^b (f(x) - g(x))^2 dx \right)^{1/2}$.

В дальнейшем эту меру близости так и будем называть «расстоянием», а чтобы не было путаницы в обозначениях с метрикой, обозначим его через ρ .

Упражнения

1.1. Выясните геометрический смысл аксиом метрики, взяв в качестве X множество точек декартовой координатной плоскости, а в качестве расстояния между двумя точками — длину отрезка, концами которого являются эти точки.

1.2. Найдите множества чисел $x \in \mathbf{R}$, удовлетворяющих условию: а) $\rho(2, x) = 0,5$, б) $\rho(2, x) \leq 0,5$, в) $\rho(2, x) > 0,5$, и изобразите эти множества на числовой прямой.

1.3. Даны два вектора: $\mathbf{a} = (-1, 2, 0)$ и $\mathbf{b} = (1, 3, -2)$. Определите расстояние между ними способами (1.2) – (1.4) и выясните, для каких из этих расстояний верно неравенство $\rho(\mathbf{a}, \mathbf{b}) \leq 2$.

1.4. Докажите, что для любой пары векторов формула (1.4) дает наименьшее расстояние из всех трех перечисленных.

У к а з а н и е: Сначала убедитесь в справедливости неравенства $\max\{a, b\} \leq \sqrt{a^2 + b^2}$ для любых двух чисел $a, b \geq 0$.

1.5. Выясните геометрический смысл расстояний, определенных в (1.2) – (1.4). Для этого рассмотрите ситуацию в случае двумерных векторов $\mathbf{a} = (x_1, y_1)$, $\mathbf{b} = (x_2, y_2)$, которым на координатной плоскости xOy соответствуют точки $A(x_1, y_1)$ и $B(x_2, y_2)$. Построив соответствующий прямоугольный треугольник с гипотенузой AB , убедитесь, что ρ_2 определяет расстояние «по прямой», ρ_1 — как сумму длин катетов, а ρ_∞ — как наибольшую из длин катетов.

1.6. Убедитесь, что геометрически расстояние (1.7) равно наибольшей из длин отрезков, заключенных между графиками функций и проведенных параллельно оси Oy .

1.7. Найдите геометрически и аналитически расстояние между заданными на отрезке $[0; 1]$ функциями $f(x) = 2 - 2x$ и $g(x) = x^3$.

1.8. Найдите с помощью формул (1.2) – (1.4) «расстояния» между заданными на отрезке $[-1; 1]$ функциями f и g , если $f(x) = x^2$ и известна таблица значений g :

x	-1	-0,4	0	0,5	1
$g(x)$	0,5	0,1	0	0,4	2

1.9. Покажите, что все расстояния (1.1) – (1.4) и (1.7) на множествах их определения удовлетворяют аксиомам 1 – 3 метрики.

У к а з а н и е: При доказательстве аксиомы треугольника для расстояния (1.3) примените неравенство:

$$\sqrt{\sum_{i=1}^n (a_i + b_i)^2} \leq \sqrt{\sum_{i=1}^n a_i^2} + \sqrt{\sum_{i=1}^n b_i^2}.$$

1.3. Абсолютная погрешность

Пусть A — точное значение некоторой числовой, векторной или функциональной величины, a — известное приближение к нему (т. е. приближенное значение для A). Обозначаем: $A \approx a$ или $a \approx A$.

В зависимости от типа величины принято называть A *точным* числом (вектором, функцией), а его приближение a — *приближенным* числом (вектором, функцией). Например, в соотношениях $\pi \approx 3,14$, $\pi \approx 3,142$ число π является точным, а числа $3,14$, $3,142$ — приближенными (приближениями к π)¹⁾.

О п р е д е л е н и е 1.1. Разность $A - a$ (или $a - A$) между точным и приближенным значениями величины называется *погрешностью* значения a .

Степень точности приближений удобно характеризовать с помощью неотрицательных чисел, поэтому понятие погрешности для этой цели не годится. Действительно, в случае чисел погрешность может быть как положительным, так и отрицательным числом, а в случае векторов или функций она также будет вектором или функцией.

Знакомство с понятием расстояния наводит на мысль, что $\rho(A, a)$ дает неотрицательную меру погрешности приближения a . Однако здесь необходимо учитывать следующие обстоятельства.

Если точное значение A известно, то расстояние $\rho(A, a)$ найти можно, хотя бывает и так, что его трудно использовать (пример 1.5). Если же A неизвестно, как чаще всего оказывается в реальных ситуациях, то найти расстояние не удастся, и тогда оно само по себе теряет практический смысл как характеристика точности.

В то же время при определении исходных данных, а затем и в процессе вычислений, расстояние между точным и приближенным значениями часто удается *оценить сверху* некоторым неотрицательным числом. При измерениях это число зависит от степени точности измерительных приборов.

П р и м е р 1.4. Для измерения длины l болта использованы метровая линейка с делениями $0,5$ см и линейка с делениями 1 мм. В обоих случаях получен результат $l \approx 3,5$ см. Ясно, что в первом случае отклонение найденной длины $3,5$ см от истинной не должно по модулю превышать $0,5$ см, во втором случае $-0,1$ см. Если этот же результат получится при измерении штангенциркулем, то $\rho(l; 3,5) = |l - 3,5| \leq 0,01$. •

Именно такая оценка, представляющая собой *верхнюю границу множества тех значений, которые расстояние $\rho(A, a)$ может принимать в данных условиях*, и используется на практике для характеристики точности приближенных значений.

¹⁾ Отсюда не следует, что числа $3,14$ и $3,142$ не могут оказаться точными в каких-то других ситуациях.

Определение 1.2. Любое неотрицательное число Δ_a , удовлетворяющее неравенству

$$\rho(A, a) \leq \Delta_a, \quad (1.9)$$

называется *абсолютной погрешностью* приближенного значения a (относительно ρ)¹⁾.

З а м е ч а н и е. Число Δ_a из определения 1.2 часто называют *предельной абсолютной погрешностью*, причем здесь слово «предельная» не имеет никакого отношения к понятию предела функции или последовательности. Смысл этого слова в том, что число Δ_a устанавливает «верхний предел» для значений расстояния и что с его помощью можно найти границы, т.е. «пределы от ... и до ...», в которых находится неизвестное точное значение A .

Как число Δ_a , так и само неравенство (1.9) называют также *оценкой погрешности (точности)* значения a или приближенного равенства $A \approx a$. Чтобы отличать эту оценку от приближенных оценок вида $\rho(A, a) \approx \Delta$, иногда ее называют *строгой оценкой*.

Если a — приближенное число, то неравенство (1.9) примет вид

$$|A - a| \leq \Delta_a.$$

Абсолютная погрешность определяется неоднозначно. Если неравенство (1.9) выполняется при каком-то $\Delta_a = c$, то любое число, превышающее c , также подойдет под определение 1.2 и потому будет абсолютной погрешностью. На практике стараются в качестве Δ_a выбрать возможно меньшее при данных обстоятельствах число, удовлетворяющее (1.9).

Если точное значение A известно, можно принять $\Delta_a = \rho(A, a)$. Когда a является точным значением, т.е. $a = A$, будет $\rho(A, a) = 0$ (аксиома 1 метрики). Понятно, что тогда и $\Delta_a = 0$.

Абсолютная погрешность дает ценную информацию о неизвестном точном значении A : оно находится от известного приближения a на расстоянии, не большем чем Δ_a . В случае числовых величин имеем

$$|A - a| \leq \Delta_a \Leftrightarrow a - \Delta_a \leq A \leq a + \Delta_a.$$

Следовательно, найдя приближенное значение a и его абсолютную погрешность Δ_a , узнаем, что точное значение A располагается на отрезке $[a - \Delta_a; a + \Delta_a]$, т.е. находится «в пределах от $a - \Delta_a$ до $a + \Delta_a$ ».

Пример 1.4 показывает зависимость абсолютной погрешности и границ, в которых находится точный результат, от точности измерительных приборов. В одном случае $\Delta_l = 0,5$ и, следовательно, $3 \leq l \leq 4$;

¹⁾ Словами «относительно ρ » указывается зависимость абсолютной погрешности от расстояния. В случае векторов ее можно определить относительно трех расстояний.

в другом — $\Delta_l = 0,1$ и $3,4 \leq l \leq 3,6$; в третьем — $\Delta_l = 0,01$ и $3,49 \leq l \leq 3,51$.

Иногда о точном значении A ничего не известно, кроме границ, между которыми оно находится. Если $a_1 \leq A \leq a_2$, то за приближенное число a можно взять середину отрезка $[a_1; a_2]$, т. е. $a = \frac{a_1 + a_2}{2}$. Тогда $\Delta_a = \frac{a_2 - a_1}{2}$.

Абсолютная погрешность является основной характеристикой точности вычислений. Во всех последующих правилах учета погрешностей речь будет идти прежде всего о ней.

О п р е д е л е н и е 1.3. Если известна абсолютная погрешность Δ_a приближенного значения a , то a называют *приближением к A с точностью до Δ_a* .

Когда говорят, что надо получить какой-то результат с заданной точностью $\varepsilon > 0$, это означает, что абсолютная погрешность его должна быть не больше ε .

В дальнейшем нам понадобится следующее понятие.

О п р е д е л е н и е 1.4. Все цифры десятичной записи числа, начиная с первой ненулевой слева, называются *значащими цифрами* этого числа.

Нули в конце числа всегда считаются значащими, в противном случае их не пишут. Так, число 0,05020 содержит четыре значащие цифры: 5, 0, 2 и 0.

Абсолютную погрешность не следует записывать с большим количеством значащих цифр. Основной информацией, содержащейся в ней, являются значение первой ненулевой цифры и десятичный разряд, в котором эта цифра расположена.

П р а в и л о. В записи абсолютной погрешности обычно оставляют только одну или две значащие цифры. Для сохранения условия (1.9) округление при этом всегда производится с избытком.

П р и м е р 1.5. Пусть требуется оценить погрешность округления числа $e = 2,7182818\dots$ до трех значащих цифр: $e \approx 2,72$.

Для этого найдем расстояние от числа e до приближения 2,72:

$$\rho(e; 2,72) = |e - 2,72| = 0,00171817\dots$$

Оно является иррациональным числом и потому неудобным для использования. Введем абсолютную погрешность $\Delta_{2,72}$. При записи с двумя значащими цифрами с соблюдением условия (1.9) следует взять $\Delta_{2,72} = 0,0018$, но не 0,0017, как получилось бы при обычных правилах округления. Приемлемым значением $\Delta_{2,72}$ будет и 0,002 (с одной значащей цифрой). •

Как и всякое действительное число, абсолютную погрешность можно записывать в так называемой *плавающей форме* (с плавающей десятичной запятой) $m \cdot 10^p$. Здесь m называется *мантиссой* числа, а p — его *порядком*. Например, $\Delta_{2,72} = 0,0018 = 0,18 \cdot 10^{-2} = 1,8 \cdot 10^{-3}$.

Упражнения

1.10. Укажите границы, в которых находится точное число D , если его приближенное значение $d = 42,36$ найдено с точностью: а) до 0,7; б) до 0,01.

1.11. С помощью двухметровой рулетки с делениями 1 см найдена длина провода $s \approx 8,56$ м. Определите границы, в которых находится точная длина s .

1.4. Десятичная запись приближенных чисел: основные понятия и правила

1.4.1. Округление чисел

При вычислениях часто приходится иметь дело с числами, содержащими большое количество значащих цифр. Независимо от того, точные эти числа или приближенные, часть цифр иногда целесообразно отбрасывать. Минимальную погрешность округления дает следующее правило.

Правило округления чисел. Чтобы округлить число до n значащих цифр, отбрасывают все его цифры, стоящие справа от n -й значащей цифры, или, если это нужно для сохранения разрядов чисел, заменяют их нулями. При этом:

1) *если первая (слева) отбрасываемая цифра меньше 5, то все сохраняемые цифры остаются без изменения;*

2) *если первая отбрасываемая цифра больше 5 или если она равна 5, но среди остальных отбрасываемых цифр есть ненулевые, то к последней сохраняемой цифре прибавляется единица;*

3) *если первая отбрасываемая цифра равна 5 и все остальные отбрасываемые цифры являются нулями, то последняя сохраняемая цифра остается неизменной, если она четная, и увеличивается на единицу, если она нечетная.*

Пусть в результате округления числа a получилось число \bar{a} . Оно имеет погрешность $a - \bar{a}$, вызванную этой операцией. Правило округления гарантирует, что $|a - \bar{a}|$ не будет превышать половины единицы разряда, где находится последняя оставленная цифра. Обратившись к примеру 1.5, в котором было использовано это правило, видим, что $|e - 2,72| \leq 0,002 \leq 0,005$, где $0,005 = \frac{0,01}{2}$ — это и есть половина единицы последнего сохраненного разряда (сотых долей).

Потребность замены цифр нулями для сохранения разрядов возникает при округлении целых чисел. Например, округляя число 56 998 до трех значащих цифр, получим в результате 57 000. Первый

из трех нулей является значащим, остальные два сохраняют разрядность числа — это *нули округления*.

Для того чтобы по записи таких чисел можно было узнать, какой нуль значащий, а какой нет, их записывают в виде $m \cdot 10^p$, оставляя значащие нули в мантиссе m . Приведенное выше число следует представить как $570 \cdot 10^2$.

1.4.2. Верные значащие цифры

С помощью абсолютных погрешностей определяют так называемые верные значащие цифры приближенных чисел.

Пусть приближенное число записано в виде десятичной дроби: $a = a_n \dots a_1 a_0, a_{-1} a_{-2} \dots a_{-m}$, т.е.

$$a = a_n \cdot 10^n + \dots + a_1 \cdot 10^1 + a_0 \cdot 10^0 + a_{-1} \cdot 10^{-1} + \dots + a_{-m} \cdot 10^{-m}.$$

О п р е д е л е н и е 1.5. Значащая цифра приближенного значения a , находящаяся в разряде, в котором выполняется условие: *абсолютная погрешность Δ_a не превосходит половину единицы этого разряда*, называется *верной*. Значащие цифры разрядов, где не выполняется данное условие, называются *сомнительными*.

Следовательно, значащая цифра a_k ($k = n, \dots, 1, 0, -1, \dots, -m$) верная, если $\Delta_a \leq 0,5 \cdot 10^k$. Понятно, что все значащие цифры, расположенные слева от верной, также будут верными, а расположенные справа от сомнительной — сомнительными.

Соответствующие десятичные разряды также называем верными или сомнительными.

П р и м е р 1.6. Для приближенного числа $x = 72,356$ известна абсолютная погрешность $\Delta_x = 0,04$. Требуется определить его верные значащие цифры.

Проверим цифру 7. Половина единицы ее разряда: $\frac{10}{2} = 5 \geq 0,04$.

Значит, она верная. Цифра 2: $\frac{1}{2} = 0,5 \geq 0,04$ — тоже верная. Верной будет и цифра 3 (проверьте!), а вот цифры 5 и 6 — сомнительные.

Действительно, для 5: $\frac{0,01}{2} = 0,005 < 0,04$, т.е. требуемое условие нарушено. •

Результат, полученный в примере 1.6, приводит к мысли о том, что цифры приближенного числа a верны во всех тех разрядах, где им соответствуют нули абсолютной погрешности. Однако это предположение верно лишь частично, ибо последняя такая цифра может оказаться сомнительной. Будет ли она верной, зависит от величины значащих цифр Δ_a .

Для выявления верных цифр числа a без проверки каждой из них «по определению» рекомендуется следующее правило.

Правило. Абсолютная погрешность округляется с избытком до одной значащей цифры (обозначим эту цифру буквой d). Если цифра $d \leq 5$, то все значащие цифры числа a левее того разряда, где находится d , будут верными. В противном случае последнюю (самую правую) из этих цифр следует признать сомнительной.

Пример 1.7. Даны числа a , b , c и их абсолютные погрешности:

$$a = 2,645$$

$$b = 0,81726$$

$$c = 3968$$

$$\Delta_a = 0,003$$

$$\Delta_b = 0,0052$$

$$\Delta_c = 49$$

Видим, что цифры 2, 6, 4 числа a верные, так как соответствующая разряду тысячных долей цифра $d = 3$ абсолютной погрешности Δ_a меньше 5. Число b имеет только одну верную значащую цифру 8. Действительно, при округлении с избытком его абсолютной погрешности получим число 0,006, содержащее в разряде тысячных долей значащую цифру $d = 6 > 5$, которая «портит» разряд сотых долей числа b . У целого числа c цифры верны в разрядах тысяч и сотен: это 3 и 9. •

Верная цифра приближенного числа не обязана буквально совпадать с цифрой соответствующего разряда точного числа. Например, пусть $A = 1,999$ — точное число, $a = 2,000$ — его приближение. Тогда $\Delta_a = 0,001$ и, следовательно, три первые цифры числа a верные, хотя ни одна из них не совпадает с соответствующей цифрой числа A .

Нередко бывает так, что исходные числовые данные приводятся без оценки их погрешностей, но с известными верными цифрами. Возникает задача: *найти абсолютные погрешности этих чисел, необходимые для последующего учета погрешностей.*

Решение следует из определения верной цифры. Если все три цифры числа $a = 4,06$ верные, это означает, что $\Delta_a \leq 0,005$. Во избежание искусственного завышения степени точности числа мы не имеем права взять конкретное $\Delta_a < 0,005$, например, $\Delta_a = 0,004$, поэтому принимаем $\Delta_a = 0,005$. Отсюда следует правило.

Правило. За абсолютную погрешность приближенного числа с известными верными значащими цифрами принимается половина единицы того разряда, где находится последняя верная цифра.

Обратим внимание на информационную значимость нулей, записанных в конце числа. Так, если известно, что все цифры чисел 3,2 и 3,20 верные, то эти записи не равноценны. За абсолютную погрешность первого числа можно взять 0,05, а второго — 0,005.

Когда в конце числа получаются верные значащие нули округления, их следует сохранять. Пусть $a = -17,2986$, $\Delta_a = 0,002$ и требуется округлить a до верных цифр. Тогда пишем $a \approx -17,30$, но не $a \approx -17,3$. Округление целого числа c из примера 1.7 до верных цифр дает результат: $c \approx 40 \cdot 10^2$.

В приближенных вычислениях часто используется другое определение верной значащей цифры.

Определение 1.6. Если абсолютная погрешность числа не превосходит *единицы* того разряда, где находится значащая цифра, то эта цифра называется верной в *нестрогом (широком) смысле*.

Таким образом, для верной в нестрогом смысле цифры a_k должно выполняться неравенство $\Delta_a \leq 1 \cdot 10^k$. Так, у числа 5,6307 с абсолютной погрешностью 0,006 цифра 3 в разряде сотых долей верна в нестрогом смысле, ибо $0,006 \leq 10^{-2} = 0,01$. Понятно, что верными будут и предыдущие цифры 5 и 6.

Для отличия цифру, верную в смысле определения 1.5, называют верной в *строгом (узком) смысле*. Нетрудно проверить: верная в строгом смысле цифра будет верной и в нестрогом смысле, а обратное утверждение не имеет места (упр. 1.14).

В дальнейшем, если не оговорено противное, будем иметь в виду только первоначальную трактовку понятия верной цифры.

1.4.3. Правило записи приближенных чисел

При решении задач приближенными методами прежде всего надо обеспечивать необходимую точность результатов. Однако элементарная вычислительная культура требует не забывать при этом об экономии времени и объема работы независимо от того, как проводятся расчеты — вручную с помощью простейших калькуляторов или с использованием мощных компьютерных систем автоматизации вычислений.

Эти вопросы тесно связаны с количеством цифр в представлении числовых данных. Во многих случаях для достижения требуемой точности нет нужды пользоваться всеми имеющимися ресурсами вычислительных устройств или выписывать все предоставляемые ими цифры результатов. Следует иметь в виду: *точность вычислений зависит не от количества значащих цифр приближенных чисел, а от количества их верных значащих цифр*.

В то же время сохранять всегда только верные цифры неразумно. Во-первых, оценки погрешностей обычно проводятся с завышением, поэтому некоторые значащие цифры, которые с точки зрения теории должны считаться сомнительными, на самом деле могут оказаться верными. Во-вторых, если в процессе вычислений каждый раз округлять до верных цифр, то погрешности округлений приведут к тому, что последние цифры станут неверными.

Правило записи приближенных чисел. *В промежуточных результатах вычислений обычно сохраняются одна-две сомнительные цифры, а окончательные результаты округляют с сохранением не более одной сомнительной цифры*.

Первая рекомендация позволяет избегать накопления погрешностей округлений в верных разрядах. Если вычислений немного, достаточно одной запасной цифры. Напротив, при больших расчетах

иногда оказывается оправданным сохранение в промежуточных результатах трех таких цифр.

В ответах часто оставляют только верные цифры. Это удобно, так как по записи числа сразу видно, какие цифры у него верные. Однако здесь надо учесть то, что при округлении некоторых приближенных чисел до верных цифр последняя цифра может оказаться верной лишь в нестрогом смысле, ибо тогда к погрешности числа прибавляется погрешность округления. Взяв, например, число $a = 3,6159$ с $\Delta_a = 0,004$, после округления до верных цифр получим $\bar{a} = 3,62$. Тогда

$$\Delta_{\bar{a}} = \Delta_a + |a - \bar{a}| = 0,004 + 0,0041 = 0,0081.$$

Последняя сохраненная цифра 2 числа \bar{a} верна в нестрогом смысле, остальные цифры верны в строгом смысле, ибо $\Delta_{\bar{a}} > 0,005$, но в то же время $\Delta_{\bar{a}} \leq 0,01$ и $\Delta_{\bar{a}} \leq 0,05$.

В подобных случаях ради сохранения качества последней верной цифры целесообразно записывать ответы с дополнительной цифрой. Лишнюю цифру следует оставлять и в том случае, когда известно, что она могла оказаться сомнительной из-за заведомо грубых оценок.

1.4.4. Правила записи знаков точного и приближенного равенств

Замена словесных выкладок символьными соотношениями в вычислительной математике нередко приводит к ситуациям, когда выбор любого из знаков точного или приближенного равенства не дает возможности соблюсти полную строгость и определенность в обозначениях и записях, причем использование вместо них еще каких-либо знаков, как правило, не улучшает положения. Отмеченная проблема особенно актуальна при подробном описании вычислительных процессов, когда приходится связывать величины и их значения, а также различные значения одной и той же величины. Она в значительной степени может быть снята, если в подобных случаях вводить дополнительные обозначения и избегать многих общепринятых символьных записей. Однако обычно так не делают, поскольку считается, что написанное будет правильно понято по контексту. А для этого, в свою очередь, необходимы специальные соглашения.

Надо отметить, что в литературе по приближенным вычислениям встречаются разные варианты записи знаков «=» и «≈», в том числе и такие, которые приводят к противоречиям между смыслом знака и реальным смыслом соотношения. Так, иногда принято связывать точное и приближенное значения функции знаком равенства, если в приближенном значении все цифры верные. Пишется, например, $\cos 0,502 = 0,87662$, хотя понятно, что здесь на самом деле имеет место приближенное равенство.

Рассмотрим правила, которых будем придерживаться далее в данном вопросе.

Представляется естественным и методически важным с точки зрения различения понятий «точное значение» и «приближенное значение» применять знак « \approx » *каждый раз, когда точное значение величины или выражения заменяется его приближенным значением.* В приведенном выше примере с косинусом, как и в других аналогичных случаях, далее обязательно ставим знак приближенного равенства: $\cos 0,502 \approx 0,87662$; $\sqrt{\pi} \approx \sqrt{3,14} \approx 1,772$. Второй из знаков « \approx » в последнем соотношении связывает приближение к $\sqrt{\pi}$ с его приближенным значением, но, поскольку у $\sqrt{3,14}$ существует точное значение, этот знак оправдан.

В то же время *если величина по определению неоднозначна, то между ней и ее значением, а также между ее различными значениями ставим знак « $=$ »*, даже если имеются переходы к приближениям. Данное правило используем, в частности, при вычислении абсолютной и относительной погрешностей (понятие относительной погрешности вводится в 1.5). Ясно, что оно будет иногда приводить к противоречивым соотношениям, однако здесь исходим из того, что, например, абсолютной погрешностью Δ_a приближения a к точному числу A с формально равным правом может быть названо любое число $c \geq 0$, удовлетворяющее неравенству $|A - a| \leq c$. Если числа c и d являются возможными значениями Δ_a и $c \neq d$, то запись $\Delta_a = c = d$ следует понимать как краткую запись двух не вызывающих сомнения равенств: $\Delta_a = c$ и $\Delta_a = d$.

В примере 1.10 (см. 1.6) отыскивается абсолютная погрешность числа $1,625 \approx e^{0,5}$ как значение функции $V: V(x) = \frac{e}{3!}x^3$ в точке $x=0,5$: $\Delta_{1,625} = V(0,5) = 0,06$. Здесь даются два значения абсолютной погрешности: сначала $V(0,5) = \frac{e \cdot 0,125}{6}$, а затем другое, полученное округлением иррационального числа $V(0,5)$. Второй из знаков точного равенства в действительности не имеет места, но если его заменить на « \approx » или на имеющий место знак « \ll », в результате получится или приближенное равенство $\Delta_{1,625} \approx 0,06$, намекающее на существование какого-то точного значения у $\Delta_{1,625}$, чьим приближением является число 0,06, или соответственно неравенство $\Delta_{1,625} < 0,6$, вообще не определяющее $\Delta_{1,625}$ ¹⁾.

¹⁾ В данном случае абсолютную погрешность можно было найти без прямого вычисления, а по ее определению. Используя верные соотношения $|e^{0,5} - 1,625| \leq V(0,5) \leq 0,06$, получаем выбранное нами значение $\Delta_{1,625} = 0,06$. Однако подобные действия «по определению» обычно сопряжены с существенно более громоздкими выкладками. Формулы прямого вычисления оценок погрешностей для того и выводятся, чтобы сократить эти выкладки.

При округлении точного числа между ним и результатом ставится знак приближенного равенства согласно общему правилу. Сложнее выбрать знак для того, чтобы связать, например, приближенное значение 2,365 с полученным после его округления числом 2,4. С одной стороны, здесь нет замены точного числа приближенным, с другой — эти числа не равны. Тем не менее *при округлении всех чисел, за исключением значений неоднозначно определяемых величин, между ними и результатами их округления будем ставить знак «≈»*. Для обоснования такой договоренности можно учесть то, что округление приводит, как правило, к менее точному результату (этого нельзя утверждать только при отбрасывании части сомнительных цифр).

Затруднения возникают и в таких естественных условиях, когда величина или выражение обозначаются каким-либо символом, а затем этот символ соединяется знаком «≈» с соответствующим приближенным значением.

Пусть, например, буквой x обозначено произведение приближенных чисел 1,06 и 7,58. Для расчетов требуется найти его с округлением до трех значащих цифр и оценить погрешность результата. Число $x = 1,06 \cdot 7,58 = 8,0348$ приближенное, его абсолютная погрешность обозначается через Δ_x . Округлив, получим $x \approx 8,03$. Принято в дальнейшем под x понимать число 8,03, а его абсолютную погрешность обозначать снова через Δ_x . Тем самым допускается определенная вольность, поскольку фактически $x \neq 8,03$. В этом случае считается, что *буква x относится не только к значению 8,0348 произведения, но и ко всем его приближениям*¹⁾.

Отмеченное соглашение будем иметь в виду и в дальнейшем при обозначениях приближенных величин.

Упражнения

1.12. Округлите числа $\pi = 3,1415926\dots$ и $\sqrt{0,5} = 0,70710678\dots$ до четырех значащих цифр, а затем найдите абсолютные погрешности результатов с одной и двумя значащими цифрами.

1.13. Каждое из чисел $x = -0,0106$ и $y = 624,5$ имеет две верные значащие цифры. Какие числа можно взять в качестве их абсолютных погрешностей?

1.14. Абсолютная погрешность числа $b = 36,02$ равна 0,07. Найдите верные значащие цифры b в строгом и нестрогом смысле.

1.15. Сформулируйте правило определения абсолютной погрешности числа с известными верными значащими цифрами в нестрогом смысле. Найдите по этому правилу Δ_a , если число $a = 12,6054$ имеет четыре верные в нестрогом смысле цифры.

¹⁾ Менее удобные, но более строгие выкладки в данной ситуации можно обеспечить, если ввести дополнительное обозначение $\bar{x} = 8,03$.

1.16. Пусть даны числа $a = 4,098$ и $b = 0,546$ с одинаковой абсолютной погрешностью $0,002$. Округлите числа до верных цифр и определите, в каком смысле будут верными последние цифры.

1.5. Относительная погрешность приближенных чисел

При приближенных измерениях и вычислениях возникает потребность в характеристике *качества* проделанной работы. Для этой цели знание только абсолютной погрешности оказывается недостаточным.

Пример 1.8. Найдена масса одного предмета $a = 510,4$ кг с точностью до $0,1$ кг и с такой же точностью определена масса $b = 0,6$ кг другого предмета. Хотя $\Delta_a = \Delta_b = 0,1$, ясно, что первое измерение выполнено лучше, чем второе. •

Для оценки качества измерений или вычислений вводится понятие относительной погрешности.

Определение 1.7. *Относительной погрешностью* приближенного числа a ($a \neq 0$) будем называть неотрицательное число

$$\delta_a = \frac{\Delta_a}{|a|}. \quad (1.10)$$

Замечание. Введенное таким образом число δ_a часто называют *предельной относительной погрешностью*. Следует также заметить, что в работах по вычислительной математике встречаются другие определения понятия относительной (предельной относительной) погрешности. Однако в плане практического использования данное здесь определение является наиболее распространенным.

Формула (1.10) связывает абсолютную и относительную погрешности чисел. Из нее, в частности, следует важное соотношение:

$$\Delta_a = |a| \cdot \delta_a.$$

Из-за неоднозначности абсолютной погрешности относительная погрешность приближенного числа также не единственна. Как и абсолютную погрешность, *относительную погрешность записывают с одной-двумя значащими цифрами и округляют при необходимости с избытком*. Она является безразмерной величиной и потому часто

выражается в процентах: $\delta_a = \frac{\Delta_a}{|a|} \cdot 100\%$.

Заметим, что при фиксированном Δ_a относительная погрешность тем меньше, чем больше абсолютная величина приближенного числа. Этот факт хорошо иллюстрируется данными из примера 1.8. Имеем: $\delta_a = 0,0002$, или $0,02\%$; $\delta_b = 0,17$, или 17% . Это означает, что в найденной массе первого предмета ошибка составляет примерно $0,02\%$, а в массе второго — 17% .

Упражнения

1.17. Найдите относительную погрешность приближенного значения 2,72 для числа e (см. пример 1.5).

1.18. Приближенное значение $x = 24,6035$ имеет относительную погрешность $\delta_x = 0,2\%$. Найдите Δ_x и верные цифры числа x .

1.19. У приближенных чисел $a = -3,540$ и $b = 0,02$ все значащие цифры верные. Найдите относительные погрешности этих чисел.

1.6. Оценка точности приближенных векторов и функций

1.6.1. Абсолютная погрешность приближенных векторов

Оценка точности вектора $a = (a_1, \dots, a_n)$, являющегося приближением к вектору $A = (A_1, \dots, A_n)$, может быть проведена относительно любого из трех расстояний (1.2) – (1.4). Если за основу взять расстояние ρ_∞ , то число $\Delta_a \geq 0$ будет *абсолютной погрешностью* вектора a при выполнении неравенства:

$$\max_{i=1, \dots, n} |A_i - a_i| \leq \Delta_a. \quad (1.11)$$

Определение Δ_a в смысле остальных расстояний предоставляет читателю.

Координаты приближенного вектора могут быть получены с разной степенью точности. В силу соотношения (1.5) абсолютная погрешность вектора является абсолютной погрешностью всех его координат и потому дает «глобальную» оценку их точности. Однако для отдельных координат эта оценка может оказаться завышенной.

Пример 1.9. Для системы из трех уравнений с тремя неизвестными найдено приближенное решение $x = (1,011; 2,999; -0,964)$ с абсолютной погрешностью $\Delta_x = 0,04$. Все координаты вектора x гарантированно имеют такую же абсолютную погрешность, и поэтому их значащие цифры верны в разрядах до десятых долей включительно. Если точным решением системы является вектор $\bar{x} = (1; 3; -1)$, то очевидно, что погрешность координаты $x_2 = 2,999$ значительно меньше, чем показывает Δ_x . •

1.6.2. Абсолютная погрешность приближенных функций и их значений

Пусть на некотором отрезке $[a; b]$ имеет место приближенная формула:

$$f(x) \approx p(x), \quad x \in [a; b]. \quad (1.12)$$

Здесь функция f такая, что ее значения находить трудно или невозможно, а p — приближение к f — достаточно просто вычисляемая функция.

Нас интересует оценка погрешностей чисел $p(x)$ как приближений к значениям $f(x)$ функции f при $x \in [a; b]$ безотносительно к тому, является ли аргумент x точным или нет, а также к тому, насколько точно вычисляются значения функции p ¹⁾. Поэтому здесь можно считать, что $f(x)$ — точные числа, а погрешности их приближенных значений $p(x)$ порождаются только заменой f на p .

Функция r : $r(x) = f(x) - p(x)$ называется *остаточным членом* формулы (1.12). Значение r в точке $x \in [a; b]$ равно погрешности числа $p(x)$, порожденной формулой. В конкретных ситуациях часто удастся получить аналитическое выражение для $r(x)$, но оно обычно содержит неизвестные числовые параметры и потому непригодно для практического применения. Для оценки погрешностей отдельных чисел $p(x)$ и степени близости функций p и f на всем отрезке $[a; b]$ подыскивается оценочная функция.

О п р е д е л е н и е 1.8. Любую заданную на отрезке $[a; b]$ конечную неотрицательную функцию V , удовлетворяющую условию

$$|f(x) - p(x)| = |r(x)| \leq V(x), \quad x \in [a; b], \quad (1.13)$$

назовем *оценочной функцией* функции p или приближенного равенства (1.12).

Видим, что оценочная функция определяется неоднозначно. Ее значения — это и есть *абсолютные погрешности значений* приближенной функции p :

$$\Delta_{p(x)} = V(x), \quad x \in [a; b]. \quad (1.14)$$

Выше предполагалось, что приближенное равенство (1.12) является единственным источником погрешностей при вычислении $f(x)$. В случае когда значение $p(x)$ при каком-либо $x \in [a; b]$ вычисляется неточно и получится приближение \bar{y} к нему, имеет место двойное приближенное равенство: $f(x) \approx p(x) \approx \bar{y}$. Если возникающая при этом дополнительная погрешность оценивается некоторым числом ε : $|p(x) - \bar{y}| \leq \varepsilon$, то абсолютную погрешность числа \bar{y} относительно точного числа $f(x)$ надо искать по формуле

$$\Delta_{\bar{y}} = V(x) + \varepsilon. \quad (1.15)$$

Однако исходную функцию f обычно заменяют такой функцией p , значения которой можно находить с любой точностью. Тогда можно добиться сколь угодно малости числа ε по сравнению с $V(x)$ и фактически прийти к предполагаемой выше ситуации. При этом

¹⁾ Учет погрешностей вычислений и оценка влияния погрешностей аргументов на точность значения функции будут рассмотрены в 1.7, 1.8.

целесообразен следующий порядок действий: сначала пайти оценку погрешности формулы (1.12) в точке x — число $V(x)$, затем с завышенной точностью подсчитать $\bar{y} \approx p(x)$ и определить верные цифры \bar{y} по $V(x)$.

Когда функции f, p, V в (1.13) непрерывны на $[a; b]$, то, применяя определение расстояния (1.7), можно указать «глобальную» оценку сверху погрешностей всех чисел $p(x)$ для $x \in [a; b]$, т. е. определить *абсолютную погрешность Δ_p функции p на отрезке $[a; b]$* . Для этого паходим наибольшее значение на $[a; b]$ обеих частей неравенства (1.13):

$$\rho(f, p) = \max_{[a; b]} |f(x) - p(x)| \leq \max_{[a; b]} V(x),$$

откуда и следует:

$$\Delta_p = \max_{[a; b]} V(x). \quad (1.16)$$

Учитывая (1.14), можно сказать, что абсолютная погрешность функции на отрезке равна наибольшей из абсолютных погрешностей ее значений для аргументов из этого отрезка. В частности, абсолютная погрешность приближенной функции p на $[a; b]$ является абсолютной погрешностью всех значений $p(x)$, $x \in [a; b]$.

Пример 1.10. Взяв частичную сумму степенного ряда функции $f: f(x) = e^x$ при $n = 2$, получим приближенную формулу:

$$e^x \approx 1 + x + \frac{x^2}{2!} \stackrel{\text{def}}{=} p(x), \quad x \in \mathbf{R}. \quad (*)$$

Ее остаточный член может быть задан выражением: $r(x) = \frac{e^{kx}}{3!} x^3$ с некоторым числом $0 < k < 1$ [3, гл. VI, § 3]. Понятно, что для подсчета погрешностей это выражение не годится, ибо оно содержит неизвестный параметр k .

Рассматривая формулу (*) на каком-либо ограниченном промежутке, нетрудно получить ее оценочную функцию, свободную от параметра k . Например, при $x \in [0; 1]$ очевидно неравенство: $|e^x - p(x)| = |r(x)| \leq \frac{e}{3!} x^3$. Следовательно, функция $V: V(x) = \frac{e}{3!} x^3$ является оценочной для формулы (*) на отрезке $[0; 1]$.

Значения функции p можно находить с произвольной точностью. При $x = 0,5$ получим точное значение $p(0,5) = 1,625$. В соотношении $e^{0,5} \approx 1,625$ погрешность числа 1,625 вызвана только приближенной формулой (*), поэтому на основании (1.14) $\Delta_{1,625} = V(0,5) = 0,06$.

Единая оценка точности формулы (*) при всех $x \in [0; 1]$ дается абсолютной погрешностью функции p на этом отрезке: $\Delta_p = \max_{[0; 1]} V(x) = e/6 = 0,46$. С помощью вычислений нетрудно убедиться-

ся, что она тем грубее, чем меньше x , но приемлема для x , близких к единице (проверьте при $x = 1$). •

Упражнения

1.20. 1. Убедитесь, что за абсолютную погрешность вектора \mathbf{a} , определяемого соотношением (1.11), можно взять наибольшую из абсолютных погрешностей его координат: $\Delta_{\mathbf{a}} = \max_{i=1, \dots, n} \Delta_{a_i}$.

2. Выразите через абсолютные погрешности координат абсолютные погрешности вектора, определяемые относительно ρ_1 и ρ_2 .

1.21. Пусть приближенный вектор $\mathbf{a} = (a_1, \dots, a_n)$ вычислен с точностью до некоторого числа $\epsilon > 0$ (в смысле ρ_{∞}). Укажите границы, в которых находятся координаты точного вектора \mathbf{A} .

1.22. Округлите координаты вектора $\mathbf{a} = (1,358; -0,6205; 3,07)$ до одной значащей цифры после десятичной запятой и для полученного вектора \mathbf{x} найдите $\Delta_{\mathbf{x}}$ и $\Delta_{\mathbf{x}}$ (относительно всех трех расстояний).

1.23. Пусть на $[a; b]$ известны приближенная функция p и ее абсолютная погрешность Δ_p . Укажите границы, в которых находятся значения точной функции f при всех $x \in [a; b]$, и изобразите на рисунке взаимное расположение графиков функций p и f .

1.24. 1. Напишите приближенную формулу для вычисления $\cos x$, $x \in \mathbf{R}$, взяв два члена степенного ряда этой функции.

2. Найдите оценочную функцию V этой формулы на \mathbf{R} (используйте свойства знакопередающихся рядов).

3. Найдите абсолютную погрешность Δ_p приближенной функции p на отрезке $[-1; 1]$.

4. Вычислите по полученной формуле приближенные значения $\cos 0,2$, $\cos(-0,5)$, $\cos 1$ с максимально возможной точностью и определите их верные значащие цифры.

1.7. Оценка влияния погрешностей аргументов на значение функции

Реализация численных методов сводится к вычислениям значений функций одной или нескольких переменных и к выполнению арифметических действий над числами. Так как любое арифметическое действие можно представить как вычисление значения функции двух переменных ($f(x, y) = x + y$ и т.д.), то вначале займемся проблемой учета погрешностей значений функций.

1.7.1. Постановка задачи

Пусть дана некоторая функция одной или нескольких переменных и вычисляется ее значение. Здесь возможны два случая:

1. Все аргументы функции являются точными числами.

2. Среди аргументов есть приближенные числа.

В первом случае погрешность значения функции фактически зависит только от способа вычисления этого значения.

Если применяется приближенная формула, надо знать оценку погрешности формулы. Об этом говорилось в п. 1.6.2. Вычисления значений посредством математических таблиц и оценка их погрешностей будут рассмотрены в гл. 3. Когда значения отыскиваются с помощью микрокалькулятора или компьютерных средств вычислений, их точность зависит от количества разрядов выводимых чисел; при этом последняя цифра обычно считается сомнительной.

Во втором случае к погрешности метода вычисления добавляется погрешность, вызванная погрешностями аргументов. Возникает проблема оценки этой погрешности.

На рис. 1.1 изображен график функции одной переменной. Видно, что одно и то же отклонение ϵ от точного значения аргумента может привести к разным ошибкам Δ_1 и Δ_2 в значениях функции: чем быстрее функция возрастает (убывает) в окрестности аргумента, тем ошибка больше. В случае дифференцируемой функции скорость ее изменения зависит от значений производной, поэтому естественно ожидать, что в формуле оценки погрешностей производные будут играть основную роль.

Для функции многих переменных ситуация аналогичная, только в этом случае в оценке погрешностей используются частные производные.

В п. 1.7.2 будет приведен способ решения поставленной задачи для дифференцируемой функции двух переменных. Его легко перенести на случай функции большего числа переменных.

1.7.2. Оценка погрешности значения дифференцируемой функции

Предварительно приведем необходимые понятия из математического анализа [4, гл. XV; 3, гл. V].

Пусть точка (a, b) является внутренней точкой области определения D_f функции двух действительных переменных $z = f(x, y)$. Предположим, что аргументам a и b даются такие приращения Δx и Δy , что точки $(a + \Delta x, b + \Delta y) \in D_f$. Функция f называется *дифферен-*

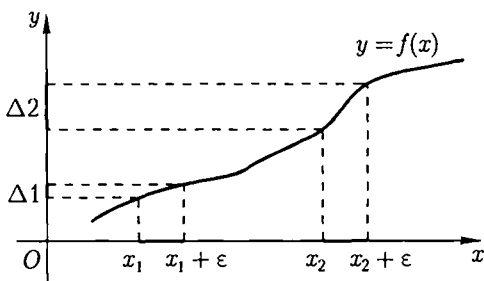


Рис. 1.1

цируемой в точке (a, b) , если ее приращение $\Delta z = f(a + \Delta x, b + \Delta y) - f(a, b)$ в этой точке может быть представлено в виде

$$\Delta z = f'_x(a, b)\Delta x + f'_y(a, b)\Delta y + \alpha\Delta x + \beta\Delta y, \quad (1.17)$$

где α и β зависят от Δx^1 , Δy и $\alpha \rightarrow 0, \beta \rightarrow 0$ при $\Delta x \rightarrow 0, \Delta y \rightarrow 0$. Линейное относительно $\Delta x, \Delta y$ выражение

$$dz = f'_x(a, b)\Delta x + f'_y(a, b)\Delta y$$

называется *дифференциалом* функции f в точке (a, b) .

Для функции одной переменной $y = f(x)$ дифференцируемость в точке $x = a$ означает возможность представления ее приращения $\Delta y = f(a + \Delta x) - f(a)$ в виде

$$\Delta y = f'(a)\Delta x + \alpha\Delta x, \quad (1.18)$$

где $\alpha \rightarrow 0$ при $\Delta x \rightarrow 0$. В этом случае дифференциалом функции f в точке a называется $dy = f'(a)\Delta x$.

Пусть теперь аргументы a, b функции двух переменных f являются приближенными числами с абсолютными погрешностями Δ_a, Δ_b и $z = f(a, b)$. Обозначим через A и B соответствующие точные значения аргументов, тогда $Z = f(A, B)$ — точное значение функции.

Ставится следующая задача: оценить погрешность приближенного равенства $Z \approx z$, вызванную погрешностями аргументов, в предположении, что f дифференцируема в точке (a, b) .

Представим A, B в виде $A = a + \Delta x, B = b + \Delta y$, трактуя тем самым погрешности чисел a, b как приращения $\Delta x, \Delta y$. Как видно из (1.17), погрешность $Z - z$ распадается на две части: дифференциал dz (назовем это *дифференциальной погрешностью*) и сумму $\alpha\Delta x + \beta\Delta y$ (является *дополнительной частью погрешности*).

Обратим внимание на следующие случаи. При малых по модулю частных производных $f'_x(a, b), f'_y(a, b)$ и близких к нулю погрешностях аргументов погрешность числа z также мала, причем ее абсолютная величина может оказаться меньше, чем $|\Delta x|$ и $|\Delta y|$. Когда $f'_x(a, b)$ и $f'_y(a, b)$ равны нулю, то погрешность z равна $\alpha\Delta x + \beta\Delta y$. Если погрешности аргументов достаточно малы, а частные производные далеки от нуля, причем по модулю во много раз больше соответствующих сомножителей $\Delta x, \Delta y$, то дифференциал представляет собой *главную часть* погрешности z .

Рассмотрим подробнее последний случай. Пусть, например, по формуле $z = xy$ вычисляется площадь прямоугольника с точными значениями сторон $A = 1,99, B = 1,02$, а при измерениях получили их приближения $a = 2,00, b = 1,00$. Тогда $f'_x(a, b) = b = 1, f'_y(a, b) = a = 2, \Delta x = -0,01, \Delta y = 0,02$, при этом погрешность $Z - z = 2,0298 - 2 =$

¹⁾ Не путайте обозначения: Δx — приращение, а Δ_x — абсолютная погрешность.

$= 0,0298$ представляется в виде суммы $dz = b\Delta x + a\Delta y = 0,03$ и небольшого дополнения $\Delta x\Delta y = -0,0002$.

Если удастся оценить обе части погрешности некоторыми числами $\bar{\Delta}_z$ и Δ_z^* : $|dz| \leq \bar{\Delta}_z$, $|\alpha\Delta x + \beta\Delta y| \leq \Delta_z^*$, получим неравенство: $|Z - z| \leq \bar{\Delta}_z + \Delta_z^*$, определяющее абсолютную погрешность значения $z = f(a, b)$:

$$\Delta_z = \bar{\Delta}_z + \Delta_z^{*1)}.$$

Число $\bar{\Delta}_z$ найти нетрудно (теорема 1.2), а вот поиск Δ_z^* во многих случаях оказывается серьезной проблемой. Ввиду этого данную формулу применяют сравнительно редко, поступая в различных ситуациях по-разному.

В случаях, когда обязателен строгий учет погрешностей, можно воспользоваться еще одним способом определения Δ_z : если функция f и ее частные производные f'_x, f'_y непрерывны на прямоугольнике

$$D = \{(x, y) : |x - a| \leq \Delta_a, |y - b| \leq \Delta_b\},$$

то справедливо соотношение

$$|Z - z| \leq \max_{(x, y) \in D} |f'_x(x, y)| \Delta_a + \max_{(x, y) \in D} |f'_y(x, y)| \Delta_b = \Delta_z^{*2)}, \quad (1.19)$$

доказательство которого основано на теореме Лагранжа (упр. 1.31).

При решении практических задач модули производных f'_x, f'_y чаще всего значительно больше абсолютных погрешностей аргументов. Если при этом небольшое завышение показателей точности результатов считается приемлемым, то ограничиваются оценкой $\bar{\Delta}_z$ главной части погрешности, а малой по величине оценкой Δ_z^* пренебрегают. Тогда число $\bar{\Delta}_z$ используется вместо Δ_z , а соответствующее ему число $\bar{\delta}_z = \frac{\bar{\Delta}_z}{|z|}$ — вместо относительной погрешности $\delta_z = \frac{\Delta_z}{|z|}$.

Далее получим формулы для вычисления $\bar{\Delta}_z, \bar{\delta}_z$ и по умолчанию будем применять в расчетах только эти приближенные оценки (включая и вопросы определения верных цифр). Для отличия от Δ_z и δ_z назовем их *абсолютной дифференциальной погрешностью* и соответственно *относительной дифференциальной погрешностью*, однако там, где смысл понятий ясен из контекста, слово «дифференциальная» и черточки в обозначениях опускаем.

¹⁾ Подчеркнем, что здесь не учитываются погрешности вычисления $f(a, b)$. Если найдены приближенное значение $\bar{z} \approx f(a, b) = z$ и оценка $|z - \bar{z}| \leq \nu$, то абсолютную погрешность приближенного числа \bar{z} относительно Z следует определять по формуле $\Delta_{\bar{z}} = \Delta_z + \nu = \bar{\Delta}_z + \Delta_z^* + \nu$.

²⁾ Вместо максимумов модулей производных можно брать неотрицательные числа M, N такие, что $|f'_x(x, y)| \leq M, |f'_y(x, y)| \leq N$ при всех $(x, y) \in D$.

Теорема 1.2. Пусть числа a, b являются приближениями к точным аргументам A, B с абсолютными погрешностями Δ_a, Δ_b и $Z = f(A, B)$, $z = f(a, b)$. Если функция f дифференцируема в точке (a, b) , то

$$\bar{\Delta}_z = |f'_x(a, b)|\Delta_a + |f'_y(a, b)|\Delta_b, \quad (1.20)$$

$$\bar{\delta}_z = \frac{1}{|f(a, b)|} (|f'_x(a, b)|\Delta_a + |f'_y(a, b)|\Delta_b) \quad (f(a, b) \neq 0). \quad (1.21)$$

○ В доказательстве нуждается только формула (1.20). Учитывая принятые выше обозначения $A = a + \Delta x$, $B = b + \Delta y$ и определение абсолютной погрешности, имеем: $|\Delta x| = |A - a| \leq \Delta_a$ и $|\Delta y| = |B - b| \leq \Delta_b$. Тогда

$$|dz| \leq |f'_x(a, b)|\Delta x + |f'_y(a, b)|\Delta y \leq |f'_x(a, b)|\Delta_a + |f'_y(a, b)|\Delta_b,$$

откуда и следует (1.20). •

З а м е ч а н и е. В [1, гл. I, § 5] показано, что при малых Δ_a и Δ_b абсолютная погрешность Δ_z из (1.19) с большой точностью приближается вычисляемой по формуле (1.20) абсолютной дифференциальной погрешностью $\bar{\Delta}_z$.

В следующем примере используем полученные в теореме 1.2 оценки.

П р и м е р 1.11. Пусть $a = -0,68$, $b = 1,134$ — приближенные числа с $\Delta_a = 0,004$, $\Delta_b = 0,0003$. Требуется найти значение z функции $f: f(x, y) = x^2 + \sin y$ при $x = a$, $y = b$ и оценить его погрешность.

Чтобы узнать, с какой точностью надо вычислять z , сначала оценим влияние погрешностей аргументов на значение функции. Поскольку f дифференцируема в точке (a, b) , причем модули частных производных $f'_x(a, b) = 2 \cdot (-0,68) \approx -1,4$, $f'_y(a, b) = \cos 1,134 \approx 0,42$ значительно больше Δ_a и Δ_b , применим для этого формулу (1.20).

