

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ	6
Введение. Задачи прикладной математики и математическое моделирование	8
Глава 1. Математические модели и вычислительные алгоритмы	12
§ 1. Математические модели	12
§ 2. Структура математической модели	
и ее место в научном и инженерном исследовании	15
1. Модели «теоретического» типа	15
2. Задача баллистики	18
3. Модели «эмпирического» типа	25
§ 3. Вычислительные алгоритмы	27
§ 4. Приближенные вычисления	29
1. Источники и классификация погрешности	29
2. Абсолютная и относительная погрешности	30
3. Теоремы о погрешностях	31
4. Правила контроля верных цифр при массовых вычислениях	
с приближенными числами	32
Глава 2. Методы аналитического представления экспериментальных	
и алгоритмически заданных зависимостей	34
§ 1. Аналитическое представление табличных функций в форме	
интерполяционного многочлена	35
1. Интерполирование табличных функций	35
2. Интерполяционный многочлен Лагранжа	37
3. Интерполяционный многочлен Ньютона	38
4. Применение интерполяционных многочленов	39
5. Недостатки интерполяционных многочленов высоких степеней	39
§ 2. Аналитическое представление табличных функций	
в виде эмпирической формулы	40
1. Структурная идентификация функции	41
2. Параметрическая идентификация функции. Метод наименьших	
квадратов	42
3. Оценка точности эмпирической формулы	44
4. Способы сведения сложных эмпирических формул к простым	46
Глава 3. Численные методы решения нелинейных уравнений	
с одним неизвестным	49
§ 1. Отделение корней уравнения	50
1. Основная теорема	50
2. Свойства действительных корней алгебраических уравнений	52
3. Графическая локализация корней	52
§ 2. Некоторые итерационные методы уточнения корней	54
1. Метод бисекции	54
2. Метод касательных	57
3. Метод секущих	62

4. Сравнительная характеристика рассмотренных итерационных методов	63
§ 3. Интерполяционные методы уточнения корней	65
Глава 4. Методы вычисления определенных интегралов	68
§ 1. Понятие определенного интеграла	68
§ 2. Формула Ньютона-Лейбница	68
§ 3. Квадратурные формулы, основанные на определении понятия интеграла	69
1. Формулы прямоугольников	70
2. Формула трапеций	71
3. Формула Симпсона (формула парабол)	72
4. Пример расчета интеграла по трем квадратурным формулам	72
§ 4. Квадратурные формулы интерполяционного типа. Формулы Гаусса	73
1. Формула Гаусса для отрезка $[-1, 1]$	73
2. Формула Гаусса для произвольного отрезка	74
3. Повышение точности квадратурной формулы Гаусса за счет разбиения отрезка интегрирования	75
§ 5. Построение первообразной функции с помощью численного интегрирования	75
Глава 5. Численное дифференцирование и методы численного интегрирования обыкновенных дифференциальных уравнений (ОДУ)	78
§ 1. Численное дифференцирование функций	78
§ 2. Численное интегрирование ОДУ первого порядка (ОДУ-1)	79
1. Обыкновенные дифференциальные уравнения первого порядка (ОДУ-1)	79
2. Задача Коши для ОДУ-1	80
3. Метод Эйлера	80
4. Методы Рунге-Кутты	82
5. Многошаговые методы (методы Адамса)	83
§ 3. Численное интегрирование систем ОДУ	84
§ 4. Численное интегрирование ОДУ второго порядка (ОДУ-2)	86
1. Задача Коши для ОДУ-2	86
2. Сведение ОДУ-2 к системе двух ОДУ-1	87
3. Краевая задача для ОДУ-2	87
Глава 6. Задачи оптимизации	90
§1. Параметры оптимизации и целевая функция	91
§2. Одномерные задачи оптимизации	92
1. Аналитические методы оптимизации	92
2. Задача о наилучшей консервной банке	93
3. Численные методы решения одномерных задач оптимизации	95
§3. Многомерные задачи оптимизации	98
1. Общие сведения	98
2. Метод сетки	100

3. Метод покоординатного спуска	100
4. Метод градиентного спуска	102
5. Метод наискорейшего спуска	103
6. Проблема «оврагов»	104
7. Проблема многоэкстремальности	105
§4. Задачи линейного программирования	106
1. Транспортная задача	106
2. Задача об оптимальном использовании ресурсов	108
§5. Общие рекомендации. Что и как оптимизировать	113

ПРЕДИСЛОВИЕ

Интенсивный процесс проникновения математических методов в различные области науки, техники, экономики, и, в частности, широкое использование во всех этих областях современной вычислительной техники выдвигают повышенные требования к подготовке высококвалифицированных специалистов. Современный инженер должен не только обладать общей математической культурой, свободно владеть современной вычислительной техникой и соответствующим программным обеспечением, но и уметь сформулировать на математическом языке техническую или экономическую проблему. Без этого умения все его математические знания останутся «мертвым грузом».

Кроме того, пользуясь тем или иным программным продуктом, инженер должен ясно представлять себе, какие вычислительные методы этот продукт реализует, уметь грамотно выбрать нужный метод и квалифицированно им воспользоваться. Одним словом, он должен ориентироваться в современных численных методах, знать их возможности, достоинства и недостатки. Очень важно также знать типичные ошибки, которые могут быть допущены при использовании этими методами, знать, каковы признаки того, что ошибка допущена, и как ее избежать. Поэтому в планы подготовки студентов технических специальностей в последние годы все чаще включаются дисциплины, знакомящие их с основами прикладной математики и методов вычислений. Названия этих дисциплин отличаются на разных факультетах, но непременно содержат слова «математические модели...», «математическое моделирование...», «вычислительная математика» и т. д.

Следует отметить, что по прикладной математике, в частности, по численным методам и математическому моделированию написано немало учебной литературы. Однако в большинстве своем эти книги ориентированы на профессиональных математиков и студентов математических специальностей. Большая часть их содержания посвящена обоснованию описываемых методов, доказательству их сходимости, выводу расчетных формул и другим специальным вопросам. Приводимые в них примеры расчетов обычно немногочисленны и носят, в основном, иллюстративный характер.

Инженеров же гораздо больше интересуют не доказательства теорем и выводы формул, а технология применения различных численных методов в реальных технических, экономических и управленческих задачах. При организации сложных вычислений у них нередко возникают вопросы, связанные с особенностями применения тех или иных методов, на которые им не всегда удастся найти ответы в вышеуказанной литературе.

Иными словами, необходимы такие учебные пособия по прикладной математике, ориентированные на студентов технических специальностей и инженеров, в которых были бы учтены все перечисленные особенности. В настоящий момент испытывается явный недостаток в литературе такого рода.

Данная книга имеет своей целью восполнить этот недостаток. Она написана на основе курса лекций, которые автор читал на протяжении ряда лет в

Уральском государственном университете путей сообщения студентам различных технических специальностей. Цель этого курса – дать студентам общее представление о прикладной математике, о проблемах, связанных с применением математических методов и вычислительной техники при решении технических, экономических и управленческих задач, в том числе задач оптимизации различных объектов и процессов.

В ходе изложения материала автор старался свести до минимума обсуждение абстрактных теоретических вопросов и сосредоточить основное внимание на технологии практического применения описываемых методов, а также на рекомендациях по их выбору и анализе наиболее типичных ошибок. Все теоретические положения курса поясняются многочисленными иллюстрациями и примерами. По мере необходимости приводится материал справочного характера.

Книга, как и курс, которому она соответствует, знакомит студентов с технологией постановки и решения математическими методами различных технических, экономических и других «нематематических» по своей изначальной сути проблем. Кроме того, она призвана помочь будущему инженеру расширить круг задач, которые он, пользуясь современной вычислительной техникой, мог бы решить самостоятельно, не привлекая профессиональных математиков.

С учетом всех приведенных соображений в книгу не включен ряд сложных вопросов, традиционно включаемых в курсы лекций по методам вычислений. Это касается, главным образом, всевозможных многомерных задач, в том числе краевых задач для дифференциальных уравнений с частными производными, интегральных уравнений и т.п. Данное решение продиктовано тем, что упомянутые задачи требуют профессиональной математической подготовки. Их должны решать математики-специалисты.

При изложении тех или иных вычислительных методов автор вовсе не стремился охватить все известные методы (да это и невозможно). В большинстве случаев изложение ограничивается двумя-тремя наиболее представительными, либо наиболее простыми, либо наиболее удобными при компьютерной реализации, и потому чаще всего используемыми сегодня методами.

ВВЕДЕНИЕ

Задачи прикладной математики и математическое моделирование

Характерной особенностью настоящего времени является то, что решение сколько-нибудь серьезных естественнонаучных, технических, экономических, социальных и многих других задач невозможно сегодня без применения математики и математических методов. Бурное развитие вычислительной техники привело не только к проникновению математических методов в различные области человеческой деятельности, но и к формированию современной *прикладной математики*. Так называется та область математики, которая занимается *приложениями* математики, т. е. применением математических методов к решению всевозможных практических, технических и других *нематематических* по своей изначальной сути задач.

Этой области математики присущи свои специфические особенности, отличающие ее от «классических» математических дисциплин. Как правило, любой математический текст, формулировка теоремы или вывод формулы начинаются со слов «Пусть дано...». Далее следует изложение исходных предпосылок на строгом математическом языке, позволяющем автору быть однозначно понятым любым математиком или другим специалистом, компетентным в соответствующей области.

Задавались ли Вы когда-нибудь вопросом «...а кем, собственно, дано?» или «почему дано именно это, а не что-то другое»? С точки зрения «классической» математики эти вопросы бессмысленны. Математик всегда формулирует ту задачу, которую он *смог* решить, при этом в качестве «дано» фигурируют те условия, при которых ему *удалось* решить эту задачу. Таким образом, он сам ставит себе задачу и сам определяет условия, при которых она имеет решение.

Иначе обстоит дело с прикладными исследованиями. В их основе лежит реальный «нематематический» объект: производственный процесс, конструкция, явление природы, экономический план, система управления и т. д. Таким образом, условие задачи первоначально бывает сформулировано на языке, характерном для сфер деятельности весьма далеких от математики и математического языка. Исследование проблемы начинается с формализации объекта, с придания задаче математической формы или, как говорят, с построения *математической модели** объекта. Только после этого мы можем воспользоваться для его изучения математическими методами.

Иногда бывает и так, что «заказчик» вообще затрудняется объяснить, чего он, собственно, ожидает. Просто у него что-то не получается с реализацией нового проекта или конструкции, не так, как задумано, идет производственный процесс и т. п. Ему необходима *математическая модель* указанного процесса, конструкции и т. д., для того, чтобы *предсказать*, как они будут вести себя в тех или иных условиях, либо *оптимизировать* их в том или

* Понятие *математической модели* будет подробно определено в Главе 1.

ином смысле (т. е. изменить их так, чтобы они наилучшим возможным образом соответствовали какому-то важному для «заказчика» условию). В этом случае задачу еще предстоит вычленить, сформулировать на обычном языке, и лишь потом приступать к ее формализации на математическом языке, построению математической модели объекта и изучения его математическими методами.

Таким образом, в постановку и решение прикладной задачи оказываются вовлеченными люди разных профессий и, даже, разных сфер деятельности, говорящие на разных «профессиональных языках». Нередко бывает так, что исходные условия (то, что «дано») и сущность проблемы (то, что требуется найти или получить) задают одни люди, а постановку и решение соответствующей математической задачи осуществляют совсем другие. В этом случае сформулированные выше вопросы совсем не бессмысленны. Особенно, если учесть, что значительное количество исходных данных приходится получать путем измерений или проведения специальных экспериментов непосредственно в процессе формализации задачи по мере осмысления их необходимости.

Отметим второе важное отличие прикладной математики от «классических» математических дисциплин. В прикладных исследованиях все результаты должны быть доведены «до числа», отсюда следует, что исходная информация тоже не может быть задана «в общем виде», т. е. в виде букв или формально записанных функций (например, $y=f(x)$). Все исходные данные должны быть представлены численно, а все функциональные зависимости должны быть конкретизированы и определены вплоть до численного значения каждого коэффициента.

Следует заметить, что численное решение прикладных задач интересовало математиков всегда. Крупнейшие математики прошлого наряду с чисто теоретическими исследованиями уделяли внимание изучению всевозможных явлений природы и построению математических моделей этих явлений. Анализ все более сложных моделей потребовал создания специальных, численных методов решения задач. Названия многих методов носят имена их создателей – Ньютона, Эйлера, Гаусса, Лагранжа, Чебышева и др.

И все же, в докомпьютерную эпоху численные методы применялись весьма ограниченно. Причина кроется в огромном объеме рутинных вычислений. Лишь с появлением современной вычислительной техники началось широкое применение численных методов. Необходимость решения новых и все более сложных задач потребовала разработки большого количества новых методов (а также переосмысления многих старых), ориентирования их на использование в высокоскоростных вычислительных машинах в автоматическом режиме, т. е. в таком режиме, когда человек не контролирует качество промежуточных этапов вычислений и не может ничего «подправить» в вычислительной процедуре.

Не следует думать, что современное состояние программного обеспечения и совершенствование вычислительной техники позволяют сразу решить любую прикладную проблему. Во многих случаях требуется доводка методов,

приспособление их к решению конкретных задач. Необходимость подобного рода действий входит в определенное противоречие с тенденцией использования таких программных средств, как *MathLab*, *MathCAD*, *Mathematica* и другие.

Вопрос о выборе численных методов и программных средств решается отдельно в каждом конкретном случае исходя из особенностей решаемой задачи и предшествующего опыта. При этом нередко приходится использовать методы, применение которых теоретически до конца не обосновано либо теоретические оценки погрешностей этих методов указывают на невозможность их практического использования^{*}. В таких случаях приходится полагаться на опыт предшествующего решения подобных задач, на интуицию и обязательное сравнение с экспериментом либо с известными точными решениями. Подобный образ действий совершенно немыслим для многих абстрактных областей классической математики, и он представляет собой третье, весьма существенное отличие прикладной математики от классических математических дисциплин.

Следует упомянуть, также, и четвертое отличие. Во всякой прикладной работе существенным моментом является необходимость получения результатов в определенный срок. Все исследования и расчеты должны быть завершены до этого срока, чтобы на их основе можно было принять конкретные решения.

Если исследования не будут завершены к оговоренному сроку, то решения все равно будут приняты, но на основании более грубого или просто «волевого» подхода. В такой ситуации лучше найти удовлетворительное решение задачи, но вовремя, чем получить полное или более точное решение задачи к тому моменту, когда оно станет бесполезным.

Описанные выше особенности прикладной математики характеризуют ее как довольно специфическую область науки, порой весьма заметно отличающуюся и предметно, и методологически от классической или, как иногда говорят, «чистой» математики. Именно эти особенности делают прикладную математику особенно полезной для инженеров, физиков, экономистов и представителей других нематематических профессий. Недаром приближенные решения, полученные по упрощенным математическим моделям, (пусть не самые точные, но зато полученные в срок!) часто называют «инженерными решениями», а соответствующие модели – «инженерными моделями».

Сказанное выше вовсе не означает, что прикладная математика является какой-то отдельной от всей остальной математики наукой, и уж тем более, она не подменяет собой другие математические дисциплины. Весь арсенал ее методов основан на понятиях и теоремах, доказанных в различных разделах математики. В этом смысле все математические науки едины.

^{*}Теоретические оценки погрешностей методов почти всегда оказываются слишком «жесткими». Эти методы сплошь и рядом удается применять (и получать вполне приемлемые для практики результаты!) при гораздо менее жестких ограничениях, т.е. в условиях, когда ни сходимость метода, ни его работоспособность, вообще говоря, не доказаны.

В то же время, очевидно, что все участники процесса постановки и решения прикладной задачи должны быть знакомы с основами прикладной математики. Не только математик, ставящий и решающий задачу в ее математической форме, но и инженер, формирующий исходные данные, должны хорошо понимать, чего они друг от друга хотят. И уж во всяком случае, на стадии построения математической модели необходимо совместное участие математиков и инженеров (или математика и инженера в одном лице), а в ряде случаев, также, экономистов и управленцев.

ГЛАВА 1

МАТЕМАТИЧЕСКИЕ МОДЕЛИ И ВЫЧИСЛИТЕЛЬНЫЕ АЛГОРИТМЫ

Знакомство с основами прикладной математики мы начинаем с понятия математической модели. Следует отметить, что в литературе, изданной в разное время, а тем более в разговорной речи в словосочетание «математическая модель» может вкладываться различный смысл. Для того чтобы избежать путаницы, мы разберем наиболее часто встречающиеся варианты трактовки этого понятия, но начнем с традиционного.

§1. Математические модели

Вы знакомы с математическими моделями и неоднократно пользовались ими, решая задачи по физике или математике, хотя, возможно, и не сталкивались прежде с самим этим термином. Типичным примером являются хорошо известные всем по школьной программе «задачи на движение». При всем разнообразии текстов этих задач и сформулированных в них обстоятельств перемещений различных объектов решаются они, как правило, с помощью уравнений, составленных на основании закона равномерного движения

$$S = vt, \quad (1.1)$$

где t – время движения того или иного объекта, v – скорость движения, S – путь, пройденный объектом за время движения.

Закон равномерного движения (1.1) предполагает, что скорость v во все время движения остается постоянной по величине. В то же время мы знаем, что любой поезд, автомобиль, велосипед и т.д., двигаясь «из пункта A в пункт B », сначала трогается, потом разгоняется до некоторой скорости, при этом он может неоднократно менять эту скорость из-за возможных подъемов и спусков, прибегать к притормаживанию и новому разгону, в зависимости от обстоятельств движения, и, наконец, постепенно снижает скорость до нуля в конце пути. Тогда почему же мы применяем закон (1.1) для описания движений, о которых нам известно, что они происходят с переменной скоростью?

Любой человек, знакомый с программой средней школы, ответит, что под скоростью v в формуле (1.1) подразумевается *средняя скорость* героя задачи (пешехода, поезда, автомобиля и т.п.) на всем участке AB или какой-то его части. Решая задачу, мы, фактически, предполагаем, что ее герой на каждом таком участке двигается с постоянной скоростью, равной средней скорости на этом участке. Поэтому мы и применяем для решения подобных задач формулу (1.1) в виде

$$S = v_{cp} t, \quad (1.2)$$

где v_{cp} – средняя скорость на каждом участке движения.

Но что такое средняя скорость на участке? Тот же, знакомый с программой средней школы человек, ответит, что средняя скорость – это величина, вычисляемая по формуле

$$v_{cp} = \frac{S}{t}, \quad (1.3)$$

где S – путь, пройденный участником движения, а t – время, за которое этот путь пройден. Но движение реальных физических объектов характеризуется мгновенной скоростью, своей для каждого момента движения и для каждой точки движущегося тела. Следовательно, средняя скорость на участке – это, скорее, не физическая характеристика движения, а некоторая абстрактная величина, придуманная нами для удобства расчетов. К тому же, вычислить ее можно лишь постфактум, когда движение уже закончено, и герой задачи прибыл в конечный (или промежуточный) пункт. До этого момента время его движения неизвестно, и в знаменатель формулы (1.3) просто нечего подставить.

Таким образом, решая «задачи на движение» мы каждый раз, сознавая или не сознавая это, используем два важных упрощения реальных физических движений, описанных в тексте задачи.

Во-первых, движение пешехода, поезда и т.п. мы рассматриваем как движение одной точки. Нас не интересуют движения рук и ног, вращение колес, деталей двигателей и т.п. Мы рассматриваем только движение центра масс или любой другой точки движущегося объекта (например, передней точки бампера автомобиля, угла передней двери автобуса и т.п.).

Во-вторых, мы считаем, что эта точка, заменяющая нам героя задачи, перемещается на каждом участке пути равномерно со скоростью v_{cp} , вычисляемой по формуле (1.3). Иными словами, мы сами определили понятие средней скорости так, чтобы выполнялся закон равномерного движения (1.2). Таким образом, мы заменяем реальный, сложный закон движения более простым, линейным. При этом, время прохождения каждой промежуточной точки дистанции, вычисленное по формуле (1.2), может отличаться от истинного, но время движения на всем участке AB или на какой-то его части совпадает со временем реального движения.

В результате этих двух серьезных упрощений мы решаем задачу с помощью линейного закона (1.1) (или, что то же самое, (1.2)). В таких случаях говорят, что мы использовали при решении задачи математическую модель равномерного движения.

Этот простой и хорошо всем известный пример мы разбираем так подробно для того, чтобы обратить внимание на ряд важных обстоятельств, характерных для всех, в том числе и гораздо более сложных математических моделей.

Прежде всего, математическая модель не адекватна самому описываемому явлению – движению реального поезда, автомобиля, пешехода и т.п., потому что не описывает всех деталей этого движения. Первое из сделанных выше упрощений оставляет из всех этих деталей лишь одну, наиболее существенную

для решения поставленной задачи – перемещение некоторой точки, связанной с объектом. Второе упрощение предписывает перемещению этой точки определенный закон – закон равномерного движения.

В результате сделанных упрощений движение транспортного средства (или пешехода) заменено абстрактной физической моделью – моделью равномерно движущейся точки. Далее для этой физической модели строится ее математическое описание в виде формулы (1.1). *Совокупность физической модели и ее математического описания* как раз и представляет собой *математическую модель*, с помощью которой мы решаем задачи на движение.

Рассмотрим еще один пример. Допустим, нужно определить площадь поверхности письменного стола. На практике поступают так: измеряют две смежные стороны поверхности стола (длину и ширину) и перемножают полученные числа. За этими элементарными действиями фактически скрывается следующее. Реальный объект – поверхность стола – заменяется абстрактной геометрической фигурой – прямоугольником (геометрическая модель). Этому прямоугольнику приписываются размеры, полученные в результате измерений, по соответствующей формуле (математическое описание) вычисляется его площадь, и ее величина принимается за площадь поверхности стола.

Выбор для поверхности стола модели прямоугольника основывается на наших зрительных ощущениях и прошлом опыте. В более серьезных случаях, когда требуется высокая точность, прежде чем воспользоваться моделью прямоугольника, ее нужно проверить. Для этого надо измерить длины противоположных сторон стола и длины диагоналей. Если с требуемой степенью точности длины противоположных сторон и длины диагоналей попарно равны, то поверхность стола действительно можно рассматривать как прямоугольник. В противном случае от модели прямоугольника придется отказаться и заменить ее другой геометрической моделью – плоским четырехугольником общего вида. В этом случае для вычисления площади придется разбить четырехугольник на два треугольника, и вычислять его площадь как сумму площадей этих треугольников, пользуясь, например, формулой Герона (в результате измерений нам уже известны длины всех сторон и диагоналей). Если требуется еще более высокая точность, то может возникнуть необходимость пойти в уточнении модели еще дальше, например, учесть закругления углов стола.

В разобранный пример математическая модель представляет собой *совокупность геометрической модели исследуемого объекта* – поверхности стола *и соответствующей вычислительной формулы* – формулы площади прямоугольника, формулы Герона или еще более сложной вычислительной процедуры, связанной с учетом закругления углов стола. Важно отметить, что *математическая модель не определяется однозначно исследуемым объектом*. Для одного и того же стола мы можем принять модель прямоугольника, либо более сложную модель плоского четырехугольника общего вида, либо еще более сложную модель четырехугольника с закругленными углами.

В приведенных выше примерах математическое описание строилось довольно легко. Во многих случаях это бывает сделать гораздо труднее. Иногда

математическое описание удастся построить только после очень серьезного упрощения физической модели*.

Бывает и так, что наши знания об изучаемом объекте недостаточны. Тогда при построении физической модели или ее математического описания приходится делать дополнительные предположения, которые носят характер *гипотез*. Модели, построенные на основе гипотез, называют *гипотетическими*. Выводы, полученные с помощью таких моделей, справедливы для изучаемого объекта настолько, насколько правильны исходные предположения. Оценить степень их правильности можно лишь, сопоставив результаты моделирования со всей имеющейся информацией об изучаемом объекте. Таким образом, вопрос о применимости и адекватности той или иной математической модели не является чисто математическим вопросом, и решить его исключительно математическими методами нельзя. Основным критерием пригодности и точности модели является эксперимент, практика в самом широком смысле этого слова.

В связи со сказанным, очень важно правильно представлять себе, какое место занимает математическое моделирование и сами математические модели в общей структуре научного и инженерного исследования.

§2. Структура математической модели и ее место в научном и инженерном исследовании

Пусть требуется изучить некоторый сложный объект, процесс или явление. Это может быть физический, химический, экономический процесс, явление природы или сложное устройство. В настоящее время выработалась технология теоретического исследования сложных объектов, допускающих математическое описание, – *вычислительный эксперимент*. Эта технология основана на построении и анализе (как правило, с помощью компьютера) математических моделей изучаемого объекта. Схема такого вычислительного эксперимента существенным образом зависит от того, на какой, теоретической или экспериментальной основе строится исследование и соответствующая математическая модель.

1. Модели «теоретического» типа

Остановимся сначала на таком распространенном случае, когда существует теория (область науки), описывающая процессы и явления, определяющие исследуемый нами объект. В этом случае схема вычислительного эксперимента выглядит так, как показано на рис. 1.1. Поясним существо этапов исследования, изображенных на схеме.

* Мы не делаем в данном случае различия между физической, геометрической либо какой бы то ни было другой моделью. Словосочетание *физическая модель* употребляется здесь в обобщающем смысле. Это дань традиции, связанной с тем, что математическое моделирование сформировалось первоначально как методика решения физических задач.

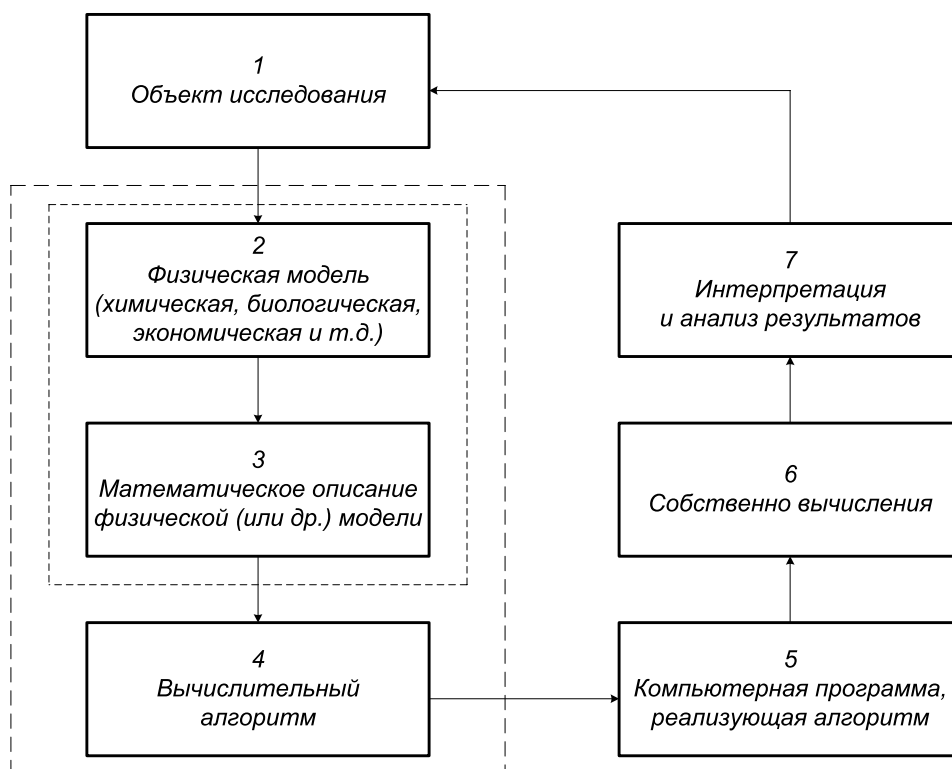


Рис. 1.1. Схема вычислительного эксперимента, использующего математическую модель «теоретического» типа

1. *Объект исследования.* Как уже говорилось, в качестве объекта исследования может выступать любой физический, химический, экономический процесс, явление природы, машина или другое сложное устройство, план перевозок, расписание движения транспортных средств и т. п. Это этап исходной постановки задачи. На нем формируется не только объект, но и цель всего исследования. Кроме того, именно на этом этапе зачастую формулируются требования к степени точности будущей модели, а также определяются ориентировочная себестоимость исследования, сроки его окончания и другие, «не имеющие прямого отношения к науке» условия.

2. *Физическая модель* объекта исследования. Название условное – это может быть химическая, биологическая, экономическая и т.д. модель. Одним словом, это модель объекта исследования, т.е. его упрощение, в рамках той области знания (или нескольких областей), в которой (в которых) рассматриваются подобные объекты. Замечания в скобках вызваны тем, что описываемый объект может представлять собой совокупность нескольких различных объектов или явлений, которые рассматриваются разными науками. На этапе формирования «физической модели» из всех свойств объекта исследования и всего множества его связей с другими объектами *выделяют самые существенные с точки зрения данного исследования.* Остальными свойствами и связями пренебрегают. При этом часто упрощают даже те свойства и связи, которые включили в рассмотрение.

3. *Математическое описание* физической модели. На данном этапе все свойства объекта исследования, сформулированные в п.2, описываются с помощью математических соотношений (как правило – уравнений, чаще всего – дифференциальных). На этом же этапе формируются такие важные условия, как границы значений переменных, в рамках которых эти математические соотношения остаются справедливыми.

Отвлекаясь на время от описания схемы вычислительного эксперимента, отметим, что совокупность «физической» модели и ее математического описания составляет математическую модель объекта в традиционном понимании этих слов. На рис. 1.1 упомянутая совокупность очерчена короткими пунктирными линиями. Такой взгляд на понятие математической модели был характерен для докомпьютерной эпохи развития прикладной математики, когда и постановка математической задачи, и ее решение (если его удавалось найти) строились в рамках одного и того же подхода – аналитического, и не было никакого смысла разделять их.

После появления современных вычислительных средств и бурного развития численных методов сложилась ситуация, когда постановка математической задачи в рамках математического описания физической модели и, собственно, решение задачи нередко разделены как методологически, так и субъектно (т. е. осуществляются разными людьми). В соответствии с этим сегодня естественно представлять математическую модель как совокупность *физической модели*, ее *математического описания* и *вычислительного алгоритма*, позволяющего построить приближенное решение соответствующей математической задачи, т.е. довести его «до числа». Эта совокупность очерчена на рис. 1.1 длинными пунктирными линиями.

Вернемся к описанию схемы вычислительного эксперимента.

4. *Вычислительный алгоритм* (или численный метод) представляет собой совокупность приемов, позволяющих свести сложную (или нерешаемую традиционными аналитическими методами) задачу к некоторой последовательности простых действий. Более подробно с понятием вычислительного алгоритма мы познакомимся в следующем параграфе.

5. *Компьютерная программа, реализующая алгоритм*, является завершением процесса создания математической модели, ее материальной реализацией. В разговорной речи нередко именно эту программу называют словами «математическая модель».

6. *Собственно вычисления*. Название этапа полностью отражает его содержание.

7. *Интерпретация и анализ результатов*. На этом этапе осуществляется сравнение результатов моделирования, полученных на предыдущем этапе, со всей имеющейся информацией об изучаемом объекте. По результатам сравнения оценивается степень точности и адекватности построенной модели. Если полученная степень точности не удовлетворяет исследователей (или «заказчика»), то модель уточняется. Уточнение начинается с пересмотра физической

модели, затем ее математического описания и т.д. После этого весь цикл вычислительного эксперимента начинается с начала.

Следует заметить, что с повышением точности модели неизбежно возрастает объем необходимых для ее реализации вычислений (и измерений). Выбор той или иной модели определяется балансом двух взаимоисключающих стремлений – к повышению точности модели и к ограничению роста объема вычислений. Рассмотрим в качестве примера хорошо известную задачу по механике.

2. Задача баллистики

Телу на Земле сообщили начальную скорость \mathbf{v}_0 , направленную под углом α к ее поверхности. Требуется найти траекторию движения тела, наивысшую точку траектории, а также расстояние по горизонтали между начальной и конечной точками траектории.

Для ответа на поставленные в задаче вопросы необходимо построить математическую модель движения тела, например, в виде уравнений его движения, т.е. определить его положение относительно некоторой системы координат в произвольный момент времени. Но построение математической модели требует конкретизации задачи, потому что под ее общую формулировку подходят слишком разные виды движений различных по величине и форме тел, с разными скоростями и дальностями полета. Естественно предположить, что различные конкретные задачи будут описываться разными математическими моделями. Рассмотрим несколько вариантов.

Вариант 1. Пусть речь идет о камне, брошенном рукой человека. В этом случае размеры и масса камня, а также его начальная скорость и дальность полета – невелики. Это позволяет построить физическую модель движения, основанную на следующих упрощающих предположениях:

- 1) камень можно считать материальной точкой;
- 2) Земля – инерциальная система отсчета;
- 3) кривизной поверхности Земли можно пренебречь и считать эту поверхность плоской;
- 4) ускорение свободного падения постоянно по величине и направлению $\mathbf{g} = \text{const}^*$;
- 5) действием воздуха на движущийся камень можно пренебречь.

Для математического описания этой физической модели введем декартову систему координат. Ее начало совместим с той точкой плоскости, на которой, условно говоря, стоит бросающий камень человек. Ось x направим горизонтально в сторону движения камня, ось y – вертикально вверх. Начальная точка движения камня A в этом случае будет иметь координаты $(0; h)$, где h – высота с которой бросают камень (рис. 1.2).

* Здесь и далее жирным шрифтом выделяются векторные величины. Обозначение **const** указывает, что данный вектор постоянен и по величине, и по направлению.

