Парадигма развития науки

Методологическое обеспечение

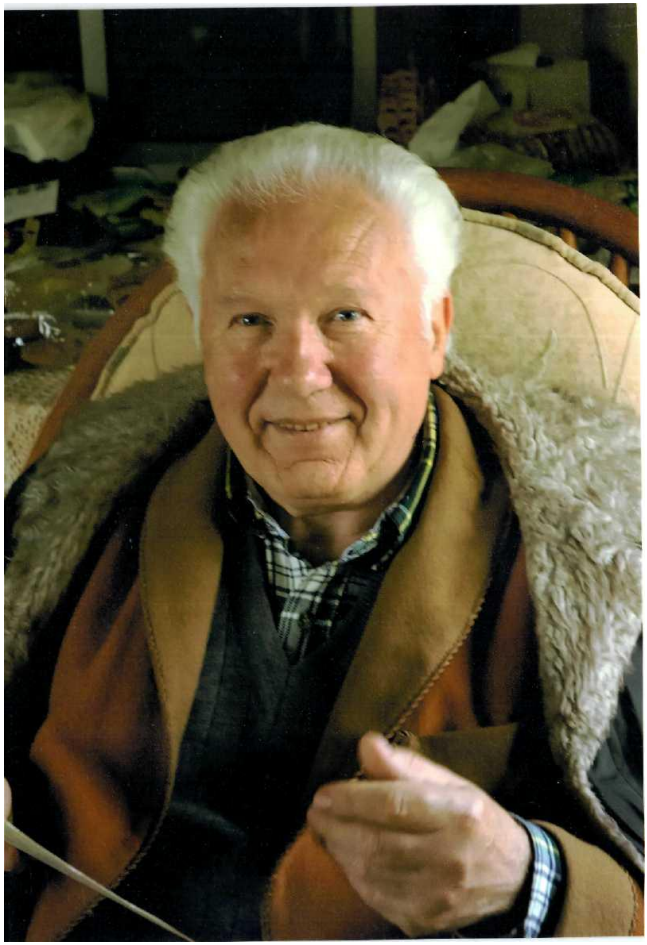
А.Е. Кононюк

ОСНОВЫ ТЕОРИИ ОПТИМИЗАЦИИ

Книга 1

Начала

Киев
Освіта України
2011




Структурная схема развивающейся панмедийной системы наук
I. Группа междисциплинарных наук
ПН Теория понятий
ЛГ Теория лингвистики
ПЗ Теория познания
СЗ Теория созидания
ПТ Общая теория понятоведения
II. Группа наук познания (фундаментальных наук)
ИН Теория информации
СТ Теория систем
РС Теория распознавания
КС Теория консалтинга (консультирования)
ПР Теория принятия решения
РР Теория реализации принятых решений
III. Группа наук созидания (прикладных наук)
СС Теория социальных систем
ЖО Теория жизнеобеспечения
КМ Теория коммуникаций и транспортирования
ОБ Теория образования (обучения)
СР Теория сохранения (защиты и охраны)
РЗ Теория развития
IV. Группа общедисциплинарных наук (обеспечивающих наук)
ДМ Теория дискретно-непрерывной математики (четкая и нечеткая)
НИ Теория научных исследований
ММ Теория моделей и моделирования
ОМ Теория оптимизации

УДК 51 (075.8)
ББК В161.я7
К 213

Рецензент: *Н.К.Печурин* - д-р техн. наук, проф. (Национальный авиационный университет).

Кононюк А.Е.
К65 Основы теории оптимизации. К.1.
Киев:"Освіта України", 2011. - 692 с.
ISBN 978-966-7599-50-8

Настоящая работа является систематическим изложением базовой теории оптимизации для конечномерных задач. Основное внимание уделяется идейным основам методов, их сравнительному анализу и примерам использования. Охвачен широкий круг задач — от безусловной минимизации до условной минимизации. Обсуждается методика постановки и решения прикладных проблем оптимизации. Приводятся условия экстремума, теоремы существования, единственности и устойчивости решения для основных классов задач. Исследуется влияние помех, негладкости функций, вырожденности минимума.

Работа предназначена для магистров, аспирантов, докторантов, инженеров, экономистов, вычислителей и всех тех, кто сталкивается с задачами оптимизации.

ББК В161.я7

ISBN 978-966-7599-50-8

Оглавление

Предисловие

Широкое распространение задач оптимизации в науке, технике, экономике, управлении требует изложения методов решения подобных задач. Однако ученому, инженеру или вычислителю трудно ориентироваться в литературе по оптимизации (большинство имеющихся книг написано «математиками для математиков»), нелегко разобраться в многообразии задач и алгоритмов. В этой работе делается попытка систематического изложения общей теории и методов оптимизации в форме, доступной как ученому, так и инженеру. Используемый математический аппарат минимален — достаточно знания начал математического анализа, линейной алгебры и теории вероятностей. Основные сведения из математического анализа приводятся в первой книге настоящей работы. Изложение построено на последовательном усложнении рассматриваемых задач. Вначале описываются наиболее простые задачи безусловной минимизации гладких функций, затем исследуется влияние различных осложняющих факторов — помех, негладкости функций, вырожденности минимума, наличия ограничений. Анализ каждого класса задач проводится единообразно — вводится требуемый математический аппарат, затем обосновываются условия экстремума, результаты о существовании, единственности и устойчивости решения, и, наконец, описываются основные методы решения и исследуются их свойства. Главное внимание уделяется идейным основам методов, их сравнительному анализу; показано, как теоретические результаты служат фундаментом при построении и изучении методов. На примерах прикладных задач оптимизации обсуждается взаимоотношение общих и специальных методов решения. Дана обширная комментированная библиография, позволяющая читателю в случае надобности обратиться к более подробным работам на интересующую его тему.

Включенный в книгу материал во многом отличается от традиционного. Нередко учебники по математическому программированию сводятся к описанию техники симплекс-метода линейного программирования. Нам этот круг вопросов не кажется центральным; ему посвящен лишь один параграф. В то же время большое внимание уделено задаче безусловной минимизации, которой посвящена полностью вся вторая книга настоящей работы, приводится богатый материал для обсуждения основных идей теории и методов оптимизации. Среди нестандартных разделов книги — задачи негладкой оптимизации, вырожденные и нестационарные задачи,

задачи с ограничениями типа равенств, условия устойчивости экстремума, влияние помех на методы оптимизации, анализ общих схем исследования сходимости итеративных методов и т. д.

Книга в основном посвящена конечномерным задачам. Это обусловлено как ограничениями на объем работы, так и предполагаемым уровнем математических знаний. Поэтому не рассматриваются такие важнейшие вопросы, как современная теория условий оптимальности в общих экстремальных задачах, задачи вариационного исчисления и оптимального управления и т. д. Вместе с тем нам кажется, что конечномерный случай очень богат идеями и результатами; он может служить прекрасной «моделью» более общих задач оптимизации. Знакомый с функциональным анализом читатель без труда заметит, что многие утверждения автоматически переносятся на задачи в гильбертовом или банаховом пространстве, однако в тексте подобные обобщения не приводятся. В отдельной книге настоящей работы рассматриваются также дискретные задачи оптимизации. Как видно из их изложения, они требуют совсем иных методов исследования, чем непрерывные, и примыкают к комбинаторике и математической логике.

Следует отметить, что у математиков, вычислителей и практиков различен подход к данному предмету. Предлагаемая работа представляет собой попытку некоторого компромиссного решения, рассчитанного на все эти категории читателей.

Введение

Обычно наши действия в условиях неоднозначности выбора определяются некоторой целью, которую мы стремимся достичь наилучшим образом. Тем самым человеческая деятельность связана с постоянным (сознательным или бессознательным) решением оптимизационных задач. Более того, многие законы природы носят вариационный характер, хотя здесь и неуместно говорить о наличии цели.

Можно было бы думать, что подобная распространенность задач оптимизации должна была найти свое отражение в математике. Однако в действительности до середины нынешнего столетия задачи на экстремум рассматривались в математике лишь эпизодически, развитая теория и методы решения подобных задач были созданы сравнительно недавно.

Наиболее простая *задача безусловной минимизации функции многих переменных* привлекла внимание математиков во времена, когда закладывались основы математического анализа. Она во многом стимулировала создание дифференциального исчисления, а необходимое условие экстремума (равенство градиента нулю), полученное Ферма в 1629 г., явилось одним из первых крупных результатов анализа. Позже в работах Ньютона и Лейбница были по существу сформулированы условия экстремума II порядка (т. е. в терминах вторых производных) для этой задачи.

Другой класс задач на экстремум, традиционно рассматривавшийся в математике, — это *задачи вариационного исчисления.* Интерес к ним проявлялся и в античной математике (разного рода изопериметрические проблемы), однако подлинное рождение вариационного исчисления произошло в конце XVIII века, когда И. Бернулли сформулировал знаменитую задачу о брахистохроне. На современном языке классическая задача вариационного исчисления представляет собой бесконечномерную задачу безусловной оптимизации с минимизируемым функционалом специального (интегрального) вида. Условия экстремума I порядка в вариационном исчислении были получены Эйлером (уравнение Эйлера), а II порядка — Лежандром и Якоби. Важный вопрос о существовании решения в вариационном исчислении был впервые поставлен Вейерштрассом во второй половине XIX века.

Обе задачи, о которых говорилось выше (конечномерная и бесконечномерная), являются примерами задач безусловной минимизации. Задачи на условный экстремум рассматривались в

классической математике лишь для *ограничений типа равенств.* Знаменитое правило множителей Лагранжа (сформулированное в XVIII веке) представляет собой необходимое условие экстремума I порядка в подобных задачах (и в конечномерных, и в задачах вариационного исчисления). Такие же условия для задач с ограничениями типа неравенств были получены лишь недавно. Сами по себе системы неравенств (вне связи с задачами минимизации) изучали Фурье, Минковский, Вейль и другие ученые; созданный ими аппарат позволял без труда получить условия экстремума в задачах с ограничениями — неравенствами.

Первые работы по *экстремальным задачам при наличии ограничений общего вида* относятся к концу 30-х — началу 40-х годов XX века. Корни этих работ были различны. Специалистов по вариационному исчислению, принадлежавших к Чикагской школе (Блисс, Больца, Макшейн, Грейвс, Хестенс и др.), стимулировал интерес возможно более широкой постановки вариационных задач. Здесь в 1937 г. появилась работа Валентайна, посвященная условиям экстремума для задач вариационного исчисления при наличии разного рода ограничений типа неравенств. Позже были созданы (Макшейн, Кокс) общие схемы анализа абстрактных экстремальных задач. Одному из аспирантов Чикагского университета, Карушу, было поручено исследовать в качестве упражнения конечномерные задачи минимизации с общими ограничениями. Каруш получил в 1939 г. условия экстремума первого и второго порядков для гладкого случая; к его работе не отнеслись серьезно, и она не была опубликована. К тем же по существу условиям экстремума несколько позже пришел американский математик Фриц Джон, занимавшийся экстремальными проблемами в геометрии (типа отыскания эллипсоида наименьшего объема, описанного вокруг заданного выпуклого тела). Работа Джона была отвергнута одним серьезным математическим журналом и была напечатана лишь в 1949 г.

Независимо от американскрх исследований оптимизационная тематика развивалась и в СССР. Пионером в этой области был Л. В. Канторович, опубликовавший в 1939 г. книгу, содержавшую математические постановки ряда экономических задач. Последние не укладывались в рамки стандартного математического аппарата, а именно, являлись задачами минимизации линейной функции на множестве, задаваемом линейными ограничениями типа равенств и неравенств. Л. В. Канторович разработал теорию подобных задач и предложил некоторые (не полностью алгоритмизованные) методы их решения. В 1940 г. появилась заметка того же автора, содержавшая общую формулировку условий экстремума при наличии ограничений в

бесконечномерном пространстве. Работы Л. В. Канторовича в то прсмя не привлекли внимания математиков и остались, по существу, незамеченными. Как видно, судьба не благоприятствовала первым исследованиям по неклассическим задачам оптимизации.

Время для них созрело несколько позже, в конце 40-х годов XX века. Под влиянием прикладной тематики, которой ему приходилось заниматься в годы войны, Данциг в США стал изучать задачи минимизации линейной функции при линейных ограничениях, получившие название *задач линейного программирования.* Он сформулировал условия оптимальности решений в линейном программировании. Под влиянием работ фон Неймана по теории игр, Данциг, Гейл, Кун и Таккер создали теорию двойственности в линейном программировании — специфическую формулировку условий экстремума.

Вскоре после разработки теории линейного программирования возникает ее естественное обобщение на нелинейный случай. Задача минимизации нелинейной функции при нелинейных ограничениях была названа *задачей математического программирования* (что вряд ли можно признать удачным, учитывая перегруженность обоих терминов). Если и минимизируемая функция, и ограничения выпуклы, то говорят о *задаче выпуклого программирования.* Условия экстремума для задач математического программирования стали широко известны после работы Куна и Таккера 1950 г.; по существу, это были те же условия Каруша — Джона. Для выпуклого случая Кун и Таккер сформулировали условия экстремума в терминах седловой точки; эта формулировка пригодна и для негладких задач.

Существенный прогресс в теории оптимизации был достигнут и при изучении так называемых *задач оптимального управления,* являющихся непосредственным обобщением классической задачи вариационного исчисления и заключающихся в оптимизации функционалов от решений обыкновенных дифференциальных уравнений, правые части которых включают подлежащие выбору функции («управления»). Необходимые условия оптимальности для этих задач были сформулированы и доказаны Л. С. Понтрягиным, В. Г. Болтянским и Р. В. Гамкрелидзе в 1956—1958 гг. в форме так называемого *принципа максимума.* В иной форме условия оптимальности для подобных задач были получены Беллманом на основе идей *динамического программирования.* Эти результаты были столь связаны со специфической формой задач оптимального управления, что не сразу было осознано их родство с условиями экстремума для задач математического программирования.

В 60-е годы XX века появился цикл работ (А. Я. Дубовицкого и А. А. Милютина, Б. Н. Пшеничного, Нейштадта, Халкина, Варги и других авторов), в которых были предложены *общие схемы получения условий экстремума* для абстрактных задач оптимизации с ограничениями, позволившие охватить как теорему Куна — Таккера, так и принцип максимума. Это дало возможность по-новому взглянуть на известные результаты и, в частности, выделить в них стандартную часть, которую можно получить с помощью общих схем, и нестандартную, связанную со спецификой задачи. Удобным аппаратом для исследования экстремальных задач оказался *выпуклый анализ* — сравнительно новый раздел математики, получивший завершенную форму в работах Р. Рокафеллара. В настоящее время техника вывода условий оптимальности развита в совершенстве.

Выше в основном говорилось о той части теории оптимизации, которая связана с условиями экстремума. Однако найти с помощью условий экстремума явное решение задачи удается лишь в редких случаях. Сложность или невозможность отыскания аналитического решения обнаружилась и в других разделах математики; постепенно стало ясно, что любая задача может считаться решенной, если указан алгоритм, позволяющий численно построить приближенное решение с требуемой точностью. Этот принципиально новый подход, подкрепленный появлением ЭВМ и приведший к возникновению вычислительной математики, существенно затронул и проблематику оптимизации. Одним из центральных направлений здесь стала разработка и обоснование *численных методов* решения.

Математиков прошлого относительно мало интересовали вычислительные проблемы, и хотя некоторые методы решения нелинейных уравнений и безусловной минимизации связывают с именами Ньютона, Гаусса, Коши, эти результаты оставались изолированными в творчестве упомянутых ученых и их последователей.

Первыми нужду в численных методах минимизации испытали статистики. В задачах оценки параметров применение метода максимального правдоподобия или метода наименьших квадратов приводило к необходимости отыскания экстремума функции многих переменных (вообще говоря, неквадратичной). Статистикам (Карри, Левенбергу, Крокету, Чернову и другим) принадлежат первые исследования по численным методам безусловной минимизации, выполненные в 40-х— 50-х годах XX века. В связи с проблемами планирования эксперимента и решения уравнений регрессии в работах Бокса, Роббинса и Монро, Кифера и Вольфовица в начале 50-х годов

XX века были предложены методы минимизации функций при наличии случайных помех.

Другим разделом математики, где происходило зарождение методов оптимизации, была *линейная алгебра.* Необходимость решения больших систем линейных уравнений, возникающих при конечно-разностной аппроксимации уравнений с частными производными, привела к развитию итеративных методов линейной алгебры. Но задача решения системы линейных уравнений эквивалентна минимизации квадратичной функции, и многие итератнвные методы удобно строить и обосновывать, опираясь на этот факт. Таковы методы покоординатного спуска, наискорейшего спуска, сопряженных градиентов и ряд других методов, мнорые были созданы в линейной алгебре к началу 50-х годов XX века. Естественным шагом было перенесение подобных методов на неквадратичный случай.

С необходимостью решения задач оптимизации столкнулись и специалисты по *теории автоматического регулирования.* Трудами В. В. Казакевича, А. А. Фельдбаума, А. А. Первозванского в 50-х годах XX века была создана *теория экстремального регулирования* и предложены специальные методы оптимизации действующих объектов в реальном масштабе времени.

Первый численный метод нелинейного программирования — *метод штрафных функций* — был введен Курантом в 1943 г. из соображений, связанных с физической природой рассматривавшейся задачи.

Наконец, мощный импульс для развития методов оптимизации дал предложенный Данцигом в конце 40-х годов XX века *симплекс-метод* для решения задач линейного программирования. Обилие приложений и наличие хороших программ и программных систем ЭВМ привели к широкой популярности симплекс-метода прежде всего среди экономистов.

До какого-то времени такого рода исследования были спорадическими и не объединялись ни единым подходом, ни аппаратом. Однако к середине 60-х годов XX века в рамках вычислительной математики сложилось самостоятельное направление, связанное с *численными методами оптимизации.* С тех пор непрерывно шло интенсивное развитие этого направления как вширь (разработка новых методов, исследование новых классов задач), так и вглубь (выработка единого аппарата для анализа сходимости и скорости сходимости, классификация и унификация методов). В настоящее время эта область вычислительной математики может считаться окончательно сформировавшейся. Разработано множество численных методов для всех основных классов задач оптимизации — безусловной минимизации гладких и негладких функций в конечномерных и

бесконечномерных пространствах, условной минимизации при ограничениях типа равенств и (или) неравенств в выпуклом или невыпуклом случае и т. д. Для большинства методов имеется строгое обоснование, выяснена скорость сходимости, установлена область применимости. Конечно, многие проблемы еще не решены до конца (построение эффективных методов для некоторых специальных типов задач, проблема оптимальных методов, подробная численная проверка имеющихся алгоритмов, создание доступных и отработанных машинных программ и т. п.). Однако, по-видимому, период наибольшей активности в области численных методов оптимизации остался позади.

В предлагаемой вниманию работе делается попытка систематического изложения современного состояния основ оптимизации.

1. Введение в теорию функций

1.1. Функция как объект оптимизации

Определение функции. Возьмем некоторое множество значений величины x, т.е. некоторое множество точек на числовой оси Ox, и обозначим его через D. Если каждому значению x из множества D по какому-нибудь правилу поставлено в соответствие одно определенное значение другой величины y, то говорят, что эта ***величина*** y есть ***функция*** величины x или что величины x и y связаны между собой *функциональной зависимостью*. При этом величина x называется *аргументом* функции y, а множество D — *областью определения* функции y. Значения аргумента x из области D определения функции y мы можем выбирать по нашему усмотрению произвольно; поэтому величина x называется *независимой переменной*. Значение же функции y, когда значение независимой переменной x уже назначено, мы выбрать произвольно не можем; это значение будет строго определенным, именно тем, которое соответствует выбранному значению независимой переменной. Значения функции зависят от значений, принимаемых независимой переменной, и обычно

изменяются при ее изменении. Поэтому ***функцию*** называют еще ***зависимой переменной***.

Определение. ***Величина y называется функцией переменной величины x в области определения D, если каждому значению x из этой области соответствует одно определенное значение величины y.***

Областью определения функции может быть любое множество точек числовой оси, но чаще всего в математическом анализе и в его применениях рассматривают лишь функции, областями определения которых служат области таких двух типов:

1) множество целых неотрицательных точек числовой оси, т. е. точек $x=0$, $x=1$, $x=2$, $x=3$, ... (или некоторая часть этого множества);

2) один или несколько интервалов (конечных или бесконечных) числовой оси.

Говорят, что в первом случае мы имеем *функцию целочисленного аргумента,* а во втором случае — *функцию непрерывного аргумента.* В первом случае аргумент может пробегать ряд чисел: $x= 0, 1, 2, 3, ...$; во втором случае аргумент пробегает один или несколько интервалов числовой оси.

Множество значений, принимаемых функцией *y,* называется *областью значений функции.*

Впрочем, слово «функция» употребляется и в ином смысле, а именно:

Закон (правило)*, по которому значениям независимых переменных отвечают (соответствуют) значения рассматриваемой зависимой переменной, называется* ***функцией.***

Таким образом, каждый раз, когда нам дан такой закон соответствия, мы можем сказать: вот функция.

Функции могут быть от одного аргумента (как в примере площади круга) или от двух и более аргументов.

Заметим, что для того, чтобы некоторая величина *y* могла рассматриваться как функция от независимой переменной *x,* нет надобности, чтобы между изменениями этих величин существовала глубокая причинная связь. Достаточно только, чтобы существовал определенный ***закон***, по которому значениям *x* отвечали бы значения *y,* этот закон может быть нам и неизвестен. Например, температуру θ в какой-либо точке можно считать функцией времени *t,* так как ясно, что значениям *t* отвечают определенные значения θ, хотя, конечно, изменение θ объясняется не просто течением времени, но рядом глубоких физических причин.

Изучение общих уравнений (дифференциальных или интегральных) связано прежде всего с уточнением и расширением понятия функции,

Термин «функция» был введен великим немецким математиком Г. Лейбницем (1646—1716), однако понятие функциональной зависимости как зависимости между переменными величинами можно обнаружить в более ранних работах французского математика Р. Декарта (1596—1650).

С течением времени смысл, вкладываемый в понятие функции, стал меняться. Во многом это объясняется успехами дифференциального и интегрального исчисления, позволившими решить задачи, казавшиеся неприступными. Поиски новых объектов для дифференцирования и интегрирования привели к тому, что на передний план стали выступать функции, задаваемые с помощью аналитических выражений, например

$$f(x) = e^{-x} (\sin x + \cos x).$$

В результате к началу XVIII века преобладающим становится взгляд на ***функцию*** как на ***аналитическое выражение***. Приведем характерное определение функции, данное выдающимся математиком, петербургским академиком Л. Эйлером (1707—1783) в классическом труде «Введение в анализ бесконечно малых» (М., Физматгиз, 1961, т. 1, с. 5): «Функция переменного количества есть аналитическое выражение, составленное каким-либо образом из этого переменного количества и чисел или постоянных количеств».

Согласно такому пониманию кривая, начерченная «свободным движением руки», не является графиком функции.

К концу XVIII века этот взгляд на функцию превращается в господствующий, и, как это часто бывает, в это же время становится ясным, что отождествление функции с аналитическим выражением сужает круг приложений математического анализа.

Первым тревожным сигналом было то, что одна и та же функция может задаваться различными аналитическими выражениями. Простые примеры когда одно аналитическое выражение сводится к другому тождественными алгебраическими преобразованиями, как, например,

$$f(x)=x^2+2x+1 \text{ и } f(x) = (x+1)^2,$$

были известны давно. Однако совершенно другую окраску эта проблема приобрела в связи с рассмотрением бесконечных сумм, например

$$f(x)=1+x+x^2+x^3+ \ldots$$

Как известно, при $|x|<1$ выполняется равенство

$$1/(1-x)=1+x+x^2+x^3+ \ldots,$$

представляющее собой формулу суммы членов бесконечно убывающей геометрической прогрессии.

В 1797 г. вышла классическая работа французского математика Ж. Л. Лагранжа (1736—1813) «Теория аналитических функций...», в

которой изучались функции, представимые степенными рядами, т. е. бесконечными суммами вида $f(x)=c_0+c_1x+c_2x^2+c_3x^3+\ldots$

Впоследствии такие функции были названы аналитическими. Этот класс функций наилучшим образом отвечал задачам дифференциального исчисления, венцом которого в то время была формула, полученная английским математиком Б. Тейлором (1685—1731). Одно из важных следствий из этой формулы таково: ***зная аналитическую функцию на сколь угодно малом интервале, можно найти ее значение в каждой точке области определения.*** Это следствие как нельзя больше играло на руку противникам функций, задаваемых «свободным движением руки». Действительно, задав аналитическую функцию на каком-либо интервале, мы уже не вправе распоряжаться ее значениями вне этого интервала — аналитически продолжить функцию можно единственным образом.

Ситуация меняется кардинальным образом, если рассматривать бесконечные суммы, отличные от степенных рядов.

В 1807 г. французский математик Ж. Фурье (1768— 1830) в цикле работ по аналитической теории теплоты впервые показал, что каждая функция, график которой может быть начерчен «свободным движением руки» (в нынешнем понимании — каждая непрерывная кусочно-гладкая функция), представляется на отрезке $[0, \pi]$ тригонометрическим рядом

$$f(x)=a_0+a_1\cos x+a_2\cos 2x+a_3\cos 3x+\ldots$$

Справедливости ради следует сказать, что еще за полвека до работ Фурье о возможности такого представления говорил швейцарский математик Д. Бернулли (1700— 1782), однако, не подкрепленная доказательством, эта точка зрения признания не получила.

Первые работы Фурье также встретили возражение, причем такого выдающегося математика, как Лагранж, однако очень скоро в правоте Фурье уже никто не сомневался. Одним из наиболее значительных открытий Фурье было то, что ***кривые, начерченные «свободным движением руки», оказались представимы аналитически***. Вновь в основу понятия функции было положено понятие зависимости между переменными величинами. Уже через три года после работ Фурье под сильным их влиянием автор известного учебника математического анализа С. Л а к р у а (1765—1843) писал: «Всякое количество, значение которого зависит от одного или нескольких других количеств, называется функцией этих последних независимо от того, знаем мы или не знаем, через какие операции нужно пройти, чтобы перейти от этих последних к первой».

К такой же трактовке понятия функции пришли в результате изучения тригонометрических рядов великий русский математик Н. И. Лобачевский (1792 — 1856) и немецкий математик П. Дирихле (1805—1859). Классическим примером функции в новом понимании является функция Дирихле:

$$D(x)=\begin{cases}1, & \text{если } x\text{—рациональное число,}\\ 0, & \text{если } x \text{ — иррациональное число.}\end{cases}$$

Безусловно, математики XVIII столетия не могли признать ее функцией, да и график ее меньше всего походит на какую-нибудь кривую. Впрочем, позже и для функции Дирихле было найдено аналитическое представление через элементарные функции с помощью дважды примененной операции предельного перехода:

$$D(x)=\lim_{n\to\infty}\lim_{m\to\infty}[\cos(2\pi x\cdot n!)]^{2m}.$$

Следующий шаг в сторону расширения понятия функции должен был заключаться в отказе от условия, что ***аргументы функций*** принимают ***числовые значения***. Однако, как это часто бывало в истории математики, в действительности события развивались в другой последовательности. Классический пример такого, казалось бы, нелогичного развития можно найти в теории чисел. Известно, что **классы чисел расширяются в такой последовательности: натуральные числа, целые числа, рациональные числа, действительные числа, комплексные числа. *Исторически* же классы чисел появлялись далеко *не в такой последовательности*.**

Подобное замечание можно отнести и к развитию понятия функции. Именно понятие функции, зависящей не от числовой переменной, появилось еще в те времена, когда функцию считали аналитическим выражением. Связано это было с возникновением вариационного исчисления.

В 1696г. появилась заметка И. Бернулли (1667— 1748), в заголовке которой содержался явный вызов: *«Probtema novutn, ad cujus solutionern mathematici invitaniur-»* («Новая задача, к решению которой приглашаются математики»). Эта новая задача формулировалась так:

В вертикальной плоскости даны две точки: *А* и *В*. Найти траекторию материальной точки *М,* которая, начав движение из точки *А,* дойдет под действием силы тяжести до точки *В* за наименьшее время.

Считается, что эта задача явилась толчком к созданию ***вариационного исчисления***. Кривая, вдоль которой движется точка, была названа ***брахистохроной*** (от греч. βραχιστος — кратчайший и χρονος— время).

Для решения задачи о брахистохроне впервые потребовалось рассмотреть зависимость между двумя переменными величинами, одна из которых принимает ***не числовые значения, а функциональные***. В самом деле, обозначим координаты точек A и B в задаче о брахистохроне соответственно (a, α) и (b, β). Предположим, что точка M движется вдоль графика некоторой функции f, заданной на отрезке $[a, b]$. . Очевидно, функция f должна удовлетворять условиям

$$f(a)=\alpha,\ f(b)=\beta. \tag{1.1}$$

Время T, требуемое точке M для прохождения кривой, зависит только от выбора функции f. Символически это записывают так:

$$T=T(f).$$

В этой записи f — независимая переменная, значениями которой являются функции, а T — зависимая переменная, принимающая числовые значения. В современной терминологии T-отображение, сопоставляющее каждой функции f из некоторого класса функций, определенных на $[a, b]$, действительное число $T(f)$. Отображение T можно было бы также назвать функцией, однако во избежание смешения понятий вместо этого термина употребляется специальный термин «функционал».

Исходя из закона сохранения энергии, нетрудно вычислить значение T по заданной функции f:

$$T(f)=\frac{1}{\sqrt{2g}}\int_a^b \frac{\sqrt{1+[f'(x)]^2}}{\sqrt{\alpha-f(x)}}\,dx, \tag{1.2}$$

где g — ускорение свободного падения, а функция f удовлетворяет условиям (1.1).

Функционал T был назван Л. Эйлером «неопределенной интегральной величиной» в том смысле, что T зависит от неизвестной функции f, входящей под знак интеграла.

Формально первым функционалом подобного типа можно было бы считать определенный интеграл от функции f:

$$I(f)=\int_a^b f(x)\,dx.$$

Этим равенством определено отображение I, которое каждой функции f сопоставляет действительное число $I(f)$.

Вместе с развитием понятия функции развивалось и понятие функционала, и к концу XIX века стало ясным, что функции и функционалы — объекты одной природы, частные случаи общего понятия ***отображения***. Для того чтобы прийти к такому выводу, оставалось сделать последний шаг: допустить, что в соотношении

$T=T(f)$ символы f и $T(f)$ могут обозначать произвольные математические объекты.

Этот шаг был подготовлен бурным развитием во второй половине XIX века теории интегральных уравнений, т. е. уравнений, в которых неизвестная функция содержится под знаком интеграла. Классический пример — интегральное уравнение Вольтерра

$$f(t) - \int_a^t K(t, s) f(s)\, ds = g(t), \qquad (1.3)$$

названное в честь итальянского математика В. Вольтерра (1860—1940). В этом уравнении $K(t, s)$ — известная функция двух переменных, называемая ядром интегрального уравнения, $g(f)$ — заданная функция, называемая правой частью, $f(t)$ — неизвестная функция, которую требуется найти.

Если ввести обозначение

$$Af(t) = \int_a^t K(t, s) f(s)\, ds, \qquad (1.4)$$

то символически уравнение Вольтерра (1.3) можно записать так:

$$f-Af=g. \qquad (1.5)$$

Проведем следующие. Перепишем последнее уравнение следующим образом:

$$(1-A)f=g.$$

Поделим обе части на $(1—A)$, т. е.

$$f = \frac{1}{1-A} \cdot g,$$

и воспользуемся формулой для суммы членов бесконечно убывающей геометрической прогрессии:

$$f=(1+A + A^2+A^3+ ...)g.$$

Итак, решение уравнения (1.5) мы получили в виде бесконечной суммы

$$f=g+Ag+A^2g+A^3g+ \ldots, \qquad (1.6)$$

в которой A^ng определяется индуктивно:

$$A^1g=Ag,\ A^ng=A\ (A^{n-1}\ g),\ n=2, 3, \ldots$$

Интересно то, что подобное манипулирование математическими символами приводит к точному результату: ряд (1.6) действительно дает решение уравнения (1.5). Этот ряд носит название ряда Неймана в честь немецкого математика К- Неймана (1832— 1925).

Другой пример — интегральное уравнение Фредгольма

$$f(t) - \int_a^b K(t, s) f(s)\, ds = g(t),$$

детально изученное шведским математиком Э. Фредгольмом (1866—1927). Замечательные результаты Фредгольма во многом объясняются удачным выбором интегрального уравнения в том виде, который наилучшим образом отвечает задачам математической физики.

Тот факт, что в формулах (1.5) и (1.6) символ A обозначает интегральный оператор (1.4), не имеет существенного значения. Те же рассуждения сохраняются и в случае, когда A — произвольное отображение, сопоставляющее каждой функции f определенную функцию Af.

Более того, в формулах (1.5) и (1.6) ***символы f и g могут обозначать и не функции, а произвольные математические объекты, для которых определена операция сложения***. Этот последний шаг и приводит нас к общему понятию ***отображения.***

Как известно, ***произвольная совокупность каких-либо объектов носит в математике название <u>множества</u>***. Примерами множеств могут служить множество натуральных чисел; множество непрерывных функций, определенных на отрезке [0, 1]; множество бесконечно малых последовательностей и т. д. Объекты, из которых состоит множество, называются, его элементами. Множества обозначают большими буквами латинского алфавита: *A, B, C, ...,* а элементы — малыми буквами: *a, b, c, ...* То, что объект x принадлежит множеству A, записывается так: $x \in A$.

Каждое множество определяется своими элементами. Это означает, что множество A считается заданным, если о каждом объекте можно сказать, принадлежит он множеству A или нет.

Например, если символом R обозначить множество всех действительных чисел, то запись $x \in R$ означает, что элемент x принадлежит множеству действительных чисел, или короче, x — действительное число.

Другой пример — множество всех функций, непрерывных на отрезке $[a, b]$. Это множество обозначается символом $C[a, b]$. Следовательно, запись $f \in C[a, b]$ означает, что f — функция, определенная на отрезке $[a, b]$ и непрерывная в каждой точке этого отрезка.

Предположим, что даны два множества P и Q и правило A, по которому каждому элементу $x \in P$ сопоставляется определенный элемент $y \in Q$. В этом случае говорят, что задано отображение A множества P в множество Q. То, что элемент x переводится

отображением A в элемент y, записывают так: $y=A(x)$, или короче $y=Ax$.

Понятие произвольного множества, введенное в математику Г. Кантором (1845—1918), является слишком общим для того, чтобы им можно было воспользоваться при решении уравнений. Необходимо предположить, что рассматриваемые множества обладают некоторыми дополнительными свойствами.

Какими же именно свойствами должны обладать эти множества? Для того чтобы выяснить это, обратимся к задаче о приближенном нахождении корней уравнений. Представляется вполне очевидным, что любое уравнение (алгебраическое, дифференциальное, интегральное и т, п.) можно привести к следующему специальному виду:

$$x=Ax. \qquad (1.7)$$

В этом уравнении x — неизвестный элемент, принадлежащий некоторому множеству M, A — заданное отображение множества M в себя. Это означает, что каждому элементу $x \in M$ отображение A сопоставляет элемент Ax, также принадлежащий M.

Будем искать решение уравнения (1.7) методом последовательных приближений. Выберем произвольно начальное приближение x_0 и определим последовательные приближения с помощью рекуррентной формулы

$$x_n=Ax_{n-1} \qquad n=1,\ 2,\ 3,... \qquad (1.8)$$

Задача заключается в отыскании условий, при которых последовательные приближения сходятся к точному решению уравнения (1.7). При этом основным вопросом для нас является следующий: как понимать сходимость x_n *к* x. Очевидно, сходимость должна означать, что элементы x_n приближаются к x на сколь угодно малое расстояние. Понятие **расстояния** и является **ключевым** в рассматриваемой теории.

Итак, решение уравнения (1.7) следует искать среди элементов такого множества, в котором введено понятие расстояния, или, как говорят, введена **метрика**. Эти множества называют ***метрическими пространствами.***

Заслуга введения в математику метрических пространств принадлежит французскому математику М. Фреше (1878—1973). Сейчас трудно переоценить роль метрических пространств в развитии современной математики, особенно того ее раздела, который называют функциональным анализом. Мы не будем касаться глубоких результатов в теории метрических пространств и ограничимся лишь знакомством с первоначальными понятиями.

1.2. Метрические пространства

Рассмотрим какое-нибудь множество M и последовательность элементов x_n, принадлежащих M. Предположим, что с увеличением номера n элементы x_n приближаются к некоторому элементу $x \in M$. Нам предстоит уточнить смысл выражения «x_n стремятся к x» и прежде всего указать тот класс множеств, в котором можно ввести понятие предела.

Впервые абстрактное определение предела было дано Фреше с помощью понятия *расстояния.*

Что такое расстояние? Из начального курса геометрии известно, что ***расстояние есть величина, имеющая размерность длины, определенным образом сопоставленная любой паре точек***. Если выбран масштаб, то расстояние можно считать безразмерной величиной, т. е. числом. Например, расстояние от точки A до точки B на рис. 1.1 равно 5 см, если же договориться, что все расстояния измеряются в сантиметрах, то можно сказать, что расстояние от A до B равно 5. Расстояние между элементами произвольного множества мы также будем считать безразмерной величиной, т. е. числом.

Для большей наглядности элементы множеств будем называть также точками, принадлежащими этим множествам.

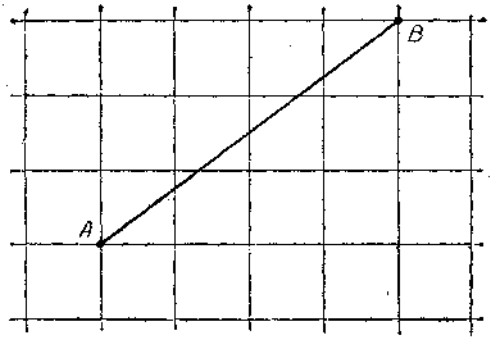


Рис. 1.1

Итак, пусть M — произвольное множество и пусть любым двум точкам $x \in M$ и $y \in M$ сопоставлено число d (так как d зависит от x и y, то будем писать $d(x, y)$). Можно ли это число $d(x, y)$ назвать расстоянием от x до у? Безусловно, мы можем давать какие угодно определения, однако каждый раз при этом желательно, чтобы новые понятия соответствовали нашим обычным представлениям.

Какие же требования следует предъявить к расстоянию? Во-первых, нужно, чтобы расстояние не было отрицательным числом, т. е. чтобы выполнялось неравенство

$$d(x, y)>0.$$

Более того, расстояние между различными точками должно быть строго положительно:

$$d(x, y)>0, \quad \text{если} \quad x\neq y.$$

Обычно мы не говорим о расстоянии от какой-нибудь точки до этой же самой точки, однако в данном случае удобно считать, что это расстояние равно нулю:

$$d(x, x)=0.$$

Второе требование также основывается на привычных представлениях: расстояние от точки x до точки y должно равняться расстоянию от y до x:

$$d(x, y)=d(y, x).$$

Это равенство отражает свойство симметрии расстояния. Такая симметрия на практике может наблюдаться не всегда. Возьмем, к примеру, несколько населенных пунктов, расположенных вдоль одной реки. «Расстоянием» от пункта x до пункта y назовем время (в часах), затрачиваемое теплоходом при движении от x до y. Тогда ясно, что расстояние от x до y не равно расстоянию от y до x (различие становится тем заметнее, чем выше скорость течения реки и ниже собственная скорость теплохода). Принимая аксиому симметрии (второе требование), мы тем самым отказываем введенному выше «расстоянию» в праве называться расстоянием.

Несколько сложнее математически формулируется третье требование: расстояние должно измеряться вдоль наикратчайшего пути. Предположим, что, двигаясь каким-то образом от точки x к точке y, мы прошли через некоторую точку z. При таком движении пройденное расстояние не меньше, чем сумма $d(x, z)+d(z, y)$. Поэтому, если расстояние $d(x, y)$ измеряется вдоль наикратчайшего пути, то для любой точки $z \in M$ должно выполняться неравенство

$$d(x, y)\leqslant d(x, z)+d(z, y).$$

Это неравенство, введенное Фреше в 1906 г., было названо им неравенством треугольника. Для объяснения этого названия отметим на плоскости три точки x, y, z и соединим их отрезками. Длины сторон получившегося треугольника равны $d(x, y)$, $d(y, z)$, $d(z, x)$. Поэтому неравенство треугольника отражает известное из начального курса геометрии свойство: ***длина любой стороны треугольника меньше суммы длин двух других его сторон***.

Дадим теперь определение ***метрического пространства***. **Множество M называют метрическим пространством, если для**

любой пары элементов $x \in M$ и $y \in M$ определено расстояние $d(x, y)$, обладающее следующими свойствами:

1° $d(x, y) \geq 0$, если $x \neq y$, и $d(x, y) = 0$, если $x = y$;
2° $d(x, y) = d(y, x)$;

3° для любых трех элементов x, y, z из M справедливо неравенство треугольника

$$d(x, y) \leqslant d(x, z) + d(z, y).$$

Приведем примеры метрических пространств.

В качестве метрического пространства возьмем множество клеток на шахматной доске. Расстоянием от клетки x до клетки y назовем наименьшее число ходов, требуемых королю, чтобы он перешел из клетки x в клетку y. Например, в позиции на рис. 1.2 расстояние между клетками, занимаемыми белым королем и черной пешкой, равно 3.

Этот пример приведен только для того, чтобы показать, что метрическое пространство может содержать конечное число точек (в данном случае 64), и расстояние может принимать только целочисленные значения (в данном случае расстояние может быть любым целым числом от 0 до 8).

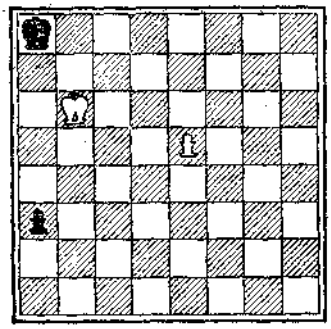


Рис. 1.2

Обратимся теперь к наиболее часто встречающемуся метрическому пространству — к числовой прямой R. Элементами R являются действительные числа. Назовем расстоянием то точки $x \in R$ до точки $y \in R$ абсолютную величину разности x и y, т. е. число

$$d(x, y) = |x - y|.$$

Возникает вопрос, является ли это число расстоянием, т. е. выполняются ли условия 1°—3°? Так как первые два условия сомнения не вызывают, то проверить нужно только третье.

Вспомним, что для любых действительных чисел a и b выполняется неравенство

$$|a+b|\leqslant|a|+|b|.$$

Положим в этом неравенстве $a=x-z$, $b=z-y$; тогда $a+b=x-y$ и неравенство принимает вид

$$|x-y|\leqslant|x-z|+|z-y|.$$

Это и есть требуемое неравенство треугольника.

Рассмотрим последний пример метрического пространства, наиболее важный для наших дальнейших исследований. Это пространство $C[a, b]$ всех числовых функций, непрерывных на отрезке $[a, b]$.

Впервые строгое математическое построение теории непрерывных функций было проведено немецким математиком К. Вейерштрассом. Ему принадлежат две наиболее известные теоремы о непрерывных функциях, которые так и называют первой и второй теоремой Вейерштрасса. Обе эти теоремы можно объединить и кратко сформулировать так: **каждая функция, непрерывная на отрезке, принимает в некоторой точке этого отрезка наибольшее значение.** Иначе говоря, если функция $f(t)$ непрерывна на отрезке $[a, b]$, то существует точка $t_0 \in [a, b]$ такая, что

$$f(t)\leqslant f(t_0)$$

для всех $t \in [a, b]$.

Значение $f(t_0)$ называют ***максимумом*** функции f на отрезке $[a, b]$ и обозначают так:

$$\max_{t\in[a,\ b]} f(t)=f(t_0).$$

Например,

$$\max_{t\in[0,\ \pi]} \sin t=\sin(\pi/2)=1.$$

Опираясь на свойства максимума, введем расстояние в пространстве $C[a, b]$. Функции, принадлежащие $C[a, b]$, т. е. непрерывные на отрезке $[a, b]$, будем обозначать не только символами $f(t)$, $g(t)$, но и символами $x(t)$, $y(t)$, $z(t)$.

Итак, пусть $x \in C[a, b]$ и $y \in C[a, b]$, т. е. функции $x(t)$ и $y(t)$ непрерывны на $[a, b]$. Тогда разность $x(t)-y(t)$ также непрерывна, и, следовательно, непрерывна функция

$$|x(t)-y(t)|.$$

Последнее вытекает из того, что абсолютная величина непрерывной функции есть непрерывная функция. По теоремам

Вейерштрасса существует максимум этой абсолютной величины, который мы и назовем расстоянием от x до y. Таким образом,

$$d(x, y) = \max_{t \in [a, b]} |x(t) - y(t)|. \qquad (1.9)$$

Графически расстояние в $C[a, b]$ иллюстрируется на рис. 1.3.

Для того чтобы установить корректность этого определения, необходимо проверить выполнение условий 1°—3°.

То, что условия 1° и 2° выполняются, проверяется очень легко, а свойство 3°следует из неравенства

$$|x(t) - y(t)| \leqslant |x(t) - z(t)| + |z(t) - y(t)|.$$

Таким образом, формула (1.9) действительно определяет расстояние в пространстве $C[a, b]$, которое тем самым становится метрическим.

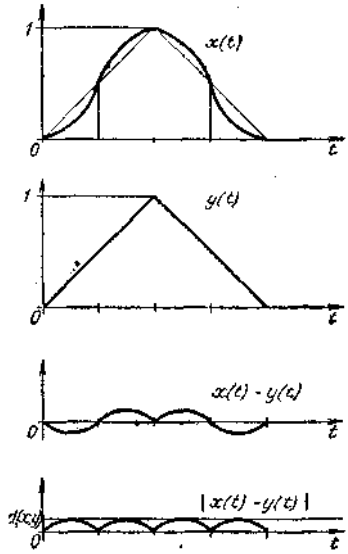


Рис. 1.3

Вернемся к изучению произвольных метрических пространств.

Пусть M — метрическое пространство; рассмотрим последовательность элементов $x_n \in M$. Понятие расстояния позволяет определить предел последовательности x_n.

Элемент $\bar{x} \in M$ называется *пределом* последовательности x_n, если расстояние от x_n до $\bar{x}$ стремится к нулю:

$$d(x_n, \bar{x}) \to 0 \text{ при } n \to \infty.$$

При этом говорят, что последовательность x_n *сходится* к $\bar{x}$ и пишут $x_n \to \bar{x}$ при $n \to \infty$. Другими словами, $x_n \to \bar{x}$, если числовая последовательность $d(x_n, \bar{x})$ является бесконечно малой. Иногда, подчеркивая, что сходимость понимается в смысле расстояния в метрическом пространстве, говорят, что последовательность сходится в *метрике M*.

Например, если в качестве метрического пространства взять числовую прямую *R*, то сходимость в метрике *R* совпадает с обычной сходимостью числовой последовательности. Действительно, в этом случае

$$d(x_n, \bar{x}) = |x_n - \bar{x}|,$$

и если $d(x_n, \bar{x}) \to 0$, то $|x_n - \bar{x}| \to 0$.

Посмотрим, что означает сходимость в метрике пространства $C[a, b]$. Пусть $x_n \in C[a, b]$, $\bar{x} \in C[a, b]$ и $d(x_n, \bar{x}) \to 0$. Тогда для всех $t \in C[a, b]$

$$|x_n(t) - \bar{x}(t)| \leqslant d(x_n, \bar{x}),$$

и, следовательно, разность функций $-x_n(t)$ и предельной функции $\bar{x}(t)$ может быть сделана по абсолютной величине меньше любого положительного числа *сразу для всех точек t из отрезка* $[a, b]$. Такая сходимость последовательности функций была введена Вейерштрассом и названа равномерной сходимостью.

Итак, сходимость в пространстве $C[a, b]$ есть равномерная сходимость последовательности непрерывных функций на отрезке $[a, b]$.

Рассмотрим теперь задачу о приближенном решении уравнений в метрических пространствах.

Пусть даны два метрических пространства *M* и *L*. Предположим, что задано также отображение *T* метрического пространства *M* в *L*. Это значит, что указано правило, по которому каждому элементу $x \in M$ сопоставляется определенный элемент *Tx* из *L*. Фиксируем произвольный элемент $h \in L$ и рассмотрим уравнение

$$Tx = h. \qquad (1.10)$$

В этом уравнении *T* — заданное отображение, называемое также оператором, *h* — известная правая часть, *x* — неизвестный элемент, который требуется найти.

Для того чтобы показать, насколько широк класс уравнений вида (1.10), рассмотрим следующий пример. Пусть метрические

пространства M в L совпадают с пространством функций, непрерывных на отрезке [0, 1], т. е. $M=L=C[0, 1]$. Определим отображение T равенством

$$Tx(t) = x(t) + \lambda \int_0^t (t-s) \sin x(s)\, ds, \tag{1.11}$$

где λ — действительное число (параметр).

Интегральный оператор T каждой непрерывной функции $x(t)$ ставит в соответствие непрерывную функцию $Tx(t)_x$ т. е. действует из $C[0, 1]$в $C[0, 1]$. Уравнение (1.10) преобразуется к виду

$$x(t) + \lambda \int_0^t (t-s) \sin x(s)\, ds = h(t), \tag{1.12}$$

где $h(t)$ — заданная функция, непрерывная на отрезке [0, 1]. Как видим, рассматриваемый нами класс уравнений (1.10) охватывает нелинейные интегральные уравнения (нелинейные, потому что подынтегральное выражение в (1.11) нелинейно зависит от неизвестной функции x).

Обычный путь получения ***уравнений*** — математическая формализация практических задач, т. е. **построение математической модели.** Не составляет исключения и интегральное уравнение (1.12).

С математической точки зрения правильная формализация должна быть такой, чтобы для уравнения (1.10) были справедливы теоремы о существовании и единственности решения при любой допустимой правой части.

Если нет теоремы единственности, то может оказаться, что уравнение (1.10) для некоторой правой части h имеет несколько решений. Как правило, ***это означает, что при составлении математической модели не были учтены какие-то важные факторы, характеризующие данное явление***.

Приведем простейший пример подобного рода. Пусть требуется найти сторону квадратной комнаты площадью 18 м^2. Построение математической модели начинается с того, что реальный объект — пол комнаты заменяется абстрактной фигурой — квадратом. Если обозначить искомую длину стороны буквой x, то для ее определения получаем уравнение

$$x^2=18.$$

Правильно ли это уравнение приближает исходную задачу? Если его рассматривать на множестве всех действительных чисел R, то у него будет два решения; $x_1 = \sqrt{18}$, $x_2 = -\sqrt{18}$, т. е. единственности нет. Для выбора единственного решения вновь обратимся к исходной задаче. Из условия ясно, что ответ должен быть положительным. Следовательно, надо сузить множество, в котором ищется решение, и

вместо всей числовой прямой R взять положительную полупрямую R_+ (т. е. множество всех положительных чисел). Теперь уравнение $x^2=18$ имеет ровно один корень, который и дает требуемое значение длины стороны.

Если нет теоремы существования, т. е. уравнение (1.10) не имеет решения, то обычно это означает, что мы предъявили слишком много требований к решению. Часто бывет достаточным расширить метрическое пространство, чтобы решение появилось.

Например, в той же задаче о нахождении длины квадрата по заданной площади мы могли бы вначале предположить, что искомая длина выражается рациональным числом. Тогда решение уравнения $x^2= 18$ следовало бы искать в множестве положительных рациональных чисел. Положительный корень должен равняться $3\sqrt{2}$, но это число иррациональное. Следовательно, в множестве рациональных чисел рассматриваемое уравнение корней не имеет.

Расширим множество, в котором ищутся корни, и вместо рациональных чисел возьмем положительные действительные числа. Теперь уравнение $x^2= 18$ имеет решение, и притом только одно.

Рассмотренная простейшая ситуация довольно точно характеризует трудности, возникающие в общей постановке задачи. Предположим, что уравнение (1.10), рассматриваемое в метрическом пространстве M, имеет единственное решение x_0. Удалим этот элемент x_0 из метрического пространства M. Оставшееся множество по-прежнему является метрическим пространством, которое мы обозначим символом M_0. Если теперь рассмотреть уравнение (1.10) в пространстве M_0, то оно уже решений иметь не будет. С абстрактной точки зрения пространства M и M_0 не отличаются, однако в первом уравнение имеет решение, а во втором — нет. Мы должны каким-то образом научиться различать эти пространства и исключать из рассмотрения те, которые получаются из «хороших» пространств удалением каких-то элементов.

Для решений этой задачи нам необходимо привлечь фундаментальные последовательности.

Последовательность элементов x_n метрического пространства называется фундаментальной, если

$$d(x_n, x_m)\to 0 \text{ при } n\to\infty \text{ и } m\to\infty.$$

Если в качестве метрического пространства M взята числовая прямая R, то это определение фундаментальной последовательности совпадает с тем, которое было дано ранее. Как и в случае числовой прямой, всякая сходящаяся последовательность элементов из M является фундаментальной. Это следует из неравенства треугольника:

$$d(x_n, x_m)\leqslant d(x_n, \bar{x})+d(\bar{x}, x_m).$$

Если $x_n \to \overline{x}$, то оба слагаемых в правой части стремятся к нулю, поэтому стремится к нулю и левая часть т. е. последовательность x_n фундаментальна.

Как известно, в случае, когда метрическое пространство M есть числовая прямая, верно и обратное: всякая фундаментальная последовательность является сходящейся. Это утверждение составляет основное содержание **критерия Коши**.

Однако в некоторых метрических пространствах обратное утверждение может оказаться и неверным. Рассмотрим, например, множество всех рациональных чисел, которое обычно обозначают буквой Q. Введем в нем естественным образом расстояние от x до y:

$$d(x, y)=|x—y|.$$

Рассмотрим последовательность рациональных приближений x_n числа $\sqrt{2}$. Последовательность x_n сходится к $\sqrt{2}$ в пространстве действительных чисел R и поэтому является фундаментальной в R. Очевидно, она фундаментальная и в пространстве Q, однако в Q она уже не является сходящейся, так как число $\sqrt{2}$ множеству Q не принадлежит.

Вообще, удалим из числовой прямой R какую-нибудь точку a и обозначим полученное пространство M_a. Тогда любая последовательность отличных от a действительных чисел, сходящаяся к a, дает пример фундаментальной, но не сходящейся последовательности в M_a. Такая ситуация оказалась возможной вследствие того, что в метрическом пространстве M_a недостает элемента a, т. е. пространство неполно. Эти наводящие соображения приводят к следующему определению полного метрического пространства.

Метрическое пространство называется полным, если всякая фундаментальная последовательность его элементов имеет предел.

Классический пример полного метрического пространства — числовая прямая. Другой, более сложный пример — пространство $C[a, b]$ функций, непрерывных на отрезке $[a, b]$. Сходимость в метрике $C[a, b]$ совпадает с равномерной сходимостью, поэтому полнота пространства $C[a, b]$ следует из теоремы, доказываемой в курсе математического анализа: предел равномерно сходящейся последовательности непрерывных функций есть непрерывная функция.

Действительно, пусть x_n — фундаментальная последовательность функкций из $C[a, b]$. Тогда для любого фиксированного t из $[a, b]$ числовая последовательность $x_n(t)$ является фундаментальной в R и, согласно критерию Коши, сходится. Ее предел, зависящий от t, обозначим через $\overline{x}(t)$:

$$\bar{x}(t) = \lim_{n\to\infty} x_n(t).$$

Так как равенство выполняется в каждой точке, то говорят, что последовательность $x_n(t)$ сходится к $\bar{x}(t)$ ***по-точечно***. Докажем теперь, что она сходится равномерно.

Воспользуемся неравенством

$$|x_n(t) - x_m(t)| \leqslant d(x_n, x_m).$$

Последовательность x_n фундаментальна в $C[a, b]$, поэтому для любого положительного числа ε, сколь бы малым оно ни было, при достаточно больших номерах n и m будет выполняться неравенство

$$d(x_n, x_m) \leqslant \varepsilon$$

Поэтому при достаточно больших n и m неравенство

$$|x_n(t) - \bar{x}_m(t)| \leqslant \varepsilon$$

справедливо для всех $t \in [a, b]$.

Отсюда, устремляя m к бесконечности, получаем:

$$|x_n(t) - \bar{x}(t)| \leqslant \varepsilon$$

для всех $t \in [a, b]$. Но это и означает, что x_n сходится к $\bar{x}$ равномерно на отрезке $[a, b]$.

Теперь можно применить уже упомянутую выше теорему о непрерывности предела равномерно сходящейся последовательности непрерывных функций и заключить, что функция $\bar{x}$ непрерывна, т. е.

$$\bar{x} \in C[a, b].$$

Обратим внимание на один существенный момент доказательства полноты $C[a, b]$. Первый шаг заключался в том, что, взяв произвольную фундаментальную последовательность $x_n \in C[a, b]$, мы нашли функцию $\bar{x}(t)$, к которой функции $x_n(t)$ сходятся поточечно. Это утверждение еще далеко от требуемого: необходимо доказать, что функция $\bar{x}(t)$ непрерывна и что x_n сходятся к $\bar{x}$ равномерно. Однако именно этот первый шаг доказательства часто бывает одним из самых трудных — требовалось найти хоть какую-нибудь функцию (непрерывность была установлена позже), к которой данная последовательность сходится в каком-либо смысле (а затем было доказано, что сходнамость равномерная).

Следует обратить также внимание на то, что для доказательства существования функции $x(t)$ мы воспользовались критерием Коши сходимости числовых последовательностей. Таким образом, полнота пространства $C[a, b]$ является следствием полноты множества действительных чисел.

1.3. Классификация функций

Имеет место следующая классификация функций (рис. 1.4)

1. Функция вида $P_n(x) = a_0x^n + a_1x^{n-1} + a_2x^{n-2} + ... + a_n$,

где $n \in N \cup \{0\}, a_0, a_1, ..., a_n \in R$, называется *целой рациональной функцией или многочленом степени n.*

2. Функция, представляющая собой отношение двух целых рациональных функций

$$\frac{P_m(x)}{Q_n(x)} = \frac{a_0x^m + a_1x^{m-1} + a_2x^{m-2} + ... + a_m}{b_0x^n + b_1x^{n-1} + b_2x^{n-2} + ... + a_n}$$

называется *дробно-рациональной.*

Совокупность дробно иррациональных и целых рациональных называется *рациональными функциями.*

3. Функция, полученная с помощью конечного числа суперпозиций и четырех арифметических действий над степенными функциями как с целыми так и с дробными показателями и не являющиеся рациональными называются *иррациональными.*

$$y = \sqrt{x}, f(x) = \frac{\sqrt[3]{x} + 2}{x^2 + 1} \text{ (пример таких функций)}$$

Рациональные и иррациональные функции образуют класс алгебраических функций.

4. Всякая функция, не являющаяся алгебраической, называется *трансцендентной.*

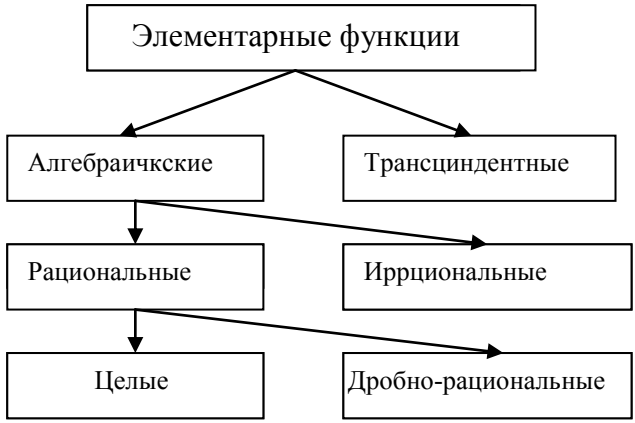


Рис. 1.4

Определение. ***Графиком*** функции (в системе декартовых прямоугольных координат) называется множество всех точек, абсциссы которых являются значениями независимой переменной, а ординаты — соответствующими значениями функции**.**

Иными словами, если взять абсциссу, равную некоторому значению независимой переменной, то ордината соответствующей точки графика должна быть равна значению функции, соответствующему данному значению независимой переменной. При этом масштабы на обеих осях координат могут быть как одинаковыми, так и различными.

Графиком функции обычно служит некоторая кривая (в частности, и прямая) линия.

Графиком постоянной величины служит прямая, параллельная оси абсцисс.

Обратно, если линия на координатной плоскости *Oxy* такова, что любая прямая, параллельная оси ординат, имеет с ней не более одной общей точки, то эта линия изображает некоторую функцию, именно ту, значения которой равны ординатам точек линии при значениях независимой переменной, равных абсциссам.

Понятия линии и функции тесно связаны. Заданием функции порождается линия — ее график; заданием линии порождается

функция — та, для которой эта линия служит графиком. Графическое задание функции состоит в задании графика этой функции.

1. Основные элементарные функция. Сложная функция.

Определение. *Основными элементарными функциями* называются следующие функции:

1) *степенная функция:* $y=x^n$, где n — действительное число;

2) *показательная функция:* $y=a^x$, где a — положительное число $a\neq 1$;

3) *логарифмическая функция:* $y=\log_a x$, где основание логарифмов a — положительное число и $a\neq 1$;

4) *тригонометрические функции:* $y=\sin x$, $y=\cos x$, $y=\operatorname{tg} x$, $y=\operatorname{ctg} x$, $y=\sec x$, $y=\operatorname{cosec} x$;

5) *обратные тригонометрические функции:* $y=\arcsin x$, $y=\arccos x$, $y=\operatorname{arctg} x$, $y=\operatorname{arcctg} x$, $y=\operatorname{arcsec} x$, $y=\operatorname{arccosec} x$.

Из основных элементарных функций можно строить другие функции при помощи арифметических действий (сложения, вычитания, умножения и деления) и новой операции взятия функции.

Аргумент функции может быть независимой переменной, а может быть промежуточной переменной, т. е. сам может являться функцией независимой переменной. В последнем случае получается сложная функция. Например, функция $y=\sin^2 x$ есть сложная функция x, а именно y есть квадратичная функция аргумента, который есть в свою очередь тригонометрическая функция — с и н у с —независимой переменной x; обозначая промежуточную переменную через u, можем записать: $y=u^2$, $u=\sin x$.

Задание сложной функции называют еще цепным заданием. Цепь функций, с помощью которых строится сложная функция, может состоять не только из двух, но из какого угодно числа звеньев. Чем их больше, тем «сложнее» функция. Часто оказывается полезным умение «расщеплять» заданную сложную функцию на отдельные звенья.

2. Элементарные функции.

I. Функции, построенные из основных элементарных функций и постоянных при помощи конечного числа арифметических действий и конечного числа операций взятия функции от функции, называются *элементарными функциями.*

Элементарные функции принято разделять на два класса: алгебраические функции и трансцендентные функции.

Определение. Функция называется *алгебраической,* если ее значения можно получить, производя над независимой переменной конечное число алгебраических действий: сложений, вычитаний, умножений, делений и возведений в степень с рациональным пока-

зателем. Функция, не являющаяся алгебраической, называется *трансцендентной.*

Алгебраические функции в свою очередь разделяются на рациональные и иррациональные.

Определение. Алгебраическая функция называется *рациональной,* если среди действий, которые производятся над независимой переменной, отсутствует извлечение корней.

Алгебраическая функция, не являющаяся рациональной, называется *иррациональной.*

Наиболее простыми рациональными функциями являются *целые* рациональные функции или *многочлены* (при этом одночлен рассматривается как частный случай многочлена),

3. Неявные функции. Многозначные функции.

Рассматривая функции, заданные аналитически, мы предполагали, что в левой части равенства, определяющего функцию, стоит только *y,* а в правой—выражение, зависящее от *x.* Такие функции мы будем называть *явными.*

Но и уравнение, связывающее две переменные, не разрешенное относительно какой-нибудь из них, может определять одну из этих переменных как функцию другой.

Например, в уравнении прямой $2x+3y=6$ ординату y можно рассматривать как функцию абсциссы $x,$ определенную на всей числовой оси. Действительно, каждому значению x соответствует одно значение $y,$ находимое из уравнения прямой. Эта функция задана в неявном виде. Чтобы перейти к ее явному заданию, разрешаем уравнение относительно y и получаем $y = 2 - \frac{2}{3}x.$

Можно сказать, что *неявной функцией y независимой переменной x называется функция, значения которой находятся из уравнения, связывающего x и y и не разрешенного относительно y.* Чтобы перейти к иному заданию функции, нужно разрешить данное уравнение относительно $y.$

Однако если мы возьмем уравнение окружности то столкнемся с тем обстоятельством, что каждому значению $x,$ заключенному между — 1 и +1, соответствует уже не одно, а два значения $y.$ В соответствии с этим мы скажем, что уравнение $x^2+y^2=1$ представляет не одну, а две функции: $y=+\sqrt{1-x^2}$ и $y= -\sqrt{1-x^2}$, каждая из которых определена в интервале [—1, 1]; если x лежит вне этого интервала, то уравнение, из которого находится $y,$ не имеет действительных решений.

С таким расширенным функции – именно с введением многозначных функций, рассматриваемых как совокупность их однозначных ветвей,— нам часто придется сталкиваться в общей теории оптимизации.

Пользуясь этим толкованием, можно сказать, что *неявной функцией у независимой переменной х называется функция, значения которой находятся из уравнения, связывающего х и у и не разрешенного относительно у.* Чтобы перейти к явному заданию функции, нужно разрешить данное уравнение относительно *у*.

1.4. Простейшие функции

1. Основные характеристики поведения функции. Изучить заданную функцию — это значит охарактеризовать ход ее изменения (или, как говорят, ее поведение) при изменении независимой переменной. При этом мы всюду (где специально не оговорено противное) будем предполагать, что *независимая переменная изменяется возрастая, причем, переходя от меньших значений к большим, она проходит через все свои промежуточные значения.*

Изначально поведение функции характеризуется следующими простейшими ее особенностями:

I. *Нули функции и знак функции в данном интервале.* II. *Четность или нечетность.* III. *Периодичность.* IV. *Рост в данном интервале.*

I. **Определение**. Значение *х,* при котором функция обращается в нуль, *f(x)*= 0, называется *нулем* функции.

В интервале положительного знака функции график ее расположен над осью *Ох,* в интервале отрицательного знака — под осью *Ох,* в нуле функции график имеет общую точку с осью *Ох*.

II. **Определение**. Функция *y=f(x)* называется *четной,* если при изменении знака у любого значения аргумента значение функции не изменяется:

$$f(-x)= f(x).$$

Функция *y = f(x)* называется *нечетной,* если при изменении знака у любого значения аргумента изменяется только знак значения функции, а абсолютная величина этого значения остается без изменения:

$$f(-x)= f-(x).$$

В этом определении предполагается, что функции определены в области, симметричной начала координат.

III. **Определение**. Функция *y =f(x)* называется *периодической,* если существует такое положительное число *а,* что для любого *х* справедливо равенство

$$f(x + a)=f(x). \qquad (1.13)$$

Наименьшее положительное число a, при котором условие (1.13) соблюдается, называется *периодом* функции.

IV. Весьма важной особенностью в поведении функции является *возрастание* и у*бывание.*

Определение. Функция называется *возрастающей* в интервале, если большим значениям аргумента соответствуют большие значения функции; она называется *убывающей,* если большим значениям аргумента соответствуют меньшие значения функции.

Таким образом, $f(x)$ возрастает в интервале $[a, b]$, если для любых значений x_1, x_2, удовлетворяющих условию $a \le x_1 < x_2 \le b$, имеет место неравенство $f(x_1) < f(x_2)$, и убывает, если для любых значений x_1, x_2, удовлетворяющих указанному условию, имеет место неравенство $f(x_1) > f(x_2)$.

Определение. Интервал независимой переменной, в котором функция возрастает, называется *интервалом возрастания* функции, а интервал, в котором функция убывает,— *интервалом убывания.* Как интервал возрастания, так и интервал убывания называют *интервалами монотонности* функции, а функцию в этом интервале — *монотонной функцией.*

Приведем определение еще одной очень важной характеристики поведения функции.

Определение. Значение функции, большее или меньшее всех других ее значений в некотором интервале, называется *наибольшим* или соответственно *наименьшим* значением функции в этом интервале.

Исследование указанных в настоящем пункте особенностей часто позволяет составить довольно ясное представление о поведении функции. Изучение функции следует начинать с отыскания области ее определения. При этом некоторые вопросы исследования могут отпасть сами собой; так, например, если функция определена только для положительных значений аргумента, то ее не нужно исследовать на четность и нечетность, и т. д.

Прямая пропорциональная зависимость и линейная функция. Приращение величины. *Прямой пропорциональной зависимостью* называется зависимость, выраженная формулой

$$y = ax,$$

где a — постоянная ($a \neq 0$).

Характерной особенностью этой зависимости является то, что величина y пропорциональна величине x. Это значит, что если x_1 и y_1, x_2 и y_2 — две пары соответственных значений x и y, то $y_2 : y_1 = x_2 : x_1$. Постоянный коэффициент a называется *коэффициентом пропорциональности.*

Если $y = ax$, то $x = \frac{1}{2}y$, и, значит, если y пропорционально x, то и x пропорционально y, но с обратным коэффициентом пропорциональности.

Прямая пропорциональная зависимость есть частный случай зависимости, устанавливаемой *линейной функцией,* т. е. функцией вида

$$y = ax+b,$$

где a, b — постоянные.

Определение. Пусть некоторая величина u переходит от одного своего (начального) значения u_1 к другому (конечному) значению u_2. Разность конечного и начального значений называется *приращением* величины u; ее обозначают через Δu, т. е.

$$\Delta u = u_2 — u_1.$$

Прямая теорема. Приращение линейной функции пропорционально приращению аргумента и не зависит от начального значения аргумента.

Найденное свойство вполне характеризует линейную функцию; другими словами, нет никаких других функций, кроме линейных, которые обладали бы тем же свойством.

Обратная теорема. Если приращение функции пропорционально приращению аргумента и не зависит от его начального значения, то эта функция — линейная.

Квадратичная функция. *Квадратичной функцией* называется функция вида

$$y = ax^2 + bx + c,$$

где a, b, c — постоянные ($a \neq 0$).

Квадратичная функция определена на всей числовой оси.

Квадратичная функция только один раз меняет характер своего изменения: она или сначала убывает, а затем возрастает, или наоборот. А именно, если $a > 0$, то функция в интервале $(-\infty, -\frac{b}{2a})$ убывает, достигая при $x = -\frac{b}{2a}$ своего наименьшего значения $y = -\frac{b^2 - 4ac}{4a}$, а затем в интервале $(-\frac{b}{2a}, +\infty)$ возрастает; наибольшего же значения функция нигде не достигает. Наоборот, при $a<0$ функция сначала в интервале $(-\infty, -\frac{b}{2a})$

возрастает, достигая при $x = -\frac{b}{2a}$ своего наибольшего значения $y = -\frac{b^2 - 4ac}{4a}$, а затем в интервале $(-\infty, -\frac{b}{2a})$ убывает; в этом случае функция нигде не достигает наименьшего значения.

Задачи отыскания наибольшего и наименьшего значении функции являются основной целью общей теории оптимизации .

Рассмотрим пример.

Из всех прямоугольников с данным периметром P найдем прямоугольник, имеющий наибольшую площадь.

Обозначим одну из сторон прямоугольника через x, тогда другая сторона будет равна $\frac{P}{2} - x$. Следовательно, площадь прямоугольника S равна

$$S = x(\frac{P}{2} - x) = \frac{P}{2}x - x^2$$

т. е. является квадратичной функцией от x. Так как $a = -1 < 0$, то функция имеет наибольшее значение; оно достигается при $x = -\frac{b}{2a} = \frac{P}{4}$ и равно

$$\frac{P}{2} \cdot \frac{P}{4} - \left(\frac{P}{4}\right)^2 = \frac{P^2}{16}$$

Но прямоугольник, периметр которого равен P, а одна из сторон имеет длину $\frac{P}{4}$, есть квадрат. Таким образом, из всех прямоугольников данного периметра наибольшую площадь имеет квадрат.

Обратная пропорциональная зависимость и дробно-линейная функция. *Обратной пропорциональной зависимостью* называется зависимость, выраженная формулой

$$y = \frac{a}{x}$$

где a — постоянная ($a \neq 0$).

Ее характерной особенностью является то, что величина y обратно пропорциональна величине x, т. е. если x_1 и y_1, x_2 и y_2—две пары соответственных значений x и y, то $y_2 : y_1 = x_1 : x_2$. Постоянный

коэффициент a называется *коэффициентом обратной пропорциональности.*

Обратная пропорциональная зависимость есть частный случай зависимости, устанавливаемой *дробно-линейной функцией,* т.е. функцией вида

$$y = \frac{ax + b}{cx + d} \tag{1.14}$$

где a, b, c, d—постоянные ($c \neq 0$).

При $a=0$, $d=0$ отсюда получается обратная пропорциональная зависимость.

Дробно-линейная функция (1.14) определена на всей оси Ox, за исключением точки $x = -\frac{d}{c}$ (точки $x = 0$ для функции $y = \frac{a}{x}$).

Будем предполагать, что $bc—ad \neq 0$, ибо в противном случае функция вырождается в постоянную.

Графиком дробно-линейной функции служит равнобочная гипербола, асимптоты которой параллельны осям координат.

Ход изменения дробно-линейной функции легко усматривается из графика, а именно:

Дробно-линейная функция (1.14) *или только возрастает, или только убывает в любом интервале оси Ox, не содержащем точки* $x = -\frac{d}{c}$.

Тот или другой характер роста (возрастание или убывание) зависит от знака выражения $bc — ad$. Если $bc—ad > 0$, то функция убывает, если $bc—ad < 0$, то функция возрастает.

Обратная функция. Пусть задана какая-либо функция $y=f(x)$; D — область ее определения (т. е. множество значений x), а G —область значений функции (т. е. множество соответствующих значений y). Будем говорить, что функция $y=f(x)$ осуществляет *отображение множества D в множество G;* из определения функции следует, что каждой точке множества D ставится в соответствие одна-единственная точка множества G. Если при этом каждой точке множества G соответствует опять-таки единственная точка множества D, то говорят, что *отображение (соответствие) взаимно однозначное.* При таком отображении различным точкам множества D соответствуют различные точки множества G.

Например, функция $y = x^3$ взаимно однозначно отображает всю ось Ox в ось Oy, так как каждому значению x соответствует одно-

единственное значение *y,* и наоборот: каждому *y* соответствует только один *x,* равный $\sqrt[3]{y}$.

Функция $y = x^3$ не осуществляет взаимно однозначного соответствия между осью *Ox* (область определения функции) и интервалом $[0, \infty)$ оси *Oy* (область значений функции), потому что любому значению $y > 0$ соответствует не одно, а два значения *x;* это $x = +\sqrt{y}$ и $x = -\sqrt{y}$.

Функция $y = \sin x$ отображает всю ось *Ox* в интервал $[-1, 1]$ оси *Oy*. Ясно видно, что это отображение не взаимно однозначное: каждому значению *y* из интервала $[-1, 1]$ соответствует бесчисленное множество значений *x*.

Если функция $y=f(x)$ осуществляет взаимно однозначное отображение множества *D* в множество *G*, то, согласно сказанному, каждому значению *y* из множества *G* ставится в соответствие одно определенное значение *x* из множества *D*. Поэтому можно сказать, что определена функция $x = \varphi(y)$; *G* является ее областью определения, а *D*—областью значений. Мы замечаем, что величины *x* и *y* как бы поменялись ролями: *y* стала независимой переменной, а *x* — функцией. В этом случае функции $y=f(x)$ и $x=\varphi(y)$ называются *взаимно обратными.*

Если отображение множества *D* в множество *G* не взаимно однозначно, то некоторым значениям *y* (а может быть, даже и всем) будет соответствовать несколько различных значений *x,* и мы опять сталкиваемся с той же трудностью, что и ранее при рассмотрении неявных функций, а именно с тем, что обратная функция оказывается многозначной.

Между графиками функций $y=f(x)$ и $y=\varphi(x)$ существует простая связь:

График обратной функции $y=\varphi(x)$ симметричен с графиком данной функции $y=f(x)$ относительно биссектрисы первого и третьего координатных углов.

Перейдем теперь к выяснению условий, при которых отображение множества значений *x* в множество значений *y* будет взаимно однозначным, т. е. условий, при соблюдении которых функция $y=f(x)$ будет иметь однозначную обратную функцию. Тот факт, что значениям *x* соответствуют единственные значения *y,* означает, что любая прямая, параллельная оси ординат, пересекает график функции не более чем в одной точке. Чтобы значениям соответствовали также единственные значения *x,* нужно чтобы этим же свойством обладали и прямые, параллельные оси абсцисс. Геометрически ясно, что *это*

условие будет соблюдаться, если данная функция монотонна; при этом монотонной будет и обратная функция.

Всюду в дальнейшем, говоря о взаимно обратных функциях, мы подразумеваем, что в областях их определения они *монотонны*.

Если данная функция не монотонна, то мы разбиваем область ее определения на интервалы монотонности и в каждом таком интервале берем соответствующую однозначную ветвь обратной функции.

Степенная функция. Степенная функция

$$x=y^n$$

при целых и положительных значениях n определена на всей числовой оси.

Показательная функция.

Показательная функция

$$y = a^x$$

рассматривается только при $a>0$ и $a \neq 1$.

Эта функция определена на всей оси Ox и всюду положительна: $a^x > 0$ при всяком x; это означает, что когда x — дробное число, мы берем только арифметическое значение корня. Поэтому график показательной функции расположен над осью Ox; так как $a^0=1$, то он проходит через точку (0,1). Поведение показательной функции существенно зависит от того, будет ли $a > 1$ или $a < 1$.

Если $a> 1$, то с увеличением показателя x увеличивается и y, причем неограниченное возрастание аргумента вызывает неограниченное же возрастание и функции. Если $a<1$, то, наоборот, при неограниченном возрастании аргумента функция убывает и неограниченно приближается к нулю.

Показательные функции встречаются в самых разнообразных задачах. При этом чаще всего имеют дело с показательной функцией, в основании которой лежит число e, играющее очень важную роль в математике; его приближенное значение равно 2,718. Часто функцию $y = e^x$ называют *экспоненциальной,* а ее график — *экспонентой.*

Логарифмическая функция. Логарифмическая функция

$$y = \log_a x$$

обратна показательной функции $y=a^x$, $a>0$, $a\neq1$.

Поведение логарифмической функции существенно зависит от того, будет ли $a> 1$ или $a <1$. В первом случае ($a>1$) логарифм во всем интервале $(0, \infty)$— возрастающая функция, притом отрицательная в интервале (0, 1) и положительная в интервале $(1, \infty)$. Во втором случае ($a < 1$) логарифм во всем интервале $(0, \infty)$ — убывающая функция, притом положительная в интервале (0, 1) и отрицательная в интервале $(1, \infty)$.

Принимая во внимание, что логарифмическая и показательная функции взаимно обратны, имеем

$$\log_a a^x = x \quad \text{и} \quad a^{\log_a x} = x;$$

первое из этих равенств справедливо при любом x, второе — при $x > 0$.

Пользуясь последним соотношением, всякую степенную функцию $y=x^n$ при $x > 0$ с любым показателем n можно представить в виде сложной функции, составленной из показательной и логарифмической функций:

$$y = x^n = (a^{\log_a x})^n = a^{n \log_a x}.$$

Тригонометрические функции. Гармонические колебания.

В качестве независимой переменной тригонометрических функций

$$y=\sin x, \quad y=\cos x, \quad y=\operatorname{tg} x,$$
$$y=\sec x, \quad y=\operatorname{cosec} x, \quad y=\operatorname{ctg} x$$

в математическом анализе всегда принимается радианная мера дуги или угла. Так, например, значение функции $y = \sin x$ при $x=x_0$ равно синусу угла в x_0 радианов.

Между шестью тригонометрическими функциями: $y=\sin x$, $y=\cos x$, $y=\operatorname{tg} x$, $y=\sec x$, $y=\operatorname{cosec} x$, $y=\operatorname{ctg} x$ — существуют пять простых независимых алгебраических соотношений, выводимых на основании определений этих функций и позволяющих по одной из них находить остальные.

Тригонометрические функции периодичны. Именно, функции $\sin x$, $\cos x$ (а потому и $\sec x$, и $\operatorname{cosec} x$) имеют период 2π, а функция $\operatorname{tg} x$ (а потому и $\operatorname{ctg} x$) — период π.

Гармонические колебания. Тригонометрические функдии имеют важные применения в математике, в естествознании и в технике. Они встречаются там, где приходится иметь дело с периодическими явлениями, т. е. явлениями, повторяющимися в одной и той же последовательности и в одном и том же виде через определенные интервалы аргумента (чаще всего — времени).

Простейшие из таких явлений — гармонические колебания, в которых расстояние s колеблющейся точки от положения равновесия является функцией времени t:

$$s = A \sin(\omega t + \varphi_0).$$

Эту функцию называют *синусоидальной.* Постоянное число A называется *амплитудой колебания.* Число A представляет собой то наибольшее значение, которого может достигнуть s (размах колебания). Аргумент синуса $\omega t + \varphi_0$ называется *фазой колебания,* а число φ_0, равное значению фазы при $t = 0$ — *начальной фазой.*

Наконец, $\frac{\omega}{2\pi}$ — называется *частотой колебания.*

Колебания, описываемые уравнением

$$s = A \sin(\omega t + \varphi_0)$$

называются *простыми гармоническими колебаниями,* а их графики *— простыми гармониками.*

Колебания, получающиеся в результате сложения нескольких простых гармонических колебаний, называются *сложными гармоническими колебаниями,* а их графики*—сложными гармониками.*

Обратные тригонометрические функции. Обратные тригонометрические функции определяются с помощью изложенных ранее. Начнем с функции, обратной для функции $y=\sin x$. Область определения $\sin x$ — всю числовую ось — разбиваем на интервалы монотонности, которых бесконечно много:

$$\ldots,\left[-\frac{3\pi}{2},-\frac{\pi}{2}\right],\left[-\frac{\pi}{2},\frac{\pi}{2}\right],\left[\frac{\pi}{2},\frac{3\pi}{2}\right],\ldots$$

Выберем в качестве основного интервал $\left[-\frac{\pi}{2},\frac{\pi}{2}\right]$ и функцию, обратную к функции $y=\sin x$ на этом интервале, обозначим через $y = \arcsin x$. Так как область значений функции $y = \sin x$ есть интервал [—1, 1], то этот же интервал есть область определения функции $y = \arcsin x$; областью ее значений является интервал $\left[-\frac{\pi}{2},\frac{\pi}{2}\right]$

$$-\frac{\pi}{2} \le \arcsin x \le \frac{\pi}{2}.$$

Аналогично определяется функция, обратная к функции $y=\cos x$. Интервалами монотонности $\cos x$ являются интервалы

..., [—2π, — π], [—π, 0], [0, π], [π, 2 π], ...

Функцию, обратную к функции $y= \cos x$ в интервале [0, π], обозначим через $y = \arccos x$. Эта функция определена в интервале [—1, 1] и принимает значения, заключенные между 0 и π:

$$0 \le \arccos x \le \pi.$$

Следовательно, равенство $y = \arccos x$ эквивалентно двум равенствам

$$\cos y = \cos(\arccos x) = x, \qquad 0 \le y \le n.$$

Функция $y = \arccos x$ убывающая, так как в интервале [0, π] убывает и $\cos x$.

Перейдем к функции, обратной для функции $y = \operatorname{tg} x$. В интервале $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ функция $\operatorname{tg} x$ возрастает и, следовательно, имеет обратную, которую мы обозначим через $y = \operatorname{arctg} x$. Из свойств функции $\operatorname{tg} x$ следует, что функция $y = \operatorname{arctg} x$ определена на всей числовой оси и является возрастающей и нечетной. Область ее значений— интервал $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$, т. е.

$$-\frac{\pi}{2} < \operatorname{arctg} x < \frac{\pi}{2}.$$

Функция $y = \operatorname{arctg} x$ называется *главным значением* $\operatorname{Arctg} x$.

Всюду в дальнейшем, если не будет оговорено противоположное, под обратными тригонометрическими функциями мы будем понимать их главные значения.

Гиперболические и обратные гиперболические функции.

Хотя функции, названные в заголовке пункта, не принадлежат к числу основных элементарных функций, мы все же рассмотрим их здесь, так как они могут быть исследованы самыми простыми средствами. Эти функции понадобятся нам в дальнейшем; кроме того, они встречаются при решении различных оптимизационных прикладных задач (в курсах электротехники, сопротивления материалов и др.).

I. **Определение**. *Гиперболическими косинусом* ($\operatorname{ch} x$), *синусом* ($\operatorname{sh} x$) и *тангенсом* ($\operatorname{th} x$) называются функции, определенные формулами

$$\operatorname{ch} x = \frac{e^x + e^{-x}}{2}, \quad \operatorname{sh} x = \frac{e^x - e^{-x}}{2},$$

$$\operatorname{th} x = \frac{\operatorname{sh} x}{\operatorname{ch} x} = \frac{e^x - e^{-x}}{e^x + e^{-x}},$$

где $e = 2{,}718....$

Эти функции определены на всей числовой оси. Они связаны рядом соотношений, аналогичных соотношениям между соответствующими тригонометрическими функциями, что и объясняет их названия. В частности, имеют место формулы:

$$\operatorname{ch}^2 x - \operatorname{sh}^2 x = 1, \quad \operatorname{ch} 2x = \operatorname{ch}^2 x + \operatorname{sh}^2 x, \quad \operatorname{sh} 2x = 2 \operatorname{sh} x \operatorname{ch} x,$$

$$\operatorname{ch}^2 x = \frac{1}{1 - \operatorname{th}^2 x}, \quad \operatorname{sh}^2 x = \frac{\operatorname{th}^2 x}{1 - \operatorname{th}^2 x}.$$

II. **Обратные гиперболические функции**. Функции, обратные к соответствующим гиперболическим функциям, обозначаются через

$$y = \operatorname{Arch} x, \quad y = \operatorname{Arsh} x, \; y = \operatorname{Arth} x$$

Функция sh x определена и возрастает в интервале (— ∞, ∞); область ее значений совпадает со всей осью Oy. Поэтому функция y=Arsh x также определена на всей числовой оси и возрастает. Ее можно выразить при помощи логарифмической функции.

$$y = \operatorname{Arsh} x = \ln\left(x + \sqrt{x^2 + 1}\right) \tag{1.15}$$

Функция ch x убывает в интервале (—∞, 0] и возрастает в интервале [0, ∞); область ее значений — интервал [1, ∞). Следовательно, функция y=Arch x состоит из двух однозначных ветвей, определенных при $x \geq 1$. Проводя такие же выкладки, как при выводе формулы (1.15), получим два выражения:

$$(\operatorname{Arch} x)_1 = \ln\left(x - \sqrt{x^2 - 1}\right), \quad (\operatorname{Arch} x)_2 = \ln\left(x + \sqrt{x^2 - 1}\right).$$

Первая функция обратна к функции y=ch x в интервале (—∞, 0], а вторая — в интервале [0, ∞). Легко заметить, что

$$x - \sqrt{x^2 - 1} = \frac{1}{x + \sqrt{x^2 - 1}},$$

т. е. что $\ln\left(x - \sqrt{x^2 - 1}\right) = -\ln\left(x + \sqrt{x^2 - 1}\right)$.

Функция y = Arth x определена в интервале (—1, 1) — это область значений функции th x — и, как нетрудно проверить, равна

$$\operatorname{Arth} x = \frac{1}{2} \ln \frac{1 + x}{1 - x}.$$

1.5. Непрерывные функции

Непрерывность функции. Напомним прежде всего, что приращением функции $y=f(x)$ в данной точке x_0 называется разность

$$\Delta y = f(x_0 + \Delta x) - f(x_0),$$

где Δx—приращение аргумента (рис. 1.5). Введем теперь следующее определение.

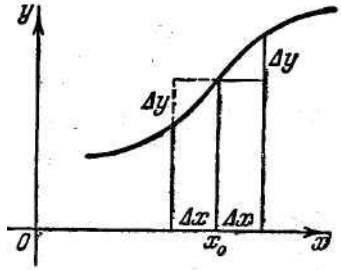


Рис. 1.5

Определение. *Функция y=f(x) называется непрерывной в точке x_0, если эта функция определена в какой-нибудь окрестности точки x_0 и если*

$$\lim_{\Delta x \to 0} \Delta y = 0,$$

т. е. если бесконечно малому приращению аргумента соответствует бесконечно малое приращение функции.

Например, функция $y = x^3$ непрерывна в любой точке x_0. Так как

$$\Delta y = (x_0 + \Delta x)^3 - x_0^3 = 3x_0^2 \Delta x + 3x_0 \Delta x^2 + \Delta x^3,$$

то ясно видно, что если $\Delta x \to 0$, то и $\Delta y \to 0$, а это и означает что функция непрерывна.

Описательно можно сказать, что функция непрерывна, если она изменяется постепенно, т. е. если малые изменения аргумента влекут за собой малые же изменения функции. Эта особенность выражает общую характерную черту многих явлений и процессов. Так, мы считаем, например, что стержень при нагревании удлиняется непрерывно, что рост организма происходит непрерывно, что температура воздуха изменяется непрерывно и т. п.

Пользуясь выражением для Δy, можно записать также, что

$$\lim_{\Delta x \to 0} f(x_0 + \Delta x) = f(x_0).$$

или, иначе,

$$\lim_{\Delta x \to 0} [f(x_0 + \Delta x) - f(x_0)] = 0,$$

Если обозначить $x_0 + \Delta x$ через x, то x при $\Delta x \to 0$ будет стремиться к x_0 и последнее равенство можно переписать так:

$$\lim_{x \to x_0} f(x) = f(x_0).$$

Таким образом, можно сказать:

Функция $y=f(x)$ *непрерывна в точке* x_0, если она определена в какой-нибудь окрестности этой точки и если предел функции при стремлении независимой переменной x к x_0 существует и равен значению функции при $x = x_0$:

$$\lim_{x \to x_0} f(x) = f(x_0). \qquad (*)$$

Заметим, что *если значение функции* $f(x)$ *в точке* x_0, *в которой она непрерывна, отлично от нуля,* $f(x_0)\neq 0$, *то значения функции* $f(x)$ *в некоторой окрестности точки* x_0 *имеют тот же знак, что и* $f(x_0)$. Действительно, вследствие непрерывности существует окрестность точки x_0, в которой значения $f(x)$ настолько мало отличаются от своего предела, т. е. от $f(x_0)$, что они остаются положительными, если $f(x_0)>0$, и остаются отрицательными, если $f(x_0) < 0$.

Определение. *Функция называется непрерывной в интервале, если она непрерывна в каждой его точке.*

Для концов интервала определения непрерывности функции в точке надо несколько изменить. Именно, для левого конца интервала приращению Δx следует придавать только положительные значения, а для правого — только отрицательные.

Геометрически непрерывность функции $y=f(x)$ в интервале означает, что ординаты ее графика, соответствующие двум точкам оси Ox, как угодно мало отличаются друг от друга, если расстояние между этими точками достаточно мало. Поэтому график непрерывной функции представляет собой сплошную линию без разрывов; такую линию можно вычертить, двигаясь в одном направлении, скажем слева направо, не отрывая карандаша от графика, так сказать, «одним росчерком».

Собственно говоря, во всем предыдущем изложении, в частности при описании свойств и графиков основных элементарных функций, мы уже предполагали, что эти функции непрерывны.

Будем считать, что *все основные элементарные функции непрерывны в тех интервалах, в которых они определены.* Дальше это общее положение будет распространено на все элементарные функции.

Точки разрыва функции. Если рассматривать график функции $y = \frac{1}{x}$ в окрестности точки x=0, то ясно видно, что он как бы «разрывается» на отдельные кривые. Говорят, что во всех указанных точках соответствующие функции становятся *разрывными.*

Определение. Если в какой-либо точке x_0 функция не является непрерывной, то точка x_0 называется *точкой разрыва* функции, а сама функция—*разрывной* в этой точке.

При этом предполагается, что функция *f(x)* определена в некоторой окрестности точки x_0; в самой же точке x_0 функция может быть как определена, так и не определена.

Наиболее характерными и часто встречающимися точками разрыва являются точки бесконечного разрыва, т. е. такие, в окрестности которых функция является неограниченной.

Важный класс точек разрыва образуют точки разрыва первого рода. Для их определения введем понятие левого и правого пределов функции.

Ясно, что если функция имеет предел при произвольном стремлении x к x_0, то существуют ее левый и правый пределы и они равны между собой. Обратное тоже справедливо: если левый и правый пределы существуют и равны между собой, то функция имеет тот же предел при произвольном стремлении x к x_0.

Определение. *Точкой разрыва первого рода* функции *f(x)* называется такая точка x_0, в которой функция имеет левый и правый пределы, не равные между собой. Геометрическая иллюстрация точки разрыва первого рода ясна из рис. 1. 6.

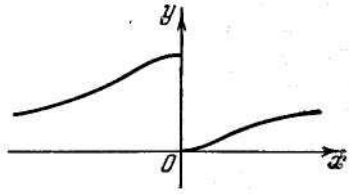


Рис. 1. 6

Все остальные точки разрыва называются *точками разрыва второго рода.*

Действия над непрерывными функциями. Непрерывность элементарных функций. Покажем сейчас, что если над непрерывными функциями произвести конечное число арифметических действий или операций взятия функции от функции, то в результате получится, как правило, также непрерывная функция. В каждом случае мы покажем, что предел соответствующей функции будет равен ее значению в предельной точке, а это и означает непрерывность функции.

Теорема I. Сумма конечного числа функций, непрерывных в некоторой точке, является функцией, непрерывной в той же точке.

Теорема II. Произведение конечного числа функций, непрерывных в некоторой точке, является функцией, непрерывной в той же точке.

Теорема III. Частное двух функций, непрерывных в некоторой точке, является функцией, непрерывной в той же точке, если только знаменатель не обращается в ней в нуль.

Теорема IV. Сложная функция, составленная из конечного числа непрерывных функций, непрерывна.

Теорема V. Функция, обратная к монотонной и непрерывной функции, непрерывна.

Так как мы уже отмечали, что все основные элементарные функции непрерывны в интервалах их определения, то в силу изложенных четырех теорем можно утверждать, что и *любая элементарная функция непрерывна в тех интервалах, где она определена.*

Точками разрыва элементарной функции могут быть только те значения независимой переменной, при которых какие-нибудь из составляющих функцию элементов делаются неопределенными или при которых обращаются в нуль знаменатели участвующих в выражении функции дробей.

Отметим, что предельное равенство

$$\lim_{x \to x_0} f(x) = f(x_0),$$

выражающее непрерывность функции при $x=x_0$, можно переписать так:

$$\lim_{x \to x_0} f(x) = f\left(\lim_{x \to x_0} x\right).$$

Значит, *символ предела и символ непрерывной функции можно переставлять между собой.*

Отсюда следует простое, практически удобное правило предельного перехода:

Для того чтобы найти предел элементарной функции, когда аргумент стремится к значению, принадлежащему области определения функции, нужно в выражение функции вместо аргумента подставить его предельное значение.

Это правило очень важное, поскольку в применениях анализа употребляются преимущественно элементарные функции.

Случаи, когда аргумент стремится к бесконечности или к какому-нибудь значению, не принадлежащему области определения функции, требуют всегда специального рассмотрения.

Свойства непрерывных функций. Непрерывность функции в замкнутом интервале обусловливает наличие у этой функции ряда важных свойств общего характера. Укажем (без доказательства) два из этих свойств.

Теорема I. Функция, непрерывная в *замкнутом* интервале, хотя бы в одной точке интервала принимает наибольшее значение и хотя бы в одной—наименьшее.

Теорема II. Функция, непрерывная в *замкнутом* интервале и принимающая на концах этого интервала значения разных знаков, хотя бы один раз обращается в нуль внутри интервала**.**

Теорему II можно формулировать в более общем виде: *Функция, непрерывная в замкнутом, интервале, принимает внутри интервала хотя бы один раз любое значение, заключенное между ее значениями на концах интервала.*

Непрерывная функция, переходя от одного своего значения к другому, обязательно проходит через все промежуточные значения; в частности,

Непрерывная в интервале функция принимает в этом интервале хотя бы один раз любое значение, заключенное между ее наименьшим и наибольшим значениями. Геометрически ясно, что если функция в данном интервале монотонна и, скажем, для определенности, возрастает, то свое наименьшее значение она принимает на левом конце интервала, а наибольшее — на правом; при этом любое промежуточное значение функция принимает в точности один раз.

2. Отображения и функции

Возвратимся к функциональному соответствию (т.е. к функции).

Если это соответствие и вдобавок еще и всюду-определенное, то оно называется ОТОБРАЖЕНИЕМ.

Если отобразить множество студентов в группе, множество фамилий в группе, то это скорее всего будет ОТОБРАЖЕНИЕ множества студентов НА множество фамилий. Т.е. сюръективное соответствие. Если же отобразить множество студентов группы на множество фамилий студентов университета, то говорят, что имеет место ОТОБРАЖЕНИЕ множества студентов В множество фамилий. Т.е., в области значений будут и "незадействованные фамилии".

Мы подошли к одному из самых фундаментальных понятий в теории множеств и теории оптимизации, в частности, - ГОМОМОРФИЗМУ.

Пример. Отобразим множество точек участка земной поверхности на множество точек карты. Сейчас оставим в стороне то, что какое-то множество точек земной поверхности отобразится в одну

точку на карте, в таких случаях неинъективность - обычное дело. Для нас существенным образом важно то, что чем выше точки земной поверхности над уровнем моря, тем в более коричневые точки карты они отображаются.

Таким образом, мы рассматриваем не просто множества элементов. В первом случае здесь между элементами множества существует отношение "выше", а во втором - "более коричневые". Где выше в первом - там более коричневые во втором. "Выше" и "более коричневые" - это отношения, которые заданы на своих множествах.

Отображение земной поверхности НА карту не просто ставит всем элементам одного множества элементы другого. Но, кроме того, если между двумя элементами первого множества существует отношение "выше", то между их образами во втором множестве имеет место отношения " более коричневые ". Очевидно, если точки земной поверхности лежат на одной высоте, то они отобразятся в точки карты с одинаковой коричневостью. Такое отображение называется ГОМОМОРФНЫМ. Или говорят, что между этими множествами существует ГОМОМОРФИЗМ.

Обратим внимание на то, что слово это не очень благозвучное, а по американским меркам и громоздкое. Поэтому по обыкновению используется более короткий (усеченный) термин - МОРФИЗМ.

Морфизмы играют в математике исключительную роль. Так как мамематику часто отождествляют с математическим моделированием, то приведем афоризм из одной умной философской книжки: КРАСИВАЯ МОДЕЛЬ ВСЕГДА ГОМОМОРФНА.

2.1. Формальное определение отображения и его свойства

Пусть X и Y — некоторые множества и $\Gamma \subseteq Y \times X$, причем $\Pr_1\Gamma = X$.

Тройка множеств (X, Y, Γ) определяет некоторое соответствие, которое обладает, однако, тем свойством, что его область определения $\Pr_1\Gamma$ совпадает с областью значений, т.е. X, и, следовательно, это соответствие определено всюду на X. Другими словами, для каждого $x \in X$ существует $y \in Y$, так что $(x,y) \in \Gamma$. Такое всюду определенное соответствие называется *отображением* X в Y, и записывается как

$$\Gamma : X \to Y \qquad (2.1)$$

Под словом «отображение» часто понимают однозначное отображение. Однако мы не будем придерживаться этого правила и

будем считать, что каждому элементу $x \in X$ отображение Г ставит в соответствие некоторое подмножество

$$\Gamma x \subseteq Y, \tag{2.2}$$

которое называют образом элемента *x*. Закон, в соответствии с которым осуществляется соответствие, определяется множеством Г.

Рассмотрим некоторые свойства отображения. Пусть $A \subseteq X$. Для любого $x \in A$ образом x будет множество $\Gamma x \subseteq Y$. Совокупность всех элементов *Y*, которые являются образами Г*x* для всех $x \in A$, назовем образом множества A и будем обозначать ГA. Согласно этому определению

$$\Gamma A = \bigcup_{x \in A} \Gamma x \cdot \tag{2.3}$$

Если A_1 и A_2 — подмножества *X*, то

$$\Gamma(A_1 \cup A_2) = \Gamma A_1 \cup \Gamma A_2 . \tag{2.4}$$

Однако соотношение

$$\Gamma(A_1 \cap A_2) = \Gamma A_1 \cap \Gamma A_2 \tag{2.5}$$

справедливо только в том случае, если отображение является однозначным. В общем же случае

$$\Gamma(A_1 \cap A_2) \subseteq \Gamma A_1 \cap \Gamma A_2 . \tag{2.6}$$

Полученные соотношения легко обобщаются и на большее число подмножеств A_i. Так, если A_1.., A_n — подмножества *X*, то

$$\Gamma\left(\bigcup_{i=1}^{n} A_i\right) = \bigcup_{i=1}^{n} \Gamma A_i ; \tag{2.7}$$

$$\Gamma\left(\bigcup_{i=1}^{n} A_i\right) \subseteq \bigcup_{i=1}^{n} \Gamma A_i \cdot \tag{2.8}$$

2.2. Типы отображений

При *отображении X* в *Y* каждый элемент *x* из *X* имеет один и только один образ *y*=Г*(x)* из *Y*. Однако совсем не обязательно, чтобы и всякий элемент из *Y* был образом некоторого элемента из *X* (рис. 2.1, *а)*. Если же любой элемент из *Y* есть образ, по крайней мере, одного элемента из *X* (рис. 2.1, б), то говорят, что имеет место *отображение X на Y (сюръекция* или *накрытие).*

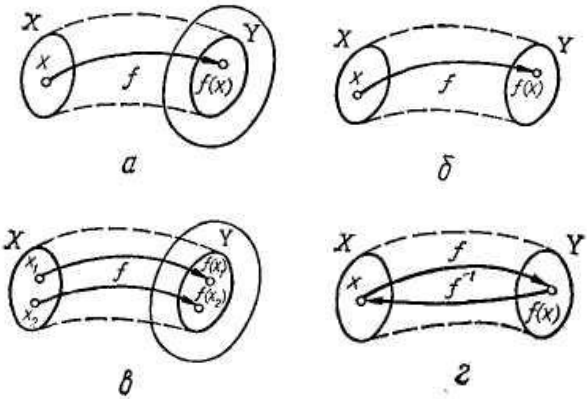


Рис. 2.1. Типы отображений:
а — отображение *X* в *Y*;
б — отображение *X* на *Y* (сюръекция);
в — взаимно-однозначное отображение *X* в *Y* (инъекция);
г — взаимно-однозначное отображение *X* на *Y* (биекция).

Если для любых двух разных элементов x_1 и x_2 из *X* их образы $y_1=\Gamma(x_1)$ и $y_2=\Gamma(x_2)$ также разные, то отображение называется *инъекцией* (рис. 2.1,*в)*. Отображение, которое является одновременно сюръективным и инъективным (рис. 2.1, *г)*, называется *биекцией (наложением)*. В этом случае говорят, что $\Gamma: X \to Y$ *есть взаимно-однозначное* отображение, а между элементами *X* и *Y* есть взаимно-однозначное соответствие. При этом, обратное отношение Γ^{-1} также взаимно-однозначное отображение, $x=\Gamma^{-1}(y)$ равносильно $y=\Gamma(x)$ и $(\Gamma^{-1})^{-1}$ совпадает с Г.

Любое отображение Г из *X* в *Y* есть элемент множества $U(X\times Y)$, которое обозначается также через Y^X (напомним, что $U(X\times Y)$ — это множество всех подмножеств прямого произведения $X\times Y$, а элементами последнего являются упорядоченные пары (x, y), где $x \in X$ и $y \in Y$). Если Г — взаимно-однозначное отображение, а множества *X* и *Y* совпадают $(X=Y)$, то $\Gamma: X \to X$ называют *отображением множества X на себя.* Элементы $(x, x) \in X\times X$ образуют *тождественное отображение е,* причем

$$\Gamma\Gamma^{-1} = \Gamma^{-1}\Gamma = e.$$

2.3. Отображения, заданные на одном множестве

Важным частным случаем отображения является случай, когда множества X и Y совпадают. При этом отображение Γ: $X \to X$ будет представлять собой отображение множества X самого в себя и будет определяться парой

$$(X, \Gamma), \tag{2.9}$$

где $\Gamma \subseteq X^2$.

Подробным изучением таких отображений занимается теория графов. Затронем здесь лишь некоторые операции над подобными отображениями.

Пусть Γ и Δ — отображение множества X в X. Композицией этих отображений назовем отображение $\Gamma\Delta$, которое определяется следующим образом:

$$(\Gamma\Delta)x = \Gamma(\Delta x). \tag{2.10}$$

В частном случае, если $\Delta=\Gamma$, получаем отображение

$$\Gamma^2 x = \Gamma(\Gamma x); \tag{2.11}$$

$$\Gamma^3 x = \Gamma(\Gamma^2 x) \text{ и т.д.} \tag{2.12}$$

Таким образом, в общем случае для любого $s \geq 2$

$$\Gamma^s x = \Gamma(\Gamma^{s-1} x). \tag{1.13}$$

Специальным определением введем соотношение

$$\Gamma^0 x = x. \tag{2.14}$$

Это дает возможность распространить соотношение (2.13) и на отрицательные s. Действительно, согласно (2.13)

$$\Gamma^0 x = \Gamma(\Gamma^{-1} x) = \Gamma\Gamma^{-1} x = x. \tag{2.15}$$

Это означает, что $\Gamma^{-1} x$ представляет собой обратное отображение.

Тогда

$$\Gamma^{-2} x = \Gamma^{-1}(\Gamma^{-1} x) \tag{2.16}$$

и т.д.

Пример 1. Пусть X — множество людей. Для каждого человека $x \in X$ обозначим через Γx множество его детей. Тогда $\Gamma^2 x$— множество внуков x; $\Gamma^3 x$— множество правнуков x; $\Gamma^{-1} x$— множество родителей x и т. д.

Изображая людей точками и рисуя стрелки, которые идут из x в Γx, получаем родословное или гениалогическое дерево (рис. 2.2).

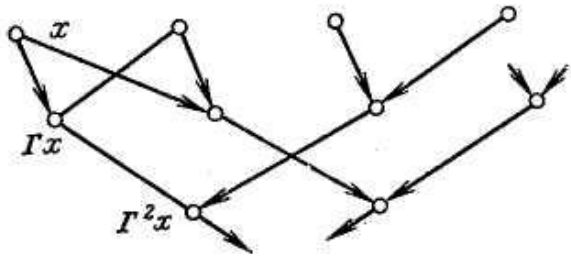


Рис. 2.2. Гениалогическое дерево

Пример 2. Рассмотрим шахматную игру. Обозначим через x некоторое положение (расположение фигур на доске), которое может создаваться в процессе игры, а через X множество всевозможных положений. Тогда Γx для любого $x \in X$ будет означать множество положений, которые можно получить из x, делая один ход при соблюдении правил игры. При этом $\Gamma x = \varnothing$, если x матовое или патовое положение; $\Gamma^3 x$ — множество положений, которые можно получить из x тремя ходами; $\Gamma^{-1} x$ — множество положений, из которых данное положение может быть получено за один ход.

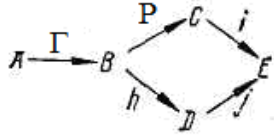
Если элементы $x \in X$ представляют собой состояние динамической системы, то отображение Γx может рассматриваться как множество состояний, в которые система может перейти из данного состояния. В этом случае удобно использовать термин *преобразования состояния динамической системы*. Для обозначения некоторых специальных видов отображений, заданных на одном том же множестве, используется также термин *отношение.*

2.4. Композиция отображений.

Если $\Gamma: X \to Y$ и $P: Y \to Z$, то их композиция $(P \circ \Gamma): X \to Z$, причем $(P \circ \Gamma)(x) = P(\Gamma(x))$. Пусть, например, $\Gamma = \sin$, $P = \ln$; тогда

$$(P \circ \Gamma)(x) = (\ln \circ \sin)x = \ln \sin x.$$

Для наглядности представления соотношений, где встречается несколько отображений, пользуются диаграммами, например:



Такая диаграмма называется *коммутативной,* если в любом случае, когда можно пройти от одного множества к другому по различным последовательностям стрелок, соответствующие композиции совпадают (в приведенном выше примере условие коммутативности $i \circ P = j \circ h$).

2.5. Подстановки как отображение.

Взаимно-однозначное отображение множества $N=\{1, 2, ..., n\}$ на себя называется *подстановкой n* чисел (или подстановкой *n*-й степени). Обычно принято записывать подстановку двумя строками, заключенными в скобки. Первая строка содержит аргументы (первые координаты) подстановки, а вторая - соответствующие им образы (вторые координаты). Например, взаимно-однозначное соответствие четырех чисел, заданное множеством упорядоченных пар {(1, 2), (2, 4), (3, 3), (4, 1)} запишется как подстановка *a* четвертой степени

$$a = \begin{pmatrix} 1\,2\,3\,4 \\ 2\,4\,3\,1 \end{pmatrix}$$

в которой 1 переходит в 2, 2 — в 4, 3 — в 3 и 4 — в 1.

Так как безразлично, в каком порядке идут упорядоченные пары отображения, то одна и та же подстановка допускает различные представления:

$$\begin{pmatrix} 1\,2\,3\,4 \\ 2\,4\,3\,1 \end{pmatrix} = \begin{pmatrix} 4\,2\,3\,1 \\ 1\,4\,3\,2 \end{pmatrix} = \begin{pmatrix} 2\,1\,4\,3 \\ 4\,2\,1\,3 \end{pmatrix} \text{ и т.д.}$$

Каждая строка в записи подстановки *n*-й степени содержит *n* различных чисел, расположенных в определенном порядке, т.е. представляет собой некоторую *перестановку n* чисел 1, 2, ..., *n.* Если обозначить *i*-е элементы перестановок через α_i и β_i (i = 1, 2,..., *n),* причем $\alpha_i, \beta_i \in N$, то подстановку *n*-й степени можно представить как

$$a = \begin{pmatrix} \alpha_1, \alpha_2, \ldots, \alpha_n \\ \beta_1, \beta_2, \ldots, \beta_n \end{pmatrix}.$$

Поскольку число всех перестановок из *n* чисел равно *n*!, то число всех различных подстановок *n*-й степени, как и число всевозможных способов записи каждой из таких подстановок, также равно *n*!

Тождественная подстановка n-й степени e_n переводит каждое число в себя. Очевидно, одной из записей e_n является следующая:

$$e_n = \begin{pmatrix} 1\,2 \ldots n \\ 1\,2 \ldots n \end{pmatrix}.$$

Если в подстановке a поменяем местами ее перестановки, то получим подстановку a^{-1}, *симметричную a.* Например

$$a = \begin{pmatrix} 1\,2\,3\,4 \\ 2\,4\,3\,1 \end{pmatrix}; \qquad a^{-1} = \begin{pmatrix} 2\,4\,3\,1 \\ 1\,2\,3\,4 \end{pmatrix}$$

Композицией подстановок n-й степени a и b называется подстановка n-й степени $c=ab$, являющаяся результатом последовательного выполнение сначала a, потом b. Например:

$$c = ab = \begin{pmatrix} 1\,2\,3\,4 \\ 2\,4\,3\,1 \end{pmatrix} \begin{pmatrix} 1\,2\,3\,4 \\ 1\,4\,3\,2 \end{pmatrix} = \begin{pmatrix} 1\,2\,3\,4 \\ 4\,2\,3\,1 \end{pmatrix}$$

так как 1 переходит в 2 и 2 — в 4, т.е. в результате 1 переходит в 4 и т.д.

Очевидно, если a — подстановка n-й степени, то

$$ae_n = e_n a = a, \quad aa^{-1} = a^{-1}a = e_n.$$

Подстановка называется *четной,* если общее число инверсий в ее строках (перестановках) четно, и *нечетной* — в противном случае. Как известно, *инверсию* образуют два числа в перестановке, когда меньшее из них расположено правее от большего. Каждой перестановке можно сопоставить число инверсий в ней, которое подсчитывается следующим образом: для каждого из чисел определяется количество стоящих правее его меньших чисел, и полученные результаты складываются. Например, подстановка

$$\begin{pmatrix} 4\,2\,5\,1\,3\,6 \\ 5\,3\,1\,4\,2\,6 \end{pmatrix}$$

нечетная, так как количество инверсий в верхней перестановке

$$3+1+2+0+0+0=6$$

и в нижней перестановке

$$4+2++0+1+0+0=7,$$

т.е. общее число инверсий 6 + 7=13.

2.6. Разложение подстановки в циклы

Всякую подстановку можно разложить в *произведение циклов,* множество элементов которых попарно не пересекаются. *Цикл* — это такая подстановка

$$\begin{pmatrix} \alpha_1, \alpha_2, \ldots, \alpha_{k-1}, & \alpha_k, \alpha_{k+1}, \ldots, \alpha_n \\ \alpha_2, \alpha_3, \ldots, \alpha_k, & \alpha_1, \alpha_{k+1}, \ldots, \alpha_n \end{pmatrix} = (\alpha_1, \alpha_2, \ldots, \alpha_k)$$

которая переводит α_1 в α_2, α_2 в α_3, ..., α_{k-1} в α_k и α_k в α_1, а другие элементы α_{k+}, ..., α_n переходят в самих себя.

Сокращенная запись цикла $(\alpha_1, \alpha_2, ..., \alpha_k)$ сводится к перечислению множества элементов, которые циклически переходят друг в друга, а количество этих элементов k определяет *длину (порядок) цикла.* Так,

$$\begin{pmatrix} 4\,2\,5\,1\,3\,6 \\ 5\,3\,1\,4\,2\,6 \end{pmatrix} = (1, 4, 5)(2, 3)(6).$$

Цикл длины 1 представляет собой тождественную подстановку и часто не записывается. Подстановка, все n элементов которой образуют цикл, называется *круговой* или *циклической.* Цикл длины 2 называют *транспозицией* (это подстановка, которая переставляет только два элемента). Всякая подстановка представляется произведением транспозиций, например:

$$\begin{pmatrix} 1\,2\,3\,4\,5 \\ 2\,5\,4\,1\,3 \end{pmatrix} = \begin{pmatrix} 1\,2\,3\,4\,5 \\ 2\,1\,3\,4\,5 \end{pmatrix} \cdot \begin{pmatrix} 2\,1\,3\,4\,5 \\ 2\,5\,3\,4\,1 \end{pmatrix} \cdot \begin{pmatrix} 2\,5\,3\,4\,1 \\ 2\,5\,4\,3\,1 \end{pmatrix} \cdot \begin{pmatrix} 2\,5\,4\,3\,1 \\ 2\,5\,4\,1\,3 \end{pmatrix} =$$

$= (1, 2)(1, 5)(3,4)(1,3).$

Заметим, что подобное разложение может содержать циклы с общими элементами и при этом оно не является единственным. В то же время разложеник подстановки на *независимые циклы* (без общих элементов) всегда можно осуществить только единственным способом.

Разность между числом всех элементов подстановки n и количеством ее циклов m (с учетом циклов длины 1) называется *декрементом* подстановки $d= n-m$. Четность подстановки совпадает с парностью ее декремента.

2.7. Функция

Рассмотрим некоторое отображение

$$f: X \to Y \qquad (2.17)$$

Это отображение называется *функцией,* если оно является однозначным, т.е. если для любых пар $(x_1, y_1) \in f$ и $(x_2, y_2) \in f$ из $x_2 = x_1$ следует $y_2 = y_1$.

Из определения отображения и из приведенных ранее примеров следует, что элементами множества X и Y могут быть объекты любой природы. **Однако в задачах теории оптимизации особый интерес представляют отображения, которые являются однозначными и множество значений которых представляет собой множество вещественных чисел *R*.** Однозначное отображение f, которое определяется (2.17) называется ***функцией с вещественными значениями***, если $Y \subseteq R$.

Напомним некоторые общие наиболее фундаментальные свойства функции, не касаясь свойств конкретных классов функций.

Пример 3. Из данного города в другой можно проехать по железной дороге, автобусом или самолетом. Стоимость билета будет соответственно 70, 90 и 120 грн. Стоимость билета в этом примере можно представить как функцию от вида транспорта. Для этого рассмотрим множества

$$X = \{\text{ж.д., авт., сам.}\}; \quad Y = \{70, 90, 120\}.$$

Функция $f : X \to Y$, получаемая из условий примера, может быть записана в виде множества

$$f = \{(\text{ж.д.}, 70), \ (\text{авт.}, 90), \ (\text{сам.}, 120)\}.$$

Значение y в любой из пар $(x, y) \in f$ называется функцией от данного x, которая записывается в виде $y=f(x)$.

Такая запись позволяет ввести следующее формальное определение функции:

$$f = \{(x, y) \in X \times Y \mid y = f(x)\}. \qquad (2.18)$$

Таким образом, символ f используется при определении функции в двух смыслах:

1) f является множеством, элементами которого являются пары (x, y), которые принимают участие в соответствии;

2) $f(x)$ является обозначением для $y \in Y$, соответсвующего данному $x \in X$.

Формальное определение функции в виде соотношения (2.18) позволяет установить способы задания функции.

1. Перечисление всех пар (x, y), составляющих множество f, как это было сделано в примере 3. Такой способ задания функции применим, если X является конечным множеством. Для большей наглядности пары (x, y) удобно располагать в виде таблицы.

2. Во многих случаях как X, так и Y представляют собой множества вещественных или комплексных чисел. В таких случаях очень часто под $f(x)$ понимают формулу, т.е. выражение, которое содержит перечень математических операций (сложение, вычитание, деление, логарифмирование и т.п.), которые нужно выполнить над $x \in X$, чтобы получить y.

Пример 4. Пусть

$$X = Y = R \quad и \quad f = \{(x, y) \in R^2 \mid y = x^2\}.$$

Тогда

$$f(x) = x^2$$

Иногда для различных подмножеств множества X функции приходится пользоваться различными формулами. Пусть $A_1, \ldots, A_n$ — попарно непересекающиеся подмножества X. Обозначим через $f_i(x)$,

i=1, 2,..., *n* формулу, которая определяет *y* при $x \in A_i$. Тогда функция *f(x)* будет определяться выражением

$$f(x) = \begin{cases} f_1(x) \; при \; x \in A_1; \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ f_n(x) \; при \; x \in A_n. \end{cases} \qquad (2.19)$$

Так, функцию $y=f(x) = |x|$ можно задать в виде

$$y = \begin{cases} x \; при \; x \geq 0; \\ -x \; при \; x < 0. \end{cases}$$

3. Если *X* и *Y*— множества вещественных чисел, то элементы $(x, y) \in f$ можно изобразить в виде точек на плоскости R^2. Полная совокупность таких точек будет представлять собой *график* функции *f(x).*

Если в выражении (2.17) $X=U \times V$, то приходим к функции от двух переменных *u* и *v*, обозначаемой через *f(u, v)*, где $u \in U$ и $v \in V$. Формальное определение функции двух вещественных переменных будет следующим :

$$f = \{(u, v, y) \in U \times V \times Y \mid y = f(u, v)\}. \qquad (2.20)$$

Рассмотрим определение функции с применением декартова произведения множеств.

Напомним, что декартовым произведением $M_a \times M_b$ множеств M_a и M_b называется множество M вида

$$M = \{(m_i, m_j) / m_i \in M_a, m_j \in M_b\}.$$

Подмножество $F \in M_x \times M_y$ называется *функцией,* если для каждого элемента *x*, $x \in M_x$, имеется не более одного элемента $y \in M_y$ вида $(x, y) \in F$; при этом если для каждого элемента *x* существует один элемент *y* вида $(x, y) \in F$, то функция называется *всюду (полностью) определенной,* в противном случае - *частично определенной (недоопределенной).* Множество M_x образует *область определения функции F,* множество M_y *— область значений функции F.* Часто вместо записи $(x, y) \in F$ используют запись $y=F(x)$; при этом элемент *x* называют *аргументом* или *независимой переменной,* а *y — значением функции F, или зависимой переменной.*

Сопоставим с декартовым произведением двух множеств прямоугольную решетку, узлы которой взаимно однозначно соответствуют элементам декартова произведения. Подмножество декартова произведения на рисунках будем отмечать штриховкой соответствующих элементов.

Пример 5. На рис. 2.3,*а* изображено подмножество декартова произведения множеств $M_x=\{x_1\ x_2,\ x_3,\ x_4\}$ и $M_y=\{y_1,\ y_2,\ y_3\}$, не являющееся функцией; на рис. 2.3,*б*, - являющееся полностью определенной функцией; на рис. 2.3, *в* — частично определенной функцией.

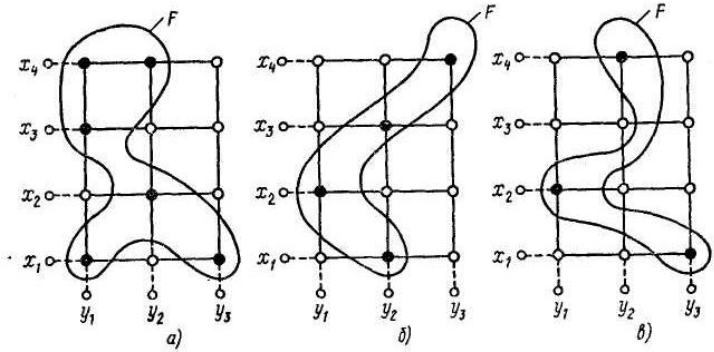


Рис. 2.3

Количество аргументов определяет *местность функции.* Выше были рассмотрены одноместные функции.

Аналогично понятию декартова произведения двух множеств определим декартово произведение *n* множеств.

Декартовым произведением

$$M_1 \times M_2 \times ... \times M_n = \prod_{i=2}^{n} M_i$$

множеств $M_1, M_2, ..., M_n$ называется множество

$$M = \{(m_{i_1}, m_{i_2}, ..., m_{i_n}) / m_{i_1} \in M_1, m_{i_2} \in M_2, ..., m_n \in M_{i_n}\}$$

Элементами декартова произведения $M_1 \times M_2 \times ... \times M_n$ являются всевозможные последовательности, каждая из которых состоит из *n* элементов, причем первый элемент принадлежит множеству M_1, второй - множеству M_2,..., *n*-й элемент — множеству M_n.

Если множество M_x в определении функции $y=F(x)$ является декартовым произведением множеств

$$M_{x_1} \times M_{x_2} \times ... \times M_{x_n},$$

то получаем определение *n*-местной *функции*

$$y = F(x_1, x_2, ..., x_n).$$

Частным случаем *n*-местной функции $y=F(x_1,x_2,...,x_n)$ является

n-местная операция. Под *n*-местной *операцией* O_n в множестве *M* понимается *n*-местная функция $y=F(x_1,x_2,...,x_n)$, в которой область определения аргументов и область значений функции совпадают:

$$M_{x_1} = M_{x_2} = ... = M_{x_n} = M_y.$$

Таким образом, *n*-местная операция по *n* элементам множества *M* определяет $(n+1)$-й элемент этого же множества.

Сужение и продолжение функции. Пусть функция $f:X \to Y$ определена на множестве *X*, а f_1 — на множестве $Q \subset X$, причем для каждого $x \in Q$ значение функций *f* и f_1 совпадают. Тогда f_1 называют *ограничением (сужением) функции* *f* на *Q*, а *f* — *продолжением функции* f_1 *на X*.

Например, функция $f(x)=x^3$ (другая запись $x \to x^3$), определенная на множестве действительных чисел *R*, отображает это множество на себя. Если ограничить область определения этой функции множеством целых чисел *Z*, то получим сужение $f_1(x)$ функции $f(x)$ на *Z*, причем $f_1(x)$ отображает множество *Z* в *Z* (**а не на** *Z*), так как не всякое число является кубом целого числа. Операцию сужения функции часто используют для табличной задачи функций с бесконечной областью определения *X*. В качестве множества *A* берут обычно выборку равнозначных значений *x* множества *X*. Получаемое при этом сужение f_A функцию *f* уже легко представить в виде таблицы. По этому принципу построены таблицы логарифмов, тригонометрических функций и другие. Функции *f* и *g* равны, если их область определения – то же самое множество *A* и для любого $a \in A$ $f(a)= g(a)$.

Пример 6. 1) Функция $f(x)=2^x$ является отображением *N* в *N* и *N* на M_{2^n}

2) Всякая нумерация счетного множества есть его отображением на *N*.

3) Функция $f(x)=\sqrt{x}$ не полностью определена, если ее тип $N \to N$, и полностью определена, если ее тип $N \to R$ или $R_+ \to R$ (R_+ положительное подмножество *R*).

4) Пусть зафиксирован список $\{a_1, ..., a_n\}$ всех элементов конечного множества *A*. Тогда любой вектор $v_i = (a_{i_1},..., a_{i_n})$ из A^n можно рассматривать как описание функции f_i: $A \to A$ (т.е. преобразование *A*), определяемой следующим образом: $f_i(a_j)=a_{ij}$, т.е. значение f_i для a_j равно *j*-й компоненте v_i. Число всех преобразований *A* равно, следовательно, $|A^n| = n^n$. Аналогично всякую функцию типа $N \to N$ можно представить бесконечной последовательностью

элементов *N*, т.е. натуральных чисел; отсюда нетрудно показать, что множество всех преобразований счетного множества континуально.

5) Каждое натуральное число *n* единственным образом разлагается на произведение простых чисел (простых делителей этого числа). Поэтому, если договориться располагать простые делители *n* в определенном порядке (например, в порядке неубывания), то получим функцию *q(n)* типа

$$N \to \bigcup_{i=1}^{\infty} N^i ,$$

которая отображает *N* в множество векторов произвольной длины. Например,

$$q(42)=(2, 3, 7),\ q(23)=23,\ q(100)=(2, 2, 5, 5).$$

Это отображение не является сюръективным, так как в область значений *q* не входят векторы, для компонентов которых не выполнено условие неубывания.

6) Каждому человеку соответствует множество его знакомых. Если зафиксировать момент времени (например, 10 января 2010 г., 5 ч. 00 мин), то это соответствие будет однозначным и является отображением множества *M* людей, которые живут в этот момент, в множество подмножеств *M*.

Пример 7. Функция sin *x* имеет тип $R \to R$. Отрезок $[-\pi/2, \pi/2]$ она взаимно-однозначно отображает на отрезок [—1, 1]. Поэтому на отрезке [—1, 1] для нее является обратная функция arcsin *x*.

Пример 8. 1) Функции sin *x* и $\sqrt{x}$ имеют тип $R \to R$, т.е. отображают одно и то же множество в себя. Поэтому их композиция возможна в произвольном порядке и дает функции $\sin\sqrt{x}$ и $\sqrt{\sin x}$. Заметим, что области определения их различны: первая функция определена на положительной полуоси; вторая функция определена на множестве отрезков $[2k\pi, (2k+1)\pi]$, где $k = 0, \pm1, \pm2$ ... Таким образом, область определения композиции может быть уже области определения обеих исходных функций и даже быть пустой.

2) Множество $K= \{k_l ..., k_m\}$ команд ЭВМ отображается в машинные коды этой ЭВМ, т.е. в натуральные числа. Кодирующая функция φ имеет тип $K \to N$. С помощью суперпозиции этой функции и арифметических функций оказываются возможными арифметические действия над командами (которые сами по себе числами не является), т.е. функции вида $\varphi(k_1)+\varphi(k_2)$, $\varphi(k_1)+4$ и т.д.

3) В функции $f_1(x_1, x_2, x_3) = x_1+2x_2+7x_3$ переименование x_3 в x_2, приводит к функции $f_1(x_1 x_2, x_2)=x_1+2x_2+7x_2$, что равно функции двух

аргументов $f_2(x_1, x_2)=x_1+9x_2$. Переименование x_1 и x_3 в x_2 приводит к одноместной функции $f_3(x_2)=10x_2$.

4) Элементарной функцией в математическом анализе называется всякая функция f, которая является суперпозицией фиксированного (т.е. не зависимого от значений аргументов f) числа арифметических функций, а также функций e^x, $\log x$, $\sin x$, $\arcsin x$. Например, функция $log^2(x_1+x_2)+3\sin\sin x_1+x_3$ элементарна, так как является результатом нескольких последовательных суперпозиций x_1+x_2, x^2, $\log x$, $3x$, $\sin x$.

5) Всякая непрерывная функция n переменных представима в виде суперпозиции непрерывных функций двух переменных.

Числовые функции. Проиллюстрируем введенные понятия на функциях, определенных на числовых множествах, элементами которых являются действительные числа. Такая функция каждому числу x из области определения ставит в соответствие число $y=f(x)$ из области ее значений. Иначе говоря, числовая функция f определяется множеством упорядоченных пар чисел (x, y).

Говоря геометрическим языком, множеству действительных чисел отвечает множество *точек прямой (числовой оси).* Пары чисел (x, y) представляются в декартовой системе координат *точками плоскости* с координатами $x \in X$ и $y \in Y$, причем первая координата x — *абсцисса,* а вторая y — *ордината* точки. Числовые оси, которые отвечают множествам X и Y, есть *осями координат,* а декарто произведение $X \times Y$, представляет собой множество точек плоскости. Таким образом, между элементами множества $X \times Y$ и точками плоскости устанавливается взаимно-однозначне соответствие.

Различные подмножества действительных чисел, на которых определяется функция, отвечают подмножествам точек прямой. В качестве таких подмножеств часто используют следующие:

отрезок (замкнутый интервал) $[a, b] = \{x \mid a \le x \le b\}$;

полуинтервал, открытый слева $(a, b] = \{x \mid a < x \le b\}$;

полуинтервал, открытый справа $[a, b) = \{x \mid a \le x < b\}$;

открытый интервал (или просто *интервал)* $(a, b) = \{x \mid a < x < b\}$.

Область определения функции может быть задана и отдельными точками числовой прямой. Множество точек плоскости, которая отвечает множеству упорядоченных пар $(x,y) \in f$, называется графиком функции f. На рис. 2.4 изображен график функции $y=f(x)$, определенной на множестве G с областью значений F.

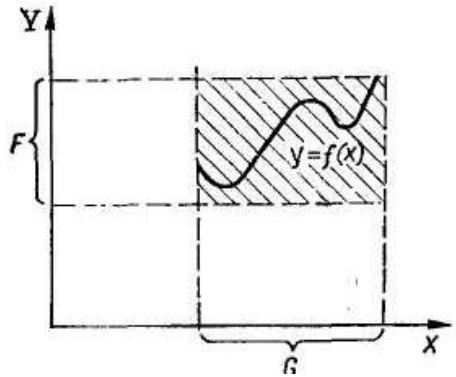


Рис. 2.4. График числовой функции *y=f(x)* (*G* — область определения; *F* — область значений).

В заключение отметим, что при более строгом рассмотрении между отображением и функцией все же имеется некоторое различие, характеризуемое способом определения этих отношений на множестве *X*. Функциональное отношение $A \subset X \times Y$ называют отношением множества *X* в *Y*, если это отношение всюду определено на *X*, т. е. его область определения $D_0(A)$ совпадает с множеством *X*.

Отношение $A \subset X \times Y$ называют функциональным, если все его элементы (упорядоченные пары) имеют различные первые координаты, т. е. каждому элементу $x \in X$, такому, что $(x, y) \in A$, соответствует один и только один элемент $y \in Y$. При этом первая координата *x* упорядоченной пары $(x, y) \in A$ является аргументом (переменной), а вторая *y* —образом (значением) функции.

Пример. Во множестве $N = \{1, 2, 3, 4, 5, 6\}$ заданы отношения:

$$\{(1, 3), (2, 4), (2, 6), (3, 5), (3, 2)\}, \quad (2.21)$$

$$\{(1, 6), (2, 2), (3, 5), (4, 5), (5, 6)\}. \quad (2.22)$$

Какие из этих отношений являются функциями и какие отображениями?

Решение. В выражениях (2.21) и (2.22) первое отношение является отображением, второе — функцией, так как для второго отношения все первые координаты отличны друг от друга, а для первого это условие не выполняется.

Рассмотрим пример конструирования печатной платы. Пусть *x* — некоторое исходное расположение конструктивных элементов на плате; *X* — множество различных расположений таких элементов на плате. Тогда Г*x* для любого $x \subset X$ —множество положений, которые можно получить из *x*, например с помощью парных перестановок конструктивных элементов, делая один шаг перестановок в на-

правлении улучшения некоторого показателя качества размещения. При этом $\Gamma^4 x$—множество перестановок конструктивных элементов, которые можно выполнить из состояния x четырьмя шагами; $\Gamma^{-1} x$ — множество положений (состояний) конструктивных элементов, из которых данное положение может быть получено за один шаг. Если из положения x перестановками с другими элементами не удается улучшить показатель качества размещения (достичь локальный оптимум показателя качества), то $\Gamma x = \varnothing$.

2.8. Обратная функция

Понятие обратной функции может быть применено для такого отображения $f:X \to Y$, которое, во-первых, является однозначным, т.е. для любых $(x_1, y_1) \in f$ и $(x_2, y_2) \in f$ из $x_2 = x_1$ следует $y_2 = y_1$ и, во-вторых, является взамно-однозначным, т.е. из $x_2 \neq x_1$ следует $y_2 \neq y_1$. При выполнении этих условий отображение $f:X \to Y$ является однозначным, т.е. определяет функцию $y=f(x)$. Обратное отображение $f^1:Y \to X$ также является однозначным и определяет функцию $y=f^1(x)$, которую называют обратной по отношению к функции $y=f(x)$. При аналитическом задании функции f принято аргумент как прямой, так и обратной функции обозначать одной и той же буквой, например, x. Поэтому для нахождения обратной функции нужно уравнение $y=f(x)$ решить относительно x и поменять обозначения, заменив x на y и y на x. При этом обратная функция запишется в виде $y=f^1(x)$.

Пусть заданы множества A, B и C и отношение σ между A и B и ρ между B и C. Определим отношение между A и C таким образом: оно действует из A в B с помощью σ, а потом из B в C с помощью ρ. Такое отношение называют *составным* и обозначают $\rho \circ \sigma$, т.е.

$$(\rho \circ \sigma)(a) = \rho(\sigma(a)).$$

Следовательно, $(x, y) \in (\rho \circ \sigma)$, если существует $z \in B$ такое, что $(x, z) \in \sigma$ и $(z, y) \in \rho$. Отсюда следует, что $G_{\rho^{\circ}\sigma} = \sigma^{-1} G_{\rho}$. Чтобы проиллюстрировать ситуацию, рассмотрим рис. 2.5.

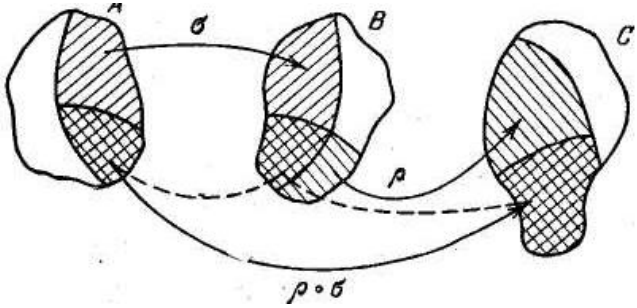


Рис. 2.5

Области определения и значений σ и ρ заштрихованы в разных направлениях. Следовательно, сегменты с двойной штриховкой на *A, B* и *C* представляют собой $G_{\rho\circ\sigma}$ $G_\rho \bigcap F_\rho$, $F_{\rho\circ\sigma}$ соответственно.

Замечание. Из записи отношений σ и ρ следует, что они применяются справа налево. Следовательно, $(\rho\circ\sigma)(a)$ означает, что вначале берется *a* и преобразуется посредством σ, а затем преобразуется посредством ρ. В алгебре это иногда записывают в виде $a\sigma\rho$. Следует обращать внимание при чтении других математических книг на то, какой порядок выполнения отношений принят в той книге.

Пример. Пусть σ и ρ — отношения на *N* такие, что

$$\sigma=\{(x, x+1): x\in N\},\quad \rho=\{(x^2, x): x\in N\}.$$

Тогда

$G_\rho = \{x^2: x\in N\}$, $G_\sigma =\{ x: x, x+1\in N = N$,

$G_{\rho\circ\sigma} = \sigma^{-1}G_\rho= \{x: x\in N$ и $x+1=y^2$, где $y\in N\} =\{3, 8, 15, 24, ...\}$ (рис. 2.6).

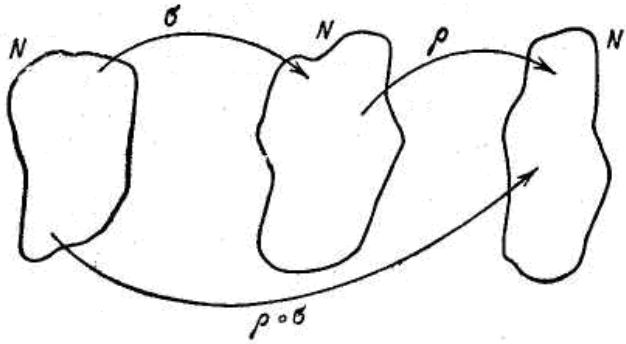


Рис. 2.6

Используя результаты, полученные выше, выполним исследование сложных функций. Пусть дана функция f: $A \to B$; в этом случае f^{-1} является функцией тогда и только тогда, когда f инъективна, а отображением тогда и только тогда, когда f биективна. В большинстве рассматриваемых нами случаев f — биекция; тогда f^{-1} — также биекция, а функции $f^{-1} \circ f$ и $f \circ f^{-1}$ являются тождественными отображениями.

Рассмотрим функции f: $A \to B$ и g: $B \to C$. Тогда:

а) если f и g инъективны, то существует $g \circ f$;

б) если f и g сюръективны, то также существует $g \circ f$.

Обратным отношением к $g \circ f$ есть $f^{-1} \circ g^{-1}$. Порядок должен быть обратным, как указано на рис. 2.7.

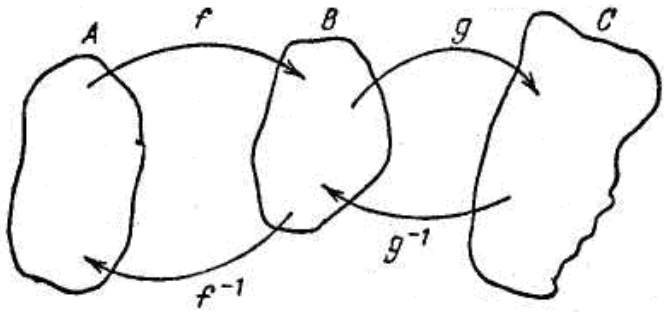


Рис. 2.7

Заметим, что если g — отображение, т.е. $G_g = B$, то $F_f \subseteq G_g$ и, следовательно, $G_{g \circ f} = F_g$. Аналогично, если $F_f \supseteq G_g$, то $F_{g \circ f} = F_g$. Если f и g инъективны, то существует $g \circ f$; следовательно, $f^{-1} \circ g^{-1}$ — функция.

Подытоживая вышесказанное, имеем: из $F_f = G_q$ следует, что $g \circ f: G_f \to F_q$ — отображение; если f и g также инъективны, то $f^{-1} \circ g^{-1}: F_f \to G_q$ — биекция. Очевидно, что эти критерии выполняются, если f и g — биекции.

2.9. Некоторые специальные классы функций

В этом разделе мы немного отойдем от основной темы обсуждения для того, чтобы коротко рассмотреть следующих три важных класса функций: *подстановки, последовательности, функционалы*.

Эти функции часто используются в методах оптимизации; особенно отметим их приложение к теории графов, к трассированию вычислений, к определению языков программирования и перевода, к машинной графике.

Начнем из подстановок и перестановок. Частично мы их уже рассматривали выше.

Понятие подстановок и последовательности

Определение. *Подстановкой* множества A называется биекция на A.

Подстановки конечных множеств представляют особый интерес в вычислениях. Когда A конечно, мы можем вычислить число разных подстановок A.

Пусть $|A|=n \in N$. Обозначим через ${}_nP_n$ число таких подстановок. Значение ${}_nP_n$ легко вычислить. Можно рассматривать задачу построения биекции на A как задачу заполнения ящиков, пронумерованных от 1 до n (рис. 2.8), объектами $a_1, ..., a_n$.

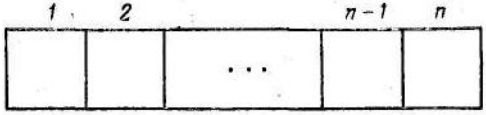


Рис. 2.8

Порядок, в котором заполняются ящики - несущественен (любой другой порядок можно получить перемешиванием ящиков). Поэтому будем заполнять их слева направо. Первый ящик может быть заполнен n способами, так как мы имеем свободный выбор из всего множества A. Убирая выбранный элемент из A, получим множество из n — 1 элементов. Следовательно, второй ящик может быть заполнен n — 1 способами, третий ящик — n — 2 способами и т.д. Продолжая этот процесс, получим, что (n — 1)-й ящик может быть заполнен двумя способами, а ящик с номером n — единственным оставшимся элементом из A. Следовательно, число различных подстановок из A равно

$$n \cdot (n — 1) \cdot (n — 2) \cdot ... \cdot 3 \cdot 2 \cdot 1.$$

Это произведение называют *факториалом n*. Следовательно, ${}_nP_n=n!$

Так как $A \sim N_n$, то можно свести наше рассмотрение к N_n. Любая подстановка на N_n должна определять образ каждого элемента в N_n (который, безусловно, должен быть единственным и отличным от других). Пусть ψ — подстановка на N_n. Тогда ψ можно определить как множество из n пар следующим образом:

$$\psi = \{(1, x_1), (2, x_2), ..., (n, x_n)\},$$

где

$$\{x_1 ..., x_n\} = N_n.$$

Не обязательно, конечно, должно быть $x_1=1$ и т.д. Можно также представить ψ следующим образом;

$$\psi = \begin{pmatrix} 1 & 2 & 3 & \dots & n \\ x_1 & x_2 & x_3 & \dots & x_n \end{pmatrix}$$

Пример. Пусть σ — подстановка на N_6 :

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 5 & 6 & 3 & 1 & 4 & 2 \end{pmatrix}$$

Тогда σ(1) = 5, σ (3)= 3 и т.д.

Достоинством этого обозначения - является простота, с которой могут быть вычислены сложные подстановки. Предположим, что ψ — подстановка на N_n, которая определена выше, а χ - другая подстановка на том же самом множестве. Тогда подстановка χ может быть записана как совокупность пар в порядке, определяемом $x_1, x_2, \dots, x_n$. Если две последовательности записать одну над другой (первая применяемая подстановка должна быть записана первой), то верхняя и нижняя строки дадут результирующую подстановку.

Пример. Пусть σ - подстановка из предыдущего примера и

$$\rho = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 2 & 6 & 1 & 4 & 5 \end{pmatrix}$$

Можно переписать ρ в виде

$$\rho = \begin{pmatrix} 5 & 6 & 3 & 1 & 4 & 2 \\ 4 & 5 & 6 & 3 & 1 & 2 \end{pmatrix}$$

Поэтому ρ◦σ может быть вычислено следующим образом:

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 5 & 6 & 3 & 1 & 4 & 2 \end{pmatrix}$$

$$\rho = \begin{pmatrix} 5 & 6 & 3 & 1 & 4 & 2 \\ 4 & 5 & 6 & 3 & 1 & 2 \end{pmatrix}$$

одинаковые

$$\rho \circ \sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 5 & 6 & 3 & 1 & 2 \end{pmatrix}$$

Следовательно, например,

ρ◦σ (2)(= ρ(σ(2))= ρ(6))=5 и т.д.

Отсюда следует, что представление обратной (конечной) подстановки выходит перестановкой строк, которые представляют исходную подстановку. Существует более простое определение, которое может употребляться непосредственно для некоторых простых подстановок и косвенно для всех конечных.

Определение. Пусть $A=\{a_1, ..., a_n\}$. Подстановку ρ называют *циклом (циклической подстановкой),* если

$$\rho = \begin{pmatrix} a_1 & a_2 & \dots & a_{n-1} & a_n \\ a_2 & a_3 & \dots & a_n & a_1 \end{pmatrix}.$$

Предположим, что $A \subseteq B$ и B конечно. Распространяя ρ на все B, можно определить подстановку σ так, что

$$\sigma : x \mapsto \begin{cases} \rho(x), & \text{если} \quad x \in A, \\ x, & \text{если} \quad x \in B \setminus A. \end{cases}$$

В этом случае σ ведет себя подобно ρ во всех случаях, когда элементы B не остаются на месте. Применение σ к A передвигает элементы по кругу циклическим образом, и, если известна область A, мы можем обозначить подстановку как $(a_1,a_2,...,a_n)$. Эта подстановка называется *циклом длины n.*

Пример. Рассмотрим опять подстановку

$$\rho = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 2 & 6 & 1 & 4 & 5 \end{pmatrix}$$

Подстановка является циклом длины 5 и может быть записана как

(1, 3, 6, 5, 4).

Не все подстановки являются циклами. Например, подстановка σ в рассмотренном ранее примере не является циклом. Напомним, что σ имела вид

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 5 & 6 & 3 & 1 & 4 & 2 \end{pmatrix}$$

Поэтому σ(1)=5, σ(5)=4, σ(4)=1, откуда следует, что σ содержит цикл (1, 5, 4). Начиная с 2, получаем другой цикл — (2, 6). Таким образом, имеем σ =(1, 5, 4)∘(2,6) и σ =(2, 6)∘(1, 5, 4).

В действительности каждая конечная подстановка может быть представлена как произведение циклов, при этом циклы могут располагаться в любом порядке. Из построения следует, что один элемент не может встретиться более чем в одном цикле, т.е. циклы *не пересекаются.*

Теорема. *Каждая подстановка* ρ *на конечном множестве A выражается в виде произведения непересекающихся циклов.*

Если все элементы $x \in N_n$ такие, что $\rho(x) \neq x$ (будем называть такие элементы *нестационарными),* содержатся в σ_1, то $\rho=\sigma_1$ — единственный цикл (который, естественно, не пересекается). В противном случае найдем следующий наименьший элемент $x_2 \in N_n$

такой, что $\rho(x_2)\neq x_2$ и x_2 не встречается в σ_1. Из x_2 строим множество различных степеней ρ:

$$\sigma_2=(x_2, \rho(x_2), \rho^2(x_2), \rho^3(x_2),\ldots,\rho^m(x_2))...$$

Это цикл длины не менее 2, и он не пересекается с σ_1. Если все нестационарные элементы исчерпаны, то $\rho=\sigma_1\circ\sigma_2=\sigma_2\circ\sigma_1$. Очевидно, что множество нестационарных элементов, которые не входят в эти циклы, можно уменьшить, и в конце концов придем к Ø. Следовательно, $\rho=\sigma_1\circ\sigma_2\circ\sigma_{3...}\circ\sigma_r$, для некоторого $r\in N$.

Рассмотрим теперь несколько другую ситуацию. Возьмем множества $A:|A|=n$ и $B\subseteq A$, $|B|=r\leq n$. Возникает вопрос: сколько биективных функций существует из A в B? Или, что эквивалентно, сколько существует инъективных отображений из B в A? Число перестановок (без повторений) из n элементов по r обозначается ${}_nP_r$, и вычисляется так же, как и ${}_nP_n$, за исключением того факта, когда процесс прекращается после заполнения r ящиков. Таким образом,

$${}_nP_r = n \bullet (n - 1)\bullet \ldots \bullet (n - r +1).$$

Легко видеть, что, продолжая процесс заполнения ящиков, оставшиеся n-r элементов можно разместить по последним n-r ящикам ${}_{n-r}P_{n-r}$ способами. Поэтому и

$${}_nP_r = \frac{{}_nP_n}{{}_{n-r}P_{n-r}} = \frac{n!}{(n-r)!}.$$

При вычислении ${}_nP_r$ мы находим число биективных функций из A в B. Подсчитаем число таких функций.

Определение. Пусть A — конечное множество и $B\subseteq A$, $|A|=n\geq r=|B|$. Множество B называется *сочетанием* (без повторений) из n элементов по r. Число таких сочетаниий обозначается через C^r_n.