Значение $\bar{\Delta}_z = 1,4 \cdot 0,004 + 0,42 \cdot 0,0003 = 0,006$ говорит о том, что верные цифры z могут быть только в разрядах до десятых долей включительно. Чтобы вычислительная погрешность не «испортила» эти разряды, подсчитаем $x^2 + \sin y$ с двумя верными цифрами после десятичной запятой и получим

$$z = (-0,68)^2 + \sin 1,134 \approx 0,46 + 0,91 = 1,37.$$

Относительная погрешность $\bar{\delta}_z = 0,006/1,37 = 0,0044$. •

Отметим, что числа, входящие в формулы для вычисления абсолютной или относительной погрешности, достаточно брать с двумя-тремя значащими цифрами. Так поступили в примере 1.11 со значениями частных производных. Это объясняется тем, что лишние цифры не влияют на итоговый результат, поскольку у оценок погрешностей все равно сохраняется не более двух значащих цифр.

Упражнения

1.25. Докажите, что если функция $y = f(x)$ дифференцируема в точке a , то дифференциальная погрешность значения $f(a)$ оценивается числами:

$$\bar{\Delta}_{f(a)} = |f'(a)|\Delta_a, \quad \bar{\delta}_{f(a)} = \frac{|f'(a)|}{|f(a)|}\Delta_a, \quad (f(a) \neq 0). \quad (1.22)$$

1.26. Докажите на основании формул (1.22) следующие утверждения и убедитесь в них по соответствующим графикам:

1. Абсолютная дифференциальная погрешность $\sin x$ и $\cos x$ не превышает абсолютной погрешности x .

2. Абсолютная дифференциальная погрешность $\operatorname{tg} x$ и $\operatorname{ctg} x$ не меньше абсолютной погрешности x .

3. Абсолютная дифференциальная погрешность e^x при фиксированном Δ_x увеличивается с ростом x , а относительная дифференциальная погрешность остается неизменной.

4. При фиксированном Δ_x оценки дифференциальной погрешности значений \sqrt{x} ($x > 0$) и $\ln x$ уменьшаются с ростом x , а с приближением x к нулю они резко увеличиваются.

1.27. Оцените дифференциальные погрешности чисел $\cos 0,854$, $\operatorname{tg} 0,517$ и запишите эти числа с верными значащими цифрами при условии, что оба аргумента являются приближенными числами с относительной погрешностью $0,004$.

1.28. Докажите, что абсолютная дифференциальная погрешность $\ln xy$ ($x, y > 0$) равна сумме относительных погрешностей x и y .

1.29. Найдите оценки погрешности значения объема предмета из примера 1.1, если измерения дали радиус дна $r \approx 20$ см с точностью до $0,5$ см и высоту $h \approx 50$ см с точностью до $0,1$ см. Вычислите объем и определите верные значащие цифры результата.

У к а з а н и е. Рассмотрите функцию $v: v(r, h) = \pi r^2 h$. Округлите число π так, чтобы погрешность его приближения не оказала существенного влияния на точность результата.

1.30. До сих пор решалась задача оценки погрешности значения функции по известной оценке погрешностей аргументов. В случае дифференцируемой функции одной переменной $y = f(x)$ нетрудно решить обратную задачу: с какой допустимой погрешностью необходимо брать аргумент a , чтобы получить значение $y = f(a)$ с наперед заданной точностью $\varepsilon > 0$.

1. Наншите формулу вычисления абсолютной погрешности Δ_a по известной абсолютной погрешности $\bar{\Delta}_y$, если f дифференцируема в точке a и $f'(a) \neq 0$.

2. Получен результат: $\sin 1,3834 \approx 0,982$. Если число $x = 1,3834$ приближенное, то сколько его верных цифр достаточно, чтобы все значащие цифры числа $0,982$ были верными?

3. Определите, до скольких цифр надо округлить число π , чтобы с точностью до 0,01 вычислить π^3 ; $\ln \pi$; e^π ; $\sqrt{\pi}$.

1.31. Докажите неравенство (1.19), воспользовавшись теоремой 6.4 (см. 6.5).

1.32. Выведите аналогичную (1.19) оценку погрешности функции одной переменной, используя для этого теорему 2.9 (см. п. 2.5.3).

1.33. Дана функция $f: f(x, y) = x^2 + y^2$ и ее приближенные аргументы $a = 1,1$, $b = 2,5$ с верными значащими цифрами.

1. Найдите по формуле (1.20) абсолютную дифференциальную погрешность $\bar{\Delta}_z$, значения $z = f(a, b)$ и оцените соответствующую дополнительную погрешность.

2. Найдите абсолютную погрешность Δ_z по формуле (1.19) и сравните ее с $\bar{\Delta}_z$.

1.8. Оценка погрешностей арифметических операций

При оценке погрешностей суммы или разности двух приближенных чисел как значений функций надо иметь в виду то, что приращения функций $z = x \pm y$ совпадают с их дифференциалами, и потому $\bar{\Delta}_z = \Delta_z$.

Теорема 1.3. 1. Абсолютная погрешность суммы и разности двух приближенных чисел a и b равна сумме абсолютных погрешностей слагаемых:

$$\Delta_{a \pm b} = \Delta_a + \Delta_b. \quad (1.23)$$

2. Относительная дифференциальная погрешность частного и произведения двух не равных нулю приближенных чисел a и b равна сумме относительных погрешностей этих чисел:

$$\bar{\delta}_{a \cdot b} = \delta_a + \delta_b; \quad \bar{\delta}_{a/b} = \delta_a + \delta_b. \quad (1.24)$$

○ Докажем вторую из формул (1.24), остальные доказываются аналогично. Функция $f: f(x, y) = \frac{x}{y}$ дифференцируема в области определения. По формуле (1.21) имеем

$$\bar{\delta}_{a/b} = \frac{|b|}{|a|} \left(\frac{1}{|b|} \Delta_a + \frac{|a|}{|b|^2} \Delta_b \right) = \frac{\Delta_a}{|a|} + \frac{\Delta_b}{|b|} = \delta_a + \delta_b. \quad \bullet$$

В следующем примере черточки в обозначениях оценок дифференциальных погрешностей опускаем.

Пример 1.12. Пусть в выражении $d = 2,63 - 1,026 \cdot \sqrt{5,40}$ все числа приближенные и записаны с верными цифрами. Требуется найти значение d и определить абсолютную и относительную погрешности.

Для оценки погрешности значения этого выражения подошел бы аналог теоремы 1.2 для функции трех переменных, примененный к функции $f: f(x, y, z) = x - y\sqrt{z}$, однако теперь есть возможность провести расчеты с учетом погрешностей результатов на каждом шагу вычислений. Для этого обозначим: $a = 5,40$, $b = \sqrt{a}$, $c = 1,026 \cdot b$. Тогда $d = 2,63 - c$.

Пока не известно, со сколькими значащими цифрами надо вычислить $b = \sqrt{5,40}$, так как абсолютная погрешность числа b неизвестна. Чтобы подсчитать Δ_b , возьмем грубо $b = \sqrt{5,40} \approx 2,3$. Поскольку $\Delta_a = 0,005$, на основании (1.22) будем иметь

$$\Delta_b = \frac{\Delta_a}{2 \cdot \sqrt{a}} = \frac{0,005}{2 \cdot 2,3} = 0,0011.$$

Следовательно, у числа b две цифры после десятичной запятой будут верными и его надо вычислить более точно: $b \approx 2,324$ (с одной сомнительной цифрой). Относительная погрешность

$$\delta_b = \frac{\Delta_b}{|b|} = \frac{0,0011}{2,3} = 0,00048 \text{ } ^1).$$

Оценим погрешность произведения $c = 1,026 \cdot b = 2,384424$, взяв для этого $c \approx 2,4$. Учитывая, что $\Delta_{1,026} = 0,0005$, подсчитаем сначала $\delta_{1,026} = 1,026 \cdot 0,0005 = 0,00052$, а затем $\delta_c = 0,00052 + 0,00048 = 0,001$ и $\Delta_c = 2,4 \cdot 0,001 = 0,0024$. Последний результат говорит о том, что в произведении три верные цифры и для дальнейших вычислений надо брать $c \approx 2,384$.

Первое число в разности $d = 2,63 - c$ имеет абсолютную погрешность $0,005$, поэтому $\Delta_d = 0,005 + 0,0024 = 0,0074$. В итоге получим результат $d \approx 0,25$ с одной верной значащей цифрой. Его относительная погрешность $\delta_d = \frac{0,0074}{0,25} = 0,03 = 3\%$. •

Отметим, что результат в примере 1.12 получен со значительно меньшей точностью по сравнению с точностью исходных данных как по количеству верных значащих цифр, так и по значению относительной погрешности (найдите ее для всех чисел исходного выражения!). Причиной этого явилось последнее действие, в котором разность двух чисел с тремя верными значащими цифрами оказалась только с одной такой цифрой.

Возникшая здесь при вычитании ситуация, называемая *потерей точности*, не случайна. Пусть $a > 0$, $b > 0$ и $x = a - b$. Тогда

$$\delta_x = \frac{\Delta_x}{|x|} = \frac{\Delta_a + \Delta_b}{|a - b|},$$

¹⁾ Напоминаем, что лишние значащие цифры при получении оценок погрешностей бесполезны, поэтому при вычислении δ_b число $2,324$ заменили на $2,3$.

откуда понятно, что даже если абсолютные погрешности Δ_a и Δ_b невелики, но числа a и b очень мало отличаются друг от друга (модуль их разности значительно меньше самих чисел), то показатель качества вычислений — относительная погрешность — может оказаться большим. При этом верные цифры у разности могут вообще исчезнуть (упр. 1.39).

Приведем полезную при учете погрешностей формулу. Если k — точное число, x — приближенное число, и они оба не равны нулю, то

$$\Delta_{kx} = |k| \Delta_x. \quad (1.25)$$

Упражнения

1.34. В теореме 1.3 для каждого из арифметических действий даны формулы вычисления либо абсолютной, либо относительной погрешности. Как найти при этом другую оценку погрешности?

1.35. Докажите формулы (1.23) и первую из формул (1.24) на основании теоремы 1.2.

1.36. Докажите формулы (1.23) без применения теоремы 1.2, пользуясь свойствами модулей чисел.

1.37. Найдите значение выражения $z = \frac{\ln x + y}{x - y}$ и оцените его погрешность, если $x = 1,23$, $y = 0,874$, $\Delta_x = 0,002$, $\Delta_y = 0,0004$.

1.38. Выполните упражнение 1.29 с помощью правил учета погрешностей арифметических действий.

1.39. Пусть у приближенных чисел $a = 5,245$, $b = 5,244$ все цифры верные. Покажите, что у разности $a - b$ нет ни одной верной значащей цифры. Сравните относительные погрешности исходных данных и полученной разности.

1.40. Докажите утверждение (1.25).

Глава 2. ПРИБЛИЖЕННОЕ РЕШЕНИЕ УРАВНЕНИЙ

2.1. Постановка задачи

В данной главе будут рассмотрены способы решения нелинейных уравнений с одним неизвестным. Каждое такое уравнение можно представить в виде

$$f(x) = 0. \quad (2.1)$$

Область определения D_f функции f называем также *областью определения уравнения*.

Число t называется *корнем уравнения* (2.1) или *нулем функции* f , если при подстановке его вместо x уравнение превращается в верное равенство. Решить уравнение — это значит найти множество всех его корней. Два уравнения называются *эквивалентными*, или *равносильными*, если у них одинаковые множества корней.

Корни уравнения могут быть действительными и комплексными. Здесь будем заниматься поиском только действительных изолированных корней. Корень называется *изолированным*, если существует такой ненулевой интервал, в котором этот корень единственный.

Школьные и вузовские курсы математики знакомят с классами уравнений, которые можно решить точно с помощью аналитических преобразований и специальных формул. Однако точные методы не являются основными в практике вычислений.

Во-первых, уравнений, для которых существуют точные методы решения, сравнительно мало. Например, не удастся решить точно уравнения $\sin x - e^x = 0$ и $P_n(x) = 0$, где P_n — многочлен общего вида степени $n \geq 5$. Во-вторых, многие точные методы настолько трудоемки, что применять их нецелесообразно. В-третьих, очень часто точное решение уравнений и не является необходимым: проблема поиска корня считается решенной, если найдено его приближение с заданной степенью точности.

Процесс приближенного решения уравнений распадается на два этапа: 1) отделение корней; 2) уточнение корней.

О п р е д е л е н и е 2.1. Будем говорить, что корень t *отделен на отрезке* $[a; b] \subset D_f (a < b)$, если $t \in (a; b)$ и других корней в этом отрезке нет. При этом $[a; b]$ называем *отрезком изоляции* корня t .

Отделить корни уравнения (2.1) — значит для каждого из корней найти свои отрезки изоляции.

О п р е д е л е н и е 2.2. Поиск приближенного значения корня с точностью до заданного достаточно малого числа $\epsilon > 0$ называется *уточнением* этого корня.

Следовательно, задача уточнения будет решена, если найдется число x такое, что $|t - x| \leq \epsilon$. Тогда $t \approx x$ с точностью до ϵ .

Если отрезок $[a; b]$ изоляции корня t найден, то любое число из него можно взять в качестве приближенного корня. Например, $t \approx a$, $t \approx b$, $t \approx x = (a + b)/2$. Пусть $h = b - a$ — длина отрезка. Поскольку $|t - a| \leq h$, $|t - b| \leq h$, концы отрезка являются приближениями к t с точностью до h . Легко проверить, что x имеет точность до $h/2$, т. е. обладает лучшей характеристикой точности¹⁾.

Чем меньше длина отрезка изоляции, тем выше точность приближения, однако с помощью используемых для отделения корней приемов (см. 2.2) получить отрезок достаточно малой длины трудно. Нужны специальные *методы уточнения корней*. Далее в главе будет рассмотрено несколько таких методов, которые реализуют следующие два способа поиска приближенного корня с заданной точностью $\epsilon > 0$:

1. Последовательно уменьшая длины отрезка изоляции корня по какому-либо правилу, отыскивается отрезок $[a; b]$ такой, что $t \in [a; b]$ и $b - a \leq 2\epsilon$. Тогда приближенным корнем требуемой точности будет середина отрезка $[a; b]$: $x = (a + b)/2$.

2. Строится последовательность чисел $x_0, x_1, \dots, x_n, \dots$, сходящаяся к корню t . Как только окажется $|t - x_n| \leq \epsilon$, можно положить $t \approx x_n$. Такая последовательность называется *последовательностью приближений*, а определяющий ее метод — *методом последовательных приближений*.

Первый способ удобен тем, что позволяет легко устанавливать завершение процесса уточнения, поскольку отрезки изоляции и их длины на каждом шаге вычислений известны.

Непосредственную проверку условия $|t - x_n| \leq \epsilon$ из второго способа проводить не удастся, ибо неизвестен точный корень t . В то же время для каждого метода последовательных приближений есть возможность получить неравенства вида

$$|t - x_n| \leq V_n \quad (n = 1, 2, \dots). \quad (2.2)$$

Здесь V_n — числовое выражение, значения которого при каждом n характеризуют степень близости приближения x_n к корню, обеспечиваемую данным методом²⁾. Из (2.2) следует, что за условие окончания процесса приближений можно взять неравенство: $V_n \leq \epsilon$.

¹⁾ На самом деле неизвестно, будет ли именно середина отрезка лучшим приближением, так как точный корень может оказаться ближе к одному из концов отрезка.

²⁾ Часто для одного и того же метода удается найти несколько выражений V_n , т. е. несколько оценок точности получаемых с его помощью приближений.

Неравенство (2.2) позволяет решать и задачу определения абсолютной погрешности каждого приближения x_n , ибо ясно, что $\Delta_{x_n} = V_n$.

З а м е ч а н и е. Для того чтобы погрешности вычислений и округлений не оказывали существенного влияния на степень точности результатов, все вычисления надо вести либо точно, либо с одной-двумя запасными цифрами. Пусть, например, требуется найти приближенный корень уравнения $x - \sqrt[3]{x} - \pi = 0$ с тремя верными цифрами после десятичной запятой ($\epsilon = 0,0005$). Если при этом использовать приближение $\pi \approx 3,14$, то заданная точность не будет достигнута. Здесь число π следует округлить по крайней мере до четырех верных цифр после десятичной запятой: $\pi \approx 3,1416$. Надо также иметь средства вычисления значений $\sqrt[3]{x}$ с не меньшей точностью.

Упражнение

2.1. Дано уравнение

$$f(x) = 1, \text{ где } f(x) = \begin{cases} 1, & x \in [0; 1]; \\ x^2, & x \in \mathbf{R} \setminus [0; 1]. \end{cases}$$

Найдите все корни этого уравнения и определите, какие из них являются изолированными, а какие нет.

2.2. Отделение корней

Отделение корней обычно производится графически и (или) аналитически.

Графический способ отделения корней уравнения (2.1) заключается в поиске таких отрезков $[a; b] \subset D_f$, внутри которых находится абсцисса точки пересечения графика функции $y = f(x)$ с осью Ox , т. е. нуль функции f . На рис. 2.1 отмечены два корня и соответствующие им отрезки.

Во многих случаях оказывается удобным заменить уравнение (2.1) равносильным ему уравнением

$$p(x) = g(x), \quad (2.3)$$

где p и g — функции, графики которых строятся легче, чем график функции f . Корнями уравнения (2.3) являются абсциссы точек пересечения линий $y = p(x)$ и $y = g(x)$ (рис. 2.2).

Графический способ применяется наиболее часто. Обычно с него

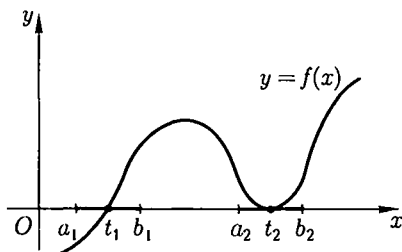


Рис. 2.1

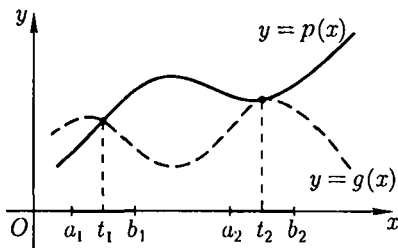


Рис. 2.2

и начинают отделение корней, однако он не обладает большой точностью, и потому делать окончательные выводы на его основе не всегда удастся.

При схематическом изображении графиков можно ошибиться в выборе отрезка изоляции корня. Это случается, когда необходим отрезок небольшой длины или корни расположены близко. Здесь

надо учесть, что последующие методы уточнения требуют от функций определенных свойств в окрестности корня. Чем мельче отрезок, тем вероятность выполнения этих свойств выше. Бывает также, что на некоторых участках области определения уравнения графики функций располагаются настолько близко, что установить факт их пересечения или количество пересечений трудно.

Таким образом, результаты графического метода необходимо проверять и уточнять. Делается это аналитическими методами.

Аналитический способ. Приведенные ниже теоремы составляют основу аналитического способа отделения корней.

Теорема 2.1 (первая теорема Больцано – Коши). Пусть функция f определена и непрерывна на отрезке $[a; b]$, причем на концах его принимает значения разных знаков, т. е. $f(a) \cdot f(b) < 0$. Тогда существует по крайней мере одна точка $t \in (a; b)$, в которой значение функции равно нулю [3, гл. IV, § 5].

Теорема 2.2. Если непрерывная на отрезке $[a; b]$ функция f имеет на $(a; b)$ конечную производную и $f'(x) > 0$ ($f'(x) < 0$) для всех $x \in (a; b)$, то она возрастает (убывает) на этом отрезке [3, гл. VII, § 1].

Теорема 2.1 гарантирует наличие корня уравнения (2.1) в отрезке $[a; b]$, но не единственность его. Как видно из рис. 2.1, на отрезке $[a_1; b_2]$ выполняются условия теоремы: f непрерывна, $f(a_1) < 0, f(b_2) > 0$, однако там два корня t_1 и t_2 .

Понятно, что корень, о котором говорится в теореме 2.1, будет единственным, если дополнительно потребовать от функции строгую монотонность на отрезке (рис. 2.1, отрезок $[a_1; b_1]$). Для проверки монотонности можно воспользоваться теоремой 2.2.

При чисто аналитическом отделении корней уравнения (2.1) полезно вычислить односторонние пределы функции f (если они существуют) в точках разрыва, в концах промежутков определения, а также в $\pm \infty$. Знаки этих пределов помогут установить факт пересечения графика функции с осью Ox (пример 2.1).

Пример 2.1. Уравнение $x^5 - 4x^2 + x = 0$ имеет по крайней мере один корень, так как функция $f: f(x) = x^5 - 4x^2 + x$ определена и непрерывна на всей действительной прямой и

$$\lim_{x \rightarrow -\infty} f(x) = -\infty, \quad \lim_{x \rightarrow +\infty} f(x) = +\infty. \bullet$$

Далее берется несколько промежуточных точек из области непрерывности функции f , подбор которых зависит от ее свойств, и вычисляются значения функции в них¹⁾. Если в двух выбранных соседних точках из одного и того же промежутка непрерывности функции произойдет смена знака, то, согласно теореме 2.1, между ними есть корень.

Следует заметить, что теорем 2.1 и 2.2 недостаточно для обнаружения и изоляции всех корней уравнения $f(x) = 0$. Корень может быть внутри $[a; b]$ и при одинаковых знаках $f(a), f(b)$ (рис. 2.1, отрезок $[a_2; b_2]$). Такими корнями могут быть точки максимума или минимума функции f . Чтобы их обнаружить, не требуется проводить полное исследование функции на экстремум, достаточно найти подозрительные на экстремум точки t и вычислить в них значение функции. Если $f(t) = 0$, то t — корень.

Полезно знать также теорему Ролля [3, гл. VI, § 1]:

Т е о р е м а 2.3. Пусть на $[a; b]$ определена функция f , причем:

- 1) f непрерывна на $[a; b]$;
- 2) на $(a; b)$ существует конечная производная f' ;
- 3) $f(a) = f(b)$.

Тогда найдется точка $c \in (a; b)$ такая, что $f'(c) = 0$.

Теорема Ролля помогает устанавливать, сколько всего различных корней может быть у уравнения.

П р и м е р 2.2. Докажем, что уравнение $e^x = x + 2$ имеет не более двух корней. Здесь $f(x) = e^x - x - 2$. Эта функция определена и дифференцируема на \mathbf{R} . Ее производная $f'(x) = e^x - 1$ равна нулю в одной точке $x = 0$. Предположим, что уравнение имеет, например, три различных корня $-t_1, t_2, t_3$. Тогда $f(t_1) = f(t_2) = f(t_3) = 0$, т. е. значения функции в трех точках равны. В этом случае, согласно теореме Ролля, между ними найдутся две точки, где производная обращается в нуль. А это противоречит установленному факту. •

Однако теорема Ролля не утверждает, что корни есть на самом деле. Чтобы проверить их наличие, надо воспользоваться теми методами, что были приведены выше. Например, для уравнения $e^x = x$ имеет место ситуация, аналогичная описанной в примере 2.2, однако это уравнение корней не имеет вообще (убедитесь графически!). А вот функция f из этого примера меняет знак на отрезках $[-2; -1,5]$ и $[1; 1,5]$, и потому уравнение $e^x = x + 2$ имеет два различных корня (но не больше).

В следующих параграфах излагаются методы уточнения корней.

¹⁾ В число этих точек обязательно надо включить точки, *подозрительные на экстремум*. Напоминаем, что таковыми называются точки, в которых выполняется *необходимое условие экстремума*: первая производная либо равна нулю, либо обращается в бесконечность, либо не существует [3, гл. VII, § 2].

Упражнения

2.2. Даны уравнения: а) $\ln x - \frac{1}{x} = 0$, б) $3\cos x = x + 1$. Отделите их корни графически и проверьте результаты аналитически.

2.3. Убедитесь, что условие непрерывности функции в теореме 2.1 является обязательным для существования корня. Для этого можно взять, например, уравнение

$$\operatorname{sgn}(\ln x) - 0,5 = 0$$

и отрезок $[0,5; 1,5]$ ¹⁾.

2.4. С помощью теоремы Ролля докажите, что если производная дифференцируемой на некотором промежутке функции f обращается в нуль n раз ($n = 0, 1, 2, \dots$), то уравнение (2.1) может иметь на этом промежутке не более $n + 1$ различных корней. Дайте графическую иллюстрацию этому факту при $n = 3$ в случаях, когда: есть четыре корня; есть только один корень; корней нет.

2.5. Дано уравнение $x^3 - 3x + 1 = 0$.

1. Докажите с помощью пределов при $x \rightarrow \pm\infty$, что оно имеет хотя бы один корень.

2. Докажите, что у него не более трех корней.

3. Методом подбора отделите все его корни.

2.3. Метод половинного деления

Пусть дано уравнение (2.1): $f(x) = 0$.

Будем считать, что его корень t отделен на отрезке $[a; b]$. Требуется найти приближенное значение корня с точностью до ϵ , где ϵ — достаточно малое положительное число.

В основе метода половинного деления лежит *теорема о вложенных отрезках*. Последовательность отрезков

$$[a_1; b_1] \supset [a_2; b_2] \supset \dots \supset [a_n; b_n] \supset \dots$$

называется *вложенной*. При условии, что длины отрезков $b_n - a_n \rightarrow 0$ при $n \rightarrow \infty$, эта последовательность называется *стягивающейся*.

Теорема 2.4 (теорема Кантора). Для всякой стягивающейся последовательности вложенных отрезков существует единственная точка t , принадлежащая всем отрезкам этой последовательности [3, гл. III, § 8].

Предположив, что функция f непрерывна на отрезке $[a; b]$ и на его концах принимает значения разных знаков, научимся определять последовательность вложенных отрезков $[a_n; b_n] \subset [a; b]$ ($n =$

¹⁾ $\operatorname{sgn}(x) = 1$ при $x > 0$; 0 при $x = 0$; -1 при $x < 0$.

$= 1, 2, \dots$), стягивающуюся к корню t . Тем самым получим метод, реализующий первый из отмеченных в 2.1 способов уточнения корней.

Итак, пусть f непрерывна на $[a; b]$ и $f(a) \cdot f(b) < 0$ (рис. 2.3).

Разделим $[a; b]$ пополам точкой $c = (a + b)/2$ и вычислим $f(c)$.

Если $f(c) = 0$, то корень t найден точно (а именно, $t = c$). Если же

нет, выберем ту половину отрезка, на концах которой значения функции разных знаков, и обозначим ее $[a_1; b_1]$ (на рис. 2.3 правая половина $[a; b]$). Затем отрезок $[a_1; b_1]$ делим пополам точкой $c_1 = (a_1 + b_1)/2$ и проводим аналогичные рассуждения. Получится либо точный корень c_1 , либо отрезок $[a_2; b_2]$ со свойством $f(a_2) \cdot f(b_2) < 0$. И так далее.

Если на каком-то шаге обнаружится точный корень (что на практике маловероятно), то процесс деления пополам закончится, если же нет, то в результате получится бесконечная последовательность вложенных отрезков $[a_1; b_1], [a_2; b_2], \dots, [a_n; b_n], \dots$, таких, что

$$f(a_n) \cdot f(b_n) < 0 \text{ при всех } n = 1, 2, \dots \quad (2.4)$$

Понятно, что $b_n - a_n = \frac{b-a}{2^n} \rightarrow 0$ при $n \rightarrow \infty$. Следовательно, полученная последовательность отрезков является стягивающейся.

Теорема Кантора гарантирует существование единственной точки, принадлежащей всем отрезкам этой последовательности. Так как f непрерывна на $[a; b]$, из (2.4) в силу теоремы 2.1 следует, что этой точкой является именно корень t .

Для того чтобы найти приближенное значение корня с точностью до $\epsilon > 0$, необходимо остановить процесс половинного деления на таком шаге n , на котором отрезок $[a_n; b_n]$ будет иметь длину

$$b_n - a_n = \frac{b-a}{2^n} \leq 2\epsilon, \quad (2.5)$$

и вычислить $x = \frac{a_n + b_n}{2}$. Ясно, что тогда можно взять $t \approx x$, причем $|t - x| \leq \epsilon$ (рис. 2.4).

Полезно иметь в виду, что вычисляемые при каждом $n = 1, 2, \dots$ середины c_n отрезков $[a_n; b_n]$ образуют на самом деле последовательность приближений к корню. При этом очевидно, что

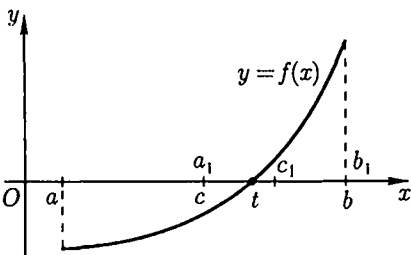


Рис. 2.3

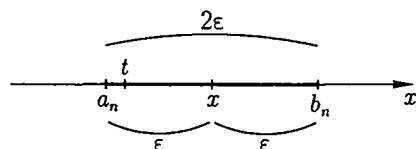


Рис. 2.4

$$\Delta_{c_n} = \frac{b_n - a_n}{2} = \frac{b - a}{2^{n+1}}.$$

Метод половинного деления дает простой и удобный алгоритм уточнения корней с любой наперед заданной степенью точности. Он требует от функции f выполнения легко проверяемых свойств: непрерывности на отрезке изоляции корня и разных знаков значений на его концах.

Отличительной чертой метода является то, что скорость приближения по нему совершенно не связана со свойствами функции в окрестности корня. Количество шагов приближения зависит лишь от отрезка $[a; b]$ и заданной точности ϵ [условие (2.5)], причем на каждом шаге абсолютная погрешность приближенного значения корня уменьшается ровно вдвое (и только!). Вследствие этого увеличение точности всегда сопряжено с пропорциональным ростом объема вычислительной работы.

В правильности сделанных выводов убеждают пример 2.3 и результаты упр. 2.7.

Пример 2.3. Единственный корень t уравнения $x^3 - x - 1 = 0$ расположен на отрезке $[1; 2]$ (проверьте графически!). Для уточнения t можно применить метод половинного деления, поскольку функция $f: f(x) = x^3 - x - 1$ непрерывна на этом отрезке и на его концах принимает значения разных знаков: $f(1) = -1 < 0$, $f(2) = 5 > 0$.

Найдем середину $c = 1,5$ отрезка $[1; 2]$ и вычислим значение функции f в этой точке: $f(1,5) = 0,875$. Значит, число 1,5 не является точным корнем. Если взять $t \approx 1,5$, то $\Delta_{1,5} = (2 - 1)/2 = 0,5$. Точность приближения невысока, но цифра 1 уже является верной.

Далее заметим, что $f(1) < 0$, $f(1,5) > 0$ и, следовательно, корень лежит на отрезке $[a_1; b_1] = [1; 1,5]$. Делим пополам полученный отрезок точкой $c_1 = 1,25$ и находим $f(1,25) \approx -0,3$. Абсолютная погрешность приближенного корня 1,25 равна 0,25. Точность повысилась, но незначительно, так как верных цифр у приближения не прибавилось.

На следующем шаге надо взять $[a_2; b_2] = [1,25; 1,5]$ и $c_2 = 1,375$. Поскольку $\Delta_{c_2} = 0,125$, снова гарантируется только одна верная цифра приближенного числа 1,375. И так далее. Видим, что уточнение идет медленно, а для отыскания приближенного корня с двумя верными цифрами необходимо провести еще два шага вычислений (выполните их!). •

З а м е ч а н и е. Если отрезок изоляции корня известен, условие непрерывности функции f перестает быть необходимым для метода половинного деления. Надо лишь, чтобы функция была определена на этом отрезке и принимала слева от корня значения одного знака, а справа — другого (упр. 2.6).

Упражнения

2.6. 1. Дано уравнение:

$$f(x) = 0, \text{ где } f(x) = \begin{cases} x - 1 & \text{при } x < 0; \\ x^2 & \text{при } x \geq 0. \end{cases}$$

Убедитесь, что, несмотря на разрыв функции f в корне $t = 0$, метод половинного деления формально можно применить в любом отрезке $[a; b]$, $a < 0 < b$.

2. Приведите пример уравнения, в котором функция разрывна в каждой точке отрезка изоляции корня, но метод применим (используйте функцию Дирихле¹⁾).

2.7. Корень некоторого уравнения отделен на отрезке $[2; 3]$. Определите, сколько шагов метода половинного деления надо выполнить для уточнения корня с точностью до $\epsilon = 0,1; 0,01; 0,001$.

2.8. Найдите с точностью до $0,05$ корень уравнения из упр. 2.5, расположенный в отрезке $[0; 1]$.

2.4. Метод хорд и метод касательных (общие вопросы)

Далее рассмотрим методы хорд и касательных уточнения корней уравнений.

Эти методы относятся к *методам последовательных приближений*. Приближения к корню находятся так: если известно предыдущее приближение x_n ($n = 0, 1, 2, \dots$), то последующее приближение x_{n+1} вычисляется по формуле

$$x_{n+1} = P(x_n), \quad (2.6)$$

где P — некоторое выражение, устанавливающее связь между предыдущим и последующим приближениями. Начинается процесс с какого-либо числа x_0 из отрезка изоляции корня — *начального приближения*.

Формула вида (2.6) называется *рекуррентной формулой*²⁾, а получаемую с ее помощью последовательность приближений называют *итерационной последовательностью*³⁾.

¹⁾ Функция Дирихле φ на некотором промежутке X определяется так: $\varphi(x) = 1$ для рациональных $x \in X$; $\varphi(x) = 0$ для иррациональных $x \in X$.

²⁾ Рекуррентная формула (от лат. *recurrens* — возвращающийся) — формула, позволяющая выразить $(n + 1)$ -й член последовательности через несколько предыдущих его членов. При наличии рекуррентной формулы (2.6) последовательность полностью определяется выбором начального приближения x_0 .

³⁾ Итерация (от лат. *iteratio* — повторение) — неоднократное применение какой-нибудь операции, в данном случае вычислений по рекуррентной формуле.

Итак, пусть дано уравнение (2.1): $f(x) = 0$, корень t которого отделен на отрезке $[a; b]$.

При уточнении корня обоими методами будем полагать, что значения $f(a)$ и $f(b)$ функции f на концах отрезка имеют разные знаки и, кроме того, функция f имеет непрерывные производные f' и f'' с отличными от нуля и сохраняющими постоянный знак значениями при всех $x \in [a; b]$.

Для выполнения приведенных условий иногда приходится сужать первоначально найденные отрезки изоляции корня.

Из курса математического анализа [3, гл. VII] известно, что при $f'(x) > 0$ на $[a; b]$ функция возрастает на этом отрезке; при $f'(x) < 0$ — убывает; при $f''(x) > 0$ график функции вогнутый, а при $f''(x) < 0$ — выпуклый.

Возможны четыре случая:

- 1) $f' > 0, f'' > 0$ — функция возрастает, график вогнутый;
- 2) $f' < 0, f'' < 0$ — функция убывает, график выпуклый;
- 3) $f' > 0, f'' < 0$ — функция возрастает, график выпуклый;
- 4) $f' < 0, f'' > 0$ — функция убывает, график вогнутый.

От того, какой из случаев имеет место, зависят рекуррентная формула (в методе хорд) и правило выбора начального приближения x_0 (в обоих методах).

Покажем, что, заменяя подходящим образом при различных комбинациях знаков производных исходную функцию f другими функциями g , можно вместо (2.1) рассматривать уравнения вида

$$g(x) = 0, \quad (2.1')$$

для которых имеет место только один из перечисленных случаев (например, первый).

С л у ч а й 1. $f'(x) > 0$ и $f''(x) > 0$ на $[a; b]$. Можно считать его основным и именно при таких условиях решать проблему уточнения корней обоими методами. Соответствующий график функции изображен на рис. 2.3.

С л у ч а й 2. $f'(x) < 0, f''(x) < 0$ на $[a; b]$. Возьмем функцию $g: g(x) = -f(x)$ (рис. 2.5, а). Тогда $g'(x) = -f'(x) > 0, g''(x) = -f''(x) > 0$ и приходим к первому случаю. Заметим, что здесь уравнения (2.1) и (2.1') равносильны.

С л у ч а й 3. $f'(x) > 0$ и $f''(x) < 0$ на $[a; b]$. Пусть $g: g(x) = -f(-x)$ (рис. 2.5, б). Тогда корнем уравнения (2.1') будет $-t$, а отрезком его изоляции является $[-b; -a]$. Для доказательства положительности производных функции g на $[-b; -a]$ обозначим $z = -x$. Очевидно, что $x \in [-b; -a] \Leftrightarrow z \in [a; b]$, поэтому

$$g'(x) = (-f(z))'_x = -f'(z)z'_x = f'(z) > 0;$$

$$g''(x) = (f'(z))'_x = f''(z)z'_x = -f''(z) > 0.$$

С л у ч а й 4. $f'(x) < 0$ и $f''(x) > 0$ на $[a; b]$. Положим $g: g(x) = f(-x)$ (рис. 2.5, в). Такими же рассуждениями, как в случае 3, показывает, что производные g' и g'' положительны на $[-b; -a]$.

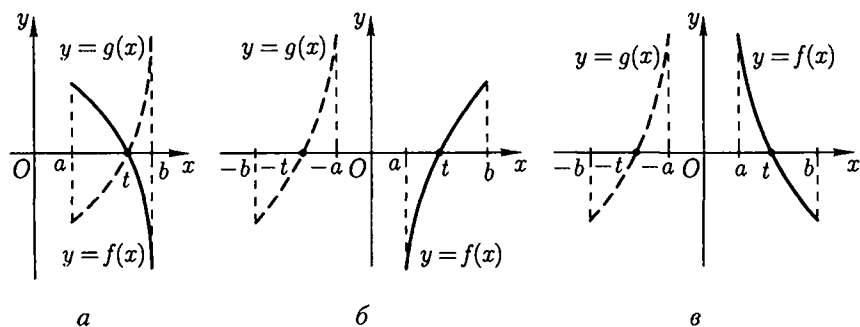


Рис. 2.5

Итак, проведя предварительные исследования, можно прийти к уравнению вида (2.1) с положительными производными $f'(x)$ и $f''(x)$ на отрезках изоляции корня. Затем в случаях 3 и 4 вместо найденного приближения к корню уравнения (2.1') надо взять это приближение с обратным знаком.

Пример 2.4. Корень уравнения $\ln x = -x$ отделен на отрезке $[0,2; 1]$. Здесь $f(x) = \ln x + x$. На данном отрезке $f'(x) = \frac{1}{x} + 1 > 0$, $f''(x) = -\frac{1}{x^2} < 0$. Следовательно, имеет место случай 3 и исходное уравнение можно заменить уравнением вида $-f(-x) = 0$:

$$-(\ln(-x) - x) = 0 \Leftrightarrow x - \ln(-x) = 0.$$

Корень полученного уравнения лежит на отрезке $[-1; -0,2]$. Нелегко проверить, что производные g' и g'' функции $g: g(x) = x - \ln(-x)$ положительны на $[-1; -0,2]$. •

Упражнение

2.9. Для уравнений: а) $\operatorname{arctg} x - 0,5x = 0$, б) $1 - x = \sin x$ найдите отрезки изоляции корней, где применимы методы хорд и касательных. Выберите для уточнения каждого корня соответствующие уравнение и отрезок так, чтобы значения первой и второй производных функции f были положительны на этом отрезке.

2.5. Метод хорд

Пусть выполнены все условия, приведенные в 2.4 относительно уравнения (2.1): $f(x) = 0$.

2.5.1. Рекуррентная формула

Предположим, что производные f' и f'' положительны на $[a; b]$. Тогда $f(a) < 0$, $f(b) > 0$ (рис. 2.6).

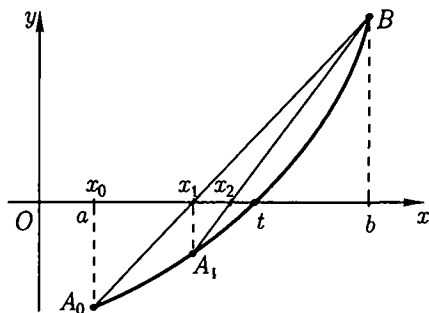


Рис. 2.6

Построим итерационную последовательность, взяв в качестве x_0 левый конец отрезка — число a . Соединим точки $A_0(x_0; f(x_0))$ и $B(b; f(b))$ отрезком (хордой). Абсциссу точки пересечения хорды A_0B с осью Ox возьмем в качестве x_1 . Уравнение хорды:

$$\frac{x - x_0}{b - x_0} = \frac{y - f(x_0)}{f(b) - f(x_0)}.$$

Положив в этом уравнении $y = 0$, получим $x = x_1$. Следовательно,

$$x_1 = x_0 - \frac{b - x_0}{f(b) - f(x_0)} f(x_0).$$

Далее напишем уравнение хорды, соединяющей точки $A_1(x_1; f(x_1))$ и $B(b; f(b))$, и при $y = 0$ получим $x = x_2$ (абсциссу точки пересечения хорды A_1B с осью Ox):

$$x_2 = x_1 - \frac{b - x_1}{f(b) - f(x_1)} f(x_1).$$

Продолжая подобным образом, получим итерационную последовательность, вычисляемую по рекуррентной формуле

$$x_{n+1} = x_n - \frac{b - x_n}{f(b) - f(x_n)} f(x_n) \quad (n = 0, 1, 2, \dots), \quad (2.7)$$

где в качестве x_0 выбран левый конец a отрезка $[a; b]$, а правый конец b этого отрезка остается неподвижным.

Пример 2.5. Корень уравнения $x^3 - x - 1 = 0$ изолирован в отрезке $[0; 2]$, причем $f(0) = -1 < 0$, $f(2) = 5 > 0$, однако на этом отрезке необходимые для применения метода хорд условия не выполняются, так как $f'(x) = 3x^2 - 1$ меняет знак в точке $x = \frac{1}{\sqrt{3}}$, а $f''(x) = 6x = 0$ при $x = 0$.

Легко проверить, что на более узком отрезке изоляции $[1; 2]$ для уточнения корня применима рекуррентная формула (2.7) с $x_0 = 1$. Запишем ее для данного уравнения и вычислим x_1 с тремя значащими цифрами. Здесь $b = 2$, $f(b) = 5$, $f(x_n) = x_n^3 - x_n - 1$, поэтому

$$x_{n+1} = x_n - \frac{(2 - x_n)(x_n^3 - x_n - 1)}{6 - x_n^3 + x_n}.$$

Подставив $x_0 = 1$ и округлив результат вычисления до трех значащих цифр, получим $x_1 = 1,17$. •

Вид рекуррентных формул и правило выбора x_0 для других комбинаций знаков f' и f'' указаны в упр. 2.10 и 2.11.

Упражнения

2.10. Выведите рекуррентную формулу метода хорд для случая $f' < 0$, $f'' < 0$ на $[a; b]$ (без преобразования уравнения) и убедитесь, что она совпадает с (2.7), причем $x_0 = a$.

2.11. Покажите, что когда знаки f' и f'' на $[a; b]$ различны, рекуррентная формула для уравнения (2.1) имеет вид

$$x_{n+1} = x_n - \frac{x_n - a}{f(x_n) - f(a)} f(x_n) \quad (n = 0, 1, 2, \dots), \quad (2.8)$$

при этом надо брать $x_0 = b$ (левый конец a неподвижен).

2.12. Отделите корни уравнений: $\cos x - x = 0$; $\cos x + x = 0$ так, чтобы на отрезках изоляции выполнялись условия метода хорд, и запишите соответствующие итерационные формулы с конкретными данными.

2.13. Убедитесь в применимости метода хорд для уравнения $\frac{1}{x} - 1 = 0$ на отрезке $[0,5; 2]$ изоляции его корня и вычислите с тремя значащими цифрами приближения x_1, x_2 .

2.5.2. Сходимость итерационной последовательности

Оказывается, что при выполнении наложенных выше на уравнение (2.1) условий итерационная последовательность метода хорд всегда сходится к корню. Чтобы доказать это, приведем сначала дополнительные сведения из математического анализа.

Числовая последовательность (x_n) называется *возрастающей* (*убывающей*), если $x_n < x_{n+1}$ ($x_n > x_{n+1}$) при всех $n = 0, 1, 2, \dots$.

Теорема 2.5. *Любая возрастающая и ограниченная сверху числом t последовательность имеет предел, не превышающий t . Любая убывающая и ограниченная снизу числом t последовательность имеет предел, не меньший t* [3, гл. III, § 3, 6].

Теорема 2.6 (предельный переход в равенствах). *Если числовые последовательности (x_n) и (y_n) имеют пределами соответственно числа a и b , причем $x_n = y_n$ для всех n , начиная с некоторого номера, то $a = b$* [3, гл. III, § 3].

Теорема 2.7. *Если две числовые последовательности имеют пределами числа a и b , то сумма, разность, произведение и частное этих последовательностей также имеют предел, равный соответственно $a + b$, $a - b$, $a \cdot b$, a / b (при $b \neq 0$)* [3, гл. III, § 4].

Далее будет применяться *метод математической индукции*. Напомним его суть. Пусть требуется доказать истинность некоторого

утверждения $Q(n)$ при всех $n = 0, 1, \dots$. Для этого вначале устанавливается истинность $Q(0)$. Затем из истинности $Q(n)$ при произвольном n выводится истинность $Q(n+1)$. Если оба этапа реализованы, то требуемое доказательство получено.

Установление сходимости итерационной последовательности метода хорд в теореме 2.8 производится в предположении, что $f' > 0$, $f'' > 0$ на $[a; b]$, однако это нетрудно сделать и при других комбинациях знаков производных.

Теорема 2.8. Пусть корень t уравнения (2.1) отделен на отрезке $[a; b]$, $f(a) \cdot f(b) < 0$ и, кроме того, производные f' , f'' непрерывны и имеют положительные значения на всем $[a; b]$. Тогда последовательность (x_n) , определяемая формулой (2.7) при начальном приближении $x_0 = a$, сходится к корню t .

○ Доказательство проведем, опираясь на геометрический факт, характеризующий взаимное расположение хорды и соответствующей части графика функции (на рис. 2.6 он очевиден):

Пусть функция f возрастает и ее график вогнут и пусть точка $M(x, y)$ лежит на части графика, определяемой двумя различными точками A и B этого графика, причем $M \neq A$ и $M \neq B$ ¹⁾. Тогда точка (x^*, y) хорды AB расположена строго левее точки M (т.е. $x^* < x$), а точка (x, y^*) этой хорды расположена строго выше точки M (т.е. $y^* > y$).

Методом математической индукции убедимся в том, что последовательность (x_n) возрастает и ограничена сверху. А именно, докажем, что $x_n < x_{n+1} < t$ при всех $n = 0, 1, 2, \dots$.

По условию теоремы имеем $x_0 < t < b$, поэтому в силу возрастания f верны неравенства $f(x_0) < f(t) < f(b)$. Следовательно, точка $(t, f(t)) = (t, 0)$ принадлежит дуге (A_0B) , а по отмеченным выше свойствам точек хорды и кривой точка $(x_1, 0)$ хорды A_0B находится строго левее точки $(t, 0)$. Значит, $x_1 < t$. С учетом неравенств $b - x_0 > 0$, $f(b) - f(x_0) > 0$, $f(x_0) < 0$, из (2.7) при $n = 0$ получим

$$x_1 - x_0 = -\frac{b - x_0}{f(b) - f(x_0)} f(x_0) > 0.$$

Таким образом, $x_0 < x_1 < t$, т.е. справедливость доказываемого утверждения при $n = 0$ получена. Пусть теперь $x_{n-1} < x_n < t$. Покажем, что тогда $x_n < x_{n+1} < t$.

Функция f возрастает, поэтому $f(x_n) < f(t) = 0$. Это означает, что точка $A_n(x_n, f(x_n))$ находится ниже оси Ox , точка $(t, 0)$ — на дуге (A_nB) , а соответствующая ей точка $(x_{n+1}, 0)$ хорды A_nB — строго левее. Следовательно, $x_{n+1} < t$. Неравенство $x_n < x_{n+1}$ доказывается так же, как неравенство $x_0 < x_1$.

Итак, метод хорд порождает возрастающую ограниченную последовательность $x_0 < x_1 < x_2 < \dots < x_n < \dots < t < b$. Согласно теоре-

¹⁾ Далее для краткости будем записывать: точка M принадлежит дуге (AB) .

ме 2.5, она сходится к некоторому числу z , причем $a < z \leq t < b$. Если перейти к пределу в равенстве (2.7), то, учитывая непрерывность f и применяя теоремы 2.6 и 2.7, будем иметь

$$z = z - \frac{b - z}{f(b) - f(z)} f(z).$$

Ввиду того что $b \neq z$, получим $f(z) = 0$. А поскольку корень уравнения (2.1) на $[a; b]$ единственный, то $t = z$. •

Упражнение

2.14. 1. Сформулируйте аналоги теоремы 2.8 для последовательностей, полученных по формулам (2.7) или (2.8) в случаях 2–4 относительно производных (без преобразования исходного уравнения).

2. Докажите эти теоремы по схеме доказательства теоремы 2.8, предварительно охарактеризовав взаимное расположение хорд и соответствующих точек графика $y = f(x)$ в данных случаях.

Обратите внимание на то, что каждый раз имеет место монотонная сходимость приближений к корню (слева или справа).

2.5.3. Оценка погрешностей приближений

Приближения x_n , найденные методом хорд, стремятся к неизвестному корню t с одной стороны. Поэтому так же легко, как в методе половинного деления, проверять условие окончания итерационного процесса здесь не удастся. Необходимы специальные формулы оценки сверху расстояния $|t - x_n|$, для вывода которых воспользуемся теоремой Лагранжа [3, гл. VI, § 1]:

Теорема 2.9. Пусть функция f определена на некотором промежутке и имеет на нем конечную производную f' . Если две точки x и x_0 находятся в данном промежутке, то существует точка c между x и x_0 такая, что

$$f(x) - f(x_0) = f'(c)(x - x_0)^1. \quad (2.9)$$

Как будет видно далее, теорема Лагранжа находит широкое применение в вопросах получения оценок погрешностей (и не только в них!). Это одна из важнейших теорем математического анализа, лежащих в основе численных методов.

Следующая теорема интересна тем, что она дает оценку точности последовательных приближений, получаемых как методом хорд, так и любым другим методом.

Теорема 2.10. Пусть корень t уравнения (2.1) отделен на отрезке $[a; b]$, и все члены некоторой последовательности (x_n) приближений к t расположены в этом отрезке. Если производная f' конечна

¹⁾ Данная формула называется формулой Лагранжа.

на $[a; b]$ и существует такое число $m > 0$, что $|f'(x)| \geq m$ для всех $x \in [a; b]$ ¹⁾, то имеет место неравенство:

$$|t - x_n| \leq \frac{|f(x_n)|}{m} \quad (n = 0, 1, 2, \dots). \quad (2.10)$$

○ Возьмем произвольное x_n . По теореме Лагранжа найдется такая точка c между x_n и t , что $|f(t) - f(x_n)| = |f'(c)||t - x_n|$. Так как t — корень, то $f(t) = 0$ и потому

$$|t - x_n| = \frac{|f(x_n)|}{|f'(c)|}.$$

Осталось сослаться на то, что $|f'(c)| \geq m$. •

Теперь приведем непосредственно связанную с методом хорд формулу, позволяющую оценивать погрешность x_n по известным приближениям x_n и x_{n-1} .