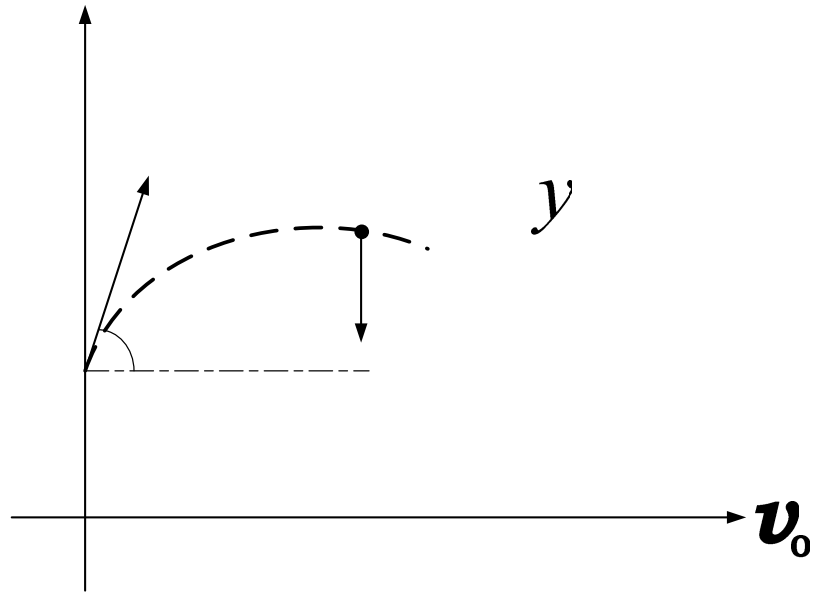


Рис. 1.2

Рассмотрим положение камня в произвольный момент времени. Согласно принятой нами физической модели он представляет собой материальную точку (на рис.1.2 это точка M с координатами (x, y)), движущуюся в плоскости xOy под действием одной только силы тяжести $\mathbf{P} = m\mathbf{g}$, направленной вертикально вниз. Движение материальной точки в инерциальной системе отсчета подчиняется второму закону Ньютона

$$m\mathbf{a} = \sum_i \mathbf{F}_i, \quad (1.4)$$

где \mathbf{a} – ускорение точки, m – ее масса, \mathbf{F}_i – силы, действующие на точку.

В рассматриваемом случае

$$m\mathbf{a} = m\mathbf{g}$$

или, учитывая, что масса точки отлична от нуля,

$$\mathbf{a} = \mathbf{g}. \quad (1.5)$$

Проектируя соотношение (1.5) на введенную нами систему координат, получаем дифференциальные уравнения движения камня как материальной точки

$$\begin{aligned} \ddot{x} &= 0, \\ \ddot{y} &= -g, \end{aligned} \quad (1.6)$$

где g – модуль вектора \mathbf{g} , а две точки над координатами обозначают двукратное дифференцирование по времени. Начальные условия движения имеют вид

$$\begin{aligned} x_0 &= 0, \\ y_0 &= h, \\ \dot{x}_0 &= v_0 \cos \alpha, \\ \dot{y}_0 &= v_0 \sin \alpha, \end{aligned} \quad (1.7)$$

где v_0 – модуль вектора \mathbf{v}_0 . Первые два из условий (1.7) представляют собой координаты начальной точки движения A , а два вторых – проекции вектора начальной скорости \mathbf{v}_0 на соответствующие оси координат.

Общее решение системы дифференциальных уравнений (1.6) имеет вид

$$\begin{aligned}x &= C_1 t + C_2, \\y &= -\frac{g}{2} t^2 + C_3 t + C_4,\end{aligned}\tag{1.8}$$

где t – время, C_1, \dots, C_4 – постоянные интегрирования. Определив значения этих постоянных из начальных условий (1.7), получаем закон движения материальной точки

$$x = t v_0 \cos \alpha,\tag{1.9}$$

$$y = -0,5 g t^2 + t v_0 \sin \alpha + h.\tag{1.10}$$

Формулы (1.9), (1.10) представляют собой математическую модель задачи при предположениях 1) – 5). Полученная модель довольно проста, и с ее помощью легко ответить на поставленные в задаче вопросы. Например, выразив из (1.9) время t через координату x

$$t = \frac{x}{v_0 \cos \alpha},$$

и подставив в (1.10), получим уравнение траектории камня

$$y = -x^2 \frac{g}{2 v_0^2 \cos^2 \alpha} + x \operatorname{tg} \alpha + h,\tag{1.11}$$

которая представляет собой параболу.

Мы не будем сейчас определять высшую точку траектории и дальность полета камня, поскольку перед нами стоит совсем другая задача. Отметим лишь, что в этом, первом варианте задачи нам удалось не только записать систему дифференциальных уравнений движения (1.6), но и проинтегрировать ее в общем виде, т. е. получить общее решение системы (1.8) и частное решение (1.9), (1.10), соответствующее начальным условиям движения (1.7).

Рассмотрим теперь другую конкретную задачу.

Вариант 2. Стрельба из старинной пушки ядрами шарообразной формы. Симметричность ядра позволяет и в этом случае использовать физическую модель материальной точки. Таким образом, предположение 1) остается в силе. Скорость и дальность полета ядра превосходят скорость и дальность полета камня в первом варианте задачи, но не настолько, чтобы серьезно нарушались предположения 2), 3) и 4). Их также можно оставить в силе. А вот от последнего, пятого предположения придется отказаться.

Значительные размеры пушечных ядер в сочетании с довольно большой скоростью полета вызывают заметное сопротивление воздуха, приводящее к отклонению их траектории от параболы (1.11). Пятое предположение мы переформулируем следующим образом: воздух оказывает сопротивление движению

ядра с силой, модуль которой пропорционален квадрату скорости, а направление противоположно направлению скорости. Таким образом, физическая модель задачи во втором случае основывается на следующих предположениях:

- 1) камень можно считать материальной точкой;
- 2) Земля – инерциальная система отсчета;
- 3) кривизной поверхности Земли можно пренебречь и считать эту поверхность плоской;
- 4) ускорение свободного падения постоянно по величине и направлению $\mathbf{g} = \text{const}$;
- 5) воздух оказывает сопротивление движению ядра с силой $\mathbf{F}_c \updownarrow \mathbf{v}$, причем

$$F_c = C_c v^2, \quad (1.12) \quad \text{где}$$

C_c – коэффициент сопротивления, зависящий от плотности воздуха, величины поперечного сечения тела и его формы.

Для математического описания этой физической модели воспользуемся той же системой координат, что и в первом случае. Единственное отличие заключается в том, что теперь на материальную точку M помимо силы тяжести действует еще и сила сопротивления воздуха \mathbf{F}_c , направление которой противоположно направлению вектора скорости точки M (рис. 1.3), а модуль задается формулой (1.12). Для вектора \mathbf{F}_c в этом случае справедливы следующие равенства

$$\mathbf{F}_c = C_c v^2 \left(-\frac{\mathbf{v}}{v} \right) \quad (1.13)$$

или

$$\mathbf{F}_c = -C_c v \mathbf{v}, \quad (1.14) \quad \text{где}$$

v – модуль вектора \mathbf{v} .

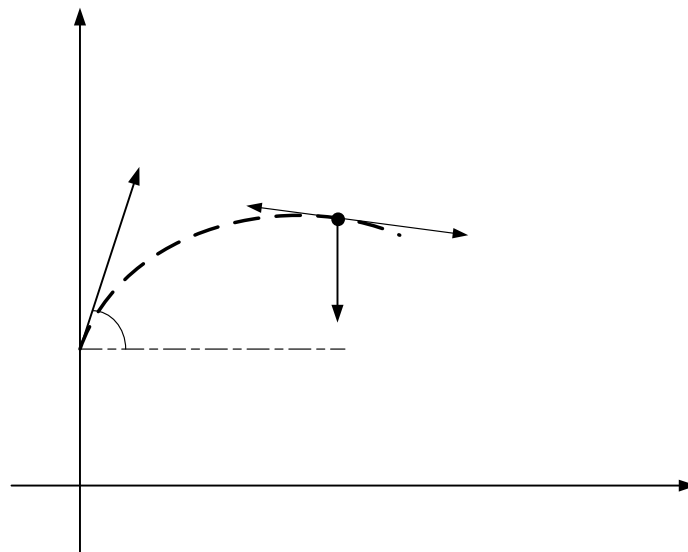


Рис. 1.3

Векторное уравнение второго закона Ньютона в этом случае запишется в виде

$$m \mathbf{a} = m \mathbf{g} - C_c v \mathbf{v}$$

или (поскольку масса ядра отлична от нуля)

$$\mathbf{a} = \mathbf{g} - \frac{C_c}{m} v \mathbf{v}. \quad (1.15)$$

Спроектируем (1.15) на оси введенной нами системы координат, учитывая, что вектор скорости \mathbf{v} в произвольный момент времени имеет координаты (\dot{x}, \dot{y}) , и тогда модуль этого вектора может быть вычислен по формуле

$$v = \sqrt{\dot{x}^2 + \dot{y}^2}. \quad (1.16)$$

В результате проектирования получаем систему дифференциальных уравнений, описывающих движение ядра как материальной точки

$$\begin{aligned} \ddot{x} &= -\frac{C_c}{m} \dot{x} \sqrt{\dot{x}^2 + \dot{y}^2}, \\ \ddot{y} &= -\frac{C_c}{m} \dot{y} \sqrt{\dot{x}^2 + \dot{y}^2} - g. \end{aligned} \quad (1.17)$$

Начальные условия движения так же, как и в первом случае, имеют вид (1.7).

Сравнение системы дифференциальных уравнений (1.17), описывающих движение ядра во втором варианте задачи, с системой (1.6) показывает, насколько сильно усложняет задачу учет сопротивления воздуха. А ведь мы ввели в первоначальную физическую модель всего одно усложнение.

Система (1.17) уже не допускает решения в виде простых аналитических формул и может быть решена только с помощью численных методов. На рис. 1.4 схематически показано, как отличается траектория полета ядра при наличии сопротивления воздуха (сплошная линия) от параболы, по которой двигалось бы ядро, если бы сопротивления воздуха не было (пунктирная линия). Хорошо заметное на рисунке различие в высоте и особенно в дальности полета ядра тем существенней, чем больше начальная скорость ядра и чем, следовательно, дольше продолжается полет ядра.

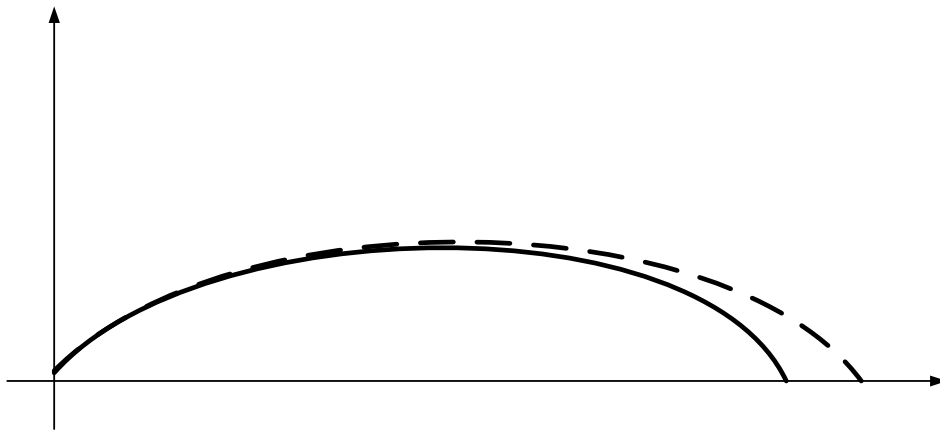


Рис. 1.4

В заключение рассмотрим еще один вариант этой задачи.

Вариант 3. Стрельба из современного нарезного дальнобойного орудия.

Дальность стрельбы в этом случае составляет десятки километров, а скорость снаряда превышает скорость звука. При таких скоростях около снаряда образуется волна сильного сжатия, которая расходится от него в виде конуса. На образование этой волны постоянно расходуется кинетическая энергия снаряда, что приводит к резкому возрастанию сопротивления воздуха. Модуль силы лобового сопротивления, как и в предыдущем случае, можно представить в виде (1.13), но коэффициент C_c для сверхзвуковых скоростей уже не будет постоянным. Он резко возрастает с увеличением скорости, т. е. является функцией последней.

Для уменьшения лобового сопротивления снаряду придают форму вытянутого тела вращения с заостренным передним концом, а для стабилизации полета ему придается вращение вокруг продольной оси. Такой снаряд уже нельзя рассматривать как материальную точку. Его необходимо рассматривать как твердое тело с шестью степенями свободы.

Большая дальность и связанная с ней большая высота полета снаряда в сочетании со значительной длительностью этого полета не позволяют пренебречь не только кривизной поверхности Земли, но и ее вращением вокруг своей оси. Следовательно, Землю в этом случае уже нельзя рассматривать как инерциальную систему отсчета. Кроме того, вектор ускорения свободного падения g уже не будет постоянным ни по величине (из-за большого перепада высот траектории полета), ни по направлению (вследствие сферичности поверхности Земли).

Таким образом, в третьем варианте задачи не выполняется ни одно из первоначально принимавшихся предположений 1) – 5). Более того, ситуация усложняется еще одним обстоятельством. Из-за вращения Земли всякое тело должно отклоняться от направления своего движения в Северном полушарии в правую сторону, а в Южном полушарии – в левую.* С учетом этого отклонения траектория центра масс снаряда перестает быть плоской линией даже при абсолютно спокойной атмосфере. Для описания такого движения уже недостаточно двух координат, да и сама система координат должна быть в этом случае не декартовой, а сферической.

Все перечисленные отличия физической модели в третьем варианте задачи приводят к очень значительному усложнению ее математического описания. Мы не будем выписывать систему дифференциальных уравнений, задающих движение снаряда в этом случае. Она слишком сложна и громоздка. Отметим лишь, что в эти уравнения входит большое количество величин, которые почти никогда не бывают известны точно. Например, распределение плотности

* С этим эффектом связаны два хорошо известных явления. У рек северного полушария правый берег обычно размывается сильнее левого, а на двухколейных железных дорогах быстрее снашивается правый рельс по ходу движения поезда.

воздуха на разных высотах, которое может сильно меняться в зависимости от температуры, влажности, давления, запыленности атмосферы и т.п. Кроме того, нельзя изготовить снаряды абсолютно одинаковыми. Они немного отличаются друг от друга по массе, по форме, по величине порохового заряда и, следовательно, по начальной скорости движения.

Наконец, невозможно дважды выстрелить под одним и тем же углом α к горизонту и в абсолютно одном и том же направлении.

Все эти и целый ряд других неконтролируемых факторов и неопределяемых точно величин приводят к тому, что два снаряда, выпущенные из орудия при одинаковых на первый взгляд условиях, никогда не попадут в одну и ту же точку. Из-за действия различных случайных факторов снаряды рассеиваются. Это значит, что любое предсказание полета снаряда, основанное на расчетах, носит не строго определенный (как говорят – детерминированный), а вероятностный характер.

Сложность подобного рода математических моделей и недостаточная определенность результата делает не только невозможным, но и бессмысленным их использование непосредственно на поле боя. Артиллеристы при ведении огня используют заранее составленные «таблицы стрельбы» для каждого конкретного типа орудий и снарядов, а также практику *коррекции огня* или *пристрелки*.

Подведем итог обсуждения математических моделей трех рассмотренных нами задач баллистики. Оно началось с простейшей модели, основанной на предположениях, которые сильно упрощают задачу. По такой модели легко вести расчеты, но она справедлива лишь в очень узком диапазоне скоростей: $v_0 \leq 30$ м/с.

Затем была рассмотрена модель, учитывающая сопротивление воздуха в виде (1.13). При этом мы считали коэффициент сопротивления C_c постоянным. Вести расчеты по такой модели значительно сложнее, но она справедлива почти во всем диапазоне дозвуковых скоростей $v_0 \leq 300$ м/с.

Наконец, при рассмотрении третьего варианта задачи мы убедились, что решать ее расчетным путем бессмысленно. Такой вывод справедлив не только по причине громоздкости и сложности получающейся в этом случае математической модели, но и потому, что неопределенность и даже случайность многих параметров, от которых зависит в данном случае полет снаряда, делает бессмысленным точное описание этого полета.

Управление огнем в дальнобойной артиллерии осуществляется по эмпирической модели, т.е. с помощью специальных таблиц, составленных опытным путем, и поправок, которые вводятся в ходе пристрелки. Схема вычислительного эксперимента, построенного на экспериментальной, эмпирической основе немного отличается от исследований, в основе которых лежит математическая модель «теоретического» типа.

3. Модели «эмпирического» типа

К этому типу моделирования прибегают в двух случаях. Во-первых, когда объект исследования недостаточно изучен и нет еще (по крайней мере, общепризнанной) теории, описывающей объекты подобного типа. И, во-вторых, в ситуациях, подобных третьему варианту задачи баллистики, т.е. когда теория есть, но объект исследования очень сложен и его математическое описание не только громоздко, но еще и зависит от большого количества случайных или неопределяемых заранее параметров. Схема вычислительного эксперимента в этом случае выглядит так, как показано на рис. 1.5. Поясним существо этапов исследования, изображенных на схеме.

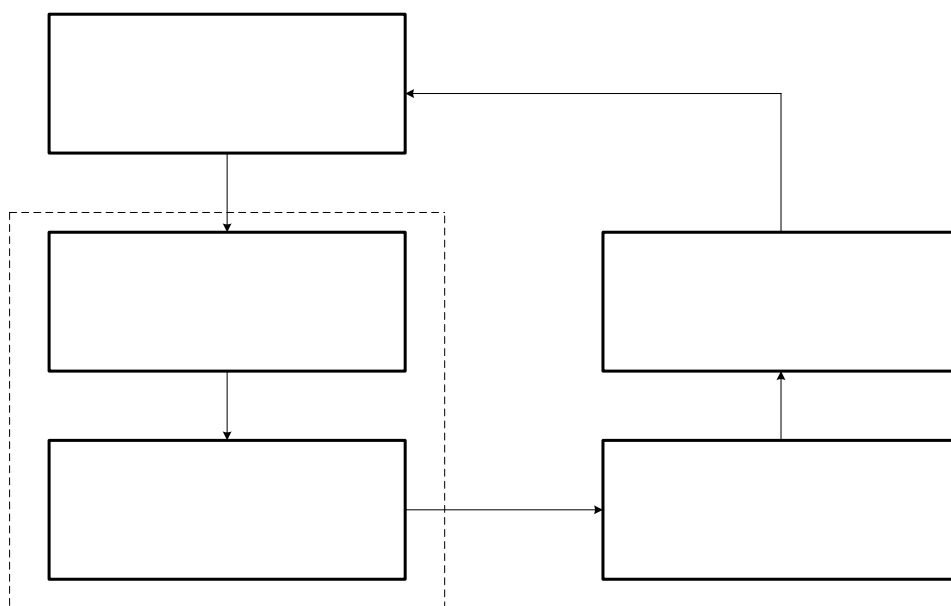


Рис. 1.5. Схема вычислительного эксперимента, использующего математическую модель «эмпирического» типа

1

1. *Объект исследования.* Содержание этого этапа вычислительного эксперимента точно такое же, как и в случае модели «теоретического» типа (см. рис. 1.1 и пояснения к нему).



Рис. 1.6. Схема объекта типа «черный ящик».

B_1, B_2, \dots, B_m – воздействия на объект
 P_1, P_2, \dots, P_n – реакции объекта

2. *Экспериментальное исследование объекта.* На этом этапе экспериментальным путем устанавливается, от чего и каким образом зависят те или иные параметры, характеризующие объект исследования. Это относится как к малоисследованным объектам, так и к сложно устроенным. В последнем случае мы отказываемся от использования известных нам теоретических знаний о свойствах объекта, а исследуем его как «черный ящик», т.е. объект, устройство которого не известно, а известно лишь, как он реагирует на те или иные воздействия (рис. 1.6).

3. *Математическая обработка результатов эксперимента.* На этом этапе исследования экспериментальные данные стараются представить по возможности в простой, аналитической форме. По сути своей – это тоже математическое описание (см. рис. 1.1 и пояснения к нему), но в данном случае математическими соотношениями описывается не сам объект или его физическая модель, а *результаты эксперимента*. Для разных экспериментов, проведенных в разных условиях и (или) по разным технологиям, могут быть получены несколько отличающиеся друг от друга результаты, и, следовательно, в процессе математической обработки этих результатов будут получены отличающиеся друг от друга математические соотношения.

Таким образом, экспериментальное исследование и математическая обработка результатов эксперимента тесно связаны друг с другом и образуют такую же совокупность, как физическая модель и ее математическое описание в моделях теоретического типа. Эту совокупность (на рис. 1.4 она очерчена короткими пунктирными линиями) называют математической моделью эмпирического типа. Поскольку объект исследования в моделях данного типа представляется в виде «черного ящика», и зависимости реакций объекта на произведенные воздействия стараются представить в виде аналитических функций (а не дифференциальных уравнений и т.п.), то необходимости в специальных вычислительных алгоритмах и реализующих их компьютерных программах обычно не возникает.

Содержание двух последних этапов вычислительного эксперимента такое же, как и в случае использования моделей теоретического типа. Заметим только, что под уточнением математической модели в данном случае понимается уточнение экспериментальных данных (в том числе увеличение их количества) и уточнение, либо пересмотр способа математического описания результатов эксперимента.

Следует заметить, что в историческом плане «эмпирическое» моделирование всегда предшествует «теоретическому». Всякое первоначальное исследование начинается с накопления информации об объекте и построения математической модели эмпирического типа. Последующее осмысление полученных зависимостей приводит к построению теорий, их обоснованию, уточнению и развитию. И лишь после этого появляется возможность строить математические модели теоретического типа.

Для более или менее изученных объектов математические модели эмпирического типа редко строятся «в чистом виде». Чаще всего встречаются моде-

ли смешанного типа. Эти модели в основном построены на теоретической основе, но содержат элементы эмпирики. Их структура и схема использования в вычислительном эксперименте близки к «теоретическому» типу.

В заключение подчеркнем еще раз, что математическое моделирование сводит изучение реального «нематематического» объекта к решению математической задачи, что позволяет использовать возможности хорошо разработанного математического аппарата и быстродействующей вычислительной техники.

§3. Вычислительные алгоритмы

После того, как сформулирована «физическая» модель рассматриваемого объекта и построено ее математическое описание, т.е. задача изучения объекта поставлена как математическая, наступает следующий важный этап исследования – поиск метода решения сформулированной математической задачи.

В большинстве задач, с которыми вы встречались до сих пор в математике, ответ давался в виде формулы. Всякая формула определяет последовательность математических операций, которую нужно выполнить для вычисления искомой величины. Например, формула корней квадратного уравнения задает последовательность действий, позволяющих либо найти эти корни по значениям коэффициентов уравнения, либо показать, что корней нет. Формулы (1.9), (1.10) задают последовательность действий, позволяющих для первого варианта задачи баллистики определить в заданной системе координат положение брошенного камня в любой момент времени в зависимости от начальной скорости и угла бросания.

В то же время, далеко не всякую последовательность действий удастся записать в виде формулы. Достаточно вспомнить правило вычисления суммы двух (или более) чисел путем поразрядного сложения столбиком. Это правило полностью решает поставленную задачу, поскольку определяет последовательность математических операций, которую нужно выполнить для вычисления искомой величины. Но попробуйте записать это правило в виде формулы!

В самом общем случае для решения математической задачи нужно указать систему правил, которая задает строго определенную последовательность математических операций, приводящих к искомому ответу. Такая система правил называется *алгоритмом*. В тех случаях, когда удастся построить алгоритм, приводящий к решению задачи, говорят, что решение получено в алгоритмическом виде. Так, например, правило сложения столбиком дает в алгоритмическом виде решение задачи о сумме любых двух (или более) чисел, записанных в виде десятичных дробей.

Понятие алгоритма относится к числу основных понятий математики. Можно сказать, что оно является обобщением понятия формулы. Действительно, всякая формула представляет собой символическую запись некоторой последовательности математических операций, то есть алгоритм, но не всякий алгоритм, как уже было сказано, может быть представлен в виде формулы.

Очень многие математические задачи, для которых не удастся получить ответ в виде формулы, могут быть решены в алгоритмическом виде. Алгоритмы решения таких задач чаще всего основаны на следующей процедуре: строится бесконечный процесс, сходящийся к искомому решению. Этот процесс обрывается на некотором шаге, т.к. вычисления нельзя продолжать бесконечно. Полученная на момент прерывания величина приближенно принимается за решение рассматриваемой задачи. Сходимость процесса гарантирует, что для любой заданной точности $\varepsilon > 0$ найдется такой номер шага N , что на этом шаге ошибка в определении решения задачи не превысит ε .

Бесконечный процесс стараются построить так, чтобы в нем все время повторялись одни и те же действия (группы действий). Такие процессы называют *итерационными*, а результаты их применения – *итерациями* (от латинского *iteratio* – повторение). Один из самых распространенных способов организации итерационного процесса – это вычисления по *рекуррентным формулам* (от латинского *resurgens* – возвращающийся). Так называются формулы, позволяющие выразить $(n + 1)$ -й член последовательности через значения ее первых n членов.

В качестве примера рассмотрим задачу об извлечении квадратного корня из произвольного положительного числа a . Эта задача может быть решена с помощью алгоритма, основанного на построении монотонной последовательности, сходящейся к \sqrt{a} . Выберем в качестве x_0 произвольное положительное число и рассмотрим последовательность $\{x_n\}$, определенную с помощью рекуррентной формулы

$$x_{n+1} = \frac{1}{2} \left(x_n + \frac{a}{x_n} \right), \quad n = 0, 1, 2, \dots, \quad (1.18) \text{ где}$$

a – положительное число, из которого нужно извлечь квадратный корень. Эта последовательность сходится к \sqrt{a} :

$$\lim_{n \rightarrow \infty} x_n = \sqrt{a}, \quad (1.19)$$

монотонно приближаясь к этому пределу сверху*.

Для иллюстрации этого алгоритма вычислим $\sqrt{10}$. За начальное приближение примем число $x_0 = 3,5$. Выпишем несколько первых членов последовательности (1.18):

$$\begin{aligned} x_1 &= 3,178\,571\,428; & x_4 &= 3,162\,277\,660; \\ x_2 &= 3,162\,319\,422; & x_5 &= 3,162\,277\,660. \\ x_3 &= 3,162\,277\,660; \end{aligned}$$

* Мы опускаем доказательство этого факта, как и большинство доказательств в дальнейшем, поскольку их рассмотрение не входит в цели, поставленные перед настоящим курсом. Желающие могут найти это доказательство, например, в книге А.Н. Тихонова, Д.П. Костомарова «Вводные лекции по прикладной математике».

Обратите внимание на то, что после третьего шага числа x_n перестали изменяться. Процесс «остановился». В этом проявляется принципиальная особенность вычислений с конечным числом значащих цифр. После третьего шага различие между двумя последовательными приближениями стало меньше, чем 10^{-9} . Чтобы продолжить расчет дальше нужно перейти к вычислениям с большим числом значащих цифр.

Этот алгоритм сходится очень быстро и не сильно зависит от начального приближения. Однако так бывает не всегда. Мы продолжим знакомство с алгоритмами в третьей главе, рассматривая задачу решения нелинейных уравнений.

§4. Приближенные вычисления

В предыдущем параграфе мы рассмотрели пример того, как бесконечный сходящийся процесс может быть использован для приближенного решения задачи. Нас не должны пугать слова «приближенное решение». Не следует воспринимать их как «решение второго сорта». Даже если в какой-нибудь задаче удастся найти ответ в виде формулы, то при попытке подсчитать по ней численное значение искомой величины, мы все равно получим приближенное значение. Это вызвано разными причинами. Во-первых, величины, входящие в нашу «точную» формулу, известны нам лишь приближенно (с той точностью, с какой их удалось измерить). Во-вторых, в любом вычислительном устройстве числа представляются в виде конечных десятичных (двоичных) дробей и при отбрасывании бесконечного «хвоста» дроби вносится некоторая погрешность. И, наконец, вспомним, что в прикладных исследованиях решение математической задачи дает нам лишь приближенную информацию об изучаемом объекте. Это следствие тех упрощений (а иногда и гипотетических предположений), которые мы приняли на стадии построения «физической» модели и ее математического описания.

Итак, вычислительные погрешности неизбежны в реальных расчетах, поэтому любому инженеру необходимо знакомство с правилами приближенных вычислений.

1. Источники и классификация погрешности

Погрешность решения задачи обусловлена следующими причинами:

1. Неточность физической модели и ее математического описания, в частности, неточность исходных данных задачи (как правило, вследствие неточности измерений).

2. Применяемый для решения метод требует неограниченного или неприемлемо большого числа арифметических операций, поэтому вместо получения точного решения задачи приходится прибегать к приближенному.

3. При вводе данных в вычислительное устройство, при выполнении арифметических операций и при выводе данных производятся округления.

Погрешности, соответствующие этим причинам, называют:

1) *неустранимой погрешностью*,

- 2) погрешностью метода,
- 3) погрешностью вычислений.

Очевидно, что нет никакого смысла применять метод решения задачи с погрешностью существенно меньшей, чем величина неустранимой погрешности.

2. Абсолютная и относительная погрешности

Пусть a – точное значение некоторой величины, а a^* – известное приближение той же величины. Разница между точным числом a и его приближенным значением a^* называется *погрешностью* приближенного числа a^* . Проблема в том, что определить величину погрешности невозможно. Для этого надо знать точное значение числа a . Но, если бы мы его знали, нам не понадобилось бы приближенное значение.

Итак, величина погрешности нам не известна. Но, как правило, из тех или иных соображений бывает можно оценить абсолютную величину погрешности сверху, т.е. указать такое число, которого абсолютная величина погрешности не превосходит

$$|a - a^*| \leq \Delta(a^*). \quad (1.20)$$

В этом случае число $\Delta(a^*)$ называют *предельной абсолютной погрешностью* приближенного числа a^* . Очень часто в качестве $\Delta(a^*)$ выступает цена деления шкалы того измерительного прибора, которым измерено значение a^* .

Величину

$$\delta(a^*) = \frac{\Delta(a^*)}{|a^*|} \quad (1.21)$$

называют *предельной относительной погрешностью* приближенного числа a^* . С учетом (1.20)

$$\delta(a^*) \geq \left| \frac{a - a^*}{a^*} \right|. \quad (1.22)$$

Предельную относительную погрешность часто выражают в процентах.

Для краткости слово «предельная» в названиях обеих погрешностей нередко опускают.

Тот факт, что a^* является приближенным значением числа a с предельной абсолютной погрешностью $\Delta(a^*)$, часто записывают в виде

$$a = a^* \pm \Delta(a^*). \quad (1.23)$$

Числа a^* и $\Delta(a^*)$ принято записывать с одинаковым количеством знаков после запятой, например

$$\begin{aligned} a &= 1,234 \pm 0,005; \\ a &= 1,234 \pm 5 \cdot 10^{-3}. \end{aligned}$$

Это означает, что

$$1,234 - 0,005 \leq a \leq 1,234 + 0,005.$$

Информацию, что a^* является приближенным значением числа a с предельной относительной погрешностью $\delta(a^*)$, записывают в виде

$$a = a^* (1 \pm \delta(a^*)). \quad (1.24)$$

Например, записи

$$\begin{aligned} a &= 1,234(1 \pm 0,003); \\ a &= 1,234(1 \pm 3 \cdot 10^{-3}); \\ a &= 1,234(1 \pm 0,3\%) \end{aligned}$$

означают, что

$$(1 - 0,003) \cdot 1,234 \leq a \leq (1 + 0,003) \cdot 1,234.$$

3. Теоремы о погрешностях

Результат арифметических действий над приближенными числами представляет собой также приближенное число. Погрешность результата может быть выражена через погрешности первоначальных данных при помощи следующих теорем.

Теорема 1. *Предельная абсолютная погрешность алгебраической суммы равна сумме предельных абсолютных погрешностей слагаемых.*

$$\begin{aligned} \Delta(a + b) &= \Delta(a) + \Delta(b); \\ \Delta(a - b) &= \Delta(a) + \Delta(b). \end{aligned} \quad (1.25)$$

Из этой теоремы вытекает важное следствие.

Следствие. Желательно избегать вычитания близких по значению чисел, поскольку разность в этом случае мала, а предельные абсолютные погрешности уменьшаемого и вычитаемого складываются (вторая из формул (1.25)). При неблагоприятных условиях предельная абсолютная погрешность результата может оказаться соизмеримой с самим результатом и даже превзойти его по абсолютной величине.

Пример

$$\begin{aligned} a &= 1004 \pm 1; \quad b = 1005 \pm 1; \\ b - a &= 1; \quad \Delta(b - a) = \Delta(b) + \Delta(a) = 2, \end{aligned}$$

таким образом, $b - a = 1 \pm 2$, т.е. величина предельной абсолютной погрешности разности чисел вдвое превосходит эту разность.

Теорема 2. *Предельная относительная погрешность суммы заключена между наибольшей и наименьшей из предельных относительных погрешностей слагаемых.*

Теорема 3. *Предельная относительная погрешность произведения и частного равна сумме предельных относительных погрешностей сомножителей или делимого и делителя.*

$$\begin{aligned} \delta(ab) &= \delta(a) + \delta(b); \\ \delta\left(\frac{a}{b}\right) &= \delta(a) + \delta(b). \end{aligned} \quad (1.26)$$

Теорема 4. *Предельная относительная погрешность p -й степени приближенного числа в p раз больше предельной относительной погрешности основания (как для целых, так и для дробных p).*

$$\delta(a^p) = p\delta(a). \quad (1.27)$$

Пользуясь этими теоремами, можно определить погрешность результата любой комбинации арифметических действий над приближенными числами. Однако при массовых вычислениях этот путь оказывается слишком трудоемким. Кроме того, следует помнить, что таким способом могут быть оценены только *предельные* погрешности, т.е. величины, заведомо превосходящие истинную погрешность. При этом все время предполагается, что различные погрешности усиливают друг друга, тогда как практически это бывает не часто. При массовых вычислениях не находят погрешность каждого промежуточного результата, а пользуются специальными *правилами подсчета верных цифр*.

4. Правила контроля верных цифр при массовых вычислениях с приближенными числами

Для изложения этих правил введем несколько новых терминов.

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

Пример. У чисел $a^* = 0,0123$, $a^* = 0,012300$ значащими цифрами являются подчеркнутые цифры. Число *значащих цифр* в первом случае равно 3, во втором 5.

Значащую цифру называют *верной*, если абсолютная погрешность числа не превосходит единицы разряда, соответствующего этой цифре.

Примеры. $a^* = 1,2345$, $\Delta(a^*) = 0,00002$; $a^* = 1,2345$, $\Delta(a^*) = 0,001$; подчеркнутые цифры являются верными.

Если все значащие цифры верные, то говорят, что число записано *со всеми верными цифрами*.

Пример. $a^* = 0,0123$, $\Delta(a^*) = 0,00001$; число a^* записано со всеми верными цифрами.

Иногда употребляется термин *число верных цифр после запятой*: подсчитывается число цифр от запятой до последней верной цифры. В последнем примере четыре верных цифры после запятой.

Правила контроля верных цифр при массовых вычислениях можно сформулировать так:

1. При сложении и вычитании приближенных чисел в результате следует сохранять столько *десятичных знаков*, сколько их в приближенном данном с наименьшим числом десятичных знаков.

2. При умножении и делении в результате следует сохранять столько *значащих цифр*, сколько их имеет приближенное данное с наименьшим числом значащих цифр.