Вычисление C^r_n производится следующим образом. Положим $|A|=n$. Возьмем произвольное подмножество $B\subseteq A$ такое, что $|B|=r$. Тогда B является образом подстановки из n элементов по r. Число инъективных функций на A, которые имеют B своим образом, является ${}_nP_r$. Если f является такой функцией и g — другая такая функция, которая имеет ту же самую область значений, то g связана с f соотношением $g=\varphi\circ f$, где φ — подстановка на B. Функции g и f определяют одну и ту же комбинацию, и в действительности число функций, которые определяют эту комбинацию, равно числу подстановок φ на B. Следовательно,

$${}_nP_r = C^r_n \bullet {}_rP_r$$

откуда

$$C^r_n = \frac{{}_nP_r}{{}_rP_r} = \frac{n!}{r!(n-r)!}.$$

Поскольку относительные дополнения единственны и $|A\backslash B|=n\text{-}r$, то отсюда следует, что $C^r_n = C^{n\text{-}r}_n$.

Вернемся теперь к математическим объектам, которые упоминались нами раньше, но которые не рассматривались как функции.

Определение. *Последовательностью* на множестве S называют отображение $N \to S$.

Если σ: $N \to S$ заданная последовательность и $\sigma(n)=s_n$, то, обычно, обозначают последовательность не σ, а (s_n) или $(s_1, s_2, .., s_n, ...)$. В этом случае s_n называют n-м членом последовательности.

Часто при изучении свойств последовательностей возникает понятие «расстояние» между соседними элементами последовательности (скажем, s_n и s_{n+1}) и между элементами s_n при $n \geq n_0$ (где n_0 — некоторый фиксированный элемент N) и фиксированным элементом из S_1.

Мы возвратимся к этим вопросам чуть позже, поскольку в данный момент у нас в общем случае нет понятия расстояния.

2.10. Понятие функционала

Говоря об отображении $f: X \to Y$ как о функции с вещественными значениями, мы не накладывали на характер элементов множества X каких-либо особых ограничений. В простейших задачах множество X, как и множество Y представляет собой множество вещественных чисел. В этом случае каждая пара $(x, y) \in f$ ставит в соответствие одному вещественному числу x другое вещественное число y. Однако важным в теории оптимизации есть случай, когда множество X представляет собой множество функций, а множество Y — множество вещественных чисел. Этот случай приводит к понятию функционала, подробное рассмотрение которого удобно провести на примере.

Представим себе некоторую линию $y=f(x)$, которая соединяет фиксированные точки A и B, как показано на рис. 2.9, по которой скатывается свободно движущийся шарик. Обозначим через t время, которое шарик затратит на перемещение из точка A в точку B. Это время зависит от характера линии AB, т.е. от вида функции $f(x)$. Если обозначить через $F(x)$ множество различных функций, которые изображают линию AB, а через T множество вещественных чисел t, определяющих время движения шарика, то зависимость времени движения от вида функции может быть записана как отображение

$$J: F(x) > T.$$

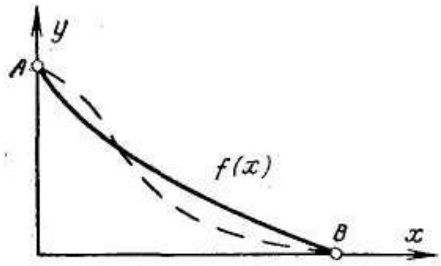


Рис. 2.9. Линия наискорейшего спуска.

Элементами множества J будут пары $(f(x), t)$), в которых $f(x) \in F(x)$, а $t \in T$. В этом случае говорят, что вещественное число $t \in T$ представляет собой функционала J от функции $f(x) \in F(x)$, и записывают это в виде

$$t = J[f(x)]$$

В оптимизационных задачах функционалы используются как параметры оптимизации. Так, в рассмотренном примере время перемещения шарика из точки A в точку B можно трактовать как параметр «оптимизации» избранной функции $f(x)$. При этом говорят об *оптимальном параметре* как о таком, при котором соответствующий параметр оборачивается в минимум. С этой точки зрения определения «оптимального» вида функции $f(x)$ сводится к выполнению условия

$$\min_{f \in F} J[f(x)],$$

при котором время t будет минимальным. Подобная линия наискорейшего спуска получила название ***брахистохроны***.

Обращение с функционалами не вызывает трудностей при условии, что ссылка делается на основной функционал (т. е. $A \rightarrow B$ или $A \rightarrow [B \rightarrow C]$). Следовательно, в дальнейшем мы будем рассматривать их просто как функции, имеющие нетривиальные области значений, и будем обращаться с ними соответствующим образом.

В заключение определим функции, которые сохраняют некоторые структуры. Из дальнейшего будет видно, что в некоторых ситуациях желательно сохранить многие из алгебраических свойств, которыми множества могут обладать. Ограничимся вначале рассмотрением простейшего случая.

Определение. Пусть X — множество, на котором задано отношение эквивалентности ρ. Тогда X *разбивается отношением* ρ *на* ρ *-эквивалентные классы;* множество классов обозначается как X/ρ.

Определение. Пусть X и Y— множества, ρ_X и ρ_Y — отношения эквивалентности на них, и пусть f: $X \to Y$—отображение. Обозначим через $\hat{f}$ *отношение*

$$\hat{f}: X/\rho_X \to Y/\rho_Y$$

такое, что

$$\hat{f} = \{([x], [f(x)]): x \in X\},$$

где $[x]$ — класс эквивалентности x. Если $\hat{f}$ — функция, то

$$x_1 \rho_X x_2 \Rightarrow \hat{f}([x_1]) = \hat{f}([x_2]),$$

и f является *отображением, сохраняющим эквивалентность.* В этом случае говорят, что f: $X \to Y$ *индуцирует* отображение

$$\hat{f}: X/\rho_X \to Y/\rho_Y. \ //$$

Наглядный способ представления такого отображения дан на рис. 2.10.

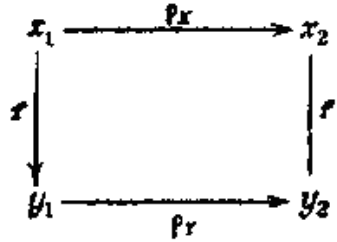


Рис. 2.10

Если рассмотреть отображение f, согласованное с отношением эквивалентности, то можно переходить от x_1 к y_1 или через x_2, используя соотношения $y_2=f(x_2)$ и $x_2 \rho_X x_1$, или через y_1, используя соотношения $y_1=f(x_1)$ и $y_2 \rho_Y y_1$.

Пример. Пусть $X = \{1, 2, 3\}$, $Y = \{1, 4, 9\}$, и пусть ρ_X и ρ_Y таковы, что

$$X/\rho_X = \{\{1\}, \{2, 3\}\}, \quad Y/\rho_Y = \{\{1\}, \{4, 9\}\},$$

и f: $X \to Y$ такое, что $x \mapsto x^2$. Тогда

$$\hat{f}([1]) = [f(1)] = [1] = \{1\},$$
$$\hat{f}([2]) = [4] = \{4, 9\},$$
$$\hat{f}([3]) = [9] = \{4, 9\}.$$

В этом случае $\{2, 3\} \in X/\rho_X \Rightarrow 2 \rho_X 3 \Rightarrow [2] = [3]$ и $\hat{f}([2]) = \hat{f}([3])$.

Поэтому $\hat{f}$ является функцией и f сохраняет отношения эквивалентности.

Пример. Пусть X, Y и f те же, что и раньше, и отношение эквивалентности σ_X и σ_Y индуцируют разбиения {{1}, {2, 3}} и {{1, 4}, {9}} соответственно. В этом случае индуцированные отношения дают

$$\hat{f}([2]) = [f(2)] = [4] = \{1, 4\},$$
$$\hat{f}([3]) = [f(3)] = [9] = \{9\}.$$

Так как $2\sigma_X 3$, то [2] = [3] в X/σ_X, но $(4, 9) \notin \sigma_Y$, поскольку $[4] \neq [9]$ в Y/σ_Y. По сравнению с рис. 2.10 этот пример дает отношения, показанные на рис. 2.11.

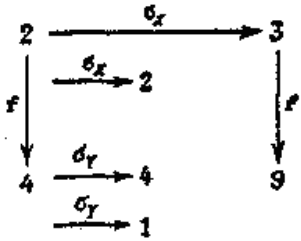


Рис. 2.11

Так как нельзя соединить стороны прямоугольника во всех случаях, то отношения эквивалентности не сохраняются. Эти диаграммы могут быть использованы для определения операций таким образом, чтобы соединить углы прямоугольника. После этого можно будет объединять диаграммы подобно строительным блокам.

2.11. Функция времени

В основе понятия функции времени лежит множество $T \subseteq R$ с элементами t, которое называют множеством моментов времени. ***Время обладает той характерной особенностью, что имеет направление***. Это означает, что если $t_1, t_2 \in T$ и $t_1 < t_2$, то момент t_1 предшествует моменту t_2. Другими словами T — **упорядоченное множество.**

Функция времени определяет отображение f множества моментов времени T на множество вещественных чисел R:

$$f: T \to R. \qquad (2.23)$$

Элементами f будут пары (t, x), которые обозначаются также через $x(t)$, где $t \in T$, $x \in R$. Каждая такая пара определяет значение функции в момент t и называется ***событием*** или ***мгновенным значением функции***. Полная совокупность пар (t, x), т.е. значений $x(t)$ для всех

$t \in T$, и представляет собой функцию времени. Дальнейшее уточнение функции времени связано с уточнением ее области определения, т.е. вида множества T.

Если $T=R$, т.е. t может принимать любое вещественное значение от $-\infty$ до $+\infty$, то функция $x(t)$ называется функцией с *непрерывным временем.* Примером может служить синусоидальная функция времени $x(t)=A\sin(\omega t+\varphi)$, описывающая напряжение в сети переменного тока.

Однако нас обычно не интересуют весьма удаленные моменты времени как в прошлому, так и в будущем. Поэтому производят сужение $x(t)$ на ограниченный интервал $t_1<t\leq t_2$, который обычно считают полузакрытым интервалом и обозначают $(t_1,t_2]$. Полузакрытые интервалы времени удобны тем, что допускают последовательное сочленение друг с другом. Так, если интервал $(t_1,t_2]$ разбить моментом t' на два интервала $(t_1, t']$ и $(t', t_2]$, то не будет сомнений, к какому интервалу отнести t'.

Сужение функции $x(t)$, заданной на интервале $-\infty < t < +\infty$, на интервал $(t_1, t_2]$ называется *отрезком функции x(t)* и обозначается $x_{(t_1,t_1]}$. Итак, по определению

$$x_{(t_1,t_1]} = \{(t) | t \in (t_1, t_2]\} \qquad (2.24)$$

Для осуществления операции сужения часто используют специальную функцию времени, которую называют *единичной функцией* или единичным скачком:

$$1(t-\lambda) = \begin{cases} 0 & при \quad t \leq \lambda; \\ 1 & при \quad t > \lambda, \end{cases} \qquad (2.25)$$

приведенную на рис. 2.12,*а*.

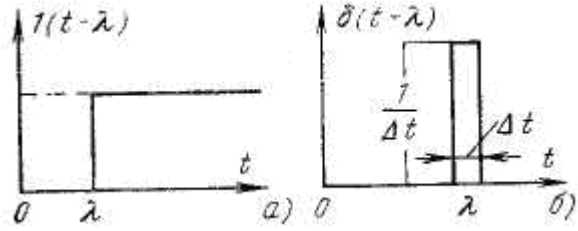


Рис. 2.12. Единичный скачок и импульсная функция .

Так, напряжение, которое подается на вход прибора, который подключается к сети в момент $t=\lambda$, будет равно:

$$u(t)=\mathbf{1}(t-\lambda)x(t)=\mathbf{1}(t-\lambda)A\sin(\omega t+\varphi).$$

Другой широко используемой функцией времени является импульсная функция $\delta(t-\lambda)$, определяемая соотношениями:

$$\delta(t-\lambda)=\begin{cases}0 & при \quad t\neq\lambda;\\ \infty & при \quad t=\lambda,\end{cases} \tag{2.26}$$

$$\int_{\lambda-\varepsilon}^{\lambda+\varepsilon}\delta(t-\lambda)\,dt=1, \quad \varepsilon>0. \tag{2.27}$$

Функцию $\delta(t-\lambda)$ можно рассматривать как предельный случай приведенного на рис. 2.12,*б* прямоугольного импульса шириной Δt и высотой $1/\Delta t$, появляющегося в момент $t=\lambda$ при $\Delta t\to 0$.

Импульсная функция позволяет выделять мгновенные значения функции $x(t)$ для фиксированных моментов времени. Так, если $t_1<\lambda<t_2$, то

$$\int_{t_1}^{t_2}x(t)\delta(t-\lambda)\,dt=x(\lambda)\int_{\lambda-\varepsilon}^{\lambda+\varepsilon}\delta(t-\lambda)\,dt=x(\lambda). \tag{2.28}$$

Если множество T представляет собой множество натуральных чисел

$$\ldots\ , -2, -1, 0,\ 1, 2,\ ,.,\ ,\ n,\ \ldots\ ,$$

то говорят о функциях *с дискретным временем.* В этом случае элементы множества T обозначают через n, так что пара (n, x), которая обозначается также $x[n]$ или x_n, определяет значение функции в момент n. На рис. 2.13 приведен пример функции с дискретным временем.

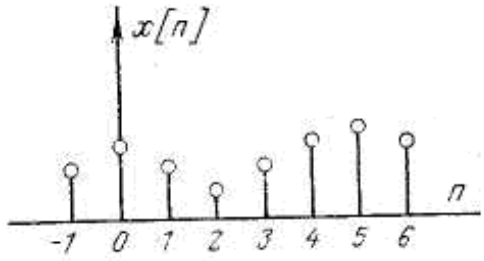


Рис. 2.13. Функция с дискретным временем.

2.12. Понятие оператора

Оператором L называется отображение

$$L:X\rightarrow Y, \qquad (2.29)$$

в котором множества X и Y являются множествами функций с элементами $x(t)$ и $y(t)$, так что элементами множества L будут пары $(x(t)$ и $y(t))$. В этом случае говорят, что оператор L преобразует функцию $x(t)$ в функцию $y(t)$, и пишут:

$$y(t)= L\,[x(t)].$$

Рис. 2.14. Представление управляющей системы в виде оператора.

Примером оператора служит оператор дифференцирования p, который ставит в соответствие функции $f(x)$ другую функцию $f'(x)=df(x)/dx$, что может быть записано в виде

$$f'(x)=p\,[f(x)].$$

В задачах управления роль оператора часто выполняет сама управляющая система, которая преобразует по некоторому закону L входной сигнал $x(t)$ в выходной сигнал $y(t)$, как это показано на рис. 2.14.

2.13. Аналитические свойства вещественных функций

Здесь содержится материал, использующий теорию множеств. Цель, которая при этом преследуется, состоит не в развитии техники вычислений, а в создании строгих утверждений типа:

«Предел $f(x)$ при x, стремящемся к 0, есть y»,

«Наклон графика f в точке a равен b»,

«f имеет гладкий график» и т. п.

(Два последних понятия относятся к графике.) Мы приведем основные определения, которые будут использоваться при получении некоторых результатов в теории оптимизации. Этого достаточно для того, чтобы проиллюстрировать доказательства большинства теорем.

Определение. *Вещественной последовательностью* называется отображение N на R.

Последовательность записывают в виде (a_n). Если при возрастании n члены a_n становятся «близкими» к некоторому фиксированному значению $a \in \mathrm{R}$, то говорят, что последовательность (a_n) имеет предел a или что a_n стремится к a при стремлении n к бесконечности. Дадим строгое определение сказанному.

Определение. Если (a_n) - вещественная последовательность и для любого $\varepsilon>0$ существует $N_\varepsilon \in \mathrm{N}$ такое, что $N > N_\varepsilon \Rightarrow |a_N - a| < \varepsilon$. то говорят, что (a_n) имеет *предел a,* и записывают это как

$$\lim_{n\to\infty} a_n = a$$

или $a_n \to a$ при $n\to\infty$. (Здесь $|x|$ обозначает модуль числа $x \in R$).

Если (a_n) имеет предел, то говорят, что последовательность *сходится.* Если последовательность не имеет предела, то говорят, что она *расходится.*

Пример.

1. Последовательность (a_n), где $a_n = 1/n$, имеет предел 0; для $\varepsilon > 0$ можно выбрать N_ε — любое натуральное число, большее $1/\varepsilon$. Тогда

$$N > N_\varepsilon \Rightarrow |a_N - 0| = 1/N < 1/N_\varepsilon < \varepsilon;$$

следовательно,

$$\lim_{n\to\infty} \frac{1}{n} = 0,$$

2. Последовательность (a_n), где $a_n = (-1)^n$, расходящаяся.

Предложение. *Если* (s_n) *и* (t_n) *—последовательности и* $\lambda \in \mathrm{R}$, *тогда* $(s_n + t_n)$, $(s_n t_n)$ *и* (λs_n) *также являются последовательностями, и если* $\lim_{n\to\infty} s_n = s$ *и* $\lim_{n\to\infty} t_n = t$,

то:

а) $\lim_{n\to\infty} (s_n + t_n) = s + t;$

б) $\lim_{n\to\infty} (s_n t_n) = st;$

в) $\lim_{n\to\infty} (\lambda s_n) = \lambda s;$

г) *если* $\varepsilon \neq 0$, *то* $\varepsilon_n/t_n \to s/t$ *при* $n \to\infty$.

Доказательство. Пусть $\varepsilon > 0$. Тогда существует $N_\varepsilon \in \mathrm{N}$ такое, что

$$|s_N - s| < \varepsilon/2 \text{ и } |t_N - t| < \varepsilon/2$$

при $N > N_\varepsilon$. Так как при $N > N_\varepsilon$,

$$|s_N + t_N - (s + t)| = |s_N - s + t_N - t| \leqslant$$
$$\leqslant |s_N - s| + |t_N - t| < \varepsilon,$$

то

$$\lim_{n\to\infty}(s_n + t_n) = s + t.$$

Аналогично для случая б)

$$|s_N t_N - st| = |s_N t_N - s_N t + s_N t - st| \leqslant$$
$$\leqslant |s_N t_N - s_N t| + |s_N t - st| \leqslant |s_N|\,|t_N - t| + |s_N - s|\,|t|.$$

Пусть задано $\varepsilon > 0$. Тогда существует $N_\varepsilon \in \mathrm{N}$ такое, что для $N > N_\varepsilon$ справедливы неравенства

$$|s_N - s| < \frac{1}{2}\frac{\varepsilon}{|t|+1},$$

$$|t_N - t| < \frac{1}{2}\frac{\varepsilon}{|s|+1},\quad |s_N| < |s| + 1.$$

Следовательно,

$$|s_N|\,|t_N - t| + |s_N - s|\,|t| \leqslant (|s| + 1)\,|t_N - t| +$$
$$+ |s_N - s|\,|t| < \frac{1}{2}\varepsilon + \frac{1}{2}\frac{\varepsilon}{|t|+1}|t| < \frac{1}{2}\varepsilon + \frac{1}{2}\varepsilon = \varepsilon,$$

откуда получаем $(s_n t_n) \to st$.

Определение. Пусть (a_n) — последовательность в R. Последовательность

$$s_n = \sum_{i=1}^{n} a_i$$

определяет *ряд* $\sum a_n$. При этом s_n называют *n-й частной суммой ряда.* Если последовательность (s_n) сходится, то говорят, что ряд *сходящийся,* и число

$$\lim_{n\to\infty} s_n$$

называют *суммой ряда.* Оно обозначается

$$\sum_{n=1}^{\infty} a_n.$$

2.14. Операции

Определение. *Операцией над множеством S* называется функция $f: S^n \to S$, $n \in \mathrm{N}$.

В этом определении есть два важных момента, которые заслуживают особого вспоминания. Во-первых, раз операция является функцией, то результат применения операции *однозначно определен.* Поэтому данный упорядоченный набор из n элементов S функция f

переводит только в один элемент *S*. Во-вторых, поскольку область значений операции лежит в *S,* на которое операция действует, будем говорить, что операция *замкнута* на *S,*

Говорят, что операция $S^n \to S$ *имеет порядок n.* Ограничимся рассмотрением ситуаций, когда порядок равен 1 или 2. В этом случае операции называют *монадическими* (или *унарными)* и *диадическими* (или *бинарными)* соответственно. Элементы набора из *n* элементов в области определения называют операндами. Операции обычно обозначают символами, которые называют операторами. В случае унарных операций обычно символ оператора ставят перед операндом.

Наиболее простым примером является операция изменения знака на *R*. В предположении, что операция сложения уже определена, *-x* определяет операцию $x \mapsto$ *y: x+y* =0 *(x* отображается в *y: x+y*=0).

Определение. Бинарные операции обозначают одним из трех способов. В первом случае оператор ставится между операндами *(infix),* во второму — перед операндами *(prefix)* и в третьему — после операндов *(postfix).*

Пример.

a+b infix,
+ab prefix,
ab+ postfix.

Переход от одной формы к другой нетруден и лучше всего описывается в терминах ориентированных графов.

В соответствии с большинством математических традиций, кроме некоторых работ по алгебре и формальной логике, мы будем использовать обозначение *infix.* Другие обозначения имеют то преимущество, что не требуют скобок при определении порядка вычислений сложных выражений, и это делает их особенно удобными для автоматической обработки. Можно проверить соответствие между следующими парами выражений, записанными в формах *infix* и *postfix* соответственно:

а) *a+b• c+(d+e• (f+g),*
abc• + defg+• ++;
б) *(a+b)• c+d+e• f+g,*
ab+c• d+ef• +g+;
в) *a+(b• (c+d)+e)• f+g,*
abed+ e+f +g+.

Пример. Рассмотрим алгебраическое выражение

$$a + b \bullet c + (d + e \bullet (f + g))$$

и его представление на рис. 2.15, которое называют деревом.

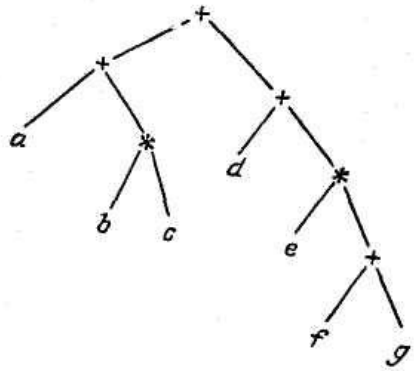


Рис. 2.15

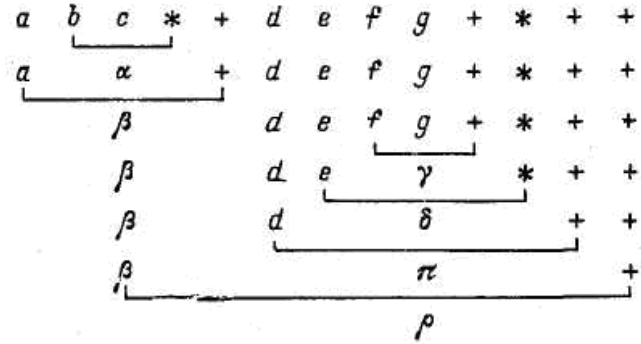
Из свойств арифметических операций мы знаем, что значение этого выражения можно вычислить многими способами. Однако если двигаться слева направо и снизу вверх, то получаем

$$\alpha \leftarrow b \bullet c,\ \beta \leftarrow a + \alpha,\ \gamma \leftarrow f + g,$$
$$\delta \leftarrow e \bullet \gamma,\ \pi \leftarrow d + \delta,\ \rho \leftarrow \beta + \pi.$$

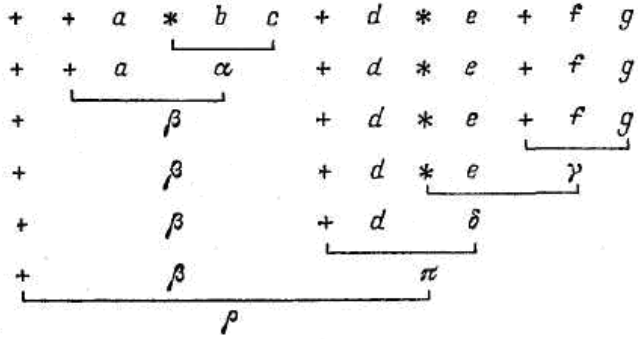
Здесь греческими буквами обозначаются промежуточные результаты, за исключением ρ - искомого результата.

Вычисление значения этого выражения с помощью дерева выполняется очень просто, однако если работать непосредственно с исходным выражением, то это можно сделать по-иному. Действительно, обычно *(infix)* выражение, как это показано в примере, нерегулярно потому, что некоторые подвыражения заключены в скобки, а некоторые нет. Особенно такая ситуация будет наблюдаться в том случае, если проинтегрировать информацию о разных символах на дереве (поскольку на самом деле его нет). Очевидно, что формы записи *prefix* и *postfix* этого выражения несут больше информации.

Вычисление значения выражения в форме *postfix* осуществляется следующим образом:



Аналогично в форме *prefix* вычисления осуществляются следующим образом:



«Переходы» по дереву показаны на рис. 2.16, *а* (форма *prefix)* на рис.2.16,*b* (форма *postfix)* и на рис.2.16, *с* (форма *infix)* со скобками:

((a + (b• c)) + (d + (e• (g + g)))).

Мы уже знакомы со многими бинарными операциями, например с арифметическими операциями +, •, —,/ и операциями над множествами — объединением ($\bigcup$) и пересечением ($\bigcap$).

Операции, которые определены на конечных множествах, часто удобнее задавать с помощью таблиц.

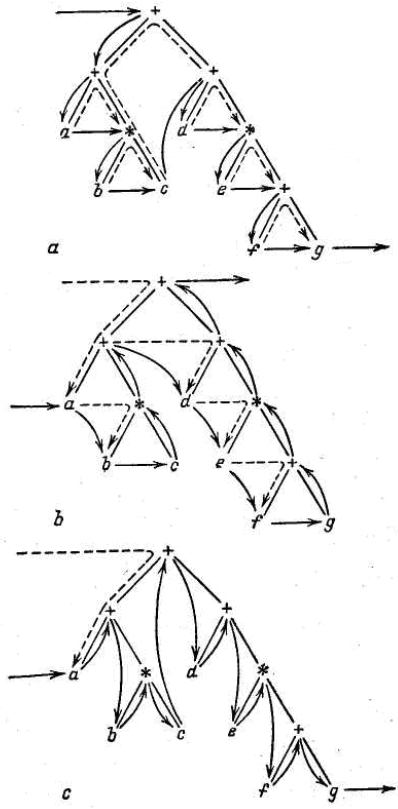


Рис.2.16.

Пример. Пусть операция $\otimes$ определена на множестве $\{a, b, c\}$ с помощью таблицы

$\otimes$	a	b	c
a	a	a	b
b	b	a	c
c	a	b	b

Следовательно,

$a \otimes b = a$,

$b \otimes b = a$,

$c \otimes b = b$, ...

Такие символы, как $\otimes$ и $\oplus$, будут использоваться для обозначения различных операций, которые будут вводиться в процессе изложения материала.

Очевидно, что использование таблиц имеет важное значение, так как некоторые операции, с которыми приходится иметь дело в теории оптитмизации, непригодны для словесного задания.

Обратим теперь внимание на свойства операций. Операции вместе со своими следствиями обеспечивают основу всех математических вопросов оптимизации, так как они определяют порядок работы с оптимизационными объектами.

Определение. Говорят, что бинарная операция $\otimes$ на множестве *A коммутативна,* если

$$a \otimes b = b \otimes a \quad \text{для всех} \quad a, b \in A.$$

Следовательно, обычная операция сложения на Z коммутативна, а вычитания — нет.

Определение. Говорят, что операция $\otimes$ на множестве *A ассоциативна,* если

$$(a \otimes b) \otimes c = a \otimes (b \otimes c) \quad \text{для всех} \quad a,b,c \in A.$$

Заметим, что в определении ассоциативности порядок операндов *a, b* и *c* сохранен (операция может быть некоммутативной!) и использованы круглые скобки, чтобы определить порядок вычислений.

Таким образом, выражение $(a \otimes b) \otimes c$ требует, чтобы сначала вычислялось $a \otimes b$ и результат этого (скажем, *x)* принимал участие в операции с *c*, т.е. давал $x \otimes c$. Если операция ассоциативна, то порядок вычислений несуществен и, следовательно, скобки не требуются.

Пример. Над Z имеем

$$(1+2)+3 = 1+2+3 = 1+(2+3),$$

но

$$(1-2)-3 = -4 \quad \text{и} \quad 1-(2-3) = 2.$$

Таким образом, операция вычитания не ассоциативна.

Коммутативность и ассоциативность являются двумя важными свойствами, которые могут быть определены для простых операций. Перед тем как описывать свойства, которые связывают две операции, определим некоторые термины, относящиеся к специальным элементам множеств, к которым эти операции применяются.

Определение. Пусть $\otimes$ — бинарная операция на множестве *A* и $l \in A$ такая, что

$$l \otimes a = a \quad \text{для всех} \quad a \in A.$$

Тогда l называется *левой единицей* относительно $\otimes$ на A. Аналогично, если существует $r \in A$ такое, что

$$r \otimes a = a \text{ для всех } a \in A,$$

то r является *правой единицей* относительно $\otimes$. Далее, если существует элемент e, который является и левой, и правой единицей, т.е.

$$e \otimes a = a \otimes e = a \text{ для всех } a \in A,$$

то e называется *(двусторонней) единицей* по отношению к $\otimes$.

Пример. Над R 0 является правой единицей по отношению к вычитанию и единицей по отношению к сложению, так как

$$a - 0 = a,$$

но

$$0 - a \neq a, \text{ если } a \neq 0;$$
$$a + 0 = a \text{ и } 0 + a = a \text{ для всех } a.$$

Определение. Пусть $\otimes$ — операция на A с единицей e и $x \otimes y = c$. Тогда говорят, что x — *левый обратный* элемент к y, а y — *правый обратный* элемент к x. Дальше, если x и y такие, что

$$x \otimes y = e = y \otimes x,$$

то y называется *обратным элементом* к x по отношению к $\otimes$, и наоборот.

Замечание. В некоторых книгах левые (правые) обратные элементы относят к левой (правой) единицы, однако, в большинстве случаев единицы являются двусторонними и, следовательно, не требуется делать никаких различий. Для решения уравнений необходимо существование и единственность единиц и обратных элементов. Менее общим свойством операций является идемпотентнисть, хотя оно используется в алгебре логики.

Определение. Пусть операция $\otimes$ на множестве A и произвольный элемент $x \in A$ таковы, что $x \otimes x = x$. Тогда говорят, что x *идемпотентен* по отношению к $\otimes$.

Очевидно, что любое подмножество идемпотентно по отношению к операциям пересечения и объединения.

Определение. Пусть дано множество A, на котором определены две операции $\otimes$ и $\oplus$. Тогда, если

$$a \otimes (b \oplus c) = (a \otimes b) \oplus (a \otimes c) \text{ для всех } a,b,c \in A,$$

то говорят, что $\otimes$ *дистрибутивна* по отношению к $\oplus$.

Если сказанное выше не совсем понятно, следует провести соответствие между этим тождеством и обычной арифметикой на R, например,

$3*(1 + 2) = (3*1)+(3*2)$.

Наиболее общеизвестная алгебра может быть построена из относительно небольшого набора основных правил. Сейчас мы продемонстрируем, как из элементарных предположений можно извлечь некоторые простые следствия.

Пример. Пусть $\otimes$ — операция на множестве A и существует единица по отношению к $\otimes$. Тогда единичный элемент единствен.

Доказательство. Предположим, что x и y — единицы по отношению к $\otimes$, т.е.

$$x \otimes a = a \otimes x = a,$$

$$y \otimes a = a \otimes y = a \quad \text{для всех } a \in A.$$

Тогда $x = x \otimes y$, так как y — единица, и $x \otimes y = y$, поскольку x — единица. Следовательно, $x = y$.

Пример 45. Пусть $\otimes$ — ассоциативная операция на множестве A и e — единица по отношению к $\otimes$. Тогда если $a \in A$ и x имеет обратный, то обратный элемент единствен по отношению к $\otimes$.

Доказательство. Допустим, что x' и x'' - обратные элементы к x, так что

$$x \otimes x' = x' \otimes x = e \text{ и } x \otimes x'' = x'' \otimes x = e.$$

Тогда

$$x' = x' \otimes e = x' \otimes (x \otimes x'') = (x' \otimes x) \otimes x'' = e \otimes x'' = x''.$$

3. Производная и дифференциал

3.1. Производная функция

Скорость изменения функции. Производная функция. Производная степенной функции.

I. Скорость изменения функции. Каждое из четырех специальных понятий: скорость движения, плотность, теплоемкость, скорость химической реакции, несмотря на существенное различие их физического смысла, является с математической точки зрения, как легко заметить, одной и той же характеристикой соответствующей функции. Все они представляют собой частные виды так называемой скорости изменения функции, определяемой, так же как и перечисленные специальные понятия, с помощью понятия предела.

Определение. *Отношение* $\frac{\Delta y}{\Delta x}$ *называется средней скоростью* $v_{ср}$ *изменения функции винтервале (x, x+Δx).*

Ясно, что чем меньше рассматриваемый интервал, тем лучше средняя скорость характеризует изменение функции, поэтому мы заставляем Δx стремиться к нулю. Если при этом существует предел средней скорости, то он принимается в качестве меры, измеряющей скорость изменения функции при данном *x,* и называется скоростью изменения функции.

Определение. *Скоростью изменения функции в данной точке x называется предел средней скорости изменения функции в интервале (x, x+Δx) при стремлении Δx к нулю:*

$$v = \lim_{\Delta x \to 0} v_{ср} = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}.$$

II. Производная функция. Скорость изменения функции *y=f(x)* определяется посредством такой последовательности действий:

1) по приращению Δx придаваемому данному значению *x,* находится соответствующее приращение функции

$$\Delta y = f(x+\Delta x) - f(x);$$

2) составляется отношение $\frac{\Delta y}{\Delta x}$;

3) находится предел этого отношения (если он существует) при произвольном стремлении Δx к нулю.

Определение. *Производной данной функции называется предел отношения приращения функции к приращению независимой переменной при произвольном стремлении этого приращения к нулю:*

$$f'(x) = \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}.$$

Значение производной функции в какой-либо данной точке x_0 обозначается обычно $f'(x_0)$ или $y'_{x=x_0}$.

Пользуясь введенным определением производной, можно сказать, что:

1) Скорость прямолинейного движения есть производная от функции $s = F(t)$ по t (производная от пути по времени).

2) Линейная плотность есть производная от функции $m = \Phi(s)$ по s (производная от массы по длине).

3.2. Дифференцирование функций

Дифференцирование результатов арифметических действий. Действие отыскания производной называется дифференцированием «Продифференцировать функцию *f(x)*» — значит найти ее производную *f ' (x)*. Для этого нужно найти

$$\lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}.$$

Такое непосредственное отыскание предела в большинстве случаев представляет собой весьма громоздкое и трудное действие. Но если знать — раз и навсегда — производные всех основных элементарных функций, а также правила, по которым следует дифференцировать сложные функции, и результаты арифметических действий, то можно находить производные любых элементарных функций, не выполняя всякий раз указанного предельного перехода.

Выведем прежде всего правила дифференцирования арифметических действий. Мы будем предполагать, что функции-компоненты, т. е. слагаемые, сомножители, делимое и делитель, непрерывны и имеют производные при рассматриваемых значениях независимой переменной. Из приводимых ниже теорем тогда вытекает, что функции-результаты, т. е. сумма, произведение, частное, также имеют производные при тех же значениях независимой переменной. Имея в виду это обстоятельство, мы не будем всякий раз на него указывать, чтобы не загромождать формулировки теорем, а будем включать в эти формулировки только вопрос о том, как найти производную результата, если известны производные компонент.

Теорема I. *Производная суммы конечного числа функций равна сумме производных слагаемых.*

Теорема II. *Производная произведения двух функций равна сумме произведений производной первой функции на вторую и производной второй на первую.*

Правило дифференцирования произведения двух функций последовательно распространяется на произведение какого угодно конечного числа функций. Так, если $y = uvw$, то

$$y' = (uvw)' = (uv)'w + uvw' = (u'v + uv')w + uvw' = u'vw + uv'w + uvw',$$

т. е. производная произведения трех функций равна сумме трех слагаемых, причем каждое из них является произведением двух из данных функций на производную третьей.

Теорема III. *Производная частного двух функций равна дроби, знаменатель которой равен квадрату делителя, а числитель— разности между произведением производной делимого на делитель и произведением делимого на производную делителя.*

Дифференцирование сложной и обратной функций.

I. Дифференцирование сложной функции. Очень важным является правило дифференцирования сложной функции, указывающее выражение для ее производной через производные функций, из которых она составлена.

Теорема. *Производная сложной функции равна производной заданной функции по промежуточному аргументу, умноженной на производную этого аргумента по независимой переменной*.

Итак, для того чтобы продифференцировать сложную функцию $y=f[\varphi(x)]$, нужно взять производную от «внешней» функции f, рассматривая ее аргумент $\varphi(x)=u$ просто как переменную, по которой совершается дифференцирование, и умножить на производную от «внутренней» функции $\varphi(x)$ по независимой переменной.

Аналогично выводится формула при любом числе промежуточных аргументов. Во всех случаях производная y' получается как произведение производной по первому промежуточному аргументу на производную от первого по второму, на производную от второго по третьему и т. д., наконец, на производную последнего (считая от y к x) промежуточного аргумента по x.

Коротко можно сказать так:

Производная сложной функции равна произведению производных от функций, ее составляющих.

II. Дифференцирование обратной функции. Пусть $y=f(x)$ и $x=\varphi(y)$ — пара взаимно обратных функций. Покажем, что если известна производная от одной из этих функций, то легко получить выражение для производной другой. Допусгнм, например, что нам известна производная

$$f'(x) = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}$$

И она не равна нулю. Чтобы найти производную $\varphi'(y)$, нужно найти предел $\lim\limits_{\Delta y \to 0} \dfrac{\Delta x}{\Delta y}$. Так как $\Delta x \to 0$ при $\Delta y \to 0$ (мы рассматриваем только непрерывные функции), то из тождества $\dfrac{\Delta x}{\Delta y} = \dfrac{1}{\dfrac{\Delta y}{\Delta x}}$ получим

$$\lim_{\Delta y \to 0} \frac{\Delta x}{\Delta y} = \frac{1}{\lim\limits_{\Delta x \to 0} \frac{\Delta y}{\Delta x}},$$

т. е. что $\varphi'(y) = \frac{1}{f'(x)}$. Аналогично, если $\varphi'(y) \neq 0$, то $f'(x) = \frac{1}{\varphi'(y)}$. Коротко можно сказать, что *производные от взаимно обратных функций обратны по величине.*

Записывают это так:

$$y'_x = \frac{1}{x'_y} \quad \text{или} \quad x'_y = \frac{1}{y'_x},$$

где индексы x и y показывают, по какой переменной производится дифференцирование, т. е. какая из переменных принята за независимую.

Параметрически заданные функции и их дифференцирование.

I. Параметрическое задание функций и линий. Зададим две функции одной и той же переменной t; обозначим их через x и y:

$$x = \varphi(t) \qquad y = \psi(t). \qquad (*)$$

Задание этих функций означает задание функциональной зависимости между переменными x и y. В самом деле, для каждого значения t (в некоторой области) из системы (*) находятся значения x и y, которые и являются соответствующими друг другу.

Определение. *Задание функциональной зависимости между двумя переменными, состоящее в том, что обе переменные определяются каждая в отдельности как функции одной и той же вспомогательной переменной, называется параметрическим, а вспомогательная переменная —* ***параметром.***

Отыскание по системе (*) непосредственной связи между переменными x и y без участия переменной t называется *исключением параметра.* В результате исключения параметра мы получаем уравнение между x и y, задающее одну из этих переменных как явную или неявную функцию другой.

Прямой путь для исключения параметра таков: из первого, например, равенства находим выражение для t через x, т. е. $t = \chi(x)$, где χ — функция, обратная функции φ, и подставляем это выражение во второе равенство:

$$y = \psi[\chi(x)].$$

Таким образом мы получаем явное выражение для y как функции от x. В конкретных случаях употребляются и другие приемы исключения параметра.

Рассматривая зависимость между x и y как уравнение соответствующей линии, можно сказать, что линия задана параметрически, или, точнее, параметрическими уравнениями.

Параметрическое задание функций и линий часто имеет преимущества по сравнению с другими формами их задания. В то время как непосредственная связь между x и y может быть весьма сложной, функции, определяющие x и y через параметр, могут оказаться простыми. Кроме того, при параметрическом задании заранее не предусматривается, какая именно из переменных принимается за независимую и какая — за функцию.

Заметим, что если хотя бы одна из функций системы (*) постоянна, то исключить параметр t нельзя. Например, уравнения $x = 1$, $y = \sin t$ не дают возможности связать x и y. Здесь точки с координатами x, y, соответствующие различным значениям t, располагаются на отрезке прямой $x=1$, заключенном между точками (1, —1) и (1, 1).

Параметр может иметь различное истолкование в соответствии с характером функциональной зависимости.

Параметрически заданные функции особенно часто встречаются в механике при задании траектории движения. Параметром служит при этом время t, а фукции $x(t)$ и $y(t)$ выражают законы движения проекций движущейся точки на оси координат.

Исключая параметр t, мы получаем уравнение траектории в виде связи между x и y — координатами движущейся точки.

III. Производные параметрически заданных функций. Укажем способ отыскания производной параметрически заданной функции. Пусть y как функция x задана параметрическими уравнениями

$$x = \varphi(t), \qquad y = \psi(t).$$

Дифференцируя вторую из заданных функций по правилу дифференцирования сложной функции, получим

$$y'_x = \psi'(t)\, t'_x$$

(индекс показывает, какая переменная считается независимой; по ней и производится дифференцирование). Производную t'_x найдем по правилу дифференцирования обратной функции

$$t'_x = \frac{1}{x'_t} = \frac{1}{\varphi'(t)}.$$

Окончательно

$$y'_x = \frac{\psi'(t)}{\varphi'(t)},$$

что можно короче записать так:

$$y'_x = \frac{y'_t}{x'_t}.$$

Например, для углового коэффициентакасательной к окружности $x = a\cos t$, $y = a\sin t$ находим выражение

$$y'_x = \frac{y'_t}{x'_t} = \frac{a\cos t}{-a\sin t} = -\operatorname{ctg} t.$$

Так как угловой коэффициент радиуса, проведенного в ту же точку окружности, равен, очевидно, tg t, то это означает, что касательная к окружности перпендикулярна к радиусу.

3.3. Дифференциал

Дифференциал и его геометрический смысл. С понятием производной теснейшим образом связано другое фундаментальное понятие математического анализа — *дифференциал функции*.

Пусть $y=f(x)$ — функция, непрерывная при рассматриваемых значениях x и имеющая производную

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = f'(x).$$

Из этого равенства следует, что

$$\frac{\Delta y}{\Delta x} = f'(x) + \varepsilon,$$

где ε — бесконечно малая величина при $\Delta x \to 0$. Отсюда находим, что

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

где $\alpha = \varepsilon \Delta x$.

Итак, *бесконечно малое приращение Δy дифференцируемой функции $y=f(x)$ может быть представлено в виде суммы двух слагаемых:* 1) *величины, пропорциональной бесконечно малому приращению независимой переменной Δx, и* 2*) бесконечно малой величины более высокого порядка, чем Δx.*

Справедливо и обратное предложение: *если для данного значения x приращение $\Delta y = f(x+\Delta x) - f(x)$ можно представить в виде*

$$\Delta y = a\Delta x + \alpha,$$

где α — бесконечно малая величина более высокого порядка, чем Δx, то функция $y=f(x)$ имеет производную и $a=f'(x)$.

Определение. *Дифференциалом функции называется величина, пропорциональная бесконечно малому приращению аргумента Δx и*

отличающаяся от соответствующего приращения функции на бесконечно малую величину более высокого порядка, чем Δx.

Дифференциал функции y обозначается через dy или $df(x)$.

Мы видим, что необходимым и достаточным условием существования дифференциала функции $y = f(x)$ служит существование производной, и тогда

$$dy = f'(x)\,\Delta x.$$

Приращение Δx *независимой переменной* x *называют ее дифференциалом* dx, т. е.

$$\Delta x = dx\,.$$

Это согласуется с тем, что дифференциал функции $y = x$ равен

$$dy = dx = (x)'\,\Delta x = \Delta x,$$

т. е. $dx = \Delta x$.

Таким образом,

Дифференциал функции равен ее производной, умноженной на дифференциал независимой переменной, т. е.

$$dy = f'(x)\,dx.$$

Зная производную, легко найти дифференциал, и обратно. Поэтому действия отыскания производной и дифференциала данной функции носят общее название дифференцирование.

Дифференциал dy *функции* $y=f(x)$ *в точке* x *изображается приращением ординаты точки касательной, проведенной к линии* $y=f(x)$ *в соответствующей ее точке* $(x, f(x))$.

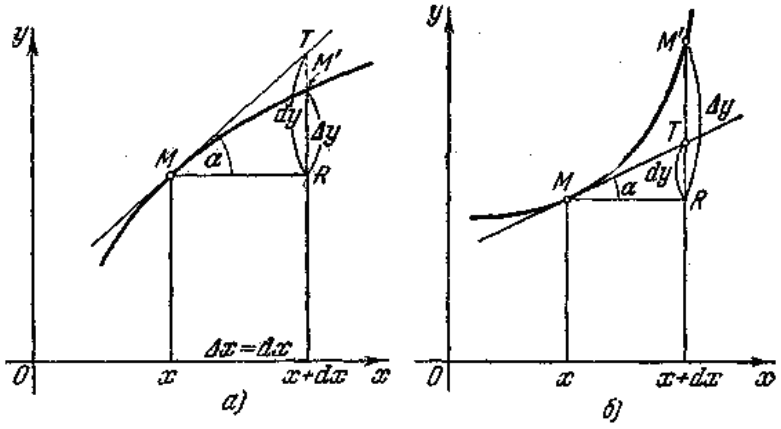


Рис. 3.1.

Дифференциал функции в данной точке может быть как больше приращения (рис. 3.1, *а),* так и меньше его (рис. 3.1,б).

Свойства дифференциала.

I. Дифференциалы основных элементарных функций. Так как дифференциал получается из производной умножением ее на дифференциал независимой переменной, то, зная производные основных элементарных функций, можем составить без всяких затруднений таблицу дифференциалов этих функций. Так, $d(x^n) = nx^{n-1}dx$, $d(a^x) = a^x \ln a\, dx$ и т. д.

II. Дифференциалы результатов арифметических действий. В соответствии с правилами отыскания производных и принятыми обозначениями придем к аналогичным правилам для отыскания дифференциалов.

а) Так как

$$(u+v+\ldots+w)' = u'+v'+\ldots+w',$$

то, умножаяобе части равенства на *dx,* получим

$$d(u+v+\ldots+w) = du+dv+\ldots+dw.$$

б) Так как

$$(uv)' = u'v+uv',$$

то, умножая обе части равенства на *dx,* получим

$$d(uv) = v\,du+u\,dv,$$

в частности

$$d(Cu) = C\,du.$$

в) Так как

$$\left(\frac{u}{v}\right)' = \frac{u'v-uv'}{v^2},$$

то, умножая обе части равенства на *dx,* получим

$$d\left(\frac{u}{v}\right) = \frac{v\,du-u\,dv}{v^2}.$$

III. Дифференциал сложной функции. Свойство инвариантности. Рассмотрим свойство дифференциала функции, вытекающее из правила дифференцирования сложной функции.

Дифференциал функции $y=f(u)$ сохраняет одно и то же выражение независимо от того, является ли ее аргумент u независимой переменной или функцией от независимой переменной.

Это свойство называется *инвариантностью* (т. е. неизменностью) *формы дифференциала.* Всегда можно, не интересуясь природой аргумента функции, записать ее дифференциал в одном и том же виде.

IV. Дифференциал как главная часть приращения. Пусть в точке x производная функции $y=f(x)$ отлична от нуля; $f(x)\neq 0$. Тогда

$$\Delta y = f'(x)\,dx + \alpha = dy + \alpha,$$

где α — бесконечно малая величина более высокого порядка, чем dx. Но при указанном условии она будет бесконечно малой величиной более высокого порядка и чем dy и Δy. Действительно, при $dx\to 0$ имеем

$$\frac{\alpha}{dy} = \frac{\alpha}{f'(x)\,dx} \to 0,$$

ибо $\frac{\alpha}{dx} \to 0$, а $f'(x) \neq 0$. Значит, Δy и dy отличаются друг от друга на бесконечно малую величину более высокого порядка, чем они сами, и следовательно, они эквивалентны:

$$dy \sim \Delta y.$$

В дальнейшем приращение функции Δy будет очень часто заменяться ее дифференциалом dy. В связи с этим коснемся общего вопроса о приближенном выражении одной переменной величины через другую при условии, что обе они стремятся к одному и тому же пределу.

Пусть функции $u(x)$ и $v(x)$ при $x\to x_0$ стремятся к одному и тому же пределу A, т. е. $\lim u = \lim v = A$. Тогда предел их разности равен нулю: $\lim (u - v) = 0$ и, следовательно, сама эта разность

$$u(x) - v(x) = \alpha(x)$$

есть бесконечно малая величина при $x\to x_0$.

Если теперь в окрестности точки x_0 заменять значения функции $u(x)$ значениями функции $v(x)$, то абсолютная ошибка $\alpha = |u - v|$ будет бесконечно малой.

Если предел $A\neq 0$, то и относительная ошибка, по которой и оценивается точность приближения, также является бесконечно малой:

$$\delta = \frac{\alpha}{|v|} = \left|\frac{u-v}{v}\right| = \left|\frac{u}{v} - 1\right| \to \left|\frac{A}{A} - 1\right| = 0.$$

Но при $A = 0$, т. е. когда обе функции $u(x)$ и $v(x)$ сами являются бесконечно малыми при $x\to x_0$, относительная ошибка может и не быть бесконечно малой. *Она будет бесконечно малой только тогда, когда $u(x)$ и $v(x)$ — эквивалентные бесконечно малые.* В самом деле, из того, что

$$\delta = \left| \frac{u-v}{v} \right| \to 0,$$

следует, что разность u—v есть бесконечно малая более высокого порядка, чем v. Это и означает, что u и v эквивалентны.

Итак, если две бесконечно малые величины эквивалентны, то значения одной из них являются приближенными значениями другой с бесконечно малой относительной ошибкой. Коротко говорят, что *каждая из двух бесконечно малых есть главная часть другой.*

Вернемся теперь к вопросу о замене приращения функции ее дифференциалом. При фиксированном значении x и приращение $\Delta y = f(x+\Delta x) - f(x)$, и дифференциал $dy = f'(x)\Delta x$ зависят только от Δx и при $\Delta x \to 0$ стремятся к нулю. Так как выше уже было показано, что Δy и dy эквивалентны, то замена Δy на dy приводит к бесконечно малой относительной ошибке. Поэтому можно сказать, что

Дифференциал функции есть главная часть приращения функции, пропорциональная дифференциалу (приращению) независимой переменной.

Заметим только, что если в данной точке $f'(x)=0$, то дифференциал $dy = 0$, и он не сравнивается ни с какой другой бесконечно малой величиной, в том числе и с приращением функции Δy.

Дифференцируемость функции.

Определение. *Функция $y=f(x)$ называется дафференцируемой в точке x, если она имеет в этой точке дифференциал.*

При этом, в точке x существует производная $f'(x)$ (и обратно). Таким образом, существование производной может быть принято в качестве условия дифференцируемости функции, чему с геометрической точки зрения соответствует существование у линии $y=f(x)$ касательной, не перпендикулярной к оси Ox.

Допустим, что функция $f(x)$ дифференцируема при некотором значении x, т. е.

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = f'(x);$$

тогда она обязательно непрерывна в этой точке. В самом деле, при этом $\Delta y = f'(x)\,\Delta x + \alpha$, и при $\Delta x \to 0$ приращение Δy также стремится к нулю, а это и является условием непрерывности функции $y=f(x)$ в точке x. Поэтому в точках разрыва функция не может иметь производной.

Итак, *если функция $y=f(x)$ дифференцируема, то она обязательно непрерывна.* Обратное не всегда справедливо, потому что можно указать примеры таких непрерывных функций, которые не во всех точках имеют производную.

Например, функция $y = |x|$, непрерывная на всей оси Ox, не имеет производной при $x = 0$. В самом деле,

$$\Delta y = |x + \Delta x| - |x|,$$

что при $x = 0$ дает

$$\Delta y = |\Delta x|$$

и, значит,

$$\frac{\Delta y}{\Delta x} = \frac{|\Delta x|}{\Delta x};$$

$$\text{при } \Delta x > 0 \text{ будет } \quad \frac{\Delta y}{\Delta x} = \frac{\Delta x}{\Delta x} = 1,$$

$$\text{а при } \Delta x < 0 \quad \text{» } \quad \frac{\Delta y}{\Delta x} = \frac{-\Delta x}{\Delta x} = -1.$$

Отсюда следует, что отношение $\frac{\Delta y}{\Delta x}$ не имеет предела при Δx, произвольно стремящемся к нулю, а это и означает, что производной не существует. Геометрически это также ясно: графиком функции $y=|x|$ служит ломаная линия (рис. 3.2), вершина которой находится в точке (0,0), и, разумеется, в этой точке линия не имеет касательной.

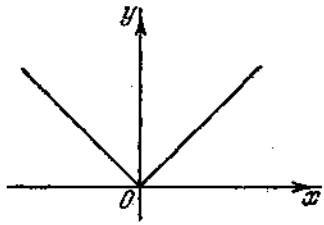


Рис. 3.2.

Заметим, что *все элементарные функции дифференцируемы всюду в интервале, в котором они определены, за исключением, может быть, лишь отдельных точек.*

II. Односторонние производные. Вернемся еще раз к функции $y=|x|$, график которой изображен на рис. 3.2. Как ми уже видели, при $x = 0$ предела отношения $\frac{\Delta y}{\Delta x}$ при произвольном стремлении Δx к нулю не существует. Однако левый и правый пределы этого отношения существуют:

$$\lim_{\substack{\Delta x \to 0 \\ \Delta x < 0}} \frac{\Delta y}{\Delta x} = -1, \qquad \lim_{\substack{\Delta x \to 0 \\ \Delta x > 0}} \frac{\Delta y}{\Delta x} = 1.$$

Эти пределы называются соответственно левой и правой производными в точке $x = 0$. Вообще, если функция $y=f(x)$ в точке x_0 непрерывна и существуют левый и правый пределы отношения

$$\frac{\Delta y}{\Delta x},$$

то они называются *односторонними производными* в точке x_0; соответственно *левая производная* ($f'_{лев}(x_0)$) *и правая производная* ($f'_{пр}(x_0)$). Ясно, что если левая и правая производные и данной точке совпадают, то функция дифференцируема.

Если функция определена на некотором замкнутом интервале, то производные на концах интервала определяются именно как односторонние: на левом конце интервала это будет правая производная, а на правом — левая. Это совершенно аналогично определению непрерывности функции на концах интервала.

3.4. Производные и дифференциалы высших порядков

Производные высших порядков.

I. Допустим, что функция $y=f(x)$ имеет производную $f'(x)$ в некотором интервале независимой переменной x. Производная от $f'(x)$ (если она существует) называется *производной второго порядка* или *второй производной* от первоначальной функции $f(x)$и обозначается через $f''(x)$:

$$f''(x) = [f'(x)]' = \lim_{\Delta x \to 0} \frac{f'(x+\Delta x) - f'(x)}{\Delta x}.$$

Таким же образом *производной третьего порядка* или *третьей производной* $f'''(x)$ от функции $y = f(x)$ называется производная от производной второго порядка. Дадим общее определение.

Определение. *Производной n-го порядка* $f^{(n)}(x)$ *называется производная от производной* $(n-1)$*-го порядка*

$$f^{(n)}(x) = [f^{(n-1)}(x)]' = \lim_{\Delta x \to 0} \frac{f^{(n-1)}(x+\Delta x) - f^{(n-1)}(x)}{\Delta x}.$$

Производные второго и вообще высших порядков оказываются существенно необходимыми для определения важных понятий математики, механики, физики и для более полного исследования функций, чем то, которое можно выполнить, применяя лишь первую производную.

Производные от данной функции в данной точке могут существовать до некоторого определенного порядка, а производных высшего порядка функция в этой точке может и не иметь. Однако всякая элементарная функция, за исключением быть может отдельных точек, имеет в своей области определения производные любых порядков.

В целях единства терминологии производную данной функции называют производной первого порядка или первой производной. Умение дифференцировать всякую элементарную функцию позволяет найти одну вслед за другой последовательные производные данной элементарной функции до любого порядка.

II. Производные неявных функций. Если y — неявная функция, то для отыскания ее высшей производной нужно соответствующее число раз дифференцировать заданное уравнение, связывающее x и y, помня всегда, что y и все ее производные суть функции независимой переменной.

III. Производные параметрическизаданных функций. Для того чтобы найти производную высшего порядка от параметрически заданной функции, нужно продифференцировать выражение для предыдущей производной, как сложную функцию независимой переменной.

IV. Формула Лейбница. Укажем имеющее практическое значение правило для отыскания производной n-го порядка от произведения функций.

Пусть u и v — некоторые функции от x и

$$y = uv;$$

выразим $y^{(n)}$ через производные функций u и v.

Имеем последовательно

$$y' = u'v + uv',$$
$$y'' = u''v + 2u'v' + uv'',$$
$$y''' = u'''v + 3u''v' + 3u'v'' + uv'''.$$

Легко подметить аналогию между выражениями для второй и третьей производных и разложением биномов соответственно во второй и в третьей степенях; эти выражения так сконструированы из производных u и v (нулевого, первого, второго и третьего порядков), как разложения для биномов — из последовательных степеней u и v (нулевой, первой, второй и третьей). Оказывается, эта аналогия справедлива в общем случае.

Формула Лейбница. При любом n справедливо равенство

$$y^{(n)} = (uv)^{(n)} = u^{(n)}v + nu^{(n-1)}v' + \frac{n(n-1)}{2!}u^{(n-2)}v'' + \ldots$$
$$\ldots + nu'v^{(n-1)} + uv^{(n)}. \quad (*)$$

Эту формулу можно получить из разложения бинома $(u+v)^n$, если в нем заменить степени u и v соответствующими производными от u и v.

Дифференциалы высших порядков. Дифференциал dy функции $y=f(x)$ есть функция двух переменных: независимой переменной x и ее дифференциала dx. Дифференциал dx независимой переменной x есть величина, не зависящая от x: при заданном значении x значения dx могут выбираться произвольно.

Рассматривая $df(x)$ как функцию x, возьмем дифференциал $d[df(x)]$. Если этот дифференциал существует, то он называется *дифференциалом второго порядка* или *вторым дифференциалом* от функции $f(x)$ и обозначается через d^2y:

$$d^2y=d(dy).$$

Таким же образом *дифференциалом третьего порядка* или *третьим дифференциалом* d^3y от функции $y= f(x)$ называется дифференциал от дифференциала второго порядка как функции x. Дадим общее определение.

Определение. *Дифференциалом n-го порядка* d^ny называется дифференциал от дифференциала $(n—1)$-го порядка как функции x:

$$d^n y = d\,(d^{n-1}y).$$

Дифференциал n-го порядка равен произведению производной n-го порядка по независимой переменной на n-ю степень дифференциала независимой переменной.

Дифференциал $df(x)$ функции $f(x)$ называется, для общности терминологии, *дифференциалом первого порядка или первым дифференциалом.*

Из формул для дифференциалов получаем выражения для производных в виде дробей:

$$y' = f'(x) = \frac{dy}{dx},$$

$$y'' = f''(x) = \frac{d^2y}{dx^2},$$

. . . . . .

$$y^{(n)} = f^{(n)}(x) = \frac{d^ny}{dx^n},$$

. . . . . .

Эти формулы, за исключением первой, верной всегда, справедливы лишь при условии, что x — независимая переменная.

Для удобства записи часто вместо $\frac{d^ny}{dx^n}$ условно пишут

$$\frac{d^n}{dx^n} y.$$

Так, например, записывают: $\frac{d^3}{dx^3}(2x^4 - x + 1) = 48x.$

4. Применение дифференциального исчисления к исследованию функций

4.1. Теоремы Ферма, Ролля, Лагранжа и Коши

Теоремы Ферма и Ролля. Изучение попросов применения дифференциального исчисления к исследованию функций мы начнем с теоремы Ферма.

Теорема Ферма. *Пусть функция y=f(x), непрерывная в некотором интервале* $[x_1, x_2]$, *принимает свое наибольшее (или наименьшее) значение во внутренней точке* ξ *этого интервала:* $x_1 < \xi < x_2$. *Если в точке* ξ *производная функции f(x) существует, то она обязательно равна нулю:* $f'(\xi) = 0$.

Геометрический смысл теоремы Ферма ясен из рис. 4.1: *касательная к графику функции в его наивысшей (или наинизшей) точке параллельна оси абсцисс.*

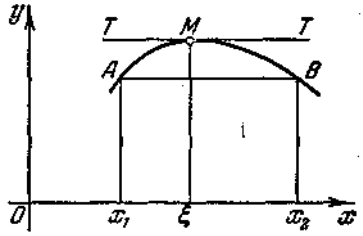


Рис. 4.1.

Если наибольшее или наименьшее значения функции принимаются ею на конце интервала (например, на рис. 4.1 наименьшее значение принимается при $x = x_1$ и $x = x_2$), то касательная в соответствующей точке не обязана быть параллельной оси абсцисс, т. е. производная может и не равняться нулю.

Отметим, что возможны случаи, когда функция принимает свое наибольшее или наименьшее значения в точках, в которых производная не существует. Так, например, функция $y = |x|$ принимает свое наименьшее значение, равное нулю, в точке $x = 0$, где она не имеет производной.

Почти непосредственным следствием теоремы Ферма является теорема Ролля.

Теорема Ролля. *Если функция f(x) непрерывна в замкнутом интервале* $[x_1, x_2]$*, дифференцируема во всех его внутренних точках и имеет на концах интервала равные значения, то в этом интервале существует хотя бы одно значение* $x = \xi$*, для которого* $f'(\xi) = 0$**.**

Геометрическое истолкование теоремы Ролля таково:

На линии y= f (x), где функция f(x) удовлетворяет условиям теоремы Ролля, найдется точка, в которой касательная параллельна оси абсцисс (рис. 4.1).

Если $f(x_1)=f(x_2)=0$, то теорему Ролля можно формулировать следующим образом:

Между всякими двумя нулями функции лежит хотя бы один нуль производной.

Именно так, причем лишь для многочленов, теорема впервые была указана самим Роллем.

Утверждение теоремы Ролля перестает быть верным, если не требовать дифференцируемости функции во всех внутренних точках интервала. Например, непрерывная функция $y = \sqrt[3]{x^2}$ на концах интервала [—1, +1] имеет равные значения (=1), а вместе с тем ее производная $y' = \frac{2}{3\sqrt[3]{x}}$ нигде в нуль не обращается. И действительно, в данном случае условия теоремы не выполнены: в точке $x = 0$, лежащей внутри интервала (—1, + 1), производной не существует. Весьма характерен график этой функции (рис. 4.2).

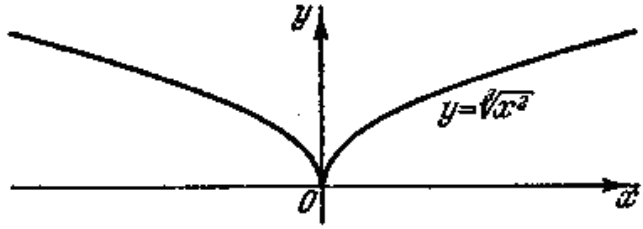


Рис. 4.2.

В точке (0, 0) имеется касательная, перпендикулярная к оси абсцисс, причем кривая в этой точке как бы заостряется. Такие точки будем называть *точками возврата* кривой.

Теорема Лагранжа. Вернемся снова к геометрическому истолкованию теоремы Ролля. Так как хорда, стягивающая концы дуги *AB* (рис. 4.1), параллельна оси абсцисс, то *касательная в точке M будет параллельна этой хорде.* Геометрически ясно, что это последнее

свойство сохранится и в том случае, когда хорда, стягивающая концы дуги, не будет параллельна оси абсцисс. Именно, пусть *AB* — линия, имеющая в каждой точке касательную и не имеющая точек возврата (рис. 4.3).

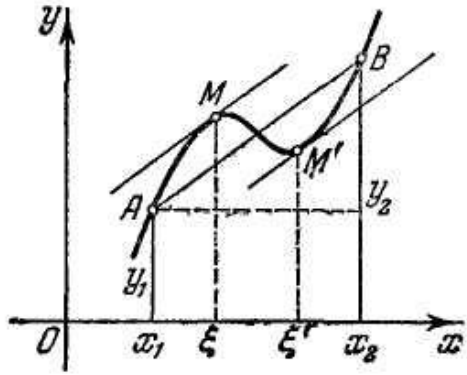


Рис. 4.3

Тогда на *ней всегда найдется точка, в которой касательная к линии будет параллельна хорде, стягивающей линию.*

Если условия, наложенные на линию *AB,* не выполняются, то такой точки может и не найтись (рис. 4.4).

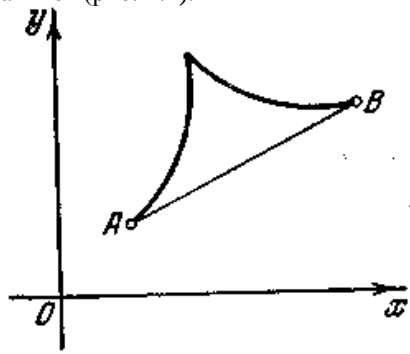


Рис. 4.4

Теорема Лагранжа. *Если функция f(x) непрерывна в замкнутом интервале* $[x_1, x_2]$ *и дифференцируема во всех его внутренних течках, то в этом интервале существует хотя бы одно значение* $x = \xi$, *для которого*

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

По теореме Лагранжа имеем

$$f(x_2)-f(x_1)=f'(\xi)(x_2-x_1), \qquad x_1<\xi<x_2.$$

Эта формула называется *формулой Лагранжа.* Она выражает тот факт, что

Приращение функции на интервале равно произведению производной в некоторой промежуточной точке интервала на приращение независимой переменной.

Формула Лагранжа позволяет дать точное выражение для приращения функции через приращение аргумента и значение производной в некоторой точке интервала. Она имеет большое теоретическое значение и лежит в основе доказательств ряда важных теорем.

Теорема Коши. Теорема Коши является обобщением теоремы Лагранжа. Приведем ее аналитическую формулировку.

Теорема Коши. *Если функции f(x) и φ(x) непрерывны в замкнутом интервале* $[x_1, x_2]$ *и дифференцируемы во всех его внутренних точках, причем φ'(x) в этих точках не обращается в нуль, то в этом интервале существует хотя бы одно значение* $x = \xi$*, для которого*

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

Заметим, что $\varphi(x_2)-\varphi(x_1)\neq 0$, так как в противном случае по теореме Ролля существовала бы внутри этого интервала точка, в которой производная φ'*(x)* обращалась бы в нуль, что противоречит условиям теоремы.

4. 2. Поведение функции в интервале

Признаки монотонности функции. Одно из самых важных назначений дифференциального исчисления — это применение его к исследованию функций (и линий), т. е. к характеристике поведения функции при изменении независимой переменной.

Применение дифференциального исчисления к исследованию функций опирается на весьма простую связь, существующую между поведением функции и свойствами ее производных, прежде всего ее первой производной.

Теорема (необходимый признак монотонности).

1) *Если функция f(x) в интервале возрастает, то ее производная f (x) неотрицательна: f ' (x) ≥ 0;*

2) *если функция f(x) в интервале убывает, то ее производная f'(x) неположительна:* $f'(x) \le 0$;

3) *если функция f(x) в интервале не изменяется (есть константа), то ее производная f (x) тождественно равна нулю.*

Геометрический смысл этой теоремы очевиден (рис. 4.5): если подвижная точка *M(x, y)* при движении по графику функции слева направо, т.е. при возрастании абсциссы, поднимается, то касательная к графику образует с осью *Ox* острый угол, тангенс которого положителен; если же точка *M(x, y)* опускается, то касательная образует с осью *Ox* тупой угол, тангенс которого отрицателен.

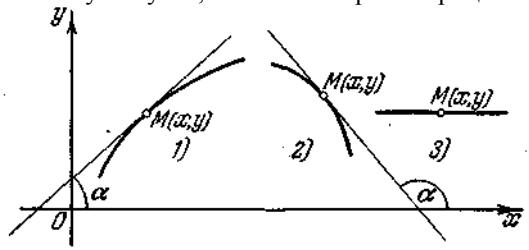


Рис. 4.5.

Следует подчеркнуть, что производная монотонной функции может в отдельных точках обращаться в нуль. Вот простой пример: $y = x^3$. Эта кубическая функция возрастает на всей оси *Ox,* но ее производная $y' = 3x^2$ обращается в нуль в точке $x = 0$, будучи положительной во всех остальных точках. Геометрический смысл этого факта таков: касательная к графику монотонной функции в отдельных точках может быть параллельна оси *Ox*.

Таким образом, *в интервале монотонности функции знак ее производной не может измениться на обратный.*

Это предложение позволяет по характеру роста монотонной функции в интервале установить знак ее производной в этом интервале.

Однако когда мы только начинаем исследовать функцию, то ее поведение обычно неизвестно, и поэтому значительно важнее обратное предложение, сводящее вопрос о характере роста функции в данном интервале к более простому вопросу о знаке ее производной в этом интервале.

Теорема (достаточный признак монотонности),

1) *Если производная f'(x) от функции f(x) всюду в интервале положительна, то функция f (x) в этом интервале возрастает;*

2) *если производная f'(x) от функции f(x) всюду в интервале отрицательна, то функция f(x) в этом интервале убывает;*

3) *если производная f (x) от функции f(x) всюду в интервале равна нулю, то функция f(x) в этом интервале не изменяется (есть константа).*

Геометрически ясно, что функция будет монотонной и в том случае, когда ее производная, сохраняя все время постоянный знак, в отдельных точках равна нулю. Например, функция $y = x$— $\sin x$ возрастает в любом интервале, так как ее производная $y' = 1 - \cos x$ все время положительна, кроме точек $x = 2k\pi$, где она равна нулю. Заметим, что те значения x, в которых производная обращается в нуль, называются *стационарными точками* функции.

Экстремумы функции. Особую роль в исследовании функций играют значения x, отделяющие интервал возрастания от интервала убывания или интервал убывания от интервала возрастания (рис. 4.6). В этих точках функция $f(x)$ меняет характер своего изменения; при переходе независимой переменной через эти точки (как всегда, слева направо) функция $f(x)$ из возрастающей становится убывающей (график *I* на рис. 4.6) или, наоборот, из убывающей—возрастающей (график *II* на рис. 4.6).

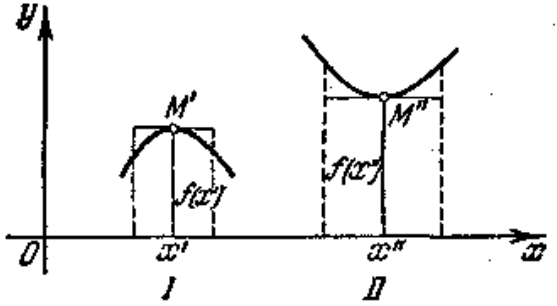


Рис. 4.6.

Если имеет место первый случай (точка x' на рис. 4.6 отделяет интервал возрастания от интервала убывания), то существует такая окрестность точки x', что значение $f(x')$ является наибольшим значением функции $f(x)$ в этой окрестности; если имеет место второй случай (точка x'' на рис. 4.6 отделяет интервал убывания от интервала возрастания), то существует такая окрестность точки x'', что значение $f(x'')$ является наименьшим значением функции $f(x)$ в этой окрестности. Дадим общее определение.

Определение. *Точка x_0 называется точкой максимума функции $f(x)$, если $f(x_0)$ есть наибольшее значение функции $f(x)$ в некоторой окрестности точки x_0.*

Аналогично

Точка x_0 называется точкой минимума, если $f(x_0)$ есть наименьшее значение функции $f(x)$ в некоторой окрестности точки x_0.

Точки максимума и минимума объединяются названием *точки экстремума.* Если x_0 — точка экстремума (или, как еще можно сказать, экстремальная точка) функции $f(x)$, то говорят, что эта функция *достигает экстремума* — соответственно *максимума* или *минимума* — в точке x_0 и что $f(x_0)$ есть *экстремальное* — соответственно *максимальное* или *минимальное* — значение функции $f(x)$. Функция на данном интервале может иметь несколько экстремумов, причем какой-нибудь минимум функции может оказаться больше какого-нибудь максимума (рис. 4.7).

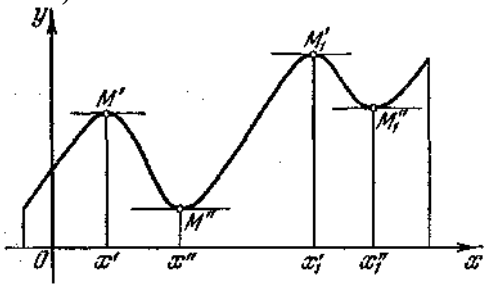


Рис. 4.7

Поэтому наибольшее и наименьшее значения функции в интервале в отличие от экстремумов, носящих относительный характер, называют иногда *абсолютными максимумом и минимумом функции.*

На графике функции точкам экстремума соответствуют вершины линии, обращенные соответственно вверх или вниз (на рис. 4.7 точки *M'* и *M''*).

Обычно встречающиеся функции имеют на заданном конечном интервале лишь конечное, определенное число точек экстремума.

Установим теперь признак, дающий *необходимое условие* того, чтобы данная точка являлась точкой экстремума.

Необходимый признак экстремума. *Если в точке x_0 функция $f(x)$ достигает экстремума, то ее производная в этой точке либо равна нулю, либо не существует.*

Если в точке x_0 функция достигает экстремума, скажем максимума, то значение функции в этой точке является наибольшим ее значением в некоторой окрестности точки x_0. Но по теореме Ферма в тех внутренних точках интервала, в которых дифференцируемая

функция достигает своего наибольшего значения, ее производная равна нулю. Совершенно аналогично проводится рассуждение и для точки минимума.

С геометрической точки зрения рассматриваемый признак означает, что *касательная к графику функции в его вершине параллельна оси Ох* (см. рис. 4.7).

Функция может также иметь экстремумы и в тех отдельных точках, в которых она недифференцируема. Случаи эти изображены на рис. 4.8.

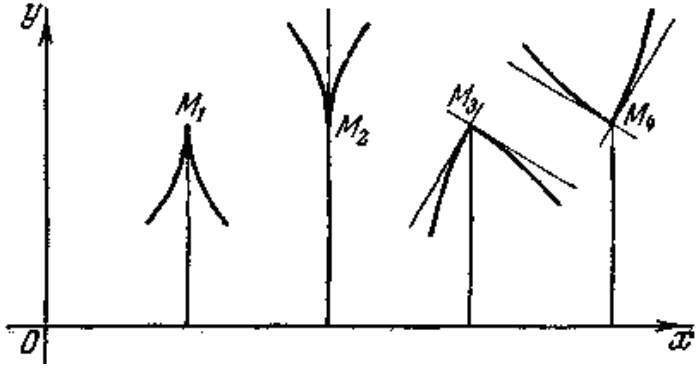


Рис. 4.8.

При этом в точках M_1 и M_2 график функции имеет касательную, перпендикулярную к оси абсцисс. Такие точки мы условились называть *точками возврата.* Характерным их признаком является то, что производная $f'(x)$ при стремлении x к абсциссе любой такой точки стремится, с одной стороны, к $+\infty$, а с другой стороны, к $-\infty$. В точках M_1 и M_2, касательная переходит внезапно от одного наклона к другому, т. е. в этих точках график не имеет определенной касательной; такие точки называются *угловыми точками* кривой. При значениях x, равных абсциссам таких точек, левая и правая производные функции $f(x)$ различны.

Важно подчеркнуть, что *необходимый признак экстремума не является достаточным,* т. е. из того, что производная в данной точке обращается в нуль (или ее не существует), еще не следует, что эта точка обязательно будет точкой экстремума. Так, например, функция $y = x^3$ имеет производную $y' = 3x^2$, обращающуюся в нуль при $x = 0$. Однако точка $x = 0$ вовсе не является точкой экстремума (рис. 4.9, *а).*

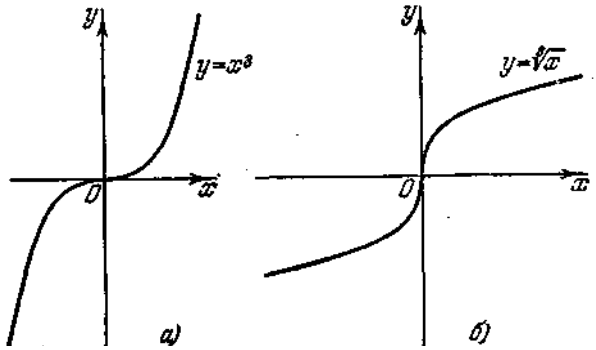


Рис. 4.9.

Функция $y = \sqrt[3]{x}$, для которой

$$y' = \frac{1}{3\sqrt[3]{x^2}},$$

не имеет производной при $x = 0$. Из рис. 4.9,б ясно видно, что и здесь точка $x = 0$ не является экстремальной. Как легко заметить, в обоих приведенных примерах точка $x=0$ не отделяет друг от друга интервалы монотонности противоположного смысла. И слева и справа от этой точки производная имеет один и тот же знак.

Для того чтобы иметь возможность судить о том, когда же данная точка будет являться точкой экстремума, мы перейдем к установлению *достаточного признака экстремума.*

Первый достаточный признак экстремума. *Точка x_0 является точкой экстремума функци f(x), если производная f '(x) при переходе x через x_0 меняет знак; при перемене знака + на — точка x_0 является точкой максимума; при перемене — на + точка x_0 является точкой минимума.*

В самой точке x_0 производная в силу необходимого признака равна нулю или не существует.

Замечание. Следует иметь в виду, что, основываясь только на перемене знака производной, нельзя еще заключить о наличии экстремума; необходимо знать еще, что в самой точке функция непрерывна. Так, например, пусть $y = \frac{1}{x^2}$. Производная этой функции $y' = -\frac{2}{x^3}$ меняет знак при переходе x через точку $x = 0$: слева от нуля $y' > 0$ и, значит, функция возрастает, справа от нуля $y' < 0$ и, значит, функция

убывает; сама же точка $x = 0$ не является точкой максимума, ибо функция при $x = 0$ имеет бесконечный разрыв.
На графике функции (рис. 4.10) это обстоятельство очевидно.

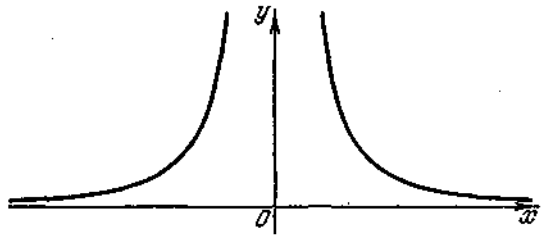


Рис. 4.10.

Схема исследования функций на экстремумы. Наибольшее и наименьшее значения функции.

I. Схема исследования. Укажем последовательность действий для изучения роста и отыскания экстремумов непрерывной функции $y=f(x)$ в заданном интервале, который может быть как конечным, так и бесконечным. Будем считать, что функция $y=f(x)$ имеет производную всюду, за исключением, быть может, отдельных точек.

1) Прежде всего нужно найти точки интервала, в которых производная равна нулю (стационарные точки), т. е. действительные корни уравнения $f'(x) = 0$, а также те точки, в которых производная не существует. Обозначим все найденные точки в порядке возрастания через $x_1, x_2, \ldots, x_n$.

Таким образом,

$$x_1 < x_2 < \ldots < x_n.$$

Это — те точки интервала, в которых функция $f(x)$ может иметь экстремумы. Их иногда называют *критическими.*

2) Затем разбиваем при помощи точек x_i весь заданный интервал $[a, b]$ (a и b могут быть равны соответственно — ∞ и + ∞) на *частичные* интервалы: (a, x_1), (x_1, x_2),…, (x_{n-1}, x_n), (x_n, b), в каждом из которых производная сохраняет постоянный знак. В самом деле, в противном случае производная была бы равна нулю (или не существовала бы) еще в точках, отличных от выделенных, а все такие точки мы уже нашли в п. 1). Следовательно, эти интервалы являются интервалами монотонности функции.

3) Находим знак производной в каждом из частичных интервалов, для чего достаточно узнать ее знак в какой-нибудь одной точке интервала. По знаку производной определяем характер

изменения функции в каждом интервале монотонности (возрастает она или убывает). Следя за переменой знака производной при переходе через границы интервалов монотонности —точки *x*, выясняем, какие из этих точек будут точками максимума и какие — минимума. При этом может оказаться, что какая-нибудь точка x_i не служит точкой экстремума функции. Это случится если в двух соседних интервалах (x_{i-1}, x_i) и (x_i, x_{i+1}), разделяемых точкой x_i функция монотонна в одинаковом смысле, т. е. производная в них имеет один и тот же знак. Тогда они объединяются в один интервал монотонности функции. В этом случае точка x_i не будет точкой экстремума (пример: $x = 0$ для функции $y = x^3$).

4) Подстановкой в выражение функции $f(x)$ критических значений $x = x_i$ находим соответствующие значения функции:

$$f(x_1),\ f(x_2),\ \ldots,\ f(x_n).$$

Как уже отмечалось, не все из этих значений могут оказаться экстремальными.

Если значения $f(x_1),\ f(x_2),\ \ldots,\ f(x_n)$ вычислены, а также найдены значения функции на концах интервала $f(a)$ и $f(b)$, то ход изменения функции легко представить и без исследования знака производной. Так как нам уже известно, что в каждом из интервалов $(a,\ x_1),\ (x_1,\ x_2),\ \ldots,\ (x_n,\ b)$ функция не имеет точек экстремума и, следовательно, монотонна, то, сравнивая между собой значения функции на концах каждого такого интервала, мы и определим, где функция возрастает и где убывает.

Какой из приемов следует употребить, зависит от конкретных обстоятельств. Может быть, что удобнее и легче использовать знаки производной, а может быть и так, что, нуждаясь все равно в значениях функции в критических точках, проще использовать эти значения и не выяснять знаки производной.

Результаты исследования целесообразно сводить в таблицу:

x	a	$a < x < x_1$	x_1	$x_1 < x < x_2$	x_2	$x_2 < x < b$	b
y y'	$f(a)$	возрастает +	$f(x_1)$ 0	убывает —	$f(x_2)$ 0	убывает —	$f(b)$

На приведенной таблице (для простоты считаем, что в интервале $[a, b]$ всего две критические точки) показана примерная схема ее заполнения. Знак производной в каждом из интервалов определяется путем вычисления значения производной в одной из точек интервала

(безразлично какой). Если производная представлена в виде произведения нескольких множителей, то достаточно найти знаки этих множителей, не вычисляя их значений; по знакам множителей определится и знак произведения.

По такой таблице можно построить схематический график функции.

II. Наибольшее и наименьшее значения функции. После того как все экстремальные значения функции найдены, легко найти наибольшее M и наименьшее m значения функции $f(x)$ в интервале $[a, b]$. Для этого нужно *сравнить между собой значения функции в экстремальных точках и на концах интервала.* Действительно, наибольшим значением функции в интервале $[a, b]$ может быть или одно какое-нибудь из максимальных ее значений, или значение на конце интервала. Аналогично наименьшее значение функции следует искать или среди минимальных ее значений, или среди значений на концах интервала. Некоторые из возможных случаев представлены на рис. 4.11.

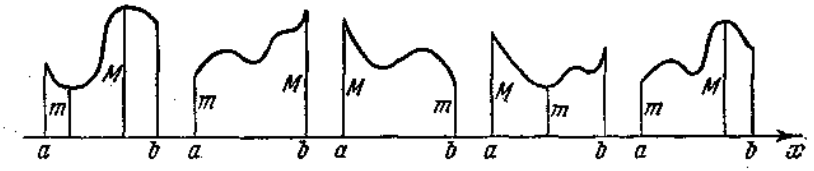


Рис. 4.11.

Ясно, что если функция в интервале $[a, b]$ монотонно возрастает, то ее наибольшее значение будет $f(b)$, а наименьшее $f(a)$; если функция монотонно убывает, то указанные значения просто поменяются местами.

Часто встречается случай, когда функция в интервале имеет только одну экстремальную точку. Если это точка максимума, то значение функции в ней, очевидно, будет наибольшим значением функции в интервале, а если минимума, то наименьшим.

III. Примеры исследования функций. 1**) Основные элементарные функции.** Хотя поведение каждой из основных элементарных функций уже описано нами ранее, однако применение первой производной позволяет очень легко аналитически обнаружить простейшие свойства этих функций.

Степенная функция: $y=x^n$. Ее производная $y'=nx^{n-1}$ при $x>0$ положительна, если $n > 0$, и отрицательна, если $n < 0$. Следовательно, на положительной полуоси Ox функция или везде возрастает $(n > 0)$, или везде убывает $(n < 0)$.

Показательная функция: $y = a^x$. Знак ее производной $y' = a^x \ln a$ совпадает со знаком $\ln a$, так как функция a^x везде положительна. Поэтому $y' > 0$, если $a > 1$, и $y' < 0$, если $a < 1$. Показательная функция монотонна на всей оси *Ox:* возрастает в первом случае, убывает во втором.

Логарифмическая функция: $y = 1n\ x$. Ее производная $y' = \frac{1}{x}$ при $x>0$ (а только для этих значений x функция определена) положительная, и значит, $1n\ x$ везде возрастает.

Подобным образом, обращаясь к производным, нетрудно восстановить ход изменения и других основных элементарных функций.

2) Рассмотрим функцию

$$y = 3x^3 + 4{,}5x^2 - 4x + 1.$$

Эта функция имеет непрерывную производную на всей оси *Ox*.

Находим производную:

$$y' = 9x^2 + 9x - 4 = (3x - 1)(3x + 4).$$

Отсюда видно, что производная равна нулю при $x = -\frac{4}{3}$ и при $x = \frac{1}{3}$

Так как при $x < -\frac{4}{3}$ оба множителя отрицательны, то производная при этих значениях x положительна и, следовательно, функция возрастает. При $-\frac{4}{3} < x < \frac{1}{3}$ производная отрицательна и функция убывает, а при $x > \frac{1}{3}$ производная снова положительна и функция возрастает. Таким образом, $x = -\frac{4}{3}$ есть точка аксимума, а $x = \frac{1}{3}$ — точка минимума, и мы имеем три интервала монотонности: от $-\infty$ до $-\frac{4}{3}$ — интервал возрастания, от $-\frac{4}{3}$ до $\frac{1}{3}$ — интервал убывания и от $\frac{1}{3}$ до $+\infty$ — интервал возрастания.

Вычислим значения функции в точках экстремума и составим таблицу поведения функции. При составлении таблицы учитываем, что

$$\lim_{x \to \pm\infty} y = \lim_{x \to \pm\infty} x^3 \left(3 + \frac{,5}{x} - \frac{4}{x^2} + \frac{1}{x^3}\right) = \pm\infty.$$

x	$-\infty$	$-\infty < x < -\frac{4}{3}$	$-\frac{4}{3}$	$-\frac{4}{3} < x < \frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3} < x < +\infty$	$+\infty$
y	$-\infty$	возрастает	$7\frac{2}{9}$ max	убывает	$\frac{5}{18}$ min	возрастает	$+\infty$
y'		+	0	−	0	+	

По таблице легко построить схематический график функции (рис. 4.12).

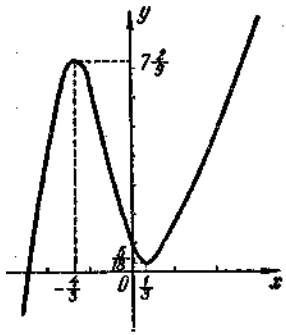


Рис. 4.12

Из графика ясно, между прочим, что уравнение

$$3x^3 + 4{,}5x^2 - 4x + 1 = 0$$

имеет только один действительный корень, т. е. остальные два его корня (напомним, что всякое алгебраическое уравнение имеет столько корней, какова его степень) — сопряженные комплексные числа.

3) Изучим функцию

$$y = \sqrt[3]{x^2}e^x$$

в интервале [—2; 0,5]. В этом интервале функция непрерывна; она будет положительна во всех точках, за исключением точки $x = 0$, где она обращается в нуль. Возьмем производную

$$y' = \frac{2}{3\sqrt[3]{x}}e^x + \sqrt[3]{x^2}e^x = \frac{2+3x}{3\sqrt[3]{x}}e^x.$$

Производная существует в каждой точке заданного интервала, кроме точки x=0. В этой же точке производная бесконечна, причем если подходить к точке $x = 0$ справа, то $y' \to +\infty$, а если слева, то

$y' \to -\infty$. Это указывает на наличие в точке (0, 0) графика функции точки возврата в виде острия (рис. 4.13).

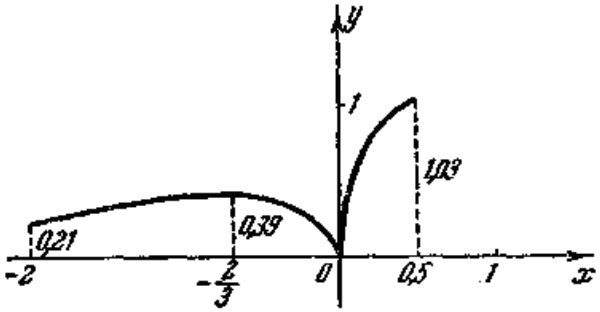


Рис. 4.13.

Производная равна нулю при $x = -\frac{2}{3}$. Следовательно, в данном интервале есть две критические точки: стационарная точка $x = -\frac{2}{3}$ и точка $x = 0$, в которой функция недифференцируема.

Мы имеем три интервала монотонности функции: $\left[-2, -\frac{2}{3}\right)$, $\left(-\frac{2}{3}, 0\right)$, $(0; 0{,}5]$.

Составляя таблицу поведения функции, видим, что в первом интервале производная положительна — функция в нем возрастает; во втором отрицательна — функция в нем убывает; в третьем снова положительна — функция в нем возрастает. Впрочем, это ясно и из сравнения значений функции в критических точках и на концах интервала:

$$y_{x=-2} \approx 0{,}21; \quad y_{x=-\frac{2}{3}} = \sqrt[3]{\frac{4}{9}} e^{-\frac{2}{3}} \approx 0{,}39; \quad y_{x=0} = 0; \quad y_{x=0{,}5} \approx 1{,}03.$$

x	-2	$-2 < x < -\frac{2}{3}$	$-\frac{2}{3}$	$-\frac{2}{3} < x < 0$	0	$0 < x < 0{,}5$	0,5
y	$\approx 0{,}21$	возрастает	$\approx 0{,}39$ max	убывает	0 min	возрастает	$\approx 1{,}03$
y'		+	0	−	∞	+	

Наибольшее значение функции в интервале [—2; 0,5] приближенно равно 1,03 и достигается на правом конце интервала, а наименьшее значение равно 0 и достигается в точке $x = 0$.

4) Методы исследования функций часто могут быть использованы для доказательства неравенств.

Докажем, например, справедливость неравенств

$$\frac{2}{\pi}x < \sin x < x, \quad \text{если } 0 < x < \frac{\pi}{2}.$$

Рассмотрим функцию $y = \frac{\sin x}{x}$. Так как ее производная

$$y' = \frac{\cos x}{x^2}(x - \operatorname{tg} x)$$

в интервале $\left(0, \frac{\pi}{2}\right)$ отрицательна ($x < \operatorname{tg} x$), то y убывает и, значит, функция $\frac{\sin x}{x}$ меньше своего значения при $x = 0$ (т. е. меньше 1) и больше своего значения при $x = \frac{\pi}{2}$ $\left(\text{т. е. больше } \frac{\sin\frac{\pi}{2}}{\frac{\pi}{2}} = \frac{2}{\pi}\right)$:

$$\frac{2}{\pi} < \frac{\sin x}{x} < 1,$$

откуда и следуют наши неравенства.

IV. Задачи о наибольших и наименьших значениях. Предположим, что даны две величины, связанные функциональной зависимостью, и требуется отыскать значение одной из них (заключенное в некотором интервале, который может быть и неограниченным), при котором другая принимает наименьшее или наибольшее возможное значение.

Для решения такой задачи прежде всего следует составить выражение для функции, с помощью которой одна величина выражается через другую, а затем найти наибольшее или наименьшее значение полученной функции в данном интервале.

Примеры. 1) Найдем наименьшую длину l забора, с помощью которого можно огородить участок в форме прямоугольника с данной площадью s, примыкающий к стене (рис. 4.14).

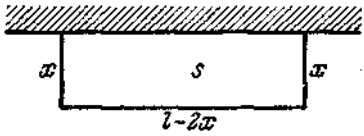


Рис. 4.14

Обозначим одну из сторон прямоугольника через x; тогда, как легко видеть, будем иметь

$$s = x(l - 2x),$$

откуда

$$l = 2x + \frac{s}{x}.$$

Наша задача сводится к отысканию наименьшего значения этой функции при изменении x от 0 до ∞. При $x \to 0$ и при $x \to \infty$ функция $f \to \infty$. Следовательно, наименьшее значение следует искать среди минимумов функции в интервале $(0, \infty)$. Возьмем производную

$$l' = 2 - \frac{s}{x^2}.$$

В интересующем нас интервале имеем одну стационарную точку: $x = \sqrt{\frac{s}{2}}$, которая является точкой минимума функции, ибо $l' < 0$, если $x < \sqrt{\frac{s}{2}}$, и $l' > 0$, если $x > \sqrt{\frac{s}{2}}$. Минимальное значение

$$l_{\min} = 2\sqrt{\frac{s}{2}} + \frac{s}{\sqrt{\frac{s}{2}}} = 2\sqrt{2s}$$

здесь служит и наименьшим значением функции во всем интервале $(0, \infty)$. Значит, какой бы забор, огораживающий прямоугольный участок с площадью s и примыкающий к стене, мы ни взяли, его длина не может быть меньше $2\sqrt{2s}$ и равна этому значению только в том случае, когда меньшая сторона прямоугольника (равная $\sqrt{\frac{s}{2}} = \frac{1}{2}\sqrt{2s}$) в два раза меньше его большей стороны (равной $2\sqrt{2s} - 2 \cdot \frac{1}{2}\sqrt{2s} = \sqrt{2s}$). В указанных условиях самый экономичный забор тот, у которого большая сторона в два раза длиннее меньшей стороны.

Применение второй производной. Точки перегиба.

I. Второй признак экстремума. С помощью второй производной $f''(x)$ исследуемой функции $f(x)$ можно установить так называемый *второй достаточный признак экстремума,* который иногда оказывается более удобным и простым, чем первый.

Второй достаточный признак экстремума. *Точка x_0 есть точка экстремума функции $f(x)$, если $f'(x_0) = 0$, а $f''(x_0) \neq 0$, причем, если $f''(x_0) > 0$, то x_0 — точка минимума, а если $f''(x_0) < 0$, то x_0 — точка максимума.*

Доказательство. Пусть $f'(x_0) = 0$ и $f''(x_0) > 0$. Предполагая вторую производную непрерывной, мы можем считать, что она сохраняет свой знак в некоторой окрестности точки x_0. Отсюда следует, что функция $f'(x)$ в этой окрестности будет возрастающей, потому что ее производная $f''(x)>0$. Так как $f'(x_0) = 0$, то слева от точки x_0 производная $f'(x)$, принимая меньшие значения, будет отрицательной: $f'(x)<0$, а справа, — принимая большие значении,— положительной: $f'(x) > 0$. Итак, функция $f'(x)$ при переходе x через x_0 меняет свой знак с — на + , и, согласно известному нам первому достаточному признаку, точка x_0 является точкой минимума.

Рассуждая аналогично, получим, что если $f''(x_0)<0$, то $f'(x)$ убывает и, переходя через точку x_0, меняет знак с + на — , т. е. точка x_0 является точкой максимума.

В том случае, когда $f'(x_0) = 0$ и $f''(x_0) = 0$, а также в случае, когда первой производной не существует, вторым признаком воспользоваться нельзя и нужно обратиться к первому признаку. Так, например, обе первые производные функции $y = x^4$ равны нулю при $x = 0$; вторым признаком воспользоваться нельзя, а по первому мы устанавливаем, что функция в этой точке достигает минимума: производная $y' = 4x^3$ меняет знак с — на + при переходе x через нуль. Функция же $y = x^3$, хотя ее первые две производные тоже равны нулю при $x = 0$, не имеет экстремума; действительно, ее первая производная $f'=3x^2$ не меняет знака при переходе x через нуль.

Однако в случае своей применимости второй признак оказывается весьма удобным: вместо рассмотрения знака функции $f'(x)$ в точках, отличных от предполагаемой точки экстремума, он позволяет дать ответ по знаку функции $f''(x)$ в той же точке.

Пример. $y — ax^2 + bx+c$. Так как $y' = 2ax + b$ равна нулю при $x=-\frac{b}{2a}$, а $y'' = 2a$, то y имеет максимум в указанной точке, если $a < 0$, и минимум, если $a > 0$.

Этот пример может наглядно убедить в большом значении теории оптимизации для практических целей. Он показывает, что с расширением средств анализа упрощаются и укорачиваются рассуждения и выкладки, необходимые для исследования конкретных функций. Когда мы еще не владели дифференциальным исчислением, изучение квадратичной функции потребовало от нас довольно длинных и специальных рассмотрений, а здесь для этого изучения нам понадобилось лишь три строки.

II. Выпуклость и вогнутость линии. Точки перегиба.

Определение. *Дуга называется выпуклой, если она пересекается с любой своей секущей не более чем в двух точках.*

Дуга *AB* на рис. 4.15 выпуклая, а на рис. 4.16 невыпуклая; у дуги на рис. 4.16 есть секущие M_1M_2, которые, кроме точек M_1 и M_2, пересекаются с дугой *AB* еще в других, отличных от этих точках M_3.

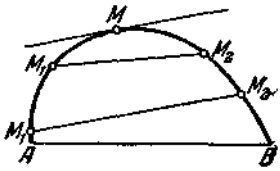


Рис. 4.15.

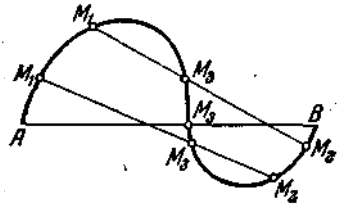


Рис. 4.16

Если дуга выпуклая, то она целиком лежит по одну сторону от касательной, проведенной в любой ее точке.

Будем теперь рассматривать линии, являющиеся графиками непрерывных функций $y = f(x)$. Если такая линия выпуклая, то ее выпуклость обращена или вверх (рис. 4.17, *I*), или вниз (рис. 4.17, *II*).

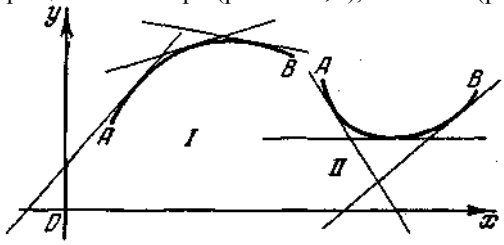


Рис. 4.17.

Линии, обращенные выпуклостью вверх, условились называть *выпуклыми,* а обращенные выпуклостью вниз — *вогнутыми.* Геометрически ясно, что выпуклая дуга лежит под любой своей касательной, а вогнутая дуга — над касательной.

Особую роль играют точки на линии, отделяющие выпуклую дугу от вогнутой; они называются точками перегиба. На рис. 4.18 точка *C* отделяет выпуклую дугу *AC* от вогнутой *CB*.

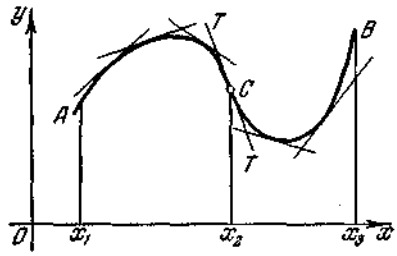


Рис. 4.18.

Определение. *Точкой перегиба называется такая точка линии, которая отделяет выпуклую дугу от вогнутой.*

При этом предполагается, что в этой точке к линии можно провести касательную. Предоставляем читателю нарисовать линию с угловой точкой, которая отделяла бы выпуклую дугу от вогнутой; такие угловые точки мы не причисляем к точкам перегиба.

В точке перегиба касательная пересекает линию; в окрестности этой точки линия лежит по обе стороны от касательной.

Установление участков выпуклости и вогнутости и точек перегиба линии $y=f(x)$ имеет важное значение для характеристики этой линии, а значит, и для характеристики поведения функции $f(x)$.

III. Признаки выпуклости и вогнутости линии. Связь между второй производной $f''(x)$ и выпуклостью или вогнутостью линии $y= f(x)$ мы установим, пользуясь лишь самыми простыми геометрическими соображениями. В основе лежит следующее предложение.

Интервалу убывания первой производной соответствует участок выпуклости графика функции, а интервалу возрастания — участок вогнутости.

В самом деле, если дуга выпуклая, то при перемещении точки касания слева направо угловой коэффициент касательной, т. е. $f'(x)$, как это видно из рис. 4.18, уменьшается: сначала он принимает все меньшие положительные значения, потом становится равным нулю, а затем и отрицательным. Таким образом, интервал $[x_1, x_2]$ есть интервал убывания функции $f'(x)$.

Верно и обратное: если функция $f'(x)$ убывает, то убывает и угловой коэффициент касательной к линии; при этом дуга кривой будет лежать под любой своей касательной, т. е. она выпуклая. (Это утверждение становится более наглядным, если представить себе кривую, как бы построенную из «бесконечно малых» отрезков касательных.)

Совершенно аналогично можно установить, что если дуга вогнутая, то функция $f'(x)$ возрастает, и наоборот. Так, для графика функции,

изображенного на.рис. 4.18, дуга *СВ*—вогнутая и $[x_2, x_3]$ — интервал возрастания *f (x).*

Воспользуемся теперь основной теоремой, устанавливающей связь между характером изменения функции и знаком ее производной, приняв за функцию — *f{x)* и за ее производную — *f"(x).* Если *f* "(*x*)> 0, то *f(x)* возрастает, а если *f* " *(x)* < 0, то *f* ' *(x)* убывает. В результате мы получаем следующую теорему.

Теорема. *Если вторая производная f"(x) всюду в интервале отрицательна, то дуга линии y=f(x), соответствующая этому интервалу, выпуклая. Если вторая производная f"(x) всюду в интервале положительна, то дуга линии y=f(x), соответствующая этому интервалу, вогнутая.*

Изложенная теорема делает особенно ясным второй достаточный признак экстремума.

Если *f* ' (x_0) = 0 и *f* " (x_0) < 0, то вершина кривой, соответствующая точке x_0, лежит на выпуклой части линии *y=f(x),* и поэтому точка x_0 является точкой максимума. Если же $f'(x_0)=0$ и $f''(x_0)>0$, то соответствующая вершина лежит на вогнутой части линии, и поэтому точка x_0 является точкой минимума.

IV. Признаки точки перегиба. Мы определили точку перегиба как такую точку графика, которая отделяет выпуклую дугу от вогнутой. Из приведенного выше рассуждения следует, что абсцисса точки перегиба (точка x_2 на рис. 4.18) разделяет два интервала монотонности первой производной *f'(x),* т. е. является *точкой экстремума для первой производной.* Применяя необходимый признак экстремума, получаем:

Если x_0 — абсцисса точки перегиба, то либо f "(x_0) = 0, либо f "(x_0) не существует.

Это и дает нам необходимый признак точки перегиба.

Однако не всякий корень уравнения *f* "*(x)*=0 является абсциссой точки перегиба. Например, функция $y = x^4$ имеет производные $y' = 4x^3$ и $y''=12x^2$. Несмотря на то, что $y''_{x=0} = 0$, точка (0,0) является вершиной кривой и не является точкой перегиба.

Для того чтобы выяснить, когда же точка x_0 будет действительно являться абсциссой точки перегиба, нужно проследить еще, меняет ли вторая производная свой знак при переходе через эту точку. Таким образом, мы и приходим к достаточному признаку точки перегиба:

Точка (x_0, y_0) есть точка перегиба линии y=f(x), если f " (x) меняет знак при переходе x через x_0. При перемене знака с — на + слева от нее лежит участок выпуклости, а справа — участок вогнутости; при перемене знака с + на —, наоборот, участок вогнутости сменяется участком выпуклости.

Приведем еще второй достаточный признак точки перегиба, аналогичный второму достаточному признаку экстремума.

Точка (x_0, y_0) есть точка перегиба линии $y=f(x)$, если $f''(x_0) = 0$, а $f'''(x_0) \neq 0$; при $f'''(x_0) > 0$ слева от нее лежит участок выпуклости, справа — участок вогнутости, а при $f'''(x_0) < 0$ слева лежит участок вогнутости, а справа — участок выпуклости.

В заключение отметим, что при отыскании точек перегиба и интервалов выпуклости и вогнутости мы пользуемся теми же правилами, что и при отыскании точек экстремума, применяя, однако, эти правила не к самой функции, а к ее первой производной. При этом *интервалу возрастания первой производной соответствует участок вогнутости графика функции, интервалу убывания — участок выпуклости и точке экстремума первой производной — абсцисса точки перегиба.*

V. Пример. Характер графиков основных элементарных функций в смысле их выпуклости теперь может быть очень легко проверен.

Так график степенной функции $y = x^n$ на положительной полуоси Ox вогнутый при $n<0$ и $n> 1$ и выпуклый при $0<n<1$. Действительно, вторая производная $y'' = n(n-1)x^{n-2}$ положительна в первых случаях и отрицательна во втором.

График показательной функции $y=a^x$ при любом $a>0$ вогнут на всей оси Ox, так как $y'' = a^x \ln^2 a > 0$.

Логарифмика $y = \log_a x$ выпукла при $a > 1$ и вогнута при $a < 1$. Действительно, вторая производная $y'' = -\frac{1}{\ln a}\frac{1}{x^2}$ отрицательна в первом случае и положительна во втором.

Так как

$$(\sin x)'' = -\sin x,$$

то участки выпуклости синусоиды расположены выше оси абсцисс, а участки вогнутости — ниже; точками перегиба синусоиды являются точки ее пересечения с осью Ox.

Выпуклые функции.

Понятие выпуклости играет огромную роль в теории экстремальных задач, и мы будем многократно обращаться к нему. Числовая функция $f(x)$ на R^n называется *выпуклой,* если

$$(\lambda x + (1-\lambda)y) \leqslant \lambda f(x) + (1-\lambda) f(y) \qquad (4.1)$$

для любых $x, y \in R^n$, $0 \leqslant \lambda \leqslant 1$. Это определение имеет наглядный геометрический смысл — график функции на отрезке $[x, y]$ лежит ниже хорды, соединяющей точки $(x, f(x))$ и $(y, f(y))$ (рис. 4.19).

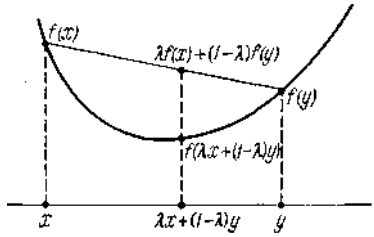


Рис. 4.19. Выпуклая функция.

В определении выпуклости фигурируют две точки *x, y* и их выпуклые комбинации. Совершенно аналогичное неравенство справедливо для любого числа точек и их выпуклых комбинаций.

Лемма 1 (неравенство Иенсена). *Пусть f(x) — выпуклая функция на* R^n. *Тогда для любых* $x^1 ..., x^k \in R^n$ *и* $\lambda_i \geq 0$,

$$i = 1, \ldots, k, \quad \sum_{i=1}^{k} \lambda_i = 1,$$

$$f(\lambda_1 x^1 + \ldots + \lambda_k x^k) \leqslant \lambda_1 f(x^1) + \ldots + \lambda_k f(x^k). \qquad (4.2)$$

Функция *f(x),* для которой —*f(x)* является выпуклой, называется *вогнутой* Очевидно, что *аффинная* функция *f(x)* =— *(a, x)* + β является и выпуклой и вогнутой.

Из определения очевидно, что если $f_i(x)$ выпуклы, *i*=1,.... *m,* то и $f(x) = \sum_{i=1}^{m} \gamma_i f_i(x), \quad \gamma_i \geqslant 0,$ и $f(x) = \max_{1 \leqslant i \leqslant m} f_i(x)$ также будут выпуклы.

Важным частным случаем выпуклых функции являются строго и сильно выпуклые функции. Функция *f(x)* на R^n называется *строго выпуклой,* если для любых $x \neq y$, $0 < \lambda < 1$,

$$f(\lambda x + (1-\lambda) y) < \lambda f(x) + (1-\lambda) f(y), \qquad (4.3)$$

и *сильно выпуклой с константой* $l > 0$, если при $0 \leq \lambda \leq 1$

$$f(\lambda x + (1-\lambda) y) \leqslant \lambda f(x) + (1-\lambda) f(y) - l\lambda(1-\lambda)\| x - y \|^2/2. \qquad (4.4)$$

Ясно, что сильно выпуклая функция строго выпукла.

Важно иметь аналитические критерии, по которым можно судить о выпуклости функций. Такие критерии существуют и очень просты для случая дифференцируемых функций. Они основываются на следующем элементарном результате.

Лемма 2. *Пусть* $\psi(\tau)$ *— дифференцируемая функция на* R^n. *Тогда выпуклость* $\psi(\tau)$ *эквивалентна монотонности производной* $(\psi'(\tau_1) \geqslant \psi'(\tau_2)$ *при* $\tau_1 \geq \tau_2$), *строгая выпуклость — строгой*

монотонности $(\psi'(\tau_1) > \psi'(\tau_2)$ *при* $\tau_1>\tau_2$*), а сильная выпуклость — сильной монотонности* $(\psi(\tau_1) - \psi(\tau_2) \geqslant l(\tau_1 - \tau_2)$, $\tau_1>\tau_2$.

Лемма 3. Для дифференцируемой функции f(x) на R^n *выпуклость эквивалентна неравенству*

$$f(x+y) \geqslant f(x) + (\nabla f(x),\ y), \tag{4.5}$$

(здесь ∇f обозначает градиент скалярной функции $f(x)$)
строгая выпуклость — неравенству

$$f(x+y) > f(x) + (\nabla f(x),\ y), \qquad y \neq 0, \tag{4.6}$$

а сильная выпуклость — неравенству

$$f(x+y) \geqslant f(x) + (\nabla f(x),\ y) + l\|y\|^2/2 \tag{4.7}$$

для любых $x, y \in R^n$.

Иначе говоря, график (строго) выпуклой функции лежит (строго) выше касательной гиперплоскости, а для сильно выпуклой функции график лежит выше некоторого параболоида (рис. 4.20).

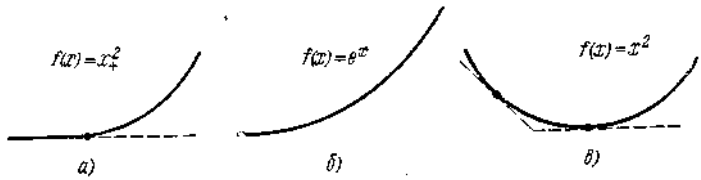


Рис. 4.20. Типы выпуклости: *а*) выпуклая функция; *б*) строго выпуклая функция; *в)* сильно выпуклая функция.

Из (4.5) получаем полезное неравенство

$$(\nabla f(x) - \nabla f(y),\ x - y) \geqslant 0, \tag{4.8}$$

являющееся обобщением условия *монотонности* производной выпуклой функции на многомерный случай. Для строго выпуклой функции справедливо условие *строгой монотонности*

$$(\nabla f(x) - \nabla f(y),\ x - y) > 0, \qquad x \neq y, \tag{4.9}$$

а для сильно выпуклой — условие *сильной монотонности*

$$(\nabla f(x) - \nabla f(y),\ x - y) \geqslant l\|x - y\|^2. \tag{4.10}$$

Наиболее просто критерий выпуклости формулируется для дважды дифференцируемых функций *f(x):* выпуклость эквивалентна выполнению условия

$\nabla^2 f(x) > 0$, (4.11)

(здесь $\nabla^2 f$ обозначает матрицу вторых производных, гессиан)

а сильная выпуклость — выполнению условия

$$\nabla^2 f(x) > lI \qquad (4.12)$$

для всех x. Если же

$$\nabla^2 f(x) > 0 \qquad (4.13)$$

для всех x, то $f(x)$ строго выпукла. Последнее условие является лишь достаточным (например, для строго выпуклой функции $f(x)=||x||^4$ будет $\nabla^2 f(0) = 0$).

Пусть x^* — точка минимума дифференцируемой сильно выпуклой (с константой l) функции $f(x)$. Такая точка заведомо существует, единственна и $\nabla f(x^*) = 0$ (см. ниже). Поэтому из неравенств (4.7), (4.10) получаем

$$f(x) \geqslant f(x^*) + l\|x - x^*\|^2/2, \qquad (4.14)$$

$$(\nabla f(x),\ x - x^*) \geqslant l\|x - x^*\|^2, \qquad (4.15)$$

$$\|\nabla f(x)\| \geqslant l\|x - x^*\|. \qquad (4.16)$$

4.3. Условия экстремума

Условия экстремума гладких функций на всем пространстве хорошо известны. Мы рассмотрим их, однако, достаточно подробно, так как они служат моделью, по которой строятся аналогичные условия в более сложных случаях.

1. Необходимое условие I порядка. Точка x^* называется *локальным минимумом* $f(x)$ на R^n, если найдется $\varepsilon > 0$ такое, что $f(x) \geq f(x^*)$ для всех x из ε-окрестности x^* (т. е. при $||x - x^*|| \leq \varepsilon$). Иногда в таком случае говорят просто о *точке минимума,* отбрасывая слово «локальный». Нужно, однако, иметь в виду разницу между локальным и *глобальным* минимумом (т. е. точкой x^* такой, что $f(x) \geq f(x^*)$ для всех x). В необходимых условиях экстремума можно говорить просто о точке минимума, поскольку если некоторое условие выполняется для локального минимума, то оно же справедливо для глобального. При формулировке достаточных условий нужно различать, какой из типов минимума подразумевается.

Теорема 1 (Ферма). *Пусть x^* — точка минимума $f(x)$ на R^n и $f(x)$ дифференцируема в x^*. Тогда*

$$\nabla f(x^*) = 0. \qquad (4.17)$$

Доказательство. Пусть $\nabla f(x^*) \neq 0$. Тогда

$$f(x^* - \tau\nabla f(x^*)) = f(x^*) - \tau\|\nabla f(x^*)\|^2 + o(\tau\nabla f(x^*)) =$$
$$= f(x^*) - \tau(\|\nabla f(x^*)\|^2 + \tau^{-1}o(\tau)) < f(x^*)$$

для достаточно малых $\tau>0$ по определению $o(\tau)$. Но это противоречит тому, что x^* — точка локального минимума.

В предположении, что условие экстремума не выполняется, показано, как построить точку с меньшим значением *f(x).* Таким образом, это доказательство указывает путь для построения метода минимизации. Такой метод (он называется градиентным) будет подробно изучаться дальше.

2. Достаточное условие I порядка. Очевидно, если какая-либо точка является *стационарной* (т. е. градиент в ней обращается в 0), то она не обязана быть точкой минимума (рис. 4.21) — например, она может быть точкой максимума или седловой точкой. Для выпуклых функций, однако, такая ситуации невозможна.

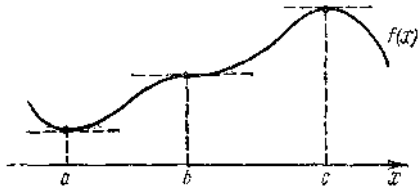


Рис. 4.21
Стационарные точки:
a — точка минимума, *b* —точка перегиба, *c* — точка максимума.

Теорема 2. *Пусть f(x) — выпуклая функция, дифференцируемая в точке x*, и* $\nabla f(x^*) = 0$. *Тогда x* точка глобального минимума f(x), на* R^n.

Доказательство следует сразу из формулы (4.5), так как $f(x) \geq f(x^*) + (\nabla f(x^*), x - x^*) = f(x^*)$ для любого $x \in R^n$.

Таким образом, для выпуклых функций необходимое условие экстремума является и достаточным. Дальше мы увидим, что эта ситуация является общей и для других типов выпуклых экстремальных задач.

3. Необходимое условие II порядка. Для невыпуклых задач можно продолжить исследование условий экстремума с помощью старших производных.

Теорема 3. *Пусть x* — точка минимума f(x) на* R" *и f(x) , дважды дифференцируема в x*. Тогда*

$$\nabla f(x^*) \geq 0. \qquad (2)$$

Доказательство. По теореме 1 $\nabla f(x^*)=0$, поэтому для . произвольного y и достаточно малых τ

$$f(x^*) \leqslant f(x^* + \tau y) = f(x^*) + \tau^2 (\nabla^2 f(x^*)\, y,\ y)/2 + o(\tau^2),$$
$$(\nabla^2 f(x^*)\, y,\ y) \geqslant o(\tau^2)/\tau^2.$$

Переходя к пределу при $\tau \to 0$, получаем $(\nabla^2 f(x^*)y, y \geq 0$. В силу произвольности y это означает, что $\nabla^2 f(x^*) \geq 0$.

Достаточное условие II порядка.

Теорема 4. *Пусть в точке x^* $f(x^*)$ дважды дифференцируема, выполнено необходимое условие* I *порядка (т. е.* $\nabla f(x^*) = 0$) и

$$\nabla^2 f(x^*) > 0. \qquad (4.19)$$

Тогда x^* — точка локального минимума.

Доказательство. Пусть y — произвольный вектор с единичной нормой. Тогда

$$f(x^* + \tau y) = f(x^*) + \tau^2 (\nabla^2 f(x^*)\, y,\ y)/2 + o(\tau^2 \| y \|^2) \geqslant$$
$$\geqslant f(x^*) + \tau^2 l/2 + o(\tau^2),$$

где $l > 0$ — наименьшее собственное значение $\nabla^2 f(x^*)$, а функция $o(\tau^2)$ не зависит от y. Поэтому найдется τ_0 такое, что при $0 \leq \tau \leq \tau_0$ будет $\tau^2 l/2 \geq o(\tau^2)$, т. е. $f(x^* + \tau y) \geq f(x^*)$.

Если в точке x^* выполняются необходимые условия I и II порядков (т. е. $\nabla f(x^*) = 0$, $\nabla^2 f(x^*) \geq 0$), но не выполняется достаточное условие II порядка (матрица $\nabla^2 f(x^*)$ не является положительно определенной), то x^* может и не являться точкой минимума (например, $f(x) = x^3$, $x \in R^1$) и в принципе анализ можно продолжить с помощью старших производных. Для одномерного случая правило действий хорошо известно (нужно найти первую отличную от 0 производную), для многомерного случая техника вычислений сложна.

Как мы уже отмечали, предлагается следующий рецепт для отыскания точек экстремума. Нужно найти все точки, удовлетворяющие необходимому условию I порядка, а затем полученные точки исследовать с помощью условий II порядка, отобрав из них точки минимума. Таким образом, создается впечатление, что условия экстремума — эффективный инструмент для решения задач оптимизации.

Но это не совсем так. Отыскать в явной форме точку минимума с помощью условий экстремума удается лишь в редких случаях, для специально построенных примеров (они обычно и приводятся в книгах). Дело в том, что решение системы уравнений $\nabla f(x) = 0$ — задача ничуть не более простая, чем исходная, и явный вид ответа в ней найти, как правило, нельзя.

Зачем же в таком случае нужны условия экстремума и почему им уделяется столь большое внимание в теории экстремальных задач? Отчасти такое внимание является данью традиции, когда численные методы оптимизации не изучались, а решением задачи считалось лишь некоторое аналитическое выражение. Нередко при этом вывод условий экстремума для различных типов экстремальных задач превращается в чисто математическую игру, где целью является получение изощренных формулировок для разного рода вырожденных ситуаций без всякой заботы о том, как пользоваться этими условиями экстремума. При чтении некоторых монографий создается впечатление, что формулировка условий оптимальности является главным (или даже единственным) объектом исследования в области экстремальных задач. На взгляд ряда авторов условия экстремума являются той основой, на которой строятся методы решения оптимизационных задач, — с этой точки зрения и нужно рассматривать вопрос о их полезности. В дальнейшем мы увидим, что, во-первых, в ряде случаев условия экстремума хотя и не дают возможности явно найти решение, но сообщают много информации о его свойствах. Во-вторых, доказательство условий экстремума или вид этих условий часто указывают путь построения методов оптимизации. Мы уже видели выше, что доказательство условия $\nabla f(x^*)=0$ естественно приводит к градиентному методу минимизации. В-третьих, при обосновании методов приходится делать ряд предположений. Обычно при этом требуется, чтобы в точке x^* выполнялось достаточное условие экстремума. Таким образом, условия экстремума фигурируют в теоремах о сходимости методов. Наконец, сами доказательства сходимости обычно строятся на том, что показывается, как «невязка» в условии экстремума стремится к нулю.

4.4. Существование, единственность, устойчивость минимума

Важной частью математической теории экстремальных задач (и в частности, задач безусловной оптимизации) являются проблемы существования, единственности и устойчивости решения.

1. Существование решения. Вопрос о существовании точки минимума обычно решается совсем просто с помощью следующей теоремы.

Теорема 1 (Вейерштрасс). *Пусть $f(x)$ непрерывна на R^n и множество $Q_\alpha = \{x - f(x) \le \alpha\}$ для некоторого α непусто и ограничено. Тогда существует точка глобального минимума $f(x)$ на R^n.*

Доказательство. Пусть

$$f(x^k) \to \inf_{x \in R^n} f(x) > \alpha,$$

тогда $x^k \in Q_\alpha$ для достаточно больших k. Множество Q_α замкнуто (в силу непрерывности $f(x)$) и ограничено, т. е. компактно, а потому у последовательности x^k существует предельная точка, $x^* \in Q_\alpha$. Из непрерывности $f(x)$ следует, что $f(x^*) = \inf_{x \in R^n} f(x)$,

т. е.

$$x^* = \arg\min_{x \in R^n} f(x).$$

Предположение об ограниченности Q_α существенно (например, функции x и $1/(1+x^2)$ непрерывны на R^1, но не имеют точки минимума). В некоторых случаях можно доказать существование решения и в ситуациях, не охватываемых теоремой 1.

2. Единственность решения. Будем называть точку минимума *локально единственной,* если в некоторой ее окрестности нет других локальных минимумов. Будем говорить, что x^* — *невырожденная точка минимума,* если в ней выполнено достаточное условие экстремума II порядка, т. е. $\nabla f(x^*) = 0$, $\nabla^2(x^*) > 0$.

Теорема 2. *Невырожденная точка минимума локально единственна.*

Для выпуклых функций ответ на вопрос об единственности минимума часто может быть получен совсем просто.

Теорема 3. *Точка минимума строго выпуклой функции (глобально) единственна.*

Доказательство следует непосредственно из определения строгой выпуклости.

3. Устойчивость решения. При практическом решении задач оптимизации постоянно приходится сталкиваться со следующими проблемами. Пусть метод оптимизации приводит к построению минимизирующей последовательности, следует ли отсюда ее сходимость к решению? Если вместо исходной задачи минимизации решается задача, близкая к ней, можно ли утверждать близость их решений? Такого типа проблемы относятся к области теории экстремальных задач, связанной с понятиями устойчивости и корректности. Мы будем пользоваться термином «устойчивость» задач оптимизации, оставляя термин «корректность» для задач, не связанных с оптимизацией (решение алгебраических, интегральных, операторных уравнений и т. п.).

Точка x^* локального минимума $f(x)$ называется *локально устойчивой,* если к ней сходится любая *локальная минимизирующая последовательность,* т. е. если найдется $\delta>0$ такое, что из $f(x^k)\to f(x^*)$, $\|x^k - x^*\| \leqslant \delta$ следует $x^k \to x^*$.

Теорема 4. *Точка локального минимума непрерывной функции f(x) локально устойчива тогда и только тогда, когда она локально единственна.*

Доказательство. Пусть x^* локально единственна. Возьмем произвольную локальную минимизирующую последовательность x^k, $\|x^k - x^*\| \leqslant \delta$, $f(x^k)\to f(x^*)$. В силу компактности шара в R^n из нее можно выбрать сходящуюся подпоследовательность $x^{k_i}\to \bar{x}$, $\|\bar{x} - x^*\| \leqslant \delta$. Из непрерывности $f(x)$ следует, что

$$f(\bar{x}) = \lim f(x^{k_i}) = f(x^*),$$

но тогда $\bar{x} = x^*$ (так как x^* локально единственная точка минимума). Поскольку это же верно для любой другой подпоследовательности, то и вся последовательность x^k сходится к x^*. Таким образом, x^* локально устойчива. Обратно, пусть x^* локально устойчива, но существует другая точка минимума $x_1^* \neq x^*$, $\|x_1^* - x^*\| \leqslant \delta$. Тогда $f(x^*_1) = f(x^*)$. Возьмем последовательность точек $x^1, x^2, \ldots$, поочередно совпадающих то с x^*, то с x^*_1. Она является минимизирующей, но не сходится, что противоречит локальной устойчивости x^*.

Аналогично доказывается следующий результат.

Теорема 5. *Пусть x^* — локально устойчивая точка минимума непрерывной функции f(x), а g(x) — непрерывная функция. Тогда для достаточно малых $\varepsilon>0$ функция $f(x)+\varepsilon g(x)$ имеет локально единственную точку минимума x_ε в окрестности x^* и $x_\varepsilon \to x^*$ при $\varepsilon \to 0$.*

Таким образом, из устойчивости следует близость точек минимума исходной и «возмущенной» функции.

Невырожденная точка минимума, как следует из теорем 2 п 4, является локально устойчивой. В этом случае результат теоремы 5 можно уточнить.

Теорема 6. *Пусть x^* — невырожденная точка минимума $f(x)$, а функция $g(x)$ непрерывно дифференцируема в окрестности x^*. Тогда для достаточно малых $\varepsilon>0$ существует x_ε — локальная точка минимума функции $f(x)+\varepsilon\, g(x)$ в окрестности x^*, причем*

$$x_\varepsilon = x^* - \varepsilon\,[\nabla^2 f(x^*)]^{-1}\nabla g(x^*) + o(\varepsilon). \qquad (4.20)$$

Можно ввести и глобальное понятие устойчивости точек минимума. Для этого нужно в определении слово «локальный» заменить на «глобальный». Именно, точка глобального минимума называется *глобально устойчивой,* если к ней сходится любая минимизирующая

последовательность. Будем в этом случае говорить о глобальной устойчивости задачи минимизации. Повторяя почти дословно доказательство теоремы 4, получаем, что если x^* — единственная точка глобального минимума непрерывной функции $f(x)$ и множество $Q_\alpha = \{x: f(x) \leqslant \alpha\}$ непусто и ограничено для некоторого $\alpha > f(x^*)$, то x^* глобально устойчива. Требование ограниченности Q_α существенно. Например, у функции $f(x) = x^2/(1+x^4)$, $x \in \mathbb{R}^1$, точка глобального минимума $x^* = 0$ единственна, но не глобально устойчива (так как минимизирующая последовательность $x^k \to \infty$ не сходится к x^*).

Можно было бы ввести следующее более широкое определение устойчивости, которое не предполагает единственности минимума, Множество X^* точек глобального минимума $f(x)$ назовем *слабо устойчивым,* если все предельные точки любой минимизирующей последовательности принадлежат X^*.

Помимо качественной характеристики (устойчива или нет точка минимума), важно иметь количественные оценки устойчивости. Такие оценки, позволяющие судить о близости точки x к решению x^*, если $f(x)$ близко к $f(x^*)$, уже были получены ранее для сильно выпуклых функций. Именно, из (4.14) получаем

$$\| x - x^* \|^2 \leqslant 2l^{-1}(f(x) - f(x^*)), \tag{4.21}$$

где l — константа сильной выпуклости. Аналогичная локальная оценка справедлива для невырожденной точки минимума:

$$\| x - x^* \|^2 \leqslant 2l^{-1}(f(x) - f(x^*)) + o(f(x) - f(x^*)), \tag{4.22}$$

где l — наименьшее собственное значение матрицы $\nabla^2 f(x^*)$.

Таким образом, число l характеризует «запас устойчивости» точки минимума. Оно, однако, не всегда удобно как мера устойчивости— например, оно меняется при умножении $f(x)$ на константу. Поэтому часто используют следующий «нормированный» показатель. Назовем *обусловленностью* точки минимума x^* число

$$\mu = \overline{\lim_{\delta \to 0}} \left(\sup_{x \in L_\delta} \| x - x^* \|^2 / \inf_{x \in L_\delta} \| x - x^* \|^2 \right),$$
$$L_\delta = \{x: f(x) = f(x^*) + \delta\}. \tag{4.23}$$

Иначе говоря, μ характеризует степень вытянутости линий уровня $f(x)$ в окрестности x^*. Ясно, что всегда $\mu \geq 1$. Если μ велико, то линии уровня сильно вытянуты — функция имеет *овражный* характер, т. е. резко возрастает по одним направлениям и слабо меняется по другим. В таких случаях говорят о *плохо обусловленных* задачах минимизации. Если же μ близко к 1, то линии уровня $f(x)$ близки к сферам; это соответствует хорошо обусловленным задачам. В дальнейшем мы увидим, что число обусловленности μ возникает во многих проблемах,

связанных с безусловной минимизацией, и может служить одним из показателей сложности задачи. Для квадратичной функции

$$f(x) = (Ax,\ x)/2 - (b,\ x), \qquad A > 0 \tag{4.24}$$

имеем $L_\delta = \{x:\ (A(x - x^*),\ x - x^*) = 2\delta\}$, поэтому максимум $||x - x^*||$ по $x \in L_\delta$ достигается на векторе $x_1 = x^* + \gamma_1 l_1$, где l_1 — нормированный собственный вектор, отвечающий наименьшему собственному значению λ_1 матрицы A, а множитель γ_1 определяется из условия $x_1 \in L_\delta$, т. е. $\lambda_1 \gamma_1^2 = 2\delta$, $\gamma_1 = (2\delta/\lambda_1)^{1/2}$. Аналогично минимум $||x - x^*||$ по $x \in L_\delta$ достигается на векторе $x_n = x^* + \gamma_n l_n$, l_n — собственный вектор, отвечающий наибольшему собственному -значению λ_n, $\gamma_n = (2\delta/\lambda_n)^{1/2}$ (рис. 4.22).

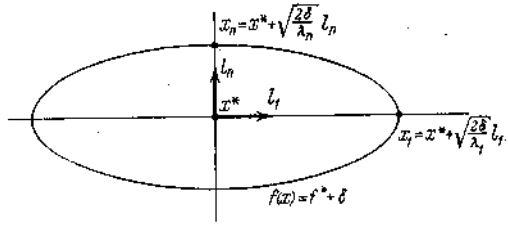


Рис. 4.22. Обусловленность квадратичной функции.

Таким образом, отношение

$$\mu(\delta) = \|x_1 - x^*\|^2 / \|x_n - x^*\|^2 = \gamma_1^2/\gamma_n^2 = \lambda_n/\lambda_1$$

фактически не зависит от δ и

$$\mu = \frac{\lambda_n}{\lambda_1} \tag{4.25}$$

Заметим, что отношение наибольшего к наименьшему собственному значению называется в линейной алгебре *числом обусловленности матрицы.*

Для случая неквадратичиой функции обусловленность задачи ее минимизации равна числу обусловленности гессиана в точке минимума. Именно, если x^* — невырожденная точка минимума, то

$$\mu = \frac{L}{l} \tag{4.26}$$

где L — наибольшее, а l — наименьшее собственное значение матрицы $\nabla^2 f(x^*)$.

5. Правило Лопиталя. Схема исследования функции

5.1. Правило Лопиталя.

I. Основные случаи. В дополнение к известным уже нам правилам предельного перехода приведем здесь еще одно очень удобное и простое правило, называемое *правилом Лопиталя.* В дальнейшем мы его также будем применять к исследованию функций. Оно может быть сформулировано в виде следующей теоремы.

Теорема Лопиталя. *Пусть функции f(x) и φ(x) при* $x \to x_0$ *(или* $x \to \infty$*) совместно стремятся к нулю или к бесконечности. Если отношение их производных имеет предел, то отношение самих функций также имеет предел, равный пределу отношения производных, т. е.*

$$\lim_{x \to x_0} \frac{f(x)}{\varphi(x)} = \lim_{x \to x_0} \frac{f'(x)}{\varphi'(x)}.$$

Доказывать эту теорему в общем виде мы не будем, а ограничимся лишь рассмотрением простейших случаев и примеров, разъясняющих суть дела.

Докажем прежде всего, что *если при* $x \to x_0$ *функции f(x) и φ(x) стремятся к нулю и их производные в точке* x_0 *существуют, причем* $\varphi'(x_0) \neq 0$, *то*

$$\lim_{x \to x_0} \frac{f(x)}{\varphi(x)} = \frac{f'(x_0)}{\varphi'(x_0)}. \qquad (*)$$

Разумеется, здесь предполагается, что функции *f(x)* и *φ(x)* определены и непрерывны в некоторой окрестности точки x_0 и знаменатель *φ (x)* не обращается в нуль в точках этой окрестности, за исключением самой точки x_0.

Поскольку $f(x_0) = \varphi(x_0) = 0$, то

$$\frac{f(x)}{\varphi(x)} = \frac{f(x) - f(x_0)}{\varphi(x) - \varphi(x_0)} = \frac{\dfrac{f(x) - f(x_0)}{x - x_0}}{\dfrac{\varphi(x) - \varphi(x_0)}{x - x_0}}.$$

Переходя к пределу при $x \to x_0$ и используя теорему о пределе дроби, получим

$$\lim_{x \to x_0} \frac{f(x)}{\varphi(x)} = \frac{\lim\limits_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}}{\lim\limits_{x \to x_0} \frac{\varphi(x) - \varphi(x_0)}{x - x_0}},$$

что в силу определенияпроизводной дает

$$\lim_{x \to x_0} \frac{f(x)}{\varphi(x)} = \frac{f'(x_0)}{\varphi'(x_0)}.$$

Примеры.

1) $\lim\limits_{x \to 0} \frac{\sin x}{x} = \frac{(\sin x)'_{x=0}}{1} = 1.$

2) $\lim\limits_{x \to 0} \frac{\ln(1+x)}{x} = \frac{\left(\frac{1}{1+x}\right)_{x=0}}{1} = 1.$

Мы видим, как просто отыскиваются теперь пределы. Конечно, простота эта в данном случае только кажущаяся, ибо мы пользуемся дифференцированием функций $\sin x$ и $\ln(1+x)$, а оно само опирается на знание пределов отношений

$$\frac{\sin x}{x} \text{ и } \frac{\ln(1+x)}{x}$$

при $x \to 0$.

3) $\lim\limits_{x \to 2} \frac{x^2 - 5x + 6}{x^2 - 3x + 2} = \frac{2 \cdot 2 - 5}{2 \cdot 2 - 3} = -1.$

4) $\lim\limits_{x \to \pi} \frac{\sin 5x}{\sin 2x} = \frac{5 \cos 5\pi}{2 \cos 2\pi} = -\frac{5}{2}.$

Может случиться, что при $x = x_0$ производная $\varphi'(x_0)$ равна нулю. Тогда формула (*) неприменима и приходится пользоваться правилом Лопиталя в его общей форме:

$$\lim_{x \to x_0} \frac{f(x)}{\varphi(x)} = \lim_{x \to x_0} \frac{f'(x)}{\varphi'(x)}. \qquad (**)$$

5) $\lim\limits_{x \to 0} \frac{x - x\cos x}{x - \sin x} = \lim\limits_{x \to 0} \frac{1 - \cos x + x \sin x}{1 - \cos x}.$

Здесь производные числителя и знаменателя в свою очередь стремятся к нулю. Применяя правило Лопиталя повторно, а в случае необходимости и далее, получим

$$\lim_{x\to 0}\frac{1-\cos x+x\sin x}{1-\cos x}=\lim_{x\to 0}\frac{2\sin x+x\cos x}{\sin x}=\lim_{x\to 0}\frac{3\cos x-x\sin x}{\cos x}=3.$$

Заметим, что формула (**) остается справедливой и тогда, когда отношение производных стремится к бесконечности; тогда и отношение самих функций тоже стремится к бесконечности.

$$6)\ \lim_{x\to 0}\frac{1-\cos x}{x-\sin x}=\lim_{x\to 0}\frac{\sin x}{1-\cos x}=\lim_{x\to 0}\frac{\cos x}{\sin x}=\infty.$$

При повторном применении правила Лопиталя рекомендуется сначала произвести все возможные упрощения, например сократить общие множители и использовать уже знакомые пределы.

$$7)\ \lim_{x\to 0}\frac{\operatorname{arctg} x-\arcsin x}{\operatorname{tg} x-\sin x}=\lim_{x\to 0}\frac{\dfrac{1}{1+x^2}-\dfrac{1}{\sqrt{1-x^2}}}{\dfrac{1}{\cos^2 x}-\cos x}=$$

$$=\lim_{x\to 0}\frac{\sqrt{1-x^2}-1-x^2}{1-\cos x}\cdot\frac{\cos^2 x}{(1+\cos x+\cos^2 x)(1+x^2)\sqrt{1-x^2}}.$$

Ясно, что предел второго множителя равен $^1/_3$; продолжая применять правило Лопиталя к первому множителю, получим

$$\frac{1}{3}\lim_{x\to 0}\frac{\dfrac{-x}{\sqrt{1-x^2}}-2x}{\sin x}=-\frac{1}{3}\lim_{x\to 0}\frac{x}{\sin x}\left(\frac{1}{\sqrt{1-x^2}}+2\right)=-1.$$

Докажем, что формула (**) остается справедливой при $x\to\infty$, предполагая, что функции $f(x)$ и $\varphi(x)$ определены и дифференцируемы для достаточно больших $|x|$. Полагая, $x=\frac{1}{z}$, придем к рассмотренному случаю, так как если $x\to\infty$, то $z\to 0$. Имеем

$$\lim_{x\to\infty}\frac{f(x)}{\varphi(x)}=\lim_{z\to 0}\frac{f\left(\frac{1}{z}\right)}{\varphi\left(\frac{1}{z}\right)}=\lim_{z\to 0}\frac{f'\left(\frac{1}{z}\right)\left(-\frac{1}{z^2}\right)}{\varphi'\left(\frac{1}{z}\right)\left(-\frac{1}{z^2}\right)};$$

сокращая на $\left(-\frac{1}{z^2}\right)$ и заменяя $\frac{1}{z}$ снова на x, получим

$$\lim_{x\to\infty}\frac{f(x)}{\varphi(x)}=\lim_{x\to\infty}\frac{f'(x)}{\varphi'(x)}.$$

Отметим, что правило применимо во всех случаях стремления x к бесконечности, т. е. при $x\to+\infty$, $x\to-\infty$ и $x\to\infty$.

II. Другие случаи. Первые рассмотренные нами примеры отыскания предела отношения $\frac{f(x)}{\varphi(x)}$, когда и $f(x) \to 0$ и $\varphi(x) \to 0$, можно условно обозначить как случаи $\frac{0}{0}$; примеры же, когда и $f(x) \to \infty$ и $\varphi(x) \to \infty$, — как случаи $\frac{\infty}{\infty}$.

С помощью правила Лопиталя очень часто удается находить пределы функций и в других случаях, отличных от случаев $\frac{0}{0}$ и $\frac{\infty}{\infty}$. Мы рассмотрим случаи: **1)** $0 \cdot \infty$, **2)** $\infty - \infty$, **3)** 1^{∞}, **4)** ∞^{0}, **5)** 0^{0}. Разумеется, эти обозначения $(0 \cdot \infty;\ \infty - \infty;\ 1^{\infty};\ \infty^{0};\ 0^{0},$ как и $\frac{0}{0}$ и $\frac{\infty}{\infty})$ могут служить исключительно для наиболее краткого указания определенного случая при отыскании предела функции. Именно, они указывают, что при заданном изменении независимой переменной x ($x \to x_0$ или $x \to \infty$) функция представляется выражением, являющимся соответственно:

1) произведением функции, стремящейся к нулю, и функции, стремящейся к бесконечности;
2) разностью двух функций, стремящихся к положительной бесконечности;
3) степенью, основание которой стремится к 1, а показатель — к бесконечности;
4) степенью, основание которой стремится к бесконечности, а показатель— к нулю;
5) степенью, и основание и показатель которой стремятся к нулю.

Заметим только, что в случаях 3), 4), 5) функция предварительно логарифмируется и, значит, сначала отыскивается предел не заданной функции, а ее логарифма, а затем уже по пределу логарифма находится предел функции (что допустимо вследствие непрерывности логарифмической функции).

5.2. Асимптоты линий

Когда мы хотим изучить функцию при стремлении аргумента к бесконечности, нам приходится иметь дело с частями графика, уходящими в бесконечность, так называемыми *бесконечными ветвями графика*. С бесконечными же ветвями нам приходится иметь

дело и тогда, когда мы рассматриваем функцию вблизи точек ее бесконечного разрыва. Знание бесконечных ветвей функции необходимо для того, чтобы правильно представить себе форму всего графика и, следовательно, характер изменения функции во всей области ее определения.

Подойдем к вопросу с геометрической точки зрения и введем в связи с этим общее определение асимптоты линии.

Определение. *Прямая линия* Г *называется асимптотой линии L, если расстояние точки линии L от прямой* Г *стремится к нулю при неограниченном удалении этой точки от начала координат.*

Следует различать случаи вертикальной и наклонной асимптот.

1) Пусть линия $y = f(x)$ имеет вертикальную асимптоту. Уравнение такой асимптоты будет $x = x_0$, а поэтому, согласно определению асимптоты, обязательно $f(x) \to \infty$ при $x \to x_0$; обратно: если точка x_0 есть точка бесконечного разрыва функции $f(x)$, то прямая $x = x_0$ служит асимптотой линии $y=f(x)$.

Итак, если

$$\lim_{x \to x_0} f(x) = \infty,$$

то линия $y = f(x)$ имеет асимптоту $x = x_0$.

Взаимное расположение бесконечной ветви линии и ее вертикальной асимптоты $x=x_0$ обнаруживается исследованием знака бесконечности $(\pm \infty)$, к которой стремится $f(x)$, когда x стремится к x_0, оставаясь меньше x_0, т. е. слева, или оставаясь больше x_0, т. е. справа. Это ясно из рис. 5.1, где показаны возможные случаи.

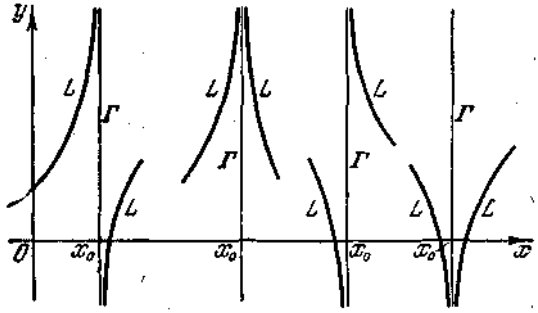


Рис. 5.1.

Может также быть, что график приближается к асимптоте только с одной ее стороны. Например, это будет для функции $y = \ln x$.

2) Пусть линия $y = f(x)$ имеет наклонную асимптоту. Уравнением такой асимптоты будет $y = ax+b$. Согласно определению асимптоты расстояние MN_1 точки M на линии L от асимптоты Г (рис. 5.2) стремится к нулю при $x\to\infty$. Удобнее вместо расстояния MN_1 рассматривать расстояние MN, т. е. разность ординат точки M и точки N, лежащей на прямой Г и имеющей ту же абсциссу, что и точка M.

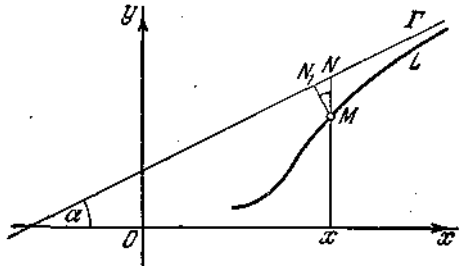


Рис. 5.2.

Из рис. 5.2 имеем

$$MN = \frac{MN_1}{\cos\alpha},\quad MN_1 = MN\cos\alpha,$$

где α — угол между асимптотой и осью Ox; поэтому расстояния MN_1 и MN стремятся к нулю одновременно.

Так как ордината точки M равна значению функции $f(x)$, а ордината точки N — значению линейной функции $ax+b$, то

$$MN = |f(x) - (ax + b)|.$$

Это значит, что *если*

$$\lim_{x\to\infty}[f(x) - (ax + b)] = 0,$$

то линия $y = f(x)$ *имеет асимптоту* $y = ax+b$; при этом говорят, что *функция* $f(x)$ *асимптотически стремится к функции* $ax + b$.

Таким образом, вопрос о существовании и нахождении наклонной асимптоты линии $y=f(x)$ сводится к вопросу о существовании и отыскании таких чисел a и b, что

$$\lim_{x\to\infty}[f(x) - ax - b] = 0. \qquad (*)$$

Если это имеет место, то

$$f(x) = ax + b + \alpha(x),$$

где $\alpha(x)$ —величина бесконечно малая при $x\to\infty$. Разделим обе части равенства на x и перейдем к пределу при $x\to\infty$:

$$\lim_{x\to\infty}\frac{f(x)}{x}=\lim_{x\to\infty}\left(a+\frac{b}{x}+\frac{\alpha(x)}{x}\right).$$

Так как $\frac{b}{x}\to 0$ и $\frac{\alpha(x)}{x}\to 0$, то

$$\lim_{x\to\infty}\frac{f(x)}{x}=a. \qquad (**)$$

Из основного условия (*) находим теперь, что

$$\lim_{x\to\infty}[f(x)-ax]=b. \qquad (***)$$

Верно и обратное: если пределы (**) и (***) существуют и числа a и b найдены, то основное условие соблюдается; таким образом,

Если

$$\frac{f(x)}{x}$$

при $x\to\infty$ *стремится к конечному пределу* a *и если* $f(x)-ax$ *при* $x\to\infty$ *стремится к конечному пределу* b, *то линия* $y=f(x)$ *имеет асимптоту* $y=ax+b$.

В частности, если функция $f(x)$ стремится к конечному пределу при $x\to\infty$:

$$\lim_{x\to\infty}f(x)=b,$$

то, очевидно, $a=0$ и линия $y=f(x)$ имеет горизонтальную асимптоту, параллельную оси Ox, именно $y=b$. Если и a и b равны нулю, то асимптотой служит сама ось Ox.

Если хотя бы один из указанных пределов не существует, то линия $y=f(x)$ наклонных асимптот не имеет.

Например, для линии $y=x+\ln x$ имеем

$$\lim_{x\to\infty}\frac{x+\ln x}{x}=\lim_{x\to\infty}\left(1+\frac{\ln x}{x}\right)=1, \quad \text{т. е.} \quad a=1,$$

но

$$\lim_{x\to\infty}(x+\ln x-1\cdot x)=\infty$$

и значит, асимптоты в данном случае не существует.

Асимптотическое изменение функции может быть различным при стремлении x к положительной или к отрицательной бесконечности, и поэтому следует раздельно рассматривать случаи $x\to+\infty$ и $x\to-\infty$.

Если существует асимптота в первом случае, то мы будем называть ее *правосторонней,* а если во втором, то *левосторонней.* Может случиться, что и при $x\to+\infty$, и при $x\to-\infty$ пределы (**) и

(***) одинаковы. Это означает, что правосторонняя и левосторонняя асимптоты являются частями одной и той же прямой.