Теорема 2.11. Пусть первая и вторая производные функции f уравнения (2.1) непрерывны и сохраняют постоянный знак на отрезке $[a; b]$ изоляции корня t , $f(a)$ и $f(b)$ разных знаков, а числа m и M такие, что $0 < m \leq |f'(x)| \leq M$, $x \in [a; b]$ ²⁾. Тогда погрешности приближений к t , найденных методом хорд, оцениваются формулой

$$|t - x_n| \leq \frac{M - m}{m} |x_n - x_{n-1}| \quad (n = 1, 2, \dots). \quad (2.11)$$

○ Доказательство проведем в предположении, что приближения вычисляются по формуле (2.7) с $x_0 = a$. В случае применения формулы (2.8) доказательство аналогично.

Сначала отметим, что все приближения находятся в $[a; b]$. Записав по (2.7) формулу для вычисления x_n , затем выделив оттуда множитель $f(x_{n-1})$ и учтя, что $f(t) = 0$, получим:

$$f(t) - f(x_{n-1}) = \frac{f(b) - f(x_{n-1})}{b - x_{n-1}} (x_n - x_{n-1}).$$

Заменив здесь приращения функции f выражениями, найденными по формуле Лагранжа:

$$f(t) - f(x_{n-1}) = f'(c_n)(t - x_{n-1}),$$

$$f(b) - f(x_{n-1}) = f'(d_n)(b - x_{n-1}),$$

где $c_n \in (x_{n-1}; t)$, $d_n \in (x_{n-1}; b)$, будем иметь

¹⁾ Например, $m = \min_{[a, b]} |f'(x)|$, если f' непрерывна и отлична от нуля на $[a; b]$.

²⁾ Можно взять $m = \min_{[a, b]} |f'(x)|$, $M = \max_{[a, b]} |f'(x)|$.

$$f'(c_n)(t - x_{n-1}) = f'(d_n)(x_n - x_{n-1}),$$

или

$$t - x_{n-1} = \frac{f'(d_n)}{f'(c_n)}(x_n - x_{n-1}).$$

С учетом того, что $t - x_n = t - x_{n-1} - \frac{f'(c_n)}{f'(c_n)}(x_n - x_{n-1})$, получим

$$|t - x_n| = \frac{|f'(d_n) - f'(c_n)|}{|f'(c_n)|} |x_n - x_{n-1}|.$$

Поскольку значения f' имеют одинаковый знак на отрезке $[a; b]$, а в последнем соотношении этот знак безразличен, то для числителя в его правой части очевидно неравенство:

$$|f'(d_n) - f'(c_n)| \leq |M - m| = M - m.$$

Чтобы получить (2.11), осталось еще учесть то, что $|f'(c_n)| \geq m$. •

С помощью формул (2.10) и (2.11) можно решать задачи определения окончания процесса приближений и отыскания абсолютной погрешности каждого приближенного корня.

Если задана абсолютная погрешность $\varepsilon > 0$, то процесс останавливается при выполнении одного из неравенств:

$$\frac{|f(x_n)|}{m} \leq \varepsilon, \quad \frac{M - m}{m} |x_n - x_{n-1}| \leq \varepsilon.$$

Если итерационный процесс прерван на каком-то шаге $n \geq 1$ и взято $t \approx x_n$, то абсолютную погрешность приближения x_n можно найти, используя неравенства (2.10) и (2.11).

Пример 2.6. Оценим обоими способами погрешность приближения $x_1 = 1,17$ к корню $t \in [1; 2]$ уравнения $x^3 - x - 1 = 0$, найденного в примере 2.5. Производная $f'(x) = 3x^2 - 1$ функции $f: f(x) = x^3 - x - 1$ положительна и возрастает на $[1; 2]$, поэтому $2 \leq f'(x) \leq$

11. Значит, $m = 2$, $M = 11$. По формуле (2.10) $\Delta_{x_1} = \frac{|f(1,17)|}{2} = 0,3$. По

формуле (2.11) $\Delta_{x_1} = \frac{M - m}{m} |x_1 - x_0| = 4,5 \cdot |1,17 - 1| = 0,77$.

В данном случае оценка (2.10) оказалась точнее, чем (2.11). Из двух значений абсолютной погрешности в качестве $\Delta_{1,17}$ следует выбрать меньшее — число 0,3. Ввиду того что корень t с точностью до трех верных цифр равен 1,32, можно найти приближенное расстояние от числа 1,17 до корня: $|t - 1,17| \approx 0,15$. Как и следовало ожидать, оно меньше, чем 0,3. •

Надо иметь в виду, что если отрезок $[a; b]$ выбран неудачно (например, недостаточно малым), дроби $\frac{1}{m}$ и $\frac{M-m}{m}$ могут оказаться столь большими, что правые части формул (2.10) и (2.11) будут определять грубо завышенные абсолютные погрешности приближений. Для (2.10) такая ситуация возникает при близких к нулю m , а для (2.11) — когда число M намного больше m .

Зависимость оценок от размера отрезка $[a; b]$ вытекает из того, что в силу наложенных нами на уравнение (2.1) условий, функция $|f'|$ принимает свои наименьшее и наибольшее значения на концах этого отрезка (убедитесь в этом!). Следовательно, числа M и m отражают свойства функции f на концах $[a; b]$, а не в непосредственной близости от корня.

Приведенные рассуждения в определенной степени объясняют разницу в оценках, полученных в примере 2.6. По формуле (2.10) получена неплохая степень точности x_1 , поскольку число $m = 2$ сравнительно большое. Вычисленная по правой части (2.11) абсолютная погрешность оказалась завышенной из-за того, что число M существенно больше m .

Если отрезок $[a; b]$ достаточно мал, неравенство (2.11) удастся заменить более простым соотношением (упр. 2.16).

Упражнения

2.15. Даны два эквивалентных уравнения: а) $0,1x^2 - 0,1 = 0$, б) $3x^2 - 3 = 0$ и отрезок изоляции $[0,1; 2]$ их корня $t = 1$.

1. Найдите отдельно для обоих уравнений приближения x_1, x_2 методом хорд и начертите соответствующие графики. Попутно убедитесь в том, что скорость сходимости итерационной последовательности метода хорд, вообще говоря, не зависит от степени выпуклости или вогнутости графика функции f (т.е. от того, насколько близко график примыкает к хордам).

2. Запишите для обоих уравнений формулы (2.10), (2.11) и найдите с их помощью абсолютные погрешности чисел x_1, x_2 . Проведите анализ зависимости полученных оценок от свойств функций на отрезке $[0,1; 2]$.

2.16. 1. Покажите, что если значения f' на $[a; b]$ обеспечивают выполнение неравенства $M \leq 2m$, для метода хорд имеет место оценка погрешностей приближений

$$|t - x_n| \leq |x_n - x_{n-1}| \quad (n = 1, 2, \dots). \quad (2.12)$$

2. Убедитесь, что на отрезке $[1,2; 1,5]$ изоляции корня уравнения из примера 2.6 применима оценка (2.12).

2.17. Отделите корень уравнения $\sin x = -x$ так, чтобы на отрезке изоляции была применима формула оценки погрешностей (2.12).

2.18. Докажите теорему 2.11 для случая, когда f' и f'' имеют разные знаки на отрезке $[a; b]$.

2.6. Метод касательных

Метод касательных называется также *методом Ньютона*.

Пусть выполнены условия, приведенные в 2.4, относительно уравнения (2.1): $f(x) = 0$.

Здесь потребуются весь аппарат математического анализа из 2.5, а также формула Тейлора с остаточным членом в форме Лагранжа [3, гл. VI, § 3].

Теорема 2.12. Пусть функция f имеет в некотором промежутке конечные производные до $(n + 1)$ -го порядка включительно. Если числа x и x_0 принадлежат этому промежутку, то найдется число c между x и x_0 такое, что

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + \dots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n + \frac{f^{(n+1)}(c)}{(n + 1)!}(x - x_0)^{n+1}. \quad (2.13)$$

Соотношение (2.13) называется *формулой Тейлора*. Очевидно, что теорема Лагранжа является частным случаем теоремы 2.12.

2.6.1. Рекуррентная формула

Пусть t — точный корень уравнения (2.1), лежащий внутри отрезка $[a; b]$, и пусть известно какое-то приближение $x_n \approx t$, $x_n \in [a; b]$ (например, начальное приближение x_0). Тогда в силу теоремы 2.12 между t и x_n найдется точка c_n такая, что

$$0 = f(t) = f(x_n) + f'(x_n)(t - x_n) + \frac{1}{2} f''(c_n)(t - x_n)^2. \quad (2.14)$$

Если x_n достаточно близко к t , то $(t - x_n)^2$ мало, поэтому третье слагаемое в (2.14) можно отбросить и получить приближенное равенство:

$$0 \approx f(x_n) + f'(x_n)(t - x_n).$$

Отсюда $t \approx x_n - f(x_n) / f'(x_n)$. Получилось новое приближение к t , которое примем за x_{n+1} . Таким образом,

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \quad (n = 0, 1, 2, \dots). \quad (2.15)$$

Вывод рекуррентной формулы (2.15) осуществлен независимо от знака производных функции f и связанных с ними особенностей графика. Следовательно, в методе касательных рекуррентная формула одинакова для всех случаев 1–4 (см. 2.4) относительно знака f', f'' на отрезке $[a; b]$.

Геометрический смысл метода заключается в том, что приближения по нему равны абсциссам точек пересечения оси Ox и касательных к графику функции $y = f(x)$.

Покажем, что это действительно так, взяв для определенности случай $f' > 0, f'' > 0$ на $[a; b]$ (рис. 2.7). Возьмем $x_0 = b$ и проведем касательную к графику функции в точке $B_0(x_0; f(x_0))$. Ее уравнение:

$$y - f(x_0) = f'(x_0)(x - x_0).$$

Касательная пересечет ось Ox при $y = 0$. Подставив $y = 0$ в уравнение, получим абсциссу точки пересечения $x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$. Записав уравнение касательной к графику в точке $B_1(x_1; f(x_1))$, при $y = 0$

получим $x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}$ и т. д. Видим, что каждый раз абсциссы то-

чек пересечения касательных с осью Ox вычисляются по формуле (2.15), причем всегда $a < t \leq x_{n+1} < x_n \leq b$.

Как и в методе хорд, здесь важен вопрос о выборе начального приближения $x_0 \in [a; b]$. При положительных производных f', f'' в качестве x_0 нами выбран правый конец b отрезка $[a; b]$. С помощью рисунков нетрудно проверить, что если в этом случае брать $x_0 = a$, то проведенная через точку $A(x_0; f(x_0))$ касательная к графику может пересечь ось Ox за пределами $[a; b]$ (см. упр. 2.19) и x_1 может оказаться в области, где не выполняются условия метода.

Общее правило выбора x_0 приведено в упр. 2.20.

Пример 2.7. Найдем методом касательных приближение x_1 к корню $t \in [1; 2]$ уравнения $x^3 - x - 1 = 0$ из примера 2.5. Поскольку $f'(x) = 3x^2 - 1 > 0$, $f''(x) = 6x > 0$ на $[1; 2]$, выбираем $x_0 = 2$. Рекуррентная формула:

$$x_{n+1} = x_n - \frac{x_n^3 - x_n - 1}{3x_n^2 - 1}.$$

Подставим в правую часть x_0 и после округления результата до трех значащих цифр получим $x_1 = 1,55$. •

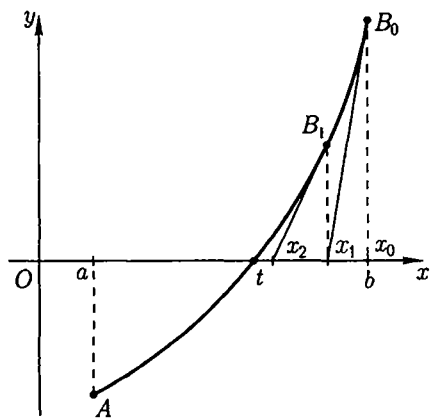


Рис. 2.7

Упражнения

2.19. Вычислите x_1 по методу касательных для уравнения б) из упр. 2.15 с начальным приближением $x_0 = 0,1$ и проверьте, лежит ли он в отрезке $[0, 1; 2]$ или нет. Какое число здесь надо брать в качестве x_0 ? Проведите геометрическую иллюстрацию.

2.20. Убедитесь графически, что когда f' и f'' одного знака на отрезке $[a; b]$, следует брать $x_0 = b$, а если они разных знаков, то $x_0 = a$ (*правило выбора начального приближения для метода Ньютона.*)

2.21. Корень уравнения $1 + x - x^4 = 0$ изолирован в отрезке $[1; 2]$. Вычислите приближения x_1, x_2 методом касательных и проиллюстрируйте вычисления геометрически.

2.22. Если производная $f'(x)$ мало изменяется на $[a; b]$, иногда принимают упрощенную рекуррентную формулу метода касательных:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_0)} \quad (n = 0, 1, 2, \dots).$$

Выясните ее геометрический смысл.

2.6.2. Сходимость итерационной последовательности

Построенная в п. 2.6.1 последовательность принималась за последовательность приближений. Покажем, что она действительно такова.

Теорема 2.13. Пусть корень t уравнения (2.1) отделен на отрезке $[a; b]$, $f(a) \cdot f(b) < 0$ и, кроме того, производные f' и f'' непрерывны и сохраняют постоянный знак на всем $[a; b]$. Тогда итерационная последовательность (x_n) , определяемая формулой (2.15) с соответствующим начальным приближением $x_0 \in [a; b]$, сходится к корню t .

○ Для определенности положим, что f' и f'' положительны на $[a; b]$ и, следовательно, $x_0 = b$. В остальных случаях доказательство проводится аналогично.

Методом математической индукции сначала покажем, что последовательность (x_n) убывает и ограничена снизу числом t .

Ясно, что $x_0 > t$. Тогда $f(x_0) > f(t) = 0$. Из (2.15) получим: $x_1 - x_0 = -\frac{f(x_0)}{f'(x_0)} < 0$, т. е. $x_1 < x_0$. По теореме 2.12 имеем

$$0 = f(t) = f(x_0) + f'(x_0)(t - x_0) + \frac{1}{2} f''(c_0)(t - x_0)^2,$$

где $t < c_0 < x_0$. Ввиду положительности f'' третье слагаемое в правой части полученного равенства больше нуля, поэтому должно быть $f(x_0) + f'(x_0)(t - x_0) < 0$. Отсюда $t < x_0 - \frac{f(x_0)}{f'(x_0)} = x_1$.

Итак, $t < x_1 < x_0$. Пусть теперь $t < x_n < x_{n-1}$. Проведенными выше рассуждениями нетрудно показать, что отсюда следует $t < x_{n+1} < x_n$, только здесь надо учесть, что $f(x_n) > f(t) = 0$, и применить формулу Тейлора к точкам t и x_n .

Таким образом, требуемое доказано. В силу теоремы 2.5 существует $\lim_{x \rightarrow \infty} x_n = z$, принадлежащий $[a; b]$. Перейдем к пределу в равенстве (2.15):

$$z = z - \frac{f(z)}{f'(z)}.$$

Следовательно, $f(z) = 0$, а так как корень на $[a; b]$ единственный, $z = t$. •

Упражнения

2.23. Докажите теорему 2.13 для остальных комбинаций знаков f' и f'' на $[a; b]$ с учетом правила выбора начального приближения из упр. 2.20. Убедитесь, что во всех случаях имеет место монотонное приближение к корню слева или справа.

2.24. На основании формулы (2.15) убедитесь, что в случае, когда в окрестности корня значения производной f' близки к нулю (график $y = f(x)$ почти горизонтален), сходимость приближений может оказаться медленной. Для проверки полученного вывода рассмотрите уравнения а) и б) из упр. 2.15 и вычислите методом касательных два приближения к корню $t = 1$ отдельно для каждого уравнения. Выполните соответствующие графические построения.

2.6.3. Оценка погрешностей приближений

В п. 2.5.3 была получена оценка (2.10), применимая и в случае использования метода касательных. Выведем еще одну формулу для оценки точности приближений x_n .

Теорема 2.14. Пусть первая и вторая производные функции f уравнения (2.1) непрерывны и имеют постоянный знак на отрезке $[a; b]$ изоляции корня t . Пусть также $f(a) \cdot f(b) < 0$, а положительные числа m и M_2 таковы, что $|f'(x)| \geq m$, $|f''(x)| \leq M_2$, $x \in [a; b]$ ¹⁾. Тогда погрешности приближений к t , найденных методом касательных, оцениваются формулой

$$|t - x_n| \leq \frac{M_2}{2m} (x_n - x_{n-1})^2 \quad (n = 1, 2, \dots). \quad (2.16)$$

¹⁾ Например, $m = \min_{[a; b]} |f'(x)|$, $M_2 = \max_{[a; b]} |f''(x)|$.

○ При соответствующем выборе начального приближения x_0 все приближения $x_n (n = 1, 2, \dots)$, вычисленные по формуле

$$x_n = x_{n-1} - \frac{f(x_{n-1})}{f'(x_{n-1})},$$

находятся на отрезке $[a; b]$. По формуле Тейлора при некотором c_n между x_n и x_{n-1} имеем

$$\begin{aligned} f(x_n) &= f(x_{n-1}) + f'(x_{n-1})(x_n - x_{n-1}) + \\ &+ \frac{1}{2} f''(c_n)(x_n - x_{n-1})^2. \end{aligned} \quad (2.17)$$

Переписав формулу для вычисления x_n в виде равенства

$$f(x_{n-1}) + f'(x_{n-1})(x_n - x_{n-1}) = 0,$$

из (2.17) и условий теоремы получим

$$|f(x_n)| \leq \frac{M_2}{2} (x_n - x_{n-1})^2.$$

Из формулы (2.10) с учетом данного неравенства следует доказываемая оценка. •

Соотношение (2.16) позволяет найти абсолютную погрешность приближения x_n и выявить условие окончания итерационного процесса метода касательных при вычислении корня с точностью до $\epsilon > 0$ (напишите соответствующие формулы!).

В п. 2.5.3 отмечалось, что общая формула (2.10) и полученная для метода хорд формула (2.11) во многих случаях дают слишком грубые оценки погрешностей приближений. Аналогичный вывод имеет место и здесь. Если на отрезке $[a; b]$ скорость убывания или возрастания функции f претерпевает резкие изменения, число M_2 может оказаться намного больше числа m .

Пример 2.8. Продолжим исследование оценок точности первых приближений к корню $t \in [1; 2]$ уравнения $x^3 - x - 1 = 0$, полученных методом хорд и методом касательных (см. примеры 2.5 – 2.7).

В примере 2.7 методом касательных найден приближенный корень $x_1 = 1,55$. Поскольку $m = 2$, $M_2 = 12$, оценка точности x_1 по формуле (2.10): $|t - 1,55| \leq \frac{f(1,55)}{2} \leq 1,18$, а по формуле (2.16): $|t - 1,55| \leq \frac{12}{4} (2 - 1,55)^2 \leq 0,61$. Ввиду того, что с точностью до трех цифр корень равен 1,32, обе оценки завышенные.

Заметим, что для двух приближений 1,17 и 1,55, находящихся слева и справа от корня примерно на одинаковом расстоянии, формула (2.10) при одном и том же знаменателе m дает значительно отличающиеся абсолютные погрешности: $\Delta_{1,17} = 0,3$ (пример 2.6) и

$\Delta_{1,55} = 1,18$. Этот факт подчеркивает зависимость оценки (2.10) от поведения функции в окрестности корня. Нетрудно убедиться, что функция f справа от корня возрастает быстрее, чем слева. •

Когда отрезок $[a; b]$ настолько мал, что на нем имеет место неравенство $M_2 \leq 2m$, погрешности приближений x_n оцениваются по формуле

$$|t - x_n| \leq (x_n - x_{n-1})^2 \quad (n = 1, 2, \dots). \quad (2.18)$$

В упр. 2.16 требовалось показать, что похожую формулу (2.12) при уточнении корня уравнения $x^3 - x - 1 = 0$ методом хорд можно применять на отрезке $[1,2; 1,5]$. Для использования оценки (2.18) требуется более узкий отрезок изоляции корня, например $[1,3; 1,4]$ (проверьте!).

Упражнения

2.25. В продолжение примеров 2.5–2.8 найдите с тремя значащими цифрами вторые приближения к корню $t \in [1; 2]$ уравнения $x^3 - x - 1 = 0$ методом хорд (обозначим x_2) и методом касательных (y_2). Определите абсолютные погрешности этих чисел по имеющимся формулам.

2.26. Проведите предварительную работу по приближенному решению уравнения $(x - 1)^4 - 0,5 = 0$ с применением методов хорд и касательных:

1. Отделите корни так, чтобы на отрезках изоляции уравнение удовлетворяло условиям применимости методов.

2. Напишите соответствующие рекуррентные формулы и выберите начальные приближения для уточнения корней обоими методами.

3. Напишите формулы оценки погрешностей приближений (2.10), (2.11) и (2.16) на выделенных отрезках изоляции корней.

4. Проиллюстрируйте геометрически поиск первых приближений к корням.

5. Найдите отрезки изоляции корней, где можно использовать оценки погрешностей (2.12) и (2.18).

2.7. Комбинированный метод хорд и касательных

В методах хорд и касательных приближения «подходят» к корню только с одной стороны, что затрудняет оценку их погрешностей. Фактическая погрешность приближений часто оказывается гораздо меньше, чем показывают формулы их оценки, что, в частности, приводит к лишним шагам при вычислениях с заданной точностью. Однако если эти методы применять не раздельно, а в сочетании друг

с другим, то от указанного недостатка можно избавиться. Более того, при этом перестают быть нужными специальные формулы оценки погрешностей.

В каждом из случаев относительно знаков f' и f'' последовательные приближения рассматриваемых методов находятся по разные стороны от корня. Если обозначить приближения метода хорд через x_n , приближения метода касательных через y_n , то всегда выполняется $x_n < t < y_n$ или $y_n < t < x_n$. Тогда, как и в методе половинного деления, корень будет находиться в каждом из вложенных отрезков с концами x_n, y_n ($n = 0, 1, 2, \dots$), причем $x_n \rightarrow t, y_n \rightarrow t$ при $n \rightarrow \infty$. Отрезки стягиваются к корню t , поэтому процесс уточнения с точностью до $\varepsilon > 0$ можно остановить сразу же, как только окажется $|x_n - y_n| \leq 2\varepsilon$, и взять в качестве приближенного корня середину отрезка между x_n и y_n : $t \approx t_n = \frac{x_n + y_n}{2}$.

Когда вычисления ведутся без заданной степени точности и на некотором шаге n в качестве приближения к корню выбрана средняя точка t_n между x_n и y_n , тогда $\Delta_{t_n} = \frac{|x_n - y_n|}{2}$.

Сочетать методы можно по-разному. Находя числа x_n и y_n независимо друг от друга по соответствующим рекуррентным формулам, будем иметь один способ. Он фактически реализуется в примерах из предыдущих двух параграфов и упр. 2.25, 2.26, только там не обращалось внимание на преимущества одновременного применения метода хорд и метода касательных. В примерах вычислены $x_1 = 1,17, y_1 = 1,55$ при $1,17 < t < 1,55$. Взяв $t \approx t_1 = (1,17 + 1,55)/2 = 1,36$, получим приближение с абсолютной погрешностью $\Delta_{t_1} = 0,19$. Аналогичные выкладки с использованием x_2, y_2 из упр. 2.25 дадут приближение t_2 .

Процесс уточнения будет более быстрым, если для вычисления x_{n+1} методом хорд вместо соответствующего неподвижного конца отрезка $[a; b]$ использовать найденное методом касательных приближение y_n , т. е. когда хорды проводятся через точки графика функции с абсциссами x_n и y_n .

Именно при таком способе вычислений есть смысл говорить о *комбинированном методе хорд и касательных*. Он проиллюстрирован на рис. 2.8 при $f' > 0, f'' > 0$ на $[a; b]$.

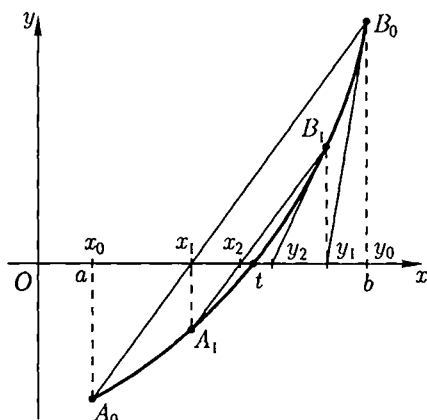


Рис. 2.8

Из сказанного понятно, что вычисление пары чисел x_{n+1} , y_{n+1} надо начинать с y_{n+1} , которое определяется по-прежнему формулой метода касательных:

$$y_{n+1} = y_n - \frac{f(y_n)}{f'(y_n)} \quad (n = 0, 1, 2, \dots) \quad (2.19)$$

при соответствующем начальном приближении y_0 . Затем отыскивается x_{n+1} . При этом благодаря комбинированию методов его вычисление упрощается, поскольку формула метода хорд становится единой, не зависящей от знаков производных (упр. 2.28).

Упражнения

2.27. На рис. 2.8 изображены два шага комбинированного метода хорд и касательных при положительных f' , f'' . Проведите аналогичную геометрическую иллюстрацию в остальных трех случаях относительно знаков производных.

2.28. Убедитесь, что для вычисления приближений x_n комбинированным методом хорд и касательных во всех случаях относительно знаков f' , f'' на отрезке $[a; b]$ имеет место одна и та же рекуррентная формула:

$$x_{n+1} = x_n - \frac{y_n - x_n}{f(y_n) - f(x_n)} f(x_n). \quad (2.20)$$

При этом если f' , f'' одного знака, то $x_0 = a$; если f' , f'' разных знаков, то $x_0 = b$.

2.29. Для корня $t \in [1; 2]$ уравнения $x^3 - x - 1 = 0$ найдите приближенный корень после двух шагов комбинированного метода и определите его абсолютную погрешность.

2.30. Напишите рекуррентные формулы для уточнения обоих корней уравнения из упр. 2.26 комбинированным методом хорд и касательных.

2.8. Метод простой итерации

Будем рассматривать уравнение вида

$$x = g(x) \quad (2.21)$$

с корнем t , отделенным на отрезке $[a; b]$. Функция g предполагается непрерывной на этом отрезке. Уравнение (2.21) можно получить из уравнения (2.1): $f(x) = 0$ путем эквивалентных преобразований.

Пример 2.9. Уравнение $x^3 - 3x + 1 = 0$ представляется в виде (2.21) разными способами. Например: а) $x = \frac{x^3 + 1}{3}$; б) $x = x^3 - 2x + 1$; в) $x = \sqrt[3]{3x - 1}$. •

Метод простой итерации¹⁾ является одним из наиболее удобных и эффективных методов приближенного решения уравнений. Как видно из его названия, он предполагает уточнение корня с использованием итерационной последовательности.

2.8.1. Рекуррентная формула

Рекуррентная формула определяется на основе самого уравнения (2.21). Если известен какой-либо член последовательности x_n (например, $x_0 \in [a; b]$), то за x_{n+1} можно взять $g(x_n)$. Соотношение

$$x_{n+1} = g(x_n) \quad (n = 0, 1, 2, \dots) \quad (2.22)$$

и является искомой рекуррентной формулой.

Если существует конечный предел $\lim_{n \rightarrow \infty} x_n = z$ и функция g непрерывна в точке z , переходом к пределу в равенстве (2.22) получим

$$z = g(z),$$

т. е. число z является корнем уравнения (2.21). Если $z \in [a; b]$, в силу единственности корня на отрезке $[a; b]$ z совпадает с t .

Вычисления по формуле (2.22) проиллюстрированы на рис. 2.9.

Построим графики функций из левой и правой частей уравнения (2.21), т. е. линии $y = x$ и $y = g(x)$. Они должны пересекаться в точке с абсциссой t . Взяв некоторое число x_0 , вычислим $g(x_0)$ и получим на кривой $y = g(x)$ точку A_0 . Линия проекции этой точки на ось Oy пересечет прямую $y = x$ в точке B_1 . Проекция B_1 на ось Ox дает x_1 ($x_1 = g(x_0)$). Вычислив $g(x_1)$ и спроецировав точку A_1 графика функции g на ось Oy , найдем точку B_2 на прямой $y = x$ и ее проекцию x_2 на ось Ox ($x_2 = g(x_1)$) и т. д.

Пример 2.10. Один из корней уравнения $x = x^3 - 2x + 1$ из примера 2.9 находится в отрезке $[0; 1]$. Рекуррентная формула (2.22) здесь имеет вид $x_{n+1} = x_n^3 - 2x_n + 1$. Возьмем различные $x_0 \in [0; 1]$ и проследим за поведением соответствующих итерационных последовательностей. При $x_0 = 1$ или $x_0 = 0$ все значения x_n ($n \geq 1$) попеременно равны 0 или 1, поэтому последовательность (x_n) расходится. Взяв

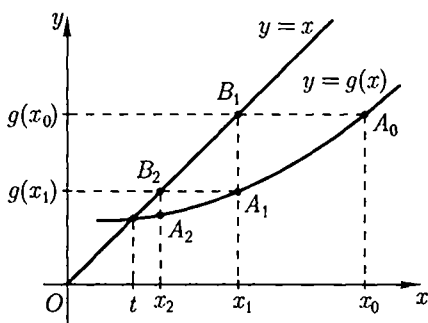


Рис. 2.9

¹⁾ В гл. 6 будет показано, что рассматриваемый здесь метод является частным случаем общего метода итерации, обосновываемого с помощью принципа сжимающих отображений (теоремы Банаха).

$x_0 = 0,5$, получим $x_1 = 0,125$, $x_2 = 0,752$, $x_3 = -0,079$. Вычислив еще несколько значений x_n , убеждаемся, что в этом случае трудно сказать что-либо определенное о сходимости или расходимости данной последовательности. •

Возникает вопрос: при каких условиях итерационная последовательность сходится к корню t , т. е. когда она является последовательностью приближений? Ответ будет дан в п. 2.8.2.

Упражнения

2.31. На рис. 2.9 приведен вариант, когда функция g возрастает в окрестности корня, причем медленнее, чем функция $y = x$. Если g дифференцируема, это означает, что $0 \leq g'(x) < 1$ в этой окрестности. Как видно, здесь последовательность (x_n) сходится к t монотонно и справа. Приведите аналогичную иллюстрацию метода в случае $-1 < g'(x) \leq 0$. Убедитесь, что тогда (x_n) также сходится к корню, но не монотонно, а «колеблясь» вокруг t .

2.32. Начертите графики для случая, когда $|g'(x)| > 1$ ($g'(x) > 1$ или $g'(x) < -1$) в окрестности корня, т. е. когда функция g возрастает (убывает) быстрее, чем $y = x$ ($y = -x$). Убедитесь, что тогда (x_n) не сходится к этому корню.

2.8.2. Сходимость итерационной последовательности

Предварительно сформулируем еще одну теорему из математического анализа.

Теорема 2.15. Пусть (a_n) и (b_n) — две числовые последовательности. Если $a_n \rightarrow 0$ и $|b_n| \leq |a_n|$ при всех $n = 1, 2, \dots$, то $b_n \rightarrow 0$. (Следует из теоремы о сжатых последовательностях [3, гл. III, § 3].)

Анализ упр. 2.31 и 2.32 наводит на мысль о справедливости нижеследующих теорем 2.16 и 2.17.

Теорема 2.16 (достаточное условие сходимости итерационной последовательности). Пусть корень t уравнения (2.21) отделен на отрезке $[a; b]$ длины h . Если на отрезке $[c; d] = [a - h; b + h]$ функция g дифференцируема и найдется число $0 \leq q < 1$ такое, что

$$|g'(x)| \leq q \tag{2.23}$$

при всех $x \in [c; d]$, то итерационная последовательность, порожденная формулой (2.22), сходится к корню t при любом выборе начального приближения $x_0 \in [a; b]$ (рис. 2.10).

○ Пусть $x_0 \in [a; b]$. Докажем по индукции, что при всех $n = 1, 2, \dots$ числа x_n принадлежат отрезку $[c; d]$ и верно неравенство

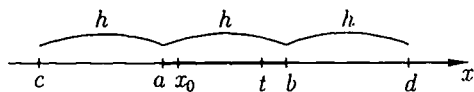


Рис. 2.10

$$|t - x_n| \leq q^n h. \quad (2.24)$$

Вычислим $x_1 = g(x_0)$. Учитывая, что $t = g(t)$, и применяя теорему Лагранжа для точек x_0 и t , получим

$$|t - x_1| = |g(t) - g(x_0)| = |g'(p_0)||t - x_0|,$$

где p_0 лежит между t и x_0 . На основании (2.23) $|t - x_1| \leq q|t - x_0| < qh$. Отсюда, в частности, $|t - x_1| < h$ и, следовательно, $x_1 \in [c; d]$.

Теперь возьмем произвольное n и предположим, что $x_n \in [c; d]$ и справедливо неравенство (2.24). Найдем $x_{n+1} = g(x_n)$, а затем с помощью теоремы Лагранжа при некотором p_n между t и x_n и условия (2.23) получим неравенство (2.24) для x_{n+1} :

$$|t - x_{n+1}| = |g'(p_n)||t - x_n| \leq q|t - x_n| \leq q^{n+1} h.$$

Поскольку $q^{n+1} < 1$ и $t \in [a; b]$, то $x_{n+1} \in [c; d]$.

Тем самым доказываемые по индукции свойства последовательности (x_n) действительно имеют место. Ввиду того что $q^n \rightarrow 0$ при $n \rightarrow \infty$, из (2.24) и теоремы 2.15 следует сходимость (x_n) к t . •

З а м е ч а н и я. 1. Требование теоремы 2.16, чтобы условие (2.23) имело место на расширенном отрезке, а не только на $[a; b]$, обеспечивает важное свойство приближений — их принадлежность отрезку $[c; d]$ и выполнение неравенства (2.24). Если этого не требовать, то при неудачном выборе x_0 последующие приближения могут выйти за пределы действия (2.23) и не будут сходиться к корню (убедитесь в этом графическими построениями!).

2. Если известно, что значения $g(x)$ находятся в $[a; b]$ при всех $x \in [a; b]$, то выполнение условия (2.23) достаточно потребовать лишь на $[a; b]$. Тогда необходимость отрезка $[c; d]$ «тройной длины» отпадает.

3. Из неравенства (2.24) видно, что чем меньше число q , тем быстрее сходится последовательность приближений.

4. Метод простой итерации является *самоисправляющимся*. Если какое-то приближение x_n найдено с ошибкой, но при этом ошибка не вывела его из отрезка $[a; b]$, то последующие члены последовательности все равно будут приближаться к корню.

Т е о р е м а 2.17. Если на отрезке $[a; b]$ функция g дифференцируема и

$$|g'(x)| \geq 1 \quad \text{при всех } x \in [a; b], \quad (2.25)$$

то определяемая формулой (2.22) итерационная последовательность не сходится к корню $t \in [a; b]$ ни при каком $x_0 \neq t$ из этого отрезка.

○ Обозначим $h = |t - x_0| \neq 0$. Применяя теорему Лагранжа, нетрудно проверить, что тогда для всех $x_n \in [a; b]$ будет $|t - x_n| \geq h$. Отсюда следует, что (x_n) не сходится к $t \in [a; b]$. •

З а м е ч а н и е. Теорема 2.17 не утверждает, что последовательность (x_n) будет вообще расходящейся. Выйдя из отрезка $[a; b]$, (x_n) может сходиться к другому корню этого же уравнения. На рис. 2.11

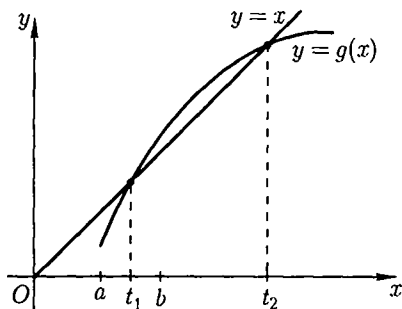


Рис. 2.11

видно, что $g'(x) > 1$ в окрестности корня t_1 . Отправляясь от любого x_0 , где $t_1 < x_0 \leq b$, последовательность будет приближаться к корню t_2 (убедитесь в этом графическими построениями!).

Пример 2.11. Проверим, можно ли уточнить методом простой итерации корень $t \in [0; 1]$ уравнения $x^3 - 3x + 1 = 0$ из примера 2.9 на основании приведенных там представлений уравнения в виде $x = g(x)$. Для этого сначала возьмем

более узкий отрезок изоляции корня: $[a; b] = [0,2; 0,5]$. Тогда $h = 0,5 - 0,2 = 0,3$ и, следовательно, $[c; d] = [-0,1; 0,8]$.

Рассмотрим уравнение а): $x = \frac{x^3 + 1}{3}$. Очевидно, что $|g'(x)| = x^2$ принимает наибольшее значение в правом конце отрезка $[c; d]$, поэтому $|g'(x)| \leq 0,64$ при всех x из этого отрезка ($q = 0,64 < 1$). На основании теоремы 2.16 последовательность, полученная по формуле $x_{n+1} = \frac{x_n^3 + 1}{3}$ ($n = 0, 1, 2, \dots$), сходится к корню при любом начальном приближении $x_0 \in [0,2; 0,5]$. Например, взяв $x_0 = 0,2$, получим приближения $x_1 = 0,336$, $x_2 = 0,346$, $x_3 = 0,347^{1)}$.

Функция $g(x) = x^3 - 2x + 1$ уравнения б) такова, что модуль ее производной $|g'(x)| = |3x^2 - 2| = 2 - 3x^2$ на отрезке $[0,2; 0,5]$ принимает значения от 1,25 до 1,88. Следовательно, в силу теоремы 2.17 построенная по уравнению б) итерационная последовательность не сходится к искомому корню ни при каком $x_0 \in [0,2; 0,5]$ (соотнесите этот вывод с примером 2.10!). •

Упражнения

2.33. В продолжение примера 2.11 покажите, что представление исходного уравнения в виде в) $x = \sqrt[3]{3x - 1}$ непригодно для уточнения корня из $[0,2; 0,5]$ методом простой итерации.

2.34. Уравнение $x = e^x - 2$ имеет два корня.

1. Выясните вопрос о сходимости итерационной последовательности $x_{n+1} = e^{x_n} - 2$ ($n = 0, 1, 2, \dots$) в окрестности этих корней.

2. В окрестностях обоих корней выполните два-три шага геометрической иллюстрации метода.

2.35. Убедитесь с помощью графических построений, что возможны ситуации, когда в окрестности корня для g' не выполняются ни

¹⁾ Об оценке их погрешностей см. в п. 2.8.4.

(2.23), ни (2.25), и что в этом случае сходимость или расходимость итерационной последовательности может зависеть от выбора x_0 .

2.36.* Докажите, что при $0 \leq g'(x) \leq q < 1$ на $[c; d]$ последовательные приближения (2.22) сходятся к t монотонно.

2.37.* Докажите, что при $-1 < -q \leq g'(x) \leq 0$ на $[c; d]$ последовательные приближения (2.22) колеблются около корня t .

У к а з а н и е: В упр. 2.36 и 2.37 следует рассмотреть отдельно случаи $x_0 < t$, $x_0 > t$ и воспользоваться теоремой Лагранжа.

2.8.3. Приведение уравнений к виду, пригодному для метода простой итерации

Доказанные в п. 2.8.2 теоремы и пример 2.11 показывают, что для метода простой итерации не безразличен способ записи уравнения в виде (2.21). Кроме того, оказывается, что используемое для уточнения одного корня представление $x = g(x)$ может оказаться непригодным в окрестности другого корня этого же уравнения, и наоборот.

Рассмотрим вновь уравнение $x^3 - 3x + 1 = 0$ и три равносильные ему уравнения а), б) и в) из примера 2.9. Пример 2.11 и упр. 2.33 показывают, что для уточнения корня из отрезка $[0,2; 0,5]$ можно использовать представление а), но не б) и в). Далее в примере 2.12 выясняется, что при поиске приближения к корню из $[1,4; 1,6]$ неприменимы уравнения а) и б), но можно пользоваться в).

Таким образом, необходимы навыки получения уравнения (2.21) с такими значениями $|g'(x)|$ в окрестности корня, чтобы выполнялось условие (2.23). Если первая (обычно самая простая и напрашивающаяся) попытка представления уравнения в требуемом виде оказалась неудачной, отчаиваться не следует. В ряде случаев можно использовать специальные приемы. Рассмотрим некоторые из них.

С п о с о б 1. Если $f(x)$ содержит в себе выражение некоторой обратной на $[c; d]$ функции $y = \psi(x)$, причем такой, что $|\psi'(x)| \geq q > 1$ на $[c; d]$, то следует попытаться заменить уравнение (2.1) на равносильное вида (2.21) с использованием обратной для ψ функции $\varphi: x = \varphi(y)$. Этот способ основан на известном соотношении между производными взаимобратных функций $\varphi'(y) = 1/\psi'(x)$ и следствии из него:

$$\text{если } |\psi'(x)| \geq q > 1, \text{ то } |\varphi'(y)| = \frac{1}{|\psi'(x)|} \leq \frac{1}{q} < 1.$$

С п о с о б 2. В случае когда способ 1 применить трудно или он не дает нужного результата, можно использовать следующий способ.

Пусть дано уравнение (2.1) с единственным корнем в $[a; b]$. Предположим, что на введенном в теореме 2.16 отрезке $[c; d]$ производная f' функции f непрерывна, не равна константе и принимает значения одного и того же знака. Будем считать, что $f'(x) > 0$, ибо в

противном случае вместо (2.1) можно рассматривать равносильное уравнение: $-f(x) = 0$.

Обозначим $m = \min_{[c; d]} f'(x)$, $M = \max_{[c; d]} f'(x)$, $k = \frac{1}{M}$ и $q = 1 - \frac{m}{M}$. Ясно, что $0 \leq q < 1$. Заменяем (2.1) эквивалентным уравнением

$$x = x - kf(x) \quad (2.26)$$

и покажем, что для функции $g(x) = x - kf(x)$ на $[c; d]$ имеет место свойство (2.23).

Для $x \in [c; d]$ справедливы неравенства: $0 < m \leq f'(x) \leq M$. Разделим их почленно на число M и для разностей между единицей и полученными дробями получим неравенства:

$$0 \leq 1 - \frac{f'(x)}{M} \leq 1 - \frac{m}{M} = q,$$

откуда и вытекает, что

$$0 \leq g'(x) = 1 - kf'(x) \leq q$$

при всех $x \in [c; d]$.

Приведенные способы 1 и 2 иллюстрируются в следующем примере.

Пример 2.12. Один из корней уравнения $x^3 - 3x + 1 = 0$ из примера 2.9 расположен в отрезке $[1,2; 1,6]$.

Рассмотрим представление а) $x = \frac{x^3 + 1}{3}$ исходного уравнения. На данном отрезке $|g'(x)| = x^2 > 1$ и потому применять метод простой итерации к уравнению а) нельзя. Аналогичный вывод имеет место и в случае представления $x = x^3 - 2x + 1$ (проверьте!).

Способ 1. Очевидно, что уравнение в) $x = \sqrt[3]{3x - 1}$ получилось реализацией первого способа преобразования, а именно, переходом от функции $y = x^3$ к обратной функции $x = \sqrt[3]{y}$:

$$x^3 = 3x - 1 \Leftrightarrow x = \sqrt[3]{3x - 1}.$$

Рассмотрим отрезок «тройной» длины $[c; d] = [0,8; 2]$. Производная $g'(x) = \frac{1}{\sqrt[3]{(3x - 1)^2}}$ функции $g(x) = \sqrt[3]{3x - 1}$ на отрезке $[c; d]$ положительна, убывает, и потому $|g'(x)| \leq g'(0,8) < 0,8$ при всех $x \in [c; d]$. Следовательно, для этого уравнения на выбранном отрезке $[a; b]$ выполнены все условия теоремы 2.16 с $q = 0,8^1$.

¹⁾ Если выбрать $[a; b] = [1; 2]$, то условия теоремы 2.16 перестают быть верными и в случае этого уравнения, поскольку g' в точке $x = 1/3 \in [0; 3] = [c; d]$ терпит разрыв. Данный факт еще раз подчеркивает значение размера отрезка изоляции корня.

Способ 2. Привести уравнение $f(x) = x^3 - 3x + 1 = 0$ к виду (2.26) и проиллюстрировать второй способ преобразования на определенном выше отрезке $[c; d] = [0,8; 2]$ не удастся, так как производная $f'(x) = 3x^2 - 3$ меняет знак в точке $x = 1 \in [c; d]$. Однако для рассматриваемого нами корня можно взять меньший отрезок изоляции $[a; b] = [1,4; 1,6]$ и соответственно $[c; d] = [1,2; 1,8]$. Тогда $f'(x) > 0$ на $[c; d]$ и $1,3 \leq f'(x) \leq 6,8$. Следовательно, $m = 1,3$, $M = 6,8$, поэтому $k = 0,15$ и $q = 0,82 < 1$.

Таким образом, для уточнения нужного нам корня методом простой итерации можно использовать уравнение

$$x = x - 0,15(x^3 - 3x + 1) \Leftrightarrow x = -0,15x^3 + 1,45x - 0,15$$

с выбором $x_0 \in [1,4; 1,6]$. Кстати, это уже четвертое представление исходного уравнения в виде $x = g(x)$. •

Упражнения

2.38.* Имеется еще один способ приведения уравнения (2.1) к виду, пригодному для метода простой итерации¹⁾. Докажите следующее утверждение:

Пусть корень t уравнения $f(x) = 0$ отделен на отрезке $[a; b]$ длиной h . Если функция f дифференцируема на $[c; d] = [a - h; a + h]$, производная $f'(x) > 0$ во всех точках $[c; d]$ и имеет на этом отрезке не равные нулю наименьшее и наибольшее значения m и M ($m \neq M$), то правая часть уравнения

$$x = x - \frac{2}{M + m} f(x) \tag{2.27}$$

обладает на $[c; d]$ свойством (2.23) с $q = \frac{M - m}{M + m}$.

Убедитесь сравнением чисел q , что итерационная последовательность, соответствующая уравнению (2.27), сходится быстрее, чем последовательность, составленная по уравнению (2.26).

2.39. Уравнение $x^3 - 3x + 1 = 0$ имеет еще один корень помимо рассмотренных в примерах 2.11 и 2.12.

1. Проверьте, пригодны ли приведенные в примере 2.9 его представления а), б), в) для уточнения этого корня.

2. Проверьте, выполняются ли в окрестности этого корня условия, требующиеся для составления уравнений вида (2.26) и (2.27). Если да, то выделите необходимые отрезки, запишите эти уравнения и определите соответствующие значения q .

2.40. Уравнение $x + \ln x = 0$ имеет один корень. Составьте всеми тремя изученными способами эквивалентные уравнения вида $x = g(x)$, пригодные для уточнения этого корня методом простой итерации.

¹⁾ Этот способ сообщил автору В. А. Попов.

2.8.4. Оценка погрешностей приближений

Теорема 2.18. В условиях теоремы 2.16 справедлива формула:

$$|t - x_n| \leq \frac{q}{1-q} |x_n - x_{n-1}| \quad (n = 1, 2, \dots). \quad (2.28)$$

○ Обозначим $f(x) = x - g(x)$. Функция f дифференцируема, причём

$$|f'(x)| = |1 - g'(x)| \geq 1 - |g'(x)| \geq 1 - q \quad (2.29)$$

на $[c; d]$. Учитывая определение функции f и то, что $f(t) = 0$, имеем

$$|x_{n+1} - x_n| = |g(x_n) - x_n| = |-f(x_n)| = |f(t) - f(x_n)|.$$

Применив сначала теорему Лагранжа с некоторым числом d_n между t и x_n , а затем неравенство (2.29), получим

$$|x_{n+1} - x_n| = |f'(d_n)| |t - x_n| \geq (1 - q) |t - x_n|. \quad (2.30)$$

Модуль разности $|x_{n+1} - x_n|$ можно оценить и сверху. Поскольку $x_{n+1} - x_n = g(x_n) - g(x_{n-1})$, на основе теоремы Лагранжа с числом p_n между x_n, x_{n-1} и неравенства (2.23) выводим:

$$|x_{n+1} - x_n| = |g'(p_n)| |x_n - x_{n-1}| \leq q |x_n - x_{n-1}|. \quad (2.31)$$

Чтобы теперь получить требуемую оценку (2.28), надо выделить $|t - x_n|$ из неравенства (2.30) и учесть (2.31). •

Таким образом, если задана точность приближенного корня $\varepsilon > 0$, то итерационный процесс необходимо закончить при выполнении условия

$$\frac{q}{1-q} |x_n - x_{n-1}| \leq \varepsilon \quad (2.32)$$

и взять $t \approx x_n$.

На каждом шаге метода простой итерации можно определить абсолютную погрешность приближения x_n :

$$\Delta_{x_n} = \frac{q}{1-q} |x_n - x_{n-1}|.$$

Пример 2.13. Применим полученную формулу для оценки точности найденного в примере 2.11 приближения $x_3 = 0,347$ к корню уравнения $x = \frac{x^3 + 1}{3}$. Поскольку $q = 0,64$ и $x_2 = 0,346$, то $\Delta_{x_3} = \frac{0,64}{0,36} \times |0,347 - 0,346| = 0,0018$. •

Как следует из упр. 2.31 и 2.37, при $-1 < -q \leq g'(x) \leq 0$ на $[c; d]$ (функция g убывает) приближения x_{n-1}, x_n находятся по разные стороны от корня t , т.е. на каждом шаге итерации корень лежит в отрезке с концами x_{n-1} и x_n . В этом случае оценка погрешности примет вид

$$|t - x_n| \leq |x_n - x_{n-1}| \quad (n = 1, 2, \dots), \quad (2.33)$$

а условием окончания итерационного процесса при заданной точности $\varepsilon > 0$ будет неравенство $|x_n - x_{n-1}| \leq \varepsilon$.

Упражнения

2.41.* Используя доказательство теоремы 2.16, покажите, что

$$|x_n - x_{n-1}| \leq q^{n-1} |x_1 - x_0| \quad (n = 1, 2, \dots),$$

и на основании этого выведите еще одну формулу оценки погрешностей:

$$|t - x_n| \leq \frac{q^n}{1 - q} |x_1 - x_0| \quad (n = 1, 2, \dots). \quad (2.34)$$

2.42. Найдите с точностью до трех верных значащих цифр корень $t \in [1,4; 1,6]$ уравнения $x^3 - 3x + 1 = 0$, используя запись этого уравнения в виде $x = -0,15x^3 + 1,45x - 0,15$ (см. пример 2.12).

2.43. Отрицательный корень уравнения из упр. 2.34 можно уточнить методом простой итерации. Выполните три шага итерации и найдите абсолютные погрешности каждого приближения по формулам (2.28) и (2.34).

2.44. 1. Выберите отрезок изоляции корня уравнения $x = \cos x$, при котором применим метод простой итерации с оценкой погрешностей приближений (2.33).