3. При возведении в степень (в том числе, дробную) в результате следует сохранять столько же *значащих цифр*, сколько их имеет основание. (При этом последняя цифра квадрата, а тем более куба и т.д., менее надежна, чем последняя цифра основания; последняя цифра квадратного, а тем более кубического и т.д. корня, более надежна, чем последняя цифра основания.)

4. Во всех *промежуточных результатах* длительных вычислений следует сохранять *одной цифрой более*, чем рекомендуют предыдущие правила. В окончательном результате эта «запасная цифра» отбрасывается.

5. Если некоторые данные имеют больше *десятичных знаков* (при сложении и вычитании) или больше *значащих цифр* (при умножении, делении или возведении в степень), чем другие, то их предварительно следует округлить, сохраняя лишь одну лишнюю цифру.

При соблюдении этих правил можно считать, что в среднем полученные результаты будут иметь все знаки верными, хотя в отдельных случаях возможна ошибка в несколько единиц последнего знака.

ГЛАВА 2

МЕТОДЫ АНАЛИТИЧЕСКОГО ПРЕДСТАВЛЕНИЯ ЭКСПЕРИМЕНТАЛЬНЫХ И АЛГОРИТМИЧЕСКИ ЗАДАНЫХ ЗАВИСИМОСТЕЙ

В предыдущей главе мы отметили, что первоначальное исследование любого объекта или процесса начинается с накопления информации о них (как правило, экспериментальным путем) и построения математической модели эмпирического типа. Соответствующий этап математического моделирования называется «математическая обработка результатов эксперимента» (рис. 1.5). Текущая глава посвящена вопросам математической обработки результатов экспериментов, точнее – описанию этих результатов *в аналитическом виде*, т.е. *в виде математической формулы*.

Допустим, мы исследуем экспериментальными методами зависимость между двумя величинами: x и y . Любой эксперимент может быть поставлен лишь конечное количество раз. Это значит, что нам может быть известна из эксперимента связь между указанными величинами только в виде конечного количества пар их значений.

Пусть, для определенности, мы изменяли в ходе экспериментов величину x и измеряли зависимость от этих изменений величины y . Эту зависимость можно рассматривать как функцию $y = y(x)$, значения которой известны только при конечном числе изолированных значений аргумента x , то есть как функцию, заданную в табличном виде. Пример такой функции приведен в табл. 2.1.

Таблица 2.1

x	1,1	1,5	2,0	2,4	3,1	3,5	4,2	4,7	5,2
y	7,11	8,40	8,91	9,22	9,53	9,56	9,57	9,56	9,52

Заметим, что, абстрагируясь от способа проведения эксперимента и смысла величин x и y , можно рассматривать эту же зависимость как функцию вида $x = x(y)$.

Для построения математической модели объекта исследования необходимо научиться определять значения функции, заданной таблично, при значениях аргумента, не входящих в таблицу. Лучший способ сделать это – найти аналитическое выражение (формулу) для зависимости $y = y(x)$ (или $x = x(y)$). В этом случае можно вычислить значение функции для любого значения аргумента из области ее определения. Кроме того, функции, заданные в аналитическом виде, удобнее всего для исследования, поскольку они позволяют пользоваться обширным аппаратом математического анализа (дифференцирование, интегрирование и т.п.).

Таким образом, перед нами стоит задача заменить табличную функцию $y(x)$ приближенной формулой, т.е. подобрать такую функцию $\varphi(x)$, которая близка, в некотором смысле, к $y(x)$ (и в то же время достаточно просто вычис-

ляется). Замечание в скобках вызвано следующими двумя важными обстоятельствами:

1. Иногда связь между параметрами x и y известна, но соответствующая функция слишком сложна для непосредственного исследования (а иногда и для вычисления). В таких случаях функцию на каком-то участке области ее определения заменяют более простой приближенной формулой. При этом действуют так, как будто проводят экспериментальное исследование свойств этой функции. Вычисляют ее значения при некотором количестве значений аргумента и для полученной табличной функции подбирают приближенную формулу.

2. Иногда зависимость параметра y от параметра x задана в виде алгоритма или группы алгоритмов и реализована в виде некоторой компьютерной программы. В этом случае можно сказать, что функция $y = y(x)$ задана в алгоритмическом виде. Мы можем построить на экране монитора или на бумаге ее график, но не можем записать в виде формулы ни саму функцию, ни ее производную или первообразную (в чем может возникнуть необходимость при построении математической модели, включающей в себя эту функцию и (или) ее производную (первообразную)).

В таких случаях независимо от того, известен алгоритм работы программы или нет, с ней работают, как с объектом типа «черный ящик» (см. рис. 1.5 и пояснения к нему), воздействие на который задается параметром x , а реакция выражается параметром y . С этим объектом проводят серию вычислительных экспериментов и для полученной табличной функции подбирают приближенную формулу.

Итак, во всех трех рассмотренных случаях перед нами стоит одна и та же задача – заменить табличную функцию приближенной формулой. Существует несколько подходов к решению этой задачи. Мы рассмотрим в данной главе два наиболее простых и самых распространенных в инженерной практике подхода: *построение интерполяционного многочлена* и *построение эмпирической формулы по методу наименьших квадратов*. Первый из них обычно применяется в тех случаях, когда количество экспериментальных значений x и y (столбцов таблицы) невелико. Второй – напротив, в тех случаях, когда экспериментальных данных много.

§ 1. Аналитическое представление табличных функций в форме интерполяционного многочлена

1. Интерполирование табличных функций

Пусть функция $y = y(x)$ задана таблицей:

Таблица 2.2

x	x_0	x_1	\dots	x_{n-1}	x_n
y	y_0	y_1	\dots	y_{n-1}	y_n

Требуется найти значения функции при значениях аргумента x , не входящих в таблицу. Будем подбирать функцию $\varphi(x)$, приближающую табличную функцию так, чтобы в точках x_i она принимала те же значения y_i , что и исходная функция. Этот процесс называют *интерполяцией* (или *интерполированием*). Функция $\varphi(x)$ называется *интерполирующей функцией*. Табличные значения аргумента x_0, x_1, \dots, x_n называют *узлами интерполяции*. Чаще всего интерполирующую функцию ищут в виде многочлена.

$$P_n(x) = a_0 + a_1x + a_2x^2 + \dots + a_{n-1}x^{n-1} + a_nx^n, \quad a_n \neq 0. \quad (2.1)$$

Количество коэффициентов многочлена на единицу больше его порядка. Многочлен второй степени – это квадратный трехчлен. Многочлен первой степени – линейная функция. Многочлен нулевой степени – это константа. Два многочлена равны, т.е. их значения совпадают при любых значениях аргумента, если попарно равны все их коэффициенты при равных степенях.

$$P_n(x) = a_0x^n + a_1x^{n-1} + \dots + a_{n-1}x + a_n, \quad a_0 \neq 0, \quad (2.2)$$

Интерполяционным многочленом будем называть многочлен, коэффициенты которого выбраны таким образом, чтобы значения многочлена в узлах таблицы в точности совпадали со значениями табличной функции.

$$\left\{ \begin{array}{l} a_0 + a_1 x_0 + a_2 x_0^2 + \dots + a_n x_0^n = y_0, \\ a_0 + a_1 x_1 + a_2 x_1^2 + \dots + a_n x_1^n = y_1, \\ \\ a_0 + a_1 x_n + a_2 x_n^2 + \dots + a_n x_n^n = y_n. \end{array} \right. \quad (2.3)$$

Получилась система линейных (относительно неизвестных a_0, \dots, a_n) уравнений $(n + 1)$ -го порядка. Если все x_i различные, то определитель этой системы отличен от нуля, и система (2.3) имеет единственное решение. Отсюда следует, что для таблицы, не содержащей одинаковых узлов, существует единственный интерполяционный многочлен вида (2.1). Решив эту систему и найдя численные значения a_0, \dots, a_n , мы полностью определим интерполяционный многочлен для заданной табличной функции.

Отметим, что решение системы (2.3) при больших значениях n – задача весьма трудоемкая. Поэтому на практике описанным выше методом пользуются редко, а предпочитают составлять интерполяционные многочлены специального вида. Они не требуют решения системы (2.3). Ниже мы рассмотрим два типа таких многочленов.

2. Интерполяционный многочлен в форме Лагранжа

При интерполировании табличных функций с небольшим числом узлов удобно использовать интерполяционный многочлен Лагранжа. Этот многочлен составляют, используя значения x_i и y_i из таблицы 2.2, по формуле

$$L_n(x) = y_0 \frac{(x - x_1)(x - x_2) \dots (x - x_n)}{(x_0 - x_1)(x_0 - x_2) \dots (x_0 - x_n)} +$$

$$+ y_1 \frac{(x - x_0)(x - x_2) \dots (x - x_n)}{(x_1 - x_0)(x_1 - x_2) \dots (x_1 - x_n)} + \dots + y_n \frac{(x - x_0)(x - x_1) \dots (x - x_{n-1})}{(x_n - x_1)(x_n - x_2) \dots (x_n - x_{n-1})}. \quad (2.4)$$

Многочлен Лагранжа может быть составлен, если в табл. 2.2 нет одинаковых значений аргумента (в противном случае в одном из знаменателей окажется нулевой сомножитель). Каждое слагаемое в правой части (2.4) представляет собой многочлен степени n , поэтому и сам многочлен Лагранжа будет многочленом n -ой степени. Выражение (2.4) представляет собой общий вид многочлена Лагранжа для любого n , поэтому кажется чересчур громоздким. Рассмотрим процедуру построения этого многочлена на конкретном примере.

Пример 2.1

Пусть задана таблица значений функции $y = y(x)$.

Таблица 2.3			
x	1	3	4
y	2	4	3

Требуется построить для этой функции интерполяционный многочлен Лагранжа.

Решение. Число узлов равно трем, следовательно, многочлен Лагранжа будет многочленом второй степени.

$$L_2(x) = 2 \cdot \frac{(x - 3)(x - 4)}{(1 - 3)(1 - 4)} + 4 \cdot \frac{(x - 1)(x - 4)}{(3 - 1)(3 - 4)} + 3 \cdot \frac{(x - 1)(x - 3)}{(4 - 1)(4 - 3)}.$$

После преобразований получаем окончательный вид многочлена

$$L_2(x) = -\frac{2}{3}x^2 + \frac{11}{3}x - 1. \quad (2.5)$$

С помощью многочлена (2.5) теперь можно вычислить приближенное значение функции $y = y(x)$ при любом значении x из промежутка $x \in (1; 4)$, для которого составлена таблица.

3. Интерполяционный многочлен в форме Ньютона

Существует несколько разновидностей интерполяционного многочлена в форме Ньютона. Рассмотрим одну из них

$$P_n(x) = C_0 + C_1(x - x_0) + C_2(x - x_0)(x - x_1) + \dots \\ \dots + C_n(x - x_0)(x - x_1)\dots(x - x_{n-1}). \quad (2.6)$$

Коэффициенты C_i , $i = 0, 1, \dots, n$ вычисляются из условий совпадения значений многочлена в узлах интерполяции с соответствующими значениями функции, т.е.

$$P_n(x_i) = y_i, \quad i = 0, 1, \dots, n. \quad (2.7)$$

Условия (2.7) определяют систему линейных уравнений $(n + 1)$ -го порядка. Система имеет удобный для решения треугольный вид. При подстановке x_0 все слагаемые в правой части (2.6), кроме первого, содержат нулевую скобку. Это позволяет легко определить C_0 . При подстановке x_1 обращаются в нуль все слагаемые, кроме первых двух, и т.д., что позволяет найти все коэффициенты многочлена, решая каждый раз линейное уравнение с одним неизвестным. Разберем эту процедуру на примере.

Пример 2.2

Рассмотрим функцию $y = y(x)$, заданную табл. 2.3, и запишем для нее многочлен Ньютона второй степени (т.к. таблица содержит три узла)

$$P_2(x) = C_0 + C_1(x - 1) + C_2(x - 1)(x - 3).$$

Найдем последовательно коэффициенты C_0, C_1, C_2 , подставляя в это равенство пары значений x и y из табл. 2.3:

$$\begin{aligned} \text{при } x = 1 \quad P_2(1) &= 2, \text{ откуда } C_0 = 2; \\ \text{при } x = 3 \quad P_2(3) &= 4, \text{ откуда } C_1 = 1; \\ \text{при } x = 4 \quad P_2(4) &= 3, \text{ откуда } C_2 = -\frac{2}{3}. \end{aligned}$$

После подстановки найденных коэффициентов в многочлен

$$P_2(x) = 2 + 1(x - 1) - \frac{2}{3}(x - 1)(x - 3),$$

раскрытия скобок и приведения подобных получаем

$$P_2(x) = -\frac{2}{3}x^2 + \frac{11}{3}x - 1,$$

т.е. точно такой же многочлен, что и (2.5). Легко убедиться, что если бы мы искали интерполяционный многочлен для функции, заданной табл. 2.3, составляя систему уравнений (2.3), то и в этом случае пришли бы к такому же результату.

Итак, различные формы интерполяционного многочлена отличаются друг от друга лишь способом организации вычислений и, следовательно, их количе-

ством. Сам многочлен (при отсутствии совпадающих узлов) полностью определяется таблицей.

До появления современной вычислительной техники интерполяционные многочлены применяли, главным образом, по их прямому назначению, т.е. для вычислений междуузловых значений табличных функций. Все функции, как сложные, так и элементарные (синусы, косинусы, логарифмы) вычисляли тогда по специальным таблицам. При этом, сам многочлен обычно не находили, а требуемые значения функции вычисляли непосредственно по формулам (2.4) или (2.6), подставляя в них аргумент.

В этих случаях имело значение, в каком порядке производить вычисления. Многочлен (2.6.) был удобен для нахождения значений функции от аргументов, близких к x_0 , и за ним закрепилось название «многочлен для интерполирования вперед». При вычислении значений функции от аргументов, близких к x_n многочлен Ньютона удобнее было составлять в виде

$$P_n(x) = C_n + C_{n-1}(x - x_n) + C_{n-2}(x - x_n)(x - x_{n-1}) + \dots \\ \dots + C_0(x - x_n)(x - x_{n-1}) \dots (x - x_1), \quad (2.8)$$

который называют многочленом «для интерполирования назад». Разумеется, оба названия весьма условны и различия между формулами (2.6) и (2.8) сказываются лишь при разовых вычислениях. После выполнения всех преобразований они приводят к одному и тому же многочлену (при отсутствии совпадающих узлов таблицы).

4. Применение интерполяционных многочленов

Интерполирование вообще и, в частности, интерполирование с помощью интерполяционных многочленов оказывается полезным при решении следующих задач.

1. Сгущение таблиц (нахождение междуузловых значений табличных функций).
2. Построение аналитической формулы для экспериментально полученных зависимостей и функций, заданных в алгоритмическом виде.
3. Упрощение аналитического представления громоздких функций.

5. Недостатки интерполяционных многочленов высоких степеней

Степень интерполяционного многочлена жестко связана с количеством узлов табличной функции. По мере увеличения количества узлов (в силу, например, увеличения количества проведенных экспериментов) увеличивается степень интерполяционного многочлена и количество вычислений, связанных с его формированием.

Возникает вопрос, будет ли повышаться точность представления табличной функции многочленом по мере роста его степени. Иными словами, будет ли стремиться к нулю *погрешность интерполирования* $f(x) - P_n(x)$, если число узлов n неограниченно увеличивать? Ответ, вообще говоря, отрицательный. При больших n графики функций вида $P_n(x)$ начинают все больше «осциллировать», совершая иногда довольно значительные колебания около приближаемой функции. С увеличением n эта тенденция только усиливается.

На рис. 2.1 в качестве иллюстрации изображен график многочлена $P_8(x)$ при $0 \leq x \leq 1$, построенного для функции $y = |x|$ по равноотстоящим узлам на отрезке $[-1; 1]$.

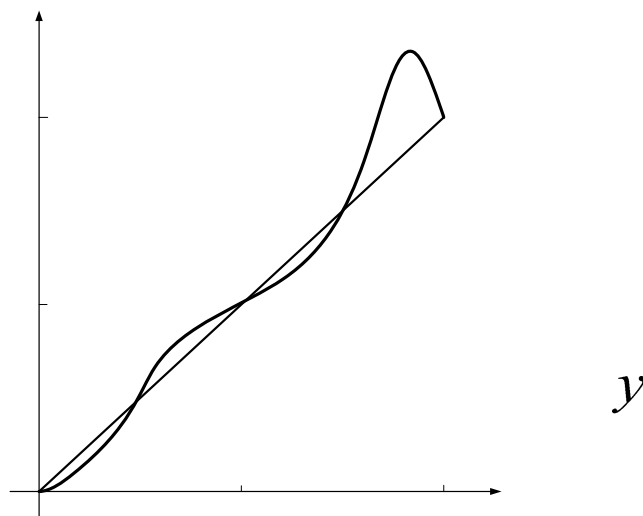


Рис. 2.1. График интерполяционного многочлена

для функции $y = |x|$

1,0

Складывается парадоксальная ситуация. Чем больше информации мы пытаемся учесть, тем хуже может стать приближение. Для того чтобы избежать больших погрешностей, весь отрезок $[a, b]$ разбивают на частичные отрезки и на каждом из частичных отрезков приближенно заменяют функцию $f(x)$ многочленом невысокой степени (так называемая кусочно-полиномиальная интерполяция).

§ 2. Аналитическое представление табличных функций в виде эмпирической формулы

Пусть в результате наблюдений некоторого процесса или при проведении эксперимента получена таблица значений двух величин x и y .

Таблица 2.3

x	x_1	x_2	\dots	x_i	\dots	x_n
y	y_1	y_2	\dots	y_i	\dots	y_n

Требуется подобрать формулу, описывающую приближенно функциональную зависимость $y = y(x)$, заданную этой таблицей. Как мы уже выяснили, метод интерполяции с его требованием совпадения значений функции с таблицей в узлах плохо подходит для приближения при большом количестве узлов.

Сформулируем задачу иначе. Будем строить такую приближающую функцию, которая не обязательно совпадает с табличными значениями в узлах, но, в некотором смысле, «не далеко отклоняется» от табличных значений. При

этом ее аналитическая формула должна содержать небольшое количество параметров, а их количество не должно зависеть от количества табличных точек. Функция, соответствующая сформулированным требованиям, называется *эмпирической формулой*.

Задача построения эмпирической формулы для табличной функции состоит из трех этапов:

1. *Структурная идентификация* эмпирической формулы (т.е. определение конкретного вида формулы).
2. *Параметрическая идентификация* эмпирической формулы (т.е. определение численных значений параметров, входящих в формулу).
3. *Оценка точности* эмпирической формулы (т.е. оценка расхождения между значениями эмпирической формулы и результатами эксперимента).

1. Структурная идентификация функции

На первом этапе построения эмпирической формулы определяют, в классе каких функций следует искать приближение. С этой целью пары значений аргумента и функции из табл. 2.3 изображают точками в некоторой системе координат. Сравнение точечного графика с различными кривыми, графики которых известны, нередко дает указание на возможный тип формулы.

Пример 2.1. На рис. 2.3 в декартовой системе координат построены точки, отражающие зависимость величины y от величины x , по результатам некоторого эксперимента.

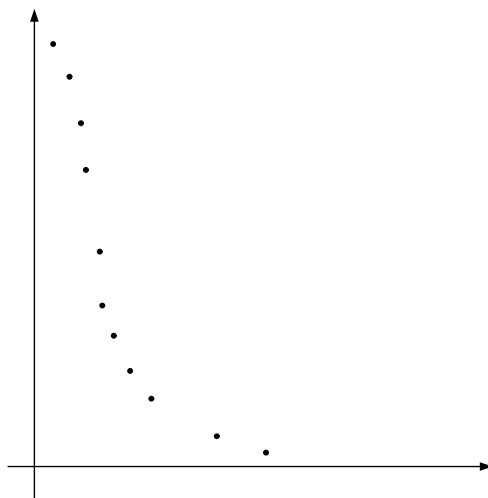


Рис. 2.3

Эту зависимость можно попытаться описать квадратичной функцией

$$y = a_0 + a_1x + a_2x^2 \quad (2.9)$$

или многочленом третьей степени

$$y = a_0 + a_1x + a_2x^2 + a_3x^3. \quad (2.10)$$

Коэффициенты a_0, a_1, \dots — неизвестные параметры этих функций. Но эту же зависимость можно приблизить функцией вида

$$y = a_0 + \frac{a_1}{x} \quad (2.11)$$

или экспонентой

$$y = a \cdot e^{bx}, \quad (2.12)$$

где b предполагается отрицательным.

Как видим, у поставленной задачи может быть несколько решений. Об оценке точности каждой из них мы поговорим подробно, когда будем рассматривать третий этап построения эмпирической формулы. Но начинать исследование надо всегда с наиболее простых функций, содержащих меньше всего параметров. В приведенном примере этому условию лучше всего удовлетворяет формула (2.11).

При рассмотрении графиков следует всегда иметь в виду, что при использовании эмпирическими формулами используется лишь часть кривой, соответствующая некоторому интервалу изменения независимой переменной. Поэтому, например, не следует думать, что формулы (2.9) и (2.10) удобны только при наличии у заданной кривой максимума или минимума.

Иногда в соответствии с экспериментом бывает можно построить грубую теорию изучаемого явления. Результаты теоретического исследования могут подсказать вид эмпирической формулы.

Пример 2.2. Изучается закон растворимости некоторого вещества в определенной жидкости. Нужно получить формулу, выражающую зависимость количества растворившегося вещества от времени.

За основу можно принять закон: количество вещества dx , растворившегося за малый промежуток времени dt , пропорционально этому промежутку и количеству нерастворенного вещества

$$dx = k(M - x)dt, \quad (2.13)$$

M – общее количество вещества, x – количество вещества, растворившегося до настоящего времени, $k > 0$ – коэффициент пропорциональности.

В качестве эмпирической формулы в этом случае можно выбрать общее решение дифференциального уравнения (2.13)

$$x(t) = M - c \cdot e^{-kt},$$

где c и k – неизвестные параметры.

В целом, этап структурной идентификации эмпирической формулы представляет собой наиболее неопределенную и творческую часть работы, которую иногда приходится повторять не один раз. Успешность проведения этого этапа во многом определяется опытом и кругозором исследователя.

2. Параметрическая идентификация функции. Метод наименьших квадратов

Рассмотрим вторую часть задачи о построении эмпирической формулы – определение численных значений входящих в формулу параметров. Эти параметры можно определять исходя из разных соображений. Один из самых распространенных методов выбора параметров – *метод наименьших квадратов*. Он заключается в таком выборе коэффициентов эмпирической функции, при

котором сумма квадратов всех отклонений значений функции от опытных данных минимальна.

Пусть эмпирическая формула имеет вид

$$y = f(x, a_0, \dots, a_m), \quad m < n, \quad (2.14)$$

где m – количество параметров эмпирической формулы, n – количество экспериментальных точек. Величина

$$\varepsilon_i = f(x_i, a_0, a_1, \dots, a_m) - y_i \quad (2.15)$$

задает отклонения при всевозможных значениях x_i (x_i, y_i взяты из табл. 2.3).

Наилучшими параметрами считаются те, для которых сумма

$$S(a_0, \dots, a_m) = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n [f(x_i, a_0, \dots, a_m) - y_i]^2 \quad (2.16)$$

будет минимальной. Условием минимума функции $S(a_0, \dots, a_m)$ является равенство нулю ее частных производных по параметрам a_0, \dots, a_m . Из этого условия получается система уравнений:

$$\left\{ \begin{array}{l} \frac{\partial S}{\partial a_0} = 2 \sum_{i=1}^n [f(x_i, a_0, ..., a_m) - y_i] \cdot \frac{\partial f}{\partial a_0} = 0, \\ \\ \frac{\partial S}{\partial a_m} = 2 \sum_{i=1}^m [f(x_i, a_0, ..., a_m) - y_i] \cdot \frac{\partial f}{\partial a_m} = 0. \end{array} \right. \quad (2.17)$$

Самый простой вид система (2.17) имеет в случае линейной зависимости $y = a_0 + a_1 x$. В ней два параметра, следовательно, в системе будет два уравнения:

$$\begin{cases} \frac{\partial S}{\partial a_0} = 2 \sum_{i=1}^n (a_0 + a_1 x_i - y_i) = 0, \\ \frac{\partial S}{\partial a_1} = 2 \sum_{i=1}^n (a_0 + a_1 x_i - y_i) \cdot x_i = 0. \end{cases} \quad (2.18)$$

После преобразований получается следующая система линейных относительно параметров a_0 и a_1 уравнений:

$$\begin{cases} a_0 n + a_1 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i, \\ a_0 \sum_{i=1}^n x_i + a_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i. \end{cases} \quad (2.19)$$

Известно, что определитель такой системы отличен от нуля, поэтому коэффициенты a_0 и a_1 вычисляются однозначно.

Если эмпирическая формула имеет вид $y = a_0 + a_1x + a_2x^2$, то параметры a_0, a_1, a_2 находят из следующей системы уравнений

$$\begin{cases} a_0 n + a_1 \sum_{i=1}^n x_i + a_2 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i, \\ a_0 \sum_{i=1}^n x_i + a_1 \sum_{i=1}^n x_i^2 + a_2 \sum_{i=1}^n x_i^3 = \sum_{i=1}^n x_i y_i, \\ a_0 \sum_{i=1}^n x_i^2 + a_1 \sum_{i=1}^n x_i^3 + a_2 \sum_{i=1}^n x_i^4 = \sum_{i=1}^n x_i^2 y_i. \end{cases} \quad (2.20)$$

3. Оценка точности эмпирической формулы

Для оценки точности эмпирической формулы используют понятие *среднеквадратичного уклонения*. Эта величина задается выражением

$$\varepsilon = \sqrt{\frac{1}{n} \sum_{i=1}^n \varepsilon_i^2}, \quad (2.21)$$

где ε_i определяются по формуле (2.15). Среднеквадратичное уклонение показывает среднюю величину отклонения опытных значений исследуемой зависимости от расчетных, полученных по эмпирической формуле.

Каждое измерение в эксперименте производится с некоторой погрешностью, и табличные значения функции $y = f(x)$ отличаются от истинных. Одна из целей построения эмпирической формулы – сглаживание случайных погрешностей измерений. Величину ε используют для определения пригодности эмпирической формулы. Если число параметров формулы намного меньше, чем точек в таблице, а значение ε примерно равно погрешности экспериментальных данных, то формулой можно пользоваться. Если величина среднеквадратичного уклонения ε намного больше или намного меньше, чем погрешности табличных значений, то следует поискать другой, более подходящий вид эмпирической формулы.

Для иллюстрации всех трех этапов построения эмпирической формулы рассмотрим следующий простой пример.

Пример 2.3. В эксперименте исследовалась реакция некоторого измерительного прибора на ударные воздействия. В ходе исследований на прибор с различной высоты h бросали без начальной скорости груз фиксированной массы и измеряли время t , через которое прибор выходил на нормальный режим работы. По результатам эксперимента получена следующая таблица значений

Таблица 2.4					
$h(\text{дм})$	1	2	3	4	5
$t(\text{сек})$	2	2,8	4,1	5	6,1

Требуется найти эмпирическую формулу, выражающую зависимость $t = f(h)$, определить методом наименьших квадратов ее параметры и вычислить среднеквадратичное уклонение.

Решение. Изобразим точки (h_i, t_i) в декартовой системе координат (рис. 2.4):

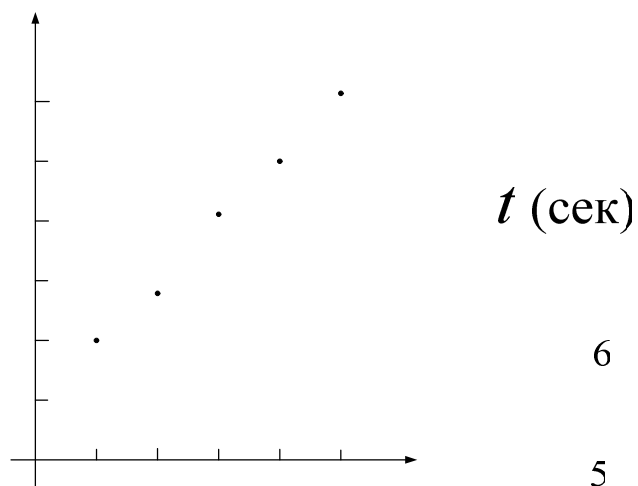


Рис. 2.4

На графике видно, что точки лежат вдоль некоторой прямой. В качестве эмпирической формулы выбираем линейную функцию $t = a_0 + a_1 h$.

Параметры a_0, a_1 , определим из системы уравнений

$$\begin{cases} a_0 \cdot 5 + a_1 \sum_{i=1}^n h_i = \sum_{i=1}^n t_i, \\ a_0 \sum_{i=1}^n h_i + a_1 \sum_{i=1}^n h_i^2 = \sum_{i=1}^n h_i t_i. \end{cases} \quad (2.22)$$

Для того чтобы найти коэффициенты этой системы, проведем предварительные расчеты, результаты которых сведем в следующую таблицу

Таблица 2.5

i	h_i	t_i	h_i^2	$h_i \cdot t_i$	$t(h_i)$	ε_i	ε_i^2
1	1	2	1	2	1,92	-0,08	0,0064
2	2	2,8	4	5,6	2,96	0,16	0,0256
3	3	4,1	9	12,3	4	-0,1	0,01
4	4	5	16	20	5,04	0,04	0,0016
5	5	6,1	25	30,5	6,08	-0,02	0,0004
$\sum_{i=1}^5$	15	20	55	70,4			0,044

Последняя строка таблицы содержит коэффициенты системы. Подставляя их в (2.22), получаем

$$\begin{cases} 5a_0 + 15a_1 = 20, \\ 15a_0 + 55a_1 = 70,4. \end{cases} \quad (2.23)$$

Решая эту систему, находим: $a_0 = 0,88, a_1 = 1,04$. Окончательный вид эмпирической формулы $t = 0,88 + 1,04h$.

Для вычисления среднеквадратичного отклонения заполним три последних столбца табл. 2.5:

$t(h_i)$ – значения, полученные по найденной эмпирической формуле в точках h_i ;

$\varepsilon_i = t(h_i) - t_i$ – отклонения между теоретическими и опытными значениями.

Суммируя значения последнего столбца, вычислим среднеквадратичное отклонение

$$\varepsilon = \sqrt{\frac{1}{5} \sum_{i=1}^5 \varepsilon_i^2} = \sqrt{\frac{0,044}{5}} = \sqrt{0,0088} \approx 0,0938.$$

В табл. 2.4. значения времени указаны с точностью $\pm 0,1$ секунды. Полученное нами среднеквадратичное отклонение имеет примерно тот же порядок. Это значит, что построенная эмпирическая формула удачна.

В заключение сделаем одно важное замечание. Если число параметров эмпирической формулы совпадает с числом точек в таблице, то эмпирическая формула переходит в интерполяционную. Поскольку исходные данные обычно бывают измерены с некоторой погрешностью, а значения интерполяционной формулы совпадают с измеренными значениями в узлах интерполяции, то подобная формула их не сглаживает, а просто повторяет. Среднеквадратичное отклонение в этом случае, очевидно, равно нулю.

4. Способы сведения сложных эмпирических формул к простым.

Системы уравнений (2.19) и (2.20) получены нами для коэффициентов эмпирических формул, представляющих собой линейную и квадратичную зависимости. В тех случаях, когда связь между исследуемыми величинами носит более сложный характер, стараются с помощью подходящей замены переменной свести эти сложные зависимости к линейной или квадратичной функции. Неизвестные параметры новых функций определяют с помощью метода наименьших квадратов. После этого возвращаются к исходным переменным.

Рассмотрим примеры таких преобразований.

1. Пусть выбрана эмпирическая формула

$$y = a_0 + \frac{a_1}{x}. \quad (2.24)$$

Вводя новую переменную $t = \frac{1}{x}$, получим из эмпирической формулы линейную функцию

$$y = a_0 + a_1 t. \quad (2.25)$$

Составим новую таблицу значений для переменных t и y , в которой значения t_i вычисляются по формуле $t_i = \frac{1}{x_i}$, а значения y_i оставлены прежние. Неизвестные параметры a_0, a_1 определяют, применяя метод наименьших квадратов к линейной зависимости (2.25). Найденные параметры используют в формуле (2.24).

2. Если выбрана эмпирическая формула

$$y = a_0 + \frac{a_1}{x} + \frac{a_2}{x^2}, \quad (2.26)$$

то введя новую переменную $t = \frac{1}{x}$, приходим к квадратичной функции $y = a_0 + a_1 t + a_2 t^2$. Далее, действуя так же, как в предыдущем случае, находим значения параметров a_0, a_1, a_2 . Найденные параметры используют в формуле (2.26).

3. Рассмотрим эмпирическую формулу $y = ax^b$. Прологарифмируем ее: $\ln y = \ln a + b \cdot \ln x$. Обозначим $z = \ln y$, $a_0 = \ln a$, $t = \ln x$. Получим линейную зависимость $z = a_0 + bt$. Составляем новую таблицу значений переменных z_i, t_i , соответствующих заданным x_i, y_i , далее, методом наименьших квадратов определяются коэффициенты a_0, b . Значение a вычисляется по формуле $a = e^{a_0}$. Это значение так же, как и b , подставляется в исходную эмпирическую формулу.

4. Пусть выбрана эмпирическая формула $y = a \cdot e^{bx}$. Логарифмируя ее, найдем $\ln y = \ln a + bx$. Обозначив $z = \ln y$, $a_0 = \ln a$, получаем линейную функцию $z = a_0 + bx$. Вычислив значения z_i , соответствующие y_i , составляем новую таблицу и находим коэффициенты линейной функции методом наименьших квадратов. Значение a вычисляется по формуле $a = e^{a_0}$.

5. Рассмотрим функцию

$$y = \frac{c}{ax + b}. \quad (2.27)$$

Переворачивая дробь, получаем равенство $\frac{1}{y} = \frac{ax + b}{c} = \frac{a}{c}x + \frac{b}{c}$. Обозначив $z = \frac{1}{y}$, $a_1 = \frac{a}{c}$, $a_0 = \frac{b}{c}$, приходим к линейной функции $z = a_0 + a_1 x$. Вычислим значения z_i , соответствующие y_i . Коэффициенты линейной зависимости $z = a_0 + a_1 x$ найдем по методу наименьших квадратов. Коэффициенты исходной функции (2.27) найдем по формулам $a = a_1 c$, $b = a_0 c$, где число c может быть выбрано произвольно. Если, например, числа a_0 и a_1 представляют собой обыкновенные дроби, то в качестве c удобно выбрать их наименьший общий знаменатель. В этом случае все три коэффициента формулы (2.27) будут целыми числами.