Вслед за тем как найдена асимптота $y = ax+b$, исследованием знака выражения $\alpha(x) = f(x) - ax - b$ ..... при $x \to \infty$ можно установить взаимное расположение бесконечной ветви линии и ее асимптоты. Ветвь линии находится, начиная с некоторого места, либо над асимптотой (a>0), либо под ней (a<<0), либо неограниченное число раз пересекает ее (а бесконечное число раз меняет знак).

5.3. Общая схема исследования функций

Теперь мы можем составить схему, по которой удобно производить исследование функций (и линий).

Пусть рассматривается функция $y=f(x)$. Предлагаемая схема состоит из четырех разделов, в которых устанавливаются:

I. 1) Область определения функции.

2) Точки разрыва и интервалы непрерывности.

3) Поведение функции в окрестностях точек разрыва; вертикальные асимптоты.

4) Точки пересечения графика с осями координат.

5) Симметрия графика (четность или нечетность функции).

6) Периодичность графика.

II. Интервалы монотонности функции; точки экстремума и экстремальные значения.

Ш. Интервалы выпуклости и вогнутости; точки перегиба.

IV. Поведение функции в бесконечности. Наклонные (в частности, горизонтальные) асимптоты.

В первом разделе дается в общих чертах описание особенностей функции и ее графика. При этом нельзя забывать о существе той конкретной задачи, которая привела к исследуемой функции. Может, например, оказаться, что достаточно ограничиться более узким интервалом изменения независимой переменной, чем область определения функции.

Выполнение четвертого раздела иногда удобно производить вместе с первым, когда выясняется общая картина поведения функции.

Можно, конечно, ограничиться словесным описанием поведения заданной функции, но правильное, учитывающее все ее особенности, графическое изображение является очень выразительным средством для наглядного представления изучаемой функциональной зависимости. Поэтому *выполнение первых четырех разделов следует*

сопровождать постепенным построением графика функции. При этом прежде всего нужно на оси *Ox* выделить характерные точки, к которым относятся: точки разрыва, нули, точки экстремума, абсциссы точек перегиба. На плоскости *Oxy* отмечаются точки графика, соответствующие этим выделенным значениям аргумента.

В промежутках между характерными точками функция не меняет резко своего поведения, ведет себя плавно, но для уточнения в иных случаях следует брать в этих промежутках обыкновенные точки и также вычислять соответствующие им значения функции, получая новые точки графика. Чем больше таких точек, тем точнее линия, проведенная через все построенные точки, будет выражать график функции.

Пример. Исследуем функцию

$$y = \frac{x^2}{1+x}.$$

I. Функция определена и непрерывна на всей оси *Ox,* за исключением точки x=—1, где она терпит бесконечный разрыв. Следовательно, прямая x= —1 является вертикальной асимптотой, причем

$$\lim_{\substack{x\to -1\\ x<-1}} y = -\infty, \qquad \lim_{\substack{x\to -1\\ x>-1}} y = +\infty.$$

Нулем функциислужит только точка x= 0.

II. Имеем

$$y' = \frac{x^2+2x}{(1+x)^2} = \frac{x(x+2)}{(1+x)^2}.$$

Производная обращается в нуль при x = 0 и x=—2; в интервале (— ∞, —2) она положительна, в интервалах (—2, —1) и (—1,0) отрицательна (в точке x = — 1 она не определена) и в интервале (0, ∞) снова положительна. Значит, функция в первом интервале возрастает, во втором и третьем убывает, в четвертом возрастает; x = — 2 является точкой максимума, причем максимальное значение функции равно —4, а x = 0 является точкой минимума, причем минимальное значение функции равно 0.

III. Имеем

$$y'' = \frac{2}{(1+x)^3}.$$

Вторая производная в нуль нигде не обращается, но при переходе x через точку x = —1 меняет свой знак с минуса на плюс. Таким образом, в интервале (—∞,—1) вторая производная отрицательна, в интервале (—1, ∞) положительна. В первом интервале график функции выпуклый, во втором вогнутый.

IV. Так как

$$\lim_{x\to\infty}\frac{y}{x}=\lim_{x\to\infty}\frac{x}{1+x}=1$$

и

$$\lim_{x\to\infty}(y-1\cdot x)=\lim_{x\to\infty}\left(-\frac{x}{1+x}\right)=-1,$$

то существует наклонная асимптота $y = x - 1$. Вследствие того что разность

$$\alpha(x)=\frac{x^2}{1+x}-(x-1)=\frac{1}{1+x}$$

положительна при $x > -1$ и отрицательна при $x < -1$, график справа от прямой $x=-1$ находится над асимптотой $y = x-1$, а слева от прямой $x=-1$ под ней. Приняв

$$y=\frac{x^2}{1+x}\approx x-1,$$

мы совершаем абсолютную ошибку $\alpha=\left|\frac{1}{1+x}\right|$ и относительную $\delta=\frac{1}{x^2}$; уже при $|x|>10$ относительная ошибка меньше одного процента. Практически функцию

$$y=\frac{x^2}{1+x}$$

при $|x| > 10$ можно считать линейной функцией $y = x-1$, что значительно упрощает ее использование.

Приняв во внимание полученные результаты, построим график (рис. 5.3), дающий правильное представление о ходе изменения функции на всей оси *Ox*.

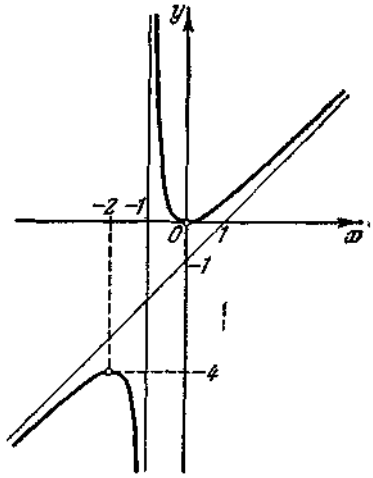


Рис. 5.3

Методами аналитической геометрии можно убедиться, что линия $y = \frac{x^2}{1+x}$ есть гипербола.

5.4. Векторная функция скалярного аргумента

I. Векторная функция. Годограф. Как известно из векторной алгебры, разложение любого вектора **A**, проекции которого на оси координат равны *x, y* и *z,* имеет вид

$$\mathbf{A} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k},$$

где **i, j** и **k** — единичные векторы, направленные по осям координат. Если проекции *x, y, z* — постоянные числа (только такой случай и рассматривается в векторной алгебре), то и вектор **A** называется *постоянным.* Пусть теперь проекции вектора являются функциями параметра *t,* изменяющегося в некотором интервале:

$$x = x(t), \quad y = y(t), \quad z = z(t).$$

Тогда и сам вектор называется *переменным;* при этом каждому значению параметра *t* будет соответствовать определенный вектор

$$\mathbf{A}(t) = x(t)\,\mathbf{i} + y(t)\,\mathbf{j} + z(t)\,\mathbf{k}.$$

Определение. *Если каждому значению параметра t соответствует определенный вектор* **A***(t), то* **A** *(t) называется векторной функцией скалярного аргумента.*

Как и в векторной алгебре, мы рассматриваем свободные векторы, т. е. такие векторы, которые считаются равными, если они имеют равные модули и одинаковые направления, или, иначе говоря, равные проекции на оси координат.

Мы будем представлять себе вектор **A** *(t)* исходящим из начала координат; тогда при изменении *t* конец вектора **A**(*t*), имеющий координаты $x(t)$, $y(t)$, $z(t)$, будет описывать некоторую линию *L*, уравнениями которой служат следующие параметрические уравнения:

$$x = x(t), \quad y = y(t), \quad z = z(t).$$

Так как вектор **A** *(t)* есть не что иное, как радиус-вектор **r** точки *M* на линии *L*, то эту линию можно задать таким одним векторным уравнением:

$$\mathbf{r} = x(t)\,\mathbf{i} + y(t)\,\mathbf{j} + z(t)\,\mathbf{k}.$$

Определение. *Линия L, описанная концом вектора* **A***, называется годографом векторной функции* **r** = **A***(t).*

Начало координат называют при этом *полюсом годографа.*

Если у вектора **A** *(t)* меняется только модуль, то годографом его будет луч, исходящий из полюса, или некоторая часть этого луча. Если модуль вектора **A** *(t)* постоянен (| **A**(*t*) | = const) и меняется только его направление, то годограф есть линия, лежащая на сфере с центром в полюсе и с радиусом, равным модулю вектора **A**(*t*).

С векторной функцией скалярного аргумента особенно часто приходится иметь дело в кинематике при изучении движения точки. Радиус-вектор движущейся точки является функцией времени: **r=A***(t);* годограф этой функции есть траектория движения. При этом уравнение **r=A***(t)* называют уравнением диижения.

II. Предел и непрерывность векторной функции. Дли векторных функций **A***(t)* скалярного аргумента вводятся основные понятия анализа аналогично тому, как это делается для скалярных функций.

Определение. *Вектор* **A** *называется пределом векторной функции* **A** *(t) при* $t \to t_0$*:*

$$\lim_{t \to t_0} \mathbf{A}(t) = \mathbf{A},$$

если для всех значений t, достаточно мало отличающихся от t_0*, модуль разности векторов* | **A** *(t)* — **A**| *будет как угодно мал.*

Если $\mathbf{A}(t) = x(t)\,\mathbf{i} + y(t)\,\mathbf{j} + z(t)\,\mathbf{k}$ и $\mathbf{A} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$, то

$$|\mathbf{A}(t)-\mathbf{A}|=\sqrt{[x(t)-a]^2+[y(t)-b]^2+[z(t)-c]^2}.$$

Ясно, что из стремления к нулю $|\mathbf{A}(t)-\mathbf{A}|$ при $t \to t_0$ следует, что $x(t) \to a$, $y(t) \to b$, $z(t) \to c$; верно и обратное. Коротко это можно выразить в виде простого правила: *проекции предела векторной функции* $\mathbf{A}(t)$ *равны пределам ее проекций.*

Определение. *Векторная функция* $\mathbf{A}(t)$ *называется непрерывной при данном значении параметра t, если она определена в окрестности точки t и если*

$$\lim_{\Delta t \to 0}|\mathbf{A}(t+\Delta t)-\mathbf{A}(t)|=\lim_{\Delta t \to 0}|\Delta \mathbf{A}(t)|=0.$$

При этом геометрический смысл разности $\Delta\mathbf{A}(t)$ ясен из рис. 5.4.

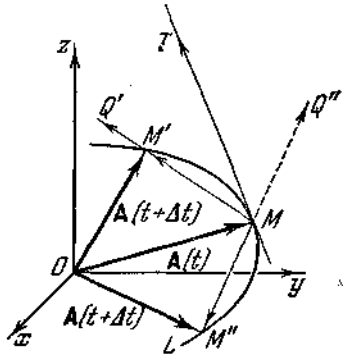


Рис. 5.4

Это будет вектор $\overline{MM'}$ или вектор $\overline{MM''}$. Пусть вектор $\mathbf{A}(t)$ имеет разложение

$$\mathbf{A}(t)=x(t)\,\mathbf{i}+y(t)\,\mathbf{j}+z(t)\,\mathbf{k}. \qquad (*)$$

Тогда

$$\mathbf{A}(t+\Delta t)=x(t+\Delta t)\,\mathbf{i}+ \\ +y(t+\Delta t)\,\mathbf{j}+z(t+\Delta t)\,\mathbf{k}.$$

В соответствии с правилами векторной алгебры

$$\Delta\mathbf{A}(t)=\mathbf{A}(t+\Delta t)-\mathbf{A}(t)= \\ =\Delta x\,\mathbf{i}+\Delta y\,\mathbf{j}+\Delta z\,\mathbf{k}, \qquad (**)$$

где $\Delta x=x(t+\Delta t)-x(t)$ и т. д.

Так как

$$|\Delta\mathbf{A}(t)|=\sqrt{\Delta x^2+\Delta y^2+\Delta z^2},$$

то из условия $|\Delta \mathbf{A}(t)| \to 0$ следует, что и $\Delta x \to 0$, $\Delta y \to 0$, $\Delta z \to 0$. Очевидно и обратное, т. е. из стремления к нулю Δx, Δy, Δz следует стремление к нулю $|\Delta \mathbf{A}(t)|$. Это означает, что если непрерывна векторная функция $\mathbf{A}(t)$, то непрерывными будут и ее проекции на оси координат $x(t)$, $y(t)$, $z(t)$, и наоборот. Ясно, что годографом непрерывной векторной функции $\mathbf{r} = \mathbf{A}(t)$ будет непрерывная линия. Мы будем в дальнейшем предполагать, что эта линия во всех точках имеет касательную.

III. Производная векторной функции. Чтобы определить производную векторной функции $\mathbf{A}(t)$ по скалярному аргументу t, составим прежде всего отношение

$$\frac{\Delta \mathbf{A}(t)}{\Delta t} = \frac{\mathbf{A}(t+\Delta t) - \mathbf{A}(t)}{\Delta t}.$$

Геометрически это отношение изображается вектором, направленным в сторону, соответствующую возрастанию t (направление от M к M' на рис. 5.4). Действительно, если $\Delta t > 0$, то вектор $\frac{\Delta \mathbf{A}(t)}{\Delta t} = \overrightarrow{MQ'}$ имеет то же направление, что и вектор $\Delta \mathbf{A}(t) = \overrightarrow{MM'}$, т. е. он направлен в ту сторону годографа, которая соответствует возрастанию параметра t. Если же $\Delta t < 0$, то вектор $\Delta \mathbf{A}(t) = \overrightarrow{MM''}$ направлен в противоположную сторону, и при делении на отрицательное число Δt мы получим вектор $\overrightarrow{MQ''}$, направленный снова в сторону возрастания t.

Рассмотрим теперь предел отношения $\frac{\Delta \mathbf{A}(t)}{\Delta t}$ при $\Delta t \to 0$.

Определение. *Предел*

$$\lim_{\Delta t \to 0} \frac{\Delta \mathbf{A}(t)}{\Delta t}$$

называется производной от векторной функции $\mathbf{A}(t)$ *по скалярному аргументу* t.

В силу определения предела производная векторной функции сама является вектором. Она обозначается $\mathbf{A}'(t)$ *или* $\frac{d\mathbf{A}(t)}{dt}$. Чтобы определить направление вектора $\mathbf{A}'(t)$, достаточно наметить, что при $\Delta t \to 0$ точка M' (M'') стремится к точке M, и поэтому секущая MM' (MM'') стремится к касательной в точке M. Поэтому *производная* $\mathbf{A}'(t)$ *является вектором* $\overrightarrow{MT}$, *касательным к годографу векторной функции* $\mathbf{A}(t)$, *направленным в сторону, соответствующую возрастанию параметра* t.

Если векторная функция $\mathbf{A}(t)$ имеет постоянный модуль, но переменное направление, то ее производная функция $\mathbf{A}'(t)$ является *вектором, перпендикулярным к вектору* $\mathbf{A}(t)$. В самом деле, годограф лежит на сфере, и поэтому производная $\mathbf{A}'(t)$, как вектор, касательный к годографу, перпендикулярна к радиусу-вектору $\mathbf{A}(t)$.

Итак, *производная вектора с постоянным модулем перпендикулярна к нему.*

Перейдем теперь к фактическому отысканию производной $\mathbf{A}'(t)$. Пусть векторная функция $\mathbf{A}(t)$ задана своим разложением

$$\mathbf{A}(t) = x(t)\,\mathbf{i} + y(t)\,\mathbf{j} + z(t)\,\mathbf{k}.$$

Тогда, согласно (**),

$$\frac{\Delta \mathbf{A}(t)}{\Delta t} = \frac{\Delta x}{\Delta t}\,\mathbf{i} + \frac{\Delta y}{\Delta t}\,\mathbf{j} + \frac{\Delta z}{\Delta t}\,\mathbf{k}.$$

Перейдя к пределу при $\Delta t \to 0$ и воспользовавшись правилом отыскания предела векторной функции, получим разложение производной $\mathbf{A}'(t)$ по единичным векторам

$$\mathbf{A}'(t) = x'(t)\,\mathbf{i} + y'(t)\,\mathbf{j} + z'(t)\,\mathbf{k}.$$

Из этого разложения следует, что

$$|\mathbf{A}'(t)| = \sqrt{x'^2(t) + y'^2(t) + z'^2(t)}.$$

Вспоминая выражение для дифференциала длины дуги ds, последнее равенство можно записать в виде

$$|\mathbf{A}'(t)| = \frac{ds}{dt}.$$

Таким образом, модуль производной векторной функции $|\mathbf{A}'(t)|$ равен *производной от длины годографа по аргументу t.* Необходимо подчеркнуть, что модуль производной $|\mathbf{A}'(t)|$ не равен производной от модуля $(|\mathbf{A}(t)|)'$. Особенно наглядно это видно на примере производной вектора с постоянным модулем: в этом случае производная модуля вектора как постоянного числа просто равна нулю; производная же самого вектора есть вектор, к нему перпендикулярный.

Пользуясь выражением для $\mathbf{A}'(t)$, легко показать, что все основные правила дифференцирования переносятся почти без изменения на векторные функции:

1) $[\mathbf{A}_1(t) + \mathbf{A}_2(t)]' = \mathbf{A}_1'(t) + \mathbf{A}_2'(t)$,

2) $[f(t)\,\mathbf{A}(t)]' = f'(t)\,\mathbf{A}(t) + f(t)\,\mathbf{A}'(t)$,

где $f(t)$ — скалярная функция; в частности, $[C\mathbf{A}(t)]' = C\mathbf{A}'(t)$, где C — скаляр.

Правила дифференцирования скалярного и векторного произведений двух векторных функций также ничем не отличаются от соответствующего правила в случае произведения скалярных функций:

$$3)\ [\mathbf{A}_1(t)\cdot\mathbf{A}_2(t)]' = \mathbf{A}_1'(t)\cdot\mathbf{A}_2(t)+\mathbf{A}_1(t)\cdot\mathbf{A}_2'(t),$$

$$4)\ [\mathbf{A}_1(t)\times\mathbf{A}_2(t)] = [\mathbf{A}_1'(t)\times\mathbf{A}_2(t)]+[\mathbf{A}_1(t)\times\mathbf{A}_2'(t)].$$

(В правиле 4 важен порядок перемножаемых векторов.)

Последовательным дифференцированием можно найти производные высших порядков от векторной функции. Так,

$$\mathbf{A}''(t) = x''(t)\,\mathbf{i}+y''(t)\,\mathbf{j}+z''(t)\,\mathbf{k}$$

и т. д.

6. Функции комплексного переменного

6.1. Понятие функции комплексного переменного

Понятие комплексного числа рассмотрено, например, в книге «Высшая математика. Дифференциальное и интегральное исчисление». Там же рассмотрены многочлены $Q_n(z)$ от комплексного переменного. Многочлен является простейшим примером функции комплексного переменного.

Комплексные числа изображают точками плоскости, где задана прямоугольная система координат.

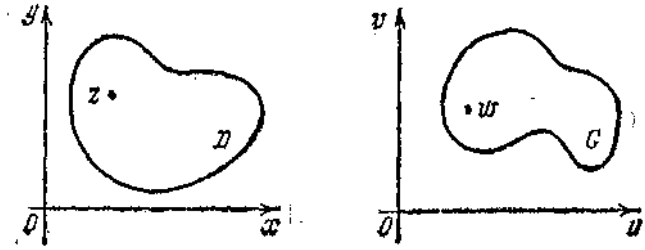


Рис. 6.1

Приведем понятие *функции от комплексного переменного.*

Пусть даны две плоскости комплексных чисел $z=x+iy$ и $w=u+iv$ (рис. 6.1). Рассмотрим некоторое множество точек D в плоскости z и множество G в плоскости w. Если каждому числу $z \in D$ по некоторому закону поставлено в соответствие определенное комплексное число $w \in G$, то говорят, что на множестве D *задана однозначная функция комплексного переменного, отображающая множество D в множество G*. Символически это обозначают так:

$$w=f(z).$$

Множество D называют *областью определения* функции. Если каждая точка множества G является значением функции, то говорят, что G — *область значений* этой функции или *образ множества D* при помощи функции $f(G=f(D))$. В этом случае говорят еще, что функция f *отображает D на G.*

Функцию $f(z)$ можно записать в виде

$$f(z)=u(x,\ y)+iv(x,\ y) \qquad ((x,\ y)\in D),$$

где

$$u(x,\ y)=\operatorname{Re} f(z), \qquad v(x,\ y)=\operatorname{Im} f(z) \qquad ((x,\ y)\in D),$$

— действительные функции от переменных x, y.

Если каждому $z \in D$ соответствует несколько разных значений w, то функция $w = f(z)$ называется *многозначной.*

Понятия предела и непрерывности функции комплексного переменного вводятся аналогично, как это делается для функции действительного переменного, необходимо лишь всюду вместо абсолютной величины писать модуль комплексного числа.

Говорят, что функция

$$w=f(z)=u(x,\ y)+iv(x,\ y)$$

имеет *предел* в точке z_0, равный числу $A=a+ib$, если

$$\lim_{|z-z_0|\to 0}|f(z)-A|=0. \tag{6.1}$$

В этом случае пишут

$$\lim_{z\to z_0} f(z)=A.$$

На языке функций u и v свойство (6.1) записывается в виде равенства

$$\lim_{|z-z_0|\to 0}\sqrt{(u-a)^2+(v-b)^2}=0 \tag{6.2}$$

или, что все равно, в виде двух равенств

$$\lim_{(x,\ y)\to(x_0,\ y_0)} u(x,\ y)=a, \qquad \lim_{(x,\ y)\to(x_0,\ y_0)} v(x,\ y)=b. \tag{6.3}$$

Для комплексных функций $f(z)$ и $g(z)$ имеют место свойства, аналогичные соответствующим свойствам действительных функций:

$$\left.\begin{aligned} &\lim_{z\to z_0}[f(z)\pm g(z)]=\lim_{z\to z_0} f(z)\pm\lim_{z\to z_0} g(z),\\ &\lim_{z\to z_0}[f(z)\,g(z)]=\lim_{z\to z_0} f(z)\lim_{z\to z_0} g(z),\\ &\lim_{z\to z_0}\frac{f(z)}{g(z)}=\frac{\lim\limits_{z\to z_0} f(z)}{\lim\limits_{z\to z_0} g(z)} \qquad \left(\lim_{z\to z_0} g(z)\neq 0\right). \end{aligned}\right\} \tag{6.4}$$

Как обычно, формулы (6.4) надо понимать в том смысле, что если пределы, стоящие в их правых частях, существуют, то существуют также пределы, стоящие в их левых частях, и выполняется соответствующее равенство.

Функция $w = f(z) = u(x, y) + iv(x, y)$ называется *непрерывной* в точке z_0, если для нее выполняется свойство

$$\lim_{z \to z_0} f(z) = f(z_0) \quad (f(z_0 + \Delta z) - f(z_0) \to 0, \ \Delta z \to 0). \tag{6.5}$$

Таким образом, непрерывная в точке z_0 функция должна быть определена в окрестности этой точки, в том числе и в ней самой и должно выполняться равенство (6.5). Равенство (6.5) эквивалентно двум равенствам:

$$\lim_{(x, y) \to (x_0, y_0)} u(x, y) = u(x_0, y_0), \quad \lim_{(x, y) \to (x_0, y_0)} v(x, y) = v(x_0, y_0).$$

Следовательно, непрерывность f в точке z_0 эквивалентна непрерывности функций u и v в точке (x_0, y_0).

Из свойств (6.4) следует, что сумма, разность, произведение и частное непрерывных в точке z_0 комплексных функций $f(z)$ и $g(z)$ есть непрерывная функция в этой точке. В случае частного надо в этой формулировке считать, что $g(z_0) \neq 0$.

Пример 1. Функция $w = |z| = \sqrt{x^2 + y^2}$ задана на всей комплексной плоскости. Ее значения — неотрицательные числа. Эта функция непрерывна во всех точках комплексной плоскости:

$$||z + \Delta z| - |z|| \leqslant |\Delta z| \to 0 \quad (\Delta z \to 0).$$

Пример 2.

$$w = \operatorname{Arg} z = \arg z + 2k\pi \quad (k = 0, \pm 1, \ldots). \tag{6.6}$$

Эта функция многозначная (бесконечнозначная); $\varphi = \arg z$ —главное значение аргумента $(0 \leqslant \varphi < 2\pi)$.

Пример 3. Функция $w = z$. Она непрерывна:

$$|z + \Delta z - z| = |\Delta z| \to 0 \quad (\Delta z \to 0).$$

Но тогда и функция z^n (n=2, 3, ...) непрерывна как произведение конечного числа непрерывных функций. Множество комплексных чисел D будем называть *областью,* если D, как множество точек плоскости, открыто и связно.

Область D называется *односвязной,* если любая непрерывная замкнутая самонепересекающаяся кривая, проведенная в D, ограничивает некоторую область G, целиком принадлежащую D. Область, не обладающую этим свойством, будем называть *многосвязной.*

Пример 4. Кольцо $r < |z| < R$— многосвязная (двусвязная) область. Кривая L (рис. 6.2) принадлежит кольцу, но ограничивает область, не входящую целиком в него.

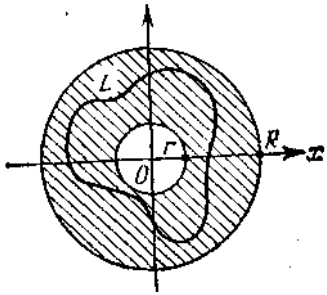


Рис. 6.2

6.2. Производная функции комплексного переменного

Пусть задана однозначная функция $w=f(z)$ на области D (открытом связном множестве) комплексной плоскости z.

Производной от функции $f(z)$ в точке z называется предел

$$\lim_{\Delta z \to 0} \frac{\Delta w}{\Delta z} = \lim_{\Delta z \to 0} \frac{f(z+\Delta z)-f(z)}{\Delta z} = f'(z) = \frac{dw}{dz}, \qquad (6.7)$$

когда Δz любым образом стремится к нулю.

Далеко не всякая функция комплексного переменного имеет производную. Существование предела (6.7) — очень сильное требование: при подходе $z+\Delta z$ к z по любому пути каждый раз должен существовать указанный в (6.7) предел.

Функцию $f(z)$, имеющую непрерывную производную в любой точке области D комплексной плоскости, называют *аналитической функцией на этой области.*

Можно доказать, что *если производная аналитической функции $f(z)$ не равна нулю на области D, то множество значений G функции $f(z)$ также есть область.* Мы будем пользоваться этим свойством.

Дадим геометрическое представление производной $f'(z)$, когда она не равна нулю. Кроме плоскости z, введем еще другую плоскость точек до. Опишем из точки z открытый круг σ радиуса $\delta > 0$ с центром в ней (рис. 6.3).

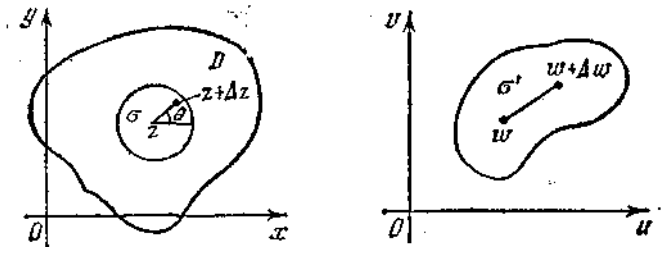


Рис. 6.3.

Произвольная точка σ имеет вид $z+\Delta z$, где Δz —произвольное комплексное число с модулем, меньшим δ: $|\Delta z|<\delta$. Запишем Δz в показательной форме

$$\Delta z = \rho e^{i\varphi} \qquad (\rho > 0). \tag{6.8}$$

При помощи функции $w=f(z)$ круг σ перейдет в некоторую область σ′ плоскости w. Область σ′ состоит из точек $w+\Delta w$, где приращения Δw соответствуют всевозможным указанным приращениям Δz $(|\Delta z| < \delta)$ (см. рис. 6.3).

Из (6.7) следует равенство

$$\frac{\Delta w}{\Delta z} = f'(z) + \alpha(\Delta z), \quad \text{где} \quad \alpha(\Delta z) \underset{\Delta z \to 0}{\longrightarrow} 0.$$

Умножая левую и правую части последнего равенства на Δz, получаем

$$\Delta w = f'(z)\,\Delta z + \Delta z \cdot \alpha(\Delta z). \tag{6.9}$$

Произведение $\Delta z \cdot \alpha(\Delta z)$ стремится к нулю при $\Delta z \to 0$ быстрее чем Δz. Поэтому, если $f(z) \neq 0$, то первый член правой части (6.9) является *главным*. Приближенно, с точностью до бесконечно малых высшего порядка (по сравнению с Δz), при достаточно малых Δz можно написать

$$\Delta w \approx f'(z)\,\Delta z.$$

Число $f'(z)$ запишем в показательной форме

$$f'(z) = re^{i\varphi} \qquad (r > 0). \tag{6.10}$$

Поэтому, учитывая (6.8), получим

$$\Delta w \approx r\rho e^{i(\varphi+\theta)} \qquad (|\Delta z| = \rho < \delta).$$

Мы видим, что модуль $|\Delta w|$, с точностью до бесконечно малой высшего порядка, в $r = |f'(z)|$ раз больше модуля $|\Delta w\ z|$:

$$|\Delta w| \approx r\rho = r\,|\Delta z|,$$

а аргумент Δw (тоже с точностью до бесконечно малой высшего порядка) получается из аргумента Δz прибавлением к нему числа φ (рис. 6.4):

$$\operatorname{Arg}(\Delta w) \approx \operatorname{Arg}(\Delta z) + \varphi.$$

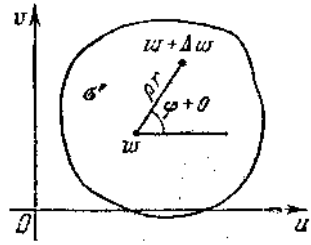


Рис.6.4

Таким образом, для того чтобы представить себе, куда перешли точки $z + \Delta z$ с $|\Delta z|<\delta$ при помощи функции $w = f(z)$ надо: 1) повернуть круг σ на угол $\varphi = \arg f'(z)$ и 2) растянуть его в $r = |f'(z)|$ раз.

Каждая точка $z+\Delta z$, $|\Delta z|< \delta$, при помощи этих двух операций перейдет в некоторую точку, которую надо еще сдвинуть на величину $\Delta z \cdot \sigma(\Delta z)$ — бесконечно малую высшего порядка чем Δz.

Пусть Γ_1 и Γ_2 — гладкие кривые, выходящие из точки z. Касательные к ним образуют с осью x углы соответственно θ_1, θ_2 (отсчитываемые от оси x против часовой стрелки). Образы этих кривых Γ'_1, Γ'_2 на плоскости w (рис. 6.5) при помощи функции $w = f(z)$ имеют касательные в точке w, образующие с осью абсцисс соответственно углы θ'_1, θ'_2 (которые отсчитываются тоже против часовой стрелки). При этом (в силу свойства 1))

$$\theta'_1 = \theta_1 + \varphi, \quad \theta'_2 = \theta_2 + \varphi,$$

откуда, следует свойство

$$\theta'_2 - \theta'_1 = \theta_2 - \theta_1,$$

выражающее, что *данное отображение сохраняет углы* и притом с сохранением направления отсчета (если $\theta_2 > \theta_1$, то $\theta'_2 > \theta'_1$.

Кроме того, как мы видели выше, данное отображение осуществляет в каждой точке z, где $f'(z)\neq 0$, растяжение, не зависящее от направления.

Отображение, обладающее (с точностью до бесконечно малых высшего порядка) свойством сохранения углов (с сохранением направления отсчета) и свойством постоянства растяжений, называется *конформным отображением.*

Из вышеизложенного следует, что *отображение с помощью аналитической функции $w=f(z)$ является конформным во всех точках, где $f'(z)\neq 0$.*

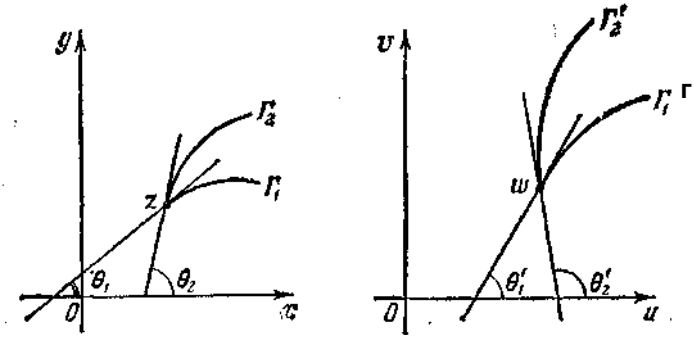


Рис. 6.5.

Замечание 1. *Если функция f (z) комплексной переменной z имеет всюду на области D производную f'*(z), *то автоматически эта производная непрерывна всюду на D, т. е. f(z) аналитическая на D.*

Замечание 2. Из равенства (6.9) следует, что если функция *f*(*z*) имеет производную в точке *z*, то она непрерывна в этой точке (т. е. $\Delta w \to 0$ при $\Delta z \to 0$).

Производная от функции *f*(*z*) порядка *k* обозначается через $f^{(k)}(z)$ и определяется по индукции

$$(f^{(k-1)}(z))' = f^{(k)}(z) \qquad (k = 1,\ 2,\ \ldots;\ f^{(0)}(z) = f(z)).$$

Зная, что у аналитической на области *D* функции *f*(*z*) производная непрерывна на *D,* нам будет в дальнейшем нетрудно заключить, что *f (z)* имеет на *D* непрерывные производные любого порядка

$$f'(z),\quad f''(z),\quad f'''(z),\quad \ldots.$$

Употребляют еще такую терминологию: функция *f(z)* называется *аналитической в точке* z_0, если она аналитическая на некоторой окрестности этой точки. Наконец, говорят, что функция *f(z)* аналитическая на замыкании *D* области *D,* если существует область *G,* содержащая в себе $\overline{D}$ $(G \supset \overline{D})$, на которой *f(z)* аналитическая.

Приведем *основные свойства производных* функций комплексного переменного, аналогичные соответствующим свойствам производных для функций действительного переменного:

$$[u(z) \pm v(z)]' = u'(z) \pm v'(z), \tag{6.11}$$

$$[u(z)\,v(z)]' = u(z)\,v'(z) + u'(z)\,v(z), \tag{6.12}$$

$$\left[\frac{u(z)}{v(z)}\right]' = \frac{u'(z)\,v(z) - u(z)\,v'(z)}{v^2(z)} \quad (v(z) \neq 0), \tag{6.13}$$

$$\frac{dw}{dz} = \frac{dw}{dv}\,\frac{dv}{dz}. \tag{6.14}$$

Формулу (6.14) надо понимать так: если *w* есть функция $w = \varphi(v)$ комплексного переменного *v*, имеющая производную

$\frac{dw}{dv} = \varphi'(v)$, а $v = \psi(z)$ — функция от комплексной переменной z, имеющая производную $\frac{dv}{dz} = \psi'(z)$, то производная сложной функции

$$w = F(z) = \varphi[\psi(z)]$$

вычисляется по формуле (6.14).

Ниже мы приводим некоторые элементарные функции комплексного переменного.

Степенная функция

$$w = z^n,$$

n—целое.

Эта функция имеет производную, вычисляемую по формуле

$$(z^n)' = nz^{n-1} \quad (n = \ldots, -2, -1, 0, 1, \ldots).$$

При $n > 0$ ее удобно вычислить как предел

$$\lim_{\Delta z \to 0} \frac{(z+\Delta z)^n - z^n}{\Delta z} = nz^{n-1},$$

применяя формулу бинома Ньютона.

При $n < 0$ теперь можно воспользоваться формулой (6.13).

Функция z^n при $n>0$ аналитическая на всей плоскости z, а при $n < 0$ на всей плоскости с выколотой из нее точкой $z = 0$.

Функции e^z, $\sin z$, $\cos z$, $\operatorname{tg} z$.

Первые три из этих функций определены как суммы степенных рядов:

$$e^z = 1 + \frac{z}{1!} + \frac{z^2}{2!} + \ldots,$$

$$\sin z = z - \frac{z^3}{3!} + \frac{z^5}{5!} - \ldots,$$

$$\cos z = 1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \ldots.$$

Радиус сходимости каждого из этих рядов равен ∞. Поэтому производные от этих функций могут быть получены для любого z почленным дифференцированием соответствующих рядов:

$$(e^z)' = 1 + \frac{z}{1!} + \frac{z^2}{2!} + \ldots = e^z,$$

$$(\sin z)' = 1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \ldots = \cos z,$$

$$(\cos z)' = -z + \frac{z^3}{3!} - \frac{z^5}{5!} + \ldots = -\sin z.$$

Формулы для тригонометрических функций суммы комплексных аргументов остаются такими же, как и в случае действительного переменного.

Функция tg z определяется по формуле

$$\operatorname{tg} z = \frac{\sin z}{\cos z}.$$

Ее производная равна

$$(\operatorname{tg} z)' = \frac{\cos^2 z + \sin^2 z}{\cos^2 z} = \frac{1}{\cos^2 z} \qquad (\cos z \neq 0),$$

что следует из формулы (6.13).

Функция a^z ($a > 0$) может быть определена по формуле

$$a^z = e^{z \ln a} = \exp(z \ln a).$$

(Ее производная вычисляется на основании формулы (6.14) о производной сложной функции:

$$(a^z)' = (e^{z \ln a})' = e^{z \ln a} \ln a = a^z \ln a.$$

Гиперболические функции sh z, ch z, th z определяются формулами

$$\operatorname{sh} z = \frac{e^z - e^{-z}}{2}, \quad \operatorname{ch} z = \frac{e^z + e^{-z}}{2}, \quad \operatorname{th} z = \frac{\operatorname{sh} z}{\operatorname{ch} z}.$$

Отсюда следует, что

$$\operatorname{sh} iz = \frac{e^{iz} - e^{-iz}}{2} = i \sin z, \qquad \operatorname{ch} iz = \cos z. \tag{6.15}$$

Заменяя в (6.15) z на iz, получаем

$$\cos iz = \operatorname{ch} z, \qquad \sin iz = i \operatorname{sh} z. \tag{6.16}$$

Отметим еще легко проверяемую формулу

$$\operatorname{ch}^2 z - \operatorname{sh}^2 z = 1.$$

Формулы сложения для гиперболических функций легко получить из (6.15) и (6.16) и соответствующих формул для тригонометрических функций от комплексного переменного. Например;

$$\operatorname{ch}(z_1 + z_2) = \cos i(z_1 + z_2) =$$
$$= \cos iz_1 \cos iz_2 - \sin iz_1 \sin iz_2 = \operatorname{ch} z_1 \operatorname{ch} z_2 + \operatorname{sh} z_1 \operatorname{sh} z_2.$$

Производные от этих функций вычисляются на основании формул (6.11), (6.13), (6.14):

$$(\operatorname{sh} z)' = \left(\frac{e^z - e^{-z}}{2}\right)' = \frac{e^z + e^{-z}}{2} = \operatorname{ch} z, \quad (\operatorname{ch} z)' = \operatorname{sh} z,$$

$$(\operatorname{th} z)' = \frac{\operatorname{ch}^2 z - \operatorname{sh}^2 z}{\operatorname{ch}^2 z} = \frac{1}{\operatorname{ch}^2 z} \qquad (\operatorname{ch} z \neq 0).$$

Пример. Выделить действительную и мнимую части у функции $w = \cos z$ и найти нули этой функции.

Пусть $z = x + iy$, $w = u(x, y) + iv(x, y)$. Имеем

$$\cos z = \cos(x + iy) = \cos x \cos yi - \sin x \sin iy =$$
$$= \cos x \operatorname{ch} y - i \sin x \operatorname{sh} y.$$

Таким образом, $u(x, y) = \cos x \operatorname{ch} y$, $v(x, y) = -\sin x \operatorname{sh} y$. Чтобы найти нули функции cos *z*, мы должны приравнять нулю ее действительную и мнимую части:

$$\left.\begin{aligned} \cos x \operatorname{ch} y &= 0, \\ \sin x \operatorname{sh} y &= 0. \end{aligned}\right\}$$

Решим эту систему. Так как ch $\neq 0$ для любого действительного *y,* то из первого уравнения получаем cos *x*=0.

Из второго уравнения при $y \neq 0$ получаем, что sin *x*=0. При действительных *x* косинус и синус не обращаются одновременно в нуль, поэтому при $y \neq 0$ система решений не имеет. Если же $y = 0$, то sh $z = 0$ и второе уравнение удовлетворяется при любых *x.* Таким образом, нули функции cos *z* расположены на действительной оси *x* и совпадают с нулями cos *x*.

Замечание 3. Из этого утверждения следует, что нули функции ch *z* совпадают с нулями функции cos *y,* где $y = \operatorname{Im} z$.

6.3. Условия Даламбера — Эйлера (Коши — Римана)

Рассмотрим комплексную функцию

$$w = f(z) = u(x, y) + iv(x, y) \quad (z \in D),$$

определенную на области *D* комплексной плоскости. Пусть она имеет производную в точке $z \in D$

$$f'(z) = \lim_{\Delta z \to 0} \frac{\Delta w}{\Delta z}, \tag{6.17}$$

$$\Delta w = [u(x + \Delta x, y + \Delta y) - u(x, y)] + i[v(x + \Delta x, y + \Delta y) - v(x, y)].$$

Таким образом, при любом способе стремления $\Delta z = \Delta x + i\Delta y$ к нулю должен существовать предел (6.17), равный одному и тому же комплексному числу $f'(z)$. В частности, это должно иметь место, если а) $\Delta z = \Delta x + i0 = \Delta x$ и $\Delta x \to 0$ или, если б) $\Delta z = 0 + i\Delta u = i\Delta u$ и $\Delta u \to 0$.

В первом случае

$$f'(z) = \lim_{\Delta z \to 0} \frac{\Delta w}{\Delta z} =$$

$$= \lim_{\Delta x \to 0} \left[\frac{u(x+\Delta x, y) - u(x, y)}{\Delta x} + i \frac{v(x+\Delta x, y) - v(x, y)}{\Delta x} \right] =$$

$$= \lim_{\Delta x \to 0} \frac{u(x+\Delta x, y) - u(x, y)}{\Delta x} + i \lim_{\Delta x \to 0} \frac{v(x+\Delta x, y) - v(x, y)}{\Delta x} =$$

$$= \frac{\partial u}{\partial x} + i \frac{\partial v}{\partial x}.$$

Во втором случае

$$f'(z) = \lim_{\Delta z \to 0} \frac{\Delta w}{\Delta z} =$$

$$= \lim_{\Delta y \to 0} \left[\frac{u(x, y+\Delta y) - u(x, y)}{i\Delta y} + \frac{v(x, y+\Delta y) - v(x, y)}{\Delta y} \right] =$$

$$= -i \lim_{\Delta y \to 0} \frac{u(x, y+\Delta y) - u(x, y)}{\Delta y} + \lim_{\Delta y \to 0} \frac{v(x, y+\Delta y) - v(x, y)}{\Delta y} =$$

$$= -i \frac{\partial u}{\partial y} + \frac{\partial v}{\partial y}.$$

Но тогда должны выполняться равенства

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}, \tag{6.18}$$

которые обычно называют *условиями Коши—Римана.* Некоторое время думали, что именно Коши и Риман впервые получили эти условия. Но выяснилось, что они были известны еще Эйлеру и Даламберу.

Итак нами доказана

Теорема 1. *Если функция*

$$f(z) = u(x, y) + iv(x, y)$$

имеет производную в точке $z=x+iy$, *то ее действительные компоненты* u *и* v *имеют в точке* (x, y) *частные производные первого порядка, удовлетворяющие условию Коши — Римана.*

Теорему 1 можно обратить, правда при добавочном предположении, что частные производные от u и v непрерывны.

Теорема 2. *Если функции* $u(x, y)$ *и* $v(x, y)$ *имеют в точке* (x, y) *непрерывные частные производные, удовлетворяющие условиям Коши —Римана, то функция комплексной переменной* $f(z)t=u+iv$ *имеет в точке* $z = x+iy$ *производную.*

Доказательство. Пусть функции u и v имеют непрерывные частные производные в точке (x, y). Тогда они дифференцируемы в этой точке, т. е. их приращения, соответствующие приращениям Δx, Δy, могут быть записаны в виде

$$\Delta u = u(x+\Delta x,\ y+\Delta y) - u(x,\ y) =$$
$$= \frac{\partial u}{\partial x}\Delta x + \frac{\partial u}{\partial y}\Delta y + o_1(\rho) \qquad (\rho \to 0),$$
$$\Delta v = v(x+\Delta x,\ y+\Delta y) - v(x,\ y) =$$
$$= \frac{\partial v}{\partial x}\Delta x + \frac{\partial v}{\partial y}\Delta y + o_2(\rho) \qquad (\rho \to 0),$$

где $\rho = |\Delta z| = \sqrt{\Delta x^2 + \Delta y^2}$, $o_1(\rho)$ и $o_2(\rho)$ $(\rho \to 0)$ — бесконечно малые функции высшего порядка малости чем ρ, т. е. $\lim\limits_{\rho \to 0} \frac{o_j(\rho)}{\rho} = 0 \quad (j = 1, 2)$. Поэтому, учитывая, что $o_1(\rho) + i o_2(\rho) = o(\rho)$ $(\rho \to 0)$, имеем (в силу (6.18))

$$\frac{\Delta w}{\Delta z} = \frac{\Delta u + i\Delta v}{\Delta x + i\Delta y} =$$
$$= \frac{\frac{\partial u}{\partial x}\Delta x + \frac{\partial u}{\partial y}\Delta y + i\left(\frac{\partial v}{\partial x}\Delta x + \frac{\partial v}{\partial y}\Delta y\right)}{\Delta x + i\Delta y} + \frac{o(\rho)}{\Delta z} =$$
$$= \frac{\frac{\partial u}{\partial x}(\Delta x + i\Delta y) + \frac{\partial v}{\partial x}(-\Delta y + i\Delta x)}{\Delta x + i\Delta y} + \frac{o(\rho)}{\rho}\cdot\frac{\rho}{\Delta z} =$$
$$= \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} + \underset{\rho \to 0}{o(1)},$$

потому что $\left|\frac{\rho}{\Delta z}\right| = \frac{\rho}{|\Delta z|} = 1$. Символ $\underset{\rho \to 0}{o(1)}$ означает бесконечно малую функцию при $\rho \to 0$. Таким образом,

$$\lim_{\Delta z \to 0} \frac{\Delta w}{\Delta z} = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x},$$

т. е. функция *f* имеет в точке *z* производную, равную

$$f'(z) = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x}. \qquad (6.19)$$

Используя условия (6.18), можно получить и другие формы для выражения производной *f' (z).* Теорема доказана.

Если учесть, что существование производной *f'(z)* на области *D* автоматически влечет за собой ее непрерывность на *D,* то из теорем 1 и 2 вытекает следующая

Теорема 3. *Для того чтобы функция*

$$f(z) = u(x,\ y) + iv(x,\ y)$$

была аналитической на области D плоскости z, необходимо и достаточно, чтобы частные производные первого порядка функций и и v были непрерывны на D и выполнялись условия Коши — Римана

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x} \qquad ((x, y) \in D).$$

Функции u и v называют *сопряженными друг к другу на* D.

Пример 1. Функции $|z| = \sqrt{x^2 + y^2}$, $\operatorname{Re} z = x$, $\operatorname{Im} z = y$ не являются аналитическими на плоскости z. Ведь каждая из них может быть записана в виде $f(z) = u + iv$, где $u \not\equiv 0$ и $v \equiv 0$ — действительные функции не удовлетворяющие условиям Коши—Римана.

Пример 2. Проверить выполнение условий Коши — Римана для действительной и мнимой частей функции $w = \cos z$.

В примере 1 п 6.2 мы показали, что

$$u(x, y) = \cos x \operatorname{ch} y, \quad v(x, y) = -\sin x \operatorname{sh} y.$$

Отсюда

$$\frac{\partial u}{\partial x} = -\sin x \operatorname{ch} y, \quad \frac{\partial u}{\partial y} = \cos x \operatorname{sh} y,$$

$$\frac{\partial v}{\partial x} = -\cos x \operatorname{sh} y, \quad \frac{\partial v}{\partial y} = -\sin x \operatorname{ch} y.$$

Таким образом,

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x} \quad (\forall x, y),$$

т. е. условия Коши—Римана выполнены.

Так как частные производные первого порядка от функций u и v непрерывны для любых точек (x, y), то функция $w = \cos z$ аналитична на всей комплексной плоскости.

Замечание. Если функцию $f(z)$ представить в виде

$$f(z) = R(x, y) \exp(i\Phi(x, y)),$$

где R—модуль, а Φ — аргумент функции $f(z)$, то условия Коши—Римана имеют вид

$$\frac{\partial R}{\partial x} = R\frac{\partial \Phi}{\partial y}, \quad \frac{\partial R}{\partial y} = -R\frac{\partial \Phi}{\partial x}.$$

6.4. Гармонические функции

Пусть на области D плоскости z задана аналитическая функция $f(z) = u(x, y) + iv(x, y)$. Тогда, как это уже было отмечено в п. 6.2, функция $f(z)$ имеет на D непрерывные производные любого порядка. Но тогда функции u и v имеют на D непрерывные частные производные любого порядка, а первые производные удовлетворяют условиям Коши—Римана

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}, \tag{6.20}$$

из которых следует

$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial^2 v}{\partial x \partial y} \quad \frac{\partial^2 u}{\partial y^2} = -\frac{\partial^2 v}{\partial x \partial y}.$$

Складывая эти равенства, получаем

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0. \tag{6.21}$$

Левую часть уравнения (6.21) обозначают символом

$$\Delta u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}.$$

Уравнение

$$\Delta u = 0 \tag{6.22}$$

называют *уравнением Лапласа.* Символ

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$$

называют *оператором Лапласа.*

Функцию *u*, имеющую непрерывные частные производные второго порядка на *D* и удовлетворяющую уравнению Лапласа (6.22), называют *гармонической на D.*

Итак, мы установили, что *действительная часть аналитической на D функции является гармонической функцией на D.*

Если первое равенство в (6.20) продифференцировать по *y,* а второе — по *x* и вычесть второе равенство из первого, то будем иметь

$$\Delta v = 0,$$

т, е. и *мнимая часть аналитической функции является гармонической функцией.*

Однако функция $f(z) = u + iv$, где *u* и *v* — произвольные гармонические на *D* функции, не всегда является аналитической на *D.* Она будет аналитической, только если функции *u* и *v* удовлетворяют на *D* условиям Коши—Римана.

Покажем, что если *D* есть односвязная область, то для всякой гармонической на *D* функции *u(x, y)* существует единственная, с точностью до произвольной постоянной, сопряженная к *u* на *D* функция *v* такая, что

$$f(z) = u(x, y) + iv(x, y)$$

аналитическая на *D.*

Пусть задана на *D* гармоническая функция *u (x, y),* Положим

$$P(x, y) = -\frac{\partial u}{\partial y}, \quad Q(x, y) = \frac{\partial u}{\partial x}.$$

Так как u имеет на D непрерывные частные производные второго порядка, удовлетворяющие уравнению Лапласа, то

$$\frac{\partial P}{\partial y} = -\frac{\partial^2 u}{\partial y^2} = \frac{\partial^2 u}{\partial x^2} = \frac{\partial Q}{\partial x}.$$

Из полученного равенства

$$\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x} \text{ на } D$$

и односвязности D следует, что криволинейный интеграл

$$\int_{(x_0,\ y_0)}^{(x,\ y)} (P\,dx + Q\,dy) = \int_{(x_0,\ y_0)}^{(x,\ y)} \left(-\frac{\partial u}{\partial y}\,dx + \frac{\partial u}{\partial x}\,dy\right) = v(x,\ y) \quad (6.23)$$

вдоль любого кусочно-гладкого пути $L \subset D$, соединяющего точки (x_0, y_0) и (x, y), зависит от этих точек, но не зависит от формы пути. При этом v есть функция, потенциальная для вектора (P, Q) на D, т, е,

$$\frac{\partial v}{\partial x} = P = -\frac{\partial u}{\partial y}, \quad \frac{\partial v}{\partial y} = Q = \frac{\partial u}{\partial x}. \quad (6.24)$$

Это показывает, что v имеет непрерывные частные производные на D, удовлетворяющие вместе с u условиям Коши—Римана. Но тогда u и v — сопряженные друг к другу функции.

Если v_1 — другая функция, сопряженная к u на D, то

$$\frac{\partial v_1}{\partial x} = -\frac{\partial u}{\partial y}, \quad \frac{\partial v_1}{\partial y} = \frac{\partial u}{\partial x}. \quad (6.25)$$

Из (6.24) и (6.25) следует:

$$\frac{\partial (v_1 - v)}{\partial x} = 0, \quad \frac{\partial (v_1 - v)}{\partial y} = 0 \text{ на } D.$$

Но тогда $v_1 - v = C$ на D, где C — постоянная. Утверждение доказано.

Пример 1. Функция $u = x^2 - y^2$ удовлетворяетуравнению $\Delta u = 0$. Найти аналитическую функцию $f(z)$, у которой $\mathrm{Re}\, f(z) = u$.

Мнимую часть этой функции ищем по формуле (6.23) (рис. 6.6):

$$v = \int_{(0,\ 0)}^{(x,\ y)} (2y\,dx + 2x\,dy) =$$

$$= \int_0^x (2 \cdot 0 + 2x \cdot 0)\,dx + \int_0^y (2y \cdot 0 + 2x)\,dy = \int_0^y 2x\,dy = 2xy + C.$$

Тогда функция $f(z) = (x^2 - y^2) + i(2xy + C) = z^2 + iC$ аналитическая во всей комплексной плоскости.

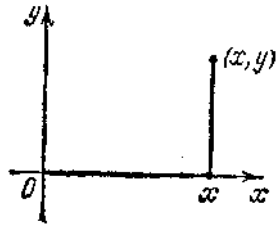


Рис. 6.6

Пример 2. Функция $w=z=x+iy$ аналитическая на плоскости z. Следовательно, функции $u=x$, $v=y$ гармонические и удовлетворяют условиям Коши— Римана на плоскости z. Это можно проверить непосредственно.

Пример 3. Функции $u=x$, $v=-y$ гармонические, но условия Коши—Римана для них не выполнены, поэтому функция $f(z)=x+i(-y)=\bar{z}$ не является аналитической.

Убедимся в этом непосредственно: $w=\bar{z}=x-iy$,

$$\Delta w=\overline{(z+\Delta z)}-\bar{z}=\bar{z}+\overline{\Delta z}-\bar{z}=\overline{\Delta z},\quad \frac{\Delta w}{\Delta z}=\frac{\overline{\Delta z}}{\Delta z}=\frac{\Delta x-i\Delta y}{\Delta x+i\Delta y}.$$

Выбираем два пути подхода точки $z+\Delta z$ к точке z, а именно, а) $\Delta x=0$, $\Delta y\to 0$; б) $\Delta x\to 0$, $\Delta y=0$. Тогда:

в случае а) $\frac{\Delta w}{\Delta z}=\frac{-i\Delta y}{i\Delta y}=-1$, т. е. $\frac{\Delta w}{\Delta z}\to -1$;

в случае б) $\frac{\Delta w}{\Delta z}=\frac{\Delta x}{\Delta x}=1$, т. е. $\frac{\Delta w}{\Delta z}\to 1$.

Таким образом, предела $\frac{\Delta w}{\Delta z}$ при $\Delta z\to 0$ не существует и функция $w=\bar{z}$ г не имеет производной в любой точке плоскости.

6.5. Обратная функция

Пусть задана аналитическая функция

$$w=f(z)\quad (z\in D), \tag{6.26}$$

отображающая область D плоскости г на область G плоскости w взаимно однозначно (или одно-однозначно). Это значит, что каждому $z\in D$ соответствует при помощи функции (6.26) одно значение $w\in G$ и при этом каждое $w\in G$ в силу этого закона соответствует только одному значению $z\in D$. Этим определена на G однозначная функция

$$z=\varphi(w)\quad (w\in G), \tag{6.27}$$

обладающая тем свойством, что

$$f[\varphi(w)] = w \qquad (w \in G).$$

Имеет место и другое равенство

$$\varphi[f(z)] = z \quad (z \in D).$$

Функция $z=\varphi(w)$ называется *обратной функцией* к функции $w = f(z)$ $(z \in D)$.

Покажем, что если

$$f'(z) \neq 0 \quad (z \in D),$$

то функция $z=\varphi(w)$ есть аналитическая функция на G.

В самом деле, пусть точки $w,\ w + \Delta w \in G$. Этим точкам соответствуют при помощи обратной функции точки $z,\ z+\Delta z$. Так как по условию функция f имеет производную в точке z, то она непрерывна в этой точке: $\Delta w \to 0$. если $\Delta z \to 0$. В силу указанной взаимной однозначности верно и обратное, $\Delta z \to 0$, если $\Delta w \to 0$. Но тогда

$$\lim_{\Delta w \to 0} \frac{\Delta z}{\Delta w} = \lim_{\Delta z \to 0} \frac{1}{\frac{\Delta w}{\Delta z}} = \frac{1}{f'(z)} \qquad (f'(z) \neq 0).$$

Это показывает, что производная от обратной функции $z=\varphi(w)$ существует в точке w и равна

$$\varphi'(w) = \frac{1}{f'(z)} \qquad (w \in G). \tag{6.28}$$

Так как w—произвольная точка G, то функция $\varphi(w)$ аналитическая на G.

Пример. Функция

$$w = az + b$$

при $a \neq 0$ отображает всю плоскость z на всю плоскость w взаимно однозначно. При этом обратная функция имеет вид

$$z = \frac{w-b}{a}.$$

Непосредственно видно, что обе эти функции аналитичны соответственно на плоскостях z и w $\left(w' = a,\ z' = \frac{1}{a}\right)$.

Функция $\sqrt[n]{z} = z^{1/n}$ *(*n—натуральное). Плоскость R точек z разрежем на n секторов лучами

$$\theta = \theta_k = \frac{2\pi}{n} k \quad (k = 0,\ 1,\ \ldots,\ n-1),$$

выходящими из нулевой точки (см. рис. 6.7, где n=3).

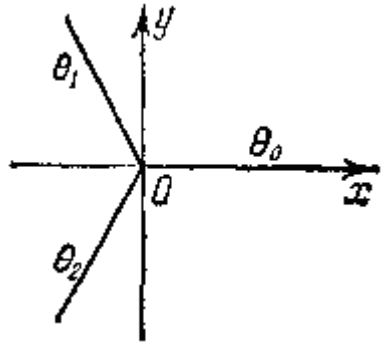


Рис. 6.7

Пусть D_k есть сектор

$$\theta_k < \theta < \theta_{k+1} \qquad (\rho > 0), \tag{6.29}$$

точнее, множество точек $z = \rho e^{t\theta}$, $\rho>0$, имеющих аргумент $\theta = \arg z$, удовлетворяющий неравенствам (6.29). Очевидно, D_k есть область. Обозначим также через D^*_k множество, получаемое добавлением к D_k луча $\theta = \theta_k$ (вместе с нулевой точкой). Точки D^*_k можно записать в виде

$$z = \rho e^{t\theta} \quad (\theta_k \leqslant \theta < \theta_{k+1},\ \rho \geqslant 0).$$

Положим еще

$$\theta = \theta_k + \psi \ (\theta_k \leqslant \theta < \theta_{k+1}).$$

Если $0 \leqslant \psi < \theta_1 = 2\pi/n$, то $\theta_k \leqslant \theta < \theta_{k+1}$ и обратно.

Функция $w = z^n$ отображает D^*_k взаимно однозначно и непрерывно на всю плоскость

$$w = re^{i\varphi} \ (0 \leqslant \varphi < 2\pi),$$

которую обозначим через R'. В самом деле,

$$re^{i\varphi} = \rho^n e^{in\psi} = \rho^n e^{in\left(\frac{2\pi}{n}k+\psi\right)} = \rho^n e^{in\psi},$$

поэтому

$$r = \rho^n,\ \varphi = n\psi \ \left(0 \leqslant \psi < \theta_1 = \frac{2\pi}{n}\right),$$

откуда

$$\rho = r^{1/n} = \sqrt[n]{r},\ \psi = \varphi/n,$$

где $\sqrt[n]{r}$ есть арифметическое значение корня n-й степени из r, т. е. неотрицательное число, n-я степень которого равна r. Из сказанного следует, что функция $w=z^n$ на множестве D^*_k имеет обратную функцию

$$z = (z)_k = \rho e^{i\theta} = r^{1/n} e^{i\frac{\varphi + 2k\pi}{n}}$$
$$(k = 0,\ 1,\ \ldots,\ n-1;\ w \in R'). \qquad (6.30)$$

Вообще же функция $w=z^n$ имеет обратную n-значную функцию

$$z = \sqrt[n]{w},$$

имеющую n непрерывных ветвей (6.30), соответствующих числам k = 0, 1, ..., n—1. Ветви (6.30), определяемые числами k = 0, 1, ..., n—1, отображают R' соответственно на $D_0^*,\ D_1^*,\ \ldots,\ D_{n-1}^*$.

Чтобы вычислить производную от k-й ветви, нам придется рассмотреть область $D_k \subset D_k^*$. Обозначим через R'_1 пространство R' без луча φ =0.

Аналитическая функция $w = z^n$ отображает взаимно однозначно D_k на R'_1. При этом соответствующая обратная функция определяется по формулам (6.30). В силу (6.28) производная от нее равна $(z \in D_k)$

$$\left(\sqrt[n]{w}\right)' = \left(\sqrt[n]{w}\right)'_k = \frac{1}{(z^n)'} = \frac{1}{nz^{n-1}} = \frac{z}{nw} =$$
$$= \frac{1}{n} \frac{r^{1/n} e^{i\frac{\varphi + 2k\pi}{n}}}{re^{i(\varphi + 2k\pi)}} = \frac{1}{n} r^{\frac{1}{n}-1} e^{i\left(\frac{1}{n}-1\right)[\varphi + 2k\pi]} = \frac{1}{n} w^{\frac{1}{n}-1}.$$

Рассматривая области D_k вместо множеств D^*_k мы исключаем из рассмотрения лучи $\theta = \theta_k$ плоскости R. Если бы нас интересовало поведение функции z^n в окрестности этих лучей, то следовало бы плоскость R разрезать лучами

$$\theta = \theta_k + \alpha \qquad (0 < \alpha < 2\pi/n,\quad k = 0,\ 1,\ \ldots,\ n-1)$$

и считать, что D_k, D^*_k суть множества точек $z \in R$, определяемых соответственно неравенствами

$$\theta_k + \alpha < \theta < \theta_{k+1} + \alpha, \quad \theta_k + \alpha \leqslant \theta < \theta_{k+1} + \alpha.$$

Функции e^z, $\ln z$. Функция

$$w = e^z = e^{x+iy} = e^x e^{iy} \qquad (z = x + iy)$$

аналитическая на плоскости R точек z. Она не равна нулю для всех $z \in R$. Это видно из того, что

$$e^x \neq 0 \quad \text{и} \quad |e^{iy}| = 1.$$

Обозначим через R' плоскость точек w, через R'_0 — эту плоскость с выкинутой из нее точкой O и через R'_1 — эту плоскость с выкинутым из нее положительным лучом оси x.

Из дальнейшего мы увидим, что образ R при помощи функции $w = e^z$ есть область R'_0. Однако отображение R на R'_0 не взаимнооднозначно — обратная функция к функции $w = e^z$, называемая *натуральным логарифмом* w и обозначаемая через

$$z = \ln w \qquad (w \in R_0'),$$

бесконечнозначна. Ниже мы определяем эту функцию. Для этого разрежем R на полосы прямыми (рис. 6.8)

$$y = y_k = 2\pi k \quad (k = 0, \pm 1, \pm 2, \ldots).$$

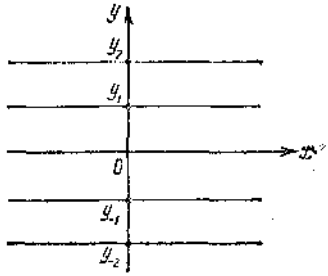


Рис. 6.8

Открытую полосу $y_k < y < y_{k+1}$ обозначим через $D_л$ и полузамкнутую полосу $y_k \le y \le y_{k+1}$ — через D^*_k.

Замена переменной y на η при помощи равенства

$$y = y_k + \eta$$

преобразовывает полосу D^*_k точек $z = x+iy$ *на* полосу D^*_1 точек $x+i\eta$ взаимно однозначно.

Рассмотрим функцию $w=e^z$ на множестве D^*_k. Полагая $z= x + iy$, $w=re^{i\varphi}$, $0 \le \varphi < 2\pi$, будем иметь

$$w = re^{i\varphi} = e^x e^{iy} = e^x e^{i(2k\pi + \eta)} = e^x e^{i\eta} \qquad (0 \leqslant \eta < 2\pi),$$

откуда

$$r = e^x, \quad \varphi = \eta.$$

Таким образом,

$$x = \ln r = \ln |w|,$$
$$y = y_k + \eta = y_k + \varphi = 2k\pi + \arg w \qquad (k = 0, \pm 1, \pm 2, \ldots).$$

Следовательно, функция $w = e^z$ имеет на полосе D^*_k обратную однозначную функцию

$$z = x + iy = (z)_k = \ln |w| + i(\arg w + 2k\pi) \qquad (6.31)$$
$$(k = 0, \pm 1, \ldots, w \in R_0').$$

Вообще же функция $w=e^z$ имеет обратную бесконечно-значную функцию

$$z = \ln w \qquad (w \in R_0'),$$

имеющую бесконечное число непрерывных ветвей (6.31), соответствующих числам $k = 0, \pm 1, \pm 2, \ldots$

Область D_k преобразуется при помощи аналитической функции $w=e^z$ на область R'_1 плоскости w взаимно однозначно. Обратная к ней однозначная функция, определяемая для *данного* k равенством (6.31), аналитическая на R'_1. Ее производную лучше всего вычислить с помощью формулы (6.28):

$$(\ln w)' = (\ln w)'_k = \frac{1}{(e^z)'} = \frac{1}{e^z} = \frac{1}{w} \qquad (w \in R'_1). \tag{6.32}$$

Мы здесь вычислили производную не от многозначной функции ln w, а от ее определенной однозначной ветви, соответствующей некоторому k.

Тот факт, что производная оказалась равной функции $1/w$, не зависящей от k, объясняется тем, что разные ветви (6.31) отличаются на постоянную.

При вычислении производной от z = ln w мы считали, что точки z принадлежат к областям D_k, исключив из рассмотрения прямые $y=y_k$ плоскости R.

Если бы мы интересовались поведением рассмотренных функций на прямых $y=y_k$, то тогда следовало бы разрезать плоскость R сдвинутыми прямыми

$$y = y_k + \alpha \qquad (0 < \alpha < 2\pi, \ y_k = 2\pi k, \quad k = 0, \pm 1, \pm 2, \ldots),$$

считая таким образом, что D_k есть область точек $z=x+iy$, для которых $y_k + \alpha < y < y_{k+1} + \alpha$.

Степенная функция z^α (α — действительное число) определяется по формуле

$$w = z^\alpha = e^{\alpha \ln z} = e^{\alpha [\ln |z| + i (\arg z + 2k\pi)]} = |z|^\alpha e^{i\alpha (\arg z + 2k\pi)}$$
$$(k = 0, \pm 1, \pm 2, \ldots) \tag{6.33}$$

или

$$w = \rho^\alpha e^{i\alpha (\theta + 2k\pi)} \quad (k = 0, \pm 1, \pm 2, \ldots),$$

где (6.34)

$$z = \rho e^{i\theta}.$$

Если α —целое, то

$$e^{i\alpha 2k\pi} = 1$$

и

$$w = (\rho e^{i\theta})^\alpha = z^\alpha,$$

где z^α понимается в обычном смысле как произведение α множителей z. Если $\alpha = \pm p/q$, где $p > 0$, $q > 0$, —целые, то числа справа в (6.34) существенно различны лишь при k = 0, 1, ..., q—1:

$$w = \rho^{\alpha} e^{i\alpha(\theta + 2k\pi)} \qquad (k = 0,\ 1,\ \ldots,\ q-1).$$

В частности, при $\alpha = 1/n$ и n натуральном мы получили уже эти факты (см. (5)).

Если же α —иррациональное число, то функции, определяемые формулой (6.33) или (6.34) для разных k, различны. Это *непрерывные ветви многозначной* (бесконечнозначной) *функции* $w = z^{\alpha}$.

Имеем далее $(z = \rho e^{i\theta})$

$$(z^{\alpha})' = (e^{\alpha \ln z})' = e^{\alpha \ln z} \cdot \frac{\alpha}{z} = \alpha \frac{e^{\alpha [\ln \rho + i(\theta + 2k\pi)]}}{e^{[\ln \rho + i(\theta + 2k\pi)]}} =$$

$$= \alpha e^{(\alpha - 1)[\ln \rho + i(\theta + 2k\pi)]} = \alpha z^{\alpha - 1},$$

т. е. равенство

$$(z^{\alpha})' = \alpha z^{\alpha - 1}, \tag{6.35}$$

верное для любой ветви z^{α}., При этом, с каким k взята ветвь z^{α} в левой части (6.35), с таким же k надо взять ветвь $z^{\alpha - 1}$ в правой части.

Замечание. Обратные функции для тригонометрических и гиперболических функций можно ввести аналогичным образом.

Например, функция $w = \operatorname{Arcsin} z$ является обратной к функции $z = \sin w$, т. е. $\sin[\operatorname{Arcsin} z] = z$.

Из уравнения

$$z = \sin w = \frac{e^{iw} - e^{-iw}}{2i} = \frac{e^{2iw} - 1}{2ie^{iw}}$$

находим

$$e^{2iw} - 2ize^{iw} - 1 = 0, \quad e^{iw} = iz \pm \sqrt{1 - z^2},$$

т. е.

$$iw = \ln|iz \pm \sqrt{1 - z^2}| + i[\arg(iz \pm \sqrt{1 - z^2}) + 2k\pi].$$

Таким образом,

$$w = \operatorname{Arcsin} z = -i[\ln|iz \pm \sqrt{1 - z^2}| + i[\arg(iz \pm \sqrt{1 - z^2}) + 2k\pi]]$$

— бесконечнозначная функция $(k = 0,\ \pm 1,\ \pm 2,\ \ldots)$. Аналогично можно получить

$$\operatorname{Arccos} z = -i[\ln|z \pm \sqrt{z^2 - 1}| + i[\arg(z \pm \sqrt{z^2 - 1}) + 2k\pi]],$$

$$\operatorname{Arctg} z = -\frac{i}{2}\left[\ln\left|\frac{1 + zi}{1 - zi}\right| + i\left(\arg\frac{1 + zi}{1 - zi} + 2k\pi\right)\right],$$

$$\operatorname{Arsh} z = \ln|z \pm \sqrt{z^2 + 1}| + i[\arg(z \pm \sqrt{z^2 + 1}) + 2k\pi],$$

$$\operatorname{Arch} z = \ln|z \pm \sqrt{z^2 - 1}| + i[\arg(z \pm \sqrt{z^2 - 1}) + 2k\pi].$$

6.6. Интегрирование функций комплексного переменного

Пусть $w = f(z) = u + iv$ — непрерывная функция комплексного z, определенная в области D и L — гладкая кривая, лежащая в D, с началом в точке A и концом в точке B (рис. 6.9), заданная уравнением

$$z = z(t) = x(t) + iy(t) \quad (\alpha \leqslant t \leqslant \beta)$$

или, что все равно, двумя уравнениями

$$\left.\begin{aligned} x &= x(t), \\ y &= y(t) \end{aligned}\right\} \quad (\alpha \leqslant t \leqslant \beta). \tag{6.36}$$

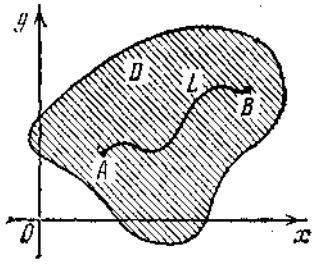


Рис. 6.9

Как всегда, направление на L соответствует измененению параметра t от α до β ($A = z(\alpha)$, $B = z(\beta)$).

Интеграл от функции $f(z)$ вдоль кривой L определяется следующим образом:

$$\int_L f(z)\,dz = \int_L (u + iv)(dx + i\,dy) = \int_L (u\,dx - v\,dy) +$$

$$+ i\int_L (v\,dx + u\,dy) = \int_\alpha^\beta [u(x(t), y(t))\,x'(t) - v(x(t), y(t))y'(t)]dt +$$

$$+ i\int_\alpha^\beta [v(x(t), y(t))\,x'(t) + u(x(t), y(t))\,y'(t)]\,dt. \tag{6.37}$$

Если учесть, что $z'(t) = x'(t) + iy'(t)$ и $u(x(t), y(t)) = u(z(t))$, то равенство (6.37) можно коротко записать так:

$$\int_L f(z)dz = \int_\alpha^\beta f[z(t)]z'(t)\,dt. \qquad (6.38)$$

Таким образом, из (6.37) видно, что интеграл по комплексному переменному есть сумма двух криволинейных интегралов, и его вычисление сводится к вычислению обыкновенных интегралов.

Интеграл (6.37) существует для любой непрерывной функции $f(z)$ (в этом случае функции $u(x, y)$ и $v(x, y)$ также непрерывны) и любой гладкой кривой L (т. е. когда $x'(t)$, $y'(t)$ непрерывны и $x'(t)^2 + y'(t^2) > 0$).

Если кривая L кусочно-гладкая и состоит из гладких ориентированных кусков $L_1, \ldots, L_n$, то по определению считаем

$$\int_L f(z)\,dz = \sum_{k=1}^{n} \int_{L_k} f(z)\,dz. \qquad (6.39)$$

На основании свойств криволинейного интеграла легко получаем

1) $$\int_L f(z)\,dz = -\int_{L_-} f(z)\,dz,$$

где L_- та же кривая, что и L, но ориентированная противоположно.

2) $$\int_L [Af(z) + B\varphi(z)]\,dz = A\int_L f(z)\,dz + B\int_L \varphi(z)\,dz,$$

где A, B—постоянные числа.

3) Если $|f(z)| \leqslant M$ при $z \in L$, то

$$\left|\int_L f(z)\,dz\right| \leqslant Ml,$$

где l — длина L.

В самом деле, на основании свойства обыкновенного интеграла имеем

$$\left|\int_L f(z)\,dz\right| = \left|\int_\alpha^\beta f[z(t)]z'(t)\,dt\right| \leqslant \int_\alpha^\beta |f[z(t)]|\cdot|z'(t)|\,dt \leqslant$$

$$\leqslant \int_\alpha^\beta M|z'(t)|\,dt = M\int_\alpha^\beta \sqrt{x'(t)^2 + y'(t)^2}\,dt = Ml.$$

Пример 1.

$$\int_L \frac{dz}{z - z_0} = 2\pi i, \qquad (6.40)$$

где L есть окружность с центром в точке z_0, ориентированная против часовой стрелки.

В самом деле, уравнение L можно записать в форме

$$z = z_0 + \rho e^{it} \qquad (0 \leqslant t < 2\pi),$$

где ρ — радиус окружности *L*. Поэтому

$$\int_L \frac{dz}{z - z_0} = \int_0^{2\pi} \frac{\rho i e^{it}\, dt}{\rho e^{it}} = i \int_0^{2\pi} 1\, dt = 2\pi i.$$

Пример 2. При целом $n \neq -1$

$$\int_L (z - z_0)^n\, dz = 0, \tag{6.41}$$

где *L*— снова окружность с центром в точке z_0, ориентированная против часовой стрелки.

В самом деле,

$$\int_L (z - z_0)^n\, dz = \int_0^{2\pi} \rho^{n+1} i e^{i(n+1)t}\, dt = i\rho^{n+1} \int_0^{2\pi} e^{i(n+1)t}\, dt =$$

$$= i\rho^{n+1} \frac{e^{i(n+1)t}}{i(n+1)} \Big|_0^{2\pi} = 0 \qquad (n + 1 \neq 0),$$

потому что $e^{i2(n+1)\pi} = 1$ для любых целых *n*.

Теорема 1 (К о ш и). *Если функция f*(z) *аналитическая на односвязной области D, то интеграл от f*(z) *по любому кусочно-гладкому замкнутому контуру* Г, *принадлежащему D, равен нулю:*

$$\int_\Gamma f(z)\, dz = 0.$$

Доказательство. Так как *f(z)=u+iv* — аналитическая на *D* функция, то функции *u(x, y)* и *v(x, y)* непрерывно дифференцируемы, и выполняются условия Коши—Римана:

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}, \tag{6.42}$$

в силу которых выражения *vdx+udy* и *udx—vdy* есть полные дифференциалы некоторых функций. Поэтому криволинейные интегралы по замкнутому контуру Г от этих выражений равны нулю. Но тогда, согласно равенству (6.37),

$$\int_\Gamma f(z)\, dz = \int_\Gamma (u\, dx - v\, dy) + i \int_\Gamma (v\, dx + u\, dy) = 0.$$

Пример 3.

$$\int_\Gamma z^n\,dz = 0 \qquad (n = 0,\ 1,\ 2,\ \ldots),$$

$$\int_\Gamma e^z\,dz = 0, \qquad \int_\Gamma a^z\,dz = 0 \qquad (a > 0),$$

$$\int_\Gamma \sin z\,dz = 0, \qquad \int_\Gamma \cos z\,dz = 0,$$

$$\int_\Gamma \operatorname{sh} z\,dz = 0, \qquad \int_\Gamma \operatorname{ch} z\,dz = 0,$$

где Γ — произвольный замкнутый кусочно-гладкий контур, потому что подынтегральные функции аналитические на плоскости *z*. Ведь они имеют непрерывную производную во всех точках *z* комплексной плоскости.

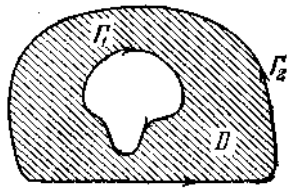


Рис. 6.10

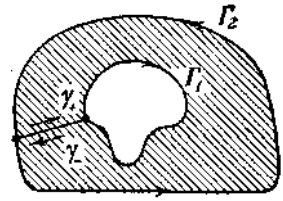


Рис. 6.11

Как следствие из теоремы 1 получаем следующую теорему.

Теорема 2. *Пусть область D комплексной плоскости ограничена сложным положительно ориентированным кусочно-гладким контуром* Γ, *т. е. при обходе по* Γ *точки D остаются слева. Тогда для функции f*(z), *аналитической на D, имеет место равенство*

$$\int_\Gamma f(z)\,dz = 0.$$

Поясним эту теорему. На рис. 6.10 изображена двусвязная область *D* с кусочно-гладким контуром $\Gamma=\Gamma_1+\Gamma_2$ ориентированным положительно.

Соединим контуры Γ_1 и Γ_2 гладким куском γ, как на рис. 6.11. Ориентируем γ двумя противоположными способами: γ_+, γ_-. В результате получим новую область D^1 односвязиую, ограниченную ориентированным контуром $\Gamma_2+\gamma_++\Gamma_1+\gamma_-$. По теореме 1

$$\int_{\Gamma_2+\gamma_++\Gamma_1+\gamma_-} f(z)\,dz = 0.$$

Но

$$\int\limits_{\gamma_- + \gamma_+} f(z)\,dz = \int\limits_{\gamma_-} f(z)\,dz + \int\limits_{\gamma_+} f(z)\,dz = 0,$$

поэтому

$$\int\limits_{\Gamma} f(z)\,dz = \int\limits_{\Gamma_1} f(z)\,dz + \int\limits_{\Gamma_2} f(z)\,dz = 0.$$

Каждый из интегралов

$$\int\limits_{\Gamma_1}, \int\limits_{\Gamma_2}$$

при этом может и не равняться нулю.

Замечание 1. Для краткости мы будем позволять себе писать «контур» вместо «замкнутый непрерывный кусочно-гладкий контур».

Из теоремы 2 как следствие вытекает

Теорема 3. *Пусть область D ограничена внешним контуром* Γ, *ориентированным против часовой стрелки, и внутренними контурами* Γ_1, Γ_2, ..., Γ_N, *ориентированными тоже против часовой стрелки* (как на рис. 6.12, где *N* = 3*), и пусть на D задана аналитическая функция f (z).*

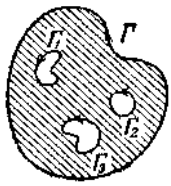


Рис. 6.12

Тогда имеет место равенство

$$\int\limits_{\Gamma} f(z)\,dz = \sum_{k=1}^{N} \int\limits_{\Gamma_k} f(z)\,dz. \tag{6.43}$$

В самомделе, если считать, что $\Gamma_{\bar{k}}$ — тот же контур, что и T_k, но ориентированный по часовой стрелке, то по теореме 2

$$\int\limits_{\Gamma} f(z)\,dz + \sum_{k=1}^{N} \int\limits_{\Gamma_k^-} f(z)\,dz = 0,$$

откуда следует (6.43), потому что

$$\int_{\Gamma_k^-} f(z)\,dz = -\int_{\Gamma_k} f(z)\,dz.$$

Отметим, что если в теореме 3 $N = 1$, то

$$\int_{\Gamma} f(z)\,dz = \int_{\Gamma_1} f(z)\,dz \tag{6.44}$$

рис. 6.13.

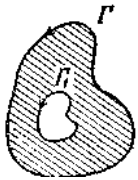


рис. 6.13

Замечание 2. Из равенства (6.44), т. е. из теоремы 3 при $N = 1$ следует, что равенства (6.40) и (6.41) остаются верными, если в них окружность L с центром в точке z_0 заменить на любой замкнутый кусочно-гладкий контур L', содержащий внутри точку z_0 и ориентированный против часовой стрелки:

$$\int_{L'} \frac{dz}{z - z_0} = 2\pi i, \tag{6.45}$$

$$\int_{L'} (z - z_0)^n\,dz = 0 \qquad (n \neq -1). \tag{6.46}$$

Формулы (6.45) и (6.46) являются *основными* в этой теории. Именно к ним, обычно, сводится вычисление криволинейных интегралов от аналитических функций.

6.7. Формула Коши

Пусть функция $f(z)$ аналитическая в односвязной замкнутой области $\bar{D}$ $(\bar{D} = D \cup \partial D)$, с кусочно-гладкой границей L, ориентированной в положительном направлении (рис. 6.14), т. е. против часовой стрелки. Тогда имеет место формула Коши

$$f(z_0) = \frac{1}{2\pi i} \int_L \frac{f(z)\,dz}{z - z_0}, \tag{6.47}$$

где z_0—любая точка внутри контура L.

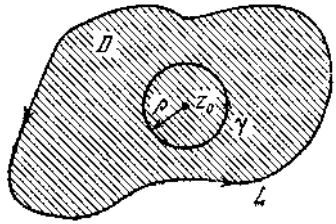


Рис. 6.14

Таким образом, аналитическую функцию достаточно определить на контуре L, а по формуле (6.47) можно автоматически получить ее значения в других точках D.

Для доказательства формулы (6.47) рассмотрим функцию

$$\varphi(z) = \frac{f(z) - f(z_0)}{z - z_0}, \tag{6.48}$$

Функция $\varphi(z)$ аналитическая во всех точках $\bar{D}$, кроме $z = z_0$. Опишем около точки z_0 окружность $\gamma \subset D$ (см. рис. 6.14), ориентированную положительно. Тогда по теореме 3 п. 6.6

$$\int_L \varphi(z)\,dz = \int_\gamma \varphi(z)\,dz \tag{6.49}$$

и значение интеграла

$$\int_\gamma \varphi(z)\,dz = \int_{|z-z_0|=\rho} \varphi(z)\,dz$$

на самом деле от ρ не зависит. Заметим, что из (6.48) следует, что

$$\lim_{z \to z_0} \varphi(z) = f'(z_0).$$

Если доопределить функцию $\varphi(z)$ в точке z_0, полагая $\varphi(z_0) = f'(z_0)$, то она становится непрерывной в замкнутой области D и, следовательно, ее модуль ограничен: $|\varphi(z)| \leqslant M \quad \forall z \in \bar{D}$. В силу этого

$$\left| \int_\gamma \varphi(z)\,dz \right| \leqslant M \cdot 2\pi\rho.$$

Так как число ρ можно взять как угодно малым и интеграл $\int_\gamma \varphi(z)\,dz$ от ρ не зависит, то

$$\int_{\gamma} \varphi(z)\,dz = 0.$$

Поэтому из (6.49) имеем

$$\int_{L} \varphi(z)\,dz = \int_{L} \frac{f(z)-f(z_0)}{z-z_0}\,dz = 0.$$

Так как

$$\int_{L} \frac{f(z_0)}{z-z_0}\,dz = f(z_0)\int_{L} \frac{dz}{z-z_0} = 2\pi i f(z_0),$$

то формула Коши доказана.

Формула Коши имеет место и для многосвязной области и доказательство ее может быть сведено к уже доказанной формуле Коши для односвязной области.

На рис. 6.15 изображена двусвязная область *D* с положительно ориентированной границей L, состоящей из двух замкнутых соответственно ориентированных контуров *(L = L_0+L_1).*

Пусть z_0 — произвольная точка *D.* Соединим контуры A_o и Lj кусочно-гладкой кривой *y,* ориентированной от *Li* к L_o,. не проходящей через точку z_0. Наряду с кривой *y* вводим совпадающую с ней кривую *y_,* но ориентированную противоположно.

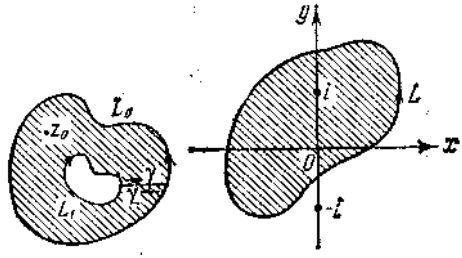


Рис. 6.15 Рис. 6.16

Если из *D* выкинуть γ, то оставшаяся область D_* будет односвязной с положительно ориентированной границей:

$$L' = L_0 + \gamma_- + L_1 + \gamma = L + \gamma_- + \gamma.$$

Функция *f(z)* аналитическая на $\overline{D}_*$ и $z_0 \in D_*$. Поэтому на основании теоремы Коши для односвязной области

$$f(z_0) = \frac{1}{2\pi i}\int_{L'} \frac{f(z)}{z-z_0}\,dz =$$

$$= \frac{1}{2\pi i}\left(\int_{L} + \int_{\gamma} + \int_{\gamma_-}\right)\frac{f(z)}{z-z_0}\,dz = \frac{1}{2\pi i}\int_{L} \frac{f(z)}{z-z_0}\,dz,$$

потому что

$$\int_{\gamma} \frac{f(z)}{z-z_0}\,dz + \int_{\gamma_-} \frac{f(z)}{z-z_0}\,dz = 0.$$

Пример. Вычислить интеграл

$$\int_{L} \frac{\sin z}{z^2+1}\,dz,$$

где L — ориентированный против часовой стрелки контур, содержащий в себе точку $z = i$ (рис. 6.16) и такой, что точка $z = -i$ находится вне него. Запишем наш интеграл в виде

$$\int_{L} \frac{\sin z\,dz}{(z+i)(z-i)}$$

и рассмотрим функцию $f(z) = \sin z/(z+i)$. В силу наших предположений о контуре L эта функция апалитична в замкнутой области, ограниченной контуром L, поэтому по формуле Коши

$$\int_{L} \frac{\sin z\,dz}{z^2+1} = \int_{L} \frac{f(z)}{z-i}\,dz = 2\pi i f(i) = 2\pi i \frac{\sin i}{2i} = \pi \sin i = \pi i \operatorname{sh} 1.$$

6.8. Интеграл типа Коши

Выражение

$$\frac{1}{2\pi i}\int_{L} \frac{f(z)\,dz}{z-z_0},$$

где $f(z)$ — аналитическая функция в замкнутой области D, ограниченной положительно ориентированным контуром L, называется *интегралом Коши.*

Если z_0 лежит внутри L, то интеграл равен $f(z_0)$, если же z_0 лежит вне L, то

$$\frac{f(z)}{(z-z_0)}$$

—аналитическая функция в D и, следовательно, интеграл Коши равен нулю.

Пусть теперь $\mathscr{L}$ — любая кусочно-гладкая ориентированная кривая, не обязательно замкнутая, и $\varphi(z)$ — непрерывная функция, определенная вдоль $\mathscr{L}$. Выражение

$$F(z_0) = \frac{1}{2\pi i} \int_{\mathscr{L}} \frac{\varphi(z)}{z - z_0} dz \qquad (6.50)$$

называется *интегралом типа Коши.* Оно представляет собой функцию $F(z_0)$, определенную вне $\mathscr{L}$ $(z_0 \overline{\in} \mathscr{L})$.

Теорема 1. *Интеграл* (6.50) *типа Коши есть аналитическая функция* $F(z_0)$ *для всех* $z_0 \overline{\in} \mathscr{L}$.

Производная порядка n от $F(z_0)$ *вычисляется по формуле*

$$F^{(n)}(z_0) = \frac{n!}{2\pi i} \int_{\mathscr{L}} \frac{\varphi(z)\, dz}{(z - z_0)^{n+1}} \qquad (n = 1, 2, \ldots). \qquad (6.51)$$

Доказательство. Пусть σ есть произвольный круг, не имеющий общих точек с кривой $\mathscr{L}$. Функция двух комплексных переменных z_0 и z

$$\Phi(z_0, z) = \frac{\varphi(z)}{z - z_0}$$

непрерывна на множестве $\sigma \times \mathscr{L}$ $(z_0 \in \sigma, \ z \in \mathscr{L})$ и имеет на нем непрерывную частную производную

$$\frac{\partial \Phi}{\partial z_0} = \frac{\varphi(z)}{(z - z_0)^2}$$

(надо учесть, что так как круг σ не пересекается с $\mathscr{L}$, то при любых $z_0 \in \sigma$ и $z \in \mathscr{L}$ разность $z - z_0 \neq 0$). Это показывает, что дифференцирование $F(z_0)$ по параметру z_0 законно произвести под знаком интеграла в (6.50):

$$F'(z_0) = \frac{1}{2\pi i} \int_{\mathscr{L}} \frac{\varphi(z)\, dz}{(z - z_0)^2}.$$

При этом производная $F'(z_0)$ непрерывна вне $\mathscr{L}$. Но тогда $F(z_0)$ аналитична вне $\mathscr{L}$.

Мы доказали формулу (6.51) в случае $n=1$. Для $n=2$ рассуждения ведутся по индукции.

Следствие. *Если функция* $w = f(z)$ *аналитическая в области D, т. е. имеет непрерывную первую производную на D, то она имеет производные всех порядков.*

Доказательство. Пусть z_0—любая точка D и σ — круг с центром в z_0, целиком лежащий в области D, а γ — окружность — граница σ, ориентированная против часовой стрелки. Тогда по формуле Коши

$$f(z_0)=\frac{1}{2\pi i}\int_{\gamma}\frac{f(z)\,dz}{z-z_0},$$

т. е. функция $f(z_0)$ изображается интегралом типа Коши при $\mathscr{L}=\gamma$ и $\varphi(z)=f(z)$. Значит, в силу теоремы 1 $f(z)$ бесконечно дифференцируема и

$$f^{(n)}(z_0)=\frac{n!}{2\pi i}\int_{\gamma}\frac{f(z)\,dz}{(z-z_0)^{n+1}} \qquad (n=1,\ 2,\ \ldots). \tag{6.52}$$

6.9. Степенной ряд

Рассмотрим степенной ряд

$$f(z)=\sum_{k=0}^{\infty}c_k(z-z_0)^k \qquad (|z-z_0|<R), \tag{6.53}$$

имеющий радиус сходимости $R > 0$.

Из теории степенных рядов мы знаем, что ряд (6.53) равномерно сходится на круге $|z|\le\rho$, где ρ —любое положительное число, меньшее R ($\rho < R$). Поэтому сумма $f(z)$ ряда (6.53) — непрерывная функция в открытом круге $|z-z_0|<R$. Больше того, $f(z)$ имеет на этом круге непрерывную производную $f^{(n)}(z)$ любого порядка, которую можно вычислить путем почленного дифференцирования ряда (6.53). Это показывает, что сумма степенного ряда есть *аналитическая функция* в круге (открытом!) его сходимости. Числа c_k вычисляются по формуле

$$c_k=\frac{f^{(k)}(z_0)}{k!} \qquad (k=0,\ 1,\ 2,\ \ldots), \tag{6.54}$$

что показывает, что степенной ряд есть ряд Тейлора своей суммы. В силу равенств (6.52) эту формулу можно заменить следующей:

$$c_k=\frac{1}{2\pi i}\int_{L}\frac{f(z)\,dz}{(z-z_0)^{k+1}} \qquad (k=0,\ 1,\ 2,\ \ldots),$$

где L — произвольный контур, ориентированный против часовой стрелки, принадлежащий к кругу сходимости ряда (6.53) и содержащий внутри точку z_0.

Но верна также

Теорема 1. *Функция $f(z)$, аналитическая в круге $|z-z_0|<R$, разлагается в сходящийся к ней степенной ряд по степеням $(z-z_0)$.*

Доказательство. Пусть $f(z)$ аналитическая в круге $|z-z_0|<R$. Обозначим через z любую точку внутри этого круга (рис. 6.17).

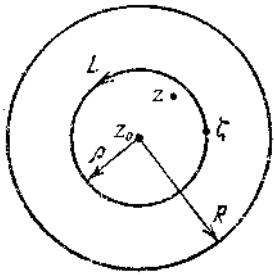


Рис. 6.17.

Опишем положительно ориентированную окружность L с центром в точке z_0 и радиуса $\rho < R$ так, чтобы точка z оказалась внутри контура L. Тогда функция $f(z)$ будет аналитической на контуре L и внутри него. Поэтому по теореме Коши

$$f(z)=\frac{1}{2\pi i}\int_L \frac{f(\zeta)\,d\zeta}{(\zeta-z)}. \tag{6.55}$$

Дробь $1/(\zeta-z)$ можно представить в виде

$$\frac{1}{\zeta-z}=\frac{1}{\zeta-z_0}\cdot\frac{1}{1-\dfrac{z-z_0}{\zeta-z_0}}. \tag{6.56}$$

Так как точка $\zeta\in L$, а z находится внутри этогр контура, то

$$\left|\frac{z-z_0}{\zeta-z_0}\right|=\frac{|z-z_0|}{\rho}<1. \tag{6.57}$$

Поэтому

$$1\Big/\left(1-\frac{z-z_0}{\zeta-z_0}\right)$$

можно рассматривать как сумму сходящейся геометрической прогрессии

$$\frac{1}{1-\dfrac{z-z_0}{\zeta-z_0}}=1+\frac{z-z_0}{\zeta-z_0}+\left(\frac{z-z_0}{\zeta-z_0}\right)^2+\dots \tag{6.58}$$

Из (6.56) и (6.58) получаем

$$\frac{1}{\zeta-z}=\frac{1}{\zeta-z_0}+\frac{z-z_0}{(\zeta-z_0)^2}+\frac{(z-z_0)^2}{(\zeta-z_0)^3}+\dots, \tag{6.59}$$

причем ряд (6.59) равномерно сходится при любых $\zeta\in L$ и постоянном z, потому что, как это видно из (6.57), выражение

$$\left|\frac{z-z_0}{\zeta-z_0}\right|$$

не зависит от $\zeta\in L$ и меньше 1.

Умножая (6.59) на $\frac{f(\zeta)}{2\pi i}$ (не нарушая его равномерной сходимости) и интегрируя вдоль L, имеем

$$\frac{1}{2\pi i}\int_L \frac{f(\zeta)\,d\zeta}{\zeta - z} = \frac{1}{2\pi i}\int_L \frac{f(\zeta)\,d\zeta}{\zeta - z_0} + \frac{z - z_0}{2\pi i}\int_L \frac{f(\zeta)\,d\zeta}{(\zeta - z_0)^2} + \ldots$$

В силу (6.55)

$$f(z) = c_0 + c_1(z - z_0) + c_2(z - z_0)^2 + \ldots, \qquad (6.60)$$

где мы обозначили

$$c_n = \frac{1}{2\pi i}\int_L \frac{f(\zeta)\,d\zeta}{(\zeta - z_0)^{n+1}} = \frac{f^{(n)}(z_0)}{n!} \qquad (n = 0, 1, 2, \ldots). \qquad (6.61)$$

Итак, мы доказали, что аналитическая функция $f(z)$ в круге $|z - z_0| < R$ изображается степенным рядом (6.60) с коэффициентами (6.61), т. е. своим рядом Тейлора.

Пример 1. При разложении функций в ряд Тейлора можно использовать известные разложения элементарных функций. Например,

$$\cos z = \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n}}{(2n)!},$$

поэтому

$$\sin^2 z = \frac{1 - \cos 2z}{2} = \frac{-1}{2}\sum_{n=1}^{\infty} (-1)^n \frac{(2z)^{2n}}{(2n)!}.$$

Пример 2. Функция $\operatorname{tg} z = \frac{\sin z}{\cos z}$ в достаточно малой окрестности $z = 0$ является аналитической функцией $\left((\operatorname{tg} z)' = \frac{1}{\cos^2 z}, \cos z \neq 0\right)$. Поэтому данную функцию можно разложить в ряд Тейлора по степеням z, хотя общий вид коэффициента трудно вычислить. Имеем по формуле (6.54): $c_0 = 0$, $c_1 = 1$, $c_2 = 0$, $c_3 = 2$, т. е.

$$\operatorname{tg} z = z + \frac{2z^3}{3!} + \ldots$$

Пример 3. Разложить в ряд Тейлора функции $w = \operatorname{sh} z$ и $w = \operatorname{ch} z$. Имеем

$$e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!},$$

поэтому

$$\operatorname{sh} z = \frac{e^z - e^{-z}}{2} = \sum_{n=0}^{\infty} \frac{z^{2n+1}}{(2n+1)!}, \quad \operatorname{ch} z = \frac{e^z + e^{-z}}{2} = \sum_{n=0}^{\infty} \frac{z^{2n}}{(2n)!},$$

6.10. Ряд Лорана

Теорема 1. *Пусть*

$$0 \leqslant r < R \leqslant \infty.$$

Всякая аналитическая в кольце

$$r < |z - z_0| < R \tag{6.62}$$

функция f(z) однозначно представляется в этом кольце в виде сходящегося ряда

$$f(z) = \sum_{n=-\infty}^{\infty} c_n (z - z_0)^n = \sum_{n=0}^{\infty} c_n (z - z_0)^n + \sum_{n=1}^{\infty} \frac{c_{-n}}{(z - z_0)^n}, \tag{6.63}$$

где

$$c_n = \frac{1}{2\pi i} \int_{\gamma} \frac{f(\zeta)\, d\zeta}{(\zeta - z_0)^{n+1}} \qquad (n = 0, \pm 1, \pm 2, \ldots), \tag{6.64}$$

а γ — любая окружность $|\zeta - z_0| = \rho$, $r < \rho < R$, ориентированная против часовой стрелки.

Ряд (6.63) называется *рядом Лорана функции f*(z) по степеням $(z - z_0)$ или *разложением Лорана функции f (z) в кольце* $z < |z - z_0| < R$.

Замечание. Когда говорят, что ряд $\sum_{-\infty}^{\infty}$ сходится, под этим подразумевается, что сходятся отдельно ряды

$$\sum_{n=0}^{\infty} \quad \text{и} \quad \sum_{n=-\infty}^{-1}.$$

Доказательство теоремы 1. Возьмем ориентированные против часовой стрелки окружности *c* и *C* радиусов r' и R' с центром в точке z_0, где $r < r' < R' < R$ (рис. 6.18).

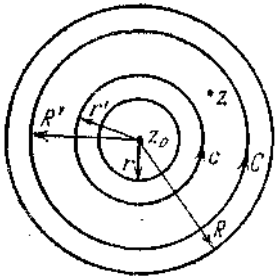


Рис. 6.18

В силу условия теоремы $f(z)$ аналитична в кольце между окружностями c и C и на самих окружностях.
Поэтому по формуле Коши для сложного контура имеем

$$f(z)=\frac{1}{2\pi i}\int_{C}\frac{f(\zeta)\,d\zeta}{\zeta-z}+\frac{1}{2\pi i}\int_{c_-}\frac{f(\zeta)\,d\zeta}{\zeta-z}$$

или

$$f(z)=\frac{1}{2\pi i}\int_{C}\frac{f(\zeta)\,d\zeta}{\zeta-z}-\frac{1}{2\pi i}\int_{c}\frac{f(\zeta)\,d\zeta}{\zeta-z}, \tag{6.65}$$

где z — точка между окружностями c и C.
В первом интеграле точка ζ обозначает точку окружности C, поэтому

$$\left|\frac{z-z_0}{\zeta-z_0}\right|=\frac{|z-z_0|}{R'}<1,$$

$$\frac{1}{\zeta-z}=\frac{1}{(\zeta-z_0)\left[1-\dfrac{z-z_0}{\zeta-z_0}\right]}=\sum_{n=0}^{\infty}\frac{(z-z_0)^n}{(\zeta-z_0)^{n+1}}, \tag{6.66}$$

причем ряд справа сходится равномерно для $\zeta\in C$ (при фиксированном z).

Во втором интеграле точка ζ обозначает точку окружности c, поэтому

$$\left|\frac{\zeta-z_0}{z-z_0}\right|=\frac{r'}{|z-z_0|}<1,$$

$$\frac{1}{\zeta-z}=\frac{-1}{(z-z_0)\left[1-\dfrac{\zeta-z_0}{z-z_0}\right]}=-\sum_{n=0}^{\infty}\frac{(\zeta-z_0)^n}{(z-z_0)^{n+1}}, \tag{6.67}$$

причем ряд справа сходится равномерно для всех $\zeta\in c$ (при фиксированном z).

Подставляя (6.66) и (6.67) в (6.65) и почленно интегрируя, получаем

$$f(z)=\sum_{n=0}^{\infty}\frac{1}{2\pi i}\int_{C}\frac{f(\zeta)\,d\zeta}{(\zeta-z_0)^{n+1}}(z-z_0)^n+$$
$$+\sum_{n=0}^{\infty}\frac{1}{2\pi i}\int_{c}\frac{f(\zeta)\,d\zeta}{(\zeta-z_0)^{-n}}(z-z_0)^{-n-1}=$$
$$=\sum_{n=0}^{\infty}\frac{1}{2\pi i}\int_{C}\frac{f(\zeta)\,d\zeta}{(\zeta-z_0)^{n+1}}(z-z_0)^n+\sum_{n=1}^{\infty}\frac{1}{2\pi i}\int_{c}\frac{f(\zeta)\,d\zeta}{(\zeta-z_0)^{-n+1}}(z-z_0)^{-n}. \tag{6.68}$$

Так как функция $f(\zeta)/(\zeta-z_0)^{n+1}$ при любом n аналитична в кольце, то в силу теоремы Коши интеграл (6.64) равен подобному интегралу по

любой другой окружности, в частности по c и C. Поэтому из (6.68) следует (6.63), где числа c_n вычисляются по формулам (6.64).

Первый ряд

$$\sum_{n=0}^{\infty} c_n (z-z_0)^n$$

в правой части (6.63), сходится в круге $|z—z_0|<R$ к некоторой аналитической в этом круге функции $f_1(z)$. Он называется *правильной частью ряда Лорана.*

Второй ряд в правой части (6.63)

$$\sum_{n=1}^{\infty} c_{-n} (z-z_0)^{-n},$$

сходится при $|z—z_0|>r$. Он определяет некоторую аналитическую функцию $f_2(z)$, называемую *главной частью ряда Лорана.*

Итак,

$$f(z) = f_1(z) + f_2(z),$$

где $f_1(z)$ — функция аналитическая в круге $|z—z_0|<R$, а $f_2(z)$ — вне круга радиуса r с центром в точке z_0 $(|z-z_0|>r)$. Внутри кольца

$$r<|z-z_0|<R$$

обе эти функции аналитичны.

Коэффициенты ряда Лорана c_n рассматриваемой функции $f(z)$ единственны, потому что они вычисляются по формулам (6.64).

Пример 1. Функция

$$f(z) = \frac{1}{(z-2)(z-3)} = \frac{1}{z-3} - \frac{1}{z-2}$$

аналитична на плоскости z, за исключением точек $z = 2$ и $z = 3$.

а) Функция $f(z)$ аналитична в круге $|z| < 2$, и потому на основании теоремы 1 п 6.9 ее можно разложить в ряд Тейлора по степеням z, сходящийся в круге $|z|<2$:

$$f(z) = \sum_{k=0}^{\infty} c_k z^k. \qquad (6.69)$$

Числа c_k можно вычислить по формуле

$$c_k = \frac{f^{(k)}(0)}{k!} \qquad (k = 0,\ 1,\ 2,\ \ldots). \qquad (6.70)$$

Однако в данном случае ряд (6.69) можно также получить, применив формулу для суммы членов убывающей геометрической прогрессии. Имеем (если $|z|<2$)

$$\frac{1}{z-3} = -\frac{1}{3}\frac{1}{1-\frac{z}{3}} = -\frac{1}{3}\left[1+\frac{z}{3}+\left(\frac{z}{3}\right)^2+\ldots\right] =$$

$$= -\frac{1}{3}\sum_{k=0}^{\infty}\left(\frac{z}{3}\right)^k,$$

$$\frac{1}{z-2} = -\frac{1}{2}\frac{1}{1-\frac{z}{2}} = -\frac{1}{2}\left[1+\frac{z}{2}+\left(\frac{z}{2}\right)^2+\ldots\right] =$$

$$= -\frac{1}{2}\sum_{k=0}^{\infty}\left(\frac{z}{2}\right)^k.$$

Поэтому для нашей функции $c_k = \frac{1}{2^{k+1}} - \frac{1}{3^{k+1}}$.

В силу единственности разложения функции в степенной ряд полученные числа c_k равны соответственно числам c_k, вычисляемым по формуле (6.70).

б) Функция $f(z)$ аналитична в кольце $2<|z|<3$. Поэтому ее можно разложить в ряд Лорана

$$f(z) = \sum_{-\infty}^{\infty} c_k z^k, \tag{6.71}$$

$$c_k = \frac{1}{2\pi i}\int_{\gamma}\frac{f(\zeta)\,d\zeta}{\zeta^{n+1}} \qquad (k=0,\ \pm 1,\ \pm 2,\ \ldots), \tag{6.72}$$

где γ — окружность $|\zeta| = \rho$, $2 < \rho < 3$, ориентированная против часовой стрелки. Но числа c_k можно получить, не прибегая к сложным формулам (6.72). Имеем для $2<|z|<3$

$$\frac{1}{z-3} = -\sum_{k=0}^{\infty}\frac{z^k}{3^{k+1}},$$

$$\frac{1}{z-2} = \frac{1}{z}\frac{1}{1-\frac{2}{z}} = \frac{1}{z}\left[1+\frac{2}{z}+\left(\frac{2}{z}\right)^2+\ldots\right] = \sum_{k=1}^{\infty}\frac{2^{k-1}}{z^k}.$$

Поэтому ряд Лорана функции $f(z)$ имеет вид

$$f(z) = -\sum_{k=1}^{\infty}\frac{2^{k-1}}{z^k} - \sum_{k=0}^{\infty}\frac{z^k}{3^{k+1}}.$$

Вследствие единственности разложения в ряд Лорана полученные коэффициенты равны соответственно числам c_k, определяемым по формулам (6.72).

в) Функция $f(z)$ аналитична также по внешности круга $|z|\leq 3$, т. е. для значений z, удовлетворяющих неравенству $|z|>3$ и обладает свойством

$$\lim_{z\to\infty} f(z) = 0. \tag{6.73}$$

Поэтому $f(z)$ можно разложить в ряд Лорана вида

$$f(z) = \sum_{k=1}^{\infty} \frac{c_{-k}}{z^k}. \tag{6.74}$$

Члены вида $c_k z^k$ (k = 0, 1, ...) не могут входить в разложение Лорана функции f, т. е. c_k = 0 для указанных k. Иначе это противоречило бы свойству (6.73).

Числа c_{-k} здесь тоже можно получить непосредственно. Имеем для $|z|>3$

$$\frac{1}{z-3} = \frac{1}{z}\frac{1}{1-\frac{3}{z}} = \frac{1}{z}\left[1+\frac{3}{z}+\left(\frac{3}{z}\right)^2+\dots\right] = \sum_{k=1}^{\infty}\frac{3^{k-1}}{z^k},$$

$$\frac{1}{z-2} = \sum_{k=1}^{\infty}\frac{2^{k-1}}{z^k}.$$

Поэтому

$$f(z) = \sum_{k=1}^{\infty}\frac{3^{k-1}-2^{k-1}}{z^k}.$$

Пример 2. Надо разложить функцию

$$f(z) = \frac{1}{(z-2)(z-3)} = \frac{1}{z-3} - \frac{1}{z-2} \tag{6.75}$$

в ряд Тейлора по степеням $z—i$ и определить радиус сходимости этого ряда.

Решение. Наибольший круг с центром в точке i, внутри которого функция $f(z)$ аналитическая, имеет радиус, равный расстоянию от точки i до ее ближайшей особой точки. Таковой является, очевидно, точка z = 2. Следовательно, указанный радиус равен

$$R = |2-i| = \sqrt{2^2+1^2} = \sqrt{5}.$$

Обозначим через σ открытый круг (без границы) с центром в точке i радиуса $R = \sqrt{5}$. Внутри круга σ функция $f(z)$ аналитическая, а любой концентрический ему круг большего радиуса содержит в себе особую точку z = 2, в которой аналитичность нарушается.

На основании теоремы 1 п 6.9 функция $f(z)$ разлагается в ряд Тейлора по степеням z — i. Этот ряд легко получить эффективно.

Имеем

$$\frac{1}{z-2}=\frac{1}{(z-i)+(i-2)}=\frac{1}{\frac{z-i}{i-2}+1}\cdot\frac{1}{i-2}=$$

$$=\frac{1}{i-2}\left(1-\frac{z-i}{i-2}+\left(\frac{z-i}{i-2}\right)^2-\dots\right), \quad (6.76)$$

и мы получили степенной ряд по степеням z—*i*, сходящийся в круге $|z-i|<R$, $R=|i-2|=\sqrt{5}$. Далее

$$\frac{1}{z-3}=\frac{1}{(z-i)+(i-3)}=\frac{1}{i-3}\frac{1}{1+\frac{z-i}{i-3}}=$$

$$=\frac{1}{i-3}\left(1-\frac{z-i}{i-3}+\left(\frac{z-i}{i-3}\right)^2-\dots\right). \quad (6.77)$$

Снова получен степенной ряд по степеням $z - i$, тоже сходящийся в круге $|z—i|<R$. На самом деле он сходится в круге радиуса $|i-3|=\sqrt{10}$, но это нам не понадобится.

Разность рядов (6.77) и (6.76) есть разложение в ряд Тейлора по степеням $z —i$ функции $f(z)$. Радиус сходимости этого ряда равен $R = \sqrt{5}$.

6.11. Классификация изолированных особых точек. Вычеты

В п. 6.10 была доказана теорема 1, утверждающая, что если $0 \leqslant r < R \leqslant \infty$ и функция $f(z)$ аналитична в кольце

$$r<|z-z_0|<R,$$

то она разлагается в сходящийся к ней ряд Лорана

$$f(z)=\sum_{k=-\infty}^{\infty} c_k (z-z_0)^k = f_1(z)+f_2(z) \quad (r<|z-z_0|<R), \quad (6.78)$$

где

$$\left.\begin{aligned} f_1(z)&=\sum_{k=0}^{\infty} c_k (z-z_0)^k \qquad (|z-z_0|<R),\\ f_2(z)&=\sum_{k=1}^{\infty}\frac{c_{-k}}{(z-z_0)^k} \qquad (|z-z_0|>r). \end{aligned}\right\} \quad (6.79)$$

Пусть r=0. Предполагается, таким образом, что функция аналитична в открытом круге $0 < |z—z_0| < R$, из которого выколота точка z_0. В самой точке z_0 функция f чаще всего бывает не определена.

Говорят в этом случае, что z_0 есть *изолированная особая точка* функции f. Ниже будет дана классификация изолированных особых точек.

Степенной ряд

$$f_1(z) = \sum_{k=0}^{\infty} c_k (z - z_0)^k \qquad (|z - z_0| < R)$$

имеет радиус сходимости $R>0$. Поэтому его сумма имеет непрерывную производную в круге $|z—z_0| < R$.

Рассмотрим три случая (при $z = 0!$).

Случай а). Функция $f(z)$ имеет вид

$$f(z) = f_1(z) = \sum_{k=0}^{\infty} c_k (z - z_0)^k, \qquad (6.80)$$

т. е. все числа $c_k = 0$ ($k = 1, 2, ...$). Так как степенной ряд (6.80) сходится для всех z с $|z—z_0| <R$, то его радиус сходимости равен R и, следовательно его сумма $f_1(z)$ определена и непрерывно дифференцируема во всех точках круга $|z—z_0|<R$, в том числе и в точке z_0. Таким образом, функция $f_1(z)$ аналитична в этом круге. Поэтому, если принять, что

$$f(z_0) = f_1(z_0) = c_0,$$

то и функция $f(z)$ будет аналитической в этом круге.

В этом случае говорят, что *особенность у функции f в точке z_0 устранима.* Достаточно положить $f(z_0)=c_0$, как функция f станет аналитической не только поблизости от точки z_0, но и в самой точке. Заметим, что в данном случае интеграл

$$\int_L f(z)\, dz = 0 \qquad (6.81)$$

для любого замкнутого контура L, содержащего внутри точку z_0 и принадлежащего к кругу $|z—z_0|<R$.

Случай б). Функция $f(z)$ имеет вид

$$f(z) = f_1(z) + \sum_{k=1}^{m} \frac{c_{-k}}{(z - z_0)^k} = \sum_{k=-m}^{\infty} c_k (z - z_0)^k \quad (c_{-m} \neq 0). \qquad (6.82)$$

Таким образом, $c_k = 0$ для $k = -(m+1), -(m+2), \ldots$

В этом случае говорят, что точка z_0 *есть полюс функции $f(z)$ порядка (кратности) m.* При $m=1$ точку z_0 называют еще *простым полюсом.*

Так как

$$\lim_{z \to z_0} f_1(z) = c_0$$

п

$$\lim_{z \to z_0} \sum_{k=1}^{m} \frac{c_{-k}}{(z-z_0)^k} =$$

$$= \lim_{z \to z_0} \frac{1}{(z-z_0)^m} [c_{-m} + c_{-(m-1)}(z-z_0) + \ldots$$

$$\ldots + c_{-1}(z-z_0)^{m-1}] = \infty, \qquad (6.83)$$

то

$$\lim_{z \to z_0} f(z) = \infty. \qquad (6.84)$$

Теперь, если L—контур, ориентированный против часовой стрелки, содержащий внутри z_0 и принадлежащий к кругу $|z-z_0| < R$, то

$$\int_L f(z)\,dz = 2\pi i\, c_{-1}. \qquad (6.85)$$

В самом деле,

$$\int_L f(z)\,dz = \int_L f_1(z)\,dz + \sum_{k=1}^{m} \int_L \frac{c_{-k}}{(z-z_0)^k} dz = 0 + 2\pi i\, c_{-1},$$

потому что

$$\int_L \frac{dz}{z-z_0} = 2\pi i, \quad \int_L \frac{dz}{(z-z_0)^k} = 0 \qquad (k = 2, \ldots, m)$$

Случай в). Функция $f(z)$ имеет вид

$$f(z) = \sum_{k=0}^{\infty} c_k (z-z_0)^k + \sum_{k=1}^{\infty} \frac{c_{-k}}{(z-z_0)^k} = f_1(z) + f_2(z), \qquad (6.84)$$

где в ряду

$$f_2(z) = \sum_{k=1}^{\infty} \frac{c_{-k}}{(z-z_0)^k}$$

не равно нулю бесконечное число коэффициентов c_{-k}.

В этом случае говорят, что *функция $f(z)$ имеет в точке z_0 существенную особенность.*

Мы знаем, что

$$\lim_{z \to z_0} f_1(z) = c_0.$$

Однако $f_2(z)$ при указанных условиях не стремится при $z \to z_0$ к какому-нибудь пределу — конечному или бесконечному.

Заметим, что рассуждения, которые приводились при доказательстве равенства (6.83) в случае полюса, в данном случае неприменимы, потому что для бесконечных сумм операция почленного предельного перехода не всегда законна.

Пример 1. Функция $e^{-1/z^2} = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} z^{-2n}$ имеет существенную особенность в точке $z = 0$. Эта функция не имеет предела в точке $z = 0$.

В самом деле, при $z = x$ (x—действительное) $\exp(-1/x^2) \to 0$, когда $x \to 0$. Однако если $z = \frac{i}{n}$, то $\exp(-1/z^2) = \exp(n^2) \to +\infty$ при $n \to \infty$. Значит, предел в точке $z = 0$ у функции $\exp(-1/z^2)$ не существует.

Для любого ориентированного против часовой стрелки контура L, принадлежащего к кругу $|z - z_0| < R$ и содержащего внутри точку z_0, так же как в случае полюса

$$\int_L f(z)\,dz = 2\pi i c_{-1}. \tag{6.87}$$

Дело в том, что интеграл по L в данном случае можно заменить на интеграл по какой-либо ориентированной против часовой стрелки окружности γ с центром в точке z_0, принадлежащей к кругу $|z - z_0| < R$. Но на γ ряды (6.86) равномерно сходятся и, следовательно, их можно почленно проинтегрировать по γ. Однако, как мы знаем,

$$\int_\gamma \frac{dz}{(z-z_0)^k} = 0 \quad (k \neq 1) \quad \text{и} \quad \int_\gamma \frac{dz}{z-z_0} = 2\pi i,$$

откуда следует равенство (6.87).

Сделаем теперь определение: *пусть z_0 есть изолированная точка функции $f(z)$, т.е. пусть функция $f(z)$ аналитическая в некотором круге*

$$|z - z_0| < R,$$

из которого выколота точка z_0. Вычетом функции f в точке z_0 называется интеграл

$$\frac{1}{2\pi i}\int_L f(z)\,dz = \operatorname*{Выч}_{z=z_0} f(z), \tag{6.88}$$

где L—контур в круге $|z - z_0| < R$, ориентированный против часовой стрелки и содержащей в себе точку z_0.

На основании сказанного выше (см. случаи а), б), в)), *если*

$$f(z) = \sum_{k=-\infty}^{\infty} c_k (z - z_0)^k \qquad (0 < |z - z_0| < R)$$

есть ряд Лорана f в точке z_0, то

$$\operatorname*{Выч}_{z=z_0} f(z) = c_{-1}. \tag{6.89}$$

Поэтому, если известно разложение функции в ряд Лорана по степеням $z - z_0$, то вычет в точке z_0 легко находится.

В частности, если z_0 — устранимая особая точка, то

$$\operatorname*{Выч}_{z=z_0} f(z) = 0.$$

Иногда разложить функцию $f(z)$ в ряд Лорана трудно, и поэтому приходится искать другие способы вычислении вычета, не разлагая функцию в ряд Лорана.

Пусть $z = z_0$ — полюс порядка $m \geq 1$. Тогда

$$f(z) = \sum_{k=0}^{\infty} c_k (z - z_0)^k + \sum_{k=1}^{m} c_{-k} (z - z_0)^{-k} \qquad (c_{-m} \neq 0). \tag{6.90}$$

Умножая левую и правую части (6.90) на $(z - z_0)^m$, имеем

$$(z - z_0)^m f(z) = c_{-m} + c_{-m+1}(z - z_0) + \ldots$$
$$\ldots + c_{-1}(z - z_0)^{m-1} + \sum_{k=0}^{\infty} c_k (z - z_0)^{k+m}. \tag{6.91}$$

Если продифференцировать равенство (6.91) $(m-1)$ раз, то свободный член справа будет равен $(m-1)!\,c_{-1}$ и, следовательно,

$$\lim_{z \to z_0} \frac{d^{m-1}[(z - z_0)^m f(z)]}{dz^{m-1}} = (m-1)!\, c_{-1},$$

откуда

$$\operatorname*{Выч}_{z=z_0} f(z) = c_{-1} = \frac{1}{(m-1)!} \lim_{z \to z_0} \frac{d^{m-1}[(z - z_0)^m f(z)]}{dz^{m-1}}. \tag{6.92}$$

Если функция

$$f(z) = \frac{\varphi(z)}{\psi(z)},$$

где $\varphi(z_0) \neq 0$, а $\psi(z)$ имеет простой нуль при $z = z_0$ $(\psi(z_0) = 0,\ \psi'(z_0) \neq 0)$, то $z = z_0$ является простым полюсом $f(z)$. На основании формулы (6.92) (при $m = 1$) имеем

$$\operatorname*{Выч}_{z=z_0} f(z) = \operatorname*{Выч}_{z=z_0} \frac{\varphi(z)}{\psi(z)} = \lim_{z \to z_0} (z - z_0) \frac{\varphi(z)}{\psi(z)} =$$
$$= \lim_{z \to z_0} \frac{\varphi(z)}{\dfrac{\psi(z) - \psi(z_0)}{z - z_0}} = \frac{\varphi(z_0)}{\psi'(z_0)}.$$

Таким образом, в данном случае

$$\operatorname*{Выч}_{z=z_0} f(z) = c_{-1} = \frac{\varphi(z_0)}{\psi'(z_0)}. \tag{6.93}$$

В случае, когда $z = z_0$ — существенно особая точка, у нас имеется только один способ вычисления вычета — разложение функции $f(z)$ в ряд Лорана.

Пример 2. Найти вычет функции $f(z)=\frac{\sin^2 z}{\cos z}$ в точке

$$z=\frac{\pi}{2}.$$

В данном случае

$$f(z)=\frac{\varphi(z)}{\psi(z)}, \text{ где } \varphi(z)=\sin^2 z, \psi(z)=\cos z.$$

Точка z = π/2 являетсяпростым полюсом функции *f*(z), так как

$$\varphi(\pi/2)=1\neq 0, \quad \psi(\pi/2)=0, \quad \psi'(z)=-\sin z,$$

$\psi'(\pi/2)=-1\neq 0$. Значит, по формуле (6.93) получаем

$$\operatorname*{Выч}_{z=\pi/2} f(z)=\frac{\varphi(\pi/2)}{\psi'(\pi/2)}=\frac{1}{-1}=-1.$$

Пример 3. Найти вычет функции exp(1/z) в точке *z*=0.

Имеем

$$\exp\left(\frac{1}{z}\right)=1+\frac{1}{z}+\frac{1}{2!\,z^2}+\ldots$$

Таким образом, точка *z* = 0 является существенно особой и

$$\operatorname*{Выч}_{z=0}\exp\left(\frac{1}{z}\right)=c_{-1}=1.$$

Пример 4. Найти вычет функции

$$f(z)=\frac{1}{(z-2)^2(z-3)}$$

относительно точки z = 2.

Данная точка является полюсом второго порядка, поэтому по формуле (6.92) имеем

$$\operatorname*{Выч}_{z=2} f(z)=\lim_{z\to 2}\frac{d}{dz}\left[(z-2)^2 f(z)\right]=\lim_{z\to 2}\frac{d}{dz}\frac{1}{z-3}=$$

$$=\lim_{z\to 2}\frac{-1}{(z-3)^2}=-1.$$

6.12. Классификация особых точек на бесконечности

Предположим теперь, что в теореме 1 п. 6.10 $z_0 = 0$ и $R = \infty$, а r — любое неотрицательное число ($0 \le r < \infty$). Тогда теорема 1 гласит: если функция *f*(z) аналитическая для всех комплексных чисел *z,* удовлетворяющих неравенству

$$|z|>r, \qquad (6.94)$$

то ее можно разложить в ряд Лорана по степеням *z:*

$$f(z)=\sum_{k=-\infty}^{\infty} c_k z^k = F_1(z)+F_2(z) \qquad (|z|>r), \quad (6.95)$$

сходящийся для всех z с $|z|>r$. Здесь

$$F_1(z)=\sum_{k=0}^{\infty}\frac{c_{-k}}{z^k}, \quad F_2(z)=\sum_{k=1}^{\infty} c_k z^k. \qquad (6.96)$$

Множество (6.94) называют *внешностью круга* $|z|\le r$. Удобно считать, что это множество есть *окрестность бесконечно удаленной точки* (точки ∞).

Таким образом, мы формально добавляем к множеству комплексных точек (чисел) еще абстрактную бесконечно удаленную точку ($z=\infty$).

Функция $f(z)$ аналитична в окрестности точки $z=\infty$, исключая саму точку ∞, которую естественно в данном случае называть *изолированной особой точкой функции f*.

В зависимости от поведения функции $f(z)$ в окрестности точки $z=\infty$ естественно ввести следующую классификацию:

а) *Особенность в точке* $z=\infty$ *устранимая,* если

$$c_k=0 \qquad (k=1,2,\ldots),$$

т. е. если

$$f(z)=F_1(z) \qquad (|z|>r).$$

В этом случае

$$\lim_{z\to\infty} f(z)=\lim_{z\to\infty} F_1(z)=c_0.$$

Очевидно также

$$\frac{1}{2\pi i}\int_{L_-} f(z)\,dz=-c_{-1},$$

где L_- — произвольный контур, ориентированный по часовой стрелке, содержащий внутри себя окружность $|z|=r$ (рис. 6.19).

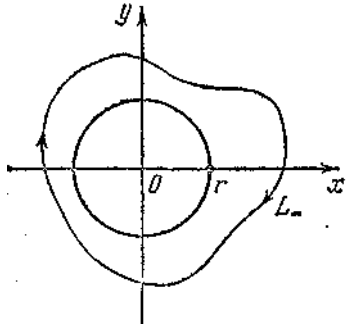


Рис. 6.19

При известном воображении можно считать, что точка ∞ находится внутри контура L_-,— если двигаться по контуру L_- по часовой стрелке, то точка ∞ остается слева.

б) *Точка* $z = \infty$ *есть по люс порядка m,* если

$$f(z) = F_1(z) + \sum_{k=1}^{m} c_k z^k$$

$$(c_m \neq 0).$$

В этом случае, очевидно, $\lim\limits_{z \to \infty} f(z) = \infty$. Далее

$$\int_{L_-} f(z)\,dz = \sum_{k=0}^{\infty} c_{-k} \int_{L_-} \frac{dz}{z^k} + \sum_{k=1}^{m} c_k \int_{L_-} z^k\,dz =$$

$$= -c_{-1} \int_{L} \frac{dz}{z} = -2\pi i c_{-1},$$

потому что

$$\int_{L_-} z^k\,dz = -\int_{L} z^k\,dz = 0 \qquad (k \neq -1).$$

в) *Точка* $z = \infty$ *есть существенно особая точка,* если

$$f(z) = F_1(z) + \sum_{k=1}^{\infty} c_k z^k \tag{6.97}$$

и имеется бесконечное множество чисел c_k, не равных нулю. Функция $F_i(z)$ стремится к конечному пределу при $z \to \infty$, в то время как функция $\sum\limits_{k=1}^{\infty} c_k z^k$, на основании теоремы Сохоцкого, не стремится ни к какому пределу при $z \to \infty$. Поэтому и функция $f(z)$ не стремится к пределу при $z \to \infty$. Далее

$$\int_{L_-} f(z)\,dz = \sum_{k=-\infty}^{\infty} c_k \int_{L_-} z^k\,dz = -2\pi i\, c_{-1}.$$

Почленное интегрирование здесь законно, потому что, как мы знаем, интегралы $\int\limits_{L_-}$ можно заменить на интегралы $\int\limits_{C}$ по окружности радиуса $\rho > r$, на которой ряд (6.97) равномерно сходится.

Введем определение.

Вычетом функции $f(z)$ *в бесконечно удаленной точке называется*

$$\frac{1}{2\pi i} \int_{L_-} f(z)\,dz = \operatorname*{Выч}_{z=\infty} f(z),$$

где L_ есть произвольный, замкнутый, контур, ориентированный по часовой стрелке, принадлежащий к множеству $|z|>r$ (где функция $f(z)$ аналитична!). В данном случае говорят, что при движении по контуру по часовой стрелке «точка ∞ остается слева».

На основании сказанного (см. а), б), в)), *если*

$$f(z)=\sum_{k=-\infty}^{\infty} c_k z^k \qquad (|z|>r),$$

ряд Лорана функции f во внешности окружности $|z|=r$, *то*

$$\operatorname*{Выч}_{z=\infty} f(z) = -c_{-1}.$$

Если $z = \infty$ — устранимая особая точка, то в ряде Лорана функции $f(z)$ отсутствуют положительные степени z, а z^{-1} может присутствовать, поэтому

$$\operatorname*{Выч}_{z=\infty} f(z)$$

в этом случае не обязательно равен нулю.

Пример. Функция

$$f(z)=\frac{z}{z^2+1}=\frac{1}{z\left(1+\frac{1}{z^2}\right)}=$$

$$=\frac{1}{z}\sum_{n=0}^{\infty}(-1)^n z^{-2n}=\sum_{n=0}^{\infty}(-1)^n z^{-1-2n} \qquad (|z|>1)$$

имеет устранимую особенность в точке $z = \infty$ и $c_{-1} = 1$, значит,

$$\operatorname*{Выч}_{z=\infty} f(z) = -1.$$

6.13. Теорема о вычетах

Теорема 1. *Пусть функция f (z) аналитическая на всей плоскости z, за исключением конечного числа точек* z_1, ,.., z_N. *Тогда имеет место равенство*

$$\sum_{k=1}^{N} \operatorname*{Выч}_{z=z_k} f(z) + \operatorname*{Выч}_{z=\infty} f(z) = 0. \qquad (6.98)$$

Доказательство. Построим окружности $\gamma_1, \gamma_2, \ldots, \gamma_N$, ориентированные по часовой стрелке, с центрами соответственно $z_1, z_2, ..., z_N$, настолько малого радиуса, чтобы они не пересекались.

Кроме того, построим окружность Г, ориентированную против часовой стрелки, с центром в нулевой точке, настолько большого

радиуса, чтобы окружности $\gamma_1^-, \ldots, \gamma_N^-$ оказались внутри Г (рис. 6.20).

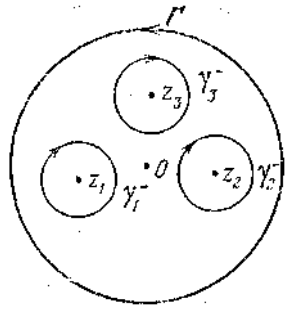


Рис. 6.20

Сложный контур $L=\gamma_1^- + \gamma_2^- + \ldots + \gamma_N^- + \Gamma$ ограничивает область Ω, внутри которой функция $f(z)$ аналитическая. Она аналитическая также на L. При этом при обходе по L область Ω остается слева.

Но тогда на основании теоремы Коши для сложного контура

$$\int_{\gamma_1^-} f(z)\,dz + \ldots + \int_{\gamma_N^-} f(z)\,dz + \int_{\Gamma} f(z)\,dz = 0 \qquad (6.99).$$

или, если помножить левую часть этого равенства на $-1/2\pi i$, то получим

$$\frac{1}{2\pi i}\int_{\gamma_1} f(z)\,dz + \ldots + \frac{1}{2\pi i}\int_{\gamma_N} f(z)\,dz + \frac{1}{2\pi i}\int_{\Gamma_-} f(z)\,dz = 0,$$

т.е.

$$\operatorname{Выч}_{z=z_1} f(z) + \ldots + \operatorname{Выч}_{z=z_N} f(z) + \operatorname{Выч}_{z=\infty} f(z) = 0.$$

Надо учесть, что внутри каждого из контуров γ_k имеется только одна особая точка z_k, а вне Г—только одна особая точка $z=\infty$. Теорема доказана.

Применение этой теоремы сводится к следующему. Если затруднительно вычислить один из интегралов, входящих в (6.99), то можно попытаться вычислить оставшиеся интегралы, входящие в (6.99), и получить искомый интеграл из (6.99).

Само вычисление этих интегралов сводится к разложению функции $f(z)$ в ряд Лорана в окрестности соответствующих особых точек. В сущности и эти разложения не надо знать полностью. Достаточно только знать члены вида $c_{-1}/(z-z_k)$ этих разложений, чтобы прийти к цели.

6.14. Вычисление интегралов при помощи вычетов

Пусть функция $f(z)$ аналитична в верхней полуплоскости, включая действительную ось, за исключением конечного числа особых точек $a_1, a_2, ..., a_N$, лежащих в верхней полуплоскости. При этих условиях мы рассмотрим способы вычисления интегралов

$$\int_{-\infty}^{\infty} f(x)\,dx, \quad \int_{-\infty}^{\infty} f(x)\,e^{ix}\,dx.$$

Теорема 1. *Пусть функция* $f(z)$ *удовлетворяет перечисленным выше условиям и, кроме того,* $|f(z)| \leqslant M/|z|^m$ *при* $|z| \geq R$, *где* $m \geq 2$ *и* R *— достаточно большое число. Тогда*

$$\int_{-\infty}^{\infty} f(x)\,dx = 2\pi i \sum_{j=1}^{N} \operatorname{Выч}_{z=a_j} f(z). \qquad (6.100)$$

Доказательство. Опишем полуокружность L (ориентированную против часовой стрелки) радиуса R с центром в точке O так, чтобы все особые точки функции $f(z)$ попали внутрь L (рис. 6.21).

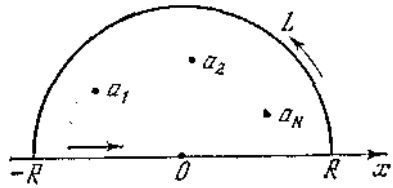


Рис. 6.21.

В силу теоремы 1 п. 6.13

$$\int_{-R}^{R} f(x)\,dx + \int_{L} f(z)\,dz = 2\pi i \sum_{j=1}^{N} \operatorname{Выч}_{z=a_j} f(z). \qquad (6.101)$$

Так как $|f(z)| \leq M/|z|^m$ при $|z| \geq R$, то

$$\left|\int_{L} f(z)\,dz\right| \leqslant \frac{M}{R^m}\pi R = \frac{\pi M}{R^{m-1}} \to 0, \quad R \to \infty \quad (m-1 \geqslant 1).$$

Переходя к пределу в равенстве (6.101) при $R \to \infty$, получим (6.100).

Пример 1. Вычислить интеграл $\int_{-\infty}^{\infty} \frac{dx}{1+x^4}$.

Функция $f(z) = \frac{1}{1+z^4}$ аналитична в верхней полуплоскости, за исключением точек

$$a_1 = e^{\pi i/4} = \frac{\sqrt{2}}{2}(1+i), \quad a_2 = e^{3\pi i/4} = \frac{\sqrt{2}}{2}(i-1),$$

в которых она имеет простые полюсы. Кроме того, $|f(z)| \leqslant 1/|z|^4 (m = 4 > 2)$. Найдем вычеты функции $f(z)$ в точках a_1, a_2. По формуле (6.936)

$$\operatorname*{Выч}_{z=a_j} f(z) = \frac{1}{\psi'(a_j)} \quad (j = 1, 2),$$

где $\psi(z) = 1 + z^4$. Имеем $\psi'(z) = 4z^3$, $\psi'(a_1) = 4e^{3\pi i/4} = -4e^{-i\pi/4} \neq 0$, $\psi'(a_2) = 4e^{3\pi i/4} = 4e^{i\pi/4} \neq 0$. Отсюда

$$\operatorname*{Выч}_{z=a_1} f(z) = -\frac{1}{4}e^{i\pi/4}, \quad \operatorname*{Выч}_{z=a_2} f(z) = \frac{1}{4}e^{-i\pi/4}.$$

По формуле (6.100) получаем

$$\int_{-\infty}^{\infty} \frac{dx}{1+x^4} = \frac{2\pi i}{4}\left(-e^{i\pi/4} + e^{-i\pi/4}\right) = \pi \frac{e^{i\pi/4} - e^{-i\pi/4}}{2i} =$$

$$= \pi \sin\frac{\pi}{4} = \pi\frac{\sqrt{2}}{2}.$$

Теорема 2. *Пусть функция f(z) удовлетворяет условиям, отмеченным в начале параграфа и*

$$\lim_{z\to\infty} f(z) = 0$$

равномерно относительно arg z = φ. *Тогда*

$$\int_{-\infty}^{\infty} f(x)\, e^{ix}\, dx = 2\pi i \sum_{j=1}^{N} \operatorname*{Выч}_{z=a_j} f(z)\, e^{iz}. \tag{6.102}$$

Доказательство. Так же как при доказательстве теоремы 1, имеем

$$\int_{-R}^{R} f(x)\, e^{ix}\, dx + \int_{L} f(z)\, e^{iz}\, dz = 2\pi i \sum_{j=1}^{N} \operatorname*{Выч}_{z=a_j} f(z)\, e^{iz} \tag{6.103}$$

(функция $f(z)e^{iz}$ имеет те же особенности, что и $f(z)$).

Нам нужно доказать, что при $R \to \infty$ интеграл

$$\int_{L} f(z)\, e^{iz}\, dz$$

стремится к нулю. Имеем

$$\Lambda = \left| \int_L f(z)\, e^{iz}\, dz \right| = \left| \int_0^{\pi} f(Re^{i\varphi})\, e^{-R\sin\varphi} e^{iR\cos\varphi} iRe^{i\varphi}\, d\varphi \right| \leqslant$$

$$\leqslant \int_0^{\pi} |f(Re^{i\varphi})|\, e^{-R\sin\varphi} R\, d\varphi.$$

В силу условия теоремы $|f(Re^{i\varphi})| \leqslant \varepsilon$; при $R > N_1$ для всех φ $(0 \leqslant \varphi \leqslant \pi)$ и достаточно большого N_1. Поэтому $(\sin\varphi \geqslant 2\varphi/\pi$ при $0 \leqslant \varphi \leqslant \pi/2)$

$$\Lambda \leqslant \varepsilon \int_0^{\pi} Re^{-R\sin\varphi}\, d\varphi =$$

$$= 2\varepsilon \int_0^{\pi/2} Re^{-R\sin\varphi}\, d\varphi \leqslant 2\varepsilon \int_0^{\pi/2} Re^{-2R\varphi/\pi}\, d\varphi =$$

$$= \left(\frac{2R}{\pi}\varphi = t \right) = \varepsilon\pi \int_0^{R} e^{-t}\, dt = \pi\varepsilon\,(1 - e^{-R}) < \pi\varepsilon \quad (R > N_1).$$

Переходя к пределу в (6.103), при $R \to \infty$ получаем (6.102).

Если функция $f(z)$ имеет особенности на действительной оси, то специальным построением контура интегрирования можно вычислить соответствующие интегралы, если они существуют.

Пример 2. Пусть $f(z) = 1/z$. Эта функция имеет простой полюс на действительной оси в точке $z = 0$. Далее, $\lim\limits_{z \to \infty} f(z) = 0$ равномерно относительно $\arg z = \varphi$.

Построим контур интегрирования, как на рис. 6.22.

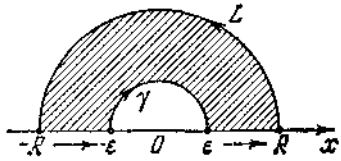


Рис. 6.22

Обход контура осуществляется по стрелкам, указанным на этом рисунке. В заштрихованной части функция e^{iz}/z аналитическая при любом R и любом ε, поэтому по теореме Коши (полуокружность γ ориентирована против часовой стрелки)

$$\left(\int\limits_{-R}^{-\varepsilon}+\int\limits_{\varepsilon}^{R}\right)\frac{e^{ix}}{x}\,dx+\int\limits_{L}\frac{e^{iz}}{z}\,dz-\int\limits_{\gamma}\frac{e^{iz}}{z}\,dz=0. \qquad (6.104)$$

Как и выше, легко показать, что

$$\lim_{R\to\infty}\int\limits_{L} z^{-1}e^{iz}\,dz=0.$$

Далее

$$\lim_{\varepsilon\to 0}\int\limits_{\gamma} e^{iz}\,\frac{dz}{z}=\lim_{\varepsilon\to 0}\int\limits_{0}^{\pi} e^{i\varepsilon(\cos\varphi+i\sin\varphi)}\,\frac{i\varepsilon e^{i\varphi}\,d\varphi}{\varepsilon e^{i\varphi}}=$$

$$=i\lim_{\varepsilon\to 0}\int\limits_{0}^{\pi} e^{i\varepsilon(\cos\varphi+i\sin\varphi)}\,d\varphi=i\int\limits_{0}^{\pi}d\varphi=\pi i.$$

Таким образом, равенство (6.104) в пределе, при $R\to\infty$ и $\varepsilon\to 0$, принимает вид

$$\pi i=\lim_{\substack{\varepsilon\to 0\\ R\to\infty}}\left(\int\limits_{-R}^{-\varepsilon}+\int\limits_{\varepsilon}^{R}\right)\frac{e^{ix}}{x}\,dx=$$

$$=\lim_{\substack{\varepsilon\to 0\\ R\to\infty}}2i\int\limits_{\varepsilon}^{R}\frac{e^{ix}-e^{-ix}}{2xi}\,dx=2i\int\limits_{0}^{\infty}\frac{\sin x}{x}\,dx,$$

т. е.

$$\int\limits_{0}^{\infty}\frac{\sin x}{x}\,dx=\frac{\pi}{2}.$$

Так как функция $\frac{\sin x}{x}$ четная, то

$$\int\limits_{-\infty}^{\infty}\frac{\sin x}{x}\,dx=\pi.$$

Замечание. Если под знаком интеграла есть сомножитель $\sin x$ или $\cos x$, то часто удобно рассматривать интеграл от функции, где $\sin x$ или $\cos x$ заменены на e^{iz}. Это объясняется тем, что $|\sin z|$ и $|\cos z|$ неограниченно возрастают при $z\to\infty$, а $|e^{iz}|=e^{-y}\to 0$ при $y\to\infty$ ($y>0$). Поэтому поведение функции

$$f(z)\begin{Bmatrix}\sin z\\ \cos z\end{Bmatrix}$$

будет другое, чем у функции $f(z)\,e^{iz}$. Затем, получив значение интеграла $\int\limits_{-\infty}^{\infty} f(x)\,e^{ix}\,dx$, выделяя действительнуюи мнимую части, мы найдем $\int\limits_{-\infty}^{\infty} f(x)\cos x\,dx$ и

$$\int\limits_{-\infty}^{\infty} f(x)\sin x\,dx.$$

Пример 3. Вычислить интеграл

$$\int\limits_{0}^{\infty}\frac{\cos\alpha x\,dx}{a^2+x^2}\quad(\alpha,\ a>0).$$

Рассмотрим функцию $e^{i\alpha z}/(a^2+z^2)$. Эта функция аналитична в верхней полуплоскости, кроме точки $z=ai$. Функция $f(z)=1/(a^2+z^2)\to 0$ при $z\to\infty$ равномерно относительно $\arg z=\varphi$. Поэтому по теореме 2

$$\int\limits_{-\infty}^{\infty}\frac{e^{i\alpha x}}{a^2+x^2}\,dx=2\pi i\operatorname*{Выч}_{z=ai}\frac{e^{i\alpha z}}{a^2+z^2}=2\pi i\,\frac{e^{i\alpha ai}}{2ai}=\frac{\pi}{a}\,e^{-\alpha a}.$$

Выделяя действительную часть, получим

$$\int\limits_{-\infty}^{\infty}\frac{\cos\alpha x}{a^2+x^2}\,dx=\frac{\pi}{a}e^{-\alpha a},\quad\int\limits_{0}^{\infty}\frac{\cos\alpha x}{a^2+x^2}\,dx=\frac{\pi}{2a}e^{-\alpha a}.$$

Пример 4. Вычислить интеграл

$$I=\int\limits_{0}^{\infty}\frac{\sin^2 x}{1+x^2}\,dx.$$

Имеем

$$I=\int\limits_{0}^{\infty}\frac{1-\cos 2x}{2(1+x^2)}\,dx=\frac{1}{2}\int\limits_{0}^{\infty}\frac{dx}{1+x^2}-\frac{1}{2}\int\limits_{0}^{\infty}\frac{\cos 2x}{1+x^2}\,dx=$$
$$=\frac{1}{2}\operatorname{arctg}x\Big|_0^{\infty}-\frac{1}{2}\,\frac{\pi}{2}\,e^{-2}=\frac{\pi}{4}-\frac{\pi}{4}e^{-2}=\frac{\pi}{4}(1-e^{-2}).$$

Итак,

$$\int\limits_{0}^{\infty}\frac{\sin^2 x\,dx}{1+x^2}=\frac{\pi}{4}(1-e^{-2}).$$

Пример 5.

$$\int_0^\infty \frac{\cos^2 x\,dx}{1+x^2} = \int_0^\infty \frac{1-\sin^2 x}{1+x^2}\,dx =$$

$$= \int_0^\infty \frac{dx}{1+x^2} - \int_0^\infty \frac{\sin^2 x}{1+x^2}\,dx = \frac{\pi}{2} - \frac{\pi}{4}(1-e^{-2}) = \frac{\pi}{4}(1+e^{-2}).$$

Пример 6. Вычислить интегралы Френеля

$$\int_{-\infty}^{\infty} \cos x^2\,dx, \quad \int_{-\infty}^{\infty} \sin x^2\,dx.$$

Рассмотрим функцию e^{iz^2}. Эта функция в заштрихованной области (рис. 6.23) аналитическая, поэтому по теореме Коши

$$\int_0^R e^{ix^2}\,dx + \int_L e^{iz^2}\,dz + \int_\gamma e^{iz^2}\,dz = 0,$$

где L — часть окружности $|z| = R$, γ — отрезок прямой $z = \rho e^{i\pi/4}$, $0 \leqslant \rho \leqslant R$ (ориентированные по стрелкам).

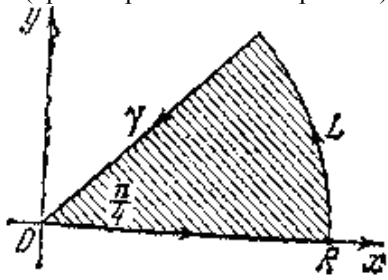


Рис. 6.23

Далее

$$\int_\gamma e^{iz^2}\,dz = \int_R^0 e^{i\rho^2 i} e^{i\pi/4}\,d\rho = -e^{i\pi/4}\int_0^R e^{-\rho^2}\,d\rho \to -e^{i\pi/4}\int_0^\infty e^{-\rho^2}\,d\rho;$$

$$\left|\int_L e^{iz^2}\,dz\right| = \left|\int_0^{\pi/4} e^{i\left(Re^{i\varphi}\right)^2} Rie^{i\varphi}\,d\varphi\right| =$$

$$= \left|\int_0^{\pi/4} Rie^{i\varphi} e^{iR^2\cos 2\varphi} e^{-R^2\sin 2\varphi}\,d\varphi\right| \leqslant \int_0^{\pi/4} Re^{-R^2\sin 2\varphi}\,d\varphi \leqslant$$

$$\leqslant R\int_0^{\pi/4} e^{-R^2\frac{4\varphi}{\pi}}\,d\varphi \leqslant \frac{c}{R}\int_0^\infty e^{-t}\,dt \to 0, \quad R \to \infty.$$

Итак, в пределе при $R\to\infty$ получаем

$$\int_0^\infty e^{ix^2}dx = e^{i\pi/4}\int_0^\infty e^{-\rho^2}d\rho = e^{\pi i/4}\sqrt{\frac{\pi}{4}}.$$

Выделяя действительную и мнимую части, получаем

$$\int_0^\infty \cos x^2\,dx = \frac{\sqrt{2}}{2}\sqrt{\frac{\pi}{4}}, \quad \int_0^\infty \sin x^2\,dx = \frac{\sqrt{2}}{2}\sqrt{\frac{\pi}{4}},$$

т. е.

$$\int_0^\infty \cos x^2\,dx = \int_0^\infty \sin x^2\,dx = \frac{1}{2}\sqrt{\frac{\pi}{2}} = \sqrt{\frac{\pi}{8}}.$$

6.15. Линейная функция. Дробно-линейная функции

Целая линейная функция. Рассмотрим три функции

$w = z+c,$ (6.105)

$w = rz,$ (6.106)

$w = e^{i\theta}z,$ (6.107)

где c—постоянное комплексное число, $z>0$, θ—произвольное действительное число.

Все три функции (6.105), (6.106), (6.107) отображают плоскость z на всю плоскость w.