2. Найдите корень с точностью до 0,005.

Глава 3. АНАЛИТИЧЕСКОЕ ПРИБЛИЖЕНИЕ ТАБЛИЧНЫХ ФУНКЦИЙ

3.1. Основные понятия

3.1.1. Табличная функция

В вычислительной практике часто приходится иметь дело с таблицами значений действительных функций одной или нескольких переменных¹⁾. В форме таблиц обычно оформляются результаты экспериментальных исследований — *эмпирические* таблицы. Существуют таблицы значений математических функций — *математические* таблицы. Таблицы используются при приближенном решении многих задач. Например, табличный вид имеет численное решение обыкновенных дифференциальных уравнений (см. гл. 5).

В дальнейшем будем рассматривать функции одной переменной.

Для построения таблицы значений функции $y = f(x)$ берется отрезок $[a; b] \subset D_f$, выбираются точки $a = x_0 < x_1 < x_2 < \dots < x_n = b$ и каким-либо способом определяются значения $y_i (i = 0, 1, \dots, n)$ функции в этих точках. Числа x_i называются *табличными аргументами*, а числа y_i — *табличными значениями* функции f .

Если табличные аргументы *равноотстоящие*, т. е. $x_{i+1} - x_i = h = \text{const} (i = 0, 1, \dots, n - 1)$, общее расстояние $h > 0$ между ними называют *шагом* таблицы. В этом случае имеет место формула

$$x_i = x_0 + ih \quad (i = 0, 1, \dots, n).$$

Математические таблицы строятся для известных функций и имеют, как правило, постоянный шаг. Среди эмпирических таблиц часто встречаются таблицы с неравноотстоящими аргументами.

Аргументы таблицы будем считать точными числами. Табличные значения находят путем вычислений или измерений и, как правило, являются приближенными числами. Если они все имеют одинаковую абсолютную погрешность, то она называется *погрешностью таблицы*. Иногда погрешность на разных участках таблицы различна.

¹⁾ Для краткости вместо «таблица значений функции» иногда будем писать «таблица».

Таблица 3.1

x	$f(x)$
x_0	y_0
x_1	y_1
x_2	y_2
\vdots	\vdots
x_n	y_n

В зависимости от количества цифр (знаков) в записи значений функции различают таблицы четырехзначные, пятизначные и т. д. В случае математических таблиц, в которых значения функции записываются только с верными цифрами, между количеством цифр и точностью таблицы существует неразрывная связь: если на каком-то участке таблицы все значения имеют k цифр после десятичной запятой, погрешность этого участка составляет $0,5 \cdot 10^{-k}$.

З а м е ч а н и е. Если числа y_i найдены для известной функции f приближенно, то они являются значениями f с некоторой погрешностью: $f(x_i) \approx y_i$. Данный факт необходимо учитывать при конкретной работе с таблицами, но так как в излагаемых далее вопросах он не является существенным, для удобства везде считаем, что $f(x_i) = y_i$.

Далее будем рассматривать не просто таблицы значений функций, а функции вместе с их таблицами значений.

О п р е д е л е н и е 3.1. Функцию f назовем *табличной функцией*, если на некотором отрезке $[a; b] \subset D_f$ задана таблица ее значений.

Необходимо различать табличную функцию и функцию, заданную таблично (табличным способом). Во втором случае речь идет только о таблице значений, которая является функцией с областью определения, состоящей из конечного множества заданных табличных аргументов. Например, областью определения табл. 3.1, понимаемой как «таблично заданная функция», будет множество $\{x_0, x_1, \dots, x_n\}$, тогда как D_f включает в себя весь отрезок $[x_0; x_n]$. График табл. 3.1 на плоскости xOy представляет собой множество точек $\{(x_i, y_i) : i = 0, 1, \dots, n\}$, лежащих на графике функции $y = f(x)$ ¹⁾.

Часто таблица вида табл. 3.1 является основным или даже единственным доступным для математических исследований носителем информации о табличной функции f . Так будет, если f выразит зависимость между какими-либо экспериментальными величинами и ее аналитическое выражение неизвестно. Но при этом всегда подразумеваем, что f задана по крайней мере на отрезке $[x_0; x_n]$.

Упражнение

3.1. Составьте трехзначную таблицу значений функции $y = x^3$ на отрезке $[1; 1,5]$ с шагом $h = 0,1$. Начертите график полученной таблицы и график функции $y = x^3$.

¹⁾ Здесь учитывается сделанное выше замечание. Если y_i являются приближениями к $f(x_i)$, то на самом деле точки могут лежать около графика.

3.1.2. Задача аналитического приближения табличных функций

Одной из основных проблем, связанных с табличными функциями, является их *аналитическое приближение на основе известной таблицы значений*, а именно, поиск такой аналитически заданной и достаточно просто вычисляемой функции p , которая в каком-то смысле близка к табличной функции f на всем отрезке $[a; b] = [x_0; x_n]$ или на некоторой его части¹⁾.

Потребность в формуле

$$f(x) \approx p(x), \quad x \in [a; b] \quad (3.1)$$

возникает в различных ситуациях. Отметим некоторые из них.

Во-первых, ее используют при вычислении значений функции f . Когда $x \in [a; b]$ является одним из табличных аргументов, найти $f(x)$ легко: оно равно y_i при некотором $i = 0, 1, \dots, n$ с той же точностью, с какой определено y_i . Если же $x \neq x_i$ ($i = 0, 1, \dots, n$), то решить эту задачу без формулы (3.1) бывает невозможно (таблица эмпирическая, и точная связь $y = f(x)$ неизвестна) или трудно (таблица математическая, но простых способов непосредственного вычисления $f(x)$ не существует). Во-вторых, аналитическое приближение является целью математической обработки эмпирических таблиц (см. 3.7). В-третьих, на основе формул вида (3.1) выводятся численные методы решения ряда задач (см. гл. 4).

Если функция p отыскивается на всем отрезке $[a; b]$, то говорят о *глобальном приближении* табличной функции. Приближение может быть и *локальным*, т. е. на некотором отрезке $[x_k; x_m]$, $0 \leq k < m \leq n$. С точки зрения методов поиска приближающих функций разницы между локальным и глобальным приближениями нет. Впредь будем заниматься приближением на основе полной табл. 3.1, имея в виду, что при локальном приближении она составлена для выбранного участка основной таблицы.

Имеется несколько способов аналитического приближения, в которых по-разному понимается близость между функциями f и p . Далее будут рассмотрены два из них: *интерполирование* и метод *наименьших квадратов*. Интерполирование обычно используется в случае таблиц с небольшим количеством аргументов, поэтому оно больше подходит для локального приближения. Метод наименьших квадратов фактически не ограничивает количество табличных данных и чаще всего применяется для установления приближенной аналитической зависимости между параметрами эмпирических таблиц.

Наиболее просто, но без четкой меры близости функций эта задача решается методом *подбора* при условии, когда точечный график таблицы находится около графика некоторой известной аналитически заданной функции (унр. 3.2).

¹⁾ Иногда приближение необходимо в более широком промежутке, включающем в себя $[x_0; x_n]$, однако в такой постановке задачу рассматривать не будем.

Упражнение

3.2. Начертите графики функций по данным таблиц и определите, какими известными функциями можно приблизить функции f_1 и f_2 .

x	-1,3	-0,5	0	0,4	1	1,7
$f_1(x)$	2,6	1,3	0,9	1,2	2	4

x	-1	-0,4	0,2	0,4	0,8	1
$f_2(x)$	0,4	0,7	1,2	1,5	2,2	2,7

3.2. Интерполирование табличных функций

3.2.1. Постановка задачи

Предположим, что табличная и приближающая функции считаются близкими, если значения этих функций в табличных аргументах совпадают. Способ аналитического приближения с выполнением указанного свойства называется *интерполированием*, или *интерполяцией*. В общем виде задача интерполирования формулируется следующим образом:

для функции f с табл. 3.1 требуется найти достаточно простую известную функцию p , удовлетворяющую соотношениям

$$p(x_i) = y_i \quad (i = 0, 1, 2, \dots, n). \quad (3.2)$$

В этом случае функция p называется *интерполирующей* функцией, а табличные аргументы x_0, x_1, \dots, x_n , в которых ее значения должны совпадать со значениями функции f , — *узлами интерполирования*.

Соотношения (3.2) представляют собой критерий, позволяющий судить, является ли функция p интерполирующей или нет. Геометрически они означают то, что график функции p проходит через точки графика функции f с абсциссами $x_i, i = 0, 1, \dots, n$ (рис. 3.1).

Если интерполирующая функция p найдена, то на отрезке $[a; b]$ будет иметь место приближенное равенство (3.1), которое в этом случае называется *интерполяционной формулой*.

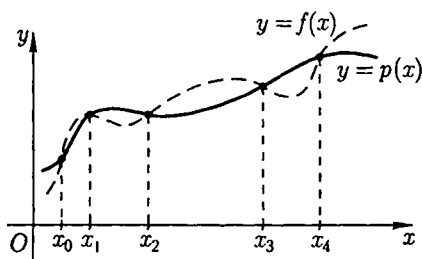


Рис. 3.1

Ее определитель

$$\begin{vmatrix} x_0^n & \dots & x_0^2 & x_0 & 1 \\ x_1^n & \dots & x_1^2 & x_1 & 1 \\ \dots & \dots & \dots & \dots & \dots \\ x_n^n & \dots & x_n^2 & x_n & 1 \end{vmatrix}$$

не равен нулю как *определитель Вандермонда* с попарно различными x_0, x_1, \dots, x_n [14, гл. I, § 6]. Следовательно, система (3.4) имеет единственное решение, определяющее коэффициенты многочлена P_n . •

Таким образом, для любой табличной функции найдется единственный интерполяционный многочлен, степень которого в общем случае на единицу меньше количества узлов интерполирования. Но на самом деле степень многочлена может оказаться еще меньше. Это будет происходить тогда, когда при решении системы (3.4) получатся нулевые коэффициенты при старших степенях x (упр. 3.6).

Замтим, что по таблице с $n + 1$ узлом можно построить бесконечное множество интерполяционных многочленов P_k степени $k > n$. Действительно, в этом случае неоднородная система вида (3.4) содержит больше неизвестных, чем уравнений, поэтому определяет бесконечное множество наборов коэффициентов a_k, \dots, a_0 [14, гл. I, § 11].

В дальнейшем будем искать только единственные многочлены, определяемые на основе теоремы 3.1. Существуют специальные правила их построения, некоторые из них будут приведены в 3.4 и 3.5.

Упражнения

3.4. Пусть табл. 3.1 является таблицей значений некоторого многочлена n -й степени H_n . Докажите, что тогда интерполяционный многочлен P_n , построенный по этой таблице, совпадает с H_n .

3.5. Теорема 3.1 фактически дает способ построения интерполяционных многочленов, коэффициенты которых определяются решением системы (3.4). Найдите интерполяционный многочлен для функции $y = f_1(x)$, проверьте выполнение равенств (3.3) и начертите графики таблицы и многочлена.

3.6. 1. Для функции f_2 с данной таблицей найдите интерполяционный многочлен на основе теоремы 3.1 и убедитесь, что он первой степени, а не второй.

x	0	1	2
$f_1(x)$	1	0	0

2. Придумайте пример таблицы с тремя узлами, для которой интерполяционный многочлен имеет нулевую степень.

x	0	1	2
$f_2(x)$	1	2	3

3.3. Оценка погрешности полиномиальной интерполяции

Пусть для заданной на $[a; b] = [x_0; x_n]$ табличной функции f получена интерполяционная формула

$$f(x) \approx P_n(x), \quad x \in [a; b] \quad (3.5)$$

с интерполяционным многочленом P_n . Оказывается, что когда функция f достаточное число раз непрерывно дифференцируема, есть возможность решить проблему оценки погрешностей формулы (3.5) независимо от способа построения интерполяционного многочлена. При этом остаточный член R_n и оценочная функция V_n формулы определяются только функцией f и выбором узлов интерполирования.

Введем в рассмотрение многочлен $(n + 1)$ -й степени, выраженный через узлы интерполирования: $Q_{n+1}(x) = (x - x_0)(x - x_1) \dots (x - x_n)$.

Теорема 3.2. Пусть дана табл. 3.1 значений функции $y = f(x)$, имеющей на $[a; b]$ непрерывные производные до $(n + 1)$ -го порядка включительно, и P_n — ее интерполяционный многочлен. Тогда для любого $x \in [a; b]$ существует точка $c \in (a; b)$ такая, что

$$f(x) - P_n(x) = R_n(x) = \frac{f^{(n+1)}(c)}{(n + 1)!} Q_{n+1}(x). \quad (3.6)$$

○ Если $x = x_i$ при некотором $i = 0, 1, \dots, n$, то соотношение (3.6) справедливо при любом $c \in (a; b)$, так как обе его части обращаются в нуль. Пусть $x = x^*$ — произвольная точка из $[a; b]$, не совпадающая с узлами интерполирования. Тогда $Q_{n+1}(x^*) \neq 0$.

Рассмотрим вспомогательную функцию Φ :

$$\Phi(x) = f(x) - P_n(x) - kQ_{n+1}(x), \quad (3.7)$$

где k — некоторый постоянный множитель. Очевидно, что Φ равна нулю в точках x_0, x_1, \dots, x_n .

Теперь подберем множитель k таким, чтобы функция Φ обращалась в нуль и в точке $x = x^*$:

$$f(x^*) - P_n(x^*) - kQ_{n+1}(x^*) = 0.$$

Так будет, если

$$k = \frac{f(x^*) - P_n(x^*)}{Q_{n+1}(x^*)}. \quad (3.8)$$

Далее считаем, что k определен по формуле (3.8).

В силу выбора множителя k функция Φ обращается в нуль по крайней мере $n + 2$ раза: $\Phi(x_0) = \dots = \Phi(x_n) = \Phi(x^*) = 0$, т. е. имеет равные значения в $n + 2$ различных точках: x_0, \dots, x_n и x^* . Тогда согласно теореме Ролля производная Φ' на интервале $(a; b)$ имеет не менее $n + 1$ различных нулей (см. также упр. 2.4).

Применяя эту теорему к Φ' , убеждаемся, что вторая производная Φ'' имеет не менее n различных нулей в $(a; b)$. Проводя далее аналогичные рассуждения, в итоге получим, что производная $\Phi^{(n+1)}$ обращается в нуль по крайней мере в одной точке из $(a; b)$.

Пусть точка $c \in (a; b)$ такова, что $\Phi^{(n+1)}(c) = 0$ ¹⁾. Если вычислить значение $(n+1)$ -й производной правой части (3.7) в точке c , то оно окажется равным нулю. Поскольку $(n+1)$ -я производная P_n равна нулю, а такая же производная многочлена $Q_{n+1}(x) = x^{n+1} + a_n x^n + \dots + a_1 x + a_0$ равна $(n+1)!$, будем иметь

$$f^{(n+1)}(c) - k(n+1)! = 0.$$

Подставляя сюда значение k , получим

$$f(x^*) - P_n(x^*) = \frac{f^{(n+1)}(c)}{(n+1)!} Q_{n+1}(x^*). \quad (3.9)$$

Для завершения доказательства осталось учесть то, что точка x^* была взята произвольным образом. •

Пусть $M_{n+1} = \max_{[a; b]} |f^{(n+1)}(x)|$. Тогда из соотношения (3.6) получаем оценочную функцию V_n интерполяционной формулы (3.5):

$$|f(x) - P_n(x)| \leq \frac{M_{n+1}}{(n+1)!} |Q_{n+1}(x)| \stackrel{\text{def}}{=} V_n(x), \quad x \in [a; b]. \quad (3.10)$$

С помощью функции V_n отыскивается абсолютная погрешность приближения $P_n(x)$ к точному числу $f(x)$ в точке $x \in [a; b]$:

$$\Delta_{P_n(x)} = V_n(x) \quad (3.11)$$

и абсолютная погрешность многочлена P_n на всем отрезке $[a; b]$

$$\Delta_{P_n} = \max_{[a; b]} V_n(x).$$

З а м е ч а н и е. В практических расчетах в качестве M_{n+1} обычно берется некоторое число q , удовлетворяющее неравенству $|f^{(n+1)}(x)| \leq q$ при всех $x \in [a; b]$.

Итак, мы научились оценивать погрешность замены функции f многочленом P_n , но для оценки точности результатов при вычислениях с помощью формулы (3.5) этого недостаточно. Как отмечалось в п. 1.6.2, надо учитывать также погрешность вычисления значения a многочлена P_n и для оценки погрешности приближенных равенств $f(x) \approx P_n(x) \approx a$ пользоваться формулой вида (1.15), тем более что точность коэффициентов интерполяционного мно-

¹⁾ Замечаем, что нули функций $\Phi, \Phi', \dots, \Phi^{(n-1)}$ зависят от выбора $x = x^*$. Следовательно, на самом деле $c = c(x)$.

гочлена, выражающихся через табличные данные, зависит от точности таблицы.

x	$\sin x$
0	0
0,5	0,5
1,0	0,8
1,5	1,0

Пример 3.1. Дана таблица значений функции $\sin x$ с верными цифрами. Обсудим вопрос о точности вычисления $\sin 0,7$ с помощью интерполяционного многочлена третьей степени P_3 .

Погрешность приближения $\sin 0,7 \approx P_3(0,7)$ оценивается числом $V_3(0,7) = M_4 \cdot Q_4(0,7)/4!$, где $Q_4(0,7) = = 0,7(0,7 - 0,5)(0,7 - 1)(0,7 - 1,5) \approx 0,034$ и $M_4 = 1$, поскольку $|f^{(4)}(x)| = \sin x \leq 1$ для любого $x \in [0; 1,5]$.

В итоге получим $V_3(0,7) = 0,0015$. Следовательно, интерполяционная формула обеспечивает две верные цифры после запятой, но, поскольку у табличных значений верных цифр мало, после вычисления $P_3(0,7)$ точность результата скорее всего окажется меньше. •

З а м е ч а н и е. Оказывается, что погрешность интерполирования на данном отрезке не обязательно стремится к нулю при неограниченном увеличении числа узлов [10, гл. 2, § 1], [20, гл. 3, § 2]. Из-за этого, в частности, на практике редко пользуются интерполяционными многочленами высокой степени, ограничиваясь $n = 1, 2, 3$.

Упражнение

3.7. Дана табличная функция.

x	0	0,2	0,4	0,6
$\cos x$	1	0,98	0,92	0,83

Найдите оценочную функцию для интерполяционного многочлена второй степени, построенного на узлах $0; 0,2; 0,4$.

3.4. Интерполяционный многочлен Лагранжа

Пусть узлы табл. 3.1 расположены произвольным образом. Интерполяционный многочлен будем искать в виде

$$L_n(x) = a_0(x - x_1) \dots (x - x_n) + a_1(x - x_0)(x - x_2) \dots (x - x_n) + \dots + a_n(x - x_0) \dots (x - x_{n-1}). \quad (3.12)$$

Коэффициенты a_0, \dots, a_n многочлена L_n подберем так, чтобы для него выполнялись условия (3.3).

При $x = x_0$ все слагаемые в (3.12), начиная со второго, равны нулю. Следовательно, $L_n(x_0) = a_0(x_0 - x_1) \dots (x_0 - x_n) = y_0$, и потому

$$a_0 = \frac{y_0}{(x_0 - x_1) \dots (x_0 - x_n)}.$$

Действуя по аналогичной схеме при $x = x_1, \dots, x_n$, выводим:

$$a_1 = \frac{y_1}{(x_1 - x_0)(x_1 - x_2) \dots (x_1 - x_n)},$$

$$\dots$$

$$a_n = \frac{y_n}{(x_n - x_0)(x_n - x_1) \dots (x_n - x_{n-1})}.$$

Подставив найденные коэффициенты в (3.12), получим общее выражение для *интерполяционного многочлена Лагранжа*:

$$L_n(x) = \sum_{i=0}^n \frac{(x - x_0) \dots (x - x_{i-1})(x - x_{i+1}) \dots (x - x_n)}{(x_i - x_0) \dots (x_i - x_{i-1})(x_i - x_{i+1}) \dots (x_i - x_n)} y_i. \quad (3.13)$$

Остаточный член и оценочная функция для *интерполяционной формулы Лагранжа*

$$f(x) \approx L_n(x), \quad x \in [a; b]$$

те же самые, что были получены в 3.3.

Пример 3.2. Найдем интерполяционный многочлен Лагранжа первой степени по узлам $x_0 = 0,5$, $x_1 = 0,7$ данной таблицы.

x	0,5	0,7	1
$f(x)$	2	2,4	3

Подставляя $y_0 = 2$, $y_1 = 2,4$, получим

$$L_1(x) = \frac{y_0}{x_0 - x_1}(x - x_1) + \frac{y_1}{x_1 - x_0}(x - x_0) =$$

$$= \frac{2}{0,5 - 0,7}(x - 0,7) + \frac{2,4}{0,7 - 0,5}(x - 0,5) = 2x + 1^1). \bullet$$

Построение многочлена Лагранжа сопряжено с большой вычислительной работой, однако он удобен тем, что его можно составлять по произвольной таблице. Далее рассмотрим более простые способы получения интерполяционных многочленов по таблицам с равноотстоящими узлами.

Упражнения

3.8. Дана таблица значений функции $y = e^x$.

1. Постройте интерполяционный многочлен Лагранжа L_2 .

2. Вычислите $e^{1,57} \approx L_2(1,57)$.

3. Оцените интерполяционную погрешность числа $L_2(1,57)$.

x	1,5	1,6	1,7
e^x	4,48	4,95	5,47

¹⁾ Кстати, таблица составлена именно для функции $y = 2x + 1$.

3.9. Убедитесь, что для производной Q'_{n+1} верны равенства $Q'_{n+1}(x_i) = (x_i - x_0) \dots (x_i - x_{i-1})(x_i - x_{i+1}) \dots (x_i - x_n)$ ($i = 0, 1, \dots, n$), и на основе этого факта выведите из (3.13) другой вид многочлена Лагранжа:

$$L_n(x) = Q_{n+1}(x) \sum_{i=0}^n \frac{y_i}{(x - x_i) Q'_{n+1}(x_i)}.$$

3.5. Интерполяционные многочлены Ньютона

Пусть для функции f построена таблица вида табл. 3.1 со значений s постоянным шагом $h > 0$. Как и раньше, обозначим $[a; b] = [x_0; x_n]$.

3.5.1. Конечные разности

В случае таблиц с равноотстоящими узлами используются величины, называемые конечными разностями. *Конечные разности первого порядка* — это разности между соседними табличными значениями: $\Delta y_0 = y_1 - y_0, \dots, \Delta y_{n-1} = y_n - y_{n-1}$. *Разности второго порядка*: $\Delta^2 y_i = \Delta y_{i+1} - \Delta y_i$ ($i = 0, 1, \dots, n-2$) и т.д.

Формула для *конечных разностей k -го порядка* ($k > 1$):

$$\Delta^k y_i = \Delta^{k-1} y_{i+1} - \Delta^{k-1} y_i \quad (i = 0, 1, \dots, n-k).$$

Для табл. 3.1 разностью наивысшего порядка будет $\Delta^n y_0$.

Можно определять разности функций в произвольной точке x при некотором шаге h : $\Delta f(x) = f(x+h) - f(x)$ (приращение функции), $\Delta^2 f(x) = \Delta f(x+h) - \Delta f(x)$ и т.д.

Конечные разности удобно располагать в форме таблиц (табл. 3.2).

Таблица 3.2

x	$f(x)$	Δy	$\Delta^2 y$...	$\Delta^n y$
x_0	y_0	Δy_0	$\Delta^2 y_0$...	$\Delta^n y_0$
x_1	y_1	Δy_1	$\Delta^2 y_1$...	
\vdots	\vdots	\vdots	\vdots	\vdots	
x_{n-2}	y_{n-2}	Δy_{n-2}	$\Delta^2 y_{n-2}$		
x_{n-1}	y_{n-1}	Δy_{n-1}			
x_n	y_n				

По таблице конечных разностей часто удается находить наилучшую степень интерполяционного многочлена. Если разности k -го

порядка на каком-то участке таблицы практически постоянны, то это значит, что здесь табличная функция близка к многочлену k -й степени [см. упр. 3.11 (2)].

Упражнения

3.10. Докажите следующие свойства конечных разностей:

1) Если $c = \text{const}$, то $\Delta c = 0$.

2) $\Delta(cf(x)) = c\Delta f(x)$.

3) $\Delta(f_1(x) + f_2(x)) = \Delta f_1(x) + \Delta f_2(x)$.

4) $\Delta(x^n) = nhx^{n-1} + \frac{n(n-1)}{2!}h^2x^{n-2} + \dots + nh^{n-1}x + h^n$.

3.11. Пусть P_n — многочлен степени n . Убедитесь, что разность k -го порядка $\Delta^k P_n(x)$ является: 1) многочленом степени $n - k$ при $k < n$; 2) ненулевой константой при $k = n$; 3) нулем при $k > n$.

У к а з а н и е: Воспользуйтесь свойствами конечных разностей из упр. 3.10.

3.12. Покажите, что табличные значения y_k ($k = 1, 2, \dots, n$) можно выразить через конечные разности:

$$\left. \begin{aligned} y_1 &= y_0 + \Delta y_0, y_2 = y_0 + 2\Delta y_0 + \Delta^2 y_0, \dots, \\ y_n &= y_0 + n\Delta y_0 + \frac{n(n-1)}{2!}\Delta^2 y_0 + \dots + \Delta^n y_0. \end{aligned} \right\} \quad (3.14)$$

3.13. Найдите значения функции $y = x^2 + 3$ для $x = 1, 2, 3, 4, 5$ и постройте таблицу конечных разностей до Δ_y^3 . Сопоставьте полученный результат и последние два вывода из упр. 3.11.

3.5.2. Первый интерполяционный многочлен Ньютона

Будем искать интерполяционный многочлен n -й степени в виде

$$\begin{aligned} P_n^{(1)}(x) &= a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1) + \\ &+ \dots + a_n(x - x_0)\dots(x - x_{n-1}). \end{aligned} \quad (3.15)$$

Коэффициенты a_0, a_1, \dots, a_n находим, исходя из условий (3.3).

Пусть $x = x_0$. Тогда $P_n^{(1)}(x_0) = a_0 = y_0 \Rightarrow a_0 = y_0$.

При $x = x_1$

$$P_n^{(1)}(x_1) = y_0 + a_1(x_1 - x_0) = y_1 \Rightarrow a_1 = \frac{y_1 - y_0}{h} = \frac{\Delta y_0}{h}.$$

При $x = x_2$

$$P_n^{(1)}(x_2) = y_0 + \frac{\Delta y_0}{h}(x_2 - x_0) + a_2(x_2 - x_0)(x_2 - x_1) = y_2,$$

или, используя выражение для y_2 из (3.14),

$$y_0 + 2\Delta y_0 + 2!h^2 a_2 = y_0 + 2\Delta y_0 + \Delta^2 y_0.$$

Отсюда имеем $a_2 = \frac{\Delta^2 y_0}{2!h^2}$. Аналогичным образом можно получить

$$a_3 = \frac{\Delta^3 y_0}{3!h^3}, \dots, a_n = \frac{\Delta^n y_0}{n!h^n}.$$

Подставим найденные значения в (3.15):

$$\begin{aligned} P_n^{(1)}(x) = & y_0 + \frac{\Delta y_0}{h}(x - x_0) + \frac{\Delta^2 y_0}{2!h^2}(x - x_0)(x - x_1) + \\ & + \dots + \frac{\Delta^n y_0}{n!h^n}(x - x_0) \dots (x - x_{n-1}). \end{aligned} \quad (3.16)$$

Многочлен (3.16) называется *первым интерполяционным многочленом Ньютона* (или *в форме Ньютона*), а приближенное равенство

$$f(x) \approx P_n^{(1)}(x), \quad x \in [a; b] \quad (3.17)$$

— *первой интерполяционной формулой Ньютона*. Видно, что коэффициенты в (3.16) содержат соответствующее узлу x_0 значение функции y_0 и числа из верхней строки таблицы конечных разностей.

Погрешности формулы (3.17) оцениваются по правилам, приведенным в 3.3.

На практике чаще всего используется другая форма записи многочлена $P_n^{(1)}$. Положим

$$t = \frac{x - x_0}{h} \quad \text{или} \quad x = x_0 + th. \quad (3.18)$$

Тогда $x - x_1 = h(t - 1)$, $x - x_2 = h(t - 2)$, ..., $x - x_{n-1} = h(t - n + 1)$. С учетом полученных соотношений многочлен (3.16) примет вид

$$\begin{aligned} P_n^{(1)}(x) = P_n^{(1)}(x_0 + th) = & y_0 + t\Delta y_0 + \frac{t(t-1)}{2!} \Delta^2 y_0 + \\ & + \dots + \frac{t(t-1) \dots (t-n+1)}{n!} \Delta^n y_0. \end{aligned} \quad (3.19)$$

Исходя из общего вида остаточного члена (3.6) полиномиальной интерполяционной формулы и ее оценочной функции (3.10), для формулы (3.17) с интерполяционным многочленом (3.19) получаем выражения остаточного члена

$$R_n^{(1)}(x) = f(n+1)(c) \frac{h^{n+1}}{(n+1)!} t(t-1) \dots (t-n) \quad (c \in (a; b)) \quad (3.20)$$

и оценочной функции

$$V_n^{(1)}(x) = M_{n+1} \frac{h^{n+1}}{(n+1)!} |t(t-1)\dots(t-n)|, \quad (3.21)$$

где $M_{n+1} = \max_{[a;b]} |f^{(n+1)}(x)|$.

С помощью первой интерполяционной формулы Ньютона обычно вычисляют значения функции f для x , близких к x_0 , т. е. находящихся в начале таблицы. Это обстоятельство следует учитывать при выборе узлов для построения $P_n^{(1)}$.

Пример 3.3. В табл. 3.3 все значения функции $y = \ln x$ приведены с верными значащими цифрами. Требуется вычислить $\ln 1,64$ с помощью первого интерполяционного многочлена Ньютона и оценить погрешность результата.

Таблица 3.3

x	$\ln x$	Δy	$\Delta^2 y$
1,5	0,405	0,065	-0,004
1,6	0,470	0,061	-0,004
1,7	0,531	0,057	-0,003
1,8	0,588	0,054	
1,9	0,642		

Поскольку вторые разности таблицы практически постоянны, хорошее приближение должен дать многочлен второй степени. Его получим из (3.19),

выбрав узлы $x_0 = 1,6$, $x_1 = 1,7$, $x_2 = 1,8$. Используя данные второй строки табл. 3.3, на отрезке $[1,6; 1,8]$ будем иметь интерполяционную формулу

$$\ln x \approx P_2(x) = 0,470 + t \cdot 0,061 - \frac{t(t-1)}{2} \cdot 0,004, \quad (*)$$

где $t = \frac{x-1,6}{0,1}$.

Оценим погрешность этой формулы в точке $x = 1,64$. Поскольку $|(\ln x)'''| \leq 0,49$ для всех $x \in [1,6; 1,8]$, берем $M_3 = 0,49$. Учитывая значение $t = 0,4$, на основе (3.21) находим

$$V_2(1,64) = \frac{0,49 \cdot (0,1)^3}{3!} |0,4(0,4-1)(0,4-2)| = 0,32 \cdot 10^{-4}.$$

Таким образом, при $x = 1,64$ погрешность интерполяции на порядок меньше погрешности таблицы $0,5 \cdot 10^{-3}$, поэтому точность результата определится точностью табличных данных. Чтобы узнать степень влияния этих данных, сначала проведем вычисления по формуле (*):

$$\ln 1,64 \approx 0,470 + 0,4 \cdot 0,061 - \frac{0,4 \cdot (-0,6)}{2} \cdot 0,004 \approx 0,4949,$$

а затем найдем оценку ν точности приближения $0,4949 \approx P_2(1,64)$: $|P_2(1,64) - 0,4949| \leq \nu$. Для этого применим правила учета погрешностей арифметических действий и формулу (1.25), имея в виду, что число $t = 0,4$ здесь точное. Абсолютные погрешности таблич-

ных данных определим по верным цифрам значений y_i : у чисел y_i они равны 0,0005, у разностей $\Delta y_i - 0,001$, у $\Delta^2 y_i - 0,002$. Тогда

$$\nu = 0,0005 + 0,001 \cdot 0,4 + 0,002 \cdot \frac{0,4 \cdot 0,6}{2} = 0,0012.$$

Как видим, полученная оценка ν действительно является основной в суммарной оценке погрешности приближения 0,4949 к $\ln 1,64$:

$$|\ln 1,64 - 0,4949| \leq V_3(1,64) + \nu \leq 0,0013.$$

Итак, теоретически число 0,4949 имеет только две верные значащие цифры. Для сравнения приведем искомое значение логарифма с пятью верными значащими цифрами: $\ln 1,64 \approx 0,49470$. Расстояние $|0,49470 - 0,4949| = 0,0002$ показывает, что при учете вычислительной погрешности получена завышенная ее оценка (в основном за счет завышения оценок погрешностей значений y_i и конечных разностей по сравнению с их реальными погрешностями). На самом деле верными являются три цифры после запятой, т. е. точность результата оказалась такой же, как и точность таблицы. •

Упражнения

3.14. Для табличной функции из упр. 3.13 постройте по формуле (3.16) многочлен $P_2^{(1)}$ с узлами $x = 1, 2, 3$. Объясните полученный результат (см. упр. 3.4).

3.15. С помощью интерполяционного многочлена из примера 3.3 найдите приближенно $\ln 1,61$, $\ln 1,76$ и оцените точность результатов.

3.16. Найдите на основе табл. 3.3 приближенное значение $\ln 1,76$ с применением интерполяционного многочлена Ньютона первой степени и оцените погрешность результата.

3.5.3. Второй интерполяционный многочлен Ньютона

При вычислении значений функции для x из конца таблицы, т. е. близких к x_n , применяется *второй интерполяционный многочлен Ньютона*:

$$P_n^{(2)}(x) = y_n + \frac{\Delta y_{n-1}}{h} (x - x_n) + \frac{\Delta^2 y_{n-2}}{2!h^2} (x - x_n)(x - x_{n-1}) + \dots + \frac{\Delta^n y_0}{n!h^n} (x - x_n) \dots (x - x_1)^1. \quad (3.22)$$

¹⁾ Он строится в точном соответствии с первым, только данные табл. 3.2 используются в обратном порядке: снизу вверх.

Если взять $t = \frac{x - x_n}{h}$ или $x = x_n + th$, многочлен (3.22) запишется в форме

$$P_n^{(2)}(x) = P_n^{(2)}(x_n + th) = y_n + t\Delta y_{n-1} + \frac{t(t+1)}{2!} \Delta^2 y_{n-2} + \dots + \frac{t(t+1)\dots(t+n-1)}{n!} \Delta^n y_0. \quad (3.23)$$

Остаточный член и оценочная функция *второй интерполяционной формулы Ньютона*

$$f(x) \approx P_n^{(2)}(x), \quad x \in [a; b],$$

с многочленом в форме (3.23) будут иметь вид

$$R_n^{(2)}(x) = f^{(n+1)}(c) \frac{h^{n+1}}{(n+1)!} t(t+1)\dots(t+n) \quad (c \in (a; b)); \quad (3.24)$$

$$V_n^{(2)}(x) = M_{n+1} \frac{h^{n+1}}{(n+1)!} |t(t+1)\dots(t+n)|, \quad (3.25)$$

где $M_{n+1} = \max_{[a; b]} |f^{(n+1)}(x)|$.

Упражнения

3.17. Выведите второй интерполяционный многочлен Ньютона в формах (3.22) и (3.23), а также формулы (3.24) и (3.25).

3.18. Выполните задание из примера 3.3 с помощью многочлена $P_2^{(2)}(x)$.

3.19. В упр. 3.5 уже был получен интерполяционный многочлен второй степени. Постройте для этой же таблицы многочлен Лагранжа и оба многочлена Ньютона. Объясните результат.

3.6. Линейное интерполирование

Остановимся отдельно на простейшем виде локального аналитического приближения таблиц, которым является линейное интерполирование. Во-первых, во многих случаях его вполне достаточно для вычисления значений табличных функций, особенно математических. Во-вторых, в этом случае можно дать простое правило определения верных значащих цифр приближенного значения функции непосредственно по свойствам исходной таблицы, без вычисления абсолютной погрешности этого значения. В-третьих, в линейном случае легко решается часто встречающаяся задача *обратного интерполирования*, заключающаяся в поиске аналитического приближения функции, обратной для данной табличной функции.

3.6.1. Формула линейного интерполирования

Пусть имеется таблица значений функции f с постоянным шагом $h > 0$ и требуется по табличным данным найти $f(x)$ при x , не совпадающем с табличными аргументами. Для этого обозначим через x_0, x_1 ($x_0 < x_1$) два соседних табличных аргумента, между которыми находится x , через y_0, y_1 — соответствующие табличные значения, и из первой интерполяционной формулы Ньютона при $n = 1$ получим

$$f(x) \approx y_0 + \frac{x - x_0}{h} \Delta y_0 \stackrel{\text{def}}{=} P_1(x). \quad (3.26)$$

Это и есть *формула линейного интерполирования*.

Слагаемое $\frac{x - x_0}{h} \Delta y_0$ называется *поправкой* к значению y_0 , соответствующей отклонению x от x_0 .

Пусть вторая производная функции f непрерывна на $[x_0; x_1]$. Тогда абсолютные погрешности приближений $P_1(x)$ к значениям $f(x)$, вычисленным по формуле (3.26) для $x \in [x_0; x_1]$, можно находить с помощью оценочной функции V_1 . Выражения для этой функции получаются из соотношений (3.10) или (3.21):

$$V_1(x) = \frac{M_2}{2} |(x - x_0)(x - x_1)| \text{ или } V_1(x) = \frac{M_2 h^2}{2} |t(t - 1)|, \quad (3.27)$$

где $t = \frac{x - x_0}{h}$ и $M_2 = \max_{[x_0; x_1]} |f''(x)|$.

В случае линейной интерполяции удобно пользоваться общей оценкой погрешностей для всех $x \in [x_0; x_1]$, которую выведем с помощью второй формулы из (3.27). Учитывая то, что $t \in [0; 1]$ при $x \in [x_0; x_1]$, а также равенство

$$\max_{[0; 1]} \left| \frac{t(t - 1)}{2} \right| = \max_{[0; 1]} \frac{t - t^2}{2} = \frac{1}{8},$$

для абсолютной погрешности функции P_1 на $[x_0; x_1]$ получим формулу

$$\Delta_{P_1} = \max_{[x_0; x_1]} V_1(x) = \frac{1}{8} M_2 h^2.$$

Следовательно,

$$|f(x) - P_1(x)| = |R_1(x)| \leq \frac{1}{8} M_2 h^2, \quad x \in [x_0; x_1]. \quad (3.28)$$

Покажем, что при оценке погрешностей линейной интерполяции можно избавиться от вычисления f'' и поиска числа M_2 . Для этого установим связь между f'' и конечной разностью второго порядка.

Возьмем $x_2 = x_1 + h$ и предположим, что производная f'' непрерывна также на $[x_1; x_2]$. По теореме Лагранжа имеем

$$\Delta y_0 = f(x_0 + h) - f(x_0) = f'(c_1)h,$$

где $c_1 \in (x_0; x_1)$. Далее с помощью этой же теоремы найдем выражение для второй разности через f'' :

$$\Delta^2 y_0 = h\Delta f'(c_1) = h(f'(c_1 + h) - f'(c_1)) = f''(c_2)h^2,$$

где число c_2 принадлежит интервалу $(c_1; c_1 + h) \subset (x_0; x_2)$.

В силу непрерывности f'' на $[x_0; x_2]$ при малом шаге h с большой степенью точности можно считать, что $f''(x) \approx f''(x_0)$ для всех $x \in [x_0; x_2]$. Отсюда вытекает, что

$$\Delta^2 y_0 \approx f''(x_0)h^2 \text{ или } f''(x_0) \approx \frac{\Delta^2 y_0}{h^2},$$

а также $M_2 \approx |f''(x_0)|$. Полученные соотношения позволяют вместо (3.28) применять достаточно хорошую приближенную оценку:

$$|R_1(x)| \approx \frac{1}{8} |\Delta^2 y_0|, \quad x \in [x_0; x_1]. \quad (3.29)$$

Пример 3.4. Чтобы вычислить приближения $a \approx \sin 1,11$ и $b \approx \sin 1,14$ по табл. 3.4 с помощью линейной интерполяции, надо взять $x_0 = 1,1$, $x_1 = 1,2$. Тогда $y_0 = 0,891$, $y_1 = 0,933$, $\Delta y_0 = 0,042$. Шаг таблицы $h = 0,1$. Для определения точности, обеспечиваемой в данном случае формулой (3.26), воспользуемся всеми тремя полученными нами оценками. Цифры табличных значений функции будем считать верными.

Таблица 3.4

x	$\sin x$	Δy	$\Delta^2 y$
1,0	0,841	0,050	-0,008
1,1	0,891	0,042	-0,011
1,2	0,933	0,031	
1,3	0,964		

Поскольку $|(\sin x)''| \leq \sin 1,2 \leq 0,94$ для всех $x \in [1,1; 1,2]$, берем $M_2 = 0,94$. Сначала применим обнисе для всех $x \in [1,1; 1,2]$ неравенства (3.28) и (3.29). На основании (3.28) получим

$$|R_1(x)| \leq \frac{1}{8} \cdot 0,94 \cdot 0,01 \leq 0,0012.$$

Так как $\Delta^2 y_0 = -0,011$, из (3.29) следует, что

$$|R_1(x)| \approx \frac{0,011}{8} \approx 0,0014.$$

Обе оценки фактически одинаковы и показывают, что не только числам a и b , но и всем другим приближениям к $\sin x$, $x \in [1,1; 1,2]$, формула (3.26) обеспечивает по крайней мере две верные значащие цифры после десятичной запятой.

Можно ожидать, что оценочные функции (3.27), характеризующие погрешность интерполяции при конкретных x , дадут нам более точные оценки. Проверим это. При $x = 1,11$ будет $t = 0,1$, поэтому

$$V_1(1,11) = \frac{0,94 \cdot 0,01}{2} \cdot 0,1 \cdot 0,9 = 0,00043.$$

Аналогичным образом получаем $V_1(1,14) = 0,0012$. Следовательно, на самом деле формула (3.26) при $x = 1,11$ дает приближение с тремя верными цифрами в дробной части, а не с двумя. В точке $x = 1,14$ оценка совпала с общей ($x = 1,14$ дальше от узла, чем $x = 1,11$, и потому точность интерполяции там меньше).

Найдем искомые приближения:

$$\sin 1,11 \approx P_1(1,11) = 0,891 + \frac{1,11 - 1,1}{0,1} 0,042 = 0,8952 = a,$$

$$\sin 1,14 \approx P_1(1,14) = 0,891 + \frac{1,14 - 1,1}{0,1} 0,042 = 0,9078 = b.$$

Без учета вычислительных погрешностей число a имеет верные цифры 8, 9 и 5, а число b — цифры 9 и 0.

Для оценки точности вычислений определяем абсолютные погрешности чисел $y_0 = 0,891$ и $\Delta y_0 = 0,042$ по известным верным цифрам табличных значений (см. аналогичную оценку в примере 3.3).

Точность значения $P_1(1,11)$ оценится числом

$$\nu = 0,0005 + 0,1 \cdot 0,001 = 0,0006.$$

Добавляем 0,0006 к подсчитанной ранее оценке 0,00043 погрешности интерполяции и получим оценку полной погрешности: $\Delta_a = 0,0011$. Значит, вычисления ухудшили качество цифры 5 числа 0,8952, и ее придется считать сомнительной.

Оценка ν погрешности вычисления $P_1(1,14)$ равна 0,0009 и незначительно увеличивает найденную выше оценку 0,0012 точности формулы. Поскольку их сумма дает $\Delta_b = 0,0021$, у числа 0,9078 также две верные значащие цифры.

Чтобы иметь представление о реальной точности полученных приближений, приведем значения синуса с пятью верными значащими цифрами:

$$\sin 1,11 \approx 0,89570, \quad \sin 1,14 \approx 0,90863. \bullet$$

Понятно, что результаты вычислений по формуле (3.26) не могут иметь больше верных цифр, чем их имеют табличные значения функции в используемых узлах. Какие цифры верны на самом деле,

можно узнать по завышенным оценкам погрешности интерполяции, найденным с помощью (3.28) или (3.29), не учитывая при этом вычислительные погрешности (в предположении, что вычисления ведутся с максимально возможной точностью).

Оценка (3.29) позволяет вывести правило определения верных цифр непосредственно по таблице конечных разностей. Пусть k — номер разряда десятичной записи числа $P_1(x)$. Если разряд в целой части числа, то $k \geq 0$, если в дробной, то $k < 0$. Цифра в k -м разряде верная, если $|R_1(x)| \leq 0,5 \cdot 10^k$. Согласно (3.29), это условие можно считать выполненным, как только окажется $|\Delta^2 y_0| \leq 4 \cdot 10^k$. Отсюда и вытекает следующее правило.

П р а в и л о. Если на каком-либо участке таблицы модули конечных разностей второго порядка имеют в k -м разряде не более четырех единиц, то у приближенных значений функции f , найденных с помощью линейного интерполирования для x из этого участка, цифры k -го разряда будут верными.

Замечасм, что в табл. 3.4 модули разностей второго порядка в разряде десятых и сотых долей ($k = -1$ и $k = -2$) имеют не более четырех единиц (0 и 1), а в разряде тысячных долей — 8 и 11 единиц. Следовательно, если искать приближения $\sin x$ для $x \in [1; 1,3]$ по формуле (3.26), то в разрядах десятых и сотых долей будут получаться верные цифры, а в разряде тысячных долей — сомнительные.

Сравните сказанное с результатами примера 3.4.

Согласно сформулированному правилу линейное интерполирование табл. 3.3 позволяет получать более точные результаты, а именно, с тремя верными цифрами после десятичной запятой.

Стандартные математические таблицы составляются так, что условие $|\Delta^2 y_i| \leq 4 \cdot 10^k$ выполняется для последнего разряда табличных значений по всей сетке аргументов (или почти по всей, в зависимости от функции). Тем самым *линейное интерполирование математических таблиц обычно обеспечивает такую же точность, какая имеется у таблицы.*

Упражнения

3.20. В упр. 3.16 использована формула линейного интерполирования. Определите верные цифры отыскиваемого там приближения по разностям табл. 3.3.

3.21. Найдите приближения $\sin 1,02$ и $\sin 1,06$ линейным интерполированием табл. 3.4 и исследуйте их погрешности по схеме из примера 3.4.

3.22. Проведите геометрическую иллюстрацию вычислений по формуле линейного интерполирования. Укажите на рисунке значения $f(x)$, $P_1(x)$, а также соответствующую поправку и погрешность при некотором $x \in (x_0; x_1)$.

3.6.2. Обратное линейное интерполирование

До сих пор с помощью интерполяционных формул вычисляли значение табличной функции f по данному аргументу x . На практике часто приходится решать обратную задачу: дано какое-то значение y функции f , не равное табличным значениям y_i , и необходимо найти соответствующий аргумент x .

Речь в этом случае идет о вычислении значений *обратной* по отношению к $y = f(x)$ функции, которую мы обозначим $x = \varphi(y)$. Если f монотонна в области вычислений, то обратная для нее функция там существует, причем она также табличная (см. табл. 3.1), только теперь y_i являются аргументами, а x_i — значениями φ ($i = 0, 1, \dots, n$).

Например, на отрезке $[1,0; 1,3]$ обратной функцией для функции $y = \sin x$ будет $x = \arcsin y$. Таблицу ее значений можно получить из табл. 3.4:

y	0,841	0,891	0,933	0,964
$\arcsin y$	1,0	1,1	1,2	1,3

Табличные аргументы y_i обратной функции будут равноотстоящими лишь в случае линейной зависимости между x и y , поэтому для обратного интерполирования в общем случае подойдет многочлен Лагранжа. Он имеет такой же вид, как в (3.13), только x везде заменяется на y , и наоборот. Аналогично поступаем при записи остаточного члена и оценочной функции, дополнительно заменяя в (3.6) производную $f^{(n+1)}(c)$, где $c \in (x_0; x_n)$, на $\varphi^{(n+1)}(c)$ при некотором c между y_0 и y_n , а в (3.10) взяв число $M_{n+1} = \max|\varphi^{(n+1)}(y)|$, где наибольшее значение находится по всем y между y_0 и y_n .

Формула *обратного линейного интерполирования*

$$x = \varphi(y) \approx x_0 + \frac{y - y_0}{\Delta y_0} h \stackrel{\text{def}}{=} P_1(y) \quad (3.30)$$

получается из (3.26). Для вычислений с ее помощью при заданном y выбираются два соседних табличных значения функции f , между которыми находится y , и предыдущее значение принимается за y_0 , а последующее — за y_1 .

Остаточный член формулы (3.30) выглядит следующим образом:

$$\varphi(y) - P_1(y) = R_1(y) = \frac{\varphi''(c)}{2} (y - y_0)(y - y_1), \quad (3.31)$$

где c — некоторое число между y_0 и y_1 . Следовательно,

$$|\varphi(y) - P_1(y)| \leq \frac{M_2}{2} |(y - y_0)(y - y_1)| \stackrel{\text{def}}{=} V_1(y), \quad (3.32)$$

где $M_2 = \max|\varphi''(y)|$, y находится между y_0 и y_1 .

Общая оценка погрешности формулы для всех y , находящихся между y_0 и y_1 , обеспечивается неравенством

$$|R_1(y)| \leq \frac{1}{8} M_2(\Delta y_0)^2. \quad (3.33)$$

Упрощенную оценку вида (3.29) с выражением M_2 через табличную разность второго порядка функции φ получить в данном случае не удастся, поскольку первые разности этой функции равны шагу h исходной таблицы и потому вторые разности заведомо равны нулю.

Пример 3.5. Пусть требуется найти приближенное значение для $e^{0,432}$ линейным интерполированием табл. 3.3. Поскольку функции $y = \ln x \stackrel{\text{def}}{=} f(x)$ и $x = e^y \stackrel{\text{def}}{=} \varphi(y)$ взаимнообратные, поставленную задачу можно переформулировать так: по значению 0,432 функции $y = \ln x$ найти такой аргумент $x = a$, чтобы выполнялось приближенное равенство $\ln a \approx 0,432$. Тогда $e^{0,432} \approx a$.

Число 0,432 находится между $y_0 = 0,405$ и $y_1 = 0,470$, поэтому $x_0 = 1,5$, $x_1 = 1,6$ и $\Delta y_0 = 0,065$. При y между y_0 и y_1 значения $|\varphi''(y)| = |(e^y)''| = e^y$ лежат между 1,5 и 1,6, значит, $M_2 = 1,6$. Найдем оценку погрешности интерполяции сначала по формуле (3.32)

$$V_1(0,432) = \frac{1,6}{2} |(0,432 - 0,405)(0,432 - 0,470)| = 0,00083,$$

а затем по формуле (3.33)

$$\Delta_a = \frac{1,6}{8} (0,065)^2 = 0,00084.$$

Тем самым вычисленное по формуле (3.30) искомое значение

$$e^{0,432} \approx 1,5 + \frac{0,432 - 0,405}{0,065} \cdot 0,1 \approx 1,5 + 0,042 = 1,542$$

имеет три верные значащие цифры. •

Упражнения

3.23. Запишите в общем виде интерполяционную формулу Лагранжа для обратного интерполирования, ее остаточный член и оценочную функцию.