Разумеется, перечень возможных замен переменных, позволяющих свести ту или иную функцию к линейной или квадратичной зависимости, можно продолжать. Круг функций, позволяющих сделать такую замену, весьма широк и ограничивается только кругозором и опытом исследователя. Важно лишь, пользуясь этим приемом, неизменно соблюдать следующее правило.

Во всех случаях, при оценке точности выбранной эмпирической формулы, надо вычислять среднеквадратичное отклонение для ее исходного вида, а не для той функции, к которой мы сводим эту формулу путем замены переменных.

ГЛАВА 3. ЧИСЛЕННЫЕ МЕТОДЫ РЕШЕНИЯ НЕЛИНЕЙНЫХ УРАВНЕНИЙ С ОДНИМ НЕИЗВЕСТНЫМ

В теоретических и прикладных исследованиях и расчетах приходится решать различные математические задачи. Одна из самых часто встречающихся среди них – это решение уравнений.

В школьном курсе математики изучаются линейные и квадратные уравнения, корни которых могут быть найдены по известным формулам. Заметим, что линейные и квадратные уравнения являются частными случаями особой группы уравнений, называемых *алгебраическими уравнениями*. К этой группе относятся уравнения вида

$$a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0 = 0, \quad a_n \neq 0. \quad (3.1)$$

Помимо уже упомянутых формул для решения линейных и квадратных уравнений существуют также формулы для решения алгебраических уравнений третьей и четвертой степеней, но они очень сложны и неудобны для практического применения (мы не будем их рассматривать, чтобы не отвлекаться от основной темы). Что же касается уравнений более высоких степеней, то найти явные выражения для их корней удастся только в очень редких частных случаях. Доказано, что в общем случае решения уравнения (3.1) при $n \geq 5$ нельзя выразить через коэффициенты с помощью арифметических действий и операций извлечения корней.

Если рассматривать неалгебраические уравнения (к ним относятся тригонометрические, показательные, логарифмические, дробно-рациональные, а также, смешанные, содержащие функции разных типов), то в этом случае найти для корней явные выражения, за редким исключением, не удастся*.

Рассмотрим в качестве примера очень простое уравнение

$$\cos x = x. \quad (3.2)$$

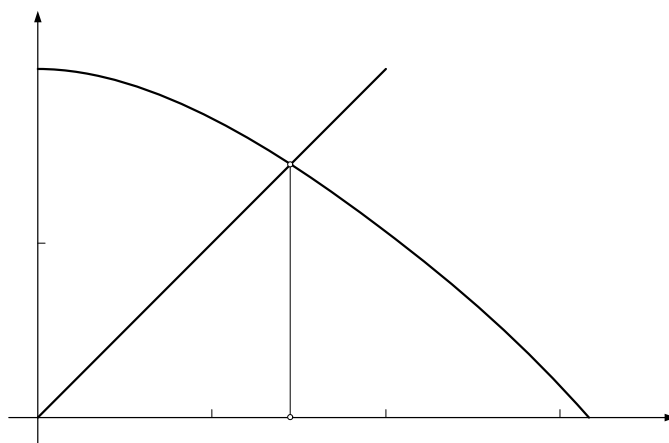


Рис. 3.1.

*Рассматриваемые в школьном курсе тригонометрические, показательные и логарифмические уравнения как раз и представляют собой эти редкие исключения.

Построим графики функций, стоящих в левой и правой части (рис. 3.1). Как видно из рисунка, графики пересекаются при $x = c$, ($0 < c < 1$). Число c – корень уравнения (3.2), однако получить формулу для его вычисления невозможно.

В подобных ситуациях, когда не удастся построить формулы для отыскания корней уравнений, большое значение приобретают вычислительные алгоритмы, позволяющие строить последовательности чисел, сходящиеся к искомым корням. Мы рассмотрим некоторые из них.

Пусть задано произвольного вида уравнение с одним неизвестным. Запишем это уравнение в виде

$$f(x) = 0, \quad (3.3)$$

где $f(x)$ – функция действительного переменного. Требуется найти действительные* корни уравнения (3.3) или, что то же самое, нули функции $f(x)$.

При численном решении уравнения (3.3) приходится решать две задачи:

1) *отделение корней*, включающее в себя определение количества корней и их локализацию, т.е. отыскание достаточно малых областей, в каждой из которых заключен один (и только один) корень уравнения;

2) *уточнение корней*, т.е. вычисление локализованных корней с заданной точностью.

Разберем подробно методы решения этих задач.

§ 1. Отделение корней уравнения

1. Основная теорема

При решении уравнения (3.3) очень важно *отделить* его действительные корни, т.е. найти достаточно малые числовые промежутки, содержащие корни этого уравнения. В этом может помочь следующая теорема.

Теорема о существовании корня непрерывной функции. *Если функция $f(x)$ непрерывна на отрезке $[a, b]$ и принимает на его концах значения разных знаков, то на этом отрезке уравнение (3.3) имеет хотя бы один корень.*

В дальнейшем, на протяжении всей главы, мы будем называть эту теорему «*основной теоремой*».

Требование непрерывности функции $f(x)$ во всех точках отрезка $[a, b]$ существенно. При наличии хотя бы одной точки разрыва (неважно, первого или второго рода) утверждение теоремы становится неверным. В качестве примеров на рис. 3.2 приведены графики двух разрывных функций. Одна из них терпит разрыв

* Уже на примере алгебраических уравнений (например, квадратного) известно, что корни уравнений могут быть как действительными, так и комплексными. Поэтому более точная постановка задачи состоит в нахождении корней уравнения (3.3), расположенных в заданной области комплексной плоскости. Мы будем рассматривать в дальнейшем наиболее важный для технических приложений случай действительных корней. Желающие ознакомиться с методами нахождения комплексных корней могут воспользоваться более полной специальной математической литературой.

первого рода, другая – разрыв второго рода. Каждая из функций принимает на концах отрезка $[a, b]$ значения разных знаков, но не имеет на этом отрезке корней.

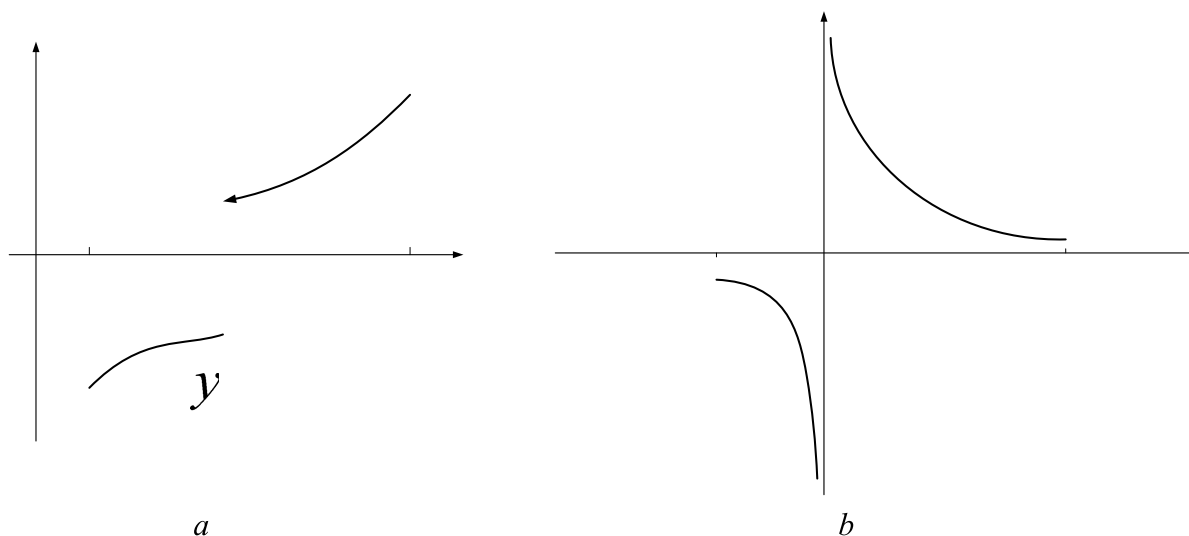


Рис. 3.2. Примеры разрывных функций, принимающих на концах отрезка $[a, b]$ значения разных знаков, но не имеющих на этом отрезке корней:
 a – разрыв первого рода; b – разрыв второго рода

Следует обратить внимание на то, что, гарантируя существование решения уравнения (3.3), теорема не позволяет определить количество его корней. На рис. 3.3 в качестве примера изображен график функции, удовлетворяющей условиям теоремы и имеющей на рассматриваемом отрезке четыре корня.

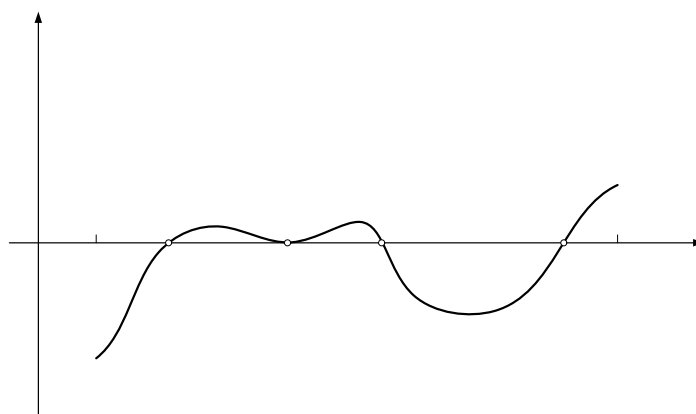


Рис. 3.3. Пример непрерывной функции, имеющей на отрезке $[a, b]$ четыре корня

В ряде случаев некоторую помощь в определении количества корней на отрезке $[a, b]$ может оказать следующее утверждение.

Если на отрезке $[a, b]$ выполнены условия основной теоремы и при этом функция $f(x)$ имеет первую производную, не меняющую знака на отрезке $[a, b]$, то уравнение (3.3) имеет на этом отрезке единственный корень.

2. Свойства действительных корней алгебраических уравнений

Из основной теоремы следуют два полезных свойства действительных корней алгебраических уравнений, т.е. уравнений вида

$$P_n(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0 = 0, \quad a_n \neq 0. \quad (3.4)$$

Теорема 1. Всякое алгебраическое уравнение нечетной степени имеет по крайней мере один действительный корень.

Действительно, функция $P_n(x)$ непрерывна на всей числовой прямой. Знак многочлена при достаточно больших по модулю значениях x совпадает со знаком его старшего члена $a_n x^n$. В силу нечетности n этот знак различен для отрицательных и положительных значений x . Отсюда следует утверждение теоремы.

Теорема 2. Если в алгебраическом уравнении четной степени n знаки коэффициентов a_n и a_0 противоположны, то это уравнение имеет по крайней мере один отрицательный и один положительный корень.

Предположим для определенности, что $a_n > 0, a_0 < 0$. Тогда при больших по модулю значениях x (как положительных, так и отрицательных) многочлен $P_n(x)$, как и его старший член $a_n x^n$, принимает положительное значение. В то же время при $x = 0$ он принимает отрицательное значение: $P_n(x) = a_0 < 0$. Отсюда следует утверждение теоремы.

Можно указать и другие свойства действительных корней алгебраических уравнений, помогающие их локализации. Например, хороший способ отыскания верхней границы положительных корней указал Ньютон. Этот способ основан на утверждении: *если при $x = a > 0$ выполняются неравенства $P_n(a) > 0, P'_n(a) > 0, P''_n(a) > 0, \dots, P_n^{(n)}(a) > 0$, то уравнение (3.4) не имеет корней, больших a .*

И все же, наибольший интерес для нас представляют неалгебраические уравнения, т.е. уравнения самого общего вида (3.3). В этом случае весьма эффективными оказываются графические методы локализации корней.

3. Графическая локализация корней

Для отыскания грубых приближенных значений действительных корней уравнения (3.3) можно построить график функции $y = f(x)$ и найти абсциссы точек пересечения графика с осью x (рис. 3.3). Иногда удобнее сначала представить уравнение в виде $\varphi(x) = \psi(x)$ и затем, построив графики функций $y = \varphi(x)$ и $y = \psi(x)$, найти абсциссы точек их пересечения, которые и будут приближенными значениями корней (именно так мы поступили с уравнением (3.2) на рис. 3.1).

Рассмотрим следующий пример. Пусть нужно найти корни уравнения

$$x \sin x - 1 = 0. \quad (3.5)$$

Поскольку $x = 0$ не является корнем этого уравнения, представим его в виде

$$\sin x = \frac{1}{x}$$

и построим графики функций, стоящих в левой и правой частях этого уравнения (рис. 3.4).

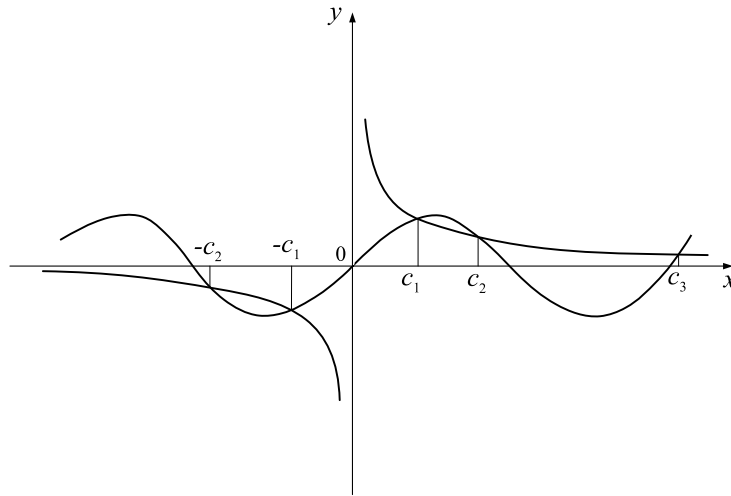


Рис. 3.4.

Корни уравнения симметричны относительно $x = 0$, поэтому мы можем рассматривать только положительные корни (соответствующие отрицательные корни отличаются от положительных только знаком). Значения x_1 , x_2 и еще несколько корней можно довольно точно определить графически, однако на графике невозможно изобразить все корни x_n . В то же время, по ходу графиков очевидно, что при больших n значения x_n будут близки к πn .

В этом случае можно считать, что для всех n , начиная, по крайней мере, с четвертого, выполняются условия

$$x_n \in [\pi n - \varepsilon; \pi n + \varepsilon], \quad (3.6)$$

где $\varepsilon > 0$ некоторое не слишком малое число (достаточное для того, чтобы на границах отрезков (3.6) левая часть уравнения (3.5) принимала значения разных знаков). Соотношения (3.6) представляют собой искомые достаточно малые области, в каждой из которых заключен один и только один корень уравнения (3.5). Таким образом, положительные корни уравнения (3.5) отделены и локализованы соотношениями (3.6).

В разобранный пример мы воспользовались нашими знаниями о свойствах функций, входящих в левую часть уравнения (3.5). Как быть в тех случаях, когда этих функций много и они сложны? В этих случаях можно предложить следующую схему действий, по сути своей весьма близкую к графическому методу отделения корней.

В практических приложениях интересующие нас величины (неизвестные, входящие в уравнения) не могут принимать сколь угодно большие или сколь угодно малые значения. Область изменения неизвестного всегда ограничена как снизу, так и сверху, причем эти ограничения никак не связаны с областью оп-

ределения функций, входящих в уравнение (этим прикладные задачи, рассматриваемые в данном курсе, отличаются от «чисто математических»). Для уравнения с одним неизвестным эта область представляет собой числовой промежуток или совокупность нескольких промежутков (если внутри области изменения неизвестного есть точки или множества точек, в которых функции, входящие в уравнение, не определены).

Разобьем эту область на участки точками x_n , $n = 1, \dots, N$ и вычислим значения $f(x_n)$ левой части уравнения (3.3) в точках разбиения. Если шаг разбиения достаточно мал*, то можно считать, что корней у уравнения ровно столько, сколько раз меняется знак $f(x_n)$. По основной теореме все эти корни будут локализованы на тех из участков разбиения, где функция $f(x)$ меняет знак. После того, как корни отделены и локализованы, приступают к их уточнению.

§ 2. Некоторые итерационные методы уточнения корней

Методы уточнения корней нелинейных уравнений основаны, как правило, на применении итерационных вычислительных алгоритмов, сходящихся к искомому корню. Мы познакомились с такими алгоритмами в первой главе (§3). Сейчас мы продолжим знакомство с ними на примере решения нелинейных уравнений общего вида (3.3).

Пусть на отрезке $[a, b]$ локализован корень уравнения (3.3). Требуется определить значение этого корня с заданной точностью $\varepsilon > 0$, т.е. найти такое число x^* , для которого выполняется условие

$$|x^* - c| < \varepsilon, \quad (3.7)$$

где c – точное решение уравнения (3.3).

Существует множество итерационных методов, реализующих эту задачу. Мы разберем лишь самые характерные из них и наиболее эффективные в том или ином смысле.

1. Метод бисекции

В литературе можно встретить и другие названия этого метода – метод *дихотомии*, или метод *деления отрезка пополам* (последнее название представляет собой перевод на русский язык первых двух). Иногда этот метод называют еще *методом вилки*. Такое название вызвано аналогией между методом бисекции и известным в артиллерии методом пристрелки: один снаряд посылают с недолетом, другой – с перелетом. При этом говорят, что цель взята в «вилку». Следующий снаряд посылают со средним значением прицела между двумя предыдущими и смотрят, как он упадет – с недолетом или перелетом. В результате

*Достаточность определяется компромиссом между двумя стремлениями: гарантировать один корень на участке за счет очень маленького шага и уменьшить количество вычислений за счет увеличения шага.

вилка сужается. Такая корректировка прицела продолжается до тех пор, пока снаряды не накроют цель.

Алгоритм метода бисекции основан на идее «артиллерийской вилки». Пусть функция $f(x)$ из уравнения (3.3) удовлетворяет условиям основной теоремы. Предположим для определенности (рис. 3.5), что на левом конце отрезка $[a, b]$ она принимает отрицательное значение, а на правом – положительное

$$f(a) < 0, \quad f(b) > 0. \quad (3.8)$$

Найдем середину отрезка $[a, b]$

$$x_1 = \frac{a + b}{2}.$$

Если $f(x_1) = 0$, то задача решена (правда, вероятность такого «точного попадания» очень близка к нулю). Если $f(x_1) \neq 0$, поступим следующим образом: рассмотрим два отрезка $[a, x_1]$ и $[x_1, b]$ и выберем один из них, исходя из условия, чтобы функция $f(x)$ принимала на его концах значения разных знаков. В ситуации, изображенной на рис. 3.5, это отрезок $[a, x_1]$. Выбранный отрезок обозначим $[a_1, b_1]$ (в нашем случае $a_1 = a; b_1 = x_1$). По построению

$$f(a_1) < 0, \quad f(b_1) > 0. \quad (3.9)$$

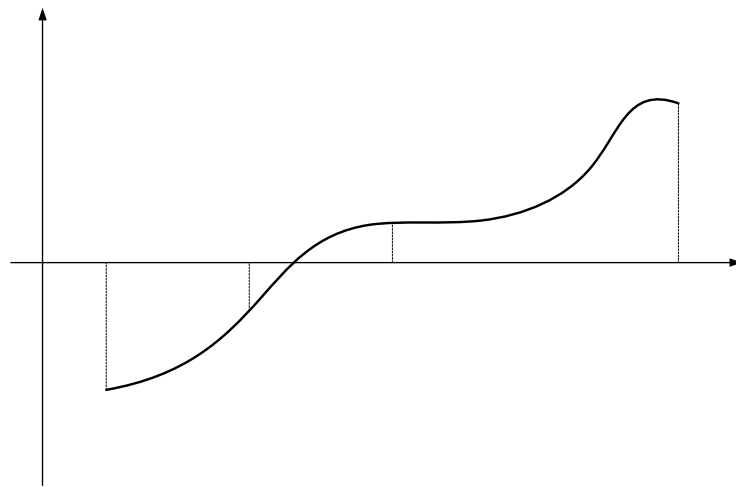


Рис. 3.5.

Затем возьмем среднюю точку отрезка $[a_1, b_1]$:

$$x_2 = \frac{a_1 + b_1}{2}.$$

Если $f(x_2) = 0$, то задача решена. Если $f(x_2) \neq 0$, снова рассмотрим два отрезка $[a_1, x_2]$ и $[x_2, b_1]$ и снова выберем тот, на концах которого функция $f(x)$ принимает значения разных знаков. На рис. 3.5 это отрезок $[x_2, x_1]$, являющийся, с учетом предыдущих обозначений, отрезком $[x_2, b_1]$. Выбранный отрезок обозначим $[a_2, b_2]$. По построению

$$f(a_2) < 0, \quad f(b_2) > 0. \quad (3.10)$$

Если продолжать этот процесс, то он либо оборвется на некотором шаге n из-за того, что выполнится условие

$$f(x_n) = 0, \quad (3.11)$$

либо будет продолжаться неограниченно. В первом случае уравнение (3.3) решено, и его корень – число x_n . Во втором случае неограниченное продолжение процесса дает последовательность вложенных друг в друга отрезков $[a, b]$, $[a_1, b_1]$, ..., $[a_n, b_n]$, ... Каждый последующий отрезок принадлежит всем предыдущим, длины отрезков стремятся к нулю как члены бесконечно убывающей геометрической прогрессии со знаменателем 0,5. При этом для всех отрезков выполняются условия основной теоремы

$$f(a_n) < 0, \quad f(b_n) > 0. \quad (3.12)$$

Отсюда следует, что корень уравнения (3.3) c принадлежит каждому из последовательности вложенных отрезков:

$$a_n < c < b_n, \quad n = 0, 1, 2, \dots \quad (3.13)$$

Построение такой последовательности можно продолжать бесконечно (если, конечно, на одном из этапов не выполнится условие (3.11)), но на практике его останавливают тогда, когда выполнится условие

$$b_n - a_n < 2\varepsilon, \quad (3.14)$$

где ε – та точность, с которой требовалось решить уравнение (3.3). При этом корнем уравнения считают число

$$x^* = \frac{a_n + b_n}{2}. \quad (3.15)$$

В силу выполнения условия (3.14) для приближения x^* , вычисленного по формуле (3.15), выполняется и условие (3.7). Итак, задача уточнения корня уравнения (3.3) решена.

Замечание! Если на левом конце отрезка $[a, b]$ функция принимает положительное значение, а на правом отрицательное, то это приводит лишь к изменению знаков в неравенствах (3.8), (3.9), (3.10), (3.12) на противоположные. Сущность алгоритма при этом не изменяется.

В качестве примера решим методом бисекции с точностью $\pm 0,00001$ уравнение (3.2), переписав его предварительно в виде

$$\cos x - x = 0. \quad (3.16)$$

В качестве начального отрезка возьмем отрезок $[0; 1]$ (см. рис. 3.1). Решение уравнения (3.16) с точностью $\pm 0,00001$ требует 16-кратного деления этого отрезка пополам.

Вычисления удобно выполнять, записывая промежуточные результаты в виде таблицы. Для метода бисекции удобнее всего использовать таблицу вида 3.1. Знаки, стоящие в скобках в заголовках второго и третьего столбцов, указывают на знаки $f(a)$ и $f(b)$. Поскольку метод бисекции не использует значения функ-

ции $f(x)$, то в последнем столбце можно указывать только знак функции. Если на одном из этапов деления отрезка выполнится условие (3.11), то в этой графе ставят 0, а соответствующее значение x_n является корнем уравнения.

Таблица 3.1.

n	a_n (+)	b_n (-)	$x_{n+1} = \frac{a_n + b_n}{2}$	Знак $f(x_{n+1})$
0	0,000 000 000	1,000 000 000	0,500 000 000	+
1	0,500 000 000	1,000 000 000	0,750 000 000	—
2	0,500 000 000	0,750 000 000	0,625 000 000	+
3	0,625 000 000	0,750 000 000	0,687 500 000	+
4	0,687 500 000	0,750 000 000	0,718 750 000	+
5	0,718 750 000	0,750 000 000	0,734 375 000	+
6	0,734 375 000	0,750 000 000	0,742 187 500	—
7	0,734 375 000	0,742 187 500	0,738 281 250	+
8	0,738 281 250	0,742 187 500	0,740 234 375	—
9	0,738 281 250	0,740 234 375	0,739 257 812	—
10	0,738 281 250	0,739 257 812	0,738 769 531	+
11	0,738 769 531	0,739 257 812	0,739 013 672	+
12	0,739 013 672	0,739 257 812	0,739 135 742	—
13	0,739 013 672	0,739 135 742	0,739 074 707	+
14	0,739 074 707	0,739 135 742	0,739 105 224	—
15	0,739 074 707	0,739 105 224	0,739 089 965	—
16	0,739 074 707	0,739 089 965	0,739 082 336	

Из табл. 3.1 видно, что $b_{16} - a_{16} < 0,00002$. При этом

$$0,739074707 < c < 0,739089965,$$

и в качестве приближенного значения корня уравнения (3.16) можно взять число $x^* \approx 0,739082336$ или, округляя с точностью $\pm 0,00001$, $x^* \approx 0,73908$.

В заключение отметим достоинства и недостатки метода бисекции.

Достоинства:

1. Простота алгоритма.
2. Гарантированная сходимость при выполнении условий основной теоремы (скорость и сам факт сходимости не зависят от свойств функции $f(x)$ и ее производных).
3. На каждом шаге для корня дается двусторонняя оценка, позволяющая определить достигнутой точности.

Недостатки:

1. Метод сходится довольно медленно.

2. Метод касательных (метод Ньютона)

Это один из наиболее эффективных численных методов решения уравнений. Идея метода очень проста. Предположим, что функция $f(x)$, имеющая корень c на отрезке $[a, b]$, дифференцируема на этом отрезке и ее производная $f'(x)$ не обращается на нем в нуль. Возьмем произвольную точку $x_0 \in [a, b]$ и запишем уравнение касательной к графику функции $f(x)$ в этой точке:

$$y = f(x_0) + f'(x_0)(x - x_0). \quad (3.17)$$

Найдем точку пересечения (3.17) касательной с осью x (рис. 3.6).

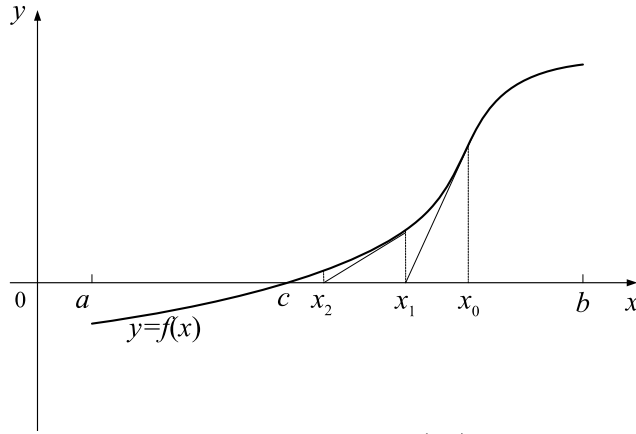


Рис. 3.6. Построение последовательности $\{x_n\}$ по методу касательных

Для определения точки x_1 имеем уравнение

$$f(x_0) + f'(x_0)(x_1 - x_0) = 0,$$

откуда

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}. \quad (3.18)$$

Повторим сделанную процедуру. Напишем уравнение касательной к графику функции $f(x)$ в точке $x = x_1$ и найдем точку ее пересечения с осью x (см. рис. 3.6)

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

Продолжая этот процесс, получим последовательность $\{x_n\}$, определенную с помощью рекуррентной формулы

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \quad n = 0, 1, 2, \dots \quad (3.19)$$

В качестве первого примера применения метода касательных рассмотрим задачу извлечения квадратного корня из произвольного положительного числа a . Будем искать его как положительный корень уравнения

$$f(x) = x^2 - a = 0. \quad (3.20)$$

Рекуррентная формула (3.19) в данном случае принимает вид

$$x_{n+1} = x_n - \frac{x_n^2 - a}{2x_n} = \frac{1}{2} \left(x_n + \frac{a}{x_n} \right). \quad (3.21)$$

Она совпадает с формулой (1.18) из §3 первой главы. Следовательно, алгоритм вычисления \sqrt{a} по формуле (1.18) основан на решении уравнения (3.20) методом касательных. Известно, что процесс сходится к \sqrt{a} при любом выборе начального приближения $x_0 > 0$, причем сходимость весьма быстрая.

Следует заметить, что метод касательных является односторонним методом, т.е. последовательность $\{x_n\}$ приближается к корню c с одной стороны. Вследствие этого, критерий остановки процедуры вычислений по формуле

(3.19) сформулировать сложнее, чем в случае метода бисекции. Чаще всего поступают так. Проводят вычисления по рекуррентной формуле до тех пор, пока последующий член не начнет отличаться от предыдущего меньше, чем на заданную величину. Именно так мы прервали вычисление квадратного корня по рекуррентной формуле (1.18).

Однако остается не ясным, на какую величину должны отличаться два приближения, чтобы выполнялось условие (3.7). Более того, сам факт сходимости последовательности $\{x_n\}$ к корню уравнения c , равно как и условия, при которых эта сходимость имеет место, требуют обоснования.

Оставляя пока открытым вопрос о сходимости метода касательных, решим этим методом, в качестве второго примера, уравнение (3.16) с той же точностью $\pm 0,00001$, с какой решали его методом бисекции. Левая часть уравнения (3.16) и ее производная имеют вид

$$f(x) = \cos x - x; \quad f'(x) = -\sin x - 1.$$

В качестве начального приближения выберем $x_0 = 1$. Результаты вычислений приведены в табл. 3.2.

Таблица 3.2.

n	x_n	$f(x_n)$	$f'(x_n)$
		-0,459697694	-1,841470984
		-0,018923072	-1,681904952
		-0,000046456	-1,673632544
0	1	-0,0000000003	-1,673612029
1	0,750363867		
2	0,739112891		
3	0,739085133		
4	0,739085132		

Как видим, для уравнения (3.16) итерационный процесс сходится очень быстро. Уже между результатами четвертой и третьей итераций разница меньше, чем 10^{-8} .

Быстрая сходимость – это главное достоинство метода касательных. Однако уже беглый анализ ситуаций, приведенных на рис. 3.7 – 3.9, показывает, что этот метод сходится не всегда. В случае, изображенном на рис. 3.7, первая же итерация выходит за рамки отрезка $[a, b]$, т.е. в область значений, при которых функция $f(x)$, вообще говоря, может и не быть определена. В таких случаях говорят: «итерационная процедура расходится».

На рис. 3.8 «выброс» происходит не сразу, а на втором шаге (это может произойти и после нескольких шагов). Если в ситуации, изображенной на рис. 3.8, точка x_2 совпадает с точкой x_0 или оказывается вблизи нее, то итерационная процедура «зацикливается». Это значит, что числа последовательности $\{x_n\}$ будут по очереди оказываться близкими то к x_0 , то к x_1 , и так может продолжаться бесконечно.

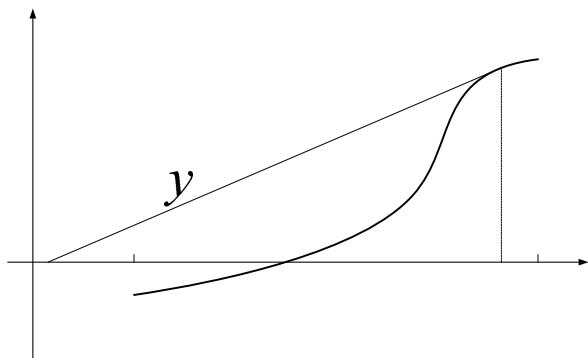


Рис. 3.7

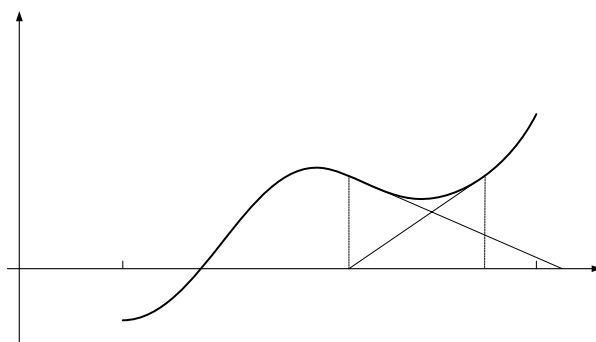


Рис. 3.8.

Может показаться, что подобные проблемы возникают только из-за неудачно выбранной начальной точки x_0 , однако это не совсем так. Действительно, в случаях, изображенных на рис. 3.7 и 3.8, выбор начальной точки в ближайшей окрестности точки c исправил бы ситуацию. Но, во-первых, в реальных расчетах мы не знаем точного значения корня (иначе, зачем было бы искать приближенное). Во-вторых, понятия «ближе» и «дальше» весьма относительны, и не всегда можно с уверенностью сказать, близко от корня мы выбрали начальную точку или нет. И, наконец, третьих, существуют такие «неудачные» функции, для которых, как бы близко к корню мы не выбирали начальное приближение x_0 , итерационная процедура метода касательных все равно расходится или закичивается.

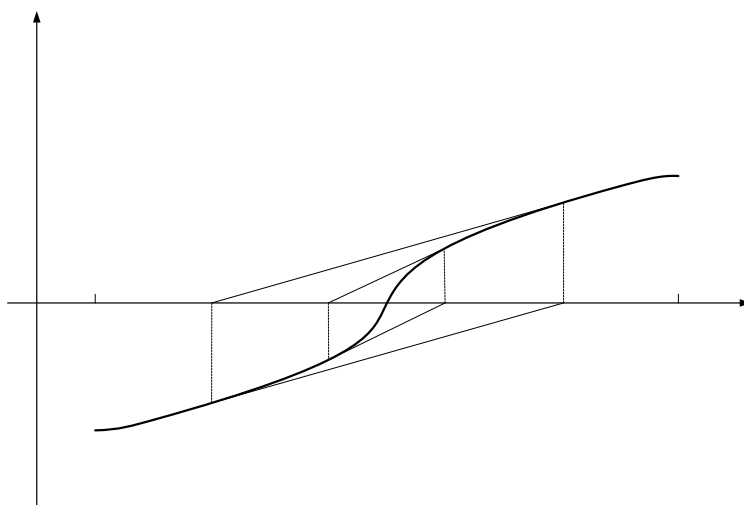


Рис. 3.9.