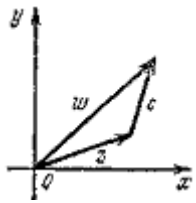


Рис. 6.24.

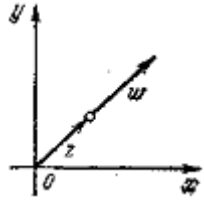


Рис. 6.25.

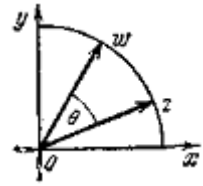


Рис. 6.26.

Функция (6.105) осуществляет сдвиг плоскости z на вектор c (рис. 6.24).

Функция (6.106) ($z > 0$!) осуществляет растяжение (при $z > 1$) и сжатие (при $z < 1$) плоскости z в r раз: $|w= r|z|$, Arg w = Arg z. На рис. 6.25 показан случай $r> 1$.

Функция (6.107) осуществляет поворот плоскости z вокруг нулевой точки на угол θ (рис. 6.26).

Функции (6.105), (6.106), (6.107) имеют соответственно производные

$$w' = 1, \quad w' = r, \quad w' = e^{i\theta},$$

не равные нулю и потому они осуществляют конформные отображения.

Все эти три функции являются частными случаями более общей *целой линейной функции*

$$w = az + b \quad (a \neq 0), \tag{6.108}$$

где a и b — постоянные комплексные числа.

Осуществляемое ею отображение можно записать в виде

$$w = a\,(z + c) = re^{i\theta}\,(z + c),$$

где $c = b/a$, $a = re^{i\theta}$.

Отсюда следует, что она сводится к (6.105), (6.106), (6.107):

$$w = ru, \quad u = e^{i\theta}v, \quad v = z + c.$$

Иначе говоря, преобразование плоскости z, осуществляемое функцией (6.108), сводится к переносу (на вектор c), затем к повороту плоскости (на угол θ) и затем к растяжению или сжатию плоскости в z раз.

Функция $w = \frac{1}{z}$. Полагая $z = re^{i\varphi}$, $w = \rho e^{i\theta}$, имеем

$$\rho = \frac{1}{r}, \quad \theta = -\varphi, \tag{6.109}$$

где второе равенство надо понимать с точностью до $2k\pi$ (k=0, ±1, ±2, ...).

Отсюда видно, что окружность | z | = 1 переходит в себя, точнее каждая ее точка переходит в точку, симметричную относительно действительной оси.

Отметим, что если окружность |z| = 1 проходится в направлении против часовой стрелки, то отображенная окружность |w|=1 проходится по часовой стрелке.

Преобразование (6.109) удобно разбить на два преобразования:

$$r' = \frac{1}{r}, \quad \varphi' = \varphi; \tag{6.110}$$

$$\rho = r', \quad \theta = -\varphi'. \tag{6.111}$$

Преобразование (6.110) называется *инверсией* относительно единичной окружности.

При инверсии относительно единичной окружности точки z и z', лежащие на луче, составляющем угол φ с осью x, переходят в точки, лежащие на этом же луче, и притом так, что

$$rr' = 1.$$

Построение точки $z' = r'e^{i\varphi}$ по известной точке $z = ze^{i\varphi}$

видно из рис. 6.27, где рассмотрен случай, когда z лежит вне окружности $|z| = 1$.

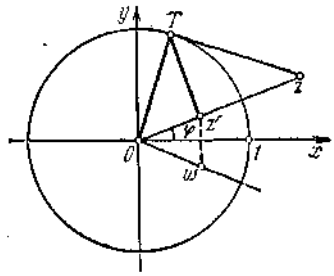


Рис. 6.27

Из точки z проводим касательную к окружности $|z|=1$, T — точка касания, $Tz' \perp Oz$. Из подобия треугольников ($\triangle OTz'$ и $\triangle OTz$)

$$\frac{OT}{r'} = \frac{r}{OT}, \quad OT = 1, \quad rr' = 1$$

Если точка z находится внутри окружности $|z|=1$, то восстанавливаем из нее перпендикуляр к Oz до пересечения с окружностью в точке T. Через последнюю проводим касательную к окружности до пересечения с лучом Oz. Точка пересечения и будет точкой z'.

Точки z и z' называют *взаимно симметричными относительно окружности* $|z| = 1$.

Отображая теперь (по (6.111)) точку z' зеркально относительно действительной оси, мы получаем точку

$$w = r'e^{-i\varphi} = \frac{1}{re^{i\varphi}} = \frac{1}{z}.$$

Из формулы $w=1/z$ видно, что при $z \to 0$ точка w имеет неограниченно возрастающий модуль, поэтому удобно считать, что при помощи этой формулы точке $z = 0$ соответствует «бесконечно удаленная точка», которую обозначают символом $w = \infty$.

Итак, функция $w=1/z$ отображает плоскость z на плоскость w при помощи преобразования инверсии относительно окружности $|z| = 1$ и зеркального отображения относительно оси x. При этом точка $z = 0$ переходит в точку $w = \infty$, а точка $z = \infty$ — в точку $w = 0$.

Далее $w' = \frac{-1}{z^2} \neq 0$ при любых z с $|z|>0$, поэтому отображение с помощью функции

$$w = \frac{1}{z} \quad (|z| > 0)$$

конформно.

Дробио-линейная функция

$$w = \frac{az+b}{cz+d}. \qquad (6.112)$$

Будем считать, что $ad—bc\neq 0$. Очевидно, что точка $z = — d/c$ $(c\neq 0)$ переходит в точку $w = \infty$.

Функцию w, выделяя ее целую часть, можно представить в виде

$$w = \frac{a}{c} + \frac{bc-ad}{c(cz+d)}, \qquad (6.113)$$

откуда видно, что

$$w' = \frac{ad-bc}{(cz+d)^2} \neq 0 \quad (cz+d\neq 0),$$

т. е. отображение с помощью функции (6.112) конформно. Из равенства (6.113) видно, что данное отображение состоит из рассмотренных выше отображений:

$$z' = cz+d, \quad z'' = \frac{1}{z'}, \quad w = Az'' + B.$$

Если считать прямую линию за окружность бесконечного радиуса, то при преобразовании (6.113) окружность переходит в окружность *(круговое свойство).*

Из геометрических соображений ясно, что при параллельном переносе, растяжении и вращении окружность переходит в окружность. Больше того, внутренность отображаемой окружности переходит на внутренность отображенной окружности. Поэтому достаточно проверить круговое свойство для преобразования $w = \frac{1}{z}$. Уравнение окружности в плоскости xOy, как нам известно, имеет вид

$$A(x^2+y^2) + mx + ny + l = 0$$

или

$$Az\cdot\bar{z} + m\frac{z+\bar{z}}{2} + n\frac{z-\bar{z}}{2i} + l = 0 \qquad \left(z = x+iy,\ \bar{z} = x-iy\right)$$

или

$$Az\bar{z} + \bar{B}z + B\bar{z} + l = 0 \quad \left(B = \frac{m+in}{2}\right). \qquad (6.114)$$

В рассматриваемом случае $z = 1/w$, $\bar{z} = 1/\bar{w}$. Следовательно, уравнение (6.114) переходит в уравнение

$$A\frac{1}{w\bar{w}} + \frac{\bar{B}}{w} + \frac{B}{\bar{w}} + l = 0$$

или в уравнение

$$A + \bar{B}\bar{w} + Bw + lw\bar{w} = 0,$$

которое описывает некоторую окружность в плоскости w.

В частности, при $l = 0$ получаем прямую линию, т.е. окружность, проходящая через начало координат в плоскости z, переходит в прямую в плоскости w.

Отметим, что отображение с помощью функции (6.113) может переводить внутренность отображаемой окружности как на внутренность, так и на внешность отображенной окружности.

Функция (6.113) в принципе зависит от трех параметров, за которые можно взять отношение чисел a, b, c, d к одному из них (не равному 0). Поэтому, чтобы определить преобразование (6.113), надо задать три условия. Обычно задают три пары соответствующих точек:

$$w_k = \frac{a}{c} + \frac{bc - ad}{c(cz_k + d)} \qquad (k = 1, 2, 3).$$

Легко подсчитать, что

$$w - w_k = \frac{(ad - bc)(z - z_k)}{(cz + d)(cz_k + d)}, \qquad w_k - w_j = \frac{(ad - bc)(z_k - z_j)}{(cz_k + d)(cz_j + d)}.$$

Отсюда

$$\frac{w - w_1}{w - w_3} : \frac{w_3 - w_1}{w_3 - w_2} = \frac{z - z_1}{z - z_2} : \frac{z_3 - z_1}{z_3 - z_2}. \qquad (6.115)$$

Это и есть преобразование (6.113), переводящее точки z в w_k (k= 1,2,3).

Пусть заданы две окружности Г и Γ^1 соответственно в плоскостях z и w. Требуется найти дробно-линейное отображение, переводящее Г на Γ^1 и внутренность Г на внутренность Г'.

На Г и Г* соответственно зададим произвольные тройки точек $\{z_1, z_2, z_3\}$ и $\{w_1, w_2, w_3\}$, следующие в положительном направлении, т. е. против часовой стрелки. Тогда преобразование (6.115) и будет решением поставленной задачи.

В самом деле, оно отображает точки z_k соответственно в точки w_k (k=1, 2, 3) и, очевидно, окружность Г на Г' (в силу кругового свойства). Тот факт, что в данном случае внутренность Г переходит на внутренность Г' следует из конформности отображения, осуществляемого дробно-линейной функцией.

В данном случае окружности Г, Г' имеют положительную ориентацию (проходятся против часовой стрелки). В силу конформности отображения внутренняя нормаль к Г (например, в точке z_1) переходит в дугу окружности (перпендикулярной Г' в точке w_1), которая находится внутри Г', а это и обеспечивает отображение внутренности Г на внутренность Г'.

Если же нужно найти дробно-линейное преобразование, отображающее Г на Г- и внутренность Г на внешность Г', то в формуле (6.115) надо взять точки $\{z_1, z_2, z_3\}$ на Г, расположенные в

положительном направлении, а точки $\{w_1, w_2, w_3\}$ на Г'— в отрицательном направлении.

Эти выводы распространяются и на случай, когда либо Г, либо Г', либо и Г и Г' являются прямыми. Однако требует пояснения, что надо понимать под внутренностью прямой Г, когда на ней отмечены точки $\{z_1, z_2, z_3\}$.

В случае рис. 6.28 это есть верхняя полуплоскость, а в случае рис. 6.29 — нижняя полуплоскость.

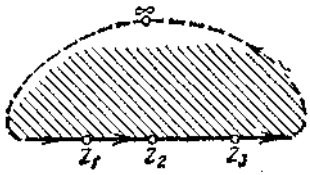


Рис. 6.28

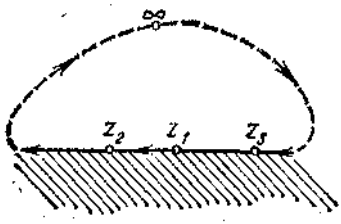


Рис. 6.29

Если прямую Г дополнить точкой ∞, то ее можно мыслить как непрерывную окружность (см. рис. 6.28, 6.29) бесконечного радиуса.

Будем двигаться по Г (возможно, через бесконечно удаленную точку) от z_1 к z_2 в таком направлении, чтобы дуга z_1z_2 не содержала в себе z_3. Этим направление обхода на Г определено и тогда внутренностью Г называется область, расположенная слева от Г при движении но этому направлению. На самом деле этой областью является верхняя или нижняя полуплоскость.

7. Решение уравнений

7.1. Общие сведения об уравнениях.

С решением уравнений приходится сталкиваться в самых разных задачах, как чисто математического, так и прикладного характера. Например, при исследовании функции *y=f(x)* для отыскания нулей функции, точек экстремума и абсцисс точек перегиба требуется решить соответственно уравнения $f(x)=0$, $f'(x)=0$, $f''(x)=0$. Далеко не всегда эти уравнения удается решить точно, и приходится отыскивать их корни приближенно. Однако, прежде чем переходить к

приближенным методам решения, мы изложим вкратце некоторые общие свойства уравнений, частично уже известных читателю.

I. Алгебраические уравнения и разложение многочленов на множители. Пусть дан многочлен

$$P(x) = a_0x^n + a_1x^{n-1} + \ldots + a_n$$

с действительными или комплексными коэффициентами, причем $a_0 \neq 0$. Тогда уравнение

$$a_0x^n + a_1x^{n-1} + \ldots + a_n = 0 \qquad (*)$$

называется *алгебраическим уравнением n-й степени.*

Основная теорема высшей алгебры утверждает, что *всякое алгебраическое уравнение степени n> 0 имеет хотя бы один корень, действительный или комплексный.* При этом, однако, теорема не указывает способов фактического отыскания корня; она говорит только об его существовании.

При делении многочлена *n*-й степени *P(x)* на двучлен $x - \alpha$ мы получаем в частном многочлен *Q(x)* степени *n*—1 и в остатке какое-то число *R.* Это можно записать в виде тождества

$$P(x) = (x - \alpha)Q(x) + R.$$

Подставляя в это тождество вместо *x* число α, получим, что $P(\alpha) = R$, т. е. что *остаток от деления многочлена на двучлен x*— α *равен значению этого многочлена при x* = α (теорема Безу)). Если α является корнем многочлена *P(x),* то $R = P(\alpha) = 0$ и

$$P(x) = (x - \alpha)Q(x).$$

Пусть x_1 является корнем уравнения (*); тогда

$$P(x) = (x - x_1)Q_1(x),$$

где $Q_1(x)$ — многочлен степени *n*—1. Легко заметить, что его старший коэффициент (коэффициент при x^{n-1}) равен старшему коэффициенту многочлена *P(x),* т. е. a_0.

Если степень $Q_1(x)$ не равна нулю, т. е. он не сводится к постоянной a_0, то к нему можно снова применить основную теорему. Пусть x_2 — корень $Q_1(x)$; тогда $Q_1(x) = (x - x_2)Q_2(x)$, где степень многочлена $Q_2(x)$ равна *n* —2, а старший коэффициент по-прежнему равен a_0. Выражение для *P(x)* примет вид

$$P(x) = (x - x_1)(x - x_2)Q_2(x).$$

Ясно, что x_2 тоже является корнем уравнения (*). Повторяя этот процесс *n* раз, мы придем к разложению многочлена *P(x)* на линейные множители

$$P(x) = a_0(x - x_1)(x - x_2)\ldots(x - x_n).$$

При таком последовательном получении корней $x_1, x_2, .. , x_n$ может оказаться, что некоторые из них будут совпадать. Если какой-то корень встретился k раз, то он называется *корнем кратности k.* Если k=1, т. е. корень встретился только один раз, то он называется *простым.*

Пусть корень x_1 имеет кратность k_1, корень x_2 — кратность k_2 и т. д. Тогда разложение многочлена $P(x)$ можно записать так:

$$P(x) = a_0 (x - x_1)^{k_1} (x - x_2)^{k_2} \dots (x - x_r)^{k_r}, \qquad (**)$$

где r — число различных корней, а сумма кратностей всех корней равна n

$$k_1 + k_2 + \dots + k_r = n.$$

Последнее равенство означает, что *алгебраическое уравнение n-й степени имеет n корней, если каждый корень считать столько раз, какова его кратность.*

Все сказанное до сих пор относилось к уравнениям (*) с любыми комплексными коэффициентами. Будем теперь считать, что все коэффициенты $a_0, a_1 ..., a_n$ - действительные числа. Тогда комплексные корни уравнения (*) (если они есть) попарно сопряжены. Действительно, если коэффициенты многочлена $P(x)$ - действительные числа, то, вычисляя его значения при комплексно сопряженных значениях x, мы снова получим комплексно сопряженные числа: $P(\alpha - i\beta) = \overline{P(\alpha + i\beta)}$. Это и значит, что если $P(\alpha + i\beta) = 0$, то и $P(\alpha - i\beta) = 0$. Легко проверить, что корни $\alpha + i\beta$ и $\alpha - i\beta$ имеют одинаковую кратность.

Отсюда, кстати, следует, что *любой многочлен нечетной степени имеет хотя бы один действительный корень.*

Объединим в разложении (**) множители, соответствующие комплексно сопряженным корням кратности l; перемножая их, получим

$$(x - \alpha - i\beta)^l (x - \alpha + i\beta)^l = (x^2 - 2\alpha x + \alpha^2 + \beta^2)^l = (x^2 + px + q)^l,$$

где

$$p = -2\alpha \text{ и } q = \alpha^2 + \beta^2.$$

Дискриминант квадратного трехчлена $p^2 - 4q < 0$, так как его корни комплексные.

Проделав указанное преобразование со всеми множителями, содержащими комплексные корни, получим разложение многочлена $P(x)$ на линейные и квадратичные действительные множители

$$P(x) = a_0 (x - x_1)^{k_1} \dots (x^2 + p_1 x + q_1)^{l_1} \dots,$$

где k_i — кратности действительных корней, а l_i — кратности комплексно сопряженных корней. Ясно, что $(k_1 + \ldots) + 2(l_1 + \ldots) = n$. Полученным разложением мы будем пользоваться в дальнейшем .

Из сказанного выше следуют простые, но важные для дальнейшего замечания.

1. *Если многочлен P(x) равен нулю при любых значениях x,* т. е. $P(x) \equiv 0$, *то все его коэффициенты равны нулю*

$$a_0 = a_1 = \ldots = a_n = 0.$$

В самом деле, если бы $a_0 \neq 0$, то $P(x)$ имел бы n корней, если бы $a_0 \neq 0$, $a_1 \neq 0$, то $P(x)$ имел бы n —1 корень и т. д.; в условии же сказано, что многочлен имеет бесчисленное множество корней.

Из приведенного рассуждения следует также, что если многочлен n-й степени имеет больше чем n корней, то он тождественно равен нулю.

2. *Если два многочлена равны друг другу при любых значениях x*

$$a_0x^n + a_1x^{n-1} + \ldots + a_n \equiv b_0x^m + b_1x^{m-1} + \ldots + b_m,$$

то равны их степени и равны, между собой коэффициенты при одинаковых степенях x.

Действительно, перенося все члены тождества в левую часть и считая, например, что $n > m$, запишем

$$a_0x^n + \ldots + a_{n-m-1}x^{m+1} + (a_{n-m} - b_0)x^m + \ldots$$
$$\ldots + (a_{n-1} - b_{m-1})x + (a_n - b_m) \equiv 0.$$

Но тогда из первого замечания следует, что во-первых $a_0 = a_1 = \ldots = a_{n-m-1} = 0$ т. е. что старший член первого многочлена есть $a_{n-m}x^m$, а во-вторых, что $a_{n-m} = b_0, \ldots, a_{n-1} = b_{m-1}, a_n = b_m$; при этих условиях оба многочлена тождественны.

Алгебраические уравнения с произвольными коэффициентами читатель умеет решать в двух случаях: $n = 1$ (линейное уравнение) и $n = 2$ (квадратное уравнение). Для случаев $n = 3$ и $n = 4$ (кубическое уравнение и уравнение четвертой степени) также имеются общие формулы решения (формулы Кардано и Феррари)), однако ввиду их громоздкости пользование ими без использования компьютеров крайне затруднительно. Для уравнений же пятой степени и выше доказано, что вообще не существует формул, пользуясь которыми можно было бы при помощи конечного числа алгебраических действий выразить корни через коэффициенты таких уравнений, или, как еще говорят, решить уравнение в радикалах; известны также и конкретные примеры таких «нерешаемых» уравнений. Лишь в отдельных частных случаях, в основном известных из школьного курса алгебры, возможно и алгебраическое решение; чаще всего это бывает тогда, когда левую

часть уравнения удается разложить на произведение многочленов меньших степеней.

В силу сказанного ясно, почему так важно уметь решать уравнения приближенно.

II. Уравнение $R(x) = 0$, где $R(x)$ — рациональная функция, всегда можно привести к виду $\frac{P(x)}{Q(x)} = 0$, где $P(x)$ и $Q(x)$многочлены. Для его решения нужно найти корни алгебраического уравнения $P(x)=0$ и взять из них те, при которых знаменатель $Q(x)$ не равен нулю. Если при некотором значении $x = \alpha$ и $P(\alpha) = 0$ и $Q(\alpha) = 0$, то условимся α считать корнем $R(x)$, если

$$\lim_{x \to \alpha} R(x) = 0,$$

и не считать в противном случае.

Иррациональные уравнения, т.е. уравнения, содержащие x под знаком корня, обычно удается свести к алгебраическим, возвышая обе части уравнения в соответствующие степени. При этом, однако, могут появиться посторонние корни, и поэтому, найдя корни полученного алгебраического уравнения, нужно взять из них только те, которые удовлетворяют исходному уравнению.

III. Трансцендентные уравнения. Уравнение $f(x) = 0$ называется трансцендентным, если функция $f(x)$ — трансцендентная. Примерами трансцендентных уравнений служат логарифмические, показательные и тригонометрические уравнения, изучаемые в курсе элементарной математики. Другие типы трансцендентных уравнений, как правило, алгебраическим путем решены быть не могут. Более того, без специального исследования вообще нельзя сказать, имеет ли данное трансцендентное уравнение корни и в каком количестве.

7.2. Признак кратности корня.

Чтобы перенести на трансцендентные уравнения определение кратности корня, рассмотрим подробнее свойства кратных корней алгебраических уравнений. Пусть многочлен $P(x)$ имеет корень α кратности k. Тогда разложение (**) из п. 7.1 можно записать в виде

$$P(x) = (x - \alpha)^k Q(x), \qquad \text{где} \quad Q(\alpha) \neq 0.$$

Производная $P'(x)$ (это тоже многочлен) равна

$$P'(x) = (x - \alpha)^{k-1} [kQ(x) + (x - \alpha) Q'(x)] = (x - \alpha)^{k-1} Q_1(x).$$

Ясно видно, что $Q_1(\alpha) = kQ(\alpha) \neq 0$. Поэтому производная $P'(x)$

имеет число α корнем кратности *k*—1; если *k*=1, то $P'(\alpha) \neq 0$ и α не является корнем производной. Рассуждая аналогично, установим, что *P″ (x)* имеет α корнем кратности *k* — 2 и т. д. Для производной $P^{(k-1)}(x)$ корень α будет простым, а для *k*-й производной $P^{(k)}$ *(x)* число α вообще не является корнем. Таким образом, если α является корнем кратности *k* для многочлена *P(x),* то

$$P(\alpha) = 0, \quad P'(\alpha) = 0, \quad \ldots, \quad P^{(k-1)}(\alpha) = 0, \text{ но } P^{(k)}(\alpha) \neq 0.$$

Именно это свойство кратных корней алгебраического уравнения мы и примем за определение кратности корня любого уравнения.

Определение. *Число* α *называется корнем кратности k уравнения f(x)* = 0*, или й-кратным нулем функции f(x)* (говорят также *нулем k-го порядка*), *если*

$$f(\alpha) = 0, \; f'(\alpha) = 0, \; \ldots, \; f^{(k-1)}(\alpha) = 0, \text{ но } f^{(k)}(\alpha) \neq 0.$$

При этом предполагается, что функция *f(x)* имеет *k* производных в точке α.

Например, функция *y* = *x* — sin *ч* имеет в точке *x* = 0 трехкратный нуль, так как

$$y_{x=0} = 0, \; y'_{x=0} = (1 - \cos x)_{x=0} = 0, \; y''_{x=0} = (\sin x)_{x=0} = 0, \text{ но } y'''_{x=0} = \cos 0 \neq 0.$$

Уравнение $\sqrt[3]{x^2} = 0$ имеет корень, равный нулю. Мы не говорим об его кратности, так как уже первая производная его левой части $\frac{2}{3\sqrt[3]{x}}$ не существует при *x* = 0.

Если α — простой корень уравнения *f(x)*=0, то график функции *y= (x)* пересекает ось абсцисс; если α — двукратный корень, то график касается оси абсцисс и в некоторой окрестности точки α целиком лежит по одну сторону от оси абсцисс (т. е. точка α является точкой экстремума функции *f(x)).* Когда α — трехкратный корень, то точка (α, 0) оси абсцисс является точкой перегиба графика и касательная в ней совпадает с осью абсцисс; нетрудно также установить, что будет, если α —корень более высокой кратности.

7. 3. Приближенное решение уравнений.

Мы рассмотрим лишь те способы приближенного решения уравнений, которые непосредственно примыкают к содержанию настоящего раздела — исследованию функций. Заметим прежде всего, что знание хотя бы примерного поведения функции *y=f(x)* уже позволяет приблизительно установить, в каких интервалах график функции пересекает ось абсцисс, т. е. где уравнение *f(x)* = 0 имеет

корни. Функция $f(x)$ предполагается непрерывной; поэтому если в точках x_1 и x_2 функция имеет разные знаки, то из свойств непрерывных функций сразу следует, что в интервале $[x_1, x_2]$ имеются нули функции. Будем считать, что интервал $[x_1, x_2]$ удалось выбрать настолько малым, что в нем лежит только один корень уравнения $f(x) = 0$; такой интервал назовем *интервалом изоляции корня.* То, что выбранный интервал является интервалом изоляции, обычно удается проверить при помощи производной $f'(x)$: если она сохраняет постоянный знак, то функция $f(x)$ в этом интервале монотонна и график ее пересекает ось абсцисс только один раз.

Итак, мы исходим из того, что нам так или иначе удалось изолировать корень x_0 уравнения $f(x) = 0$ в некотором интервале $[x_1, x_2]$, $x_1 < x_2$. Каждое из чисел x_1 и x_2 можно считать приближенным значением корня x_0: первое x_1 — с недостатком, второе x_2 — с избытком, причем разность $x_2 — x_1$ является, очевидно, предельной абсолютной ошибкой этих приближенных значений. Излагаемые здесь методы приближенного решения уравнений состоят в приемах, посредством которых по данному интервалу изоляции $[x_1, x_2]$ и по функции $f(x)$ находится такой новый интервал $[x_1, x'_2]$, что

$$x_1 \leqslant x'_1 < x_0 < x'_2 \leqslant x_2,$$

т. е. суживается интервал изоляции; значения x'_1 и x'_2 — лучшие, чем x_1 и x_2, приближенные значения корня x_0. Применяя к интервалу $[x'_1, x'_2]$ тот же или другой метод, получаем еще лучшие приближенные значения x''_1, x''_2 корня x_0.

Мы приводим три метода, самые простые и удобные: ***метод проб, метод хорд и метод касательных***, который известен еще под названием метода Ньютона, а также комбинирование этих трех методов.

Отметим еще, что иногда примерное положение корня уравнения $f(x) = 0$ удобнее находить, записав уравнение в виде $\varphi_1(x) = \varphi_2(x)$, где $\varphi_1(x) - \varphi_2(x) = f(x)$. Если графики функций $\varphi_1(x)$ и $\varphi_2(x)$ известны, то абсциссы точек их пересечения и будут являться корнями данного уравнения. Так, например, записав уравнение $\ln x + x — 2 = 0$ в виде $\ln x = 2 — x$ и построив соответствующие графики, легко заметить, что единственный корень лежит между 1 и 2.

I. Метод проб. Этот метод является самым простым, но не лучшим из методов последовательного приближения к корню уравнения. Пусть интервал $[x_1, x_2]$ есть интервал изоляции корня уравнения $f(x) = 0$. Если корень простой, то значения функции $f(x)$ на концах интервала имеют разные знаки; допустим для определенности, что $f(x_1) < 0$, а $f(x_2) > 0$.

Возьмем любое значение $x=x'$ в интервале $[x_1, x_2]$ и испробуем его, подставив в функцию $f(x)$; если $f(x') < 0$, то мы, заменяя x_1 на x', получим суженный интервал изоляции $[x', x_2]$; если же $f(x')>0$, то мы придем к суженному интервалу изоляции $[x_1, x']$, заменив x_2 через x'.

Неограниченно применяя метод проб, мы получим последовательность точек $x', x'', \ldots$, которая, как это можно доказать, имеет своим пределом корень x_0. В силу этого с помощью метода проб можно найти приближенное значение корня с любой точностью.

Пример. Рассмотрим уравнение

$$f(x) = x^3 + 1{,}1\,x^2 + 0{,}9\,x - 1{,}4 = 0.$$

Так как $f'(x) = 3x^2 + 2{,}2\,x + 0{,}9 > 0$ для всех значений x, то функция $f(x)$ монотонно возрастает и, значит, ее график лишь один раз пересекает ось Ox; кроме того, $f(0) = -1{,}4$, а $f(1) = 1{,}6$, и значит, уравнение имеет единственный действительный корень, лежащий в интервале [0,1].

Находим $f(0{,}5) = -0{,}55$; затем вычисляем $f(0{,}7) = 0{,}112$.

Это показывает, что интервал [0,5; 0,7] есть уменьшенный интервал изоляции искомого корня. Далее мы имеем

$$f(0{,}6) = -0{,}25, \quad f(0{,}65) = -0{,}076, \quad f(0{,}67) = -0{,}002.$$

Ясно, что мы таким образом приближаемся к корню; он лежит в интервале [0,67; 0,7]. Приблизим теперь к корню правую границу интервала изоляции. Испробуем 0,68; получим $f(0{,}68) = 0{,}034$.

Итак, мы нашли новый интервал изоляции [0,67; 0,68], который в 100 раз меньше первоначального [0,1]. Если взять в качестве значения корня число 0,675, то абсолютная ошибка будет меньше 0,005.

Метод проб чаще всего требует более длинных вычислений, чем излагаемые ниже методы хорд и касательных, ибо в этом методе выбор каждого следующего, более точного приближенного значения корня в значительной мере случаен, тогда как в двух следующих методах этот выбор производится целесообразно.

II. Метод хорд. Условия сохраняем те же, что и в методе проб. Соединим концы дуги M_1M_2 (рис. 7.1) линии $y=f(x)$, соответствующей интервалу $[x_1, x_2]$, хордой M_1M_2.

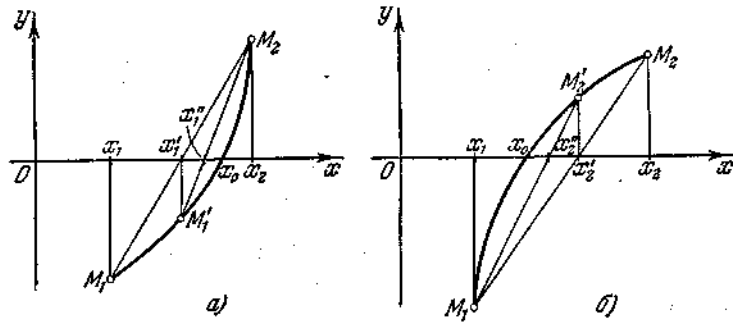


Рис. 7.1.

Очевидно, что на рис. 7.1, *а* точка $x = x'_1$ пересечения этой хорды с осью *Ox* лежит ближе к x_0, чем x_1; исходя из нового суженного интервала $[x'_1, x_2]$, получим точно так же точку x''_1, которая будет лежать еще ближе к x_0, чем x'_1. Таким образом, находим последовательность точек x_1 x'_1, x''_1, ..., стремящуюся, возрастая, к неизвестному нам корню x_0. Аналогично на рис. 7.1, *б* также получается последовательность точек x_2, x'_2, x''_2, ..., стремящаяся, убывая, к корню x_0. Написав уравнение хорды M_1M_2

$$\frac{y-f(x)}{f(x_2)-f(x_1)}=\frac{x-x_1}{x_2-x_1},$$

найдем, положив $y = 0$, выражение для абсциссы x' точки пересечения хорды с осью *Ox:*

$$x'=x_1-f(x_1)\frac{x_2-x_1}{f(x_2)-f(x_1)} \quad \text{или} \quad x'=x_2-f(x_2)\frac{x_2-x_1}{f(x_2)-f(x_1)}. \quad \text{(А)}$$

Это выражение, верное для обоих случаев, изображенных на рис. 7.1, *а* и *б* (а также и когда $f(x_1)>0$, $f(x_2)<0$), дает новое приближение x' к корню x_0 по двум предыдущим приближениям x_1 и x_2. Для того чтобы сузить интервал изоляции, нужно заменить x_1 или x_2 через x'. Какая именно из этих точек заменяется, можно установить сразу по известному нам поведению функции $f(x)$ или, когда это затруднительно, по знаку $f(x')$. Пример. Применим метод хорд к тому же уравнению

$$f(x)=x^3+1{,}1\,x^2+0{,}9\,x-1{,}4=0.$$

Полагая $x_1 = 0$, $x_2=1$, находим по формуле (А)

$$x'=1-f(1)\frac{1-0}{f(1)-f(0)}\approx 0{,}467$$

и, далее, считая $x_1 = 0{,}467$, $x_2=1$,

$$x^{II} \approx 0{,}617;$$

точно так же

$$x^{III} \approx 0{,}660, \quad x^{IV} \approx 0{,}668, \quad x^{V} \approx 0{,}670, \quad x^{VI} \approx 0{,}670.$$

Устойчивость первых трех десятичных знаков в x^V и x^{VI} указывает, как и почти всегда при подобных вычислениях, на то, что мы подошли близко к истинному значению корня. Испробуем для выяснения точности значение 0,671. Имеем: $f(0{,}671) \approx 0{,}0012$, и так как $f(0{,}670) < 0$, то новым интервалом изоляции длиной всего в 0,001 будет интервал [0,670; 0,671]. Взяв за приближенное значение корня 0,6705, мы допускаем ошибку, не превосходящую 0,0005, т. е. в 10 раз меньшую, чем та, которая была допущена в методе проб при одном и том же (примерно) объеме вычислений.

III. Метод касательных. Пусть теперь дуга M_1M_2 линии *fx),* соответствующая интервалу изоляции $[x_1, x_2]$, имеет в каждой своей точке касательную и не имеет точек перегиба, т. е. *f''(x)* не меняет знака в интервале $[x_1, x_2]$.

В то время как метод хорд основан на замене дуги линии ее хордой, метод касательных исходит из замены этой дуги ее касательной. Касательная проводится в концевой точке дуги M_1 или M_2, причем именно в той, которая лежит над осью *Ox,* если дуга вогнута *(f''(x)≥0)* (рис. 7.2, *а),* и которая лежит под осью *Ox,* если дуга выпукла (/"(л;)^0) (рис. 7.2, *б).*

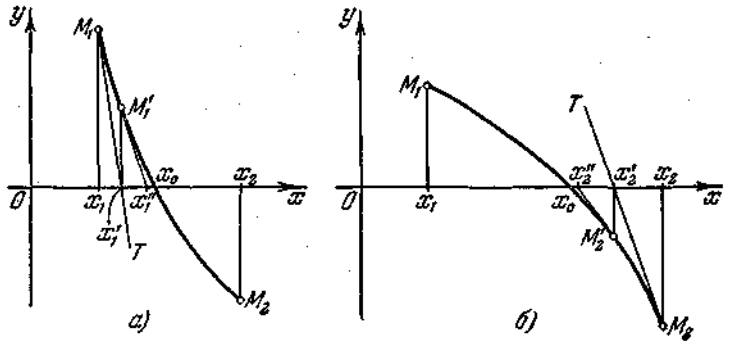


Рис. 7.2.

Эти условия обеспечивают то, что точка пересечения x'_1 (или x'_2) касательной с осью *Ox* всегда будет находиться между корнем x_0 и одним из концов $(x_1$ или $x_2)$ интервала изоляции $[x_1, x_2]$; интервал $[x'_1, x_2]$ или $[x_1, x'_2]$ будет новым суженным интервалом изоляции корня x_0. Если какое-нибудь из условий не выполнено, то новый интервал

изоляции может оказаться шире прежнего, и мы, таким образом, не приблизимся к корню, а удалимся от него. Повторное применение метода касательных приводит к последовательности, имеющей в качестве предела корень x_0. Значит, и по этому методу корень можно найти с любой точностью.

На рис. 7.2 представлены два возможных случая; остальные случаи читатель изобразит на чертеже самостоятельно.

Написав уравнение касательной M_1T или M_2T

$$y - f(x_1) = f'(x_1)(x - x_1) \quad \text{или} \quad y - f(x_2) = f'(x_2)(x - x_2),$$

найдем, положив $y = 0$, выражение для абсциссы x'_1 или x'_2 точки ее пересечения с осью Ox:

$$x'_1 = x_1 - \frac{f(x_1)}{f'(x_1)} \quad \text{или} \quad x'_2 = x_2 - \frac{f(x_2)}{f'(x_2)}. \qquad (Б)$$

Пример. Снова вернемся к уравнению

$$f(x) = x^3 + 1{,}1x^2 + 0{,}9x - 1{,}4 = 0.$$

Так как $f''(x) = 6x + 2{,}2 > 0$ в интервале [0, 1], то касательную проводим в точке с абсциссой, равной 1. По формуле (Б) имеем последовательно

$$x^{I} = 1 - \frac{f(1)}{f'(1)} \approx 0{,}738, \qquad x^{II} = 0{,}738 - \frac{f(0{,}738)}{f'(0{,}738)} \approx 0{,}674,$$

$$x^{III} \approx 0{,}671, \; x^{IV} \approx 0{,}671,$$

причем ясно из самого метода получения приближений, что $f(0{,}671) > 0$. Так как $f(0{,}670) < 0$, тоновым интервалом изоляции будет интервал [0,670, 0,671]. Этот интервал по методу касательных мы получили еще быстрее, чем по методу хорд.

IV. Комбинированные методы. Совместное использование различных методов для решения уравнений иногда называют *комбинированным методом.*

Допустим сейчас, что выполнены условия, указанные во всех рассмотренных методах: дуга линии $y=f(x)$, соответствующая интервалу изоляции $[x_1, x_2]$ корня x_0 уравнения $f(x) = 0$, в каждой своей точке имеет касательную, не имеет точек перегиба и $f(x_1)f(x_2) < 0$. Если применять совместно, например, метод хорд и метод касательных, то это приведет к двум последовательностям точек, стремящихся к точке x_0, с недостатком и с избытком. В случае *а* на рис. 7.2 метод касательных дает приближенные значения $x_1^{(n)}$ к корню x_0 с недостатком, а метод хорд — приближенные значения $x_2^{(n)}$ с избытком; в случае же *б* — наоборот. Это и ускоряет процесс вычисления корня с данной точностью.

Разумеется, комбинированный метод можно также употреблять, совмещая метод хорд или метод касательных с методом проб, как фактически уже было сделано выше при решении уравнения

$$f(x) = x^3 + 1{,}1x^2 + 0{,}9x - 1{,}4 = 0.$$

Применим к этому уравнению комбинированный метод. Имеем $x'_1 \approx 0{,}467$ (см. II) и $x'_2 \approx 0{,}738$ (см. III). По формулам (А) и (Б) находим

$$x''_1 = 0{,}467 - \frac{f(0{,}467)(0{,}738 - 0{,}467)}{f(0{,}738) - f(0{,}467)} \approx 0{,}658, \quad x''_2 \approx 0{,}674,$$

$$x'''_1 = 0{,}658 - \frac{f(0{,}658)(0{,}674 - 0{,}658)}{f(0{,}674) - f(0{,}658)} \approx 0{,}670, \quad x'''_2 \approx 0{,}671.$$

Здесь мы уже на третьем шаге достигли интервала изоляции [0,670; 0,671], в 1000 раз меньшего первоначального интервала [0, 1].

Методы приближенного решения уравнений составляют один из важнейших разделов вычислительной математики или, как еще говорят, численного анализа. Особенно подробно разработаны приемы решения алгебраических уравнений, позволяющие находить не только действительные, но и комплексные корни.

8. Функции нескольких переменных. Дифференциальное исчисление

8.1. Функции нескольких переменных

Функции двух и многих переменных.

I. Случай двух независимых переменных. До сих пор мы рассматривали функции одной независимой переменной. Рассмотрим случаи, когда какая-нибудь величина зависит не от одной независимой переменной, а от двух или большего числа независимых переменных, т. е. когда значения первой величины находятся но значениям не одной, а двух или большего числа других переменных величин. В этих случаях говорят, что указанная величина является функцией двух или соответственно большего числа независимых переменных.

Например, площадь 5 прямоугольника есть функция двух независимо друг от друга изменяющихся переменных — сторон прямоугольника *a* и *b;* выражение для этой функции таково:

$$S = ab.$$

Объем V прямоугольного параллелепипеда является функцией трех независимо друг от друга изменяющихся величин — ребер параллелепипеда a, *b, c:*

$$V=abc.$$

Работа электрического тока A на участке цепи зависит от разности потенциалов U на концах участка, силы тока f и времени t; эта функциональная зависимость дается формулой

$$A = IUt.$$

Остановимся сначала на случае двух независимых переменных, которые будем обозначать буквами x и y.

Каждой паре значений x и y соответствует точка на плоскости Oxy, координатами которой они служат. Возьмем некоторое множество точек на плоскости Oxy и обозначим его через D.

Определение. Величина z называется *функцией* переменных величин x и y на множестве D, если каждой точке этого множества соответствует одно определенное значение величины z.

Множество точек D называется *областью определения функции*. Обычно областью определения функции является некоторая часть плоскости, ограниченная одной или несколькими линиями.

Тот факт, что величина z есть функция величин x и y, записывают так:

$$f=/(x, y).$$

В такой записи за символом функции (которым может быть не только буква f, но и любая другая буква) указываются в скобках те переменные (аргументы), от которых зависит функция.

II. Способы задания функции. Рассмотрим аналитическое и графическое задания функции. Аналитическое задание функции означает, что дается формула, при помощи которой по заданным значениям независимых переменных отыскиваются значения функции.

Так, например, каждая из формул

$$z=2x+3y-5, \qquad z=\frac{xy}{x^2+y^2}, \qquad z=\frac{\sin(2x+3y)}{\sqrt{1+(x-y)^2}}$$

задает z как функцию x и y. При аналитическом задании функции за область ее определения принимают (если нет дополнительных условий) множество значений x и y, для которых формула, определяющая функцию, имеет смысл, т. е. когда в результате ее применения для z получаются определенные действительные значения. Так, например, областью определения функции

$$z=\sqrt{r^2-x^2-y^2}$$

является множество точек плоскости Oxy, координаты которых удовлетворяют соотношению

$$x^2+y^2\leqslant r^2,$$

т. е. круг радиуса r с центром в начале координат. Для функции

$$z = \ln (x^2 + y^2 - r^2)$$

областью определения служат точки, координаты которых удовлетворяют условию

$$x^2 + y^2 > r^2,$$

т. е. внешность круга.

Функция $z = \frac{x}{1 + x^2 + y^2}$ определена на всей плоскости, а функция

$$z = \frac{\sin (xy)}{x - y}$$

на всей плоскости, за исключением прямой $y = x$.

Также как и для функций одной переменной, можно рассматривать неявные функции двух независимых переменных. Уравнение, связывающее три переменные величины:

$$F(x, y, z) = 0,$$

обычно определяет каждую из этих величин как неявную функцию двух остальных.

Прежде чем перейти к графическому заданию, заметим, что графиком функции z двух независимых переменных x и y (в системе декартовых прямоугольных координат) называется множество точек, абсциссы и ординаты которых являются значениями x и y, а аппликаты — соответствующими значениями z. Графиком функции непрерывных аргументов обычно служит некоторая поверхность.

Так как уравнение $x^2 + y^2 + z^2 = R^2$ есть уравнение сферы радиуса R с центром в начале координат, то графиками функций

$$z = \sqrt{R^2 - x^2 - y^2} \quad \text{и} \quad z = -\sqrt{R^2 - x^2 - y^3}$$

служат верхняя и нижняя половины этой сферы (рис. 8.1).

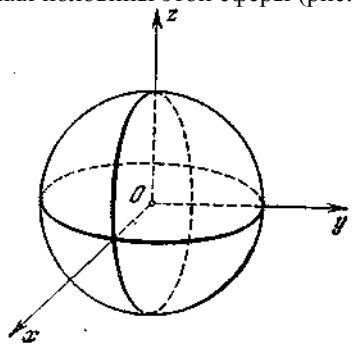


Рис. 8.1

Графиком функции

$$z = x^2 + y^2$$

ляется параболоид вращения (рис. 8.2), а графиком линейной функции

$$z = ax + by + c$$

— плоскость; в частности, графиком константы, т. е. функции $z = C$ (C— const), служит плоскость, параллельная плоскости Oxy.

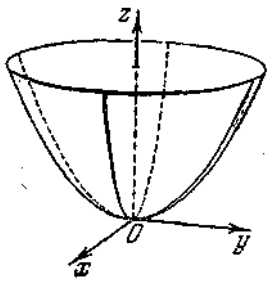


рис. 8.2

Обратно, всякая поверхность в координатном пространстве $Oxyz$ изображает некоторую функцию двух независимых переменных, именно ту, значения которой равны аппликатам точек поверхности при значениях независимых переменных, равных абсциссам и ординатам.

Графическое задание функции двух переменных как раз состоит в задании графика этой функции.

III. Случай многих независимых переменных. Распространим теперь данные выше определения на случай большего числа независимых переменных.

Пусть $x, y, z, \ldots, t$—независимые переменные, а u — величина, зависимая от них, т. е. их функция.

Определение. *Величина u называется функцией переменных величин $x, y, z, \ldots, t$, если каждой рассматриваемой совокупности этих величин соответствует одно определенное значение величины u.*

То, что величина u и есть функция величин $x, y, z, \ldots, t$, записывают так:

$$u = f(x, y, z, \ldots, t),$$

причем вместо буквы f в качестве символа функции может быть употреблена любая другая буква.

Все другие определения, относящиеся к случаю двух независимых переменных, без существенных изменений переносятся на случай многих независимых переменных, и поэтому на них останавливаться нет нужды. Заметим только, что геометрическая иллюстрация функций от n независимых переменных при $n > 2$ теряет наглядность. Уже при $n = 3$, т. е. для функции $u=f(x, y, z)$, геометрически можно представить только область определения функции в виде части трехмерного пространства. При $n>3$ даже области определения функции делаются геометрически ненаглядными. Запись

$$F(x, y, z, \ldots, t, u) = 0 \qquad (*)$$

означает в общем виде наличие функциональной связи между величинами $x, y, z, \ldots, t, u,$ т. е. тот факт, что какая-нибудь из этих величин, например $u,$ является неявной функцией остальных.

8.2. Метод сечений. Предел и непрерывность.

I. Метод сечений. Линии уровня. Как известно, в аналитической геометрии при изучении поверхностей второго порядка обычно пользуются методом сечений, который заключается в том, что определение вида поверхности по ее уравнению производится путем исследования кривых, образованных при пересечении этой поверхности плоскостями, параллельными координатным плоскостям. Этот же метод исследования применим и при изучении любых функций двух переменных. Пусть, например, задана функция

$$z=f(x, y),$$

определяющая некоторую поверхность. Если мы придадим аргументу y постоянное значение y_0 и будем изменять только $x,$ то z станет функцией одной независимой переменной x:

$$z=f(x, y_0).$$

Применив к этой функции известные методы исследования функций одной переменной, мы можем выяснить характер изменения величины z в зависимости от изменения x. Геометрически это означает, что мы рассматриваем линию пересечения поверхности $z=f(x, y)$ и плоскости $y =y_0,$ параллельной плоскости Oxz (линия CD на рис. 8.3). Придавая теперь y другое постоянное значение $y_1,$ получим линию C_1D_1 и т. д. Аналогично можно выяснить поведение z в зависимости от изменения y при различных, но постоянных значениях x.

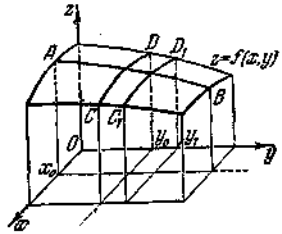


Рис. 8.3

Так, на рис, 8.3 изображена линия AB пересечения поверхности $z=f(x,y)$ и плоскости $x = x_0$. Чтобы представить себе эту линию, мы должны исследовать функцию одной переменной $z=f(x_0, y)$. Зная характер расположения полученных линий, мы сможем составить себе представление о всей поверхности, т. е. описать поведение функции z при произвольном изменении ее аргументов.

Но можно изучать функцию $z = f(x, y)$ посредством того же приема сведения функции двух переменных к функции одной переменной, придавая постоянные значения не одной из независимых переменных, а самой функции. Именно положим $z = z_0$; тогда уравнение

$$f(x,y)=z_0 \qquad \textbf{(*)}$$

дает зависимость между переменными x и y (т. е. функцию одной переменной), при которой заданная функция z сохраняет постоянное значение z_0. Геометрически придание z постоянного значения z_0 означает пересечение поверхности $z=f(x, y)$ с плоскостью $z=z_0$, параллельной плоскости Oxy.

На плоскости Oxy уравнение (*) есть уравнение проекции l линии пересечения L поверхности $z=f(x, y)$ с плоскостью $z = z_0$ (рис. 8.4).

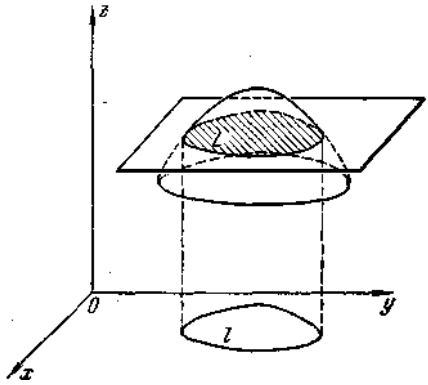


Рис. 8.4

При перемещении точки с координатами x, y вдоль линии l функция сохраняет постоянное значение, равное z_0.

Определение. *Линией уровня функции* $z = f(x, y)$ *называется линия на плоскости* Oxy, *в точках которой функция сохраняет постоянное значение*.

Совокупность линий уровня, соответствующих различным значениям z, называется *сетью линий уровня* функции $z=f(x, y)$. Сеть эта, при условии, что она проведена для мало отличающихся друг от друга значений z, довольно наглядно характеризует поведение функции.

II. Предел функции. Перейдем теперь к определению предела функции двух переменных; оно будет совершенно аналогично определению предела для функций одной переменной.

Определение. *Число* A *называется пределом функции* $z=f(x, y)$ *при* $x \to x_0$, $y \to y_0$, *если для всех значений* x *и* y, *достаточно мало отличающихся соответственно от чисел* x_0 *и* y_0, *соответствующие значения функции* $f(x, y)$ *как угодно мало отличаются от числа* A.

Записывают

$$\lim_{\substack{x \to x_0 \\ y \to y_0}} f(x, y) = A. \qquad (*)$$

В этом определении не предполагается, что функция определена в самой точке $P_0\ (x_0, y_0)$, и поэтому мы считаем, что либо $x \neq x_0$, либо $y \neq y_0$. Вполне аналогично определяется предел функции n независимых переменных при $n > 2$.

Отметим также, что все правила предельного перехода, без всяких изменений переносятся на случай функций многих переменных.

III. Непрерывность функции. Пусть точка $P_0\ (x_0, y_0)$ принадлежит области определения функции $z=f(x,y)$. Так же как и для функции одной переменной, *приращением* функции $z = f(x, y)$ в данной точке P_0 называется разность

$$\Delta z = f(x_0 + \Delta x,\ y_0 + \Delta y) - f(x_0,\ y_0),$$

где Δx и Δy — приращения аргументов.

Определение. *Функция* $z = f(x, y)$ *называется непрерывной в точке* (x_0, y_0), *если она определена в некоторой окрестности этой точки и если бесконечно малым приращениям* x *и* y *соответствует бесконечно малое приращение* z, *т. е.*

$$\lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} \Delta z = 0.$$

Обозначив $x_0+\Delta x$ через x и $y_0+\Delta y$ через y, мы можем условие непрерывности функции $f(x, y)$ записать в виде

$$\lim_{\substack{x\to x_0\\ y\to y_0}} [f(x, y) - f(x_0, y_0)] = 0 \text{ или } \lim_{\substack{x\to x_0\\ y\to y_0}} f(x, y) = f(x_0, y_0).$$

Последнее означает, что *функция непрерывна в точке (x_0, y_0), если ее предел равен значению функции в предельной точке.*

Определение. *Функция, непрерывная в каждой точке области, называется непрерывной в этой области.*

8. 3. Производные и дифференциалы. Дифференциальное исчисление

Частные производные и дифференциалы.

I. Частные производные. Пусть $z=f(x, y)$ — функция двух независимых переменных x и y. Будем считать сейчас аргумент y постоянным и рассмотрим получающуюся при этом функцию одной переменной x. Допустим, что эта функция $f(x, y)$ при данном значении x дифференцируема, т. е. имеет производную, равную

$$\lim_{\Delta x\to 0} \frac{f(x+\Delta x, y) - f(x, y)}{\Delta x}.$$

Этот предел мы обозначим через $f_x(x, y)$, указывая нижним индексом x, что производная берется по переменной x при фиксированном y.

Определение. *Частной производной по x от функции $z=f(x, y)$ называется функция переменных x и y, получающаяся при дифференцировании $f(x, y)$ по x в предположении, что y считается постоянным.*

Для частной производной по x от функции $z=f(x, y)$ употребляют следующие обозначения:

$$\frac{\partial z}{\partial x},\ z'_x,\ \frac{\partial f(x, y)}{\partial x},\ \frac{\partial}{\partial x}[f(x, y)].$$

Обращаем внимание читателя на то, что y считается постоянным только в процессе дифференцирования. После того, как выражение для $f_x(x, y)$ найдено, x и y могут принимать любые значения. Это и означает, что $f_x(x, y)$ является функцией обеих переменных. Например, если $z = x^2y$, то $z_x = 2xy$.

Разумеется, в частных случаях $f_x(x, y)$ может оказаться зависящей от одной переменной или даже быть постоянной. Например, если $z = xy$, то $z_x=y$, а если $z = 2x+y^2$, то $z_x = 2$.

Совершенно аналогично определяется *частная производная по y от функции $z=f(x,y)$:*

$$f'_y(x, y) = \lim_{\Delta y \to 0} \frac{f(x, y+\Delta y) - f(x, y)}{\Delta y}.$$

Она обозначается еще так:

$$\frac{\partial z}{\partial y}, \quad z'_y, \quad \frac{\partial f(x, y)}{\partial y}, \quad \frac{\partial}{\partial y}[f(x, y)].$$

Так как частная производная является обыкновенной производной от данной функции, взятой в предположении, что изменяется только переменная, по которой производится дифференцирование, то фактическое отыскание частных производных элементарных функций осуществляется по известным правилам дифференцирования функций одной переменной.

Аналогично определяются частные производные от функции любого числа независимых переменных.

Обозначения для частных производных такие же, как и в случае двух переменных, например:

$$\frac{\partial u}{\partial x} = f'_x(x, y, z, \ldots, t) = \lim_{\Delta x \to 0} \frac{f(x+\Delta x, y, z, \ldots, t) - f(x, y, z, \ldots, t)}{\Delta x}.$$

II. Частные дифференциалы. Приращение, которое получает функция $z=f(x, y)$, когда изменяется только одна из переменных, называется *частным приращением функции по соответствующей переменной.* Употребляются такие обозначения:

$$\Delta_x z = f(x+\Delta x, y) - f(x, y),$$
$$\Delta_y z = f(x, y+\Delta y) - f(x, y).$$

Определение. *Частным дифференциалом по x функции* $z=f(x, y)$ *называется главная часть частного приращения* $\Delta_x z = f(x+\Delta x, y) - f(x, y)$, *пропорциональная приращению* Δx *независимой переменной x.*

Аналогично определяется *частный дифференциал по y.* Дифференциалы независимых переменных x и y просто равны их приращениям, т. е.

$$dx = \Delta x, \quad dy = \Delta y.$$

Частные дифференциалы обозначаются так: $d_x z$— частный дифференциал по x, $d_y z$—частный дифференциал по y.

Так же, как для случая функции одной переменной, доказывается, что если функция $z = f(x, y)$ имеет частный дифференциал по x, то она имеет и частную производную z_x, и обратно. При этом

$$d_x z = \frac{\partial z}{\partial x} dx.$$

Аналогично, если функция $z=f(x, y)$ имеет частный дифференциал по y, то она имеет и частную производную z_y и обратно,
Причем

$$d_y z = \frac{\partial z}{\partial y} dy.$$

Таким образом, *частный дифференциал функции двух независим мых переменных равен произведению соответствующей частной производной на дифференциал этой переменной.*

Таким же образом, как для функций двух переменных, определяются частные приращения и частные дифференциалы функций любого числа независимых переменных.

При этом, если $u=f(x, y, z, \ldots, t)$, то, как и выше,

$$d_x u = \frac{\partial u}{\partial x} dx, \qquad d_y u = \frac{\partial u}{\partial y} dy, \quad \ldots,$$

т. е. *частный дифференциал функции нескольких независимых переменных по какой-нибудь из них равен произведению соответствующей частной производной на дифференциал этой переменной.*

Полный дифференциал.

I. Полное приращение и полный дифференциал. Пусть функция $z=f(x, y)$ дифференцируема по x и по y; тогда с помощью частных дифференциалов мы можем находить как угодно точные выражения для приращений функции при достаточно малых изменениях x и y в отдельности. Естественно постараться найти выражение для приращения функции $z=f(x, y)$ при произвольных совместных изменениях ее обоих аргументов. В этом случае приращение функции

$$\Delta z = f(x + \Delta x, y + \Delta y) - f(x, y)$$

называют *полным приращением.* Геометрически приращение функции при переходе от точки $P(x, y)$ к точке

$$P_1 (x + \Delta x, y + \Delta y)$$

выразится отрезком QM_1 (рис. 8.5).

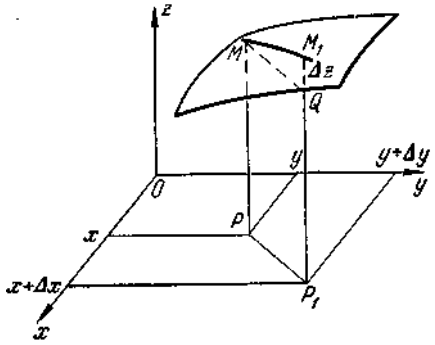


Рис. 8.5

Полное приращение функции весьма сложно выражается через приращения независимых переменных; только в одном случае эта зависимость проста, именно в случае, когда функция *f(x, y)* линейная:

$$f(x,\ y) = ax + by + c;$$

тогда, как легко видеть,

$$\Delta z = a\Delta x + b\Delta y.$$

Сумма $a\Delta x + b\Delta y$ называется *полным дифференциалом функции* $z = f(x, y)$; она обозначается через dz или $df(x_t y)$:

$$dz = a\,dx + b\,dy \qquad (*)$$

Определение. *Полным дифференциалом функции двух независимых переменных называется главная часть полного приращения функции, линейная относительно приращений независимых переменных.*

Теорема. Полный дифференциал функции двух независимых переменных равен сумме произведений частных производных функции на дифференциалы соответствующих независимых переменных.

Дифференцируемость функций.

Определение. *Функция двух независимых переменных, имеющая в некоторой точке дифференциал, называется дифференцируемой в этой точке.*

Если, помимо существования всех частных производных, добавить еще требование их непрерывности, то из этого уже будет вытекать дифференцируемость функции.

Теорема. Если функция $z=f(x, y)$ имеет в точке $P(x, y)$ непрерывные частные производные $f'_x(x, y)$ и $f'_y(x,y)$, то в этой точке функция дифференцируема**.**

Геометрический смысл полного дифференциала функции двух независимых переменных вытекает из следующего предложения.

Теорема. Полный дифференциал функции $z=f(x, y)$ при $x = x_0$, $y=y_0$ изображается приращением аппликаты точки касательной плоскости, проведенной к поверхности $z=f(x, y)$ в ее точке $M_0(x_0, y_0, z_0)$.

Применение полного дифференциала к приближенным вычислениям.

Ранее мы видели, как применяется к приближенным вычислениям дифференциал функции одной переменной. Покажем теперь, как к решению аналогичных задач для функций многих переменных применяется полный дифференциал. Будем для простоты записи говорить о функциях двух переменных; все сказанное легко перенесется на функции большею числа переменных.

Исходным для дальнейшего является *замена полного приращения функции ее полным дифференциалом,* т. е. приближенное равенство

$$f(x_0 + \Delta x,\ y_0 + \Delta y) - f(x_0,\ y_0) \approx dz,$$

справедливое при малых Δx и Δy. Заменив dz его выражением, придадимэтому равенству вид

$$f(x_0 + \Delta x,\ y_0 + \Delta y) \approx f(x_0,\ y_0) + f'_x(x_0,\ y_0)\,\Delta x + f'_y(x_0,\ y_0)\,\Delta y.$$

Если положить $x_0+\Delta x = x$, $y_0+\Delta y =y$, то это приближенное равенство можно записать и так:

$$f(x,\ y) \approx f(x_0,\ y_0) + f'_x(x_0,\ y_0)(x - x_0) + f'_y(x_0,\ y_0)(y - y_0). \quad (*)$$

Последняя формула показывает, что замена полного приращения функции ее полным дифференциалом привела к замене функции $f(x, y)$ в окрестности точки (x_0, y_0) линейной функцией. Геометрически это значит, что участок поверхности $z=f(x, y)$ заменяется соответствующим участком касательной плоскости к поверхности в точке $M_0 (x_0, y_0, z_0)$. Формула (*) сразу позволяет по известным значениям

$$f(x_0, y_0),\ f'_x(x_0,\ y_0),\ f'_y(x_0,\ y_0)$$

приближенно вычислять значения функции $f(x, y)$ при условии, что x мало отличается от x_0, а y от y_0.

Пример. Составим приближенную формулу для вычисления значений функции

$$z = \ln(xy + 2y^2 - 2x)$$

в окрестности точки (1, 1).

Здесь $x_0 = 1$ $y_0 = 1$. Последовательно вычисляем:

$$\frac{\partial z}{\partial x} = \frac{1}{xy + 2y^2 - 2x}(y - 2), \qquad \left(\frac{\partial z}{\partial x}\right)_{\substack{x=1 \\ y=1}} = -1,$$

$$\frac{\partial z}{\partial y} = \frac{1}{xy + 2y^2 - 2x}(x + 4y), \qquad \left(\frac{\partial z}{\partial y}\right)_{\substack{x=1 \\ y=1}} = 5.$$

Так как $z_0 = 0$, то

$$\ln(xy + 2y^2 - 2x) \approx -(x - 1) + 5(y - 1).$$

Покажем теперь, как вычисляется предельная абсолютная ошибка функции двух переменных по заданным предельным абсолютным ошибкам аргументов.

Пусть задана функция $z = f(x, y)$ и требуется вычислить ее значение при условии, что известны лишь приближенные значения аргументов x и y:

$$x = x_0 + dx, \quad |dx| < \varepsilon_x; \quad y = y_0 + dy, \ |dy| < \varepsilon_y,$$

где x_0 и y_0 приближенные значения, а ε_x и ε_y— предельные абсолютные ошибки. Чтобы найти предельную абсолютную ошибку функции ε_z, нужно оценить модуль разности между истинным значением $f(x, y)$ и приближенным $f(x_0, y_0)$. Из формулы (*) следует

$$|f(x, y) - f(x_0, y_0)| \approx |f'_x(x_0, y_0)\,dx + f'_y(x_0, y_0)\,dy| \leqslant$$
$$\leqslant |f'_x(x_0, y_0)|\,|dx| + |f'_y(x_0, y_0)|\,|dy| <$$
$$< |f'_x(x_0, y_0)|\,\varepsilon_x + |f'_y(x_0, y_0)|\,\varepsilon_y.$$

Таким образом, можно сказать, что

$$\varepsilon_z = |f'_x(x_0, y_0)|\,\varepsilon_x + |f'_y(x_0, y_0)|\,\varepsilon_y. \qquad (**)$$

Деля ε_z на $f(x_0, y_0)$, получим предельную относительную ошибку δ_z.

В качестве примеров найдем относительные ошибки произведения и частного.

1) Пусть $z = xy$ и x_0, y_0 — приближенные значения аргументов. По формуле (**)

$$\varepsilon_z = |y_0|\,\varepsilon_x + |x_0|\,\varepsilon_y.$$

Деля на $|z_0| = |x_0 y_0|$, находим δ_z:

$$\delta_z = \frac{\varepsilon_x}{|x_0|} + \frac{\varepsilon_y}{|y_0|} = \delta_x + \delta_y,$$

т. е. *предельная относительная ошибка произведения равна сумме предельных относительных ошибок сомножителей.* Разумеется, это правило справедливо для любого числа сомножителей.

2) Пусть $z = \frac{y}{x}$ и $z_0 = \frac{y_0}{x_0}$. Тогда

$$\varepsilon_z = \left| -\frac{y_0}{x_0^2} \right| \varepsilon_x + \left| \frac{1}{x_0} \right| \varepsilon_y,$$

$$\delta_z = \frac{\varepsilon_x}{|x_0|} + \frac{\varepsilon_y}{|y_0|} = \delta_x + \delta_y,$$

т. е. *предельная относительная ошибка частного равна сумме предельных относительных ошибок делимого и делителя.* (Таким образом, предельные относительные ошибки, получающиеся при делении двух величин и при их произведении равны между собой.)

Формула (**) позволяет решать и обратную задачу: по известным приближенным значениям x_0, y_0 и заданной допустимой ошибке ε_z в определении функции $f(x, y)$ рассчитать, каковы должны быть ε_x и ε_y. Пользуясь неопределенностью задачи (две величины связаны только одним соотношением), мы можем подбирать величины ε_x и ε_y в зависимости от того, какую из них легче сделать меньше, т. е. какую из величин x и y легче измерить с большей точностью. Если при этом окажется, что приближенные значения x_0, y_0 измерены с недостаточной точностью, то надо заново произвести их измерения уже с требуемой точностью и затем снова вычислить значение функции.

Производные и дифференциалы высших порядков.

I. Допустим, что функция $z=f(x, y)$ имеет частные производные

$$\frac{\partial z}{\partial x} = f'_x(x, y), \qquad \frac{\partial z}{\partial y} = f'_y(x, y);$$

эти производные в свою очередь являются функциями независимых переменных x и y. Частные производные от этих функций называются *вторыми частными производными* или *частными производными второго порядка* от данной функции $f(x, y)$. Каждая производная *первого порядка* имеет две частные производные; таким образом, мы получаем четыре частные производные второго порядка, которые обозначаются так:

$$\frac{\partial}{\partial x}\left(\frac{\partial z}{\partial x}\right)=\frac{\partial^2 z}{\partial x^2}=f''_{xx}=z''_{xx}, \qquad \frac{\partial}{\partial y}\left(\frac{\partial z}{\partial x}\right)=\frac{\partial^2 z}{\partial x\,\partial y}=f''_{xy}=z''_{xy},$$

$$\frac{\partial}{\partial x}\left(\frac{\partial z}{\partial y}\right)=\frac{\partial^2 z}{\partial y\,\partial x}=f''_{yx}=z''_{yx}, \qquad \frac{\partial}{\partial y}\left(\frac{\partial z}{\partial y}\right)=\frac{\partial^2 z}{\partial y^2}=f''_{yy}=z''_{yy}.$$

Производные f''_{xy} и f''_{yx} называются *смешанными;* одна из них получается дифференцированием функции сначала по *x,* затем по *y,* другая, наоборот, — сначала по *y,* затем по *x.*

Теорема. Если вторые смешанные производные функции $z=f(x, y)$ непрерывны, то они равны между собой:

$$f''_{xy}(x, y)=f''_{yx}(x, y).$$

Условие непрерывности частных производных является существенным; если оно не выполнено, теорема может оказаться несправедливой.

Таким образом, функция двух переменных $z=f(x, y)$ имеет при указанных условиях фактически не четыре, а только три частные производные второго порядка:

$$\frac{\partial^2 z}{\partial x^2}, \qquad \frac{\partial^2 z}{\partial x\,\partial y}=\frac{\partial^2 z}{\partial y\,\partial x}, \qquad \frac{\partial^2 z}{\partial y^2}.$$

Частные производные от частных производных второго порядка называются *частными производными третьего порядка* (или *третьими частными производными)* и т. д.

Теорема о равенстве вторых смешанных производных позволяет доказать общее предложение:

Результат повторного дифференцирования функции двух независимых переменных не зависит от порядка дифференцирования (предполагается, что рассматриваемые частные производные *непрерывны).*

Элементарные функции двух независимых переменных, как правило (за исключением отдельных точек и отдельных линий), имеют в своей области определения частные производные любых порядков.

Частные производные высшего порядка для функций любого числа независимых переменных определяются аналогично. Здесь также имеет место теорема о независимости результата от порядка дифференцирования. Так, например, если $u = f(x, y, z)$, то

$$\frac{\partial^3 u}{\partial x\,\partial y\,\partial z}=\frac{\partial^3 u}{\partial x\,\partial z\,\partial y}=\frac{\partial^3 u}{\partial y\,\partial x\,\partial z}=\frac{\partial^3 u}{\partial y\,\partial z\,\partial x}=\frac{\partial^3 u}{\partial z\,\partial x\,\partial y}=\frac{\partial^3 u}{\partial z\,\partial y\,\partial x}.$$

Повторное дифференцирование функции нескольких переменных фактически производится последовательным нахождением, одной производной вслед за другой.

II. Полный дифференциал функции $z = f(x, y)$

$$dz = f'_x(x,\ y)\,dx + f'_y(x,\ y)\,dy = \frac{\partial z}{\partial x}\,dx + \frac{\partial z}{\partial y}\,dy$$

зависит, во-первых, от независимых переменных *x, y* и, во-вторых, от их дифференциалов *dx, dy*. Дифференциалы *dx* и *dy* являются величинами, не зависящими от *x* и *y*.

Полным дифференциалом второго порядка называется полный дифференциал от полного дифференциала первого порядка dz при условии, что dx и dy считаются постоянными. Согласно этому определению

$$d^2z = d\left(\frac{\partial z}{\partial x}\,dx + \frac{\partial z}{\partial y}\,dy\right) =$$
$$= \frac{\partial}{\partial x}\left(\frac{\partial z}{\partial x}\,dx + \frac{\partial z}{\partial y}\,dy\right)dx + \frac{\partial}{\partial y}\left(\frac{\partial z}{\partial x}\,dx + \frac{\partial z}{\partial y}\,dy\right)dy.$$

Отыскание функции по ее полному дифференциалу.

I. Случай двух независимых переменных. Пусть *x* и *y* — независимые переменные, а *P(x,y)* и *Q(x, y)* — функции от них, непрерывные вместе со своими первыми частными производными.

Говорят, что *дифференциальное выражение P(x, y)dx+Q(x, y)dy является полным дифференциалом, если существует такая функция u(x, y), полный дифференциал которой равен данному выражению*

$$du = P(x,\ y)\,dx + Q(x,\ y)\,dy.$$

Теорема. Для того чтобы выражение *P(x, y)dx + Q(x, y)dy* было полным дифференциалом, необходимо и достаточно соблюдение тождества

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

II. Случай трех независимых переменных. Метод отыскания функции двух независимых переменных по ее полному дифференциалу почти без всяких изменений переносится на функции трех переменных. Мы поэтому ограничимся краткими указаниями. Пусть *P(x, y, z)*, *Q(x, y, z)* и *R (x, y, z)* — функции независимых переменных *x, y, z*, непрерывные вместе со своими частными производными. Имеет место теорема.

Теорема. Для того чтобы выражение

$$P(x,\ y,\ z)\,dx + Q(x,\ y,\ z)\,dy + R(x,\ y,\ z)\,dz$$

было полным дифференциалом некоторой функции *u(x, y, z)*, необходимо и достаточно соблюдение тождеств

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

Дифференцирование сложных функций. Правила для отыскания дифференциала функций.

I. Дифференцирование сложных функций. Пусть задана дифференцируемая функция $z = f(u, v)$. Тогда ее приращение Δz можно представить в виде

$$\Delta z = \frac{\partial z}{\partial u}\Delta u + \frac{\partial z}{\partial v}\Delta v + \alpha, \qquad (*)$$

где α — бесконечно малая более высокого порядка, чем $\sqrt{\Delta u^2 + \Delta v^2}$, т.е.

$$\lim_{\substack{\Delta u \to 0 \\ \Delta v \to 0}} \frac{\alpha}{\sqrt{\Delta u^2 + \Delta v^2}} = 0.$$

Предположим теперь, что u и v в свою очередь являются дифференцируемыми функциями независимой переменной x, т. е.

$$u = \varphi(x) \quad \text{и} \quad v = \psi(x).$$

Таким образом,

$$z = f[\varphi(x), \psi(x)] = F(x),$$

т. е. z является сложной функцией переменной x. Выразим теперь производную $\frac{dz}{dx}$ через частные производные $\frac{\partial z}{\partial u}$ и $\frac{\partial z}{\partial v}$ и производные от функций u и v по x. Придадим аргументу x приращение Δx. Тогда u и v получат соответственно приращения Δu и Δv, через которые Δz выразится по формуле (*). Разделим обе части этой формулы на Δx:

$$\frac{\Delta z}{\Delta x} = \frac{\partial z}{\partial u}\frac{\Delta u}{\Delta x} + \frac{\partial z}{\partial v}\frac{\Delta v}{\Delta x} + \frac{\alpha}{\Delta x}$$

и перейдем к пределу при $\Delta x \to 0$. Согласно условию

$$\lim_{\Delta x \to 0} \frac{\Delta u}{\Delta x} = \frac{du}{dx}$$

и $\lim_{\Delta x \to 0} \frac{\Delta v}{\Delta x} = \frac{dv}{dx}$. Отношение $\frac{\alpha}{\Delta x}$ представим в виде

$$\frac{\alpha}{\Delta x} = \frac{\alpha}{\sqrt{\Delta u^2 + \Delta v^2}} \frac{\sqrt{\Delta u^2 + \Delta v^2}}{\Delta x} = \frac{\alpha}{\sqrt{\Delta u^2 + \Delta v^2}} \sqrt{\left(\frac{\Delta u}{\Delta x}\right)^2 + \left(\frac{\Delta v}{\Delta x}\right)^2}.$$

Первый множитель, согласно определению α, стремится к нулю, а второй к определенному числу: $\sqrt{\left(\frac{du}{dx}\right)^2 + \left(\frac{dv}{dx}\right)^2}$. Следовательно,

$$\lim_{\Delta x \to 0} \frac{\alpha}{\Delta x} = 0.$$

Замечая еще, что значения

$$\frac{\partial z}{\partial u} \text{ и } \frac{\partial z}{\partial v}$$

зависят только от выбранного значения x, определяющего значения u и v, и не зависят от Δx, окончательно получим

$$\frac{dz}{dx} = \frac{\partial z}{\partial u}\frac{du}{dx} + \frac{\partial z}{\partial v}\frac{dv}{dx}.$$

Ясно, что эта формула является обобщением правила дифференцирования сложной функции одной переменной.

Пусть теперь z является сложной функцией двух независимых переменных x и y, т. е.

$$z = f(u, v),$$

где

$$u = \varphi(x, y) \quad \text{и} \quad v = \psi(x, y).$$

Таким образом,

$$z = f[\varphi(x, y), \psi(x, y)] = F(x, y).$$

Все функции предполагаются дифференцируемыми.

Чтобы найти z , мы должны считать y постоянным, но тогда и u и v становятся функциями только одной переменной x, и мы приходим к уже рассмотренному случаю. Разница состоит только в том, что обыкновенные производные $\frac{du}{dx}$ и $\frac{dv}{dx}$ заменятся частными производными $\frac{\partial u}{\partial x}$ и $\frac{\partial v}{\partial x}$, и мы получим

$$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial z}{\partial v}\frac{\partial v}{\partial x}$$

и аналогично

$$\frac{\partial z}{\partial y} = \frac{\partial z}{\partial u}\frac{\partial u}{\partial y} + \frac{\partial z}{\partial v}\frac{\partial v}{\partial y}.$$

Итак,

Частная производная сложной функции равна сумме произведений частных производных заданной функции по промежуточным аргументам (и и v) на частные производные этих аргументов (и и v) по соответствующей независимой переменной (x или y).

Сформулированное правило дифференцирования сложной функции остается справедливым для функций любого числа независимых переменных и при всяком числе промежуточных аргументов.

Пусть z задана как функция аргументов $u, v, ..., w$, которые являются функциями независимых переменных $x, y, ..., t$. Тогда

$$\frac{\partial z}{\partial x}=\frac{\partial z}{\partial u}\frac{\partial u}{\partial x}+\frac{\partial z}{\partial v}\frac{\partial v}{\partial x}+\ldots+\frac{\partial z}{\partial w}\frac{\partial w}{\partial x},$$
$$\frac{\partial z}{\partial y}=\frac{\partial z}{\partial u}\frac{\partial u}{\partial y}+\frac{\partial z}{\partial v}\frac{\partial v}{\partial y}+\ldots+\frac{\partial z}{\partial w}\frac{\partial w}{\partial y},$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$\frac{\partial z}{\partial t}=\frac{\partial z}{\partial u}\frac{\partial u}{\partial t}+\frac{\partial z}{\partial v}\frac{\partial v}{\partial t}+\ldots+\frac{\partial z}{\partial w}\frac{\partial w}{\partial t}.$$

В частном случае аргументы *u, v, ..., w* могут быть функциями одной независимой переменной *x.* Значит, фактически *z* является функцией только *x.* При этом ее обыкновенная производная — ее называют в этом случае полной производной — имеет вид

$$\frac{dz}{dx}=\frac{\partial z}{\partial u}\frac{du}{dx}+\frac{\partial z}{\partial v}\frac{dv}{dx}+\ldots+\frac{\partial z}{\partial w}\frac{dw}{dx}.$$

Если в этой формуле положить $u = x$, то получим

$$\frac{dz}{dx}=\frac{\partial z}{\partial x}+\frac{\partial z}{\partial v}\frac{dv}{dx}+\ldots+\frac{\partial z}{\partial w}\frac{dw}{dx}.$$

Обращаем внимание читателя на различие между двумя производными по *x,* содержащимися в этой формуле. В то время как $\frac{dz}{dx}$ есть полная производная, т. е. обыкновенная производная от *z* как функции *x,* $\frac{\partial z}{\partial x}$ есть частная производная от *z* по аргументу *x,* входящему в выражение функции непосредственно, значит, при условии, что остальные аргументы, хотя они также зависят от *x,* при дифференцировании остаются постоянными.

II. Инвариантность формы первого дифференциала и правила для отыскания дифференциала.

Теорема. Полный дифференциал функции $z=f(u,v)$ сохраняет один и тот же вид независимо от того, являются ли ее аргументы *u* и *v* независимыми переменными или функциями от независимых переменных.

Если *u* и *v* — независимые переменные, то полный дифференциал равен

$$dz=\frac{\partial z}{\partial u}du+\frac{\partial z}{\partial v}dv.$$

Таким образом, и для функций двух независимых переменных имеет место свойство инвариантности формы первого дифференциала функции**.**

Совершенно аналогично, если $z = f(u, v, .. ., w)$, то всегда

$$dz = \frac{\partial z}{\partial u}\,du + \frac{\partial z}{\partial v}\,dv + \ldots + \frac{\partial z}{\partial w}\,dw$$

независимо от того, являются ли непосредственные аргументы этой функции независимыми переменными или нет. При нахождении дифференциала функции нескольких независимых переменных можно также пользоваться простыми правилами, аналогичными правилам, имеющим место в случае одной независимой переменной. Пусть $u, v, \ldots, w$ — функции любого числа независимых переменных. Тогда имеют место правила, выраженные следующими формулами:

1) $d(u + v + \ldots + w) = du + dv + \ldots + dw$;

2) $d(uv) = v\,du + u\,dv$, в частности $d(Cu) = C\,du$, C — const;

3) $d\left(\frac{u}{v}\right) = \frac{v\,du - u\,dv}{v^2}$.

Теорема существования неявной функции. Мы уже неоднократно встречались с неявными функциями, т. е. функциями, определяемыми при помощи уравнений, связывающих переменные величины. Напомним, что неявная функция одного переменного определяется уравнением

$$F(x, y) = 0. \qquad (*)$$

Выясним сейчас, при каких условиях можно утверждать, что заданное уравнение $F(x,y) = 0$ действительно определяет одну из переменных как функцию другой. Имеет место теорема.

Теорема существования неявной функции. Пусть функция $F(x, y)$ непрерывна вместе со своими частными производными в какой-нибудь окрестности точки $M_0(x_0, y_0)$. Если

$$F(x_0, y_0) = 0 \quad \text{и} \quad F'_y(x_0, y_0) \neq 0,$$

то уравнение

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

при значениях x, близких к x_0, имеет единственное непрерывно зависящее от x решение $y = \varphi(x)$ такое, что $\varphi(x_0) = y_0$. Функция $\varphi(x)$ имеет также непрерывную производную.

Неявная функция двух переменных определяется уравнением

$$F(x, y, z) = 0,$$

связывающим три переменные величины. На вопрос о существовании такой функции отвечает следующая теорема, аналогичная приведенной выше.

Теорема. Пусть функция $F(x, y, z)$ непрерывна вместе со своими частными производными в какой-нибудь окрестности точки $M_0(x_0, y_0, z_0)$. Если

$$F(x_0,\ y_0,\ z_0)=0 \text{ и } F'_z(x_0,\ y_0,\ z_0)\neq 0,$$

то уравнение $F(x, y, z) = 0$ в некоторой окрестности точки (x_0, y_0) имеет единственное, непрерывно зависящее от x и y решение $z = \varphi(x, y)$ такое, что $\varphi(x_0, y_0) = z_0$. Функция $\varphi(x, y)$ имеет также непрерывные частные производные.

Отметим еще, что если в точке M_0 производная $F'_z = 0$, а, скажем, $F'_y \neq$, то уравнение $F(x, y, z) = 0$ может не определять z как функцию x и y, но определяет y как функцию x и z.

Дифференцирование неявных функций.

Пусть условия теоремы существования неявной функции, сформулированной в предыдущем пункте, выполнены и уравнение

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

определяет y как некоторую функцию $y=\varphi(x)$. Если в уравнение подставить вместо y функцию $\varphi(x)$, то получим тождество

$$F[x,\ \varphi(x)] \equiv 0.$$

Следовательно, производная по x от функции $F(x, y)$, где $y=\varphi(x)$, также должна быть равной нулю. Дифференцируя по правилу дифференцирования сложной функции, найдем

$$F'_x + F'_y \frac{dy}{dx} = 0,$$

откуда

$$y' = \frac{dy}{dx} = -\frac{F'_x}{F'_y}.$$

Эта формула выражает производную неявной функции $y=\varphi(x)$ через частные производные заданной функции $F(x, y)$. Производная y' не существует при тех значениях x и y, при которых $F'_y = 0$. Но при этом, как было отмечено, мы и не можем гарантировать существования самой функции $\varphi(x)$.

II. Пусть теперь уравнение

$$F(x, y, z) = 0$$

определяет z как некоторую функцию $z = \varphi(x, y)$ независимых переменных x и y. Если в уравнение подставить вместо z функцию $\varphi(x, y)$, то получим тождество

$$F[x, y, \varphi(x, y)] = 0.$$

Следовательно, частные производные по x и по y от функции $F(x, y, z)$, где $z = \varphi(x, y)$, также должны быть равны нулю. Дифференцируя, найдем

$$F'_x + F'_z \frac{\partial z}{\partial x} = 0, \qquad F'_y + F'_z \frac{\partial z}{\partial y} = 0,$$

откуда

$$\frac{\partial z}{\partial x} = -\frac{F'_x}{F'_z}, \qquad \frac{\partial z}{\partial y} = -\frac{F'_y}{F'_z}.$$

Эти формулы выражают частные производные неявной функции $z = \varphi(x, y)$ через производные заданной функции $F(x, y, z)$. При $F_z=0$ эти формулы теряют смысл.

В общем случае, когда уравнение

$$F(x, y, z, \ldots, u) = 0$$

определяет u как некоторую функцию от $x, y, z, \ldots$, аналогично предыдущему найдем

$$\frac{\partial u}{\partial x} = -\frac{F'_x}{F'_u}, \quad \frac{\partial u}{\partial y} = -\frac{F'_y}{F'_u}, \quad \frac{\partial u}{\partial z} = -\frac{F'_z}{F'_u}, \quad \ldots$$

8. 4. Экстремумы функций нескольких переменных

Необходимые условия экстремума.

Начнем с рассмотрения функций двух переменных и дадим определение точки экстремума; оно совершенно аналогично соответствующему определению для функций одной переменной.

Определение. *Точка $P_0(x_0, y_0)$ называется точкой экстремума (максимума или минимума) функции $z = f(x, y)$, если $f(x_0, y_0)$ есть соответственно наибольшее или наименьшее значение функции $f(x,y)$ в некоторой окрестности точки P_0 (x_0,y_0).* При этом значение $f(x_0, y_0)$ называется *экстремальным значением* функции (соответственно *максимальным* или *минимальным).*

Говорят также,что функция $f(x, y)$ имеет в точке $P_0(x_0, y_0)$ экстремум (или достигает в точке P_0 экстремума).

Заметим, что в силу определения точка экстремума функции обязательно лежит внутри области определения функции, так что функция определена в некоторой (хотя бы и малой) области, содержащей эту точку. Вид поверхностей, изображающих функции в окрестности точек экстремума, показан на рис. 8.6.

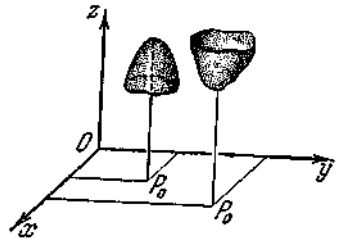


Рис. 8.6

Установим сначала необходимые условия, при которых функция $z=f(x, y)$ достигает в точке $P_0(x_0, y_0)$ экстремума; будем пока рассматривать дифференцируемые функции.

Необходимый признак экстремума. Если в точке $P_0(x_0, y_0)$ дифференцируемая функция $z=f(x,y)$ имеет экстремум, то ее частные производные в этой точке равны нулю:

$$\left(\frac{\partial z}{\partial x}\right)_{\substack{x=x_0\\y=y_0}}=0, \qquad \left(\frac{\partial z}{\partial y}\right)_{\substack{x=x_0\\y=y_0}}=0.$$

Доказательство. Допустим, что функция $z=f(x, y)$ имеет в точке $P_0\ (x_0, y_0)$ экстремум.

Согласно определению экстремума функция $z=f(x, y)$ при постоянном $y = y_0$, как функция одного x, достигает экстремума при $x = x_0$. Как известно, необходимым условием этого является обращение в нуль производной от функции $f(x, y_0)$ при $x = x_0$, т. е.

$$\left(\frac{\partial z}{\partial x}\right)_{\substack{x=x_0\\y=y_0}}=0.$$

Аналогично функция $z=f(x, y)$ при постоянном $x = x_0$, как функция одного y, достигает экстремума при $y=y_0$. Значит,

$$\left(\frac{\partial z}{\partial y}\right)_{\substack{x=x_0\\y=y_0}}=0,$$

что и требовалось доказать.

Точка $P_0(x_0, y_0)$, координаты которой обращают в нуль обе частные производные функции $z = f(x, y)$, называется *стационарной точкой функции* $f(x, y)$.

Уравнение касательной плоскости к поверхности $z=f(x, y)$:

$$z-z_0=\left(\frac{\partial z}{\partial x}\right)_0(x-x_0)+\left(\frac{\partial z}{\partial y}\right)_0(y-y_0)$$

для стационарной точки $P_0\ (x_0, y_0)$ принимает вид

$$z = z_0.$$

Следовательно, *необходимое условие достижения дифференцируемой функцией z=f(x, y) экстремума в точке $P_0(x_0, y_0)$ геометрически выражается в том, что касательная плоскость к поверхности — графику функции в соответствующей ее точке — параллельна плоскости независимых переменных.*

Для отыскания стационарных точек функции $z=f(x, y)$ нужно приравнять нулю обе ее частные производные

$$\frac{\partial z}{\partial x}=0, \qquad \frac{\partial z}{\partial y}=0 \qquad (*)$$

и решить полученную систему двух уравнений с двумя неизвестными.

Точками экстремума могут быть стационарные точки функции и точки, в которых функция недифференцируема.

Вполне аналогично определяется понятие экстремума функции любого числа независимых переменных

$$u = f(x, y, z, \ldots, t)$$

и устанавливаются необходимые условия экстремума. Именно:

Дифференцируемая функция n переменных может иметь экстремумы только при тех значениях x, y, z, . . ., t, при которых равны нулю все ее n частных производных первого порядка:

$$f'_x(x, y, z, \ldots, t)=0, \quad f'_y(x, y, z, \ldots, t)=0,$$
$$f'_z(x, y, z, \ldots, t)=0, \ \ldots, \ f'_t(x, y, z, \ldots, t)=0.$$

Эти равенствао бразуют систему n уравнений с n неизвестными.

Достаточные условия экстремума для функций двух переменных.

Так же как и для функций одной переменной, необходимый признак экстремума в случае многих переменных не является достаточным. Это значит, что из равенства нулю частных производных в данной точке вовсе не следует, что эта точка обязательно является точкой экстремума. Возьмем функцию $z=xy$. Ее частные производные $z_x = y$, $z_y = x$ равны нулю в начале координат, однако функция экстремума не достигает. В самом деле, функция $z = xy$, будучи равной нулю в начале координат, имеет в любой близости к началу координат как положительные значения (в первом и третьем координатных углах), так и отрицательные (во втором и четвертом координатных углах), и значит, нуль не является ни наибольшим, ни наименьшим значением этой функции.

Достаточные условия экстремума для функций нескольких переменных носят значительно более сложный характер, чем для

функций одной переменной. Мы приведем эти условия только для функций двух переменных.

Пусть точка $P_0\ (x_0,\ y_0)$ является стационарной точкой функции $z=f(x, y)$, т. е.

$$\left(\frac{\partial z}{\partial x}\right)_0 = 0 \text{ и } \left(\frac{\partial z}{\partial y}\right)_0 = 0.$$

Вычислим в точке P_0 значения вторых частных производных функции $f(x, y)$ и обозначим их для краткости буквами A, B и C:

$$A = \left(\frac{\partial^2 z}{\partial x^2}\right)_0, \quad B = \left(\frac{\partial^2 z}{\partial x\,\partial y}\right)_0, \quad C = \left(\frac{\partial^2 z}{\partial y^2}\right)_0.$$

Если $B^2 - AC < 0$, то функция $f(x, y)$ имеет в точке $P_0(x_0, y_0)$ экстремум: максимум при $A<0$ и $C<0$ и минимум при $A>0$ и $C>0$ (из условия $B^2 - AC<0$ следует, что A и C обязательно имеют одинаковые знаки).

Если $B^2 - AC>0$, то точка P_0 не является точкой экстремума.

Если $B^2 - AC = 0$, то никакого заключения о характере стационарной точки сделать нельзя и требуется дополнительное исследование.

Задачи о наибольших и наименьших значениях.

Пусть требуется найти наибольшее и наименьшее значение функции $z = f(x, y)$ в некоторой области (рассматриваемой вместе со своей границей). Если какое-либо из этих значений достигается функцией внутри области, то оно, очевидно, является экстремальным. Но может случиться, что наибольшее или наименьшее значение принимается функцией в некоторой точке, лежащей на границе области.

Из сказанного следует правило:

Для того чтобы найти наибольшее или наименьшее значение функции $z=f(x, y)$ в замкнутой области, нужно найти все максимумы или минимумы функции, достигаемые внутри этой области, а также наибольшее или наименьшее значение функции на границе области. Наибольшее из всех этих чисел и будет искомым наибольшим значением, а наименьшее — наименьшим.

Пример. Найдем точку на плоскости Oxy, сумма квадратов расстояний которой до трех точек $P_x(0, 0)$, $P_2(1, 0)$, $P_3\ (0, 1)$ имеет наименьшее значение, и точку треугольника с вершинами в P_1, P_2, P_3, сумма квадратов расстояний которой до вершин имеет наибольшее значение.

Возьмем на плоскости какую-нибудь точку $P(x, y)$. Сумма квадратов ее расстояний до заданных точек P_1, P_2, P_3 (обозначим эту сумму через z) выражается так:

$$z = x^2 + y^2 + (x-1)^2 + y^2 + x^2 + (y-1)^2$$

или

$$z = 3x^2 + 3y^2 - 2x - 2y + 2.$$

Первая часть задачи сводится к нахождению наименьшего значения этой функции во всей плоскости, вторая часть — к нахождению наибольшего значения функции при условии, что точка $P(x, y)$ принадлежит замкнутой области D, ограниченной треугольником $P_1P_2P_3$.

Найдем экстремумы функции $z = 3x^2 + 3y^2 - 2x - 2y + 2$. Из уравнений

$$\frac{\partial z}{\partial x} = 6x - 2 = 0 \quad \text{и} \quad \frac{\partial z}{\partial y} = 6y - 2 = 0$$

получаем

$$x = \frac{1}{3}, \quad y = \frac{1}{3}.$$

Значит, существует лишь одна стационарная точка $P\left(\frac{1}{3}, \frac{1}{3}\right)$. Во всей плоскости функция не имеет наибольшего значения, так как ясно, что существуют точки, для которых указанная сумма больше любого наперед заданного числа. А так как, с другой стороны, очевидно, что эта сумма должна достигать наименьшего значения, то именно стационарная точка P только и может быть точкой, в которой функция получает свое наименьшее значение $\left(= 1\frac{1}{3}\right)$. Между прочим, точка

$$P\left(\frac{1}{3}, \frac{1}{3}\right)$$

служит центром тяжести треугольника с вершинамив P_1, P_2, P_3.

Переходим ко второй части задачи. Вследствие того, что заданная функция не имеет максимума, ее наибольшим значением в области D является наибольшее из значений, принимаемых на границе, т. е. на сторонах треугольника.

На стороне P_1P_2 имеем $y = 0$ и, значит,

$$z = 3x^2 - 2x + 2;$$

эта функция достигает в интервале [0, 1] наибольшего значения (= 3) при x=1, т. е. в точке P_2.

На стороне P_1P_3 имеем $x = 0$ и, значит,

$$z = 3y^2 - 2y + 2;$$

наибольшее значение этой функции в интервале [0, 1] также равно 3 и достигается при y=1, т. е. в точке P_3.

Наконец, на стороне P_2P_3 имеем x-+y=1 и, значит,

$$z = 3x^2 + 3(1-x)^2 - 2x - 2(1-x) + 2 = 6x^2 - 6x + 3;$$

эта функция достигает в интервале [0, 1] наибольшего значения (= 3) при $x = 0$ и при $x=1$, т. е. наибольшее значение на стороне P_2P_3 функция z получает в тех же точках P_2 и P_3. Итак, во второй части задачи искомых точек треугольника имеется две: P_2 и P_3. Среди всех точек треугольника сумма квадратов расстояний этих точек от вершин P_1, P_2, P_3 имеет наибольшее значение.

Условные экстремумы.

При отыскании экстремумов функций нескольких переменных часто возникают задачи, связанные с так называемым *условным экстремумом*. Разъясним это понятие на примере функции двух переменных.

Пусть задана функция $z=f(x,y)$ и линия L на плоскости Oxy. Задача состоит в том, чтобы на линии L найти такую точку $P(x,y)$, в которой значение функции $f(x,y)$ является наибольшим или наименьшим по сравнению со значениями этой функции в точках линии L, находящихся вблизи точки P. Такие точки P называются *точками условного экстремума функции f(x, y) на линии L*. В отличие от обычной точки экстремума значение функции в точке условного экстремума сравнивается со значениями функции не во всех точках некоторой ее окрестности, а только в тех, которые лежат на линии L.

Совершенно ясно, что точка безусловного экстремума является и точкой условного экстремума для любой линии, проходящей через эту точку. Обратное же неверно: точка условного экстремума может и не быть точкой безусловного экстремума. Поясним сказанное простым примером. Графиком функции $z = \sqrt{1 - x^2 - y^2}$ является верхняя полусфера (рис. 18.7).

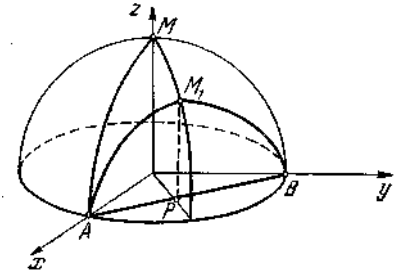


Рис. 18.7

Эта функция имеет максимум в начале координат; ему соответствует вершина M полусферы. Если линия L есть прямая, проходящая через точки A и B (ее уравнение $x+y-1=0$), то геометрически ясно, что для

точек этой линии наибольшее значение функции достигается в точке

$$P\left(\frac{1}{2}, \frac{1}{2}\right),$$

лежащей посередине между точками *A* и *B*. Это и есть точка условного экстремума (максимума) функции $z=\sqrt{1-x^2-y^2}$ на данной линии; ей соответствует точка M_1 на полусфере, и из рисунка видно, что ни о каком обычном экстремуме здесь не может быть речи.

Отметим, что в заключительной части задачи об отыскании наибольшего и наименьшего значений функции в замкнутой области нам приходится находить экстремальные значения функции на границе этой области, т. е. на какой-то линии, и тем самым решать задачу на условный экстремум.

Приступим теперь к практическому отысканию точек условного экстремума функции $z=f(x,y)$ при условии, что переменные z и y связаны уравнением $\varphi(x,y) = 0$. Это последнее соотношение будем называть *уравнением связи.* Если из уравнения связи y можно явно выразить через x: $y=\psi(x)$, то, подставляя в выражение функции $z=f(x, y)$ вместо y функцию $\psi(x)$, мы получим функцию одной переменной

$$z = f[x, \psi(x)] = \Phi(x).$$

Найдя значения x, при которых эта функция достигает экстремума, и определив затем из уравнения связи соответствующие им значения y, мы и получим искомые точки условного экстремума.

Так в вышеприведенном примере из уравнения связи $x+y - 1=0$ имеем $y=1-x$. Отсюда

$$z = \sqrt{1-x^2-(1-x)^2} = \sqrt{2x-x^2}.$$

Легко проверить, что z достигает максимума при $x = 0{,}5$; но тогда из уравнения связи $y = 0{,}5$, и мы получаем как раз точку P, найденную из геометрических соображений.

Очень просто решается задача на условный экстремум и тогда, когда уравнение связи можно представить параметрическими уравнениями: $x= x(t)$, $y=y(t)$. Подставляя выражения для x и y в данную функцию, снова приходим к задаче отыскания экстремума функции одной переменной.

Если уравнение связи имеет более сложный вид и нам не удается ни явно выразить одну переменную через другую, ни заменить его параметрическими уравнениями, то задача отыскания условного экстремума становится более трудной. Будем по-прежнему считать, что в выражении функции $z=f(x, y)$ переменная y является функцией от

x, определенной неявно уравнением связи *y(x, y)*=0. Полная производная от функции *z = f(x, y)* по *x* равна

$$\frac{dz}{dx} = f'_x(x, y) + f'_y(x, y)\frac{dy}{dx} = f'_x(x, y) - \frac{\varphi'_x(x, y)}{\varphi'_y(x, y)} f'_y(x, y),$$

где производная *y'* найдена по правилу дифференцирования неявной функции. В точках условного экстремума найденная полная производная должна равняться нулю; это дает одно уравнение, связывающее *x* и *y.* Так как они должны удовлетворять еще и уравнению связи, то мы получаем систему двух уравнений с двумя неизвестными

$$f'_x - \frac{\varphi'_x}{\varphi'_y} f'_y = 0, \qquad \varphi(x, y) = 0.$$

Преобразуем эту систему к гораздо более удобной, записав первое уравнение в виде пропорции и введя новую вспомогательную неизвестную λ:

$$\frac{f'_x}{\varphi'_x} = \frac{f'_y}{\varphi'_y} = -\lambda \qquad (*)$$

(знак минус перед λ поставлен для удобства). От этих равенств легко перейти к следующей системе:

$$f'_x(x, y) + \lambda\varphi'_x(x, y) = 0, \quad f'_y(x, y) + \lambda\varphi'_y(x, y) = 0, \qquad (**)$$

которая вместе с уравнением связи *φ(x, y)* = 0 образует систему трех уравнений с неизвестными *x, y* и λ.

Уравнения (**) легче всего запомнить при помощи следующего правила: *для того чтобы найти точки, которые могут быть точками условного экстремума функции z=f(x,y) при уравнении связи φ(x, y)*= 0, *нужно образовать вспомогательную функцию*

$$\Phi(x, y) = f(x, y) + \lambda\varphi(x, y),$$

где λ *— некоторая постоянная, и составить уравнения для отыскания точек экстремума этой функции.*

Указанная система уравнений доставляет, как обычно, только необходимые условия, т. е. не всякая пара значений *x* и *y,* удовлетворяющая этой системе, обязательно является точкой условного экстремума. Достаточные условия для точек условного экстремума мы приводить не будем; очень часто конкретное содержание задачи само подсказывает, чем является найденная точка. Описанный прием решения задач на условный экстремум называется *методом множителей Лагранжа.*

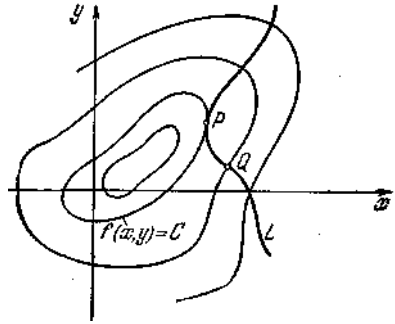


Рис. 8.8.

Метод Лагранжа имеет наглядный геометрический смысл, который мы сейчас и выясним. Предположим, что на рис. 8.8 изображены линии уровня функции $z=f(x, y)$ и линия L, на которой отыскиваются точки условного экстремума. Если в точке Q линия L пересекает линию уровня, то эта точка не может быть точкой условного экстремума, так как по одну сторону от линии уровня функция $f(x,y)$ принимает большие значения, а по другую — меньшие. Если же в точке P линия L не пересекает соответствующую линию уровня и, значит, в некоторой окрестности этой точки лежит по одну сторону от линии уровня, то точка P будет как раз являться точкой условного экстремума. В такой точке линия L и линия уровня $f(x, y) = C$ касаются друг друга (мы предполагаем, что линии гладкие) и угловые коэффициенты касательных к ним должны быть равны. Из уравнения связи $\varphi(x, y) = 0$ имеем $y' = -\frac{\varphi'_x}{\varphi'_y}$, а из уравнения линии уровня $y' = -\frac{f'_x}{f'_y}$. Приравнивая производные и произведя простейшее преобразование, мы и получим уравнения (*).

Пример. Найти наибольшее значение функции $z = xy$, если x и y положительны и подчиняются уравнению связи $\frac{x^2}{8} + \frac{y^2}{2} = 1$.

Здесь уравнение связи простое (оно представляет эллипс) и можно было бы сразу прийти к отысканию экстремума функции одной переменной, выразив y через x или взяв параметрические уравнения эллипса. Рекомендуем читателю сделать это самостоятельно; мы же для иллюстрации решим пример методом Лагранжа. Составим вспомогательную функцию

$$\Phi(x, y) = xy + \lambda\left(\frac{x^2}{8} + \frac{y^2}{2}\right).$$

(Постоянное слагаемое в левой части уравнения связи

$$\varphi(x, y) \equiv \frac{x^2}{8} + \frac{y^2}{2} - 1 = 0$$

при дифференцировании дает нуль, и поэтому мы его просто не пишем.) Приравнивая частные производные по x и по y нулю, получим

$$\frac{\partial \Phi}{\partial x} = y + \frac{\lambda x}{4} = 0, \qquad \frac{\partial \Phi}{\partial y} = x + \lambda y = 0.$$

Исключая λ, приходим к уравнению $4y^2 - x^2 = 0$, решая которое совместно с уравнением связи находим $x = 2$ и $y = 1$ (по условию задачи x и y положительны).

Функция $z = xy$ при рассматриваемых значениях x и y положительна и в точках пересечения эллипса с осями координат $(2\sqrt{2}, 0)$ и $(0, \sqrt{2})$ равна нулю. Поэтому найденная единственная точка P (2, 1) будет точкой условного максимума; в этой точке $z_{max}=2$. Легко проверить, что в точке P эллипс касается линии уровня функции $z = xy$, проходящей через эту точку, т. е. гиперболы $xy = 2$.

Задачи на условный экстремум для функций трех переменных допускают большее разнообразие. Пусть задана функция $u = f(x, y, z)$ и переменные x, y и z связаны одним уравнением $\varphi(x, y, z) = 0$; это есть уравнение некоторой поверхности. Тогда точка условного экстремума является такой точкой поверхности, в которой функция принимает наибольшее или наименьшее значение по сравнению с ее значениями во всех близлежащих точках этой же поверхности.

Можно искать условный экстремум функции $f(x, y, z)$ и при двух уравнениях связи: $\varphi_1(x, y, z) = 0$ и $\varphi_2(x, y, z) = 0$. Эти уравнения определяют линию в пространстве. Таким образом, задача сводится к отысканию такой точки линии, в которой функция принимает экстремальное значение, причем сравниваются значения функции только в точках рассматриваемой линии.

Метод множителей Лагранжа в случае двух уравнений связи применяется следующим образом: строим вспомогательную функцию

$$\Phi(x, y, z) = f(x, y, z) + \lambda_1\varphi_1(x, y, z) + \lambda_2\varphi_2(x, y, z),$$

где λ_1 и λ_2 — новые дополнительные неизвестные, и составляем систему уравнений для отыскания экстремумов этой функции

$$\frac{\partial \Phi}{\partial x} = \frac{\partial f}{\partial x} + \lambda_1 \frac{\partial \varphi_1}{\partial x} + \lambda_2 \frac{\partial \varphi_2}{\partial x} = 0,$$
$$\frac{\partial \Phi}{\partial y} = \frac{\partial f}{\partial y} + \lambda_1 \frac{\partial \varphi_1}{\partial y} + \lambda_2 \frac{\partial \varphi_2}{\partial y} = 0,$$
$$\frac{\partial \Phi}{\partial z} = \frac{\partial f}{\partial z} + \lambda_1 \frac{\partial \varphi_1}{\partial z} + \lambda_2 \frac{\partial \varphi_2}{\partial z} = 0.$$

Добавляя сюда два уравнения связи, получаем систему пяти уравнений с пятью неизвестными $x, y, z, \lambda_1, \lambda_2$. Искомыми точками условного экстремума могут быть только те, координаты x, y, z которых являются решениями этой системы.

В самом общем виде задача ставится так: требуется найти экстремумы функции n переменных $u = f(x, y, z, ..., t)$ при условии, что эти переменные подчинены m уравнениям связи ($m < n$)

$$\varphi_1(x, y, z, \ldots, t) = 0, \quad \varphi_2(x, y, z, \ldots, t) = 0, \ \ldots, \ \varphi_m(x, y, z, \ldots, t) = 0.$$

Вспомогательная функция зависит от n переменных и содержит m дополнительных неизвестных

$$\Phi = f + \lambda_1 \varphi_1 + \lambda_2 \varphi_2 + \ldots + \lambda_m \varphi_m.$$

Уравнения для отыскания точек экстремума этой функции и уравнения связи составят систему $m+n$ уравнений, из которой определяются координаты $x, y, z.....t$ возможных точек условного экстремума.

Примеры. 1) Найти прямоугольный параллелепипед наибольшего объема, если его полная поверхность равна заданной величине S.

Обозначим стороны параллелепипеда через x, y, z. Тогда требуется найти наибольшее значение функции $V=xyz$ при условии, что $xy + yz + zx = \frac{S}{2}$.

Вспомогательная функция $\Phi(x; y, z) = xyz + \lambda(xy + yz + zx)$. Уравнения для отыскания точек экстремума этой функции имеют вид

$$yz + \lambda(y + z) = 0, \quad xz + \lambda(x + z) = 0, \quad xy + \lambda(x + y) = 0.$$

Вычитая эти уравнения друг из друга, получим

$$(z + \lambda)(y - x) = 0, \ (x + \lambda)(z - y) = 0, \ (y + \lambda)(x - z) = 0.$$

Отсюда ясно, что $x=y = z$, т. е. что искомый параллелепипед — куб. Размеры его определим из уравнения связи:

$$x = y = z = \sqrt{\frac{S}{6}} \quad \text{и} \quad V = \frac{S\sqrt{S}}{6\sqrt{6}}.$$

2) Найти наибольшее расстояние от начала координат до точек линии пересечения параболоида вращения $z=x^2+y^2$ с плоскостью $x+2y$ — z= 0. (Наименьшее расстояние равно нулю, так как линия пересечения проходит через начало координат.)

Расстояние r от начала координат до точки (x, y, z) равно

$$\sqrt{x^2+y^2+z^2}.$$

Наибольшие значения корня и подкоренного выражения достигаются в одной и той же точке; поэтому будем искать условный максимум функции $u=x^2+y^2+z^2$ при двух уравнениях связи

$$\varphi_1(x, y, z) \equiv x^2+y^2-z=0 \quad \text{и} \quad \varphi_2(x, y, z) \equiv x+2y-z=0.$$

Составляя вспомогательную функцию

$$\Phi = x^2+y^2+z^2+\lambda_1(x^2+y^2-z)+\lambda_2(x+2y-z)$$

и приравнивая нулю ее частные производные, получим

$$2x+2\lambda_1 x+\lambda_2=0, \quad 2y+2\lambda_1 y+2\lambda_2=0, \quad 2z-\lambda_1-\lambda_2=0.$$

Выразим отсюда координаты x, y, z через вспомогательные неизвестные

$$x=-\frac{\lambda_2}{2(1+\lambda_1)}, \quad y=-\frac{\lambda_2}{1+\lambda_1}, \quad z=\frac{\lambda_1+\lambda_2}{2}.$$

Подставляя в уравнения связи

$$\frac{5}{4}\left(\frac{\lambda_2}{1+\lambda_1}\right)^2=\frac{\lambda_1+\lambda_2}{2}, \quad -\frac{5}{2}\left(\frac{\lambda_2}{1+\lambda_1}\right)=\frac{\lambda_1+\lambda_2}{2}$$

и приравнивая левые части обоих уравнений, получаем квадратное уравнение относительно $\frac{\lambda_2}{1+\lambda_1}$. Его первый корень приводит к значениям $\lambda_2 = 0$ и $\lambda_1 = 0$; им соответствует начало координат — точка минимума. Второй корень равен $\frac{\lambda_2}{1+\lambda_1}=-2$. Даже не отыскивая λ_1 и λ_2, находим $x=1$, $y=2$ и $z = 5$. Ясно, что эта точка наиболее удалена от начала координат и $r_{max}=\sqrt{30}$.

8. 5. Скалярное поле

Скалярное поле. Поверхности уровня.

Предположим, что в каждой точке P некоторой области D нам задано значение скалярной физической величины u, т. е. такой величины, которая полностью характеризуется своим числовым значением. Например, это может быть температура точек неравномерно нагретого тела, плотность распределения электрических зарядов в изолированном наэлектризованном теле, потенциал электрического поля и т. д. При этом u называется *скалярной функцией точки;* записывается это так: $u = u(P)$.

Область D, в которой определена функция $u(P)$, может совпадать со всем пространством, а может являться некоторой его частью.

Определение. *Если в области D задана скалярная функция точки u(P), то говорят, что в этой области задано скалярное поле.*

Мы будем считать, что скалярное поле *стационарное,* т. е. что величина *u(P)* не зависит от времени *t.* В будущем нам придется сталкиваться и с нестационарными полями. Тогда величина *u* будет зависеть не только от точки *P,* но и от времени *t.*

Если физическая величина векторная, то ей будет соответствовать *векторное поле,* например силовое поле, электрическое поле напряженности, магнитное поле и др.

Если скалярное поле отнесено к системе координат *Oxyz,* то задание точки *P* равносильно заданию ее координат *x, y, z;* и тогда функцию *u(P)* можно записать в обычном виде функции трех переменных: *u (x, y, z).* Мы пришли, таким образом, к физическому толкованию функций трех переменных.

Определение. Поверхностью уровня скалярного поля называется геометрическое место точек, в которых функция *u* принимает постоянное значение, т. е.

$$u(x,y,z) = C.$$

В курсе физики при рассмотрении поля потенциала поверхности уровня называют обычно *эквипотенциальными* поверхностями (т. е. поверхностями равного потенциала).

Уравнение поверхности уровня, проходящей через данную точку $M_0 (x_0, y_0, z_0)$, записывается так:

$$u(x, y, z) = u(x_0, y_0, z_0).$$

Если в частном случае скалярное поле плоское, т. е. мы изучаем распределение значений физической величины в какой-то плоской области, то функция *u* зависит от двух переменных, например *x* и *y.* Линиями уровня этого поля будут линии уровня функции *u (x, y)*:

$$u(x, y) = C.$$

Производная по направлению.

Важной характеристикой скалярного поля является скорость изменения поля в заданном направлении. Пусть задано скалярное поле, т. е. задана функция *u(x, y, z).* Возьмем точку *P(x, y, z)* и какой-нибудь луч λ, из нее выходящий. Направление этого луча зададим углами α, β, γ, которые он образует с направлениями осей *Ox, Oy, Oz* (рис. 8.9).

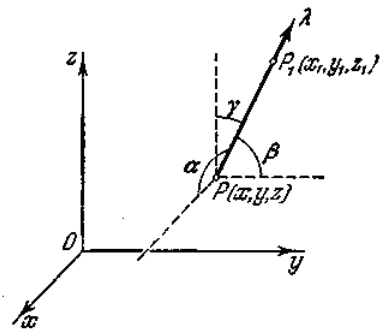


Рис. 8.9

Если $\mathbf{e}_\lambda$ единичный вектор, направленный по лучу λ, то его проекциями будут направляющие косинусы

$$\mathbf{e}_\lambda \{\cos\alpha,\ \cos\beta,\ \cos\gamma\}.$$

Пусть точка $P_1(x_1, y_1, z_1)$ лежит на луче λ; расстояние PP_1 обозначим через ρ. Проекции вектора $\overline{PP_1}$ на оси координат будут, с одной стороны, равны ρ cos α, ρ cos β, ρ cos γ, а с другой стороны, — разностям $x_1 - x$, $y_1 - y$ и $z_1 - z$. Следовательно,

$$x_1 = x + \rho\cos\alpha,\ y_1 = y + \rho\cos\beta,\ z_1 = z + \rho\cos\gamma.$$

Рассмотрим теперь приращение функции u при переходе из точки P в точку P_1:

$$u(P_1) - u(P) = u(x + \rho\cos\alpha,\ y + \rho\cos\beta,\ z + \rho\cos\gamma) - u(x,\ y,\ z).$$

Если точка P_1 будет изменять свое положение на луче λ, то в выражении для разности $u(P_1) - u(P)$ будет меняться только величина ρ. Составим отношение

$$\frac{u(P_1) - u(P)}{\rho}$$

и перейдем к пределу при $\rho \to 0$, предполагая, что этот предел существует.

Определение. Предел

$$\lim_{\rho\to 0}\frac{u(P_1) - u(P)}{\rho} = \lim_{\rho\to 0}\frac{u(x+\rho\cos\alpha,\ y+\rho\cos\beta,\ z+\rho\cos\gamma) - u(x,y,z)}{\rho} \qquad (*)$$

называется производной от функции $u(x, y, z)$ по направлению λ в точке P.

Этот предел будем обозначать символом $\dfrac{\partial u}{\partial \lambda}$ или $u'_\lambda(x, y, z)$.

Величина его зависит от выбранной точки $P(x, y, z)$ и от направления луча λ, т. е. от α, β и γ.

Если точка P фиксирована, то величина производной u'_λ будет зависеть только от направления луча λ.

Из определения производной по направлению следует, что если направление λ совпадает с положительным направлением оси Ox, т. е.

$$\alpha = 0, \ \beta = \gamma = \frac{\pi}{2},$$

то предел (*) будет просто равен частной производной от функции $u(x, y, z)$ по x:

$$\lim_{\rho \to 0} \frac{u(x+\rho, y, z) - u(x, y, z)}{\rho} = \frac{\partial u}{\partial x}.$$

Аналогичную картину мы получим, если направление λ будет совпадать с направлениями осей Oy и Oz.

Подобно тому как частные производные

$$u'_x, \ u'_y \text{ и } u'_z$$

характеризуют скорость изменения функции u в направлении осей координат, так и производная по направлению u'_λ будет являться скоростью изменения функции $u(x, y, z)$ в точке P по направлению луча λ. Абсолютная величина производной u'_λ по направлению λ определяет величину скорости, а знак производной — характер изменения функции u (возрастание или убывание).

Вычисление производной по направлению производится при помощи следующей теоремы:

Теорема. Если функция $u(x, y, z)$ дифференцируема, то ее производная u'_λ по любому направлению λ существует и равна

$$\frac{\partial u}{\partial \lambda} = \frac{\partial u}{\partial x}\cos\alpha + \frac{\partial u}{\partial y}\cos\beta + \frac{\partial u}{\partial z}\cos\gamma, \qquad (**)$$

где $\cos\alpha$, $\cos\beta$, $\cos\gamma$ — направляющие косинусы луча b.

Доказательство. Так как функция $u(x, y, z)$ дифференцируема, то ее полное приращение можно записать в виде

$$\Delta u = u(x+\Delta x, \ y+\Delta y, \ z+\Delta z) - u(x, \ y, \ z) =$$
$$= \frac{\partial u}{\partial x}\Delta x + \frac{\partial u}{\partial y}\Delta y + \frac{\partial y}{\partial z}\Delta z + \varepsilon,$$

где ε — бесконечно малая величина более высокого порядка, чем

$$\rho = \sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2}.$$

Полагая $\Delta x = \rho\cos\alpha$, $\Delta y = \rho\cos\beta$, $\Delta z = \rho\cos\gamma$, представим разность $u(P_1) - u(P)$ в виде

$$u(P_1) - u(P) = \frac{\partial u}{\partial x}\rho\cos\alpha + \frac{\partial u}{\partial y}\rho\cos\beta + \frac{\partial u}{\partial z}\rho\cos\gamma + \varepsilon,$$

причем $\frac{\varepsilon}{\rho} \longrightarrow 0$ при $\rho \to 0$. Отношение $\frac{u(P_1)-u(P)}{\rho}$ будет равно

$$\frac{u(P_1)-u(P)}{\rho} = \frac{\partial u}{\partial x}\cos\alpha + \frac{\partial u}{\partial y}\cos\beta + \frac{\partial u}{\partial z}\cos\gamma + \frac{\varepsilon}{\rho}.$$

Так как значения частных производных $\frac{\partial u}{\partial x}$, $\frac{\partial u}{\partial y}$ и $\frac{\partial u}{\partial z}$ в точке P, а также α, β и γ от ρ не зависят, то, переходя к пределу при $\rho \to 0$, получим

$$\frac{\partial u}{\partial \lambda} = \lim_{\rho \to 0} \frac{u(P_1)-u(P)}{\rho} = \frac{\partial u}{\partial x}\cos\alpha + \frac{\partial u}{\partial y}\cos\beta + \frac{\partial u}{\partial z}\cos\gamma,$$

что и требовалось доказать.

Из этой формулы непосредственно следует сделанное выше замечание, что если направление λ совпадает с положительным направлением одной из осей координат, то производная по этому направлению равна соответствующей частной производной, например, если $\alpha = 0$, $\beta = \gamma = \frac{\pi}{2}$, то $\frac{\partial u}{\partial \lambda} = \frac{\partial u}{\partial x}$.

Из формулы (**) видно, что производная по направлению λ', противоположному направлению λ, равна *производной по направлению λ, взятой с обратным знаком.* Действительно, при перемене направления углы α, β и γ изменятся на π и

$$\frac{\partial u}{\partial \lambda'} = \frac{\partial u}{\partial x}\cos(\alpha+\pi) + \frac{\partial u}{\partial y}\cos(\beta+\pi) + \frac{\partial u}{\partial z}\cos(\gamma+\pi) = -\frac{\partial u}{\partial \lambda}.$$

Это означает, что при перемене направления на противоположное абсолютная величина скорости изменения функции u не меняется, а изменяется только характер ее изменения; если, например, в направлении λ функция возрастает, то в направлении λ' она убывает, и наоборот.

Пример. Дана функция $u = xyz$. Найдем ее производную в точке $P(5, 1, 2)$ в направлении, идущем от этой точки к точке $Q(7, -1, 3)$. Находим частные производные функции $u = xyz$

$$\frac{\partial u}{\partial x} = yz, \qquad \frac{\partial u}{\partial y} = xz, \qquad \frac{\partial u}{\partial z} = xy$$

и вычисляем их значения в точке P:

$$\left(\frac{\partial u}{\partial x}\right)_P = 2, \qquad \left(\frac{\partial u}{\partial y}\right)_P = 10, \qquad \left(\frac{\partial u}{\partial z}\right)_P = 5.$$

Так как проекции вектора $\overline{PQ}$ равны 2, —2 и 1, то его направляющими косинусами будут

$$\cos\alpha = \frac{2}{\sqrt{2^2+(-2)^2+1^2}} = \frac{2}{3}, \qquad \cos\beta = \frac{-2}{3}, \qquad \cos\gamma = \frac{1}{3}.$$

Следовательно,

$$\frac{\partial u}{\partial \lambda} = 2\cdot\frac{2}{3} + 10\cdot\left(-\frac{2}{3}\right) + 5\cdot\frac{1}{3} = -\frac{11}{3}.$$

Знак минус указывает, что в данном направлении функция *u* убывает.

Если поле плоское, то направление луча λ вполне определяется углом α его наклона к оси абсцисс. Формулу для производной по направлению в случае плоского поля можно получить из общей формулы, положив $\gamma = \frac{\pi}{2}$ и $\beta = \frac{\pi}{2} - \alpha$.

Тогда

$$\frac{\partial u}{\partial \lambda} = \frac{\partial u}{\partial x}\cos\alpha + \frac{\partial u}{\partial y}\sin\alpha.$$

Если $\alpha = 0$, то $\frac{\partial u}{\partial \lambda} = \frac{\partial u}{\partial x}$, а если $\alpha = \frac{\pi}{2}$, то $\frac{\partial u}{\partial \lambda} = \frac{\partial u}{\partial y}$.

Градиент.

Рассмотрим снова формулу для производной по направлению

$$\frac{\partial u}{\partial \lambda} = \frac{\partial u}{\partial x}\cos\alpha + \frac{\partial u}{\partial y}\cos\beta + \frac{\partial u}{\partial z}\cos\gamma.$$

Вторые множители в каждом слагаемом являются, как мы уже отмечали, проекциями единичного вектора e_λ, направленного по лучу λ:

$$\mathbf{e}_\lambda\{\cos\alpha,\ \cos\beta,\ \cos\gamma\}.$$

Возьмем теперь вектор, проекциями которого на оси координат будут служить значения частных производных

$$\frac{\partial u}{\partial x},\ \frac{\partial u}{\partial y}\ \text{и}\ \frac{\partial u}{\partial z}$$

в выбранной точке *P(x, y, z).* Назовем этот вектор *градиентом* функции *u (x, y, z)* и будем обозначать его символами grad *u* или ∇u,

Определение. *Градиентом* функции *u(x, y, z)* называется вектор, проекциями которого служат значения частных производных этой функции, т. е.

$$\mathbf{grad}\ u = \frac{\partial u}{\partial x}\mathbf{i} + \frac{\partial u}{\partial y}\mathbf{j} + \frac{\partial u}{\partial z}\mathbf{k}.$$

Подчеркнем, что проекции градиента зависят от выбора точки *P(x, y, z)* и изменяются с изменением координат этой точки. Таким образом, каждой точке скалярного поля, определяемого функцией поля *u(x, y, z)*, соответствует определенный вектор — градиент этой функции. Отметим, что градиент линейной функции $u = ax+by+cz+d$ есть постоянный вектор: $\mathbf{grad}\ u = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$.

Пользуясь определением градиента, формуле для производной по направлению можно придать такой вид:

$$\frac{\partial u}{\partial \lambda} = \mathbf{grad}\, u \cdot \mathbf{e}_\lambda. \qquad (*)$$

Следовательно:

Производная функции по данному направлению равна скалярному произведению градиента функции на единичный вектор этого направления.

Так как скалярное произведение равно модулю одного вектора, умноженному на проекцию другого вектора на направление первого, то можно еще сказать, что

Производная функции по данному направению равна проекции градиента функции на направление дифференцирования, т. е.

$$\frac{\partial u}{\partial \lambda} = |\mathbf{grad}\, u|\ \cos\varphi,$$

где φ — угол между вектором grad ии лучом λ (рис. 8.10).

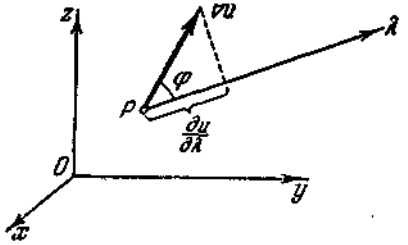


Рис. 8.10

Отсюда сразу следует, что производная по направлению достигает наибольшего значения, когда cos φ =l, т. е. при φ = 0. Это наибольшее значение равно |grad *u*|.

Итак, | grad *u*| есть наибольшее возможное значение производной u'_λ в данной точке *P,* а направление grad *u* совпадает с направлением луча, выходящего из точки *P,* вдоль которого функция меняется быстрее всего, т. е. направление градиента есть направление наискорейшего

возрастания функции. Ясно, что в противоположном направлении функция *и* будет быстрее всего убывать.

Докажем теперь теорему, устанавливающую связь между направлением градиента функции и поверхностями уровня скалярного поля.

Теорема. Направление градиента функции *u(x, y, z)* в каждой точке совпадает с направлением нормали к поверхности уровня скалярного поля, проходящей через эту точку (рис. 8.11).

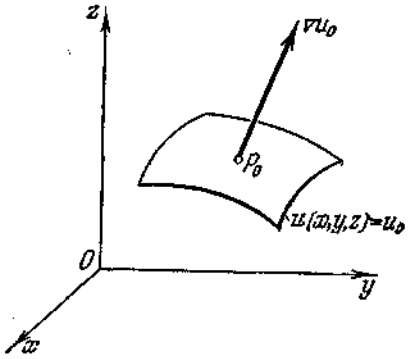


Рис. 8.11

Доказательство. Выберем произвольную точку $P_0(x_0, y_0, z_0)$. Уравнение поверхности уровня, проходящей через точку P_0, запишется в виде $u(x, y, z) = u_0$, где $u_0 = u(x_0, y_0, z_0)$.

Составим уравнение нормали к этой поверхности в точке P_0:

$$\frac{x-x_0}{\left(\frac{\partial u}{\partial x}\right)_0}=\frac{y-y_0}{\left(\frac{\partial u}{\partial y}\right)_0}=\frac{z-z_0}{\left(\frac{\partial u}{\partial z}\right)_0}.$$

Отсюда и следует, что направляющий вектор нормали, имеющий проекции

$$\left(\frac{\partial u}{\partial x}\right)_0,\ \left(\frac{\partial u}{\partial y}\right)_0,\ \left(\frac{\partial u}{\partial z}\right)_0,$$

является градиентом функции $u(x, y, z)$ в точке u_0, что и требовалось доказать.

Таким образом, градиент в каждой точке перпендикулярен касательной плоскости к поверхности уровня, проходящей через данную точку, т. е. его проекция на эту плоскость равна нулю. Следовательно:

Производная по любому направлению, касательному к поверхности уровня, проходящей через данную точку, равна нулю.

Укажем теперь некоторые свойства градиента функции, часто облегчающие его вычисление.

1) $\operatorname{grad}(u_1 + u_2) = \operatorname{grad} u_1 + \operatorname{grad} u_2$.

2) $\operatorname{grad} Cu_1 = C \operatorname{grad} u_1$, где C — постоянная.

3) $\operatorname{grad} u_1 u_2 = u_2 \operatorname{grad} u_1 + u_1 \operatorname{grad} u_2$. В самом деле,

$$\operatorname{grad} u_1 u_2 = \frac{\partial (u_1 u_2)}{\partial x}\,\mathbf{i} + \frac{\partial (u_1 u_2)}{\partial y}\,\mathbf{j} + \frac{\partial (u_1 u_2)}{\partial z}\,\mathbf{k} =$$

$$= u_2\left(\frac{\partial u_1}{\partial x}\,\mathbf{i} + \frac{\partial u_1}{\partial y}\,\mathbf{j} + \frac{\partial u_1}{\partial z}\,\mathbf{k}\right) + u_1\left(\frac{\partial u_2}{\partial x}\,\mathbf{i} + \frac{\partial u_2}{\partial y}\,\mathbf{j} + \frac{\partial u_2}{\partial z}\,\mathbf{k}\right) =$$

$$= u_2 \operatorname{grad} u_1 + u_1 \operatorname{grad} u_2.$$

4) $\operatorname{grad} f(u) = f'(u) \operatorname{grad} u$. Действительно,

$$\operatorname{grad} f(u) = \frac{\partial}{\partial x}[f(u)]\,\mathbf{i} + \frac{\partial}{\partial y}[f(u)]\,\mathbf{j} + \frac{\partial}{\partial z}[f(u)]\,\mathbf{k} =$$

$$= f'(u)\frac{\partial u}{\partial x}\,\mathbf{i} + f'(u)\frac{\partial u}{\partial y}\,\mathbf{j} + f'(u)\frac{\partial u}{\partial z}\,\mathbf{k} = f'(u) \operatorname{grad} u.$$

Перечисленные свойства градиента показывают, что правила его отыскания совпадают с правилами отыскания производной функции.

Примеры. 1) Пусть $r = \sqrt{x^2 + y^2 + z^2}$ - расстояние от точки до начала координат. Тогда

$$\operatorname{grad} r = \frac{\partial r}{\partial x}\,\mathbf{i} + \frac{\partial r}{\partial y}\,\mathbf{j} + \frac{\partial r}{\partial z}\,\mathbf{k} = \frac{x\mathbf{i} + y\mathbf{j} + z\mathbf{k}}{r} = \frac{\mathbf{r}}{r}.$$

grad r направлен по радиусу-вектору $\mathbf{r}$, и модуль его равен единице.

2) Пусть скалярное поле определено функцией $\frac{q}{r}$, где r определено в примере 1. Тогда по свойству 4)

$$\operatorname{grad}\frac{q}{r} = -\frac{q}{r^2}\operatorname{grad} r = -\frac{q}{r^2}\frac{\mathbf{r}}{r}.$$

В плоском поле $u = u(x, y)$ градиент

$$\operatorname{grad} u = \frac{\partial u}{\partial x}\,\mathbf{i} + \frac{\partial u}{\partial y}\,\mathbf{j}$$

лежит в плоскости Oxy и перпендикулярен к линии уровня.

Если в плоском поле построена достаточно густая сетка линий уровня (рис. 8.12), то можно с некоторым приближением графически определить модуль и направление градиента.

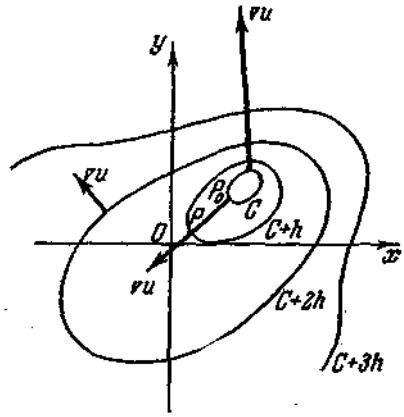


Рис. 8.12

Направление градиента будет перпендикулярно к линии уровня. Производная в этом направлении будет при достаточно малом h приближенно равна

$$\frac{\partial u}{\partial \lambda} \approx \frac{u(P) - u(P_0)}{P_0 P} = \frac{h}{P_0 P},$$

где P_0 - точка линии уровня $u(x, y)=C$, а P— точка линии уровня $u(x, y) = C+h$. Величина h известна, а длина отрезка P_0P может быть измерена на чертеже как расстояние по нормали между соседними линиями уровня. Производная же по направлению градиента равна его модулю, и поэтому

$$|\operatorname{grad} u| \approx \frac{h}{P_0 P}.$$

Построение grad u изображено на рис. 8.12.

9. Дифференциальные уравнения

В этом разделе кратко напоминаются основные сведения из теории обыкновенных дифференциальных уравнений, описываются возможные методы приближенного решения задачи Коши и постановка задачи о разностных методах решения задачи Коши.

Несколько слов о курсе "Обыкновенные дифференциальные уравнения"

Если не оговорено противное, мы будем рассматривать *систему обыкновенных дифференциальных уравнений первого порядка, разрешенных относительно старшей производной*, т. е. систему вида

$$x' = f(t, x), \qquad \text{(E)}$$

где f: $\mathbf{R}\times\mathbf{R}^m \to \mathbf{R}^m$ — непрерывная функция.

На протяжении всей работы $\mathbf{R}^m$ — m-мерное линейное вещественное пространство. Будем всегда считать, что в $\mathbf{R}^m$ фиксирован базис и отождествляем $\mathbf{R}^m$ с координатным m-мерным вещественным пространством: точки из $\mathbf{R}^m$ — это упорядоченные наборы m вещественных чисел $x = (x_1, ..., x_m)$. Таким образом, (E) — это сокращенная форма записи системы дифференциальных уравнений

$x'_1 = f_1(t, x_1, ..., x_m)$,

...

$x'_m = f_m(t, x_1, ..., x_m)$,

в которой $x = (x_1, ..., x_m)$, а $f(t, x) = (f_1(t, x), ..., f_m(t, x)) = (f_1(t, x_1, ..., x_m), ..., f_m(t, x_1, ..., x_m))$. Кроме того, предполагается, что в $\mathbf{R}^m$ фиксирована некоторая норма $\|\cdot\|$.

Решением уравнения (E) на промежутке $[a, b]$ называется функция φ: $[a, b] \to \mathbf{R}^m$, обращающая (E) в тождество на $[a, b]$:

$\varphi'(t) \equiv f(t, \varphi(t)), \quad t \in [a, b]$.

График решения (лежащий, по определению, в *расширенном фазовом пространстве* $\mathbf{R}\times\mathbf{R}^m$) называется *интегральной кривой* I_φ. Проекция интегральной кривой на *фазовое пространство* $\mathbf{R}^m$ параллельно $\mathbf{R}$ называется *траекторией* T_φ. Здесь же отметим, что независимую переменную t мы будем трактовать как время.

В общей ситуации, уравнение (E) имеет бесконечное множество решений (точнее, m-параметрическое семейство решений). Для того чтобы выделить одно из них, нужны дополнительные условия (уравнения). Такие условия могут быть различными. Мы будем рассматривать только один вид дополнительных условий, так называемые *начальные условия* — требование, чтобы решение в заданной точке принимало заданное значение:

$$x(t_0) = x^0. \qquad \text{(C)}$$

Задача о нахождении решения уравнения (E), удовлетворяющего начальному условию (C), называется *задачей Коши*. Обозначение "задача Коши (E) – (C)" будет универсальным, т.е. действовать на протяжении всей книги; универсальные обозначения и предположения будут заключаться в рамку:

Задача Коши (E) – (C).

Говорят, что функция f удовлетворяет *условию Липшица* по второму аргументу, если

$$\exists\,(L)\ \forall\,(t \in \mathbf{R};\ x, y \in \mathbf{R}^m)\ [\|f(t, x) - f(t, y)\| \le L\|x - y\|]$$

(число L называется *константой Липшица*).

Фундаментальное в теории обыкновенных дифференциальных уравнений утверждение о разрешимости задачи Коши — следующая

Теорема Коши — Пикара. *Если функция f непрерывна по первому аргументу и удовлетворяет условию Липшица по второму, то задача Коши (E) – (C) на любом отрезке вида $[t_0, t_0 + T]$ имеет единственное решение.*

Всюду дальше мы будем предполагать, что *выполнены условия теоремы Коши — Пикара. При этом обозначение L для константы Липшица будет универсальным.*

Простым достаточным условием выполнения условия Липшица является следующее утверждение. *Если функция f дифференцируема по второму аргументу и ее производная равномерно ограничена некоторой константой L:* $\|\partial f(t, x)/\partial x\| \le L$ *при всех* $(t, x) \in \mathbf{R}\times\mathbf{R}^m$, *то она удовлетворяет условию Липшица с константой L.*

Это же утверждение в "координатной форме": *если функции* $(t, x) \to f_i(t, x) = f_i(t, x_1, ..., x_m)$ *дифференцируемы по последним m аргументам и все частные производные ограничены некоторой константой K:* $|\partial f_i(t, x_1, ..., x_m)/\partial x_j| \le K$ *при всех* $(t, x_1, ..., x_m) \in \mathbf{R}\times\mathbf{R}^m$ *и* $i, j = 1, ..., m$, *то функция f удовлетворяет условию Липшица с некоторой константой L.* (Найдите *L* через *K* и *m*.)

Несколько слов о геометрической трактовке уравнения (E). В каждой точке расширенного фазового пространства это уравнение задает направление касательной к интегральной кривой, поскольку предписывает, чему должна равняться производная $\varphi'(t)$ в точке $(t, x) = (t, \varphi(t))$. Если в каждой точке $\mathbf{R}\times\mathbf{R}^m$ вектором $(1, f(t, x))$ указать направление касательной, то получившийся объект называют *полем направлений*, отвечающим уравнению (E)

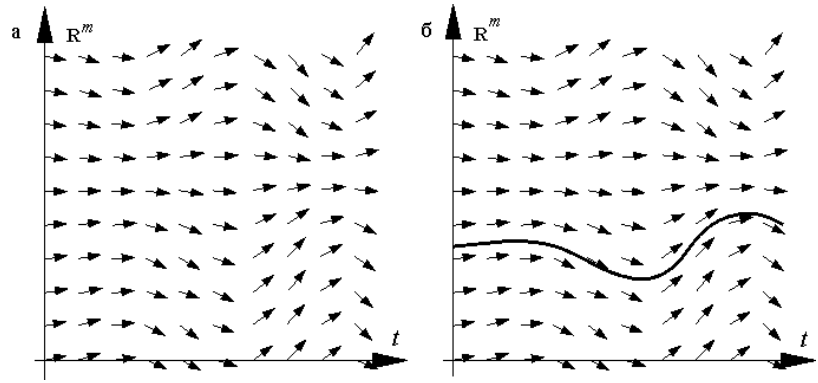


Рис. 9.1.

(рис. 9.1, а). Интегральная кривая должна "касаться" векторного поля в каждой своей точке (рис. 9.1, б). Поэтому расширенное фазовое

пространство можно представлять как парк, часто заполненный стрелками-указателями, а решение — как прогулку по этому парку в соответствии со стрелками (в направлении, указываемом стрелками).

Утверждение о гладкости решений. *Если в задаче (E) – (C) функция f является k раз непрерывно дифференцируемой, то ее решение φ непрерывно дифференцируемо k* + 1 раз.

Для чего нужны дифференциальные уравнения и чего мы от них хотим?

Дифференциальные уравнения являются инструментом познания мира и, как всякий инструмент, они развиваются и самосовершенствуются. Поэтому с точки зрения теории это самодостаточный объект. С прикладной же точки зрения дифференциальные уравнения описывают окружающий нас мир, и с их помощью мы можем узнавать о нем новое. "Познание мира" с помощью дифференциальных уравнений обычно состоит из двух этапов: составление модели (дифференциального уравнения, описывающего то или иное явление) и исследование получившейся модели. Нас интересует в данной работе второй этап. Поскольку именно решения описывают те или иные природные процессы, в конечном счете прикладнику важна информация именно о них. Больша́я часть теории обыкновенных дифференциальных уравнений посвящена изучению решений в случаях, когда оно точно не известно. Это так называемая качественная теория обыкновенных дифференциальных уравнений. К ней относится, например, теория устойчивости, позволяющая, не зная решения, по свойствам уравнения указать свойства устойчивости решений. В конкретных задачах часто возникает необходимость найти решение или иметь возможность вычислить решение в каждой точке. Иногда решение некоторых дифференциальных уравнений удается выписать в явном виде. В то же время множество дифференциальных уравнений, решения которых можно в явном виде выразить через элементарные функции, весьма и весьма бедное. Уже простейшее нелинейное уравнение первого порядка $x' = t^2 + x^2$ не допускает решений в квадратурах. Поэтому нужны методы, позволяющие вычислять решения произвольных дифференциальных уравнений приближенно.

9. 1. Дифференциальные уравнения первого порядка

Общие понятия. Теорема существования. При изучении интегрального исчисления функций одной переменной мы сталкивались с необходимостью отыскивать неизвестную функцию y по ее производной или дифференциалу.

Уравнение

$$y' = f(x) \quad \text{или} \quad dy = f(x)\,dx, \qquad (*)$$

где y— неизвестная функция от x, а $f(x)$ — заданная функция, является простейшим *дифференциальным уравнением.* Для его решения, т. е. для отыскания неизвестной функции y, нужно проинтегрировать данную функцию $f(x)$. При этом, как известно, мы получим бесчисленное множество функций, каждая из которых будет удовлетворять условию (*). В этом разделе нам удобнее будет под интегралом $\int f(x)dx$ понимать какую-либо одну первообразную. Тогда любое решение уравнения (*) запишется в виде

$$y = \int f(x)\,dx + C.$$

Далее мы увидим, что гораздо чаще приходится иметь дело с уравнениями более сложного вида. Именно в эти уравнения, помимо производной y' и независимой переменной x, может входить и сама неизвестная функция y. Примером тому служат уравнения

$$y' + x^2y = 0, \quad y' = \frac{y}{x}, \quad xy' = y + x \quad \text{и т. д.}$$

Заменяя y' через $\frac{dy}{dx}$, можно эти самые уравнения переписать в дифференциальной форме:

$$dy + x^2y\,dx = 0, \quad x\,dy - y\,dx = 0, \quad x\,dy - (y + x)\,dx = 0.$$

Определение. *Дифференциальным уравнением первого порядка называется уравнение, связывающее независимую переменную, неизвестную функцию и ее производную.*

Так как производную можно представить в виде отношения дифференциалов, то уравнение может содержать не производную, а дифференциалы неизвестной функции и независимой переменной.

Дифференциальные уравнения *второго* и *высших* порядков будут рассмотрены в п. 9. 2.

Мы будем рассматривать только такие уравнения, в которых неизвестная функция зависит от одного аргумента. Такие уравнения называются обыкновенными

Дифференциальное уравнение первого порядка в общем виде записывается так:

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

В частных случаях в левую часть уравнения могут не входить x или y, но всегда обязательно входит y'. Нам придется в основном иметь дело с уравнениями, разрешенными относительно производной, т. е. вида

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

Определение. *Решением дифференциального уравнения называется функция, которая при подстановке ее вместе с производной в это уравнение превращает его в тождество.*

Простейшие примеры показывают, что дифференциальное уравнение может иметь бесчисленное множество решений. Мы наблюдали это уже на примере уравнения (*). Простой проверкой легко убедиться также, что уравнение $y' = \frac{y}{x}$ имеет решениями функции $y = Cx$, а уравнение $y' = -\frac{y}{x}$ — функции $y = \frac{C}{x}$, где C — любое число.

Уравнение $y' = \frac{y+x}{x}$ имеет решениями функции $y = x \ln x + Cx$.

В самом деле, найдя производную $y' = \ln x + 1 + C$ и подставив ее в уравнение, получим тождество

$$\ln x + 1 + C = \frac{x \ln x + Cx + x}{x}.$$

Как мы видим, в решения приведенных дифференциальных уравнений входит произвольная постоянная C; придавая ей различные значения, мы будем получать разные решения.

Несмотря на то, что рассмотренные примеры носят частный характер, мы все-таки, не приводя доказательства, сделаем следующий общий вывод.

Любое дифференциальное уравнение $y' = f(x, y)$ имеет бесчисленное множество решений, которые определяются формулой, содержащей одну произвольную постоянную. Эту совокупность решений будем называть общим решением дифференциального уравнения первого порядка и записывать так:

$$y = \varphi(x, C).$$

Придавая произвольной постоянной C определенные числовые значения, мы будем получать *частные решения.*

В дальнейшем при решении конкретных задач нас будут интересовать преимущественно частные решения. Необходимо выяснить, каким же образом из общего решения можно выделить требуемое решение. Зададим для этого начальное условие. *Задать начальное условие дифференциального уравнения первого порядка это значит*

указать пару соответствующих друг другу значений независимой переменной (x_0) и функции (y_0). Записывают это так:

$$y\big|_{x=x_0} = y_0.$$

Покажем на примере, как по общему решению и заданному начальному условию можно отыскивать соответствующее этому условию частное решение.

Выше мы видели, что уравнение $y' = \frac{y}{x}$ имеет общее решение $y = Cx$. Зададим начальное условие $y|_{x=2} = 6$. Подставив эти значения x и y в общее решение, получим $6 = 2C$, откуда $C=3$. Следовательно, функция $y = 3x$ удовлетворяет как дифференциальному уравнению, так и начальному условию.

Вопрос о том, в каком случае можно утверждать, что частное решение дифференциального уравнения, удовлетворяющее данному начальному условию, существует, а также что оно будет единственным, выясняется следующей теоремой.

Теорема существования и единственности решения. Если *функция $f(x,y)$ непрерывна в области, содержащей точку $P_0(x_0, y_Q)$, то уравнение $y' = f(x,y)$ имеет решение $y = y(x)$ такое, что $y(x_0) = y_0$*

Если, кроме того, непрерывна и частная производная $\frac{\partial f}{\partial y}$, то это решение уравнения единственно.

Интересно отметить, что в условии теоремы не требуется существования производной $\frac{\partial f}{\partial x}$.

Теорема эта впервые была сформулирована и доказана Коши. Поэтому часто задачу отыскания частного решения по начальным условиям называют *задачей Коши.*

Перейдем теперь к геометрической иллюстрации введенных понятий.

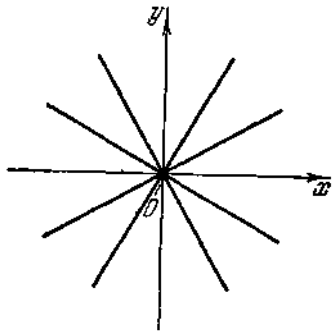


Рис. 9.2

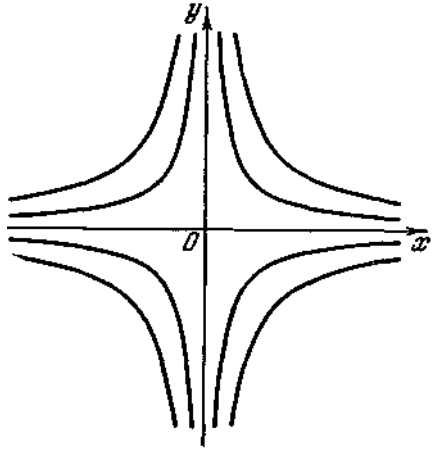


Рис. 9.3

График любого частного решения дифференциального уравнения называется *интегральной кривой.* Общему решению соответствует *семейство интегральных кривых.* Так как мы уже проверили, что уравнение $y' = \frac{y}{x}$ имеет общее решение $y = Cx$, то соответствующее ему семейство интегральных кривых — пучок прямых, проходящих через начало координат (рис. 9.2). Уравнение $y' = -\frac{y}{x}$ имеет общее решение $y = \frac{C}{x}$. Ему соответствует семейство равнобочных гипербол, асимптотами которых являются оси координат (рис. 9.3), а также прямая $y = 0$.

Задание начального условия $y|_{x=x_0} = y_0$ означает задание точки $P_0\ (x_0, y_0)$, через которую должна проходить интегральная кривая, соответствующая искомому частному решению. Таким образом, отыскание частного решения по начальному условию $y|_{x=x_0} = y_0$ геометрически означает, что из семейства интегральных кривых мы выбираем ту, которая проходит через точку $P_0(x_0, y_0)$. Согласно теореме существования и единственности решения через каждую точку, в которой функции $f(x, y)$ и $\frac{\partial f}{\partial y}$ непрерывны, проходит одна-единственная интегральная кривая. Если в данной точке эти условия нарушены, то это означает, что через эту точку либо вообще не проходит ни одна интегральная кривая, либо проходит несколько.

Возьмем, например, уравнение $y' = \frac{y}{x}$; из рис. 9.2 видно, что через начало координат проходит бесчисленное множество его интегральных кривых. Это не противоречит теореме, так как в точке (0, 0) условия теоремы существования нарушены: правая часть уравнения становится неопределенной.