3.24. Проиллюстрируйте графически формулу (3.30), отметив значение $x = \varphi(y)$ и соответствующую поправку при некотором y между y_0 и y_1 (на основе рисунка из упр. 3.22).

3.25. Выведите оценку (3.33).

3.26. По табл. 3.4 обратным линейным интерполированием найдите $\arcsin 0,856$ и определите верные значащие цифры полученного приближенного значения.

3.7. Приближение табличных функций по методу наименьших квадратов

3.7.1. Постановка задачи

Пусть в результате экспериментов получена табл. 3.1 с произвольным расположением аргументов. Аналитическое выражение табличной функции f может быть неизвестным. На основе этой таблицы требуется найти формулу $y = p(x)$, приближенно описывающую зависимость между экспериментальными данными таблицы.

Полученное для этой цели соотношение $y = p(x)$ называется *эмпирической формулой*, а функция p — *эмпирической функцией*.

Возможным вариантом решения задачи является интерполирование. Однако этот способ, требующий обязательного совпадения значений табличной и приближающей функций во всех табличных аргументах, в данном случае малоприменим. При большом количестве узлов он является неудобным и сложным, ибо потребует отыскания либо многочлена большой степени, либо другой громоздкой функции с «извилистым» графиком, проходящим через все табличные точки. В этой связи следует учесть замечание из 3.3 о том, что высокая степень интерполяционного многочлена не оправдана.

Кроме того, экспериментальные данные в силу ряда причин могут иметь трудно учитываемые случайные или систематические ошибки. В этих условиях интерполирование вообще становится сомнительным. Вместо того чтобы «сглаживать» случайные ошибки, интерполирующая функция включает в себя вместе с данными все их погрешности и может в результате оказаться слишком грубым приближением. Часто с помощью какой-либо простой функции с проходящим около табличных точек графиком удается добиться эффекта сглаживания ошибок и получить более точное приближение.

Иллюстрацией сказанного является упр. 3.2. Для первой таблицы хорошей эмпирической формулой будет $y = x^2 + 1$, а для второй — формула $y = e^x$. Параболическая интерполяция в обоих случаях приводит к многочлену 5-й степени и искажает существующую между табличными данными зависимость.

По этим причинам впрямь не будем требовать от функции p обязательного выполнения равенств $p(x_i) = y_i$. Главное, чтобы она была достаточно простой, учитывала характер табличной функции и в точках x_i имела близкие к y_i значения при всех $i = 0, \dots, n$.

Поиск эмпирической формулы начинается с определения класса функций, которые лучше всего отражают связь между табличными данными. Эффективным методом для этого являются графические соображения. На координатной плоскости отмечаются опреде-

ляемые данной таблицей точки, а затем по характеру их расположения подбирается вид приближения из числа известных элементарных функций.

В перечень наиболее часто используемых классов функций входят, например, многочлены, тригонометрические функции $y = \sin kx$, $y = \cos kx$ и их линейные комбинации, логарифмические функции $y = a \ln x + b$, показательные функции $y = ae^x + b$, дробно-линейные функции $y = \frac{1}{ax + b} + c$ и т. д.

На рис. 3.2 вблизи точек некоторой таблицы проведена линия $y = p(x)$, напоминающая часть квадратной параболы. В этом случае искомой эмпирической формулой будет $y = ax^2 + bx + c$ при некоторых значениях коэффициентов a , b , c . Как видно из рисунка, для данной таблицы можно выбрать и более простое приближение в виде линейной функции $y = ax + b$.

Пусть тип эмпирической формулы $y = p(x)$ выбран, причем, как мы заметили, функция p на самом деле зависит от одного или нескольких числовых параметров. Чтобы найти функцию из выбранного класса, график которой в каком-либо смысле ближе всех расположен к табличным точкам, надо определить соответствующие значения этих параметров.

Есть несколько способов поиска. Мы рассмотрим следующий: возьмем в качестве меры близости функций p и f на отрезке $[x_0, x_n]$ «расстояние» $\bar{\rho}(p, f)$ между ними (п. 1.2.4) и наилучшей функцией p будем считать ту, для которой $\bar{\rho}(p, f)$ имеет наименьшее значение.

При определении «расстояния» $\bar{\rho}$ будем пользоваться евклидовым расстройением (1.3) между векторами-значениями (y_0, y_1, \dots, y_n) и $(p(x_0), p(x_1), \dots, p(x_n))$:

$$\bar{\rho}(p, f) = \sqrt{\sum_{i=0}^n (p(x_i) - y_i)^2}. \quad (3.34)$$

Числа $v_i = p(x_i) - y_i$ называют *уклонениями*, а вычисленное по формуле (3.34) «расстояние» $\bar{\rho}(p, f)$ — *среднеквадратичным уклонением* функции p от табличной функции f на отрезке $[x_0, x_n]$. Абсолютные величины уклонений равны длинам отрезков между точками (x_i, y_i) и соответствующими точками графика $y = p(x)$ (на рис. 3.2 эти отрезки отмечены сплошными вертикальными линиями).

Ясно, что задача минимизации «расстояния» (3.34) сводится к минимизации подкоренного выражения.

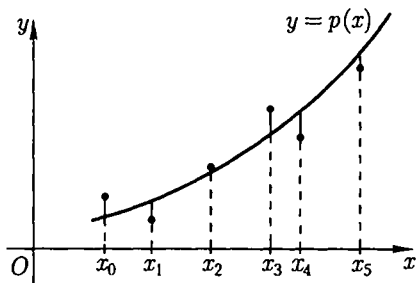


Рис. 3.2

Теперь, после всех уточнений, можно сформулировать задачу приближения по методу наименьших квадратов.

Пусть связь между аргументами x_i и значениями y_i , табл. 3.1 приближенно описывается формулой $y = p(x, a_1, \dots, a_k)$ с числовыми параметрами a_1, \dots, a_k . Требуется определить такие значения этих параметров, при которых сумма квадратов отклонений

$$\sum_{i=0}^n v_i^2 = \sum_{i=0}^n (p(x_i, a_1, \dots, a_k) - y_i)^2 \quad (3.35)$$

будет наименьшей.

Если $\bar{a}_1, \dots, \bar{a}_k$ — искомые значения параметров, то их называют наилучшими параметрами, соотношение $y = p(x, \bar{a}_1, \dots, \bar{a}_k)$ — наилучшей эмпирической формулой данного класса, а функцию p — наилучшей функцией из данного класса функций¹⁾.

Здесь не ставится задача оценки погрешностей приближенного равенства

$$f(x) \approx p(x, \bar{a}_1, \dots, \bar{a}_k)$$

во всех точках между x_0 и x_n . При интерполировании для этого были получены соответствующие формулы, но там предполагалось, что функция f имеет достаточно хорошее аналитическое выражение. В случае эмпирических таблиц, с которыми имеем дело в данном параграфе, никакой информации об f , кроме ее значений в точках x_i , может не оказаться.

По этой причине будем довольствоваться тем, что научимся из некоторого класса функций выбирать наилучшую функцию p , и тем, что сумеем найти модули погрешностей ее значений в точках x_i , равные $|v_i|$, а также глобальную характеристику близости p к табличной функции — среднеквадратичное отклонение $\bar{\rho}(f, p)$.

Если для таблицы можно указать несколько классов эмпирических функций, то сначала из каждого класса отыскивается наилучшая функция, а затем из них выбирается та, которая дает наименьшее среднеквадратичное отклонение.

Упражнения

3.27. Между данными табл. 3.5 можно установить линейную зависимость $y = ax + b$. Рассмотрите функции, получающиеся при: 1) $a = 0,8, b = -1$; 2) $a = 1,2, b = -0,9$, и определите, которая из них является лучшим приближением к табличной функции в смысле «расстояния» $\bar{\rho}$. Начертите соответствующие графики.

3.28. В результате экспериментов получена следующая таблица:

Таблица 3.5

x	y
1	0,2
1,5	0,5
2	1,1
3	2,2

¹⁾ Не подчеркиваем каждый раз, что они найдены по методу наименьших квадратов.

x	-0,2	0,5	1	1,2	1,5	2	2,5	2,9	3,3	4	4,5	5
$f(x)$	-0,3	0,6	0,9	1,1	1,4	0,8	0,5	-0,1	-0,3	-0,8	-0,4	0,1

Начертите график функции по данным таблицы и подберите два класса функций, которыми можно приблизить функцию f .

3.7.2. Полиномиальное приближение по методу наименьших квадратов

Познакомимся с основными идеями поиска наилучших эмпирических функций на примере многочленов. Особенности этой работы в случае некоторых других классов функций можно найти в [8] (см. также упр. 3.31 – 3.33).

Итак, возьмем многочлен k -й степени

$$P_k(x) = a_k x^k + \dots + a_1 x + a_0 \quad (3.36)$$

и выведем правило нахождения его наилучших коэффициентов.

Сначала скажем несколько слов о степени многочлена (3.36). На самом деле, значение k не влияет на способ решения задачи, но приведенные ниже рассуждения полезны в чисто практическом плане.

Если взять $k \geq n$, то наименьшее значение $\bar{\rho}(f, P_k) = 0$ дают интерполяционные многочлены, ибо для них $P_k(x_i) = y_i$ при всех $i = 0, 1, \dots, n$. При $k = n$ такой многочлен единственный, а при $k > n$ их бесконечное множество (замечание после теоремы 3.1). Поскольку методы интерполяции нам уже известны и, кроме того, от идеи интерполирования мы отказались, будем считать $k < n$. Независимо от количества табличных данных на практике обычно ограничиваются многочленами степени не выше трех.

Сумма квадратов уклонений для многочлена (3.36) представляет собой неотрицательную функцию от переменных a_0, \dots, a_k . Обозначим ее через F :

$$F(a_0, \dots, a_k) = \sum_{i=0}^n (a_k x_i^k + \dots + a_1 x_i + a_0 - y_i)^2. \quad (3.37)$$

Требующиеся нам наилучшие коэффициенты многочлена должны давать минимум функции F . Из курса математического анализа известно, что необходимым условием экстремума дифференцируемой функции многих переменных является равенство нулю ее частных производных по всем переменным. Вычислим их и приравняем нулю¹⁾:

$$\begin{cases} F'_{a_0} = 2 \sum (a_k x_i^k + \dots + a_1 x_i + a_0 - y_i) = 0, \\ F'_{a_1} = 2 \sum (a_k x_i^k + \dots + a_1 x_i + a_0 - y_i) x_i = 0, \\ \dots \\ F'_{a_k} = 2 \sum (a_k x_i^k + \dots + a_1 x_i + a_0 - y_i) x_i^k = 0. \end{cases} \quad (3.38)$$

¹⁾ Для упрощения записи в тех случаях, когда это очевидно, пределы суммирования по $i = 0, 1, \dots, n$ будем опускать.

$$\begin{cases} 7,5a + 4b = 4, \\ 16,25a + 7,5b = 9,75. \end{cases}$$

Ее решение $\bar{a} = \frac{36}{35}, \bar{b} = -\frac{13}{14}$.

Следовательно, в данном случае наилучшей линейной эмпирической функцией является $y = \frac{36}{35}x - \frac{13}{14}$. Округлим ее коэффициенты до двух цифр после десятичной запятой: $y = 1,03x - 0,93$. Уклонения последней функции от табличных значений: $v_0 = -0,1, v_1 = 0,115, v_2 = 0,03, v_3 = -0,04$; среднеквадратичное уклонение по всей таблице 0,16. •

Упражнения

3.29. Запишите систему (3.41) и ее коэффициенты (3.40) для эмпирической формулы $y = ax^2 + bx + c$.

3.30. Найдите для табл. 3.5 по методу наименьших квадратов эмпирическую функцию в виде квадратного трехчлена и проверьте, будет ли она лучшим приближением, чем линейная функция из примера 3.6.

3.31. Схема поиска параметров эмпирических формул по методу наименьших квадратов в общем случае такая же, как и при полиномиальном приближении. Напишите систему уравнений, аналогичную (3.38), для произвольной эмпирической формулы $y = p(x, a_1, \dots, a_k)$.

3.32. На основе результата упр. 3.31 напишите системы уравнений для поиска параметров следующих формул: 1) $y = ae^{bx}$, 2) $y = ax^b$,

3) $y = a \ln x + b$, 4) $y = \frac{1}{ax + b}$, 5) $y = \frac{a}{x} + b$.

3.33.* Пусть табличная функция приближается одной из приведенных в упр. 3.32 функций. Покажите, что задачу поиска их наилучших параметров можно свести к поиску коэффициентов некоторой линейной функции. Для этого следует либо ввести новые переменные, либо перейти к новым функциям (построив соответствующим образом новые таблицы на основе исходных).

У к а з а н и е: Равенства 1) и 2) сначала прологарифмируйте в предположениях, что $a > 0$ в 1); $x > 0, a > 0$ в 2). Вместо формулы 4)

рассмотрите формулу: $\frac{1}{y} = ax + b$.

Глава 4. ЧИСЛЕННОЕ ИНТЕГРИРОВАНИЕ И ДИФФЕРЕНЦИРОВАНИЕ

4.1. Задача приближенного вычисления определенных интегралов

В приложениях математики одной из наиболее часто встречающихся задач является вычисление интеграла

$$J = \int_a^b f(x) dx \quad (4.1)$$

от некоторой интегрируемой на отрезке $[a; b]$ функции f . Далее нижний предел интегрирования всегда будем считать меньшим верхнего, поскольку случай $a > b$ сводится к $a < b$ переходом к интегралу с противоположным знаком, а при $a = b$ интеграл равен нулю.

В курсе математического анализа доказывается [3, гл. IX], что когда f непрерывна на $[a; b]$, для нее существует первообразная F ($F'(x) = f(x)$ для всех $x \in [a; b]$), причем

$$\int_a^b f(x) dx = F(b) - F(a). \quad (4.2)$$

Формула (4.2), называемая *формулой Ньютона – Лейбница*, представляет собой точный метод вычисления определенного интеграла. Однако в реальности использовать ее удастся не всегда. Она теряет практический смысл в случае так называемых *неберущихся* интегралов, т. е. таких, у которых подынтегральные функции не имеют выражающихся через элементарные функции первообразных (например,

$\int_a^b e^{x^2} dx$, $\int_a^b \frac{\sin x}{x} dx$). Неэффективно ее применение и тогда, когда первообразная существует, но ее отыскание требует сложной процедуры. Кроме того, на практике часто приходится вычислять интегралы от табличных функций. Если при этом аналитическое выражение f неизвестно, формула Ньютона – Лейбница также бесполезна.

В таких случаях применяются приближенные методы интегрирования, которые можно подразделить на аналитические и численные.

Аналитические методы. Суть аналитических методов состоит в замене подынтегральной функции f на отрезке $[a; b]$ некоторой аналитически заданной функцией p , чью первообразную нетрудно найти: $f(x) \approx p(x)$, $x \in [a; b]$. Тогда имеет место формула

$$\int_a^b f(x) dx \approx \int_a^b p(x) dx. \quad (4.3)$$

Например, если удастся разложить функцию f на отрезке $[a; b]$ в степенной или тригонометрический ряд, то в качестве $p(x)$ можно взять частичную сумму этого ряда. Интеграл от нее легко вычисляется, так как она является либо многочленом, либо линейной комбинацией функций $\sin kx$ и $\cos kx$.

В случае, когда для приближенного равенства $f \approx p$ известна интегрируемая на $[a; b]$ оценочная функция V , можно получить строгую оценку погрешности формулы (4.3):

$$\left| J - \int_a^b p(x) dx \right| \leq \int_a^b V(x) dx. \quad (4.4)$$

Существует еще одна разновидность аналитического метода. Разложение подынтегральной функции в равномерно сходящийся функциональный ряд, каждый член которого легко интегрируется, сводит задачу вычисления интеграла к вычислению суммы числового ряда, а задачу приближенного вычисления интеграла — к вычислению частичной суммы этого ряда [4, гл. XXI].

Представление о практическом применении аналитических методов можно получить, решив упр. 4.2 – 4.4. Далее не будем ими заниматься.

Численные методы позволяют обходиться без аналитических построений. Приближение к интегралу отыскивается по числовому выражению на основе значений подынтегральной функции в конечном множестве точек из отрезка интегрирования. Такой способ вычислений часто называется *механической квадратурой*, соответствующие приближенные формулы называют *формулами численного интегрирования* или *квадратурными формулами*, а используемые при этом аргументы функции — *узлами квадратуры*.

В следующих параграфах будет рассмотрено несколько несложных методов численного интегрирования, в которых используется такая общая схема вычислений: берут число $n \in N$, отрезок $[a; b]$ разбивают на n равных частей одинаковой длины $h = \frac{b-a}{n}$ (шаг разбиения), из каждой части выбирают узлы x_i , вычисляют $f(x_i)$, а затем эти данные подставляют в правую часть квадратурной формулы вида

$$J = \int_a^b f(x) dx \approx h \sum_i A_i f(x_i) = J_n. \quad (4.5)$$

Числовые коэффициенты A_i определяют при выводе каждой конкретной формулы. Суммы состоят из $n + 1$ или n слагаемых. Остаточный член $R_n = J - J_n$ формул (4.5) таков, что $R_n \rightarrow 0$ при $n \rightarrow \infty$. Это означает, что чем меньше шаг разбиения, тем квадратурная формула точнее, и, следовательно, достаточно мелким разбиением отрезка интегрирования можно обеспечить сколь угодно малость погрешности квадратурных формул.

Выражения для остаточного члена и строгая оценка погрешностей квадратурных формул $|J - J_n| \leq V_n$ выводятся в предположении, что функция f имеет на $[a; b]$ достаточное число непрерывных производных.

В тех случаях, когда вычисление и исследование производных доставляют значительные трудности или когда они вообще не могут быть найдены (подынтегральная функция табличная и ее аналитическое выражение неизвестно), применяется приближительная оценка погрешностей $|J - J_n| \approx V_n$ методом двойного пересчета (см. 4.5).

На самом деле, интеграл отыскивается посредством двух приближенных равенств $J \approx J_n \approx \bar{J}_n$, поскольку точные значения функции f обычно неизвестны и поэтому удается найти лишь приближенное значение \bar{J}_n выражения J_n . При этом часто возникает ситуация, когда узлы x_i приходится брать с округлением. Естественно, что погрешности x_i также повлияют на точность $f(x_i)$.

В результате сложения погрешностей квадратурной формулы и вычисления значения J_n образуется погрешность числа \bar{J}_n относительно точного интеграла J . Пусть ν_n — оценка погрешности вычисления \bar{J}_n : $|J_n - \bar{J}_n| \leq \nu_n$. Тогда абсолютная погрешность приближенного интеграла \bar{J}_n относительно J определяется по формуле

$$\Delta_{\bar{J}_n} = V_n + \nu_n^{(1)}. \quad (4.6)$$

Степень влияния двух отмеченных источников погрешностей на результат зависит от подынтегральной функции.

Если f задана аналитически и $f(x_i)$ можно находить с любой точностью, основное внимание обращают на погрешность квадратурной формулы. Действительно, тогда при каждом $n \in \mathbb{N}$ вычисления с достаточно большим количеством цифр обеспечат незначительность ν_n по сравнению с V_n .

Другой крайний случай возникает в ситуации, когда нет ограничений на количество точек разбиения отрезка интегрирования, но $f(x_i)$ находятся с малой точностью, например экспериментальным

¹⁾ Числа V_n и ν_n являются абсолютными погрешностями, определяемыми неоднозначно. Согласно принятой в п. 1.4.4 договоренности, соотношения при их вычислении записывают со знаком точного равенства.

путем. Тогда на первое место выходит задача учета погрешностей вычислений.

Оба источника существенны, если данные для квадратурных формул берутся из таблицы значений подынтегральной функции, ибо тогда точность вычислений ограничена точностью таблицы, а точность формулы — шагом таблицы.

При решении задачи численного интегрирования с заданной точностью $\varepsilon > 0$ необходимо найти \bar{J}_n так, чтобы оказалось $|J - \bar{J}_n| \leq \varepsilon$. Для этого допустимую погрешность обычно делят между двумя источниками с учетом степени их влияния на результат: $\varepsilon = \varepsilon_1 + \varepsilon_2$. Сначала определяют число n , удовлетворяющее неравенству $V_n \leq \varepsilon_1$, что равносильно выбору такого шага разбиения, при котором квадратурная формула обеспечивает точность ε_1 . Затем при найденном n вычисляют значение J_n с точностью до ε_2 .

Ясно, что в случае хорошо вычисляемых функций f надо выбирать ε_2 существенно меньшим, чем ε_1 . Это позволит сократить объем вычислений за счет уменьшения количества разбиений отрезка $[a; b]$.

В заключение обратим внимание на некоторые вопросы, связанные с интегрированием табличных функций.

Пусть f — известная математическая функция с таблицей значений. Если интеграл от f легко вычисляется по формуле Ньютона — Лейбница, то, естественно, таблица не нужна. В противном случае целесообразна квадратура, точность которой, как было сказано выше, зависит от точности и шага таблицы.

Когда вся информация о функции f сосредоточена в таблице, сама постановка задачи интегрирования предполагает использование приближенных методов. Действительно, тогда точный интеграл от f заведомо не может быть найден, а понятие определенного интеграла от таблицы лишено смысла. В то же время на основе таблицы методами, описанными в гл. 3, можно отыскать простое аналитическое приближение p к функции f и вычислить приближенное значение интеграла по формуле (4.3). Если таблица имеет на $[a; b]$ достаточно густую сетку аргументов, хороший результат могут дать численные методы. Чтобы иметь право говорить о точном интеграле, во всех случаях обычно считают, что табличная функция интегрируема на $[a; b]$.

Упражнения

4.1. Выведите формулу (4.4), пользуясь определением оценочной функции и свойствами определенного интеграла.

4.2. Приближенное равенство $e^x \approx 1 + x + \frac{x^2}{2!}$ на отрезке $[0; 1]$ имеет оценочную функцию V : $V(x) = \frac{e}{3!} x^3$ (см. пример 1.10).

1. Вычислите приближенное значение интеграла $\int_0^1 e^x dx$ по формуле (4.3) и найдите его абсолютную погрешность.

2. Найдите интеграл по формуле Ньютона – Лейбница и сравните его с полученным выше приближенным значением.

4.3. Найдите приближенные значения интегралов $\int_{-1}^0 e^{-x^2} dx$ и $\int_{0,1}^1 \frac{\sin x}{x} dx$, взяв три члена разложения подынтегральных функций в степенной ряд.

4.4. Пусть известные значения некоторой функции f заданы табл. 3.5. Вычислите приближенно интеграл $\int_1^3 f(x) dx$, используя полученное в примере 3.6 лучшее приближение к этой функции.

4.2. Формулы прямоугольников

4.2.1. Вывод формул

Интегрируемость функции f на отрезке $[a; b]$ означает, что если этот отрезок разбивать точками $a = x_0 < x_1 < \dots < x_n = b$ на n частей ($n \in \mathbf{N}$), выбирать числа $t_i \in [x_{i-1}; x_i]$ и составлять интегральные суммы

$$S_n = \sum_{i=1}^n f(t_i) \Delta x_i, \quad (\Delta x_i = x_i - x_{i-1}),$$

то будет существовать конечный предел этих сумм при $h \rightarrow 0$ ($h = \max_{i=1, \dots, n} \Delta x_i$ — шаг разбиения), не зависящий ни от способа разбиения отрезка на части, ни от выбора точек t_i из этих частей [3, гл. IX, § 2]. Этот предел и есть определенный интеграл

$$\int_a^b f(x) dx = \lim_{h \rightarrow 0} S_n. \quad (4.7)$$

Поскольку способ разбиения отрезка может быть любым, в дальнейшем будем разбивать $[a; b]$ на равные части. В этом случае

$$h = \frac{b-a}{n} = \Delta x, \quad (i = 1, 2, \dots, n),$$

а условие $h \rightarrow 0$ равносильно условию $n \rightarrow \infty$. Следовательно, при достаточно больших n можно положить:

$$\int_a^b f(x) dx \approx h \sum_{i=1}^n f(t_i).$$

Конкретизируем правило выбора t_i . Если взять их равными левым концам отрезков $[x_{i-1}; x_i]$, $y_{i-1} = f(x_{i-1})$, то получим

$$\int_a^b f(x) dx \approx h(y_0 + y_1 + \dots + y_{n-1}) \stackrel{\text{def}}{=} J_n^{(1)}. \quad (4.8)$$

Приняв $t_i = x_i$ — правые концы отрезков $[x_{i-1}; x_i]$, $y_i = f(x_i)$, будем иметь

$$\int_a^b f(x) dx \approx h(y_1 + y_2 + \dots + y_n) \stackrel{\text{def}}{=} J_n^{(2)}. \quad (4.9)$$

Квадратурная формула (4.8) называется *формулой прямоугольников с левыми ординатами*, а формула (4.9) — *формулой прямоугольников с правыми ординатами*. Взяв в качестве t_i середины частичных отрезков, получим формулу прямоугольников с *центральными ординатами* (упр. 4.7).

Эти названия объясняются геометрическим смыслом формул. Как известно, определенный интеграл от неотрицательной интегрируемой функции f на отрезке $[a; b]$ равен площади соответствующей криволинейной трапеции. Если криволинейную трапецию заменить ступенчатой фигурой D , составленной из прямоугольников с основаниями $[x_{i-1}; x_i]$ и с высотами, равными ординатам точек (x_{i-1}, y_{i-1}) графика $y = f(x)$ ($i = 1, 2, \dots, n$), то формула (4.8) выражает замену площади криволинейной трапеции на площадь фигуры D . На рис. 4.1 взято $n = 2$, фигура D с «левыми» ординатами заштрихована.

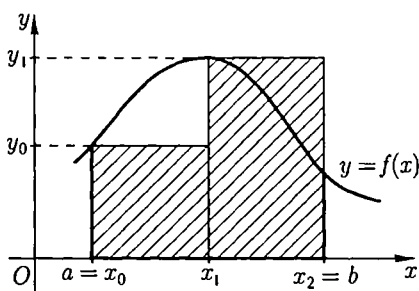


Рис. 4.1

Аналогичный смысл имеет и формула (4.9).

Упражнения

4.5. Приведите геометрическую иллюстрацию квадратурной формулы (4.9).

4.6. Найдите приближенные значения интеграла $\int_0^2 x^2 dx$ по формулам (4.8) и (4.9) при $n = 4$ и сравните их с точным значением интеграла. Проиллюстрируйте графически вычисления по обеим формулам.

4.7. Выведите формулу прямоугольников с *центральными ординатами*

$$J = \int_a^b f(x) dx \approx h \sum_{i=1}^n f(\bar{x}_i) \stackrel{\text{def}}{=} J_n^{(3)}, \quad (4.10)$$

где $\bar{x}_i = (x_{i-1} + x_i)/2$ — середины отрезков разбиения. Выясните ее геометрический смысл.

4.8. Выполните упр. 4.6 с квадратурной формулой (4.10).

4.2.2. Оценка погрешностей

Здесь и далее в данной главе нам понадобятся следующие теоремы.

Теорема 4.1 (вторая теорема Больцано – Коши). Пусть функция f непрерывна на отрезке $[a; b]$, а числа m и M — ее наименьшее и наибольшее значения на $[a; b]$. Тогда для любого числа C , заключенного между m и M , найдется точка $c \in [a; b]$ такая, что $f(c) = C$ [3, гл. IV, § 5].

Теорема 4.2 (аддитивность интеграла) [3, гл. IX, § 4]. Если $a = x_0 \leq x_1 \leq x_2 \leq \dots \leq x_n = b$, то

$$\int_a^b f(x) dx = \sum_{i=1}^n \int_{x_{i-1}}^{x_i} f(x) dx.$$

Теорема 4.3 (обобщенная теорема о среднем значении интеграла [21, гл. IX, § 2]). Пусть:

1) функция f непрерывна на отрезке $[a; b]$, а функция g интегрируема на этом отрезке;

2) функция g не меняет знак на всем отрезке $[a; b]$.

Тогда существует точка $c \in [a; b]$ такая, что

$$\int_a^b f(x)g(x)dx = f(c) \int_a^b g(x)dx.$$

Погрешности формул (4.8) и (4.9) оцениваются одинаково, поэтому далее правые части этих формул $J_n^{(i)}$ и их остаточные члены $R_n^{(i)}$ ($i = 1, 2$) для простоты будем записывать без верхних индексов.

Теорема 4.4. Если подынтегральная функция f имеет на $[a; b]$ непрерывную производную f' , то оценка погрешностей формул (4.8) и (4.9) дается неравенством

$$|J - J_n| = |R_n| \leq M_1 \frac{(b-a)^2}{2n}, \quad (4.11)$$

где $M_1 = \max_{[a; b]} |f'(x)|^1$.

○ Доказательство проведем для формулы (4.8). Пусть отрезок $[a; b]$ разбит на n равных частей $[x_0; x_1], \dots, [x_{n-1}; x_n]$ одинаковой длины $h = \frac{b-a}{n}$.

¹⁾ На практике за M_1 обычно принимают число, удовлетворяющее неравенству $|f'(x)| \leq M_1$ для всех $x \in [a; b]$. Подобное замечание справедливо и далее при аналогичных оценках.

Возьмем любой отрезок $[x_{i-1}; x_i]$ ($i = 1, 2, \dots, n$). Для всякого x из него найдется зависящее от x число $c_i^* \in [x_{i-1}; x]$ такое, что $f(x) = f(x_{i-1}) + f'(c_i^*)(x - x_{i-1})$ (теорема Лагранжа). Тогда

$$\int_{x_{i-1}}^{x_i} f(x) dx = hf(x_{i-1}) + \int_{x_{i-1}}^{x_i} f'(c_i^*)(x - x_{i-1}) dx.$$

Функция f' непрерывна, а функция $g: g(x) = x - x_{i-1}$ интегрируема и неотрицательна на $[x_{i-1}; x_i]$. Следовательно, к интегралу в правой части полученного соотношения можно применить теорему 4.3 с некоторым числом $c_i \in [x_{i-1}; x_i]$:

$$\int_{x_{i-1}}^{x_i} f(x) dx = hf(x_{i-1}) + f'(c_i) \int_{x_{i-1}}^{x_i} (x - x_{i-1}) dx = hf(x_{i-1}) + f'(c_i) \frac{h^2}{2}.$$

Сложив левые и правые части при $i = 1, 2, \dots, n$ и воспользовавшись теоремой 4.2, получим

$$J = h(y_0 + \dots + y_{n-1}) + \frac{(b-a)^2}{2n} \cdot \frac{f'(c_1) + \dots + f'(c_n)}{n}.$$

Первое слагаемое справа есть J_n , тогда второе — R_n . Число

$$C = \frac{f'(c_1) + \dots + f'(c_n)}{n}$$

представляет собой среднее арифметическое значений функции f' , находящееся между ее наименьшим и наибольшим значениями на $[a; b]$, и потому равно $f'(c)$ для некоторого $c \in [a; b]$ (теорема 4.1). Следовательно,

$$R_n = \frac{(b-a)^2}{2n} f'(c) \quad (c \in [a; b]). \quad (4.12)$$

Для получения оценки (4.11) осталось учесть определение M_1 . •

Таким образом, погрешность первого из приближенных равенств $J \approx J_n \approx \bar{J}_n$, порождаемая формулами (4.8) и (4.9), оценивается числом $V_n = M_1 \frac{(b-a)^2}{2n}$. Общее правило (4.6) вычисления абсолютной погрешности приближенных интегралов примет вид

$$\Delta J_n = M_1 \frac{(b-a)^2}{2n} + \nu_n, \quad (4.13)$$

где ν_n — оценка точности вычисления значения J_n .

Как видно из выражения для V_n , оценка погрешностей формул (4.8) и (4.9) зависит от подынтегральной функции, от величины отрезка интегрирования и количества n частей его разбиения.

ния. Поскольку для каждого конкретного интеграла числа M_1 и $(b - a)^2$ постоянны, можно сказать, что погрешность обратно пропорциональна n .

Точность этих формул невысока, в чем можно убедиться на основе примера 4.1. Формула прямоугольников с центральными ординатами (4.10) является более точной: ее погрешность обратно пропорциональна n^2 (упр. 4.12). При уменьшении шага разбиения отрезка $[a; b]$ вдвое погрешность формул (4.8) и (4.9) уменьшается примерно в два раза, а погрешность формулы (4.10) — в четыре раза.

В следующем примере проведем подробные исследования вычисления приближенного интеграла при заданном n и оценки его погрешности, а также наметим план вычисления интеграла с заданной точностью. Аналогичные выкладки могут иметь место в случае других интегралов и квадратурных формул.

Пример 4.1. Рассмотрим интеграл $J = \int_0^{0.5} e^x dx$.

1. Определим, с какой точностью можно получить приближение к этому интегралу по формуле (4.8) при $n = 5$, а затем найдем его.

В данном случае $f(x) = e^x$, $h = 0,1$, узлы квадратуры: $x_0 = 0$, $x_1 = 0,1$, ..., $x_4 = 0,4$. Обозначив $y_i = f(x_i)$, будем иметь

$$J \approx J_5 = 0,1 \cdot (y_0 + y_1 + y_2 + y_3 + y_4). \quad (*)$$

Для оценки погрешности формулы (*) с помощью (4.11) найдем число M_1 . Поскольку $|f'(x)| \leq e^{0,5} \leq 1,65$ для $x \in [0; 0,5]$, можно взять

$$M_1 = 1,65. \text{ Тогда получим } V_5 = 1,65 \cdot \frac{0,5^2}{10} = 0,042.$$

Следовательно, формула (*) позволяет находить приближенный интеграл с верными цифрами до первого разряда после десятичной запятой включительно. Выясним, в каких разрядах следует оставлять цифры у чисел y_i , чтобы сохранить эту степень точности. Учтем при этом, что узлы x_i здесь точные и не влияют на точность $f(x_i)$.

Обозначим через S результат вычисления суммы $y_0 + \dots + y_4$, тогда $\bar{J}_5 = 0,1 \cdot S$. Округляя y_i до двух верных цифр после запятой, гарантируем $\Delta_{y_i} = 0,005$ ($i = 0, \dots, 4$) и, значит, $\Delta_S = 5 \cdot 0,005 = 0,025$. В выражении \bar{J}_5 множитель $0,1$ точный, поэтому, согласно формуле (1.25), оценка погрешности вычисления \bar{J}_5 равна $v_5 = 0,1 \cdot \Delta_S = 0,0025$. Тогда полная погрешность \bar{J}_5 оценится числом $\Delta_{\bar{J}_5} = V_5 + v_5 = 0,045^1$.

Отсюда следует, что выбранная точность вычисления $f(x_i)$ достаточна и что в этом случае вычислительная погрешность мало влияет на определяемую квадратурной формулой точность результата.

¹⁾ Проведенные выкладки по оценке погрешности вычисления \bar{J}_n носят общий характер. Когда число n небольшое, можно найти погрешность округления каждого y_i и получить более точную оценку.

В итоге получим

$$J \approx 0,1 \cdot (1,00 + 1,10 + 1,22 + 1,35 + 1,49) \approx 0,62$$

с одной верной значащей цифрой. Сравнение числа 0,62 с найденным по формуле Ньютона – Лейбница приближением $0,649 \approx J$, имеющим три верные значащие цифры, говорит о справедливости такого вывода.

2. Приведем возможный план вычислений данного интеграла по формуле (4.8) с точностью до $\varepsilon = 0,005$.

Поскольку значения f можно находить с любой точностью, данную абсолютную погрешность разделим так: $\varepsilon_1 = 0,004$ – на погрешность формулы, $\varepsilon_2 = 0,001$ – на погрешность вычисления \bar{J}_n .

Сначала находим наименьшее натуральное число n , удовлетворяющее неравенству

$$V_n = 1,65 \cdot \frac{0,5^2}{2n} = \frac{0,4125}{2n} \leq \varepsilon_1.$$

Это будет $n = 52$ ¹⁾. Далее следует разбить отрезок $[0; 0,5]$ на 52 части с шагом $h = \frac{0,5}{52}$ и найти $y_i = f(x_i)$ ($i = 0, 1, \dots, 51$) так, чтобы

вычислительная погрешность оказалась не больше ε_2 .

Обозначим $S = y_0 + \dots + y_{51}$ и предположим, что все y_i имеют три верные цифры после десятичной запятой. Тогда $\Delta_{y_i} = 0,0005$, поэтому $\Delta_S = 52 \cdot 0,0005 = 0,026$ и $\nu_{52} = h \cdot 0,026 = 0,00025 < \varepsilon_2$. Значит, трех верных цифр после запятой достаточно. Нетрудно проверить, что две такие цифры могут не обеспечить требуемую точность.

Следует учесть еще влияние погрешностей x_i на $f(x_i)$. При делении отрезка $[0; 0,5]$ на 52 части шаг разбиения h , а значит и x_i , придется округлять. Узнаем, какая точность должна быть у узлов x_i , чтобы $f(x_i)$ имели три верные цифры после запятой. Это задача оценки погрешности аргумента по известной оценке погрешности функции (упр. 1.30).

В силу (1.22), для функции $y = e^x$ имеем $\Delta_y = e^x \cdot \Delta_x$. Наибольшее значение абсолютной погрешности Δ_y при постоянном Δ_x , равное $e^{0,5} \cdot \Delta_x \approx 1,65 \cdot \Delta_x$, должно быть не больше 0,0005. Отсюда получим требование на точность x_i ($i = 0, 1, \dots, 51$): $\Delta_{x_i} \leq 0,00031$, которое заведомо выполнится при округлении x_i до четырех верных цифр после запятой. •

З а м е ч а н и е. Как видим, проблема установления количества верных цифр у чисел x_i и $f(x_i)$ не проста. Ею можно не заниматься, если имеется возможность находить узлы и значения подынтегральной функции с запасными верными значащими цифрами.

¹⁾ Оно велико вследствие малой точности формулы (4.8).

Упражнения

4.9. Проведите доказательство теоремы 4.4 для формулы (4.9).

4.10. Убедитесь с помощью соотношения (4.11) и графически, что в случае $f(x) \equiv C = \text{const}$ в формулах (4.8) и (4.9) имеют место точные равенства.

4.11. 1. Оцените погрешность полученных в упр. 4.6 приближенных значений интеграла $\int_0^2 x^2 dx$.

2. Определите, на сколько частей надо разделить отрезок $[1; 2]$ для того, чтобы найти этот интеграл по формулам (4.8) и (4.9) с точностью до $\epsilon = 0,01$, считая при этом вычислительные погрешности несущественными.

4.12.* В предположении, что f'' непрерывна на $[a; b]$, выведите для формулы прямоугольников с центральными ординатами соотношения:

$$R_n^{(3)} = \frac{(b-a)^3}{24n^2} f''(c) \quad (c \in [a; b]) \text{ и } |J - J_n^{(3)}| \leq M_2 \frac{(b-a)^3}{24n^2}, \quad (4.14)$$

где $M_2 = \max_{[a; b]} |f''(x)|$.

Указание. В соответствии с формулой Тейлора на каждом из отрезков $[x_{i-1}; x_i]$ можно взять

$$f(x) = f(\bar{x}_i) + (x - \bar{x}_i)f'(\bar{x}_i) + \frac{1}{2}(x - \bar{x}_i)^2 f''(c_i^*),$$

где число c_i^* лежит между x и \bar{x}_i .

4.13. 1. Покажите с помощью соотношений (4.14) и графически, что формула (4.10) точна в случае линейной функции $f: f(x) = ax + b$.

2. Для проверки вывода найдите интеграл $\int_1^2 (3x - 1) dx$ по формуле Ньютона – Лейбница и по формуле (4.10) при $n = 2$.

3. Вычислите данный интеграл по формуле (4.8) при $n = 2$ и убедитесь, что результат не будет точным.

4.14. Определите план приближенного вычисления интеграла из примера 4.1 с точностью до $\epsilon = 0,005$ по формуле прямоугольников с центральными ординатами.

4.3. Формула трапеций

4.3.1. Вывод формулы

В этом и следующем параграфах для получения квадратурных формул подынтегральная функция на частичных отрезках заменяется соответствующими интерполяционными многочленами.

Разобьем отрезок $[a; b]$ точками $a = x_0 < x_1 < \dots < x_n = b$ на n равных частей одинаковой длины $h = \frac{b-a}{n}$ и найдем $y_i = f(x_i)$ ($i = 0, 1, \dots, n$). На каждом из отрезков $[x_{i-1}; x_i]$ функцию f заменим по формуле линейного интерполирования (3.26).

На отрезке $[x_0; x_1]$

$$f(x) \approx P_1(x) = y_0 + t\Delta y_0,$$

где $t = \frac{x-x_0}{h}$, $\Delta y_0 = y_1 - y_0$. Тогда

$$\int_{x_0}^{x_1} f(x) dx \approx \int_{x_0}^{x_1} (y_0 + t\Delta y_0) dx.$$

Найдем правый интеграл переходом к переменной t :

$$\int_{x_0}^{x_1} f(x) dx \approx h \int_0^1 (y_0 + t\Delta y_0) dt = h \left(y_0 + \frac{\Delta y_0}{2} \right) = h \frac{y_0 + y_1}{2}. \quad (4.15)$$

На остальных отрезках аналогичные выкладки дают

$$\int_{x_1}^{x_2} f(x) dx \approx h \frac{y_1 + y_2}{2}, \dots, \int_{x_{n-1}}^{x_n} f(x) dx \approx h \frac{y_{n-1} + y_n}{2}. \quad (4.16)$$

Складываем почленно приближенные равенства (4.15), (4.16) и в силу аддитивности интеграла получим *формулу трапеций*:

$$\int_a^b f(x) dx \approx h \left(\frac{y_0 + y_n}{2} + y_1 + \dots + y_{n-1} \right) \stackrel{\text{def}}{=} J_n^{(\tau)}. \quad (4.17)$$

На рис. 4.2 показан геометрический смысл этой формулы при $n = 2$. Линейная интерполяция приводит к замене графика функции f ломаной, соединяющей точки (x_0, y_0) , (x_1, y_1) , ..., (x_n, y_n) этого графика. Затем вместо криволинейной трапеции рассматривают фигуру D , составленную из прямолинейных трапеций с основаниями y_i и y_{i+1} и высотой h . Правые части соотношений (4.15) и (4.16) равны площадям этих прямолинейных трапеций, а (4.17) означает замену площади криволинейной трапеции площадью фигуры D .

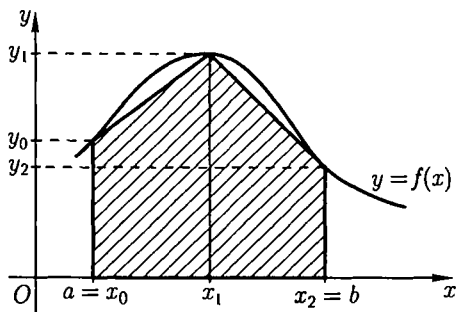


Рис. 4.2

Упражнения

4.15. Выполните задания упр. 4.6 с формулой трапеций.

x	1	1,5	2	2,5	3
$f(x)$	0,02	1,23	2,98	5,23	8,01

4.16. Найдите по формуле трапеций при $n = 2$ и $n = 4$ интеграл $\int_1^3 f(x) dx$ от функции с данной таблицей значений. Проведите геометрическую иллюстрацию.

4.3.2. Оценка погрешностей

Теорема 4.5. Если вторая производная функции f непрерывна на $[a; b]$, для квадратурной формулы (4.17) имеет место неравенство

$$|J - J_n^{(r)}| = |R_n^{(r)}| \leq M_2 \frac{(b-a)^3}{12n^2}, \quad (4.18)$$

где $M_2 = \max_{[a; b]} |f''(x)|$.

○ Пусть $[x_{i-1}; x_i]$ — произвольный отрезок из разбиения $[a; b]$ на n равных частей с шагом $h = \frac{b-a}{n}$ ($i = 1, 2, \dots, n$). Пользуясь формулой остаточного члена из теоремы 3.2 для интерполяционного многочлена первой степени, построенного на узлах x_{i-1} и x_i , получим

$$f(x) = P_1(x) + \frac{f''(c_i^*)}{2} (x - x_{i-1})(x - x_i), \quad (4.19)$$

где c_i^* — зависящее от x число между x_{i-1} и x_i . Проинтегрируем левую и правую части (4.19) на отрезке $[x_{i-1}; x_i]$. При этом учтем следующее: во-первых, интеграл от $P_1(x)$ равен $h \frac{y_{i-1} + y_i}{2}$ (соотношения (4.15) и (4.16)); во-вторых, к интегралу от второго слагаемого правой части можно применить обобщенную теорему о среднем (f'' непрерывна, а выражение $(x - x_{i-1})(x - x_i)$ не меняет знак); в-третьих,

$$\int_{x_{i-1}}^{x_i} (x - x_{i-1})(x - x_i) dx = -\frac{h^3}{6}.$$

Следовательно,

$$\int_{x_{i-1}}^{x_i} f(x) dx = h \frac{y_{i-1} + y_i}{2} - \frac{f''(c_i)}{12} h^3 \quad (c_i \in [x_{i-1}; x_i]).$$

Просуммировав левые и правые части полученных равенств при $i = 1, 2, \dots, n$, по аналогии с окончанием доказательства теоремы 4.4 сначала получим

$$R_n^{(r)} = -\frac{(b-a)^3}{12n^2} f''(c) \quad (c \in [a; b]),$$

а затем, с учетом определения M_2 , и требуемую оценку (4.18). •

Итак, формула трапеций порождает погрешности, оцениваемые числом $V_n = M_2 \frac{(b-a)^3}{12n^2}$. Ее точность практически такая же, как и у формулы прямоугольников с центральными ординатами.

Вычислив значение \bar{J}_n выражения $J_n^{(r)}$ с точностью до ν_n , находим оценку его полной погрешности относительно J :

$$\Delta_{\bar{J}_n} = M_2 \frac{(b-a)^3}{12n^2} + \nu_n. \quad (4.20)$$

Пример 4.2. Найдём приближенное значение интеграла $J = \int_0^{0,5} e^x dx$ из примера 4.1 по формуле трапеций при $n = 5$ и определим его верные цифры. Шаг разбиения и узлы квадратуры остаются прежними.

Оценим погрешность квадратурной формулы

$$J \approx J_5 = 0,1 \cdot \left(\frac{y_0 + y_5}{2} + y_1 + y_2 + y_3 + y_4 \right). \quad (**)$$

Поскольку $f''(x) = e^x$, можно взять $M_2 = 1,65$ (см. пример 4.1). Тогда $V_5 = 1,65 \frac{0,5^3}{12 \cdot 5^2} = 0,0007$. Следовательно, формула (**) обеспечивает верные цифры результата во всех разрядах слева от разряда тысячных долей.

Чтобы приближенный интеграл \bar{J}_5 имел верные цифры в этих же разрядах, должно быть $\Delta_{\bar{J}_5} \leq 0,005$. Значит, правую часть формулы (**) надо вычислить с точностью до 0,005 — $V_5 = 0,0043$. Покажем, что для этого достаточно округлять y_i до двух верных цифр после десятичной запятой: $\Delta_{y_i} = 0,005$ ($i = 0, 1, \dots, 5$). Действительно, обозначив через S выражение в скобках из (**) и $\bar{J}_5 = 0,1 \cdot S$, будем иметь

$$\nu_5 = 0,1 \cdot \Delta_S = 0,1 \cdot \left(\frac{0,01}{2} + 4 \cdot 0,005 \right) = 0,0025 < 0,0043.$$

Подставив в формулу (**) значения $f(x_i)$, получим

$$J \approx 0,1 \cdot \left(\frac{1,00 + 1,65}{2} + 1,10 + 1,22 + 1,35 + 1,49 \right) = 0,6485$$

с верными цифрами 6 и 4. Поскольку $J \approx 0,64872$ с пятью верными значащими цифрами, на самом деле точность результата выше. •

Упражнения

4.17. Убедитесь, что формула трапеций, как и формула прямоугольников с центральными ординатами, дает точное значение интеграла от $f(x) = ax + b$ (см. упр. 4.13).

4.18. По формуле трапеций вычисляется интеграл $\int_0^{\pi} \sin x dx$.

1. Оцените погрешность формулы при: а) $n = 5$; б) $n = 10$.

2. Выясните, до каких разрядов необходимо округлять значения x_i и $\sin x_i$, чтобы приближенные интегралы \bar{J}_5, \bar{J}_{10} сохранили обеспечиваемые формулой верные разряды.

4.19. Определите, на сколько частей следует разбить отрезок $[-1; 1]$ для того, чтобы при вычислении интеграла $\int_{-1}^1 (x + \ln(x + 2)) dx$ формула трапеций обеспечила точность до $\epsilon = 0,001$.

4.4. Формула Симпсона

От квадратурной формулы следует ожидать большей точности, если для приближения подынтегральной функции f на частичных отрезках использовать квадратичное интерполирование.

Снова разобьем отрезок $[a; b]$ на n равных частей точками $a = x_0 < x_1 < \dots < x_n = b$ с общей длиной $h = \frac{b-a}{n}$ и обозначим $y_i = f(x_i)$ ($i = 0, 1, \dots, n$), но теперь возьмем *четное* число n . Тогда можно рассматривать «сдвоенные» отрезки $[x_0; x_2], [x_2; x_4], \dots, [x_{n-2}; x_n]$ с тремя известными узлами и на них функцию f заменять интерполяционными многочленами Ньютона второй степени (на каждом отрезке свой многочлен).

Для $x \in [x_0; x_2]$ имеем

$$f(x) \approx y_0 + t\Delta y_0 + \frac{t(t-1)}{2} \Delta^2 y_0, \text{ где } t = \frac{x-x_0}{h}. \quad (4.21)$$

Вычислим интеграл от правой части на отрезке $[x_0; x_2]$ с заменой переменной $x = x_0 + ht$:

$$\begin{aligned} & \int_{x_0}^{x_2} \left(y_0 + t\Delta y_0 + \frac{t(t-1)}{2} \Delta^2 y_0 \right) dx = \\ & = h \int_0^2 \left(y_0 + t\Delta y_0 + \frac{t^2-t}{2} \Delta^2 y_0 \right) dt = h \left(2y_0 + 2\Delta y_0 + \frac{1}{3} \Delta^2 y_0 \right) = \\ & = \frac{h}{3} (6y_0 + 6(y_1 - y_0) + (y_0 - 2y_1 + y_2)) = \frac{h}{3} (y_0 + 4y_1 + y_2). \end{aligned}$$

Следовательно,

$$\int_{x_0}^{x_2} f(x) dx \approx \frac{h}{3} (y_0 + 4y_1 + y_2). \quad (4.22)$$

Аналогично на остальных отрезках:

$$\int_{x_i}^{x_{i+2}} f(x) dx \approx \frac{h}{3} (y_i + 4y_{i+1} + y_{i+2}) \quad (i = 2, 4, \dots, n-2). \quad (4.23)$$

Результатом суммирования всех полученных приближенных равенств и будет формула Симпсона:

$$\int_a^b f(x) dx \approx \frac{h}{3} (y_0 + y_n + 4(y_1 + y_3 + \dots + y_{n-1}) + 2(y_2 + y_4 + \dots + y_{n-2})), \quad (4.24)$$

правую часть которой обозначим $J_n^{(C)}$.