На рис. 3.9 показан пример функции, для которой при любом x_0 происходит закичивание (чтобы не загромождать рисунок, на нем показаны только две серии закичивающихся последовательностей, но ситуация будет аналогичной, сколь бы близко к корню мы не выбирали начальное приближение).

По аналогии с рис.3.9 можно начертить график, для которого система точек $\{x_n\}$ будет удаляться от корня с ростом n , как бы близко к корню не было выбрано начальное приближение x_0 .

Очевидно, что во всех вышеперечисленных случаях не были выполнены некоторые условия, при которых итерационный процесс метода касательных сходится к корню уравнения $f(x) = 0$. Эти условия могут быть сформулированы следующим образом.

Теорема. Если уравнение $f(x) = 0$ имеет на отрезке $[a, b]$ единственный корень $x = c$, при этом функция $f(x)$ удовлетворяет условиям основной теоремы и, кроме того, имеет непрерывные производные $f'(x)$ и $f''(x)$, не обращающиеся в нуль на отрезке $[a, b]$, то существует такая окрестность точки c : $[c - \delta, c + \delta] \subset [a, b]$, $\delta > 0$, что если начальное приближение x_0 взято из этой окрестности, то последовательность (3.19) сходится к $x = c$. Начальное приближение x_0 следует выбирать так, чтобы было

$$f(x)f''(x) > 0. \quad (3.22)$$

Действительно, в ситуации, изображенной на рис. 3.8, производная $f'(x)$ дважды обращается в нуль на отрезке $[a, b]$ – в точке максимума и в точке минимума. В ситуации, изображенной на рис. 3.9, обращается в нуль вторая производная $f''(x)$, и происходит это в точке пересечения графика функции $f(x)$ с осью x . В случае, изображенном на рис. 3.7, на отрезке $[a, b]$ выполнены все условия сходимости метода касательных, тем не менее, начальное приближение оказалось неудачным. Все дело в том, что приведенная теорема, задающая условия сходимости метода касательных, лишь гарантирует существование такой окрестности корня, что если начальное приближение взято из этой окрестности, то итерационная процедура сходится к корню. Но ни эта теорема, ни какие бы то ни было другие, не предлагают способа определения этой окрестности.

Отметим еще одно важное обстоятельство. Для того чтобы проверить выполнение условий сходимости метода, надо решить два уравнения:

$$f'(x) = 0; \quad f''(x) = 0; \quad x \in [a, b].$$

Иными словами, попытка проверить выполнение условий сходимости приводит к необходимости решения двух задач, эквивалентных исходной.

Кроме того, функция $f(x)$ может не быть задана в виде формулы. Ее значения могут находиться в результате численного решения некоторой математической задачи, получаться из измерений и т.д., что исключает или затрудняет нахождение производных.

Из всего сказанного совершенно ясно, что при решении конкретных практических задач трудно, а зачастую, невозможно проверить выполнение тех ограничений, которые накладывает на функцию $f(x)$ и ее производные приведенная выше теорема. В таких случаях сходимость метода проверяют «экспериментально»: начинают расчет и следят за поведением первых членов последовательности $\{x_n\}$. Если по ним видно, что процесс сходится, то расчет продолжают, пока не достигнут нужной точности. В противном случае вычисления прекращают и либо начинают их с другого начального значения x_0 , либо выбирают другой метод решения уравнения.

Разумеется, подобную «экспериментальную» тактику можно применять только при «ручном» счете, когда мы контролируем каждый промежуточный результат и имеем время на размышления. Именно при «ручном» счете скорость сходимости является решающим фактором.

В то же время, непредсказуемость поведения алгоритма метода касательных делает нецелесообразным его использование в качестве элемента сложных компьютерных программ, работающих в автоматическом режиме, поскольку в этом случае решающим фактором является стабильность алгоритма, его независимость от свойств конкретной функции $f(x)$.

Следует заметить, что иногда в качестве условий сходимости метода касательных (помимо условий основной теоремы) называют условия

$$|f'(x)| \geq m > 0, \quad |f''(x)| \leq M, \quad x \notin [a, b], \quad (3.23)$$

т.е. необращение в нуль первой и ограниченность второй производных функции $f(x)$. Но в случае, изображенном на рис. 3.9, условия (3.23) соблюдены, в то же время метод не сходится при практически любом выборе начальной точки. (Точнее говоря, окрестность сходимости $[c - \delta, c + \delta]$ в этом случае оказывается исчезающе малой.). Принципиальным оказывается обращение в нуль на отрезке $[a, b]$ второй производной $f''(x)$.

В заключение отметим еще раз достоинства и недостатки метода касательных.

Достоинство:

1. Если метод сходится, то сходится очень быстро.

Недостатки:

1. Помимо вычисления функции $f(x)$, метод требует вычисления ее производной.

2. Метод односторонний, и у него нет надежного критерия окончания итерационной процедуры.

3. Метод сходится не всегда (при этом, проверка критерия сходимости, в лучшем случае, представляет собой задачу более сложную, чем решение исходного уравнения).

3. Метод секущих

Если в рекуррентной формуле метода касательных (3.19) заменить $f'(x_n)$ разделенной разностью

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

вычисленной по известным значениям x_n и x_{n-1} , то мы получаем новый итерационный метод

$$x_{n+1} = x_n - \frac{(x_n - x_{n-1})f(x_n)}{f(x_n) - f(x_{n-1})}, \quad n = 1, 2, \dots, \quad (3.24)$$

который называется методом секущих. В отличие от ранее рассмотренных методов этот метод является *двухшаговым*, т.е. новое приближение определяется

двумя предыдущими итерациями. Для того чтобы начать вычисления по формуле (3.24), необходимо задать два начальных приближения x_0 и x_1 . Но в дальнейшем каждая итерация требует лишь однократного вычисления функции $f(x)$ и не требует вычисления производной.

Геометрическая интерпретация метода секущих состоит в следующем. Через точки $(x_{n-1}, f(x_{n-1}))$, $(x_n, f(x_n))$ проводится прямая (рис. 3.10). Абсцисса точки пересечения этой прямой с осью Ox является новым приближением x_{n+1} . Иными словами, функция $f(x)$ интерполируется на отрезке $[x_{n-1}, x_n]$ многочленом первой степени и за очередное приближение x_{n+1} принимается корень этого многочлена.

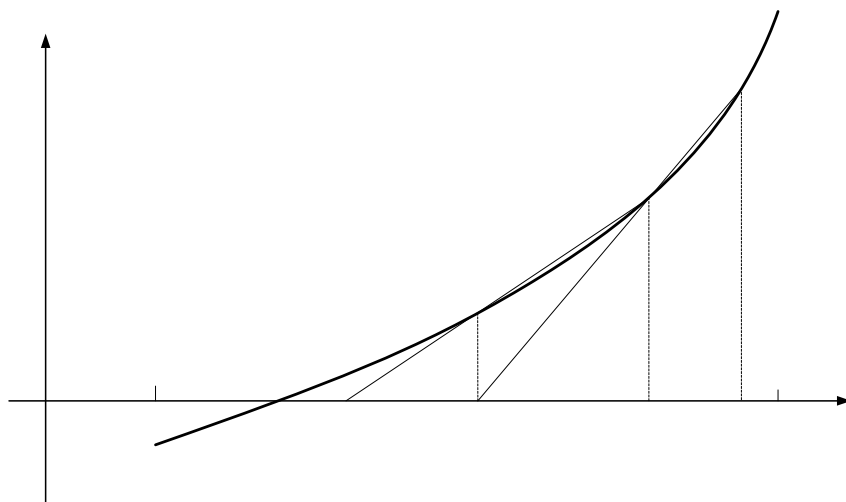


Рис. 3.10. Построение последовательности $\{x_n\}$ по методу секущих

Достоинства метода секущих:

1. Не требуется вычисления производной. Вычисляется только сама функция $f(x)$ и только один раз за шаг.
2. Метод довольно быстро сходится (лишь немного уступая в скорости сходимости методу касательных).

Недостатки метода секущих:

1. Этот метод, точно так же, как и метод касательных, сходится не всегда. (Ситуации, в которых метод секущих расходится, равно как и условия его сходимости, аналогичны таковым для метода касательных).
2. Метод односторонний и у него нет надежного критерия окончания итерационной процедуры.
3. Несколько более сложная (чем у предыдущих двух методов) вычислительная формула.

4. Сравнительная характеристика рассмотренных итерационных методов

Мы познакомились с тремя итерационными методами решения уравнений. Существует еще целый ряд методов, на которых мы не останавливались. Ситуация, когда одну и ту же математическую задачу можно решать с помо-

щью разных методов, довольно типична. Естественно возникает необходимость сравнения их между собой.

При оценке эффективности численных методов существенное значение имеют следующие их свойства:

- 1) универсальность;
- 2) простота организации вычислительного процесса и контроля его точности;
- 3) скорость сходимости.

Посмотрим с этих точек зрения на разобранные нами методы решения уравнений.

1. Наиболее универсальным является метод бисекции. Он требует только непрерывности функции $f(x)$. Два других метода накладывают более сильные ограничения. Во многих случаях это преимущество метода бисекции оказывается решающим.

2. С точки зрения организации вычислительного процесса все три метода достаточно просты. Однако и здесь метод бисекции обладает весьма серьезным преимуществом. Вычисления этим методом можно начинать с любого отрезка $[a, b]$, на концах которого непрерывная функция $f(x)$ принимает значения разных знаков. Процесс будет сходиться к корню уравнения $f(x) = 0$, причем на каждом шаге метод дает для корня двустороннюю оценку, по которой легко определить достигнутую точность. Сходимость же методов касательных и секущих зависит от того, насколько удачно выбраны нулевые приближения.

3. Наибольшей скоростью сходимости обладает метод касательных. Метод секущих, лишь немного уступая в этой скорости методу касательных, не требует вычисления производной. В случае, когда подсчет значений функции $f(x)$ и ее производной связан с большим объемом вычислений, это преимущество может стать определяющим.

Итак, мы видим, что ответ на вопрос о наилучшем численном методе решения уравнений не однозначен. Он существенно зависит от того, какую дополнительную информацию о функции $f(x)$ мы имеем, частью какой более общей проблемы выступает задача о решении уравнения и, в соответствии с этим, каким свойствам метода мы придаем наибольшее значение.

Опыт применения различных численных методов позволяет дать следующие общие рекомендации.

Методы секущих и касательных лучше всего применять в тех случаях, когда мы имеем возможность контролировать промежуточные результаты расчетов, например, при «ручном» счете. Применение этих методов в качестве элементов больших программных комплексов, работающих в автоматическом режиме с функциями, свойства которых заранее неизвестны, крайне нежелательно, поскольку может привести к «зацикливанию» алгоритма и, вследствие этого, к «зависанию» программы.

Для работы в таком режиме лучше всего подходит метод бисекции. Он сходится надежно (при выполнении условий основной теоремы), не дает сбоев,

а скорость сходимости в этом случае не так важна. К тому же, с ростом производительности вычислительной техники она играет все меньшую роль.

§ 3. Интерполяционные методы уточнения корней

Идея интерполяционных методов состоит в том, что нахождение корней нелинейного (или громоздкого) уравнения $f(x) = 0$ заменяется нахождением корней интерполяционного многочлена, построенного для $f(x)$. Интерполяционный метод первого порядка приводит к методу секущих. Существует несколько различных вариантов интерполяционных методов второго порядка. В частности, в книге А.А. Самарского и А.В. Гулина «Численные методы» (Самарский А.А., Гулин А.В. Численные методы. – М.: Наука, 1989.–432 с.) рассматривается вариант интерполяционного метода второго порядка, названный *методом парабол*. Он основан на построении некоторой последовательности парабол и удобен тем, что позволяет находить не только действительные, но и комплексные корни уравнения (3.3), пользуясь вещественными начальными приближениями.

Рассмотрим другой вариант итерационного метода нахождения простых (или нечетной кратности) вещественных корней уравнения (3.3), использующий одновременно и квадратичную интерполяцию функции $f(x)$, и процедуру сжатия отрезка знакопеременности функции $f(x)$.

Пусть дано уравнение вида

$$f(x) = 0, \quad (3.3)$$

где функция $f(x)$ – однозначно определена и непрерывна на отрезке $[a, b]$.

Пусть также

$$f(a) \cdot f(b) < 0, \quad (3.25)$$

т. е. выполнены условия основной теоремы. Следовательно, на отрезке $[a, b]$ существует хотя бы один корень уравнения (3.3).

Будем считать, что процедура отделения корней уже осуществлена, и на отрезке $[a, b]$ расположен только один корень $x = x^*$ уравнения (3.3). Требуется найти его значение с заданной точностью ε .

Пусть для определенности $f(a) < 0$, $f(b) > 0$ (рис. 3.11). Положим

$$c = 0,5(a + b) \quad (3.26)$$

и вычислим $f(c)$. Если $f(c) = 0$, то корень уравнения (3.3) найден. Если $f(c) \neq 0$, то, приняв обозначения

$$y_a = f(a), \quad y_b = f(b), \quad y_c = f(c), \quad (3.27)$$

построим по трем точкам (a, y_a) , (b, y_b) , (c, y_c) интерполяционный многочлен второго порядка

$$P_2(x) = Ax^2 + Bx + C, \quad (3.28)$$

где коэффициенты A, B, C определяются соотношениями

$$\begin{aligned}
A &= \frac{2}{d^2}(y_a - 2y_c + y_b), \\
B &= \frac{2}{d}(y_b - y_c) - A(b + c), \\
C &= y_a - Aa^2 - Ba,
\end{aligned} \tag{3.29}$$

где

$$d = b - a. \tag{3.30}$$

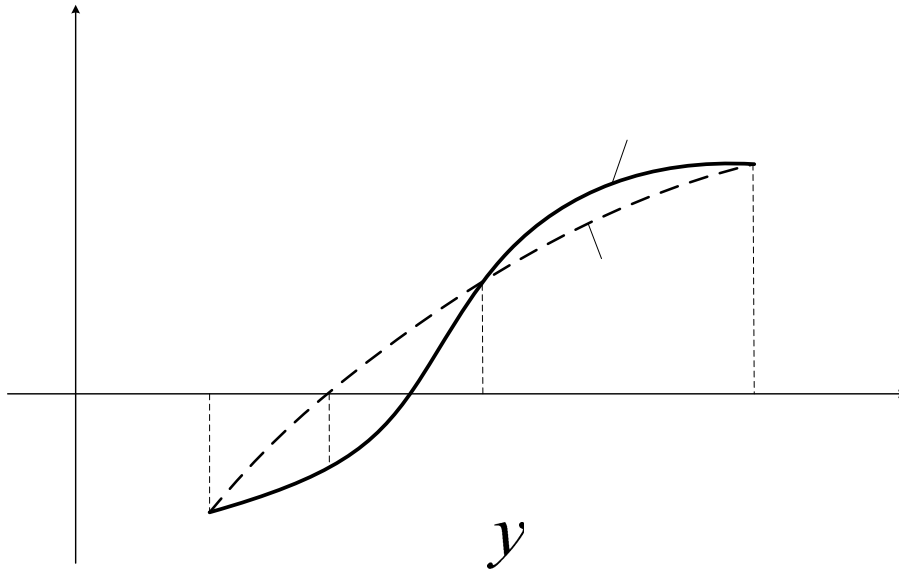


Рис. 3.11.

Функция $P_2(x)$ непрерывна на $[a, b]$, а значения $P_2(a)$ и $P_2(b)$ совпадают с y_a и y_b соответственно, т. е. $P_2(x)$ принимает на концах отрезка $[a, b]$ значения разных знаков. Следовательно, уравнение

$$Ax^2 + Bx + C = 0 \tag{3.31}$$

имеет на отрезке $[a, b]$ ровно один корень (поскольку не может иметь более двух). Если $A = 0$, то единственный корень уравнения (3.31) определяется как

$$x_1 = -\frac{C}{B} \tag{3.32}$$

Здесь учтено, что если $A = 0$, то $B \neq 0$, поскольку, если $A = B = 0$, то функция $P_2(x)$ принимает постоянное значение $P_2(x) = C$ при любых значениях x , что противоречит условию (3.25).

Если $A \neq 0$, то уравнение (3.31) заведомо имеет положительный дискриминант и, как следствие, два различных корня

$$\tilde{x}_{1,2} = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A}, \quad a \quad x_1 \tag{3.33}$$

один из которых принадлежит отрезку $[a, b]$. Действительно, если бы дискриминант уравнения (3.31) не был положительным, то функция $P_2(x)$ была бы знакопостоянной во всей своей области определения \mathcal{U}_a за исключением разве что

нулевого значения при $\tilde{x} = -\frac{B}{2A}$ в случае нулевого дискриминанта), а это противоречит условию (3.25).

Заметим, что условию (3.25) противоречат также и предположения, что ни один из корней не принадлежит отрезку $[a, b]$ и, что оба корня принадлежат ему. Таким образом, в случае $A \neq 0$ только один из корней (3.33) принадлежит отрезку $[a, b]$, а какой именно – определяется простым перебором.

Из сказанного следует, что при любых значениях параметров A, B и C , которые не противоречат условию (3.25), на отрезке $[a, b]$ имеется один (и только один) корень уравнения (3.31) (обозначим его x_1), который определяется либо соотношением (3.32), либо одним из соотношений (3.33).

Вычислим $y_1 = f(x_1)$. Если $y_1 = 0$, то x_1 – корень уравнения (1). Если $y_1 \neq 0$ (как на рис.1), то точками c и x_1 отрезок $[a, b]$ оказался разделенным на три отрезка. При сделанных выше предположениях (на отрезке $[a, b]$ расположен только один корень уравнения (3.3)) лишь один из этих отрезков характеризуется тем, что на его концах функция $f(x)$ меняет знак. Этот отрезок обозначается как $[a_1, b_1]$.

В ситуации, которая изображена на рис.3.11, $a_1 = x_1, b_1 = c$. В общем случае в качестве a_1 выбирается наибольшее из чисел a, c, x_1 , значение функции $f(x)$ в которых имеет тот же знак, что и y_a , а в качестве b_1 – наименьшее из чисел c, x_1, b , в которых значение функции $f(x)$ того же знака, что и y_b .

Таким образом, в результате совершенных действий мы либо находим корень x^* уравнения (3.3), либо формируем новый отрезок $[a_1, b_1]$ такой, что на его концах функция $f(x)$ принимает значения разных знаков, т. е. выполняются условия основной теоремы, причем

$$\begin{aligned} x^* &\in [a_1, b_1], \\ [a_1, b_1] &\subset [a, b]. \end{aligned} \quad (3.34)$$

Для нового отрезка $[a_1, b_1]$ вся процедура повторяется. В результате либо находится корень уравнения (3.3) в точках c_1 или x_2 , либо формируется отрезок $[a_2, b_2]$ такой, что $x^* \in [a_2, b_2], [a_2, b_2] \subset [a_1, b_1]$, и снова выполняются условия основной теоремы.

Процедура повторяется до тех пор, пока либо не будет явно найден корень x^* уравнения (3.3), либо на каком-то i -м этапе отрезок $[a_i, b_i]$ не окажется таким, что

$$d_i = b_i - a_i < 2\varepsilon. \quad (3.35)$$

В этом случае за решение уравнения принимается величина

$$\xi = 0,5(a_i + b_i), \quad (3.36)$$

что гарантирует выполнение условия $|\xi - x^*| < \varepsilon$.

ГЛАВА 4.

МЕТОДЫ ВЫЧИСЛЕНИЯ ОПРЕДЕЛЕННЫХ ИНТЕГРАЛОВ

Решение многих задач физики, геометрии, техники и экономики связано с вычислением определенного интеграла от некоторой функции. Для удобства дальнейшего изложения весьма полезно вспомнить, как вводится это важнейшее понятие математического анализа.

§ 1. Понятие определенного интеграла

Пусть на отрезке $[a, b]$ задана некоторая функция $f(x)$. Разобьем этот отрезок точками x_i : $a = x_0 < x_1 < \dots < x_{n-1} < x_n = b$ на отрезки $[x_{i-1}, x_i]$, $i = 1, \dots, n$. Вычислим длины полученных отрезков $\Delta x_i = x_i - x_{i-1}$ и обозначим $\Delta = \max \{\Delta x_i\}$.

Выберем на каждом из полученных отрезков какую-нибудь точку $\xi_i \in [x_{i-1}, x_i]$ и вычислим в ней значение функции $f(x)$. Составим из найденных величин сумму, которую называют *интегральной суммой*

$$J(x_i, \xi_i) = f(\xi_1)\Delta x_1 + f(\xi_2)\Delta x_2 + \dots + f(\xi_n)\Delta x_n = \sum_{i=1}^n f(\xi_i)\Delta x_i. \quad (4.1)$$

Будем неограниченно измельчать разбиение так, чтобы $\Delta \rightarrow 0$ (при этом, естественно, $n \rightarrow \infty$). Если существует конечный предел

$$J = \lim_{\Delta \rightarrow 0} \sum_{i=1}^n f(\xi_i)\Delta x_i, \quad (4.2)$$

то функцию $f(x)$ называют *интегрируемой* на отрезке $[a, b]$, а число J – *определенным интегралом* от этой функции по отрезку $[a, b]$ и обозначают символом

$$J = \int_a^b f(x) dx. \quad (4.3)$$

Свойством интегрируемости обладает достаточно широкий класс функций. К таковым, в частности, относятся функции, непрерывные на отрезке $[a, b]$; функции, имеющие на отрезке $[a, b]$ конечное число точек разрыва и ограниченные; функции, монотонные на отрезке $[a, b]$.

§ 2. Формула Ньютона – Лейбница

Один из способов вычисления определенного интеграла основан на применении формулы Ньютона – Лейбница

$$\int_a^b f(x) dx = F(b) - F(a), \quad (4.4)$$

где $F(x)$ – любая первообразная подынтегральной функции $f(x)$, определенная на отрезке $[a, b]$. Эта формула позволяет вычислять интегралы от элемен-

тарных функций, первообразные которых также являются элементарными функциями. Если первообразная найдена в явном виде, то вычисление определенного интеграла сводится к подсчету разности ее значений в точках a и b .

При всей своей внешней простоте формула Ньютона – Лейбница имеет один очень серьезный недостаток. Нахождение первообразной подынтегральной функции нередко представляет собой очень трудоемкую и даже творческую задачу*. Более того, существует много функций, первообразные которых не выражаются через элементарные функции. И, наконец, формула Ньютона-Лейбница не позволяет вычислять интегралы от функций, которые заданы графиком или таблицей, а также от функций, заданных алгоритмически (например, в виде компьютерной программы).

Все это позволяет сделать вывод, что формула Ньютона-Лейбница не дает общего, универсального метода нахождения определенного интеграла от произвольной функции $f(x)$ по ее значениям на отрезке $[a, b]$, она не является алгоритмом решения рассматриваемой задачи. (Напомним, что в определении понятия определенного интеграла первообразная не фигурирует. Там речь идет только о самой подынтегральной функции $f(x)$.)

Ниже мы рассмотрим некоторые универсальные вычислительные алгоритмы решения задачи определенного интегрирования, которые позволяют подсчитывать интегралы непосредственно по значениям подынтегральной функции $f(x)$ и не зависят от способа ее задания. Соответствующие формулы обычно называют *формулами численного интегрирования* или *квадратурными формулами* (т.е. формулами вычисления площадей).

§ 3. Квадратурные формулы, основанные на определении понятия интеграла

Проще всего к идее численного интегрирования можно подойти, используя определение интеграла как предела сумм (4.2). Если взять достаточно мелкое разбиение отрезка $[a, b]$ и построить для него интегральную сумму (4.1), то ее значение можно приближенно принять за значение соответствующего интеграла. Приближенное равенство

$$\int_a^b f(x) dx \approx \sum_{i=0}^{n-1} c_i f(\xi_i), \quad (4.5)$$

где c_i – числовые коэффициенты, а ξ_i – точки отрезка $[x_i, x_{i+1}]$, называется *квадратурной формулой*, а сумма, стоящая в (4.5) справа – *квадратурной суммой*. Точки ξ_i называются *узлами квадратурной формулы*, а числа c_i – *коэффициентами квадратурной формулы*.

* Успех в нахождении первообразной нередко определяется удачным выбором замены переменной. Сам же этот выбор, ввиду отсутствия общей методики, зависит, главным образом, от опыта и интуиции вычислителя.

Разность

$$\gamma_n = \int_a^b f(x) dx - \sum_{i=0}^{n-1} c_i f(\xi_i) \quad (4.6)$$

называется *погрешностью квадратурной формулы*. Погрешность зависит как от расположения узлов, так и от выбора коэффициентов.

1. Формулы прямоугольников

Пусть, например, отрезок $[a, b]$ разбит на n равных частей длины

$$h = \frac{b-a}{n}. \quad (4.7)$$

Если в качестве точек ξ_i выбрать левые концы соответствующих отрезков: $x_0, x_1, x_2, \dots, x_{n-1}$, то получим формулу «левых» прямоугольников (рис. 4.1)

$$J \approx (f(a) + f(x_1) + \dots + f(x_{n-1})) \frac{b-a}{n}. \quad (4.8)$$

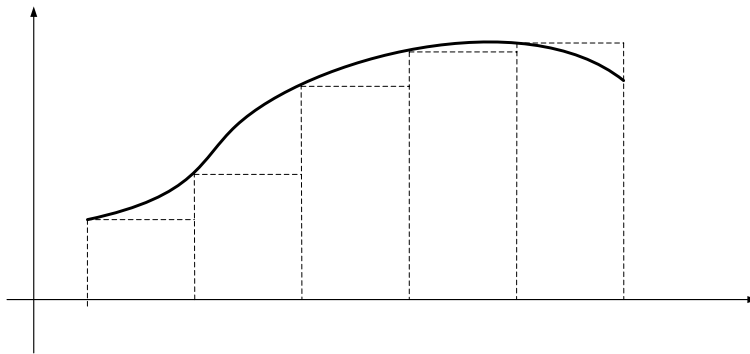


Рис. 4.1. Геометрическая интерпретация формулы «левых» прямоугольников

Если выбрать в качестве точек ξ_i правые концы отрезков: $x_1, x_2, \dots, x_{n-1}, x_n$, то получим формулу «правых» прямоугольников (рис. 4.2)

$$J \approx (f(x_1) + \dots + f(x_{n-1}) + f(b)) \frac{b-a}{n}. \quad (4.9)$$

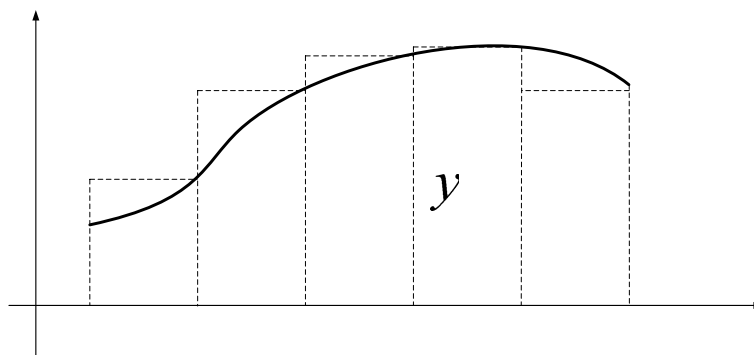


Рис. 4.2. Геометрическая интерпретация формулы «правых» прямоугольников

В случае монотонно возрастающей функции $f(x)$ первая из этих формул будет давать приближение интеграла с недостатком, а вторая – с избытком (в случае монотонно убывающей функции – наоборот). Поэтому чаще всего ис-

пользуют формулу «средних» прямоугольников, в которой в качестве точек ξ_i берут средние точки отрезков разбиения (рис. 4.3)

$$J \approx (f(0,5(x_0 + x_1)) + f(0,5(x_1 + x_2)) + \dots + f(0,5(x_{n-1} + x_n))) \frac{b-a}{n}. \quad (4.10)$$

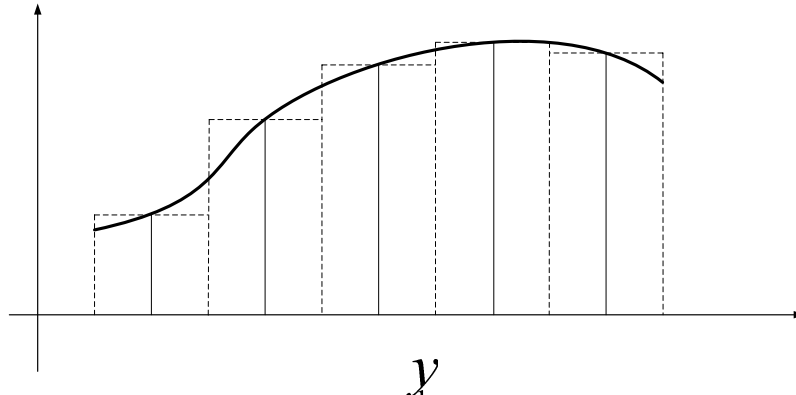


Рис. 4.3. Геометрическая интерпретация формулы «средних» прямоугольников

Слово «средних» в названии формулы (4.10) обычно опускают и называют ее просто формулой прямоугольников.

Погрешность (4.6) при вычислении интеграла по формуле (4.10) с ростом n (или, что то же самое, с уменьшением шага деления $h = \frac{b-a}{n}$) убывает, как $1/n^2$.

2. Формула трапеций

Если сложить формулы (4.8) и (4.9) и поделить сумму пополам, то получится формула трапеций

$$J \approx (\frac{1}{2} f(a) + f(x_1) + \dots + f(x_{n-1}) + \frac{1}{2} f(b)) \frac{b-a}{n}. \quad (4.11)$$

Причина такого названия, как и в предыдущем случае, связана с геометрической интерпретацией формулы. Она приближенно представляет площадь криволинейной трапеции, соответствующей интегралу J , в виде суммы площадей обычных трапеций (рис. 4.4).

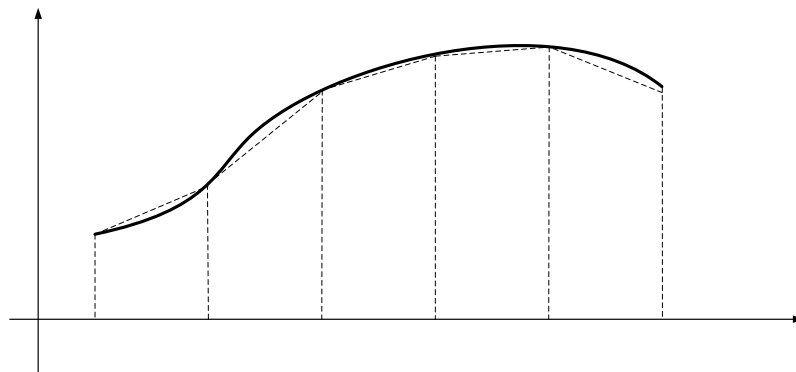


Рис. 4.4. Геометрическая интерпретация формулы трапеций

Погрешность формулы трапеций с ростом n убывает практически с той же скоростью, что и у формулы (4.10). Порядок убывания $-1/n^2$.

Анализируя рис. 4.4, можно сказать, что в формуле трапеций кривая $y = f(x)$ заменяется ломаной линией, т.е. некоторой вспомогательной функцией, для которой определенный интеграл вычисляется легко. Эту идею можно развивать дальше и использовать для получения более точных формул численного интегрирования. Например, если в качестве вспомогательной функции использовать не кусочно-линейную, а кусочно-квадратичную функцию, то мы придем к формуле Симпсона.

3. Формула Симпсона (формула парабол)

Разделим отрезок $[a, b]$ на n равных частей, где число n – **четное!** Квадратурная формула Симпсона имеет вид

$$J \approx (f(a) + 4f(x_1) + 2f(x_2) + 4f(x_3) + \dots + 2f(x_{n-2}) + 4f(x_{n-1}) + f(b)) \frac{b-a}{3n}. \quad (4.12)$$

Значения функции $f(x)$ в нечетных точках разбиения x_1, x_3, \dots, x_{n-1} входят в формулу с коэффициентом 4, в четных точках x_2, x_4, \dots, x_{n-2} – с коэффициентом 2, а в двух граничных точках $x_0 = a, x_n = b$ – с коэффициентом 1.

С возрастанием n погрешность формулы Симпсона убывает как $1/n^4$, т.е. быстрее, чем в формулах прямоугольников и трапеций.

4. Пример расчета интеграла по трем квадратурным формулам

Для иллюстрации изложенного материала вычислим интеграл

$$J = \int_1^2 \frac{1}{x} dx \quad (4.13)$$

по формулам прямоугольников (4.10), трапеций (4.11) и Симпсона (4.12), взяв во всех трех случаях $n = 10$. В табл. 4.1 приведены значения подынтегральной функции $f(x) = \frac{1}{x}$ в точках $\xi_i = 1 + h(i - 0,5)$; $h = 0,1$; $(i = 1, 2, \dots, 10)$, которые используются в формуле прямоугольников (4.10).

Таблица 4.1			Таблица 4.2		
i	ξ_i	$f(\xi_i)$	i	x_i	$f(x_i)$
			0	1,0	1,000000000
1	1,05	0,952380952	1	1,1	0,909090909
2	1,15	0,869565217	2	1,2	0,833333333
3	1,25	0,800000000	3	1,3	0,769230769
4	1,35	0,740740740	4	1,4	0,714285714
5	1,45	0,689655172	5	1,5	0,666666666
6	1,55	0,645161290	6	1,6	0,625000000
7	1,65	0,606060606	7	1,7	0,588235294
8	1,75	0,571428571	8	1,8	0,555555555
9	1,85	0,540540540	9	1,9	0,526315789
10	1,95	0,512820512	10	2,0	0,500000000

Подставив эти значения в формулу, (4.10) получаем приближенное значение интеграла по формуле прямоугольников

$$J \approx I_{10} = 0,692835360.$$

В табл. 4.2 приведены значения подынтегральной функции $f(x) = \frac{1}{x}$ в точках $x_i = 1 + hi$; $h = 0,1$; $(i = 0,1,...,10)$, которые используются в формулах трапеций и Симпсона.

Вычислим значение интеграла по формуле трапеций (4.11)

$$J \approx T_{10} = 0,693771403$$

и по формуле Симпсона (4.12)

$$J \approx S_{10} = 0,693150230.$$

В данном случае мы можем воспользоваться тем, что нам известно точное значение интеграла (4.13)

$$J = \ln 2 = 0,693147180.$$

Вычислим погрешности, которые дают все три формулы:

$$J - I_{10} = 0,000311820;$$

$$J - T_{10} = -0,000624223; \quad (4.14)$$

$$J - S_{10} = -0,000003050.$$

Погрешность формулы Симпсона, как и следовало ожидать, наименьшая.