Точки, в которых условия теоремы существования и единственности решения нарушаются, называются *особыми точками.* Вопрос о том, как ведут себя интегральные кривые в окрестности особой точки, будет рассмотрен ниже**;** пока же предполагается, что если мы отыскиваем частное решение уравнения $y'=f(x,y)$ по заданному начальному условию $y|_{x=x_0} = y_0$, то в точке (x_0, y_0) выполняются условия теоремы существования и единственности. Такие начальные условия будем называть *возможными.*

Теперь мы можем указать основное свойство общего решения:

Общее решение $y = \varphi(x, C)$ *дифференциального уравнения* $y'=f(x, y)$ *обладает тем свойством, что из него по любому заданному возможному начальному условию* $y|_{x=x_0} = y_0$ *может быть найдено частное решение, удовлетворяющее этому условию.*

Это означает, что, подставляя в общее решение значения x_0 и y_0, мы получаем уравнение относительно C: $y_0 = \varphi(x_0, C)$, из которого всегда может быть найдено одно-единственное значение $C=C_0$. Функция $y = \varphi(x, C_0)$ и будет искомым частным решением.

Отметим еще, что отыскание решения дифференциального уравнения часто называют интегрированием уравнения. При этом действие интегрирования функций называют *квадратурой.*

Перейдем теперь к приемам решения отдельных типов дифференциальных уравнений.

Уравнения с разделяющимися переменными. Рассмотрим уравнение вида

$$f_1(y)dy=f_2(x)dx, \qquad (*)$$

где $f_1(y)$ и $f_2(x)$ — заданные функции. В этом дифференциальном уравнении переменные разделены, т. е. каждая из переменных содержится только в той части уравнения, где находится ее дифференциал. Уравнение $dy=f(x)dx$ является частным случаем рассматриваемого уравнения.

В обеих частях уравнения (*) стоят дифференциалы некоторых функций; справа этот дифференциал выражен прямо через независимую переменную x, а слева через промежуточный аргумент y, который является функцией от x. Именно эта зависимость y от x и

является искомой. Произведя интегрирование, мы получим связь между переменными x и y, освобожденную от их дифференциалов:

$$\int f_1(y)\,dy = \int f_2(x)\,dx + C.$$

Напомним, что под символом $\int$ мы условились понимать какую-то одну первообразную. Ясно также, что произвольную постоянную C можно писать в любой части равенства.

Если задано начальное условие $y|_{x=x_0} = y_0$, то, определяя постоянную C, получим частное решение, удовлетворяющее данному условию. Воспользовавшись определенными интегралами, можно сразу записать искомое частное решение:

$$\int_{y_0}^{y} f_1(y)\,dy = \int_{x_0}^{x} f_2(x)\,dx.$$

При этом значения x_0 и y_0 действительно соответствуют друг другу, так как обе части равенства обращаются в нуль при замене верхних пределов y и x на y_0 и x_0.

Выполняя фактически интегрирование, мы обычно получаем неизвестную функцию y в неявном виде. Иногда, если решение уравнения дано не в явном виде: $y = \varphi(x)$, а в неявном: $\Phi(x, y) = 0$, то его называют *интегралом уравнения*.

Очень часто встречаются уравнения, в которых переменные еще не разделены, но их можно разделить, производя простые арифметические операции.

Определение. *Дифференциальные уравнения, в которых переменные можно разделить посредством умножения обеих частей уравнения на одно и то же выражение, называются дифференциальпыми уравнениями с разделяющимися переменными.*

Таким будет, например, уравнение

$$\frac{dy}{dx} = \frac{f_2(x)}{f_1(y)}.$$

В нем переменные еще не разделены, однако, умножив обе его части на $f_1(y)dx$, мы приходим к уравнению с разделенными переменными.

Легко также разделить переменные, если уравнение записано в дифференциальной форме и имеет вид

$$f_1(x)\,f_2(y)\,dx + f_3(x)\,f_4(y)\,dy = 0. \qquad (**)$$

Деля обе части уравнения на произведение $f_2(y)f_3(x)$, получим

$$\frac{f_1(x)}{f_3(x)}\,dx + \frac{f_4(y)}{f_2(y)}\,dy = 0.$$

Нам теперь даже не обязательно переносить одно из слагаемых в правую часть. Интегрируя, запишем

$$\int \frac{f_1(x)}{f_3(x)}\,dx + \int \frac{f_4(y)}{f_2(y)}\,dy = C.$$

Следует заметить, что при делении на $f_2(y)f_3(x)$ может произойти потеря некоторых частных решений. Пусть, например, при $y=y_0$ имеем $f_2(y_0)=0$. Тогда функция $y=y_0$ (постоянная) является решением уравнения. Действительно, $dy = 0$ и подстановка в уравнение (**) приводит к тождеству. Считая x и y равноправными и рассуждая аналогично, получим, что если $f_3(x_0)=0$, то $x=x_0$ тоже является решением уравнения.

В связи со сказанным сделаем еще такое замечание. Уравнение $y' = \frac{y}{x}$ имеет, как уже отмечено выше, общее решение $y = Cx$, т. е. совокупность прямых, проходящих через начало координат, за исключением прямой $x = 0$ — оси ординат. Записав это же самое уравнение в виде $x\,dy—ydx = 0$, мы получим и решение $x = 0$. Поэтому общим решением считают всю совокупность указанных прямых, включая и ось ординат. Такое же замечание можно сделать и при решении многих других примеров; например, уравнение $y' = -\frac{y}{x}$, семейство интегральных кривых которого изображено на рис. 9.3, имеет еще решения $y = 0$ и $x = 0$, т. е. оси координат, а уравнение примера 1—решение $y = 0$.

Однородные и линейные уравнения первого порядка. Прежде всего рассмотрим простые и важные классы уравнений первого порядка, приводящихся к уравнениям с разделяющимися переменными.

I. Однородные уравнения.

Определение. *Уравнение*

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

называется однородным, если функция $f(x, y)$ может быть представлена как функция отношения своих аргументов:

$$f(x, y) = \varphi\left(\frac{y}{x}\right).$$

Например, уравнение

$$(xy - y^2)\,dx - (x^2 - 2xy)\,dy = 0$$

однородное, так как его можно записать в виде

$$\frac{dy}{dx}=\frac{xy-y^2}{x^2-2xy}=\frac{\frac{y}{x}-\left(\frac{y}{x}\right)^2}{1-2\frac{y}{x}}.$$

В общем случае переменные в однородном уравнении не разделяются. Однако, вводя вспомогательную неизвестную функцию u по формуле

$$\frac{y}{x}=u \text{ или } y=xu,$$

мы сможем преобразовать однородное уравнение в уравнение с разделяющимися переменными. Действительно, имеем

$$y'=u+xu',$$

и уравнение $y'=\varphi\left(\frac{y}{x}\right)$ принимает вид

$$u+xu'=\varphi(u), \text{ т. е. } x\frac{du}{dx}=\varphi(u)-u.$$

Отсюда

$$\frac{du}{\varphi(u)-u}=\frac{dx}{x};$$

после интегрирования получаем

$$\int\frac{du}{\varphi(u)-u}=\ln|x|+C.$$

Найдя отсюда выражение для u как функции от x и возвращаясь к переменной $y = xu$, получим искомое решение однородного уравнения.

Чаще всего не удается просто найти явное выражение для u. Тогда после интегрирования следует в левую часть вместо u подставить $\frac{y}{x}$; в результате мы получим решение уравнения в неявном виде.

Разумеется, мы предполагаем, что $\varphi(u)-u\not\equiv 0$. Если $\varphi(u)\equiv u$, то $\varphi\left(\frac{y}{x}\right)\equiv\frac{y}{x}$ и не нужно делать никаких преобразований, ибо само заданное уравнение $y'=\frac{y}{x}$ — с разделяющимися переменными.

Если же знаменатель $\varphi(u)—u$ обращается в нуль лишь при каком-то значении u_0, то, как уже отмечено, функция $u = u_0$ является решением преобразованного уравнения, а функция $y = u_0x$ — исходного.

II. Линейные уравнения. Вторым часто встречающимся типом уравнений первого порядка является линейное уравнение.

Определение. *Уравнение вида*

$$y'+p(x)y = q(x), \qquad (*)$$

т. е. линейное относительно искомой функции и ее производной, называется линейным. Здесь $p(x)$ и $q(x)$ — известные функции независимой переменной x.

Уравнение (*) сводится к двум уравнениям с разделяющимися переменными путем следующего искусственного приема. Запишем функцию y в виде произведения двух функций: $y = uv$. Одной из них мы можем распорядиться совершенно произвольно; при этом вторая должна быть определена в зависимости от первой таким образом, чтобы их произведение удовлетворяло данному линейному уравнению. Свободой выбора одной из функций u и v мы воспользуемся для максимального упрощения уравнения, получающегося после замены. Из равенства $y = uv$ находим производную y':

$$y' = u'v+uv'.$$

Подставляя это выражение в уравнение (*), имеем

$$u'v + uv' + p(x)\,uv = q(x), \text{ или } u'v + u(v' + p(x)\,v) = q(x).$$

Выберем в качестве v *какое-нибудь* частное решение уравнения

$$v' + p(x)v = 0. \qquad (**)$$

Тогда для отыскания u получим уравнение

$$u'v=q(x). \qquad (***)$$

Сначала найдем v из уравнения (**). Разделяя переменные, имеем

$$\frac{dv}{v} = -p(x)\,dx,$$

откуда

$$\ln v = -\int p(x)\,dx \text{ и } v = e^{-\int p(x)\,dx}.$$

Под неопределенным интегралом здесь понимается *какая-нибудь* одна первообразная от функции $p(x)$, т. е. v является вполне определенной функцией от x.

Зная v, находим далее u из уравнения (***):

$$\frac{du}{dx} = \frac{q(x)}{v} = q(x)\,e^{\int p(x)\,dx}, \quad du = q(x)\,e^{\int p(x)\,dx}\,dx,$$

и значит,

$$u = \int q(x)\,e^{\int p(x)\,dx}\,dx + C.$$

Здесь мы уже берем для u все первообразные. По u и v найдем искомую функцию y:

$$y = uv = e^{-\int p(x)\,dx}\left[\int q(x)\,e^{\int p(x)\,dx}\,dx + C\right].$$

Полученная формула дает общее решение линейного уравнения (*).

Положение не изменится, если мы прибавим произвольную постоянную к интегралу в показателе. В самом деле, эта вторая

произвольная постоянная в конечном счете исчезнет, так как один множитель будет содержать ее в знаменателе, а другой — в числителе.

К линейным уравнениям часто приводятся уравнения более сложного вида. Рассмотрим, например, так называемое уравнение Бернулли

$$y' + p(x)\,y = q(x)\,y^n.$$

При $n = 0$ — это линейное уравнение, а при $n=1$ можно разделить переменные. При других значениях n оно сводится к линейному при помощи следующего приема: делим обе части уравнения на y^n и записываем его так:

$$y^{-n}y' + p(x)\,y^{-n+1} = q(x).$$

Если ввести вспомогательнуюнеизвестную функцию $y^{-n+1}=z$, то $(-n+1)y^{-n}y' = z'$, и уравнение примет вид

$$z' + (-n+1)p(x)z=(-n+1)q(x).$$

Это линейное уравнение; решая его и переходя от z снова к y, мы и получим решение исходного уравнения.

Уравнения в полных дифференциалах

Возьмем уравнение первого порядка, записанное в дифференциальной форме:

$$P(x,\ y)\,dx + Q(x,\ y)\,dy = 0. \qquad \text{(A)}$$

Определение. *Если левая часть уравнения* (A) *является полным дифференциалом некоторой функции и (x, y), то это уравнение называется уравнением в полных дифференциалах.*

Выражение же $Pdx + Qdy$, как известно, есть полный дифференциал, если $\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x}$.

Пользуясь методами отыскания функции по ее полному дифференциалу, находим такую функцию $u(x, y)$, что

$$du(x,\ y) = P(x,\ y)\,dx + Q(x,\ y)\,dy.$$

Тогда уравнение (A) можно записать так:

$$du(x, y)=0.$$

Последнее равенство означает, что между переменными x и y существует зависимость вида

$$u(x, y) = C,$$

где C —произвольная постоянная. Полученная зависимость и дает общее решение уравнения (A). Следовательно, интегрирование уравнения (A) сводится к отысканию первообразной от левой части. Воспользовавшись выражениями для этой первообразной, получаем общее решение уравнения (A) в виде

$$\int_{x_0}^{x} P(x,\ y)\,dx + \int_{y_0}^{y} Q(x_0,\ y)\,dy = C$$

или в виде

$$\int_{x_0}^{x} P(x, y_0)\,dx + \int_{y_0}^{y} Q(x,\ y)\,dy = C.$$

9.2. Теорема существования решения дифференциального уравнения первого порядка

Класс дифференциальных уравнений, которые мы можем эффективно решить, весьма узок. Например, решение простого на первый взгляд дифференциального уравнения

$$\frac{dy}{dx} = x^2 + y^2$$

оказывается не может быть сведено даже к квадратурам (интегралам). Поэтому в большинстве елучаев приходится решать дифференциальные уравнения приближенно.

Но прежде чем применять какой-либо приближенный метод, надо знать, существует ли на самом деле решение дифференциального уравнения. Очень важно также знать заранее, единственно ли оно.

Ниже формулируются условия, которые гарантируют существование и единственность решения дифференциального уравнения первого порядка

$$\frac{dy}{dx} = f(x,\ y) \tag{1}$$

при начальном условии

$$y(x_0)=y_0. \tag{2}$$

Имеет место следующая теорема.

Теорема 1. *Пусть функция f(x, y) непрерывна на прямоугольнике*

$$D = \{x_0 - a \leqslant x \leqslant x_0 + a,\ y_0 - b \leqslant y \leqslant y_0 + b\}$$

и имеет на нем ограниченную производную $\frac{\partial f}{\partial y}$, *удовлетворяющую неравенству*

$$\left|\frac{\partial f}{\partial y}\right| \leqslant N. \tag{3}$$

Тогда на отрезке $\sigma = [x_0 - \delta,\ x_0 + \delta]$, *где*

$$\delta < \min\left\{a,\ \frac{1}{N},\ \frac{b}{M}\right\},\quad M = \max_{(x,\ y)\in D} |f(x,\ y)|, \tag{4}$$

существует и притом единственное решение уравнения (1), *удовлетворяющее начальному условию* (2). *При этом выполняется неравенство*

$$|y(x)-y_0|\leqslant b \quad (\forall x\in\sigma).$$

Решение y (x) непрерывно дифференцируемо на σ. *А если f (x, y) на самом деле имеет непрерывные частные производные по x и y порядка p, то y(x) имеет на* σ *непрерывные производные до порядка p* +1 *включительно.* На рис. 9.4 в плоскости *(x, y)* изображен прямоугольник *D* и принадлежащий к нему прямоугольник

$$D_1=\{x_0-\delta\leqslant x\leqslant x_0+\delta,\ y_0-b\leqslant y\leqslant y_0+b\}.$$

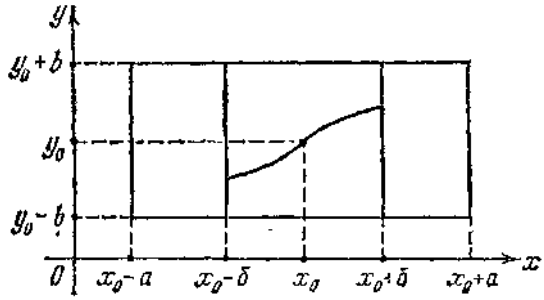


Рис. 9.4

Теорема утверждает, что если на прямоугольнике *D* функция *f(x, y)* непрерывна и имеет ограниченную частную производную $\frac{\partial f}{\partial u}$, удовлетворяющую неравенству (3), то через эту точку (x_0, y_0) проходит единственная интегральная кривая *y=y(x),* определенная для всех значений

$$x\in[x_0-\delta,\ x_0+\delta].$$

Она полностью принадлежит к прямоугольнику D_1. Число δ удовлетворяет соотношениям (4).

Подчеркнем, что теорема 1 гарантирует существование определенного отрезка

$$\sigma=[x_0-\delta,\ x_0+\delta],$$

на котором заведомо существует решение

$$y = y(x)$$

уравнения (1), проходящее через точку (x_0, y_0).

Если бы нам понадобилось найти это решение приближенно, то при наличии указанной информации мы организовали бы нахождение приближенного решения именно на этом отрезке 0, потому что нельзя ручаться, что указанное решение определено вне σ.

Ниже приводится доказательство теоремы.

Пусть задано уравнение

$$\frac{dy}{dx} = f(x,\ y), \tag{5}$$

где функция $f(x, y)$ непрерывна на прямоугольнике

$$D = \{x_0 - a \leqslant x \leqslant x_0 + a,\ \ y_0 - b \leqslant y \leqslant y_0 + b\}$$

и имеет ограниченную частную производную $\frac{\partial f}{\partial y}$, удовлетворяющую неравенству $\left|\frac{\partial f}{\partial y}\right| \leqslant N$.

Нам надо доказать, что на отрезке $\sigma = [x_0 - \delta,\ x_0 + \delta]$ существует и притом единственное решение $y = y(x)$ дифференциального уравнения (5), удовлетворяющее начальному условию

$$y(x_0) = y_0, \tag{6}$$

где

$$\delta < \min\left\{a,\ \frac{1}{N},\ \frac{b}{M}\right\}, \qquad M = \max_{(x,\ y) \in D} |f(x,\ y)|. \tag{7}$$

При этом $y(x)$ непрерывно дифференцируема на σ. Дифференциальное уравнение (5) с условием (6) эквивалентно следующему интегральному уравнению:

$$y(x) = y_0 + \int_{x_0}^{x} f(t,\ y(t))\, dt. \tag{8}$$

В самом деле, пусть непрерывная функция $y(x)$ является решением (8), тогда, дифференцируя тождество (8), получим

$$\frac{dy}{dx} = f(x,\ y(x)) \text{ и, очевидно, } y(x_0) = y_0.$$

Таким образом, функция $y(x)$ удовлетворяет уравнению (5) с условием (6),

Обратно, пусть $y(x)$ является решением (5):

$$\frac{dy(t)}{dt} = f(t,\ y(t)) \qquad (x_0 \leqslant t \leqslant x);\quad y(x_0) = y_0.$$

Тогда, интегрируя это тождество в пределах от x_0 до x, получим

$$\int_{x_0}^{x} \frac{dy(t)}{dt}\, dt = \int_{x_0}^{x} f(t,\ y(t))\, dt$$

или

$$y(x)-y_0=\int_{x_0}^{x} f(t,\ y(t))\,dt,$$

т. е. *y (x)* является решением (8).

В дальнейшем мы будем исследовать уравнение (8).

Обозначим через $\mathfrak{M}$ множество непрерывных функций *y=y(x),* заданных на отрезке $\sigma=[x_0-\delta,\ x_0+\delta]$ и удовлетворяющих на нем неравенству $|\ y(x)-y_0\ |\leqslant b$.

В $\mathfrak{M}$ введем расстояние

$$\rho(y,\ z)=\max_{x\in\sigma}|\ y(x)-z(x)\ | \qquad (y,\ z\in\mathfrak{M}).$$

Таким образом, $\mathfrak{M}$ есть метрическое пространство. Это полное пространство. В самом деле, если последовательность функций $y_n\in\mathfrak{M}$ удовлетворяет в смысле введенной метрики условию Коши (является фундаментальной последовательностью), то, как мы знаем, эта последовательность сходится равномерно на отрезке σ к некоторой непрерывной на этом отрезке функции *y=y(x).*

Для функций $y_n = y_n(x)$ выполняется неравенство

$$|\ y_n(x)-y_0\ |\leqslant b \qquad (x\in\sigma,\ n=1,\ 2,\ \ldots),$$

которое после перехода к пределу при $n\to\infty$ сохраняется:

$$|\ y(x)-y_0\ |\leqslant b.$$

Но тогда $y=y\ (x)\in\mathfrak{M}$, что показывает, что $\mathfrak{M}$—полное пространство. Равенство

$$z(x)=y_0+\int_{x_0}^{x} f(t,\ y(t))\,dt, \qquad y(x_0)=y_0, \tag{9}$$

приводит в соответствие каждой функции $y=y(x)\in\mathfrak{M}$ некоторую функцию $z=z(x)\in\mathfrak{M}$. В самом деле, если $y\in\mathfrak{M}$, то *y = y(t)* есть непрерывная функция, график которой принадлежит к прямоугольнику

$$D_1=\{x_0-\delta\leqslant x\leqslant x_0+\delta,\ y_0-b\leqslant y\leqslant y_0+b\},$$

поэтому в силу непрерывности *f (x, y)* на D_x правая часть равенсва (9) есть непрерывная функция от *x,* т. е. *z= z(x)* есть непрерывная функция на σ. Далее,

$$|\ z(x)-y_0\ |=\left|\int_{x_0}^{x} f(t,\ y(t))\,dt\right|\leqslant M\,|\ x-x_0\ |\leqslant M\delta<M\cdot\frac{b}{M}=b,$$

чю показывает, что $z\in\mathfrak{M}$.

Итак мы можем считать, что равенство (9) определяет оператор

$$z=Fy \qquad (y\in\mathfrak{M},\ z\in\mathfrak{M}),$$

отображающий полное пространство $\mathfrak{M}$ в полное пространство $\mathfrak{M}$. Этот оператор сжимающий, потому что, если

$$z_1 = Fy_1, \quad z_2 = Fy_2 \quad (y_1, \ y_2 \in \mathfrak{M}),$$

то

$$|z_1(x) - z_2(x)| = \left| \int_{x_0}^{x} [f(t, \ y_1(t)) - f(t, \ y_2(t))]\, dt \right| =$$

$$= \left| \int_{x_0}^{x} [y_1(t) - y_2(t)]\, f'_y(t, \ \lambda(t))\, dt \right| \leqslant \left| \int_{x_0}^{x} \rho(y_1, \ y_2)\, N\, dt \right| \leqslant$$

$$\leqslant \rho(y_1, \ y_2)\, \delta N = \alpha \rho(y_1, \ y_2), \quad (10)$$

где число $\alpha = \delta N$ удовлетворяет неравенству $0 \leq \alpha < 1$, потому что по условию $\delta < 1/N$. Из (10) следует, что

$$\rho(z_1, \ z_2) = \max_{x \in \sigma} |z_1(x) - z_2(x)| \leqslant \alpha \rho(y_1, \ y_2).$$

Но тогда, как мы знаем, в $\mathfrak{M}$ существует единственная функция (неподвижная точка) $y = y(x) \in \mathfrak{M}$, для которой

$$y = Fy,$$

иначе говоря, которая удовлетворяет уравнению (8), а следовательно, уравнению (5) и условию (6).

Применяя метод итераций, можно получить приближенное решение уравнения (5):

$$y_n(x) = Fy_{n-1}(x) = y_0 + \int_{x_0}^{x} f(t, \ y_{n-1}(t))\, dt \qquad (n = 1, 2, \ \ldots), \quad (11)$$

где $y_0(x) = y_0 \in \mathfrak{M}$.

На основании формулы

$$\rho(x^n, \ \bar{x}) \leqslant \frac{\alpha}{1-\alpha} \rho(x^{n-1}, \ x^n) \qquad (0 \leqslant \alpha < 1).$$

оценка приближения имеет вид

$$|y(x) - y_n(x)| \leqslant \frac{N\delta}{1 - N\delta} \max_{x \in \sigma} |y_n(x) - y_{n-1}(x)|.$$

Существуют и другие приближенные методы решения задачи Коши.

Весьма простым является метод Эйлера.

Остается еще доказать, что если $f(x, y)$ имеет непрерывные производные по x и y до p-го порядка на D, то указанное решение $y(x)$ уравнения (5) имеет непрерывные производные по x до $(p+1)$-го порядка на σ.

В самом деле, имеет место тождество

$$y'(x) = f(x, \ y(x)) \qquad (x \in \sigma). \quad (12)$$

Так как функция $y(x)$ удовлетворяет дифференциальному уравнению (5), то она всюду на σ имеет производную по x и потому непрерывна. Далее по условию $f(x, y)$ непрерывна по x и y на D, поэтому правая часть (12) непрерывна по x на σ. Значит, $y'(x)$ также непрерывна на σ Если $p \geq 1$, то правая часть (12) имеет непрерывную производную по переменной x, значит, и левая часть тождества имеет непрерывную производную по x. Следовательно, функция $y(x)$ имеет непрерывную производную второго порядка. Из тождества (12) находим

$$y''(x) = f'(x, y(x)) + f'_y(x, y(x))\, y'(x). \qquad (13)$$

Применяя к тождеству (13) те же рассуждения, что и выше, найдем, что при $p \geq 2$ функция $y(x)$ имеет непрерывную производную третьего порядка на а и т. д.

Рассмотрим пример.

Пример. Уравнение

$$\frac{dy}{dx} = -y^2 \qquad (14)$$

есть частный случай дифференциального уравнения (1).

Правая его часть не зависит от x. В данном случае функция $f(x, y)$ равна $-y^2$ при любом x.

Так как функция $-y^2$ при любом y непрерывна вместе со своей производной по y, то определяемая ею функция $f(x, y)$ непрерывна вместе со своей частной производной $\dfrac{\partial f}{\partial y}$ на всей плоскости (x, y).

Поэтому, не решая уравнение (14), можно заключить на основании теоремы существования, что через любую точку (x_0, y_0) проходит и притом единственная интегральная кривая уравнения (14).

Пусть $x_0=3$, $y_0=1$. Зададим произвольный прямоугольник

$$D = \{3 - a \leqslant x \leqslant 3 + a,\ 1 - b \leqslant y \leqslant 1 + b\} \qquad (0 < a,\ b).$$

Для него

$$M = \max_{(x,\ y) \in D} |f(x,\ y)| = \max_{y \in [1-b,\ 1+b]} |-y^2| = \max_{1-b \leqslant y \leqslant 1+b} y^2 = (1+b)^2;$$

$$N = \max_{(x,\ y) \in D} \left|\frac{\partial f}{\partial y}\right| = \max_{1-b \leqslant y \leqslant 1+b} |-2y| = 2(1+b).$$

Следовательно,

$$\delta < \min\left\{a,\ \frac{1}{2(1+b)},\ \frac{b}{(1+b)^2}\right\} < \frac{1}{2}. \qquad (15)$$

Уравнение (14) легко решается. Общий его интеграл в верхней полуплоскости ($y > 0$) и в нижней полуплоскости ($y<0$) определяется равенством

$$y = \frac{1}{x - C}. \tag{16}$$

Имеется еще одно решение $y \equiv 0$.

Среди решений (16) выберем то, которое проходит через точку (3, 1). Очевидно, это есть решение

$$y = \frac{1}{x - 2}.$$

Его график изображен на рис. 9.5.

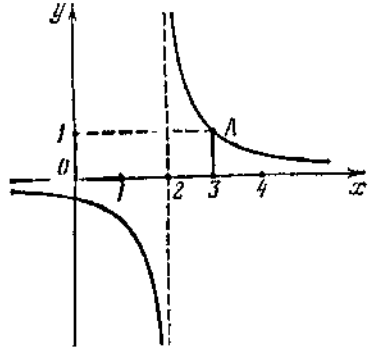


Рис. 9.5.

Мы видим, что интегральная кривая уравнения (14), проходящая через точку $A = (3, 1)$, уходит в бесконечность при $x \to 2$.

Наибольший интервал с центром в точке x = 3, на котором определена наша интегральная кривая, есть интервал (2, 4). Соотношение (16), полученное из общей теоремы существования, дает несколько меньший интервал.

9.3. Полное метрическое пространство.

Метрическое пространство M называется *полным,* если в нем всякая фундаментальная последовательность сходится к элементу этого же пространства.

Мы знаем, что одномерное пространство R_1 (чисел) полно (критерий Коши!). Можно доказать, что и пространство R_n полно при $\forall\, n \geq 1$.

Теорема 1. *Пространство* $C[a, b]$ *полное.*

Доказательство. Пусть элементы этого пространства $\{f_n(t)\}$ образуют фундаментальную последовательность в смысле метрики

$$\rho(f, g) = \max_{a \leqslant t \leqslant b} |f(t) - g(t)|. \tag{1}$$

Для всякого $\varepsilon > 0$ $\exists N$ такое, что

$$\rho(f_n, f_m) = \max_{a \leqslant t \leqslant b} |f_n(t) - f_m(t)| < \varepsilon \tag{2}$$

при $m, n > N$.

Из (2) следует, что при фиксированном $t \in [a, b]$

$$|f_n(t) - f_m(t)| < \varepsilon \quad (n, m > N) \tag{3}$$

Последнее означает, что числовая последовательность $\{f_n(t)\}$ фундаментальна, поэтому на основании критерия Коши она сходится к некоторому действительному числу, которое мы обозначаем $f(t)$:

$$\lim_{n \to \infty} f_n(t) = f(t) \qquad (\forall t \in [a, b]). \tag{4}$$

Переходя к пределу в неравенстве (3) при $m \to \infty$, получаем

$$|f_n(t) - f(t)| \leqslant \varepsilon \quad (n > N, \ \forall t \in [a, b]). \tag{5}$$

Отсюда

$$\sup_{a \leqslant t \leqslant b} |f_n(t) - f(t)| \leqslant \varepsilon \quad (n > N), \tag{6}$$

Это показывает, что последовательность $\{f_n(t)\}$ сходится *равномерно* к $f(t)$ на $[a, b]$, и так как функции $f_n(t)$ непрерывны на $[a, b]$, то и предельная функция $f(t)$ непрерывна на $[a, b]$, т. е. $f(t) \in C[a, b]$. Теорема доказана.

9.4. Принцип сжатых отображений

Изучение различных процессов, происходящих в окружающем нас мире, расчет хода этих процессов, получение их количественных характеристик приводят к необходимости решения различных математических уравнений — алгебраических, дифференциальных, интегральных.

При этом очень важно знать, существует ли решение, единственно оно или нет, и если оно есть, то найти его хотя бы приближенно.

Один из наиболее употребительных методов доказательства существования и единственности решений уравнений и построения приближенных решений основан на так называемом принципе сжатых отображений.

Хотя рассматриваемые в математике уравнения отличаются большим разнообразием, каждое из них можно привести к специальному виду:

$$x=A(x).$$

В этой записи x обозначает некоторый математический объект, например число, вектор или функцию, а A — некоторое преобразование этих объектов.

Обозначим множество всех значений, которые может принимать x, буквой M. Тогда отображением A множества M называют соответствие, которое каждому x из M сопоставляет $A(x)$ из M. Например, если x принимает числовые значения, то A — обычная функция числового аргумента. Если x — функция, то A может быть дифференциальным или интегральным оператором.

Наглядно действие отображения A можно представить себе как перемещение точек, лежащих в M, при котором точка x переходит в точку $A(x)$. Равенство

$$A(x)=x$$

при такой интерпретации означает, что в результате отображения точка x осталась неподвижной. Таким образом, вопрос о решении уравнения $x=A(x)$ является вопросом об отыскании всех неподвижных точек отображения A.

Ниже рассматривается задача приближенного нахождения неподвижных точек одного класса отображении, широко используемых в теории оптимизации. Основное свойство этих отображений заключается в том, что они уменьшают расстояния между точками. Они как бы сжимают множество M и поэтому называются сжатыми отображениями (правильнее было бы «сжимающими»). В 1920 г. польский математик Стефан Банах в своей докторской диссертации (опубликована в 1922 г.) доказал, что каждое сжатое отображение A имеет неподвижную точку, т. е. у уравнения $x=A(x)$ существует решение. Это утверждение получило в математике название принципа сжатых отображений.

Задача об отыскании корней числовых уравнений

Принцип сжатых отображений формулируется для отображений абстрактных множеств, называемых полными метрическими пространствами. Частным случаем этих пространств является отрезок числовой прямой, и поэтому принцип сжатых отображений может быть применен для исследования числовых уравнений. Однако мы поступим по-другому: сначала изучим числовые уравнения, а затем,

отталкиваясь от них, исследуем сжатые отображения произвольных метрических пространств.

В этом пункте мы рассмотрим функции f, определенные на некотором отрезке $[a, b]$ и принимающие числовые значения. Это означает, что каждому числу $x \in [a, b]$ функция f сопоставляет действительное число $f(x)$.

С каждой функцией f, заданной на некотором отрезке $[a, b]$, может быть связано уравнение

$$f(x)= 0.$$

Число c из отрезка $[a, b]$ называют корнем этого уравнения, если $f(c)=0$.

Для приближенного нахождения корня уравнения $f(x)=0$ преобразуем это уравнение так, чтобы оно приняло вид

$$x=\varphi(x), \qquad (1)$$

где φ — функция, заданная на отрезке $[a, b]$.

Разумеется, уравнение $f(x)=0$ может быть преобразовано к виду (1) многими способами, и очень часто успех в решении задачи определяется удачным выбором преобразования. Например, уравнение

$$x-\cos x=0$$

естественно записать в виде

$$x=\cos x. \qquad (2)$$

В этом примере нужное нам преобразование отыскать было очень легко. В общих же случаях эта задача совсем не проста.

Итак, пусть задано уравнение (1). Требуется найти корни этого уравнения.

Конечно, хотелось бы получить для корней уравнения (1) явное выражение через элементарные функции, однако даже в случае простейших уравнений такая задача оказывается неразрешимой.

Возьмем в качестве примера: $x=\cos x$.

Построим графики функций, стоящих в левой и правой частях. Как видно на рис. 9.6, они пересекаются при некотором $x=c$, $0<c<1$. Число c является корнем уравнения (2), однако получить для него формулу невозможно. Как говорят - «Хоть видит око, да зуб неймет».

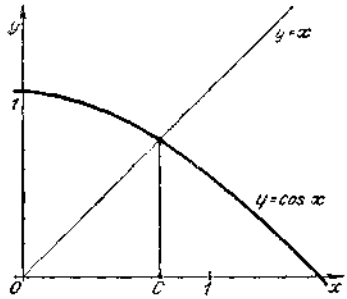


Рис. 9.6.

В условиях, когда формулы «не работают», когда рассчитывать на них можно только в самых простейших случаях, важное значение приобретают универсальные вычислительные алгоритмы.

Одним из наиболее употребительных алгоритмов является метод последовательных приближений. Именно для того чтобы можно было воспользоваться этим методом, мы записали наше уравнение в виде (1).

Метод последовательных приближений заключается в следующем. Задается произвольно число x_0 из отрезка $[a, b]$, на котором ищутся корни. Число x_0 называется нулевым приближением и определяется обычно прикидкой (или, как говорят в математической физике, из физических соображений). Далее в качестве первого приближения x_1 берется число

$$x_1=\varphi(x_0).$$

Затем в качестве второго приближения x_2 берут число

$$x_2=\varphi(x_1).$$

И вообще, если найдены приближения $x_1, x_2, \ldots, x_{n-1}$, то в качестве n-го приближения берут число

$$x_n=\varphi(x_{n-1}),\ n=1, 2, 3, \ldots \qquad (3)$$

Полученные числа $x_1, x_2, \ldots$ называют последовательными приближениями корня уравнения (1).

Предположим, что с увеличением номера n числа x_n приближаются к некоторому числу c из отрезка $[a, b]$:

$$x_n \to c.$$

Если значения $\varphi(x_n)$ стремятся к $\varphi(c)$, то, устремляя в равенстве (3) номер n к бесконечности, получаем в пределе

$$c=\varphi(c),$$

т. е. c — корень уравнения (1).

Математическое обоснование метода последовательных приближений заключается в доказательстве того, что

а) для любого номера n можно построить n-е приближение x_n;

б) последовательные приближения x_n стремятся к некоторому числу c;

в) значения $\varphi(x_n)$ стремятся к значению $\varphi(c)$.

Если эти три условия выполнены, то из равенства (3) следует, что c — корень уравнения (1).

Рассмотрим, какие требования необходимо наложить на функцию φ, чтобы выполнялись условия а)—в).

Условие а) может оказаться невыполненным, если для некоторого номера n число $\varphi(x_{n-1})$ выйдет за пределы отрезка $[a, b]$. Тогда для числа $x_n = \varphi(x_{n-1})$ нельзя вычислить $\varphi(x_n)$, так как функция φ, вообще говоря, не определена вне отрезка $[a, b]$. Следовательно, нельзя построить $(n+1)$-е приближение

$$x_{n+1} = \varphi(x_n),$$

так как значение $\varphi(x_n)$ не определено.

Рассмотрим, например, на отрезке [1,2] функцию $\varphi(x) = \sqrt{x-1}$ и возьмем в качестве нулевого приближения число $x_0=2$. Тогда

$$x_1 = \varphi(x_0) = \varphi(2) = \sqrt{2-1} = \sqrt{1} = 1;$$

$$x_2 = \varphi(x_1) = \varphi(1) = \sqrt{1-1} = \sqrt{0} = 0;$$

$$x_3 = \varphi(x_2) = \varphi(0) = \sqrt{0-1} = \sqrt{-1} = ?,$$

т. е. уже третьего приближения не существует.

Для выполнения условия а) потребуем, чтобы на всем отрезке $[a, b]$ выполнялось двойное неравенство

$$a \leqslant \varphi(x) \leqslant b \quad \text{при} \quad x \in [a, b]. \tag{4}$$

Это неравенство означает, что каково бы ни было число $x \in [a, b]$, число $\varphi(x)$ лежит на отрезке $[a, b]$. Иначе говоря, *функция* φ *переводит отрезок* $[a, b]$ *в себя.*

Очевидно, при выполнении неравенства (4) для любого номера n можно по формуле (3) построить приближение x_n.

Для выполнения условия в) потребуем, чтобы функция φ была непрерывной в каждой точке отрезка $[a, b]$. Напомним, что функция φ называется *непрерывной* в точке $x \in [a, b]$, если для любой последовательности чисел $x_n \in [a, b]$ из $x_n \to x$ следует $\varphi(x_n) \to \varphi(x)$.

Можно дать наглядную интерпретацию понятия непрерывности: если функция непрерывна в каждой точке отрезка $[a, b]$, то ее график можно начертить, не отрывая карандаша от бумаги.

Таким образом, если мы хотим получить решение уравнения (1) методом последовательных приближений, необходимо потребовать, чтобы функция φ была непрерывной и удовлетворяла неравенствам (4).

Обратимся теперь к условию б). Каковы должны быть требования, накладываемые на функцию φ, чтобы последовательные приближения x_n стремились к некоторому числу c? Этот вопрос является намного более трудным, и для ответа на него нам придется коснуться основных положений математического аназиза.

Итак, пусть дана последовательность действительных чисел x_n. Как узнать, стремятся ли эти числа x_n к какому-нибудь действительному числу?

Рассмотрим, например, последовательность $x_n=1/n$. (Обычно для обозначения последовательности используют фигурные скобки $\{x_n\}$, но мы будем их опускать, так как здесь это не приведет к недоразумениям.)

Ясно, что с увеличением номера n числа x_n стремятся к нулю. Математически это утверждение формулируют так: *увеличивая номер n, числа x_n можно сделать по абсолютной величине меньше любого наперед заданного положительного числа.*

Числовые последовательности, обладающие таким свойством, называют *бесконечно малыми.*

Примером бесконечно малой последовательности может служить последовательность членов геометрической прогрессии со знаменателем q, по абсолютной величине меньшим единицы:

$$x_n=q^n, \quad |q|<1.$$

Если же $|q|\geq 1$, то последовательность q^n не является бесконечно малой. Проиллюстрируем сказачное на примере последовательности

$$x_n=(-1)^n$$

т. е. в случае $q=-1$.

Эта последовательность не является бесконечно малой, так как абсолютные величины чисел x_n равны 1,

$$|x_n|=|(-1)^n|=1,$$

и они не могут быть сделаны меньше любого положительного числа. Например, сколько бы мы ни увеличивали номер n, мы не добьемся того, чтобы выполнялось неравенство

$$|x_n|<0{,}1.$$

Пусть теперь дана произвольная числовая последовательность x_n. Говорят, что последовательность x_n *сходится* к числу c, если последовательность разностей

$$x_n-c$$

является бесконечно малой (т. е. стремится к нулю). Число c называют *пределом* последовательности x_n и обозначают это так:

$$x_n \to c \text{ при } n \to \infty.$$

Например, последовательность

$$x_n = n/(n+1)$$

имеет предел, равный 1. Действительно,

$$x_n - 1 = n/(n+1) - 1 = -1/(n+1) \to 0$$

т. е. последовательность $(x_n - 1)$ — бесконечно малая. Поэтому $x_n \to 1$.

Примером последовательности, не имеющей предела, может служить последовательность

$$x_n = (-1)^n.$$

Всякая последовательность x_n, имеющая предел, называется *сходящейся.*

Итак, как проверить, является ли данная последовательность x_n сходящейся? Решение этой проблемы было дано французским математиком О. К о ш и (1789—1857).

В чем состоит трудность этой проблемы? Для лучшего понимания, предположив, что дана последовательность чисел x_n, сравним следующие два вопроса:

1) верно ли, что последовательность x_n сходится к данному числу c?
2) верно ли, что последовательность x_n сходится?

Казалось бы, эти два вопроса почти не отличаются друг от друга. Однако различие обнаруживается сразу же, как только вы начинаете искать ответ на них.

Для ответа на первый вопрос надо взять *данное* число c, составить разность $x_n - c$ и проверить, является ли эта разность бесконечно малой.

Если ответ положительный, то $x_n \to c$, если нет, то x_n не сходится к c. Для ответа на второй вопрос надо брать *каждое* действительное число c, составлять разности $x_n - c$ и проверять, являются ли они бесконечно малыми. Если, перебирая все действительные числа, мы найдем такое значение c, при котором последовательность $x_n - c$ бесконечно малая, то x_n сходится. Если же, перебрав все значения c, мы так и не получим бесконечно малую последовательность $x_n - c$, то x_n не сходится, т. е. расходится.

Можно ли дать ответ на второй вопрос, не перебирая все действительные значения c, а пользуясь только заданными значениями x_n? Да, это можно сделать, если применить критерий Коши.

Предположим, что последовательность x_n сходится, т. е. существует такое число c, что $(x_n - c)$ — бесконечно малая последовательность. Это означает, что для достаточно больших номеров n числа

$$|x_n - c|$$

могут быть сделаны сколь угодно малыми. Отсюда следует, что для достаточно больших номеров n и m числа

$$|x_n—x_m|$$

будут сколь угодно малыми.

Действительно,

$$|x_n—x_m|= | (x_n—c)+(c-x_m)|\leq |x_n—c|+ |c—x_m|,$$

и если $|x_n—c|$ и $|x_m—c|$ малы, то и величина $|x_n—x_m|$ мала.

Это свойство числовой последовательности получило особое название. Числовую последовательность x_n называют *фундаментальной,* если

$$|x_n-x_m|\to 0, \qquad (5)$$

когда номера n и m стремятся к бесконечности.

Приведенные выше рассуждения означают, что всякая сходящаяся последовательность фундаментальна. Замечательным оказалось то, что верно и обратное утверждение.

Фундаментальность последовательности означает, что с увеличением номера ее члены «уплотняются», сгущаясь около некоторой точки. Следовательно, всякая фундаментальная последовательность сходится. Это утверждение и составляет основу критерия Коши, который звучит так: *числовая последовательность является сходящейся тогда и только тогда, когда она фундаментальна.*

Главное достоинство критерия Коши заключается в том, что он позволяет решить вопрос о сходимости последовательности x_n, не прибегая к перебору возможных значений предела c.

Рассмотрим для примера снова последовательность

$$x_n=(-1)^n.$$

Выше уже отмечалось, что эта последовательность расходится, т е. не имеет предела. Применение же критерия Коши позволяет доказать расходимость сразу: будем стремить n и m к бесконечности так, чтобы n было четным числом, а m — нечетным.

Тогда $x_n=(-1)^n=1$, $x_m=(-1)^m=-1$. Поэтому

$$|x_n—x_m|=|1—(-1)|=2,$$

и значит $|x_n—x_m|$ не стремится к нулю, т. е. последовательность x_n не фундаментальна и, следовательно, расходится.

Вернемся теперь к вопросу о приближенном решении уравнения (1). Какие требования необходимо наложить на функцию φ, чтобы последовательность x_n, определенная соотношениями (3), была фундаментальной?

К сожалению, нет критерия, который позволил бы для каждой функции φ дать однозначный ответ на этот вопрос. Однако имеются многочисленные достаточные условия, выполнение которых автоматически обеспечивает фундаментальность (а значит, и

сходимость) последовательных приближений. Одно из таких условий, так называемое условие Липшица, будет интересовать нас больше, чем все остальные, поэтому только его мы и рассмотрим.

Будем говорить, что функция φ, определенная на отрезке [*a, b*], удовлетворяет условию Липшица с константой α, если для всех *x* и *y* из отрезка [*a, b*] выполняется неравенство

$$|\varphi(x)-\varphi(y)|\leqslant\alpha|x-y|,\ x\in[a,\ b],\ y\in[a,\ b]. \qquad (6)$$

Это условие было введено немецким математиком Р. Липшицем (1832—1903) в 1864 г. как достаточное условие сходимости ряда Фурье функции φ. Важно, что положительная постоянная α в неравенстве (6) не зависит от *x* и *y*.

Заметим, что каждая функция, удовлетворяющая условию Липшица, непрерывна на отрезке [*a, b*]. Действительно, пусть *x* — любая точка отрезка [*a, b*] и пусть $x_n \to x$, т. е. $|x_n - x| \to 0$. Тогда из неравенства (6) получаем

$$|\varphi(x_n)-\varphi(x)|\leqslant\alpha|x_n-x|\to 0,$$

т. е. $\varphi(x_n) \to \varphi(x)$. Это и означает, что φ непрерывна в каждой точке отрезка [*a, b*].

Особое значение имеет условие Липшица (6) в случае, когда постоянная $\alpha<1$. Это и есть то самое условие, которое обеспечивает фундаментальность последовательных приближений.

Поясним значение условия $\alpha<1$ на примере линейной функции

$$\varphi(x)=kx,$$

где *k* — постоянная. Убедимся сначала в том, что эта функция удовлетворяет условию Липшица. Для этого оценим разность $\varphi(x) - \varphi(y)$. Из очевидного равенства

$$|\varphi(x)-\varphi(y)|=|k|\cdot|x-y|$$

следует, что условие Липшица выполняется с постоянной $\alpha=|k|$.

Уравнение (1) для рассматриваемой функции φ принимает вид

$$x=kx.$$

Если $k\neq1$, то уравнение имеет единственный корень $x=0$ (см. рис. 9.7—9.10, где k=—1,5; —0,5; 0,5; 1,5).

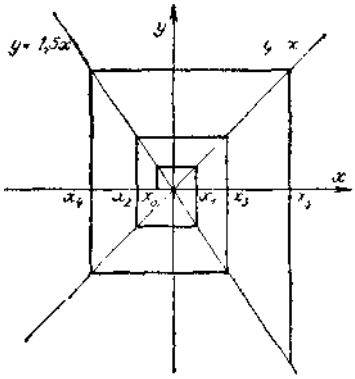


Рис. 9.7

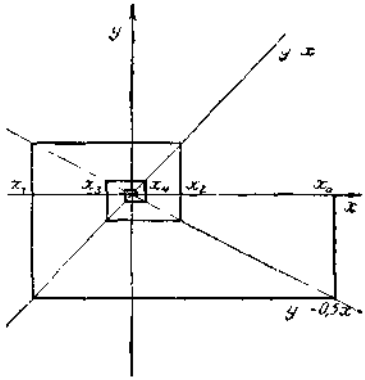


Рис. 9.8

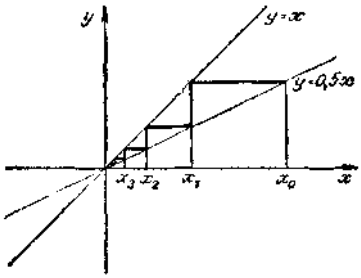


Рис. 9. 9

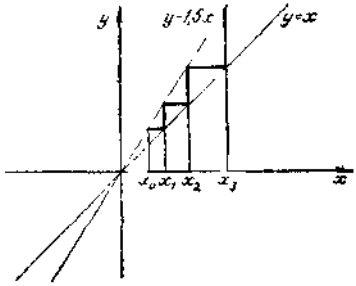


Рис. 9.10

Выясним, при каких k этот корень можно найти методом последовательных приближений. Из рис. 9.8 и 9.9 можно заключить, что последовательные приближения сходятся к корню при $|k|<1$, а из рис. 9.7 и 9.10 — что они неограниченно увеличиваются при $|k|>1$, каково бы ни было начальное приближение x_0.

Рассмотрев этот простой пример, естественно ожидать, что для сходимости последовательных приближений в общем случае следует потребовать, чтобы константа Липшица была меньше 1.

Условию Липшица с константой $\alpha<1$ можно дать наглядную интерпретацию. Для этого привычную фразу «значение функции φ в точке x равно φ (x)» заменим фразой «функция φ отображает точку x в точку φ (x)». Тогда функция φ, удовлетворяющая неравенствам (4), может рассматриваться как отображение отрезка $[a, b]$ в себя. Точка $\overline{x}$ называется образом точки x, если выполняется равенство $x = \varphi(x)$.

Если отображение φ удовлетворяет условию Липшица с константой $\alpha < 1$, то расстояние между образами двух любых точек из отрезка $[a, b]$ меньше расстояния между этими точками по крайней мере в $1/\alpha$ раз. Отображение φ как бы сжимает отрезок $[a, b]$, поэтому его называют *сжатым.* Правильнее было бы назвать его сжимающим отображением, однако мы будем придерживаться традиционной терминологии.

Итак, функция φ, определенная на отрезке $[a, b]$, есть сжатое отображение отрезка $[a, b]$, если выполняется неравенство (6) с некоторой положительной постоянной $\alpha < 1$.

Докажем теперь следующее утверждение.

Принцип сжатых отображений. *Пусть функция* φ *есть сжатое отображение отрезка* $[a, b]$ *в себя. Тогда независимо от выбора нулевого приближения* $x_0 \in [a, b]$ *последовательность* x_n, *определяемая формулой*

$$x_n = \varphi(x_{n-1}), \quad n = 1, 2, 3, \ldots,$$

сходится к корню уравнения

$$x = \varphi(x).$$

Доказательство этого утверждения проведем в несколько шагов.

1°. Заметим прежде всего, что по определению последовательности x_n справедливы равенства

$$x_n = \varphi(x_{n-1}), \quad x_{n+1} = \varphi(x_n), \quad n = 1, 2, 3\ldots$$

Из этих равенств получаем

$$x_n - x_{n+1} = \varphi(x_{n-1}) - \varphi(x_n). \tag{7}$$

Воспользуемся тем, что функция φ удовлетворяет условию Липшица с константой α:

$$|\varphi(x_{n-1}) - \varphi(x_n)| \leqslant \alpha |x_{n-1} - x_n|.$$

Применяя эту оценку, из формулы (7) получаем следующее основное неравенство:

$$|x_n - x_{n+1}| \leqslant \alpha |x_{n-1} - x_n|. \tag{8}$$

Это означает, что длина каждого шага $|x_n - x_{n+1}|$ по сравнению с предыдущим уменьшается по крайней мере в $1/\alpha$ раз. Поэтому мы получаем следующую цепочку неравенств:

$$|x_{n+1} - x_{n+2}| \leqslant \alpha^2 |x_{n-1} - x_n|,$$
$$|x_{n+2} - x_{n+3}| \leqslant \alpha^3 |x_{n-1} - x_n|,$$
$$|x_{n+k-1} - x_{n+k}| \leqslant \alpha^k |x_{n-1} - x_n|.$$

Так как абсолютная величина разности $x_n - x_{n+k}$ не превосходит суммы величин, стоящих в левой части написанных неравенств, то для любого натурального k выполняется оценка

$$|x_n - x_{n+k}| \leqslant (\alpha + \alpha^2 + \ldots + \alpha^k) |x_{n-1} - x_n|. \tag{9}$$

2°. Заметим теперь, что в скобках стоит сумма k членов геометрической прогрессии со знаменателем α. Эта сумма равна

$$\alpha + \alpha^2 + \cdots + \alpha^k = \frac{\alpha - \alpha^{k+1}}{1 - \alpha},$$

и так как $0 < \alpha < 1$, то

$$\alpha + \alpha^2 + \ldots + \alpha^k < \alpha/(1-\alpha).$$

Поэтому из оценки (9) мы получаем следующее важное неравенство:

$$|x_n - x_{n+k}| \leqslant \frac{\alpha}{1-\alpha} |x_{n-1} - x_n|. \qquad (10)$$

3°. Применим теперь для оценки правой части неравенства (10) формулу (8). После (n—1)-кратного применения этой оценки получим

$$|x_{n-1} - x_n| \leqslant \alpha |x_{n-2} - x_{n-1}| \leqslant \alpha^2 |x_{n-3} - x_{n-2}| \leqslant \alpha^3 |x_{n-4} - x_{n-3}| \leqslant \ldots \leqslant \alpha^{n-1} |x_0 - x_1|.$$

Отсюда и из (10) следует неравенство

$$|x_n - x_{n+k}| < \frac{\alpha^n}{1-\alpha} |x_0 - x_1|.$$

Вспомним, что $x_1 = \varphi(x_0)$. Тогда последнее неравенство можно переписать так:

$$|x_n - x_{n+k}| < \frac{\alpha^n}{1-\alpha} |x_0 - \varphi(x_0)|. \qquad (11)$$

4°. Докажем теперь, что последовательность x_n фундаментальна. Пусть n и m — любые натуральные числа. Без ограничения общности можно считать, что $m > n$, т. е. $m = n + k$, где k — натуральное число. Тогда в силу доказанного выше неравенства (11)

$$|x_n - x_m| < \frac{\alpha^n}{1-\alpha} |x_0 - \varphi(x_0)|. \qquad (12)$$

Теперь следует воспользоваться предположением о том, что $\alpha < 1$, где α — константа Липшица. Если n стремится к бесконечности, то $\alpha^n \to 0$, и поэтому величина в правой части (12) также стремится к нулю. Но тогда и величина в левой части стремится к нулю, т. е. при $n \to \infty$ и $m \to \infty$

$$|x_n - x_m| \to 0.$$

Это и означает, что последовательность x_n фундаментальна.

5°. Согласно критерию Коши последовательность x_n сходится к некоторому числу c из отрезка $[a, b]$:

$$x_n \to c, \quad c \in [a, b].$$

Так как функция φ удовлетворяет условию Липшица, то она непрерывна, и поэтому

$$\varphi(x_n) \to \varphi(c).$$

Устремим в равенстве

$$x_n = \varphi(x_{n-1})$$

номер n к бесконечности и получим

$$c = \varphi(c),$$

т е. c — корень уравнения (1). Тем самым справедливость принципа сжатых отображений полностью установлена.

Отметим наиболее важные особенности принципа сжатых отображений.

Во-первых, следует обратить внимание на то, что в формулировке этого принципа не делалось предположения о наличии корня уравнения (1). Более точно: мы доказали, что если функция φ удовлетворяет сформулированным выше условиям, то уравнение

$$x = \varphi(x)$$

непременно имеет корень на отрезке $[a, b]$. Тем самым доказана *теорема существования* корня уравнения (1). Если бы исходя из каких-либо других соображений мы установили, что уравнение (1) имеет корень c на отрезке $[a, b]$, то доказательство сходимости последовательных приближений x_n к c можно было бы провести в несколько строк. Приведем доказательство. Итак, пусть

$$x_n = \varphi(x_{n-1}), \quad c = \varphi(c).$$

Тогда

$$|x_n - c| = |\varphi(x_{n-1}) - \varphi(c)| \leqslant \alpha |x_{n-1} - c|,$$

т. е.

$$|x_n - c| \leqslant \alpha |x_{n-1} - c| \leqslant \alpha^2 |x_{n-2} - c| \leqslant \ldots \leqslant \alpha^n |x_0 - c|.$$

Так как $\alpha^n \to 0$ при $n \to \infty$, то $x_n \to c$, что и требовалось доказать.

Еще раз подчеркнем, что, не предполагая заранее наличия корня уравнения (1), мы доказали выше, что такой корень обязательно существует. Только этим объясняется некоторая сложность доказательства (включающего доказательство фундаментальности последовательности x_n с последующим применением критерия Коши).

Второе обстоятельство, на которое следует обратить внимание, — это вопрос о количестве корней уравнения (1) на отрезке $[a, b]$. Оказывается, при сформулированных выше условиях уравнение (1) не может иметь больше одного корня. Доказывается это просто. Если допустить, что, кроме найденного корня c, имеется еще один корень c_1, то из равенств

$$c = \varphi(c), \quad c_1 = \varphi(c_1)$$

получаем

$$c-c_1=\varphi(c)-\varphi(c_1).$$

Так как φ — сжатое отображение, то

$$|\varphi(c)-\varphi(c_1)|\leqslant\alpha|c-c_1|,$$

поэтому

$$|c-c_1|\leqslant\alpha|c-c_1|,$$

т. е.

$$(1-\alpha)|c-c_1|\leqslant 0.$$

Левая часть отрицательной быть не может, следовательно, она равна нулю, и так как $\alpha<1$, то $c-c_1=0$, т. е. $c=c_1$. Это и означает, что уравнение (1) не может иметь больше одного корня. Тем самым доказана *теорема единственности* корня уравнения (1).

Третье важное обстоятельство: принцип сжатых отображений носит конструктивный характер. Мы установили не только существование и единственность корня уравнения (1), но и указали конкретный способ приближенного нахождения корня. Следует обратить внимание на то, что последовательные приближения сходятся к точному решению уравнения (1) независимо от выбора начального приближения x_0. Это говорит об устойчивости вычислительного алгоритма по начальному приближению.

Остановимся подробнее на применении принципа сжатых отображений для приближенного решения уравнения (1) с помощью ЭВМ. Прежде всего отметим, что значительному упрощению вычислений благоприятствует способ задания последовательности x_n с помощью равенств

$$x_n=\varphi(x_{n-1}),\ n=1,\ 2,\ 3,\ldots$$

Формула, позволяющая выразить n-й член последовательности через предыдущие, называется рекуррентной (от латинского *recurrens* — возвращающийся). В нашем случае рекуррентная формула имеет простейший вид, так как значение x_n вычисляется только по x_{n-1}. При таком способе задания последовательности достаточно знать начальное приближение x_0, все остальные члены последовательности определяются после этого однозначно. Достоинства рекуррентного способа задания последовательности особенно четко выявляются при счете на ЭВМ, так как процесс счета является итерационным.

Численный метод называют итерационным (от латинского *iteratio* — повторение), если в нем производится последовательное, шаг за шагом, уточнение первоначального грубого приближения. Каждый шаг в таком методе называется итерацией. В нашем случае итерацией является вычисление с помощью рекуррентной формулы значения x_n по найденному ранее значению x_{n-1}.

Поясним кратко, почему итерационные методы наиболее удобны при счете на ЭВМ. Предположим, что нам задана функция $\varphi(x)$ и требуется найти корень уравнения (1) с точностью до ε. Это означает, что, взяв некоторое нулевое приближение, мы должны вести счет до тех пор, пока не получим приближение x_n, удовлетворяющее неравенству

$$|x_n - c| \leqslant \varepsilon, \tag{13}$$

где c — искомый корень.

Приведем один из алгоритмов вычисления приближенного решения уравнения (1).

Схематически этот алгоритм может быть описан так.

1. Ввести в ячейку x значение x_0.
2. Вычислить $\varphi(x)$ и результат ввести в ячейку $\tilde{x}$.
3. Сравнить $|\tilde{x} - c|$ и ε. Если $|\tilde{x} - c| < \varepsilon$, то перейти к пункту 6, иначе перейти к п. 4.
4. Ввести значение $\tilde{x}$ в ячейку x.
5. Перейти к пункту 2.
6. Выдать значение $\tilde{x}$ и остановить счет.

Пункты 2—5 образуют цикл, выполнение которого представляет собой итерацию. При таком алгоритме необходимое число итераций определяется самой ЭВМ: как только достигается требуемая точность, счет прекращается. Это становится возможным благодаря тому, что в ЭВМ имеется оператор условного перехода (см. пункт 3), предназначенный для итерационных алгоритмов.

Разумеется, приведенный выше алгоритм представляет собой грубое упрощение, недостатки которого сразу бросаются в глаза. Главный из этих недостатков заключается в следующем: машина должна сравнивать приближенное значение с точным, однако точное значение нам неизвестно. Если бы мы знали точное значение, отпала бы необходимость в счете.

Для того чтобы избавиться от этого недостатка, постараемся получить оценку погрешности без привлечения точного значения искомого корня. Воспользуемся установленной выше оценкой (11):

$$|x_n - x_{n+k}| \leqslant \frac{\alpha^n}{1-\alpha} |x_0 - \varphi(x_0)|.$$

Зафиксируем номер n и устремим номер k к бесконечности. Так как последовательные приближения сходятся к корню c уравнения (1), то $x_{n+k} \to c$ при $k \to \infty$. Поэтому из оценки (11) получаем

$$|x_n - c| \leqslant \frac{\alpha^n}{1-\alpha} |x_0 - \varphi(x_0)|, \tag{14}$$

Это и есть требуемая нам оценка погрешности, выраженная через начальное приближение и число итераций. Исходя из этого, можно найти число итераций, достаточное для достижения требуемой точности.

Обозначим через δ погрешность начального приближения:

$$\delta = |x_0 - \varphi(x_0)|.$$

Тогда искомое число итераций n должно быть таким, чтобы выполнялось неравенство

$$\frac{\alpha^n}{1-\alpha} \cdot \delta \leqslant \varepsilon,$$

где ε — заданная погрешность. Отсюда получаем

$$\alpha^n \leqslant \frac{\varepsilon}{\delta}(1-\alpha),$$

т. е.

$$\frac{1}{\alpha^n} \geqslant \frac{\delta}{\varepsilon(1-\alpha)}.$$

Следовательно,

$$n \geqslant \frac{1}{\log \frac{1}{\alpha}} \cdot \log \frac{\delta}{\varepsilon(1-\alpha)}. \qquad (15)$$

Обозначим через $N(\varepsilon)=N(\varepsilon, \delta, \alpha)$ наименьшее целое число n, удовлетворяющее неравенству (15). Теперь приближенное вычисление корня уравнения (1) можно вести по такому алгоритму:

1. Ввести в ячейку x число x_0.
2. Ввести в ячейку n число 1.
3. Вычислить $\varphi(x)$ и результат ввести в ячейку $\tilde{x}$.
4. Сравнить n с $N(\varepsilon)$. Если $n \geq N(\varepsilon)$, то перейти к пункту 8, иначе перейти к п. 5.
5. Ввести значение $\tilde{x}$ в ячейку x.
6. Увеличить число в ячейке n на 1.
7. Перейти к пункту 3.
8. Выдать значение $\tilde{x}$ и остановить счет.

В приведенном алгоритме пункты 3—7 образуют цикл, выполнение которого представляет собой итерацию. В данном случае число итераций устанавливают заранее, опираясь на оценку (15).

По этому алгоритму уже можно вести счет, так как в нем не требуется располагать точным значением корня уравнения (1). Однако и этот алгоритм практически не используется по следующим двум причинам.

Во-первых, для определения числа итераций необходимо найти возможно более точно константу Липшица α, что представляет собой трудную задачу.

Во-вторых, для вычисления решения с заданной погрешностью ε (см. неравенство (13)) часто требуется гораздо меньше итераций *n*, чем для выполнения неравенства (14). Иначе говоря, требуемая точность уже будет достигнута, а счет все еще будет продолжаться.

Обычно программа составляется так, чтобы счет прекращался после достижения неравенства

$$|x_n - x_{n-1}| \leqslant \varepsilon. \qquad (16)$$

Приведем аргументы в пользу того, что этот способ является в достаточной мере удовлетворительным. Для этого обратимся к установленной ранее оценке (10):

$$|x_n - x_{n+k}| \leqslant \frac{\alpha}{1-\alpha} |x_n - x_{n-1}|.$$

Аналогично тому, как это было сделано выше, зафиксируем номер *n* и устремим номер *k* к бесконечности. Так как последовательные приближения сходятся к точному значению корня *c*, то $x_{n+k} \to c$ при $k \to \infty$. Поэтому из оценки (10) получаем:

$$|x_n - c| \leqslant \frac{\alpha}{1-\alpha} |x_n - x_{n-1}|.$$

Это и есть требуемая оценка погрешности. Если выполняется условие (16), то из него получаем

$$|x_n - c| \leqslant \frac{\alpha}{1-\alpha} \cdot \varepsilon. \qquad (17)$$

Конечно, эта оценка несколько отличается от требуемой оценки (13), однако тем не менее является вполне приемлемой. Более того, при $\alpha < 1/2$ оценка (17) дает лучшее приближение, чем оценка (13).

Приведем алгоритм, составленный на основе неравенства (16).

1. Ввести в ячейку *x* число x_0.
2. Вычислить $\varphi(x)$ и ввести результат в ячейку $\tilde{x}$.
3. Сравнить $|x - \tilde{x}|$ и ε. Если $|x - \tilde{x}| \leq \varepsilon$, то перейти к п. 6, иначе перейти к п. 4.
4. Ввести значение $\tilde{x}$ в ячейку *x*,
5. Перейти к п. 2.
6. Выдать значение $\tilde{x}$ и остановить счет.

После того как мы познакомились с принципом сжатых отображений для общих числовых уравнений, рассмотрим конкретный пример его применения. Найдем корни уравнения

$$\cos x = ax \sin x, \qquad (18)$$

в котором a — заданная положительная постоянная. К этому уравнению приводит задача об определении температуры стержня с теплоизолированной боковой поверхностью, один конец которого теплоизолирован, а на другом конце происходит конвективный теплообмен со средой, имеющей постоянную температуру Уравнение (18) имеет бесконечно много положительных корней, однако при описании процесса изменения температуры наиболее существенным является наименьший положительный корень этого уравнения. Нетрудно догадаться, что этот корень лежит внутри интервала $(0, \pi/2)$.

Итак, требуется найти корни уравнения (18), расположенные в интервале $(0, \pi/2)$.

1°. Преобразуем уравнение (18) к виду (1). Считая, что x лежит в интервале $(0, \pi/2)$, поделим обе части уравнения (18) на $\sin x$, в результате получим

$$\operatorname{ctg} x = ax. \tag{19}$$

Формально уравнение

$$\frac{1}{a}\operatorname{ctg} x = x$$

можно считать приведенным к виду (1), однако исследование этого уравнения осложняется тем обстоятельством, что котангенс неограничен на рассматриваемом интервале. Поэтому преобразуем уравнение (19), переходя к обратным тригонометрическим функциям. Нам потребуется следующее свойство: равенство

$$\operatorname{ctg} \alpha = k$$

при $0<\alpha<\pi/2$ и $k>0$ эквивалентно равенству

$$\alpha = \pi/2 - \operatorname{arctg} k.$$

Так как по условию $a>0$, то, применяя данное свойство, из равенства (19) получаем

$$x = \pi/2 - \operatorname{arctg} ax. \tag{20}$$

Таким образом, уравнение (18) приведено к виду (1) с функцией φ, равной

$$\varphi(x) = \pi/2 - \operatorname{arctg} ax. \tag{21}$$

Обратим внимание на следующее очевидное обстоятельство: хотя мы считали при переходе от (18) к (20), что x лежит в интервале $(0, \pi/2)$, уравнения (18) и (20) являются равносильными и на отрезке $[0, \pi/2]$. Это замечание важно для применения принципа сжатых отображений, относящегося к уравнениям на замкнутых отрезках.

2°. Найдем те значения a, при которых функция φ отображает отрезок $[0, \pi/2]$ в себя. Вспомним, что при любых $a>0$ и $x\geq 0$ выполняются неравенства

$$0 \leqslant \operatorname{arctg} ax < \pi/2.$$

Следовательно, при $x \geq 0$ верны неравенства

$$0 < \pi/2 - \operatorname{arctg} ax \leqslant \pi/2.$$

Эти неравенства тем более справедливы при $x \in [0, \pi/2]$, т. е. функция φ, определенная равенством (21), при каждом $a>0$ отображает отрезок $[0, \pi/2]$ в себя.

3°. Найдем теперь те значения a, при которых функция φ удовлетворяет условию Липшица с константой $\alpha<1$.

Пусть x и y — произвольные числа из отрезка $[0, \pi/2]$. Предположим, не ограничивая общности, что

$$0 \leqslant x < y \leqslant \pi/2. \qquad (22)$$

Оценим разность $\varphi(x) - \varphi(y)$. Согласно определению (21) эта разность равна

$$\varphi(x) - \varphi(y) = \operatorname{arctg} ay - \operatorname{arctg} ax.$$

По формуле тангенса разности отсюда получаем

$$\operatorname{tg}[\varphi(x) - \varphi(y)] = \frac{ay - ax}{1 + ay \cdot ax} = \frac{a(y-x)}{1 + a^2 \cdot xy}. \qquad (23)$$

Воспользуемся еще одним соотношением из теории тригонометрических функций: для любого t из интервала $0<t<\pi/2$ выполняется неравенство

$$t < \operatorname{tg} t.$$

Положив в этом неравенстве $t = \varphi(x) - \varphi(y)$, из формулы (23) получаем

$$\varphi(x) - \varphi(y) < \operatorname{tg}[\varphi(x) - \varphi(y)] = \frac{a(y-x)}{1 + a^2 xy}.$$

Следовательно, для всех x и y, удовлетворяющих условию (22), выполняется неравенство

$$\varphi(x) - \varphi(y) < a(y-x).$$

Если же x и y — любые числа из отрезка $[0, \pi/2]$, то из этого неравенства следует оценка

$$|\varphi(x) - \varphi(y)| \leqslant a|x-y|. \qquad (24)$$

Это означает, что функция φ удовлетворяет условию Липшица с константой a. Таким образом, функция φ, определенная равенством (21), при $0<a<1$ является сжатым отображением отрезка $[0, \pi/2]$ в себя. Приведем несколько последовательных приближений, (см. рис. 9.11), которые для уравнения (20) опредеяются следующей рекуррентной формулой:

$$x_n = \pi/2 - \operatorname{arctg} a x_{n-1},$$

$$n = 1, 2, 3, \ldots$$

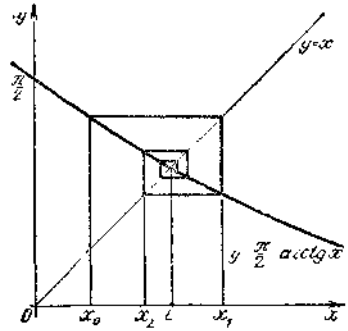


Рис. 9.11

Вычисления производились с точностью до шестого знака после запятой для a=0,5 и x_0=1.

Как видно из табл. 1, последовательные приближения с нечетными номерами уменьшаются, а с четными номерами увеличиваются и сходятся к точному значению корня c.

Таблица 1

x_1	1,107149	x_2	1,065213
x_3	1,081405	x_4	1,075119
x_5	1,077554	x_6	1,076610
x_7	1,076976	x_8	1,076834
x_9	1,076889	x_{10}	1,076868
x_{11}	1,076876	x_{12}	1,076873
x_{13}	1,076874	x_{14}	1,076874

Следовательно, числа x_n с нечетными номерами являются приближениями корня с избытком, а числа x_n с четными номерами — приближениями с недостатком. Это обстоятельство позволяет дать оценку погрешности найденных приближений. Именно для любого номера n для погрешности справедлива оценка

$$|x_n - c| \leqslant |x_n - x_{n-1}|.$$

При вычислениях на ЭВМ мы можем пользоваться этой оценкой до тех пор, пока величина $|x_n - x_{n-1}|$ не сравняется с ошибками округления. Например, из таблицы видно, что $x_{13}=x_{14}$, однако отсюда не следует, что точное значение корня c равно x_{14}. Мы можем лишь утверждать, что $c = 1{,}076\,874 \pm 10^{-6}$

После четырнадцатой итерации становится ясно, что дальнейшие вычисления с шестью знаками после запятой проводить бессмысленно.

В действительности же для расчета температуры стержня уже достигнутая точность является чрезмерной, так как полученная погрешность ничтожно мала по сравнению, например, с погрешностью, связанной с отбрасыванием других, корней уравнения (18).

Из полученной выше оценки (14) видно, что с уменьшением константы Липшица (т. е., как это следует из (24), с уменьшением a) скорость сходимости возрастает. Для иллюстрации сказанного приведем вычисления корня при a=0,1. В качестве нулевого приближения вновь возьмем x_0=1. Результаты счета приведены в табл. 2.

Таблица 2

x_1	1,471128	x_2	1,424731
x_3	1,429276	x_4	1,428830
x_5	1,428874	x_6	1,428870
x_7	1,428870		

Сравнение табл. 1 и 2 показывает, что для вычисления корня с точностью до 10^{-6} при a=0,5 потребовалось 13 итераций, а при a=0,1 — 6 итераций.

Разумеется, сделанный вывод относится только лишь к уравнению (20). Если преобразовать исходное уравнение (18) к другому уравнению вида $x=\varphi(x)$, то может оказаться, что скорость сходимости будет возрастать с увеличением a. В частности, принцип сжатых отображений может оказаться применимым и для $a \geq 1$.

Рассмотрим задачу о приближенном решении уравнения

$$x=Ax, \qquad (25)$$

где x — неизвестный элемент, принадлежащий некоторому множеству M, A — заданное отображение множества M в себя. Наглядно действие отображения A можно представить себе как перемещение точек множества M. При этом точка x переходит в точку Ax.

Предположим, что некоторая точка $\bar{x}$ никуда не переместилась, т. е. осталась неподвижной. Так как, с другой стороны, точка $\bar{x}$ должна перейти в точку $A\bar{x}$, то формально это означает, что $A\bar{x} = \bar{x}$. Теперь становится понятным следующее определение: элемент $\bar{x}$ называется *неподвижной точкой* отображения A, если выполняется равенство

$$A\bar{x} = \bar{x}.$$

Сравнение этого равенства с уравнением (25) показывает, что неподвижные точки отображения A — это те и только те элементы, которые являются решениями уравнения (25).

Итак задача о решении уравнения (25) есть задача об отыскании всех неподвижных точек отображения A.

Уравнение (25) напоминает своим видом уравнение (1) для определения корней числовых функций. Поэтому естественно попытаться применить для его решения тот же метод последовательных приближений.

Выберем произвольно элемент $x_0 \in M$ и определим последовательные приближения с помощью рекуррентной формулы

$$x_n = Ax_{n-1}, \quad n=1,2,3,\ldots \tag{26}$$

Дальнейшие рассуждения практически не отличаются от тех, которые были приведены ранее. Предположим, что последовательность x_n сходится к некоторому элементу $\bar{x} \in M$. Будет ли этот элемент неподвижной точкой отображения A?

Непосредственно из рекуррентной формулы (26) видно что для положительного ответа на этот вопрос достаточно, чтобы последовательность Ax_n сходилась к элементу $A\bar{x}$. В самом деле, устремляя в (26) n к бесконечности, в этом случае получаем

$$\bar{x} = A\bar{x}.$$

Это и означает, что $\bar{x}$ — неподвижная точка отображения A.

По аналогии с числовыми функциями отображение A называют непрерывным в точке $x \in M$, если из сходимости x_n к x следует сходимость Ax_n к Ax.

Итак, если мы хотим, чтобы последовательные приближения x_n сходились к неподвижной точке $\bar{x}$, следует потребовать, чтобы отображение A было непрерывным в точке $\bar{x}$ Поскольку мы не знаем, какая именно точка является неподвижной (в отыскании ее и состоит задача), необходимо предположить, что отображение A непрерывно в каждой точке метрического пространства M.

Однако одной лишь непрерывности недостаточно для того, чтобы отображение имело неподвижную точку. Рассмотрим, например, отображение A числовой прямой R в себя, определенное равенством

$$Ax = x+1.$$

Очевидно, что это отображение непрерывно, однако неподвижных точек у него нет, так как нет такого действительного числа x, для которого выполнялось бы равенство $x=x+1$.

Этот пример показывает, что для сходимости последовательных приближений необходимо наложить на отображение A дополнительные требования. Одно из таких требований, относящееся к числовым функциям, было приведено ранее: условие Липшица с константой $\alpha<1$. Это условие мы интерпретировали как уменьшение

расстояний между точками по крайней мере в $1/\alpha$ раз. Такая интерпретация, апеллирующая только к понятию расстояния, наводит на мысль о том, как дать соответствующее определение для произвольных отображений метрических пространств.

Отображение A метрического пространства M в себя называется сжатым, если существует положительная постоянная $\alpha < 1$, такая, что для любых элементов $x \in M$ и $y \in M$ выполняется неравенство

$$d(Ax, Ay) \leqslant \alpha d(x, y). \qquad (27)$$

Простейший пример сжатого отображения A числовой прямой в себя дается формулой

$$Ax = \frac{1}{2}x.$$

Из этой формулы получаем

$$|Ax - Ay| = \frac{1}{2}|x - y|.$$

Так как для $x \in R$ и $y \in R$ расстояние определяется равенством

$$d(x, y) = |x - y|,$$

то

$$d(Ax, Ay) = \frac{1}{2}d(x, y),$$

т. е. неравенство (27) выполняется с константой $\alpha = {}^1/_2$. Отображение A уменьшает расстояние между точками в 2 раза.

В общем случае всякое сжатое отображение можно представлять себе как сжатие метрического пространства, при котором расстояния уменьшаются по крайней мере в $1/\alpha$ раз.

Нетрудно проверить, что всякое сжатое отображение A является непрерывным в каждой точке $x \in M$. Действительно, пусть $x_n \to x$. Тогда из неравенства (27) получаем

$$d(Ax_n, Ax) \leqslant \alpha d(x_n, x).$$

Так как

$$d(x_n, x) \to 0, \text{ то } d(Ax_n, Ax) \to 0, \text{ т. е. } Ax_n \to Ax.$$

Это и означает, что отображение A непрерывно в точке x.

Всякое ли сжатое отображение метрического пространства в себя имеет неподвижную точку? Ответ на этот вопрос зависит от свойств метрического пространства, точнее, от того, является ли это пространство полным. В самом деле, если пространство неполное, то в нем может недоставать именно того элемента x, который мог быть неподвижной точкой сжатого отображения. Если же добавить требование полноты, то на сформулированный выше вопрос можно дать положительный ответ.

Именно справедливо следующее утверждение.

Всякое сжатое отображение полного метрического пространства в себя имеет неподвижную точку, и притом только одну. При любом выборе начального приближения последовательные приближения сходятся к неподвижной точке.

Приведем доказательство этого принципа, которое, как это принято в математике, разобьем на два этапа.

Итак, пусть A — сжатое отображение полного метрического пространства M в себя.

1. *Существование неподвижной точки.* Мы начнем доказательство с заключительного утверждения принципа сжатых отображений. Возьмем произвольный элемент x_0 в качестве нулевого приближения и построим последовательные приближения по рекуррентной формуле (26), т. е.

$$x_n = Ax_{n-1}, \qquad n = 1, 2, 3, \ldots$$

Докажем, что последовательность x_n фундаментальна. Из равенств

$$x_n = Ax_{n-1}, \qquad x_{n+1} = Ax_n$$

получим

$$d(x_n, x_{n+1}) = d(Ax_{n-1}, Ax_n).$$

Так как отображение A сжатое, то оно с некоторой положительной постоянной $\alpha<1$ удовлетворяет условию (27). Поэтому

$$d(x_n, x_{n+1}) \leqslant \alpha d(x_{n-1}, x^n).$$

Применяя это неравенство n раз, получаем

$$d(x_n, x_{n+1}) \leqslant \alpha d(x_{n-1}, x_n) \leqslant \alpha^2 d(x_{n-2}, x_{n-1}) \leqslant \ldots \leqslant \alpha^n d(x_0, x_1).$$

Итак, для любого номера n справедлива следующая важная оценка расстояния между соседними членами последовательности :

$$d(x_n x_{n+1}) \leqslant \alpha^n d(x_0, x_1). \tag{28}$$

Возьмем теперь любые номера n и p и оценим расстояние от x_n до x_{n+1}. Применим для этого p—1 раз неравенство треугольника:

$$d(x_n, x_{n+p}) \leqslant d(x_n x_{n+1}) + d(x_{n+1}, x_{n+2}) + \ldots + d(x_{n+p-1}, x_{n+p}).$$

Каждое слагаемое в правой части неравенства представляет собой расстояние между соседними членами последовательности x_n, поэтому для их оценки можно воспользоваться неравенством (28). В результате получим

$$d(x_n, x_{n+p}) \leqslant \alpha^n d(x_0, x_1) + \alpha^{n+1} d(x_0, x_1) + \alpha^{n+2} d(x_0, x_1) + \ldots$$
$$\ldots + \alpha^{n+p-1} d(x_0, x_1) = (1 + \alpha + \alpha^2 + \ldots + \alpha^{p-1}) \alpha^n d(x_0, x_1).$$

Следовательно,

$$d(x_n, x_{n+p}) \leqslant \frac{\alpha^n}{1-\alpha} d(x_0, x_1). \qquad (29)$$

При этом мы воспользовались формулой суммы p членов геометрической прогрессии и вытекающим из определения сжатого отображения неравенством $0<\alpha<1$:

$$1 + \alpha + \alpha^2 + \cdots + \alpha^{p-1} = \frac{1-\alpha^p}{1-\alpha} < \frac{1}{1-\alpha}.$$

Числовая последовательность α^n является бесконечно малой при $n\to\infty$, поэтому согласно неравенству (29) $d(x_n, x_{n+p})\to 0$, при $n\to\infty$ и любом p. Это означает, что

$$d(x_n, x_m)\to 0 \text{ при } n\to\infty, \ m\to\infty,$$

т. е. последовательность x_n фундаментальна.

По условию рассматриваемое метрическое пространство M полное, поэтому последовательность x_n сходится к некоторому элементу $\bar{x}$. Так как всякое сжатое отображение непрерывно, то $Ax_n\to A\bar{x}$. В таком случае, устремляя в равенстве

$$x_n = Ax_{n-1}$$

номер n к бесконечности, получаем

$$\bar{x} = A\bar{x}.$$

Тем самым доказано существование неподвижной точки и сходимость к ней последовательных приближений.

2. *Единственность неподвижной точки.* Доказательство проведем методом от противного. Если бы отображение A имело две неподвижные точки, то оно сохранило бы расстояние между ними. Но этого не может быть, так как отображение A уменьшает расстояние между любыми двумя различными точками. Принцип сжатых отображений полностью доказан.

9.5. Применение принципа сжатых отображений

В качестве первого применения принципа сжатых отображений рассмотрим вопрос о разрешимости нелинейного интегрального уравнения

$$x(t) - \lambda \int_0^t \sin x(s)\, ds = h(t), \qquad (30)$$

в котором λ — числовой параметр, $h(t)$ — заданная правая часть, $x(t)$ — неизвестная функция.

Предположим, что функция *h(t)* непрерывна на отрезке [0,1]. Выясним, при каких значениях параметра λ уравнение (30) имеет решение *x(t),* определенное на отрезке [0, 1].

Применим принцип сжатых отображений. Начнем с того, что выберем метрическое пространство. Так как функция *h(t)* непрерывна, то следует ожидать, что решение *x(t)* также непрерывно, поэтому естественно в качестве метрического пространства взять пространство *C*[0, 1] всех функций, непрерывных на отрезке [0, 1].

Теперь приведем интегральное уравнение (30) к виду

$$x=Ax,$$

где *A* — отображение пространства *C*[0, 1] в себя. Наиболее просто это можно сделать, положив

$$Ax(t) = h(t) + \lambda \int_0^t \sin x(s)\, ds. \qquad (31)$$

Итак, задача о решении интегрального уравнения (30) сведена к задаче об отыскании неподвижных точек отображения *A* определенного равенством (31). Найдем те значения параметра *X,* при которых отображение *A* является сжатым.

Пусть *x* и *y* — две произвольные функции из *C*[0, 1]. Напомним, что расстояние в *C*[0, 1] равно

$$d(x, y) = \max_{t \in [0,\ 1]} |x(t) - y(t)|.$$

Оценим расстояние *d(Ax, Ay).* Из равенства (31) получаем

$$Ax(t) - Ay(t) = \lambda \int_0^t [\sin x(s) - \sin y(s)]\, ds.$$

Отсюда

$$|Ax(t) - Ay(t)| \leqslant |\lambda| \int_0^t |\sin x(s) - \sin y(s)|\, ds. \qquad (32)$$

Для оценки разности синусов воспользуемся формулой

$$|\sin x - \sin y| \leqslant |x - y|. \qquad (33)$$

Теперь мы можем оценить интеграл в (32) следующим образом:

$$|Ax(t) - Ay(t)| \leqslant |\lambda| \int_0^t |x(s) - y(s)|\, ds.$$

По определению расстояния в *C*[0, 1] выполняется неравенство

$$|x(s) - y(s)| \leqslant d(x, y).$$

Следовательно,

$$|Ax(t) - Ay(t)| \leqslant |\lambda| d(x, y) \int_0^t ds = t\,|\lambda| \cdot d(x, y).$$

Взяв максимум левой и правой части этого неравенства по всем t из отрезка [0,1], получим

$$d(Ax, Ay) \leqslant |\lambda| d(x, y).$$

Из этого неравенства видно, что при $|\lambda|<1$ отображение A является сжатым.

В таком случае из принципа сжатых отображений следует, что при $|\lambda|<1$ уравнение (30) имеет и притом единственное решение при любой правой части $h \in C[0, 1]$. Это решение можно найти с помощью метода последовательных приближений, причем последовательные приближения сходятся к точному решению равномерно на отрезке [0,1].

Разумеется, при некоторых $h(t)$ решение уравнения (30) можно выразить явно через элементарные функции. Например, если $h(t) \equiv \pi/2$, то

$$x(t) = 2\operatorname{arc\,tg}\, e^{\lambda t}.$$

Однако ввиду того что уравнение (30) нелинейно, отыскание частных решений нисколько не помогает найти общее решение для какого-нибудь нетривиального семейства правых частей. В то же время метод последовательных приближений позволяет сколь угодно точно найти решение для любой непрерывной правой части.

Предположим теперь, что функция $h(t)$ из уравнения (30), непрерывно дифференцируема на отрезке [0, 1]. Тогда решение $x(t)$ также должно быть непрерывно дифференцируемым (это видно непосредственно из уравнения (30)), и, взяв производную левой и правой частей уравнения, получаем

$$\frac{dx}{dt} - \lambda \sin x(t) = \frac{dh}{dt}.$$

Обозначим $\frac{dh}{dt}$ через $f(t)$, тогда последнее равенство можно записать так:

$$\frac{dx}{dt} - \lambda \sin x(t) = f(t). \qquad (34)$$

Итак, функция $x(t)$ является решением нелинейного дифференциального уравнения первого порядка (34). Как известно, для определения единственного решения дифференциального уравнения первого порядка необходимо задать одно дополнительное условие, например

$$x(0)=a, \qquad (35)$$

где a — заданная постоянная. Если переменную t интерпретировать как время, то задание значения $x(0)$ можно считать начальным условием. Задача об отыскании решения уравнения (34), удовлетворяющего начальному условию (35), называется задачей Коши.

Приведенные выше соображения наводят на мысль о том, что принцип сжатых отображений можно использовать для решения задачи Коши (34)—(35). Покажем, что это действительно можно сделать.

Пусть заданы непрерывная на отрезке [0,1] функция $f(t)$ и действительное число a. Положим далее

$$h(t) = a + \int_0^t f(s)\,ds.$$

Тогда, очевидно, функция $h(t)$ непрерывно дифференцируема на отрезке [0, 1], и выполняются равенства

$$\frac{dh}{dt} = f(t), \quad h(0) = a.$$

Обозначим через $x(t)$ решение интегрального уравнения (30) с выбранной функцией $h(t)$. Тогда из приведенных выше рассуждений следует, что $x(t)$ является решением дифференциального уравнения (34). Для проверки условия (35) положим в интегральном уравнении (30) $t=0$:

$$x(0) = h(0) = a.$$

Следовательно, $x(t)$ является искомым решением задачи Коши. Все эти рассуждения можно провести в обратном порядке, в результате чего получаем следующее утверждение.

Если $|\lambda|<1$, то задача Коши (34)—(35) для любой правой части $f \in C[0, 1]$ и любого начального значения $a \in R$ имеет, и притом единственное, решение.

Естественно, возникает вопрос о том, нельзя ли применить принцип сжатых отображений для более сложных дифференциальных уравнений. Рассмотрим одно из таких уравнений, описывающих движение математического маятника.

Математический маятник представляет собой материальную точку массы m, подвешенную на невесомой нерастяжимой нити длины l. Обозначим через φ угол, на который отклоняется маятник от вертикали (см. рис. 9.12).

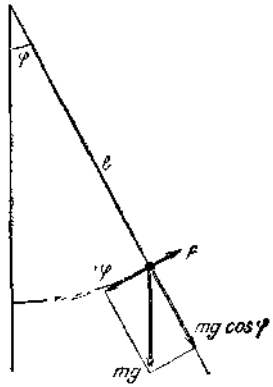


Рис. 9.12

Предположим, что на точку m действует сила $F(t)$, известным образом зависящая от времени t и направленная по касательной к окружности, по которой движется точка. Кроме того, точка находится под действием силы тяжести mg, направленной вертикально вниз. Разложим эту силу на две составляющие: силу $mg\sin\varphi$, действующую вдоль касательной, и силу $mg\cos\varphi$, действующую вдоль направления нити. Так как из-за нерастяжимости нити движение может происходить только по окружности, то сила, направленная вдоль нити, уравновешивается ее натяжением.

Ускорение точки равно $l\ddot{\varphi}$ (мы используем стандартные обозначения $\dot{\varphi} = \frac{d\varphi}{dt}$, $\ddot{\varphi} = \frac{d^2\varphi}{dt^2}$), поэтому согласно второму закону Ньютона

$$ml\ddot{\varphi} = F - mg\sin\varphi. \tag{36}$$

Это и есть уравнение движения математического маятника под действием заданной силы $F(t)$. Перепишем его, введя следующие, более удобные для наших целей обозначения:

$$x(t) = \varphi(t), \quad \lambda = g/l, \; f(t) = F(t)/m.$$

Тогда уравнение (36) примет следующий вид:

$$\ddot{x}(t) + \lambda\sin x(t) = f(t). \tag{37}$$

К этому уравнению необходимо добавить начальные условия

$$x(0) = \alpha, \qquad \dot{x}(0) = \beta. \tag{38}$$

Задача Коши для уравнения (37) заключается в отыскании решения этого уравнения, удовлетворяющего начальным условиям (38). При ее

исследовании одной из наиболее трудных проблем является доказательство существования решения. Обычно предполагается, что маятник при своем движении отклоняется на малый угол, и поэтому можно приближенно записать

$$\sin x \approx x.$$

В результате такого упрощения, называемого линеаризацией, уравнение (37) заменяется линейным дифференциальным уравнением с постоянными коэффициентами

$$\ddot{x}(t)+\lambda x(t)=f(t),$$

описывающим гармонические колебания.

Попытаемся тем не менее доказать существование решения задачи Коши для исходного нелинейного уравнения (37). Аналогично тому, как это было сделано для уравнения (34), попробуем свести уравнение (37) к интегральному уравнению. Проинтегрируем для этого обе части уравнения (37) 2 раза от 0 до t. После первого интегрирования получим

$$\dot{x}(t)-\dot{x}(0)+\lambda\int_0^t \sin x(s)\,ds=\int_0^t f(s)\,ds.$$

Так как

$$\dot{x}(0)=\beta,$$

то полученное равенство можно записать так:

$$\dot{x}(t)+\lambda\int_0^t \sin x(s)\,ds=\beta+\int_0^t f(s)\,ds.$$

Второе интегрирование приведет к следующему равенству:

$$x(t)-x(0)+\lambda\int_0^t du\int_0^u \sin x(s)\,ds=\beta t+\int_0^t du\int_0^u f(s)\,ds. \quad (39)$$

Учитывая, что $x(0)=\alpha$, введем новую функцию

$$h(t)=\alpha+\beta t+\int_0^t du\int_0^u f(s)\,ds. \quad (40)$$

Преобразуем последнее слагаемое в левой части (39), для чего поменяем порядок интегрирования:

$$\int_0^t du\int_0^u \sin x(s)\,ds=\int_0^t \sin x(s)\,ds\int_s^t du=\int_0^t (t-s)\sin x(s)\,ds.$$

Тогда с учетом (40) равенство (39) можно записать в следующем виде:

$$x(t) + \lambda \int_0^t (t-s) \sin x(s)\, ds = h(t). \qquad (41)$$

Таким образом, мы показали, что каждое решение задачи Коши (37)—(39) является решением интегрального уравнения (40) с правой частью *h,* определенной равенством (40). Замечательно то, что верно и обратное: если функция *h(t)* определена равенством (40), то решение интегрального уравнения (41) является решением задачи Коши (37)—(38). Это проверяется непосредственно двукратным дифференцированием уравнения (41).

Итак, нами установлена эквивалентнось задачи Коши (37)—(38) интегральному уравнению (41). Для исследования уравнения (41) применим принцип сжатых отображений.

Предположим, что требуется найти решение интегрального уравнения (41) для всех значений *t,* принадлежащих отрезку [0, *T*], где *T* — заданное число. Возьмем в качестве метрического пространства C[0, *T*] пространство всех функций, непрерывных на отрезке [0, *T*]. Приведем уравнение (41) к виду $x=Ax$. Проще всего сделать это, положив

$$Ax(t) = h(t) - \lambda \int_0^t (t-s) \sin x(s)\, ds. \qquad (42)$$

Определенное этим равенством отображение *A,* очевидно, переводит метрическое пространство *C*[0, *T*] в себя. Выясним, при каких *K* отображение *A* является сжатым.

Возьмем две произвольные функции *x(t)* и *y(f)* из *C*[0, *T*]. Из равенства (42) получаем

$$Ax(t) - Ay(t) = \lambda \int_0^t (t-s) [\sin y(s) - \sin x(s)]\, ds.$$

Отсюда

$$|Ax(t) - Ay(t)| < |\lambda| \int_0^t (t-s) |\sin y(s) - \sin x(s)|\, ds.$$

Для оценки подынтегрального выражения применим неравенство (33):

$$|\sin y(s) - \sin x(s)| \leqslant |y(s) - x(s)|.$$

В результате получим

$$|Ax(t) - Ay(t)| \leqslant |\lambda| \int_0^t (t-s) |y(s) - x(s)|\, ds.$$

Так как $|y(s)-x(s)|\leqslant d(x, y)$, то

$$|Ax(t)-Ay(t)|\leqslant|\lambda|\cdot d(x,y)\int_0^t(t-s)\,ds=|\lambda|\cdot\frac{t^2}{2}\cdot d(x,y).$$

Взяв максимум левой и правой части, окончательно получаем

$$d(Ax, Ay)\leqslant\frac{|\lambda|T^2}{2}d(x,y).$$

Из этого неравенства следует, что при

$$|\lambda|<2/T^2 \tag{43}$$

отображение A является сжатым. Следовательно, при выполнении условия (43) интегральное уравнение (41) для любой правой части $h\in C[0, T]$ имеет, и притом единственное, решение непрерывное на отрезке $[0, T]$.

В силу установленной выше эквивалентности, аналогичное заключение можно сделать и относительно задачи Коши для дифференциального уравнения (37). Именно если выполняется условие (43), то задача Коши (37)—(38) для любой функции $f\in C[0, T]$ и любых начальных $\alpha\in R$ и $\beta\in R$ имеет, и притом единственное, решение, дважды непрерывно дифференцируемое на отрезке $[0, T]$.

Проанализируем внимательней условие (43). Для этого перепишем его в виде

$$T<\sqrt{\frac{2}{|\lambda|}}. \tag{44}$$

Записанное в таком виде, оно может быть интерпретировано следующим образом: для любого $\lambda\neq0$ задача Коши (37)—(38) имеет решение на отрезке $[0, T]$, где T удовлетворяет неравенству (44). Казалось бы, никакой новой информации при этом мы не получаем, однако, как это ни парадоксально, такой взгляд на условие (43) позволяет для *любого* $\lambda\in R$ доказать существование решения на *всей* полупрямой $t\geq0$.

Действительно, из приведенных выше рассуждений следует, что задача Коши для уравнения (37) имеет решение на любом отрезке длины T, т. е. для любого отрезка вида $[a, a+T]$ существует решение уравнения (37), удовлетворяющее условиям

$$x(a)=\bar{\alpha},\qquad \dot{x}(a)=\bar{\beta},$$

где $\bar{\alpha}$ и $\bar{\beta}$ — произвольные действительные числа. Отсюда следует, что решение исходной задачи Коши (37)—(38), которое существует на отрезке $[0, T]$, можно продолжить так, что оно будет определено и на

отрезке [0, 2T], а значит, и на отрезках [0, 3T], [0, 4T], ..., т. е. решение можно определить на всей полупрямой [0, ∞].

Чрезвычайно важно подчеркнуть, что в условие (44) не входят начальные данные α и β. Именно благодаря этому счастливому обстоятельству оказалось возможным продолжить решение на всю полупрямую. Если же длина интервала, на котором удается построить решение, зависит от начальных данных, то далеко продолжить решение не всегда удается.

Рассмотрим, например, задачу Кошн для дифференциального уравнения

$$x(t)—x^2(t)=0 \qquad (16)$$

с начальным условием

$$x(0)=a. \qquad (17)$$

Решение этой задачи находится элементарным интегрированием.

$$x(t)=a/(1—at).$$

Других решений, удовлетворяющих начальному условию $x(0)=a$, нет. Непосредственно видно, что решение определено только при $0\le t<1/a$ и не может быть продолжено за точку $t=1/a$. Обратите внимание на то, что с увеличением начального значения a интервал, на котором определено решение, укорачивается.

Приведенные примеры показывают широту области применения принципа сжатых отображений. В некоторых случаях этот принцип применяется сразу (допустим, для решения интегральных уравнений), а в других требуется предварительная работа. Например, непосредственно для решения дифференциальных уравнений принцип сжатых отображений неприменим. Это объясняется тем, что дифференциальный оператор, как правило, увеличивает расстояние между функциями, и поэтому определяемое им отображение метрического пространства в себя не может быть сжатым. Поэтому дифференциальное уравнение вначале заменяют эквивалентным ему интегральным, а затем уже применяют принцип сжатых отображений.

Впервые метод последовательных приближений для получения решения интегрального уравнения был применен К. Нейманом в 70-х годах позапрошлого столетия. Им было доказано, что последовательные приближения сходятся к точному решению интегрального уравнения вида

$$x—Tx=h,$$

где T — линейный интегральный оператор с достаточно малым ядром. В современной терминологии это означает, что отображение T является сжатым.

Через несколько лет появилась работа французского математика Э. П и к а р а (1856—1941), в которой задача Коши для нелинейного

дифференциального уравнения сводилась к нелинейному интегральному уравнению вида

$$x=Ax.$$

В этой работе Пикаром были найдены условия, обеспечивающие выполнение неравенства

$$\max_t |Ax(t)-Ay(t)| \leqslant \alpha \max_t |x(t)-y(t)|$$

с некоторой постоянной $\alpha<1$. По существу, это было первое явно сформулированное условие сжатости отображения A. При выполнении этого условия Пикар доказал существование и единственность решения задачи Коши и сходимость последовательных приближений к решению.

Стремление найти общий подход к различным способам приближенного решения интегральных уравнений привел М. Фреше к понятию абстрактного метрического пространства и к задаче об отыскании неподвижных точек отображений метрических пространств. Под влиянием этих работ в начале прошлого столетия было получено много изящных теорем о неподвижных точках, но среди них принципа сжатых отображений не было.

В 1922 г. вышла, ставшая классической, работа польского математика С. Банаха (1892—1945). В этой работе были заложены основы теории линейных полных метрических пространств, в которых расстояние между точками не менялось при их сдвиге.

$$d(x, y)=d(x+z, y+z).$$

Сейчас эти пространства называются банаховыми. Трудно переоценить значение работ Банаха для развития современной математики. Мы отметим лишь одно утверждение в упомянутой работе Банаха, относящееся к рассматриваемой нами проблеме: если T — сжатое отображение банахова пространства в себя, то уравнение

$$x-Tx=h$$

имеет единственное решение при любой правой части h. Для линейного отображения T решение уравнения дается рядом

$$x=h+Th+T^2h+T^3h+\ \ldots,$$

который носит название ряда Неймана.

После работ Банаха для многих математиков стало ясно, что принцип сжатых отображений справедлив не только в линейных, но и в произвольных полных метрических пространствах.

Принцип сжатых отображений относится к тем математическим утверждениям, которые удивляют контрастом между чрезвычайной простотой доказательства и глубиной получаемых с их помощью результатов. Одним из замечательных достоинств этого принципа

является его конструктивный характер: ***указывается способ построения приближенных решений со сколь угодно высокой точностью***. Вместе с тем важно подчеркнуть, что метод последовательных приближений может оказаться применимым и в тех случаях, когда отображение не является сжатым.

Вообще изучение неподвижных точек произвольного непрерывного отображения (не обязательно сжатого) оказывается несравненно более трудным, чем для сжатого отображения. Пример тождественного отображения говорит о том, что единственности в этом случае ожидать не приходится. Доказательство же существования требует чрезвычайно тонких рассуждений. Первый результат в этом направлении принадлежит голландскому математику Л. Брауэру (1881—1966). Им было доказано, что каждое непрерывное отображение конечномерного шара в себя имеет хотя бы одну неподвижную точку. Это утверждение получило название принципа Брауэра неподвижной точки.

Приведем одну из наглядных интерпретаций теоремы Брауэра, получившую, пожалуй, наибольшее распространение. Предположим, что к каждой точке поверхности шара прикреплен отрезок единичной длины. Отрезок может располагаться под любым углом к касательной плоскости, важно только, чтобы при переходе от одной точки к другой угол менялся непрерывно. Тогда теорема Брауэра утверждает, что всегда найдется отрезок, расположенный перпендикулярно к касательной плоскости. Действительно, если x любая точка на поверхности шара, а точка y — проекция на поверхность шара конца отрезка, выходящего из x, то отображение, переводящее x в y, непрерывное. Точка x является неподвижной точкой этого отображения только в том случае, когда отрезок, выходящий из x, направлен вдоль радиуса, соединяющего x с центром шара, т. е. когда отрезок перпендикулярен плоскости, касающейся поверхности шара в точке x. Кратко эти рассуждения можно резюмировать так: ***при любом расчесывании шара будут оставаться вихры***.

Сейчас имеется несколько доказательств теоремы Брауэра, однако ни одно из них не может быть изложено в популярной форме. Исключение составляет одномерный вариант теоремы Брауэра, который формулируется так: каждое непрерывное отображение отрезка числовой прямой в себя имеет по крайней мере одну неподвижную точку.

Приведем доказательство этого утверждения. Пусть φ — непрерывное отображение отрезка [0,1] в себя. Иначе говоря, φ — непрерывная функция, определенная на отрезке [0, 1] и удовлетворяющая неравенствам

$$0 \le \varphi(x) \le 1. \qquad (47)$$

Рассмотрим функцию

$$f(x)=x-\varphi(x).$$

Из неравенств (47) получаем

$$f(0)=-\varphi(0)\le 0;$$
$$f(1)=1-\varphi(1)\ge 0.$$

Так как функция f непрерывна на отрезке [0,1] и принимает на левом конце значение $f(0)\le 0$, а на правом конце — значение $f(1)\ge 0$, то на отрезке [0, 1] найдется точка $\bar{x}$, в которой $f(\bar{x})=0$. Следовательно

$$\bar{x} - \varphi(\bar{x})=0.$$

т. е. $\bar{x}$ — неподвижная точка отображения φ.

В приведенном доказательстве существенно использовалась упорядоченность множества действительных чисел. Уже в двухмерном случае доказательство теоремы Брауэра становится значительно сложнее.

Тот факт, что в теореме Брауэра говорится об отображениях шара, является несущественным. Вместо шара можно рассматривать эллипсоид, куб, тетраэдр и вообще любое ***выпуклое тело***. ***Условие выпуклости играет существенную роль,*** и просто его отбросить нельзя. Например, тор (бублик) — невыпуклое тело. Повернув его на угол, меньший 360°, получим непрерывное отображение тора в себя, не имеющее неподвижных точек.

В отличие от принципа сжатых отображений теорема Брауэра носит неконструктивный характер в том смысле, что в ней не указан способ приближенного нахождения решения. Это так называемая чистая теорема существовании, примером которой может служить утверждение: в лесу есть дерево с наибольшим количеством листьев. Однако и такие теоремы имеют важное прикладное значение; знание того, что решение существует, позволяет применить известные методы приближенного нахождения решений. Другой, более существенный недостаток принципа Брауэра состоит в том, что он применяется только для отображений конечномерных пространств, в то время как наиболее употребительные функциональные пространства являются бесконечномерными. Например, рассмотренное выше пространство $C[a, b]$ всех функций, непрерывных на отрезке $[a, b]$, бесконечномерно. Это означает, что существует бесконечная последовательность непрерывных функций $f_1(t), f_2(t), f_3(t), \ldots$, таких, что ни одна из $f_n(t)$ не может быть представлена в виде суммы

$$f_n(t) = c_1 f_1(t) + c_2 f_2(t) + \ldots + c_{n-1} f_{n-1}(t)$$

с некоторыми постоянными $c_1, c_2, \ldots, c_{n-1}$. В качестве такой последовательности можно взять, например, последовательность, образованную функциями $f_n(t)=t^n$. Действительно, предположим, что

эта последовательность не обладает нужными свойствами. Тогда для некоторого номера n найдутся постоянные c_1, c_2, ..., c_{n-1}, такие, что равенство

$$t^n = c_1 t + c_2 t^2 + \ldots + c_{n-1} t^{n-1}$$

будет выполняться для всех t из отрезка $[a, b]$. Следовательно, для всех $t \in [a, b]$будет выполняться равенство

$$t^n - c_1 t - c_2 t^2 - \ldots - c_{n-1} t^{n-1} = 0$$

Иначе говоря, многочлен степени n, стоящий в левой части, имеет бесконечно много корней, чего не может быть, так как в силу известный теоремы алгебры любой многочлен степени n не может иметь более n корней. Полученное противоречие показывает, что пространство $C[a, b]$ бесконечномерно.

Формально теорема Брауэра на бесконечномерные пространства не распространяется, так как существуют непрерывные отображения бесконечномерного шара в себе, не имеющие неподвижных точек. С другой стороны, бесконечномерные пространства (например, $C[a, b]$), с необходимостью возникают при решении различных дифференциальных и интегральных уравнений. Для распространения теоремы Брауэра на бесконечномерные пространства потребовалось понятие *компактного* множества, введенное в начале 20-х годов прошлого столетия советскими математиками П. С. Александровым и П. С. Урысоном. Данное множество элементов банахова пространства компактно, если его можно сколь угодно точно аппроксимировать конечномерными множествами. В 1927 году польский математик Ю. П. Шаудер доказал существование неподвижной точки у каждого непрерывного отображения выпуклого компактного подмножества банахова пространства в себя. Окончательный результат в этом направлении принадлежит А. Н. Тихонову, доказавшему в 1935 г. существование неподвижной точки у непрерывного отображения выпуклого компактного множества в любом локально-выпуклом линейном топологическом пространстве. Эта на первый взгляд слишком абстрактная теорема была успешно применена А. Н. Тихоновым для исследования систем бесконечного числа дифференциальных уравнений, связанных с задачами математической физики.

Последний пример служит наглядным свидетельством того, что, как показывает вся история развития математики, именно те математические задачи, которые возникли в результате формализации реальных естественнонаучных проблем, оказываются наиболее интересными объектами математического исследования.

9.6. Приближенное решение конечных уравнений

1. Введение. Мы будем рассматривать здесь численное решение уравнений вида

$$f(x) = 0, \qquad (1)$$

где f—заданная функция. Такие уравнения могут быть *алгебраическими,* если функция f алгебраическая, или *трансцендентными* в противном случае; как те, так и другие называются *конечными* в отличие, например, от дифференциальных уравнений. Мы укажем лишь несколько наиболее универсальных методов решения уравнений вида (1); другие методы читатель может найти в курсах приближенных вычислений.

Численное решение уравнения (1) обычно начинают с нахождения грубого, совсем приближенного решения, так называемого *нулевого приближения.* Если решается физическая задача, то это решение может быть известно из физического смысла задачи. Можно начать с примерного, хотя бы довольно грубого построения графика функции f. Если при этом обнаружится, что на каком-нибудь интервале a, b эта функция всюду определена, не имеет точек разрыва и принимает в точках a и b значения противоположных знаков, то в силу свойств непрерывных функций f должна иметь на этом интервале хотя бы один нуль, т. е. уравнение (1) имеет там по крайней мере один корень. Если к тому же функция f на этом интервале монотонна, то такой корень здесь только один, т. е. этот корень *отделен* от остальных. Если обозначить этот, пока неизвестный, корень через α, то можно ручаться, что $a<\alpha<b$. Для дальнейшего уточнения значения α применяются различные методы (см. п. 2).

Часто бывает удобнее переписать уравнение (1) в форме $\varphi(x) = \psi(x)$, после чего искать пересечение графиков $y= \varphi(x)$ и $y = \psi(x)$. При этом левую часть уравнения (1) стараются разбить так, чтобы получились хорошо известные или, во всяком случае, более или менее простые графики; иногда помогает замена неизвестной x,

Рассмотрим, например, уравнение

$$\operatorname{tg} ax^2 - bx^2 = 0, \qquad (2)$$

где a, b — заданные положительные постоянные. Замена $ax^2 = s$ приводит к уравнению

$$\operatorname{tg} s = ks \qquad \left(k = \frac{b}{a}\right). \qquad (3)$$

Графики левой и правой частей показаны на рис. 1, причем ясно, что нас интересуют лишь значения ≥ 0.

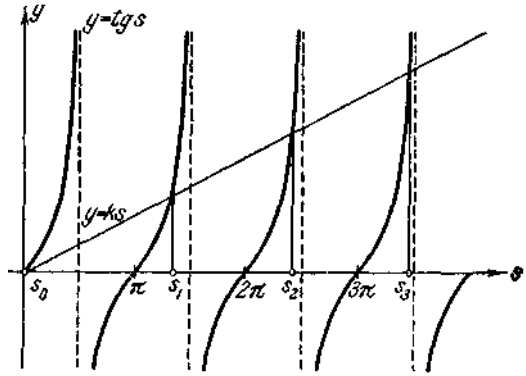


Рис. 1

Мы видим, что уравнение (3), а с ним и (2), имеет бесконечное число решений $s_0 = 0 < s_1 < s_2 < \ldots$, причем на графике хорошо видна зависимость решений от k, т. е. от a и b. В частности, видно, что при $k > 1$ на интервале

$$0 < s < \frac{\pi}{2}$$

появляется новое решение.

Легко получить *асимптотическое выражение* для решения s_n уравнения (3), пригодное для больших n. Пусть для определенности $k < 1$. Тогда из рис. 1 получаем искомое выражение $s_n = n\pi + \frac{\pi}{2} - \alpha_n$, где $\alpha_n \to 0$ при $n \to \infty$; в других обозначениях:

$$s_n = n\pi + \frac{\pi}{2} + o(1).$$

Если мы хотим уточнить это разложение, то надо его подставить в (3), что даст

$$\operatorname{tg}\left(n\pi + \frac{\pi}{2} - \alpha_n\right) = k\left(n\pi + \frac{\pi}{2} - \alpha_n\right)$$

и после преобразований

$$\cos\alpha_n = k\left(n\pi + \frac{\pi}{2} - \alpha_n\right)\sin\alpha_n.$$

Отсюда (4)

$$\alpha_n \sim \sin\alpha_n = \frac{\cos\alpha_n}{k\left(n\pi + \frac{\pi}{2} - \alpha_n\right)} \sim \frac{1}{k\pi n}, \text{ т. е. } \alpha_n = \frac{1}{k\pi n} + o\left(\frac{1}{n}\right).$$

Если нужно дальнейшее уточнение, то можно, например, в формуле (4) обозначить $\frac{1}{n} = t \to 0$, $\alpha_n = \alpha(t) \xrightarrow[t \to 0]{} 0$, что даст

$$t \cos \alpha = k \left[\pi + \left(\frac{\pi}{2} - \alpha \right) t \right] \sin \alpha \qquad (\alpha = \alpha(t);\ \alpha(0) = 0),$$

после чего написать первые члены разложения $\alpha(t)$ в ряд Маклорена (вида

$$f(x) = f(0) + \frac{f'(0)}{1!} x + \frac{f''(0)}{2!} x^2 + \dots,$$

), но по степеням t).

Вычисления, которые мы предоставляем читателю, дают

$$\alpha = \frac{1}{k\pi} t - \frac{1}{2k\pi} t^2 + \frac{1}{k\pi} \left(\frac{1}{4} + \frac{1}{k\pi^2} - \frac{1}{3k^2\pi^2} \right) t^3 + \dots = \frac{1}{k\pi n} - \frac{1}{2k\pi n^2} + \dots$$

Применение формулы

$$(1+x)^a = 1 + \binom{a}{1} x + \binom{a}{2} x^2 + \binom{a}{3} x^3 + \dots$$

дает асимптотическое выражение для положительных решений уравнения (2) при больших n:

$$x_n = \sqrt{\frac{s_n}{a}} = a^{-1/2} \sqrt{n\pi + \frac{\pi}{2} - \frac{1}{k\pi n} + \frac{1}{2k\pi n^2} - \dots} =$$

$$= \left(\frac{n\pi}{a} \right)^{1/2} \left(1 + \frac{1}{2n} - \frac{1}{k\pi^2 n^2} + \dots \right)^{1/2} =$$

$$= \left(\frac{n\pi}{a} \right)^{1/2} \left[1 + \frac{1}{4n} - \left(\frac{1}{2k\pi^2} + \frac{1}{32} \right) \frac{1}{n^2} + \dots \right].$$

2. Методы проб, хорд и касательных. *Метод проб,* с которого часто начинают, состоит в следующем. Пусть для определенности $f(a)<0$; $f(b)>0$. Тогда берут произвольно значение c между a и b и вычисляют $f(c)$, причем тут существен *только знак* $f(c)$. Допустим, что получится $f(c)>0$. Это значит, что произошел «перелет», следовательно, $a < \alpha < c$. Тогда берут какое-либо значение d между a и c, вычисляют $f(d)$; если $f(d)<0$, то произошел «недолет», т.е. $d < \alpha < c$, и т. д. При этом значения c, d, ... берутся более или менее произвольными, удобными для вычисления; правда, если, например, $|f(a)|$ значительно меньше, чем $f(b)$, то довольно вероятно, что α окажется ближе к a, чем к b, и поэтому c следует взять поближе к a и т. п.

Метод хорд состоит в том, что в качестве c берется не произвольная точка, а (рис. 2) точка пересечения оси x с хордой графика, проведенной через точки $M[a; f(a)]$ и $N[b; f(b)]$.

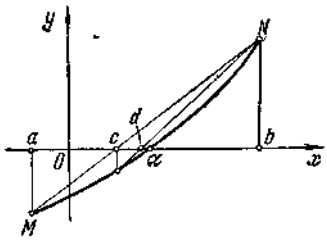


Рис. 2

Другими словами, мы как бы приближенно принимаем дугу графика за отрезок прямой, т. е. производим *линейную интерполяцию,* что является достаточно обоснованным, если интервал *a, b* не слишком велик. Для нахождения точки *c* напишем уравнение хорды *MN*:

$$\frac{y-f(a)}{f(b)-f(a)}=\frac{x-a}{b-a},$$

а затем, положив $y=0$, найдем соответствующее значение $x=c$:

$$c=a-\frac{f(a)(b-a)}{f(b)-f(a)}=b-\frac{f(b)(b-a)}{f(b)-f(a)}. \qquad (5)$$

Если необходимо, это построение можно повторить (см. рис. 2).

В *методе касательных* (он же называется *методом Ньютона)* за *c* берется точка пересечения оси *x* с касательной, проведенной к графику в одном из концов рассматриваемой дуги. Уравнение касательной, изображенной на рис. 3, имеет вид

$$y-f(b)=f'(b)(x-b),$$

откуда, полагая $y = 0$, найдем

$$c=b-\frac{f(b)}{f'(b)}. \qquad (6)$$

И здесь построение можно повторить (см. рис. 3).

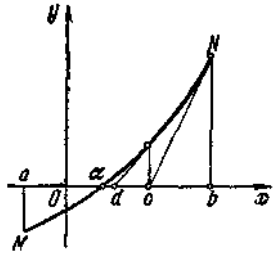


Рис. 3

Метод Ньютона можно истолковать независимо от его геометрического смысла. Обозначим нулевое приближение решения через x_0 и разложим левую часть (1) по степеням $x—x_0$ в силу формулы Тейлора

$$f(x)=f(a)+\frac{f'(a)}{1!}(x-a)+\frac{f''(a)}{2!}(x-a)^2+\ldots$$

мы получим уравнение

$$f(x_0)+\frac{f'(x_0)}{1!}(x-x_0)+\frac{f''(x_0)}{2!}(x-x_0)^2+\ldots=0.$$

Если произвести линеаризацию, т. е. отбросить члены выше первого порядка малости, получим *линеаризованное уравнение* (1):

$$f(x_0)+f'(x_0)(x-x_0)=0.$$

Решение этого линеаризованного уравнения

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

можно принять за первое приближение решения уравнения (1); мы приходим к той же формуле (6). Из первого приближения можно получить второе по формуле

$$x_2=x_1-\frac{f(x_1)}{f'(x_1)} \qquad (7)$$

и т. д. Метод Ньютона всегда приводит к цели, если только нулевое приближение не лежит слишком далеко от искомого решения.

Иногда применяется следующий вариант метода Ньютона: знаменатель формулы (7), а также формул для дальнейших приближений заменяют на $f(x_0)$; геометрически это означает, что все наклонные прямые на рис. 3 проводят параллельно касательной в исходной точке *N*. Метод в этом варианте сходится несколько хуже, но подсчет каждого приближения, естественно, упрощается.

Комбинированный метод основан на том соображении, что если рассматриваемый участок графика не имеет ни изломов, ни точек перегиба, то метод хорд и метод касательных дают точки, расположенные по разные стороны от искомого корня. Если, например, график расположен, как на рис. 4, то, отправляясь от интервала a, b, можно построить точку a_1 по методу хорд, а точку b_1 по методу касательных, в результате чего получится новый интервал a_1, b_1, на котором лежит искомый корень α.

Проделав аналогичное построение на интервале a_1, b_1, получим новый интервал a_2, b_2, содержащий искомый корень, и т. д.

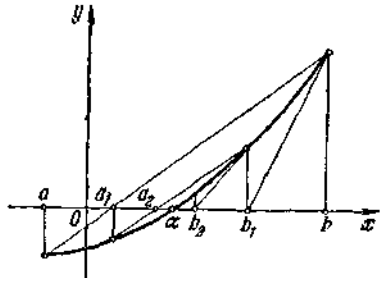


Рис. 4

При этом получается *двустороннее приближения* к этому корню, которое обрывается при достижении требуемой точности. Рассмотрим, например, уравнение

$$x^3+x^2-3=0, \qquad (8)$$

коэффициенты которого будем считать совершенно точными. Исследование производных показывает (проверьте!), что при $-\infty < x < -\frac{2}{3}$ левая часть, которую мы обозначим через *f(x),* возрастает от

$$-\infty \text{ до } -2\frac{23}{27},$$

затем при

$$-\frac{2}{3} < x < 0$$

убывает до —3 и далее возрастает до ∞ и имеет единственную точку перегиба при $x=-\frac{1}{3}$. Значит, уравнение имеет единственный вещественный и притом положительный корень α. Так как *f*(0) = —3, *f*(1)=—1, *f*(2) = 9 (рис. 5), то $1 < \alpha < 2$.

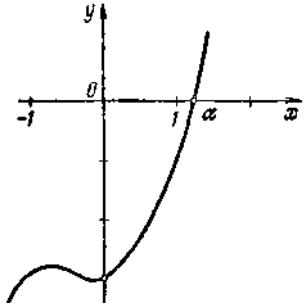


Рис. 5

Согласно методу проб вычисляем $f(1,1)$=—0,459; $f(1,2)$= 0,168, т. е. $1,1 < \alpha < 1,2$ (грубая *«прикидка»* корня проводится с помощью метода проб). Полагая $a = 1,1$; $b = 1,2$, применяем формулы (5) и (6) согласно комбинированному методу:

$$a_1 = 1,2 - \frac{0,168 \cdot 0,1}{0,168 + 0,459} \approx 1,173,$$

$$b_1 = 1,2 - \frac{0,168}{6,72} = 1,175.$$

Таким образом, с точностью до 0,001 можно положить $\alpha = 1,174$. Если эта точность недостаточна, то можно провести дальнейшее вычисление: $f(1,174)$=—0,003628 («недолет»; вычисление с точностью до 10^{-6}); $f(1,175) = 0,002859$. Приняв a= 1,174, b=1,175, получаем по комбинированному методу после вычислений с точностью до 10^{-7}: $a_2 = 1,1745593$; b_2= 1,1745596. Таким образом, с точностью до 0,000001 можно положить α= 1,174559.

3. Метод итераций. Методы, описанные в п. 2, принадлежат к числу итерационных методов (иначе говоря, методов последовательных приближений), в которых некоторый единообразный процесс последовательно повторяется *(«итерируется»,* от латинского «итерацио» — повторение), в результате чего получаются все более точные приближенные решения. Это единообразие имеет многочисленные удобства, в частности, в применении ЭВМ.

В общем виде в применении к уравнению (1) метод итераций выглядит так: уравнение переписывайся в равносильной форме:

$$x = \varphi(x). \qquad (9)$$

Затем выбирается некоторое значение $x = x_0$ в качестве нулевого приближения; желательно, чтобы оно было по возможности ближе к искомому решению, если о последнем что-либо известно. Последующие приближения вычисляются по формулам $x_1 = \varphi(x_0)$, $x_2 = \varphi(x_1)$, . . . , вообще

$$x_{n+1} = \varphi(x_n). \qquad (10)$$

При этом может быть два случая.

1) *Процесс может сходиться,* т. е. последовательные приближения x_n при $n \to \infty$ стремятся к некоторому пределу x; в этом случае, переходя в формуле (10) к пределу при $n \to \infty$, видим, что $x = \bar{x}$ является решением уравнения (9).

2) *Процесс может расходиться,* т. е. конечного предела построенных «приближений» существовать не будет. Из этого отнюдь не следует, что и решения уравнения (9) не существует, просто могло оказаться, что процесс итераций выбран неудачно. (Впрочем, и в случае сходимости бывает, что в пределе получается не то решение,

около которого мы выбрали x_0, а другое, быть может, даже не имеющее физического смысла.)

Поясним сказанное на простом примере уравнения, которое можно решить без всякой «науки»,

$$x = \frac{x}{2} + 1 \tag{11}$$

С очевидным решением $\bar{x} = 2$. Если положить $x_0 = 0$ и вычислять с точностью до 0,001, то получим $x_1 = 1{,}000$; $x_2 = 1{,}500$; $x_3 = 1{,}750$; $x_4 = 1{,}875$; $x_5 = 1{,}938$; $x_6 = 1{,}969$; $x_7 = 1{,}984$; $x_8 = 1{,}992$; $x_9 = 1{,}996$; $x_{10} = 1{,}998$; $x_{11} = 1{,}999$; $x_{12} = 2{,}000$; $x_{13} = 2{,}000$, т.е. процесс практически сошелся.

Если взамен (11) рассмотреть уравнение

$$x = \frac{x}{10} + 1$$

и принять $x_0 = 0$, то с точностью до 0,001 будет $x_1 = 1{,}000$; $x_2 = 1{,}100$; $x_3 = 1{,}110$; $x_4 = 1{,}111$; $x_5 = 1{,}111$, т. е. процесс практически сошелся уже после четырех итераций.

Если уравнение (11) разрешить относительно x, стоящего в правой части, т. е. переписать в равносильной форме

$$x = 2x - 2, \tag{12}$$

и начать с $x_0 = 0$, то мы получим последовательность $x_1 = -2$, $x_2 = -6$, $x_3 = -14$ и т. д., т. е. процесс сходиться не будет. Это можно было предвидеть, заметив, что из (10) вытекает равенство

$$x_{n+1} - x_n = \varphi(x_n) - \varphi(x_{n-1}), \tag{13}$$

т. е. $x_2 - x_1 = \varphi(x_1) - \varphi(x_0)$; $x_3 - x_2 = \varphi(x_2) - \varphi(x_1)$ и т. д.

Если значения функции меняются медленнее, чем значения аргумента, точнее, если

$$|\varphi(x) - \varphi(\tilde{x})| \leqslant k\,|x - \tilde{x}| \qquad (k = \text{const} < 1), \tag{14}$$

то расстояния между последовательными приближениями будут быстро стремиться к нулю и процесс итераций сходится, притом тем быстрее, чем меньше k. Неравенство (14) должно выполняться для всех x, $\bar{x}$ либо, во всяком случае, вблизи искомого корня $\bar{x}$ уравнения (9). В п. 4 будет показано, что неравенстов (14) выполняется, если $|\varphi'(x)| \leq k\varphi$

Мы видим, что уравнения (11) и (12) полностью равносильны, но порождают различные итерационные процессы. И в других случаях уравнение (1) можно переписать в форме (9) многими способами, каждый их которых порождает свой итерационный процесс, причем одни из этих процессов могут оказаться быстросходящимися и потому наиболее удобными, другие — медленно сходящимися, а третьи —

даже вовсе расходящимися. В частности, легко проверить, что если уравнение (1) записать в виде

$$x = x - \frac{f(x)(b-x)}{f(b)-f(x)},$$

то, если начинать с $x_0 = a$ [см. формулу (5)], получится метод хорд, а если уравнение (1) записать в виде

$$x = x - \frac{f(x)}{f'(x)}, \qquad (15)$$

то получится метод касательных.

В более сложных примерах, чем были разобраны выше, чаще всего не проводят теоретического доказательства сходимости процесса итераций, а просто вычисляют несколько приближений и по их виду делают вывод о сходимости или расходимости процесса. Если сочтут, что какое-либо приближение достаточно мало отличеется от предыдущего, например, если отличие выходит за рамки принятой степени точности, то процесс итераций обрывают. Во всяком случае, это свидетельствует о том, что достигнутое приближение удовлетворяет уравнению (9) с хорошей точностью, так как если $|x_n - x_{n+1}| < h$, то и $|x_n - \varphi(x_n)| < h$.

4. Формула конечных приращений. Неравенство (14) можно проверить с помощью так называемой формулы конечных приращений, которую мы сейчас выведем. Допустм, что функция $y = \varphi(x)$ на интервале a, b (включая концы) пепреритш вместе со своей производной. Рассмотрим (рис. 6) график этой функции на интервале $a \le x \le b$, проведем хорду MN, стягивающею его концы, и допустим для определенности, что график хотя бы частично расположен над этой хордой.

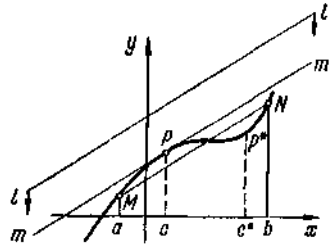


Рис. 6

Проведем тогда выше графика прямую $ll \| MN$ и станем ее непрерывно опускать, оставляя параллельной MN. Тогда в некотором положении она коснется графика в точке P, т. е. *на гладкой дуге непременно найдется по крайней мере одна точка, в которой каса-*

тельная параллельна хорде, стягивающей эту дугу. Если приравнять угловые коэффициенты хорды и касательной, то мы получим

$$\frac{\varphi(b)-\varphi(a)}{b-a}=\varphi'(c), \quad \text{т. е.} \quad \varphi(b)-\varphi(a)=\varphi'(c)(b-a), \quad (16)$$

где c — некоторая точка между a и b. Формула (16) называется *формулой конечных приращений* (так как расстояние a от b может не быть малым) или *теоремой Лагранжа* по имени выдающегося французского математика и механика Ж. Лагранжа (1736—1813). Отметим, что значение c, фигурирующее в формуле (16), для данной функции и данного интервала a, b никак не является произвольным, хотя для c может получиться несколько пригодных значений. Например, для рис. 6 в формуле (16) вместо c можно взять c^*, так как в точке P^* касательная также параллельна хорде MN. При применении формулы (16) значение c обычно неизвестно, однако о c часто достаточно знать, что оно находится где-то между a и b.

Например, пусть дано, что на некотором интервале $|\varphi'(x)|\leq k$. Тогда, применяя формулу (16) к любым двум точкам x и $\bar{x}$ этого интервала, увидим, что для них

$$|f(x)-f(\tilde{x})|\leqslant k|x-\tilde{x}|$$

[см. формулу (14)].

Из формул (13) и (16) вытекает также, что если последовательные приближения находятся недалеко от точного решения $\bar{x}$, так что $\varphi'(x)$ меняется мало, то *процесс итераций сходится* примерно *со скоростью геометрической прогрессии* со знаменателем $\varphi'(x)$. Если бы разности между последовательными приближениями образовывали точно геометрическую прогрессию, как в примере (11), то ее первый член $a = x'_1 - x_0$, а знаменатель $q=\frac{x_2-x_1}{x_1-x_0}$. Поэтому сумма всей прогрессии, т. е. $\bar{x}=x_0$, равнялась бы

$$\frac{a}{1-q}=\frac{x_1-x_0}{1-\frac{x_2-x_1}{x_1-x_0}}=\frac{(x_1-x_0)^2}{2x_1-x_0-x_2},$$

откуда

$$\bar{x}=x_0+\frac{(x_1-x_0)^2}{2x_1-x_0-x_2}=\frac{x_1^2-x_0x_2}{2x_1-x_0-x_2}. \quad (17)$$

В более сложных примерах последовательные разности лишь напоминают геометрическую прогрессию. Тогда формула (17) не дает точного решения, но дает возможность «перескочить» через несколько приближений и получить приближенное значение решения, от которого можно вновь начать итерации. Особую роль играет

итерационный процесс Ньютона. В самом деле, производная от правой части (15), т. е.

$$1-\frac{f'f'-ff''}{f'^2}=\frac{ff''}{f'^2},$$

обращается в нуль при $x = \bar{x}$, так как $f(\bar{x}) = 0$. Значит, в силу предыдущего *метод Ньютона сходится быстрее геометрической прогрессии с любым знаменателем.* Скорость этой сходимости легко установить на следующем простом типичном примере. Пусть рассматриваются приближения по способу Ньютона к нулевому корню уравнения $x+x^2=0$. Эти приближения связаны друг с другом соотношением

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

Для оценки скорости сходимости заменим это приближенное равенство на точное; тогда последовательно получим,

$$x_1=x_0^2,\ x_2=x_1^2=x_0^4,\ x_3=x_2^2=x_0^8 \text{ и т. д.}$$

вообще $x_n=x_0^{2^n}$. При $|x_0|<1$ правая часть с увеличением n стремится к нулю быстрее любой экспоненты.

5. Метод малого параметра. *Метод малого параметра,* он же *метод возмущений,* как и метод итераций, представляет собой один из наиболее универсальных методов в математике и заключается в следующем. Пусть формулировка некоторой задачи, помимо основных неизвестных величии, содержит некоторый параметр α, причем эта задача при каком-то значении $\alpha = \alpha_0$ может быгь более или менее легко решена *(невозмущенное решение).* Тогда решение задачи при α, близких к α_0 *(возмущенное решение),* во многих случаях можно получить разложенным по степеням $\alpha - \alpha_0$ с той или иной степенью точности. При этом первый член разложения, не содержащий $\alpha - \alpha_0$, получается при и $\alpha=\alpha_0$, т. е. дает невозмущенное решение. Дальнейшие же члены дают поправки на «возмущение» решения; эти поправки имеют первый, второй и т. д. порядки малосги (по сравнению с $\alpha - \alpha_0$). Эти члены обычно находятся *по методу неопределенных коэффициентов,* т. е. коэффициенты при $(\alpha - \alpha_0)$, $(\alpha - \alpha_0)^2$ и т. д. обозначаются какими-то буквами, которые находятся затем из условий задачи. Этот метод дает хороший результат только при α, близких к α_0, при этом чем $|\alpha - \alpha_0|$ меньше, тем меньше членов разложения нужно вычислять; так как часто принимают $\alpha_0=0$, то отсюда и происходит название метода. Следует иметь также в виду, что при больших $|\alpha - \alpha_0|$ метод можег привести к принципиальным ошибкам, так как может получиться, что отбрасываемые члены более существенны, чем оставляемые.

Таким образом, метод малого параметра дает возможность, исходя из решения некоторых «узловых» задач, получить решение задач, формулировка которых близка к этим узловым, если, конечно, изменение формулировки не влечет за собой принципиального, качественного изменения решения. Во многих задачах уже вид первого члена, содержащего параметр, дает возможность сделать полезные выводы о зависимости решения от параметра при его малом изменении.

Пример. Найдем решение уравнения

$$x^3 - \alpha x^2 + 1 = 0 \qquad (18)$$

при малых | α | с точностью до величин поряда α^3 включительно. Для этого заметим, что при α = 0 получается уравнение $x^3+1=0$ с очевидным решением $x_0 = -1$. Поэтому пишем

$$x_\alpha = -1 + a\alpha + b\alpha^2 + c\alpha^3 + \text{члены высшего порядка малости.}$$

Подставляя это выражение в (18) и выписывая члены только до α^3, получим (проверьте!)

$$[-1 + 3a\alpha + 3b\alpha^2 - 3a^2\alpha^2 - 6ab\alpha^3 + 3c\alpha^3 + a^3\alpha^3] -$$

$$-\alpha(1 - 2a\alpha - 2b\alpha^2 + a^2\alpha^2) + 1 + \text{члены высшего порядка малости} = 0.$$

Отсюда, приравнивая коэффициенты при одинаковых степенях α, получим $3a - 1 = 0$, $3b - 3a^2 + 2a = 0$, $-6ab + 3c + a^3 + 2b - a^2 = 0$ и последовательно находим $a = \frac{1}{3}$, $b = -\frac{1}{9}$, $c = \frac{2}{81}$, т. е. решение уравнения (18)

$$x_\alpha = -1 + \frac{\alpha}{3} - \frac{\alpha^2}{9} + \frac{2\alpha^3}{81} \qquad (19)$$

с точностью до величин высшего порядка малости относительно α.

Тог же результат можно получить, *применяя непосредственно формулу Тейлора* с измененными обозначениями:

$$x_\alpha = x_0 + \left(\frac{dx}{d\alpha}\right)_0 \alpha + \frac{1}{2!}\left(\frac{d^2x}{d\alpha^2}\right)_0 \alpha^2 + \frac{1}{3!}\left(\frac{d^3x}{d\alpha^3}\right)_0 \alpha^3; \qquad (20)$$

индекс «нуль» указывает на подстановку α=0. Для этого дифференцируем равенство (18) по α:

$$3x^2\frac{dx}{d\alpha} - x^2 - 2\alpha x\frac{dx}{d\alpha} = 0,$$

$$6x\left(\frac{dx}{d\alpha}\right)^3 + 3x^2\frac{d^2x}{d\alpha^2} - 4x\frac{dx}{d\alpha} - 2\alpha\left(\frac{dx}{d\alpha}\right)^2 - 2\alpha x\frac{d^2x}{d\alpha^2} = 0,$$

$$6\left(\frac{dx}{d\alpha}\right)^3 + 18x\frac{dx}{d\alpha}\frac{d^2x}{d\alpha^2} + 3x^2\frac{d^3x}{d\alpha^3} - 6\left(\frac{dx}{d\alpha}\right)^3 - 6x\frac{d^2x}{d\alpha^2} - 6\alpha\frac{dx}{d\alpha}\frac{d^2x}{d\alpha^2} - 2\alpha x\frac{d^3x}{d\alpha^3} = 0.$$

Подставляя α = 0, x=—1, получим

$$3\left(\frac{dx}{d\alpha}\right)_0 - 1 = 0, \quad -6\left(\frac{dx}{d\alpha}\right)_0^2 + 3\left(\frac{d^2x}{d\alpha^2}\right)_0 + 4\left(\frac{dx}{d\alpha}\right)_0 = 0,$$
$$6\left(\frac{dx}{d\alpha}\right)_0^3 - 18\left(\frac{dx}{d\alpha}\right)_0\left(\frac{d^2x}{d\alpha^2}\right)_0 + 3\left(\frac{d^3x}{d\alpha^3}\right)_0 - 6\left(\frac{dx}{d\alpha}\right)_0^2 + 6\left(\frac{d^2x}{d\alpha^2}\right)_0 = 0,$$

откуда получаем последовательно

$$\left(\frac{dx}{d\alpha}\right)_0 = \frac{1}{3}; \qquad \left(\frac{d^2x}{d\alpha^2}\right)_0 = -\frac{2}{9}; \qquad \left(\frac{d^3x}{d\alpha^3}\right)_0 = \frac{4}{27}.$$

Отсюда из формулы (20) вытекает разложение (19), дающее хорошую точность при малых $|\alpha|$.

Метод малого параметра непосредственно связан с методом итераций п. 3, что мы продемонстрируем на том же примере (18). Удобно, чтобы невозмущенное решение было нулевым; это достигается с помощью подстановки $x = -1+y$, откуда

$$-1 + 3y - 3y^2 + y^3 - \alpha + 2\alpha y - \alpha y^2 + 1 = 0,$$

т. е.

$$y = \frac{1}{3}\alpha - \frac{2}{3}\alpha y + y^2 + \frac{1}{3}\alpha y^2 - \frac{1}{3}y^3.$$

Если теперь проводить итерации, начиная от значения $y_0=0$ и отбрасывая в разложении члены выше третьего порядка малости, то за три шага мы придем к требуемому разложению. При этом легко проверить, что в каждом приближении можно отбрасывать члены, порядок малости которых выше номера приближения.

Ниже мы укажем некоторые *методы построения приближенных формул* для решения, аналогичные описанным ранее методам решения конечных уравнений, а также *методы численного решения,* в которых искомое частное решение строится в табличном виде. Мы будем для простоты рассматривать уравнения первого порядка, но те же методы естественно переносятся на уравнения любого порядка и на системы уравнений.

Метод итераций. Рассмотрим дифференциальное уравнение первого порядка с заданным начальным условием

$$y' = f(x,\ y), \quad y(x_0) = y_0. \tag{21}$$

Если взять интегралы от обеих частей уравнения, получим

$$\int_{x_0}^{x} y'\,dx = y - y_0 = \int_{x_0}^{x} f(x,\ y)\,dx = \int_{x_0}^{x} f(x,\ y(x))\,dx.$$

Изменяя обозначение переменной интегрирования, напишем

$$y(x) = y_0 + \int_{x_0}^{x} f(s,\ y(s))\,ds. \tag{22}$$

Уравнение (22) равносильно сразу обоим равенствам (21), так как после его дифференцирования получится первое равенство, а после подстановки $x = x_0$ — второе. Уравнение (22) является *интегральным уравнением,* так как в нем неизвестная функция стоит под знаком интеграла.

Вид уравнения (22) удобен для применения метода итераций, хотя сейчас неизвестной является не число, а функция. Выбрав некоторую функцию $y_0(x)$ в качестве нулевого приближения (желательно, чтобы она была по возможности ближе к искомому решению; если о последнем ничего не известно, то можно положить хотя бы $y_0(x) \equiv (y_0)$, находим первое приближение по формуле

$$y_1(x) = y_0 + \int_{x_0}^{x} f(s,\ y_0(s))\, ds.$$

Подставляя результат в правую часть (22), находим второе приближение и т. д.; вообще

$$y_{n+1}(x) = y_0 + \int_{x_0}^{x} f(s,\ y_n(s))\, ds \qquad (n = 0,\ 1,\ 2,\ \ldots). \quad (23)$$

Если процесс итераций сходится, т. е. если последовательные приближения стремятся с ростом n к некоторой предельной функции, то она удовлетворяет уравнению (22); для проверки этого надо в равенстве (23) перейти к пределу при $n \to \infty$. Замечательно, что *метод итераций для уравнения* (22), *как правило, сходится для всех x, достаточно близких к* x_0; так будет, во всяком случае, если выполнены условия теоремы Коши из п. 3. Это связано с тем, что при вычислении последующих приближений надо интегрировать предыдущие, а при последовательном интегрировании функции в целом «сглаживаются» и всякие неправильности, происходящие из-за выбора нулевого приближения, погрешностей округления и т. п., постепенно устраняются. В отличие от этого при последовательном дифференцировании функции, как правило, ухудшаются, первоначальные неправильности разрастаются и поэтому итерационный метод, основанный на последовательном дифференцировании, не дал бы сходимости.

Различие между интегрированием и дифференцированием показано на рис. 7.

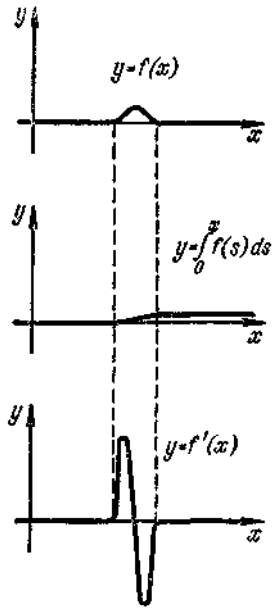


Рис. 7

Нарисованный «горбик», добавленный к какой-либо функции, значительно портит ее производную, не говоря уже о последующих производных (какой вид имеет вторая производная $y = f''(x)$?) и почти не сказывается на интеграле.

Рассмотрим, например, частный вид уравнения Риккати

$$y' = x^2 + y^2$$

при начальном условии $y(0) = 0$. После интегрирования получим

$$y(x) = \frac{x^3}{3} + \int_0^x y^2(s)\,ds.$$

Выберем в качестве нулевого приближения для искомого решения, о котором мы пока ничего не знаем, нулевую функцию $y_0(x) = 0$, так как она удовлетворяет хотя бы начальному условию. Тогда получим

$$y_1(x) = \frac{x^3}{3}, \quad y_2(x) = \frac{x^3}{3} + \int_0^x \left(\frac{s^3}{3}\right)^2 dx = \frac{x^3}{3} + \frac{x^7}{63},$$

$$y_3(x) = \frac{x^3}{3} + \frac{x^7}{63} + \frac{2x^{11}}{2079} + \frac{x^{15}}{59535}$$

и т. д. Видно, что при небольших x, например при $|x| < 1$, процесс хорошо сходится; так, с точностью до 0,001 при $|x| < 1$ можно положить $y = \frac{x^3}{3} + \frac{x^7}{63}$, а при $|x| < \frac{1}{2}$ даже просто

$$y = \frac{x^3}{3}.$$

Как обычно на практике, вопрос о том, на каком приближении нужно остановиться, решается с помощью сравнения последующих приближений с предыдущими.

Применение ряда Тейлора. Из уравнения и начального условия (21) можно с помощью дифференцирования найти значения $y'(x_0)$, $y''(x_0)$ и т. д., после чего составить разложение решения в степенной ряд Тейлора. Необходимое количество членов определяется с помощью их последовательного вычисления и сравнения с выбранной степенью точности.

Рассмотрим, например, задачу

$$y' = x^2 + y^2, \quad y(0) = 1.$$

Подстановкой в правую часть уравнения находим, что

$$y'(0) = 0^2 + 1^2 = 1.$$

Если продифференцировать обе части уравнения, получим $y'' = 2x + 2yy'$ и, подставив $x=0$, найдем

$$y''(0) = 2 \cdot 0 + 2 \cdot 1 \cdot 1 = 2.$$

Аналогично находим

$$y''' = 2 + 2y'^2 + 2yy''; \quad y'''(0) = 8; \quad y^{IV} = 6y'y'' + 2yy''', \quad y^{IV}(0) = 28$$

и т. д. Подставляя это в формулу Маклорена, получим

$$y = y(0) + \frac{y'(0)}{1!}x + \frac{y''(0)}{2!}x^2 + \ldots = 1 + x + x^2 + \frac{4}{3}x^3 + \frac{7}{6}x^4 + \ldots$$

Этой формулой можно пользоваться для небольших $|x|$

Применение степенных рядов с неопределенными коэффициентами. Этот метод тесно связан с предыдущим методом и состоит в том, что решение уравнения ищется в форме ряда с неизвестными коэффициентами

$$y = a + b(x - x_0) + c(x - x_0)^2 + d(x - x_0)^3 + \ldots, \qquad (24)$$

которые находятся с помощью подстановки в уравнение и последующего приравнивания коэффициентов при одинаковых степенях x (и применения начального условия, если оно задано). Рассмотрим, например, уравнение второго порядка

$$y'' + xy = 0. \qquad (25)$$

Будем искать решение разложенным по степеням x:

$$y = a_0 + a_1x + a_2x^2 + a_3x^3 + \ldots \qquad (26)$$

После дифференцирования и подстановки в уравнение получим

$$(1\cdot 2a_2 + 2\cdot 3a_3 x + 3\cdot 4a_4 x^2 + \ldots) + x(a_0 + a_1 x + a_2 x^2 + \ldots) = 0.$$

Приравнивание коэффициентов при одинаковых степенях x дает

$$1\cdot 2a_2 = 0; \quad 2\cdot 3a_3 + a_0 = 0; \quad 3\cdot 4a_4 + a_1 = 0;$$
$$4\cdot 5a_5 + a_2 = 0; \quad 5\cdot 6a_6 + a_3 = 0, \ \ldots,$$

откуда последовательно находим

$$a_2 = 0; \quad a_3 = -\frac{a_0}{2\cdot 3}; \quad a_4 = -\frac{a_1}{3\cdot 4}; \quad a_5 = -\frac{a_2}{4\cdot 5} = 0;$$
$$a_6 = -\frac{a_3}{5\cdot 6} = \frac{a_0}{2\cdot 3\cdot 5\cdot 6}; \quad a_7 = -\frac{a_4}{6\cdot 7} = \frac{a_0}{3\cdot 4\cdot 6\cdot 7}; \quad a_8 = -\frac{a_5}{7\cdot 8} = 0;$$
$$a_9 = -\frac{a_6}{8\cdot 9} = -\frac{a_0}{2\cdot 3\cdot 5\cdot 6\cdot 8\cdot 9} \text{ и т. д.}$$

Подстановка этих результатов в формулу (26) дает общее решение уравнения (25):

$$y = a_0 + a_1 x - \frac{a_0}{2\cdot 3}x^3 - \frac{a_1}{3\cdot 4}x^4 + \frac{a_0}{2\cdot 3\cdot 5\cdot 6}x^6 +$$
$$+ \frac{a_1}{3\cdot 4\cdot 6\cdot 7}x^7 - \frac{a_0}{2\cdot 3\cdot 5\cdot 6\cdot 8\cdot 9}x^9 - \ldots =$$
$$= a_0\left(1 - \frac{x^3}{2\cdot 3} + \frac{x^6}{2\cdot 3\cdot 5\cdot 6} - \frac{x^9}{2\cdot 3\cdot 5\cdot 6\cdot 8\cdot 9} + \ldots\right) +$$
$$+ a_1\left(x - \frac{x^4}{3\cdot 4} + \frac{x^7}{3\cdot 4\cdot 6\cdot 7} - \frac{x^{10}}{3\cdot 4\cdot 6\cdot 7\cdot 9\cdot 10} + \ldots\right).$$

Константы a_0 и a_1 остаются в качестве произвольных постоянных. Ряды, стоящие в скобках, представляют два линейно независимых частных решения уравнения (25).

Описанный прием всегда применим, в частности, к линейным уравнениям

$$a_0(x)\,y^{(n)} + a_1(x)\,y^{(n-1)} + \ldots + a_n(x)\,y = f(x), \qquad (27)$$

если все функции $a_0(x)$, $a_1(x)$, $\ldots$, $f(x)$ представляют собой многочлены относительно x или, в более общем случае, суммы рядов по степеням $x - x_0$, причем $a_0(x_0) \neq 0$.

Если $a_0(x_0)=0$, то значение x_0 называется *особой точкой* для уравнения (27); тогда найти решение в форме (24) возможно не всегда. В этом случае иногда удается найти решение в форме

$$y = (x - x_0)^\rho [a + b(x - x_0) + c(x - x_0)^2 + d(x - x_0)^3 + \ldots], \qquad (28)$$

где постоянная ρ также подбирается. При этом можно считать, что $a \neq 0$, так как в противном случае можно вынести за скобку некоторую степень $x - x_0$, так что дело сведется к изменению ρ.