Формула Симпсона выглядит более громоздкой по сравнению с формулами прямоугольников и трапеций, но она значительно точнее их и может привести к требуемому результату при меньших n .

Теорема 4.6. Если производная четвертого порядка $f^{(4)}$ подынтегральной функции непрерывна на $[a; b]$, то

$$R_n^{(C)} = -\frac{(b-a)^5}{180n^4} f^{(4)}(d) \quad (d \in [a; b]), \quad (4.25)$$

и поэтому

$$|J - J_n^{(C)}| \leq M_4 \frac{(b-a)^5}{180n^4}, \quad \text{где } M_4 = \max_{[a; b]} |f^{(4)}(x)|. \quad (4.26)$$

Доказательство теоремы 4.6 можно построить на тех же идеях, что были использованы при доказательстве теоремы 4.5, но оно связано с более сложными выкладками, поэтому его опускаем¹⁾.

Как видно по оценке $V_n = M_4 \frac{(b-a)^5}{180n^4}$, точность формулы Симпсона на два порядка выше точности формулы трапеций и формулы прямоугольников с центральными ординатами. Она является одной из самых употребительных в практике вычисления определенных интегралов.

Пример на использование формулы Симпсона приведен в 4.5 (см. пример 4.3).

¹⁾ Упомянутое доказательство приведено в [6, гл. XVI, § 4, 7], [9, гл. 11, § 2]. По другой схеме (с привлечением интерполяционных многочленов Эрмита) теорема доказана в [23, п. 327].

Упражнения

4.20. Формула Симпсона называется также *формулой парабол*. Объясните данное название, приведя геометрическую иллюстрацию формулы.

4.21. 1. Процесс вывода формулы Симпсона убеждает, что она дает точное значение интегралу от квадратного многочлена. Докажите, что на самом деле формула Симпсона точна для всех многочленов $P_m(x)$, $m \leq 3$.

2. Проверьте справедливость полученного вывода на примере, вычислив интеграл $\int_0^1 x^3 dx$ по формуле Симпсона при $n = 2$ и по формуле Ньютона – Лейбница.

4.22. Определите, на сколько частей надо разделить отрезок интегрирования, чтобы найти интеграл $\int_1^2 \frac{dx}{x}$ с точностью до $\varepsilon = 0,0005$ по формулам трапеций и Симпсона (вычислительные погрешности не учитываем).

4.23. Выполните задания упр. 4.18 (при $n = 10$) и 4.19 с формулой Симпсона.

4.5. Учет погрешностей квадратурных формул методом двойного пересчета

Выше были получены формулы строгой оценки погрешностей квадратурных формул. Все они пригодны лишь в случае аналитически заданной подынтегральной функции и требуют нахождения максимума модуля производных, что подчас представляет собой далеко не простую задачу.

Существует не связанный с вычислением производных способ ориентировочной оценки погрешностей, применимый и для интегралов от табличных функций. Это так называемый *метод двойного пересчета* или *метод Рунге*.

Рассмотрим его сначала для формул прямоугольников (4.8) и (4.9).

Выбираем некоторое натуральное число n и проводим вычисления по одной из этих формул дважды: при разбиениях отрезка на n и на $2n$ частей (с шагом $h_n = \frac{b-a}{n}$ и $h_{2n} = \frac{b-a}{2n}$ соответственно). Обозначим полученные результаты через J_n и J_{2n} ¹⁾. Ясно, что лучшим

¹⁾ На самом деле это числа \bar{J}_n и \bar{J}_{2n} , но, поскольку здесь речь идет только о погрешностях, вызванных квадратурными формулами, черточки не пишем.

приближением будет J_{2n} , которое и считаем в дальнейшем приближенным значением интеграла.

Для остаточного члена R_{2n} , равного погрешности числа J_{2n} , справедливо равенство (4.12) с некоторым числом $c^* \in [a; b]$: $R_{2n} = \frac{(b-a)^2}{4n} f'(c^*)$. Пусть значения производной f' мало изменяются на отрезке $[a; b]$. Тогда

$$R_n = \frac{(b-a)^2}{2n} f'(c) \approx \frac{(b-a)^2}{2n} f'(c^*) = 2R_{2n}. \quad (4.27)$$

С учетом (4.27) и очевидных равенств $J = J_n + R_n = J_{2n} + R_{2n}$, получим сначала

$$R_{2n} = J - J_{2n} = J_n + R_n - J_{2n} \approx J_n - J_{2n} + 2R_{2n},$$

а затем

$$|R_{2n}| = |J - J_{2n}| \approx |J_n - J_{2n}|. \quad (4.28)$$

Таким образом, соотношение (4.28) дает приближительную оценку погрешности числа J_{2n} , полученного по формуле прямоугольников с левыми (или правыми) ординатами.

Аналогичные оценки имеют место в случае других квадратурных формул. Для формулы трапеций и формулы прямоугольников с центральными ординатами

$$|J - J_{2n}| \approx \frac{|J_n - J_{2n}|}{3}, \quad (4.29)$$

для формулы Симпсона

$$|J - J_{2n}| \approx \frac{|J_n - J_{2n}|}{15}. \quad (4.30)$$

Они тем точнее, чем менее значительно изменяются на $[a; b]$ вторая и соответственно четвертая производные подынтегральной функции.

Хотя соотношения (4.28) – (4.30) требуют двойного счета по квадратурным формулам, они практически удобны, особенно при компьютерных вычислениях.

Пример 4.3. Найдем по формуле Симпсона приближения J_2 и J_4 к интегралу $J = \int_1^2 \ln x dx$, а затем сравним оценки погрешности

формулы для J_4 , полученные методом Рунге и с помощью строгой оценки (4.26). Ввиду большой точности формулы вычисления будем вести с пятью верными цифрами после десятичной запятой.

При $n = 2$ шаг разбиения $h = 0,5$, $x_0 = 1$, $x_1 = 0,5$, $x_2 = 2$, поэтому

$$J_2 = \frac{h}{3} (y_0 + y_2 + 4y_1) \approx \frac{0,5}{3} (0 + 0,69315 + 4 \cdot 0,40547) \approx 0,38584.$$

Разбивая отрезок $[1; 2]$ на 4 части с шагом $h = 0,25$, получим

$$J_4 = \frac{h}{3} (y_0 + y_4 + 4(y_1 + y_3) + 2y_2) \approx 0,38626.$$

Взяв $J \approx J_4 \approx 0,38626$, оценим погрешность по формуле (4.30):

$$|J - 0,38626| \approx \frac{|J_2 - J_4|}{15} \approx 0,000028.$$

Теперь вычислим $M_4 = \max_{[0;1]} |(\ln x)^{(4)}| = 6$ и по (4.26) найдем строгую оценку погрешности:

$$|J - 0,38626| \leq \frac{6}{180 \cdot 4^4} \leq 0,00013.$$

Полученные оценки существенно отличаются. Оценка по правилу Рунге показывает, что результат содержит четыре верные значащие цифры (без учета вычислительных погрешностей), а по строгой оценке получаются только три такие цифры. Находим интеграл по формуле Ньютона—Лейбница с шестью верными значащими цифрами — $J \approx 0,386294$ — и видим, что строгая оценка, как и можно было ожидать, оказалась завышенной. •

Для применения метода Рунге в случае, когда данные для квадратурных формул берутся из таблицы значений подынтегральной функции, необходимо, чтобы таблица была с постоянным шагом и имела на отрезке $[a; b]$ сетку аргументов, достаточную для выбора узлов квадратуры, соответствующих шагам h_n и h_{2n} . Чтобы выяснить, что означает при этом условие «производная k -го порядка подынтегральной функции мало изменяется на отрезке интегрирования», предположим, что таблица составлена для некоторой достаточно точное число раз дифференцируемой на $[a; b]$ функции f , и установим связь между производными $f^{(k)}$ и конечными разностями $\Delta^k f$.

Аналогично тому, как это делалось в п. 3.6.1, нетрудно убедиться, что при любом $k = 1, 2, \dots$ для x таких, что $x, x + kh \in [a; b]$, имеют место приближенные равенства

$$\Delta^k f(x) \approx f^{(k)}(x)h^k, \quad \text{или} \quad f^{(k)}(x) \approx \frac{\Delta^k f(x)}{h^k}. \quad (4.31)$$

Чем меньше h , тем приближенные равенства точнее. В узлах x_0, x_1, \dots, x_n таблицы получим

$$\left. \begin{aligned} f'(x_i) &\approx \frac{\Delta y_i}{h} \quad (i = 0, 1, \dots, n-1), \\ f''(x_i) &\approx \frac{\Delta^2 y_i}{h^2} \quad (i = 0, 1, \dots, n-2), \\ &\dots \dots \dots \\ f^{(n)}(x_0) &\approx \frac{\Delta^n y_i}{h^n}. \end{aligned} \right\} \quad (4.32)$$

Таким образом, производная k -го порядка ($k \leq n$) в точках x_0, x_1, \dots, x_{n-k} приближенно равна конечным разностям этого же порядка, поделенным на h^k . Так как $h = \text{const}$, отсюда следует, что *если табличные разности некоторого порядка практически постоянны, можно считать, что производная функции f этого же порядка мало изменяется на $[a; b]$.*

З а м е ч а н и е. Правая формула в (4.31) и формулы (4.32) дают способ приближенного вычисления производных табличных функций. Однако их точность невысока. В следующем параграфе будут получены более точные формулы *численного дифференцирования*.

Упражнения

4.24. Выведите приближенные равенства (4.29) и (4.30) для соответствующих квадратурных формул.

4.25. 1. Убедитесь, что в упр. 4.16 для оценки погрешностей приближений интеграла $\int_3^3 f(x) dx$ можно применять метод Рунге.

2. Вычислите приближения J_2 и J_4 этого интеграла по формуле трапеций и оцените их погрешности.

4.26. Методом математической индукции докажите, что соотношения (4.31) имеют место при любом натуральном k , если только $x, x + kh \in [a; b]$ и на $[a; b]$ существует конечная производная $f^{(k)}$.

4.6. Численное дифференцирование

4.6.1. Постановка задачи

Рассматриваемые в курсе математического анализа задачи нахождения производных функции одной переменной обычно доставляют меньше затруднений, нежели вычисление интегралов. Конечный набор формул и правил дифференцирования являет собой точный метод их решения. Потребность в приближенных методах возникает, например, когда нахождение точных производных требует сложных выкладок, а результат содержит громоздкие выражения, или когда надо найти производную табличной функции с неизвестным аналитическим выражением. Ясно, что в последнем случае речь может идти только о приближенном дифференцировании¹⁾.

Пусть функция $y = f(x)$ задана на некотором отрезке $[a; b]$ и имеет таблицу значений в точках $a = x_0 < x_1 < \dots < x_n = b$. *Численным дифференцированием* можно решать задачи двух видов:

¹⁾ Относительно производных табличных функций справедливы замечания, аналогичные тем, что были сделаны в конце 4.1 по поводу интегрирования таких функций.

1. Найти приближенное значение производной k -го порядка функции f в точке $x \in [a; b]$.

2. Найти таблично заданную функцию ($a \leq x_0^* < x_1^* < \dots < x_m^* \leq b$), являющуюся приближением к производной $f^{(k)}$ на отрезке $[x_0^*; x_m^*] \subset [a; b]$. Здесь таблица может строиться на исходных табличных аргументах ($m = n$ и $x_i^* = x_i$), но это не обязательно.

x	x_0^*	x_1^*	...	x_m^*
y	y_0^*	y_1^*	...	y_m^*

Первая задача основная, так как после определения аргументов x_i^* вторая сводится к первой. По этой причине далее будем заниматься только вычислением приближенных значений производных в точках $x \in [a; b]$ и оценкой их погрешностей.

Для решения этой задачи функцию f обычно заменяют достаточным простым k раз дифференцируемым аналитическим приближением p :

$$f(x) \approx p(x), \quad x \in [a; b], \quad (4.33)$$

а затем производную $f^{(k)}(x)$ в точке $x \in [a; b]$ принимают приближенно равной $p^{(k)}(x)$:

$$f^{(k)}(x) \approx p^{(k)}(x), \quad x \in [a; b]. \quad (4.34)$$

Обсудим вопросы, связанные с погрешностью приближенной формулы (4.34) и оценкой этой погрешности.

Если функция f имеет k -ю производную и известен остаточный член r : $r(x) = f(x) - p(x)$ формулы (4.33), справедливо соотношение

$$r^{(k)}(x) = f^{(k)}(x) - p^{(k)}(x), \quad x \in [a; b]. \quad (4.35)$$

Следовательно, погрешность приближения $p^{(k)}(x)$ равна k -й производной остаточного члена r в точке x . Но, поскольку в выражении остаточного члена обычно присутствуют неизвестные числовые параметры, эту погрешность удается находить только приближенно (см., например, п. 4.6.2).

Строгая оценка погрешностей формулы (4.34) представляет собой серьезную проблему уже при вычислении первой производной. В 4.1 показано, что оценочная функция V приближенного равенства $f \approx p$ позволяет получить абсолютную погрешность интеграла от p — неравенство (4.4). Здесь дело обстоит иначе.

Пусть V — оценочная функция для приближенной формулы (4.33). Оказывается, что найти абсолютную погрешность числа $p'(x)$ с помощью функции V не удается, если даже V дифференцируема. А именно, из неравенства $|r(x)| = |f(x) - p(x)| \leq V(x)$ не следует, что $|r'(x)| \leq V'(x)$ и $\Delta_{p'(x)} = V'(x)$. На рис. 4.3 приведен график функции $y = r(x)$, когда $r(x) \geq 0$ и $r'(x) \geq 0$ на $[a; b]$ (значит, $|r(x)| = r(x)$ и $|r'(x)| = r'(x)$). Видно, что $r(x) \leq V(x)$ на $[a; b]$, но в то же время $r'(x) \geq V'(x)$ при всех $x \in [a; b]$.

Надо также заметить, что при дифференцировании приближенных формул вида (4.33) может происходить существенная потеря

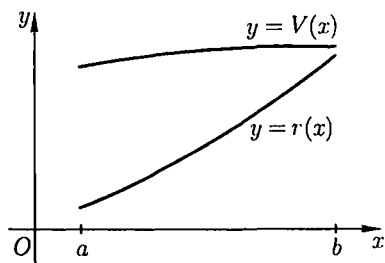


Рис. 4.3

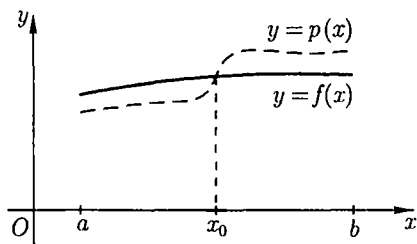


Рис. 4.4

точности. Вследствие этого погрешность формулы (4.34) часто оказывается значительно больше погрешности исходного приближения (4.33). На рис. 4.4 видно, что в достаточно малой окрестности точки x_0 значения функций f и p близки, в точке x_0 они совпадают, однако первые производные в этой окрестности значительно отличаются (самое большое различие между $f'(x)$ и $p'(x)$ имеет место в точке x_0).

Отсюда следует, что задача численного дифференцирования относится к числу *некорректных задач*, т. е. таких, решение которых при малых погрешностях исходных данных может привести к большим погрешностям результатов. При вычислении производных высших порядков возможен еще больший рост погрешностей (упр. 4.28), поэтому на практике приближенные методы редко используются для нахождения производных порядка $k > 2$.

Естественно, что дополнительные погрешности возникают при вычислении значений $p^{(k)}(x)$, но, поскольку они не так существенны, как погрешности формул, и их оценка не вызывает затруднений, далее о них говорить не будем.

Конкретный вид формул численного дифференцирования зависит от способа приближения функции f и выбранного класса приближающих функций p . В п. 4.6.2 они выводятся для случая, когда f приближается интерполяционным многочленом Ньютона.

Следует отметить, что основанные на полиномиальном интерполировании формулы позволяют легко находить приближенные значения производных непосредственно по табличным данным. Однако они могут обладать малой точностью ввиду того, что отмеченная выше некорректность задачи численного дифференцирования при таком способе приближения функции f проявляется весьма значительно. Представим себе, что на отрезке $[a; b]$ функция f близка к линейной (например, как на рис. 4.4), и на узлах $a = x_0 < x_1 < \dots < x_n = b$ построен интерполяционный многочлен $P_n(x)$ высокой степени n , приближающий f с точностью до некоторого числа $\epsilon > 0$:

$$|f(x) - P_n(x)| \leq \epsilon \text{ для всех } x \in [a; b].$$

Соответствующая часть графика многочлена представляет собой извилистую линию, пересекающую график функции f во всех его точках с абсциссами $x_i, i = 0, 1, \dots, n$ и расположенную в области

$$\{(x, y): x \in [a; b], f(x) - \varepsilon \leq y \leq f(x) + \varepsilon\}.$$

Тогда приведенная на рис. 4.4 ситуация, когда в окрестности точки x_0 погрешность приближенного равенства $f'(x) \approx p'(x)$ оказывается существенно большей по сравнению с погрешностью соотношения $f(x) \approx p(x)$, будет иметь место практически на всем отрезке $[a; b]$ (нарисуйте).

Упражнения

4.27. В примере 3.6 для функции с табл. 3.5 получена линейная приближающая функция. Найдите с ее помощью приближенную производную табличной функции в точках $x = 1$ и $x = 2,5$.

4.28. Пусть $f(x) = 1, p(x) = 10x^2 + 1$. Покажите, что на отрезке $[-0,1; 0,1]$ имеет место приближенное равенство $f \approx p$ с абсолютной погрешностью $\Delta_p = 0,1$. Убедитесь, что при дифференцировании этого приближенного равенства происходит рост абсолютной погрешности: $\Delta_{p'} = 20\Delta_p, \Delta_{p''} = 200\Delta_p$.

4.6.2. Формулы численного дифференцирования на основе интерполяционного многочлена Ньютона

Будем считать, что функция f имеет таблицу значений с постоянным шагом h , а приближение ее производится первым интерполяционным многочленом Ньютона P_n (см. п. 3.5.2). Узлы интерполирования выбираем таким образом, чтобы точка x , где вычисляются производные, находилась между узлами x_0 и x_1 .

Первую интерполяционную формулу Ньютона можно записать так:

$$f(x) \approx P_n(x) = y_0 + t\Delta y_0 + \frac{t^2 - t}{2!} \Delta^2 y_0 + \frac{t^3 - 3t^2 + 2t}{3!} \Delta^3 y_0 + \dots, \quad (4.36)$$

где $t = \frac{x - x_0}{h}$. Продифференцируем левую и правую части (4.36), используя соотношение

$$\frac{dP_n}{dx} = \frac{dP_n}{dt} \cdot \frac{dt}{dx} = \frac{1}{h} \frac{dP_n}{dt},$$

и получим

$$f'(x) \approx P'_n(x) = \frac{1}{h} \left(\Delta y_0 + \frac{2t-1}{2} \Delta^2 y_0 + \frac{3t^2-6t+2}{6} \Delta^3 y_0 + \dots \right). \quad (4.37)$$

Если производная вычисляется в табличном аргументе, выбираем его в качестве начального узла. Тогда $x = x_0$, $t = 0$, и формула (4.37) значительно упрощается:

$$f'(x_0) \approx \frac{1}{h} \left(\Delta y_0 - \frac{\Delta^2 y_0}{2} + \frac{\Delta^3 y_0}{3} - \dots \right). \quad (4.38)$$

Полученные соотношения позволяют достаточно легко вычислять приближенные значения производной первого порядка. Аналогично выводятся формулы для производных более высоких порядков.

З а м е ч а н и я. 1. Чтобы избежать громоздких выражений, формулы (4.37) и (4.38) записаны только с несколькими первыми слагаемыми (фактически, для многочлена степени $n = 3$). При необходимости нетрудно указать дополнительные слагаемые.

2. Приближенные равенства (4.31) и (4.32) из предыдущего параграфа являются частными случаями получаемых здесь формул.

Как было сказано в п. 4.6.1, погрешность числа $P'_n(x)$ равна производной остаточного члена формулы (4.36). Предположим, что функция f имеет на $[a; b]$ непрерывную производную $(n + 1)$ -го порядка, и воспользуемся остаточным членом полиномиальных интерполяционных формул общего вида (см. 3.3):

$$R_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!} Q_{n+1}(x). \quad (4.39)$$

Здесь $c = c(x)$ — зависящее от x число, не равное узлам интерполирования и x , а $Q_{n+1}(x) = (x - x_0)(x - x_1) \dots (x - x_n)$.

Поскольку находить и оценивать производную от множителя $f^{(n+1)}(c)$ сложно, при малых h обычно ее заменяют:

$$f^{(n+1)}(c) \approx f^{(n+1)}(x_0) \approx \frac{\Delta^{n+1} y_0}{h^{n+1}} \quad (4.40)$$

[см. формулы (4.32)]¹⁾. Проведем в (4.39) эту замену и дифференцированием получим следующее приближенное значение искомой погрешности:

$$R'_n(x) \approx \frac{\Delta^{n+1} y_0}{h^{n+1} (n+1)!} Q'_{n+1}(x). \quad (4.41)$$

Погрешность производной, найденной в точке $x = x_0$ по формуле (4.38), отыскивается более простым способом. Так как

¹⁾ Для отыскания $\Delta^{n+1} y_0$ надо взять дополнительный узел x_{n+1} и построить полную таблицу конечных разностей.

$$Q'_{n+1}(x_0) = (x_0 - x_1)(x_0 - x_2) \dots (x_0 - x_n) = (-1)^n n! h^n$$

(см. упр. 3.9), вместо (4.41) будем иметь

$$R_n'(x_0) \approx (-1)^n \frac{\Delta^{n+1} y_0}{h(n+1)}. \quad (4.42)$$

Пример 4.4. Дана таблица значений (табл. 4.1) функции $f: f(x) = \operatorname{tg} x$:

Таблица 4.1

x	$\operatorname{tg} x$	Δy	$\Delta^2 y$	$\Delta^3 y$	$\Delta^4 y$
0,8	1,027	0,233	0,064	0,047	0,041
0,9	1,260	0,297	0,111	0,088	0,136
1,0	1,557	0,408	0,199	0,224	
1,1	1,965	0,607	0,423		
1,2	2,572	1,030			
1,3	3,602				

Найдем приближенное значение производной этой функции в точке $x = 0,8$ по формуле численного дифференцирования при $n = 2$. Применяя формулу (4.38) с шагом $h = 0,1$ и $x_0 = 0,8$, получим

$$f'(0,8) \approx \frac{1}{h} \left(\Delta y_0 - \frac{\Delta^2 y_0}{2} \right) = 10 \left(0,233 - \frac{0,064}{2} \right) = 2,01.$$

Определяем погрешность числа 2,01 по формуле (4.42):

$$R_2'(0,8) \approx (-1)^2 \frac{\Delta^3 y_0}{0,1 \cdot 3} = \frac{0,047}{0,3} \approx 0,16.$$

Теперь вычислим значение производной непосредственно по формуле $(\operatorname{tg} x)' = \frac{1}{\cos^2 x}$. С точностью до трех верных значащих цифр после десятичной запятой имеем

$$f'(0,8) = \frac{1}{\cos^2 0,8} \approx 2,060.$$

Видим, что приближенное значение 2,01 производной совпадает с точным значением в двух значащих цифрах. Реальная погрешность приближения примерно равна 0,05, т. е. она оказалась меньше, чем показывает формула (4.42). •

З а м е ч а н и е. Здесь ограничились рассмотрением формул численного дифференцирования на основе первого интерполяционного многочлена Ньютона. Аналогичные формулы можно полу-

чить для случая произвольных таблиц и на основе других интерполяционных многочленов. Их погрешности находят в результате вычисления производных остаточного члена (4.39). Заметим, что упрощенные формулы (4.41) и (4.42) выведены для таблиц с равноотстоящими узлами, но без ограничения на вид интерполяционного многочлена.

Упражнения

4.29. В продолжение проведенных в п. 4.6.2 выкладок выведите формулы приближенного вычисления производных второго порядка и их погрешностей.

4.30. Точность формулы численного дифференцирования при одном и том же шаге h таблицы и степени n интерполяционного многочлена зависит от функции и от точки, где вычисляется производная. Чтобы убедиться в этом, выполните следующие действия:

а) вычислите при $n = 2$ по формуле (4.38) приближенные значения производных: функции $\sin x$ в точке $x = 1,0$ с использованием табл. 3.4 (см. 3.6); функции $\operatorname{tg} x$ в точке $x = 1,0$ (см. табл. 4.1);

б) найдите по формуле (4.42) погрешности приближений и сравните их с аналогичной погрешностью из примера 4.4;

в) вычислите точные производные функций в указанных точках, сравните их с приближенными значениями и найдите реальные погрешности.

4.31. Выполните действия примера 4.4, взяв промежуточную точку $x = 0,93$ и $n = 3$.

4.32. Выведите формулы численного дифференцирования и формулы учета их погрешностей на основе второго интерполяционного многочлена Ньютона.

Глава 5. ЧИСЛЕННОЕ РЕШЕНИЕ ОБЫКНОВЕННЫХ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ

5.1. Необходимые сведения о дифференциальных уравнениях первого порядка¹⁾

Задача Коши. Пусть дано обыкновенное дифференциальное уравнение первого порядка вида

$$y' = f(x, y) \quad (5.1)$$

в предположении, что функция f непрерывна как функция двух переменных в области своего определения D_f .

О п р е д е л е н и е 5.1. Функция $y = \varphi(x)$, непрерывно дифференцируемая на некотором конечном или бесконечном промежутке из \mathbf{R} и обращающая на нем уравнение (5.1) в тождество

$$\varphi'(x) \equiv f(x, \varphi(x)), \quad (5.2)$$

называется *решением* уравнения (5.1) на этом промежутке²⁾.

Различают *общее решение* дифференциального уравнения, которое записывается в виде функции

$$y = \varphi(x, C) \quad (5.3)$$

с произвольной числовой постоянной C , и *частное решение* $y = \varphi(x, C_0)$, получающееся из общего решения при конкретном (допустимом) значении числового параметра $C = C_0$.

Для выделения частного решения обычно ставится условие, которому должно удовлетворять это решение:

$$y = y_0 \text{ при } x = x_0, \quad (5.4)$$

или, короче, $y(x_0) = y_0$. Понятно, что здесь $(x_0, y_0) \in D_f$.

¹⁾ См. [4, гл. XXV, § 2, 3], [16, гл. I, § 1–3].

²⁾ В данной главе слово «решение» будет появляться в двух смыслах: во-первых, это функция, удовлетворяющая некоторому дифференциальному уравнению; во-вторых, это процесс поиска такой функции. О чем конкретно идет речь в той или иной ситуации, нетрудно будет понять по контексту.

Соотношение (5.4) называют *начальным условием*, числа x_0 и y_0 — *начальными данными*, а точку (x_0, y_0) — *начальной точкой*.

Определение 5.2. Задача нахождения частного решения уравнения (5.1), удовлетворяющего начальному условию (5.4), называется *задачей Коши*.

Задачу Коши можно решить без выявления общего решения. Если же общее решение (5.3) известно и начальное условие (5.4) дано, то число C_0 , определяющее искомое частное решение, находят из уравнения относительно C : $y_0 = \varphi(x_0, C)$.

Пример 5.1. Общим решением дифференциального уравнения

$$y' = y$$

на всем \mathbf{R} является функция $y = Ce^x$. При $C = 2$ получится частное решение $y = 2e^x$, а число $C = 0$ дает частное решение $y = 0$.

Пусть требуется решить задачу Коши с начальным условием $y(2) = 1$. Для этого сначала найдем $C = e^{-2}$ из уравнения $1 = Ce^2$, а затем, подставив это значение C в общее решение, получим искомое частное решение $y = e^{x-2}$. •

График решения дифференциального уравнения называется *интегральной кривой* этого уравнения. Общее решение (5.3) определяет семейство интегральных кривых (каждому конкретному значению C соответствует своя кривая), имеющих, как правило, одинаковую или похожую конфигурацию. Например, графики всех найденных в примере 5.1 решений, за исключением прямой $y = 0$, являются экспоненциальными кривыми.

Геометрический смысл задачи Коши заключается в нахождении интегральной кривой, проходящей через начальную точку (x_0, y_0) . В частности, через точку $(2, 1)$ из примера 5.1 пройдет кривая $y = e^{x-2}$.

Существование и единственность решения задачи Коши. Прежде чем приступить к решению задачи Коши, предварительно надо выяснить, существует ли решение уравнения, удовлетворяющее данному начальному условию, а если да, то сколько их может быть. Справедлива следующая теорема [4, гл. XXV, § 2].

Теорема 5.1 (теорема Пикара). Пусть точка (x_0, y_0) является внутренней точкой замкнутой прямоугольной области

$$D = \{(x, y): a_1 \leq x \leq b_1; a_2 \leq y \leq b_2\} \subset D_f$$

($a_1 < b_1, a_2 < b_2$), и на области D выполнены условия:

- 1) функция f непрерывна как функция двух переменных;
- 2) частная производная f'_y существует и ограничена как функция двух переменных (в частности, непрерывна в этой области).

Тогда найдется такой отрезок $[x_0 - \delta; x_0 + \delta] \subset [a_1; b_1]$ ($\delta > 0$), на котором уравнение (5.1) имеет единственное решение $y = \varphi(x)$, удовлетворяющее начальному условию (5.4).

Заметим, что в качестве начальной точки в теореме Пикара можно взять любую внутреннюю точку области D . Следовательно, че-

рез каждую такую точку в достаточно малой ее окрестности проходит единственная интегральная кривая из семейства $y = \varphi(x, C)$. Понятно, что когда f и f'_y непрерывны на \mathbb{R}^2 , задача Коши имеет единственное решение при любой начальной точке $(x_0, y_0) \in \mathbb{R}^2$.

Далее займемся поиском приближений к решению задачи Коши, считая его единственным. Примем следующее соглашение.

С о г л а ш е н и е 5.1. Будем считать, что все дальнейшие рассуждения и вычисления, связанные с приближенным решением задачи Коши, проводятся внутри некоторой замкнутой прямоугольной области D , где выполняются условия теоремы Пикара.

Геометрический смысл правой части дифференциального уравнения (5.1). Возьмем произвольную точку $A(x, y)$ на интегральной кривой $y = \varphi(x)$ (рис. 5.1). Так как функция φ является решением, справедливо соотношение

$$\varphi'(x) = f(x, \varphi(x)) = f(x, y). \quad (5.5)$$

С другой стороны, по геометрическому смыслу производной, $\varphi'(x) = \operatorname{tg} \alpha$, где α — угол между касательной, проведенной к дан-

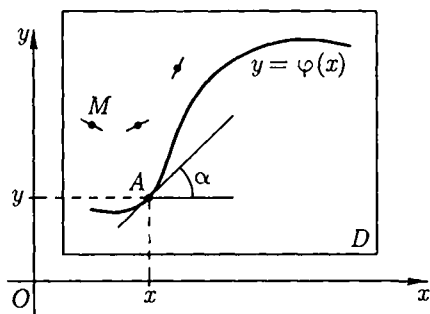


Рис. 5.1

ной кривой в точке A , и положительным направлением оси Ox . Отсюда следует, что $\operatorname{tg} \alpha = f(x, y)$, т.е. значение $f(x, y)$ функции f равно угловому коэффициенту касательной, проведенной в точке A к интегральной кривой $y = \varphi(x)$.

Полученный результат справедлив для любой внутренней точки $M(x, y)$ из области D . Вычислив $f(x, y)$, получим угловой коэффициент касательной к некоторой проходящей через эту

точку интегральной кривой, при этом сама кривая может быть неизвестна. Можно сказать, что значения $f(x, y)$ определяют *направления интегральных кривых* уравнения (5.1) в тех точках, где они вычислены.

Направление кривой в точке $M(x, y)$ обычно указывается с помощью отрезка небольшой длины с центром в M и с углом наклона $\alpha = \operatorname{arctg} f(x, y)$ к положительному направлению оси Ox . Проведя достаточно большое число таких отрезков, получаем некоторое представление о конфигурации интегральных кривых уравнения.

Упражнения

5.1. 1. Проиллюстрируйте геометрически полученные в примере 5.1 результаты.

2. Отметьте на графике направления интегральных кривых уравнения из примера 5.1 в точках $(-2, -1)$ и $(3, 1)$.

5.2. Пусть даны дифференциальное уравнение $y' = f(x)$ и внутренняя точка x_0 области определения функции f .

1. Убедитесь, что функция $y = \int_{x_0}^x f(t) dt + C$ является решением уравнения при всех $C \in \mathbb{R}$.

2. Запишите в интегральной форме частное решение этого уравнения, удовлетворяющее начальному условию $y(x_0) = y_0$.

5.3. Дано уравнение $y' = 2(x - 1)$. Докажите, что задача Коши с начальным условием $y(0) = 1$ имеет единственное решение, и найдите это решение двумя способами: выделением параметра C из общего решения и с помощью результата из упр. 5.2 (задание 2).

5.4. Докажите, что задача Коши для уравнения $y' = x^2 + y^2$ имеет единственное решение при любой начальной точке $(x_0, y_0) \in \mathbb{R}^2$.

5.5. Дано уравнение $y' = \frac{x}{y}$. Приведите примеры замкнутых прямоугольных областей D , где выполняются условия теоремы Пикара для задачи Коши с начальным условием $y(2) = 2$.

5.2. Понятие численного решения задачи Коши

Потребность в приближенном решении задачи Коши возникает прежде всего в том случае, когда дифференциальное уравнение не принадлежит ни одному из классов уравнений, для которых известны точные методы интегрирования. К таким уравнениям относится, например, уравнение из упр. 5.4. Приближенные методы часто применяют и тогда, когда точные методы оказываются неэффективными ввиду больших затрат времени и усилий для их реализации.

На практике наиболее распространенными являются *численные методы* приближенного интегрирования дифференциальных уравнений, дающие решение задачи Коши в виде таблицы приближенных значений точного решения φ . Эту таблицу [таблично заданную функцию $y = y(x)$] и называют *численным решением задачи Коши*. Для выполнения начального условия (5.4) таблица должна содержать данные x_0, y_0 .

Если задан конечный промежуток, на котором ищется решение, и точка x_0 лежит внутри этого промежутка, приближенное решение имеет вид табл. 5.1. Вначале в ней известны x_0 и y_0 , затем отыскиваются остальные табличные данные. Так как правила определения «верхних» и «нижних» (относительно x_0, y_0)

Таблица 5.1

x	y
x_{-m}	y_{-m}
\vdots	\vdots
x_{-1}	y_{-1}
x_0	y_0
x_1	y_1
\vdots	\vdots
x_n	y_n

Таблица 5.2 данных одинаковы, будем искать это решение в виде табл. 5.2.

x	y
x_0	y_0
x_1	y_1
x_2	y_2
\vdots	\vdots
x_n	y_n

Для построения табл. 5.2 выбирается шаг h и вычисляются табличные аргументы $x_i = x_0 + ih$ ($i = 1, 2, \dots, n$). Затем последовательно находят числа y_i , близкие к значениям точного решения φ в точках x_i :

$$\varphi(x_i) \approx y_i \quad (i = 1, 2, \dots, n). \quad (5.6)$$

Точность приближенных равенств (5.6) зависит от способа вычисления y_i и от шага таблицы h . Чем меньше шаг, тем выше должна быть точность таблицы. Заданное значение y_0 считается точным числом. Погрешности появляются при вычислении y_1 , а далее обычно происходит их накопление.

В параграфах 5.3 и 5.4 излагаются методы определения y_n , а вопросы оценки их погрешностей рассматриваются в 5.5.

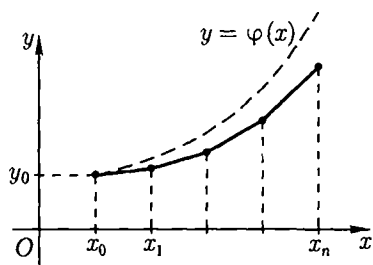


Рис. 5.2

Табл. 5.2 является приближением к решению φ . Если на плоскости xOy построить точки таблицы $(x_0, y_0), \dots, (x_n, y_n)$ и соединить их отрезками, получится так называемая *ломаная Эйлера* (рис. 5.2). Она будет приближением интегральной кривой $y = \varphi(x)$.

Для вывода формул вычисления y_i понадобится теорема 5.2, являющаяся частным случаем утверждения из [17, п. 144].

Теорема 5.2. Если все частные производные функции f до k -го ($k \geq 1$) порядка непрерывны в прямоугольной области D , то всякое решение $y = \varphi(x)$ уравнения (5.1), график которого проходит через внутреннюю точку (x_0, y_0) этой области, имеет производную $(k + 1)$ -го порядка, непрерывную в некоторой окрестности x_0 .

Упражнения

5.6. Численное приближение точного решения φ задачи Коши наиболее просто можно получить в случае уравнений вида $y' = f(x)$. Убедитесь, что формулы $y_{i+1} = \int_{x_i}^{x_{i+1}} f(t) dt + y_i$ ($i = 0, 1, \dots, n - 1$) позволяют находить точные значения $y_i = \varphi(x_i)$. Выясните, когда вычисленные по этим формулам значения y_i могут оказаться приближенными.

5.7. Применяя формулы из упр. 5.6, найдите численное решение задачи Коши из упр. 5.3 на отрезке $[0; 3]$ с шагом $h = 1$. Постройте ломаную Эйлера и график точного решения.

5.3. Метод Эйлера

Пусть требуется найти численное решение $y = y(x)$ уравнения (5.1), удовлетворяющее начальному условию (5.4). В дополнение к соглашению 5.1 считаем, что частные производные f'_x и f'_y непрерывны в области D .

Обозначим через φ точное решение поставленной задачи Коши. В силу сделанных предположений производные второго порядка функции φ и всех других решений уравнения (5.1), графики которых проходят через используемые в процессе дальнейших вычислений точки $(x_i, y_i) \in D$, существуют и непрерывны в окрестности абсцисс этих точек (теорема 5.2).

Численное решение ищем в виде табл. 5.2. Для этого выберем шаг h и число $n \in \mathbf{N}$, вычислим $x_i = x_0 + ih$ ($i = 0, 1, \dots, n$) и проставим эти аргументы в таблицу.

Далее надо найти y_1 , исходя из известных данных x_0, y_0 и x_1 . Воспользуемся формулой Тейлора:

$$\varphi(x_1) = \varphi(x_0) + \varphi'(x_0)h + \frac{\varphi''(c_1)}{2!}h^2, \text{ где } c_1 \in (x_0; x_1). \quad (5.7)$$

В силу непрерывности φ'' при малых h можно отбросить последнее слагаемое из правой части. Тогда, с учетом (5.5) и того, что $\varphi(x_0) = y_0$, получим

$$\varphi(x_1) \approx y_0 + hf(x_0, y_0).$$

Значит, можно взять

$$y_1 = y_0 + hf(x_0, y_0). \quad (5.8)$$

Теперь известны табличные данные x_1, y_1 и x_2 . Для вывода формулы вычисления y_2 требуется решение, удовлетворяющее начальному условию $y = y_1$ при $x = x_1$. Функция φ здесь может не подойти. Обозначим через φ_1 то решение, для которого верно равенство $\varphi_1(x_1) = y_1$ и снова воспользуемся формулой Тейлора:

$$\varphi_1(x_2) = \varphi_1(x_1) + \varphi_1'(x_1)h + \frac{\varphi_1''(c_2)}{2!}h^2, \text{ где } c_2 \in (x_1; x_2). \quad (5.9)$$

Отбрасываем слагаемое с h^2 и благодаря равенству $\varphi_1'(x_1) = f(x_1, y_1)$ получаем

$$\varphi_1(x_2) \approx y_1 + hf(x_1, y_1).$$

Если шаг h мал и решение φ мало отличается от φ_1 , полученное приближение к $\varphi_1(x_2)$ можем считать приближением и к $\varphi(x_2)$. Тогда

$$y_2 = y_1 + hf(x_1, y_1). \quad (5.10)$$

Далее действуем аналогично, используя при поиске y_{i+1} по известной точке (x_i, y_i) соответствующее частное решение φ_i .

В итоге получим формулы метода Эйлера

$$y_{i+1} = y_i + hf(x_i, y_i) \quad (i = 0, 1, \dots, n-1), \quad (5.11)$$

позволяющие последовательно находить значения y_i табл. 5.2.

Вычисления методом Эйлера имеют простой геометрический смысл (рис. 5.3). Через начальную точку A_0 штриховой линией проведена интегральная кривая $y = \varphi(x)$ (она может быть неизвестной).

Число $f_0 = f(x_0, y_0)$ определит касательную к этой кривой в точке A_0 . Взяв на этой касательной точку A_1 с абсциссой x_1 , можно рассмотреть прямоугольный треугольник A_0CA_1 с углом α при вершине A_0 и параллельным оси Ox катетом A_0C . Поскольку $f_0 = \operatorname{tg} \alpha$, произведение hf_0 даст длину катета CA_1 , а формула (5.8) — ординату y_1 точки A_1 .

Проследим еще один шаг вычислений. Через точку A_1 проходит некоторая интегральная кривая $y = \varphi_1(x)$. Касательная к ней в этой точке имеет угловой коэффициент $f_1 = f(x_1, y_1)$. По формуле (5.10) определяется ордината y_2 точки A_2 . Замечаем, что f_1 представляет собой приближенное значение углового коэффициента касательной к точной кривой $y = \varphi(x)$ в точке B_1 .

Таким образом, имея некоторую точку $A_i(x_i, y_i)$ (сначала это A_0) и проведя через нее касательную к соответствующей интегральной кривой, можно найти y_{i+1} как ординату точки A_{i+1} этой касательной, соответствующей абсциссе x_{i+1} . Отрезки прямых между последовательными точками $A_i (i = 0, 1, \dots, n)$ образуют ломаную Эйлера.

В геометрических рассуждениях существенно используется то, что область вычислений заполнена родственными интегральными кривыми с одинаковой или похожей конфигурацией. Ввиду этого при каждом $i = 0, 1, \dots, n-1$ число $f(x_i, y_i) = f_i$ определяет не только направление кривой $y = \varphi_i(x)$ в точке (x_i, y_i) , но с той или иной степенью точности задает направление других близких к ней кривых в точках с абсциссой x_i (в частности, направление точной интегральной кривой). Иными словами, с помощью f_i можно узнать характер направлений интегральных кривых вдоль прямой $x = x_i$.

Метод Эйлера задает простой алгоритм вычислений, но определяет табличные значения y_i с невысокой точностью. Как видно из геометрического смысла формул, при поиске y_{i+1} используются направления интегральных кривых, характерные только для левого конца отрезка $\{x_i; x_{i+1}\}$, и не учитываются изменения в поведении этих кривых

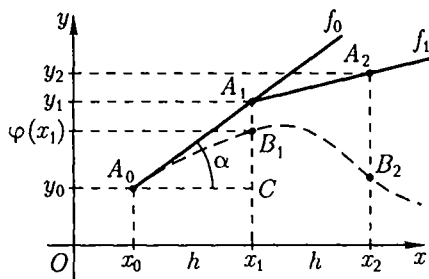


Рис. 5.3

Таблица 5.3

вых на всем отрезке. На рис. 5.3 видно, что по этой причине число y_2 оказывается очень грубым приближением к $\varphi(x_2)$.

Пример 5.2. Даны уравнение $y' = 2x$ и начальное условие $y(0) = 0$. Найдем методом Эйлера численное решение задачи Коши на отрезке $[a; b] = [0; 2]$ с шагом $h = 0,4$.

x	y	$\varphi(x)$	ε
0	0	0	0
0,4	0	0,16	0,16
0,8	0,32	0,64	0,32
1,2	0,96	1,44	0,48
1,6	1,92	2,56	0,64
2,0	3,20	4,00	0,80

Используя $x_0 = 0, y_0 = 0$, находим $y_1 = 0 + 0,4 \cdot 2 \cdot 0 = 0$. Затем $y_2 = 0 + 0,4 \cdot 2 \cdot 0,4 = 0,32$, и т.д. В результате получим таблично заданную функцию $y = y(x)$, размещенную в первых двух столбцах табл. 5.3. В третьем столбце таблицы приведены значения точного решения $\varphi: \varphi(x) = x^2$, а в четвертом столбце — погрешности $\varepsilon_i = \varphi(x_i) - y_i$ чисел y_i .

Поскольку шаг h выбран большим, погрешности велики. Замечаем, что они растут по мере приближения к концу таблицы. •

Упражнения

5.8. Начертите точную интегральную кривую из примера 5.2 и ломаную Эйлера, соответствующую полученному там численному решению.

5.9. Дано дифференциальное уравнение $y' = 2xy$. Найдите методом Эйлера на отрезке $[0; 1]$ с шагом $h = 0,2$ численное решение задачи Коши с начальным условием $y(0) = 1$ (вычисления вести с двумя цифрами после десятичной запятой). Определите погрешности y_i , используя точное решение $y = e^{x^2}$. Постройте ломаную Эйлера и точную интегральную кривую.

5.10. Формулы метода Эйлера можно вывести разными способами. Получите формулу (5.8):

1) заменив точную интегральную кривую на $[x_0; x_1]$ касательной к ней, проведенной в точке (x_0, y_0) (геометрический вывод);

2) проинтегрировав тождество (5.2) по отрезку $[x_0; x_1]$ (учтите при этом, что ввиду непрерывности функций f и φ при малых h можно взять $f(x, \varphi(x)) \approx f(x_0, y_0)$).

5.4. Усовершенствования метода Эйлера

Рассмотрим две модификации метода Эйлера. При этом в дополнение к соглашению 5.1 будем считать, что все частные производные второго порядка функции f непрерывны в области вычислений. Тогда в силу теоремы 5.2 решения уравнения имеют непрерывную третью производную.

Вычислительные формулы будем выводить двумя способами: геометрически и аналитически. Геометрические рассуждения наглядны, а аналитические выкладки позволят охарактеризовать точность методов (см. 5.5).

5.4.1. Метод Эйлера – Коши

Геометрический вывод. Пусть известны данные x_i , y_i и x_{i+1} . Более точное приближение $y_{i+1} \approx \varphi(x_{i+1})$ можно получить, если учитывать направления интегральных кривых, характерные для начала и конца отрезка $[x_i; x_{i+1}]$. Иллюстрация вычисления y_{i+1} при $i = 0$ приведена на рис. 5.4.

Точка $A_i(x_i, y_i)$ лежит на некоторой интегральной кривой (при $i = 0$ это точное решение задачи Коши $y = \varphi(x)$), касательная к которой в точке A_i имеет угловой коэффициент $f_i = f(x_i, y_i)$. По формуле (5.11) находим ординату точки на этой касательной, соответствующей абсциссе x_{i+1} :

$$\bar{y}_{i+1} = y_i + hf_i. \quad (5.12)$$

Заметим, что в методе Эйлера число \bar{y}_{i+1} принималось за искомое приближение к $\varphi(x_{i+1})$. Здесь же оно используется лишь для определения «концевой» точки $\bar{A}_{i+1}(x_{i+1}; \bar{y}_{i+1})$.

Вычислив $\bar{f}_{i+1} = f(x_{i+1}; \bar{y}_{i+1})$, узнаем направление проходящей через \bar{A}_{i+1} интегральной кривой в этой точке. Теперь найдем «усредненное» направление кривых на рассматриваемом отрезке:

$$\bar{f}_i = \frac{f_i + \bar{f}_{i+1}}{2} \quad (5.13)$$

и возьмем в качестве y_{i+1} число

$$y_{i+1} = y_i + h\bar{f}_i. \quad (5.14)$$

Геометрический смысл формулы (5.14) следующий. Если через исходную точку A_i провести прямую с угловым коэффициентом \bar{f}_i и взять на ней точку A_{i+1} с абсциссой x_{i+1} , то (5.14) определяет ординату этой точки.

Формулы (5.12) – (5.14) в совокупности задают алгоритм последовательного вычисления y_i ($i = 1, 2, \dots, n$) методом Эйлера – Коши.

Аналитический вывод. Для простоты ограничимся получением формул (5.12) – (5.14) при $i = 0$.

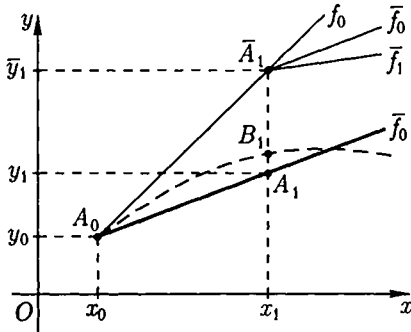


Рис. 5.4

В основе вычислений методом Эйлера лежит линейное относительно h усечение формулы Тейлора. Можно ожидать более высокой точности от y_i , если вместо линейного усечения брать квадратичное.

Имеет место равенство

$$\varphi(x_1) = \varphi(x_0) + \varphi'(x_0)h + \frac{\varphi''(x_0)}{2!}h^2 + \frac{\varphi'''(c)}{3!}h^3 \quad (c \in (x_0; x_1)), \quad (5.15)$$

из которого отбрасыванием слагаемого с h^3 получим

$$\varphi(x_1) \approx y_0 + \varphi'(x_0)h + \frac{\varphi''(x_0)}{2!}h^2. \quad (5.16)$$

Подставим в (5.16) вместо $\varphi''(x_0)$ его приближенное значение

$$\varphi''(x_0) \approx \frac{\varphi'(x_1) - \varphi'(x_0)}{h},$$

наличие которого следует из определения второй производной. Тогда

$$\varphi(x_1) \approx y_0 + h \frac{\varphi'(x_0) + \varphi'(x_1)}{2}. \quad (5.17)$$

Здесь известно число $\varphi'(x_0) = f(x_0, y_0)$. Чтобы выразить $\varphi'(x_1)$ через значение функции f , найдем «грубое» приближение к $\varphi(x_1)$:

$$\bar{y}_1 = y_0 + hf(x_0, y_0)$$

(это формула (5.12) при $i = 0$) и обозначим через ψ решение уравнения (5.1), удовлетворяющее начальному условию: $\psi(x_1) = \bar{y}_1$. Понятно, что $\psi'(x_1) = f(x_1, \psi(x_1)) \stackrel{\text{def}}{=} \bar{f}_1$. Поскольку φ и ψ являются родственными решениями, можно взять

$$\varphi'(x_1) \approx \psi'(x_1) = \bar{f}_1.$$

Если теперь подставить полученные выражения для производных $\varphi'(x_0)$ и $\varphi'(x_1)$ в (5.17), затем обозначить дробь через \bar{f}_0 , а найденное таким образом приближение к $\varphi(x_1)$ через y_1 , будем иметь формулы (5.13) и (5.14) при $i = 0$.

Вывод формул для вычисления y_{i+1} ($i = 1, 2, \dots, n - 1$) производится по аналогичной схеме с использованием вместо φ решений φ_i , удовлетворяющих условиям $\varphi_i(x_i) = y_i$.