§ 4. Квадратурные формулы интерполяционного типа. Формулы Гаусса

Не будем, в отличие от предыдущего параграфа, разбивать отрезок интегрирования $[a, b]$ на частичные отрезки, а построим квадратурные формулы вида (4.5) путем замены подынтегральной функции интерполяционным многочленом сразу на всем отрезке $[a, b]$. Полученные таким образом формулы называются *квадратурными формулами интерполяционного типа*. Как правило, точность таких формул возрастает с увеличением числа узлов интерполирования. Однако за счет удачного расположения узлов можно, при одном и том же их количестве, повысить степень многочленов, для которых квадратурная формула верна.

Иными словами, может быть сформулирована следующая задача. При заданном числе узлов n построить квадратурную формулу, точную для многочленов наиболее высокого порядка. Такие квадратурные формулы существуют. Их называют *формулами Гаусса*.

1. Формула Гаусса для отрезка $[-1, 1]$

Рассмотрим сначала определенный интеграл по отрезку $[-1, 1]$. Формула Гаусса в этом случае имеет вид

$$\int_{-1}^1 f(t) dx = \sum_{i=1}^n \alpha_i f(t_i), \quad (4.15)$$

где t_i – узлы квадратурной формулы, а α_i – весовые коэффициенты. Их значения зависят от n и вычисляются по специальным формулам (мы их рассматривать не будем) либо берутся из таблиц. В табл. 4.3 приведены узлы α_i и весовые коэффициенты t_i для нескольких наиболее употребительных значений n .

Таблица 4.3

n	α_1 t_1	α_2 t_2	α_3 t_3	α_4 t_4	α_5 t_5	α_6 t_6
2	1,0 -0,577350269	1,0 0,577350269				
3	5/9 -0,774596669	8/9 0,0	5/9 0,774596692			
4	0,347854845 -0,861136312	0,652145155 -0,339981044	0,652145155 0,339981044	0,347854845 0,861136312		
5	0,236926885 -0,906179846	0,478628671 -0,538469310	0,568888889 0,0	0,478628671 0,538469310	0,236926885 0,906179846	
6	0,171324492 -0,932469514	0,360761573 -0,661209386	0,467913935 -0,238619186	0,467913935 0,238619186	0,360761573 0,661209386	0,171324492 0,932469514

2. Формула Гаусса для произвольного отрезка

Для произвольного отрезка $[a, b]$ квадратурная формула Гаусса имеет вид

$$\int_a^b f(x) dx = \frac{b-a}{2} \sum_{i=1}^n \alpha_i f(x_i), \quad (4.16)$$

где
$$x_i = \frac{b-a}{2} t_i + \frac{b+a}{2}. \quad (4.17)$$

Значения t_i и α_i в формулах (4.16) и (4.17) берутся из табл. 4.3.

В качестве иллюстрации вычислим интеграл (4.13) по формулам Гаусса (4.16), (4.17) при $n = 4$.

В этом случае

$$\int_1^2 \frac{1}{x} dx = \frac{1}{2} \sum_{i=1}^4 \frac{\alpha_i}{x_i}, \quad (4.18)$$

где узловые точки x_i , в соответствии с (4.17), принимают значения:

$$x_1 = 1,069\,431\,844; \quad x_2 = 1,330\,009\,478; \quad x_3 = 1,669\,990\,522; \quad x_4 = 1,930\,568\,156.$$

Подставляя эти значения и весовые коэффициенты, взятые из табл. 4.3 при значении $n = 4$, в формулу (4.18), получаем приближенное значение интеграла

$$J \approx G_4 = 0,693\,146\,417. \quad (4.19)$$

Погрешность этого значения

$$J - G_4 = 0,000\,000\,763. \quad (4.20)$$

Следует обратить особое внимание на то, что значение интеграла (4.19) было получено на основе всего четырех вычислений подынтегральной функции. При этом погрешность полученного значения существенно меньше, чем погрешности (4.14) вычисления этого же интеграла по формулам прямоугольников, трапеций и Симпсона с десятикратным вычислением подынтегральной функции.

3. Повышение точности квадратурной формулы Гаусса за счет разбиения отрезка интегрирования

Итак, формулы Гаусса представляются наиболее экономичными квадратурными формулами с точки зрения количества вычислений подынтегральной функции. Однако при их использовании следует учитывать одно важное обстоятельство. Во многих случаях возникает задача вычисления интегралов, где подынтегральная функция или ее производные имеют участки резкого изменения, например, обращаются в бесконечность. Такие функции плохо приближаются многочленами сразу на всем отрезке интегрирования (ведь метод Гаусса – интерполяционный метод), и это может привести к серьезной потере точности квадратурной формулы даже при больших значениях n .

В таких случаях имеет смысл разбить исходный отрезок на части и в каждой из них применять формулы Гаусса с небольшим n , либо другие квадратурные формулы. Количество частей и конкретный вид разбиения определяются свойствами подынтегральной функции и опытом вычислителя.

§ 5. Построение первообразной функции с помощью численного интегрирования

Формула Ньютона – Лейбница (4.4) позволяет вычислить значение определенного интеграла от функции $f(x)$ через ее первообразную $F(x)$ (и мы отметили это в §2 как один из ее недостатков). В математическом анализе устанавливается и прямо противоположная возможность: первообразная функции $f(x)$, непрерывной на отрезке $[a, b]$, может быть записана в виде определенного интеграла с переменным верхним пределом

$$F(x) = \int_{x_0}^x f(t)dt. \quad (4.21)$$

Здесь нижний предел интегрирования x_0 предполагается фиксированным, а верхний x – переменным. В случае непрерывной на отрезке $[a, b]$ функции

$f(x)$ функция $F(x)$, определенная с помощью (4.21), дифференцируема и ее производная равна $f(x)$:

$$F'(x) = \frac{d}{dx} \left(\int_{x_0}^x f(t) dt \right) = f(x). \quad (4.22)$$

Формула (4.22) в сочетании с какой-нибудь формулой численного интегрирования представляет собой универсальный алгоритм построения первообразной. Рассмотрим пример, иллюстрирующий этот алгоритм.

Функция $f(x) = \frac{\sin x}{x}$ непрерывна и, следовательно, имеет первообразные. Эти первообразные не могут быть выражены через элементарные функции, однако для них справедливо представление в виде интеграла с переменным верхним пределом. Одну из первообразных мы получим, выбирая в качестве нижнего предела $x_0 = 0$. Ее называют интегральным синусом и обозначают

$$\text{Si } x = \int_0^x \frac{\sin t}{t} dt. \quad (4.23)$$

Интегральный синус определен на всей числовой прямой, является нечетной функцией x и имеет конечные предельные значения на бесконечности

$$\lim_{x \rightarrow \pm \infty} \text{Si } x = \pm \frac{\pi}{2}.$$

В соответствии с (4.22)

$$(\text{Si } x)' = \frac{\sin x}{x}.$$

Методы численного интегрирования позволяют вычислить значение $\text{Si } x$ при любом значении x . График интегрального синуса при $x \geq 0$ приведен на рис. 4.5.

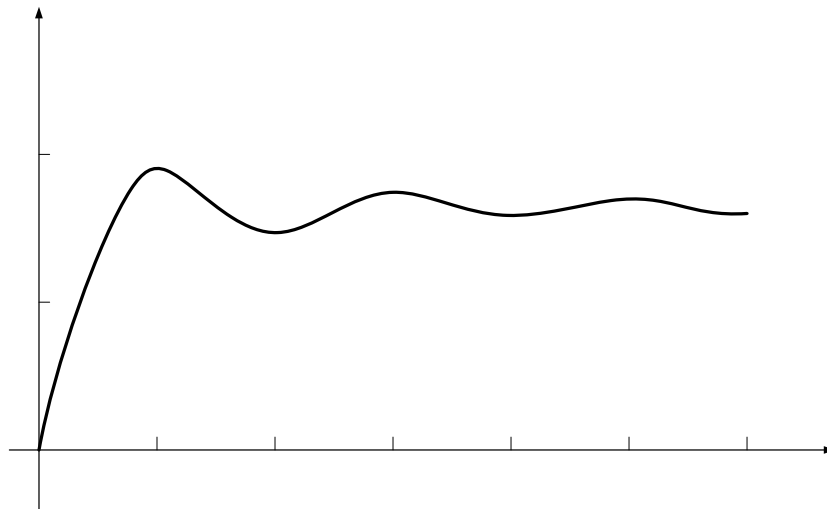


Рис. 4.5. График интегрального синуса

Существуют и другие специальные функции, которые вводятся как интегралы с переменным верхним пределом. Не останавливаясь на их описании, отметим лишь, что деление функций на элементарные и не элементарные весьма условно. Для того чтобы работать с той или иной функцией, достаточно иметь алгоритм ее вычисления при любом значении аргумента. С этой точки зрения применение интегрального синуса или другой специальной функции ничем не отличается от применения более привычных элементарных функций.

ГЛАВА 5

ЧИСЛЕННОЕ ДИФФЕРЕНЦИРОВАНИЕ И МЕТОДЫ ЧИСЛЕННОГО ИНТЕГРИРОВАНИЯ ОБЫКНОВЕННЫХ ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЙ (ОДУ)

§ 1. Численное дифференцирование функций

Задача численного дифференцирования состоит в приближенном вычислении производной функции $f'(x)$ по заданным в конечном числе точек значениям функции $f(x)$.

По определению производной функции $f(x)$ в точке x^* называется предел

$$f'(x^*) = \lim_{\Delta x \rightarrow 0} \frac{f(x^* + \Delta x) - f(x^*)}{\Delta x}, \quad (5.1)$$

если он существует и конечен (Δx может принимать как положительные, так и отрицательные значения).

Приближенное значение производной может быть получено, если в формуле (5.1) не переходить к пределу

$$f'(x^*) \approx \frac{f(x^* + \Delta x) - f(x^*)}{\Delta x}. \quad (5.2)$$

Пусть функция $y = f(x)$ задана таблицей

Таблица 5.1

x	x_1	\dots	x_{i-1}	x_i	x_{i+1}	\dots	x_n
y	y_1	\dots	y_{i-1}	y_i	y_{i+1}	\dots	y_n

В качестве приближенного значения функции $y'_i = f'(x_i)$ можно взять любое из следующих разностных отношений:

$$y'_i \approx \frac{y_{i+1} - y_i}{x_{i+1} - x_i}, \quad y'_i \approx \frac{y_i - y_{i-1}}{x_i - x_{i-1}}, \quad y'_i \approx \frac{y_{i+1} - y_{i-1}}{x_{i+1} - x_{i-1}}. \quad (5.3)$$

Первую из этих формул называют «правой», вторую – «левой», а третью – «центральной». Формулы (5.3.) немного упрощаются, если функция $y = f(x)$ задана таблицей с равноотстоящими узлами, т.е. если при любом i $x_{i+1} - x_i = h$, где h – шаг таблицы:

$$y'_i \approx \frac{y_{i+1} - y_i}{h}, \quad y'_i \approx \frac{y_i - y_{i-1}}{h}, \quad y'_i \approx \frac{y_{i+1} - y_{i-1}}{2h}. \quad (5.4)$$

Вторую производную в точке x_i в этом случае можно заменить отношением

$$y''_i = f''(x_i) \approx \frac{y_{i+1} - 2y_i + y_{i-1}}{h^2}. \quad (5.5)$$

Замечание! В крайних точках таблицы вторая производная *не определена*, а первую производную можно вычислить только по односторонним формулам («левой» или «правой»).

§ 2. Численное интегрирование ОДУ первого порядка (ОДУ-1)

Замена производной тем или иным разностным отношением лежит в основе большинства численных методов решения обыкновенных дифференциальных уравнений с начальными условиями (задачи Коши). Это – классическая область применения численных методов, часть из которых была разработана еще в докомпьютерную эпоху. Мы рассмотрим только основные методы, которые широко применяются на практике и чаще всего используются в современных вычислительных средах и стандартных пакетах программ.

1. Обыкновенные дифференциальные уравнения первого порядка (ОДУ-1)

Напомним некоторые определения, известные из курса математического анализа.

Дифференциальным уравнением порядка n называется уравнение

$$F(x, y, y', y'', \dots, y^{(n)}) = 0,$$

связывающее независимую переменную x , искомую функцию y и ее производные $y, y', y'', \dots, y^{(n)}$. Если функция $y = \varphi(x)$ есть функция *одного* независимого переменного, то дифференциальное уравнение называется *обыкновенным*. (Наряду с обыкновенными дифференциальными уравнениями существуют еще *дифференциальные уравнения в частных производных*, в которых неизвестная функция зависит от нескольких независимых переменных.)

Решением дифференциального уравнения называется всякая функция $y = \varphi(x)$, которая, будучи подставлена в уравнение, превращает его в тождество (т.е. равенство, выполняемое при любых значениях x).

Обыкновенное дифференциальное уравнение *первого* порядка (ОДУ-1) имеет вид

$$F(x, y, y') = 0.$$

Если это уравнение удастся разрешить относительно производной, то его записывают в виде

$$y' = f(x, y). \quad (5.6)$$

Общим решением ОДУ-1 называется функция

$$y = \varphi(x, C), \quad (5.7)$$

которая удовлетворяет уравнению (5.7) при любых значениях *постоянной интегрирования* C . Функция, получающаяся из (5.7) при каждом конкретном численном значении C , называется *частным решением* ОДУ-1.

Таким образом, общее решение ОДУ-1 (5.7) представляет собой совокупность бесконечного множества частных решений уравнения (5.6), соответствующих различным значениям постоянной C .

Иногда при отыскании общего решения ОДУ-1 получают соотношение вида

$$\Phi(x, y, C) = 0, \quad (5.8)$$

которое не удастся разрешить ни относительно x , ни относительно y . В этом случае равенство (5.8) называют *общим интегралом* ОДУ-1.

Процесс отыскания общего решения или общего интеграла ОДУ называют *интегрированием* ОДУ.

2. Задача Коши для ОДУ-1

Условие, что при $x = x_0$ функция y должна равняться заданному числу y_0 , называется *начальным условием*. Его обычно записывают одним из следующих способов:

$$y|_{x=x_0} = y_0; \quad y(x_0) = y_0; \quad y_0 = \varphi(x_0). \quad (5.9)$$

Задача отыскания частного решения ОДУ-1, соответствующего заданному начальному условию, называется *задачей Коши* для дифференциального уравнения первого порядка и записывается, например, в виде:

$$y' = f(x, y), \quad y(x_0) = y_0. \quad (5.10)$$

Аналитический метод решения задачи Коши для ОДУ-1 состоит из двух этапов. Первый этап заключается в отыскании общего решения уравнения (5.7) (или общего интеграла (5.8)). Если это удалось сделать, то на втором этапе в полученное общее решение (или общий интеграл) подставляют начальные условия. В результате приходят к обыкновенным (не дифференциальным) уравнениям, из которых определяют то значение постоянной интегрирования, при котором соответствующее частное решение удовлетворяет начальным условиям.

Следует заметить, что даже для ОДУ первого порядка, разрешенных относительно производной, найти общее решение или общий интеграл (иначе говоря, *проинтегрировать в аналитическом виде*) удастся далеко не всегда. В тех случаях, когда ОДУ-1 не удастся проинтегрировать аналитически, прибегают к численным методам решения задачи Коши.

Наибольшее распространение приобрели *разностные методы* решения задачи Коши, позволяющие построить приближенное решение ОДУ-1 в виде функции, заданной табличным способом. Рассмотрим некоторые из них.

3. Метод Эйлера

Пусть требуется решить задачу Коши (5.10). Будем строить решение этой задачи в виде табличной функции с шагом h . Заменим уравнение $y' = f(x, y)$ разностным уравнением

$$\frac{y_{i+1} - y_i}{h} = f(x_i, y_i), \quad i = 0, 1, \dots, \quad y_0 = y(x_0). \quad (5.11)$$

Решение этого уравнения находится явным образом по рекуррентной формуле

$$y_{i+1} = y_i + hf(x_i, y_i), \quad i = 0, 1, \dots, \quad y_0 = y(x_0). \quad (5.12)$$

Пример

Построить методом Эйлера численное решение задачи Коши

$$y' = \sqrt{xy}, \quad y(1) = 2$$

на отрезке $x \in [1, 2]$ с шагом $h = 0,1$.

Решение

Запишем для заданного ОДУ рекуррентную формулу метода Эйлера

$$y_{i+1} = y_i + 0,1 \cdot \sqrt{x_i \cdot y_i}. \quad (5.13)$$

Индекс i принимает целые значения от 0 до 10. По условию $x_0 = 1, y_0 = 2$. Вычисления по формуле (5.13) приведены в табл. 5.2. Поскольку шаг метода довольно велик ($h = 0,1$) во всех промежуточных результатах оставлены два десятичных знака после запятой.

Таблица 5.2.

i	x_i	y_i	$\sqrt{x_i y_i}$
0	1	2	1,41
1	1,1	2,14	1,53
2	1,2	2,29	1,66
3	1,3	2,46	1,79
4	1,4	2,64	1,92
5	1,5	2,83	2,06
6	1,6	3,04	2,20
7	1,7	3,26	2,35
8	1,8	3,49	2,51
9	1,9	3,74	2,67
10	2,0	4,01	2,83

Явная формула Эйлера – самая простая, однако точность решения, построенного по ней, невысока. Точность каждого шага можно повысить, уменьшая h , но при этом увеличивается количество шагов, и ошибки каждого шага накапливаются.

Рассмотрим различные способы повышения точности разностных методов решения задачи Коши. Начнем с неявной формулы Эйлера.

Помимо разностного уравнения в форме (5.11), допускающего явное решение (5.12), можно записать неявное разностное уравнение

$$\frac{y_{i+1} - y_i}{h} = f(x_{i+1}, y_{i+1}), \quad i = 0, 1, \dots, \quad y_0 = y(x_0) \quad (6.14)$$

или

$$\frac{y_i - y_{i-1}}{h} = f(x_i, y_i), \quad i = 1, 2, \dots, \quad y_0 = y(x_0). \quad (6.15)$$

Эти уравнения задают новое значение неизвестной функции в неявном виде. Для определения каждого нового значения неизвестной функции приходится решать (часто нелинейные!) уравнения (6.15) или (6.14), что представляет собой дополнительную проблему (значительно увеличивается объем вычислений,

т.к. нелинейные уравнения приходится решать одним из итерационных методов).

Преимуществом неявного метода является его более высокая точность. Однако неудобства, связанные с необходимостью решать на каждом шаге построения искомой функции нелинейные уравнения, приводят к тому, что этот метод применяют редко (по крайней мере, для интегрирования обыкновенных дифференциальных уравнений). Гораздо больший интерес представляют явные методы повышенной точности.

4. Методы Рунге-Кутты

Порядок точности можно повысить путем усложнения разностной схемы. Весьма распространены в практических вычислениях методы Рунге-Кутты второго и четвертого порядков точности. Вычисления по методам Рунге-Кутты второго порядка точности проводятся в два этапа. Их часто называют методами типа *предиктор – корректор*, потому что они работают по принципу *предсказание – поправка*. На первом этапе находится промежуточное значение y_i^* по формуле Эйлера с шагом h или $\frac{h}{2}$. На втором этапе строится значение y_{i+1} по той или иной уточняющей формуле. Приведем два варианта формул типа предиктор – корректор:

$$\begin{aligned} & y_i^* = y_i + h \cdot f(x_i, y_i), \\ \text{I)} \quad & y_{i+1} = y_i + \frac{h}{2} \cdot (f(x_i, y_i) + f(x_{i+1}, y_i^*)). \end{aligned} \quad (5.16)$$

$$\begin{aligned} & y_i^* = y_i + \frac{h}{2} \cdot f(x_i, y_i), \\ \text{II)} \quad & y_{i+1} = y_i + h \cdot f(x_i + \frac{h}{2}, y_i^*). \end{aligned} \quad (5.17)$$

Наиболее широкое применение нашли методы Рунге-Кутты четвертого порядка, которых тоже несколько разновидностей. Чаше других применяются следующие два:

$$\begin{aligned} & k_1 = f(x_i, y_i), \\ & k_2 = f(x_i + \frac{h}{2}, y_i + \frac{hk_1}{2}), \\ \text{I)} \quad & k_3 = f(x_i + \frac{h}{2}, y_i + \frac{hk_2}{2}), \\ & k_4 = f(x_i + h, y_i + hk_3), \\ & y_{i+1} = y_i + \frac{h}{6}(k_1 + 2k_2 + 2k_3 + k_4). \end{aligned} \quad (5.18)$$

$$\begin{aligned}
k_1 &= f(x_i, y_i), \\
k_2 &= f\left(x_i + \frac{h}{4}, y_i + \frac{hk_1}{4}\right), \\
\text{II)} \quad k_3 &= f\left(x_i + \frac{h}{2}, y_i + \frac{hk_2}{2}\right), \\
k_4 &= f(x_i + h, y_i + h(k_1 - 2k_2 + 2k_3)), \\
y_{i+1} &= y_i + \frac{h}{6}(k_1 + 4k_3 + k_4).
\end{aligned} \tag{5.19}$$

В современных вычислительных средах, как правило, используются варианты методов Рунге – Кутты с автоматическим выбором шага по независимой переменной (величина шага может изменяться в процессе интегрирования в зависимости от оценки полученной точности).

Все методы Рунге – Кутты являются *явными* (для определения y_{i+1} надо провести вычисления по явным формулам) и *одношаговыми* (для определения y_{i+1} надо сделать один шаг по независимой переменной от x_i до x_{i+1}).

Существуют и *многошаговые* методы, которые для вычисления нового значения функции y_{i+1} используют *несколько* ее предыдущих значений.

5. Многошаговые методы (методы Адамса)

Общая формула явного m -шагового метода Адамса может быть записана в виде

$$\frac{y_{i+1} - y_i}{h} = b_1 f_i + b_2 f_{i-1} + \dots + b_m f_{i-m+1}, \tag{5.20}$$

где

$$f_{i-l} = f(x_{i-l}, y_{i-l}), \quad l = 0, 1, \dots, m-1. \tag{5.21}$$

При $m = 2, 3, 4$ получаем соответственно следующие методы:

$$\begin{aligned}
y_{i+1} &= y_i + \frac{h}{2} \cdot (3f_i - 2f_{i-1}), \quad m = 2, \\
y_{i+1} &= y_i + \frac{h}{12} \cdot (23f_i - 16f_{i-1} + 5f_{i-2}), \quad m = 3, \\
y_{i+1} &= y_i + \frac{h}{24} \cdot (55f_i - 59f_{i-1} + 37f_{i-2} - 9f_{i-3}), \quad m = 4,
\end{aligned} \tag{5.22}$$

обозначения f_{i-l} имеют смысл (5.21).

Главный недостаток методов Адамса заключается в том, что для того чтобы начать вычисления, недостаточно одних начальных условий. Первые $m-1$ шагов должны быть сделаны каким-либо другим методом (чаще всего – одним из методов Рунге – Кутты). Второй недостаток – невозможность менять в ходе расчетов величину шага по независимой переменной. Коэффициенты в формулах (5.22) получены в предположении постоянности шага h .

Главное достоинство методов Адамса заключается в их большей устойчивости по сравнению с методами Рунге – Кутты.

$$\begin{aligned} y_{i+1} &= y_i + \frac{h}{2} \cdot (f_{i+1} + f_i), \quad m=2, \\ y_{i+1} &= y_i + \frac{h}{12} \cdot (5f_{i+1} + 8f_i - f_{i-1}), \quad m=3, \\ y_{i+1} &= y_i + \frac{h}{24} \cdot (9f_{i+1} + 19f_i - 5f_{i-1} + f_{i-2}), \quad m=4. \end{aligned} \tag{5.23}$$

$$\begin{aligned} y_{i+1} &= y_i + \frac{h}{2} \cdot (f_{i+1} + f_i), \quad m=2, \\ y_{i+1} &= y_i + \frac{h}{12} \cdot (5f_{i+1} + 8f_i - f_{i-1}), \quad m=3, \\ y_{i+1} &= y_i + \frac{h}{24} \cdot (9f_{i+1} + 19f_i - 5f_{i-1} + f_{i-2}), \quad m=4. \end{aligned} \tag{5.23}$$

§ 3. Численное интегрирование систем ОДУ

[illegible]

$$\begin{aligned} y_1(x_0) &= y_1^0, \\ &\dots\dots\dots \\ y_n(x_0) &= y_n^0. \end{aligned} \tag{5.25}$$

Для сокращения объема текста и уменьшения громоздкости выкладок задачу (5.24), (5.25) иногда записывают в виде

$$y'_k = f_k(x, y_1, \dots, y_n), \quad y_k(x_0) = y_k^0, \quad k = 1, 2, \dots, n. \quad (5.26)$$

Пример

Проинтегрировать на отрезке $x \in [0, 1]$ явным методом Эйлера с шагом $h = 0,1$ систему дифференциальных уравнений

$$\begin{cases} y' = (x+y)\sqrt{y^2+u^2}, \\ u' = y \cdot u, \end{cases} \quad (5.27) \text{ c}$$

84

$$y(0) = 1, \quad u(0) = 0,5. \quad (5.28)$$

Решение

В системе (5.27) для удобства (чтобы избежать в дальнейшем двойной индексации) неизвестные функции независимой переменной x обозначены как $y(x)$ и $u(x)$. Система явных разностных уравнений Эйлера для (5.27) имеет вид

$$\begin{cases} \frac{y_{i+1} - y_i}{h} = (x_i + y_i)\sqrt{y_i^2 + u_i^2}, \\ \frac{u_{i+1} - u_i}{h} = y_i \cdot u_i, \end{cases}$$

откуда следуют явные рекуррентные формулы для определения значений неизвестных функций y_{i+1} и u_{i+1} :

$$\begin{cases} y_{i+1} = y_i + h(x_i + y_i)\sqrt{y_i^2 + u_i^2}, \\ u_{i+1} = u_i + h \cdot y_i \cdot u_i. \end{cases} \quad (5.29)$$

Начальные значения всех величин, стоящих в правых частях (5.29), определяются начальными условиями (5.28):

$$x_0 = 0, \quad y_0 = 1, \quad u_0 = 0,5. \quad (5.30)$$

Результаты вычислений по формулам (5.29) с шагом $h = 0,1$ и начальными значениями переменных (5.30) приведены в табл. 5.3. Поскольку шаг h довольно велик, во всех промежуточных результатах оставлены два десятичных знака после запятой.

После того, как таблица заполнена, в ней содержатся обе искомые функции, заданные в табличном виде. Вторая и третья колонки табл. 5.3 задают искомую функцию $y(x)$, а вторая и четвертая колонки – функцию $u(x)$.

Таблица 5.3.

i	x_i	y_i	u_i	$f_{1i} = (x_i + y_i)\sqrt{y_i^2 + u_i^2}$	$f_{2i} = y_i \cdot u_i$
0	0	1	0,5	1,12	0,5
1	0,1	1,11	0,55	1,50	0,61
2	0,2	1,26	0,61	2,04	0,77
3	0,3	1,46	0,69	2,84	1,01
4	0,4	1,74	0,79	4,09	1,37
5	0,5	2,15	0,96	6,24	2,06
6	0,6	2,77	1,17	10,13	3,23
7	0,7	3,78	1,49	18,20	5,63
8	0,8	5,60	2,05	26,71	11,48
9	0,9	8,27	3,20	81,32	26,45
10	1,0	16,40	5,84		

Разумеется, обе функции вычислены в данном случае с невысокой точностью, но мы, тем не менее, можем построить их приближенный график на заданном отрезке, исследовать их и т.д.

Отличия в применении методов Рунге – Кутты или Адамса при интегрировании систем ОДУ носят, в сущности, такой же характер, как и при применении метода Эйлера в разобранный примере.

Следует заметить, что все современные вычислительные среды и программные комплексы обязательно оснащены (как правило, несколькими) программами численного решения задачи Коши для систем ОДУ-1. Это одна из самых часто решаемых задач прикладной математики.

§ 4. Численное интегрирование ОДУ второго порядка (ОДУ-2)

Обыкновенное дифференциальное уравнение второго порядка (ОДУ-2) имеет вид

$$F(x, y, y', y'') = 0.$$

Если это уравнение удастся разрешить относительно производной, то его записывают в виде

$$y'' = f(x, y, y'). \quad (5.31)$$

Общим решением ОДУ-2 называется функция

$$y = \varphi(x, C_1, C_2), \quad (5.32)$$

которая удовлетворяет уравнению (5.31) при любых значениях *постоянных интегрирования* C_1, C_2 . Функция, получающаяся из (5.32) при каждом конкретном сочетании численных значений C_1, C_2 , называется *частным решением* ОДУ-2.

Таким образом, общее решение ОДУ-2 (5.32) представляет собой совокупность бесконечного множества частных решений уравнения (5.31), соответствующих различным значениям постоянных C_1, C_2 .

Если при отыскании общего решения ОДУ-2 получают соотношение вида

$$\Phi(x, y, C_1, C_2) = 0,$$

которое не удастся разрешить ни относительно x , ни относительно y , то его называют *общим интегралом* данного уравнения.

1. Задача Коши для ОДУ-2

Задача Коши для обыкновенного дифференциального уравнения второго порядка заключается в отыскании частного решения этого дифференциального уравнения, соответствующего заданным начальным условиям. Начальных условий в этом случае должно быть два – начальное значение искомой функции и начальное значение ее первой производной. Таким образом, задача Коши для ОДУ-2 может быть записана в следующем виде

$$\begin{aligned} y'' &= f(x, y, y'), \\ y(x_0) &= y_0, \quad y'(x_0) = y'_0. \end{aligned} \quad (5.33)$$

Классический метод решения задачи (5.33) состоит из двух этапов. На первом этапе находят общее решение ОДУ-2 в виде (5.32). Затем в общее решение и его производную подставляют начальные условия, получая, таким об-

разом, систему двух уравнений (в общем случае – нелинейных) с двумя неизвестными C_1, C_2 :

$$\begin{cases} y_0 = \varphi(x_0, C_1, C_2), \\ y'_0 = \varphi'(x_0, C_1, C_2). \end{cases} \quad (5.34)$$

Как на первом, так и на втором этапе решения этой задачи могут возникнуть значительные (а часто и непреодолимые) математические трудности. Поэтому в случае дифференциальных уравнений второго порядка (особенно, нелинейных) еще большее значение приобретают численные методы решения. Наиболее простым из них следует признать метод сведения ОДУ-2 к системе двух ОДУ-1.

2. Сведение ОДУ-2 к системе двух ОДУ-1

Пусть дана задача Коши (5.33):

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

Введем новую функцию $p(x) = y'(x)$. Тогда $p'(x) = y''(x)$ и задачу (5.33) можно свести к задаче

$$\begin{cases} y' = p, \\ p' = f(x, y, p), \\ y(x_0) = y_0, \quad p(x_0) = p_0. \end{cases} \quad (5.35)$$

Это задача Коши для системы ОДУ-1, т.е. задача, численное решение которой мы уже умеем строить (см. предыдущий параграф), и для которой, как уже было сказано, разработано обширное и разнообразное программное обеспечение.

3. Краевая задача для ОДУ-2

В задаче Коши для ОДУ-2 помимо самого дифференциального уравнения второго порядка фигурируют два начальных условия – значения искомой функции и ее производной в одной и той же точке. Помимо этой задачи для ОДУ-2 может быть сформулирована другая, когда требуется *отыскать частное решение дифференциального уравнения второго порядка, принимающее заданные значения при двух различных значениях аргумента*. Такая задача называется *краевой задачей для ОДУ-2*, а задаваемые значения функции при двух различных значениях аргумента – *граничными условиями* задачи.

Итак, пусть требуется найти частное решение дифференциального уравнения

$$y'' = f(x, y, y'), \quad (5.36)$$

удовлетворяющее граничным условиям

$$y(x_0) = y_0, \quad y(x_k) = y_k. \quad (5.37)$$

Обычно считается, что частное решение строится на отрезке $x \in [x_0, x_k]$.

Классический метод решения этой задачи, так же, как и в случае задачи Коши, состоит из двух этапов. На первом этапе находят общее решение ОДУ-2

в виде (5.32). Затем в общее решение подставляют граничные условия (5.37). Это дает систему двух уравнений (в общем случае – нелинейных) с двумя неизвестными C_1, C_2 :

$$\begin{cases} y_0 = \varphi(x_0, C_1, C_2), \\ y_k = \varphi(x_k, C_1, C_2). \end{cases} \quad (5.38)$$

Пример

Найти частное решение дифференциального уравнения

$$y'' = 3x, \quad (5.39)$$

удовлетворяющее граничным условиям

$$y(0) = 2, \quad y(10) = 12. \quad (5.40)$$

Решение

Разделяя переменные и интегрируя два раза уравнение (5.39), получаем

$$y' = \frac{3x^2}{2} + C_1,$$

$$y = 0,5x^3 + C_1x + C_2.$$

Последнее выражение является общим решением дифференциального уравнения (5.39). Подставляя в него граничные условия (5.40), получаем систему двух линейных уравнений, относительно постоянных интегрирования C_1, C_2 :

$$\begin{cases} 2 = 0 + 0 + C_2, \\ 12 = 500 + 10C_1 + C_2. \end{cases}$$

Решая эту систему, получаем: $C_1 = -49, C_2 = 2$. Решением краевой задачи (5.39) – (5.40) является функция

$$y = 0,5x^3 - 49x + 2.$$

В рассмотренном примере уравнение интегрировалось легко, а постоянные интегрирования входили в общее решение линейно. Вследствие этого при подстановке граничных условий была получена линейная, относительно постоянных интегрирования, система уравнений. Для более сложных уравнений, так же, как в случае задачи Коши, на любом из двух этапов решения задачи могут возникнуть значительные (а часто и непреодолимые) математические трудности. И здесь тоже приобретают большое значение численные методы интегрирования ОДУ-2 (точнее, решения краевой задачи для ОДУ-2).

Следует заметить, что для произвольного нелинейного ОДУ-2 нет теорем, гарантирующих существование и единственность решения краевой задачи. Нередко возникает ситуация, когда мы вынуждены применять численный метод построения решения задачи, не будучи уверены в существовании этого решения.

На практике в таких случаях поступают следующим образом. Строят численное решение, исходя из предположения, что оно существует. Если решение удалось построить, то это доказывает его существование. Если же решение

построить не удалось, то вопрос о его существовании остается открытым (может быть, его удастся построить другим способом, а может быть, оно не существует).