Функции Бесселя. Рассмотрим важный пример уравнения Бесселя

$$x^2y'' + xy' + (x^2 - p^2)\,y = 0 \qquad (p = \text{const} \geqslant 0,\ 0 < x < \infty). \tag{29}$$

Решения этого уравнения называются *функциями Бесселя,* хотя они применялись Эйлером с 1766 г., т. е. до рождения Бесселя. Эти функции называются также *цилиндрическими* функциями, так как они широко применяются при решении уравнений математической физики в круглом цилиндре. Значение $x = 0$ является для уравнения (29) особой точкой, поэтому в силу формулы (28), в которой надо положить $x_0 = 0$, его решение можно искать в виде

$$y = ax^\rho + bx^{\rho+1} + cx^{\rho+2} + \ldots \tag{30}$$

Дифференцирование и подстановка в уравнение (29) дают

$$\begin{aligned} x^2[a\rho(\rho-1)x^{\rho-2} + b(\rho+1)\rho x^{\rho-1} + c(\rho+2)(\rho+1)x^\rho + \ldots] + \\ + x[a\rho x^{\rho-1} + b(\rho+1)x^\rho + c(\rho+2)x^{\rho+1} + \ldots] + \\ + x^2(ax^\rho + bx^{\rho+1} + cx^{\rho+2} + \ldots) - p^2(ax^\rho + bx^{\rho+1} + cx^{\rho+2} + \ldots) = 0. \end{aligned}$$

После приравнивания нулю коэффициентов при одинаковых степенях x получим цепочки равенств:

$$\begin{aligned} &a\rho(\rho-1) + a\rho - ap^2 = 0, \text{ т. е. } a(\rho^2 - p^2) = 0; \\ &b(\rho+1)\rho + b(\rho+1) - bp^2 = 0, \text{ т. е. } b(\rho^2 + 2\rho + 1 - p^2) = 0; \\ &c(\rho+2)(\rho+1) + c(\rho+2) + a - cp^2 = 0, \text{ т. е. } c(\rho^2 + 4\rho + 4 - p^2) + a = 0; \\ &d(\rho+3)(\rho+2) + d(\rho+3) + b - dp^2 = 0, \text{ т. е. } d(\rho^2 + 6\rho + 9 - p^2) + b = 0; \\ &e(\rho+4)(\rho+3) + e(\rho+4) + c - ep^2 = 0, \text{ т. е. } e(\rho^2 + 8\rho + 16 - p^2) + c = 0 \end{aligned}$$

и т. д. Из первого равенства, поскольку $a \neq 0$, мы видим, что $\rho^2 = p^2$, т. е. $\rho = \pm p$. Подставляя этот результат в остальные равенства, получим последовательно

$$\begin{aligned} &b = 0, \quad c = \frac{-a}{4\rho+4} = -\frac{a}{2^2(\rho+1)}, \quad d = 0, \\ &e = -\frac{c}{8\rho+16} = \frac{a}{2^4 \cdot 2(\rho+1)(\rho+2)}, \quad f = 0, \\ &g = -\frac{a}{2^6 \cdot 2 \cdot 3(\rho+1)(\rho+2)(\rho+3)}, \quad i = 0, \\ &j = \frac{a}{2^8 \cdot 2 \cdot 3 \cdot 4(\rho+1)(\rho+2)(\rho+3)(\rho+4)} \quad \text{и т. д.} \end{aligned}$$

Отсюда в силу формулы (30) находим решение

$$\begin{aligned} y = ax^\rho - \frac{a}{2^2(\rho+1)}x^{\rho+2} + \frac{a}{2^4 \cdot 2!(\rho+1)(\rho+2)}x^{\rho+4} - \\ - \frac{a}{2^6 \cdot 3!(\rho+1)(\rho+2)(\rho+3)}x^{\rho+6} + \ldots, \end{aligned} \tag{31}$$

где a — произвольная постоянная Удобно выбрать

$$a = \frac{1}{2^\rho \Gamma(\rho+1)}$$

Если учесть, что в силу формулы $\Gamma(p+1) = p\Gamma(p)$

$$\Gamma(\rho+1)(\rho+1)=\Gamma(\rho+2);\quad \Gamma(\rho+1)(\rho+1)(\rho+2)=$$
$$=\Gamma(\rho+2)(\rho+2)=\Gamma(\rho+3)\ \text{и т. д.},$$

то формула (31) при таком a даст

$$y=\frac{1}{\Gamma(\rho+1)}\left(\frac{x}{2}\right)^{\rho}-\frac{1}{1!\Gamma(\rho+2)}\left(\frac{x}{2}\right)^{\rho+2}+\frac{1}{2!\Gamma(\rho+3)}\left(\frac{x}{2}\right)^{\rho+4}-$$
$$-\frac{1}{3!\Gamma(\rho+4)}\left(\frac{x}{2}\right)^{\rho+6}+\ldots=\sum_{n=0}^{\infty}\frac{(-1)^n}{n!\Gamma(\rho+n+1)}\left(\frac{x}{2}\right)^{\rho+2n}. \quad (32)$$

Эта сумма называется *функцией Бесселя* 1*-го рода порядка* ρ и обозначается через $J_{\rho}(x)$. Так как ρ =±p, то общее решение уравнения (29) можно записать в виде

$$y=C_1J_p(x)+C_2J_{-p}(x). \quad (33)$$

Решение (171) не годится при целом p = 0, 1, 2, 3, ... Действительно, для таких p при ρ = — p будет

$$\Gamma(-p+1)=\Gamma(-p+2)=\ldots=\Gamma(-p+p)=\pm\infty,$$

и потому формула (34) даст

$$J_{-p}(x)=\sum_{n=p}^{\infty}\frac{(-1)^n}{n!(-p+n)!}\left(\frac{x}{2}\right)^{-p+2n}=|\,n-p=n'\,|=$$
$$=\sum_{n'=0}^{\infty}\frac{(-1)^p(-1)^{n'}}{(p+n')!(n')!}\left(\frac{x}{2}\right)^{p+2n'}=(-1)^pJ_p(x)\quad(p=0,\ 1,\ 2,\ \ldots).$$

Значит, в этом случае решения $J_p(x)$ и $J_{-p}(x)$ линейно зависимые, а потому формула (33) не дает общего решения.

Чтобы получить общее решение уравнения (29), пригодное для всех p, поступают так, сначала считают, что p не целое, и образуют функцию

$$Y_p(x)=\operatorname{ctg}p\pi J_p(x)-\frac{1}{\sin p\pi}J_{-p}(x)=\frac{\cos p\pi J_p(x)-J_{-p}(x)}{\sin p\pi}.$$

Как линейная комбинация решений она также является решением уравнения (29) и называется *функцией Бесселя* 2*-го рода порядка p,* она иногда обозначается также через $N_p(x)$. Если p становится целым, то в правой часги получается неопределенность (почему?). Ее можно раскрыть по правилу Лопиталя, чего мы здесь делать не будем. Отметим только, что в итоге получится сумма, для которой при $x\to0$ старшим членом будет

$$-\frac{(p-1)!\,2^p}{\pi x^p}\quad(p=1,\ 2,\ 3,\ \ldots);\qquad\frac{2}{\pi}\ln x\quad(p=0).$$

Итак, формула

$$y=C_1J_p(x)+C_2Y_p(x) \quad (34)$$

дает общее решение уравнения (29) для всех $p\geq0$, нецелых или целых, при $0<x<\infty$. При этом $J_p(+0)$ конечно, тогда как $Y_p(+0)=-\infty$.

Поэтому если $y(+0)$ по условиям задачи должно быть конечным, то в правой части формулы (34) надо оставить лишь первое слагаемое.

Функции Бесселя подробно изучены и затабулированы. Наибольшее значение для приложений имеют

$$\left.\begin{aligned} J_0(x) &= 1 - \frac{x^2}{1^2 2^2} + \frac{x^4}{2^2 2^4} - \frac{x^6}{3^2 2^6} + \dots, \\ J_1(x) &= \frac{x}{2} - \frac{x^3}{1^2 2^3} + \frac{x^5}{2^2 3^2 2^5} - \frac{x^7}{3^2 4^2 2^7} + \dots \end{aligned}\right\} \qquad (35)$$

Примерные графики этих функций показаны на рис. 8.

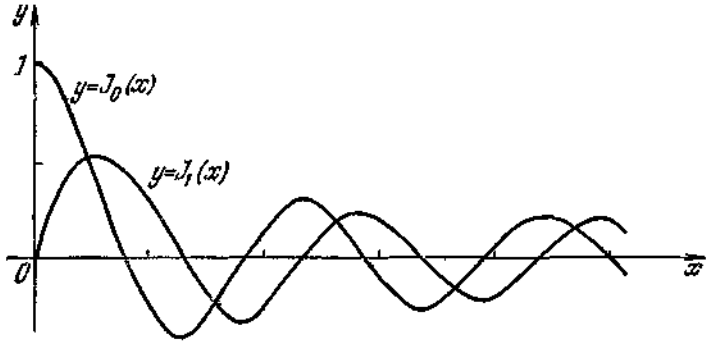


Рис. 8

Эти функции, как и все функции Бесселя 1-го и 2-го рода при возрастании x бесконечное число раз меняют знак и стремятся к нулю. Из формул (35) легко вывести, что $J_0'(x) = -J_1(x)$ (проверьте!). Имеются и другие соотношения между функциями Бесселя.

Метод малого параметра. Этот метод, описанный ранее, применяется и при решении дифференциальных уравнений. Приведем простые примеры

Задача

$$y' = \frac{x}{1 + 0{,}1xy}, \quad y(0) = 0 \qquad (36)$$

не содержит параметров. Однако можно рассмотреть более общую задачу

$$y' = \frac{x}{1 + \alpha xy}, \quad y(0) = 0, \qquad (37)$$

из которой (36) получается при $\alpha = 0{,}1$. Задача (37) легко решается при $\alpha = 0$. тогда получается $y = \frac{x^2}{2}$. Поэтому ищем решение задачи разложенным в ряд по степеням α, т. е.

$$y=\frac{x^2}{2}+\alpha u+\alpha^2 v+\alpha^3 w+\ldots, \tag{38}$$

где $u = u(x)$, $v = v(x)$ и т д.— пока неизвестные функции x.

Подстановка (38) в (37) дает после умножения на знаменатель

$$(x+\alpha u'+\alpha^2 v'+\alpha^3 w'+\ldots)\left(1+\frac{\alpha}{2}x^2+\alpha^2 xu+\alpha^3 xv+\ldots\right)=x; \tag{39}$$

$$\alpha u(0)+\alpha^2 v(0)+\ldots=0, \text{ т. е. } u(0)=0, \quad v(0)=0, \quad w(0)=0, \ldots \tag{40}$$

Раскрывая скобки в (39) и приравнивая нулю коэффициенты при степенях α, получим последовательно

$$u'+\frac{1}{2}x^4=0, \quad v'+\frac{x^3}{2}u'+x^2u=0, \quad w'+\frac{x^3}{2}v'+xuu'+x^2v=0 \text{ и т. д.},$$

откуда с учетом равенств (40) найдем (проверьте')

$$u=-\frac{x^5}{10}, \quad v=\frac{7}{160}x^8, \quad w=\frac{71}{1760}x^{11} \text{ и т. д.}$$

Поэтому формула (38) дает

$$y=\frac{x^2}{2}-\frac{\alpha}{10}x^5+\frac{7\alpha^2}{160}x^8-\frac{71\alpha^3}{1760}x^{11}+\ldots$$

В частности, для уравнения (36) получим

$$y=\frac{x^2}{2}-\frac{x^5}{100}+\frac{7x^8}{16\,000}-\frac{71x^{11}}{1\,760\,000}+\ldots$$

Этот ряд прекрасно сходится при $|x|<1$ и неплохо при $1<|x|<2$.

Рассмотрим в качестве другого примера задачу

$$y'=\sin(xy), \qquad y(0)=\alpha \tag{41}$$

В отличие от предыдущего примера здесь параметр входит в начальное условие. При $\alpha=0$ задача (41) имеет, очевидно, решение $y=0$. Поэтому при малых $|\alpha|$ ищем решение в форме

$$y=\alpha u+\alpha^2 v+\alpha^3 w+\ldots \qquad (u=u(x), \ v=v(x), \ \ldots). \tag{42}$$

Подстановка значения $x = 0$ дает

$$u(0)=1, \quad v(0)=0, \quad w(0)=0, \ \ldots \tag{43}$$

С другой стороны, подставив (42) в дифференциальное уравнение (41), получим с учетом степенного ряда для синуса:

$$\alpha u'+\alpha^2 v'+\alpha^3 w'+\ldots=$$
$$=\frac{(\alpha xu+\alpha^2 xv+\alpha^3 xw+\ldots)}{1!}-\frac{(\alpha xu+\alpha^2 xv+\alpha^3 xw+\ldots)^3}{3!}+\ldots$$

Приравнивание коэффициентов при одинаковых степенях α дает

$$u'=xu, \quad v'=xv, \quad w'=xw-\frac{x^3u^3}{3!}, \ \ldots$$

Интегрируя эти уже *линейные* уравнения с учетом начальных условий (43), найдем (проверьте!)

$$u = e^{\frac{x^2}{2}}, \quad v = 0, \quad w = \frac{1}{12}(1 - x^2)\, e^{\frac{3}{2}x^2} - \frac{1}{2}\, e^{\frac{x^2}{2}}.$$

Подстановка этих выражений в (42) дает разложение искомого решения, пригодное для небольших $|x|$ и $|\alpha|$.

В более сложных случаях при применении метода малого параметра часто бывает полезно найти хотя бы первый содержащий параметр член разложения.

Общие замечания о зависимости решения от параметра. В связи с предыдущим пунктом выскажем несколько общих соображений. Часто бывает, что изучаемое дифференциальное уравнение или система таких уравнений содержат один или несколько параметров, которые могут принимать различные постоянные значения. Рассмотрим для простоты уравнение первого порядка

$$\frac{dy}{dx} = f(x, y; \lambda) \qquad (44)$$

(λ — параметр) при определенных начальных условиях $x = x_0$, $y = y_0$. Будем считать, что точка $(x_0; y_0)$ — неособая, т. е. при заданных условиях существует единственное решение уравнения (44). Тогда из геометрического смысла уравнения (44) следует, что если его правая часть зависит от λ непрерывно, то при малом изменении λ поле направлений будет меняться мало, а потому и решение $y(x; \lambda)$ будет зависеть от λ непрерывно. Аналогичный вывод получается, если от λ зависит не только уравнение, но и начальное условие, т. е. если $x_0 = x_0(\lambda)$, $y_0 = y_0(\lambda)$.

Пусть решение $y(x; \lambda)$ уравнения (44) известно при некотором, как говорят, «невозмущенном» значении λ; пусть, далее, значение параметра изменилось и стало равным $\lambda+\Delta\lambda$, где $|\Delta\lambda|$ мало. Тогда и y изменится и получит приращение $\Delta_\lambda y$, главную линейную часть, т. е. дифференциал которого мы обозначим через δy и назовем *вариацией* решения.

Таким образом, вариация — это частный дифференциал, взятый по параметру; новое название и новое обозначение применяются, чтобы отличить дифференциал по независимой переменной от дифференциала по параметру. В тех случаях, когда малыми высшего порядка можно пренебречь, можно сказать просто, что вариация решения — это бесконечно малое его изменение, полученное за счет изменения параметра. При выбранном значении λ, величина δy, как и y, зависит от x, т. е. $\delta y = \delta y(x)$ и прямо пропорциональна $\Delta\lambda$,.

Чтобы составить дифференциальное уравнение для δy, надо приравнять дифференциалы обеих частей равенства (44) по λ:

$$\delta \frac{dy}{dx} = \delta\,(f\,(x,\ y;\ \lambda)) = \frac{\partial f}{\partial y}\,\delta y + \frac{\partial f}{\partial \lambda}\,\delta\lambda \qquad (\delta\lambda = \Delta\lambda),$$

т. е.

$$\frac{d\,(\delta y)}{dx} = f'_y\,(x,\ y;\ \lambda)\,\delta y + f'_\lambda\,(x,\ y;\ \lambda)\,\delta\lambda. \qquad (45)$$

При этом в левой части мы переставили знаки d и δ, как знаки дифференциалов по разным переменным, а в правой части воспользовались формулой для производной сложной функции. Уравнение (45) называется *уравнением в вариациях* для исходного уравнения (44). Так как в правой части взамен y надо подставить «невозмущенное решение» $y\ (x;\ \lambda)$, то уравнение (45) линейное и потому легко интегрируется. Для уравнений высших порядков и для систем уравнений соответствующие уравнения в вариациях в общем случае не интегрируются в квадратурах, но они *всегда являются линейными.*

Выведем *начальное условие для* δy. В общем случае, когда $x_0 = x_0\,(\lambda)$, $y_0 = y_0(\lambda)$, получаем при значении параметра $\lambda+\delta\lambda$, отбрасывая малые высшего порядка, что при $x = x_0\,(\lambda+\delta\lambda) = x_0 + x'_0\delta\lambda$ будет $y = y_0\ (\lambda+\delta\lambda) = y_0 + y_0\delta\lambda$, где $x'_0 = x'_0\,(\lambda)$, $y'_0 = y'_0\,(\lambda)$. Отсюда при значении параметра $\lambda+\delta\lambda$, и при $x = x_0$ будет

$$y\,|_{x=x_0} = y\,|_{x=x_0+x'_0\,\delta\lambda} - \partial_x y\,|_{dx=x'_0\delta\lambda} = y_0 + y'_0\delta\lambda - \frac{dy}{dx}\,x'_0\delta\lambda = y_0 + y'_0\delta\lambda - f_0 x'_0\delta\lambda,$$

где $f_0 = f\,(x_0,\ y_0;\ \lambda)$. Но то же значение y равно

$$(y\,|_\lambda + \delta y)_{x=x_0} = y_0 + (\delta y)_{x=x_0}.$$

Значит, начальное условие для δy таково:

$$(\delta y)_{x=x_0} = (\,y'_0 - f_0 x'_0\,)\,\delta\lambda.$$

В том частном случае, когда x_0 и y_0 не зависят от λ, будет $x'_0 = y'_0 = 0$ и потому начальное условие имеет вид $(\delta y)_{x=x_0} = 0$.

Иногда параметр входит в дифференциальное уравнение таким способом, что при некоторых значениях этого параметра уравнение понижает свой порядок, т. е. вырождается. При этом возникают новые обстоятельства, которые мы поясним на примере.

Рассмотрим задачу

$$\lambda y' + y = 0, \quad y\,|_{x=0} = 1 \qquad (46)$$

с решением $y = e^{-\frac{x}{\lambda}}$. При $\lambda = 0$ получается вырождение. Пусть решение рассматривается при $x \geq 0$ и $\lambda \to +0$; это решение показано на рис. 9.

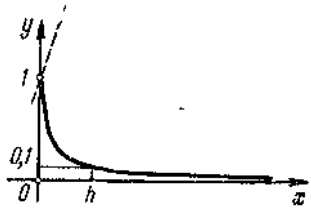


Рис. 9

Уравнение (46) в пределе переходит в равенство $y=0$, но мы видим, что при малом λ, решение близко к нулю не сразу от $x = 0$, а только от некоторого $x = h$. Промежуток $0 < x < h$, называемый *пограничным слоем,* нужен решению для того, чтобы от единичного начального значения (45) перейти к значению, близкому к нулю.

Ширина пограничного слоя условна, так как теоретически решение нигде не становится точно равным нулю. Если, например, принять за ширину пограничного слоя значение $x = h$, при котором решение уменьшается в 10 раз по сравнению с исходным значением, то для задачи (46) мы получим

$$e^{-h/\lambda} = 0{,}1; \qquad h = \ln 10 \cdot \lambda,$$

т. е. *ширина пограничного слоя прямо пропорциональна значению* λ.

Если $\lambda \to -0$, то получающееся решение, изображенное на рис. 9 пунктиром, стремится к бесконечности при любом $x > 0$.

В более сложных случаях часто наблюдается аналогичное явление. Пусть, например, рассматривается решение уравнения второго порядка, удовлетворяющее двум начальным или краевым условиям, причем при некотором значении параметра $\lambda = \lambda_0$ порядок уравнения понижается до первого. Тогда бывает, что если решение $y_0(x)$ при $\lambda = \lambda_0$ остается конечным, то оно, как решение уравнения первого порядка, удовлетворяет только одному условию, а другому не обязательно. При λ, близком к λ_0, для решения $y(x)$ имеется «пограничный слой» (ширина которого пропорциональна $|\lambda - \lambda_0|$), на протяжении которого $y(x)$ переходит от этого другого условия к $y_0(x)$. Подобная ситуация может возникнуть и для систем дифференциальных уравнений.

Методы улучшения невязки. Эти методы основаны на том, что неизвестная функция ищется в виде, включающем несколько параметров, т. е. в виде

$$y = \varphi(x, \lambda_1, \lambda_2, \ldots, \lambda_m). \qquad (47)$$

При этом правая часть обычно выбирается так, чтобы для любых значений этих параметров удовлетворялись поставленные начальные

или граничные условия. После подстановки выражения (47) в заданное дифференциальное уравнение невязка h, т е. разность между левой и правой частями, будет содержать эти параметры:

$$h = h(x;\ \lambda_1,\ \lambda_2,\ \ldots,\ \lambda_m)$$

Если бы решение (47) было точным, то невязка h тождественно равнялась бы нулю. Поэтому для нахождения параметров $\lambda_1, \lambda_2, \ldots, \lambda_m$ на невязку накладывают m условий, которые заведомо выполняются для тождественно нулевой функции. Например, можно приравнять $h = 0$ при m значениях x — это *метод коллокации.* Можно минимизировать интеграл

$$\int_a^b h^2\,dx$$

на том интервале $a \le x \le b$, на котором строится решение, —это *метод наименьших квадратов* Можно приравнять нулю интегралы

$$\int_a^b h\psi_1(x)\,dx,\quad \int_a^b h\psi_2(x)\,dx,\quad \ldots,\quad \int_a^b h\psi_m(x)\,dx$$

где $\psi_1(x),\ \psi_2(x),\ \ldots,\ \psi_m(x)$ — какая-либо выбранная система функций,— это *метод моментов,* так как подобные интегралы называются *моментами.*

Чем больше введено параметров λ_i, тем болеа «гибкой» является формула (47), т. е. тем точнее можно представить этой формулой искомое решение, но тем сложнее получаются вычисления. Большое искусство заключается в том, чтобы правильно предугадать вид искомого решения с помощью формулы, содержащей небольшое число параметров. О правильности результата можно судить, сравнивая результаты повторных вычислений по разным методам или с разным числом параметров и т. п

Если правая часть формулы (47) удовлетворяет не всем поставленным начальным или граничным условиям, то требование, чтобы эти условия удовлетворялись, соответственно уменьшает число условий, накладываемых на невязку.

Рассмотрим простой пример, в котором возможно сравнение с точным решением Пусть надо решить краевую задачу

$$y' + y = 0 \quad (0 \le x \le 1),\quad y(0) = 0,\quad y(1) = 1.$$

Будем искать решение в виде

$$y = \lambda x + \mu x^2. \qquad (48)$$

При этом первое граничное условие удовлетворяется автоматически, а второе дает $\lambda + \mu = 1$, откуда $y = \lambda x + (1 - \lambda)x^2$, и у нас остается всего

одна степень свободы, т. е. возможность поставить лишь одно условие для улучшения невязки, которая равна

$$h = y'' + y = 2(1-\lambda) + \lambda x + (1-\lambda)x^2.$$

Коллокация при $x = \frac{1}{2}$ дает значение $\lambda = \frac{9}{7}$; метод наименьших квадратов для интервала $0 \leqslant x \leqslant 1$ дает значение $\lambda = \frac{257}{202}$; метод моментов с функцией $\psi(x) \equiv 1$ дает значение $\lambda = \frac{14}{11}$ (проверьте!). Подстановка этих значений в формулу (48) дает приближенные решения, которые неплохо аппроксимируют точное

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

например, при $x = x0{,}5$ оно равно 0,5699, тогда как приближенные решения равны соответственно 0,5714, 0,5681 и 0,5682, ошибка $\pm 0{,}3\%$.

Метод упрощения. Этот метод широко применяется на практике, особенно при грубых прикидочных расчетах. Он состоит в том, что само исходное уравнение упрощается путем отбрасывания сравнительно малых членов, замены медленно меняющихся коэффициентов постоянными и т. п. После такого упрощения может получиться уравнение одного из интегрируемых типов и, интегрируя, мы получим функцию, которая может считаться приближенным решением исходного, полного уравнения; во всяком случае, она часто правильно передает характер поведения точного решения. Найдя это «нулевое приближение», иногда удается с его помощью внести поправки, учитывающие упрощение, и тем самым найти «первое приближение» и т д.

Если уравнение содержит параметры (например, массы, линейные размеры исследуемых объектов и т. п.), то нужно иметь в виду, что при одних значениях этих параметров относительно малыми могут быть одни члены уравнения, а при других значениях — другие, так что упрощение будет при разных значениях параметров производиться по-разному. Кроме того, иногда приходится разбивать интервал изменения независимой переменной на части, в каждой из которых упрощение проводится по-своему.

Особенно полезно такое упрощение уравнения в случаях, когда при самом выводе (написании) дифференциального уравнения делались существенные упрощающие предположения или когда точность, с которой известны рассматриваемые величины, невелика. Так, члены уравнения, меньшие допускаемой погрешности в других его членах, надо безусловно отбросить.

Рассмотрим, например, задачу

$$y'' + \frac{1}{1+0,1x}\, y + 0,2y^3 = 0, \quad y(0) = 1, \quad y'(0) = 0; \qquad 0 \leqslant x \leqslant 2. \quad (49)$$

Так как коэффициент при y меняется медленно, заменим этот коэффициент его средним значением:

$$\frac{1}{2-0}\int_0^2 \frac{1}{1+0,1x}\, dx = \frac{1}{2}\,\frac{\ln(1+0,1x)}{0,1}\Big|_0^2 = \frac{\ln 1,2}{0,2} = 0,911.$$

Кроме того, сравнительно малое третье слагаемое отбросим.

Получим уравнение $y'' + 0,911y = 0$ с решением при данных начальных условиях:

$$y = \cos 0,954x. \quad (50)$$

Вид этого приближенного решения подтверждает правомерность отбрасывания последнего слагаемого в уравнении, поскольку отношение третьего члена ко второму порядка $0,2y^2 < 0,2$, и потому первый член должен «почти взаимно уничтожиться» со вторым. Внесем поправку на последнее слагаемое, для чего подставим в него приближенное решение (50), оставив коэффициент осредненным:

$$y'' + 0,911y = -0,2\cos^3 0,954x = -0,05\cos 2,86x - 0,15\cos 0,954x$$

По изложенным ранее методам получаем при заданном начальном условии

$$y = 0,993\cos 0,954x - 0,079x\sin 0,954x + 0,007\cos 2,86x.$$

Разница по сравнению с нулевым приближением (50) невелика, так что вывод о значении отдельных слагаемых в уравнении (49) остается в силе; в то же время третий член уравнения (49) внес свой вклад в решение.

Подобные рассуждения зачастую не блещут строгостью и иногда приводят к ошибкам; однако если они проводятся в соответствии со здравым смыслом, то все же, и притом довольно часто, дают решение, которым можно пользоваться на практике.

Метод Эйлера. Мы переходим к изложению некоторых методов численного интегрирования дифференциальных уравнений. Эти методы применяются, если ни один из описанных выше методов «приближенного интегрирования», т. е. получения приближенных формул для решения, не является достаточно эффективным, в частности, если решение требуется с большой точностью на большом интервале изменения аргумента.

Часто целесообразно комбинировать методы приближенного и численного интегрирования. Например, для уравнения

$$y'' + (1 + e^{-x})\, y = 0$$

при каком-либо заданном начальном условии можно для малых x применить формулу Тейлора, при средних x в зависимости от требуемой точности — один из методов численного интегрирования, а при больших x — просто отбросить член e^{-x}.

Мы изложим четыре наиболее известных метода численного интегрирования уравнений первого порядка; эти методы очень просто переносятся на системы уравнений первого порядка, к которым приводятся и уравнения высших порядков.

Метод Эйлера прост и нагляден, хотя и недостаточно практически эффективен. Однако его надо хорошо понять, так как многие важные и эффективные методы в различных разделах математики являются, по существу, его развитием.

Метод Эйлера состоит в непосредственной замене производной в дифференциальном уравнении разностным отношением. Пусть рассматривается начальная задача

$$y' = f(x, y), \quad y(x_0) = y_0. \tag{51}$$

Будем считать для простоты шаг h по x постоянным и обозначим

$$x_0 + h = x_1, \quad x_0 + 2h = x_2, \quad x_0 + 3h = x_3, \quad \ldots,$$

а приближенные значения $y(x_k)$ обозначим y_k. Чтобы найти эти значения, заменим в уравнении производную разностным отношением

$$\frac{\Delta y_k}{\Delta x} = f(x_k, y_k), \text{ т. е. } \frac{y_{k+1} - y_k}{h} = f(x_k, y_k)$$

и

$$y_{k+1} = y_k + f(x_k, y_k)\, h. \tag{52}$$

По последней формуле можно, начиная от y_0 и полагая последовательно $k = 0, 1, 2, \ldots$, найти значения

$$y_1 = y_0 + f(x_0, y_0)\, h, \quad y_2 = y_1 + f(x_1, y_1)\, h, \quad \ldots$$

Метод Эйлера имеет простой геометрический смысл, показанный на рис. 10, где изображены также интегральные линии.

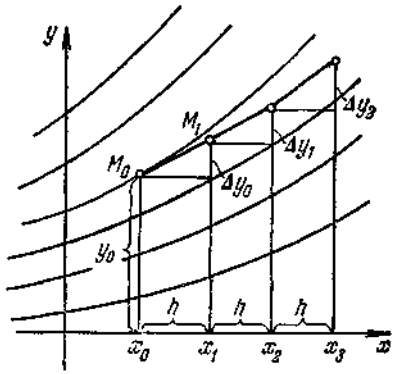


Рис. 10

Он состоит в том, что через заданную точку M_0 мы проводим не искомую интегральную линию, которая нам не известна, а отрезок M_0M_1 касательной к этой линии, руководствуясь направлением поля в точке M_0. Через M_1 мы проводим отрезок, руководствуясь направлением поля в M_1; и т. д. Полученная *ломаная Эйлера* приближенно изображает требуемую интегральную линию, которая получилась бы, если бы шаг h был бесконечно малым, т. е. если бы мы непрерывно подправляли направление ломаной.

Легко оценить порядок ошибки в методе Эйлера. Пользуясь формулой (52), мы заменяем приращение решения его дифференциалом $y'_k\Delta x = f(x_k, y_k)\,h$. При этом делается ошибка порядка h^2. Если мы строим решение на некотором интервале x_0, x и разбиваем его на n частей, то $h = \frac{x-x_0}{n}$, и суммарная ошибка будет иметь порядок $nh^2 = \frac{(x-x_0)^2}{n}$. Значит, для повышения точности в 10 раз, т. е. для вычисления одного дополнительного десятичного знака, требуется увеличить число точек деления также в 10 раз, что значительно увеличит объем вычислительной работы. В этом недостаток метода.

Отметим еще одну особенность метода Эйлера, свойственную и другим методам численного интегрирования дифференциальных уравнений. Мы уже отмечали, что решение такого уравнения может, при своем продолжении, обратиться в бесконечность при конечном значении x. В то же время ясно, что решение, построенное по методу Эйлера, остается конечным при всех значениях x. Чтобы правильно передать поведение решения в таких случаях, можно поступить следующим образом: если в результате численного интегрирования

будет обнаружено значительное возрастание решения $y(x)$ по абсолютной величине, совершить в дифференциальном уравнении замену вида

$$y = \frac{1}{z}.$$

Если тогда при дальнейшем интегрировании окажется, что z переходит через нулевое значение при некотором $x = \alpha$, это и будет означать, что $|y(\alpha)| = \infty$.

Метод Рунге — Кутта. Покажем сначала этот метод, уточняющий метод Эйлера, в более простом варианте. Пусть y_k, т. е. приближенное значение решения при $x = x_k$, уже построено; тогда y_{k+1} можно найти с помощью следующего вычисления:

$$f_k = f(x_k, y_k), \quad \alpha_k = f\left(x_k + \frac{h}{2}, \quad y_k + \frac{f_k h}{2}\right), \quad y_{k+1} = y_k + \alpha_k h. \quad (53)$$

Геометрический смысл этого вычисления показан на рис. 11.

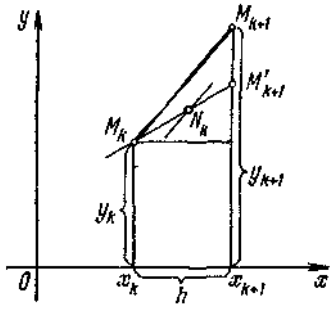


Рис. 11

Именно, каждое очередное звено M_kM_{k+1} ломаной, аппроксимирующей интегральную линию, строится так. Сначала проводим отрезок $M_kM'_{k+1}$, руководствуясь направлением f_k поля в точке M_k. Не ограничиваясь этим (как в методе Эйлера), определяем направление α_k поля в середине N_k этого отрезка и проводим отрезок M_kM_{k+1} именно под этим новым направлением. Таким образом, мы производим *пересчет,* уточнение углового коэффициента звеньев ломаной, аппроксимирующей интегральную кривую. Уже из геометрического смысла ясно, что данный метод точнее метода Эйлера, так как здесь учитывается поворот поля на интервале

$$x_k \leqslant x \leqslant x_{k+1}.$$

Это можно подтвердить и вычислением. По формуле Тейлора

$$\alpha_k = f(x_k, y_k) + f'_x(x_k, y_k)\frac{h}{2} + f'_y(x_k, y_k)\frac{f_k h}{2} + O(h^2)$$

(под $O(h^2)$ понимается величина, ограниченная по сравнению с h^2.

Подобно описанного ранее метода Тейлора, находим

$$y_{k+1} = y_k + \alpha_k h = y_k + f(x_k, y_k)h + f'_x(x_k, y_k)\frac{h^2}{2} + \\ + f'_y(x_k, y_k) f_k \frac{h^2}{2} + O(h^3) = y_k + y'_k h + \frac{y''_k}{2} h^2 + O(h^3). \quad (54)$$

Но точное значение решения при условии $y(x_k) = y_k$ равняется

$$y(x_k + h) = y_k + y'_k h + \frac{y_k}{2} h^2 + O(h^3) \quad (55)$$

Сравнивая формулы (54) и (55), видим, что значения $y(x_k+h)$ и y_{k+1} могут различаться только в величинах порядка не менее h^3. Отсюда легко заключить, что суммарная ошибка метода имеет порядок $\frac{1}{n^2}$ или, что то же, порядок h^2. Значит, если число точек деления увеличить в 10 раз, то точность повысится в 100 раз.

Еще более точный результат получится, если вычислять по схеме:

$$f_k = f(x_k, y_k),\ \alpha_k = f\left(x_k + \frac{h}{2},\ y_k + \frac{f_k h}{2}\right),\ \beta_k = f\left(x_k + \frac{h}{2},\ y_k + \frac{\alpha_k h}{2}\right),$$

$$\gamma_k = f(x_k + h,\ y_k + \beta_k h),\ y_{k+1} = y_k + \frac{1}{6}(f_k + 2\alpha_k + 2\beta_k + \gamma_k)h.$$

Вычисления, подобные (54), показывают, что ошибка на каждом шаге здесь имеет порядок h^5, а потому суммарная ошибка — порядок h^4. Значит, если число точек деления увеличится в 10 раз, то точность повысится в 10 000 раз.

Метод Адамса. Этот метод, предложенный в 1883 г. английским астрономом Дж. Адамсом (1819—1892), основан на второй интерполяционной формуле Ньютона, которую мы применим для производной $y'(x)$ от решения, начиная от некоторого значения $x_k = x_0 + kh$:

$$y'(x) = y'_k + \Delta y'_{k-1}\frac{x - x_k}{h} + \frac{\Delta^2 y'_{k-2}}{2!}\frac{x - x_k}{h}\left(\frac{x - x_k}{h} + 1\right) + \\ + \frac{\Delta^3 y'_{k-3}}{3!}\frac{x - x_k}{h}\left(\frac{x - x_k}{h} + 1\right)\left(\frac{x - x_k}{h} + 2\right). \quad (56)$$

При этом в указанной формуле мы вместо $t = x_{k+1}$ — x подставили $x_k - x = -(x - x_k)$ и соответственно вместо y_{k+1} — значение y'_k; кроме того, мы заменили знак приближенного равенства знаком точного равенства, хотя, конечно, формула (56) приближенная и ее ошибка имеет порядок $\Delta^4 y'$, т. е. h^4. Интегрирование формулы (56) от x_k до $x_{k+1} = x_k + h$ дает после подстановки $\frac{x - x_k}{h} = s$

$$y_{k+1} = y_k + \left(y'_k + \frac{1}{2}\Delta y'_{k-1} + \frac{5}{12}\Delta^2 y'_{k-2} + \frac{3}{8}\Delta^3 y'_{k-3}\right)h. \quad (57)$$

Погрешность формулы (57) получается в результате интегрирования погрешности формулы (56), т. е. имеет порядок h^5 (почему?).

Применяется формула (57) следующим образом. Сначала каким-либо способом, например с помощью формулы Тейлора или с помощью метода Руте — Кутта находим значения

$$y_1 = y(x_0 + h), \quad y_2 = y(x_0 + 2h) \quad \text{и} \quad y_3 = y(x_0 + 3h).$$

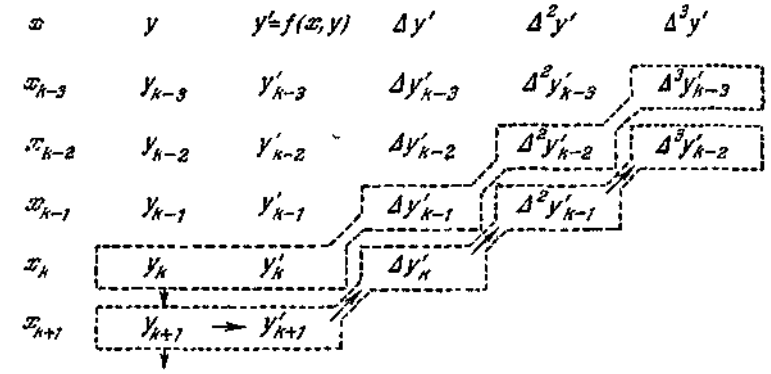
Затем вычисляем соответствующие значения

$$y'_0 = f(x_0, y_0), \quad y'_1 = f(x_1, y_1), \quad y'_2 = f(x_2, y_2), \quad y'_3 = f(x_3, y_3),$$

с помощью которых находим

$$\Delta y'_0 = y'_1 - y'_0, \ \Delta y'_1, \ \Delta y'_2, \ \Delta^2 y'_0 = \Delta y'_1 - \Delta y'_0, \ \Delta^2 y'_1, \ \Delta^3 y'_0 = \Delta^2 y'_1 - \Delta^2 y'_0.$$

Далее, полагая в формуле (57) $k = 3$, вычисляем y_4, а с его помощью $y'_4 = f(x_4, y_4)$, $\Delta y'_3 = y'_4 - y'_3$, $\Delta^2 y'_2$, $\Delta^3 y'_1$. Затем, полагая в формуле (57) $k = 4$, вычисляем y_5, далее с его помощью $y'_5 = f(x_5, y_5)$ и т. д. Вычисления проходят по схеме



Метод Милна. С помощью первой интерполяционной формулы Ньютона можно получить еще один метод, который является одним из наиболее эффективных. Мы приведем лишь окончательный результат. Вычисления в методе Милна (1926) проходят по формулам

$$\left.\begin{aligned} \bar{y}_{k+1} &= y_{k-3} + \frac{4h}{3}(2y'_{k-2} - y'_{k-1} + 2y'_k) \\ &\qquad (\text{где } y'_i = f(x_i, y_i)), \\ \bar{y}'_{k+1} &= f(x_{k+1}, \bar{y}_{k+1}), \\ y_{k+1} &= y_{k-1} + \frac{h}{3}(y'_{k-1} + 4y'_k + \bar{y}'_{k+1}) \end{aligned}\right\} \quad (k = 3, 4, 5, \ldots). \quad (58)$$

При этом, как и в методе Адамса, значения y_0, y_1, y_2, y_3 должны быть найдены каким-либо иным способом. После этого, полагая в формулах

(58) $k = 3$, находим последовательно $\bar{y}_4$, $\bar{y}'_4$, y_4. Затем, полагая $k = 4$, находим $\bar{y}_5$, $\bar{y}'_5$, y_5 и т. д. Найденные значения $y_4, y_5, y_6, \ldots$ и являются приближенными значениями решения $y(x)$ при $x = x_4, x_5, x_6, \ldots$, где $x_i = x_0 + ih$.

Оказывается, что абсолютная погрешность, получающаяся при вычислении y_{k+1} по данному методу, приблизительно равна

$$\frac{|y_{k+1} - \bar{y}_{k+1}|}{29}.$$

Поэтому при вычислениях можно попутно проверять, не выходит ли эта погрешность за рамки принятой степени точности вычислений. Если это где-либо произойдет, то, начиная с соответствующего значения x, надо уменьшить шаг, имея в виду, что суммарная ошибка данного метода имеет порядок h^4.

9.7. Уравнения, не разрешимые относительно производной

Чтобы решить дифференциальное уравнение

$$F(x,y,y')=0, \tag{1}$$

можно попытаться сначала решить его относительно y'. Если это удается, то мы получим одно или много дифференциальных уравнений вида

$$\frac{dy}{dx} = f(x, y). \tag{2}$$

Любое решение каждого из уравнений (2) будет решением уравнения (1). Все же следует попытаться выяснить, исчерпывают ли они все решения (1). Например, чтобы решить уравнение

$$(y')^2 - (2x + y)\, y' + 2xy = 0, \tag{3}$$

тождественными преобразованиями его левой части приведем его к виду

$$(y' - 2x)\,(y' - y) = 0. \tag{3'}$$

Рассмотрим два дифференциальных уравнения первого порядка

$$y' = 2x, \quad y' = y.$$

Общие их интегралы имеют соответственно вид

$$y = x^2 + C_1, \quad y = C_2 e^x, \tag{4}$$

где C_1 и C_2 — произвольные постоянные. Для частных значений C_1 и C_2 функции (4) суть частные решения уравнения (3).

Но из указанных частных решений последних двух уравнений можно строить и другие частные решения уравнения (3). Например, функция

$$y=\begin{cases} x^2+1, & x\leqslant 1, \\ 2e^{x-1}, & x>1, \end{cases}$$

является решением уравнения (3). Эта интегральная кривая составлена из двух интегральных кривых, принадлежащих разным семействам (4) (рис. 12).

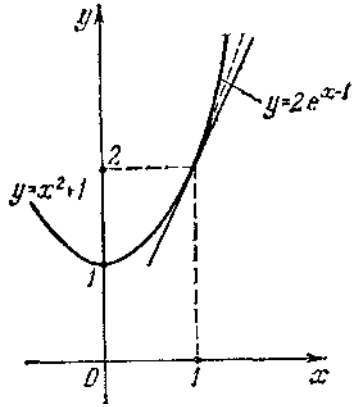


Рис. 12

Ниже рассматриваются два частных вида дифференциального уравнения (1), для которых можно указать иные пути их решения.

1°. Левая часть уравнения (1) не содержит x и y:

$$F(y')=0. \qquad (5)$$

Будем считать, что функция F непрерывна и имеет конечное число нулей.

Пусть $y=y(x)$ есть решение уравнения, имеющее непрерывную производную. Тогда $y'(x)$ равняется одному из корней уравнения (5), которое обозначим через k. Итак, $y'=k$, откуда $y=kx+C$, где C— постоянная и

$$F\left(\frac{y-C}{x}\right)=0. \qquad (6)$$

Обратно, из того, что для непрерывно дифференцируемой функции $y(x)$ при некоторой постоянной C выполняется равенство (6), следует, что

$$\frac{y-C}{x}=k \qquad (\forall x\neq 0),$$

где k—некоторый корень функции F. Но тогда $y = kx+C$, $\forall\, x$, $y' = k$ и $F(y') = 0$.

Мы доказали, что *общее (любое) решение дифференциального уравнения* (5) *определяется равенством* (6), *где С— произвольная постоянная.*

2°. Левая часть уравнения (1) не содержит x:

$$F(y,y') = 0. \qquad (7)$$

Если уравнение (7) можно разрешить относительно y', то $y' = \varphi(y)$ — уравнение с разделяющимися переменными, которое решать мы умеем.

Допустим, что уравнение (7) нельзя или трудно решить относительно y', но легко можно решить относительно y:

Введем в рассмотрение параметр $p = \frac{dy}{dx}$, тогда

$$y = \varphi(p), \qquad dy = \varphi'(p)\,dp, \qquad dx = \frac{dy}{p} = \frac{\varphi'(p)\,dp}{p},$$

откуда

$$x = \int \frac{\varphi'(p)\,dp}{p} + C,$$

или

$$x = \psi(p) + C.$$

Теперь, исключая из системы

$$x = \psi(p) + C, \quad y = \varphi(p) \qquad (8)$$

параметр p, мы и получим общий интеграл $\Phi(x, y, C)=0$ дифференциального уравнения (7).

Систему (8) можно также рассматривать как параметрическое задание решения уравнения (7). Параметр p можно вводить и произвольным образом $y'=\omega(p)$, но так, чтобы уравнение (7) проще решалось относительно y, $y=\varphi(p)$, и чтобы проще находился соответствующий интеграл для определения функции

$$x = \int \frac{\varphi'(p)\,dp}{\omega(p)} + C.$$

Пример. Решить уравнение

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

Если ввести параметр $p=y'$, то получаются довольно сложные интегралы. Здесь лучше положить $y' = \operatorname{tg} p$ $(-\pi/2 < p < \pi/2)$. Тогда

$$x = \frac{2\operatorname{tg} p}{\sqrt{1+\operatorname{tg}^2 p}} = 2\sin p, \qquad dx = 2\cos p\,dp,$$

$$dy = \operatorname{tg} p\,dx = 2\sin p\,dp, \qquad y = -2\cos p + C.$$

Из системы

$$x = 2\sin p, \quad y = -2\cos p + C$$

получаем

$$x^2 + (y - C)^2 = 4, \tag{9}$$

т. е. любое решение нашего дифференциального уравнения есть решение уравнения (9) при некоторой постоянной C. Это семейство окружностей радиуса 2 с центром в точках $(0, C)$ (рис. 13).

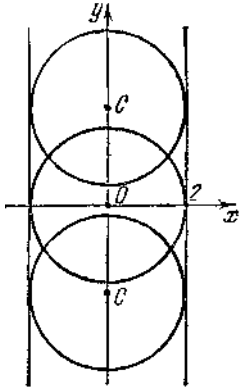


Рис. 13.

Можно доказать, что равенство (4) есть общий интеграл для решений вида $y(x)$ и вида $x(y)$.

В данном случае можно также параметр ввести по формуле $y' = \operatorname{sh} p$.

Замечание. Аналогичным образом рассматривается дифференциальное уравнение вида $F(x, y') = 0$.

Рассмотрим особые решения

Пусть дано дифференциальное уравнение

$$\frac{dy}{dx} = f(x, y). \tag{10}$$

Если в окрестности точки (x_0, y_0) плоскости xOy выполняются условия теоремы существования и единственности, то через нее проходит и притом только одна интегральная кривая.

Если условия теоремы существования не соблюдаются, но тут могут быть различные случаи. Через точку (x_0, y_0) все же может проходить одна интегральная кривая или несколько или бесконечное множество, или же нет интегральной кривой, которая проходила бы через точку (x_0, y_0). Интересен тот случай, когда дифференциальное уравнение (1) имеет особое решение.

Решение дифференциального уравнения первого порядка называется *особым,* если соответствующая интегральная кривая обладает тем свойством, что через любую ее точку проходит, кроме нее, еще и другая касающаяся ее интегральная кривая данного уравнения.

Нередко приходится иметь дело с дифференциальными уравнениями вида (10), где функция $f(x, y)$ непрерывна на некоторой области Ω, а ее частная производная $\frac{\partial f}{\partial y}$ конечна и непрерывна не всюду на Ω. Имеются на Ω такие точки, где $\frac{\partial f}{\partial y}=\infty$. В каждой такой точке, вообще говоря, нарушаются условия существования и единственности решения дифференциального уравнения (10), а если такие точки образуют гладкие линии, то последние могут представлять особые решения дифференциального уравнения.

Пример. Рассмотрим простейшее уравнение Бернулли $y' = y^{\alpha}$, $\alpha>0$, $y\geqslant 0$. Здесь $f(x, y) = y^{\alpha}$ — непрерывная функция на верхней полуплоскости. Функция $\frac{\partial f}{\partial y}=\alpha\ y^{\alpha-1}$ при $0 < \alpha < 1$ не ограничена в окрестности y=0. Функция y=0 является решением уравнения. Для начального условия $y(x_0) = 0$ есть еще одно решение

$$y = [(x - x_0)(1 - \alpha)]^{1/(1-\alpha)},$$

удовлетворяющее этому уравнению и проходящее через точку $(x_0, 0)$. Касательная к этой кривой в точке $(x_0, 0)$, очевидно, есть ось $x(y = 0)$. Поэтому $y = 0$ есть особое решение.

Пример. $y' = y^{\alpha} + 1$, $y>0$.

Здесь

$$\frac{\partial f}{\partial y} = \alpha y^{\alpha-1}$$

и при $0 < \alpha < 1$

$$\frac{\partial f}{\partial y} = \alpha y^{\alpha-1}$$

эта функцияне ограничена в окрестности $y = 0$. Однако $y = 0$ не является решением уравнения. Решение уравнения, например при α=1/2, определяетсянеявно равенством

$$x + C = 2(\sqrt{y} - \ln(\sqrt{y} + 1)) \quad (y \geqslant 0),$$

т. е. через каждую точку $(x_0, 0)$ проходит единственная

интегральная кривая $x - x_0 = 2(\sqrt{y} - \ln(\sqrt{y}+1))$.

Пример. Функции $y = C(x-C)^2$ при любом C (рис. 14) являются решениями уравнения $F(x, y, y') = 4xyy' - y'^3 - 8y^2 = 0$.

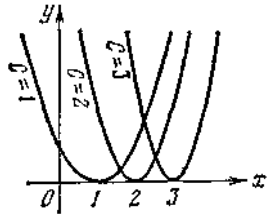


Рис. 14.

Функция $y = 0$ является особым решением данного уравнения.

9.8. Огибающая семейства кривых

Пусть дано семейство гладких кривых Γ_α, определяемых уравнением

$$\Phi(x, y, \alpha) = 0, \qquad (1)$$

где α — произвольная постоянная (параметр) и функция $\Phi(x, y, \alpha)$, непрерывно дифференцируемая на некоторой области точек (x, y, α).

Кривая E называется *огибающей* семейства кривых (1), если она касается каждой кривой Γ_α семейства и при этом вся состоит из этих точек касания.

Точнее, *огибающей* E семейства кривых Γ_α, зависящих от параметра α, где $a < \alpha < b$, называется глгдкая кривая

$$\left.\begin{aligned} x &= x(\alpha), \\ y &= y(\alpha), \end{aligned}\right\} \quad (a < \alpha < b), \qquad (2)$$

касающаяся при любом значении параметра α, соответствующей кривой Γ_α. Кривые Γ_α будем называть *огибаемыми* (рис. 15).

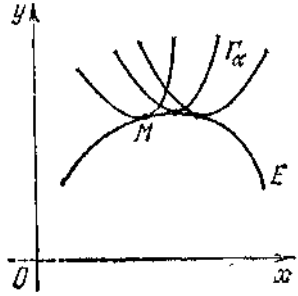


Рис. 15

Найдем уравнение огибающей. Зададим значение α. Ему соответствует на Γ_α точка $x = x(\alpha)$, $y = y(\alpha)$, принадлежащая одновременно Γ_α и E, в которой E и Γ_α имеют общую касательную.

Очевидно, имеет место тождество

$$\Phi(x(\alpha),\ y(\alpha),\ \alpha) \equiv 0 \qquad (a < \alpha < b).$$

Продифференцируем его по α:

$$\Phi'_x x'(\alpha) + \Phi'_y y'(\alpha) + \Phi'_\alpha \equiv 0 \qquad (a < \alpha < b).$$

Вектор $(x'(\alpha), y'(\alpha))$ направлен по касательной к E, которая совпадает с касательной к Γ_α в точке $(x(\alpha), y(\alpha))$. Но тогда

$$\Phi'_x x'(\alpha) + \Phi'_y y'(\alpha) = 0,$$

Следовательно,

$$\Phi'_\alpha(x(\alpha),\ y(\alpha),\ \alpha) = 0,$$

т. е.

$$\Phi'_\alpha(x,\ y,\ \alpha) = 0 \tag{3}$$

в точке (x, y) касания E с Γ_α. Для этой точки выполняется также равенство

$$\Phi(x,\ y,\ \alpha) = 0. \tag{4}$$

Таким образом, уравнение огибающей семейства (1) определяется двумя уравнениями

$$\left.\begin{aligned} \Phi(x,\ y,\ \alpha) &= 0, \\ \frac{\partial \Phi(x,\ y,\ \alpha)}{\partial \alpha} &= 0. \end{aligned}\right\} \tag{5}$$

Равенства (5) дают необходимое условие существования огибающей, т. е. если семейство (1) имеет огибающую, то ее уравнение задается системой (5).

Если же мы составим систему (5) и решим ее, то решение системы не обязательно дает огибающую семейства (1).

Пусть теперь дано дифференциальное уравнение $F(x, y, y') = 0$ и $\Phi(x, y, C) = 0$ — его общий интеграл.

Если семейство интегральных кривых $\Phi(x, y, C) = 0$ имеет *огибающую*, то ясно, что она также является интегральной кривой и, следовательно, *особым решением*.

Если же формально исключить C из системы

$$\left.\begin{aligned} \Phi(x,\ y,\ C) &= 0, \\ \frac{\partial \Phi(x,\ y,\ C)}{\partial C} &= 0, \end{aligned}\right\} \tag{6}$$

то в некоторых случаях получим особое решение.

Если общий интеграл дифференциального уравнения имеет вид

$$\Psi(x,\ y) = C,$$

где $\Psi(x, y)$ — непрерывно дифференцируемая функция, то определяемое им семейство (всех решений дифференциального уравнения) на соответствующей области ни имеет огибающей.

Если же функция $\Psi(x, y))$ не является непрерывно дифференцируемой в некоторых точках (x, y), то совокупность этих точек может дать огибающую ссмсйсша.

Пример. $y' = y^{2/3}$.

Это уравнение Бернулли. Разделяя переменные, получаем $y^{-2/3}\,dy = dx\ (y \neq 0)$, $3y^{1/3} = x - C$, $27y = (x-C)^3$, $\Phi(x, y, C) \equiv 27y - (x-C)^3 = 0$ — общийи нтеграл, где функция $\Phi(x, y, C)$ непрерывно дифферсицируема.

Составим систему (6):

$$\left.\begin{aligned} 27y - (x-C)^3 &= 0, \\ 3(x-C)^2 &= 0. \end{aligned}\right\}$$

Исключая C, получаем $y = 0$. Проверкой убеждаемся, что $y = 0$ — решение исходного уравнения. Это особое решение и огибающая семейства кривых $27y - (x-C)^3 = 0$.

Если рассматривать общий интеграл в разрешенной относительно C форме: $C = x - 3y^{1/3}$, то функция $\Psi(x, y) = x - 3y^{1/3}$ не является непрерывно дифференцируемой в точках оси $y = 0$, которая, как мы убедились, является огибающей семейства кубических парабол.

Уравнение для экспоненты.

Остановимся на очень простом, но крайне важном уравнении

$$\frac{dy}{dx} = ky \qquad (k = \text{const}). \tag{7}$$

Оно означает, что скорость изменения величины y, взятая по отношению к величине x, пропорциональна текущему значению y. Такая пропорциональность с $k > 0$, если y возрастает, и с $k < 0$, если y убывает (мы для простоты считаем, что $y > 0$), часто принимается в первом приближении при исследовании многих процессов, а иногда она оправдывается с большой точностью.

В уравнении (7) разделяются переменные, откуда

$$\frac{dy}{y} = k\,dx, \quad \ln|y| = kx + \ln C, \quad y = Ce^{kx}.$$

Если имеется также начальное условие $y(x_0) = y_0$, то получаем

$$y_0 = Ce^{kx_0}, \quad C = y_0 e^{-kx_0}, \quad \text{т. е.} \quad y = y_0 e^{k(x-x_0)}. \tag{8}$$

Итак, решение уравнения (7) представляет собой экспоненту, т. е. показательную функцию. Для решения характерно, что если придавать x значения, образующие арифметическую прогрессию с разностью Δx, то соответствующие значения y образуют геометрическую прогрессию

со знаменателем $e^{k\Delta x}$. Легко найти, каково должно быть Δx, чтобы y менялся (увеличивался или уменьшался) каждый раз вдвое. Для этого должно быть

$$|k\,\Delta x| = \ln 2, \qquad \text{т. е.} \qquad \Delta x = \frac{\ln 2}{|k|}. \tag{9}$$

Если $k > 0$, то формула (9) показывает *экспоненциальное нарастание* величины y. Так получится, например, при исследовании процесса размножения бактерий в питательной среде, пока их там не слишком много. Примем, что все они размножаются более или менее независимо друг от друга; это — так называемый *закон органического роста,* характерный для всевозможных *цепных реакций.* Тогда получаем, что скорость нарастания количества u этих бактерий, измеренного в каких-то единицах, пропорциональна этому количеству, т. е.

$$\frac{du}{dt} = ku; \qquad u = u_0 e^{k(t-t_0)}.$$

Аналогично исследуются задача о непрерывном нарастании вклада в банке и другие подобные задачи.

Если $k < 0$, то формула (8) показывает *экспоненциальное убывание* величины y. Так получится, например, при исследовании процесса радиоактивного распада. Если принять, что различные участки распадаются независимо друг or друга, то получаем, что скорость убывания еще не распавшейся массы m радиоактивного вещества пропорциональна текущему значению этой массы, т. е.

$$\frac{dm}{dt} = -pm, \qquad m = m_0 e^{-p(t-t_0)}.$$

Отметим, в частности, что в силу формулы (8) за время

$$\Delta t = \frac{\ln 2}{p}$$

значение m уменьшается наполовину; это — *период полураспада.* Так, для радия он приближенно равен $1{,}8 \cdot 10^3$ лет; другими словами, если бы запасы радия не пополнялись, то через $1{,}8 \cdot 10^3$ лет осталась бы половина начального запаса, еще через $1{,}8 \cdot 10^3$ лет — четверть начального запаса и т. д.

Аналогично исследуются убывание атмосферного давления с высотой, процесс разрядки конденсатора через сопротивление и многие другие задачи.

Иногда рассматриваемое уравнение можно более или менее просто преобразовать к виду (7). Например, при включении постоянного напряжения u в цепь, обладающую сопротивлением R и индуктивностью L, ток I удовлетворяет уравнению

$$L\frac{di}{dt}+Ri=u. \tag{10}$$

Это — линейное неоднородное уравнение, которое можно проинтегрировать (решить) по известному методу. Но проще преобразовать уравнение так:

$$L\frac{di}{dt}=-Ri+u=-R\left(i-\frac{u}{R}\right), \quad \frac{d\left(i-\frac{u}{R}\right)}{di}=-\frac{R}{L}\left(i-\frac{u}{R}\right),$$

откуда

$$i-\frac{u}{R}=\left(i_0-\frac{u}{R}\right)e^{-\frac{R}{L}(t-t_0)}, \quad i=\frac{u}{R}+\left(i_0-\frac{u}{R}\right)e^{-\frac{R}{L}(t-t_0)}.$$

Особенно просто получится, если в начальный момент, за который мы примем t =0, тока в цепи не было. Тогда $t_0 = 0$, $i_0 = 0$ и

$$i=\frac{u}{R}\left(1-e^{-\frac{R}{L}t}\right). \tag{11}$$

График полученной зависимости показан на рис. 16.

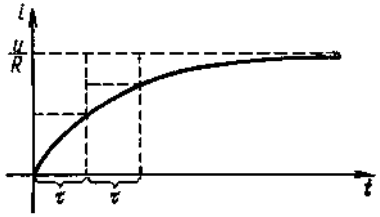


Рис. 16

Мы видим, что ток при $t \to \infty$ экспоненциально приближается к предельному стационарному значению $\frac{u}{R}$. Это же значение легко найти из самого уравнения (10), если заметить, что в процессе установления тока, при $t \to \infty$, будет $\frac{di}{dt} \to 0$, и потому в пределе

$$Ri=u, \quad i=\frac{u}{R},$$

т. е. когда ток практически установился, все напряжение расходуется только на сопротивление *R*. Отклонение тока от предельного значения уменьшается в два раза за время

$$\tau=\frac{\ln 2}{\frac{R}{L}}=\frac{L}{R}\ln 2.$$

То, что в формуле (7) в основании получается именно число *e,* и есть основная причина значения этой константы в математике и ее приложениях.

9.9. Интегрирование полного дифференциала

Дифференциальные уравнения первого порядка часто рассматривают взамен формы

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

в *симметричной форме*

$$P(x, y)\,dx + Q(x, y)\,dy = 0, \qquad (1)$$

где $P(x, y)$ и $Q(x,y)$ — заданные функции, а функциональная зависимость между x и y неизвестна. Легко перейти от одной формы к другой: например, чтобы перейти от (1) к форме $y'=f(x, y)$, надо обе части (1) разделить на Qdx, а затем перенести $\frac{P}{Q}$ в правую часть.

Форма (1) предпочтительнее в тех случаях, когда переменные x и y более или менее равноправны и заранее не требуется, чтобы именно y считался функцией x, а не наоборот.

В частном случае, когда левая часть уравнения (1) предстазляет собой полный дифференциал некоторой функции, т. е.

$$P\,dx + Q\,dy \equiv du(x, y), \qquad (2)$$

это уравнение (1) легко проинтегрировать. Действительно, тогда его можно переписать в виде $du=0$ и, интегрируя, получим общее решение

$$u(x, y) = C, \qquad (3)$$

где C — как всегда, произвольная постоянная.

В данном случае зависимости от z нет и $R\equiv 0$, так что:

$$\frac{\partial P}{\partial y} \equiv \frac{\partial Q}{\partial x}. \qquad (4)$$

Это условие необходимо и достаточно для того, чтобы, левая часть уравнения (1) *была полным дифференциалом.*

Если область многосвязная, то функция u получится, вообще говоря, многозначной. Однако формула (3) и в этом случае дает общее решение уравнения (1).

Рассмотрим, например, уравнение

$$(x^2+2xy)\,dx + (x^2-y^2)\,dy = 0. \qquad (5)$$

Здесь

$$\frac{\partial P}{\partial y} = \frac{\partial (x^2+2xy)}{\partial y} = 2x, \quad \frac{\partial Q}{\partial x} = \frac{\partial (x^2-y^2)}{\partial x} = 2x,$$

т. е. условие (4) выполнено. Для построения функции u по формуле

$$u(M) = \int_{M_0 M} (P\,dx + Q\,dy + R\,dz),$$

(рис. 17)

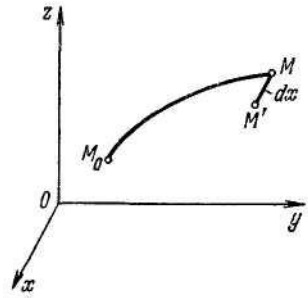


Рис. 17

выберем для определенности точку M_0 в начале координат, а путь, соединяющий M_0 с текущей точкой $M\ (x; y)$, — как на рис. 18.

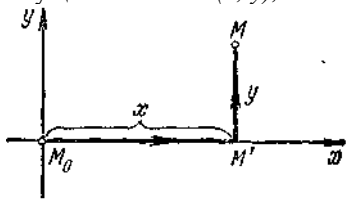


Рис. 18

Получаем

$$u(x,\ y)=\int\limits_{M_0M}[(x^2+2xy)\,dx+(x^2-y^3)\,dy]=$$
$$=\int\limits_{M_0M'}[(x^2+2xy)\,dx+(x^2-y^3)\,dy]+$$
$$+\int\limits_{M'M}[(x^2+2xy)\,dx+(x^2-y^3)\,dy].$$

В первом интеграле надо положить $y = 0$, $dy=0$, тогда как во втором считать x = const, $dx = 0$. Отсюда

$$u(x,\ y)=\int\limits_0^x x^2\,dx+\int\limits_0^y(x^2-y^3)\,dy=\frac{x^3}{3}+x^2y-\frac{y^4}{4}. \tag{7}$$

Итак, общее решениеуравнения (5) имеет вид

$$\frac{x^3}{3}+x^2y-\frac{y^4}{4}=C.$$

Проведем аналогичное исследование уравнения

$$-\frac{y\,dx}{x^2+y^2}+\frac{x\,dy}{x^2+y^2}=0, \tag{8}$$

которое, впрочем, легко проинтегрировать непосредственно, так как переменные в нем разделяются. Это уравнение приходится рассматривать на всей плоскости, за исключением начала координат, как говорят, на плоскости с *выколотым* началом координат, так как при $x = 0$, $y = 0$ оба коэффициента, P и Q, имеют разрыв. Такая область неодносвязная (двусвязная).

Для уравнения (8) условие (4) также выполняется. Для построения функции u по формуле

$$u(M) = \int_{\cup M_0 M} (P\,dx + Q\,dy + R\,dz),$$

выберем точку M_0 где угодно, но, конечно, не в начале координат, например $M_0(1; 0)$. Проведя выкладки, аналогичные (6), (7), и считая сначала, что $x > 0$, получим $u = \operatorname{arctg}\frac{y}{x}$. Эта же функция удовлетворяет соотношению (2) и при $x < 0$; однако если ее рассмотреть во всей плоскости $x, y,$ то она будет иметь разрыв на прямой $x = 0$. Чтобы избавиться от него, можно положить

$$u = \operatorname{Arctg}\frac{y}{x} = \varphi \text{ (полярному углу)}.$$

Правда, эта функция неоднозначна: даже если в некоторой точке $M \neq O$ выбрать какое-либо одно значение φ, а затем заставить M обойти вокруг начала координат, то φ получит приращение 2π. Тем не менее общее решение уравнения (8) имеет вид

$$\operatorname{Arctg}\frac{y}{x} = C, \text{ т. е. } \frac{y}{x} = \operatorname{tg} C = C_1, \quad y = C_1 x,$$

где C_1 — произвольная постоянная; геометрически получаем семейство всевозможных прямых, проходящих через начало координат.

Бывает так, что для уравнения (1) условие (4) не выполнено, т. е. левая часть этого уравнения не является полным дифференциалом, но становится им после умножения на некоторый известный множитель. Например, левая часть уравнения $-y\,dx + x\,dy = 0$ не удовлетворяет условию (4), но начинает удовлетворять после умножения обеих частей на множитель

$$\frac{1}{x^2 + y^2}$$

(см. (8)).

Такой множитель называется *интегрирующим множителем* для рассматриваемого уравнения (1). Никаких общих способов для его нахождения нет; интегрирующий множитель используется в некоторых теоретических исследованиях.

Метод предварительного дифференцирования. В некоторых случаях уравнение $F(x, y, y')=0$ удается проинтегрировать после его *предварительного дифференцирования.* Рассмотрим, например, уравнение

$$x=f(y') \tag{9}$$

или, как принято записывать,

$$x=f(p) \qquad (p = y'). \tag{10}$$

Если продифференцировать обе части, получим

$$dx = f'(p)\,dp.$$

С помощью этого равенства и формулы $\frac{dy}{dx}=p$ находим выражение для *dy:*

$$dy = p\,dx = pf'(p)\,dp,$$

откуда

$$y = \int pf'(p)\,dp + C. \tag{11}$$

Равенства (10) и (11) вместе определяют функциональную зависимость между x и y в параметрическом виде, причем параметром служит p. Мы получили общее решение уравнения (9) в параметрическом виде. Аналогично решается уравнение $y=f(y')$.

Несколько более сложным является *уравнение Лагранюа*

$$y = f(y')x + g(y'), \quad \text{т.е.} \quad y = f(p)x + g(p) \qquad (p=y'). \tag{12}$$

линейное относительно x и y, но нелинейное в основном значении этого слова. После дифференцирования получаем

$$dy = p\,dx = f'(p)\,dp\,x + f(p)\,dx + g'(p)\,dp,$$

т.е.

$$[p - f(p)]\frac{dx}{dp} = f'(p)\,x + g'(p).$$

Если $f(p) \neq p$, то после деления на $p—f(p)$ получается линейное уравнение, в котором x рассматривается как функция от p. После интегрирования этого уравнения получим равенство вида $x = x(p; C)$, которое вместе с (12) даст общее решение исходного уравнения в параметрическом виде

В частном случае, когда $f(p)=p$, уравнение (12) называется *уравнением Клеро* по имени французского математика А. Клеро (1713—1765), впервые рассмотревшего его в 1734 г.; оно имеет вид $y = xy'+g(y')$, т. е.

$$y = xp + g(p) \qquad (p = y'). \tag{13}$$

Предварительное дифференцирование дает

$$p\,dx = p\,dx + x\,dp + g'(p)\,dp$$

т е.

$$dp\,[x + g'(p)] = 0. \tag{14}$$

Приравнивая нулю первый множитель, получим в силу (13)

$$p = C, \text{ т. е. } y = Cx + g(C). \tag{15}$$

Это — общее решение уравнения (13).

Приравнивая нулю второй множитель в левой части (14), получим

$$x = -g'(p), \quad y = xp + g(p) = -pg'(p) + g(p). \tag{16}$$

Значит, получилось еще одно, особое решение уравнения (13), определенное в параметрическом виде. Геометрически формула (15) задает семейство прямых, а формулы (16) — огибающую. (Проверьте последнее утверждение, исходя из уравнения (15).)

Например, уравнение $y = xy' - y'^2$ имеет общее решение

$$y = Cx - C^2 \tag{17}$$

и особое решение, графиком которого служит огибающая семейства прямых (17). Для ее отыскания продифференцируем по C обе часги (17), что даст

$$0 = x - 2C.$$

Исключая C из двух последних формул, получим $C = \frac{x}{2}$, т. е

$$y = \frac{x}{2}x - \left(\frac{x}{2}\right)^2 = \frac{x^2}{4}.$$

Соответствующие интегральные линии показаны на рис. 19.

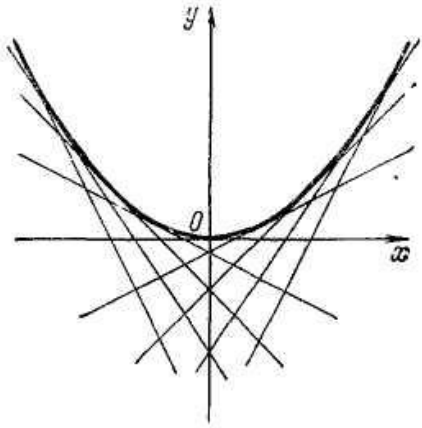


Рис. 19

10. Уравнения высших порядков и системы уравнений

10.1. Основные определения

Чаще всего *дифференциальное уравнение получается как уравнение, связывающее аргумент* или аргументы, *неизвестную функцию* (или функции) и *ее производные;* даже если первоначально было соотношение между дифференциалами, то можно перейти к соотношению между производными. *Если искомая функция зависит от одного аргумента, то дифференциальное уравнение называется обыкновенным;* в противном случае оно называется *уравнением с частными производными*.

Наивысший порядок производной от искомой функции, входящий в уравнение, называется порядком этого уравнения.

В общем случае уравнение *n*-го порядка имеет вид

$$F(x,\ y,\ y',\ y'',\ \ldots,\ y^{(n)}) = 0, \tag{1}$$

где $y = y(x)$ — искомая функция. Конечно, при этом функция F может фактически зависеть не от всех выписанных величин.

Решением дифференциального уравнения называется функция, которая, будучи подставлена в это уравнение, обращает его в тождество. Уже на простейших примерах легко убедиться в том, что *дифференциальное уравнение имеет бесконечное количество решений.* Например, из простейшего уравнения

$$y' = x^2, \quad y = y(x) \tag{2}$$

сразу найдем с помощью интегрирования

$$y = \frac{x^3}{3} + C. \tag{2}$$

Это — *общее решение* уравнения (2); оно включает *произвольную постоянную* C и является записью *всего многообразия решений.* Придавая произвольной постоянной конкретные численные значения, мы получим конкретные, *частные решения* уравнения (2):

$$y = \frac{x^3}{3},\quad y = \frac{x^3}{3} + 6,\ y = \frac{x^3}{3} - \frac{\sqrt{2}}{3} \text{ и т. п.}$$

В общем случае (1) решение находится в результате n последовательных интегрирований, так что *общее решение уравнения n-го порядка содержит n произвольных постоянных,* т. е. имеет вид

$$y = y(x;\ C_1,\ C_2,\ \ldots,\ C_n). \tag{4}$$

Особенно часто общее решение получается в неявной форме:

$$\Phi(x,\ y;\ C_1,\ C_2,\ \ldots,\ C_n) = 0. \qquad (5)$$

Соотношения (4) и (5) называются также *общими интегралами* уравнения (1). Частные решения получаются, если придать каждой произвольной постоянной C_1, C_2, ..., C_n конкретное численное значение. График каждою частного решения называется *интегральной линией рассматриваемого дифференциального уравнения.* Уравнение этой линии — это уравнение (4) и (5) с конкретными C.

Чтобы из общего решения выделить одно частное, требуется, помимо дифференциального уравнения, поставить некоторые дополнительные условия. Чаще всего ставятся *начальные условия,* которые при исследовании процесса, развивающегося во времени, являются *математической записью начального состояния процесса.*

Например, при рассмотрении процесса колебаний пружины из физических соображений ясно, что конкретное колебание полностью определяется, если заданы начальное отклонение и начальная скорость колеблющейся точки. Поэтому начальные условия для уравнения (1) имеют вид

$$\text{при}\quad t = t_0\quad \text{заданы}\quad y = y_0\ \text{и}\ \frac{dy}{dt} = v_0. \qquad (6)$$

В общем случае для уравнения (1) начальные условия имеют следующий вид:

$$\text{при}\ x = x_0\ \text{заданы}\ y = y_0,\quad y' = (y')_0,\ \ldots,\ y^{(n-1)} = (y^{(n-1)})_0. \quad (7)$$

Так как общее решение (5) содержит n произвольных постоянных, то наложенных n соотношений как раз достаточно, во всяком случае принципиально, для нахождения этих постоянных и тем самым для нахождения частного решения. И физически естественно, что если известны дифференциальный закон, управляющий развитием процесса, а также начальное состояние этого процесса, то сам процесс является полностью определенным.

Для уравнения первого порядка (2) условие (7) означает, что при некотором значении $x = x_0$ должно быть задано значение $y = y_0$. Пусть, например, требуется, чтобы $y(1) = 2$. Тогда из (3) получаем

$$2 = \frac{1^3}{3} + C, \qquad C = \frac{5}{3},$$

т. е. искомое частное решение имеет вид

$$y = \frac{x^3 + 5}{3}.$$

Задача о нахождении частного решения дифференциального уравнения при заданном начальном условии называется *задачей Коши.*

10.2. Уравнения высших порядков.

Общие понятия, относящиеся к таким уравнениям, были приведены в п. 10.1 (уравнение (1), общее решение (4) или (5), начальное условие (7)). Впрочем, как и для первого порядка, уравнение порядка *n* обычно бывает проще исследовать, если оно задано в форме, разрешенной относительно старшей производной:

$$y^{(n)} = f(x,\ y,\ y',\ \ldots,\ y^{(n-1)}).$$

В частности, на эту форму непосредственно распространяется теорема Коши: *начальные значения* (7) *определяют одно и только одно решение, если при этих значениях функция f непрерывна и имеет конечные производные первого порядка по y, y', ... , $y^{(n-1)}$.*

Рассмотрим вопрос об интегрировании этих уравнений в квадратурах; в случае уравнений высшего порядка интегрирование удается довести до конца еще реже, чем для уравнений первою порядка. Основным способом формального интегрирования нелинейных уравнений высшего порядка (о линейных уравнениях мы будем говорить особо) является *метод понижения порядка,* т. е. переход к равносильному уравнению низшего порядка. Как правило, чем ниже порядок уравнения, тем оно проще. Кроме того, бывает, что после одного или нескольких понижений порядка мы переходим к уравнению первого порядка одного из интегрируемых типов; тогда интегрирование удается довести до конца. Рассмотрим некоторые частные способы понижения порядка.

1. Пусть, например, задано уравнение второго порядка

$$y'^2 + y\,y'' = 0$$

Для его интегрирования заметим, что левую часть можно переписать в виде

$$y'^2 + yy'' \equiv (yy')',$$

откуда

$$(yy')' = 0;\ \ yy' = C_1;\ \ ydy = C_1 dx;\ \ \frac{y^2}{2} = C_1 x + C_2 \quad \text{(общее решение)}.$$

Другое уравнение

$$y'^2 - yy'' = 0$$

легко проинтегрировать аналогичным образом, если предварительно разделить обе части на y^2:

$$\frac{y'^2 - yy''}{y^2} = 0;\ \ -\left(\frac{y'}{y}\right)' = 0;\ \ \frac{y'}{y} = C_1;\ \ \frac{dy}{y} = C_1 dx;$$

$$\ln|y| = C_1 x + \ln C_2;\ \ y = C_2 e^{C_1 x} \quad \text{(общее решение)}.$$

Как говорят, после деления на y^2 мы получили *интегрируемую комбинацию: «точная производная» приравнена к нулю.* Аналогичный прием иногда применяется и в других примерах. Дальнейшие случаи понижения порядка мы будем для простоты излагать для уравнений второго порядка общего вида

$$F(x,\ y,\ y',\ y'') = 0. \tag{1}$$

2. *Пусть в уравнении* (1) *не присутствует y,* а только производные от него, т. е. мы имеем уравнение вида

$$F(x, y', y'') = 0. \tag{2}$$

Тогда вводят обозначение $y'=p = p(x)$ и из (2) получится

$$F(x, p, p') = 0,$$

т. е. уравнение первого порядка. Если нам повезет и удасгся его проинтегрировать, то получим общее решение уравнения (2)

$$p = \varphi(x;\ C_1); \quad y' = \varphi(x;\ C_1); \quad y = \int \varphi(x;\ C_1)\,dx + C_2.$$

3. *Пусть в уравнении* (1) *не присутствует x,* т. е. мы имеем уравнение

$$F(y, y', y'') = 0. \tag{3}$$

Тогда также обозначают $y' = p,$ но рассматривают p как функцию от $y.$ При этом в (3) нельзя подставлять просто $y''=p',$ так как тогда p' означало бы производную от p по $x,$ а не по $y.$ Поэтому пишут

$$y'' = \frac{d(y')}{dx} = \frac{dp}{dx} = \frac{dp}{dy}\cdot\frac{dy}{dx} = p\frac{dp}{dy}.$$

Из уравнения (3) получим

$$F\left(y,\ p,\ p\,\frac{dp}{dy}\right) = 0,$$

т. е. уравнение первого порядка. Если его удастся проинтегрировать, то мы сможем найти общее решение уравнения (3)

$$p = \varphi(y;\ C_1); \quad \frac{dy}{dx} = \varphi(y;\ C_1); \quad \int \frac{dy}{\varphi(y;\ C_1)} = x + C_2.$$

4. *Пусть левая часть уравнения* (1) *однородна относительно неизвестной функции и ее производных,* т. е.

$$F(x,\ ty,\ ty',\ ty'') \equiv t^k F(x,\ y,\ y',\ y''). \tag{4}$$

В этом случае порядок понижается после подстановки

$$\frac{y'}{y} = u = u(x),$$

откуда

$$y' = uy, \quad y'' = u'y + uy' = u'y + u\cdot uy = (u' + u^2)\,y,$$
$$F(x,\ y,\ yu,\ y(u' + u^2)) = 0, \quad F(x,\ 1,\ u,\ u' + u^2) = 0;$$

в последнем переходе мы воспользовались свойством (4). Если полученное уравнение первого порядка удастся проинтегрировать, то

$$u = \varphi(x;\ C_1);\quad \frac{y'}{y} = \varphi(x;\ C_1);$$

$$\ln|y| = \int \varphi(x;\ C_1)\,dx + \ln C_2;\quad y = C_2 e^{\int \varphi(x;\ C_1)\,dx}.$$

Связь уравнений высшего порядка с системами уравнений первого порядка. От уравнения высшего порядка (1 п.10.1) всегда можно перейти к равносильной системе *n* уравнений первого порядка с *n* неизвестными функциями. Для этого надо обозначить

$$y = y_1,\ y' = y_2,\ \ldots,\ y^{(n-1)} = y_n. \tag{5}$$

Тогда в силу уравнения (1 п.10.1) можно написать

$$\left.\begin{aligned} &y_1' = y_2,\\ &y_2' = y_3,\\ &\cdots\cdots\cdots\cdots\cdots\\ &y_{n-1}' = y_n,\\ &F(x,\ y_1,\ y_2,\ \ldots,\ y_n,\ y_n') = 0. \end{aligned}\right\} \tag{6}$$

Система (6) имеет специальный вид. Общий вид системы первого порядка, например, из трех уравнений с тремя неизвестными функциями таков:

$$\left.\begin{aligned} &F_1(x,\ y_1,\ y_2,\ y_3,\ y_1',\ y_2',\ y_3') = 0,\\ &F_2(x,\ y_1,\ y_2,\ y_3,\ y_1',\ y_2',\ y_3') = 0,\\ &F_3(x,\ y_1,\ y_2,\ y_3,\ y_1',\ y_2',\ y_3') = 0. \end{aligned}\right\} \tag{7}$$

Обратно, из системы (7) можно перейти к одному уравнению (в данном случае третьего порядка) с одной неизвестной функцией, например y_1. Поэтому общее решение системы (7) содержит три произвольные постоянные:

$$y_1 = \varphi_1(x;\ C_1,\ C_2,\ C_3),\ y_2 = \varphi_2(x;\ C_1,\ C_2,\ C_3),$$
$$y_3 = \varphi_3(x;\ C_1,\ C_2,\ C_3).$$

Для указанного перехода надо каждое из уравнений (7) продифференцировать два раза и из полученных девяти равенств, считая вместе с равенствами (7), исключить восемь величин y_2, y_3, y'_2, y'_3, y''_2, y''_3, y'''_2, y'''_3. После интегрирования получается общее решение $y_1 = \varphi_1(x;\ C_1,\ C_2,\ C_3)$, из которого y_2 и y_3 находятся уже без интегрирований на основе упомянутых равенств.

Систему уравнений любого порядка можно преобразовать в систему уравнений первого порядка с помощью обозначений, аналогичных (5). Например, система двух уравнений, третьего порядка относительно одной неизвестной функции и второго порядка относительно другой,

эквивалентна системе из пяти уравнений первого порядка с пятью неизвестными функциями; общее решение в этом примере содержит пять произвольных постоянных.

10.3. Геометрический смысл системы уравнений первого порядка

Рассмотрим для простоты случай системы из двух уравнений первого порядка с двумя неизвестными функциями $y_1(x)$ и $y_2(x)$:

$$\left.\begin{aligned} F_1(x,\ y_1,\ y_2,\ y_1',\ y_2') &= 0, \\ F_2(x,\ y_1,\ y_2,\ y_1',\ y_2') &= 0. \end{aligned}\right\} \qquad (1)$$

Если эту систему до интегрирования удастся разрешить относительно y'_1 и y'_2, то она примет более простой вид

$$\left.\begin{aligned} y_1' &= f_1(x,\ y_1,\ y_2), \\ y_2' &= f_2(x,\ y_1,\ y_2); \end{aligned}\right\} \qquad (2)$$

тогда говорят, что система записана в *нормальной форме.*

Решением сиетсмы (1) или, что то же, (2) называется, конечно, *пара* функций

$$y_1 = y_1(x), \quad y_2 = y_2(x), \qquad (3)$$

обращающая *оба* уравнения в тождества. В *общее решение* входят две произвольные постоянные, т. е. Оно имеет вид

$$y_1 = y_1(x;\ C_1,\ C_2), \quad y_2 = y_2(x;\ C_1,\ C_2).$$

Система уравнений (2) и ее решения (3) имеют простой геометрический смысл, для выяснения которого надо рассмотреть трехмерное пространство x, y_1, y_2. Тогда формулы (3) определят линию в параметрическом виде, причем здесь параметром служит сам x (можно дописать равенство $x = x$); она называется *интегральной линией* системы уравнений (2). Если для произвольной точки M в пространстве (рис. 1) подсчитать значения правых частей системы (2), то мы будем знать направления касательных к линиям $y_1 = y_1(x)$ и $y_2 = y_2(x)$, т. е. к проекциям интегральной линии, и тем самым сможем узнать направление касательной к самой интегральной линии, если она проходит через M.

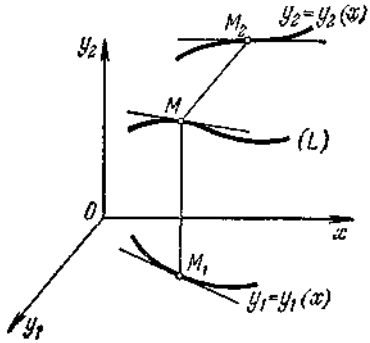


Рис. 1

Значит, система (2) задает *поле направлений в пространстве x, y_1, y_2,* а интегральная линия — это линия, в каждой своей точке идущая «вдоль поля», т. е. линия, в каждой точке которой касательная имеет направление, заданное этим полем.

В системе (2) переменные y_1 и y_2 равноправны, а переменная x имеет иное значение. Бывает, что все три переменные равноправны, так что любую из них можно принять за независимую, тогда систему уравнений предпочитают записывать в *симметричной форме,* например,

$$\frac{dx}{P(x,\ y,\ z)} = \frac{dy}{Q(x,\ y,\ z)} = \frac{dz}{R(x,\ y,\ z)}. \qquad (4)$$

От формы (4) легко перейти к форме (2) и наоборот.

Геометрический смысл системы (4) аналогичен описанному выше. Так как вектор $d\mathbf{r} = dx\mathbf{i}+dy\mathbf{j}+dz\mathbf{k}$ в любой заданной точке $M\ (x;\ y;\ z)$ в силу соотношений (4) должен быть параллелен известному вектору $P\mathbf{i}+Q\mathbf{j}+R\mathbf{k}$, то задача об интегрировании системы (4) — это задача о построении линий в пространстве, имеющих в каждой своей точке заданное направление.

Из геометрического смысла системы (2) вытекает, что для однозначного определения интегральной линии надо задать точку $M_0(x_0; y_{10}; y_{20})$ в пространстве, через которую эта линия должна пройти. Другими словами, *начальное условие*

$$y_1(x_0) = y_{10},\quad y_2(x_0) = y_{20}$$

однозначно определяет решение системы (2). Конечно, и здесь возможны особые точки и особые линии которые распознаются в общем подобно методам особых точек и особых решений. В частности, *для системы* (2) *особой точкой является всякая точка, в которой все три знаменателя обращаются в нуль,* т. е.

вектор $P\mathbf{i}+Q\mathbf{j}+R\mathbf{k}$ обращается в нуль-вектор, не имеющий определенного направления.

Система первого порядка в нормальной форме с любым числом уравнений имеет общий вид

$$\left.\begin{aligned} y_1' &= f_1(x,\ y_1,\ y_2,\ \ldots,\ y_n),\\ y_2' &= f_2(x,\ y_1,\ y_2,\ \ldots,\ y_n),\\ &\ldots\ldots\ldots\ldots\\ y_n' &= f_n(x,\ y_1,\ y_2,\ \ldots,\ y_n). \end{aligned}\right\} \qquad (5)$$

Решение ее — это система функций

$$y_1 = y_1(x),\quad y_2 = y_2(x),\ \ldots,\ y_n = y_n(x). \qquad (6)$$

Общее решение содержит n произвольных постоянных.

Для однозначного определения частного решения можно задать начальное условие

$$y_1(x_0) = y_{10};\quad y_2(x_0) = y_{20},\ \ldots,\ y_n(x_0) = y_{n0}. \qquad (7)$$

Коши доказал, что условиям (7) удовлетворяет ровно одно решетие системы (5), если при значениях $x = x_0$, $y_1 = y_{10}$, ..., $y_n = y_{n0}$ правые части системы (5) непрерывны, а их производные по переменным $y_1, y_2 \ldots . y_n$ конечны.

Геометрический смысл системы (5), решения (6) и условий (7) — это соответственно поле направлений, интегральная линия и точка, через которую должна пройти эта линия в (n+1)-мерном пространстве $x, y_1, y_2, \ldots, y_n$.

Если правые части системы (5) не содержат независимой переменной x, то эта система называется *автономной;* оказывается, что ее решения удобнее рассматривать в n-мерном пространстве $y_1, y_2, \ldots, y_n$, называемом *фазовым пространством.* Мы ограничимся для простоты случлем $n = 2$, будем обозначать независимую переменную буквой t и *истолковывать ее как время,* а искомые функции взамен y_1, y_2 будем обозначать x, y, так что $x = x(t)$, $y = y(t)$. Вместо (5) тогда получится система уравнений

$$\frac{dx}{dt} = P(x,\ y),\quad \frac{dy}{dt} = Q(x,\ y).$$

Если умножить первое уравнение на **i**, второе — на **j**, а затем произвести почленное сложение, мы получим векторное дифференциальное уравнение

$$\frac{d\mathbf{r}}{dt} = \mathbf{A}(x,\ y)\quad (= \mathbf{A}(\mathbf{r})), \qquad (8)$$

где $\mathbf{A} = P(x, y)\,\mathbf{i} + Q(x, y)\,\mathbf{j}$ — заданное векторное поле на *фазовой плоскости* x, y. Так как $\frac{d\mathbf{r}}{dt}$ — это вектор скорости, то на

плоскости x, y оказывается заданным *поле скоростей,* а решение $\mathbf{r}(t) = x(t)\,\mathbf{i} + y(t)\,\mathbf{j}$ определяет закон движения точки на плоскости, при котором эта точка в каждом своем положении имеет скорость, заданную для этого положения. Несколько вольно можно представлять себе, что уравнение (8) задает на фазовой плоскости поток жидкости, а решениям отвечают законы движения частиц этой жидкости. Автономность уравнения (8) означает, что рассматриваемый поток стационарный, а потому различные траектории не имеют друг с другом общих точек.

Запишем, например, уравнение (4) в виде автономной системы первого порядка

$$\frac{dy}{dt} = v, \quad M\frac{dv}{dt} = -ky; \tag{9}$$

здесь y и v—координата и скорость колеблющейся точки. В курсе физики выводится выражение для полной энергии колеблющейся точки

$$E = \frac{Mv^2}{2} + \frac{ky^2}{2} \tag{10}$$

При свободных колебаниях без трения энергия должна сохраняться. И действительно, в силу (9)

$$\frac{dE}{dt} = Mv\frac{dv}{dt} + ky\frac{dy}{dt} = -kyv + kyv = 0;$$

это — математическое доказательство *закона сохранения энергии* в данном примере. Таким образом, E = const для любого решения системы (9), т. е. движения в фазовой плоскости y, v происходят по эллипсам, причем разным эллипсам отвечают колебания вокруг положения равновесия с различной амплитудой (рис. 2).

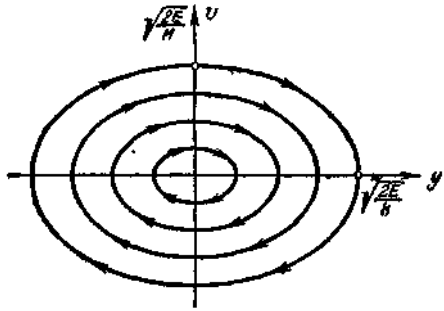


Рис. 2

Первые интегралы. Рассмотрим для определенности систему из трех уравнений первого порядка вида (7 п.10.1). Всякое соотношение вида

$$\Phi(x, y_1, y_2, y_3; C) = 0, \quad (11)$$

обязанное тождественно удовлетворяться для любого решения системы, называется *первым интегралом* этой системы уравнений; здесь C — постоянная, вообще говоря, различная для различных решений. *Знание первого интеграла дает возможность понизить число уравнений в системе на единицу:* например, если из (11) выразить y_3 через все остальное и подставить результат в первые два уравнения (7 п.10.1), то получится система из двух уравнений первого порядка с двумя неизвестными функциями y_1 и y_2. Если ее проинтегрировать, т. е. найти $y_1(x)$ и $y_2(x)$, то $y_3(x)$ можно будет найти без интегрирований из равенства (11).

Аналогичным образом знание двух независимых первых интегралов позволяет понизить число уравнений на два, а три независимых первых интеграла (т. е. таких, что ни один и них не является следствием остальных)

$$\left.\begin{aligned} \Phi_1(x, y_1, y_2, y_3; C_1) &= 0, \\ \Phi_2(x, y_1, y_2, y_3; C_2) &= 0, \\ \Phi_3(x, y_1, y_2, y_3; C_3) &= 0 \end{aligned}\right\}$$

дают общее решение системы (7 п.10.1), записанное в неявной форме.

Иногда первые интегралы удается найти, выводя из заданных уравнений оистемы *интегрируемые комбинации.* Например, для системы .

$$\left.\begin{aligned} y' &= y + z, \\ z' &= -y + z \end{aligned}\right\} \quad (12)$$

легко получить такую комбинацию:

$$yy' + zz' = y(y+z) + z(-y+z) = y^2 + z^2.$$

т. е.

$$\frac{1}{2}(y^2+z^2)' = y^2+z^2, \quad \frac{d(y^2+z^2)}{y^2+z^2} = 2dx, \quad \ln(y^2+z^2) = 2x + \ln C.$$

и окончательно имеем первый интеграл

$$y^2 + z^2 = Ce^{2x}.$$

Из него видно, например, что при $x \to \infty$ решение уходит в бесконечность, а при $x \to -\infty$ решение стремится к нулю; и в других случаях бывает возможно сделать существенные выводы о поведении решений без полного интегрирования системы. Еще один первый интеграл для системы (12) можно получить, разделив одно из уравнений (12) на другое.

В некоторых случаях первые интегралы подсказываются физическими соображениями, чаще всего теми или иными законами сохранения.

Например, формула (10), в которой E играет роль произвольной постоянной C, служит первым интегралом системы (9). Выразив из него v через y и подставив результат в первое уравнение (9), легко довести интегрирование до конца.

Подчеркнем в заключение, что, как видно из предыдущего, наиболее естественно рассматривать системы, в которых число уравнений равно числу неизвестных функций; такие системы принято называть *замкнутыми.* Если уравнений меньше, чем искомых функций, то система называется *незамкнутой (недоопределенной);* у такой системы избыточное количество неизвестных функций можно задавать произвольно. Чаще всего незамкнутость системы свидетельствует о том, что просто не все необходимые соотношения выписаны Если уравнений больше, чем неизвестных функций, то система называется *переопределенной;* такая система обычно противоречива, т. е. ***не имеет решений***. Переопределенность системы обычно свидетельствует либо о ее зависимости, т. е. о том, что некоторые из уравнении являются следствиями остальных и потому излишни, либо об ошибке при ее составлении.

10.4. Дифференциальное уравнение второго порядка

Уравнение

$$F(x, y, y', y'')=0 \qquad (1)$$

называется *дифференциальным уравнением второго порядка.*

Предполагается, что $F\ (u, v, w, g)$ — заданная непрерывно дифференцируемая функция от точек (u, v, w, g) некоторой области Ω четырехмерного пространства.

Любая функция $y=y(x),$ имеющая на некотором интервале непрерывную производную второго порядка и удовлетворяющая уравнению (1), называется *решением* этого уравнения или его *интегральной кривой.*

Каждое из них $y=y(x)$ определено, вообще говоря, на некотором своем интервале $a < x < b.$ Конечно, для любого x из этого интервала точка

$$(x, y(x), y'(x), y''(x)) \in \Omega.$$

Нередко на решение, которое ищут, накладывают дополнительные условия. Особый интерес представляют такие условия, которые гарантируют единственное решение уравнения. Обычно эти условия имеют вид

$$y(x_0) = y_0, \quad y'(x_0) = y'_0 \qquad (2)$$

и называются *начальными условиями.* Задача нахождения решения уравнения (1), удовлетворяющего начальным условиям (2), называется *задачей Коши.* С геометрической точки зрения условия (2) означают, что из семейства интегральных кривых, проходящих через точку (x_0, y_0), мы выделяем определенную интегральную кривую, имеющую заданный угол наклона ($\operatorname{tg}\alpha = y'(x_0) = y'_0$, рис. 3).

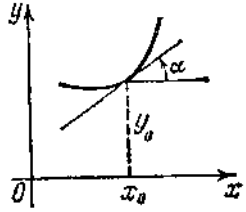


Рис. 3

В уравнение (1) могут не входить явно все переменные x, y, y', но y'' должно входить, иначе это уравнение не будет дифференциальным уравнением второго порядка, например,

$$x^2 + y' + y'' = 0, \quad y'y'' + 1 = 0.$$

Разрешим уравнение (1) относительно y''. Будем предполагать это возможным. Из теории неявных функций известно, что если функция $F(u, v, w, g)$ равна нулю в некото рой точке (u_0, v_0, w_0, g_0), имеет непрерывные частные производные в окрестности этой точки и частная производная $\dfrac{\partial F}{\partial g} \neq 0$ в этой точке, то уравнение $F(u, v, w, g) = 0$ имеет в некоторой окрестности указанной точки решение $g = f(u, v, w)$ и притом единственное. Тогда уравнение (1) примет вид

$$y'' = f(x, y, y'), \qquad (3)$$

где функция $f(u, v, w)$ задана на некоторой области ω трехмерного пространства точек (u, v, w), непрерывна на ней и имеет непрерывные частные производные. Функция f может и не зависеть явно от некоторых из переменных x, y, y'. Например, это имеет место для уравнений $y'' = \varphi(x)$, $y'' = y' + y$, $y'' = y$, $y'' = y'$.

Пусть некоторая интегральная кривая $y = y(x)$ проходит через точку (x_0, y_0) и имеет в этой точке угловой коэффициент касательной, равный заданному числу y'_0 (т. е. $y(x_0) = y_0$, $y'(x_0) = y'_0$).

Этим однозначно определяется вторая производная от $y(x)$ в точке x_0, равная

$$y''_0 = f(x_0, y_0, y'_0) \qquad (y''_0 = y''(x_0)).$$

Однако возникает вопрос, если мы зададим $x = x_0$ и произвольные числа y_0, y_0, то существует ли на самом деле интегральная кривая $y = y(x)$ уравнения (3), для которой $y(x_0) = y_0$ и $y'(x_0) = y'_0$, и как много таких интегральных кривых. Следующая теорема показывает, что функция f в окрестности точки (x_0, y_0, y'_0) достаточно гладкая, то такая интегральная кривая существует и притом одна.

Теорема 1. *Пусть правая часть уравнения* (3), *рассматриваемая как функция трех переменных (x, y, y'), заданная на трехмерной области* ω, *непрерывна и имеет на этой области непрерывные частные производные*

$$\frac{\partial f}{\partial y}, \frac{\partial f}{\partial y'}.$$

Тогда, какова бы ни была точка $(x_0, y_0, y'_0) \in \omega$, *существует интервал (a, b) и определенная на нем дважды непрерывно дифференцируемая функция y=y(x), удовлетворяющая дифференциальному уравнению* (3) *и начальным условиям* $y(x_0) = y_0$, $y'(x_0) = y'_0$.

Функция, обладающая указанными свойствами, единственная.

Про функцию $y(x)$ говорят, что она есть *решение (интегральная кривая)* дифференциального уравнения (3), удовлетворяющее *начальным условиям* (2). Или еще говорят, что она решает задачу Коши для указанных начальных условий.

Каждое такое решение удобно записывать в виде

$$y(x) = y(x, y_0, y'_0),$$

где y_0, y'_0 — параметры решения. Они независимы — их можно взять какими угодно, лишь бы точка $(x_0, y_0, y'_0) \in \omega$.

Если зафиксировать x_0, то каждой системе чисел $C_1 = y_0$, $C_2 = y'_0$, $(x_0, C_1, C_2) \in \omega$, соответствует решение дифференциального уравнения (3), которое можно записать (при фиксированном x_0) в виде

$$y = y(x, C_1, C_2),$$

где C_1, C_2 — произвольные постоянные — параметры.

Пример. Найти интегральную кривую уравнения $y''+y = 0$, проходящую через точку (0, 1) и имеющую в ней угловой коэффициент касательной $y'(0)=0$.

Легко проверить, что функция $y = C_1 \cos x + C_2 \sin x$ является решением данного уравнения при любых постоянных C_1 и C_2. Далее $y(0) = C_1$, $y'(0) = C_2$. Чтобы выполнялись начальные условия, необходимо положить $C_1 = 1$, $C_2 = 0$. Итак, искомая интегральная кривая имеет вид: $y = \cos x$.

10.5. Система из двух дифференциальных уравнений первого порядка

В уравнении

$$\frac{d^2y}{dx^2} = f(x,\ y,\ y') \tag{1}$$

наряду с его решением $y = y(x),\ x \in (a,\ b)$, введем функцию $z = z(x)$, полагая $y' = z$. Тогда оно будет эквивалентно следующей системе из двух дифференциальных уравнений первого порядка;

$$\left.\begin{aligned} y' &= z, \\ z' &= f(x,\ y,\ z) \end{aligned}\right\} \tag{2}$$

относительно двух неизвестных функций y и z.

В самом деле, пусть $y(x),\ x \in (a,\ b)$, есть решение дифференциального уравнения (1). Оно имеет вторую непрерывную производную на $(a,\ b)$. Тогда $z(x) = y'(x)$ имеет первую непрерывную производную на (a, b). Таким образом, функции $y(x)$ и $z(x)$ имеют непрерывную производную и удовлетворяют системе дифференциальных уравнений (2).

Обратно, если две функции $y(x),\ z(x),\ x \in (a,\ b)$, имеют непрерывные производные на $(a,\ b)$ и удовлетворяют системе (2), то из первого уравнения системы (2) следует, что $y(x)$ имеет вторую непрерывную производную на $(a,\ b)$, а подставляя z из первого уравнения во второе, получим, что $y(x)$ есть решение дифференциального уравнения (1).

Система (2) есть частный случай системы

$$\left.\begin{aligned} \frac{dy}{dx} &= \varphi(x,\ y,\ z), \\ \frac{dz}{dx} &= \psi(x,\ y,\ z) \end{aligned}\right\} \tag{3}$$

относительно известных функций y и z.

Эта последняя, очевидно, есть частный случай системы

$$\left.\begin{aligned} F(x,\ y,\ z,\ y',\ z') &= 0, \\ \Phi(x,\ y,\ z,\ y',\ z') &= 0, \end{aligned}\right\} \tag{4}$$

где мы будем предполагать, что функции F и Φ непрерывны и имеют непрерывные частные производные по $y,\ y',\ z,\ z'$ в некоторой области точек $x,\ y,\ z,\ y',\ z'$.

Пара функций $y(x),\ z(x)$ называется *решением системы дифференциальных уравнений* (4), если эти функции определены на некотором интервале $(a,\ b)$, зависящем от этих функций, имеют непрерывные производные и удовлетворяют на (a, b) системе (4).

Если решить уравнения (4) относительно y' и z', то получим систему вида (3) (конечно, предполагаем, что решение системы (4)

относительно y' и z' возможно, что, как известно, обычно связано с неравенством нулю якобиана

$$\frac{D(F, \Phi)}{D(y', z')}\Big) .$$

Уравнения (3) (или (4)) образуют систему двух дифференциальных уравнений первого порядка относительно двух неизвестных функций y и z.

Система (3), разрешенная относительно $\frac{dy}{dx}$, $\frac{dz}{dx}$, называется *нормальной.*

Для нормальной системы (3) справедлива

Теорема 1 (существования). *Пусть функции* $\varphi \in (x, y, z)$ *и* $\psi(x, y, z)$ *непрерывны и имеют непрерывные частные производные по* y *и* z *на области* ω *точек* (x, y, z), *и пусть задана произвольная точка* (x_0, y_0, z_0) ω.

Тогда существует интервал (a, b) *и определенные на нем непрерывно дифференцируемые функции* $y=y(x)$, $z=z(x)$, *удовлетворяющие системе* (3) *и начальным условиям*

$$y(x_0) = y_0, \quad z(x_0) = z_0. \tag{5}$$

Указанные функции $y=y(x)$, $z = z(x)$ *единственны.*

При этом, если функции φ *и* ψ *имеют непрерывные частные производные порядка* p, *то решения* $y(x)$ *и* $z(x)$ *непрерывно дифференцируемы* $p+1$ *раз на указанном интервале* (a, b).

Выше было показано, что решение уравнения (1) второго порядка относительно одной функции может быть сведено к решению двух уравнений первого порядка относительно двух неизвестных функций (система (2)).

Но общая система дифференциальных уравнений первого порядка вида (3) тоже сводится к решению одного дифференциального уравнения второго порядка. В самом деле, подставив в систему (3) вместо y и z некоторое ее решение $y=y(x)$, $z=z(x)$ и продифференцировав по x первое уравнение, получим

$$\frac{d^2y}{dx^2} = \frac{\partial\varphi}{\partial x} + \frac{\partial\varphi}{\partial y}\cdot\varphi + \frac{\partial\varphi}{\partial z}\cdot\psi = \Phi(x, y, z). \tag{6}$$

Наряду с (6) будем рассматривать также первое уравнение (3)

$$\frac{dy}{dx} = \varphi(x, y, z), \tag{7}$$

в котором подставлены $y=y(x)$, $z = z(x)$.

Найдем z из (7) ($z = \chi(x, y, y')$) и подставим в (6), тогда получим дифференциальное уравнение второго порядка

$$\frac{d^2y}{dx^2} = \Phi(x,\ y,\ \chi(x,\ y,\ y')) \equiv \Lambda(x,\ y,\ y') \qquad (8)$$

относительно рассматриваемой функции $y = y(x)$.

Мы получили, что если $y(x)$, $z(x)$ — решения системы (3), то $y(x)$ — решение уравнения второго порядка.

Конечно, для того чтобы было возможным проделать эти выкладки, потребовались новые свойства от функций φ и ψ: непрерывная дифференцируемость φ по x, y, z и возможность разрешить первое уравнение (3) относительно z.

10.6. Линейные уравнения общего вида

Линейные однородные уравнения. Исследование линейных уравнений любого порядка во многом аналогично исследованию линейных уравнений первого порядка, хотя теперь уже получить решение в квадратурах в общем случае не удастся. Рассмотрим сначала для простоты уравнение второго порядка. Уравнение

$$z'' + p(x)\,z' + q(x)\,z = 0, \qquad (1)$$

левая часть которого линейна относительно неизвестной функции и ее производных, называется *линейным однородным уравнением.*

Обозначим для краткости левую часть уравнения (1) через $L[z]$, т. е. в данном случае

$$L[z] \equiv z'' + p(x)\,z' + q(x)\,z \ \text{(по определению)}.$$

Тогда уравнение (1) можно переписать в виде $L[z] = 0$. Выражение $L[z]$ обладает следующими свойствами:

$$L[z_1 + z_2] = (z_1 + z_2)'' + p(x)(z_1 + z_2)' + q(x)(z_1 + z_2) =$$
$$= (z_1'' + p(x)z_1' + q(x)z_1) + (z_2'' + p(x)z_2' + q(x)z_2) = L[z_1] + L[z_2],$$

$$L[Cz] = CL[z] \quad (C = \text{const}) \ \text{(проверяется аналогично)}.$$

О таких выражениях, называемых линейными операторами, мы упоминали ранее.

Легко доказать следующие свойства уравнения (1).

1. *Сумма решений уравнения* (1) *будет решением того же уравнения.* Действительно, если z_1 и z_2 — два таких решения, т. е.

$$L[z_1] = 0 \ \text{и} \ L[z_2] = 0, \ \text{то} \ L[z_1 + z_2] = L[z_1] + L[z_2] = 0.$$

Аналогично проверяется свойство 2:

2. *Если решение уравнения* (1) *умножить на константу, то получится решение того же уравнения.*

Свойства 1 и 2 можно объединить так: линейная комбинация решений уравнения (1) будет решением того же уравнения. Например, если $z_1(x)$ и $z_2(x)$ удовлетворяют уравнению (1), то и

$$z = C_1 z_1(x) + C_2 z_2(x) \qquad (2)$$

удовлетворяет тому же уравнению при любых постоянных C_1, C_2.

3. *Тождественно нулевая функция удовлетворяет уравнению* (1).

4. *Если известно ненулевое решение уравнения* (1), *то его порядок можно понизить на единицу.* Действительно, пусть $z_1(x)$ — такое решение; сделаем подстановку $z = z_1 u$, где $u = u(x)$ — новая неизвестная функция. По лучим

$$(z_1'' u + 2z_1' u' + z_1 u'') + p(z_1' u + z_1 u') + q z_1 u = 0,$$

т. е.

$$z_1 u'' + (2z_1' + p z_1) u' + (z_1'' + p z_1' + q z_1) u = 0.$$

Но так как $L[z_1] = 0$, то последний член отпадает и после подстановки $u' = v$ получаем

$$z_1 v' + (2z_1' + p z_1) v = 0,$$

т. е. линейное однородное уравнение на единицу низшего, чем было, порядка. Доведем интегрирование до конца:

$$\frac{dv}{v} = -\frac{2z_1' + p z_1}{z_1}\,dx, \quad \ln|v| = -2\ln|z_1| - \int p(x)\,dx + \ln C_2,$$

$$v = \frac{C_2}{z_1^2} e^{-\int p(x)\,dx}, \quad u = C_2 \int \frac{1}{z_1^2} e^{-\int p(x)\,dx}\,dx + C_1,$$

$$z = C_1 z_1 + C_2 z_1 \int \frac{1}{z_1^2} e^{-\int p(x)\,dx}\,dx. \qquad (3)$$

Функция, при которой стоит множитель C_2, является одним из частных решений уравнения (1), так как она получается из общего решения (3), если положить $C_1 = 0$, $C_2 = 1$. Поэтому если обозначить ее через z_2, то мы приходим к свойству 5:

5. *Общее решение уравнения* (1) *имеет вид* (2), где C_1 и C_2 — произвольные постоянные, а z_1 и z_2 — два частных решения этого уравнения.

В этом свойстве *в качестве* z_1, z_2 *могут быть взяты только два линейно независимых решения,* а не любая пара решений. Понятие линейной зависимости функций вводится подобно аналогичному понятию для векторов: именно, несколько функций называются *линейно зависимыми* друг от друга, если одна из них является линейной комбинацией остальных. В частности, две функции $z_1(x)$ и $z_2(x)$ линейно зависимы, если $z_2(x) \equiv C z_1(x)$, т. е. если они пропорциональны. Тогда формула (2) не дает общего решения, так как

$$C_1 z_1 + C_2 z_2 \equiv C_1 z_1 + C_2 C z_1 \equiv (C_1 + C_2 C)\, z_1(x) \equiv D z_1(x),$$

где $D = C_1 + C_2 C$ — постоянная; значит, хотя формально справа имеются две произвольные постоянные, но они не являются существенными, т. е. их число можно уменьшить на единицу.

Все указанные свойства справедливы и для линейного однородного уравнения любого порядка

$$z^{(n)} + p(x)\, z^{(n-1)} + q(x)\, z^{(n-2)} + \ldots + s(x)\, z = 0, \qquad (4)$$

за тем исключением, что общее решение, взамен (2), имеет вид

$$z = C_1 z_1(x) + C_2 z_2(x) + \ldots + C_n z_n(x). \qquad (5)$$

Здесь все C — произвольные постоянные, а z_1, z_2, ..., z_n — какие-либо линейно независимые решения уравнения (4). Совокупность n линейно независимых решений уравнения (4) порядка n называется *фундаментальной системой решений.* Таким образом, *общее решение уравнения* (4) *есть линейная комбинация решений из фундаментальной системы с произвольными коэффициентами.* Можно сказать, что *совокупность всех решений уравнения* (4) *образует n-мерное линейное пространство; фундаментальная система решений — это базис в этом пространстве.*

Отметим в заключение, что у уравнения (4) можно понизить порядок на единицу, но это делают редко, так как после понижения порядка уравнение становится нелинейным.

Неоднородные уравнения. Рассмотрим теперь *линейное неоднородное уравнение*

$$y'' + p(x)\, y' + q(x)\, y = f(x). \qquad (6)$$

Обозначим левую часть через $L[y]$.

1. *Знание какого-либо частного решения уравнения* (6) *позволяет свести задачу об интегрировании этого уравнения к задаче об интегрировании соответствующего* (т. е. с отброшенной правой частью) *однородного уравнения* (1).

Действительно, если $Y(x)$ — такое решение, то, сделав замену

$$y = Y(x) + z, \qquad (7)$$

где $z = z(x)$ — новая неизвестная функция, получим

$$L[Y + z] = f(x), \quad L[Y] + L[z] = f(x).$$

Однако $L[Y] = f(x)$ и мы получаем уравнение (1) для z. Итак, *общее решение линейного неоднородного уравнения* (6) *есть сумма какого-либо его частного решения и общего решения соответствующего однородного уравнения*.

2. *Если правая часть f(x) равна линейной комбинации, например, двух функций, т. е.* $f(x) = \alpha f_1(x) + \beta f_2(x)$ $(\alpha, \beta = \text{const})$, *и известны*

какие-либо частные решения $Y_1(x)$ и $Y_2(x)$ уравнения (6) *с правыми частями $f_1(x)$ и $f_2(x)$, то функция*

$$Y(x) = \alpha Y_1(x) + \beta Y_2(x)$$

служит частным решением уравнения (6) *с правой частью f(x).*

3. *Если известно общее решение однородного уравнения* (1), *то общее решение уравнения* (6) *можно найти с помощью квадратур.* Это делается с помощью найденного Лагранжем *метода вариации произвольных постоянных* следующим образом. Как мы знаем, общее решение уравнения (1) имеет вид (2). Мы ищем решение уравнения (6) в виде

$$y = \varphi_1(x) z_1(x) + \varphi_2(x) z_2(x), \tag{8}$$

где φ_1, φ_2 — некоторые неизвестные пока функции. Так как их две, а уравнение (6) одно, то для нахождения этих функций мы наложим на них еще одно дополнительное соотношение ((10)). Дифференцируя равенство (8), получим

$$y' = (\varphi_1 z_1' + \varphi_2 z_2') + (\varphi_1' z_1 + \varphi_2' z_2). \tag{9}$$

Потребуем, чтобы вторая скобка обратилась в нуль:

$$\varphi_1' z_1 + \varphi_2' z_2 = 0. \tag{10}$$

Тогда при дифференцировании равенства (9) надо принимать во внимание только первую скобку, т. е.

$$y'' = (\varphi_1 z_1'' + \varphi_2 z_2'') + (\varphi_1' z_1' + \varphi_2' z_2'). \tag{11}$$

Подставляем все полученные результаты (8), (9) и (11) в уравнение (6), конечно, не выписывая нулевой суммы. Это даст

$$\varphi_1 (z_1'' + p z_1' + q z_1) + \varphi_2 (z_2'' + p z_2' + q z_2) + (\varphi_1' z_1' + \varphi_2' z_2') = f(x).$$

Поскольку функции z_1, z_2 удовлетворяют уравнению (1), то в последнем уравнении первые две скобки отпадают и оно превращается в равенство

$$\varphi_1' z_1' + \varphi_2' z_2' = f(x). \tag{12}$$

Итак, для нахождения φ_1, φ_2 у нас остались два соотношения: (10) и (12). Так как z_1, z_2 и $f(x)$ считаются известными, то получается система двух алгебраических уравнений первой степени с двумя неизвестными: φ'_1, φ'_2. Решая систему, мы находим эти неизвестные, а интегрируя, находим φ_1, φ_2.

Рассмотрим, например, простейшее *уравнение вынужденных колебаний,* которое получится, если в правой части уравнения добавить внешнюю силу *P(t).* Разделив обе части уравнения на *M,* получим

$$y'' + \omega_0^2 y = f(t), \tag{13}$$

где обозначено $\omega_0^2 = \frac{k}{M}$, $f(t) = \frac{P(t)}{M}$.

Соответствующее однородное уравнение

$$z'' + \omega_0^2 z = 0 \tag{14}$$

имеет, как легко непосредственно проверить, два решения,

$$z_1 = \cos \omega_0 t, \quad z_2 = \sin \omega_0 t,$$

и тем самым общее решение

$$z = C_1 \cos \omega_0 t + C_2 \sin \omega_0 t. \tag{15}$$

Отсюда видно, в частности, что ω_0 — это *собственная частота* колебаний рассматриваемой системы, т. е. частота колебаний при отсутствии внешних сил.

Согласно формуле (8) решение уравнения (13) ищем в виде

$$y = \varphi_1(t) \cos \omega_0 t + \varphi_2(t) \sin \omega_0 t. \tag{16}$$

Тогда уравнения (10) и (12) приобретают вид

$$\left.\begin{aligned} &\varphi_1' \cos \omega_0 t + \varphi_2' \sin \omega_0 t = 0, \\ &\varphi_1' (-\omega_0 \sin \omega_0 t) + \varphi_2' \omega_0 \cos \omega_0 t = f(t). \end{aligned}\right\}$$

Отсюда непосредственно находим

$$\varphi_1'(t) = -\frac{1}{\omega_0} f(t) \sin \omega_0 t, \quad \varphi_2'(t) = \frac{1}{\omega_0} f(t) \cos \omega_0 t.$$

При интегрировании здесь неудобно воспользоваться неопределенным интегралом из-за наличия в нем неуточняемой произвольной постоянной; лучше нижний предел интеграла зафиксировать, например, положив его равным моменту $t = 0$ начала отсчета времени;

$$\varphi_1(t) = -\frac{1}{\omega_0} \int_0^t f(t) \sin \omega_0 t \, dt + C_1,$$

где C_1 — произвольная постоянная. Поскольку в правой части t имеет два смысла — переменная интегрирования и верхний предел, то лучше воспользоваться независимостью определенного интеграла от обозначения переменной интегрирования и написать

$$\varphi_1(t) = -\frac{1}{\omega_0} \int_0^t f(\tau) \sin \omega_0 \tau \, d\tau + C_1;$$

$$\varphi_2(t) = \frac{1}{\omega_0} \int_0^t f(\tau) \cos \omega_0 \tau \, d\tau + C_2.$$

Подставляя в (16), получим

$$y = -\frac{1}{\omega_0}\cos\omega_0 t\int_0^t f(\tau)\sin\omega_0\tau\,d\tau + \\ + \frac{1}{\omega_0}\sin\omega_0 t\int_0^t f(\tau)\cos\omega_0\tau\,d\tau + C_1\cos\omega_0 t + C_2\sin\omega_0 t.$$

Теперь внесем $\cos\omega_0 t$ и $\sin\omega_0 t$ под знак интеграла (этого нельзя было бы сделать, не переменив обозначения переменной интегрирования t на τ) и объединим оба интеграла:

$$y = \frac{1}{\omega_0}\int_0^t [-f(\tau)\cos\omega_0 t\sin\omega_0\tau + f(\tau)\sin\omega_0 t\cos\omega_0\tau]\,d\tau + \\ + C_1\cos\omega_0 t + C_2\sin\omega_0 t.$$

Отсюда получаем общее решение уравнения (13):

$$y = \frac{1}{\omega_0}\int_0^t \sin\omega_0(t-\tau)f(\tau)\,d\tau + C_1\cos\omega_0 t + C_2\sin\omega_0 t. \qquad (17)$$

Произвольные постоянные C_1 и C_2 можно определить, например, из начальных условий

$$y(0) = y_0, \quad y'(0) = v_0. \qquad (18)$$

Подставляя в обе части (17) значение $t = 0$, получим $y_0 = C_1$. Чтобы использовать второе условие (18), надо продифференцировать равенство (17) по t, после чего подставить $t=0$. При дифференцировании интеграла надо иметь в виду, что t входит в него дважды, т. е.

$$y' = \frac{1}{\omega_0}\int_0^t \omega_0\cos\omega_0(t-\tau)f(\tau)\,d\tau + \\ + \left[\frac{1}{\omega_0}\sin\omega_0(t-\tau)f(\tau)\right]_{\tau=t} - C_1\omega_0\sin\omega_0 t + C_2\omega_0\cos\omega_0 t = \\ = \int_0^t \cos\omega_0(t-\tau)f(\tau)\,d\tau - C_1\omega_0\sin\omega_0 t + C_2\omega_0\cos\omega_0 t;$$

$$v_0 = C_2\omega_0; \quad C_2 = \frac{v_0}{\omega_0}.$$

Отсюда решение уравнения (13) при начальных условиях (18)

$$y = \frac{1}{\omega_0}\int_0^t \sin\omega_0(t-\tau)f(\tau)\,d\tau + y_0\cos\omega_0 t + \frac{v_0}{\omega_0}\sin\omega_0 t.$$

Все указанные свойства справедливы и для уравнения

$$y^{(n)} + p(x)\,y^{(n-1)} + q(x)\,y^{(n-2)} + \ldots + s(x)\,y = f(x). \qquad (19)$$

Метод вариации произвольных постоянных будет выглядеть так: в формулу (5) вместо $C_1, C_2, ..., C_n$ надо подставить $\varphi_1(x), \varphi_2(x), ..., \varphi_n(x)$, после чего последовательно дифференцировать эту формулу, приравнивая на каждом шаге, вплоть до $(n-1)$-го, получающуюся группу членов с φ'_k нулю; n-е соотношение получится из подстановки всех полученных выражений в (19).

4. *Любое решение уравнения* (4), *а также* (19) *можно продолжить на любой интервал, на котором коэффициенты и правая часть не обращаются в бесконечность.* Для нелинейных уравнений может получиться, что решение или его производные при таком продолжении уходят в бесконечность для конечного значения x.

Краевые задачи. Для того чтобы выделить частное решение из общего, мы пользовались начальными условиями, согласно которым искомая функция и ее производные задаются при каком-либо одном значении аргумента. Имеются и другие способы выделения частного решения из общего, которые встречаются в практических задачах. Все эти способы объединяет то, что количество дополнительных равенств, накладываемых на искомое решение, должно равняться числу степеней свободы в общем решении рассматриваемого уравнения, т. е. порядку этого уравнения.

Эти дополнительные равенства порядка n можно записать в виде

$$G_k[y] = \alpha_k \qquad (k = 1, 2, \ldots, n), \tag{20}$$

где $G_k[y]$ — какая-либо заданная комбинация значений искомой функции $y(x)$ и ее производных при, вообще говоря, различных значениях аргумента (точнее, $G_k[y]$ — это какой-либо заданный *функционал*, а α_k — заданные числа.

Если известно общее решение заданного уравнения, то для нахождения требуемого частного решения надо выражение для общего решения подставить в условия (20), в результате чего получится система n уравнений с n неизвестными $C_1, C_2, ..., C_n$.

Если

$$G_k[C_1y_1 + C_2y_2] \equiv C_1G_k[y_1] + C_2G_k[y_2] \qquad (C_1, C_2 = \text{const}),$$

то условия (20) называются *линейными;* если к тому же все $\alpha_k = 0$, то они называются *линейными однородными.* Если какие-нибудь функции, не обязательно решения дифференциального уравнения, удовлетворяют линейным однородным условиям, то и их любая линейная комбинация тоже удовлетворяет этим условиям. Действительно, если, например, $G_k[y_1] = 0$ и $G_k[y_2] = 0$, то

$$G_k[C_1y_1 + C_2y_2] = C_1G_k[y_1] + C_2G_k[y_2] = C_1 \cdot 0 + C_2 \cdot 0 = 0.$$

Разность двух функций, удовлетворяющих одинаковым неоднородным линейным условиям, удовлетворяет соответствующим однородным условиям.

В дальнейшем мы рассмотрим решение уравнения

$$y'' + p(x)\,y' + q(x)\,y = f(x) \qquad (a \leqslant x \leqslant b) \tag{21}$$

при дополнительных условиях

$$y(a) = \alpha_1, \quad y(b) = \alpha_2, \tag{22}$$

хотя *все полученные общие выводы справедливы для линейных дифференциальных уравнений любого порядка n при линейных дополнительных условиях* (20) *любого вида.* Интервал (*a*, *b*), а также функции *p(x), q(x)* и *f(x)* будем считать конечными, что дает возможность считать любое решение продолженным на весь этот интервал, включая концы. Условия вида (22), наложенные на концах интервала, в котором строится решение, называются *краевыми условиями,* а задача о решении дифференциального уравнения при заданных краевых условиях называется *краевой задачей.*

Для решения краевой задачи мы исходим из вида общего решения уравнения (21)

$$y(x) = Y(x) + C_1 z_1(x) + C_2 z_2(x) \tag{23}$$

где *Y (x)* — некоторое частное решение уравнения (21), а z_1 и z_2—два линейно независимых решения соответствующего однородного уравнения. Подставляя формулу (23) в условия (22), получим два соотношения для нахождения C_1 и C_2:

$$\left.\begin{aligned} C_1 z_1(a) + C_2 z_2(a) &= \alpha_1 - Y(a), \\ C_1 z_1(b) + C_2 z_2(b) &= \alpha_2 - Y(b). \end{aligned}\right\} \tag{24}$$

При решении этой системы двух алгебраических уравнений первой степени с двумя неизвестными могут представиться два случая.

1. *Основной случай:* определитель системы отличен от нуля. В этом случае система (24) имеет вполне определенное решение и потому *уравнение* (21) *при условиях* (22) *имеет одно и только одно решение при любом неоднородном члене f (x) и любых числах* α_1, α_2.

2. *Особый случай:* определитель системы равен нулю. В этом случае система (24), как правило, противоречива, но при некоторых правых частях она имеет бесконечное количество решений. Значит, и *уравнение* (21) *при условиях* (22) *при произвольном выборе функции f (x) и чисел* α_1, α_2*, как правило, не имеет ни одного решения, однако при некоторых таких выборах задача имеет бесконечное количество решений.* Например, можно проверить, что если *f (x)* и α_1 уже выбраны, то бесконечное количество решений получится лишь при

одном значении α_2, а при остальных значениях задача не будет иметь ни одного решения.

Подчеркнем, что то, какой именно случай имеет место, зависит от вида левых частей уравнения (21) и условий (22).

Для того чтобы имел место основной случай, необходимо и достаточно, чтобы соответствующая однородная задача, в которой положено $f(x) = 0$, $\alpha_1 = \alpha_2 = 0$, *имела только нулевое решение.* В особом случае однородная задача имеет бесконечное количество решений, а если неоднородная задгча имеет хотя бы одно решение, то общее решение получится, если к этому частному решению прибавить общее решение соответствующей однородной задачи.

При решении начальной задачи, т. е. задачи Коши, всегда имеет место основной случай, так как такое решение всегда существует и единственно. При решении краевой задачи может представиться и особый случай.

Например, рассмотрим задачу с параметром $\lambda = \text{const}$,

$$y'' + \lambda y = f(x) \quad (0 \leqslant x \leqslant l), \quad y(0) = \alpha_1, \quad y(l) = \alpha_2, \tag{25}$$

причем будем считать сначала, что $\lambda > 0$. Тогда линейно независимыми решениями соответствующего однородного дифференциального уравнения служат функции $z_1(x) = \cos \sqrt{\lambda}\, x$, $z_2(x) = \sin \sqrt{\lambda}\, x$ и определитель системы (24) равен

$$\begin{vmatrix} z_1(0) & z_2(0) \\ z_1(l) & z_2(l) \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ \cos \sqrt{\lambda}\, l & \sin \sqrt{\lambda}\, l \end{vmatrix} = \sin \sqrt{\lambda}\, l.$$

Приравнивая его нулю, получим значения

$$\lambda = \left(\frac{\pi}{l}\right)^2, \quad \left(\frac{2\pi}{l}\right)^2, \quad \left(\frac{3\pi}{l}\right)^2, \ldots, \tag{26}$$

при которых для задачи (25) имеет место особый случай, т. е. нарушается либо сущесгвование, либо единственность решения.

Набор значений параметра, входящего в формулировку задачи, при которых задача в том или ином смысле вырождается, называется *спектром* этой задачи. При $\lambda \leq 0$ для задачи (25) всегда имеет место основной случай и тем самым набор значений (26) представляет собой ее спектр.

Полученный результат имеет важное приложение к исследованию устойчивости упругого стержня при его сжатии. Пусть однородный (одинаковый по всей длине) упругий невесомый стержень расположен вдоль оси x и сжимается вдоль нее силой P, причем оба конца стержня удерживаются на оси x, но могут свободно вращаться вокруг точек закрепления (рис. 4, *а*).

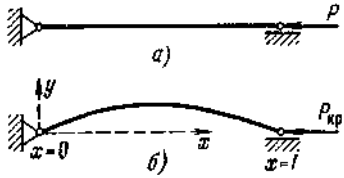


Рис. 4

Тогда при достижении силой некоторого *критического значения* $P_{кр}$ стержень выпучивается, принимая положение, изображенное на рис. 4, *б*. Если обозначить через y поперечное отклонение точки стержня от ее исходного положения, то, как доказывается в курсах сопротивления материалов, функция $y(x)$ с достаточной точностью удовлетворяет дифференциальному уравнению и краевым условиям:

$$y'' + \frac{P}{EJ} y = 0, \quad y(0) = y(l) = 0; \tag{27}$$

здесь E и J — так называемые *модуль Юнга* (Т. Юнг, 1773—1829,— английский физик, врач и астроном, один из создателей волновой теории света) и *момент инерции* поперечного сечения стержня. Как вытекает из (27), при

$$\frac{P}{EJ} < \left(\frac{\pi}{l}\right)^2 \tag{28}$$

для задачи (27) имеет место основной случай, т. е. она имеет только нулевое решение; выпучивания не происходит. Как только с увеличением P неравенство (28) переходит в равенство, то наступает особый случай и задача (27) наряду с нулевым решением приобретает решение вида $y = C \sin\frac{\pi}{l} x$, где C — произвольная постоянная. Но тогда стержень ничем не удерживается в прямолинейном состоянии и как угодно малые внешние воздействия могут привести к конечным отклонениям от этого состояния: стержень *теряет устойчивость*. Получающееся выражение для $P_{кр}$

$$P_{кр} = EJ \left(\frac{\pi}{l}\right)^2$$

было найдено Эйлером в 1757 г. Могло бы показаться, что при $P > P_{кр}$ стержень должен опять выпрямиться. Однако это не так. Уравнение (27) описывает отклонение стержня точно лишь в пределе при малых отклонениях, а анализ более точного нелинейного уравнения, справедливого при любых отклонениях, показывает, что при переходе P через $P_{кр}$ наряду с неустойчивой прямолинейной возникает искривленная форма равновесия, которая и является устойчивой. С

ростом P кривизна этой формы быстро возрастает и стержень разрушается.

К решению неоднородного уравнения при однородных краевых условиях

$$y''+p(x)y'+q(x)y=f(x) \quad (a\leqslant x\leqslant b), \quad y(a)=0, \quad y(b)=0, \quad (29)$$

в основном (неособом) случае можно применить функцию влияния. Действительно, функцию $f(x)$ можно истолковать как «внешнее воздействие», а $y(x)$ — как его результат, т. е. $y(x)=\tilde{f}(x)$. При этом имеет место принцип суперпозиции.

Если через $G(x;\ \xi)$ обозначить решение задачи (29), в которой вместо $f(x)$ взята дельта-функция $\delta(x-\xi)$, то при произвольной функции $f(x)$ решение задачи (29) получится по формуле

$$y(x)=\int_a^b f(\xi)\,G(x;\ \xi)\,d\xi. \quad (30)$$

Приведем простой пример. Пусть рассматривается задача

$$y''=f(x) \quad (0\leqslant x\leqslant l), \quad y(0)=y(l)=0 \quad (31)$$

Если взамен $f(x)$ поставить $\delta(x-\xi)$, то при $0\leq x<\xi$ и при $\xi<x\leq l$ получаем просто $y''=0$, т. е. решение

$$y=ax+b \quad (0\leqslant x<\xi), \quad y=cx+d \quad (\xi<x\leqslant l),$$

где a, b, c, d — какие-то постоянные. Применение краевых условий показывает, что $b=0$ и $cl+d=0$, т. е.

$$y=ax \quad (0\leqslant x<\xi), \quad y=c(x-l) \quad (\xi<x\leqslant l). \quad (32)$$

Если равенство $y''=\delta(x-\xi)$ проинтегрировать от $x=\xi-0$ до $x=\xi+0$, то получится, что $y'(\xi+0)-y'(\xi-0)=1$; кстати, для левой части уравнения (29) получился бы такой же результат, так как интегрирование конечной функции по отрезку нулевой длины дает нуль. При вторичном интегрировании дельта-функции получается уже непрерывная функция, так что $y(\xi-0)=y(\xi+0)$, и из (32) получаем $c-a=1$, $a\xi=c(\xi-l)$, откуда

$$a=-\frac{l-\xi}{l}, \quad c=\frac{\xi}{l}.$$

Подставляя в (32), находим функцию влияния для задачи (31):

$$G(x,\ \xi)=\begin{cases} -\dfrac{(l-\xi)x}{l} & (0\leqslant x<\xi); \\ -\dfrac{\xi(l-x)}{l} & (\xi<x\leqslant l). \end{cases}$$

Эта функция изображена на рис. 5.

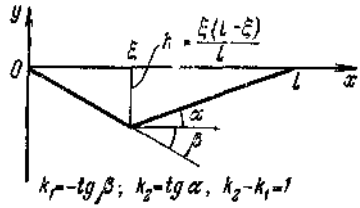


Рис. 5.

В силу формулы (30) получаем решение задачи (31) при любой функции $f(x)$:

$$y=\int_0^l G(x;\ \xi)\,f(\xi)\,d\xi=\int_0^x G(x;\ \xi)\,f(\xi)\,d\xi+\int_x^l G(x;\ \xi)\,f(\xi)\,d\xi=$$

$$=-\frac{(l-x)}{l}\int_0^x \xi f(\xi)\,d\xi-\frac{x}{l}\int_x^l (l-\xi)\,f(\xi)\,d\xi.$$

10.7. Линейные уравнения с постоянными коэффициентами

Линейные уравнения с постоянными коэффициентами составляют важнейший класс уравнений, интегрирование которых сравнительно легко доводится до конца.

Однородные уравнения. Рассмотрим для определенности уравнение третьего порядка

$$z'''+a_1z''+a_2z'+a_3z=0 \qquad (a_1,\ a_2,\ a_3=\text{const}). \tag{1}$$

Эйлер предложил искать частные решения этого уравнения в форме

$$z=e^{px}, \tag{2}$$

где p— постоянная, которую нужно подобрать. Подставляя (2) в (1), получим, что

$$e^{px}(p^3+a_1p^2+a_2p+a_3)=0,$$

и так как первый множитель отличен от нуля, то

$$p^3+a_1p^2+a_2p+a_3=0. \tag{3}$$

Итак, *для того чтобы функция* (2) *удовлетворяла уравнению* (1), *необходимо и достаточно, чтобы p удовлетворяло уравнению* (109). *Алгебраическое* относительно *p уравнение* (3) называется *характеристическим уравнением* для уравнения (1), а левая часть уравнения (3) называется *характеристическим многочленом* для уравнения (1).

Степень характеристического уравнения равна порядку соответствующего дифференциального уравнения.

Уравнение (3) имеет три корня: p_1, p_2, p_3. При этом могут быть различные случаи.

1. *Пусть все корни вещественные и простые,* т. е. различные. Тогда в силу формулы (2) мы имеем три частных решения уравнения (1)

$$z_1 = e^{p_1 x},\ z_2 = e^{p_2 x}\ z_3 = e^{p_3 x}.$$

Так как они являются независимыми, т. е. ни одно из них не равно линейной комбинации остальных, то общее решение уравнения (1) имеет вид

$$z = C_1 e^{p_1 x} + C_2 e^{p_2 x} C_3 e^{p_3 x} \qquad (4)$$

2. *Пусть все корни простые, но среди них имеются мнимые.* Тогда в правой части формулы (4) оказывается комплексная функция от вещественного аргумента. Но вся теория линейных уравнений автоматически распространяется на случай, когда все коэффициенты и решения являются такими функциями. Поэтому *и при указанных корнях уравнения* (3) *можно пользоваться формулой* (4); конечно, тогда произвольные постоянные будут, вообще говоря, комплексными.

Однако если все рассмотрения производятся над вещественными функциями, то часто предпочитают и ответ получить в вещественной форме. Для этого можно воспользоваться следующим замечанием: *если линейное однородное уравнение с вещественными коэффициентами имеет комплексное частное решение, то его вещественная и мнимая части также являются решениями того же уравнения.* Действительно, если $L[y_1 + iy_2] = 0$, то $L[y_1] + iL[y_2] = 0$, откуда $L[y_1] = 0$ и $L[y_2] = 0$.

Значит, если коэффициенты уравнения (1) вещественные и оно имеет частное решение

$$e^{(r+is)x} = e^{rx}\cos sx + ie^{rx}\sin sx$$

то функции

$$e^{rx}\cos sx, \quad e^{rx}\sin sx \qquad (5)$$

также служат решениями уравнения (1). Если вспомнить, что у алгебраического уравнения с вещественными коэффициентами сопряженные корни присутствуют парами, то получаем, что в рассматриваемом случае 2 корни уравнения (3) имеют вид

$$p_1 = r + is, \quad p_2 = r - is \qquad (p_3 \text{ — вещественное}),$$

и потому решение можно вместо (4) записать в вещественной форме:

$$z = C_1 e^{rx} \cos sx + C_2 e^{rx} \sin sx + C_3 e^{p_3 x}. \qquad (6)$$

Например, для уравнения свободных колебаний (14 п.10.6) получаем характеристическое уравнение $p^2 + \omega_0^2 = 0$ с корнями $p_{1,2} = \pm i\omega_0 = 0 \pm i\omega_0$ и аналогично формуле (6) пишем общее решение

$$z = C_1 e^{0t} \cos \omega_0 t + C_2 e^{0t} \sin \omega_0 t,$$

т. е. как раз решение (15 п.10.6).

Формулу (6) иногда записывают в ином виде, преобразовав

$$C_1 \cos sx + C_2 \sin sx = M \sin (sx + \alpha),$$

для чего надо положить

$$C_1 = M \sin \alpha, \quad C_2 = M \cos \alpha, \quad M = \sqrt{C_1^2 + C_2^2}, \quad \operatorname{tg} \alpha = \frac{C_1}{C_2}$$

Тогда взамен (6) получим

$$z = Me^{rx} \sin (sx + \alpha) + C_3 e^{p_3 x}, \qquad (7)$$

где произвольными постоянными являются уже M, α и C_3.

3. *Пусть среди корней характеристического уравнения* (3) *имеются кратные, например,* $p_2 = p_1$, , $p_3 \neq p_1$. Тогда формула (4), конечно, не даст общего решения и в виде (2) мы получим лишь два решения, $e^{p_1 x}$ и $e^{p_2 x}$.

Чтобы найти третье решение, рассмотрим сначала случай, когда $p_2 = p_1 + \Delta p$, причем $|\Delta p|$ мало́. Тогда уравнение (1) наряду с решением $e^{p_1 x}$ имеет решение

$$e^{p_2 x} = e^{p_1 x} e^{\Delta p \cdot x} = e^{p_1 x} \left(1 + \Delta p \cdot x + \frac{(\Delta p)^2 x^2}{2!} + \dots\right)$$

а потому служат решениями и их линейные комбинации

$$e^{p_2 x} - e^{p_1 x} = e^{p_1 x} \left(\Delta p \cdot x + \frac{(\Delta p)^2 x^2}{2!} + \dots\right),$$

$$\frac{e^{p_2 x} - e^{p_1 x}}{\Delta p} = e^{p_1 x} \left(x + \frac{\Delta p \cdot x^2}{2!} + \dots\right). \qquad (8)$$

Это деление на Δp дает возможность перейти к пределу при $\Delta p \to 0$. Тогда в правой части все члены, содержащие Δp, отпадут, и потому при $\Delta p = 0$, т. е. $p_2 = p_1$ решением служит функция $x e^{p_1 x}$.

Значит в рассматриваемом случае решением уравнения (1) будет

$$z = C_1 e^{p_1 x} + C_2 x e^{p_1 x} + C_3 e^{p_3 x}.$$

Подобным образом в случае $p_1 = p_2 = p_3$ частными решениями уравнения (1) наряду с $e^{p_1 x}$ служат функции $x e^{p_1 x}$ и $x^2 e^{p_1 x}$; при

доказательстве этого надо взамен (8) рассмотреть вторую разделенную разность. Поэтому в данном случае общее решение имеет вид

$$z = C_1 e^{p_1 x} + C_2 x e^{p_1 x} + C_3 x^2 e^{p_1 x}.$$

Рассмотрение уравнений любого порядка проходит аналогично. Если какой-либо корень p характеристического уравнения имеет кратность k, то функции

$$e^{px},\ xe^{px},\ \ldots,\ x^{k-1}e^{px}$$

будут частными решениями рассматриваемого дифференциального уравнения. Если какая-либо пара корней $r \pm is$ имеет кратность k, то частными решениями будут функции

$$e^{rx}\cos sx,\quad e^{rx}\sin sx,\quad xe^{rx}\cos sx,\quad xe^{rx}\sin sx,\ \ldots$$
$$\ldots,\ x^{k-1}e^{rx}\cos sx,\ x^{k-1}e^{rx}\sin sx.$$

Итак, практическая трудность при решении линейного однородного уравнения с постоянными коэффициентами состоит единственно в решении соответствующего характеристического уравнения.

В качестве примера рассмотрим свободные колебания материальной точки при линейном законе упругости и при дополнительном *вязком трении,* пропорциональном первой степени скорости. В этом случае в правой части уравнения

$$M\frac{d^2y}{dt^2} = F = -ky$$

надо добавить слагаемое $-f\dfrac{dy}{dt}$, где f *—коэффициент трения.* После переноса всех членов налево и деления на M получим уравнение

$$z'' + 2hz' + \omega_0^2 z = 0, \quad \text{где} \quad 2h = \frac{f}{M}, \quad \omega_0^2 = \frac{k}{M}. \qquad (9)$$

При решении характеристического уравнения

$$p^2 + 2hp + \omega_0^2 = 0 \qquad (10)$$

возникают два основных случая. Если $h<\omega_0$, т. е. если трение сравнительно мало, уравнение (10) имеет решение

$$p_{1,2} = -h \pm \sqrt{h^2 - \omega_0^2} = -h \pm i\sqrt{\omega_0^2 - h^2},$$

а потому общее решение уравнения (9) имеет вид, подобный (7),

$$z(t) = Me^{-ht}\sin(\omega t + \alpha), \quad \text{где} \quad \omega = \sqrt{\omega_0^2 - h^2}.$$

Мы видим, что наличие небольшого трения делает колебания затухающими по экспоненциальному закону (множитель e^{-ht}) и уменьшает частоту (так как $\omega< \omega_0$). График решения показан на рис. 6.

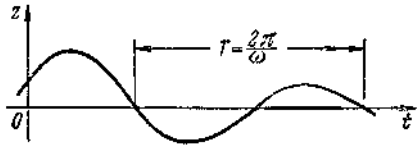


Рис. 6

Нули решения определяются множителем sin (ωt+α) и потому находятся на равном расстоянии друг от друга. Через каждый промежуток времени $T = \frac{2\pi}{\omega}$, когда синус повторяет свои значения, из e^{-ht} выделяется множитель $e^{-h\frac{2\pi}{\omega}}$, из-за чего и происходит затухание. Значение T часто называется «периодом» колебания, хотя $z(t)$ здесь непериодическая функция, так как на каждом следующем «периоде» колебание, не меняясь по форме, уменьшается по размаху в одно и то же число раз.

Если $h > \omega_0$, т, е. если трение сравнительно велико, то уравнение (10) имеет вещественные корни, а уравнение (9) имеет общее решение

$$z(t) = C_1 e^{p_1 t} + C_2 e^{p_2 t} = C_1 e^{-\left(h - \sqrt{h^2 - \omega_0^2}\right) t} + C_2 e^{-\left(h + \sqrt{h^2 - \omega_0^2}\right) t}.$$

При больших t здесь существенно только первое слагаемое, т. е. мы получаем затухание по экспоненциальному закону без колебаний; это — так называемое *апериодическое затухание.*

Теоретически возможен «пограничный случай» $h=\omega_0$. Тогда уравнение (10) имеет двойной корень. И здесь получится апериодическое затухание.

Неоднородные уравнения с правыми частями специального вида. Рассмотрим теперь линейное неоднородное уравнение с постоянными коэффициентами, например, третьего порядка:

$$y''' + a_1 y'' + a_2 y' + a_3 y = f(x) \qquad (a_1,\ a_2,\ a_3 = \text{const}). \quad (11)$$

Поскольку соответствующее однородное уравнение всегда решается, то нам остается лишь найти какое-либо частное решение уравнения (11). Для правой части общего вида это делается по методу вариации произвольных постоянных. Но для важного довольно широкого класса правых частей специального вида частное решение можно найти значительно быстрее по методу неопределенных коэффициентов. Рассмотрим сначала уравнение

$$y''' + a_1 y'' + a_2 y' + a_3 y = Ke^{\lambda x} \qquad (K,\ \lambda = \text{const}). \quad (12)$$

Естественно искать частное решение этого уравнения в форме

$$y = Ae^{\lambda x}, \tag{13}$$

где постоянная A пока неизвестна. Подстановка в (12) даст

$$A\lambda^3 e^{\lambda x} + Aa_1\lambda^2 e^{\lambda x} + Aa_2\lambda e^{\lambda x} + Aa_3 e^{\lambda x} = Ke^{\lambda x},$$

или, после сокращения,

$$A = \frac{K}{\lambda^3 + a_1\lambda^2 + a_2\lambda + a_3}, \qquad y = \frac{K}{P(\lambda)} e^{\lambda x}, \tag{14}$$

где через $P(\lambda)$ обозначен характеристический многочлен.

Полученный результат годится, если $P(\lambda) \neq 0$, т. е. если λ не является корнем характеристического уравнения. Если $P(\lambda) = 0$, то функция (13) удовлетворяет однородному уравнению (1), т. е. уравнению (12) удовлетворить в такой форме невозможно.

Пусть $P(\lambda) = 0$, $P'(\lambda) \neq 0$, т. е. λ является простым корнем характеристического уравнения. Тогда рассуждаем подобно тому, как мы рассматпривали случай кратных корней. Заменим в правой части (12) λ на $\lambda_1 = \lambda + \alpha$, где $|\alpha| \neq 0$, но мало́. Тогда в силу формулы (14), так как λ_1 уже не будет корнем характеристического уравнения, уравнение (12) будет обладать частным решением

$$\frac{K}{P(\lambda_1)} e^{\lambda_1 x} = \frac{K}{P(\lambda+\alpha)} e^{(\lambda+\alpha)x} = \frac{K}{P(\lambda+\alpha)} e^{\lambda x}\left(1 + \alpha x + \frac{\alpha^2 x^2}{2!} + \ldots\right) =$$
$$= \frac{K}{P(\lambda+\alpha)} e^{\lambda x} + K\frac{\alpha}{P(\lambda+\alpha)} e^{\lambda x}\left(x + \frac{\alpha x^2}{2!} + \ldots\right).$$

Однако первое из полученных слагаемых удовлетворяет соответствующему однородному уравнению; значит, второе слагаемое также является частным решением уравнения (12), в котором λ пока еще заменено на λ_1. Если в этом втором слагаемом перейти к пределу при $\alpha \to 0$, вычислив

$$\lim_{\alpha \to 0} \frac{\alpha}{P(\lambda+\alpha)}$$

по правилу Лопиталя, то мы в пределе получим частное решение уравнения (12) при исходном значении λ:

$$y = \frac{K}{P'(\lambda)} xe^{\lambda x}. \tag{15}$$

Подобным образом, если λ является двойным корнем характеристического уравнения, то частное решение уравнения (12) имеет вид

$$y = \frac{K}{P''(\lambda)} x^2 e^{\lambda x}$$

и т. д.