Упражнения

5.11. Дополните рис. 5.4 построениями для определения y_2 .

5.12. Вычислите методом Эйлера – Коши y_1 и y_2 из упр. 5.9 и сравните их по точности с полученными по методу Эйлера.

$$\varphi''(x_0) \approx \frac{2}{h} (\varphi'(x_1) - \varphi'(x_0)),$$

правую часть которого подставим в (5.16) вместо $\varphi''(x_0)$. После упрощений будем иметь

$$\varphi(x_1) \approx y_0 + h\varphi'(x_{\frac{1}{2}}). \quad (5.21)$$

Чтобы выразить $\varphi'(x_{\frac{1}{2}})$ приближенно через значение функции f , вычислим

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

Обозначив через ψ решение дифференциального уравнения, удовлетворяющее начальному условию $\psi(x_{\frac{1}{2}}) = y_{\frac{1}{2}}$, на основе тех же рассуждений, что были использованы в предыдущем пункте, получаем

$$\varphi'(x_{\frac{1}{2}}) \approx \psi'(x_{\frac{1}{2}}) = f(x_{\frac{1}{2}}, y_{\frac{1}{2}}) \stackrel{\text{def}}{=} f_{\frac{1}{2}}.$$

Теперь в (5.21) заменим $\varphi'(x_{\frac{1}{2}})$ на найденное приближение, обозначим правую часть через y_1 и получим формулу (5.20) при $i = 0$.

Упражнения

5.14. Выполните геометрические построения, соответствующие второму шагу метода срединных точек.

5.15. Выполните упр. 5.12 методом срединных точек.

5.16. Чем ближе к $\varphi(x_{i+1})$ грубое приближение \bar{y}_{i+1} в методе Эйлера – Коши, тем лучше число f_{i+1} отражает направление точной интегральной кривой в точке с абсциссой x_{i+1} , а значит, тем точнее найдется y_{i+1} .

1. Напишите формулы комбинирования методов: сначала ищется грубое приближение \bar{y}_{i+1} методом срединных точек, а затем оно уточняется методом Эйлера – Коши.

2. Проиллюстрируйте графически вычисления y_1 по получаемым формулам.

5.5. Точность метода Эйлера и его модификаций

Анализ вывода формул в 5.3 и 5.4 показывает, что при построении численного решения задачи Коши наслаиваются погрешности, вызванные разными источниками. По этой причине строгие оценки погрешностей чисел y_i вида $|\varphi(x_i) - y_i| \leq \Delta_{y_i}$ оказываются слож-

ными и малопригодными (см., например, [11, гл. II]). На практике ими не пользуются, а ограничиваются ориентировочной характеристикой точности методов и нахождением приближенных значений для расстояний $|\varphi(x_i) - y_i|$ по правилу Рунге.

В методе Эйлера имеют место следующие основные переходы к приближенным значениям:

$$\varphi(x_{i+1}) \stackrel{(1)}{\approx} \varphi_i(x_{i+1}) \stackrel{(2)}{\approx} \varphi_i(x_i) + \varphi'_i(x_i)h = y_i + f(x_i, y_i)h = y_{i+1}.$$

Погрешность приближенного равенства (1) вызвана заменой точного решения φ на некоторое другое решение φ_i . Значение этой погрешности зависит от дифференциального уравнения. В одних случаях это значение может быть существенным, в других — пренебрежимо малым.

Приближенное равенство (2) обусловлено отбрасыванием остаточного члена формулы Тейлора; возникающая при этом *погрешность усечения* является основной характеристикой точности метода. Поскольку при линейном усечении остаточный член равен $\frac{\varphi_i''(c_i)}{2!} h^2$, правомерен вывод: *метод Эйлера при каждом переходе от y_i к y_{i+1} порождает погрешность, пропорциональную h^2* . Говорят также, что *локальная погрешность* метода имеет порядок h^2 или равна $O(h^2)$ (читается «о большое от h^2 ») ¹⁾.

Ясно, что от y_n следует ожидать меньшей точности, чем от y_1 . Как отмечалось в 5.2, при движении к концу таблицы происходит накопление погрешностей, поэтому их следует суммировать. Учитывая, что длина отрезка $[x_0; x_n]$ равна $L = nh$, вычислим порядок погрешности числа y_n , найденного после n шагов:

$$nO(h^2) = \frac{L}{h} O(h^2) = O(h).$$

Отсюда следует вывод, что *глобальная погрешность* метода Эйлера по всей таблице равна $O(h)$.

Аналогичными рассуждениями убеждаемся, что локальная погрешность метода Эйлера – Коши и метода серединных точек равна $O(h^3)$, а их глобальная погрешность — $O(h^2)$. Таким образом, эти методы на порядок точнее метода Эйлера.

Полученные характеристики достаточно грубы, но они позволяют узнать влияние шага таблицы на точность численного решения. При уменьшении шага h вдвое можно ожидать в среднем примерно двукратного уменьшения погрешностей найденных методом Эйлера

¹⁾ При желании узнать, сколь велико влияние на погрешность коэффициента пропорциональности $\frac{\varphi_i''(c_i)}{2!}$, можно вычислить его приближенное значение (см. упр. 5.19).

ра табличных значений y_i ; если же y_i вычислялись по модифицированным алгоритмам, то их точность возрастает вчетверо.

З а м е ч а н и е. Отметим, что изученные нами метод Эйлера и его модификации принадлежат к семейству так называемых методов *Рунге – Кутты*. Для численного решения задачи Коши на практике используется еще один метод из этого семейства, описание которого можно найти в [8, § 6.4] и [15, § 31]. Его локальная и глобальная точности характеризуются порядками h^5 и h^4 соответственно.

Теперь приведем без доказательства используемые на практике формулы приблизительной оценки точности значений численного решения методом *двойного пересчета* (методом Рунге). Они основаны на учете только погрешностей усечения формулы Тейлора, т. е. при этом пренебрегают другими источниками погрешностей, включая и вычислительные (см., например, [7, гл. II, § 15]).

Пусть строится таблица с отстоящими на шаг h аргументами x_0, x_1, \dots, x_n . Сначала одним из методов отыскивают значения y_i ($i = 1, 2, \dots, n$) с шагом h , а затем проводятся вычисления с шагом $h/2$. Понятно, что в последнем случае при каждом переходе от данного аргумента к следующему потребуется двукратное применение метода. Соответствующие аргументам x_i новые табличные значения обозначим через y_i^* ($y_0^* = y_0$). Это улучшенные приближения к $\varphi(x_i)$, и поэтому таблицу с данными x_i, y_i^* ($i = 0, 1, 2, \dots, n$) возьмем в качестве искомого численного решения с шагом h .

Расстояния между y_i^* и точными числами $\varphi(x_i)$ ($i = 1, 2, \dots, n$) вычисляются по приближенным формулам

$$|\varphi(x_i) - y_i^*| \approx |y_i - y_i^*| \quad (5.22)$$

для метода Эйлера и

$$|\varphi(x_i) - y_i^*| \approx \frac{|y_i - y_i^*|}{3} \quad (5.23)$$

для обеих модификаций метода Эйлера.

П р и м е р 5.3. Пусть численное решение дифференциального уравнения $y' = x^2 + y^2$, удовлетворяющее начальному условию $y(1) = 1$, ищется методом Эйлера в виде таблицы с шагом $h = 0,2$. Требуется найти y_1 и y_1^* , соответствующие аргументу $x_1 = 1,2$, и оценить погрешность y_1^* .

При $h = 0,2$ получим

$$y_1 = y_0 + hf(x_0, y_0) = 1 + 0,2 \cdot 2 = 1,4.$$

Теперь проведем двукратные вычисления с шагом $h/2 = 0,1$. Сначала находим $x_{1/2} = 1,1$ и

$$y_{1/2} = y_0 + \frac{h}{2} f(x_0, y_0) = 1 + 0,1 \cdot 2 = 1,2.$$

Затем вычисляем значение y_1^* :

$$y_1^* = y_{1/2} + \frac{h}{2} f(x_{1/2}, y_{1/2}) = 1,2 + 0,1 \cdot 2,65 = 1,465.$$

С помощью (5.22) найдем приблизительную оценку погрешности:

$$|\varphi(1,2) - 1,465| \approx |1,4 - 1,465| = 0,065. \bullet$$

Упражнения

5.17. В продолжение примера 5.3 найдите значения y_2^* , y_3^* для аргументов $x_2 = 1,4$, $x_3 = 1,6$ и оцените их погрешности (при вычислениях сохраняйте три цифры после десятичной запятой).

5.18. 1. Напишите все формулы, необходимые для вычисления y_i^* по начальным табличным данным с шагом $h/2$ обоими модифицированными методами Эйлера, и отобразите на рисунке геометрический смысл вычислений.

2. Выполните задания примера 5.3, применяя вместо метода Эйлера метод Эйлера – Коши.

5.19.* Выведите приближенное равенство

$$\frac{\varphi_i''(c_i)}{2} \approx \frac{f(x_{i+1}, y_{i+1}) - f(x_i, y_i)}{2h}$$

для коэффициента пропорциональности погрешности линейного усечения формулы Тейлора.

У к а з а н и е. Примените идеи, использованные при выводе формул метода Эйлера – Коши.

5.6. Решение систем дифференциальных уравнений первого порядка

Изученные численные методы легко распространяются на случай систем обыкновенных дифференциальных уравнений первого порядка.

Пусть дана система из двух уравнений

$$\begin{cases} y' = f_1(x, y, z), \\ z' = f_2(x, y, z) \end{cases} \quad (5.24)$$

относительно неизвестных функций y и z от переменной x . *Решением* ее на некотором конечном или бесконечном промежутке $X \subset \mathbf{R}$ называется пара непрерывно дифференцируемых на X функций $y = \varphi(x)$, $z = \psi(x)$, которые при подстановке в систему превращают уравнения в тождества на X .

Задача Коши для системы (5.24) заключается в поиске таких решений φ и ψ , которые удовлетворяют начальным условиям:

$$\varphi(x_0) = y_0, \psi(x_0) = z_0. \quad (5.25)$$

Будем считать, что в достаточно большой окрестности начальных данных выполнены условия существования и единственности решения задачи Коши, а также необходимые для вывода вычислительных формул свойства функций f_1 и f_2 — непрерывность всех их частных производных первого и второго порядка. В этом случае имеет место теорема, гарантирующая непрерывность первой и второй производных функций φ и ψ , аналогичная теореме 5.2. Все эти вопросы подробно изложены в [17, гл. IV, V].

Численное решение задачи Коши ищется в виде двух таблично заданных функций $y = y(x)$ и $z = z(x)$, определенных на одном и том же множестве точек $x_i = x_0 + ih$ ($i = 0, 1, \dots, n$). Начальными данными этих таблиц являются x_0, y_0 и x_0, z_0 соответственно.

Алгоритм метода Эйлера состоит из последовательных вычислений по формулам

$$\begin{aligned} y_{i+1} &= y_i + hf_1(x_i, y_i, z_i), \\ z_{i+1} &= z_i + hf_2(x_i, y_i, z_i) \end{aligned} \quad (5.26)$$

для $i = 0, 1, \dots, n - 1$.

Формулы для модифицированных методов Эйлера можно вывести самостоятельно.

Упражнения

5.20. Опишите алгоритмы построения численного решения задачи Коши для системы

$$\begin{cases} y' = f_1(x, y, z, u), \\ z' = f_2(x, y, z, u), \\ u' = f_3(x, y, z, u) \end{cases}$$

с начальными условиями $y(x_0) = y_0$, $z(x_0) = z_0$, $u(x_0) = u_0$ методом Эйлера и обеими его модификациями.

5.21. Найдите на отрезке $[0; 1]$ с шагом $h = 0,5$ численное решение системы

$$\begin{cases} y' = y + z, \\ z' = 2xy, \end{cases}$$

удовлетворяющее начальным условиям: $y(0) = 1$ и $z(0) = 2$, используя: а) метод Эйлера; б) метод срединных точек.

5.7. Численное решение дифференциальных уравнений высших порядков

Основная идея численного решения обыкновенного дифференциального уравнения n -го порядка ($n > 1$) заключается в замене этого уравнения на соответствующую систему из n дифференциальных уравнений первого порядка.

Пусть дано уравнение третьего порядка

$$y''' = f(x, y, y', y''), \quad (5.27)$$

и пусть $y = \varphi(x)$ — точное решение уравнения (5.27), удовлетворяющее начальным условиям

$$y(x_0) = y_0, y'(x_0) = z_0, y''(x_0) = t_0. \quad (5.28)$$

Требуется построить таблично заданную функцию $y = y(x)$, являющуюся приближением к φ .

Чтобы решить эту задачу, сделаем замену $z = y'$, $u = y''$ и представим уравнение (5.27) в виде системы

$$\begin{cases} y' = z, \\ z' = u, \\ u' = f(x, y, z, u). \end{cases} \quad (5.29)$$

Начальные условия (5.28) переписутся так:

$$y(x_0) = y_0, z(x_0) = z_0, u(x_0) = u_0. \quad (5.30)$$

Решение задачи Коши для системы (5.29) с условиями (5.30) даст три функции $y = \varphi(x)$, $z = \psi(x)$, $u = \Phi(x)$, одна из которых, а именно $y = \varphi(x)$, является точным решением задачи Коши для исходного уравнения (5.27) [17, п. 113].

Таким образом, если найти удовлетворяющее начальным условиям (5.30) численное решение системы (5.29), которое состоит из трех таблично заданных функций $y = y(x)$, $z = z(x)$, $u = u(x)$, то таблица $y = y(x)$ — искомая.

Упражнение

5.22. Дано уравнение $y'' = 3x + y'$ и начальные условия $y(1) = 1$, $y'(1) = 0$. Применяя метод Эйлера, найдите численное решение задачи Коши на отрезке $[1; 2]$ с шагом $h = 0,5$.

Глава 6. СЖИМАЮЩИЕ ОТОБРАЖЕНИЯ И МЕТОД ИТЕРАЦИИ

Одним из важнейших инструментов математического анализа является теорема Банаха, называемая также *принципом сжимающих отображений*. В этой главе теорема используется для обоснования метода итерации приближенного решения уравнений с одним неизвестным и систем уравнений.

6.1. Основные определения

В 1.2 было введено понятие метрического пространства как некоторого множества X с заданной на нем метрикой (расстоянием) ρ . Тем самым метрическое пространство представляет собой пару (X, ρ) , состоящую из множества и расстояния. Изменение любого компонента этой пары означает изменение метрического пространства. Для краткости метрические пространства будем обозначать так же, как и множества, за исключением случаев, когда указание расстояния необходимо для конкретизации пространств.

Так же в 1.2 рассмотрены примеры метрических пространств, среди которых пространство \mathbf{R} действительных чисел с метрикой, определяемой по формуле (1.1), и три пространства n -мерных действительных числовых векторов \mathbf{R}^n с метриками (1.2) – (1.4). Когда необходимо подчеркнуть выбор расстояния в пространствах векторов, пространство \mathbf{R}^n с метрикой ρ_1 будем обозначать через \mathbf{R}_1^n , с метрикой ρ_2 – \mathbf{R}_2^n , с метрикой ρ_∞ – \mathbf{R}_∞^n . В противном случае нижние индексы опускаем.

При $n = 1$ указанные метрические пространства векторов являются не чем иным, как пространством \mathbf{R} , поскольку тогда векторы имеют одну числовую координату, а все эти три расстояния между одномерными векторами $X = (x)$ и $Y = (y)$ равны $|y - x|$.

Пространства \mathbf{R} , \mathbf{R}_1^n , \mathbf{R}_2^n , \mathbf{R}_∞^n являются *основными* в данной главе. Помимо них понадобятся и другие пространства, определяемые на их базе по следующей схеме. Пусть X – метрическое пространство. Взяв какое-либо подмножество множества X , на нем естественно ввести ту же метрику ρ , что имеется на X . Тогда получится новое метрическое пространство, называемое *метрическим подпространством* пространства X .

Например, промежутки $(0; 1)$ и $[0; 3]$ можно считать метрически-ми подпространствами пространства \mathbf{R} , поскольку расстояние между точками этих промежутков также определяется по формуле (1.1). В свою очередь, интервал $(0; 1)$ является метрическим подпространством отрезка $[0; 3]$.

Из курса математического анализа известно понятие функции одной или нескольких числовых переменных, значениями которой являются числа. Аналогично определяется функция $F: M \rightarrow N$ в случае произвольных множеств M и N . При этом вместо термина «функция» часто используются термины «отображение» или «оператор». Если $x \in M$, то элемент $F(x) \in N$ называют *образом* элемента x при отображении F . Как и в случае функций, x можно называть *аргументом*, а $F(x)$ — *значением* отображения F , соответствующим x .

Множество M называется *областью определения* отображения F (обозначим его D_F), а множество $R_F = \{F(x): x \in M\} \subset N$ — *множеством значений* этого отображения.

В данной главе речь пойдет о решении уравнений вида

$$x = F(x), \quad (6.1)$$

где F является отображением, заданным на одном из основных метрических пространств или на каком-либо их подпространстве. Это уравнение имеет смысл, если x и $F(x)$ являются либо числами, либо числовыми векторами одной и той же размерности. Более того, задачи поиска приближенных решений и оценки их погрешностей требуют, чтобы расстояния между аргументами и между соответствующими им значениями измерялись одинаковым образом. Следовательно, *множество значений отображения F должно быть подмножеством того же основного метрического пространства, в котором находится его область определения*. Не может быть, например, уравнения вида (6.1) с таким отображением F , что $D_F \subset \mathbf{R}_1^3$, а $R_F \subset \mathbf{R}_2^3$, или $D_F \subset \mathbf{R}^2$, а $R_F \subset \mathbf{R}$.

В случае, когда F — действительная функция одной переменной, уравнение (6.1) представляет собой уравнение с одним неизвестным. Рассматриваемые в 6.4 и 6.5 приведенные системы из n уравнений с n неизвестными являются частными случаями уравнения (6.1) с отображением типа «вектор \rightarrow вектор».

О п р е д е л е н и е 6.1. Элемент $x = x^* \in D_F$, при котором получается верное равенство

$$x^* = F(x^*),$$

называется *решением (корнем) уравнения (6.1)*¹⁾.

Решение x^* уравнения (6.1) называют также *неподвижной точкой отображения F* , поскольку его образ $F(x^*)$ совпадает с x^* .

¹⁾ Как и в гл. 5, слово «решение» здесь будет иметь два смысла. Во-первых, это элемент соответствующего метрического пространства, удовлетворяющий уравнению (6.1); во-вторых, это процесс поиска такого элемента.

Пример 6.1.1. Функции $f: x \rightarrow x^2 + 0,25$ соответствует уравнение вида (6.1): $x = x^2 + 0,25$. Корень $x = 0,5$ этого уравнения является неподвижной точкой функции f .

2. Уравнение $x = e^x$ не имеет корней, следовательно, у функции $y = e^x$ неподвижных точек нет.

3. При отображении $F: x \rightarrow 2 \cdot x$, действующем в пространстве \mathbf{R}^2 , получим уравнение $x = 2 \cdot x$ (в векторной форме), которое можно записать в виде системы уравнений (в координатной форме)

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

Неподвижной точкой данного отображения является вектор $(0, 0)$.

4. Пусть на \mathbf{R}^3 определено отображение $F: x \rightarrow \rho(\mathbf{0}, x)$, где $\mathbf{0} = (0, 0, 0)$ — нулевой вектор. В этом случае уравнение вида (6.1) составить нельзя, ибо в его левой части будет трехмерный вектор, а в правой — неотрицательное число. Понятно, что у данного отображения не может быть и неподвижной точки. •

Определение 6.2. Пусть X — метрическое пространство. Говорят, что F отображает X в себя, если $F: X \rightarrow X$, т.е. образ $F(x)$ любого элемента $x \in X$ принадлежит этому же пространству.

Отображение, переводящее в себя некоторое основное пространство или его подпространство, обладает и отмеченным выше свойством, при котором имеет смысл уравнение (6.1). Обратное, вообще говоря, неверно. Действительно, если ранее от заданного на множестве X отображения F требовалось лишь, чтобы X и множество его значений $Y = \{F(x): x \in X\}$ находились в одном и том же основном пространстве, то теперь должно быть $Y \subset X$.

Специфику отображения метрического пространства в себя показывает и такой факт: из $F: X \rightarrow X$ не следует, что F отображает в себя каждое подмножество $D \subset X$.

Пример 6.2. Действительные функции $f_1: x \rightarrow 0,5x$ и $f_2: x \rightarrow 2x$ отображают в себя свою область определения \mathbf{R} .

Пусть теперь $X = [-1; 1]$. Поскольку $f_1: X \rightarrow [-0,5; 0,5] \subset X$, можем сказать, что функция f_1 отображает X в себя. А вот функция f_2 не отображает X в себя, поскольку множество ее значений $\{f_2(x): x \in X\}$ равно отрезку $[-2; 2]$. Более того, f_2 не отображает в себя ни один отрезок $[a; b]$, кроме одноточечного отрезка $[0; 0]$ (убедитесь в этом!). •

Центральную роль в дальнейшем будут играть следующие отображения.

Определение 6.3. Пусть X — метрическое пространство. Заданное на X отображение F называется *сжимающим отображением*, или *отображением сжатия* на этом пространстве, если

- 1) F отображает X в себя;
- 2) F удовлетворяет на X условию Липшица:

существует число q , $0 \leq q < 1$ такое, что

$$\rho(F(x_1), F(x_2)) \leq q\rho(x_1, x_2) \text{ для всех } x_1, x_2 \in X. \quad (6.2)$$

Число q при этом называется *коэффициентом сжатия*.

Замечаем, что неравенство (6.2) требует от отображения сжатия более сильного свойства, чем то, при котором образы любых двух различных элементов $x_1, x_2 \in X$ оказывались на меньшем расстоянии друг от друга, чем x_1 от x_2 . Очевидно также, что коэффициент сжатия определяется неоднозначно. Если, например, неравенство (6.2) верно при $q = 0,6$, то оно верно и при любом числе q из полуинтервала $[0,6; 1)$.

Выполнение приведенных в определении 6.3 условий зависит как от особенностей отображения F , так и от пространства, на котором оно рассматривается. По этой причине свойство сжимаемости отображения, имеющее место на X , обычно не распространяется на многие его подпространства. И наоборот, если отображение F обладает этим свойством на каких-то подпространствах X , оно может не оказаться сжимающим на всем X .

Пример 6.3. Проверим, будет ли отображение $f: f(x) = e^{-x}$ сжимающим на нескольких выбранных нами метрических пространствах.

Пусть $X = \mathbf{R}$. Понятно, что функция f отображает \mathbf{R} в себя. Выясним, удовлетворяет ли она условию Липшица. По теореме Лагранжа для любых двух различных точек $x_1, x_2 \in \mathbf{R}$ имеем

$$|f(x_1) - f(x_2)| = |e^{-x_1} - e^{-x_2}| = e^{-c}|x_1 - x_2|, \quad (6.3)$$

где число c лежит между x_1 и x_2 . Если взять $x_1, x_2 \in [-2; -1]$, то, поскольку $e^{-c} > e$ при всех $c \in (-2; -1)$, получим неравенство $|e^{-x_1} - e^{-x_2}| > e|x_1 - x_2|$. Это говорит о нарушении условия Липшица на \mathbf{R} и, следовательно, о том, что данная функция не является сжимающей на \mathbf{R}^1 .

Теперь возьмем пространство $X = [0,2; 1]$. Функция f отображает выбранный отрезок в себя, ибо для $x \in [0,2; 1]$ значения $f(x) \in [e^{-1}; e^{-0,2}] \subset [0,2; 1]$. Используя равенство (6.3) для произвольных x_1, x_2 из этого отрезка, будем иметь

$$|f(x_1) - f(x_2)| \leq e^{-0,2}|x_1 - x_2| \leq 0,82|x_1 - x_2|.$$

Значит, отображение $y = e^{-x}$ на данном пространстве сжимающее, причем коэффициентом сжатия можно считать $q = 0,82$.

На отрезке $[1; 2]$ эта функция не будет сжимающей, хотя и удовлетворяет на нем условию Липшица (проверьте последнее утверждение!). Дело в том, что она не отображает в себя этот отрезок, поскольку ни одно значение $f(x)$, $x \in [1; 2]$, не принадлежит $[1; 2]$. •

Наконец, нам понадобится понятие сходимости последовательности (x_k) элементов метрического пространства X .

¹⁾ Нетрудно убедиться графически, что образы любых двух различных точек из $(-\infty; 0]$ при таком отображении находятся на большем расстоянии друг от друга, чем сами эти точки.

Определение 6.4. Говорят, что последовательность (x_k) *сходится к элементу* $x \in X$ (записывается: $x = \lim_{k \rightarrow \infty} x_k$) *относительно метрики* ρ , если

$$\lim_{k \rightarrow \infty} \rho(x_k, x) = 0. \quad (6.4)$$

Элемент x называется при этом *пределом последовательности* (x_k) *относительно метрики* ρ .

В случае числовых последовательностей условие (6.4) означает: $|x_k - x| \rightarrow 0$ при $k \rightarrow \infty$.

Когда (x_k) — последовательность n -мерных векторов, в обозначении k -го вектора индекс k будем писать сверху: $x^{(k)} = (x_1^{(k)}, \dots, x_n^{(k)})$. Можно показать, что последовательность векторов в рассматриваемых нами метрических пространствах сходится к некоторому вектору x тогда и только тогда, когда все последовательности координат $(x_i^{(k)})$, $i = 1, \dots, n$, сходятся к соответствующим координатам этого вектора (упр. 6.7). Ввиду этого, например,

$$\lim_{k \rightarrow \infty} \left(1 - \frac{1}{k}, \frac{2}{k}, 3 + \frac{1}{k} \right) = (1, 0, 3)$$

относительно любой из трех метрик $\rho_1, \rho_2, \rho_\infty$ на \mathbf{R}^3 .

Упражнения

6.1. Покажите, что функция $y = e^x$ отображает в себя \mathbf{R} , но не отображает в себя ни один отрезок $[a; b]$, где $a < b$.

6.2. Убедитесь, что функция $y = 1 + \ln x$ не отображает в себя область своего определения, однако имеются промежутки, которые данная функция переводит в себя. Найдите наибольший промежуток с таким свойством.

6.3. Найдите множество значений заданного на \mathbf{R}^2 отображения $F: (x_1, x_2) \rightarrow (4x_1, \sin x_2)$ и убедитесь, что F отображает \mathbf{R}^2 в себя.

6.4. Покажите, что отображения:

$$F_1: \mathbf{R} \rightarrow \mathbf{R}, \text{ где } F_1(x) = 4 + 0,3 \sin^2 x;$$

$$F_2: [0; 1] \rightarrow [0; 1], \text{ где } F_2(x) = \frac{1}{x+2};$$

$$F_3: \mathbf{R}_\infty^2 \rightarrow \mathbf{R}_\infty^2, \text{ где } F_3(x) = 0,2 \cdot x$$

являются сжимающими на своих областях определения.

6.5. 1. Покажите, что функция $y = x^2$ не является сжимающим отображением на отрезке $[-1; 1]$.

2. Найдите хотя бы один отрезок, на котором данная функция будет сжимающей.

6.6. Пусть дано отображение $F: x \rightarrow Ax + b$ множества двумерных числовых векторов \mathbf{R}^2 в себя, где

$$A = \begin{pmatrix} 0,2 & 0,5 \\ 0,3 & -0,1 \end{pmatrix}, \mathbf{b} = (-1, 2).$$

1. Напишите формулы вычисления координат вектора $\mathbf{y} = F(\mathbf{x})$ и найдите образ вектора $\mathbf{x} = (1, 2)$ при этом отображении.

2. Напишите соответствующую систему уравнений вида (6.1).

3. Покажите, что F является отображением сжатия на $\mathbf{R}_\infty^2; \mathbf{R}_1^2$.

6.7. Пусть дана последовательность n -мерных числовых векторов $\mathbf{x}^{(k)} = (x_1^{(k)}, \dots, x_n^{(k)})$, $k = 0, 1, 2, \dots$. Докажите, что сходимость последовательности $(\mathbf{x}^{(k)})$ к вектору $\mathbf{x} = (x_1, \dots, x_n)$ в пространствах $\mathbf{R}_1^n, \mathbf{R}_2^n, \mathbf{R}_\infty^n$ равносильна покоординатной сходимости:

$$\lim_{k \rightarrow \infty} \rho(\mathbf{x}^{(k)}, \mathbf{x}) = 0 \Leftrightarrow \lim_{k \rightarrow \infty} |x_i^{(k)} - x_i| = 0 \text{ при всех } i = 1, 2, \dots, n.$$

6.8. Найдите предел последовательности двумерных векторов $\mathbf{x}^{(k)} = \left(\frac{k+1}{k}, \frac{1}{k^2} \right)$, $k = 1, 2, \dots$, и определите номер k , начиная с которого члены $\mathbf{x}^{(k)}$ становятся ближе чем на $\epsilon = 0,05$ к этому пределу относительно расстояний ρ_∞ и ρ_1 .

6.2. Принцип сжимающих отображений

Приведем в общем виде один из основных способов приближенного решения рассмотренного в 6.1 уравнения (6.1): $x = F(x)$. Как известно из 2.1, приближенное решение уравнений выполняется в два этапа: отделение решений и поиск их приближений с требуемой степенью точности. Первый этап обязателен в том случае, когда кроме искомого решения есть и другие.

Для отделения решений системы из n целлинейных уравнений с n неизвестными (см. 6.5) используем множества $D \subset \mathbf{R}^n$:

$$D = \{(x_1, \dots, x_n): a_1 \leq x_1 \leq b_1, \dots, a_n \leq x_n \leq b_n\}, \quad (6.5)$$

где $a_1 < b_1, \dots, a_n < b_n$. Для уравнений с одним неизвестным множеством вида (6.5) является отрезок $D = [a; b]$, $a < b$. Разумеется, что множество D берется из области определения отображения F .

Соглашение 6.1. *Далее под метрическим пространством X будем подразумевать либо одно из пространств $\mathbf{R}, \mathbf{R}_1^n, \mathbf{R}_2^n, \mathbf{R}_\infty^n$, либо некоторое их метрическое подпространство вида (6.5).*

Теперь сформулируем принцип сжимающих отображений.

Теорема 6.1 (Банаха). *Пусть отображение F является сжимающим на полном¹⁾ метрическом пространстве с коэффициентом сжатия $0 \leq q < 1$. Тогда:*

¹⁾ Здесь не будем раскрывать суть понятия полноты метрических пространств, отметим лишь, что указанные в соглашении 6.1 пространства обладают этим свойством.

1) на X существует одно и только одно решение x^* уравнения (6.1) (единственная неподвижная точка отображения F);

2) решение x^* равно пределу итерационной последовательности (x_n) точек из X , определяемой рекуррентной формулой

$$x_{n+1} = F(x_n) \quad (n = 0, 1, \dots), \quad (6.6)$$

где x_0 — произвольный элемент из X ;

3) при каждом $n = 1, 2, \dots$ справедливо неравенство

$$\rho(x_n, x^*) \leq \frac{q^n}{1-q} \rho(x_1, x_0) \quad (6.7)$$

([12, гл. II, § 4], [9, гл. 12, дополнение 2]).

В примере 6.3 показано, что на отрезке $[0, 2; 1]$ функция $y = e^{-x}$ удовлетворяет условиям теоремы Банаха. Следовательно, на этом отрезке имеется один и только один корень уравнения $x = e^{-x}$ (проиллюстрируйте это графически), который можно найти как предел последовательности $x_{n+1} = e^{-x_n}$ ($n = 0, 1, \dots$) с произвольным x_0 из этого отрезка.

Если пространство X из теоремы 6.1 является только частью области определения отображения F , нельзя утверждать, что уравнение (6.1) имеет всего одно решение. Кроме решения $x^* \in X$ могут быть и другие решения в множестве $D_F \setminus X$.

В доказательстве неравенства (6.7) показатель n степени q^n появляется после n шагов, требующихся для вычисления x_n . Если предыдущее для x_n приближение x_{n-1} считать начальным, из (6.7) получится еще одно неравенство:

$$\rho(x_n, x^*) \leq \frac{q}{1-q} \rho(x_n, x_{n-1}) \quad (n = 1, 2, \dots). \quad (6.8)$$

Итак, в условиях теоремы Банаха имеем следующее. Во-первых, уравнение (6.1) обязательно имеет решение $x^* \in X$, причем единственное на этом множестве. Во-вторых, формула (6.6) определяет последовательность приближений к решению x^* , при этом за начальное приближение x_0 можно взять любой элемент из X . В-третьих, если решение x^* заменить приближенным значением x_n , то абсолютную погрешность приближения $x_n^{(1)}$ можно вычислить по одной из следующих формул:

$$\Delta_{x_n} = \frac{q^n}{1-q} \rho(x_1, x_0); \quad \Delta_{x_n} = \frac{q}{1-q} \rho(x_n, x_{n-1}). \quad (6.9)$$

¹⁾ Понятия абсолютных погрешностей приближенных чисел и векторов даны в параграфах 1.3 и 1.6.

Из последних соотношений следует, что для вычисления приближенного решения x_n с точностью до $\varepsilon > 0$ необходимо добиться выполнения хотя бы одного из неравенств

$$\frac{q^n}{1-q} \rho(x_1, x_0) \leq \varepsilon \text{ или } \frac{q}{1-q} \rho(x_n, x_{n-1}) \leq \varepsilon. \quad (6.10)$$

По неравенству (6.7) видно, что от величины коэффициента сжатия зависит скорость сходимости итерационной последовательности. Чем меньше q , тем быстрее можно достичь желаемой точности приближенного решения.

Учитывая все сказанное, заключаем, что теорема Банаха фактически дает метод уточнения приближенного решения уравнения (6.1), который называется *методом итерации*, или *методом простой итерации*.

Сделаем несколько важных замечаний, характеризующих метод итерации и условия его применения¹⁾.

Как и любой другой метод приближенного решения уравнений, метод итерации не универсален. Он применим лишь к уравнениям вида (6.1) и гарантирует успех в случае, когда F является отображением сжатия на полном метрическом пространстве X . В то же время он широко используется для решения не только рассматриваемых нами уравнений и систем уравнений, но и различных классов дифференциальных и интегральных уравнений и т. п. Привлекательность его объясняется тем, что он порождает процесс однотипных вычислений и нечувствителен к небольшим ошибкам. Если, к примеру, при отыскании k -го приближения была допущена ошибка и вместо x_k получили x'_k , однако при этом x'_k остался в множестве X , то итерационный процесс можно продолжать, он все равно приблизит нас к точному решению x^* .

В конкретных задачах перед применением метода итерации обычно уже известно, что решение уравнения (6.1) на множестве X существует и оно там единственное. Тогда главным смыслом накладываемых теоремой 6.1 требований на отображение F (отображать X в себя и удовлетворять на X условию Липшица) становится обеспечение сходимости итерационной последовательности к решению.

Благодаря свойству F быть отображением пространства X в себя, приближения x_n ни на каком шаге не выходят за пределы X , где выполняется условие Липшица. Однако в случае, когда условие Липшица верно в некотором более широком подпространстве $X' \supset X$ основного пространства, нарушение этого свойства не всегда приводит к расходящемуся итерационному процессу. Если какие-то x_n выходят за пределы X , оставаясь при этом в X' , то последующие члены последовательности могут снова оказаться в X и продолжать при-

¹⁾ Они во многом аналогичны тем, что были приведены в п. 2.8.2.

ближаться к x^* . Здесь многое зависит от выбора начального приближения (упр. 6.16).

Когда отображение не удовлетворяет условию Липшица, возможны разные ситуации относительно существования решения и сходимости итерационного процесса. Если образы любых двух точек из X оказываются на расстоянии, не меньшем, чем сами точки, т. е.

$$\rho((F(x_1), F(x_2))) \geq \rho(x_1, x_2) \text{ для любых } x_1, x_2 \in X, \quad (6.11)$$

то итерационная последовательность заведомо не сходится к x^* (упр. 6.10). При замене неравенства (6.2) на более слабое

$$\rho(F(x_1), F(x_2)) < \rho(x_1, x_2) \text{ для всех } x_1, x_2 \in X, x_1 \neq x_2, \quad (6.12)$$

утверждения теоремы Банаха тоже могут оказаться неверными. Так, свойство F быть отображением X в себя и неравенство (6.12) в совокупности не гарантируют существование решения на X (упр. 6.11).

Таким образом, если F отображает X в себя и на X верно неравенство (6.2), то итерационный процесс сходится; если же имеет место неравенство (6.11), то процесс расходится независимо от того, отображается X в себя, или нет. Интересен следующий вопрос: *как ведет себя итерационная последовательность в промежуточных случаях [в частности, когда вместо неравенства (6.2) выполняется (6.12)]?* Ответ неоднозначен (выполните упр. 6.12 и 6.13).

Упражнения

6.9. 1. Покажите, что функция $f: x \rightarrow 0,5x$ удовлетворяет условиям теоремы Банаха на любом отрезке $[-a; a]$ ($a > 0$), и выясните смысл утверждений этой теоремы в данном случае.

2. Выберите конкретный отрезок $[-a; a]$, возьмите $x_0 = a$ и найдите число x_3 , а затем вычислите Δ_{x_3} с помощью формул (6.9).

6.10. Пусть решение x^* уравнения (6.1) находится в метрическом пространстве X и выполняется неравенство (6.11). Докажите, что в этом случае при начальном приближении $x_0 \in X$, $x_0 \neq x^*$, итерационная последовательность (6.6) не сходится к x^* .

6.11. Покажите, что для отображающей $X = \mathbf{R}$ в себя функции $f: x \rightarrow \frac{\pi}{2} + x - \operatorname{arctg} x$ верно неравенство (6.12), между тем неподвижных точек у этого отображения нет.

6.12. Как отмечено после формулировки теоремы 6.1, функция $f: x \rightarrow e^{-x}$ из примера 6.3 имеет неподвижную точку x^* в отрезке $[0, 2; 1]$. Убедитесь, что:

1) f отображает в себя $X = [0; 1]$, вместо (6.2) на X имеет место неравенство (6.12), но итерационная последовательность сходится к x^* при любом начальном приближении из этого отрезка;

2) f не отображает в себя $X = [-1; 1]$, на этом отрезке не выполняется ни одно из неравенств (6.2), (6.11), (6.12), однако итерационная последовательность снова сходится к x^* при любом $x_0 \in [-1; 1]$.

6.13. Заданное на \mathbf{R} отображение f такое, что

$$f(x) = \begin{cases} 0,5x + 0,5, & x \leq 1; \\ x^2, & x > 1, \end{cases}$$

имеет одну неподвижную точку $x = 1$. Покажите, что f не отображает в себя $X = [0; 2]$, на этом отрезке не выполняется ни одно из неравенств (6.2), (6.11), (6.12), и при этом (в отличие от имеющей место в задании 2 упр. 6.12 ситуации) сходимость итерационного процесса зависит от выбора начального приближения $x_0 \in X$.

У к а з а н и е. Для выявления факта сходимости или расходимости итерационной последовательности в упр. 6.12 и 6.13 используйте соответствующие методу простой итерации геометрические построения (см. 2.8).

6.3. Метод итерации для уравнений с одним неизвестным

Убедимся, что изложенный в 2.8 метод простой итерации уточнения корней уравнений с одним неизвестным является частным случаем описанного в 6.2 итерационного процесса.

Рассмотрим уравнение

$$x = g(x) \tag{6.13}$$

с действительной функцией g . Корень t уравнения (6.13) является неподвижной точкой отображения $g: x \rightarrow g(x)$. В качестве метрического пространства X берем \mathbf{R} или отрезок $[a; b]$ из области определения функции g (например, отрезок изоляции корня t).

Согласно определению 6.3, функция g сжимающая на X с коэффициентом сжатия q , $0 \leq q < 1$, если $f: X \rightarrow X$ и

$$|g(x_1) - g(x_2)| \leq q|x_1 - x_2| \text{ при всех } x_1, x_2 \in X. \tag{6.14}$$

Следующая теорема соответствует теореме 6.1. В ней дополнительно приводится достаточно легко проверяемое свойство функции g , при наличии которого g удовлетворяет условию Линшица (6.14).

Теорема 6.2. Пусть:

- 1) функция g отображает X в себя;
- 2) функция g дифференцируема на X и существует число q , $0 \leq q < 1$, такое, что

$$|g'(x)| \leq q \text{ при всех } x \in X. \tag{6.15}$$

Тогда в X существует единственный корень t уравнения (6.13), который можно найти как предел последовательности (x_n) точек из X , определяемой рекуррентной формулой

$$x_{n+1} = g(x_n) \quad (n = 0, 1, \dots) \quad (6.16)$$

с любым начальным приближением $x_0 \in X$. При этом погрешности приближений $x_n \approx t$ ($n = 1, 2, \dots$) оцениваются по формулам:

$$|t - x_n| \leq \frac{q^n}{1 - q} |x_1 - x_0|, \quad (6.17)$$

$$|t - x_n| \leq \frac{q}{1 - q} |x_n - x_{n-1}|. \quad (6.18)$$

○ В силу теоремы Банаха для доказательства следует лишь убедиться в том, что из неравенства (6.15) следует неравенство (6.14). Действительно, пусть x_1, x_2 — любые две точки из X . Тогда, используя теорему Лагранжа и соотношение (6.15), получим

$$|g(x_1) - g(x_2)| = |g'(c)| |x_1 - x_2| \leq q |x_1 - x_2|,$$

где c — некоторое число, лежащее между x_1 и x_2 . •

Таким образом, ссылка на теорему Банаха позволила доказать новое (по сравнению с теоремой 2.16 из п. 2.8.2) достаточное условие применимости метода итерации для приближенного решения уравнений с одним неизвестным и получить формулы оценки погрешностей приближений, которые были независимо выведены в 2.8 (см. теорему 2.16, доказываемую в ее условиях теорему 2.18 и упр. 2.41).

Сравнивая теорему 6.2 с полученными в 2.8 результатами, полезно обратить внимание на следующее. В отличие от теоремы 6.2 в теореме 2.16 не требуется, чтобы функция g отображала в себя какой-либо отрезок: выполнение неравенства (6.15) на отрезке $[c; d]$ «тройной длины» и выбор начального приближения из «серединного отрезка» $[a; b]$ обеспечивают принадлежность итерационной последовательности отрезку $[c; d]$ и ее сходимости к корню $t \in [a; b]$. Следовательно, условия параграфа 2.8 являются более простыми и удобными для проверки.

Но, с другой стороны, результаты, приведенные в 2.8, получены лишь для случая, когда метрическим пространством X является отрезок, что существенно сужает возможности их применения. Теперь дано обоснование метода и при $X = \mathbf{R}$.

С учетом последнего факта получим следствие из теоремы 6.2.

С л е д с т в и е 6.1. Если функция g определена и дифференцируема на \mathbf{R} и ее производная g' обладает свойством (6.15) на всем \mathbf{R} , то уравнение (6.13) имеет один-единственный корень, а итерационная последовательность (6.16) сходится к нему, какое бы число $x_0 \in \mathbf{R}$ ни было взято в качестве начального приближения.

$$\begin{vmatrix} a_{11} & \dots & a_{1n} \\ \dots & \dots & \dots \\ a_{n1} & \dots & a_{nn} \end{vmatrix}$$

системы не равен нулю. Как известно, это обеспечивает существование и единственность решения на всем \mathbf{R}^n .

В курсе алгебры изучается ряд точных методов решения системы (6.19). Наиболее известными среди них являются правило Крамера и метод Гаусса. Однако при большом числе неизвестных схемы точного решения становятся сложными и требующими большого объема вычислений. В таких случаях обычно используются различные приближенные (итерационные) методы.

Чтобы воспользоваться принципом сжимающих отображений и построить итерационный процесс приближения к точному решению, систему (6.19) предварительно приводят к эквивалентной системе¹⁾ вида

$$\begin{cases} x_1 = c_{11}x_1 + \dots + c_{1n}x_n + d_1, \\ x_2 = c_{12}x_1 + \dots + c_{2n}x_n + d_2, \\ \dots \\ x_n = c_{n1}x_1 + \dots + c_{nn}x_n + d_n. \end{cases} \quad (6.20)$$

Назовем систему (6.20) *приведенной*. Обозначив

$$C = \begin{pmatrix} c_{11} & \dots & c_{1n} \\ \dots & \dots & \dots \\ c_{n1} & \dots & c_{nn} \end{pmatrix}, \quad \mathbf{x} = (x_1, \dots, x_n), \quad \mathbf{d} = (d_1, \dots, d_n),$$

ее можно записать в матричной форме:

$$\mathbf{x} = C\mathbf{x} + \mathbf{d}. \quad (6.20')$$

Покажем два способа преобразования системы (6.19) к виду (6.20).

Пример 6.4. Пусть дана система общего вида:

$$\begin{cases} 5x_1 - x_2 + x_3 = 4, \\ 0,5x_1 + 2x_2 - x_3 = 1, \\ x_1 - x_2 + 4x_3 = 2. \end{cases} \quad (*)$$

Способ 1. Поделив уравнения на соответствующие диагональные коэффициенты (они не равны нулю) и решив каждое из уравнений относительно полученных на диагонали неизвестных с коэффициентом 1, получим приведенную систему

¹⁾ Системы уравнений эквивалентны, если они имеют одинаковые множества решений.

носильна по координатной сходимости. Значит, если удастся построить последовательность приближений к \bar{x} на основе одного из расстояний, то эта последовательность будет сходиться к \bar{x} и относительно остальных двух расстояний.

Пусть далее $X = \mathbf{R}_\infty^n$, т.е. предполагаем, что расстояние между векторами $x = (x_1, \dots, x_n)$ и $y = (y_1, \dots, y_n)$ вычисляется по формуле

$$\rho(x, y) = \max_{i=1, \dots, n} |x_i - y_i|. \quad (6.24)$$

Теорема 6.3. Если матрица C коэффициентов при неизвестных в правой части системы (6.20) удовлетворяет условию

$$q = \max_{i=1, \dots, n} \sum_{j=1}^n |c_{ij}| < 1, \quad (6.25)$$

то эта система имеет единственное решение \bar{x} , а построенная по формулам (6.23) последовательность $(x^{(k)})$ сходится к \bar{x} при любом начальном приближении $x^{(0)} \in X$. Для оценки погрешности приближенного вектора $x^{(k)} \approx \bar{x}$ ($k = 1, 2, \dots$) имеют место формулы

$$\rho(x^{(k)}, \bar{x}) \leq \frac{q^n}{1-q} \max_i |x_i^{(1)} - x_i^{(0)}|, \quad (6.26)$$

$$\rho(x^{(k)}, \bar{x}) \leq \frac{q}{1-q} \max_i |x_i^{(k)} - x_i^{(k-1)}|. \quad (6.27)$$

○ Чтобы сослаться на теорему Банаха, надо убедиться, что указанное в (6.21) отображение F является сжимающим на X при выполнении условия (6.25). То, что F переводит X в себя, уже отмечалось. Проверим условие Липшица.

Пусть $x = (x_1, \dots, x_n)$ и $x' = (x'_1, \dots, x'_n)$ — два произвольных вектора из X . Обозначим $y = F(x)$, $y' = F(x')$. Координаты этих векторов вычисляются по формулам

$$y_i = \sum_{j=1}^n c_{ij} x_j + d_i, \quad y'_i = \sum_{j=1}^n c_{ij} x'_j + d_i \quad (i = 1, 2, \dots, n).$$

Для разностей соответствующих координат с учетом свойств модуля получим неравенства

$$|y_i - y'_i| \leq \sum_{j=1}^n |c_{ij}| |x_j - x'_j| \quad (i = 1, 2, \dots, n).$$

Поскольку $|x_j - x'_j| \leq \rho(x, x')$ при всех $j = 1, \dots, n$,

$$|y_i - y'_i| \leq \sum_{j=1}^n |c_{ij}| \rho(x, x') = \rho(x, x') \sum_{j=1}^n |c_{ij}| \quad (i = 1, 2, \dots, n). \quad (6.28)$$

Теперь заменим в (6.28) суммы на число q из (6.25):

$$|y_i - y'_i| \leq q \cdot \rho(x, x') \quad (i = 1, 2, \dots, n). \quad (6.29)$$

В силу справедливости неравенств (6.29) при всех $i = 1, 2, \dots, n$, имеем

$$\rho(y, y') = \max_i |y_i - y'_i| \leq q \cdot \rho(x, x').$$

Следовательно, условие Липшица выполняется, а число q из соотношения (6.25) является коэффициентом сжатия отображения F .

Формулы (6.26) и (6.27) вытекают из (6.7) и (6.8), поскольку расстояние в нашем случае определено по (6.24). •

Условие (6.25) означает, что наибольшая из сумм модулей коэффициентов при неизвестных в правой части системы (6.20), вычисленных по каждой строке, должна быть меньше единицы. Разумеется, что это возможно лишь тогда, когда все элементы матрицы C достаточно малы по модулю. Например, система (**) из примера 6.4 удовлетворяет условию (6.25), так как для нее

$$q = \max_{i=1,2,3} \sum_{j=1}^3 |c_{ij}| = 0,75 < 1$$

(проверьте!), поэтому можно быть уверенным, что последовательность векторов из примера 6.5 сходится к решению. А вот элементы матрицы C системы (***) оказались настолько большими, что уже первая ее строка дает: $|-4| + |1| + |-1| = 6$.

Существуют разные приемы преобразования системы (6.19) к виду (6.20) с выполнением свойства (6.25). Среди них показанные в примере 6.4 способы, возможность применения которых зависит от коэффициентов при неизвестных исходной системы.

Для применения первого способа необходимо, чтобы диагональные коэффициенты системы (6.19) не равнялись нулю и были намного больше по модулю остальных коэффициентов соответствующих уравнений. В примере 6.4 имеет место такая ситуация.

Когда диагональные коэффициенты близки к единице, а остальные коэффициенты малы по модулю, могут подойти оба способа. Но при этом второй предпочтительнее, поскольку он не связан с делением. Если коэффициенты исходной системы были точными, то при втором способе такими же будут и коэффициенты приведенной системы. Деление же обычно сопряжено с округлением, значит и с проблемой определения количества оставляемых цифр, необходимого для достижения требуемой точности приближенного решения.

Эти способы можно применять как в целом к системе, так и к отдельным уравнениям, переставляя, при необходимости, уравнения системы.

Пример 6.6. Пусть требуется преобразовать систему

$$\begin{cases} 0,4x + 0,7y = 1,1, \\ 5x - y = 4 \end{cases}$$

к приведенному виду с выполнением условия (6.25). Здесь пока ни одно из предварительных условий применения указанных выше способов преобразования не имеет места. Переставим уравнения:

$$\begin{cases} 5x - y = 4, \\ 0,4x + 0,7y = 1,1. \end{cases}$$

Теперь видно, что к первому уравнению можно применить первый способ, а ко второму — второй (диагональный коэффициент 0,7 второго уравнения близок к единице и представляется в виде разности $1 - 0,3$). После преобразований будем иметь

$$\begin{cases} x = 0,2y + 0,8, \\ y = -0,4x + 0,3y + 1,1. \end{cases}$$

Правая часть полученной системы задает сжимающее отображение на \mathbf{R}_∞^2 с коэффициентом сжатия $q = 0,7$. •

Поскольку начальным приближением может быть любой вектор, часто в качестве $x^{(0)}$ берется вектор $\mathbf{d} = (d_1, \dots, d_n)$ свободных членов системы (6.20). Однако для более быстрой сходимости итерационного процесса целесообразно в качестве начального вектора выбирать приближенное решение системы, найденное грубой прикидкой.