Одним из наиболее часто применяемых методов численного решения краевых задач для ОДУ-2 является метод «пристрелки» или, иначе говоря, метод пробных задач Коши. Он заключается в следующем. Пусть дана краевая задача (5.36), (5.37):

$$\begin{aligned}y'' &= f(x, y, y'), \\ y(x_0) &= y_0, \quad y(x_k) = y_k.\end{aligned}$$

Вместо нее формируют задачу Коши, с начальным условием для искомой функции в виде $y(x_0) = y_0$. Начальное условие для ее производной записывают в виде $y'(x_0) = a$, где a – переменная величина, подбираемая таким образом, чтобы выполнялось второе граничное условие ($y(x_k) = y_k$).

Этот подбор можно осуществлять по-разному. Например, решить задачу Коши для нескольких разных значений a . Если среди полученных численных решений найдутся два таких (назовем их y_1 и y_2), что у одного из них значение при $x = x_k$ окажется больше, чем y_k , а у другого – меньше, то тогда можно решить задачу Коши при $a_3 = \frac{a_1 + a_2}{2}$, где a_1 – то значение a , при котором было получено решение y_1 , а a_2 – то значение a , при котором было получено решение y_2 .

Определив значение нового решения при $x = x_k$, мы либо получаем готовое решение (если это значение совпало с y_k), либо сужаем вдвое вилку между теми двумя значениями a , при которых функция проходит «выше» и «ниже» точки (x_k, y_k) . Дальнейшие действия абсолютно аналогичны знакомому нам методу бисекции (вилки) для решения уравнений.

К сожалению, в случае нелинейного ОДУ-2 описанная выше процедура может не привести к успеху. Дело в том, что зависимость значения функции, являющейся решением задачи Коши, при $x = x_k$ от значения числа a в случае нелинейного ОДУ-2 может не оказаться непрерывной. В этом случае сходимость метода бисекции не гарантируется, задача не может быть решена методом пристрелки, и ее следует попытаться решить другими методами.

ГЛАВА 6

ЗАДАЧИ ОПТИМИЗАЦИИ

Предыдущие главы были посвящены вопросам построения математических моделей (ММ) всевозможных объектов (процессов, явлений, конструкций и т. д.). При этом мы не акцентировали внимание на том, для чего, собственно, нужно математическое моделирование реальных объектов и процессов. Почему на их разработку тратится столько времени и усилий? По-крупному, таких причин три:

1. ММ позволяют заменить натурный эксперимент (который может быть дорог, небезопасен, требовать длительной подготовки или перестройки чего-либо) безопасным и быстроосуществимым численным экспериментом.

2. ММ позволяют исследовать не только реальные, но и несуществующие, в частности, проектируемые объекты.

3. Наконец, ММ значительно упрощают задачу определения таких параметров объекта, при которых он функционирует наилучшим, в том или ином смысле, образом.

Данная глава посвящена постановке и методам решения последнего из трех указанных типов задач. Их называют задачами оптимизации. Явно или неявно мы встречаемся с оптимизацией в любой сфере человеческой деятельности, от бытового до самого высокого общегосударственного уровня. Экономическое планирование, управление, распределение ресурсов, анализ производственных процессов, проектирование сложных объектов всегда должно быть направлено на поиск наилучшего варианта *с точки зрения намеченной цели*.

Последние слова выделены не случайно. Природа вещей устроена так, что почти никогда не удастся улучшить одновременно всё. Совершенствуя интересующий нас объект или явление с одной точки зрения, мы неизбежно делаем его менее совершенным с каких-то других точек зрения. Рассмотрим следующую ситуацию: требуется проехать на автомобиле определенное расстояние между двумя пунктами по вполне определенному маршруту, например, с загородной дачи до городской квартиры.

Допустим, мы ставим перед собой цель доехать за минимальное время. В ситуации, когда маршрут определен однозначно, единственный ресурс, которым мы можем управлять по своему усмотрению, это скорость движения. Для достижения поставленной нами цели скорость должна быть максимальной, какая только возможна.

В то же время, скорость движения всегда ограничена, причем ограничения эти могут иметь различную природу. Помимо ограничения чисто технического характера, связанного с конструктивными особенностями и техническим состоянием конкретного автомобиля, существуют ограничения, вызванные состоянием дороги, и ограничения, предписанные дорожными знаками и правилами движения. Ограничения последнего типа, конечно, могут быть нарушены,

но возможные в таком случае разбирательства с дорожным инспектором приведут к значительным потерям времени, что противоречит поставленной цели.

Скорость, удовлетворяющую всем ограничениям, действующим на данном участке дороги, назовем *допустимой*. Тогда для того, чтобы совершить поездку за минимально возможное время мы должны на каждом участке дороги двигаться с максимально допустимой для этого участка скоростью. Такая стратегия позволяет нам построить график движения *оптимальный с точки зрения затраченного времени*.

Представим теперь, что перед нами стоит другая цель: совершить точно такую же поездку, затратив как можно меньше горючего. Прежняя стратегия – двигаться все время с максимально допустимой скоростью – в этом случае не годится. При движении с большими скоростями увеличивается сопротивление движению, и для поддержания высокой скорости приходится сжигать гораздо больше горючего, чем при движении с умеренными скоростями. Для каждого типа автомобилей существует своя оптимальная скорость, при которой на единицу пути расходуется наименьшее количество горючего. Для реализации поставленной цели (экономии горючего) надо двигаться с этой скоростью. Такая стратегия позволяет нам построить график движения *оптимальный с точки зрения затрат горючего*.

Как видим, ставя перед собой разные цели, мы получили различные оптимальные графики движения. Причем нельзя одновременно уменьшить и время движения, и расход горючего. Эти цели противоречат друг другу. Если мы двигаемся оптимально с точки зрения времени, то проигрываем в горючем, если оптимизируем движение с точки зрения горючего, то проигрываем во времени. Поэтому, когда говорят, что *получено оптимальное решение* какой-либо проблемы, то *обязательно следует указывать, с какой точки зрения это решение оптимально, по какому именно критерию*. Оптимизация в абсолютном смысле неосуществима.

§1. Параметры оптимизации и целевая функция

При огромном разнообразии задач оптимизации только математика может дать общие методы их решения. Однако для того, чтобы воспользоваться математическим аппаратом, необходимо сначала сформулировать интересующую нас проблему как математическую задачу, придав количественные оценки возможным вариантам и количественный смысл словам «лучше» и «хуже».

После того, как сформулирована цель оптимизации, необходимо *выразить численно* степень близости к этой цели каждого конкретного варианта решения. Среди всех факторов, влияющих на результат, следует выделить те, которые можно активно изменять по своему усмотрению (в примере с автомобилем таким фактором была скорость). Разумеется, ресурс изменения каждого фактора ограничен. Численные значения этих факторов (количественные выражения в той или иной системе измерения) называют параметрами оптимизации (ПО). Каждому набору ПО соответствует некоторый проект решения задачи.

На параметры всегда бывают наложены некоторые ограничения: экономические, технические, физические, геометрические и т.п. Совокупность значений ПО, удовлетворяющих всем ограничениям, называется *множеством допустимых значений параметров оптимизации* (МДЗПО), а соответствующий проект – *допустимым проектом*.

Если каждому набору параметров оптимизации из МДЗПО (допустимому проекту) ставится в соответствие некоторое число, выражающее степень близости проекта к поставленной цели, то на МДЗПО задана некоторая функция параметров оптимизации, которую называют *целевой функцией*. Тогда математическая сторона задачи оптимизации сводится к отысканию наименьшего (или наибольшего) значения целевой функции на множестве допустимых значений параметров оптимизации. Постановка задачи и методы исследования существенно зависят от свойств целевой функции и той информации о ней, которая может считаться доступной в процессе решения, а также которая известна априори (т.е. заранее, до начала решения задачи).

С математической точки зрения проще всего задачи, в которых целевая функция задается явной формулой и является при этом дифференцируемой функцией. В этом случае для исследования ее свойств, определения направлений возрастания и убывания, поиска точек локального экстремума может быть использована производная. Но часто бывает так, что целевая функция не задается формулой, ее значения могут получаться в результате сложных расчетов, определяться экспериментально и т.д. Подобные задачи решать гораздо сложнее, потому что для них нельзя провести исследование целевой функции с помощью производной. Для этого типа задач разработаны специальные методы решения, рассчитанные на широкое применение вычислительной техники.

Следует также иметь в виду, что сложность задачи существенно зависит от ее размерности, т.е. от числа аргументов целевой функции. Мы начнем рассмотрение методов решения задач оптимизации с одномерных задач.

§2. Одномерные задачи оптимизации

Выделение и подробный разбор одномерных задач имеет определенный смысл. Эти задачи наиболее просты, решая их легче понять постановку вопроса, методы решения и возникающие трудности. В ряде случаев одномерные задачи имеют самостоятельный практический интерес. Однако самое главное их значение заключается в том, что алгоритмы решения многомерных задач оптимизации часто сводятся к последовательному многократному решению одномерных задач и не могут быть поняты без умения решать такие задачи.

Начнем с простейшего случая, когда целевые функции задаются явными дифференцируемыми формулами.

1. Аналитические методы оптимизации

С математической точки зрения одномерную задачу оптимизации можно сформулировать следующим образом. Найти наименьшее (или наибольшее)

значение целевой функции $f(x)$, заданной на множестве X , т.е. определить значение $x \in X$, при котором она принимает свое экстремальное значение.

В том случае, когда множество X представляет собой отрезок, а целевая функция непрерывна на этом отрезке, существование решения сформулированной задачи гарантируется теоремой Вейерштрасса.

Т е о р е м а В е й е р ш т р а с с а. *Всякая функция $f(x)$, непрерывная на отрезке $[a, b]$, принимает на этом отрезке свое наименьшее и наибольшее значения, т. е. на отрезке $[a, b]$ существуют такие точки x_1, x_2 , что для любого $x \in [a, b]$ выполняются неравенства*

$$f(x_1) \leq f(x) \leq f(x_2).$$

Не исключается, в частности, возможность того, что наименьшее или наибольшее значение достигается сразу в нескольких точках. В этом легко убедиться, рассмотрев в качестве примера функцию $y = \sin x$ на любом достаточно большом отрезке.

Функция может достигнуть своего наименьшего (наибольшего) значения либо в одной из граничных точек отрезка $[a, b]$, либо в одной из точек экстремума. Мы не будем останавливаться на методике нахождения наибольшего и наименьшего значения функции одной переменной, поскольку эта задача подробно рассматривается в школьном курсе математики и в классическом курсе высшей математики для высших технических учебных заведений.

Несколько сложнее обстоит дело тогда, когда область определения функции $f(x)$ не является отрезком. Для этих случаев теорема Вейерштрасса не справедлива, и функция $f(x)$ может не достигать своих наибольшего и наименьшего значений. В частности, она может достигать наименьшее значение и не достигать наибольшего (как в случае параболы, оси которой направлены вверх) или наоборот. В таких случаях решением задачи являются экстремальные точки целевой функции. Рассмотрим следующий пример.

2. Задача о наилучшей консервной банке

Перед вами поставили задачу: указать наилучший вариант консервной банки фиксированного объема V , имеющей обычную форму прямого кругового цилиндра.

Получив такое задание, в первую очередь нужно спросить: «По какому признаку следует сравнивать банки между собой, какая банка считается наилучшей?» Иными словами, требуется указать цель оптимизации.

Рассмотрим два варианта этой задачи.

1. Наилучшая банка должна при заданном объеме V иметь наименьшую поверхность S . (В этом случае на ее изготовление пойдет наименьшее количество жести.)

2. Наилучшая банка должна при заданном объеме V иметь наименьшую длину швов l . (Швы нужно сваривать, и естественно попытаться сделать эту работу минимальной.)

Для решения задачи запишем формулы для объема банки, площади ее поверхности и длины швов:

$$V = \pi r^2 h, \quad S = 2\pi r^2 + 2\pi r h, \quad l = 4\pi r + h. \quad (6.1)$$

Объем банки задан, это устанавливает связь между радиусом r и высотой h . Выразим высоту через радиус

$$h = \frac{V}{\pi r^2} \quad (6.2)$$

и подставим полученное выражение в формулы для поверхности и длины швов. В результате получим

$$S(r) = 2\pi r^2 + \frac{2V}{r}, \quad 0 < r < \infty, \quad (6.3)$$

$$l(r) = 4\pi r + \frac{V}{\pi r^2}, \quad 0 < r < \infty. \quad (6.4)$$

Таким образом, с математической точки зрения, задача о наилучшей консервной банке сводится к определению такого значения r , при котором достигает своего наименьшего значения в первом случае функция $S(r)$, во втором – функция $l(r)$.

Рассмотрим первый вариант задачи. Вычислим производную функции $S(r)$:

$$S'(r) = 4\pi r - \frac{2V}{r^2} = \frac{2}{r^2}(2\pi r^3 - V). \quad (6.5)$$

Она обращается в нуль при

$$r_1 = \sqrt[3]{\frac{V}{2\pi}}. \quad (6.6)$$

При $0 < r < r_1$ производная $S'(r)$ отрицательна, а при $r_1 < r < \infty$ – положительна. Следовательно, в точке $r = r_1$ функция $S(r)$ достигает своего наименьшего значения. По формуле (6.2) вычислим соответствующую высоту

$$h_1 = \sqrt[3]{\frac{4V}{\pi}} = 2r_1. \quad (6.7)$$

Итак, наилучшей банкой с точки зрения минимальности поверхности S оказывается банка, у которой высота равна диаметру.

Рассмотрим теперь задачу во второй постановке. Продифференцируем функцию $l(r)$:

$$l'(r) = 4\pi - \frac{2V}{\pi r^3} = \frac{2}{\pi r^3}(2\pi^2 r^3 - V). \quad (6.8)$$

Она обращается в нуль при

$$r_2 = \sqrt[3]{\frac{V}{2\pi^2}}. \quad (6.9)$$

Как и в предыдущем случае, при $0 < r < r_2$ производная $l'(r)$ отрицательна, а при $r_2 < r < \infty$ – положительна. Следовательно, в точке $r = r_2$ функция $l(r)$ дос-

тигает своего наименьшего значения. По формуле (6.2) вычислим соответствующую высоту

$$h_2 = \sqrt[3]{4\pi V} = 2\pi r_2. \quad (6.10)$$

Наилучшей банкой с точки зрения минимальности длины сварного шва оказывается банка, у которой высота в π раз превосходит диаметр.

3. Численные методы решения одномерных задач оптимизации

Рассмотрим следующий пример. Химический завод производит некоторое вещество. Выход продукта определяется температурой: $y = f(T)$. Температуру можно варьировать в некоторых пределах: $T_1 \leq T \leq T_2$. Вид функции f заранее не известен, он зависит от используемого сырья. Получив очередную партию сырья, нужно найти температуру T , при которой наиболее выгодно вести производство, т.е. ту температуру, при которой функция $f(T)$ достигает своего наибольшего значения.

С математической точки зрения мы имеем типичную одномерную задачу оптимизации, сформулированную в начале параграфа. В то же время между этой задачей и, скажем, задачей о консервной банке имеется существенное различие. В данном случае нет никакой формулы для целевой функции $f(T)$. Чтобы определить ее значение при некоторой температуре T , нужно провести опыт либо в лаборатории (если это возможно), либо непосредственно в производственных условиях.

Совершенно ясно, что реально провести лишь конечное число измерений. Тем самым функция $f(T)$ будет известна нам только в конечном числе точек. Значений ее производной мы вообще определить не можем. Более того, мы даже не знаем, существует ли у нее производная. Добавим, что каждое измерение требует времени, а задерживать производство нельзя. Поэтому необходимо получить ответ на поставленный вопрос после небольшого числа измерений, т.е. по значениям функции $y = f(T)$ в нескольких точках.

Возможны также задачи оптимизации, в которых целевая функция $y = f(x)$ находится в результате численного решения некоторой математической задачи. Данный случай по своему характеру близок к предыдущему: мы не имеем явной формулы для целевой функции, но можем определить ее значение для любого аргумента из области допустимых значений. Ясно, что при этом в ходе решения задачи нам окажется непосредственно доступной информация о целевой функции в конечном числе точек.

Обсудим математические вопросы, связанные со следующей постановкой одномерной задачи оптимизации: определяя значения непрерывной функции $f(x)$ в конечном числе точек некоторого отрезка $[a, b]$, нужно приближенно найти ее наименьшее (наибольшее) значение на данном отрезке.

Возможны разные подходы к решению этой задачи. Начнем с самого простого.

Метод равномерного распределения точек по отрезку (метод сетки)

Возьмем некоторое целое число n , вычислим шаг $h = (b - a) / n$ и определим значения функции $f(x)$ в точках $x_k = a + kh$ ($k = 0, 1, \dots, n$): $y_k = f(x_k)$. После этого найдем среди полученных чисел наименьшее

$$m_n = \min(y_0, y_1, \dots, y_n). \quad (6.11)$$

Число m_n можно приближенно принять за наименьшее значение функции $f(x)$ на отрезке $[a, b]$. Очевидно, что с увеличением числа точек n ошибка, которую мы допускаем, принимая m_n за минимум функции $f(x)$, стремится к нулю.

Как всегда, при построении приближенного решения встает вопрос: какое n нужно взять, чтобы погрешность в определении наименьшего значения функции не превышала заданной точности ε ? Вспомним, что, например, для метода бисекции, благодаря двухсторонней оценке корня уравнения, мы легко получили условие достижения точности ε в виде неравенства (3.14).

Для данной задачи ситуация оказывается более сложной. Если нам известно только то, что функция $f(x)$ непрерывна на отрезке $[a, b]$, то ответить на поставленный вопрос нельзя. Эта трудность не связана с предложенным способом выбора точек x_k , она носит принципиальный характер. Какое бы n мы ни взяли, всегда можно указать такую непрерывную функцию, минимум которой будет отличаться от m_n больше, чем на ε .

Справедливость этого утверждения иллюстрирует рис. 6.1. На нем приведен график непрерывной функции, которая имеет узкий «язык», опускающийся гораздо ниже точки (x_{\min}, y_{\min}) , полученной методом сетки.

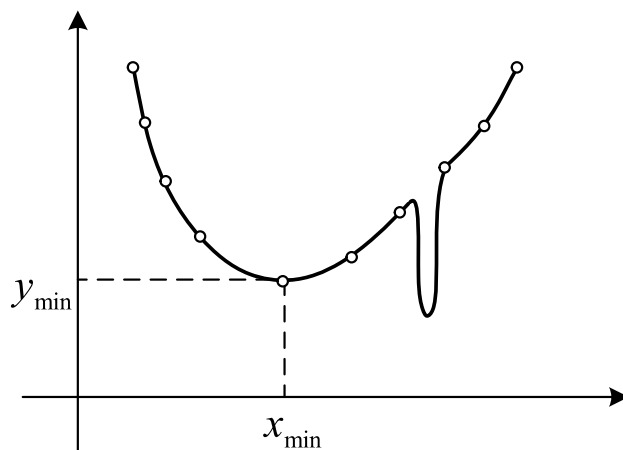


Рис. 6.1. Пример, иллюстрирующий трудности, которые могут возникнуть при приближенном определении наименьшего значения функции по ее значениям в нескольких точках

Если у нас нет перед глазами рис 6.1 (а так и бывает на практике), а известны только значения в узлах, мы приняли бы y_{\min} за наименьшее значение

функции. Разумеется, если взять большее n , то данный «язык» обнаружится, но может оказаться незамеченным другой, еще более узкий «язык».

При отсутствии дополнительной информации о свойствах функции $f(x)$ (о том, насколько «резкими» могут быть ее изменения) сомнения останутся, какое бы большое число точек мы не взяли. Поэтому при решении вопроса о числе точек и точности важно максимально полно использовать всю дополнительную информацию о свойствах целевой функции и степени ее гладкости, вытекающую из характера и особенностей задачи. Не последнюю роль играют, при этом, опыт и интуиция исследователя.

Во многих случаях из характера задачи вытекает какая-то дополнительная информация о свойствах целевой функции. Это может быть использовано для разработки специальных алгоритмов. Такой подход позволяет существенно сократить объем вычислений и получить ответ наиболее эффективным способом.

Метод сгущаемой сетки

Пусть нам известно заранее, что целевая функция $y = f(x)$ имеет на отрезке $[a, b]$ только один минимум (график такой функции показан на рис. 6.2). Для решения задачи в этом случае можно воспользоваться следующим методом.

Возьмем некоторый шаг h и будем последовательно вычислять значения функции $f(x)$ в точках $x_0 = a$, $x_1 = a + h$, $x_2 = a + 2h, \dots$, сравнивая получаемые числа y_0, y_1, y_2, \dots . Сначала они будут убывать, потом возрастать, так что обязательно найдется такая точка $x_k = a + kh$, что значение функции в ней $y_k = f(x_k)$ окажется меньше значений функции во всех остальных точках. Это значит, что наименьшее значение функции достигается на отрезке $[x_{k-1}, x_{k+1}]$ и его приближенно можно принять равным $y_k = f(x_k)$.

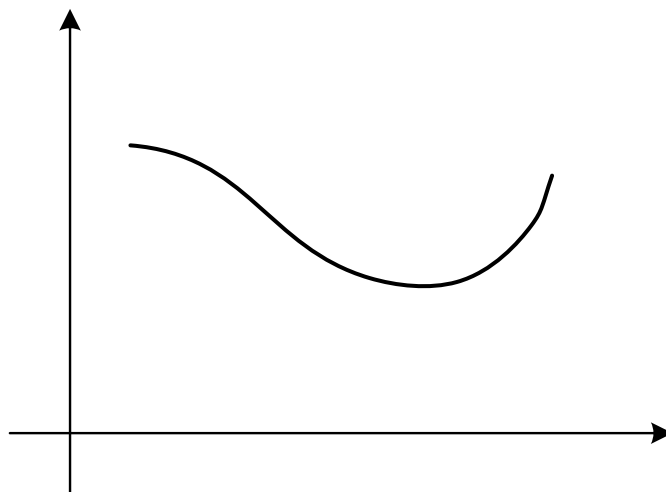


Рис. 6.2. Пример функции, имеющей один минимум

Если требуемая точность еще не обеспечена, то нужно уменьшить шаг h и повторить описанную процедуру для отрезка $[x_{k-1}, x_{k+1}]$ с новым, уменьшенным шагом. Такое последовательное сгущение сетки в окрестности текущего приближения позволяет начинать поиск наименьшего значения функции с не слишком маленьких h . В результате удастся найти решение с высокой точностью, вычислив целевую функцию в гораздо меньшем количестве точек, чем это потребовалось бы при равномерном распределении точек по отрезку.

§3. Многомерные задачи оптимизации

До сих пор мы обсуждали одномерные задачи оптимизации, в которых целевая функция зависела только от одного параметра. Однако подавляющее число реальных задач оптимизации, представляющих практический интерес, являются многомерными. В них целевая функция зависит от нескольких параметров оптимизации, причем их число может быть весьма большим.

1. Общие сведения

Математическая постановка многомерных задач оптимизации аналогична их постановке в одномерном случае: ищется наименьшее (наибольшее) значение целевой функции

$$y = f(x_1, x_2, \dots, x_n), \quad (6.12)$$

заданной на некотором множестве D допустимых значений параметров оптимизации x_1, x_2, \dots, x_n .

Каждому набору параметров оптимизации x_1, x_2, \dots, x_n можно поставить в соответствие точку некоторого n -мерного пространства с координатами (x_1, x_2, \dots, x_n) . Совокупность всех наборов параметров оптимизации, образующих множество D , занимает в этом пространстве некоторую область (эту область мы также будем обозначать буквой D). Введем следующие определения.

1. Точка M называется *внутренней точкой многомерной области D* , если существует такая окрестность этой точки, в которой содержатся точки, принадлежащие области D , и не содержатся точки, не принадлежащие этой области.

2. Точка M называется *внешней точкой по отношению к области D* , если существует такая окрестность этой точки, в которой содержатся точки, не принадлежащие области D , и не содержатся точки, принадлежащие этой области.

3. Точка M называется *граничной точкой области D* , если в любой ее окрестности содержатся как точки, принадлежащие области D , так и точки, не принадлежащие этой области.

4. Область D называется *замкнутой* (или *закрытой*), если она содержит все свои граничные точки.

5. Будем называть n -мерную область D *ограниченной*, если она целиком может быть помещена в n -мерную сферу конечного радиуса.

Т е о р е м а Вейерштрасса

Если функция $y = f(x_1, x_2, \dots, x_n)$ непрерывна в замкнутой ограниченной области D , то она достигает на этой области своего наибольшего и наименьшего значений.

Заметим, что так же, как и в случае функции одной переменной, своего наибольшего и наименьшего значений функция может достигать либо в точках экстремума, либо на границе области D .

Следует обратить особое внимание на важность такого требования в определении замкнутости области, как содержание в ней *всех* (без исключения) граничных точек. Нарушение этого требования хотя бы в одной точке может привести к тому, что теорема Вейерштрасса не будет выполняться.

Пример.

Рассмотрим функцию

$$f(x, y) = \frac{1}{x^2 + y^2}, \quad x \in (0, 10], \quad y \in (0, 10]. \quad (6.13)$$

Область определения этой функции представляет собой квадрат, две стороны которого лежат на координатных осях (рис. 6.3). Эта область включает в себя все точки ограничивающих квадрат отрезков за исключением точки $(0, 0)$. В результате для функции (6.13) не выполняется теорема Вейерштрасса: эта функция неограниченно возрастает при приближении к точке $(0, 0)$ и не достигает, таким образом, своего наибольшего значения на области определения.

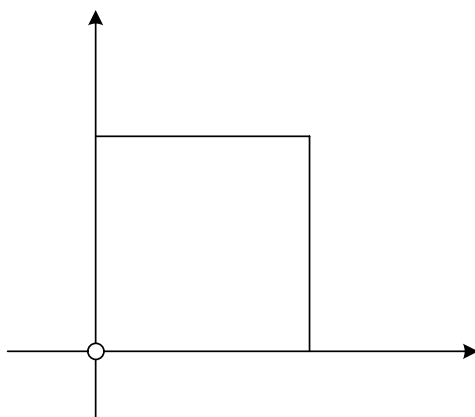


Рис. 6.3. Пример области, не содержащей одну граничную точку

Теорема Вейерштрасса выделяет класс задач оптимизации, для которых гарантировано существование решения. В дальнейшем, не оговаривая этого особо, мы всегда будем предполагать, что рассматриваемые задачи принадлежат этому классу.

Если целевая функция задается аналитической формулой и дифференцируема по всем аргументам, то тогда можно вычислить ее частные производные, получить явное выражение для градиента, определяющего в каждой точке направления возрастания и убывания функции, и использовать эту информацию для решения задачи. В других случаях никакой формулы для целевой функции нет, а имеется лишь возможность определить ее значение в любой точке рас-

сматриваемой области с помощью расчетов, в результате эксперимента и т.п. В таких задачах мы можем найти значение целевой функции лишь в конечном числе точек, и по этой информации приближенно установить ее наименьшее значение для всей области.

Отметим еще одно важное обстоятельство. При решении задач оптимизации нас интересует не только, и не столько само наименьшее (или наибольшее) значение целевой функции, сколько тот набор параметров оптимизации, при которых это наименьшее значение достигается. Поэтому, говоря о точности решения задачи, мы будем иметь в виду точность, с которой найдены значения параметров, обеспечивающие наименьшее значение целевой функции.

2. Метод сетки

Рассмотрим сначала самый простой по своей идее приближенный метод поиска наименьшего значения функции, который уже обсуждался для одномерных задач. Покроем рассматриваемую область сеткой с шагом h (рис. 6.4) и определим значения целевой функции в ее узлах. Сравнивая полученные числа между собой, найдем среди них наименьшее и примем его приближенно за наименьшее значение функции для всей области.

Этот метод – самый надежный. При достаточно малом h он позволяет найти наименьшее значение функции даже в том случае, когда она имеет несколько локальных минимумов. Он применяется для решения двумерных и трехмерных задач. Однако для задач большей размерности он практически непригоден из-за слишком быстро возрастающего по мере роста размерности количества вычислений и необходимого для их проведения времени.

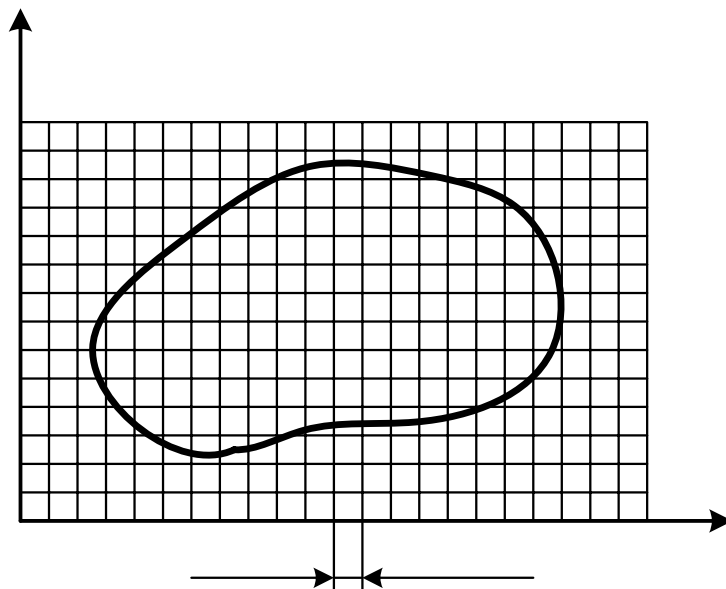


Рис. 6.4. построение сетки с шагом h

Перейдем к обсуждению методов, позволяющих вести поиск наименьшего значения функции целенаправленно.

3. Метод покоординатного спуска

Пусть нужно найти с заданной точностью ε наименьшее значение функции $y = f(M) = f(x_1, \dots, x_n)$. Здесь M – точка n -мерного пространства с координатами

татами (x_1, x_2, \dots, x_n) . Выберем какую-нибудь начальную точку $M_0 = (x_{10}, x_{20}, \dots, x_{n0})$ и рассмотрим функцию f при фиксированных значениях всех переменных, кроме первой: $f(x_1, x_{20}, \dots, x_{n0})$. Тогда она превратится в функцию одной переменной x_1 . Изменяя эту переменную, будем двигаться от начальной точки $x_1 = x_{10}$ в сторону убывания функции с выбранным шагом h , пока не дойдем до некоторого значения $x_1 = x_{11}$, после которого она начинает возрастать. Точку с координатами $(x_{11}, x_{20}, \dots, x_{n0})$ обозначим как M_1 , при этом $f(M_1) \leq f(M_0)$.

Теперь зафиксируем переменные $x_1 = x_{11}$, $x_3 = x_{30}, \dots, x_n = x_{n0}$ и рассмотрим функцию f как функцию одной переменной x_2 . Будем опять двигаться от начального значения $x_2 = x_{20}$ в сторону убывания функции с шагом h , пока не дойдем до некоторого значения $x_2 = x_{21}$, после которого она начинает возрастать. Точку с координатами $(x_{11}, x_{21}, \dots, x_{n0})$ обозначим как M_2 , при этом $f(M_2) \leq f(M_1)$.

Проведя такую же минимизацию целевой функции по остальным переменным x_3, x_4, \dots, x_n , снова вернемся к x_1 и продолжим процесс. Рано или поздно возникнет ситуация, когда движение с шагом h не приводит к уменьшению функции ни по одной переменной. В этом случае уменьшают величину шага (редукция шага) и продолжают минимизацию с меньшим значением h .

Процесс заканчивается, когда после очередной остановки процедуры для редукции шага оказывается выполненным условие

$$h < \varepsilon. \quad (6.14)$$

Таким образом, метод покоординатного спуска сводит задачу поиска наименьшего значения функции нескольких переменных к многократному решению одномерных задач минимизации.

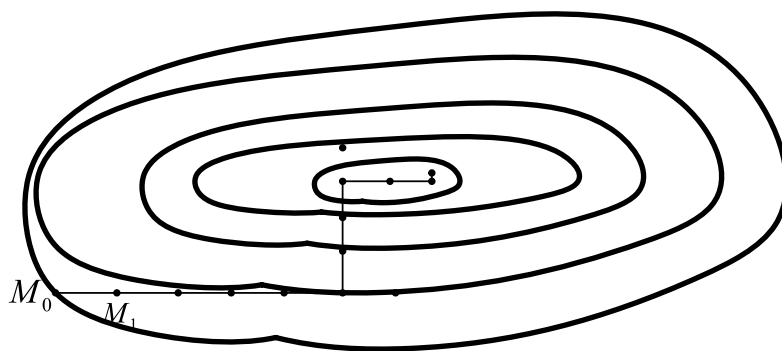


Рис. 6.5. Поиск наименьшего значения функции методом покоординатного спуска

На рис. 6.5 изображены линии уровня некоторой функции двух переменных. Показана траектория поиска ее наименьшего значения с помощью метода

покоординатного спуска. При этом следует понимать, что рисунок служит только для иллюстрации метода. Когда мы приступаем к решению реальной задачи оптимизации, такого рисунка, содержащего в себе готовый ответ, у нас, конечно, нет.

Если целевая функция задана явной формулой и дифференцируема, то мы можем вычислить ее частные производные и использовать их для определения направления убывания функции. Рассмотрим два метода, в которых используется такая возможность.

4. Метод градиентного спуска

Рассмотрим функцию f , считая для определенности, что она зависит от трех переменных x, y, z . Вычислим ее частные производные и образуем с их помощью вектор, который называют *градиентом* функции

$$\mathbf{grad} f(x, y, z) = \frac{\partial f(x, y, z)}{\partial x} \mathbf{i} + \frac{\partial f(x, y, z)}{\partial y} \mathbf{j} + \frac{\partial f(x, y, z)}{\partial z} \mathbf{k}. \quad (6.15)$$

Здесь $\mathbf{i}, \mathbf{j}, \mathbf{k}$ – единичные векторы, параллельные координатным осям. Направление вектора градиента является направлением наиболее быстрого возрастания функции в данной точке.

Введем для удобства следующие обозначения

$$a = \frac{\partial f}{\partial x}, \quad b = \frac{\partial f}{\partial y}, \quad c = \frac{\partial f}{\partial z}. \quad (6.16)$$

Тогда

$$\mathbf{grad} f = (a, b, c); \quad -\mathbf{grad} f = (-a, -b, -c), \quad (6.17)$$

а единичный вектор в направлении антиградиента будет иметь координаты:

$$\mathbf{e} = \left(\frac{-a}{\sqrt{a^2 + b^2 + c^2}}; \frac{-b}{\sqrt{a^2 + b^2 + c^2}}; \frac{-c}{\sqrt{a^2 + b^2 + c^2}} \right). \quad (6.18)$$

Перейдем к описанию метода градиентного спуска. Выберем начальную точку $M_0(x_0, y_0, z_0)$ и вычислим в ней градиент целевой функции. Сделаем шаг величины h в направлении антиградиента. В результате мы придем в точку M_1 , координаты которой

$$\left(x_0 - \frac{ah}{\sqrt{a^2 + b^2 + c^2}}; y_0 - \frac{bh}{\sqrt{a^2 + b^2 + c^2}}; z_0 - \frac{ch}{\sqrt{a^2 + b^2 + c^2}} \right), \quad (6.19)$$

где обозначения a, b и c задаются по формулам (6.16).