С помощью аналогичного, но более громоздкого рассуждения можно доказать, что уравнение

$$y''' + a_1 y'' + a_2 y' + a_3 y = Q_m(x)\, e^{\lambda x}, \qquad (16)$$

где $Q_m(x)$ — заданный многочлен степени m, если λ не является корнем характеристического уравнения, обладает частным решением вида

$$y = R_m(x)\, e^{\lambda x}, \qquad (17)$$

где $R_m(x)$ — некоторый другой многочлен степени m. Его можно найти по методу неопределенных коэффициентов, т. е. написать его сначала с буквенными коэффициентами, подставить в (16) и найти эти коэффициенты из условия тождественного равенства левой и правой частей. Если же λ является корнем характеристического уравнения кратности k, то имеется частное решение вида

$$y = x^k R_m(x)\, e^{\lambda x}.$$

При этом не исключен случай $\lambda = 0$, когда в правой части уравнения (16) стоит «чистый» многочлен. Можно также рассмотреть уравнение

$$y''' + a_1 y'' + a_2 y' + a_3 y = Q_m(x)\, e^{\mu x} \cos \nu x.$$

Так как правую часть можно переписать в виде

$$Q_m(x)\, e^{\mu x} \frac{e^{i\nu x} + e^{-i\nu x}}{2} = \frac{Q_m(x)}{2} e^{(\mu + i\nu)x} + \frac{Q_m(x)}{2} e^{(\mu - i\nu)x},$$

то в силу (17), если $\lambda = \mu \pm i\nu$ не является корнем характеристического уравнения, частное решение можно искать в виде

$$\begin{aligned} y &= R_m(x)\, e^{(\mu + i\nu)x} + \tilde{R}_m(x)\, e^{(\mu - i\nu)x} = \\ &= R_m(x)\, e^{\mu x}(\cos \nu x + i \sin \nu x) + \tilde{R}_m(x)\, e^{\mu x}(\cos \nu x - i \sin \nu x) = \\ &= [R_m(x) + \tilde{R}_m(x)]\, e^{\mu x} \cos \nu x + [iR_m(x) - i\tilde{R}_m(x)]\, e^{\mu x} \sin \nu x. \end{aligned}$$

Вводя новые обозначения, получим частное решение вида

$$y = T_m(x)\, e^{\mu x} \cos \nu x + S_m(x)\, e^{\mu x} \sin \nu x, \qquad (18)$$

где $T_m(x)$ и $S_m(x)$ — многочлены степени m, которые можно найти по методу неопределенных коэффициентов. В виде (18) ищутся частные решения уравнений и с правыми частями

$$\begin{gathered} Q_m(x)\, e^{\mu x} \sin \nu x, \quad Q_m(x)\, e^{\mu x} \cos(\nu x + \alpha), \\ Q_m(x)\, e^{\mu x} \sin(\nu x + \alpha) \qquad (\alpha = \text{const}). \end{gathered}$$

Если $\lambda = \mu \pm i\nu$ является корнем характеристического уравнения кратности k, то правую часть формулы (18) надо умножить еще на x^k.

Рассмотрим, например, уравнение вынужденных колебаний при синусоидальном внешнем воздействии с частотой ω:

$$y'' + \omega_0^2 y = K \sin \omega t. \qquad (19)$$

Согласно формуле (18), если $\lambda = \pm i\omega$ не является корнем характеристического уравнения, т. е. если $\omega \neq \omega_0$, то частное решение надо искать в виде

$$y = A \cos \omega t + B \sin \omega t.$$

Подстановкав уравнение (19) дает

$$-A\omega^2 \cos \omega t - B\omega^2 \sin \omega t + A\omega_0^2 \cos \omega t + B\omega_0^2 \sin \omega t = K \sin \omega t.$$

Так как это равенство должно быть тождеством, то

$$-A\omega^2 + A\omega_0^2 = 0, \quad -B\omega^2 + B\omega_0^2 = K,$$

откуда

$$A = 0, \quad B = \frac{K}{\omega_0^2 - \omega^2}, \quad y = \frac{K}{\omega_0^2 - \omega^2} \sin \omega t. \qquad (20)$$

Общее решение уравнения (19) получится, если добавить общее решение соответствующего однородного уравнения. Итак, если частота внешнего воздействия не равна собственной частоте колебаний, то получается наложение двух гармонических колебаний. Одно, обычно называемое *вынужденным,* происходит с частотой внешнего воздействия и имеет вполне определенные амплитуду и начальную фазу; другое происходит с собственной частотой, и его амплитуда и начальная фаза зависят от начальных данных.

Из формулы (20) видно, что если ω близко к ω_0, то амплитуда вынужденного колебания становится очень большой. Если же $\omega = \omega_0$, то согласно общей теории частное решение уравнения (19) надо искать в форме $y = At \cos \omega_0 t + B\, t \sin \omega_0 t$.
Подсчет дает

$$y = -\frac{K}{2\omega_0} t \cos \omega_0 t;$$

это можно вывести также из (20) аналогично формуле (15).

Мы видим, что если частота внешнего воздействия равна собственной частоте колебаний, то амплитуда вынужденного колебания возрастает по линейному закону. Это важное явление хорошо известно в физике и технике и называется *резонансом.*

Уравнение Эйлера. Так называется линейное уравнение вида

$$(ax+b)^n y^{(n)} + a_1 (ax+b)^{n-1} y^{(n-1)} + \ldots + a_{n-1} (ax+b) y' + a_n y = f(x),$$

в котором все коэффициенты $a_1, a_1, \ldots, a_n$ постоянны.

Уравнение Эйлера легко приводится к линейному уравнению с постоянными коэффициентами при помощи замены независимой переменной

$$|ax+b| = e^t, \qquad t = \ln |ax+b|.$$

Будем для простоты считать, что $ax + b>0$, а уравнение однородно и имеет второй порядок:

$$(ax+b)^2 y'' + a_1 (ax+b) y' + a_2 y = 0. \qquad (21)$$

После замены независимой переменной получим

$$ax+b = e^t, \qquad t = \ln (ax+b), \qquad (22)$$

$$y' = \frac{dy}{dx} = \frac{dy}{dt} \cdot \frac{dt}{dx} = \frac{dy}{dt} \cdot ae^{-t};$$

$$y'' = \frac{dy'}{dx} = \frac{dy'}{dt} \cdot \frac{dt}{dx} = \left(\frac{d^2y}{dt^2} ae^{-t} - \frac{dy}{dt} ae^{-t} \right) ae^{-t}.$$

Подстлвим эти результаты в уравнение (21):

$$a^2 \left(\frac{d^2y}{dt^2} - \frac{dy}{dt} \right) + a_1 a \frac{dy}{dt} + a_2 y = 0.$$

Это — уравнение с постоянными коэффициентами, которое надо решать по методам однородных уравнений, т. е. положить

$$y = e^{pt}, \qquad (23)$$

решить характеристическое уравнение и т д., после чего вернуться от t к x. Можно не делать замены (22), заметив, что из (22) и (23) следует

$$y = (ax+b)^p. \qquad (24)$$

Поэтому можно путем непосредственной подстановки в (21) искать частные решения вида (24), причем для нахождения p получится характеристическое уравнение, степень которого равна порядку уравнения (21). Надо только иметь в виду, что при наличии кратных корней характеристического уравнения, помимо решений вида (24), уравнение (21) будет обладать решением вида

$$y = te^{pt} = (ax+b)^p \ln (ax+b)$$

и т. д., в зависимости от кратности корня.

Операторы и операторное решение уравнений. Ранее нами был введен *дифференциальный оператор* L. Другими распространенными операторами являются: *оператор сдвига* T и *оператор образования разности* Δ, действующие по формулам

$$Tf(x) = f(x+h); \quad \Delta f(x) = f(x+h) - f(x) \qquad (25)$$

при заданном шаге h; оператор C умножения на заданное число C, в том числе *единичный оператор* 1, оставляющий функции без изменения, и *нулевой оператор* 0, переводящий все функции в тождественно нулевую функцию; оператор-умножения на какую-либо заданную функцию и т. п.

Операторы можно складывать друг с другом и умножать на числа по естественному правилу: если A и B — операторы, а α—число, то по определению

$$(A+B)f = Af + Bf, \quad (\alpha A)f = \alpha(Af).$$

Например, из равенств (25) видно, что

$$\Delta = T - 1, \qquad T = 1 + \Delta.$$

При этом выполняются все аксиомы линейных действий.

Операторы можно умножать друг на друга, что дает новый оператор, действующий по следующему правилу:

$$(AB)f = A(Bf),$$

т. е. на функцию f действует сначала оператор B, а затем на результат — оператор A. Нетрудно проверить правила

$$A(BC) = (AB)C, \quad (\alpha A + \beta B)C = \alpha AC + \beta BC \qquad (\alpha, \beta = \text{const}) \quad (26)$$

Однако далеко не всегда AB = BA, т. е. результат выполнения двух операций может зависеть от порядка действий. Если все же AB=BA, то операторы A и B называются *перестановочными (коммутирующими)* друг с другом. Например, все приведенные выше операторы D, T, Δ, C перестановочны друг с другом, так как

$$DTf(x) = D(Tf(x)) = D(f(x+h)) = f'(x+h), \quad TDf(x) = T(f'(x)) = f'(x+h) \text{ и т. п.}$$

С другой стороны, операторы дифференцирования и умножения на *функцию* неперестановочны.

Первое свойство (26) дает возможность взамен написанных там выражений писать просто ABC, получается оператор, состоящий в последовательном применении действий C, B и A. Аналогично определяется оператор ABCD и т. д Если взять множители равными, то получатся *степени оператора:* A^2, A^3, A^4 и т. д., которые означают результат повторения оператора A. Например, $D^2f = f''$; $\Delta^2 f$ — это вторая разность и т. п.

Оператор A называется *линейным*, если

$$A(f_1 + f_2) = A(f_1) + A(f_2) \qquad \text{и} \qquad A(\alpha f) = \alpha A(f) \qquad (\alpha = \text{const}). \quad (27)$$

Первое свойство можно истолковать как принцип суперпозиции, а второе возможно вывести из первого. Поэтому даже если явный вид оператора неизвестен, то при выполнении принципа суперпозиции можно заключить о линейности этого оператора, что дает возможность сделать некоторые полезные выводы, например построить ***функцию влияния.***

Оба свойства (27) можно написать вместе так:

$$A(\alpha f_1 + \beta f_2) = \alpha A f_1 + \beta A f_2 \qquad (\alpha, \beta = \text{const}).$$

Для линейного оператора A нетрудно проверить свойство

$$A(\alpha B + \beta C) = \alpha AB + \beta AC; \tag{28}$$

для этого надо обе части применить к любой функции f и показать, что получится одно и то же, именно, $\alpha A(Bf) + \beta A(Cf)$.

Все операторы, приведенные выше в качестве примеров, линейные, так как производная суммы равна сумме производных и т. п. Нелинейными являются, например, операторы возведения функции в квадрат или образования абсолютной величины и т. п. Для линейных операторов в силу (26) и (28) при линейных действиях и умножении можно пользоваться обычными правилами школьной алгебры, следя за порядком множителей; например,

$$(A+B)^2 = (A+B)(A+B) = A^2 + AB + BA + B^2$$

и т. п.

Если к тому же операторы перестановочны, то и порядок множителей несуществен, т. е. $(A+B)^2 = A^2 + 2AB + B^2$ и т. п.

Можно рассматривать и степенные ряды от операторов; например,

$$e^A = 1 + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \dots \tag{29}$$

и т. п. Применять такой ряд можно, как и любой оператор, вообще говоря, не ко всем функциям, а только к тем, для которых он имеет смысл. В примере (29) это значит, что чем больше взять членов, тем точнее должен получиться результат; совершенно точный результат теоретически получается лишь в пределе, а практически — при достаточно большом числе членов ряда.

Ряд Тейлора

$$f(x+h) = f(x) + \frac{f'(x)}{1!}h + \frac{f''(x)}{2!}h^2 + \dots$$

можнозаписать в виде

$$Tf = \left(1 + \frac{D}{1!}h + \frac{D^2}{2!}h^2 + \dots\right) f = e^{hD} f,$$

откуда мы видим связь междуоператорами T, Δ и D:

$$T = e^{hD}, \quad \Delta = e^{hD} - 1.$$

Обратная формула

$$D = \frac{1}{h}\ln T = \frac{1}{h}\ln(1+\Delta) = \frac{1}{h}\left(\Delta - \frac{\Delta^2}{2} + \frac{\Delta^3}{3} - \dots\right)$$

Можно рассматривать *операторное уравнение*

$$Ay = f, \tag{30}$$

где функция f задана, а функция y ищется. Если решение имеется, то его естественно обозначить $y = A^{-1}f$. Если оператор А линейный, то и уравнение (30) называется линейным. На линейные уравнения немедленно распространяются свойства 1—3 линейных однородных уравнений и свойство 1 неоднородных уравнений; однако надо иметь в виду, что бывают случаи, когда однородное уравнение имеет бесконечное количество линейно независимых решений, а также случаи, когда неоднородное уравнение не имеет ни одного решения.

Покажем применение оператора дифференцирования к решению *(операторный метод решения)* линейного дифференциального уравнения с постоянными коэффициентами (11). Уравнение перепишем в виде

$$(D^3 + a_1 D^2 + a_2 D + a_3)\, y = f(x).$$

В левой части в скобках стоит линейный дифференциальный оператор третьего порядка с постоянными коэффициентами. По правилам алгебры разлагаем его на множители:

$$(D - p_1)(D - p_2)(D - p_3)\, y = f(x), \tag{31}$$

где p_1, p_2, p_3 — корни характеристического уравнения. Заметив, что

$$D(e^{-px}y) = (e^{-px}y)' = -pe^{-px}y + e^{-px}y' = e^{-px}(y' - py) = e^{-px}(D - p)\, y$$

и потому

$$(D - p)\, y = e^{px} D(e^{-px}y),$$

перепишем уравнение (31) в виде

$$e^{p_1 x} D(e^{-p_1 x}\, e^{p_2 x} D(e^{-p_2 x}\, e^{p_3 x} D(e^{-p_3 x} y))) = f(x).$$

Перенося множители из левой части в правую и пользуясь тем, что равенство $Dy = z$ равносильно $y = \int z\,dx$, получим общее решение исходного уравнения

$$y = e^{p_3 x} \int e^{(p_2 - p_3)x} \left(\int e^{(p_1 - p_2)x} \left(\int e^{-p_1 x} f(x)\,dx \right) dx \right) dx.$$

В более сложных задачах применение операторов может принести существенную пользу. Отметим, что для линейных дифференциальных операторов с переменными коэффициентами и тем более для нелинейных операторов такое простое разложение *(факторизацию)* оператора на произведение нескольких операторов более низкого порядка эффективно удается осуществить лишь в очень редких случаях.

Приведем еще один простой пример. Пусть надо найти решение уравнения

$$y'' + \omega^2 y = f(x),$$

причем все величины считаются вещественными. Пишем

$$(D^2+\omega^2)\,y=f(x),\quad (D-i\omega)(D+i\omega)\,y=f(x),$$
$$e^{i\omega x}D\left(e^{-i\omega x}(D+i\omega)\,y\right)=f(x);$$
$$(D+i\omega)\,y=e^{i\omega x}\int e^{-i\omega x}f(x)\,dx,$$
$$\omega y=\operatorname{Im}\left(e^{i\omega x}\int e^{-i\omega x}f(x)\,dx\right)$$

(Im означает мнимую часть. Окончательно,

$$y=\frac{1}{\omega}\operatorname{Im}\left(e^{i\omega x}\int e^{-i\omega x}f(x)\,dx\right).$$

10.8. Системы линейных уравнений

Системы линейных уравнений. Рассмотрим для определенности *линейную однородную систему* трех уравнений первого порядка с тремя искомыми функциями *y(x), z(x)* и *u(x),* разрешенных относительно производных от этих функций:

$$\left.\begin{aligned} y' &= a_1(x)\,y+b_1(x)\,z+c_1(x)\,u,\\ z' &= a_2(x)\,y+b_2(x)\,z+c_2(x)\,u,\\ u' &= a_3(x)\,y+b_3(x)\,z+c_3(x)\,u. \end{aligned}\right\} \qquad (1)$$

Напомним, что систему уравнений любого порядка легко преобразовать в систему первого порядка, а разрешение системы относительно производных осуществляется алгебраически, без решения самих дифференциальных уравнений.

Так как от системы (1) легко перейти к равносильному уравнению третьего порядка, которое также получается линейным и однородным, то все свойства линейных однородных уравнений распространяются на систему (1). При этом суммой двух решений $y = y_1$, $z = z_1$, $u = u_1$ и $y = y_2$, $z=z_2$, $u = u_2$ считается решение

$$y=y_1+y_2,\quad z=z_1+z_2,\quad u=u_1+u_2,$$

а произведением решения $y=y_1$ $z=z_1$, $u=u_1$ на число C считается решение $y=Cy_1$, $z=Cz_1$, $u=Cu_1$, т. е. линейные действия над решениями осуществляются так же, как над векторами.

В частности, *общее решение системы* (1) *имеет вид*

$$\left.\begin{aligned} y &= C_1y_1+C_2y_2+C_3y_3,\\ z &= C_1z_1+C_2z_2+C_3z_3,\\ u &= C_1u_1+C_2u_2+C_3u_3, \end{aligned}\right\} \qquad (2)$$

где C_1, C_2, C_3— произвольные постоянные, а (y_1, z_1, u_1), (y_2, z_2, u_2) и (y_3, z_3, u_3) — три *линейно независимых решения* системы (1), т. е. таких, что ни одно из них не является линейной комбинацией остальных.

Остановимся на свойстве 4 линейных однородных уравнений. Если известно ненулевое решение (y_1, z_1, u_1) системы (1), то, сделав подстановку $y = y_1\bar{y}$, $z = z_1\bar{z}$, $u = u_1\bar{u}$, а затем

$$\bar{y} = \bar{u} + v, \quad \bar{z} = \bar{u} + w,$$

нетрудно получить систему *двух* уравнений первого порядка с *двумя* неизвестными функциями, $v(x)$ и $w(x)$, из которой u находится с помощью одного интегрирования.

На *линейные неоднородные* системы вида

$$\left.\begin{aligned} y' &= a_1(x)\,y + b_1(x)\,z + c_1(x)\,u + f_1(x), \\ z' &= a_2(x)\,y + b_2(x)\,z + c_2(x)\,u + f_2(x), \\ u' &= a_3(x)\,y + b_3(x)\,z + c_3(x)\,u + f_3(x) \end{aligned}\right\} \qquad (3)$$

также распространяются все свойства, указанные неоднородных уравнений. В частности, метод вариации произвольных постоянных (свойство 3) выглядит так. Пусть известно общее решение (2) соответствующей однородной системы (1). Тогда решение системы (3) ищется в виде

$$\left.\begin{aligned} y &= \varphi_1(\lambda)\,y_1 + \varphi_2(x)\,y_2 + \varphi_3(x)\,y_3, \\ z &= \varphi_1(x)\,z_1 + \varphi_2(x)\,z_2 + \varphi_3(x)\,z_3, \\ u &= \varphi_1(x)\,u_1 + \varphi_2(x)\,u_2 + \varphi_3(x)\,u_3. \end{aligned}\right\}$$

После подстановки в (3) получается система

$$\left.\begin{aligned} \varphi_1'y_1 + \varphi_2'y_2 + \varphi_3'y_3 &= f_1(x), \\ \varphi_1'z_1 + \varphi_2'z_2 + \varphi_3'z_3 &= f_2(x), \\ \varphi_1'u_1 + \varphi_2'u_2 + \varphi_3'u_3 &= f_3(x), \end{aligned}\right\}$$

из которой алгебраическим способом находим φ_1', φ_2', φ_3', а затем, интегрируя, находим φ_1, φ_2, φ_3.

Особое значение имеют *линейные однородные системы с постоянными коэффициентами*. Рассмотрим, например, систему

$$\left.\begin{aligned} y' &= a_1y + b_1z + c_1u, \\ z' &= a_2y + b_2z + c_2u, \\ u' &= a_3y + b_3z + c_2u, \end{aligned}\right\} \qquad (4)$$

в которой все коэффициенты $a_1, b_1, \ldots, c_3$ постоянны. Ненулевые частные решения ищут в виде

$$y = \lambda e^{px}, \quad z = \mu e^{px}, \quad u = \nu e^{px}, \qquad (5)$$

где λ, μ, ν, p— пока неизвестные постоянные. Подстановка в (4) дают после сокращения на e^{px} и переноса всех членов в одну

сторону:

$$\left.\begin{aligned}(a_1-p)\lambda+b_1\mu+c_1\nu=0,\\ a_2\lambda+(b_2-p)\mu+c_2\nu=0,\\ a_3\lambda+b_3\mu+(c_3-p)\nu=0.\end{aligned}\right\} \qquad (6)$$

Эти равенства можно рассматривать как систему трех алгебраических однородных уравнений первой степени с тремя неизвестными λ, μ, ν. Чтобы она имела ненулевое решение, а только такое решение в силу (5) нас и интересует, необходимо и достаточно, чтобы определитель системы был равен нулю:

$$\begin{vmatrix} a_1-p & b_1 & c_1 \\ a_2 & b_2-p & c_2 \\ a_3 & b_3 & c_3-p \end{vmatrix} = 0. \qquad (7)$$

Это — *характеристическое уравнение* для системы (4), из которого мы находим возможные значения *p*.

Так как уравнение (7) имеет третью степень относительно *p*, то оно имеет три корня, p_1, p_2, p_3. Если все эти корни простые, то можно любой из них подставить в систему (6), найти какое-либо ненулевое решение λ, μ, ν и по формуле (5) получить соответствующее решение $y(x)$, $z(x)$, $u(x)$. Из построенных таким образом трех частных решений (при $p=p_1$, p_2 и p_3) согласно формуле (2) получаем общее решение еистемы (4):

$$\left.\begin{aligned} y=C_1\lambda_1 e^{p_1x}+C_2\lambda_2 e^{p_2x}+C_3\lambda_3 e^{p_3x},\\ z=C_1\mu_1 e^{p_1x}+C_2\mu_2 e^{p_2x}+C_3\mu_3 e^{p_3x},\\ u=C_1\nu_1 e^{p_1x}+C_2\nu_2 e^{p_2x}+C_3\nu_3 e^{p_3x},\end{aligned}\right\} \qquad (8)$$

где C_1, C_2, C_3 — произвольные постоянные.

Если уравнение (7) имеет мнимые корни, то решение можно либо оставить в форме (8), либо записать в вещественной форме.

Случай, когда уравнение (7) имеет кратные корни, более сложный. В конкретных примерах можно поступать так. Если, например, корень p_1 двойной, то частные решения системы (4) надо при $p = p_1$ взамен (5), искать в форме

$$y=(\lambda x+\overline{\lambda})e^{p_1x},\quad z=(\mu x+\overline{\mu})e^{p_1x},\quad u=(\nu x+\overline{\nu})e^{p_1x}, \qquad (9)$$

после подстановки в (4) приравнять коэффициенты при одинаковых степенях *x* и из полученной системы уравнений найти *два* независимых варианта значений λ, $\overline{\lambda},\ldots,\overline{\nu}$ и тем самым два независимых решения вида (9). Если корень характеристического уравнения имеет высшую кратность, то соответственно усложнится и форма частного решения системы (4).

В теории систем линейных дифференциальных уравнений широко применяется матричная запись. Для этого систему (1) обычно переписывают в виде

$$\left.\begin{aligned} y_1' &= a_{11}(x)\,y_1 + a_{12}(x)\,y_2 + a_{13}(x)\,y_3, \\ y_2' &= a_{21}(x)\,y_1 + a_{22}(x)\,y_2 + a_{23}(x)\,y_3, \\ y_3' &= a_{31}(x)\,y_1 + a_{32}(x)\,y_2 + a_{33}(x)\,y_3 \end{aligned}\right\} \qquad (10)$$

и вводят *матрицу коэффициентов* и *векторное решение* (т. е. столбцовую матрицу)

$$\mathbf{A}(x) = \begin{pmatrix} a_{11}(x) & a_{12}(x) & a_{13}(x) \\ a_{21}(x) & a_{22}(x) & a_{23}(x) \\ a_{31}(x) & a_{32}(x) & a_{33}(x) \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$$

Заметим, что если дана некоторая матрица $\mathbf{B}(x)= (b_{ij}(x))$, то из правил линейных действий с матрицами следует, что

$$\frac{\Delta \mathbf{B}}{\Delta x} = \left(\frac{\Delta b_{ij}}{\Delta x}\right)$$

и тем самым $\mathbf{B}'(x) = (b_{ij}'(x))$, т. е. *чтобы продифференцировать матрицу, надо продифференцировать все ее элементы.* При этом легко проверить, что основные правила дифференцирования, такие как формулы для производной суммы или произведения и т. п., остаются в силе. Отсюда

$$\mathbf{y}' = \begin{pmatrix} y_1' \\ y_2' \\ y_3' \end{pmatrix}$$

и потому систему (10) можно записать в *матричной форме*

$$\mathbf{y}' = \mathbf{A}(x)\,\mathbf{y}, \qquad (11)$$

а линейную неоднородную систему (3) — в аналогичной форме

$$\mathbf{y}' = \mathbf{A}(x)\,\mathbf{y} + \mathbf{f}(x)$$

Решение (2) можно переписатьв векторном виде

$$\mathbf{y} = C_1\mathbf{y}^1 + C_2\mathbf{y}^2 + C_3\mathbf{y}^3,$$

где индексы сверху означают номера линейно независимых частных векторных решений уравнения (11) Система (4) приобретает вид

$$\mathbf{y}' = \mathbf{A}\mathbf{y}, \qquad (12)$$

а решение (5) — вид

$$\mathbf{y} = e^{px}\boldsymbol{\alpha}, \qquad (13)$$

где

$$\boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}$$

— некоторый постоянный вектор. Подставляя (13) в (12), получаем

$$pe^{px}\alpha = Ae^{px}\alpha, \quad \text{т. е.} \quad A\alpha = p\alpha$$

Мы видим, что α и p должны быть собственным вектором и соответствующим собственным значением матрицы **A**; последнее находится из уравнения

$$\det(A - pI) = 0,$$

которое есть не что иное, как уравнение (7)

При решении систем вида (4) с постоянными коэффициентами, а также соответствующих неоднородных систем можно поименять операторный метод. Для этого записываем $y' = Dy$, $z' = Dz$, $u' = Du$, затем решаем полученную систему уравнений как алгебраическую относительно *y, z, u*, однако решение до формул вида $P(D)\,y = f(x)$, после чего применяем методы операторного решения уравнений. По существу, это сводится к указанию правила перехода от системы уравнений первого порядка к одному уравнению высшего порядка.

10.9. Фазовое пространство

При изучении закона движения материальной точки с массой m удобно пользоваться векторной формой записи уравнений. Итак, пусть $\boldsymbol{r} = \boldsymbol{r}(t)$ — закон движения материальной точки в пространстве R^3, где t — время. Это значит, что в момент времени t точка имеет радиус-вектор $\boldsymbol{r}(t)$, или, что все равно, координаты $\{x(t), y(t), z(t)\}$.

Если точка массы m движется под действием заданной силы (вектора) $\boldsymbol{F}(t, \boldsymbol{r}, \dot{\boldsymbol{r}})$, то по закону Ньютона и механическому смыслу второй производной функция $\boldsymbol{r}(t)$ должна удовлетворять уравнению движения

$$m\ddot{\boldsymbol{r}} = \boldsymbol{F}(t, \boldsymbol{r}, \dot{\boldsymbol{r}}). \tag{1}$$

Векторное уравнение (1) эквивалентно системе трех скалярных уравнений

$$\left.\begin{aligned} m\frac{d^2x}{dt^2} &= X(t, x, y, z, \dot{x}, y, \dot{z}), \\ m\frac{d^2y}{dt^2} &= Y(t, x, y, z, \dot{x}, \dot{y}, \dot{z}), \\ m\frac{d^2z}{dt^2} &= Z(t, x, y, z, \dot{x}, \dot{y}, \dot{z}), \end{aligned}\right\} \tag{2}$$

где *X, Y, Z*—проекции вектора *F* на оси координат *x, y, z.*

Если считать неизвестными не только координаты точки *x, y, z*, но и проекции скорости

$$\frac{d\boldsymbol{r}}{dt} = \{\dot{x},\ \dot{y},\ \dot{z}\},$$

то мы получим систему из шести уравнений первого порядка

$$\left.\begin{aligned} \dot{x} &= u, \quad m\frac{du}{dt} = X(t, x, y, z, u, v, w), \\ \dot{y} &= v, \quad m\frac{dv}{dt} = Y(t, x, y, z, u, v, w), \\ \dot{z} &= w, \quad m\frac{dw}{dt} = Z(t, x, y, z, u, v, w). \end{aligned}\right\} \tag{3}$$

Векторное уравнение (1) можно также записать в виде системы двух векторных уравнений, если скорость $\boldsymbol{V} = \frac{d\boldsymbol{r}}{dt}$ считать неизвестной векторной функцией:

$$\frac{d\boldsymbol{r}}{dt} = \boldsymbol{V}, \quad m\frac{d\boldsymbol{V}}{dt} = \boldsymbol{F}(t, \boldsymbol{r}, \boldsymbol{V}), \tag{1′}$$

где **V** — вектор с проекциями *u, v, w.*

Если ввести в рассмотрение вектор

$$\boldsymbol{R}(t) = \{x(t),\ y(t),\ z(t),\ \dot{x}(t),\ \dot{y}(t),\ \dot{z}(t)\},$$

то уравнение (1) или система (3) квивалентны одному векторному уравнению первого порядка

$$\frac{d\boldsymbol{R}}{dt} = \Phi(t, x, y, z, u, v, w) \tag{4}$$

в шестимерном пространстве, причем вектор

$$\Phi = \left\{u,\ v,\ w,\ \frac{1}{m}X,\ \frac{1}{m}Y,\ \frac{1}{m}Z\right\}.$$

Шестимерное пространство точек

$$(x,\ y,\ z,\ \dot{x},\ \dot{y},\ \dot{z}) \equiv (r_x,\ r_y,\ r_z,\ V_x,\ V_y,\ V_z)$$

в физике называют *фазовым,* а кривую ***R(t)*** в шестимерном пространстве, являющуюся решением (4), называют *фазовой траекторией.*

<u>Фазовое пространство</u> — это пространство состояний движения точки по кривой.

Первые три координаты ***R(t)*** характеризуют положение точки в трехмерном пространстве ***(r(t)),*** а остальные три координаты ***R(t)*** характеризуют ее скорость $\dot{r}$ ***(t).***

Приведенная терминология дает так называемую *кинематическую интерпретацию системы уравнений.*

Систему (3), или, что то же самое, (4) называют *динамической* системой.

Для выделения одной троектории необходимо задать начальные условия: $\boldsymbol{R}(t_0) = \boldsymbol{R}_0 = (x_0,\ y_0,\ z_0,\ \dot{x}_0,\ \dot{y}_0,\ \dot{z}_0)$, т. е. начальное положение точки и ее начальную скорость. Другими слонами, интегральная кривая $\boldsymbol{R}(t)$ должна проходить через точку $\boldsymbol{R}_0$ шестимерного пространства.

Таким образом, физические задачи приводят нас к необходимости рассмотрения систем дифференциальных уравнений.

Рассмотрим произвольную систему дифференциальных уравнений первого порядка вида

$$\frac{dy_k}{dt} = f_k(t,\ y_1,\ \ldots,\ y_n) \qquad (k = 1,\ \ldots,\ n), \tag{5}$$

где $y_k(t)$ — искомые функции, а $f_k(t, y_1, \ldots, y_n)$ — известные функции, заданные на некотором множестве точек $(t, y_1, \ldots, y_n)$ $(n+1)$-мерного пространства.

Нас будут интересовать решения $y_1(t), \ldots, y_{tt}(t)$ системы (5), удовлетворяющие начальным условиям

$$y_1(t_0) = y_{10},\ \ldots,\ y_n(t_0) = y_{n0}, \tag{6}$$

где $(y_{10}, \ldots, y_{n0})$ - заданная точка n-мерного проетранства.

Систему (5) (решенную относительно производных искомых функций!) называют *нормальной*.

Если функции f_k не зависят явно от независимого переменного t, то система (5) называется *автономной*

$$\dot{y}_k = f_k(y_1,\ \ldots,\ y_n) \quad (k = 1,\ \ldots,\ n).$$

Если ввести векторы в n-мерном пространстве

$$\boldsymbol{y} = \{y_1(t),\ \ldots,\ y_n(t)\},\quad \boldsymbol{y}_0 = \{y_{10},\ \ldots,\ y_{n0}\},$$
$$\boldsymbol{F}(t, \boldsymbol{y}) = \{f_1(t, y_1, \ldots, y_n),\ \ldots,\ f_n(t, y_1, \ldots, y_n)\},$$

то систему (5) можно записать в виде

$$\dot{\boldsymbol{y}} = \boldsymbol{F}(t, \boldsymbol{y}), \tag{5'}$$

а начальные условия (6) — в форме

$$\boldsymbol{y}(t_0) = \boldsymbol{y}_0. \tag{6'}$$

Автономную систему можно записать так:

$$\dot{\boldsymbol{y}} = \boldsymbol{F}(\boldsymbol{y}),\quad \boldsymbol{F} = \{f_1(\boldsymbol{y}),\ \ldots,\ f_n(\boldsymbol{y})\}. \tag{7'}$$

Автономную систему можно интерпретировать следующим образом. В каждой точке $(y_1, \ldots, y_n)$ некоторого множества n-мерного пространства определен вектор

$$\boldsymbol{F}(\boldsymbol{y}) = \{f_1(y_1,\ \ldots,\ y_n),\ \ldots,\ f_n(y_1,\ \ldots,\ y_n)\}.$$

Этим определено на указанном множестве *поле векторов.*

Решение $y(t)$ описывает определенную траекторию движения точки в n-мерном пространстве, причем вектор скорости $\dot{\boldsymbol{y}}(t)$ в момент ее прохождения через $(y_1, \ldots, y_n)$ совпадает с вектором $\boldsymbol{F}(y)$.

Пространство размерности n точек $(y_1, \ldots, y_n)$, в котором интерпретируются решения автономной системы $(7')$ в виде траекторий, называется *фазовым пространством системы.*

Траектории $\boldsymbol{y}(t)$ называются *фазовыми траекториями,* векторы $\boldsymbol{F}(y)$*—фазовыми скоростями.*

11. Операционное исчисление

11.1. Изображение Лапласа

В этой разделе мы, как правило, будем рассматривать функции $f(t)$ действительного переменного t, заданные на $[0, \infty)$. Иногда будем считать, что $f(t)$ определена на $(-\infty,\infty)$, но при $t<0$ функция $f(t) \equiv 0$. Кроме того, будем предполагать, что функция $f(t)$ кусочно-непрерывна и на каждом конечном промежутке имеет конечное число точек разрыва первого рода. Пусть $p = a + ib$ — комплексное число.
Рассмотрим функцию

$$F(p) = \int_0^\infty e^{-pt} f(t)\,dt. \tag{1}$$

Если

$$|f(t)| \leqslant M \exp(s_0 t), \tag{2}$$

где $a > s_0$, то функция $F(p)$ аналитическая в полуплоскости $\mathrm{Re}\, p > s_0$ (рис. 1).

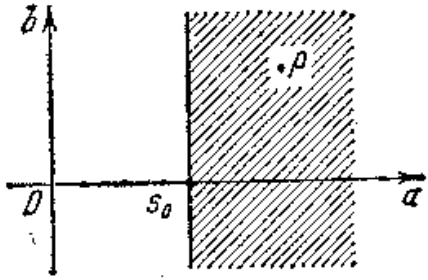


Рис. 1

В самом деле,

$$|F'(p)| = \left| \int_0^\infty t \exp(-at - ibt) f(t)\, dt \right| \leqslant \int_0^\infty t \exp(-at) |f(t)|\, dt \leqslant$$

$$\leqslant \int_0^\infty \exp(-at) \cdot tM \exp(s_0 t)\, dt =$$

$$= M \int_0^\infty t \exp(-(a - s_0) t)\, dt < \infty, \tag{3}$$

так как $a > s_0$. Законность дифференцирования по p под знаком интеграла следует из неравенства (3) и того факта, что функция $tf(t) \exp(-pt)$ кусочно-непрерывна.

Функция *F(p)* называется *изображением Лапласа* функции *f*, *L-изображением* или *преобразованием Лапласа.*

Мы будем употреблять обозначения

$$F(p) = L(f(t); p), \quad f(t) \doteqdot F(p), \quad F(p) \doteqdot f(t).$$

Функцию *f(t)* в этом случае называют *начальной функцией* или *оригиналом.* Число s_0 $(s_0 = s_0(f))$ называется *показателем роста функции f(t)* (ниже, если особо не оговорено, то мы считаем, что показатель роста *f* равен s_0).

Процесс нахождения изображения для заданного оригинала и обратно, нахождение оригинала по известному изображению называется ***операционным исчислением***, начало которому положил Хевисайд. Разработав операционное исчисление, Хевисайд не дал ему обоснования. Отметим, что он рассматривал преобразование

$$\tilde{F}(p) = p \int_0^\infty \exp(-pt) f(t)\, dt,$$

т. е. $\tilde{F}(p) = pF(p)$.

В одних вопросах удобным является преобразование Лапласа, в других — преобразование Хевисайда. Мы будем рассматривать преобразование Лапласа.

Обоснование операционного исчисления было дано в двадцатых годах прошлого века в работах ряда математиков.

Теорема 1 (единственности). *Если две непрерывные функции f(t) и g(t) имеют одно и то же L-изображение F(p), то они тождественно равны.*

На основании теоремы 1 мы можем сказать, что для непрерывной функции *f(t),* тождественно не равной нулю, изображение не может быть периодической функцией.

В самом деле, если $\forall p \quad F(p) = F(p+\omega)$, где $\omega \neq 0$, то

$$\int_0^\infty \exp(-pt) f(t)\, dt = \int_0^\infty \exp(-(p+\omega)t) f(t)\, dt.$$

По теореме1

$$f(t) = \exp(-\omega t) f(t),$$

т. е. $\exp(-\omega t) \equiv 1 \quad (\omega \neq 0)$, чего быть не может.

11.2. Изображение простейших функций и свойства изображений

Единичной функцией или *функцией Хевисайда* называется функция

$$\sigma_0(t) = \begin{cases} 1, & t \geqslant 0, \\ 0, & t < 0. \end{cases}$$

Очевидно, что показатель роста этой функции s_0=0. Найдем L - изображение этой функции в области $\operatorname{Re} p > 0$:

$$L(\sigma_0(t);\ p) = \int_0^\infty \exp(-pt)\, dt = -\frac{1}{p}\exp(-pt)\Big|_0^\infty = \frac{1}{p}.$$

Таким образом,

$$\sigma_0(t) \doteq 1/p. \tag{1}$$

Аналогично для функции $f(t)$=cos t интегрированием по частям находим

$$L(\cos t;\ p) = \int_0^\infty \exp(-pt)\cos t\, dt =$$

$$= \exp(-pt)\sin t\Big|_0^\infty + p\int_0^\infty \exp(-pt)\sin t\, dt =$$

$$= p\int_0^\infty \exp(-pt)\sin t\, dt = p\left[-\exp(-pt)\cos t\Big|_0^\infty - \right.$$

$$\left. - p\int_0^\infty \exp(-pt)\cos t\, dt\right] = p - p^2 L[\cos t;\ p].$$

Отсюда

$$L(\cos t;\ p) = \frac{p}{1+p^2} \quad (\operatorname{Re} p > 0), \text{ т. е.}$$

$$\cos t \doteq p/(1+p^2). \tag{2}$$

Попутно мы доказали, что

$$L(\cos t;\ p) = pL(\sin t;\ p),$$

откуда

$$\sin t \doteq 1/(1+p^2). \tag{3}$$

Теорема1 (подобия)

$$\doteq \frac{1}{\alpha} F\left(\frac{p}{\alpha}\right) \quad (\alpha > 0,\ \operatorname{Re} p > \max\{s_0,\ \alpha s_0\}).$$

В самом леле,

$$L[f(\alpha t);\ p] = \int_0^\infty \exp(-pt) f(\alpha t)\, dt = \left\{\alpha t = u,\ dt = \frac{du}{\alpha}\right\} =$$

$$= \frac{1}{\alpha}\int_0^\infty \exp\left(-\frac{p}{\alpha}u\right) f(u)\, du = \frac{1}{\alpha} L\left[f(u);\ \frac{p}{\alpha}\right].$$

На основании теоремы 1 получаем

$$\cos \alpha t \doteq \frac{1}{\alpha}\frac{p/\alpha}{1+(p/\alpha)^2} = \frac{p}{p^2+\alpha^2}; \tag{4}$$

$$\sin \alpha t \doteq \frac{1}{\alpha}\frac{1}{1+(p/\alpha)^2} = \frac{\alpha}{p^2+\alpha^2}. \tag{5}$$

Теорема 2 (свойство линейности). *Имеет место равенство*

$$L[Af(t) + Bg(t);\ p] = AL[f(t);\ p] + BL[g(t);\ p],$$

где A, B—постоянные числа.

Это свойство вытекает из соответствующего свойства несобственного интеграла. Отметим, что если показатели роста функций f и g соответственно равны s_0 и $\overline{s_0}$, то изображение $Af+Bg$ существует в полуплоскости

$$\operatorname{Re} p > \max\{s_0,\ \overline{s_0}\}.$$

Пример1. Найти изображение функции

$$f(t) = 3\sigma_0(t) + 2\cos 3t.$$

В силу (1), (4) и теоремы 2 имеем

$$L[f(t);\ p] = 3L[\sigma_0(t);\ p] + 2L[\cos 3t;\ p] = \frac{3}{p} + 2\frac{p}{p^2+9}.$$

Пример 2. Найти оригинал изображения

$$F(p) = \frac{2}{p} + \frac{2}{p^2+16}.$$

Представим изображение *F(p)* в виде

$$F(p) = 2\cdot\frac{1}{p} + \frac{1}{2}\frac{4}{p^2+4^2}.$$

Имеем

$$1/p \doteq \sigma_0(t), \quad 4/(p^2+4^2) \doteq \sin 4t.$$

Следовательно, оригинал (по теореме 1 п. 11.1)

$$f(t) = 2\sigma_0(t) + \frac{1}{2}\sin 4t.$$

Теорема 3 (смещение изображения).

$$L[f(t)\exp(-\alpha t);\ p] = L[f(t);\ p+\alpha], \quad \operatorname{Re}(p+\alpha) > s_0.$$

Пример 3. Найти изображение функций

$$e^{-\alpha t}\cos\beta t,$$
$$e^{-\alpha t}\sin\beta t,\ e^{-\alpha t}.$$

Так как $\cos\beta t \doteq p/(p^2+\beta^2)$, то по теореме 3

$$L[\exp(-\alpha t)\cos\beta t;\ p] = \frac{p+\alpha}{(p+\alpha)^2+\beta^2}. \tag{6}$$

Совершенно аналогично, используя формулы (5) и (1), имеем

$$L[\exp(-\alpha t)\sin\beta t;\ p] = \beta/((p+\alpha)^2+\beta^2), \tag{7}$$
$$L[\exp(-\alpha t);\ p] = L[\sigma_0(t);\ p+\alpha] = 1/(p+\alpha). \tag{8}$$

Пример 4. Найти $L[\operatorname{ch}\alpha t;\ p]$, $L[\operatorname{sh}\alpha t;\ p]$.

Используя теорему 2 и равенство (8), имеем

$$L[\operatorname{ch}\alpha t;\ p] = \frac{1}{2}L[e^{\alpha t};\ p] + \frac{1}{2}L[e^{-\alpha t};\ p] = \frac{p}{p^2-\alpha^2}.$$

Аналогично $L[\operatorname{sh}\alpha t;\ p] = \alpha/(p^2-\alpha^2)$.

Пример 5. Найти оригинал для изображения

$$F(p) = 1/(p^2+2p+5)$$

Имеем

$$F(p) = \frac{2}{2[(p+1)^2+2^2]}, \quad \frac{2}{p^2+2^2} \doteq \sin 2t,$$
$$\frac{2}{(p+1)^2+2^2} \doteq e^{-t}\sin 2t, \quad f(t) = \frac{1}{2}e^{-t}\sin 2t.$$

Теорема 4 (дифференцирование изображения).

$$(-1)^n \frac{d^n}{dp^n} L[f(t);\ p] = L[t^n f(t);\ p].$$

Доказательство. Если Re $p > s_0$, где s_0 — показатель роста функции $f(t)$, то интеграл

$$\int_0^\infty t^n f(t)\exp(-pt)\,dt$$

существует при любом $n = 1, 2, \ldots$ Далее, очевидно, что

$$\frac{d^n}{dp^n}\int_0^\infty \exp(-pt)\,f(t)\,dt = \int_0^\infty (-t)^n \exp(-pt)\,f(t)\,dt.$$

Отсюда

$$(-1)^n L[f(t)\,t^n;\ p] = \frac{d^n}{dp^n} L[f(t);\ p].$$

Пример 6. Так как

$$1/p \doteqdot \sigma_0(t),$$

то в силу теоремы 4 получаем

$$(-1)\frac{d}{dp}\left(\frac{1}{p}\right) \doteqdot 1\cdot t, \quad \text{т. е. } t \doteqdot \frac{1}{p^2}.$$

Продолжая дифференцирование, получим

$$t^n \doteqdot n!/p^{n+1} \qquad (n = 1, 2, \ldots). \tag{9}$$

Если n не целое, то

$$t^n \doteqdot \frac{\Gamma(n+1)}{p^{n+1}},$$

где $\Gamma(a+1) = \int_0^\infty e^{-t} t^a\,dt = L[t^a;\ 1]$.

При натуральном n имеем $\Gamma(n+1) = n!$.

Пример 7. Найти изображение функции $t\cos\alpha t$. Имеем

$$\frac{p}{p^2+\alpha^2} \doteqdot \cos\alpha t, \quad -\left(\frac{p}{p^2+\alpha^2}\right)' \doteqdot t\cos\alpha t$$

или

$$\frac{p^2-\alpha^2}{(p^2+\alpha^2)^2} \doteqdot t\cos\alpha t. \tag{10}$$

Теорема 5 (о дифференцировании оригинала). *Справедлива формула*

$$L[f'(t);\ p] = pL[f(t);\ p] - f(0) \qquad (\mathrm{Re}\,p > s_0) \tag{11}$$

в предположении, что функция f(t) непрерывна, имеет кусочно-непрерывную производную f (t) на [0, ∞) *с разрывами первого рода и показатели роста f(t) и f '(t) равны* s_0.

(Бывают случаи, когда функция *f(t)*, о которой говорится в теореме, задана на интервале (0, оо) и существует предел справа

$$f(0+0) = \lim_{\substack{t\to 0\\ t>0}} f(t).$$

Тогда в формуле (11) надо заменить *f*(0) на *f*(0+0))

Доказательство. Имеем ($\mathrm{Re}\,p > s_0$)

$$L[f'(t);\ p] = \lim_{N\to\infty}\int_0^N \exp(-pt)\,f'(t)\,dt =$$

$$= \lim_{N\to\infty}\left[e^{-pt}f(t)\Big|_0^N + p\int_0^N e^{-pt}f(t)\,dt\right] = -f(0) + pL[f(t);\ p],$$

потому что

$$|e^{-pN}f(N)| \leqslant e^{-aN}Me^{s_0 N} = Me^{-(a-s_0)N} \xrightarrow[N\to\infty]{} 0$$

Следствие 1. *Справедлива формула* ($\mathrm{Re}\ p > s_0$)

$$L[f^{(n)}(t);\ p] =$$
$$= p^n L[f(t);\ p] - p^{n-1}f(0) - \ldots - pf^{(n-2)}(0) - f^{(n-1)}(0) \quad (12)$$

при условии, что $f(t), \ldots, f^{(n-1)}(t)$ *непрерывны,* $f^{(n)}$ *кусочно-непрерывна на* $[0, \infty)$, *а показатель роста функции* f *вместе с ее производными до порядка* n *включительно равен* s_0.

В частности, при

$$f(0) = f'(0) = \ldots = f^{(n-1)}(0) = 0 \quad (13)$$

имеет место

$$L[f^{(n)}(t);\ p] = p^n L[f(t);\ p]. \quad (14)$$

Пример 8. Найти изображение функции $f(t) = \cos^2 t$

Пусть $F(p) \doteq \cos^2 t = f(t)$. Тогда $f'(t) \doteq pF(p) - f(0)$.

Но $f(0) = \cos^2 0 = 1$, $f'(t) = -\sin 2t \doteq -2/(p^2+4)$

Следовательно, $pF(p) - 1 = -2/(p^2+4)$, откуда

$$F(p) = \frac{1}{p} \times \left[1 - \frac{2}{p^2+4}\right] = \frac{p^2+2}{p(p^2+4)}.$$

Этот же результат мы получим, если воспользуемся равенством

$$\cos^2 t = \frac{1}{2} + \frac{\cos 2t}{2}, \quad L[\cos^2 t;\ p] = \frac{1}{2p} + \frac{1}{2}\frac{p}{p^2+4} = \frac{p^2+2}{p(p^2+4)}.$$

Теорема 6 (интегрирование оригинала).

$$\int_0^t f(\tau)\,d\tau \doteq \frac{F(p)}{p}.$$

В самом деле, изменяя порядок интегрирования, имеем

$$\int_0^\infty e^{-pt}\int_0^t f(\tau)\,d\tau\,dt = \int_0^\infty\int_\tau^\infty e^{-pt}f(\tau)\,d\tau\,dt = \int_0^\infty f(\tau)\frac{-e^{-pt}}{p}\Big|_\tau^\infty d\tau =$$

$$= \frac{1}{p}\int_0^\infty e^{-p\tau}f(\tau)\,d\tau = \frac{1}{p}L[f(\tau);\ p] = \frac{F(p)}{p}.$$

По определению полагаем $(p = \alpha + i\beta)$

$$\int_p^\infty F(q)\,dq = \int_\alpha^\infty F(\xi + i\beta)\,d\xi = \lim_{N\to\infty} \int_\alpha^N F(\xi + i\beta)\,d\xi.$$

Теорема 7 (интегрирование изображения). *Пусть заданная на* $(0, \infty)$ *кусочно-непрерывная функция* $f(t)$ *удовлетворяет условию* (2) п.11.1, $f(t) \doteq F(p)$, $p = \alpha + i\beta$,

$\alpha > s_0$ *и* $\int_0^1 \frac{f(t)}{t}\,dt$ *сходится. Тогда, если* $\int_p^\infty F(q)\,dq$ *сходится, то*

$$\frac{f(t)}{t} \doteq \int_p^\infty F(q)\,dq.$$

Доказательство. Изменяя порядок интегрирования, получаем

$$\int_p^\infty F(q)\,dq = \lim_{N\to\infty} \int_\alpha^N \left(\int_0^\infty f(t)\, e^{-(\xi + i\beta)t}\,dt \right) d\xi =$$

$$= \lim_{N\to\infty} \int_0^\infty f(t) \left(\int_\alpha^N e^{-(\xi+i\beta)t}\,d\xi \right) dt = \lim_{N\to\infty} \int_0^\infty f(t)\, \frac{-e^{-(\xi+i\beta)t}}{t} \Big|_{\xi=\alpha}^{\xi=N} dt =$$

$$= \int_0^\infty \frac{f(t)}{t}\, e^{-(\alpha+i\beta)t}\,dt = L\left[\frac{f(t)}{t};\ p\right],$$

потому что

$$I_N = \left| \int_0^\infty \frac{f(t)}{t}\, e^{-(N+i\beta)t}\,dt \right| \longrightarrow 0, \qquad N \longrightarrow \infty.$$

В самом деле, пусть s_0 — показатель роста функции f и N — достаточно большое число $(N > s_0)$, тогда в силу (2) из п.11.1

$$I_N \leqslant \left| \int_0^\eta \frac{f(t)}{t} e^{-(N+i\beta)t}\,dt \right| + \frac{M}{\eta} \int_\eta^\infty e^{-(N-s_0)t}\,dt \leqslant$$

$$\leqslant \left| \int_0^\eta \frac{f(t)}{t} e^{-(N+i\beta)t}\,dt \right| + \frac{M}{\eta\,(N - s_0)},$$

где число $\eta > 0$, которое мы определим ниже, пока произвольно. Введем в рассмотрение функцию

$$\varphi(t) = \int_0^t \frac{f(\xi)}{\xi}\,d\xi.$$

В силу условий теоремы функция $\varphi(t) \to 0$ при $t \to 0$; $\varphi(t)$ непрерывна на (0, 1]; $\varphi(t)$ дифференцируема на (0, 1) за исключением конечного числа точек; $|\varphi(t)| \leqslant K \quad \forall 0 < t \leqslant 1$.

Интегрируя по частям, имеем

$$\Lambda_N = \int_0^\eta \frac{f(t)}{t} e^{-(N+i\beta)t}\, dt =$$

$$= \varphi(t) e^{-(N+i\beta)t} \Big|_0^\eta + \int_0^\eta (N+i\beta)\varphi(t) e^{-(N+i\beta)t}\, dt =$$

$$= \varphi(\eta) e^{-(N+i\beta)\eta} + (N+i\beta) \int_0^\eta \varphi(t) e^{-(N+i\beta)t}\, dt.$$

Зададим теперь произвольное число $\varepsilon > 0$ и подберем числа $\eta > 0$ так, чтобы $|\varphi(t)| < \varepsilon$ при $0 < t \leq \eta$). Тогда при $N > \beta$

$$|\Lambda_N| \leqslant \varepsilon + \varepsilon \sqrt{N^2+\beta^2} \int_0^\eta e^{-Nt}\, dt \leqslant \varepsilon \left(1 + \sqrt{1+\frac{\beta^2}{N^2}}\right) \leqslant$$

$$\leqslant (1+\sqrt{2})\varepsilon.$$

Отсюда, для интеграла f_N, мы получаем следующую оценку:

$$I_N \leqslant (1+\sqrt{2})\varepsilon + \frac{M}{\eta(N-s_0)}$$

и в силу произвольности $\varepsilon > 0$ заключаем, что

$$\lim_{N \to \infty} I_N = 0.$$

Теорема доказана.

Замечание 1. Мы применяем здесь и ниже изменение порядка интегрирования. Согласно теореме Фубини, которую мы здесь не доказываем, эта операция законна, если полученный после изменения кратный интеграл абсолютно сходится.

Следствие 2. *Пусть функция f удовлетворяет условиям теоремы 7 при* $s_0 = 0$ *(т. е.* $|f(t)| \leq M$ *для всех* $t \geq 0$*) и пусть дополнительно несобственный интеграл* $\int_1^\infty \frac{f(t)}{t}\, dt$ *сходится. Тогда, если интеграл* $\int_0^\infty F(q)\, dq$ *сходится, то*

$$\int_0^\infty \frac{f(t)}{t}\, dt = \int_0^\infty F(q)\, dq.$$

Пример 9. Найти изображение функции $\int_0^t \sin 2\tau\, d\tau$.

Имеем $\sin 2\tau \doteqdot 2/(p^2+4)$. По теореме 6

$$\int_0^t \sin 2\tau\, d\tau \doteqdot \frac{2}{p(p^2+4)}.$$

Пример 10. Найти изображение функции $\frac{\sin t}{t}$.

Нам известно, что

$$\sin t \doteqdot 1/(p^2+1).$$

Поэтому по теореме 7

$$\frac{\sin t}{t} \doteqdot \int_p^\infty \frac{dq}{q^2+1} = \operatorname{arctg} q \Big|_p^\infty = \frac{\pi}{2} - \operatorname{arctg} p.$$

Пример 11. Найти интеграл $\int_0^\infty \frac{\sin t}{t}\, dt$.

Используя пример 10 и следствие 2, получаем

$$\int_0^\infty \frac{\sin t}{t}\, dt = \int_0^\infty \frac{dq}{q^2+1} = \operatorname{arctg} q \Big|_0^\infty = \frac{\pi}{2}.$$

Теорема 8 (запаздывание оригинала). *Пусть* $f(t) = 0$ *при* $t < 0$, *тогда*

$$L[f(t-t_0);\ p] = e^{-pt_0} L[f(t);\ p],$$

где t_0*—некоторая точка* ($t_0 \geq 0$).

Доказательство. Имеем

$$L[f(t-t_0);\ p] = \int_0^{t_0} e^{-pt} f(t-t_0)\, dt + \int_{t_0}^\infty e^{-pt} f(t-t_0)\, dt =$$

$$= \int_{t_0}^\infty e^{-pt} f(t-t_0)\, dt = \{t - t_0 = u,\ dt = du\} =$$

$$= \int_0^\infty e^{-p(u+t_0)} f(u)\, du = e^{-pt_0} L[f(u);\ p].$$

Пример 12. Так как $L[\sigma_0(t);\ p] = \frac{1}{p}$, то (рис. 2)

$$L[\sigma_0(t-h);\ p] = e^{-ph}\frac{1}{p}.$$

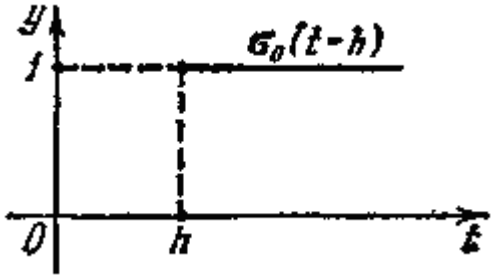


Рис. 2

Пример 13. Пусть (рис. 3)

$$f(t)=\sigma_0(t-h)-\sigma_0(t-h_1) \qquad (h<h_1).$$

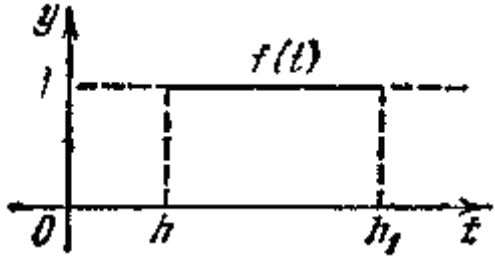


Рис. 3

По теореме 2 и теореме 8 имеем

$$L[f(t);\ p]=L[\sigma_0(t-h);\ p]-L[\sigma_0(t-h_1);\ p]=$$
$$=\frac{e^{-ph}-e^{-ph_1}}{p}.$$

Пример 14. Найти изображение функции $f(t)$ (рис. 4), определенной на отрезке $[0, 2a]$ равенствами

$$f(t)=\begin{cases} A, & 0\leqslant t\leqslant a,\\ 0, & a<t<2a \end{cases}$$

и продолженной затем на весь луч $t > 0$ с периодом $2a$.

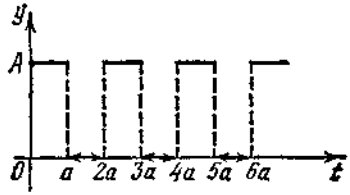


Рис. 4

Имеем

$$L[f(t);\ p]=\int_0^{\infty} e^{-pt}f(t)\,dt=\sum_{k=0}^{\infty}\int_{ka}^{(k+1)a} e^{-pt}f(t)\,dt=$$

$$=\sum_{k=0}^{\infty}\int_{2ka}^{(2k+1)a} e^{-pt}A\,dt=\sum_{k=0}^{\infty}\frac{A}{p}\left[e^{-p2ka}-e^{-p(2k+1)a}\right]=$$

$$=\frac{A(1-e^{-pa})}{p}\sum_{k=0}^{\infty}e^{-2kpa}=\frac{A(1-e^{-pa})}{p(1-e^{-2pa})}=\frac{A}{p(1+e^{-pa})}.$$

Выражение

$$\int_0^t f_1(\tau)\,f_2(t-\tau)\,d\tau$$

называется *сверткой функций* $f_1(t)$ и $f_2(t)$ и обозначается символом $f_1 * f_2$.

Легко проверить, что

$$\int_0^t f_1(\tau)\,f_2(t-\tau)\,d\tau=\int_0^t f_1(t-\tau)\,f_2(\tau)\,d\tau$$

(надо сделать замену переменной $t-\tau=u$).

Теорема 9. *Преобразование Лапласа от свертки равно произведению преобразований Лапласа от функций*

$$f_1(t)\ \text{и}\ f_2(t)\ \ (s_0(f_1)=s_0(f_2));$$

$$L\left[\int_0^t f_1(\tau)\,f_2(t-\tau)\,d\tau;\ p\right]=$$

$$=L[f_1(t);\ p]\cdot L[f_2(t);\ p].$$

Доказательство. Напомним, что мы считаем, что $f_1(t)=f_2(t)=0$ при $t<0$.

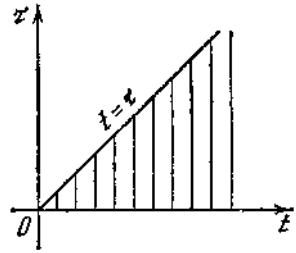


Рис. 5

Изменяя порядок интегрирования, (рис. 5) и учитывая, что $F_2(t-\tau) = 0$ для $0<t<\tau$, имеем

$$L[f_1 * f_2;\ p] = \int_0^\infty e^{-pt} \int_0^t f_1(\tau)\, f_2(t-\tau)\, d\tau\, dt =$$

$$= \int_0^\infty \int_\tau^\infty e^{-pt} f_1(\tau)\, f_2(t-\tau)\, dt\, d\tau =$$

$$= \int_0^\infty f_1(\tau) \left(\int_\tau^\infty e^{-pt} f_2(t-\tau)\, dt \right) d\tau = \{t-\tau = z,\ dt = dz\} =$$

$$= \int_0^\infty f_1(\tau) \left(\int_0^\infty e^{-p(z+\tau)} f_2(z)\, dz \right) d\tau =$$

$$= \int_0^\infty f_1(\tau)\, e^{-p\tau}\, d\tau \int_0^\infty f_2(z)\, e^{-pz}\, dz = L[f_1(\tau);\ p] \cdot L[f_2(z);\ p].$$

Отметим, что двойной интеграл по бесконечному сектору $\{0 < \tau \leqslant t,\ 0 < t < \infty\}$ от функции

$$e^{-pt} f_1(\tau)\, f_2(t-\tau)$$

абсолютно сходится при $\mathrm{Re}\, p > s_0$.

Пример 15.

$$L\left[\int_0^t e^{t-\tau} \operatorname{ch} \alpha\tau\, d\tau;\ p \right] = L[e^t;\ p]\, L[\operatorname{ch} \alpha t;\ p] = \frac{1}{p-1} \cdot \frac{p}{p^2-\alpha^2}.$$

Следствие 3. *Пусть* $F(p) \doteqdot f(t)$, $G(p) \doteqdot g(t)$, *тогда имеет место формула Дюамеля*

$$pF(p)\, G(p) \doteqdot f(t)\, g(0) + \int_0^t f(\tau)\, g'(t-\tau)\, d\tau. \qquad (15)$$

Доказательство. Имеем

$$F(p)\,G(p) \doteqdot \int_0^t f(\tau)\,g(t-\tau)\,d\tau.$$

Отсюда, по теореме 5 о дифференцировании оригинала, получаем

$$pF(p)\,G(p) \doteqdot \left(\int_0^t f(\tau)\,g(t-\tau)\,d\tau\right)' = f(t)\,g(0) + \int_0^t f(\tau)\,g'(t-\tau)\,d\tau.$$

Теорема 10. *Если* $\mathcal{F}(f(x);\ y)$ *и* $L[f(x);\ p]$ *— соответственно преобразования Фурье и Лапласа функции* f ($s_0(f)<0$), *то*

$$2\pi\mathcal{F}(f(x);\ y) = L[f(x);\ iy] + L[f(-x);\ -iy]. \quad (16)$$

В самом деле,

$$2\pi\mathcal{F}(f(x);\ y) = \int_{-\infty}^{\infty} e^{-ixy} f(x)\,dx = \int_0^{\infty} e^{-ixy} f(x)\,dx + \int_{-\infty}^{0} e^{-ixy} f(x)\,dx =$$

$$= L[f(x);\ iy] + \int_0^{\infty} e^{ixy} f(-x)\,dx =$$

$$= L[f(x);\ iy] + L[f(-x);\ -iy].$$

По формуле (16) легко найти изображение Фурье, если известно преобразование Лапласа функции f.

Пример 16. Пусть

$$f(x) = \begin{cases} e^{-\alpha x}\cos\beta x, & x \geqslant 0, \quad \alpha > 0, \\ 0, & x < 0. \end{cases}$$

Найти преобразование Фурье этой функции.

Преобразование Лапласа функции f существует ($s_0(f) = -\alpha$), поэтому

$$2\pi\mathcal{F}(f(x);\ y) = L[f(x);\ iy] = L[e^{-\alpha x}\cos\beta x;\ iy] = \frac{iy+\alpha}{(iy+\alpha)^2+\beta^2}.$$

Приведем без доказательства ряд теорем о нахождении оригинала по известному изображению.

Теорема 11. *Пусть* $F(p)$ *— аналитическая функция на расширенной комплексной плоскости и точка* $p = \infty$ *правильная и* $F(\infty)=0$, *т. е. ее ряд Лорана имеет, вид*

$$F(p) = \sum_{k=1}^{\infty} \frac{c_k}{p^k}.$$

Тогда оригинал этого изображения дается формулой

$$f(t)=\begin{cases}0, & t<0,\\ \sum\limits_{n=0}^{\infty} c_{n+1}\dfrac{t^n}{n!}, & t>0.\end{cases} \tag{17}$$

В самом деле,

$$\int\limits_0^{\infty} f(t)\,e^{-pt}\,dt=\sum_{n=0}^{\infty}\frac{c_{n+1}}{n!}\int\limits_0^{\infty}e^{-pt}t^n\,dt=\sum_{n=0}^{\infty}\frac{c_{n+1}}{p^{n+1}}=\sum_{k=1}^{\infty}\frac{c_k}{p^k},$$

В силу теоремы 1 п. 11.1 (единственности) теорема доказана.

Теорема 12. *Пусть F (p) — дробно-рациональная функция с полюсами* $p_1, p_2, \ldots, p_m$. *Тогда*

$$f(t)=\sum_{k=1}^{m}\operatorname{Выч}_{p=p_k}[F(p)\,e^{pt}]. \tag{18}$$

Если p_k *— простые полюсы и F (p) = A(p)/B(p), где A(p), B(p) — многочлены без общих корней, то*

$$f(t)=\sum_{k=1}^{m}\frac{A(p_k)}{B'(p_k)}e^{p_k t}. \tag{19}$$

Теорема 13 (формула Меллина)). *Если F (p) — аналитическая функция в* $\operatorname{Re} p\geqslant s_0$, $F(p)\to 0$ *равномерно относительно* arg *p, при* $|p|\to\infty$, $\int\limits_{x-i\infty}^{x+i\infty}|F(p)|\,dy<M$, *то F (p) является изображением функции*

$$f(t)=\frac{1}{2\pi i}\int\limits_{x-i\infty}^{x+i\infty}e^{pt}F(p)\,dp \qquad (x>s_0). \tag{20}$$

Пример 17. Найти оригинал функции

$$F(p)=1/(p-1)(p^2+1).$$

Будем пользоваться теоремой 12. Здесь $A(p)\equiv 1$, $B(p)=(p-1)(p^2+1)$. Точки $p=1$, $p=\pm i$ являются простыми полюсами функции $F(p)$. По формуле (19) имеем

$$(B'(p)=3p^2-2p+1)$$

$$f(t)=\frac{e^t}{2}-\frac{e^{it}}{2(1+i)}+\frac{e^{-it}}{2(i-1)}=\frac{1}{2}[e^t-\cos t-\sin t].$$

Пример 18. Найти оригинал $f(t)$, если $F(p)=\sin(1/p)$. Имеем

$$\sin\frac{1}{p}=\frac{1}{p}-\frac{1}{3!p^3}+\frac{1}{5!p^5}-\ldots,$$

т. е. $F(p)$ удовлетворяет условию теоремы 11. Поэтому

$$f(t) = 1 - \frac{t}{3!}\frac{t^2}{2!} + \frac{t}{5!}\frac{t^4}{4!} - \dots .$$

Для удобства пользования сведем все полученные изображения элементарных функций в единую таблицу.

Номер по порядку	Оригинал	Изображение
1	1	$\frac{1}{p}$
2	$\sin \alpha t$	$\frac{\alpha}{p^2+\alpha^2}$
3	$\cos \alpha t$	$\frac{p}{p^2+\alpha^2}$
4	$\cos \alpha (t-t_0)$	$\frac{pe^{-pt_0}}{p^2+\alpha^2}$
5	$e^{-\alpha t}$	$\frac{1}{p+\alpha}$
6	$\operatorname{sh} \alpha t$	$\frac{\alpha}{p^2-\alpha^2}$

Продолжение

Номер по порядка	Оригинал	Изображение
7	$\operatorname{ch} \alpha t$	$\frac{p}{p^2-\alpha^2}$
8	$e^{-\alpha t} \sin \beta t$	$\frac{\beta}{(p+\alpha)^2+\beta^2}$
9	$e^{-\alpha t} \cos \beta t$	$\frac{p+\alpha}{(p+\alpha)^2+\beta^2}$
10	t^n	$\frac{\Gamma(n+1)}{p^{n+1}}$
11	$t^n f(t)$	$(-1)^n \frac{d^n F(p)}{dp^n}$
12	$te^{-\alpha t}$	$\frac{1}{(p+\alpha)^2}$
13	$t \sin \alpha t$	$\frac{2p\alpha}{(p^2+\alpha^2)^2}$
14	$t \cos \alpha t$	$\frac{p^2-\alpha^2}{(p^2+\alpha^2)^2}$
15	$f^{(n)}(t),\ f(0)=\ldots$ $\ldots=f^{(n-1)}(0)=0$	$p^n F(p)$
16	$\int_0^t f(\tau)\, d\tau$	$\frac{1}{p} F(p)$
17	$\frac{f(t)}{t}$	$\int_p^\infty F(q)\, dq$
18	$f(t-t_0)$	$e^{-pt_0} F(p)$
19	$\sigma_0(t-h)$	$e^{-ph} \frac{1}{p}$
20	$f_1 * f_2$	$L[f_1; p] \cdot L[f_2; p]$
21	$f(t)\, g(0) + \int_0^t f(\tau)\, g'(t-\tau)\, d\tau$	$pL[f; p] \cdot L[g; p]$
22	$\sum_{k=0}^{\infty} c_{k+1} \frac{t^k}{k!} \quad (t > 0)$	$\sum_{k=1}^{\infty} \frac{c_k}{p^k}$

11.3. Приложения операционного исчисления

11.3.1. Операторное уравнение.

Пусть дано линейное дифференциальное уравнение n-го порядка с постоянными коэффициентами

$$a_n x^{(n)}(t) + \ldots + a_1 x'(t) + a_0 x(t) = f(t). \quad (1)$$

Требуется найти решение уравнения (1) для $t \geq 0$ при начальных условиях

$$x(0) = x_0,\ x'(0) = x_0',\ \ldots,\ x^{(n-1)}(0) = x_0^{(n-1)}. \quad (2)$$

Пусть $x(t)$ является решением (1), удовлетворяющее начальным условиям (2). Тогда после подстановки этой функции в (1) мы получим тождество. Значит, функция, стоящая в левой части (1), и функция $f(t)$ имеют одно и то же L-изображение:

$$L\left[\sum_{k=0}^{n} a_k \frac{d^k x}{dt^k};\ p\right] = L[f(t);\ p].$$

В силу следствия 1 п.11.2

$$L\left[\frac{d^k x}{dt^k};\ p\right] = p^k L[x;\ p] - p^{k-1} x(0) - \ldots$$
$$\ldots - p x^{(k-2)}(0) - x^{(k-1)}(0).$$

Поэтому, используя свойство линейности изображения, получаем

$$a_n L\left[\frac{d^n x}{dt^n};\ p\right] + \ldots + a_0 L[x;\ p] = L[f;\ p];$$

$$a_n[p^n L[x;\ p] - p^{n-1}x_0 - p^{n-2}x_0' - \ldots - p x_0^{(n-2)} - x_0^{(n-1)}] +$$
$$+ a_{n-1}[p^{n-1}L[x;\ p] - p^{n-2}x_0 - \ldots - p x_0^{n-3} - x_0^{(n-2)}] + \ldots$$
$$\ldots + a_1[L[x;\ p] - x_0] + a_0 L[x;\ p] = L[f;\ p].$$

Для краткости записи обозначим $L[x;\ p] = \bar{x}(p)$, $L[f;\ p] = F(p)$. Тогда

$$\bar{x}(p)\cdot[a_n p^n + a_{n-1}p^{n-1} + \ldots + a_1 p + a_0] =$$
$$= a_n[p^{n-1}x_0 + p^{n-2}x_0' + \ldots + x_0^{(n-1)}] +$$
$$+ a_{n-1}[p^{n-2}x_0 + p^{n-3}x_0' + \ldots + x_0^{(n-2)}] + \ldots$$
$$\ldots + a_2[p x_0 + x_0'] + a_1 x_0 + F(p). \quad (3)$$

Уравнение (3) будем называть *вспомогательным уравнением* или *изображающим уравнением,* или *операторным уравнением.*

Отметим, что коэффициент при $\bar{x}(p)$ в (3) получается из левой части (1) формальной заменой производных $\frac{d^k x}{dt^k}$ на степени p^k. Обозначим этот коэффициент через

$$R_n(p) = a_n p^n + a_{n-1} p^{n-1} + \ldots + a_1 p + a_0.$$

Легко видеть, что этот коэффициент является левой частью характеристического уравнения для дифференциального уравнения (1). Тогда изображение решения находим в виде

$$\bar{x}(p) = \frac{F(p)}{R_n(p)} + \frac{\psi_{n-1}(p)}{R_n(p)}, \qquad (4)$$

где

$$\psi_{n-1}(p) = a_1 x_0 + a_2 (p x_0 + x_0') + a_3 (p^2 x_0 + p x_0' + x_0'') + \ldots$$
$$\ldots + a_n [p^{n-1} x_0 + p^{n-2} x_0' + \ldots + p x_0^{(n-2)} + x_0^{(n-1)}].$$

Если начальные условия нулевые, т. е. $x_0 = \ldots = x_0^{(n-1)} = 0$, то формула (4) запишется

$$\bar{x}(p) = \frac{F(p)}{R_n(p)}. \qquad (4')$$

Если теперь по изображению (4) или (4') мы найдем оригинал, то в силу теоремы единственности это и будет искомое решение *x(t).*

Пример 1. Решить уравнение

$$\ddot{x} + 4x = 2, \qquad x_0 = x_0' = 0.$$

По формуле (4') имеем

$$\bar{x}(p) = \frac{2}{p(p^2+4)},$$

так как

$$2 \doteq 2/p.$$

Разложим изображение на простейшие дроби

$$\bar{x}(p) = \frac{1}{2}\left[\frac{1}{p} - \frac{p}{p^2+4}\right].$$

Отсюда

$$x(t) = \frac{1}{2}\sigma_0(t) - \frac{1}{2}\cos 2t = \frac{1}{2} - \frac{1}{2}\cos 2t.$$

Мы получили решение $(1-\cos 2t)/2$ только для $t \geq 0$. Легко проверить, что оно удовлетворяет нашему уравнению и при $t < 0$. Впрочем, этот факт следует из общих соображений, на которых мы не останавливаемся. Это замечание относится и к примерам 2—4. Можно также воспользоваться теоремой 12 п.11.2 ($A \equiv 2$, $B = p(p^2+4)$, $B' = 3p^2+4$; $0, \pm 2i$ — простые нули многочлена $B(p)$):

$$x(t) = 2\left[\frac{e^{0t}}{B'(0)} + \frac{e^{2it}}{B'(2i)} + \frac{e^{-2it}}{B'(-2i)}\right] =$$
$$= 2\left[\frac{1}{4} + \frac{1}{-8}e^{2it} + \frac{1}{-8}e^{-2it}\right] = \frac{1}{2} - \frac{1}{2}\cos 2t.$$

Пример 2. $y'' + 2y' + 5y = \sin x,\ y(0) = 0,\ y'(0) = 1.$

Составим вспомогательное уравнение:

$$[p^2\bar{y}(p) - py(0) - y'(0)] + 2[p\bar{y}(p) - y(0)] + 5\bar{y}(p) = \frac{1}{p^2+1},$$
$$\bar{y}(p)(p^2 + 2p + 5) = \frac{1}{p^2+1} + 1.$$

Отсюда

$$\bar{y}(p) = \frac{1}{(p^2+1)(p^2+2p+5)} + \frac{1}{p^2+2p+5} = \frac{p^2+2}{(p^2+1)(p^2+2p+5)}.$$

Многочлен $B(p) = (p^2+1)(p^2+2p+5)$ имеет простые нули $p = \pm i,\ p = -1 \pm 2i.$ На основании теоремы 12 п.11.2

$$(A = p^2 + 2,\ B'(p) = 2p(p^2+2p+5) + 2(p+1)(p^2+1))$$

имеем:

$$A(\pm i) = 1,\quad B'(i) = 4i(2+i),\quad B'(-i) = -4i(2-i),$$
$$A(-1+2i) = -1-4i,\quad B'(-1+2i) = -8i(2i+1),$$
$$A(-1-2i) = 4i-1,\quad B'(-1-2i) = -8i(2i-1),$$
$$y(x) = \frac{e^{ix}}{4i(2+i)} + \frac{e^{-ix}}{-4i(2-i)} + \frac{-(1+4i)e^{(-1+2i)x}}{-8i(2i+1)} +$$
$$+ \frac{(4i-1)e^{-(1+2i)x}}{-8i(2i-1)} = \frac{\sin x}{5} - \frac{\cos x}{10} + e^{-x}\left[\frac{\cos 2x}{10} + \frac{9}{20}\sin 2x\right].$$

При решении дифференциального уравнения иногда удобно использование формулы Дюамеля.

Будем рассматривать уравнение (1) при нулевых начальных условиях: $x(0) = \ldots = x^{(n-1)}(0) = 0.$ К этому случаю всегда можно свести задачу заменой искомой функции по формуле

$$x(t) = y(t) + \sum_{k=0}^{n-1} \frac{y^k}{k!} x^{(k)}(0).$$

Допустим, известно решение уравнения (1) при правой части, равной единице, и нулевых начальных условиях. Операторное уравнение для данной задачи имеет вид

$$R_n(p)\,\bar{x}_1(p) = \frac{1}{p}, \qquad (5)$$

где $\bar{x}_1(p)$ — изображение решения $x_1(t)$ указанной задачи. Из равенств (4') и (5) находим

$$\overline{x}(p) = \frac{F(p)}{R_n(p)} = p\overline{x}_1(p)\,F(p). \qquad (6)$$

Согласно формуле Дюамеля

$$pF(p)\,\overline{x}_1(p) \doteq f(t)\,x_1(0) + \int_0^t f(\tau)\,x_1'(t-\tau)\,d\tau$$

или учитывая, что $x_1(0)=0$, получаем

$$\overline{x}(p) = pF(p)\,\overline{x}_1(p) \doteq \int_0^t f(\tau)\,x_1'(t-\tau)\,d\tau.$$

Отсюда решение уравнения (1) при нулевых начальных условиях будет иметь вид

$$x(t) = \int_0^t f(\tau)\,x_1'(t-\tau)\,d\tau, \qquad (7)$$

где $x_1(t)$ — решение уравнения (1) при $f(t)\equiv 1$ и нулевых начальных условиях.

Пример 3. Решить уравнение

$$x'' - x = \frac{1}{1+e^t}, \quad x(0) = x'(0) = 0.$$

Решим вначале задачу Коши для уравнения

$$x_1'' - x_1 = 1, \quad x_1(0) = x_1'(0) = 0.$$

Составим операторное уравнение:

$$p^2\overline{x}_1(p) - \overline{x}_1(p) = \frac{1}{p}, \quad \overline{x}_1(p) = \frac{1}{p(p^2-1)} = \frac{p}{p^2-1} - \frac{1}{p}.$$

Отсюда

$$x_1(t) = \operatorname{ch} t - 1.$$

Замечание. Так как правая часть уравнения $x_1'' - x_1' = 1$ имеет специальный вид, то решение этого уравнения можно проводить и обычным образом.

По формуле (7)

$$x(t)=\int_0^t \frac{1}{1+e^{\tau}}\operatorname{sh}(t-\tau)\,d\tau=\int_0^t \frac{e^{t-\tau}-e^{-t+\tau}}{2(1+e^{\tau})}\,d\tau=$$

$$=\frac{e^t}{2}\int_0^t \frac{e^{-\tau}\,d\tau}{1+e^{\tau}}-\frac{e^{-t}}{2}\int_0^t \frac{d(e^{\tau}+1)}{1+e^{\tau}}=$$

$$=-\frac{e^{-t}}{2}\ln\frac{e^t+1}{2}-\frac{e^t}{2}\int_0^t \frac{e^{-\tau}de^{-\tau}}{e^{-\tau}+1}=$$

$$=-\frac{e^{-t}}{2}\ln\frac{e^t+1}{2}-\frac{e^t}{2}(e^{-t}-1)+\frac{e^t}{2}\int_0^t \frac{d(e^{-\tau}+1)}{e^{-\tau}+1}=$$

$$=-\frac{e^{-t}}{2}\ln\frac{e^t+1}{2}-\frac{1}{2}+\frac{e^t}{2}+\frac{e^t}{2}\ln\frac{e^{-t}+1}{2}=$$

$$=\operatorname{sh} t\ln\frac{e^t+1}{2}+\frac{1}{2}[-te^t+e^t-1].$$

11.3.2. Решение систем дифференциальных уравнений.

Рассмотрим этот вопрос на конкретном примере.

Пример 4. Пусть требуется найти решение линейной системы

$$\left.\begin{aligned}2\dot{x}+\dot{y}+x&=1,\\ \dot{x}+3\dot{y}+2y&=0\end{aligned}\right\}$$

при начальных условиях $y(0) = x(0) = 0$.

Обозначим $\bar{x}(p)$, $\bar{y}(p)$ изображения искомых функций. Составим вспомогательные уравнения:

$$\left.\begin{aligned}2p\bar{x}(p)+p\bar{y}(p)+\bar{x}(p)&=\frac{1}{p},\\ p\bar{x}(p)+3p\bar{y}(p)+2\bar{y}(p)&=0.\end{aligned}\right\}$$

Таким образом, для изображений мы получили линейную систему алгебраических уравнений. Определитель системы

$$\Delta=\begin{vmatrix}2p+1 & p\\ p & 3p+2\end{vmatrix}=5p^2+7p+2.$$

Решая систему алгебраических уравнений, находим

$$\bar{x}(p)=\frac{3p+2}{p(5p^2+7p+2)},\quad \bar{y}(p)=\frac{-1}{5p^2+7p+2}.$$

Изображение $\bar{y}(p)$ запишем в виде

$$\bar{y}(p) = -\frac{1}{5}\frac{1}{(p+0,7)^2-0,09} = -\frac{2}{3}\frac{0,3}{(p+0,7)^2-(0,3)^2},$$

откуда

$$y(t) = -\frac{2}{3}e^{-0,7t}\,\mathrm{sh}(0,3)\,t.$$

Далее

$$\bar{x}(p) = \frac{3}{5[(p+0,7)^2-(0,3)^2]} + \frac{1}{p} - \frac{5p+7}{5p^2+7p+2} =$$
$$= 2\frac{0,3}{(p+0,7)^2-(0,3)^2} + \frac{1}{p} - \frac{p+0,7}{(p+0,7)^2-(0,3)^2} -$$
$$- \frac{0,7}{(p+0,7)^2-(0,3)^2},$$

т. е.

$$x(t) =$$
$$= 2e^{-0,7t}\,\mathrm{sh}\,(0,3t) + 1 - e^{-0,7t}\,\mathrm{ch}\,(0,3t) - \frac{7}{3}e^{-0,7t}\,\mathrm{sh}\,(0,3t) =$$
$$= 1 - \frac{1}{3}e^{-0,7t}\,\mathrm{sh}\,(0,3t) - e^{-0,7t}\,\mathrm{ch}\,(0,3t).$$

11.3.3. Вычисление интегралов

Пример 5. Вычислить интеграл $I(x) = \int_0^\infty \frac{1-\cos xt}{t^2}\,dt.$

Найдем изображение этого интеграла:

$$L[I(x); p] =$$
$$= \int_0^\infty e^{-px}\int_0^\infty \frac{1-\cos xt}{t^2}\,dt\,dx = \int_0^\infty\int_0^\infty e^{-px}(1-\cos xt)\,dx\,\frac{dt}{t^2} =$$
$$= \int_0^\infty L[1-\cos xt;\ p]\frac{dt}{t^2} = \int_0^\infty\left[\frac{1}{p} - \frac{p}{p^2+t^2}\right]\frac{dt}{t^2} =$$
$$= \int_0^\infty \frac{dt}{p(p^2+t^2)} = \frac{1}{p^2}\,\mathrm{arctg}\,\frac{t}{p}\Big|_{t=0}^{\infty} = \frac{\pi}{2p^2}.$$

Таким образом,

$$I(x) = \frac{\pi}{2}x.$$

12. Обобщенные функции

12.1. Понятие обобщенной функции

В математике и ее приложениях получили большое применение обобщенные функции. Само понятие обобщенная функция возникло в работах П. Дирака.

Общая математическая теория обобщенных функций заложена в работах С. Л. Соболева и Л. Шварца.

Ниже излагаются элементарные сведения из теории обобщенных функций, заданных на всей бесконечной действительной оси $(-\infty,\infty)$.

В основе этой теории лежит пространство S, состоящее из функций $\varphi(x)$, вообще говоря комплекснозначных ($\varphi(x) = \varphi_1(x) + i\varphi_2(x)$. φ_1 и φ_2 — действительные функции). Каждая функция $\varphi \in S$ обладает следующими свойствами:

1) $\varphi(x)$ непрерывная на оси $(-\infty,\infty)$ бесконечно дифференцируемая функция;

2) для любого неотрицательного целого числа k и любого многочлена произвольной степени n

$$P_n(x) = a_0x^n + a_1x^{n-1} + \ldots + a_{n-1}x + a_n$$

произведение k-й производной от $\varphi(x)$ на многочлен $P_n(x)$ стремится к нулю при $x \to \infty$:

$$\lim_{x \to \infty} \varphi^{(k)}(x) P_n(x) = 0.$$

Из этих свойств вытекает, что для каждой функции $\varphi \in S$ существует конечный несобственный интеграл

$$\int_{-\infty}^{\infty} |\varphi^{(k)}(x)|\, dx < \infty \quad (k = 0, 1, 2, \ldots). \tag{1}$$

В самом деле, по условию, например, существует предел

$$\lim_{x \to \infty} [(1 + x^2)^m \varphi^{(k)}(x)] = 0,$$

где m — любое натуральное число. Следовательно, для числа $\varepsilon = 1$ существует такое число $N > 0$, что

$$|(1 + x^2)^m \varphi^{(k)}(x)| < 1 \tag{2}$$

для $\forall x \in |x| > N$. И так как функция $(1 + x^2)^m \varphi^{(k)}(x)$ непрерывна на отрезке $[-N, N]$, то по теореме Вейерштрасса (ее модуль ограничен на $[-N, N]$ некоторым числом M;

$$|(1 + x^2)^m \varphi^{(k)}(x)| \leq M \quad (x \in [-N, N]).$$

Но тогда

$$|(1 + x^2)^m \varphi^{(k)}(x)| \leq M + 1 \quad (\forall x \in (-\infty, \infty))$$

и, следовательно,

$$|\varphi^{(k)}(x)| \leqslant \frac{M+1}{(1+x^2)^m} \qquad (x \in (-\infty, \infty)), \tag{3}$$

откуда, на основании признака сравнения

$$\int_{-\infty}^{\infty} |\varphi^{(k)}(x)|\,dx \leqslant (M+1) \int_{-\infty}^{\infty} \frac{dx}{(1+x^2)^m} < \infty.$$

Этим мы доказали, что всякая функция $\varphi \in S$, вместе со своими производными $\varphi^{k}(x)$, принадлежит пространству $L' = L'(-\infty, \infty)$.

Примером функции $\varphi \in S$ может служить функция $\varphi(x) = \exp(-x^2)$. Это бесконечно дифференцируемая на $(-\infty,\infty)$ функция. Ее производные соответственно равны:

$$\varphi'(x) = -2x\exp(-x^2), \quad \varphi''(x) = (4x^2-2)\exp(-x^2),$$
$$\varphi'''(x) = (-8x^3+12x)\exp(-x^2).$$

По индукции нетрудно показать, что

$$\varphi^{(k)}(x) = Q_k(x)\exp(-x^2),$$

где $Q_k(x)$ есть некоторый многочлен степени k.

Если тепеоь $P_n(x)$ есть произвольный многочлен степени n, то произведение $P_n(x)\,Q_k(x) = R_{n+k}(x)$ есть, очевидно, некоторый многочлен степени $n+k$ и

$$\lim_{x\to\infty} \varphi^{(k)}(x) P_n(x) = \lim_{x\to\infty} R_{n+k}(x)\exp(-x^2) =$$
$$= \lim_{x\to\infty} (c_0 x^{n+k} + \ldots + c_{n+k-1}x + c_{n+k})\exp(-x^2) = 0,$$

так как при любом $l > 0$

$$\lim_{x\to\infty} x^l \exp(-x^2) = 0$$

Вторым примером функции $\varphi \in S$ является так называемая *финитная на* $(-\infty,\infty)$ *функция*. Это бесконечно дифференцируемая функция, равная нулю вне некоторого отрезка $[a, b]$ (рис. 1).

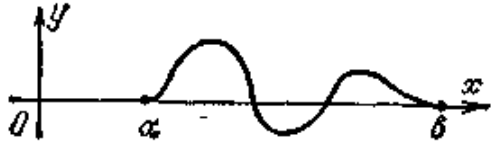


Рис. 1

Любая ее производная тоже равна нулю вне $[a, b]$ и, следовательно, для любого многочлена $P_n(x)$ степени n

$$\lim_{x\to\infty} \varphi^{(k)}(x)\,P_n(x) = 0.$$

Во множестве S вводится понятие предельного перехода.

Последовательность $\{\varphi_k(x)\}$ функций из S называется *сходящейся к функции* $\varphi \in S$ *в смысле* S, если для любого неотрицательного целого числа l и любого многочлена $P_n(x)$ имеет место равенство

$$\lim_{k \to \infty} (\varphi_k^{(l)}(x) - \varphi^{(l)}(x)) P_n(x) = 0$$

равномерно относительно всех $x \in (-\infty, \infty)$. Иначе говоря, для любого неотрицательного целого числа l, любого многочлена $P_n(x)$ и любого $\varepsilon > 0$ найдется число k_0 такое, что $|\varphi_k^{(l)}(x) - \varphi^{(l)}(x)| \cdot |P_n(x)| < \varepsilon$, для $\forall k > k_0$ и $\forall x \in (-\infty, \infty)$.

Если последовательность $\{\varphi_k\}$ сходится к φ в смысле S, то пишут

$$\varphi_k(x) \to \varphi(x) \quad (S) \quad \text{или} \quad \lim_{k \to \infty} \varphi_k(x) = \varphi(x) \quad (S).$$

Множество функций $\varphi \in S$ обладает следующим свойством: если $\varphi \in S$, $\psi \in S$ и α, β — произвольные числа, вообще комплексные, то

$$\alpha\varphi + \beta\psi \in S.$$

Благодаря этому свойству множество S называется *линейным множеством* (или *пространством).*

Введем определение: *если каждой функции* $\varphi \in S$, *в силу некоторого закона, приведено в соответствие число* y, *то говорят, что на* S *определен функционал* F *и пишут*

$$y = F\varphi = (F, \varphi).$$

Функционал F называется *линейным,* если он обладает свойством: каковы бы ни были функциии

$$\varphi \in S \text{ и } \psi \in S$$

и комплексные числа α, β, справедливо равенство

$$F(\alpha\varphi + \beta\psi) = \alpha F\varphi + \beta F\psi$$

или, в другой записи, $(F, \alpha\varphi + \beta\psi) = \alpha(F, \varphi) + \beta(F, \psi)$.

Функционал F называется *непрерывным,* если для любой последовательности $\{\varphi_k\}$ функций $\varphi_k \in S$, сходящейся в смысле S к некоторой функции φ, имеет место равенство

$$\lim_{\varphi_k \to \varphi\ (S)} F(\varphi_k) = F(\varphi).$$

Линейный и непрерывный функционал, определенный на S,

$$F(\varphi) = (F, \varphi) \quad (\varphi \in S)$$

называется обобщенной функцией над S.

Совокупность всех обобщенных функций над S обозначается через S'. Приведем примеры обобщенных функций.

Пусть $F(x)$ есть кусочно-непрерывная функция, удовлетворяющая неравенству

$$|F(x)| \leqslant c(1 + x^2)^l, \tag{4}$$

где l — некоторое натуральное число. Покажем, что интеграл

$$(F, \varphi) = \int_{-\infty}^{\infty} F(x)\varphi(x)\,dx \quad (\varphi \in S)$$

есть обобщенная функция $F \in S'$, т. е. *линейный непрерывный функционал над S*. В самом деле, на основании неравенства (4) и неравенства (3), в котором надо положить $k=0$, $m=l+1$, интеграл (5) сходится, и притом абсолютно:

$$\int_{-\infty}^{\infty} |F(x)\varphi(x)|\,dx \leqslant$$

$$\leqslant c\int_{-\infty}^{\infty} (1+x^2)^l \frac{M+1}{(1+x^2)^{l+1}}\,dx = c(M+1)\int_{-\infty}^{\infty} \frac{dx}{1+x^2} < \infty.$$

Линейность функционала (5) очевидна. Функционал (5) является также непрерывным в смысле S. В самом деле, пусть

$$\varphi_n \to \varphi \quad (S).$$

Тогда, в частности, стремится к нулю величина

$$\max_x (1+x^2)^{l+1} |\varphi_n(x) - \varphi(x)| \xrightarrow[n\to\infty]{} 0.$$

Поэтому для любого $\varepsilon > 0$ найдется $N > 0$ такое, что при $n > N$

$$\max_x (1+x^2)^{l+1} |\varphi_n(x) - \varphi(x)| < \varepsilon. \qquad (6)$$

Но тогда, в силу (4) и (6),

$$|(F, \varphi_n) - (F, \varphi)| = \left|\int_{-\infty}^{\infty} F(x)[\varphi_n(x) - \varphi(x)]\,dx\right| \leqslant$$

$$\leqslant c\int_{-\infty}^{\infty} (1+x^2)^{l+1} |\varphi_n(x) - \varphi(x)| \frac{dx}{1+x^2} \leqslant c\varepsilon \int_{-\infty}^{\infty} \frac{dx}{1+x^2} = c_1\varepsilon \quad (n > N),$$

и мы доказали непрерывность функционала (F, φ).

Важно отметить, что *для того чтобы две кусочно-непрерывные функции $F_1(x)$ и $F_2(x)$, удовлетворяющие при некотором l неравенству* (4), *представляли при помощи равенства* (5) *равные обобщенные функции $F_1 = F_2 \in S'$, необходимо и достаточно, чтобы имело место равенство $F_1(x) = F_2(x)$ во всех точках непрерывности $F_1(x)$ и $F_2(x)$.*

Достаточность условия очевидна, так как величина интеграла не изменится, если подынтегральную функцию изменить в конечном числе точек. Но можно доказать, что это условие является также и необходимым.

В связи со сказанным обобщенную функцию, представимую при помощи интеграла (5) кусочно-непрерывной функцией $F(x)$, отож-

дествляют с этой обычной функцией. Например, $\sin x$, $\frac{\sin x}{x}$,

$$\exp(-x^2), \quad \sum_{k=0}^{n} a_k x^k, \quad \sigma_0(x) = \begin{cases} 1, & x \geqslant 0, \\ 0, & x < 0 \end{cases}$$

— функция Хевисайда,

обычные функции, но также и обобщенные, принадлежащие S'. Для них справедливы неравенства типа (4):

$$|\sin x| \leqslant 1, \quad \left|\frac{\sin x}{x}\right| \leqslant 1, \quad \exp(-x^2) \leqslant 1,$$

$$\left|\sum_{k=0}^{n} a_k x^k\right| \leqslant c(1+x^2)^n, \quad |\sigma_0(x)| \leqslant 1.$$

Имеется много и других обычных функций $F(x)$, которые определяют при помощи равенства (5) обобщенную функцию $F\ (F \in S')$, хотя они и не удовлетворяют неравенству (4). Например, нетрудно показать, что функция $\psi(x) = \ln|x|$, хотя и не удовлетворяет неравенству (4), все же порождает обобщенную функцию $(\psi \in S')$.

Однако существуют элементарные функции $F(x)$, для которых интеграл (5) не является линейным непрерывным функционалом над S. Функция $F(x) = \exp(x^2)$ является примером такой функции. Ведь $\varphi(x) = \exp(-x^2) \in S$, но

$$\int_{-\infty}^{\infty} F(x)\varphi(x)\,dx = \int_{-\infty}^{\infty} \exp(x^2)\exp(-x^2)\,dx = \int_{-\infty}^{\infty} 1\,dx = \infty.$$

Обобщенную функцию $F \in S'$, порожденную обычной функцией $F(x)$ в виде интеграла (5), называют *регулярной обобщенной функцией.*

Однако в S' входят также и другие обобщенные функции.

Важным примером обобщенной нерегулярной функции является *дельта-функция Дирака,* обозначаемая через $\delta(x)$.

Функция $\delta(x)$ есть функционал, определенный на функциях $\varphi \in S$ при помощи равенства $(\delta, \varphi) = \varphi(0) \quad (\varphi \in S)$.

δ-функция приводит в соответствие каждой функции $\varphi \in S$ ее значение в точке $x=0$. Можно доказать, что не существует обычной функции $F(x)$, которая представляла бы δ-функцию в виде интерала (5), т. е. функция Дирака — это подлинно обобщенная функция.

12.2. Операции над обобщенными функциями

Производная от обобщенной функции $F \in S'$ по определению есть обобщенная функция F', определяемая равенством

$$(F', \varphi) = -(F, \varphi') \quad (\varphi \in S). \tag{1}$$

Так как из того, что $\varphi \in S$, следует, что $\varphi' \in S$, и из того, что $\varphi_n \to \varphi$ (S), следует, что $\varphi'_n \to \varphi'$ (S), то функционал (F, φ') является непрерывным функционалом над S. Линейность его очевидна. Определение (1) естественно, потому что, если, например, обычная функция $F(x) \in S$, то

$$\int_{-\infty}^{\infty} F'(x)\varphi(x)\,dx = F(x)\varphi(x)\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} F(x)\varphi'(x)\,dx =$$

$$= -\int_{-\infty}^{\infty} F(x)\varphi'(x)\,dx.$$

Ведь всякая функция из S стремится к нулю при $x \to \infty$. Очевидно, что любая обобщенная функция $F \in S'$ имеет производную (обобщенную) какого угодно порядка, определяемую по индукции $F^{(k)} = (F^{(k-1)})'$. Таким образом,

$$(F^{(k)}, \varphi) = (-1)^k (F, \varphi^{(k)}).$$

Например,

$$(\delta^{(k)}, \varphi) = (-1)^k (\delta, \varphi^{(k)}) = (-1)^k \varphi^{(k)}(0);$$

$$(\sigma_0', \varphi) = -(\sigma_0, \varphi') =$$

$$= -\int_{-\infty}^{\infty} \sigma_0(x)\varphi'(x)\,dx = -\int_{0}^{\infty} \varphi'(x)\,dx = -\varphi(x)\Big|_{0}^{\infty} = \varphi(0) = (\delta, \varphi)$$

Таким образом, производная от регулярной обобщенной функции Хевисайда $\sigma_0(x)$ равна $\delta(x)$, т. е. подлинно обобщенной функции $(\sigma_0'(x) = \delta(x))$.

По определению *последовательность обобщенных функций* $F_n \in S'$ *сходится к функции* $F \in S'$ $(F_n \longrightarrow F\ (S'))$, если

$$\lim_{n \to \infty} (F_n, \varphi) = (F, \varphi) \quad (\forall \varphi \in S).$$

Отсюда автоматически также следует, что последовательность производных F'_n сходится к производной F', потому что

$$(F'_n, \varphi) = -(F_n, \varphi') \longrightarrow -(F, \varphi') = (F', \varphi) \quad (n \longrightarrow \infty).$$

Можно рассматривать ряд

$$F = u_1 + u_2 + u_3 + \ldots \tag{2}$$

функций $u_k \in S'$, имеющий своей суммой функцию $F \in S'$, что надо понимать в том смысле, что

$$\sum_{k=1}^{N} u_k \longrightarrow F \ (S') \quad (N \longrightarrow \infty).$$

Из сказанного, очевидно, следует, что ряд (2) можно почленно дифференцировать

$$F' = u_1' + u_2' + u_3' + \ldots, \tag{3}$$

т. е. ряд (3) сходится в смысле (S'). Но тогда его можно тоже почленно дифференцировать:

$$F'' = u_1'' + u_2'' + u_3'' + \ldots.$$

Для обобщенной функции F по определению вводится операция умножения на бесконечно дифференцируемую функцию $X(x)$ с помощью равенства $(\lambda F, \varphi) = (F, \lambda\varphi)$.

Отметим еще, что если $F(x) \in S'$, $\mu \neq 0$ — действительное число, то обобщенные функции

$$F(\mu - x), \ F(\mu x)$$

определяются при помощи равенств

$$(F(\mu - x), \varphi(x)) = (F(x), \varphi(\mu - x)),$$

$$(F(\mu x), \varphi(x)) = |\mu|^{-1} \left(F(x), \varphi\left(\frac{x}{\mu}\right)\right).$$

Естественность данных определений легко выясняется на обычных функциях из пространства S.

12.3. Преобразование Фурье обобщенных функций

Отметим, что если функция φ принадлежит S, то ее преобразование Фурье

$$\tilde{\varphi}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \varphi(t)\, e^{-ixt}\, dt$$

также принадлежит S.

При этом преобразование $\varphi = \tilde{\varphi}$ отображает S на S линейно и непрерывно.

Непрерывность заключается в том, что если какая-либо последовательность функций $\varphi_n \in S$ сходится в смысле S к функции φ, то и $\tilde{\varphi}_n$ сходится к $\tilde{\varphi}$:

$$\tilde{\varphi}_n \longrightarrow \tilde{\varphi} \quad (S) \quad \text{при} \quad \varphi_n \longrightarrow \varphi \quad (S).$$

Подобные факты имеют место и для обратного преобразования Фурье $\hat{\varphi}$ функции $\varphi \in S$ $\left(\hat{\varphi}(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} \varphi(t)\, e^{ixt}\, dt\right)$.

После сделанных замечаний естественно определить *преобразование Фурье обобщенной функции* $F \in S'$ с помощью следующих равенств.

$$(\tilde{F},\ \varphi) = (F,\ \tilde{\varphi}), \quad (\hat{F},\ \varphi) = (F,\ \hat{\varphi}). \tag{1}$$

Так как для функции $\varphi \in S$ имеет место равенство

$$\hat{\tilde{\varphi}} = \tilde{\hat{\varphi}} = \varphi,$$

то подобные равенства верны также для обобщенных функций: $\hat{\tilde{F}} = \tilde{\hat{F}} = F, \quad F \in S'$. В самом деле, например,

$(\hat{\tilde{F}},\ \varphi) = (\tilde{F},\ \hat{\varphi}) = (F,\ \tilde{\hat{\varphi}}) = (F,\ \varphi)$, откуда следует, что $\tilde{F} = F$.

Отметим еще, что если $\varphi(t) \in S$, то

$$\tilde{\varphi}(t), \quad t\tilde{\varphi}(t), \quad \varphi'(t)$$

также принадлежат S, и имеет место равенство

$$\varphi'(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} it\tilde{\varphi}(t)\, e^{itx}\, dt$$

$$\left(\varphi'(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty} -it\hat{\varphi}(t)\, e^{-itx}\, dt\right)$$

или, коротко,

$$\varphi' = \widehat{it\tilde{\varphi}} \quad (\varphi' = -\widetilde{it\hat{\varphi}}). \tag{2}$$

Для обобщенных функций имеет место подобный факт:

$$F' = \widehat{ix\tilde{F}} = -\widetilde{ix\hat{F}}. \tag{3}$$

В самом деле, например з силу (1) и (2), получаем

$$(-\widetilde{ix\hat{F}},\ \varphi) = (-ix\hat{F},\ \tilde{\varphi}) = (\hat{F},\ -ix\tilde{\varphi}) = (F,\ -\widehat{ix\tilde{\varphi}}) = -(F,\ \varphi') = (F',\ \varphi),$$

т. е. $F' = -\widetilde{ix\hat{F}}$ и (3) доказано.

По индукции легко выводим, что

$$F^{(k)} = \widehat{(ix)^k \tilde{F}} = (-\widetilde{ix)^k \hat{F}} \quad (k = 0,\ 1,\ 2,\ \ldots).$$

Пример. Найти преобразование Фурье обобщенной функции Дирака.

Р е ш е н и е. По определению имеем

$$(\tilde{\delta}, \varphi) = (\delta, \tilde{\varphi}) = \tilde{\varphi}(0) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \varphi(t)\, e^{-ixt}\, dt|_{x=0} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \varphi(t)\, dt.$$

Отсюда $\tilde{\delta} = \frac{1}{\sqrt{2\pi}}$. Аналогично можно получить, что $\hat{\delta} = \frac{1}{\sqrt{2\pi}}$.

Преобразование Фурье обобщенных функций обладает свойствами преобразований Фурье обычных функций. Например, если

$$F \in S', \text{ то } \widehat{e^{i\mu t} F} = F(x+\mu)$$

при любом действительном $\mu \neq 0$.

В самом деле, на основании подобного свойства для функций из S имеем

$$(\widehat{e^{i\mu t} F},\ \varphi) = (e^{i\mu t} \tilde{F},\ \hat{\varphi}) = (\tilde{F},\ e^{i\mu t} \hat{\varphi}) =$$

$$= (F,\ \widetilde{e^{i\mu t} \hat{\varphi}}) = (F,\ \varphi(x-\mu)) = (F(x+\mu),\ \varphi),$$

откуда $\widehat{e^{i\mu t} F} = F(x+\mu)$.

13. Числа и последовательности Фибоначчи

Древняя история богата выдающимися математиками. Многие достижения древней математической науки до сих пор вызывают восхищение остротой ума их авторов, а имена Евклида, Архимеда, Герона известны каждому образованному человеку.

Иначе обстоит дело с математикой средневековья. Кроме Виеты, жившего, впрочем, уже в шестнадцатом столетии, и математиков более близких нам времен школьный курс математики не называет ни одного имени, относящегося к средним векам. Это, конечно, не случайно. Математика в эту этоху развивалась чрезвычайно медленно, и крупных математиков тогда было очень мало.

Тем больший интерес представляет для нас сочинение «Liber abacci» («Книга об абаке»), написанная знаменитым итальянским математиком Леонардо из Пизы, который известен больше по своему прозвищу Фибоначчи (Fibonacci — сокращенное filius Bonacci, т. е. сын Боначчи). Эта книга, написанная в 1202 г., дошла до нас во втором своем варианте, который относится к 1228 г.

«Liber abacci» представляет собой объемистый труд, содержащий почти все арифметические и алгебраические сведения того времени и сыгравший заметную роль в развитии математики в Западной Европе в

течение нескольких следующих столетий. В частности, именно по этой книге европейцы познакомились с индусскими («арабскими») цифрами.

Сообщаемый в «Liber abacci» материал поясняется на большом числе задач, составляющих значительную часть этого трактата.

Рассмотрим одну такую задачу, помещенную на стр. 123—-124 рукописи 1228 г.

«Сколько пар кроликов в один год от одной пары рождается?»

«Некто поместил пару кроликов в некоем месте, огороженном со всех сторон стеной, чтобы узнать, сколько пар кроликов родится при этом в течение года, если природа кроликов такова, что через месяц пара кроликов производит на свет другую пару, а рождают кролики со второго месяца после своего рождения. Так как первая пара в первом месяце дает потомство, удвой, и в этом месяце окажутся 2 пары; из них одна пара, а именно первая, рождает и в следующем месяце, так что во втором месяце оказывается 3 пары; из них в следующем месяце 2 пары будут давать потомство, так что в третьем месяце родятся еще 2 пары кроликов, и число пар кроликов в этом месяце достигнет 5; из них в этом же месяце будут давать потомство 3 пары, и число пар кроликов в четвергом месяце достигнет 8; из них 5 пар произведут другие 5 пар, которые, сложенные с 8 парами, дадут в пятом месяце 13 пар; из них 5 пар, рожденных в этом месяце, не дают в том же месяце потомства, а остальные 8 пар рождают, так что в шестом месяце оказывается 21 пара; сложенные с 13 парами, которые родятся в седьмом месяце, они дают 34 пары; сложенные с 21 парой, рожденной в восьмом месяце, они дают в этом месяце 55 пар; сложенные с 34 парами, рожденными в девятом месяце, они дают 89 пар; сложенные вновь с 55 парами, которые рождаются в десятом месяце, они дают в этом месяце 144 пары; снова сложенные с 89 парами, которые рождаются в одиннадцатом месяце, они дают в этом месяце 233 пары; сложен-ные вновь с 144 парами, рожденными в последнем месяце, они дают 377 пар; столько пар произвела первая пара в данном месте к концу одного года. Действительно, на этих полях ты можешь увидеть, как мы это делаем; именно, мы складываем первое число со вторым, т. е. 1 и 2; и второе с третьим; и третье с четвертым; и четвертое с пятым; и так одно за другим, пока не сложим десятое с одиннадцатым, т. е. 144 с 233; и мы получим общее число упомянутых кроликов, т. е. 377; и так можно делать по порядку до бесконечного числа месяцев».

Перейдем теперь от кроликов к числам и рассмотрим следующую числовую последовательность:

$$u_1, u_2, \ldots, u_n, \qquad (1)$$

в которой каждый член равен сумме двух предыдущих членов, т. е. при всяком $n > 2$

$$u_n = u_{n-1}+u_{n-2}. \qquad (2)$$

Такие последовательности, в которых каждый член определяется как некоторая функция предыдущих, часто встречаются в математике и называются *рекуррентными* или, по-русски, *возвратными* последовательностями. Сам процесс последовательного определения элементов таких последовательностей называется *рекуррентным процессом,* а равенство (2) — *возвратным (рекуррентным) уравнением.*

Заметим прежде всего, что по одному только условию (2) члены последовательности (1) вычислять нельзя. Можно составить сколько угодно различных числовых последовательностей, удовлетворяющих этому условию; например,

2, 5, 7, 12, 19, 31, 50, ...,

1, 3, 4, 7, 11, 18, 29.....

— 1, -5, -6, —11, -17, ... и т. д.

Значит, для однозначного построения последовательности (1) условия (2) явно недостаточно, и нам следует указать некоторые дополнительные условия. Например, мы можем задать несколько первых членов последовательности (1). Сколько же первых членов последовательности (1) мы должны задать, чтобы можно было вычислять все следующие ее члены, пользуясь при этом только условием (2)?

Начнем с того, что не всякий член последовательности (1) может быть получен при помощи (2) уже хотя бы потому, что не у каждого члена (1) имеется два предшествующих; например, перед первым членом последовательности вообще не стоит ни одного члена, а перед вторым ее членом стоит только один. Значит, вместе с условием (2) для определения последовательности (1) нам нужно знать два ее первых члена.

Этого, очевидно, уже достаточно для того, чтобы иметь возможность вычислить любой член последовательности (1). В самом деле, u_3 можно вычислить как сумму заданных нам u_1 и u_2; u_4 — как сумму u_2 и уже вычисленного ранее u_3; u_5 — как сумму уже вычисленных u_3 и u_4 и т. д. «по порядку до бесконечного числа членов». Переходя таким образом от двух соседних членов последовательности к непосредственно следующему за ними члену, мы можем дойти до члена с любым наперед заданным номером и вычислить его.

Обратимся теперь к важному частному случаю последовательности (1), когда $u_1 = 1$ и $u_2 = 1$. Условие (2), как было только что отмечено, дает нам возможность вычислять последовательно один за другим все

члены этого ряда. Нетрудно проверить, что в этом случае первыми четырнадцатью его членами будут числа

1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, 377,

которые уже встречались нам в задаче о кроликах.

В честь автора этой задачи вся последовательность (1) при $u_1 — u_2 = 1$ называется *рядом Фибоначчи,* а члены ее — *числами Фибоначчи.*

Числа Фибоначчи обладают целым рядом интересных и важных свойств, простейшие из которых будут рассмотрены ниже

1. Вычислим сначала сумму первых n чисел Фибоначчи. Именно, докажем, что

$$u_1 + u_2 + \ldots + u_n = u_{n+2} - 1. \qquad (3)$$

В самом деле, мы имеем:

$$\begin{aligned} u_1 &= u_3 - u_2, \\ u_2 &= u_4 - u_3, \\ u_3 &= u_5 - u_4, \\ &\cdots\cdots \\ u_{n-1} &= u_{n+1} - u_n, \\ u_n &= u_{n+2} - u_{n+1}. \end{aligned}$$

Сложив все эти рааенства почленно, мы получим

$$u_1 + u_2 + \ldots + u_n = u_{n+2} - u_2,$$

и нам остается вспомнить, что $u_2 = 1$.

2. Сумма чисел Фибоначчи с нечетными номерами:

$$u_1 + u_3 + u_5 + \ldots + u_{2n-1} = u_{2n}. \qquad (4)$$

Для доказательства этого равенства напишем

$$\begin{aligned} u_1 &= u_2, \\ u_3 &= u_4 - u_2, \\ u_5 &= u_6 - u_4, \\ &\cdots\cdots \\ u_{2n-1} &= u_{2n} - u_{2n-2}. \end{aligned}$$

Сложив эти равенства почленно, мы и получим требуемое.

3. Сумма чисел Фибоначчи с четными номерами:

$$u_2 + u_4 + \ldots + u_{2n} = u_{2n+1} - 1. \qquad (5)$$

На основании п. 1 мы имеем

$$u_1 + u_2 + u_3 + \ldots + u_{2n} = u_{2n+2} - 1;$$

вычтя почленно из этого равенства равенство (4),

мы получим

$$u_2 + u_4 + \ldots + u_{2n} = u_{2n+2} - 1 - u_{2n} = u_{2n+1} - 1,$$

а это нам и требовалось.

Вычитая, далее, почленно (5) из (4), получаем

$$u_1 - u_2 + u_3 - u_4 + \ldots + u_{2n-1} - u_{2n} = -u_{2n-1} + 1. \quad (6)$$

Прибавим теперь к обеим частям (6) по u_{2n+1}:

$$u_1 - u_2 + u_3 - u_4 + \ldots - u_{2n} + u_{2n+1} = u_{2n} + 1. \quad (7)$$

Объединяя (6) и (7), получаем выражение для знакопеременной суммы чисел Фибоначчи:

$$u_1 - u_2 + u_3 - u_4 + \ldots + (-1)^{n+1} u_n = \\ = (-1)^{n+1} u_{n-1} + 1. \quad (8)$$

4. Формулы (3) и (4) были выведены при помощи почленного сложения целой серии очевидных равенств. Еще одним примером применения этого приема может служить вывод формулы для суммы квадратов первых n чисел Фибоначчи:

$$u_1^2 + u_2^2 + \ldots + u_n^2 = u_n u_{n+1}. \quad (9)$$

Заметим для этого, что

$$u_k u_{k+1} - u_{k-1} u_k = u_k (u_{k+1} - u_{k-1}) = u_k^2.$$

Сложив равенства

$$u_1^2 = u_1 u_2,$$
$$u_2^2 = u_2 u_3 - u_1 u_2,$$
$$u_3^2 = u_3 u_4 - u_2 u_3,$$
$$\ldots \ldots \ldots \ldots$$
$$u_n^2 = u_n u_{n+1} - u_{n-1} u_n$$

почленно, мы получаем формулу (9).

5. Многие соотношения между числами Фибоначчи удобно доказывать при помощи метода *полной индукции.*

Сущность метода полной индукции (называемого часто также методом математической индукции) состоит в следующем: для доказательства, что некоторое утверждение справедливо для всякого натурального числа, достаточно установить, что:

а) оно имеет место для числа 1;

б) из справедливости доказываемого утверждения для какого-либо произвольно выбранного натурального числа n следует его справедливость для числа и + 1.

Всякое индуктивное доказательство утверждения, справедливого для любого натурального числа, состоит поэтому из двух частей.

В первой части (обычно сравнительно простой) устанавливается справедливость доказываемого утверждения для единицы. Справедливость доказываемого утверждения для единицы называют иногда *основанием индукции.* Во второй части доказательства (как правило, более сложной) делается предположение о справедливости доказываемого утверждения для некоторого произвольного (но фиксированного) числа n, и из этого предположения, которое часто называют *индуктивным предположением,* выводится, что и для числа $n+1$ доказываемое утверждение имеет место. Вторая часть доказательства называется *индуктивным переходом.*

Иногда применяется индуктивное рассуждение, которое можно назвать переходом «от всех чисел, меньших n, к n». При этом необходимость в специальном доказательстве основания индукции отпадает, так как, говоря формально, доказательство для случая $n = 1$ и есть переход от «всех» целых положительных чисел, меньших единицы (которых просто нет), к единице.

Именно таким является доказательство возможности разложения любого натурального числа на простые множители.

Предположим, что каждое из чисел, меньших некоторого n, разложимо в произведение простых множителей. Если число n оказывается простым, то оно само и является своим разложением. Если же число n составное, то его, по определению, можно представить в виде произведения хотя бы двух сомножителей: $n = n_1 n_2$, где $n_1 \neq 1$ и $n_2 \neq 1$. Но тогда $n_1 < n$ и $n_2 < n$, а по индуктивному предположению как n_1, так и n_2 разлагаются на простые множители. Тем самым и n разложимо на простые множители.

6. Простейшей реализацией идеи индукции в применении к числам Фибоначчи является само определение чисел Фибоначчи. Оно, как разъяснялось выше, состоит в указании двух первых чисел Фибоначчи: $u_1 = 1$ и $u_2 = 1$ и в индуктивном переходе от u_n и u_{n+1} к u_{n+2}, даваемым рекуррентным соотношением

$$u_n + u_{n+1} = u_{n+2}\,.$$

В частности, отсюда автоматически следует, что *если некоторая последовательность чисел начинается с двух единиц, а каждое из следующих получается сложением двух предыдущих, то эта последовательность является последовательностью чисел Фибоначчи.*

В качестве примера рассмотрим так называемую «задачу о прыгуне». Она состоит в следующем.

Прыгун может прыгать в одном направлении вдоль разделенной на клетки полосы, перемещаясь при каждом прыжке либо в соседнюю клетку, либо через клетку. Сколькими способами может он сдвинуться

на n— 1 клетку и, в частности, переместиться из первой клетки в n-ю? (Способы прыгания считаются одинаковыми, если в ходе каждого из них прыгун побывает в одних и тех же клетках.)

Обозначим искомое число через x_n. Очевидно, $x_1 = 1$ (ибо переход из первой клетки в первую же осуществляется только одним способом — отсутствием прыжков) и $x_2 = 1$ (переход из первой клетки во вторую также единствен: он состоит в одном непосредственном прыжке на соседнюю клетку). Пусть целью прыгуна является достижение $n + 2$-й клетки. Общее число способов осуществления этой цели в наших обозначениях равно x_{n+2}. Но с самого начала эти способы разбиваются на два класса: начинающиеся с прыжка во вторую клетку и начинающиеся с прыжка в третью клетку. Из второй клетки прыгун может переместиться в n+2-ю x_{n+1} способами, а из третьей x_n способами. Таким образом, последовательность чисел $x_1, x_2, \ldots, x_n, \ldots$ удовлетворяет рекуррентному соотношению

$$u_n + u_{n+1} = u_{n+2}$$

и поэтому совпадает с последовательностью чисел Фибоначчи: $x_n = u_n$.

7. Докажем по индукции следующую важную формулу:

$$u_{n+m} = u_{n-1}u_m + u_n u_{m+1}. \qquad (10)$$

Доказательство этой формулы будем вести индукцией по m. При $m = 1$ эта формула принимает вид $u_{n+1} = u_{n-1}u_n + u_n u_2$, что очевидно. При $m = 2$ формула (10) также верна, потому что

$$u_{n+2} = u_{n-1}u_2 + u_n u_3 = u_{n-1} + 2u_n =$$
$$= u_{n-1} + u_n + u_n = u_{n+1} + u_n.$$

Основание индукции, таким образом, доказано. Индуктивный переход докажем в следующей форме: предполагая, что формула (10) справедлива при $m = k$ и при $m = k + 1$, докажем, что она имеет место и при $m = k + 2$.

Итак, пусть

$$u_{n+2} = u_{n+1} + u_n = u_{n-1} + u_n + u_n =$$
$$= u_{n-1} + 2u_n = u_{n-1}u_2 + u_n u_3.$$

Сложив последние два равенства почленно, мы получим

$$u_{n+k+2} = u_{n-1}u_{k+2} + u_n u_{k+3},$$

а это и требовалось.

Формулу (10) легко интерпретировать (и даже доказать) в терминах задачи о прыгуне.

Именно, общее число способов перемещения прыгуна из первой клетки в $n + m$-ю равно u_{n+m}. Среди этих способов будут как те, при

которых прыгун перепрыгнет через n-ю клетку, так и те, при которых он побывает в ней.

При способах первого класса прыгун обязан достичь n—1-й клетки (он может сделать это u_{n-1} способами), затем совершить прыжок на n+1-ю клетку и, наконец, сместиться на оставшиеся $(n+m)—(n+1)=m—1$ клеток (это осуществимо u_m способами). Следовательно, первый класс насчитывает $u_{n-1}u_m$ способов. Аналогично, при способах второго класса прыгун достигает n-й клетки (это возможно u_n способами), после чего переходит в $n+m$-ю клетку (одним из u_{m+1} способов). Поэтому во втором классе имеется $u_n u_{m+1}$ способов, и формула (10) доказана.

8. Полагая в формуле (10) $m = n$, мы получаем

$$u_{2n} = u_{n-1}u_n + u_n u_{n+1},$$

или

$$u_{2n} = u_n (u_{n-1} + u_{n+1}). \qquad (11)$$

Из написанного равенства видно, что u_{2n} делится на u_n.

Так как

$$u_n = u_{n+1} - u_{n-1},$$

формулу (11) можно переписать так:

$$u_{2n} = (u_{n+1} - u_{n-1})(u_{n+1} + u_{n-1}),$$

или

$$u_{2n} = u_{n+1}^2 - u_{n-1}^2,$$

т. е. разность квадратов двух чисел Фибоначчи, номера которых отличаются на два, есть снова число Фибоначчи.

Аналогично (полагая $m = 2n$) можно показать, что

$$u_{3n} = u_{n+1}^3 + u_n^3 - u_{n-1}^3.$$

9. В дальнейшем нам пригодится следующая формула:

$$u_n^2 = u_{n-1}u_{n+1} + (-1)^{n+1}. \qquad (12)$$

Докажем ее индукцией по n. Для $n = 2$ (12) принимает вид

$$u_2^2 = u_1 u_3 - 1,$$

что очевидно.

Предположим теперь формулу (12) доказанной для некоторого n. Прибавим к обеим частям ее по $u_n u_{n+1}$. Мы получим

$$u_n^2 + u_n u_{n+1} = u_{n-1}u_{n+1} + u_n u_{n+1} + (-1)^{n+1},$$

или

$$u_n (u_n + u_{n+1}) = u_{n+1}(u_{n-1} + u_n) + (-1)^{n+1},$$

или

$$u_n u_{n+2} = u_{n+1}^2 + (-1)^{n+1},$$

или

$$u_{n+1}^2 = u_n u_{n+2} + (-1)^{n+2}.$$

Этим индуктивный переход обоснован, и формула (12) доказана для любого n.

10. Аналогично только что доказанным свойствам чисел Фибоначчи можно установить еще и такие свойства этих чисел:

$$u_1 u_2 + u_2 u_3 + u_3 u_4 + \ldots + u_{2n-1} u_{2n} = u_{2n}^2,$$

$$u_1 u_2 + u_2 u_3 + u_3 u_4 + \ldots + u_{2n} u_{2n+1} = u_{2n+1}^2 - 1,$$

$$n u_1 + (n-1) u_2 + (n-2) u_3 + \ldots + 2u_{n-1} + u_n = u_{n+4} - (n+3),$$

$$u_1 + 2u_2 + 3u_3 + \ldots + n u_n = n u_{n+2} - u_{n+3} + 2.$$

Доказательство предоставляется провести читателю.

11. Не менее замечательными, чем числа Фибоначчи, являются другие числа, называемые *биномиальными коэффициентами.*

Биномиальными коэффициентами называются коэффициенты при степенях x в разложениях степеней $(1+x)^n$:

$$(1+x)^n = C_n^0 + C_n^1 x + C_n^2 x^2 + \ldots + C_n^n x^n. \qquad (13)$$

Очевидно, числа C_n^k однозначно определены при всех целых неотрицательных n и всех целых неотрицательных k, не превосходящих n.

Использование биномиальных коэффициентов оказывается весьма удобным во многих математических рассуждениях. Пригодятся они нам и при изучении свойств чисел Фибоначчи. Кроме того, биномиальные коэффициенты связаны с числами Фибоначчи и непосредственно, и мы выявим некоторые закономерности, связывающие эти два класса чисел.

Предварительно установим некоторые свойства биномиальных коэффициентов.

Положив в (13) $n = 1$, мы видим, что

$$C_1^0 = C_1^1 = 1;$$

кроме того, имеет место следующая лемма.

Лемма. $C_n^k + C_n^{k+1} = C_{n+1}^{k+1}$.

Доказательство. Мы имеем

$$(1+x)^{n+1} = (1+x)^n (1+x),$$

или, пользуясь определением биномиальных коэффициентов,

$$C_{n+1}^0 + C_{n+1}^1 x + \ldots + C_{n+1}^{k+1} x^{k+1} + \ldots + C_{n+1}^{n+1} x^{n+1} =$$
$$= (C_n^0 + C_n^1 x + \ldots + C_n^k x^k + C_n^{k+1} x^{k+1} + \ldots$$
$$\ldots + C_n^n x^n)(1 + x) = C_n^0 + (C_n^0 + C_n^1) x + \ldots$$
$$\ldots + (C_n^k + C_n^{k+1}) x^{k+1} + \ldots + (C_n^{n-1} + C_n^n) x^n + C_n^n x^{n+1}.$$

Но слева и справа в этом равенстве стоит *один и тот же* полином. Поэтому и коэффициенты при одинаковвых степенях x слева и справа должны быть равны, В частности, должно быть

$$C_{n+1}^{k+1} = C_n^k + C_n^{k+1},$$

а это и требовалось.

Из доказанной леммы следует, что биномиальные коэффициенты можно вычислять при помощи некоторого рекуррентного процесса, подобного процессу получения чисел Фибоначчи, только значительно более сложной природы. Это же обстоятельство дает нам возможность доказывать по индукции разного рода утверждения о биномиальных коэффициентах.

12. Расположим биномиальные коэффициенты в виде следующей таблицы, называемой *треугольником Паскаля:*

$$C_0^0$$
$$C_1^0 \; C_1^1$$
$$C_2^0 \; C_2^1 \; C_2^2$$
$$\ldots\ldots\ldots$$
$$C_n^0 \; C_n^1 \; C_n^2 \ldots C_n^n$$
$$\ldots\ldots\ldots$$

т. е.

1
1 1
1 2 1
1 3 3 1
1 4 6 4 1
1 5 10 10 5 1
1 6 15 20 15 6 1
. . . . . .

Строки треугольника Паскаля принято нумеровать сверху вниз, причем верхняя строка, состоящая из единственной единицы, считается нулевой.

Из предыдущего вытекает, что крайние члены в каждой из строк треугольника Паскаля равны единице, а каждый из остальных членов таблицы получаются путем сложения двух других, стоящих непосредственно над ним.

13. Формула (13) позволяет сразу вывести два важных соотношения, связывающих биномиальные коэффициенты, составляющие одну строку треугольника Паскаля.

Полагая в (13) $x = 1$, получаем

$$2^n = C_n^0 + C_n^1 + C_n^2 + \ldots + C_n^n.$$

Если же принять $x = -1$, то получим

$$0 = C_n^0 - C_n^1 + C_n^2 + \ldots + (-1)^n C_n^n.$$

14. Докажем индукцией по n, что

$$C_n^k = \frac{n(n-1)\ldots(n-k+1)}{1 \cdot 2 \cdot \ldots \cdot k}. \qquad (14)$$

Эта формула часто принимается за определение биномиальных коэффициентов. Она характеризует биномиальный коэффициент C_n^k как число сочетаний из n элементов по k. Мы пошли здесь по иному, более формальному пути, который в данном случае предпочтительнее.

Если согласиться считать, что произведение нулевого числа сомножителей всегда равно единице, то при $k=0$ из (14) получаем уже известное нам $C_n^0 = 1$. Имея это в виду, мы можем ограничиться случаем $k \geq 1$.

При $n = 1$ мы имеем

$$C_1^1 = \frac{1}{1} = 1.$$

Пусть теперь при некотором данном n формула (14) справедлива при любом значении $k = 0, 1, \ldots, n$.

Рассмотрим число C_{n+1}^k. Так как $k \geq 1$, мы можем написать

$$C_{n+1}^k = C_n^{k-1} + C_n^k,$$

или, воспользовавшись индуктивным предположением (14),

$$C_n^{k-1} + C_n^k =$$
$$= \frac{n(n-1)\dots(n-k+2)}{1\cdot 2\cdot\dots\cdot(k-1)} + \frac{n(n-1)\dots(n-k+2)(n-k+1)}{1\cdot 2\cdot\dots\cdot(k-1)\,k} =$$
$$= \frac{n(n-1)\dots(n-k+2)}{1\cdot 2\cdot\dots\cdot(k-1)}\left(1 + \frac{n-k+1}{k}\right) =$$
$$= \frac{n(n-1)\dots(n-k+2)}{1\cdot 2\cdot\dots\cdot(k-1)}\,\frac{k+n-k+1}{k} =$$
$$= \frac{(n+1)\,n(n-1)\dots(n-k+2)}{1\cdot 2\cdot\dots\cdot(k-1)\,k} = C_{n+1}^k.$$

Последнее равенство является формулой (14) для биномиальных коэффициентов из следующей, n+1-й строки треугольника Паскаля.

15. Проведем через числа треугольника Паскаля линии, идущие под углом 45 градусов к его строкам, и назовем их *восходящими диагоналями* треугольника Паскаля. Восходящими диагоналями будут, например, прямые, проходящие через числа 1, 4, 3 или 1, 5, 6, 1.

Покажем, что сумма чисел, лежащих на некоторой восходящей диагонали, есть число Фибоначчи.

В самом деле, первая, самая верхняя восходящая диагональ треугольника Паскаля состоит только из единицы. Только из единицы состоит и вторая его диагональ. Для доказательство интересующего нас предложения достаточно показать, что сумма всех чисел, составляющих n-ю и n+1-ю диагонали треугольника Паскаля, равна сумме чисел, составляющих его n + 2-ю диагональ.

Но на n-й диагонали расположены числа

$$C_{n-1}^0,\quad C_{n-2}^1,\quad C_{n-3}^2,\dots,$$

а на n + 1-й — числа

$$C_n^0,\quad C_{n-1}^1,\quad C_{n-2}^2,\dots$$

Сумму всех этих чисел запишем так:

$$C_n^0 + \left(C_{n-1}^0 + C_{n-1}^1\right) + \left(C_{n-2}^1 + C_{n-2}^2\right) + \dots,$$

или, принимая во внимание лемму в п. 11,

$$C_{n+1}^0 + C_n^1 + C_{n-1}^2 + \dots$$

Последнее выражение есть сумма чисел, лежащих на n + 2-й восходящей диагонали треугольника.

Из только что доказанного на основании формулы (3) мы получаем: сумма всех биномиальных коэффициентов, лежащих выше n-й восходящей диагонали треугольника Паскаля (включая саму эту диагональ), равна u_{n+2}— 1.

Используя формулы (4), (5), (6) и подобные им, читатель без труда может получить дальнейшие тождества, связывающие числа Фибоначчи с биномиальными коэффициентами.

16. До сих пор мы определяли число Фибоначчи рекуррентно, т. е. индуктивно, по их номеру. Оказывастся, однако, что любое число Фибоначчи можно определить и непосредственно, как некоторую функцию его номера.

Исследуем для этого различные последовательности $u_1, u_2, \ldots, u_n, \ldots$, удовлетворяющие соотношению

$$u_n = u_{n-2} + u_{n-1}. \tag{15}$$

Все такие последовательности мы будем называть *решениями уравнения* (15).

Будем обозначать буквами V, V' и V'' соответственно последовательности

$$v_1, \quad v_2, \quad v_3, \quad \ldots$$
$$v'_1, \quad v'_2, \quad v'_3, \quad \ldots$$
$$v''_1, \quad v''_2, \quad v''_3, \quad \ldots$$

Сначала докажем две леммы.

Лемма 1. *Если V есть решение уравнения* (15), *а c — произвольное число, то последовательность cV (т. е. последовательность $cv_1, cv_2, cv_3, \ldots$) есть также решение уравнения* (15).

Доказательство. Умножив соотношение

$$v_n = v_{n-2} + v_{n-1}$$

почленно на c, мы получаем

$$cv_n = cv_{n-2} + cv_{n-1},$$

а это и требовалось.

Лемма 2. *Если последовательности V' и V'' являются решениями уравнения* (15), *то и их сумма $V' + V''$ (т. е. последовательность $v'_1 + v''_1, \ v'_2 + v''_2, \ v'_3 + v''_3, \ldots$) также является решением уравнения* (15).

Доказательство. Из условия леммы мы имеем

$$v'_n = v'_{n-2} + v'_{n-1}$$

и

$$v''_n = v''_{n-2} + v''_{n-1}.$$

Сложив эти два равенства почленно, мы получим

$$v'_n + v''_n = (v'_{n-2} + v''_{n-2}) + (v'_{n-1} + v''_{n-1}).$$

Этим лемма доказана.

Пусть теперь V' и V'' — два непропорциональных решения уравнения (15) (т. е. два таких решения уравнения (15), что при любом постоянном c найдется такой номер n, для которого $\frac{v'_n}{v''_n} \neq c$). Покажем, что всякую последовательность V, являющуюся решением уравнения (15), можно представить в виде

$$c_1 V' + c_2 V'', \tag{16}$$

где c_1 и c_2 — некоторые постоянные. Поэтому принято говорить, что (16) является *общим решением* уравнения (15).

Предварительно докажем, что если решения V' и V'' уравнения (15) непропорциональны, то

$$\frac{v'_1}{v''_1} \neq \frac{v'_2}{v''_2} \tag{17}$$

(т. е. что эта непропорциональность обнаруживается уже в первых двух членах последовательностей V' и V'').

Доказательство (17) ведется от противного. Пусть для непропорциональных решений V' и V'' уравнения (15)

$$\frac{v'_1}{v''_1} = \frac{v'_2}{v''_2}. \tag{18}$$

Написав производную пропорцию, мы получаем

$$\frac{v'_1 + v'_2}{v''_1 + v''_2} = \frac{v'_2}{v''_2}$$

или, принимая, во внимание, что V' и V'' являются решениями уравнения (15),

$$\frac{v'_3}{v''_3} = \frac{v'_2}{v''_2}.$$

Аналогично убеждаемся (индукция!) в том, что

$$\frac{v'_3}{v''_3} = \frac{v'_4}{v''_4} = \ldots = \frac{v'_n}{v''_n} = \ldots$$

Таким образом, из (18) следует, что последовательности V' и V'' пропорциональны, а это противоречит предположению. Значит, справедливо (17).

Возьмем теперь некоторую последовательность V, являющуюся решением уравнения (15). Эта последовательность, как уже было выяснено ранее, вполне определена, если заданы два ее первых члена, v_1 и v_2.

Найдем такие c_1 и c_2, чтобы имело место

$$\begin{aligned} c_1 v_1' + c_2 v_1'' &= v_1, \\ c_1 v_2' + c_2 v_2'' &= v_2. \end{aligned} \tag{19}$$

Тогда на основании лемм 1 и 2 $c_1V'+c_2V''$ даст нам последовательность V.

Ввиду условия (17) система уравнений (19) разрешима относительно c_1 и c_2, каковы бы ни были числа v_1 и v_2:

$$c_1 = \frac{v_1 v_2'' - v_2 v_1''}{v_1' v_2'' - v_1'' v_2'}, \qquad c_2 = \frac{v_1' v_2 - v_2' v_1}{v_1' v_2'' - v_1'' v_2'}.$$

(Условие (17) означает, что общий знаменатель этих дробей отличен от нуля.) Подставив вычисленные значения c_1 и c_2 в (16), мы и получим требуемое представление последовательности V.

Значит, для описания *всех* решений уравнения (15) нам достаточно найти *какие-нибудь два* его непропорциональных решения.

Будем искать эти решения среди геометрических прогрессий. В соответствии с леммой 1 достаточно ограничиться рассмотрением только таких прогрессий, у которых первый член равен единице. Итак, возьмем прогрессию

$$1, q, q^2, \ldots$$

Чтобы она была решением уравнения (15), необходимо, чтобы при всяком n выполнялось

$$q^{n-2} + q^{n-1} = q^n,$$

или, сокращая на q^{n-2},

$$1+q = q^2. \tag{20}$$

Корни этого квадратного уравнения, т. е. $\frac{1+\sqrt{5}}{2}$ и $\frac{1-\sqrt{5}}{2}$, и будут искомыми знаменателями прогрессий. Мы будем их обозначать соответственно через α и β. Подчеркнем, что для чисел α и β, как для корней уравнения (20), должно иметь место

$$1+\alpha = \alpha^2, \; 1+\beta = \beta^2 \text{ и } \alpha\beta = -1.$$

Мы получили, таким образом, две геометрические прогрессии, являющиеся решениями уравнения (15). Поэтому все последовательности вида

$$c_1 + c_2, \quad c_1\alpha + c_2\beta, \quad c_1\alpha^2 + c_2\beta^2, \ldots \tag{21}$$

являются решениями уравнения (15). Так как найденные прогрессии имеют разные знаменатели и потому непропорциональны, формула (21) при различных c_1 и c_2 дает все решения уравнения (15).

В частности, при некоторых c_1 и c_2 формула (21) должна дать и ряд Фибоначчи. Для этого, как указывалось выше, нужно определить c_1 и c_2 из уравнений

$$c_1 + c_2 = u_1$$

и

$$c_1\alpha + c_2\beta = u_2,$$

т. е. из системы

$$c_1 + c_2 = 1,$$
$$c_1 \frac{1+\sqrt{5}}{2} + c_2 \frac{1-\sqrt{5}}{2} = 1.$$

Решив эту систему, мы получаем

$$c_1 = \frac{1+\sqrt{5}}{2\sqrt{5}}, \quad c_2 = -\frac{1-\sqrt{5}}{2\sqrt{5}},$$

откуда

$$u_n = c_1\alpha^{n-1} + c_2\beta^{n-1} =$$
$$= \frac{1+\sqrt{5}}{2\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^{n-1} - \frac{1-\sqrt{5}}{2\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^{n-1},$$

т. о.

$$u_n = \frac{\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n}{\sqrt{5}}. \qquad (22)$$

Формула (22) называется *формулой Бине* (по имени математика, который ее вывел). Очевидно, подобные формулы можно указать и для других решений (15).

17. Мы видели, что $\alpha^2 = \alpha + 1$. Ясно поэтому, что любую целую положительную степень числа α можно представить в виде $a\alpha+b$ с целыми коэффициентами a и b. Так,

$$\alpha^3 = \alpha\alpha^2 = \alpha(\alpha+1) = \alpha^2 + \alpha = \alpha + 1 + \alpha = 2\alpha + 1,$$
$$\alpha^4 = \alpha\alpha^3 = \alpha(2\alpha+1) = 2\alpha^2 + \alpha = 2\alpha + 2 + \alpha = 3\alpha + 2$$

и т. д.

Покажем (по индукции), что

$$\alpha^n = u_n\alpha + u_{n-1}.$$

Действительно, для n=2, 3 это справедливо. Предположим, что

$$\alpha^k = u_k\alpha + u_{k-1},$$
$$\alpha^{k+1} = u_{k+1}\alpha + u_k.$$

Сложив эти равенства, получим

$$\alpha^k + \alpha^{k+1} = (u_k + u_{k+1})\alpha + (u_{k-1} + u_k),$$

или

$$\alpha^{k+2} = u_{k+2}\alpha + u_{k+1},$$

и индуктивный переход обоснован. Аналогично из $\beta^2 = \beta + 1$ следует

$$\beta^n = u_n\beta + u_{n-1}.$$

18. При помощи формулы Бине удобно суммировать многие ряды, связанные с числами Фибоначчи.

Найдем, например, чему равна сумма

$$u_3 + u_6 + u_9 + \ldots + u_{3n}.$$

Мы имеем

$$u_3 + u_6 + \ldots + u_{3n} = \frac{\alpha^3 - \beta^3}{\sqrt{5}} + \frac{\alpha^6 - \beta^6}{\sqrt{5}} + \ldots + \frac{\alpha^{3n} - \beta^{3n}}{\sqrt{5}} =$$

$$= \frac{1}{\sqrt{5}}(\alpha^3 + \alpha^6 + \ldots + \alpha^{3n} - \beta^3 - \beta^6 - \ldots - \beta^{3n})$$

или, суммируя встретившиеся нам геометрические прогрессии,

$$u_3 + u_6 + \ldots + u_{3n} = \frac{1}{\sqrt{5}}\left(\frac{\alpha^{3n+3} - \alpha^3}{\alpha^3 - 1} - \frac{\beta^{3n+3} - \beta^3}{\beta^3 - 1}\right).$$

Но

$$\alpha^3 - 1 = \alpha + \alpha^2 - 1 = \alpha + \alpha + 1 - 1 = 2\alpha,$$

и аналогично $\beta^2 - 1 = 2\beta$. Поэтому

$$u_3 + u_6 + \ldots + u_{3n} = \frac{1}{\sqrt{5}}\left(\frac{\alpha^{3n+3} - \alpha^3}{2\alpha} - \frac{\beta^{3n+3} - \beta^3}{2\beta}\right),$$

или, произведя сокращения,

$$u_3 + u_6 + \ldots + u_{3n} = \frac{1}{\sqrt{5}}\left(\frac{\alpha^{3n+2} - \alpha^2 - \beta^{3n+2} + \beta^2}{2}\right) =$$

$$= \frac{1}{2}\left(\frac{\alpha^{3n+2} - \beta^{3n+2}}{\sqrt{5}} - \frac{\alpha^2 - \beta^2}{\sqrt{5}}\right) = \frac{1}{2}(u_{3n+2} - u_2) = \frac{u_{3n+2} - 1}{2}$$

19. В качестве следующего примера применения формулы Бине вычислим сумму кубов первых n чисел Фибоначчи.

Заметим предварительно, что

$$u_k^3 = \left(\frac{\alpha^k - \beta^k}{\sqrt{5}}\right)^3 = \frac{1}{5}\,\frac{\alpha^{3k} - 3\alpha^{2k}\beta^k + 3\alpha^k\beta^{2k} - \beta^{3k}}{\sqrt{5}} =$$

$$= \frac{1}{5}\left(\frac{\alpha^{3k} - \beta^{2k}}{\sqrt{5}} - 3\alpha^k\beta^k\frac{\alpha^k - \beta^k}{\sqrt{5}}\right) =$$

$$= \frac{1}{5}\left(u_{3k} - (-1)^k 3u_k\right) = \frac{1}{5}\left(u_{3k} + (-1)^{k+1} 3u_k\right).$$

Поэтому

$$u_1^3 + u_2^3 + \ldots + u_n^3 =$$

$$= \frac{1}{5}\left((u_3 + u_6 + \ldots + u_{3n}) + 3(u_1 - u_2 + u_3 - \ldots + (-1)^{n+1} u_n)\right),$$

или, пользуясь формулой (18) и результатами предыдущего пункта,

$$u_1^3 + u_2^3 + \ldots + u_n^3 = \frac{1}{5}\left(\frac{u_{3n+2} - 1}{2} + (-1)^{n+1} 3u_{n-1} + 3\right) =$$

$$= \frac{u_{3n+2} + (-1)^{n+1} 6u_{n-1} + 5}{10}.$$

20. Поставим вопрос о том, как быстро растут числа Фибоначчи при увеличении их номеров. Формула Бине дает достаточно исчерпывающий овет и на этот вопрос.

Докажем следующую теорему.

Теорема. *Число Фибоначчи u_n есть ближайшее целое число к $\frac{\alpha^n}{\sqrt{5}}$, т. е. к n-му члену a_n геометрической прогрессии, первый член которой есть*

$$\frac{\alpha}{\sqrt{5}},$$

а знаменатель равен α.

Доказательство. Очевидно, достаточно установить, что абсолютная величина разности между u_n и a_n всегда меньше $\frac{1}{2}$. Но

$$|u_n - a_n| = \left|\frac{\alpha^n - \beta^n}{\sqrt{5}} - \frac{\alpha^n}{\sqrt{5}}\right| = \left|\frac{\alpha^n - \alpha^n - \beta^n}{\sqrt{5}}\right| = \frac{|\beta|^n}{\sqrt{5}}.$$

Так как $\beta = -0{,}618...$, то $|\beta| < 1$, а значит, $|\beta|^n < 1$ при любом n и тем более (так как $\sqrt{5} > 2$) должно быть $\frac{|\beta|^n}{\sqrt{5}} < \frac{1}{2}$. Теорема доказана.

Используя теорию пределов, легко сможет, несколько видоизменив доказательство этой теоремы, показать, что

$$\lim_{n \to \infty} |u_n - a_n| = 0.$$

Пользуясь доказанной теоремой, можно вычислять числа Фибоначчи при помощи таблиц логарифмов.

Вычислим, например, u_{14} (u_{14}, как легко сообразить, должно являться ответом задачи Фибоначчи о кроликах):

$$\sqrt{5} = 2{,}2361, \qquad \lg\sqrt{5} = 0{,}34949;$$

$$\alpha = \frac{1+\sqrt{5}}{2} = 1{,}6180, \qquad \lg\alpha = 0{,}20898;$$

$$\lg\frac{\alpha^{14}}{\sqrt{5}} = 14 \cdot 0{,}20898 - 0{,}34949 = 2{,}5762,$$

$$\frac{\alpha^{14}}{\sqrt{5}} = 376{,}9.$$

Ближайшим целым числом к 376,9 является 377; это и есть u_{14}.

При вычислении чисел Фибоначчи с большими номерами мы уже не сможем по таблицам логарифмов определить все цифры числа, а сможем указать только несколько первых цифр его, так что вычисление оказывается приближенным.

В виде упражнения читатель может доказать, что в десятичной системе счисления u_n при $n \geq 17$ имеет не более $\frac{1}{4}$ и не менее $\frac{n}{5}$ цифр. А из скольких цифр состоит u_{100}?

21. Результат предыдущего пункта можно уточнить. Следующая теорема пригодится нам в дальнейшем.