Вычисляя последовательно векторы-приближения $x^{(k)}$ к точному решению \bar{x} , на каждом шаге можно найти их абсолютную погрешность по правым частям неравенств (6.26) и (6.27). Например,

$$\Delta_{x^{(k)}} = \frac{q}{1-q} \max_i |x_i^{(k)} - x_i^{(k-1)}|.$$

Взяв $x^{(k)} \approx \bar{x}$, получим приближения $x_i^{(k)}$ соответствующих координат \bar{x}_i ($i = 1, 2, \dots, n$) искомого решения. Число $\Delta_{x^{(k)}}$ будет также абсолютной погрешностью координат $x_i^{(k)}$, $i = 1, 2, \dots, n$ (см. упр. 1.10).

Перепишав неравенства (6.10) с конкретным расстоянием (6.24), найдем условия окончания итерационного процесса при поиске решения с точностью до $\epsilon > 0$ (предоставляем сделать это читателю!).

З а м е ч а н и е. Ранее предполагалось рассматривать лишь системы с единственными решениями. В этом случае смысл неравенства (6.25) заключается только в том, что оно дает *достаточное условие сходимости итерационной последовательности к решению системы*. Между тем это условие не единственное, ибо оно обусловлено выбором расстояния. Построенная по формулам (6.23) последовательность также оказывается сходящейся к точному решению, причем снова при произвольном начальном приближении из \mathbf{R}^n , если матрица C удовлетворяет хотя бы одному из приведенных в (6.30) условий (упр. 6.22).

Упражнения

6.19. Опишите последовательность всех действий, необходимых для уточнения решения системы (6.19) методом простой итерации с заданной точностью $\epsilon > 0$.

6.20. Определите абсолютные погрешности найденных в примере 6.5 векторов $x^{(1)}$ и $x^{(2)}$.

6.21. Даны две системы:

$$\begin{cases} 1,1x_1 - 0,3x_2 + 0,2x_3 = 2, \\ 0,1x_1 - 0,1x_2 - 1,2x_3 = 0, \\ 0,4x_1 + 0,8x_2 - 0,1x_3 = 3; \end{cases} \quad \begin{cases} -8x_2 - x_3 = 0, \\ 5x_1 + x_2 + x_3 = -1, \\ -x_1 + x_2 - 4x_3 = -2. \end{cases}$$

Пробразуйте их к виду (6.20) с выполнением условия (6.25).

6.22.* Сформулируйте и докажите аналоги теоремы 6.3 в случаях, когда в качестве X взяты метрические пространства \mathbf{R}_1^n или \mathbf{R}_2^n .

У к а з а н и е. При замене расстояния ρ_∞ на ρ_1 или ρ_2 вместо условия (6.25) следует брать неравенства

$$q = \max_{j=1, \dots, n} \sum_{i=1}^n |c_{ij}| < 1 \quad \text{или} \quad q = \sqrt{\sum_{i=1}^n \sum_{j=1}^n c_{ij}^2} < 1 \quad (6.30)$$

соответственно.

6.23. Достаточное условие сходимости итерационной последовательности для системы (6.20) можно выразить через коэффициенты при неизвестных исходной системы (6.19). Докажите, что итерационный процесс сходится, если выполнено хотя бы одно из неравенств

$$|a_{ii}| > \sum_{j \neq i} |a_{ij}| \quad (i = 1, \dots, n) \quad \text{или} \quad |a_{jj}| > \sum_{i \neq j} |a_{ij}| \quad (j = 1, \dots, n).$$

6.5. Метод итерации для систем нелинейных уравнений

Решение систем уравнений вида

$$\begin{cases} \Phi_1(x_1, \dots, x_n) = 0, \\ \dots \dots \dots \\ \Phi_n(x_1, \dots, x_n) = 0, \end{cases}$$

где Φ_1, \dots, Φ_n — нелинейные действительные функции, в большинстве случаев сопряжено с серьезными трудностями. Для них не существует общих схем точного решения, таких, как методы Крамера и Гаусса для систем линейных алгебраических уравнений, поэтому обычно их решают приближенно различными итерационными методами. Описанный в 6.2 общий метод итерации применим и здесь,

причем основные рассуждения и выкладки во многом аналогичны линейному случаю.

Для простоты ограничимся рассмотрением системы из двух уравнений с двумя неизвестными

$$\begin{cases} \Phi_1(x, y) = 0, \\ \Phi_2(x, y) = 0, \end{cases} \quad (6.31)$$

в которой хотя бы одна из функций Φ_1, Φ_2 не является линейной.

Двумерный числовой вектор $\mathbf{t} = (x, y)$ назовем *решением* системы (6.31), если при подстановке его координат оба уравнения системы превращаются в верные равенства. В отличие от линейных систем здесь координатами решения могут быть комплексные числа. Например, система

$$\begin{cases} x + y^2 + 1 = 0, \\ x - y = 0 \end{cases}$$

имеет два комплексных решения (найдите их!).

Впредь нас будут интересовать только решения, принадлежащие пространству \mathbf{R}^2 , т. е. далее под словом «решение» будем подразумевать «действительное решение».

Решение системы (6.31) называется *изолированным*, если найдется такой *открытый прямоугольник* $\{(x, y): a < x < b, c < y < d\}$, в котором это решение единственное. Предположим, что система (6.31) имеет изолированные решения. Ставится задача нахождения приближений к ним методом итерации.

Для этого систему (6.31) представим в форме *приведенной системы*

$$\begin{cases} x = f_1(x, y), \\ y = f_2(x, y), \end{cases} \quad (6.32)$$

по так, чтобы эти системы были эквивалентны. Здесь f_1, f_2 — действительные функции двух переменных с областями определения D_{f_1} и D_{f_2} из \mathbf{R}^2 . Полагаем, что множество $D_{f_1} \cap D_{f_2}$ — *область определения системы* — не является пустым.

Пара функций из правой части системы (6.32) задает отображение

$$F: (x, y) \rightarrow (f_1(x, y), f_2(x, y)), \quad (6.33)$$

переводящее каждый двумерный вектор $\mathbf{t} = (x, y)$ из $D_F = D_{f_1} \cap D_{f_2}$ в единственный двумерный вектор $\mathbf{z} = F(\mathbf{t})$ с координатами $z_1 = f_1(x, y)$ и $z_2 = f_2(x, y)$. Следовательно, область определения и множество значений отображения F находятся в \mathbf{R}^2 . Любое решение $\bar{\mathbf{t}} = (\bar{x}, \bar{y})$ системы (6.32) является неподвижной точкой этого отображения, поскольку равенства $\bar{x} = f_1(\bar{x}, \bar{y}), \bar{y} = f_2(\bar{x}, \bar{y})$ и означают то, что

$\bar{t} = F(\bar{t})$. Рекуррентное соотношение (6.6) в данном случае превращается в формулы

$$x_{n+1} = f_1(x_n, y_n), \quad y_{n+1} = f_2(x_n, y_n) \quad (n = 0, 1, \dots), \quad (6.34)$$

с помощью которых последовательно находятся координаты членов итерационной последовательности $(t_n) = ((x_n, y_n))$.

Как и в случае нелинейных уравнений с одним неизвестным, перед уточнением решений системы (6.32) производится их отделение графическим методом. Для этого на плоскости xOy вычерчиваем кривые, задаваемые уравнениями системы. Координаты точек пересечения кривых дают решение. Затем для каждой изолированной точки пересечения подбираем такой содержащий эту точку замкнутый прямоугольник $D \subset D_{f_1} \cap D_{f_2}$:

$$D = \{(x, y): a \leq x \leq b, c \leq y \leq d\} \quad (a < b, c < d), \quad (6.35)$$

который не содержит других решений системы.

Пример 6.7. Представим систему уравнений

$$\begin{cases} 2y + x - 4 = 0, \\ y - e^x = 0 \end{cases} \quad (+)$$

в форме (6.32), отделим ее решения и построим соответствующую итерационную последовательность с каким-либо начальным вектором.

Наиболее очевидным вариантом приведенной системы здесь может быть такой:

$$\begin{cases} x = 4 - 2y, \\ y = e^x. \end{cases} \quad (++)$$

Нарисовав линии $y = e^x$ и $x = 4 - 2y$, нетрудно увидеть, что они пересекаются в единственной точке, расположенной внутри квадрата $D_0 = \{(x, y): 0 \leq x \leq 1, 1 \leq y \leq 2\}$.

Системе $(++)$ соответствуют рекуррентные формулы

$$\begin{cases} x_{n+1} = 4 - 2y_n, \\ y_{n+1} = e^{x_n}, \end{cases}$$

пользуясь которыми и найдем несколько членов итерационной последовательности, взяв, например, в качестве t_0 вершину $(1, 1)$ квадрата D_0 : $t_0 = (1, 1)$. Координаты получаемых векторов округляем до двух цифр после десятичной запятой.

Вычислим координаты t_1 : $x_1 = 4 - 2 \cdot 1 = 2$, $y_1 = e^1 \approx 2,72$. Значит, берем $t_1 = (2, 2,72)$. Продолжая подобным образом, находим: $t_2 = (-1,44, 7,39)$, $t_3 = (-10,78, 0,24)$ и т. д. •

Замечаем, что точки составленной для системы $(++)$ итерационной последовательности сразу же вышли из квадрата изоляции

решения, хотя начальная точка была взята из этого квадрата. Вызывает сомнение и сходимость этой последовательности к решению. Ниже увидим, что данная ситуация не случайна, поскольку система $(++)$ не удовлетворяет достаточному условию сходимости итерационного процесса.

Для получения варианта теоремы Банаха, соответствующего системе (6.32), возьмем в качестве метрического пространства X замкнутый прямоугольник (6.35) с введенной на нем метрикой $\rho = \rho_\infty$. При этом нам потребуется следующий двумерный аналог теоремы Лагранжа.

Теорема 6.4. Пусть функция $z = f(x, y)$ непрерывна и имеет непрерывные частные производные на прямоугольнике D . Тогда для любых двух точек (x_1, y_1) и (x_2, y_2) из D найдется такая точка $(x_c, y_c) \in D$ с координатами, лежащими между x_1 и x_2 , y_1 и y_2 соответственно, что

$$f(x_1, y_1) - f(x_2, y_2) = \frac{\partial f(x_c, y_c)}{\partial x} (x_1 - x_2) + \frac{\partial f(x_c, y_c)}{\partial y} (y_1 - y_2). \quad (6.36)$$

Теорема 6.4 следует из формулы Тейлора для функций двух переменных [4, гл. XVI, § 1].

Теорема 6.5. Пусть решение $\bar{t} = (\bar{x}, \bar{y})$ системы (6.32) изолировано в пространстве $X = D$. Если:

1) функции f_1 и f_2 такие, что определяемое ими отображение F переводит X в себя;

2) функции f_1, f_2 имеют непрерывные частные производные на X , и существует число $q, 0 \leq q < 1$, такое, что при всех $(x, y) \in X$ верны неравенства

$$\begin{aligned} \left| \frac{\partial f_1(x, y)}{\partial x} \right| + \left| \frac{\partial f_1(x, y)}{\partial y} \right| &\leq q, \\ \left| \frac{\partial f_2(x, y)}{\partial x} \right| + \left| \frac{\partial f_2(x, y)}{\partial y} \right| &\leq q, \end{aligned} \quad (6.37)$$

то итерационная последовательность, порожденная формулами (6.34), сходится к решению \bar{t} при любом начальном приближении $t_0 = (x_0, y_0) \in X$. Для оценки погрешности приближения $t_n \approx \bar{t}$ ($n = 1, 2, \dots$) имеют место формулы

$$\rho(\bar{t}, t_n) \leq \frac{q^n}{1 - q} \max \{|x_1 - x_0|, |y_1 - y_0|\}, \quad (6.38)$$

$$\rho(\bar{t}, t_n) \leq \frac{q}{1 - q} \max \{|x_n - x_{n-1}|, |y_n - y_{n-1}|\}. \quad (6.39)$$

○ Для доказательства покажем, что при выполнении условия (2) теоремы отображение F удовлетворяет на X условию Липшица с коэффициентом сжатия q .

Возьмем два произвольных вектора $\mathbf{t} = (x, y)$ и $\mathbf{t}' = (x', y')$ из X и обозначим

$$\mathbf{z} = F(\mathbf{t}) = (f_1(x, y), f_2(x, y)), \mathbf{z}' = F(\mathbf{t}') = (f_1(x', y'), f_2(x', y')).$$

Расстояние между \mathbf{z} и \mathbf{z}' в нашем случае выразится формулой

$$\rho(\mathbf{z}, \mathbf{z}') = \max \{|f_1(x, y) - f_1(x', y')|, |f_2(x, y) - f_2(x', y')|\}. \quad (6.40)$$

Оценим первый модуль из правой части равенства (6.40). Применяя к функции f_1 теорему 6.4 с соответствующей точкой $(x_c, y_c) \in X$ и учитывая свойства модуля, получим

$$\begin{aligned} |f_1(x, y) - f_1(x', y')| &= \left| \frac{\partial f_1(x_c, y_c)}{\partial x} (x - x') + \frac{\partial f_1(x_c, y_c)}{\partial y} (y - y') \right| \leq \\ &\leq \left| \frac{\partial f_1(x_c, y_c)}{\partial x} \right| |x - x'| + \left| \frac{\partial f_1(x_c, y_c)}{\partial y} \right| |y - y'|. \end{aligned}$$

Ввиду того что $\rho(\mathbf{t}, \mathbf{t}') = \max \{|x - x'|, |y - y'|\}$, можем заменить модули разностей между координатами на расстояние $\rho(\mathbf{t}, \mathbf{t}')$:

$$|f_1(x, y) - f_1(x', y')| \leq \left(\left| \frac{\partial f_1(x_c, y_c)}{\partial x} \right| + \left| \frac{\partial f_1(x_c, y_c)}{\partial y} \right| \right) \rho(\mathbf{t}, \mathbf{t}').$$

Используя первое из условий (6.37), имеем

$$|f_1(x, y) - f_1(x', y')| \leq q \cdot \rho(\mathbf{t}, \mathbf{t}').$$

Аналогично выводится оценка для второго модуля из (6.40)

$$|f_2(x, y) - f_2(x', y')| \leq q \cdot \rho(\mathbf{t}, \mathbf{t}').$$

В итоге получаем требуемое неравенство из условия Липшица:

$$\rho(\mathbf{z}, \mathbf{z}') \leq q \cdot \rho(\mathbf{t}, \mathbf{t}'),$$

что вместе с условием 1) даст возможность закончить доказательство ссылкой на теорему Банаха. В частности, соотношения (6.38) и (6.39) соответствуют оценкам (6.7) и (6.8). •

Пример 6.8. Обратимся снова к системе $(++)$ из примера 6.7 и проверим для нее выполнение условий теоремы 6.5 на квадрате D_0 . Здесь отображение F определяется функциями $z_1 = f_1(x, y) = 4 - 2y$ и $z_2 = f_2(x, y) = e^x$.

1. Выясним, будут ли точки (z_1, z_2) принадлежать квадрату D_0 при всех $(x, y) \in D_0$. Ответ отрицательный, поскольку тогда координаты $z_1 = 4 - 2y$ лежат на отрезке $[2; 4]$, а не на $[0; 1]$, а координаты $z_2 = e^x$ находятся на $[1; e]$ вместо $[1; 2]$. Следовательно, F не отображает D_0 в себя.

2. Частные производные $\frac{\partial f_1}{\partial x} = 0$, $\frac{\partial f_1}{\partial y} = -2$, $\frac{\partial f_2}{\partial x} = e^x$, $\frac{\partial f_2}{\partial y} = 0$ непрерывны на D_0 . Первая из сумм модулей частных производных из не-

равенств (6.37) равна 2, вторая — e^x . Понятно, что эти неравенства не имеют места не только на квадрате D_0 , но и на любом другом прямоугольнике изоляции решения. •

Пример 6.8 показывает, что запись системы (+) в форме (++) оказалась неудачной для применения метода итерации. Однако ее можно преобразовать в другую эквивалентную приведенную систему с выполнением условий теоремы 6.5 (упр. 6.24).

Формулы для вычисления абсолютных погрешностей приближений t_n к решению \bar{t} и условие окончания итерационного процесса при нахождении приближенного решения с заданной точностью $\epsilon > 0$ легко получаются из оценок (6.38) и (6.39).

З а м е ч а н и я. 1. В теореме 6.5 предварительное условие существования единственного решения на пространстве X можно было не ставить. В силу теоремы Банаха оно автоматически следует из остальных двух условий.

2. Ввиду того что теорема 6.4 справедлива не только на прямоугольниках вида (6.35), но также и на \mathbb{R}^2 , теорема 6.5 и ее доказательство остаются верными, если в качестве метрического пространства X взять все пространство \mathbb{R}_∞^n .

Из теоремы 6.5 с учетом сделанных замечаний вытекает важное следствие.

С л е д с т в и е 6.2. Если областью определения системы (6.32) является все пространство \mathbb{R}^2 и на \mathbb{R}^2 выполняется условие (2) теоремы 6.6, то система (6.32) имеет одно и только одно решение, к которому итерационная последовательность будет сходиться при любом начальном приближении $(x_0, y_0) \in \mathbb{R}^2$.

Упражнения

6.24. Приведем систему (+) к виду (6.32) другим способом:

$$\begin{cases} x = \ln y, \\ y = -\frac{1}{2}x + 2. \end{cases}$$

Покажите, что условия теоремы 6.5 для этой системы верны на прямоугольнике $\{(x, y): 0 \leq x \leq 1, 1,2 \leq y \leq 2\}$, а на квадрате D_0 — нет.

6.25. Найдите прямоугольник изоляции решения системы

$$\begin{cases} x = e^{-y}, \\ y = \sqrt{x}, \end{cases}$$

где выполняются условия теоремы 6.5.

6.26. Докажите, что кроме очевидного решения $x = 0, y = 0$, система

$$\begin{cases} 0,3 \sin(x + y) - x = 0, \\ 4y - x = 0 \end{cases}$$

не имеет других решений.

6.27. Найдите приближенные решения систем из упр. 6.24 и 6.25 с точностью до $\epsilon = 0,05$.

6.28. Перенесите все рассуждения и результаты данного параграфа с двумерного случая на случай системы n уравнений с n неизвестными.

6.29. Системы уравнений из 6.4 на самом деле являются частными случаями рассматриваемых здесь систем. Убедитесь, что теорему 6.3 (при $n = 2$) можно получить из следствия 6.2 и оценочных формул (6.38), (6.39). Обратите также внимание на аналогии в доказательствах теорем 6.3 и 6.5.

Приложение

Лабораторные работы

Целью лабораторного практикума является усвоение и закрепление теоретического материала, приобретение практических навыков приближенного решения математических задач с помощью соответствующих численных методов. Здесь приведены описания девяти лабораторных работ, охватывающих все основные разделы изучаемого в пособии курса. Их можно выполнять на аудиторных занятиях и самостоятельно. Каждое описание содержит перечень необходимых вопросов теории, индивидуальные задания на подгруппу до 15 человек и порядок выполнения работы.

Задания несложные и в своей теоретической части являются дополнениями к системе примеров и упражнений из соответствующих параграфов пособия. При их выполнении каждый шаг качественного характера следует подкреплять теоретическими положениями. Предполагается, что общее время на исследования, составление компьютерных программ (если они необходимы)¹⁾, вычисления и подготовку письменного отчета по работе не должно превышать 2 – 4 часов.

Требования к вычислительным средствам минимальны. Ручные расчеты можно выполнять на микрокалькуляторе, для автоматизации вычислений достаточно простейшей ПЭВМ с системой программирования Бейсик или Паскаль.

В работах, где основные расчеты организуются в программном режиме, но крайней мере один шаг метода рекомендуется выполнить вручную. Это поможет лучше понять алгоритм вычислений, а затем описать его в виде компьютерной программы. В программах для итерационных процессов целесообразно предусмотреть вывод таблицы, отражающей результаты каждого шага процесса. Она явится хорошим наглядным материалом для анализа вычислений и беседы преподавателя с обучающимся.

В письменном отчете по лабораторной работе необходимо отразить следующее: тему работы и задание с учетом предложенного ва-

¹⁾ Говоря о составлении программы, мы имеем в виду и возможную разработку ее схемы (блок-схемы). Этот предварительный этап важен при слабой программистской подготовке обучающихся и может быть включен преподавателем в разряд обязательных видов работы.

рианта; теоретические исследования (в краткой форме) и вычисления согласно порядку выполнения работы; программу для расчетов (если она необходима); выводимую программой таблицу (если она предусмотрена); итоговые результаты в требуемой форме.

Лабораторная работа № 1 Вычисления с учетом погрешностей

Необходимые сведения из теории (см. гл. 1).

1. Абсолютная и относительная погрешности приближенных чисел и правило их записи.
2. Верные значащие цифры приближенных чисел.
3. Нахождение абсолютной погрешности по верным цифрам.
4. Правило округления чисел.
5. Правило записи приближенных чисел.
6. Оценка влияния погрешностей аргументов на значение функции.
7. Оценка погрешностей арифметических действий.

Задание

Пусть a, b, y — приближенные числа с верными в строгом смысле значащими цифрами, x — точное число. Вычислите

$$z = \frac{ab - e^x}{\sin y}$$

и оцените погрешность результата. Для вычисления значений функций e^x и $\sin y$ используйте либо математические таблицы, либо микрокалькулятор, либо компьютер.

Данные по вариантам

Вариант	a	b	x	y
1	2,03	-1,670	0,970	0,504
2	0,971	3,26	0,035	-1,061
3	1,510	-1,84	1,115	0,234
4	-0,193	-5,97	0,871	2,060
5	3,112	0,786	2,06	-2,541
6	-1,745	1,090	1,836	-2,541
7	10,7	0,0836	0,755	-1,43
8	3,07	-1,247	0,601	0,967
9	-0,812	2,19	1,64	0,367
10	2,410	-0,794	2,019	1,96
11	8,345	0,16	0,967	-2,112
12	-1,050	2,47	1,318	0,840
13	0,189	-9,375	1,08	1,05
14	-14,1	0,781	0,542	0,641
15	3,56	1,086	2,12	-2,396

Порядок выполнения работы

Результаты расчетов расположите в таблицах:

a		b		x		y	
Δ_a		Δ_b		Δ_x		Δ_y	
δ_a		δ_b		δ_x		δ_y	

z_1		z_2		z_3		z_4		z	
Δ_{z_1}		Δ_{z_2}		Δ_{z_3}		Δ_{z_4}		Δ_z	
δ_{z_1}		δ_{z_2}		δ_{z_3}		δ_{z_4}		δ_z	

где $z_1 = ab$, $z_2 = e^x$, $z_3 = z_1 - z_2$, $z_4 = \sin y$, $z = z_3/z_4$.

1. Заполните первую таблицу, определив абсолютные погрешности исходных данных по известным верным значащим цифрам.

2. Оцените погрешности $z_1 = ab$, взяв для этого две-три значащие цифры произведения. Затем найдите верные значащие цифры z_1 и запишите ответ с одной сомнительной цифрой.

3. Вычислите $z_2 = e^x$ и округлите его при необходимости так, чтобы погрешность округления не оказала существенного влияния на точность дальнейших расчетов.

Продолжите таким же образом.

Лабораторная работа № 2 Метод половинного деления

Необходимые сведения из теории (см. 2.1 – 2.3).

1. Этапы приближенного решения уравнений с одним неизвестным.
2. Отделение корней. Графическое отделение корней.
3. Условия применения метода половинного деления.
4. Алгоритм метода половинного деления.
5. Условие окончания процесса деления при заданной допустимой погрешности.

Задание

Отделите корни данного уравнения и уточните их методом половинного деления с точностью до $\epsilon = 0,5 \cdot 10^{-4}$.

Уравнения по вариантам

Вариант	Уравнение	Вариант	Уравнение
1	$x^2 + e^x = 2$	5	$x^2 + \cos(2 + x) = 1$
2	$3 \sin(x + 0,7) - 0,5x = 0$	6	$x \ln(x + 1) = 1$
3	$\cos x - (x - 1)^2 = 0$	7	$\ln(x + 1) - (x - 2)^2 = 0$
4	$5 \sin x = x$	8	$2 \ln x - 0,5x + 1 = 0$

Вариант	Уравнение	Вариант	Уравнение
9	$(x - 2)\ln x = 1$	13	$x \ln(x + 2) = 2$
10	$\sin(x - 0,5) - 2x + 0,5 = 0$	14	$x^3 - 0,5 - \sin x = 0$
11	$\cos(x + 0,3) = x^2$	15	$\sin(x + 1) = 0,2x$
12	$x^2 - 3\sin x = 0$		

Порядок выполнения работы

1. Отделите графически все корни уравнения $f(x) = 0$ так, чтобы на отрезках изоляции корней функция f удовлетворяла условиям метода половинного деления.

2. Выполните один шаг метода для одного из корней вручную и проверьте условие окончания вычислений.

3. Составьте программу уточнения корня с точностью до ϵ , выводящую результаты в таблицу:

n	a_n	b_n	E_n
...

где a_n, b_n — концы вложенных отрезков, E_n — их длины.

4. Найдите все приближенные корни уравнения и выпишите их с верными значащими цифрами.

Лабораторная работа № 3

Комбинированный метод хорд и касательных

Необходимые сведения из теории (см. 2.1, 2.2, 2.5 – 2.7).

1. Отделение корней уравнений аналитическим способом.

2. Условия, при которых для уточнения корней применяются методы хорд и касательных.

3. Правила выбора начальных приближений для методов хорд и касательных.

4. Алгоритм уточнения корней комбинированным методом хорд и касательных.

5. Условие окончания процесса вычислений при заданной допустимой погрешности.

Задание

Отделите аналитически один из корней данного уравнения и определите его с точностью до $\epsilon = 0,5 \cdot 10^{-5}$ комбинированным методом хорд и касательных.

Уравнения по вариантам

Вариант	Уравнение	Вариант	Уравнение
1	$2x^3 - 3x^2 - 12x - 5 = 0$	2	$x^3 + 3x^2 - 24x - 10 = 0$

Вариант	Уравнение	Вариант	Уравнение
3	$x^3 - 3x^2 + 3 = 0$	10	$2x^3 + 9x^2 - 21 = 0$
4	$x^3 + 3x^2 - 2 = 0$	11	$x^3 + 3x^2 - 3,5 = 0$
5	$2x^3 - 3x^2 - 12x + 12 = 0$	12	$x^3 - 4x^2 + 2 = 0$
6	$x^3 + 3x^2 - 1 = 0$	13	$x^3 + 3x^2 - 24x + 1 = 0$
7	$x^3 - 3x^2 - 24x - 3 = 0$	14	$2x^3 - 3x^2 - 12x + 8 = 0$
8	$x^3 - 12x + 6 = 0$	15	$2x^3 + 9x^2 - 6 = 0$
9	$x^3 - 3x^2 + 2,5 = 0$		

Порядок выполнения работы

1. Отделите корни уравнения аналитически и выберите один из отрезков изоляции, на котором выполняются условия применимости метода.

2. Возьмите соответствующие начальные приближения и найдите вручную первые приближения. Проверьте условие окончания процесса вычислений.

3. Составьте программу уточнения корня с точностью до ϵ , которая выводила бы результаты в таблицу

n	x_n	y_n	E_n
...

где x_n и y_n — приближения к корню, найденные методами хорд и касательных соответственно, E_n — расстояния между ними.

4. Найдите приближенный корень и выпишите его с верными значащими цифрами.

Лабораторная работа № 4

Уточнение корней уравнений методом простой итерации

Необходимые сведения из теории (см. 2.1, 2.2, 2.8).

1. Методы отделения корней уравнений.
2. Алгоритм построения итерационной последовательности, порождаемой уравнением $x = g(x)$.
3. Достаточное условие сходимости итерационной последовательности.
4. Оценка погрешности n -го приближения к корню.
5. Условие окончания итерационного процесса при заданной допустимой погрешности.
6. Способы приведения уравнения $f(x) = 0$ к равносильному уравнению $x = g(x)$ с требуемыми для метода свойствами.

Задание

Отделите графически один из корней уравнения и определите его с точностью до $\epsilon = 0,5 \cdot 10^{-5}$ методом простой итерации.

Уравнения по вариантам

Вариант	Уравнение	Вариант	Уравнение
1	$x - 5\sin x - 1 = 0$	9	$3x + \cos x + 1 = 0$
2	$\ln x + 2x = 0$	10	$2 - x - \ln x = 0$
3	$4\sin x + 2x = -1$	11	$\sin x - 0,2x + 0,5 = 0$
4	$2x + \ln x + 0,5 = 0$	12	$x - 2\ln x = 2$
5	$x + 2 - e^x = 0$	13	$0,25x + \cos x = 0$
6	$2\cos x = 1 - x$	14	$\ln(x + 3) - x = 0$
7	$x = (x + 1)^3$	15	$x + \ln(1 + x) = 2$
8	$x^3 - 2x + 2 = 0$		

Порядок выполнения работы

1. Найдите графически отрезок $[a; b]$ небольшой длины h , изолирующий один из корней, и проверьте результат аналитически.
2. Приведите исходное уравнение к виду $x = g(x)$, пригодному для метода простой итерации на отрезке $[c; d] = [a - h; b + h]$.
3. Вычислите вручную x_1 , определите его абсолютную погрешность и проверьте условие окончания итерационного процесса.
4. Напишите программу вычисления приближений до достижения требуемой точности ϵ с выводом результатов в таблицу

n	x_n	E_n
...

где E_n — абсолютная погрешность приближения x_n .

5. Найдите приближенный корень и выпишите его с верными значащими цифрами.

Лабораторная работа № 5

Метод простой итерации приближенного решения систем линейных алгебраических уравнений

Необходимые сведения из теории (см. 1.2, 1.3, 1.6, 6.1, 6.4).

1. Способы определения расстояния в пространстве \mathbf{R}^n .
2. Абсолютная погрешность числового вектора и его координат.
3. Сходимость последовательности векторов в \mathbf{R}^n .
4. Приведенная система уравнений, способы преобразования систем к приведенному виду.
5. Построение итерационной последовательности.
6. Достаточное условие сходимости итерационной последовательности.
7. Оценка погрешности приближенного решения.
8. Условие окончания итерационного процесса при нахождении решения с заданной точностью.

Задание

Дана система уравнений, коэффициенты при неизвестных и свободные члены которой являются точными числами. Найдите ее приближенное решение с точностью до $\epsilon = 0,5 \cdot 10^{-3}$.

Системы уравнений по вариантам

Исходная система:

$$\begin{cases} Mx_1 - 0,04x_2 + 0,21x_3 - 18x_4 = -1,24, \\ 0,25x_1 - 1,23x_2 + Nx_3 - 0,09x_4 = P, \\ -0,21x_1 + Nx_2 + 0,80x_3 - 0,13x_4 = 2,56, \\ 0,15x_1 - 0,31x_2 + 0,06x_3 + Px_4 = M. \end{cases}$$

Вариант	M	N	P	Вариант	M	N	P
1	-0,77	0,16	1,12	9	-1,13	0,14	0,87
2	0,93	0,07	-0,84	10	0,91	-0,23	-1,04
3	-1,14	-0,17	0,95	11	-0,88	0,10	0,91
4	1,08	0,22	-1,16	12	1,25	-0,14	-1,09
5	0,87	-0,19	1,08	13	0,79	0,18	-0,86
6	-1,21	0,20	0,88	14	-1,19	-0,21	1,21
7	1,09	-0,16	0,84	15	0,89	0,12	-1,15
8	0,89	0,08	-1,21				

Порядок выполнения работы

1. Преобразуйте систему к приведенному виду с выполнением условия сходимости итерационной последовательности.

2. Взяв в качестве начального приближения вектор свободных членов приведенной системы, найдите вручную первое приближение, затем определите его абсолютную погрешность и проверьте условие окончания итерационного процесса.

3. Составьте программу вычисления приближений до достижения требуемой точности с выводом результатов в таблицу

k	x_1	x_2	x_3	x_4	E_k
...

где x_1, x_2, x_3, x_4 — координаты векторов-приближений, E_k — абсолютные погрешности этих векторов.

4. Найдите приближенное решение системы и выпишите его координаты с верными значащими цифрами.

Лабораторная работа № 6
Интерполирование математических таблиц

Необходимые сведения из теории (см. 3.1, 3.2, 3.5, 3.6).

1. Табличная функция.
2. Задача интерполирования табличной функции.
3. Теорема о единственности задачи полиномиального интерполирования.
4. Конечные разности таблиц.
5. Первый и второй интерполяционные многочлены Ньютона. Оценка погрешностей интерполяционных формул Ньютона.
6. Формула линейного интерполирования и способы оценки ее погрешности.
7. Обратное линейное интерполирование.

Задание

Дана таблица значений функции $f: f(x) = e^x - \sin x$ с верными цифрами:

x	$f(x)$	x	$f(x)$	x	$f(x)$	x	$f(x)$	x	$f(x)$
0	1	0,4	1,1024	0,8	1,5082	1,2	2,3881	1,6	3,9536
0,1	1,0053	0,5	1,1693	0,9	1,6763	1,3	2,7057	1,7	4,4823
0,2	1,0227	0,6	1,2575	1,0	1,8768	1,4	3,0696	1,8	5,0758
0,3	1,0543	0,7	1,3695	1,1	2,1130	1,5	3,4842	1,9	5,7396

1. Вычислите приближенное значение $f(a)$ с помощью первого интерполяционного многочлена Ньютона второй степени, определите его абсолютную погрешность и верные значащие цифры.
 2. Линейным интерполированием найдите значения функции f для аргументов a, b и определите их верные значащие цифры с помощью таблицы конечных разностей.
 3. Вычислите значения обратной для f функции φ для аргументов c, d по формуле обратного линейного интерполирования и запишите ответы с двумя цифрами после десятичной запятой.
- Все исходные данные a, b, c, d считаются точными числами.

Данные по вариантам

Вариант	a	b	c	d
1	0,38	0,35	1,0059	2,3770
2	1,02	1,07	2,6456	1,9245
3	1,15	1,18	2,8775	1,2236
4	1,22	1,24	1,0023	1,3240
5	1,36	1,31	1,1232	1,1601

Вариант	a	b	c	d
6	0,59	0,54	1,5222	2,2557
7	0,63	0,68	1,7092	3,3587
8	0,71	0,75	2,0988	1,0460
9	0,85	0,83	1,1847	2,9650
10	0,96	0,92	1,2775	1,0049
11	0,12	0,18	1,4892	1,3764
12	0,23	0,26	2,1232	1,6058
13	1,58	1,55	3,2323	1,8334
14	0,44	0,47	1,0323	2,4590
15	0,06	0,02	1,0974	1,0608

Порядок выполнения работы указан в задании.

Лабораторная работа № 7
Квадратичное приближение табличных функций
по методу наименьших квадратов

Необходимые сведения из теории (см. 3.1, 3.7)

1. Задача аналитического приближения табличных функций.
2. Задача приближения по методу наименьших квадратов.
3. Алгоритм построения наилучшего многочлена по данному методу.
4. Уклонения, среднеквадратичное уклонение.

Задание

По данной таблице найдите многочлен второй степени $P_2(x)$, являющийся наилучшим приближением к соответствующей табличной функции по методу наименьших квадратов. Начертите графики таблицы и найденного многочлена. Найдите все уклонения от табличных значений и среднеквадратичное уклонение.

Таблицы по вариантам

Вариант	Таблица						
1	x	0,10	0,30	0,40	0,60	0,70	0,80
	y	0,25	0,50	0,65	0,55	0,42	0,30
2	x	-2,00	-1,80	-1,70	-1,60	-1,40	-1,30
	y	5,10	4,00	3,20	3,90	4,80	6,10
3	x	1,30	1,40	1,60	1,70	2,00	2,10
	y	2,40	1,80	1,20	1,40	2,30	2,90

Вариант	Таблица						
4	x	0,40	0,70	0,90	1,10	1,40	1,60
	y	0,15	0,83	1,65	1,52	0,90	0,31
5	x	2,00	2,50	2,70	2,90	3,20	3,40
	y	-0,11	-0,81	-1,05	-0,90	-0,23	-0,05
6	x	-0,50	-0,30	-0,20	0,10	0,40	0,80
	y	2,30	1,20	1,05	0,90	1,20	2,10
7	x	1,10	2,00	2,50	2,90	3,50	4,00
	y	0,32	0,05	-0,10	-0,12	0,12	0,27
8	x	0,30	0,50	0,80	0,90	1,20	1,40
	y	1,10	0,60	0,40	0,38	0,65	0,90
9	x	-0,40	-0,10	0,10	0,20	0,50	0,70
	y	1,30	3,50	4,20	4,00	2,80	1,60
10	x	1,20	1,40	1,50	1,60	1,80	2,10
	y	0,90	3,30	4,10	3,90	2,80	1,10
11	x	-0,90	-0,80	-0,50	-0,40	-0,20	-0,10
	y	0,15	0,61	1,20	1,10	0,70	0,22
12	x	-1,00	-0,80	-0,70	-0,40	-0,30	-0,20
	y	1,40	0,90	0,65	0,51	0,78	1,30
13	x	0,20	0,30	0,50	0,70	0,90	1,20
	y	-2,10	-0,50	1,15	1,30	-0,60	-2,70
14	x	2,20	2,50	2,60	2,80	3,10	3,20
	y	1,70	0,80	0,52	0,30	0,91	1,50
15	x	-0,30	-0,10	0,20	0,30	0,70	0,90
	y	-2,10	1,30	3,00	2,40	-2,30	-8,00

Порядок выполнения работы

1. На координатной плоскости постройте точки таблицы и убедитесь, что они располагаются вблизи некоторой квадратной параболы.

2. Напишите в общем виде систему уравнений для определения коэффициентов многочлена $P_2(x)$ и выражения для коэффициентов системы.

3. Составьте программу вычисления коэффициентов и решения системы по правилу Крамера.

4. Найдите $P_2(x)$ (округлив коэффициенты до двух цифр в дробной части) и постройте ее график на той же координатной плоскости, где отмечены точки таблицы.

5. Найдите все уклонения и среднеквадратичное уклонение многочлена $P_2(x)$ от табличной функции.

Лабораторная работа № 8

Приближенное вычисление определенных интегралов

Необходимые сведения из теории (см. 4.1 – 4.5).

1. Численный метод приближенного вычисления определенных интегралов.

2. Квадратурные формулы трапеций и Симпсона.

3. Строгая оценка погрешностей этих формул.

4. Оценка погрешностей методом двойного пересчета.

5. Определение шага разбиения отрезка интегрирования, при котором квадратурная формула обеспечивает заданную точность.

6. Вопросы оценки точности приближенного интеграла с учетом вычислительных погрешностей.

Задание

1. Вычислите данный интеграл вручную по формуле трапеций при $n = 3$ и $n = 6$. Оцените погрешность приближения $J_6^{(n)}$ методом двойного пересчета, а затем найдите абсолютную погрешность этого же приближения по формуле строгой оценки погрешностей.

2. Вычислите данный интеграл по формуле Симпсона с точностью до $\varepsilon = 0,5 \cdot 10^{-4}$.

3. Вычислите интеграл по формуле Ньютона – Лейбница с максимальной точностью, которая возможна при используемых вычислительных средствах.

4. Сравните полученные разными способами результаты по их точности.

Интегралы по вариантам

Вариант	Интеграл	Вариант	Интеграл	Вариант	Интеграл
1	$\int_0^{\pi/2} \cos(1-2x) dx$	3	$\int_0^2 e^{2x} dx$	5	$\int_0^{\pi/2} \sin 2x dx$
2	$\int_0^{1,5} \cos x dx$	4	$\int_{\pi/6}^{\pi/2} \cos 3x dx$	6	$\int_{-1}^1 (x - e^{2x}) dx$

Вариант	Интеграл	Вариант	Интеграл	Вариант	Интеграл
7	$\int_0^2 \sqrt{1+x} dx$	10	$\int_0^2 \sin(x+1) dx$	13	$\int_0^2 \ln(2x+3) dx$
8	$\int_{-1}^1 (3x + \cos x) dx$	11	$\int_0^{1,5} (1+x+x^4) dx$	14	$\int_1^3 \sqrt{x-1} dx$
9	$\int_{-1}^2 e^{x/2} dx$	12	$\int_0^3 e^{-3x} dx$	15	$\int_0^{\pi} (\sin x + x^2) dx$

Порядок выполнения работы указан в задании. При вычислениях по формуле Симпсона сначала надо определить число n , при котором формула обеспечивает точность ε , затем составить программу реализации формулы и с ее помощью найти $J_n^{(C)}$. Для того чтобы не учитывать вычислительные погрешности, шаг разбиения и значения функций следует брать с двумя запасными цифрами.

Лабораторная работа № 9

Численное решение обыкновенных дифференциальных уравнений методом Эйлера – Коши

Необходимые сведения из теории (см. 5.1 – 5.6).

1. Общее и частное решения обыкновенного дифференциального уравнения 1-го порядка. Интегральные кривые.
2. Задача Коши. Теорема Пикара.
3. Геометрический смысл правой части дифференциального уравнения, разрешенного относительно производной.
4. Понятие численного решения. Ломаная Эйлера.
5. Метод Эйлера – Коши, его геометрический смысл.
6. Оценка погрешности численного решения методом двойного пересчета.

Задание

Используя метод Эйлера – Коши, найдите численное решение дифференциального уравнения на отрезке $[a; b]$ с шагом $h = 0,1$, удовлетворяющее начальному условию $y(x_0) = y_0$ (в таблицу представлять улучшенные значения y_i^* , найденные двукратными вычислениями с шагом $h/2 = 0,05$). Оцените погрешности чисел y_i^* методом двойного пересчета и определите верные значащие цифры этих чисел. Начертите ломаную Эйлера.

Уравнения по вариантам

Вариант	Уравнение	x_0	y_0	$[a; b]$
1	$y' = x + y$	0	0,8	$[0; 1]$
2	$y' = x + \cos y$	1,8	2	$[1,8; 2,8]$

Вариант	Уравнение	x_0	y_0	$[a; b]$
3	$y' = e^x + y$	0	1,2	[0; 1]
4	$y' = xy + \sin x$	0	2	[0; 1]
5	$y' = x + 3 \sin y/3$	1,6	2	[1,6; 2,6]
6	$y' = e^{x \cdot y}$	0	-1	[0; 1]
7	$y' = xy + e^x$	-1	0,5	[-1; 0]
8	$y' = x + y^2$	-2	0	[-2; -1]
9	$y' = \sin(x - y)$	1	3	[1; 2]
10	$y' = \cos(x + y)$	2	0	[2; 3]
11	$y' = y + \cos x$	2	0	[2; 3]
12	$y' = x^2 + y$	1	0	[1; 2]
13	$y' = x + e^y$	1	-1	[1; 2]
14	$y' = x + \sin y$	1,5	3	[1,5; 2,5]
15	$y' = x^2 + y^2$	0	0	[0; 1]

Порядок выполнения работы

1. Убедитесь в существовании и единственности решения поставленной задачи Коши.
2. Вычислите вручную y_1^* и оцените его погрешность.
3. Составьте программу вывода таблицы

x_i	y_i^*	y_i	E_i
...

где y_i – приближение к значению точного решения в точке x_i , найденное однократным вычислением по методу Эйлера – Коши (с шагом $h = 0,1$), E_i – оценка погрешности значения y_i^* .

4. Получите искомое численное решение, выписывая табличные значения с верными значащими цифрами.
5. Постройте соответствующую ломаную Эйлера.

Литература

1. Бахвалов Н. С. Численные методы. — М.: Наука, 1973.
2. Березин И. С., Жидков Н. П. Методы вычислений. Т. 1. — М.: Наука, 1966; Т. 2. — М.: Физматгиз, 1962.
3. Бохан К. А., Егорова И. А., Лащенко К. В. Курс математического анализа. Т. 1. — М.: Просвещение, 1972.
4. Бохан К. А., Егорова И. А., Лащенко К. В. Курс математического анализа. Т. 2. — М.: Просвещение, 1972.
5. Вычислительная математика / Н. И. Данилина, Н. С. Дубровская, О. П. Кваша, Г. Л. Смирнов. — М.: Высшая школа, 1985.
6. Демидович Б. П., Марон И. А. Основы вычислительной математики. — М.: Наука, 1970.
7. Демидович Б. П., Марон И. А., Шувалова Э. З. Численные методы анализа. — М.: Наука, 1967.
8. Заварыкин В. М., Житомирский В. Г., Лапчик М. П. Численные методы. — М.: Просвещение, 1990.
9. Ильин В. А., Садовничий В. А., Сендов Бл. Х. Математический анализ. — М.: Изд-во МГУ, 1985.
10. Калиткин Н. Н. Численные методы. — М.: Наука, 1978.
11. Коллатц Л. Численные методы решения дифференциальных уравнений. — М.: Иностранная литература, 1953.
12. Колмогоров А. Н., Фомин С. В. Элементы теории функций и функционального анализа. — М.: Наука, 1972.
13. Копченова И. В., Марон И. А. Вычислительная математика в примерах и задачах. — М.: Наука, 1972.
14. Курош А. Г. Курс высшей алгебры. — М.: Наука, 1975.
15. Пулькин С. П., Никольская Л. Н., Дьячков А. С. Вычислительная математика. — М.: Просвещение, 1980.
16. Матвеев Н. М. Дифференциальные уравнения. — М.: Просвещение, 1988.
17. Матвеев Н. М. Методы интегрирования обыкновенных дифференциальных уравнений. — М.: Высшая школа, 1963.
18. Мысовских И. П. Лекции по методам вычислений. — М.: Физматгиз, 1962.
19. Носач В. В. Решение задач аппроксимации с помощью персональных компьютеров. — М.: Изд-во МИКАП, 1994.
20. Самарский А. А., Гулин А. В. Численные методы. — М.: Наука, 1989.
21. Турчак Л. И. Основы численных методов. — М.: Наука, 1987.
22. Фихтенгольц Г. М. Курс дифференциального и интегрального исчисления. Т. 1. — М.: Наука, 1969.
23. Фихтенгольц Г. М. Курс дифференциального и интегрального исчисления. Т. 2. — М.: Наука, 1969.

24. *Форсайт Дж., Малькольм М., Моулер К.* Машинные методы математических вычислений. — М.: Мир, 1980.

25. *Форсайт Дж., Моулер К.* Численное решение систем линейных алгебраических уравнений. — М.: Мир, 1969.

26. *Шун Т. Е.* Прикладные численные методы в физике и технике. — М.: Высшая школа, 1990.

Предметный указатель

- Абсолютная дифференциальная погрешность 35
Абсолютная погрешность 19
– – вектора 29
– – функции 31
Аналитическое приближение табличной функции 78
- Дифференциальная погрешность 34
- Задача Коши 131, 144
Значащие цифры 20
– – верные 22
– – – в нестрогом смысле 24
– – сомнительные 22
- Интегральная кривая 131
Интерполирование 79
– обратное 96
– полиномиальное 80
Интерполирующая функция 79
Интерполяционная формула 80
– – Лагранжа 85
– – Ньютона, первая 88
– – –, вторая 91
Интерполяционный многочлен 80
– – Лагранжа 85
– – Ньютона, первый 88
– – –, второй 90
Источники погрешностей 12
Итерационная последовательность 49
Итерация 49
- Квадратура механическая 105
Квадратурная формула 105
- – прямоугольников с левыми ординатами 109
– – – с правыми ординатами 109
– – – с центральными ординатами 109
– – Симпсона (парабол) 119
– – трапеций 115
Конечные разности 86
Корень уравнения 41
– – изолированный 41
– – отделенный 41
– –, уточнение 42
Коэффициент сжатия 150
- Ломаная Эйлера 134
- Математическая модель задачи 10
Метод двойного пересчета (Рунге) 120, 143
– итерации (простой итерации) 67, 154
– касательных (Ньютона) 59
– математической индукции 54
– наименьших квадратов 100
– половинного деления 46
– последовательных приближений 42, 49
– серединных точек 140
– хорд 51
– хорд и касательных, комбинированный 65
– Эйлера 135
– Эйлера – Коши 138
Метрика (расстояние) 14
– $\rho_1, \rho_2, \rho_\infty$ 14
Метрическое подпространство 147

- пространство 14
- – $\mathbf{R}_1^1, \mathbf{R}_2^2, \mathbf{R}_\infty^n$ 147

Направление интегральных кривых 132

Начальная точка 131

Начальное приближение 49

Начальное условие 131

Начальные данные 131

Некорректная задача 125

Неподвижная точка отображения 148

Нули округления 22

Нуль функции 41

Остаточный член приближенной формулы 30

Отделение корней 41

Относительная дифференциальная погрешность 35

Относительная погрешность 28

Отображение 148

– переводящее множество в себя 149

– сжатия 149

Отрезок изоляции корня 41

Оценка погрешности (точности) 19

Опеночная функция 30

Погрешность 18

– вычислительная 12

– метода 12

– таблицы 77

– усечения формулы Тейлора 142

Последовательность приближений 42

Правило записи приближенных чисел 24

– округления абсолютной и относительной погрешностей 20, 28

– – чисел 21

– определения абсолютной погрешности по верным цифрам 23

– – верных цифр при линейной интерполяции 95

Предел последовательности в метрическом пространстве 151

Приближение с заданной точностью 20

Приближенное число (вектор, функция) 18

Приближенный численный метод 12

Приведенная система 185, 193

Принцип сжимающих отображений 177

Расстояние (метрика) 14

– между векторами 14

– – значениями функций 15

– – функциями 16

– – числами 14

Ресуррентная формула 49

Решение обыкновенного дифференциального уравнения 130

– системы обыкновенных дифференциальных уравнений 144

– – уравнений 158, 166

– – –, изолированное 166

Сжимающее отображение 149

Среднеквадратичное уклонение 99

Сходимость в метрическом пространстве 151

Таблица конечных разностей 186

Таблица значений функции 76

– – – математическая 76

– – – эмпирическая 76

Табличная функция 77

Табличные аргументы 76

– – равноотстоящие 76

– значения 77

Теорема Банаха 152

– Больцапо – Коши, первая 44

– –, вторая 110

– Вейерштрасса, вторая 16

– Кантора 46

– Лагранжа 55, 168

– обобщенная о среднем значении интеграла 110

– Пикара 131

- Ролля 45
- Точное число (вектор, функция) 18
- Точный численный метод 11

- Узлы** интерполирования 79
 - квадратуры 105
- Уклонения 99
- Условие Липшица 149
- Уточнение корня уравнения 42

- Формула Лагранжа** 55
 - линейного интерполирования 92
 - Ньютона – Лейбница 104

- обратного линейного интерполирования 96
- Тейлора 59

- Численный метод** 10
- Численное решение задачи Коши 133, 145

- Шаг** разбиения отрезка интегрирования 108
 - таблицы значений функции 76

- Эмпирическая таблица** 76
 - формула 98
 - функция 98