В новой точке вся процедура повторяется. Продолжая этот процесс, мы будем двигаться в сторону убывания функции. За счет специального выбора направления движения на каждом шаге приближение к наименьшему значению функции в этом случае будет более быстрым, чем в методе покоординатного спуска.

После того, как очередной шаг не приведет к уменьшению значения целевой функции, делают редукцию шага и уточняют найденное решение. Про-

цесс заканчивается, когда после очередной остановки процедуры оказывается выполненным условие (6.14).

Метод градиентного спуска требует вычисления градиента целевой функции на каждом шаге. Если она задана аналитически, то для частных производных, определяющих градиент, как правило, удастся получить явные формулы. В противном случае частные производные в нужных точках можно вычислить приближенно, заменяя их соответствующими разностными соотношениями

$$\begin{aligned} a &= \frac{\partial f}{\partial x} = \frac{f(x + \Delta x, y, z) - f(x, y, z)}{\Delta x}, \\ b &= \frac{\partial f}{\partial y} = \frac{f(x, y + \Delta y, z) - f(x, y, z)}{\Delta y}, \\ c &= \frac{\partial f}{\partial z} = \frac{f(x, y, z + \Delta z) - f(x, y, z)}{\Delta z}. \end{aligned} \quad (6.20)$$

5. Метод наискорейшего спуска

Вычисление градиента на каждом шаге, позволяющее все время двигаться в направлении наиболее быстрого убывания целевой функции, может в то же время замедлить вычислительный процесс. Дело в том, что подсчет градиента – обычно гораздо более трудоемкая операция, чем подсчет самой функции. Поэтому часто пользуются модификацией градиентного метода, получившей название метода наискорейшего спуска.

Согласно этому методу после вычисления в начальной точке градиента функции делают в направлении антиградиента не один шаг, а несколько шагов до тех пор, пока функция продолжает убывать. Достигнув наименьшего значения на выбранном направлении, снова вычисляют градиент функции и повторяют описанную процедуру. При этом градиент вычисляется гораздо реже, только при смене направлений движения. Редукция шага и остановка вычислительной процедуры осуществляются так же, как и в методе градиентного спуска.

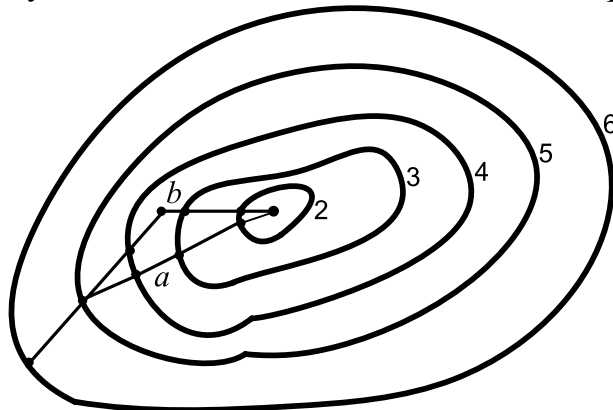


Рис. 6.6. Поиск наименьшего значения функции:

a – методом градиентного спуска,
 b – методом наискорейшего спуска

На рис. 6.6 показаны траектории поиска наименьшего значения функции двух переменных методами градиентного спуска и наискорейшего спуска. Хотя траектория второго метода ведет к цели не так быстро, экономия времени счета благодаря более редкому вычислению градиента может в некоторых случаях оказаться весьма существенной.

6. Проблема «оврагов»

Мы рассмотрели три варианта методов спуска. Их работа проиллюстрирована на рис. 6.5 и 6.6. Могло сложиться впечатление, что любой из этих методов всегда и без всяких проблем приводит к решению задачи минимизации функции с заданной точностью. На самом деле все было так хорошо потому, что на упомянутых рисунках были выбраны «удобные» функции. Но посмотрите на рис. 6.7. На нем также показаны линии уровня некоторой функции, однако их конфигурация отличается от тех, что показаны на рис. 6.5 и 6.6. Линии уровня сильно вытянуты в одном направлении и сплюснены в другом. К тому же, сама эта сплюсненная конфигурация изогнута. Картина напоминает рельеф местности с оврагом. Случай «оврага» (этот нематематический термин прочно закрепился в литературе) крайне неудобен для описанных методов спуска.

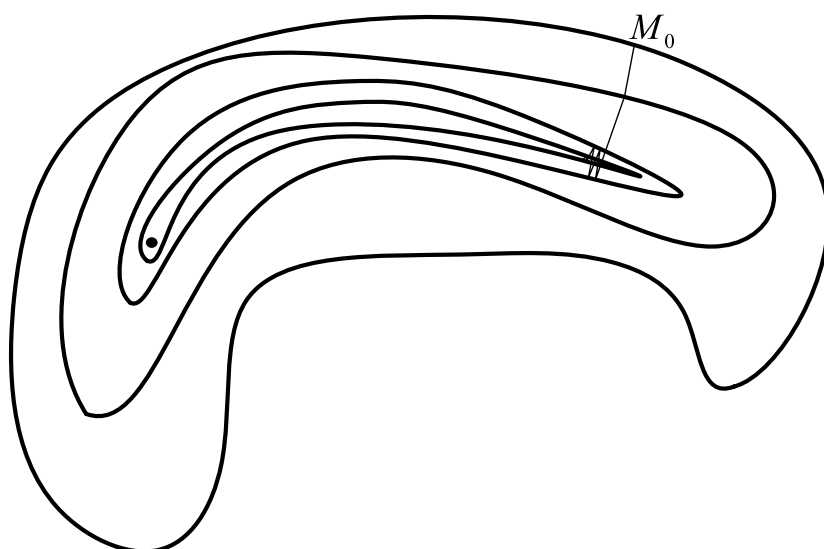


Рис. 6.7. Поиск наименьшего значения функции в случае «оврага»

Действительно, попытаемся найти наименьшее значение такой функции с помощью метода градиентного спуска. Двигаясь, например, из точки M_0 все время в направлении антиградиента (рис. 6.7), мы быстро спустимся на дно «оврага» и, поскольку движение идет хотя и маленькими, но конечными шагами, проскочим его. Оказавшись на противоположной стороне «оврага» и вычислив там градиент функции, мы вынуждены будем развернуться и сделать один или несколько шагов почти в обратном направлении. При этом мы почти наверняка снова проскочим дно «оврага» и вернемся на его первоначальную сторону.

Продолжая этот процесс, мы вместо того, чтобы двигаться по дну «оврага» в сторону его понижения, будем совершать зигзагообразные скачки поперек

«оврага», почти не приближаясь к цели (на рис. 6.7 она помечена точкой). Таким образом, в случае «оврага» методы спуска оказываются неэффективными.

Для решения таких задач приходится разрабатывать специальные методы. Мы не будем на них останавливаться. Сделаем лишь одно замечание. Многие исследователи, обладающие опытом решения практических задач оптимизации, утверждают, что подобные целевые функции с узкими «оврагами» возникают обычно в случае, когда математическая модель рассматриваемого объекта построена неудачно. Это утверждение трудно обосновать, но в ряде случаев пересмотр математической модели и (или) способа задания целевой функции позволяет избавиться от проблемы «оврагов» или, по крайней мере, значительно смягчить ее.

7. Проблема многоэкстремальности

Посмотрите на рис. 6.5 – 6.7 и сравните их с рис. 6.8. Первые три рисунка относятся к функциям, имеющим только один минимум. Поэтому, откуда бы мы ни начали поиск, мы придем в конце концов к одной и той же точке минимума. На рис. 6.8 приведены линии уровня функции с двумя локальными минимумами в точках O_1 и O_2 . Такие функции принято называть *многоэкстремальными*. Сравнивая между собой значения функции в точках $O_1, O_2 : f_1 = 3, f_2 = 1$, находим, что наименьшее значение функция достигает в точке O_2 .

Представьте теперь, что, не имея перед глазами рис. 6.8 и не зная о многоэкстремальности функции, мы начали поиск наименьшего значения с помощью метода градиентного спуска из точки A_1 . Поиск приведет нас в точку O_1 , которую можно ошибочно принять за искомый минимум. С другой стороны, если мы начнем поиск с точки A_2 , то окажемся на правильном пути и быстро придем в точку O_2 .

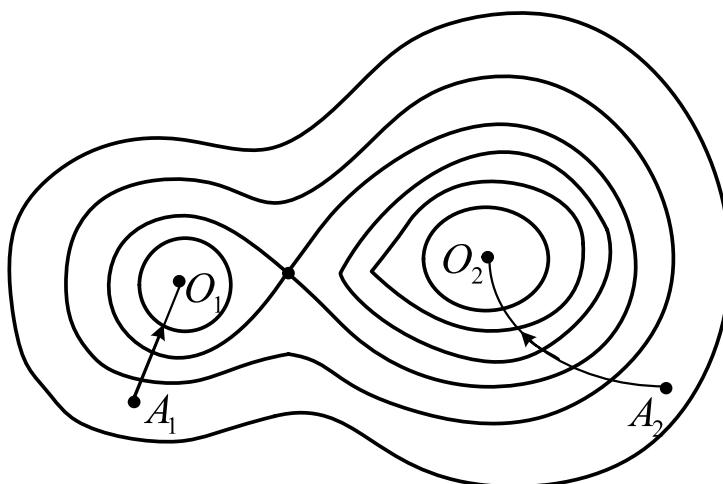


Рис. 6.8. Пример функции с двумя локальными минимумами в точках O_1 и O_2

Как бороться с многоэкстремальностью? Универсального ответа на этот вопрос нет. Самый простой прием состоит в том, что поиск проводят несколько

раз, начиная его с разных точек. Если в результате получаются разные ответы, то сравнивают полученные значения целевой функции и выбирают наименьшее. Для того чтобы не пропустить какой-нибудь локальный минимум, начальные точки поиска должны покрывать всю область допустимых значений параметров. Выбор начальных точек поиска, обоснованность прекращения расчетов в значительной степени зависят от опыта и интуиции специалистов, решающих задачу.

§4. Задачи линейного программирования

В данном параграфе мы познакомимся с линейным программированием. Так называются задачи оптимизации, в которых целевая функция является линейной функцией своих аргументов, а ограничения, наложенные на параметры оптимизации, имеют вид линейных уравнений или неравенств.

Отличительная особенность этих задач заключается в том, что у линейных функций не бывает экстремумов, поэтому наибольшего и наименьшего своих значений они достигают всегда на границе области допустимых значений параметров. Задача линейного программирования, таким образом, сводится к отысканию этой точки (точек) на границе ОДЗП.

Линейное программирование начало развиваться в первую очередь в связи с задачами экономики, с поиском способов оптимального распределения и использования ресурсов. Оно послужило основой широкого использования математических методов в экономике. Следует подчеркнуть, что в реальных экономических задачах число переменных обычно бывает очень большим (тысячи, десятки тысяч). Поэтому практическая реализация алгоритмов решения таких задач принципиально невозможна без использования современной вычислительной техники.

В качестве примеров мы рассмотрим две типичные задачи линейного программирования: транспортную задачу и задачу об оптимальном использовании ресурсов.

1. Транспортная задача

Познакомимся с этим типом задач линейного программирования на примере конкретной задачи. В городе имеются два склада муки и два хлебозавода. Ежедневно с первого склада вывозится 50 т муки, со второго – 70 т. Мука доставляется на хлебозаводы, причем первый завод получает 40 т, второй – 80 т. Допустим, что перевозка одной тонны муки с первого склада на первый завод стоит 120 р., с первого склада на второй завод – 160 р., со второго склада на первый завод – 80 р. и со второго склада на второй завод – 100 р. Как нужно спланировать перевозки, чтобы их стоимость была минимальной?

Р е ш е н и е

Параметрами оптимизации в этой задаче являются объемы перевозок по каждому из четырех возможных маршрутов. Обозначим через x_1 и x_2 количество муки, которое следует перевезти с первого склада на первый и второй за-

воды соответственно, а через x_3 и x_4 – количество муки, которую нужно перевезти со второго склада на первый и второй заводы (рис. 6.9).

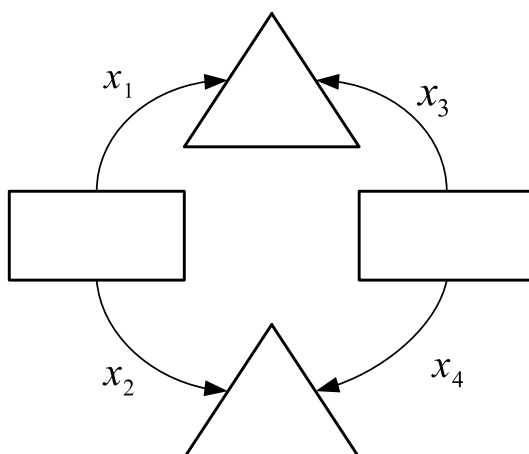


Рис. 6.9. Схема транспортной задачи и параметры оптимизации

По условию задачи на параметры оптимизации наложены следующие ограничения:

$$\begin{aligned} x_1 + x_2 &= 50, \\ x_3 + x_4 &= 70, \\ x_1 + x_3 &= 40, \\ x_2 + x_4 &= 80, \\ x_i &\geq 0, \quad i = 1, 2, 3, 4. \end{aligned} \quad \text{Склад 1}^{(6.21)} \quad (6.22)$$

Первые два уравнения системы (6.21) определяют, сколько муки нужно вывезти с каждого склада, два других уравнения показывают, сколько муки нужно привезти на каждый завод. Неравенства (6.22) означают, что в обратном направлении с заводов на склады муку не возят. Общая стоимость всех перевозок определяется формулой

$$f = 120x_1 + 160x_2 + 80x_3 + 100x_4. \quad (6.23)$$

С математической точки зрения задача заключается в том, чтобы найти числа x_1, x_2, x_3, x_4 , удовлетворяющие условиям (6.21), (6.22) и минимизирующие стоимость перевозок (6.23).

Рассмотрим систему (6.21). Это система четырех линейных уравнений с четырьмя неизвестными. Однако независимыми в ней оказываются только первые три. Если сложить два первых уравнения и вычесть третье, получится четвертое уравнение. Это значит, что четвертое уравнение является следствием первых трех. Таким образом, нужно рассмотреть следующую систему, эквивалентную (6.21):

$$\begin{aligned} x_1 + x_2 &= 50, \\ x_3 + x_4 &= 70, \\ x_1 + x_3 &= 40. \end{aligned} \quad (6.24)$$

В ней число уравнений на единицу меньше числа неизвестных, так что мы можем выбрать какую-нибудь неизвестную, например x_1 , и выразить через нее с помощью уравнений (6.24) три остальных. Соответствующие формулы имеют вид

$$\begin{aligned}x_2 &= 50 - x_1, \\x_3 &= 40 - x_1, \\x_4 &= 30 + x_1.\end{aligned}\tag{6.25}$$

Учитывая (6.22), получаем систему неравенств

$$\begin{aligned}x_1 &\geq 0, \\50 - x_1 &\geq 0, \\40 - x_1 &\geq 0, \\30 + x_1 &\geq 0,\end{aligned}\tag{6.26}$$

из которой

$$0 \leq x_1 \leq 40.\tag{6.27}$$

Таким образом, задавая любое x_1 , удовлетворяющее (6.27), и вычисляя x_2, x_3, x_4 по формулам (6.25), мы получим один из возможных планов перевозки. При реализации каждого такого плана с каждого склада будет вывезено и на каждый завод доставлено нужное количество муки. При этом стоимость реализации каждого плана будет, конечно же, разной.

Вычислим стоимость перевозок. Для этого подставим (6.25) в формулу (6.23). В результате получим

$$f = 14200 - 20x_1.\tag{6.28}$$

Формула (6.28) определяет величину f как функцию одной переменной x_1 , которую можно выбирать произвольно в пределах условий (6.27). Стоимость окажется минимальной, если мы придадим величине x_1 наибольшее возможное значение: $x_1 = 40$. Значения остальных параметров оптимизации находятся по формулам (6.25).

Итак, оптимальный по стоимости план перевозок имеет вид

$$x_1 = 40, \quad x_2 = 10, \quad x_3 = 0, \quad x_4 = 70.\tag{6.29}$$

Стоимость перевозок по этому плану составляет 13400 р. При любом другом допустимом плане перевозок стоимость окажется выше: $f > f_{\min} = 13400$.

2. Задача об оптимальном использовании ресурсов

С этим типом задач также познакомимся на конкретном примере. Мебельная фабрика выпускает стулья двух типов. На изготовление одного стула первого типа, стоимостью 400 р., расходуется 2 м досок стандартного сечения, $0,5 \text{ м}^2$ обивочной ткани и 2 чел./ч рабочего времени. Для стульев второго типа аналогичные данные составляют: 600 р., 4 м, $0,25 \text{ м}^2$ и 2,5 чел./ч.

Допустим, что в распоряжении фабрики имеется 440 м досок, 65 м^2 обивочной ткани, 320 чел./ч. рабочего времени. Какое количество стульев надо из-

готовить, чтобы в рамках имеющихся ресурсов стоимость произведенной продукции была максимальной?

Решение

Обозначим через x_1 и x_2 запланированное к производству число стульев первого и второго типов соответственно. Ограниченный запас сырья и трудовых ресурсов означает, что x_1 и x_2 должны удовлетворять неравенствам

$$\begin{aligned} 2x_1 + 4x_2 &\leq 440, \\ 0,5x_1 + 0,25x_2 &\leq 65, \\ 2x_1 + 2,5x_2 &\leq 320. \end{aligned} \quad (6.30)$$

Кроме того, по смыслу задачи они должны быть неотрицательными:

$$x_1 \geq 0, \quad x_2 \geq 0. \quad (6.31)$$

Стоимость запланированной к производству продукции определяется формулой

$$f(x_1, x_2) = 400x_1 + 600x_2. \quad (6.32)$$

Итак, с математической точки зрения задача составления оптимального по стоимости выпущенной продукции плана производства сводится к определению пары целых чисел x_1 и x_2 , удовлетворяющих линейным неравенствам (6.30), (6.31) и дающих наибольшее значение линейной функции (6.32). Это типичная задача линейного программирования. По своей постановке она немного отличается от транспортной задачи, но это различие несущественно.

Для анализа сформулированной задачи рассмотрим плоскость и введем на ней декартову систему координат x_1, x_2 . Найдем на этой плоскости множество точек, координаты которых удовлетворяют (6.30), (6.31). Неравенства (6.31) означают, что это множество лежит в первой четверти. Выясним смысл ограничений, которые задаются неравенствами (6.30).

Проведем на плоскости прямую, определяемую уравнением

$$2x_1 + 4x_2 = 0. \quad (6.33)$$

Она делит плоскость на две полуплоскости (рис. 6.10). На одной из них, располо-

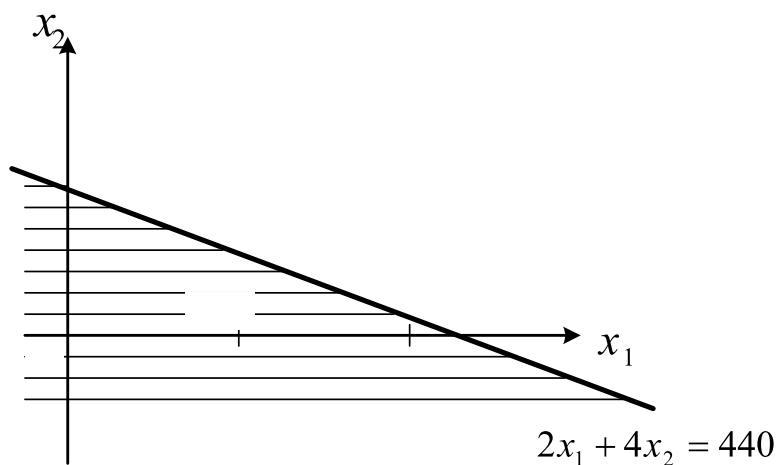


Рис. 6.10. Решение неравенства $2x_1 + 2x_2 \leq 440$

женной ниже прямой (6.33), функция $F_1(x_1, x_2) = 2x_1 + 4x_2 - 440$ принимает отрицательные значения, на другой, расположенной выше прямой (6.33), – положительные. Таким образом, первое из неравенств (6.30) выполняется на множестве точек, которое включает в себя прямую (6.33) и полуплоскость, расположенную ниже этой прямой. На рис. 6.10 соответствующая часть плоскости заштрихована.

Совершенно аналогично можно найти множества точек, удовлетворяющих второму и третьему неравенствам из системы (6.30). Они показаны на рис. 6.11, 6.12 штриховкой.

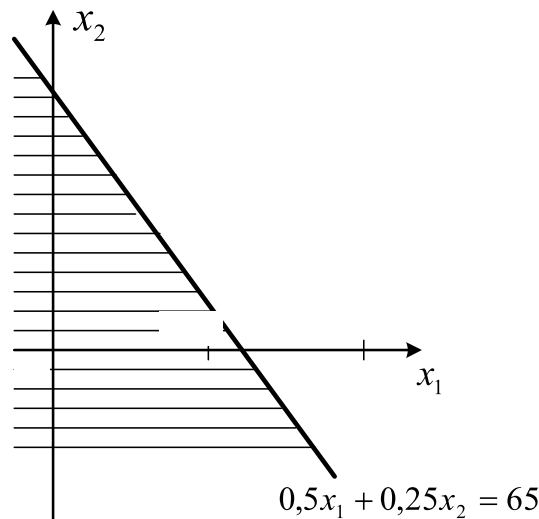


Рис. 6.11. Решение неравенства $0,5x_1 + 0,25x_2 \leq 65$

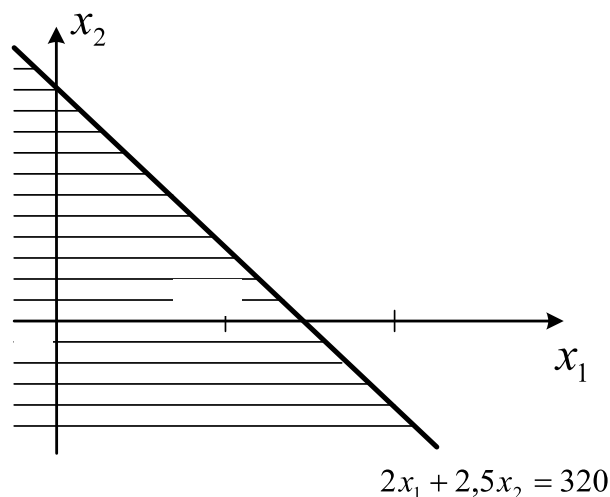


Рис. 6.12. Решение неравенства $2x_1 + 2,5x_2 \leq 320$

Возьмем пересечение трех найденных множеств и выделим его часть, расположенную в первой четверти. В результате получим множество точек, удовлетворяющих всей совокупности ограничений (6.30), (6.31). Данное множество имеет вид пятиугольника, показанного на рис. 6.13. Его вершинами являются точки пересечения прямых, на которых неравенства (6.30), (6.31) пере-

ходят в точные равенства. Координаты вершин пятиугольника указаны на рис. 6.13.

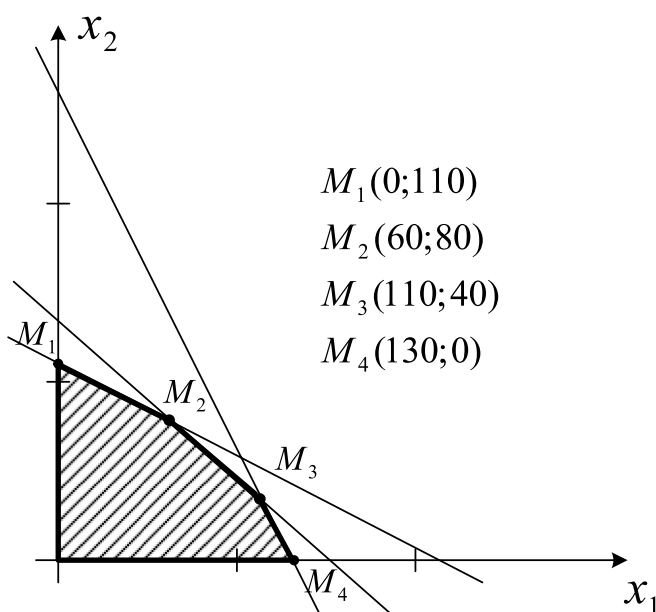


Рис. 6.13. Пятиугольник $OM_1M_2M_3M_4$, координаты которого удовлетворяют системе неравенств (6.30), (6.31)

Любой точке M с целочисленными координатами (x_1, x_2) , принадлежащей данному пятиугольнику, соответствует план выпуска стульев, который может быть выполнен при имеющихся запасах сырья и трудовых ресурсов (реализуемый план). Наоборот, если точка M не принадлежит пятиугольнику, то соответствующий план не может быть выполнен (нереализуемый план).

Рассмотрим на плоскости x_1, x_2 линии

$$400x_1 + 600x_2 = C. \quad (6.34)$$

Уравнение (6.34) описывает семейство прямых, параллельных прямой

$$400x_1 + 600x_2 = 0. \quad (6.35)$$

При параллельном переносе этой прямой вверх параметр C возрастает, при переносе вниз – убывает.

Свойства функции (6.32) тесно связаны с прямыми (6.34). Вдоль каждой из них она сохраняет постоянное значение равное C , а при переходе с одной прямой на другую ее значение меняется. Иными словами, прямые (6.34) – это линии уровня функции (6.32). Значение функции вдоль линии уровня (6.34) тем больше, чем больше число C , т. е. чем дальше от начала координат расположена эта линия. Отсюда следует важный вывод: оптимальный план должен располагаться на прямой семейства (6.34), наиболее удаленной от начала координат.

Этот вывод позволяет закончить решение задачи. Посмотрите на рис. 6.14. На нем воспроизведен пятиугольник реализуемых планов и проведена прямая семейства (6.34), проходящая через точку M_2 с координатами $(60, 80)$. Она является предельной прямой семейства, имеющей общую точку с пятиугольником.

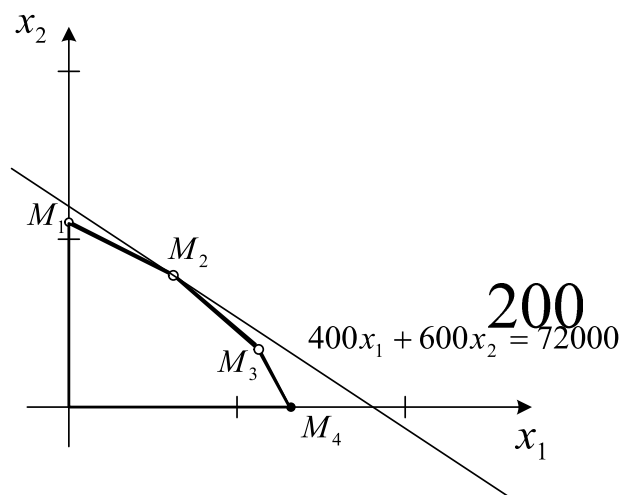


Рис. 6.14. Определение оптимального плана производства стульев

Итак, оптимальный план найден. Он предписывает производить 60 стульев первого типа и 80 стульев второго типа. Стоимость этой продукции 72000 р. На выполнение данного плана нужно затратить: 440 м досок, 50 м² обивочной ткани, 320 чел./ч рабочего времени. Оптимальный план требует полного использования запаса досок и трудовых ресурсов, в то время как обивочная ткань будет израсходована не полностью – останется 15 м².

Последний вывод ясен из рис. 6.14. Точка M_2 , определяющая оптимальный план, является вершиной пятиугольника, лежащей на пересечении прямых

$$\begin{aligned} 2x_1 + 4x_2 &= 440, \\ 2x_1 + 2,5x_2 &= 320. \end{aligned} \quad (6.36)$$

Уравнения прямых (6.36) получаются из первого и третьего условий системы (6.30) при замене их на строгие равенства. Это означает полный расход досок и трудовых ресурсов. Однако точка M_2 лежит ниже прямой

$$0,5x_1 + 0,25x_2 = 65,$$

так что второе условие системы (6.30), связанное с ограниченностью запаса ткани, имеет в ней форму неравенства $50 < 65$.

Проведенный анализ показывает, что дальнейшее увеличение стоимости продукции регламентируется запасом досок и трудовыми ресурсами.

В заключение сделаем несколько замечаний.

1. При практическом решении подобных задач вовсе не обязательно строить прямые семейства (6.34), как на рис. 6.14. Достаточно вычислить значения целевой функции (6.32) в точках пересечения прямых, ограничивающих ОДЗП (в рассмотренном случае это точки M_1, M_2, M_3, M_4) и выбрать среди них точку, в которой это значение максимально.

2. Если координаты точек пересечения получились дробными, задача решается точно так же, только в окончательном решении отбрасываются дробные части полученных значений параметров оптимизации.

3. Если в двух соседних точках (например, M_2, M_3) значения целевой функции оказались одинаковыми, это значит, что отрезок M_2M_3 лежит на одной линии уровня. В качестве решения задачи может быть выбрана любая точка этого отрезка, имеющая целочисленные координаты. Если точек с целочисленными координатами на отрезке нет, выбирают любую точку, в которой округление, вызванное отбрасыванием дробных частей, приводит к меньшему изменению значений параметров. (Таких точек может быть несколько.)

Мы рассмотрели две задачи линейного программирования. Небольшое количество переменных (4 маршрута и 2 вида продукции) позволило просто и наглядно получить их решение.

В настоящее время разработано большое количество компьютерных программ, реализующих общие алгоритмы решения задач линейного программирования для практически неограниченного количества параметров. Использование вычислительной техники при решении задач линейного программирования послужило основой широкого применения математических методов в экономике.

§5. Общие рекомендации. Что и как оптимизировать?

Мы познакомились с постановкой задач оптимизации и рассмотрели ряд способов минимизации функций (именно к этому сводится математическая сторона вопроса). Нами охвачена лишь небольшая часть известных методов решения подобных задач. Вместо дальнейшего изучения методов обратим внимание на другие, не менее важные вопросы.

Решение задач оптимизации состоит из следующих этапов:

1. Создание математической модели явления (объекта, процесса).
2. Определение целевой функции и важнейших параметров, подлежащих оптимизации.
3. Непосредственная минимизация некоторой функции (обычно большого числа переменных).
4. Внедрение результатов исследования.

Было бы ошибочно считать, что первые три из них относятся к исключительной компетенции математиков, и лишь на последнем этапе должны подключаться специалисты конкретной отрасли. Участие этих специалистов крайне необходимо и на первом, и особенно на втором этапе. В противном случае возможны серьезные просчеты. Н.С. Бахвалов описывает такую реально произошедшую ситуацию (Бахвалов Н.С. Численные методы. – М.: Наука, 1973. – 631с.):

«Производство сахара складывается из следующих этапов: выращивание сахарной свеклы, транспортировка, переработка на заводах. Математики, поставившие перед собой задачу оптимизировать производство сахарной свеклы, в первую очередь решили заняться минимизацией транспортных расходов. Эта задача была успешно решена, но хозяйственные руководители отказались принять полученный план перевозок к действию. Основой послу-

жил довод о том, что при оптимизации транспортировки свеклы ухудшались условия транспортировки других сельскохозяйственных культур, а также усложнялась организация переработки: некоторые заводы оказывались перегруженными, другие недогруженными.

За этими возражениями стояли такие соображения. Себестоимость сахара складывается из следующих составляющих (числа условные):

80% – стоимость сахарной свеклы,

5% – транспортные расходы,

15% – расходы по переработке.

Ожидаемая экономия от минимизации транспортных расходов составляет 10% этих расходов, т.е. всего 0,5% себестоимости сахара. В то же время эта перестройка требует больших хлопот и неизвестно, не произойдет ли более существенных потерь за счет каких-то неучтенных моментов.

Из описанной выше структуры себестоимости видно, что за счет оптимизации расходов по транспортировке и переработке трудно достигнуть существенного повышения эффективности. Внимательный анализ проблемы позволил обнаружить следующее обстоятельство. Выход сахара из тонны свеклы пропорционален ее сахаристости, которая изменяется по определенному закону в процессе созревания и падает в процессе хранения. В частности, при позднем начале уборки происходят большие потери из-за того, что на сахарные заводы поступает одновременно большое количество свеклы, и она там долго хранится.

В результате этого анализа была поставлена новая цель – оптимизация выхода сахара (при заданном оборудовании) и построена математическая модель, учитывающая изменение сахаристости в процессе копki и хранения свеклы.

Дальнейший анализ показал, что и эта модель обладает определенным дефектом. При оптимальном (согласно этой модели) графике работы копка должна начинаться, когда свекла еще находится в стадии интенсивного роста, что экономически невыгодно для непосредственных производителей. Для преодоления этого противоречия потребовалось, чтобы модель учитывала вопросы экономической заинтересованности производителей.»

Описанный случай весьма поучителен и позволяет сделать ряд полезных выводов.

Во-первых, никогда не следует браться за первую попавшуюся задачу только потому, что хорошо умеешь ее решать. Следует рассмотреть явление со всех сторон, выявить те факторы, которые оказывают наибольшее влияние на интересующий нас результат, и оптимизировать, в первую очередь, именно эти факторы. Если мы ставим целью снизить себестоимость чего-либо, то в первую очередь надо оптимизировать самую дорогостоящую операцию.

Во-вторых, улучшая что-то одно, мы неизбежно ухудшаем что-то другое. Поэтому, ставя задачу оптимизации, надо учитывать не только локальный выигрыш, но и все возможные проигрыши. Например, удешевление производства или строительства может привести к некоторому снижению качества, что требует более частых ремонтов. В результате эксплуатация таких «более дешевых» изделий или объектов обойдется дороже.

В-третьих, оптимальные производственные планы или графики работ часто оказываются невыгодными для отдельных участников общей работы. Например, оптимизация транспортных расходов неизбежно снижает доходы транспортников. Поэтому результатом решения оптимизационной задачи должен быть не только производственный план или график работ, но и рекомендации по наилучшей организации оплаты труда и ценовой политике.