Теорема.

$$\frac{\alpha^{n-\frac{1}{n}}}{\sqrt{5}} \leqq u_n \leqq \frac{\alpha^{n+\frac{1}{n}}}{\sqrt{5}}.$$

Доказательство. Мы ограничимся доказательством левой стороны неравенства: правая доказывается аналогично. Поскольку согласно формуле Бине

$$u_n = \frac{1}{\sqrt{5}}(\alpha^n - \beta^n),$$

а $\alpha\beta = -1$, для наших целей будет достаточно показать, что

$$\alpha^{n-\frac{1}{n}} \leqq \alpha^n - \frac{1}{\alpha^n},$$

или

$$\alpha^{2n-\frac{1}{n}} \leqq \alpha^{2n} - 1,$$

или, возводя в степень n,

$$\alpha^{2n^2-1} \leqq (\alpha^{2n} - 1)^n. \qquad (23)$$

Будем доказывать это неравенство по индукции. При $n = 1$ оно превращается в

$$\alpha \leqq \alpha^2 - 1,$$

что действительно имеет место (именно, со знаком равенства). При $n = 2$ (23) означает

$$\alpha^7 \leqq (\alpha^4 - 1)^2. \tag{24}$$

Это неравенство можно проверить и прямым вычислением. Однако его можно и доказать, воспользовавшись соотношением, выведенным в п. 17. В данном случае мы имеем

$$\alpha^4 = 3\alpha + 2,$$

$$(\alpha^4 - 1)^2 = (3\alpha + 1)^2 = 9\alpha^2 + 6\alpha + 1 = 15\alpha + 10,$$

и (24) переписывается как

$$\alpha^7 = 13\alpha + 8 \leqq 15\alpha + 10,$$

что очевидно. Наконец, при $n = 3$ (23) переписывается как

$$\alpha^{17} \leqq (\alpha^6 - 1)^3,$$

что проверяется аналогично предыдущему.

Предположим теперь, что $n>2$ и (23) имеет место, и докажем, что

$$\alpha^{2(n+1)^2-1} \leqq (\alpha^{2n+2} - 1)^{n+1}.$$

Для этого достаточно показать, что при увеличении n на единицу правая часть (23) растет быстрее левой части. Но левая часть, очевидно, возрастает в α^{4n+2} раз. Оценим увеличение правой части.

Мы имеем

$$\frac{(\alpha^{2(n+1)} - 1)^{n+1}}{(\alpha^{2n} - 1)^n} = (\alpha^{2(n+1)} - 1)\left(\frac{\alpha^{2(n+1)} - 1}{\alpha^{2n} - 1}\right)^n.$$

Последняя дробь больше, чем α^2, и притом на

$$\frac{\alpha^{2(n+1)} - 1}{\alpha^{2n} - 1} - \alpha^2 = \frac{\alpha^{2n+2} - 1 - \alpha^{2n+2} + \alpha^2}{\alpha^{2n} - 1} = \frac{\alpha^2 - 1}{\alpha^{2n} - 1} =$$

$$= \frac{1}{\alpha^{2n-2} + \alpha^{2n-4} + \ldots + \alpha^2 + 1} > \frac{1}{\alpha^{2n-1}}.$$

Следовательно, пользуясь формулой бинома,

$$\left(\frac{\alpha^{2(n+1)} - 1}{\alpha^{2n} - 1}\right)^n > \left(\alpha^2 + \frac{1}{\alpha^{2n-1}}\right)^n = \alpha^{2n} + n\,\frac{\alpha^{2n-2}}{\alpha^{2n-1}} + \ldots,$$

где точки стоят вместо положительных слагаемых.

Ввиду того, что $n > 2$, написанное выражение больше, чем $\alpha^{2n} + 1$. Значит,

$$\frac{(\alpha^{2(n+1)}-1)^{n+1}}{(\alpha^{2n}-1)^n} > (\alpha^{2(n+1)}-1)(\alpha^{2n}+1) =$$
$$= \alpha^{4n+2} + \alpha^{2n+2} - \alpha^{2n} - 1 = \alpha^{4n+2} + \alpha^{2n}(\alpha^2-1) - 1 =$$
$$= \alpha^{4n+2} + \alpha^{2n+1} - 1 > \alpha^{4n+2},$$

и теорема доказана.

22. Рассмотрим еще один класс последовательностей, основанных на числах Фибоначчи. Пусть x — произвольное число. Вычислим сумму

$$s_n(x) = u_1 x + u_2 x^2 + \ldots + u_n x^n.$$

Для этого воспользуемся прежде всего формулой Бине:

$$s_n(x) = \frac{\alpha-\beta}{\sqrt{5}} x + \frac{\alpha^2-\beta^2}{\sqrt{5}} x^2 + \ldots + \frac{\alpha^n-\beta^n}{\sqrt{5}} x^n =$$
$$= \frac{1}{\sqrt{5}}(\alpha x + \alpha^2 x^2 + \ldots + \alpha^n x^n) -$$
$$- \frac{1}{\sqrt{5}}(\beta x + \beta^2 x^2 + \ldots + \beta^n x^n). \tag{25}$$

Здесь в скобках написаны суммы двух геометрических прогрессий со знаменателями αx и βx. Известная формула, выражающая сумму геометрической прогрессии, справедлива в том случае, когда знаменатель прогрессии отличен от единицы. Если же он ранен единице, то все члены прогрессии равны друг другу, и их сумма вычисляется совсем просто.

В соответствии со сказанным рассмотрим сначала случай, когда $\alpha x \neq 1$ и $\beta x \neq 1$, т. е. когда

$$x \neq \frac{1}{\alpha} = -\beta \text{ и } x \neq \frac{1}{\beta} = -\alpha.$$

В этих случаях, суммируя в (25) геометрические прогрессии, мы получаем

$$s_n(x) = \frac{1}{\sqrt{5}} \frac{\alpha^{n+1}x^{n+1} - \alpha x}{\alpha x - 1} - \frac{1}{\sqrt{5}} \frac{\beta^{n+1}x^{n+1} - \beta x}{\beta x - 1},$$

или, выполняя естественные преобразования,

$$s_n(x) = \frac{1}{\sqrt{5}} \frac{(\alpha^{n+1}x^{n+1} - \alpha x)(\beta x - 1) - (\beta^{n+1}x^{n+1} - \beta x)(\alpha x - 1)}{(\alpha x - 1)(\beta x - 1)}$$

и далее —

$$s_n(x) = \frac{1}{\sqrt{5}}\left(\frac{\alpha^{n+1}\beta x^{n+2} - \alpha^{n+1}x^{n+1} + \alpha x}{\alpha\beta x^2 - (\alpha+\beta)x + 1} - \right.$$
$$\left. - \frac{\alpha\beta^{n+1}x^{n+2} - \beta^{n+1}x^{n+1} + \beta x}{\alpha\beta x^2 - (\alpha+\beta)x + 1}\right).$$

Вспоминая, что

$$\alpha\beta = -1, \quad \alpha + \beta = 1, \quad \text{а} \quad \alpha - \beta = \sqrt{5},$$

имеем

$$s_n(x) = \frac{1}{\sqrt{5}} \frac{x\sqrt{5} - (\alpha^n - \beta^n)x^{n+2} - (\alpha^{n+1} - \beta^{n+1})x^{n+1}}{1 - x - x^2}$$

и окончательно

$$s_n(x) = \frac{x - u_n x^{n+2} - u_{n+1}x^{n+1}}{1 - x - x^2}. \qquad (26)$$

В частности, полагая в этой формуле x=1, получим

$$s_n(1) = u_1 + u_2 + \ldots + u_n = \frac{1 - u_n - u_{n+1}}{-1} = u_{n+2} - 1,$$

что соответствует сказанному в п. 1.

При x= — 1 имеем

$$s_n(-1) = u_1 - u_2 + \ldots + (-1)^{n-1}u_n =$$
$$= \frac{-1 - u_n(-1)^{n+2} - u_{n+1}(-1)^{n+1}}{-1} = (-1)^{n+1}u_{n-1} + 1$$

(ср. формулу (8)).

Рассмотрим теперь оставшиеся «особые» случаи.

Пусть $x = \frac{1}{\alpha} = -\beta$. Тогда в (25) каждый член первой прогрессии равен единице, и сумма этой прогрессии равна n. Во второй же прогрессии знаменатель оказывается равным —β^2,

Таким образом,

$$s_n\left(\frac{1}{\alpha}\right) = \frac{1}{\sqrt{5}}\left(n - \left(\beta^2 - \beta^4 + \ldots + (-1)^{n-1}\beta^{2n}\right)\right) =$$
$$= \frac{1}{\sqrt{5}}\left(n - \frac{\beta^2 - (-1)^n\beta^{2n+2}}{1+\beta^2}\right) =$$
$$= \frac{1}{\sqrt{5}}\left(n - \frac{\beta^2}{1+\beta^2} + (-1)^n\beta^{2n}\frac{\beta^2}{1+\beta^2}\right).$$

Замечая, что

$$1 + \beta^2 = 2 + \beta = 2 + \frac{1-\sqrt{5}}{2} = \frac{5-\sqrt{5}}{2},$$

а

$$\frac{\beta^2}{1+\beta^2} = \frac{1+\beta}{2+\beta} = \frac{3-\sqrt{5}}{5-\sqrt{5}} = \frac{(3-\sqrt{5})(5+\sqrt{5})}{(5-\sqrt{5})(5+\sqrt{5})} = \frac{10-2\sqrt{5}}{20},$$

мы получаем окончательно

$$s_n\left(\frac{1}{\alpha}\right) = \frac{n}{\sqrt{5}} - \frac{\sqrt{5}-1}{10} + (-1)^n \beta^{2n} \frac{\sqrt{5}-1}{10}. \qquad (27)$$

Наконец, пусть $x = \frac{1}{\beta}$. В этом случае в (25) единице равен знаменатель второй прогрессии, а знаменатель первой прогрессии равен — α^2. Мы имеем

$$s_n\left(\frac{1}{\beta}\right) = \frac{1}{\sqrt{5}}\left((\alpha^2 - \alpha^4 + \ldots + (-1)^{n-1}\alpha^{2n}) - n\right).$$

Аналогично предыдущему получаем

$$s_n\left(\frac{1}{\beta}\right) = \frac{1}{\sqrt{5}}\left(\frac{\alpha^2 - (-1)^n \alpha^{2n+2}}{1+\alpha^2} - n\right) = \frac{1}{\sqrt{5}}\left((-1)^{n+1}\alpha^{2n}\frac{\alpha^2}{1+\alpha^2} + \frac{\alpha^2}{1+\alpha^2} - n\right)$$

и в итоге

$$s_n\left(\frac{1}{\beta}\right) = (-1)^{n+1}\frac{1+\sqrt{5}}{10}\alpha^{2n} + \frac{1+\sqrt{5}}{10} - \frac{n}{\sqrt{5}}. \qquad (28)$$

23. Посмотрим, как ведет себя сумма $s_n(x)$ при фиксированном x и неограниченно возрастающем n.,

Переходя в равенстве (25) к пределу по n, получаем

$$\lim_{n\to\infty} s_n(x) = \lim_{n\to\infty}\frac{1}{\sqrt{5}}((\alpha x + \alpha^2 x^2 + \ldots + \alpha^n x^n) - (\beta x + \beta^2 x^2 + \ldots + \beta^n x^n)) =$$
$$= \frac{1}{\sqrt{5}}\lim_{n\to\infty}(\alpha x + \alpha^2 x^2 + \ldots + \alpha^n x^n) - \frac{1}{\sqrt{5}}\lim_{n\to\infty}(\beta x + \beta^2 x^2 + \ldots + \beta^n x^n).$$

Здесь под знаками двух последних пределов стоят суммы геометрических прогрессий. Поэтому сами пределы являются суммами соответствующих бесконечных геометрических прогрессий. Но, как известно, для того, чтобы можно было говорить о сумме бесконечной геометрической прогрессии, необходимо и достаточно, чтобы ее знаменатель по абсолютной величине был меньше единицы. В

имеющихся у нас прогрессиях знаменатели равны αx и βx. Здесь $|\alpha|>|\beta|$ Поэтому из $|\alpha x|<1$ следует $|\beta x|<1$. Таким образом, выполнение неравенства $|\alpha x|<1$ будет обеспечивать существование всех интересующих нас в данный момент пределов. Итак, предел

$$\lim_{n\to\infty} s_n(x) \tag{29}$$

существует, если $|x| < \frac{1}{\alpha}$. Обозначим этот предел через $s(x)$. Для его вычисления мы можем воспользоваться формулой (26).

Заметим для этого, что на основании сказанного в п. 20

$$u_n \leqq \frac{\alpha^n}{\sqrt{5}} + 1.$$

Поэтому

$$\lim_{n\to\infty} u_n x^{n+2} \leqq \lim_{n\to\infty} \left(\frac{\alpha^n}{\sqrt{5}} + 1\right) x^{n+2} =$$

$$= \frac{x^2}{\sqrt{5}} \lim_{n\to\infty} (\alpha x)^n + \lim_{n\to\infty} x^{n+2}.$$

Ввиду $|\alpha x|<1$ должно быть и $|x|<1$, так что оба написанных предела равны нулю. По тем же причинам и

$$\lim_{n\to\infty} u_{n+1} x^{n+1} = 0.$$

Следовательно, переходя в формуле (26) к пределу по n при неограниченном возрастании n, получаем

$$s(x) = \lim_{n\to\infty} s_n(x) = \lim_{n\to\infty} \frac{x - u_n x^{n+2} - u_{n+1} x^{n+1}}{1 - x - x^2} =$$

$$= \frac{1}{1 - x - x^2} \left(x - \lim_{n\to\infty} u_n x^{n+2} - \lim_{n\to\infty} u_{n+1} x^{n+1}\right) = \frac{x}{1 - x - x^2}.$$

Найденный результат можно переписать в развернутом виде как

$$u_1 x + u_2 x^2 + \ldots + u_n x^n + \ldots = \frac{x}{1 - x - x^2}. \tag{30}$$

Придавая переменной x те или иные значения, будем получать различные конкретные формулы. Например, полагая $x = \frac{1}{2}$, обнаружим, что

$$\frac{u_1}{2} + \frac{u_2}{2^2} + \ldots + \frac{u_n}{2^n} + \ldots = 2.$$

24. Формулу (30) можно получить также при помощи несколько иных рассуждений. Напишем

$$u_1x + u_2x^2 + \ldots + u_nx^n + \ldots = s(x) \qquad (31)$$

(помня при этом, что выражение $s(x)$ имеет смысл лишь при $|x| < \frac{1}{\alpha}$)
и умножим это равенство почленно на x и на x^2:

$$u_1x^2 + u_2x^3 + \ldots + u_nx^{n+1} + \ldots = xs(x), \qquad (32)$$

$$u_1x^3 + u_2x^4 + \ldots + u_nx^{n+2} + \ldots = x^2s(x). \qquad (33)$$

Вычитая из равенства (31) оба равенства (32) и (33), мы после приведения подобных членов получим

$$u_1x + (u_2 - u_1)x^2 + (u_3 - u_2 - u_1)x^3 + (u_4 - u_3 - u_2)x^4 + \ldots + (u_n - u_{n-1} - u_{n-2})x^n + \ldots = (1 - x - x^2)s(x).$$

Все заключенные в скобках выражения в левой части равенства, кроме первого равны нулю, и это равенство превращается в

$$x = (1 - x - x^2)s(x),$$

откуда и следует (30).

25. Говоря о числе Фибоначчи u_n, мы пока все время предполагали, что его номер n является целым положительным числом. Однако основное рекуррентное соотношение, определяющее числа Фибоначчи, может быть записано и как

$$u_{n-2} = u_n - u_{n-1}. \qquad (34)$$

При этом оно будет служить для выражения чисел Фибоначчи с меньшими номерами через числа с большими.

Полагая последовательно в (34) n = 2, 1, 0, — 1, ..., мы можем вычислить

$$u_0 = 0, \qquad u_{-1} = 1, \qquad u_{-2} = -1, \qquad u_{-3} = 2, \ldots$$

и вообще, как легко убедиться (пусть читатель убедится сам!),

$$u_{-n} = (-1)^{n+1}u_n. \qquad (35)$$

Это простое выражение числа Фибоначчи с произвольным целым номером позволяет сводить все задачи о таких числах Фибоначчи к задачам об обычных числах Фибоначчи с натуральными номерами.

Например, для вычисления суммы n «первых назад» чисел Фибоначчи

$$u_{-1} + u_{-2} + \ldots + u_{-n}$$

достаточно переписать ее в соответствии с (35):

$$u_1 - u_2 + \ldots + (-1)^{n-1}u_n,$$

и вспомнить формулу (8):

$$u_{-1} + u_{-2} + \ldots + u_{-n} = (-1)^{n+1}u_{n-1} + 1 = -u_{-n+1} + 1.$$

Опирающееся на основное рекуррентное соотношение индуктивное рассуждение о числах Фибоначчи типа переходов «от n и n+1 к n+2» можно в связи с соотношением (34) проводить по схеме «от n и n — 1 к n — 2». В частности, таким образом без труда доказывается для любых целых n и m формула (10)

$$u_{n+m} = u_{n-1}u_m + u_n u_{m+1}.$$

26. Основные уравнения для чисел α и β:

$$\alpha^{n+2} = \alpha^n + \alpha^{n+1},$$
$$\beta^{n+2} = \beta^n + \beta^{n+1},$$

справедливы не только для положительных, но и для любых целых значений n (для дробных значений n эти равенства тоже в известном смысле остаются в силе, но мы на этом не будем останавливаться). Отсюда легко получить, что формула Бине

$$u_n = \frac{\alpha^n - \beta^n}{\sqrt{5}}$$

имеет место для любого целого n.

Заметим в заключение, что и результат п. 17 можно (по индукции «назад») перенести на отрицательные значения номера:

$$\alpha^{-n} = u_{-n}\alpha + u_{-n-1}. \qquad (36)$$

Это равенство переписывается как

$$(-1)^n\beta^n = (-1)^n u_n \frac{1}{\beta} + (-1)^n u_{n+1},$$

т. е.

$$\beta^{n+1} = u_{n+1}\beta + u_n.$$

Кроме того, (36) можно представить в виде

$$\alpha^{-n} = (-1)^{n-1} u_n\alpha + (-1)^n u_{n+1},$$

т. е.

$$(-1)^n \alpha^{-n} = u_{n+1} - u_n\alpha,$$

или, иначе,

$$\frac{u_{n+1}}{u_n} - \alpha = (-1)^n \alpha^{-n} \frac{1}{u_n}. \qquad (37)$$

27. Числа Фибоначчи могут составить основу своеобразной «фибоначчиевой» системы счисления, т. е., представления любого натурального числа a в виде некоторой последовательности «цифр» $\varphi_1\varphi_2 \ldots \varphi_r$. Эта последовательность может быть получена следующим (индуктивным!) образом.

Отнимем от заданного числа $a=a_0$ наибольшее из не превосходящих его чисел Фибоначчи u_n и, написав цифру $\varphi_1 = 1$ и разность $a_1 = a_0 - u_n$ будем считать это первым шагом нашего построения.

Предположим, что k шагов построения уже выполнены, в результате чего появилась последовательность цифр

$$\varphi_1\varphi_2 \ldots \varphi_k \qquad (38)$$

состоящая из нулей и единиц, а также некоторое число n_k. Тогда k+1-й шаг построения будет состоять в следующем: сравним число a_k с числом Фибоначчи u_{n-k}, и если окажется $a_k < u_{n-k}$, то припишем к последовательности (38) $\varphi_{k+1} = 0$ и фиксируем число $a_{k+1} = a_k$, а если будет $a_k \geq u_{n-k}$, то припишем к (38) $\varphi_{k+1}=1$ и положим $a_{k+1} = a_k - u_{n-k}$. Мы выполним n — 1 шагов этого процесса, в результате чего, очевидно, придем к n—1-членной последовательности (38) и числу $a_{n-1} = 0$.

Фактически описанный процесс является последовательным выделением из числа a слагаемых, равных наибольшим возможным числам Фибоначчи, т. е. представлением a в виде суммы различных чисел Фибоначчи.

Окончательную соответствующую числу a последовательность (38) будем называть его *фибоначчиевой записью* и обозначать через $\Phi(a)$. Составляющие $\Phi(a)$ нули и единицы назовем *фибоначчиевыми цифрами* числа a. Ясно, что если $\Phi(a) = \varphi_1\varphi_2 \ldots \varphi_{n-1}$,то

$$a = u_n\varphi_1 + u_{n-1}\varphi_2 + \ldots + u_2\varphi_{n-1}. \qquad (39)$$

Поясним сказанное на примере. Пусть a = 19.

Тогда

$$\begin{aligned}
&u_n = 13 \ (n = 7), && \varphi_1 = 1, && a_1 = 19 - 13 = 6;\\
&u_6 = 8 > a_1, && \varphi_2 = 0, && a_2 = a_1 = 6;\\
&u_5 = 5 \leqq a_2, && \varphi_3 = 1, && a_3 = 6 - 5 = 1;\\
&u_4 = 3 > a_3, && \varphi_4 = 0, && a_4 = a_3 = 1;\\
&u_3 = 2 > a_4, && \varphi_5 = 0, && a_5 = a_4 = 1;\\
&u_2 = 1 \leqq a_5, && \varphi_6 = 1, && a_6 = 1 - 1 = 0.
\end{aligned}$$

Таким образом, $\Phi(19)$= 101001, и 19 = $u_7 + u_5 + u_2$.

Ясно, что каждое число a имеет единственную фибоначчиеву запись $\Phi(a)$. Однако не всякая начинающаяся с единицы последовательность нулей и единиц обязана быть фибоначчиевой записью $\Phi(a)$ для некоторого числа a. Например, в $\Phi(a)$ не могут стоять две единицы подряд.

Действительно, пусть в $\Phi(a)$ две единицы встречаются подряд после некоторого нуля:

$$\varphi_k = 0, \ \varphi_{k+1} = \varphi_{k+2} = 1.$$

Это значит, что

$$a_{k-1} < u_{n-k+1}, \tag{40}$$

$$a_{k-1} = a_k, \quad a_k - u_{n-k} = a_{k+1}, \quad a_{k+1} \geqq u_{n-k-1}.$$

Но почленное сложение всех составляющих вторую строку соотношений дает нам

$$a_{k-1} \geqq u_{n-k} + u_{n-k-1} = u_{n-k+1},$$

что противоречит (40).

Значит, две единицы подряд могли бы встретиться в Ф*(а)* лишь в том случае, когда впереди них вовсе не было бы нулей, т. е. было бы $\varphi_1=\varphi_2=1$. Но тогда по определению процесса $a_1 = a_0 - u_n \geqq$ $\geqq u_{n-1}$, так что

$$a_0 \geqq u_n + u_{n-1} = u_{n+1},$$

и u_n не является наибольшим числом Фибоначчи, не превосходящим a.

Вместе с тем ограничение, заключающееся в отсутствии двух стоящих рядом единиц, оказывается уже достаточным: всякая последовательность из n— 1 нулей и единиц, начинающаяся с единицы и не содержащая двух единиц подряд, есть фибоначчиева запись Ф*(а)* некоторого числа a, для которого

$$u_n \leqq a < u_{n+1}. \tag{41}$$

В этом можно убедиться, воспользовавшись, например, результатом задачи о прыгуне из п. 6. Пусть имеется n — 1 клетка, по которым прыгун прыгает доступными для него способами (т. е. на соседнюю клетку или через клетку). После выполнения прыжков все клетки, в которых прыгун побывал, помечаются нулями, а остальные клетки — единицами. Так как всего возможно u_{n-1} способов выполнения прыжков, различных способов пометок будет тоже u_{n-1}.

Если к каждой из них приписать впереди единицу, то мы получим запись, которая может быть фибоначчиевым значением для числа a, удовлетворяющего (40). Но таких чисел ровно u_{n-1} и каждое из них должно иметь свою запись; поэтому каждой записи соответствует хотя бы одно число.

14. Интерполяция, сглаживание, аппроксимация

14.1. Задачи интерполяции, сглаживания, аппроксимации

Одной из задач теории оптимизации является оптимизация графического представления кривых и поверхностей. Ниже будут рассмотрены некоторые математические методы, из которых пользователь может выбрать наиболее подходящие для решения конкретных задач оптимизации с учетом имеющихся ресурсов вычислительной техники.

Мы рассмотрим три типа задач: интерполяцию, сглаживание и аппроксимацию. Введем следующие обозначения:

Ω - n-мерное множество, принадлежащее пространству R^n ($n \geq 1$);

$\{\delta_i\}_{i \in I}$ - множество функционалов, т.е. отображений, которые каждой функции f, определенной на Ω, ставят в соответствие действительное число $\delta_i(f)$. Наиболее типичный пример: для данной точки P пространства R^n функционал δ_n будет равен $\delta_n(f)=f(P)$ - значению f в точке P; $\{z_i\}_{i \in I}$ - заданное множество действительных чисел; V-множество функций, определенных на Ω, которые могут быть функциями интерполяции, сглаживания или аппроксимации (полиномы, кусочно-полиномиальные и другие функции).

Уточним постановку задач, рассматриваемых в этом разделе.

Задача интерполяции

Найти элемент v из V, такой что для всех i из I

$$\delta_i(v) = z_i.$$

Задача сглаживания

Найти элемент v из V, такой что множество величин $\{\delta_i(v)\}_{i \in I}$ не очень удалено от множества $\{z_i\}_{i \in I}$. Критерий удаления может быть математически выражен достаточно строго (при использовании B-сплайнов или функций Безье), но всегда полезно его уточнить в общем виде. Он определяется через полунорму φ на R^1, и задачу можно сформулировать так: найти элемент v из V, такой что полунорма

$$\varphi(\{\delta_i(v)\}_{i \in I} - \{z_i\}_{i \in I})$$

будет минимальной. (Полунорма φ отличается от нормы тем, что в ней из условия $\varphi(a) = 0$ не следует равенство $a = 0$).

Задача аппроксимации

Исходные данные для задачи:

W - множество функций, определенных на Ω со значениями в *R; f* -элемент *W; V*⊂ *W;* φ - полунорма, определенная в *W.*

Задача состоит в том, чтобы найти элемент *v* из *V,* такой что полунорма φ *(f— v)* будет минимальной.

Для функций одной переменной мы приводим классическое описание решений перечисленных задач. Это поможет читателю оценить метод, который он предполагает использовать, и обратиться к другим разделам этой работы или к специальной литературе для применения в конкретном случае. Мы представили широкий спектр существующих методов и подробно рассмотрели вопросы линейности, выбора базиса, геометрических свойств данных (в частности, в случае поверхностей).

Линейность

Будем отдавать предпочтение методам, которые обладают одним из следующих свойств:

- результат линейно зависит от данных;
- решение задачи сводится к решению системы линейных уравнений.

Выбор базисных функций

Речь идет о корректном выборе параметров множества функций *V,* на котором отыскивается решение. Эти параметры неявно определяют выбор базиса и могут изменяться в процессе решения задачи.

Геометрические свойства данных

В тех случаях, когда для решения задачи требуется выполнить разбиение области данных, равномерное разбиение может существенно упростить реализацию метода, работу с данными, вычисление функционалов и т. д.

14.2. Кривые

В методах, к описанию которых мы приступаем, обычно используется ортонормированный базис (на плоскости или в пространстве). Будем предполагать, если не оговорено противоположное, что рассматриваемые кривые являются однозначными функциями в этом базисе: одной абсциссе *t* соответствует одна и только одна ордината *f* (*t*).

14.2.1. Интерполяция

Интерполяционная формула Лагранжа. Ранее мы познакомились с линейной интерполяцией, которая состоит в приближенной замене рассматриваемой функции *y=f(x)* линейной функцией *y=ax+b,*

совпадающей с $f(x)$ в некоторых двух точках. Очевидно, что если вместо линейной функции использовать многочлен n-й степени

$$P(x) = P_n(x) = a_0 x^n + a_1 x^{n-1} + \ldots + a_{n-1} x + a_n,$$

то точность такой замены можно повысить.

Так как многочлен $P_n(x)$, *аппроксимирующий* (приближающий, приближенно заменяющий) функцию $f(x)$, содержит $n+1$ параметров, которыми служат коэффициенты, то при его подборе можно поставить $n+1$ условий. Рассмотрим для простоты случай многочлена второй степени

$$P(x) = P_2(x) = ax^2 + bx + c$$

(общий случай разбирается аналогично); при его подборе можно поставить три условия. Часто *требуют, чтобы этот многочлен совпадал с функцией f в некоторых трех заданных точках:*

$$P(x_1) = f(x_1); \quad P(x_2) = f(x_2); \quad P(x_3) = f(x_3). \qquad (1)$$

Эти значения также считаются известными.

Прежде всего ясно, что искомый многочлен *может быть только один:* если бы другой многочлен второй степени $Q(x)$ удовлетворял условиям (1), то разность $P(x) - Q(x)$ — также многочлен второй степени — равнялась бы нулю при $x = x_1$; $x = x_2$; $x = x_3$. Но если многочлен второй степени равен нулю в трех точках, то он равен нулю тождественно, т. е. все коэффициенты равны нулю; итак, $Q(x) \equiv P(x)$.
Лагранж предложил искать многочлен $P(x)$ в форме

$$P(x) = A(x-x_2)(x-x_3) + B(x-x_1)(x-x_3) + C(x-x_1)(x-x_2), \qquad (2)$$

где A, B, C — постоянные, пока неизвестные. Ясно, что это — многочлен второй степени. Для выбора постоянных A, B, C воспользуемся условиями (1), заметив, что при подстановке $x = x_1$, x_2 или x_3 в правой части формулы (2) остается лишь одно слагаемое. Получим

$$f(x_1) = A(x_1-x_2)(x_1-x_3), \quad f(x_2) = B(x_2-x_1)(x_2-x_3),$$
$$f(x_3) = C(x_3-x_1)(x_3-x_2).$$

Найдя отсюда A, B, C и подставляя их в (2), получим *интерполяционную формулу Лагранжа*

$$f(x) \approx P_2(x) = f(x_1)\frac{(x-x_2)(x-x_3)}{(x_1-x_2)(x_1-x_3)} + f(x_2)\frac{(x-x_1)(x-x_3)}{(x_2-x_1)(x_2-x_3)} + f(x_3)\frac{(x-x_1)(x-x_2)}{(x_3-x_1)(x_3-x_2)}. \qquad (3)$$

При применении этой формулы желательно, чтобы ни одна из разностей $x_1—x_2$, $x_1— x_3$, $x_2—x_3$ не была чрезмерно малой. Взамен (1) можно поставить, например, такие три условия:

$$P(x_1) = f(x_1); \quad P'(x_1) = f'(x_1); \quad P(x_2) = f(x_2).$$

Тогда многочлен $P(x)$ вместо (2) можно искать в виде

$$P(x) = A(x - x_2)(x - 2x_1 + x_2) + B(x - x_1)(x - x_2) + C(x - x_1)^2.$$

Интерполяционные формулы Ньютона. Если расстояние h между соседними значениями x, при которых задается функция f, является постоянным, то можно пользоваться более удобными формулами, чем формула (3). Пусть, например, известны значения

$$f(x_0) = y_0; \quad f(x_1) = y_1; \quad f(x_2) = y_2; \quad f(x_3) = y_3,$$

где $X_1 = x_o + h$, $x_2 = x_0$-\-$2h$, $x_3 = x_0$-{-$3h$. Тогда многочлен $P(x)$, принимающий те же значения при указанных значениях x, будет иметь третью степень (см. п. 6). Ньютон предложил искать его в виде

$$P(x) = A + Bs + Cs(s - h) + Ds(s - h)(s - 2h), \qquad (4)$$

где $s = x — x_0$. Согласно условию должно быть

$$y_0 = P(x_0) = P|_{s=0} = A; \quad y_1 = P(x_1) = P|_{s=h} = A + Bh;$$
$$y_2 = P(x_2) = P|_{s=2h} = A + B \cdot 2h + C \cdot 2h^2;$$
$$y_3 = P(x_3) = P|_{s=3h} = A + B \cdot 3h + C \cdot 3 \cdot 2h^2 + D \cdot 3 \cdot 2h^3.$$

Составляя разности для левых и правых частей, получим

$$\Delta y_0 = Bh; \quad \Delta y_1 = Bh + C \cdot 2h^2; \quad \Delta y_2 = Bh + C \cdot 2 \cdot 2h^2 + D \cdot 3 \cdot 2h^3.$$

Вторично составляя разности, а затем и в третий раз, найдем

$$\Delta^2 y_0 = C \cdot 2h^2; \quad \Delta^2 y_1 = C \cdot 2h^2 + D \cdot 3 \cdot 2h^2; \quad \Delta^3 y_0 = D \cdot 3 \cdot 2h^2.$$

Отсюда находим

$$A = y_0; \; B = \frac{\Delta y_0}{h}; \; C = \frac{\Delta^2 y_0}{2!h^2}; \; D = \frac{\Delta^3 y_0}{3!h^3}.$$

Подставляя эти значения в oamcinu, что вместо x_0 можно было начинать от любого табличного значения x_k, получим *формулу Ньютона*

$$f(x) \approx P(x) =$$
$$= y_k + \Delta y_k \frac{s}{h} + \frac{\Delta^2 y_k}{2!} \frac{s}{h}\left(\frac{s}{h} - 1\right) + \frac{\Delta^3 y_k}{3!} \frac{s}{h}\left(\frac{s}{h} - 1\right)\left(\frac{s}{h} - 2\right), \qquad (5)$$

Аналогичный вид имеет формула для интерполяционных многочленов других степеней. Увеличивая эту степень, можно перейти к бесконечному ряду

$$f(x) = y_k + \Delta y_k \frac{s}{h} + \frac{\Delta^2 y_k}{2!} \frac{s}{h}\left(\frac{s}{h} - 1\right) + \\ + \frac{\Delta^3 y_k}{3!} \frac{s}{h}\left(\frac{s}{h} - 1\right)\left(\frac{s}{h} - 2\right) + \ldots, \quad (6)$$

причем не выписанные члены содержат соответственно разности четвертого, пятого и т. д. порядков и потому имеют четвертый, пятый и т. д. порядок малости по сравнению с шагом *h*. На практике эту формулу, конечно, обрывают, доводя ее до места, начиная с которого слагаемыми можно пренебречь. Если шаг велик или если мы находимся вблизи от конца интервала, на котором задана функция *f(x),* то может оказаться, что такого места достичь нельзя: тогда и формулой (6) пользоваться нельзя.

Формулы Ньютона (5) или (6) легко применять, если функция *f* задана таблично, так как в этом случае легко подсчитывать разности. Особенно часто они применяются *в начале таблицы* (например, если $k = 0$, x_0 — первое табличное значение аргумента, а $x_0 < x < x_1$). Степень интерполяционного многочлена *P(x)* выбирают, руководствуясь значениями разностей; например, если третьи разности очень малы, то последний член в формуле (5) мал и его можно отбросить, т. е. ограничиться многочленом второй степени. В формуле (6) можно положить и $k = 0$, $x < x_0$, если $|x - x_0|$ невелико, что приведет *к экстраполяции таблицы назад.* Подобно (6) выводится *другая формула Ньютона*:

$$f(x) = y_{k+1} - \Delta y_k \frac{t}{h} + \frac{\Delta^2 y_{k-1}}{2!} \frac{t}{h}\left(\frac{t}{h} - 1\right) - \\ - \frac{\Delta^3 y_{k-2}}{3!} \frac{t}{h}\left(\frac{t}{h} - 1\right)\left(\frac{t}{h} - 2\right) + \ldots, \quad (7)$$

где $t = x_{k+1} - x$, которая применяется, в частности, *в конце таблицы,* например, если x_{k+1} — последнее табличное значение аргумента, а $x_k < x < x_{k+1}$. Эта же формула применяется для *экстраполяции таблицы вперед.*

При интерполяции в середине таблицы желательно иметь формулу, использующую в равной мере табличные значения функции как впереди, так и позади рассматриваемого значения *x.* Одной из таких формул служит *формула Бесселя,* которая получается, если взять полусумму правых частей (6) и (7):

$$f(x) = \frac{y_k + y_{k+1}}{2} + \Delta y_k \left(\frac{s}{h} - \frac{1}{2}\right) + \\ + \frac{\Delta^2 y_{k-1} + \Delta^2 y_k}{2 \cdot 2!} \frac{s}{h}\left(\frac{s}{h} - 1\right) + \frac{\Delta^3 y_{k-1}}{3!} \frac{s}{h}\left(\frac{s}{h} - 1\right)\left(\frac{s}{h} - \frac{1}{2}\right) + \ldots, \quad (8)$$

где $s = x—x_k$. Эта формула, обладающая высокой точностью, названа по имени немецкого астронома Ф. Бесселя (1784—1846), хотя фактически она принадлежит Ньютону.

Интерполяционные формулы применяются также к задаче *обратного интерполирования,* которая состоит в отыскании значения аргумента по заданному значению функции. Будем исходить, например, из формулы (5). Приняв равенство за точное, эту формулу можно разрешить относительно второго слагаемого в правой части, что после деления на Δy_k даст

$$\frac{s}{h} = \frac{y-y_k}{\Delta y_k} - \frac{\Delta^2 y_k}{2!\,\Delta y_k}\,\frac{s}{h}\left(\frac{s}{h}-1\right) - \frac{\Delta^3 y_k}{3!\,\Delta y_k}\,\frac{s}{h}\left(\frac{s}{h}-1\right)\left(\frac{s}{h}-2\right). \quad (9)$$

Если задано y, то для нахождения s можно применить метод последовательных приближений. Для этого в качестве нулевого приближения можно положить $\left(\frac{s}{h}\right)_0 = \frac{y-y_k}{\Delta y_k}$; подставив это значение в правую часть (9), находим

$$\left(\frac{s}{h}\right)_1$$

и т. д. При малом h процесс хорошо сходится.

При интерполяции разрывной функции или функции с разрывной производной надо иметь в виду, что вблизи точек разрыва качество аппроксимации может значительно понизиться, так как интерполяционный многочлен разрывов не имеет. Для имитации разрыва можно значительно сближать узлы интерполяции вблизи точек разрыва, но обычно предпочитают проводить интерполяцию только на интервалах между точками разрыва.

Интерполяционные полиномы Лагранжа

Исходные данные:

1) $\Omega = [a, b] \subset R$;

2) $\{\delta_i\}_{i \in I}$; $I = \{0, 1, \ldots, m\}$, $\delta_i(f) = f(t_i)$,

где $a \leqslant t_0 < t \ldots < t_m \leqslant b$;

3) $z_i \in R$, $i \in I$;

4) $V = P^m$ - множество полиномов степени, меньшей или равной m, т.е. V- векторное пространство размерностью $m + 1$.

Первым этапом решения задачи интерполяции является выбор базиса для V.

Очевидно, в качестве простейшей базы можно взять степенные функции вида $1, t, t^2, \ldots, t^m$, т. е. задача сводится к нахождению полинома

$$v(t) = \sum_{j=0}^{m} v_j t^j,$$

такого, что $\delta_i(v) = v(t_i) = z_i, \quad i \in I$.

Для вычисления коэффициентов v_j требуется решить систему $m + 1$ линейных уравнений с $m + 1$ неизвестными:

$$\sum_{j=0}^{m} v_j t_i^j = z_i, \quad i = 0, 1, \ldots, m.$$

Матрица этой системы $A = (a_{ij})$, где $a_{ij} = t_i^j$, называется матрицей Ван-дермонда. Поскольку все t_i различны, она является обратимой; следовательно, задача допускает единственное решение, но вычисление коэффициентов довольно трудоемко: оно требует решения системы или обращения матрицы A, не содержащей нулевых элементов.

Можно показать, что выбор одночленного базиса не является лучшим. На практике предпочтение отдается другим базисным полиномам, использование которых приводит к простым системам уравнений.

Базисные полиномы Лагранжа. Возьмем в качестве базиса P^m полиномы Лагранжа

$$L_i(t) = \prod_{\substack{j=0 \\ j \neq i}}^{m} \frac{t - t_j}{t_i - t_j}, \quad i \in I.$$

Легко убедиться, что в этом базисе решением является

$$v(t) = \sum_{i=0}^{m} z_i L_i(t),$$

поскольку $L_i(t_i) = 1$ и $L_i(t_j) = 0$ для $i \neq j$ (т.е. матрица системы диагональна). Для определения значения функции v в произвольной точке t необходимо предварительно вычислить значения всех L_i в этой точке.

Если обозначить $\Pi(t) = \prod_{j=0}^{m} (t - t_j)$, то

$$L_i(t) = \frac{1}{t - t_i} \frac{\Pi(t)}{\Pi'(t_i)}.$$

Базис Лагранжа удобен, когда нужно строить большое число интерполяционных полиномов для разных исходных данных z, на одних и тех же точках t_i. Полиномы L_i зависят только от t_i, поэтому достаточно построить таблицы L_i и ограничиться их однократным вычислением.

Базисные полиномы Ньютона. Рассмотрим базис P^m, состоящий из полиномов Ньютона:

$$N_0(t) = 1, \quad N_i(t) = \prod_{j=0}^{i-1} (t - t_j), \quad i = 1, \ldots, m.$$

Легко показать, что система уравнений в этом случае имеет нижнюю треугольную матрицу. Действительно, из вида интерполяционного полинома

$$v(t) = \sum_{j=0}^{m} c_j N_j(t)$$

следует система уравнений для вычисления коэффициентов c_j

$$v(t_i) = \sum_{j=0}^{m} c_j N_j(t_i) = \sum_{j=0}^{i} c_j N_j(t_i) = z_i, \quad i \in I,$$

$$N_j(t_i) = 0 \text{ для } j > i.$$

Решение системы записывается в виде

$$c_0 = z_0, \quad c_i = [z_i - \sum_{j=0}^{i-1} c_j N_j(t_i)]/N_i(t_i), \quad i = 1, \ldots, m.$$

Величины c_i называются разделенными разностями порядка i на значениях $z_0, \ldots, z_i$.

Важное преимущество этого базиса состоит в том, что очень просто добавляются новые точки интерполяции, при этом все вычисленные ранее полиномы N_i и их коэффициенты не меняются. Это следует из классического определения разделенных разностей. Для заданных точек $t_0, \ldots, t_m$ и функции f разделенная разность определяется с помощью рекуррентных соотношений:

$$d[t_i] = f(t_i), \quad d[t_i, t_j] = \frac{f[t_i] - f[t_j]}{t_i - t_j}, \quad i \neq j,$$

$$d[t_0, \ldots, t_m] = \frac{d[t_0, \ldots, t_{m-2}, t_{m-1}] - d[t_0, \ldots, t_{m-2}, t_m]}{t_{m-1} - t_m},$$

где

$$d[t_0, \ldots, t_m]$$

- разделенная разность порядка m. Заметим, что значения разделенных разностей не зависят от нумерации точек t_i. Если σ обозначает множество точек и x и y - две дополнительные точки, то

$$d[\sigma, x, y] = \frac{d[\sigma, x] - d[\sigma, y]}{x - y}.$$

Следует отметить однако, что добавление новой точки приводит к появлению значительных осцилляции (рис. 1-3).

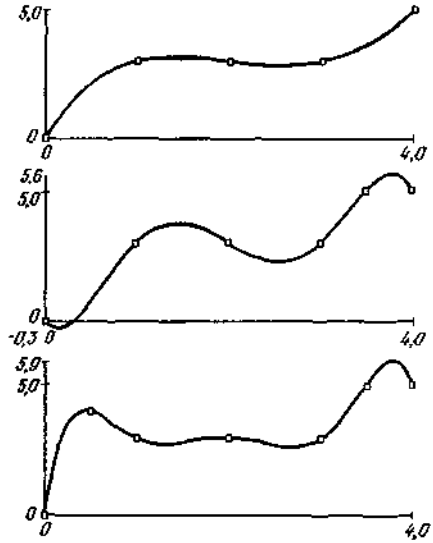


Рис. 1. Интерполяция с помощью полиномов Ньютона на 5-7 точках.

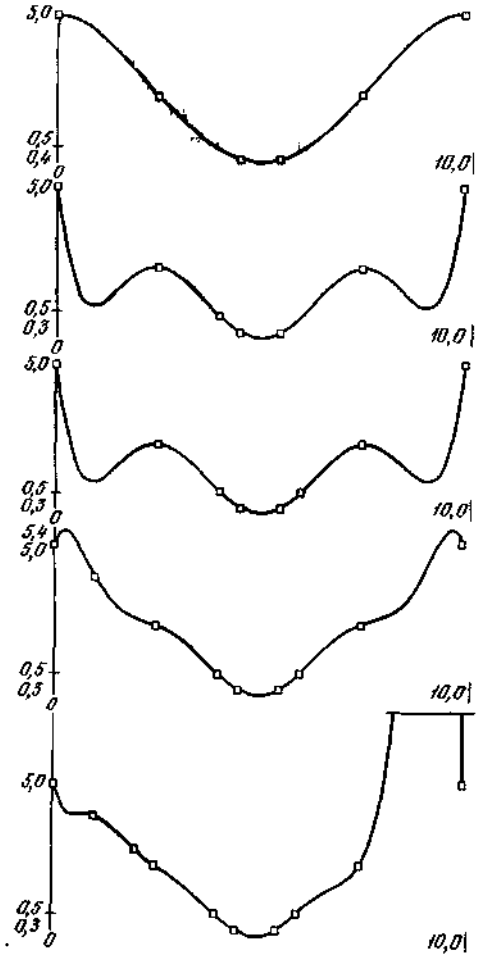


Рис. 2. Интерполяция с помощью полиномов Ньютона на 6-10 точках. Положение точек задается функцией $|x - 5|$

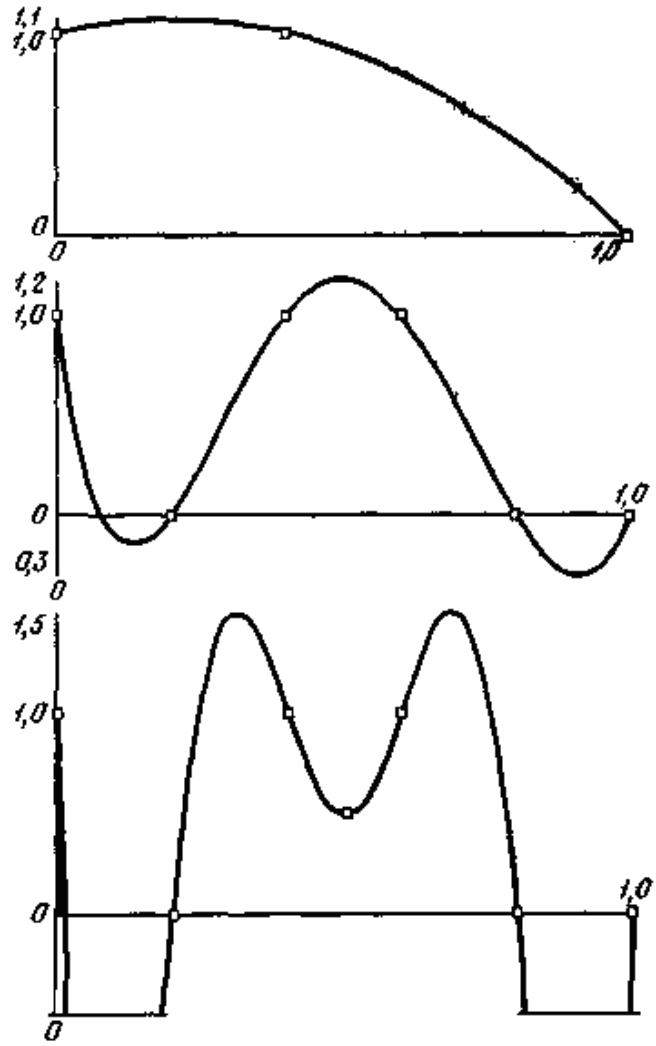


Рис. 3. Интерполяция с помощью полиномов Ньютона на 3, 6 и 7 точках. Добавление одной точки приводит к появлению значительных осцилляции

Итак, мы двумя способами построили интерполяционный полином Лагранжа. Очевидно, выбор базисных полиномов оказывает существенное влияние на последующие вычисления. Из двух рассмотренных методов обычно отдают предпочтение базисным

полиномам Ньютона с нахождением коэффициентов с помощью разделенных разностей. Для вычисления интерполяционного полинома на большом числе точек следует использовать интерполяционные сплайны.

Погрешность и сходимость. При решении задачи интерполяции нельзя упускать из виду такие важные свойства применяемого метода, как погрешность и сходимость. Дадим их общее определение и проиллюстрируем на примере интерполяционных полиномов Лагранжа. Воспользуемся прежними обозначениями Ω, $\{\delta_i\}_{i \in I}$, V и др. Обозначим через P - оператор интерполяции, который каждой функции f из W ставит в соответствие функцию Pf из V, удовлетворяющую следующим условиям:

$$\delta_i(Pf) = \delta_i(f), \quad i \in I;$$
$$P^2 f = P(Pf) = Pf.$$

Оператор интерполяции P называют точным на полиномах степени m, если для каждого полинома p из P справедливо соотношение $Pp = p$. Интерполяционный метод Лагранжа для $m + 1$ точек исходных данных является точным на полиномах степени, меньшей или равной m.

Погрешность интерполяции в общем виде определяется выражением

$$Rf(t) = f(t) - Pf(t).$$

Для нахождения погрешности интерполяционной формулы Ньютона воспользуемся легко проверяемым соотношением

$$f(t) = \sum_{i=0}^{m} d[t_0, \ldots, t_i] N_i(t) + d[t_0, \ldots, t_m, t] \Pi(t).$$

Отсюда получим выражение для ошибки

$$Rf(t) = d[t_0, t_1, \ldots, t_m, t] \Pi(t).$$

В общем случае, если для полиномов степени m интерполяция осуществляется точно, для произвольной функции ошибку находят по формуле Тейлора, предполагая функцию f дифференцируемой $m + 1$ раз. При этом получают два выражения

$$Rf(t) = \frac{f^{(m)}(\xi)}{(m+1)!} \Pi(t), \quad a \leqslant \xi \leqslant b, \quad Rf(t) = \int_a^b K_m(x, t) f^{(m+1)}(t)\, dt,$$

где K - функция ядра ошибки.

Рассмотрим теперь проблему сходимости. Пусть P_m - интерполяционный полином для функции f по $m + 1$ точкам. Должна ли ошибка интерполяции Rf стремиться к нулю, если m стремится к бесконечности? Для положительного ответа на этот вопрос надо наложить сильные ограничения на f (непрерывности самой функции и ее производных недостаточно) - функция f должна быть аналитической на $[a, b]$. Тогда

$$\lim_{m\to\infty} \max_{x\in[a,b]} |f(t) - P_m(t)| = 0.$$

Напомним, что функция f называется аналитической на $[a,b]$, если существуют такие a_i и c, что f допускает разложение в степенной ряд, сходящийся на $[a,b]$:

$$f(t) = a_0 + a_1(t-c) + \ldots + a_i(t-c)^i + \ldots .$$

Достаточное условие аналитичности функции f состоит в том, что все ее производные должны быть ограничены на $[a,b]$. Например, e^t, sin t, cos t аналитичны на любом интервале, tg t - на интервале $[a,b] \subset (-\pi/2, \pi/2)$. Следует отметить, что исследование сходимости различных методов интерполяции, аппроксимации или сглаживания позволяет довольно объективно оценить их эффективность. Однако для этого используются сложные математические приемы, описание которых выходит за рамки данной книги.

Интерполяционные полиномы Эрмита

Этот тип полиномов является обобщением интерполяционных полиномов Лагранжа и используется тогда, когда заданы не только значения функции, но также и касательные в данных точках. Таким образом, число условий удваивается и интерполяционный полином минимальной степени будет принадлежать пространству, размерность которого в два раза больше, чем в предыдущем случае. Здесь будут рассмотрены только первые производные, хотя обобщение можно распространить и на случай производных более высокого порядка. Исходные данные:

1) $\Omega = [a,b] \subset R$;

2) $\{\delta_i\}_{i\in I}$, $I = \{0, 1, \ldots, 2m+1\}$, причем

$$\delta_i(f) = f(t_i), \quad i = 0, 1, \ldots, m,$$

$$\delta_i(f) = f'(t_i), \quad i = m+1, \ldots, 2m+1,$$

где $$a = t_0 < t_1 < \ldots < t_m = b;$$

3) $V = P^{2m+1}$ - множество полиномов степени, меньшей или равной $2m+1$. Размерность этого векторного пространства равна $2m+2$.

4) $z_i \in R$, $i \in I$, где первые $m+1$ значений соответствуют ординатам точек с абсциссами t_i, остальные - коэффициентам наклона касательных в этих точках.

Рассмотрим сначала частный случай: $m = 1$, $t_0 = 0$, $t_1 = 1$. Введем базисные полиномы следующего вида:

$$\varphi_0(t) = (2t+1)(1-t)^2, \qquad \varphi_2(t) = t(1-t)^2,$$

$$\varphi_1(t) = (3-2t)t^2, \qquad \varphi_3(t) = (t-1)t^2.$$

Это полиномы третьей степени, удовлетворяющие следующим соотношениям:

$$\varphi_i(j) = \delta_{i,j}, \quad \varphi_i'(j) = \delta_{i-2,j}, \quad i = 0, 1, 2, 3, \quad j = 0, 1,$$

где $\delta_{i,j} = 1$, если $i = j$, и 0, если $i \neq j$.

Интерполяционный полином Эрмита для этого случая имеет вид

$$v(t) = z_0 \varphi_0(t) + z_1 \varphi_1(t) + z_2 \varphi_2(t) + z_3 \varphi_3(t),$$

что эквивалентно выражению с использованием оператора интерполяции

$$Pf(t) = \delta_0(f) \varphi_0(t) + \delta_1(f) \varphi_1(t) + \delta_2(f) \varphi_2(t) + \delta_3(f) \varphi_3(t).$$

Переходя к общему случаю, попытаемся записать интерполяционный полином Эрмита в виде

$$v(t) = \sum_{i=0}^{2m+1} z_i \varphi_i(t), \quad \varphi_i \in V,$$

где полиномы φ_i должны удовлетворять следующим условиям:

$$\varphi_i(t_j) = \delta_{i,j}, \quad \varphi_i'(t_j) = \delta_{i-m,j}, \quad i = 0, 1, \ldots, 2m+1, \quad j = 0, 1, \ldots, m,$$

которые позволяют их искать в виде

$$\varphi_i(t) = u_i(t) L_i^2(t), \quad \varphi_{i+m+1}(t) = v(t) L_i^2(t), \quad i = 0, 1, \ldots, m,$$

где L_i - полином Лагранжа ($L_i(t_j) = \delta_{i,j}$). Проводя вычисления, получаем

$$v(t) = \sum_{i=0}^{m} z_i [1 - 2(t - t_i) L_i'(t_i)] L_i^2(t) + \sum_{i=0}^{m} z_{i+m+1} (t - t_i) L_i^2(t).$$

Интерполяционный полином Эрмита является единственным и интерполяция является точной на полиномах степени $2m + 1$. Ошибка интерполяции равна

$$Rf(t) = \frac{f^{(2m+2)}(\xi)}{(2m+2)!} [\Pi(t)]^2, \text{ где } a \leqslant \xi \leqslant b.$$

В заключение отметим, что полиномиальную интерполяцию имеет смысл применять лишь для небольшого числа точек (не больше пятнадцати) из-за того, что вместе с числом точек растет степень полинома и имеют место большие осцилляции в промежутках между точками (рис. 1-3). Базис Лагранжа должен применяться в тех случаях, когда вычисляется большое число интерполяционных полиномов на одних и тех же точках, поскольку система уравнений получается диагональной. В других случаях обычно используют базис Ньютона, приводящий к реугольной системе, а использование разделенных разностей позволяет легко добавлять новые точки интерполяции.

Кусочная полиномиальная интерполяция

С помощью кусочной полиномиальной интерполяции можно обойти упомянутые выше затруднения и строить интерполирующие функции на большом числе точек. Идея состоит в том, чтобы строить

независимо друг от друга полиномы на каждом интервале $[t_i, t_{i+1}]$, сшивая их между собой. Для этой цели можно использовать также и сплайны, но они будут рассмотрены в других разделах.

В качестве исходных данных возьмем следующие:

1) $\Omega = [a, b] \subset R$;
2) $\delta_i(f) = f(t_i)$, $a = t_0 < \ldots < t_m = b$;
3) $z_i \in R$.

Пример 1. Рассмотрим наиболее простой метод интерполяции: на отрезке $[t_i, t_{i+1}]$ Pf является интерполяционным полиномом Лагранжа, построенным для двух точек t_i и t_{i+1}:

$$Pf(t) = \frac{t - t_i}{t_{i+1} - t_i} z_{i+1} + \frac{t - t_{i+1}}{t_i - t_{i+1}} z_i, \; t \in [t_i, t_{i+1}],$$

а на всем отрезке $[a, b]$ Pf является непрерывной функцией, состоящей из таких кусков. Достоинством метода является его равномерная сходимость при увеличении числа точек:

$$\|f - Pf\|_\infty = \max_{t \in [a, b]} |f(t) - Pf(t)| \leqslant C \max_i |t_{i+1} - t_i|$$

(C - постоянная, зависящая только от f'). Недостаток метода состоит в том, что решение представляет собой недифференцируемую функцию.

Пример 2. Чтобы устранить этот недостаток, можно потребовать, чтобы производные были непрерывны в точках сшивки t_i. Проведем кусочный полином таким образом, чтобы для $t \in [t_0, t_2]$ он представлял собой интерполяционный полином Лагранжа второй степени, проведенный по точкам t_0, t_1, t_2, а на каждом следующем участке для $t \in [t_i, t_{i+i}]$, $i = 2, \ldots, m$ —1, состоял бы из полиномов P_i, таких, что

$$P_i(t_i) = z_i, \; P_i(t_{i+1}) = z_{i+1} \text{ и } P_i'(t_i) = P_{i-1}'(t_i),$$

$$P_i(t) = z_i \left(1 - \left(\frac{t - t_i}{t_{i+1} - t_i}\right)^2\right) + z_{i+1} \left(\frac{t - t_i}{t_{i+1} - t_i}\right) + \\ + P_{i-1}'(t_i) \left(\frac{t - t_i}{t_{i+1} - t_i}\right) \left(1 - \frac{t - t_i}{t_{i+1} - t_i}\right).$$

На отрезке $[a, b]$ Pf является дифференцируемой, кусочно-полиномиальной функцией второй степени. При использовании метода нужно соблюдать осторожность, поскольку возможно накопление ошибок при вычислении производных.

Пример 3. Предположим, что кроме значений функции в точках известны также значения ее производной. В этом случае на отрезке $[t_i, t_{i+1}]$ можно построить интерполяционный полином Эрмита. Обозначим через λ_i производную функции f в точке t_i и положим

$$\tau = \frac{t - t_i}{t_{i+1} - t_i},\ t \in [t_i,\ t_{i+1}].$$

Тогда

$$Pf(t) = \varphi_0(\tau)f(t_i) + \varphi_1(\tau)f(t_{i+1}) + \varphi_2(\tau)(t_{i+1} - t_i)\lambda_i + \varphi_3(\tau)(t_{i+1} - t_i)\lambda_{i+1},$$

где функции φ_0, φ_1, φ_2, φ_3 определены выше.

Если производная неизвестна, можно попытаться вычислить ее приближенное значение. Это можно сделать двумя способами; один (общий) связан с применением сплайнов и рассмотрен ниже, другой (локальный) рассмотрен в следующем примере.

Пример 4. Производную в точках t_i оценивают по значению производной интерполяционного полинома Лагранжа, построенного на точках $t_{i-p}, \ldots, t_{i+p}$, где p - малое целое (для точек t_0 и t_m берут $2p+1$ точек с одной стороны). Например, для $p = 2$ имеем

$$\lambda_i = \frac{t_i - t_{i+1}}{(t_{i-1} - t_i)(t_{i-1} - t_{i+1})} f(t_{i-1}) + \frac{2t_i - t_{i-1} - t_{i+1}}{(t_i - t_{i-1})(t_i - t_{i+1})} f(t_i) +$$

$$+ \frac{t_i - t_{i-1}}{(t_{i+1} - t_i)(t_{i+1} - t_{i-1})} f(t_{i+1}).$$

В случае когда точки эквидистантны, формулы упрощаются. Если $h = t_{i+1} - t_i$, то

$$\lambda_i = \frac{1}{h}[f(t_{i-1}) + 2f(t_i) + f(t_{i+1})].$$

Таким образом, на каждом интервале $[t_i,\ t_{i+1}]$ строится полином третьей степени, интерполяционная функция на отрезке $[t_0,\ t_m]$ непрерывна и дифференцируема, а метод интерполяции обладает сходимостью.

Интерполяция сплайнами

Со времени введения сплайнов Шёнбергом им посвящено большое число работ, в которых даны их различные представления. Мы будем использовать представление для кубических сплайнов, и приведем некоторые обобщения.

В качестве исходных данных возьмем следующие:

1) $\Omega = [a, b]$;
2) $\delta_i(f) = f(t_i)$, $a < t_0 < t_1 < \ldots < t_n < b$;
3) $(z_i)_{0 \leq i \leq n}$.

Кубические сплайны. Вернемся к примеру 3, где предполагались известными производные λ_i функции f в точках t_i. Обозначим $\tau = \dfrac{t - t_i}{t_{i+1} - t_i}$ для $t_i \leqslant t \leqslant t_{i+1}$, тогда

$$f_\lambda(t) = Pf(t) = \varphi_0(\tau) f(t_i) + \varphi_1(\tau) f(t_{i+1}) + \\ + \varphi_2(\tau)(t_{i+1} - t_i)\lambda_i + \varphi_3(\tau)(t_{i+1} - t_i)\lambda_{i+1}.$$

Функция f_λ является непрерывной и дифференцируемой на отрезке $[t_0, t_n]$.

Если производная функции f в точке t_i неизвестна, f_λ можно рассматривать как семейство функций, зависящих от параметра $\lambda = (\lambda_0, \lambda_1, \ldots, \lambda_n)$. Тогда λ_i можно определить с помощью двух критериев.

Первый критерий:

$f''_\lambda(t_0) = 0;$

f_λ непрерывна и дважды дифференцируема на $(t_0, t_n);$

$f''_\lambda(t_n) = 0.$

Второй критерий: из всех функций $f_\lambda(t)$ выбирается такая функция, которая обращает в минимум

$$J(\lambda) = \int_{t_0}^{t_n} (f''_\lambda(t))^2 \, dt.$$

Первый критерий соответствует выбору регулярной функции, второй позволяет выбрать функцию с минимальной кривизной; последнее условие является основным свойством кубического сплайна. В действительности оба критерия эквивалентны, так как приводят к одному результату: из первого следует система $n + 1$ линейных уравнений, идентичная системе, к которой приводит минимизация выражения $J(\lambda)(J'(\lambda) = 0)$.

Этот метод интересен тем, что получаемая система уравнений имеет много нулевых коэффициентов, т. е не требует много машинного времени и места в памяти. Другим его достоинством является вычислительная устойчивость. Сходимость метода пропорциональна h^2, где $h = \max |t_{i+1} - t_i|$. Можно получить сходимость порядка h^4, изменяя граничные условия $f''(t_0) = 0$ и $f''(t_n) = 0$. Примеры интерполяции кубическими сплайнами приведены на рис. 4-6.

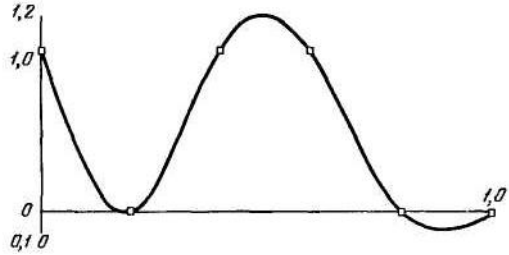


Рис 4. Интерполяция сплайном на 6 точках [ср. с рис. 3].

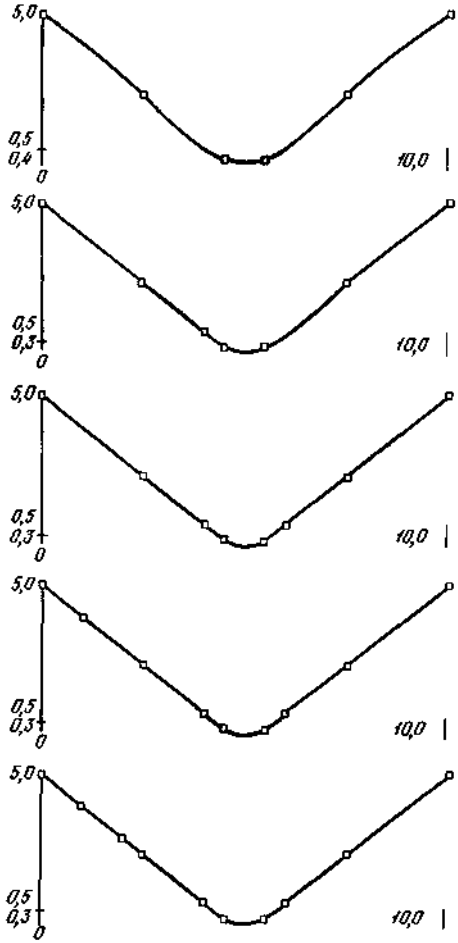


Рис. 5. Интерполяция сплайном на 6-10 точках. Положение точек задается функцией $|x - 5|$ (ср. с рис. 2)

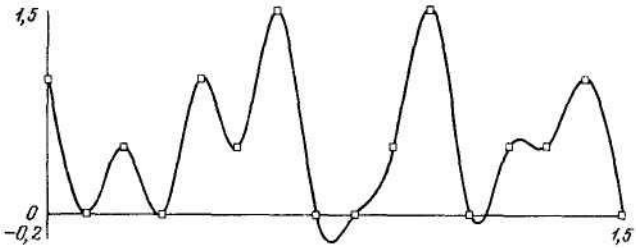


Рис. 6. Интерполяция сплайном на 16 точках. Интерполяционная кривая хорошо описывает произвольные данные

Полиномиальные сплайны. Полиномиальным сплайном степени q на n точках t_i называют функцию s, которая является полиномом степени $2q - 1$ на каждом интервале (t_i, t_{i+1}), $(i = 1, \ldots, n-1)$ и полиномом степени $q - 1$ на интервалах $[a, t_1)$ и $(t_n, b]$ и имеет непрерывную производную $s^{(2q-2)}$. Обозначим через S множество функций s, которые могут быть представлены в виде

$$s(t) = \sum_{j=0}^{q-1} \alpha_j t^j + \sum_{i=1}^{n} d_i \frac{(t - t_i)_+^{2q-1}}{(2q-1)!},$$

где

$$(t - t_i)_+ = t - t_i \text{ при } t - t_i > 0 \text{ и } (t - t_i) = 0 \text{ при } t - t_i \leqslant 0$$

$$\text{и } \sum_{i=1}^{n} d_i (t_i)^k = 0, \; k = 0, \ldots, q-1.$$

Отметим, что на каждом интервале функция s полностью определяется своим значением и значением производных до порядка $q - 1$ на обоих концах интервала. Для известных данных z, функцию s отыскивают из условий $s(t_i) = z_i$, а также из условия минимизации

$$\int_a^b [s^q(t)]^2 \, dt$$

На практике чаще всего встречается рассмотренный выше случай $q = 2$. Для $q = 3$ существующие методы дают удовлетворительные результаты только для ограниченного числа точек ($n \leq 30$).

Помимо интерполяции, сплайны можно использовать и в других задачах. Ниже будут рассмотрены *B*-сплайны, метод наименьших квадратов и др. Во всех случаях придается большое значение выбору «хороших» базовых функций на множестве S.

Сплайны общего типа. Из изложенного выше ясно, что сплайн можно определить двумя способами: исходя из взаимного согласования простых функций и из решения задачи минимизации. Третье определение использует понятие «воспроизводящего ядра». Это определение мы используем ниже в примере, когда будем рассматривать задачи с двумя пространственными переменными.

В теории сплайнов показывается, что второе определение допускает обобщение в двух направлениях:

- обобщение критерия минимизации: если D - дифференциальный оператор, можно искать минимум $\int_a^b [Df(t)]^2\,dt$;
- обобщение используемых функционалов: $\varphi(f) = \max\limits_{a<t<b} |f(t)|$, либо $\varphi(f) = \max\limits_{a<t<b} |f'(t)|$, либо $\varphi(f) = \max\limits_{a<t<b} |\alpha f(t) + \beta f''(t)|$, либо $\varphi(f) = \int_a^b |f(t)|\,dt$.

Большой интерес в теории оптимизации может представлять также сплайн «под напряжением», критерием для определения которого является минимизация выражения

$$\int_a^b \{[f''(t)]^2 + u^2[f'(t)]^2\}\,dt.$$

На каждом интервале получают функцию вида

$$\alpha + \beta x + \gamma \operatorname{sh} px + \delta \operatorname{ch} qx.$$

14.2.2. Сглаживание

Методы сглаживания используются в тех случаях, когда необходимо найти функцию, проходящую вблизи большого числа заданных точек. Мы рассмотрим три метода сглаживания:

- метод наименьших квадратов;
- сглаживающие сплайны;
- функции Безье и *B*-сплайны.

Метод наименьших квадратов

Рассмотрим два векторных пространства V и W конечных размерностей. Пусть n и m соответственно размерности V и W, $n < m$. Обозначим через $(v_i)_{1\leqslant i\leqslant n}$ и $(w_i)_{1\leqslant i\leqslant m}$ базисные функции пространств V и W. Предположим, что в W определено скалярное произведение, обозначаемое $\langle\ ,\ \rangle$, т. е. операция, которая каждой паре (x, *y)* элементов из W ставит в соответствие действительное число $\langle x, y\rangle$ из R, такое что

1) $\forall x \in W\ \langle x, x\rangle \geqslant 0$, $\langle x, x\rangle = 0$ тогда и только тогда, когда $x = 0$;
2) $\forall x, y \in W\ \langle x, y\rangle = \langle y, x\rangle$;
3) $\forall \lambda, \mu \in R$, $\forall x, y, z \in W\ \langle \lambda x + \mu y, z\rangle = \lambda\langle x, z\rangle + \mu\langle y, z\rangle$.

(В качестве упражнения возьмите $W = R^2$ и определите скалярное произведение в евклидовой геометрии.)

С помощью скалярного произведения можно определить понятие расстояния $\varphi(x-y) = \sqrt{\langle x-y, x-y\rangle}$ и довольно просто производить вычисления.

При использовании метода наименьших квадратов задача сглаживания формируется следующим образом:

1) пусть T-линейный оператор преобразования V в W полного ранга n, т. е. обладает свойством: $Tv = 0$ тогда и только тогда, когда $v = 0$;
2) пусть w - элемент W;
3) найти элемент $\bar{v}$ из V, такой что

$$\varphi(T\bar{v} - w) \leqslant \varphi(Tv - w), \quad \forall v \in V.$$

Предположим, существует такое $\bar{v}$, что

$$\varphi(T\bar{v} - w) = \min_{v \in V} \varphi(Tv - w).$$

Тогда решение $\bar{v}$ должно удовлетворять вариационному уравнению

$$\langle T\bar{v}, Tv\rangle = \langle w, Tv\rangle, \quad \forall v \in V.$$

Поскольку V имеет конечную размерность, это выражение можно записать в виде

$$\langle T\bar{v}, Tv_i\rangle = \langle w, Tv_i\rangle, \quad i = 1, \ldots, n.$$

Пусть

$$v = \sum_{j=1}^{n} \lambda_j v_j.$$

Введем обозначения:

- вектор $\lambda = (\lambda_j)$, $1 \leqslant j \leqslant n$;
- $\tau = (\tau_{ij})$, где $\tau_{ij} = \langle Tv_i, Tv_j\rangle$, $1 \leqslant i, j \leqslant n$;
- $\beta = (\beta_i)$, где $\beta_i = \langle w, Tv_i\rangle$, $1 \leqslant i \leqslant n$.

Тогда решение задачи сводится к решению системы n уравнений с n неизвестными:

$$\tau\lambda = \beta.$$

Таким же образом могут быть сформулированы другие задачи и методы их решения. Здесь V - пространство аппроксимирующих функций, W - пространство наблюдений. Поясним изложенное выше на примере.

Пусть

- $V = P^{n-1}$ - множество полиномов степени, меньшей или равной n — 1;
- $W = R^m$ со скалярным произведением

$$\langle x, y \rangle = \sum_{i=1}^{m} x_i y_i;$$

• t_i - действительные числа ($i = 1, ..., m$) и T - линейный оператор преобразования V в W,

$$Tp = (\delta_i p)_{1 \leqslant i \leqslant m} = (p(t_i)), \quad p \in V;$$

• z - элемент R^m.

Решение задачи сглаживания методом наименьших квадратов состоит в нахождении полинома p степени не больше n — 1, который минимизирует выражение

$$\sum_{i=1}^{m} (p(t_i) - z_i)^2.$$

Если через v_j обозначить базисные полиномы V, получим

$$\tau_{ij} = \langle Tv_i, Tv_j \rangle = \sum_{K=1}^{m} v_i(t_k) v_j(t_k), \quad \beta_i = \sum_{K=1}^{n} z_k v_k(t_i).$$

Допустим также, что

$$A = (a_{ij})_{\substack{1 \leqslant i \leqslant n \\ 1 \leqslant j \leqslant m}},$$

где $a_{ij} = v_i(t_j)$, тогда $\tau = A^t A$ и $\beta = A^t z$, где A' - транспонированная матрица A, и система линейных уравнений запишется в виде

$$A^t A\lambda = A^t z.$$

Эта система может быть решена с помощью известных методов. Однако вычисление матрицы A^tA (порядка n) в процессе решения системы может снизить точность решения. Поэтому предпочитают использовать другой метод, который мы кратко опишем здесь. В его основе лежит метод Гивенса, в котором матрица A представляется в виде произведения двух матриц $A = QS$, где Q - ортогональная матрица порядка n, а матрица R имеет следующую структуру:

$$S = \left| \frac{\tilde{S}}{0} \right|,$$

где S - верхняя треугольная матрица порядка n. Тогда уравнение $A^t A\lambda = A^t z$ можно записать в виде $S^t Q^t QS\lambda = S^t Q^t z$, что эквивалентно

$$S^t S\lambda = S^t Q^t z.$$

Обозначим через $[Q^t z]_n$ вектор, принадлежащий R^n и состоящий из n первых компонентов $Q^t z$. Тогда

$$\tilde{S}^t \tilde{S}\lambda = \tilde{S}^t [Q^t z]_n,$$

где S - обратимая матрица, а $\tilde{S}^t \tilde{S}$ является разложением Холецкого матрицы A^tA. Отсюда

$$\tilde{S}\lambda = [Q'z]_n.$$

Метод Холецкого

При разложении симметричных матриц по методу Гаусса можно уменьшить число выполняемых арифметических операций и требуемый объем памяти. Более того, существование *RU*-разложения следует из свойства положительной определенности симметричной матрицы *A*. Можно показать, что для симметричной положительно определенной матрицы *A* существует разложение

$$A = LL^t$$

(разложение Холецкого). Матрицу *L* легко получить методом Гаусса, но на практике предпочитают выполнять вычисления путем прямого сопоставления матриц *A* и LL^t:

$$l_{jj} = \sqrt{a_{jj} - \sum_{k=1}^{j-1} l_{jk}^2}, \quad l_{ij} = \frac{a_{ij} - \sum_{k=1}^{j-1} l_{ik} l_{jk}}{l_{jj}}, \quad j+1 \leqslant i \leqslant n$$

$$j = 1, 2, \ldots, n.$$

Полная стоимость разложения составляет половину стоимости метода Гаусса плюс *n* вычислений квадратного корня. Метод обладает вычислительной устойчивостью.

Матрица полученной системы является верхней треугольной, и систему легко решить. Примеры решения приведены на рис. 7 и 8.

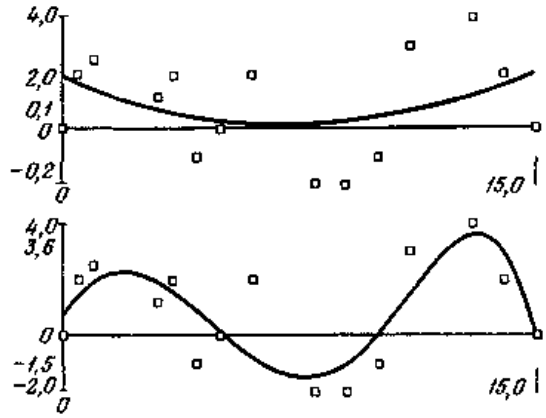


Рис. 7. Сглаживание методом наименьших квадратов полиномами 3 и 5 степени

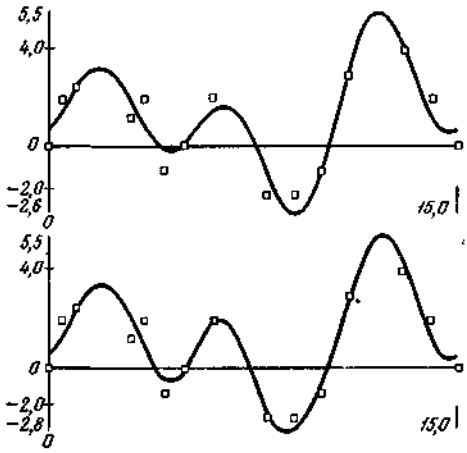


Рис. 8. Сглаживание методом наименьших квадратов тригонометрическими полиномами 3 и 5степени:

$$\sum x_i \sin(2i\pi(x-a)/(b-a)) + \\ + \sum y_i \cos(2i\pi(x-a)/(b-a)).$$

Описанный метод прост, эффективен, обладает высокой вычислительной устойчивостью. Кроме того, можно показать, что

$$\sum_{i=1}^{m} (\hat{p}(t_i) - z_i)^2 = \sum_{i=n+1}^{m} ((Q^t z)_i)^2,$$

и тем самым получить информацию об ошибках сглаживания. Отметим также, что если $m = n$ и удачно выбраны точки t_i и базовые функции v_i, то можно получить систему уравнений с диагональной матрицей. В рассмотренном выше примере вместо полиномов можно использовать другие функции, например B-сплайны, которые описаны ниже. Задачу можно обобщить, рассматривая пространство функций с дополнительными ограничениями линейного типа. В таких случаях систему уравнений решают методом неопределенных множителей Лагранжа.

Сглаживающие сплайны

Рассмотрим пример применения кубических сплайнов для сглаживания (рис 9-11). Пусть ρ - действительное положительное число. Тогда на данных $(t_i, z_i)_{1 \leqslant i \leqslant n}$ сглаживающий сплайн f_λ будет функцией, которая минимизирует критерий

$$J_\rho(f) = \int_b^a (f''(t))^2\, dt + \rho \sum_{i=1}^{n} (f(t_i) - z_i)^2.$$

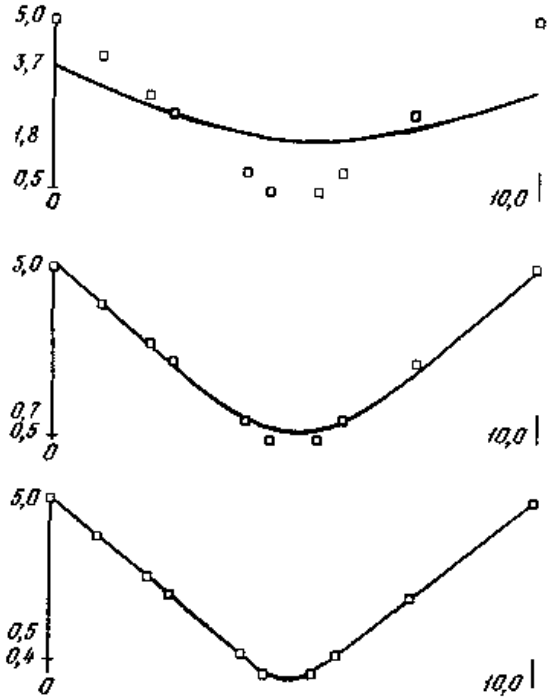


Рис. 9. Сглаживание кубическим сплайном на 10 точках с различными параметрами сглаживания ρ. Положение точек задается функцией $|x - 5|$:

ρ = 10 (результат близок к прямой, полученной по методу наименьших квадратов),

ρ = 100 (приемлемый результат для произвольных данных),

ρ = 500 (результат близок к интерполяционному сплайну)

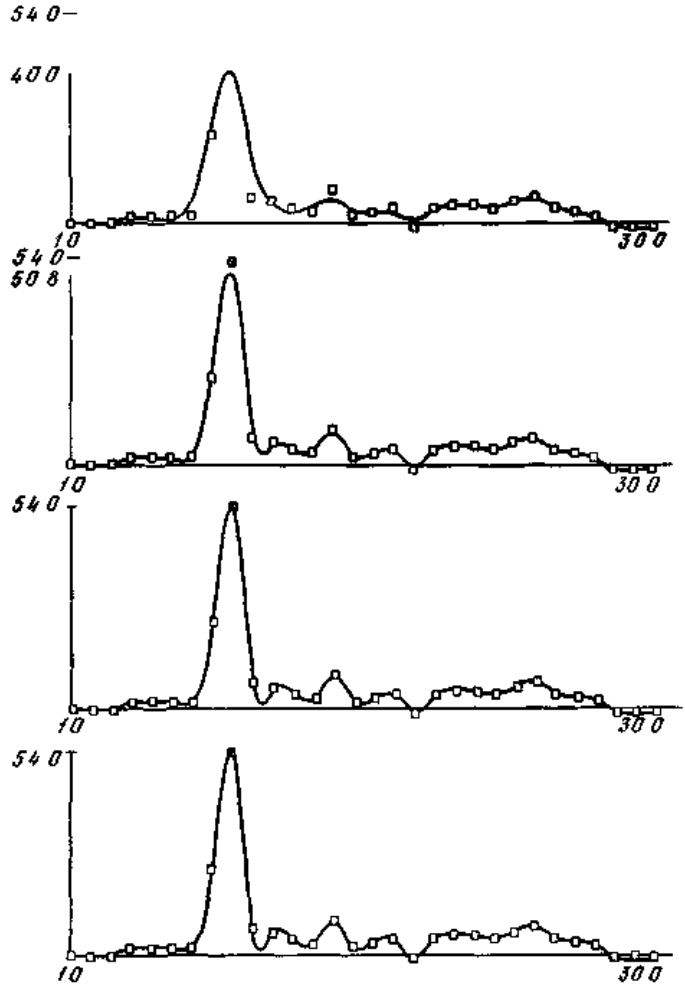


Рис. 10. Сглаживание кубическим сплайном произвольных данных с различными значениями параметра сглаживания ρ (10, 100, 500, 10000)

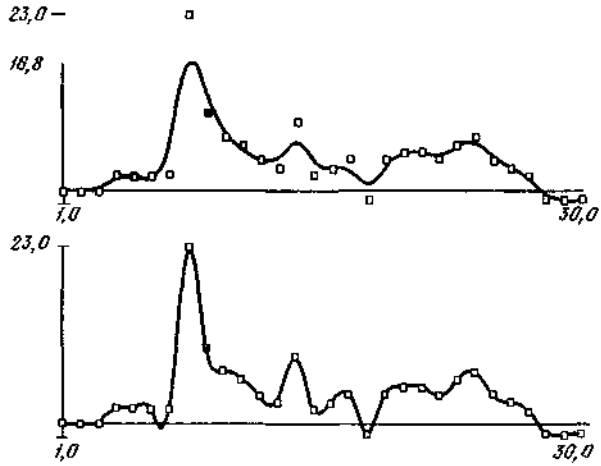


Рис. 11. Сглаживание кубическим сплайном данных, приведенных на рис. 10, за исключением точки, помеченной *x,* ордината которой уменьшена с 54 до 10, для двух значений параметра сглаживания ρ [10, 500].

Решение будет кубическим сплайном (кусочный полином третьей степени на каждом интервале, дважды дифференцируемый с непрерывной второй производной).

От выбора ρ существенно зависят свойства сплайна. Для больших значений ρ сплайн пройдет близко к точкам, а для малых значений ρ сплайн будет ближе к прямой, проведенной методом наименьших квадратов. Существует метод автоматического выбора ρ (метод возрастающего признания), но при его использовании возникают трудности в процессе построения сплайна на неэквидистантных точках.

Функции Безье (рис. 12, 13)

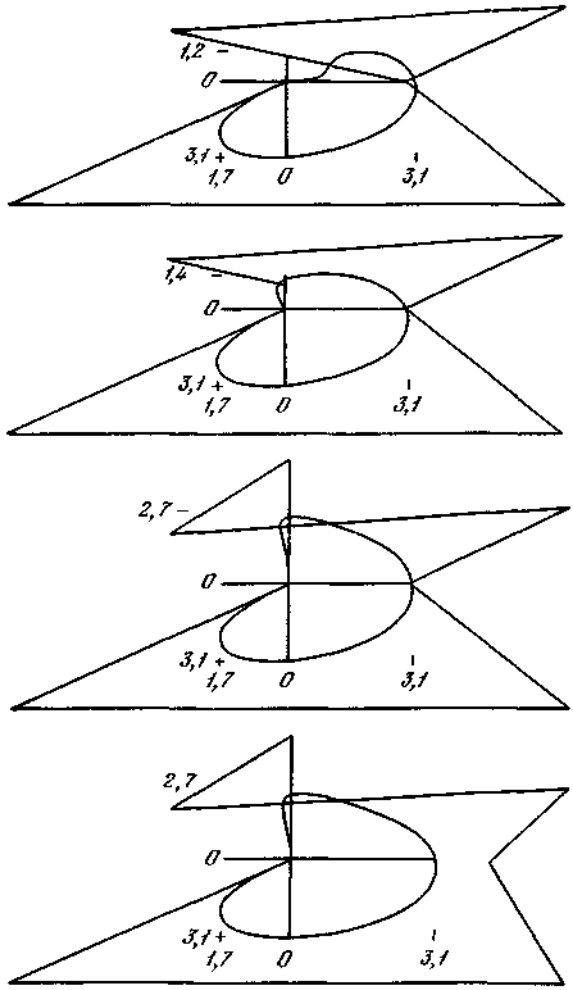


Рис. 12. Сглаживание с помощью кривых Безье на многоугольниках с координатами вершин:

(0,0), (3,0), (−3,2), (7,3), (3,0), (7,−5), (−7,−5), (0,0),
(0,0), (0,1), (−3,2), (7,3), (3,0), (7,−5), (−7,−5), (0,0),
(0,0), (0,5), (−3,2), (7,3), (3,0), (7,−5), (−7,−5), (0,0),
(0,0), (0,5), (−3,2), (7,3), (5,0), (7,−5), (−7,−5), (0,0).

Отметим касательные к кривой в крайних точках многоугольника, изменение всей кривой при изменении положения одной точки и образование петли за счет расположения данных.

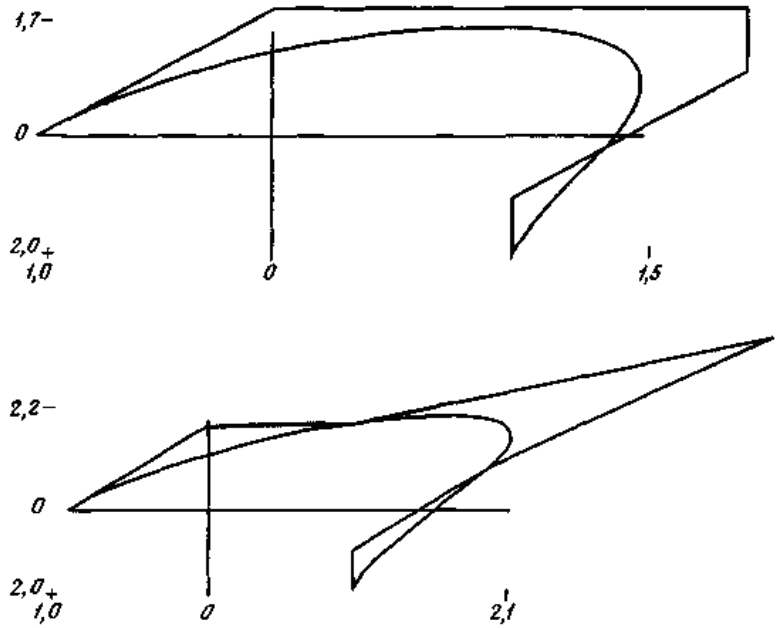


Рис. 13. Сглаживание с помощью кривых Безье на незамкнутых многоугольниках, отличающихся в одной точке.

Пусть f - действительная функция, определенная на отрезке [0, 1] Полиномом Бернштейна степени n, связанным с функцией f, называют

$$B_n(f; t) = \sum_{i=0}^{n} f\left(\frac{i}{n}\right) \Phi_{n,i}(t),$$

где $\Phi_{n,i}(t) = C_n^i t^i (1-t)^{n-i}$. Напомним, что

$$C_n^i = \frac{n!}{i!(n-i)!}$$

Функции $\Phi_{n,t}$ (t) удовлетворяют следующим свойствам:

- $\Phi_{n,i}(t) \geqslant 0, \ \forall t \in [0, 1]$,
- $\sum_{i=0}^{n} \Phi_{n,i}(t) = 1; \quad \sum_{i=0}^{n} \left(\frac{i}{n}\right) \Phi_{n,i}(t) = t$

и могут служить базисными полиномами степени, меньшей или равной n. Любой полином P_n степени n может быть представлен в виде

$$P_n(t) = \sum_{i=0}^{n} b_i \Phi_{n,i}(t).$$

Если принять

$$\Delta^0 b_i = b_i, \quad \Delta b_i = b_{i+1} - b_i, \quad \Delta^{k+1} b_i = \Delta(\Delta^k b_i),$$

получим

$$P_n(t) = \sum_{k=0}^{n} C_n^k \Delta^k b_0 t^k.$$

Если b является последовательностью действительных чисел $(b_0, b_1, ..., b_n)$, функцией Безье, связанной с b, называется полином

$$B_n(b; t) = \sum_{i=0}^{n} b_i \Phi_{n,i}(t).$$

Вычисление функций Безье осуществляется по рекуррентным формулам

$$b_{i,0} = b_i, \quad 1 \leqslant i \leqslant n,$$
$$b_{i,j}(t) = t b_{i,j-1}(t) + (1 - t) b_{i-1,j-1}(t), \quad t \in [0, 1],$$
$$B_n(b, t) = b_{n,n}(t).$$

Подобные формулы существуют и для производных:

$$b_{i,0}^{(k)} = \Delta^k b_i,$$
$$b_{i,j}^{(k)}(t) = t b_{i,j-1}^{(k)}(t) + (1 - t) b_{i-1,j-1}^{(k)}(t),$$
$$B_n^{(k)}(b, t) = b_{i-k,\, n-k}^{(k)}(t).$$

B-сплайны. Пусть $\Pi=(t_i)_{i \in Z}$ - возрастающая последовательность действительных чисел и

$$g_m(s, t) = (s - t)_+^{m-1} = \begin{cases} (s - t)^{m-1}, & \text{если } s \geqslant t, \\ 0, & \text{если } s < t. \end{cases}$$

B-сплайн (степени m) $M_{i,m}(t)$ определяется как разделенная разность порядка m функции $g_m(s, t)$ по отношению к аргументу s на точках

$$t_i, t_{i+1}, \ldots, t_{i+m}$$

$$d[t_i] = g_m(t_i, t), \ d[t_i, t_j] = (g_m(t_i, t) - g_m(t_j, t))/(t_i - t_j),$$
$$M_{i,m}(t) = d[t_i, \ldots, t_{i+m}] = \frac{d[t_i, \ldots, t_{i+m-1}] - d[t_i, \ldots, t_{i+m-2}, t_{i+m}]}{t_{i+m-1} - t_{i+m}}.$$

Нормализованный *B*-сплайн $N_{i,m}(t)$ определяется следующим образом:

$$N_{i,m}(t) = (t_{i+m} - t_i) M_{i,m}(t).$$

Если $t_i = t_{i+1} = \ldots = t_{i+k_i-1}$, $(1 \leqslant k_i \leqslant m)$, обе функции $N_{i,m}$ и $M_{i,m}$ имеют производные лишь до порядка $m - k_i - 1$. Таким образом, наличие совмещенных точек может привести к нарушению непрерывности некоторых производных.

Если $m > 1$ и Π не имеет более $m - 1$ совмещенных точек, то $M_{i,m}$ и $N_{i,m}$ являются непрерывными функциями. Ниже рассматривается этот случай. Элементы Π называются *узлами B-сплайна.*

Из рекуррентных формул для вычисления разделенных разностей можно получить формулы для нахождения функции $N_{i,m}(t)$

$$N_{i,1} = \begin{cases} 1, & \text{если} \quad t \in [t_i, t_{i+1}]; \\ 0, & \text{если} \quad t \notin [t_i, t_{i+1}]; \end{cases}$$

$$N_{i,m}(t) = \frac{t - t_i}{t_{i+m-1} - t_i} N_{i,m-1}(t) + \frac{t_{i+m} - t}{t_{i+m} - t_{i+1}} N_{i+1,m-1}(t).$$

Эти функции обладают следующими свойствами:

1) $N_{i,m} > 0, \; t \in (t_i, t_{i+m})$;
2) $N_{i,m} = 0, \; t \notin (t_i, t_{i+m})$;
3) $\sum_{i=1}^{m} N_{i,m}(t) = 1$;
4) существуют такие числа

$$\xi_1, \ldots, \xi_m, \text{ что } \sum_{i=1}^{m} \xi_i N_{i,m}(t) = t.$$

Точки ξ_i называются *узловыми точками B-сплайна.*

Сплайном степени m, определенным на отрезке $[a, b]$, называют функцию $S_m(\alpha, t)$, определяемую следующими соотношениями:

1) $S_m(\alpha, t) = \sum_{j=0}^{m+n+1} \alpha_j N_{j,m+1}(t), \; t \in [a, b]$;

2) $\Pi = \{\underbrace{t_0, \ldots, t_0}_{m+1}, t_1, t_2, \ldots, t_{n-1}, \underbrace{t_n, \ldots, t_n}_{m+1}\}$;
$t_0 = a; \; t_n = b; \; t_i \neq t_{i+1}, \; i = 0, 1, \ldots, n-1$;
3) $\alpha = (\alpha_i)_{0 \leqslant j \leqslant m+n+1}; \; \alpha_j \in R$.

Кубический сплайн, определенный выше, является сплайном третьей счепени. S_m является полиномом степени, меньшей или равной m, и на каждом интервале $[t_i, t_{i+1}]$ имеет производные до порядка $m - 1$.

Отметим, что если $\Pi = \{\underbrace{0, 0, \ldots, 0}_{m+1}, \underbrace{1, \ldots, 1}_{m+1}\}$, то функции $N_{j,m+1}$ являются полиномами Берштейна.

Если требуется найти сплайн не на отрезке $[a, b]$, а на всем множестве R, он может быть определен следующим образом:

$$\Pi = \{t_i\}_{i \in Z},$$

$$S_m(\alpha, t) = \sum_{j=-\infty}^{+\infty} \alpha_j N_{j,m+1}(t), \quad t \in R, \ \alpha_j \in R.$$

Для вычисления этих функций можно воспользоваться следующими формулами:

$$S_m(\alpha, t) = \sum_{i=l-m}^{l} \alpha_i N_{i,m+1}(t), \ t \in [t_l, t_{l+1}];$$

$$\alpha_{i,0} = \alpha_i, \ l - m \leqslant i \leqslant l;$$

$$\alpha_{i,j}(t) = \frac{t - t_i}{t_{i+m-j+1} - t_i} \alpha_{i,j-1}(t) + \frac{t_{i+m-j+1} - t}{t_{i+m-j+1} - t_i} \alpha_{i-1,j-1}(t);$$

$$1 \leqslant j \leqslant m, \quad l - m + j \leqslant i \leqslant l;$$

$$S_m(\alpha, t) = \alpha_{l,m}(t).$$

Существуют другие алгоритмы вычисления функций S_m и ее производных.

Таким образом, B-сплайны $N_{j,m+1}$ для фиксированного m являются базисными функциями для пространства сплайнов степени m. Их практическая польза состоит в том, что для вычисления можно использовать приведенные выше рекуррентные формулы (подобные тем, которые использовались при вычислении функций Безье).

Аппроксимирующие функции Безье и сплайны.

Аппроксимирующей функцией Безье степени $n \geq 1$ для функции f на отрезке $[0, 1] \in R$ называют полином Бернштейна степени n, связанный с f выражением

$$B_n(f, t) = \sum_{i=1}^{n} f\left(\frac{i}{n}\right) \Phi_{n,i}(t).$$

Аппроксимирующим сплайном функций f на $[0, 1] \in R$ называют следующую функцию: $S_m(f, t) = \sum_{i=0}^{m+n+1} f(\xi_i) N_{i,m+1}(t)$,

где $\Pi = \{t_0, \ldots, t_0, t_1, \ldots, t_{n-1}, t_n, \ldots, t_n\}$, $t_0 = a$, $t_n = b$ и ξ_i определяются из уравнения

$$\sum_{i=0}^{m+n+1} \xi_i N_{i,m+1}(t) = t.$$

Аппроксимирующим сплайном функции *f* на всем множестве *R* называют функцию

$$S_m(f, t) = \sum_{i=-\infty}^{+\infty} f(\xi_i) N_{i,m+1}(t).$$

Перечислим несколько общих свойств определенных выше функций. Для функции *f*, заданной на [0, 1] $\in R$ и непрерывной на нем, введем следующие обозначения:

V(f)- число изменений знака;

S(*f*)-число изменений знака монотонности;

T(f)-число изменений вогнутости.

Аппроксимирующие функции Безье и сплайны по отношению к функции *f* обладают следующими свойствами:

- число изменений знака уменьшается:

$$V(B_n(f)) \leqslant V(f) \text{ и } V(S_m(f)) \leqslant V(f);$$

- число изменений знака монотонности уменьшается:

$$S(B_n(f)) \leqslant S(f) \text{ и } S(S_m(f)) \leqslant S(f);$$

- число изменений вогнутости уменьшается:

$$T(B_n(f)) \leqslant T(f) \text{ и } T(S_m(f)) \leqslant T(f).$$

Кроме указанных выше геометрических свойств аппроксимирующие функции Безье и сплайны обладают и другими полезными свойствами. В частности, можно показать, что на *B*-сплайн в отличие от функции Безье оказывает влияние малое число точек. Иными словами, изменение одного значения $f\left(\frac{i}{n}\right)$ оказывает влияние на $B_n(f, t)$ для всех $t \in [0, 1]$, в то время как для $S_m(f, t)$ это влияние ограничивается одной функцией

14.2.3. Аппроксимация

По теории аппроксимации существует обширная литература.

Приведем вкратце одну из формулировок задачи аппроксимации. Введем следующие обозначения:

Ω - множество, принадлежащее *R;*

f - функция, определенная на Ω, со значениями в *R;*

V - конечномерное векторное пространство действительных функций, определенных на Ω, со значениями в *R;*

W - векторное пространство, порожденное *f* и *V;*

$\|\cdot\|$ - норма в *W.*

Напомним, что норма является таким отображением *W* в R^+, что $\|q\| = 0$ тогда итолько тогда, когда $\zeta = 0$, кроме того,

$$\forall g \in W,\ \|\lambda g\| = |\lambda| \cdot \|g\|,\ \forall \lambda \in R,\ \forall g \in W,$$
$$\|g + h\| \leqslant \|g\| + \|h\|,\ \forall g, h \in W.$$

В частности, норма может определяться скалярным произведением

$$\|g\| = (\langle g, g \rangle)^{1/2}.$$

Тогда задача аппроксимации заключается в следующем: найти (если существует) $\overline{f}$ -элемент пространства *C,* являющегося частью *W,* наиболее близкий к *f* в смысле нормы, т.е. найти функцию $\overline{f} \in C$, такую что

$$\|f - \overline{f}\| = \min_{g \in C} \|f - g\|.$$

Основные трудности, возникающие при теоретическом анализе задачи и ее решении, связаны со структурой пространства *C* и геометрическими свойствами нормы. Рассмотрим случай, когда *C* = *V*.

Наилучшая равномерная аппроксимация

Для этой задачи имеем

1) Ω = [0, 1];

2) *V* - подпространство векторного пространства непрерывных функций, отображающих Ω, на *R;*

3) $\|g\| = \max_{t \in [0,1]} |g(t)|.$

Данная норма называется нормой максимума или равномерной нормой. Она определяет расстояние между двумя функциями как максимальное отклонение между их графиками. Задача имеет по крайней мере одно решение, но оно может быть неединственным, что приводит к затруднениям при составлении алгоритма решения. Единственность может быть обеспечена, если *V* удовлетворяет определенному условию, называемому условием Хаара. (Условие Хаара сводится к тому, чтобы любая функция $f(t) \in V$ имела не более n — 1 корня при $t \in Q$).

Предположим, что *V* представляет собой множество полиномов степени, меньшей или равной *n*, т.е. будем искать решение в виде полинома $\overline{f} \in P_n$ расстояние которого от *f* минимально. Если *f* является решением, то существуют $n + 2$ точки t_i, в которых

$$|f(t_i) - \overline{f}(t_i)| = \max_{t \in [0,1]} |f(t) - \overline{f}(t)|$$

и

$$f(t_{i+1}) - \overline{f}(t_{i+1}) = -(f(t_i) - \overline{f}(t_i)),\ i = 1, 2, \ldots, n + 1.$$

Из изложенного видно, что наилучшая равномерная аппроксимация представляет собой весьма сложную задачу и вообще следует избегать такого типа норм, в выражениях для которых появляются операторы max или sup. Тем не менее, если функция f должна быть вычислена в большом числе точек за ограниченное время, может оказаться выгоднее искать именно наилучшую равномерную аппроксимацию на множестве полиномов. Подобные методы, как правило, используются при вычислении элементарных функций в алгоритмических языках высокого уровня.

Метод наименьших квадратов с непрерывными переменными

Условия задачи записываются следующим образом:

1) $\Omega = [a, b]$;

2) $\|g\| = \left(\int_a^b (g(t))^2 p(t)\, dt\right)^{1/2}$,

где непрерывная функция p отображает Ω в R^+. В основе данной нормы лежит скалярное произведение

$$\langle g, h \rangle = \int_a^b g(t) h(t) p(t)\, dt.$$

Решение задачи $\overline{f}$ существует и единственно. Его нахождение сводится к вычислению компонент $\overline{f}$ в базовых функциях пространства V. Для этого необходимо решить систему из n уравнений с n неизвестными. Если

$$\{v_i\}_{i=1, 2, \ldots, n}$$

являются базовыми функциями V, то

$$\overline{f} = \sum_{i=1}^{n} \bar{\alpha}_i v_i,$$

где $\overline{\alpha}_i$ являются решениями системы

$$\sum_{j=1}^{n} \bar{\alpha}_j \int_a^b v_i(t) v_j(t) p(t)\, dt = \int_a^b f(t) v_i(t) p(t)\, dt, \; i = 1, \ldots, n.$$

Для решения системы можно использовать один из известных методов. Интерес представляет частный случай, когда базисные функции ортогональны:

$$\int_a^b v_i(t) v_j(t) p(t)\, dt = 0, \; i \neq j.$$

В этом случае матрица системы диагональна и $\overline{\alpha}_i$ легко определяются. Одна из таких систем базисных функций при $p = 1$ представляет собой

$$\{1, \sin \pi t, \cos \pi t, \sin 2\pi t, \ldots, \cos n\pi t\}.$$

Для этой ортогональной базы $\overline{\alpha}_i$ являются коэффициентами разложения и ряд Фурье функции *f*.

Если *V* представляет собой пространство полиномов степени не выше *n*, для него также можно построить ортогональный базис, пользуясь результатами, полученными в теории ортогональных полиномов (с другими *p*). Интерес к ним связан с наличием рекуррентных формул, помогающих вычислять отдельные полиномы. Определенная с их помощью функция $\overline{f}$ также представляет собой равномерную аппроксимацию *f*, которая если и не является лучшей, то по крайней мере удовлетворительна (полиномы Чебышева).

Приведем основные типы ортогональных полиномов и рекуррентные формулы для их вычисления.

Полиномы Чебышева: $p(t) = \dfrac{1}{\sqrt{1-t^2}}$, $a = -1$, $b = 1$,

$$p_{n+1}(x) - 2xp_n(x) + p_{n-1}(x) = 0, \quad p_0(x) = 1, \quad p_1(x) = x.$$

Полиномы Лагерра: $p(t) = -e^{-t}$, $a = 0$, $b = \infty$,

$$p_{n+1}(x) - (1 + 2n - x)p_n(x) + n^2 p_{n-1}(x) = 0, \; p_0(x) = 1, \; p_1(x) = 1 - x.$$

Полиномы Лежандра: $p(t) = 1$, $a = -1$, $b = 1$,

$$(n+1)p_{n+1}(x) - (2n+1)xp_n(x) + np_{n-1}(x) = 0, \; p_0(x) = 1, \; p_1(x) = x.$$

Полиномы Эрмита: $p(t) = e^{-t^2}$, $a = -\infty$, $b = \infty$,

$$p_{n+1}(x) - 2xp_n(x) + 2np_{n-1}(x) = 0, \; p_0(x) = 1, \; p_1(x) = 2x.$$

Мы рассмотрели метод наименьших квадратов с нормой лишь одного типа. На самом деле любая норма, основанная на скалярном произведении, в задачах аппроксимации приводит к решению симметричной системы линейных уравнений. Что касается ортогональных базисных полиномов, то не очевидно, что их применение оправдано во иеех случаях.

Сформулируем задачу аппроксимации в общем случае. Найти такую функцию $\overline{f}$ что

$$\| f - \overline{f} \| = \min_{g \in C} \| f - g \|.$$

В этом случае *C* уже являетсячастью *W*. Могут быть введены ограничения на класс искомых функций, например:

$$C = \{g \in W;\ g'(t) \geqslant 0,\ \forall t\}.$$

В литературе описываются попытки решить эту задачу главным образом для случая выпуклого С:

$$\forall g_1,\ g_2 \in C,\ \forall \lambda,\ \mu \geqslant 0,\ \lambda + \mu = 1 \text{ влечет за собой } \lambda g_1 + \mu g_2 \in C.$$

Такимобразом, задача сводится к оптимизации.

Обширная литература посвящена решению задачи аппроксимации с помощью сплайнов с ограничениями на форму. Основная трудность здесь состоит в поиске эффективных алгоритмов.

Замечания

Практическая реализация рассмотренных алгоритмов обычно требует большого времени вычислений, большого объема памяти ЭВМ, важную роль играют также вопросы вычислительной устойчивости. Один из путей устранения указанных трудностей состоит в сегментации задачи, т. е. поочередном решении задачи на небольших областях с последующей сшивкой результатов, которая может быть осуществлена несколькими методами.

В методе подгонки общая задача разбивается на множество локальных задач, решаемых на непересекающихся интервалах. В методе сшивки границы локальных интервалов $[a_i, b_i]$ удовлетворяют соотношениям

$$a_i < b_{i-1} < a_{i+1} < b_i < a_{i+2} < b_{i+1} < \ldots.$$

Обозначим через f_i результат решения локальной задачи на отрезке $[a_i, b_i]$. Тогда окончательный результат на отрезке $[b_{i-1}, a_{1+2}]$ определяется следующим образом:

$$f(t) = \begin{cases} f_i(t),\ \forall t \in [b_{i-1}, a_{i+1}], \\ \varphi_0\left(\dfrac{t - a_{i+1}}{b_i - a_{i+1}}\right) f_i(t) + \varphi_1\left(\dfrac{t - a_{i+1}}{b_i - a_{i+1}}\right) f_{i+1}(t),\ \forall t \in [a_{i+1}, b_i], \\ f_{i+1}(t),\ \forall t \in [b_i, a_{i+2}], \end{cases}$$

где функции φ_0 и φ_1 - базисные полиномы Эрмита. Решение f является непрерывной дифференцируемой функцией, если таковыми являются функции f_i. Можно получить и другие варианты сшивки, если вместо φ_0 и φ_1 использовать другие функции.

Метод Франка с большим успехом применяется для функций многих переменных, но его можно использовать и в одномерном случае. В этом методе область Ω разбивается на локальные области Ω_i, для каждой из которых определяется неотрицательная функция w_i, равная нулю вне Ω_i. На каждой Ω_i локальной процедурой отыскивается решение f_i. Окончательный результат вычисляется по формуле

$$f(t) = \left(\sum_i w_i(t) f_i(t)\right) / \sum_i w_i(t).$$

Очевидно, что функция $f(t)$ удовлетворяет тем же условиям, что $f_i(t)$, и если w_i и f_i дифференцируемы, то это относится и к f.

Однако возможны случаи, когда сама функция или одна из ее производных имеет разрыв. Если точка разрыва известна, ее учитывают при решении задачи в качестве дополнительного условия и задача в целом остается линейной. Если известна только область, где находится точка разрыва, то решают задачу по обе стороны от этой области и вычисляют точку пересечения двух кривых, решая возникающие при этом нелинейные уравнения. В настоящее время разработаны более общие методы решения таких задач.

14.2.5. Кривые на плоскости и в пространстве

Будем считать, что кривые на плоскости и в пространстве задаются в параметрической форме и параметр, меняющийся на отрезке [0, 1], пображается в пространство R^2 или R^3. Замкнутой будет такая кривая c, для которой $c(0) = c(1)$.

Общий случай

Для данного случая интерполяция, сглаживание или аппроксимация состоят, вообще говоря, в применении изложенных выше методов к каждой координате. Поэтому здесь мы ограничимся лишь некоторыми замечаниями. Для замкнутых кривых используются периодические функции интерполяции, сглаживания или аппроксимации. Если можно воспользоваться критерием минимизации, его можно применить ко всем компонентам (координатам), например общий критерий представить как сумму критериев по отдельным компонентам.

Единственная важная проблема состоит в том, что решение будет зависеть от способа выбора параметра. Большую помощь может оказать удачный выбор системы координат: например, если искомая кривая близка к окружности или эллипсу, нужно отдать предпочтение полярным координатам. Для исходных точечных данных параметр должен изменяться в пределах [0, 1]. Например, в качестве начальных параметров для кривой на R^2, проходящей через точки $P_0, P_1, \ldots, P_n$, можно выбрать

$$t_0^0 = 0,\ t_1^0 = \frac{|P_0P_1|}{\sum\limits_i |P_iP_{i+1}|} = \frac{\text{Длина } P_0P_1}{\sum\limits_i \text{Длина } P_iP_{i+1}},$$

$$t_j^0 = \frac{\sum\limits_{k=0}^{j-1} |P_kP_{k+1}|}{\sum\limits_{i=0}^{n-1} |P_iP_{i+1}|}\ (j = 1, \ldots, n),\ t_n^0 = 1.$$

После этого можно выбрать метод решения, найти кривую G_1 и ее использовать для вычисления параметров в следующем приближении:

$$t_0^1 = 0,$$

$$t_j^1 = \frac{\sum\limits_{k=0}^{j-1} |P_kP_{k+1}|_{C_1}}{\sum\limits_{i=0}^{n-1} |P_iP_{i+1}|_{C_1}}$$ **(где длины дуг измеряются вдоль кривой C_1).**

Затем вычисляют кривую C_2 и т.д., выполняя необходимое число итераций. Процессы такого типа быстро сходятся (из 5-6 итераций) и параметры определяются из самой кривой.

Кривые Безье и *B*-сплайны

Благодаря своим свойствам в теории оптимизации часто применяются кривые Безье и *B*-сплайны (рис. 14-16).

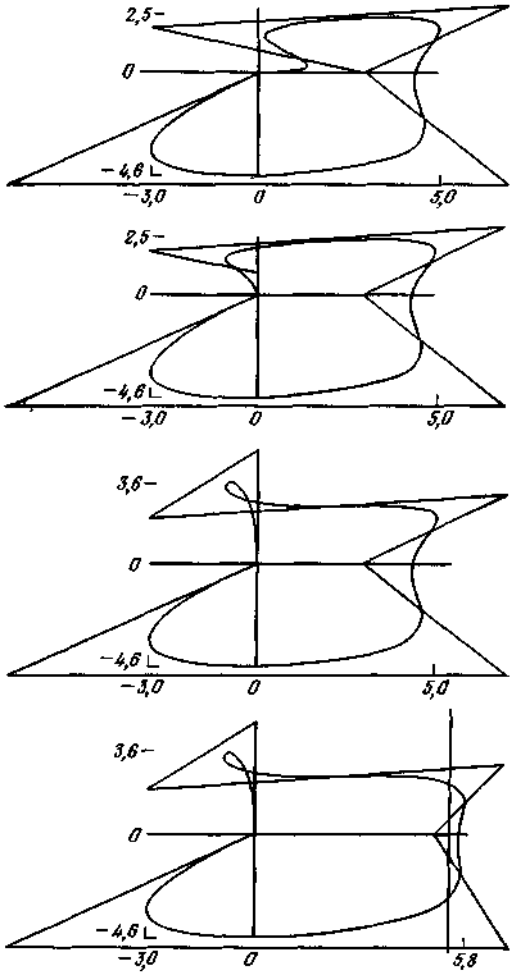


Рис. 14. Сглаживание *B*-сплайнами 3-й степени с эквидистантными узлами в интервале [0,1] на многоугольниках с координатами вершин:

(0,0), (3,0), (−3,2), (7,3), (3,0), (7,−5), (−7,−5), (0,0);
(0,0), (0,1), (−3,2), (7,3), (3,0), (7,−5), (−7,−5), (0,0);
(0,0), (0,5), (−3,2), (7,3), (5,0), (7,−5), (−7,−5), (0,0).

Отметим касательные к кривым в крайних точках многоугольников, локальное изменение кривой при изменении положения одной точки и образование петли за счет расположения данных. Сравнение с кривой Безье (рис. 12).

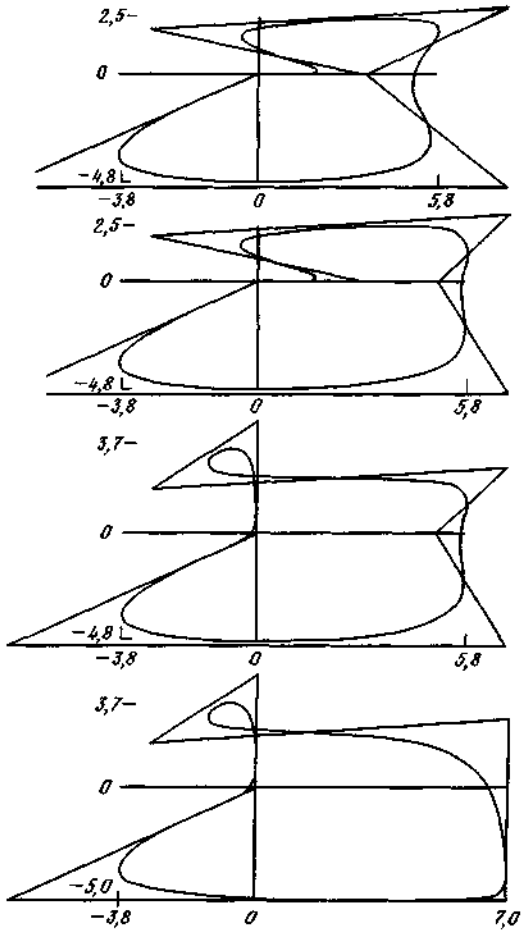


Рис. 15. Сглаживание *B*-сплайнами 3-й степени с эквидистантными узлами, периодическими на *R*, на многоугольниках с координатами вершин:

(0,0), (3,0), (−3,2), (7,3), (3,0), (7,−5), (−7,−5), (0,0);
(0,0), (3,0), (−3,2), (7,3), (5,0), (7,−5), (−7,−5), (0,0);
(0,0), (0,5), (−3,2), (7,3), (5,0), (7,−5), (−7,−5), (0,0);
(0,0), (0,5), (−3,2), (7,3), (7,−5), (7,−5), (−7,−5), (0,0).

Отметим дифференцируемость кривых, локальное изменение кривой при изменении положения одной точки, образование петли за счет расположения данных. Сравните с кривыми Безье (рис. 12) и *B*-сплайнами (рис. 14).

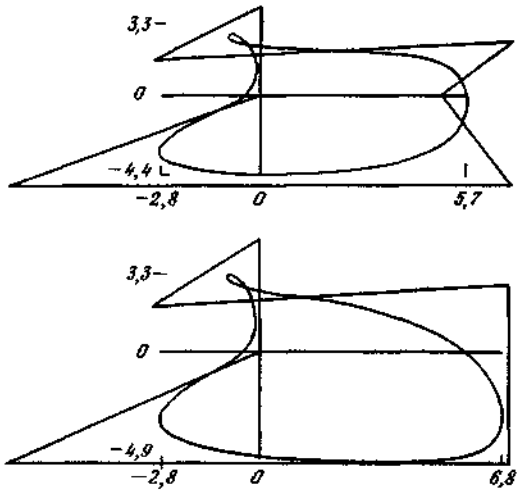


Рис. 16. Сглаживание *B*-сплайнами 5-й степени с эквидистантными узлами на *R* на многоугольниках с координатами вершин:

(0,0), (0,5), (−3,2), (7,3), (3,0), (7,−5), (−7,−5), (0,0);
(0,0), (0,5), (−3,2), (7,3), (5,0), (7,−5), (−7,−5), (0,0).

Пусть Q - множество точек в R^2 (или R^3):

$$Q = \{ Q_i;\ 0 \leqslant i \leqslant N \}.$$

Кривой Безье, связанной с Q, называется кривая, определяемая следующим образом:

$$B[Q](t) = \sum_{i=0}^{N} \Phi_{N,i}(t) Q_i,$$

т. е., если (x_i, y_i, z_i) являются координатами Q_i, то

$$x(t) = \sum_{i=0}^{N} \Phi_{N,i}(t) x_i,\ \ y(t) = \sum_{i=0}^{N} \Phi_{N,i}(t) y_i,\ \ z(t) = \sum_{i=0}^{N} \Phi_{N,i}(t) z_i.$$

В дополнение к известным свойствам (разд. 14.2.2) заметим, что кривая Безье начинается в точке Q_0 и заканчивается в точке Q_N. В точке Q_0 (соответственно Q_N) она является касательной к отрезку Q_0Q_1 (соответственно к $Q_{N-1}\ Q_N$):

$$B[Q](0) = Q_0, \quad B[Q](1) = Q_N.$$

Более того, она целиком содержится в выпуклой оболочке точек Q_i. Существенное влияние на решение оказывает пространственное расположение и порядок точек Q_i. Последовательность отрезков Q_iQ_{i+1}

называется многоугольником Безье, связанным с Q, на котором строится кривая Безье.

Из общих свойств функций Безье вытекают следующие особенности их применения в данном случае:

изменение положения хотя бы одной точки приводит к заметному изменению кривой;

добавление хотя бы одной точки приводит к необходимости решения задачи заново;

при большом числе точек вычисление функций Безье затрудняется из-за большой степени полиномов.

После приобретения практических навыков вычисления функций Безье можно довольно быстро располагать точки Q_i так, чтобы получить желаемую форму кривой после небольших модификаций.

Аналогичная кривая, связанная с Q и построенная с помощью B-сплайнов, может быть представлена в следующем виде:

$$S[Q](t) = \sum_{i=0}^{N} N_{i,n}(t) Q_i .$$

Основная их особенность состоит в локальном характере определения. Это приводит к более простому учету изменения положения отдельных точек Q. Кроме того, они обладают более высокой вычислительной устойчивостью, что связано с небольшими степенями используемых полиномов.

14.3. Поверхности

В задачах обработки поверхностей возникают затруднения, связанные с особенностями геометрических свойств исходных данных. Мы рассмотрим наиболее благоприятные случаи, когда, используя свойства тензорного произведения, можно воспользоваться уже рассмотренными методами функции одной переменной. Затем дадим представление о других, более общих методах.

14.3.1. Тензорное произведение

Рассмотрим простой случай, когда искомую функцию можно представить в виде линейной комбинации произведений функций одной переменной вида $\Phi(x)\Psi'(y)$.

Метод Гордона и Кунса

Один из методов решения такой задачи состоит в следующем.

Пусть P_1 и P_2 - два интерполяционных оператора, определенных на множестве функций, отображающих отрезок [0, 1] в R. $P_1 f$ и $P_2 f$ могут

быть, например, интерполяционными полиномами Лагранжа или Эрмита для функции f. Если R_1 и R_2 являются операторами ошибки интерполяции для P_1 и P_2 соответственно (разд. 14.2.1), то имеем следующие тождества:

$$I = P_1 + R_1; \quad I = P_2 + R_2,$$

где I- тождественный оператор.

Если g - функция двух переменных, отображающая область [0, 1]×[0,1] на R, то в дальнейшем будем применять операторы P_1 или P_2 к g, как к функции одной переменной $g(., y)$ (соответственно $g(x, .)$), рассматривая другую переменную как параметр. Поступая таким образом, легко получить следующие тождества:

$$g = P_1 P_2 g + (P_1 R_2 + P_2 R_1 + R_1 R_2) g,$$
$$g = (P_1 + P_2 - P_1 P_2) g + R_1 R_2 g.$$

Член $P_1+P_2 — P_1P_2$ называется булевой суммой и обозначается $P_1 \oplus P_2$. В каждом из этих двух тождеств присутствует интерполяционный оператор для g. Один из них $P_1 P_2 g$, другой $(P_1 \oplus P_2) g$. Приведем три примера, в которых показано их применение и различие между ними.

Интерполяция Лагранжа. Предположим, что P_1 и P_2 являются операторами интерполяции Лагранжа на точках 0 и 1:

$$P_1 f(t) = P_2 f(t) = (1 - t) f(0) + t f(1).$$

Тогда два приведенных выше оператора будут иметь вид

$$P_1 P_2 g(x, y) = (1 - x)(1 - y) g(0, 0) + (1 - x) y g(0, 1) + x(1 - y) g(1, 0) + xy g(1, 1),$$
$$(P_1 \oplus P_2) g(x, y) = (1 - x) g(0, y) + x g(1, y) + (1 - y) g(x, 0) + y g(x, 1) - (1 - x)(1 - y) g(0, 0) - (1 - x) y g(0, 1) - x(1 - y) g(1, 0) - xy g(1, 1).$$

Отсюда видно, что оба оператора выполняют интерполяцию по-разному. Для определения $P_1 P_2 g$ необходимо знать значения g только в вершинах квадрата, тогда как $P_1 \oplus P_2 g$ требует значений g на сторонах квадрата.

Интерполяция Эрмита. Рассмотрим следующий пример, в котором операторы P_1 и P_2 являются операторами интерполяции Эрмита (разд. 14.2.1):

$$P_1 f(t) = P_2 f(t) = \varphi_0(t) f(0) + \varphi_1(t) f(1) + \varphi_2(t) f'(0) + \varphi_3(t) f'(1).$$

Тогда

$$P_1 P_2 g(x, y) = [\varphi_0(x)\ \varphi_1(x)\ \varphi_2(x)\ \varphi_3(x)]\, G \begin{vmatrix} \varphi_0(y) \\ \varphi_1(y) \\ \varphi_2(y) \\ \varphi_3(y) \end{vmatrix},$$

где

$$G = \begin{vmatrix} g(0,0) & g(0,1) & \frac{\partial}{\partial y} g(0,0) & \frac{\partial}{\partial y} g(0,1) \\ g(1,0) & g(1,1) & \frac{\partial}{\partial y} g(1,0) & \frac{\partial}{\partial y} g(1,1) \\ \frac{\partial}{\partial x} g(0,0) & \frac{\partial}{\partial x} g(0,1) & \frac{\partial^2}{\partial x \partial y} g(0,0) & \frac{\partial^2}{\partial x \partial y} g(0,1) \\ \frac{\partial}{\partial x} g(1,0) & \frac{\partial}{\partial x} g(1,1) & \frac{\partial^2}{\partial x \partial y} g(1,0) & \frac{\partial^2}{\partial x \partial y} g(1,1) \end{vmatrix}$$

и

$$\begin{aligned} (P_1 \oplus P_2) g(x, y) = {} & \varphi_0(y) g(x, 0) + \varphi_1(y) g(x, 1) + \\ & + \varphi_2(y) \frac{\partial}{\partial x} g(x, 0) + \varphi_3(y) \frac{\partial}{\partial x} g(x, 1) \Big\} P_2 g(x, y) \\ & + \varphi_0(x) g(0, y) + \varphi_1(x) g(1, y) + \\ & + \varphi_2(x) \frac{\partial}{\partial y} g(0, y) + \varphi_3(x) \frac{\partial}{\partial y} g(1, y) \Big\} P_1 g(x, y) \\ & - P_1 P_2 g(x, y). \end{aligned}$$

Относительно этих операторов можно сделать те же замечания, что и в предыдущем случае. В дополнение к ним должны быть известны значения перекрестных производных $\frac{\partial^2}{\partial x \partial y}$ в вершинах квадрата.

Метод Кунса. Наконец, в третьем примере допустим, что

$$P_1 f(t) = P_2 f(t) = \varphi_0(t) f(0) + \varphi_1(t) f(1).$$

Учитывая, что $(P_1 f)'(0) = (P_1 f)'(1)$, можно записать

$$\begin{aligned} (P_1 \oplus P_2) g(x, y) = {} & \varphi_0(x) g(0, y) + \varphi_1(x) g(1, y) + \varphi_0(y) g(x, 0) + \\ & + \varphi_1(y) g(x, 1) - \varphi_0(x) \varphi_0(y) g(0, 0) - \varphi_0(x) \varphi_1(y) g(0, 1) - \\ & - \varphi_1(x) \varphi_0(y) g(1, 0) - \varphi_1(x) \varphi_1(y) g(1, 1). \end{aligned}$$

Этот метод чрезвычайно прост потому, что в нем требуется, чтобы нормальная производная функции была равна 0 на каждой стороне квадрата.

Замечания. В приведенных выше примерах куски поверхностей сшиваются непрерывным образом, а в последних двух случаях - вместе с нормальными производными по непрерывной границе. На их основе можно разрабатывать и другие методы построения поверхностей, но при этом резко возрастает сложность вычисления функций. Если в методе предусматривается использование значений функции g или ее

производных на границах квадрата, то их можно получить, решая предварительно линейную задачу для функции одной переменной.

Тензорное произведение и интерполяционные сплайны

Если известны значения функции на равномерной решетке точек $(ih, jk)_{0 \leqslant i \leqslant N, 0 \leqslant j \leqslant M}$, можно определить интерполяционный сплайн с помощью тензорного произведения. Критерию минимизации в этом случае будет удовлетворять бикубический интерполяционный сплайн. На каждой прямой, параллельной одной из осей решетки, его значения равны значениям кубического сплайна как функции одной переменной. Таким образом, вычисление бикубического сплайна сводится к вычислению $2N + M$ или $2M + N$ сплайнов одной переменной. Это свойство распространяется и на другие типы интерполяционных функций.

Возможности применения методов, использующих тензорное произведение, ограничиваются геометрическими свойствами данных. Однако эти методы настолько эффективны и просты, что нередко прибегают к преобразованию области задания исходных данных или к специальной параметризации, чтобы обеспечить необходимые для них условия.

Сплайны, являющиеся решением задачи сглаживания, обладают другими свойствами.

Поверхности Безье и *B*-сплайна

Для обработки поверхностей также широко применяются функции Безье и *B*-сплайны и метод тензорного произведения. Пусть *Q*-решетка точек в R^3:

$$Q = \{Q_{i,j},\ 0 \leqslant i \leqslant N,\ 0 \leqslant j \leqslant M\}$$

Поверхностью Безье, связанной с *Q*, называется поверхность, определяемая следующим образом

$$B[Q](s, t) = \sum_{i=0}^{N} \sum_{j=0}^{M} \Phi_{N,i}(s)\, \Phi_{M,j}(t)\, Q_{i,j},$$

где *s* и *t* принадлежат отрезку [0, 1] Аналогично поверхностью *B*-сплайна, связанной с *Q*, называется поверхность, определяемая следующим образом:

$$S[Q](s, t) = \sum_{i=0}^{N} \sum_{j=0}^{N} N_{i,N}(s)\, N_{j,M}(t)\, Q_{i,j},$$

где функции Ф и *N* определены в разд. 14 2.2 Необязательно, чтобы узлы и узловые точки совпадали в обоих направлениях. При переходе от кривых к поверхностям сохраняется важное свойство: поверхности Безье и *B*-сплайна расположены в выпуклой оболочке своей решетки На практике используются главным образом билинейные

(N=M=n=m=1), биквадратные и бикубические поверхности. Отметим, наконец, что возможно обобщение этих представлений для правильных решеток, отличных от прямоугольных.

14.3.2. Методы интерполяции для произвольно расположенных точек

Для произвольно расположенных точек в случае функции нескольких переменных не существует общей теории интерполяции Обычно при решении подобных задач налагают дополнительные условия на геометрическое расположение точек. Подход к решению задачи заключается либо в построении кусочно-полиномиальных функций (в общем случае они строятся на треугольной сетке), либо в использовании сплайнов.

Интерполяция на треугольнике

Во многих работах предпринимались попытки разработать методы, аналогичные методам Гордона и Кунса, но для треугольного элемента поверхности, причем во всех случаях стремятся свести двумерную задачу к одномерной. Приведем несколько примеров.

Рассмотрим треугольник T с вершинами $A(0,0)$, $B(1,0)$ и $C(0, 1)$ (рис. 17).

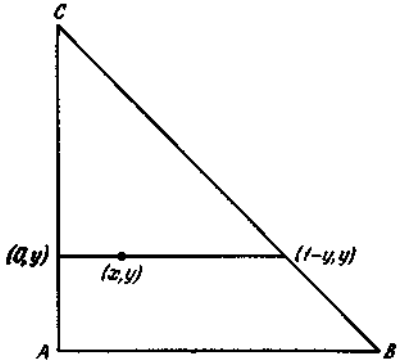


Рис. 17. Интерполяция Лагранжа функции двух переменных

Вычисления будем проводить в декартовых координатах Построим на этих точках интерполяционную поверхность Лагранжа

$$P_1 g(x, y) = \frac{1 - x - y}{1 - y} g(0, y) + \frac{x}{1 - y} g(1 - y, y)$$

Аналогично определяем P_2 вдоль направления AC. Тогда интерполяци-

онный оператор $P_1 \oplus P_2$ позволяет осуществить непрерывную сшивку (в пространстве C^0). Подобным образом можно осуществить непрерывную и дифференцируемую сшивку (в пространстве C^1), если в качестве исходных операторов взять

$$P_1 g(x, y) = \varphi_0\left(\frac{x}{1-y}\right) g(0, y) + \varphi_1\left(\frac{x}{1-y}\right) g(1-y, y) +$$
$$+ \varphi_2\left(\frac{x}{1-y}\right)(1-y)\frac{\partial}{\partial x} g(0, y) + \varphi_3\left(\frac{x}{1-y}\right)\frac{\partial}{\partial x} g(1-y, y)$$

и P_2, который определяется вдоль направления AC. Булева сумма приводит к появлению перекрестных производных, при вычислении которых возникают большие трудности. Вместо использования направлений, параллельных сторонам треугольника, можно взять направление из вершины треугольника на противоположную сторону. Разработан другой метод, в котором каждой вершине треугольника ставятся в соответствие координатные оси, образованные прилегающими сторонами, и в каждой паре осей осуществляется интерполяция с помощью булевой суммы, а окончательное решение берется в виде комбинации полученных функций. При вычислениях в каждом треугольнике вместо декартовых координат обычно используют так называемые барицентрические координаты, в которых с произвольной точкой P внутри треугольника $A_1A_2A_3$ (рис. 2.18) связываются три положительных или равных нулю числа

$$\lambda_i = \frac{\text{Площадь треугольника } PA_jA_k}{\text{Площадь треугольника } A_1A_2A_3}; \quad i, j, k \in \{1, 2, 3,\}, \quad i \neq j \neq k.$$

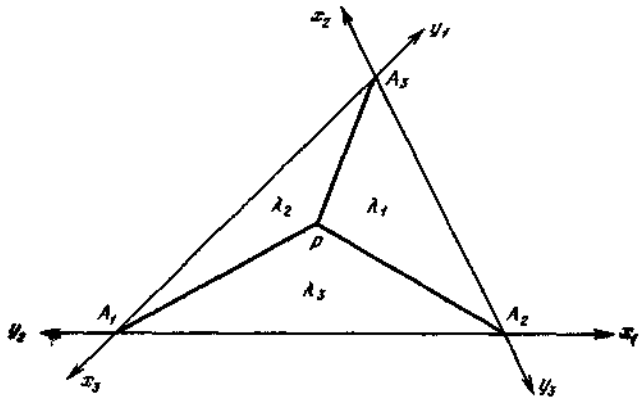


Рис. 18. Интерполяция с использованием барицентрических координат.

Затем вычисляют интерполяционную функцию класса C^0, оставляя инвариантными полиномы первой степени:

$$Pg(\lambda_1, \lambda_2, \lambda_3) = \lambda_1 [g(1-\lambda_2, \lambda_2, 0) + g(1-\lambda_3, 0, \lambda_3) - g(1, 0, 0)] + \lambda_2 [g(\lambda_1, 1-\lambda_1, 0) + g(0, 1-\lambda_3, \lambda_3) - g(0, 1, 0)] + \lambda_3 [g(\lambda_1, 0, 1-\lambda_1) + g(0, \lambda_2, 1-\lambda_2) - g(0, 0, 1)].$$

Достоинством этих методов является возможность непрерывной сшивки самих поверхностей или поверхностей вместе со своими нормальными производными. При их использовании могут возникнуть трудности, связанные с разбиением области на треугольники и с выбором метода сшивки.

Описанные методы позволяют строить некоторые конечные элементы.

Конечные треугольные элементы

Вкратце суть проблемы состоит в следующем. Конечный элемент определяется как триплет *(K, P,* $\sum$), где

1) *K* - замкнутый многогранник из R^2 (в данном случае - треугольник);

2) $P \subset C^s(K)$ - пространство действительных функций, отображающих *K* на *R*, непрерывных и *s* раз дифференцируемых;

3) $\sum$ - конечное множество линейных, линейно независимых функционалов, определенных на $C^s(k)$ и таких, что $\forall l_i \in \Sigma$, $\forall z \in R^N$, $\exists$ единственный $p \in P$, $l_i(p) = z_i$, $i = 1, \ldots, N$.

Конечный элемент считается принадлежащим классу *k*, если интерполирующая функция, построенная на данном разбиении, непрерывна вместе со своими производными до *k*-го порядка. Обычно в качестве *P* берут пространство полиномов, базис которого $\{p_i\}$ удовлетворяет условию

$$l_i\{p_j\} = \delta_{ij}.$$

Полиномы здесь также представляются в барицентрических координатах.

Приведем два примера конечных элементов.

Пример 1. Конечный элемент Лагранжа первой степени класса C^0:

P- пространство полиномов первой степени;

$$\Sigma = \{l_1, l_2, l_3\}; \quad l_i(f) = f(A_i); \quad p_i(\lambda_1, \lambda_2, \lambda_3) = \lambda_i,$$

где A_i - вершины треугольника $K = A_1A_2A_3$.

Пример 2. Конечный элемент Эрмита класса C^1. Если использовать только полиномы, для его построения потребуется 18-мерное пространство, определяемое так, что в вершинах треугольника задаются значения самой функции и ее первой и второй производных.

В качестве P могут быть использованы также пространства рациональных или кусочно-полиномиальных функций. Для сохранения границ области иногда используют криволинейные треугольники. В теории изопараметрических элементов изучаются преобразования перехода от обычных к криволинейным конечным элементам с сохранением необходимых свойств.

Интерполяционные сплайны и нерегулярные данные

Для неравномерно расположенных исходных данных поиск интерполяционных сплайнов от нескольких переменных осуществляется теми же способами, что и для одной переменной, среди которых выше были рассмотрены:

а) применение кусочно-полиномиальных функций с определенными условиями сшивки;

б) минимизация критерия.

Первый способ разрабатывался многими авторами, однако до сих пор нет такого метода, который можно было бы действительно применить на практике. Разработка метода минимизации критерия основывается на теории воспроизводящих ядер. Разработаны методы решения задач на ограниченных областях R^2. В некоторых работах эта теория распространена на неограниченные множества.

Таким образом, минимизация критерия (аналог энергии изгиба бесконечной пластины)

$$\iint_{R^2} \left(\frac{\partial^2}{\partial x^2} \mu(x, y) \right)^2 + 2 \left(\frac{\partial^2}{\partial x \partial y} \mu(x, y) \right)^2 + \left(\frac{\partial^2}{\partial y^2} \mu(x, y) \right)^2 \partial x \partial y$$

с интерполяционными условиями

$$\mu(x_i, y_i) = z_i, \quad i = 1, \ldots, n,$$

приводит к определению сплайна следующего вида:

$$\sigma(x, y) = \sum_{i=1}^{n} \lambda_i K(x - x_i, y - y_i) + \alpha x + \beta y + \gamma,$$

где функция K (полугильбертово ядро) связана с выбранным критерием

$$K(x, y) = \frac{1}{2}(x^2 + y^2) \ln (x^2 + y^2),$$

а λ_i, α, β, γ определяются из решения линейной системы уравнений:

$$\sum_{j=1}^{n} K(x_i - x_j, y_i - y_j)\lambda_j + \alpha x_i + \beta y_i + \gamma = 0, \quad i = 1, \ldots, n$$

$$\sum_{j=1}^{n} \lambda_j x_j = 0,$$

$$\sum_{j=1}^{n} \lambda_j y_j = 0,$$

$$\sum_{j=1}^{n} \lambda_j = 0.$$

Существует много численных методов решения этой задачи, в частности и такие, в которых не надо упорядочивать интерполяционные точки. Если n велико (>130), решение системы и вычисление сплайнов затруднительно. Для $n < 80$ метод Гаусса дает вполне удовлетворительные результаты, если первые точки выбирать удаленными друг от друга.

Возможно применение и других критериев минимизации, которые приводят к различным видам ядер: $K(x, y) = (x^2 + y^2)^{\theta/2}$ (0 - действительное положительное нецелое число);

$$K(x, y) = \frac{1}{2}(x^2 + y^2)^k \ln (x^2 + y^2).$$

В зависимости от значений 0 и k полученный сплайн будет иметь большее или меньшее число производных.

Сшивка

Интерполяционные процедуры могут быть выполнены на расположенных рядом треугольных и прямоугольных элементах так, чтобы обеспечить построение непрерывных и дифференцируемых поверхностей при условии, что эти процедуры должны быть класса C^1 с общими данными на границах. Как и в случае одной переменной, решение задачи сразу для всей области может быть весьма трудоемким, поэтому часто разбивают задачу на несколько локальных задач с последующей сшивкой решений. Предположим, например, что поверхности S_{ij} были вычислены в результате решения локальных задач на областях $R_{ij}((a_i, b_i) \times (c_j, d_j))$ (рис. 19).

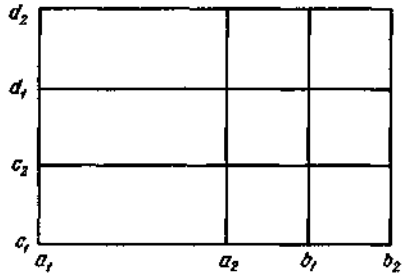


Рис. 19. Разбиение области интерполяции.

Тогда окончательное решение определяется следующей формулой:

$$S(x, y) = \sum_{i,j=1}^{2} \Psi_i(x)\Phi_j(y)S_{ij}(x, y),$$

где

$$\Psi_1(x) = \begin{cases} 1 & a_1 \leqslant x \leqslant a_2, \\ \varphi_0\left(\dfrac{x - a_2}{b_1 - a_2}\right) & a_2 \leqslant x \leqslant b_1, \\ 0 & b_1 \leqslant x \leqslant b_2, \end{cases} \quad \Phi_1(y) = \begin{cases} 1 & c_1 \leqslant y \leqslant c_2, \\ \varphi_0\left(\dfrac{y - c_2}{d_1 - c_2}\right) & c_2 \leqslant y \leqslant d_1, \\ 0 & d_1 \leqslant y \leqslant d_2, \end{cases}$$

$$\Psi_2(x) = 1 - \Psi_1(x), \quad \Phi_2(y) = 1 - \Phi_1(y).$$

Можно также использовать метод Франка (разд. 14.2.4), определяя весовые функции для случая R^2. Если выбрать весовые функции для каждой точки в задаче интерполяции таким образом, что

$$w_i(x, y) = (R_i^2 - ((x - x_i)^2 + (y - y_i)^2))_+,$$

тогда можно вычислить локальные поверхности на дисках с центром (x_i, y_i) и радиусом R_i.

Рассмотренные методы сшивки предполагают сегментацию поверхности с целью упрощения вычислений и предполагают заданными значения функции и ее производной. Однако часто в задачах интерполяции известны только значения самой функции, но не ее производных (экспериментальные результаты, численные расчеты и т. п.). В этом случае можно предложить следующий алгоритм вычислений:

- произвести разбиение области на треугольники, взяв в качестве вершин треугольников точки, в которых известны значения функции;
- выбрать метод интерполяции на области треугольной формы. Ввести дополнительное множество точек, необходимое для реализации метода;

• вычислить в каждой дополнительной точке P недостающие данные с помощью интерполяционного сплайна, определенного по известным значениям в вершинах треугольника, содержащего P;
• провести интерполяционную поверхность на каждом треугольнике.

Если позволяют условия, следует выбрать равномерное разбиение. В этом случае необходимо:
• определить равномерное разбиение;
• выбрать метод интерполяции и определить множество точек, как и раньше;
• каждой точке P поставить в соответствие точки данных θ_p, лежащие на многогранниках, формирующих сетку разбиения и содержащих P.

На θ_p строят локальную поверхность, с помощью которой вычисляют недостающие точки.

14.3.3. Автоматическое разбиение на треугольники

Рассматриваемая задача разбиения на треугольники аналогична задаче, возникающей в методе конечных элементов с неравномерно расположенными данными. При этом возможны два случая:
• автоматическое разбиение области, заданной своими границами;
• автоматическое разбиение на заданных точках.

Решение определяется выбором критерия оптимального разбиения. Рассмотрим два метода предварительного разбиения, а затем их усовершенствование.

Разбиение на области

Для автоматического разбиения заданной области Ω из R^2 предположим, что она является связной (без дыр, в противном случае достаточно ввести фиктивные вспомогательные границы) и известна ее граница. Это равносильно тому, что заранее известны вершины разбиения на границе. Идея метода состоит в том, чтобы, начиная от известных вершин $I_1, \ldots, I_n$, строить остальные точки разбиения, постепенно продвигаясь в глубь области. Можно выделить следующие этапы этой работы:

• *Начальный этап*: выбираются три последовательные точки, расположенные на границе, и упорядочивается их нумерация в направлении по часовой стрелке. Область Ω представим в виде 2 частей: τ-часть, где разбиение уже проведено, D - оставшаяся часть. В исходном состоянии имеем

$$F = \{I_1, I_2, I_3\}, \quad \tau = \varnothing, \quad D = \Omega.$$

• Пока область D не сведена к отрезку, для

$$F = \{I_1, \ldots, I_n\}$$

отыскиваем такое p, чтобы $\theta = \widehat{I_{p-1} I_p I_{p+1}} = \inf \widehat{I_{i-1} I_i I_{i+1}}$, где $\widehat{I_{p-1} I_p I_{p+1}}$ обозначает угол, образованный отрезками

$$I_{p-1} I_p \text{ и } I_p I_{p+1}.$$

- Находим k, удовлетворяющее условию

$$k\frac{\pi}{3} < \theta \leqslant (k+1)\frac{\pi}{3}.$$

- Если $k = 0$:
- строим треугольник T с вершинами $I_{p-1} I_p I_{p+1}$,
- $\tau \to \tau \cup T$, $D = D - T$;
- исключаем точку I_p из F.
- Если $k > 0$:
- строим точки $P_1, ..., P_k$ таким образом, чтобы

$$\widehat{I_{p-1} I_p P_j} = \frac{j\theta}{k+1}$$

и треугольники $T_0 = I_{p-1} I_p P_1$,

$$T_i = P_i I_p P_{i+1}, \quad T_k = P_k I_p I_{p+}$$

были бы максимально близки к равносторонним;

- строим треугольники $T_0, \ldots, T_k$;
- $\tau \to \tau \cup T_0 \cup \ldots \cup T_k$, $D = D - (T_0 \cup \ldots \cup T_k)$;
- исключаем I_p из F и добавляем в F $P_{1, ...,} P_k$;
- если $p = 2$, добавляем к F граничную точку, предшествующую I_1;
- если $p = n$ — 1, добавляем к F граничную точку, следующую за I_n.

Разбиение на заданных точках

Автоматическое разбиение на известных точках заключается в построении треугольников с вершинами в данных точках. Множество этих точек обозначим через $P = \{Q_1, \ldots, Q_2\}$ и выпуклую замкнутую границу множества P- через C. Часть точек n_f будет лежать на границе, а часть n_i - внутри C.

Тогда

$$n = n_f + n_i,$$
$$n_t = n_f + 2(n_i - 1) \leqslant 2n,$$
$$n_c = 2n_f + 3(n_i - 1) \leqslant 3n,$$

где n_i - число треугольников, n_c - число сторон. Укажем на один из возможных алгоритмов решения задачи, близкий по своей идее к рассмотренному выше.

Согласно этому алгоритму, построение начинается с точек, лежащих на границе. Методы определения выпуклой граничной оболочки рассмотрены в известных работах. Единственное отличие от описанного выше алгоритма состоит в построении точек, поскольку используются уже известные точки. Как и раньше, строятся точки P_1, а затем внутри окружности заданного радиуса определяется, имеются ли еще неиспользованные точки Если их несколько, берут наиболее близкую к границе, если их нет - увеличивают радиус окружности.

Существуют также другие алгоритмы, использующие иные принципы разбиения. В них предусматривается построение начального треугольника, от которого разбиение распространяется на всю область

Оптимизация разбиения на треугольники

Для оптимизации разбиения необходимо строить треугольники, максимально близкие к равносторонним. Решение подобной задачи ни в общем виде, ни для локальных четырехугольных областей не получено. На практике обычно задают начальное разбиение, а затем его последовательно оптимизируют в четырехугольных областях. Положение точек можно немного изменять (разд. 14.2.3). Для выпуклого четырехугольника отметим две возможности. Первая заключается в максимизации минимального внутреннего угла каждого треугольника, вторая возможность заключается в следующем Определим область

$$D_i = \{P \in R^2 / d(P, Q_i) < d(P, Q_j),\ j \neq i,\ \forall j = 1, 2, 3, 4\},\ i = 1, \ldots 4.$$

Если соединить точки Q_i и Q_k и при этом D_i и D_k имеют общую границу, то такое разбиение называется разбиением Делоне. Оно может быть проведено для множества произвольных точек.

15. Сходимость

15.1. Введение в сходимость

Сходимость, математическое понятие, означающее, что некоторая переменная величина имеет *предел*. В этом смысле говорят о сходимости последовательности, сходимости ряда, сходимости бесконечного произведения, сходимости непрерывной дроби, сходимости интеграла и т. д. Понятие сходимость возникает, например, когда при изучении того или иного математического объекта строится последовательность более простых в известном смысле объектов, приближающихся к данному, то есть имеющих его своим пределом (так, для вычисления длины окружности используется

последовательность длин периметров правильных многоугольников, вписанных в окружность; для вычисления значений функций используются последовательности частичных сумм рядов, которыми представляются данные функции, и т. п.).

Сходимость последовательности $\{a_n\}$, $n = 1, 2,...$, означает существование у неё конечного предела

$$\lim_{n\to\infty} a_n = a;$$

сходимость ряда $\sum_{n=1}^{\infty} u_k$ - конечного предела (называемого суммой ряда) у последовательности его частичных сумм $S_n = \sum_{k=1}^{n} u_k$, $n = 1,2,\ldots,C$; сходимость бесконечного произведения $b_1b_2...b_n$ - конечного предела, не равного нулю, у последовательности конечных произведений $p_n=b_1b_2...b_n$, $n = 1, 2,...$; сходимость интеграла $\int_a^b f(x)dx$ от функции $f(x)$, интегрируемой по любому конечному отрезку $[a, b]$,- конечного предела у интегралов при $b \to +\mu$, называется *несобственным интегралом*

$$\int_a^{+\infty} f(x)dx.$$

Свойство сходимости тех или иных математических объектов играет существенную роль как в вопросах теории оптимизации. Например, часто используется представление каких-либо величин или функций с помощью сходящихся рядов; так, для основания натуральных логарифмов e имеется разложение его в сходящийся ряд

$$e = 1 + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \ldots + \frac{1}{n!} + \ldots$$

для функции $\sin x$ - в сходящийся при всех x ряд

$$\sin a = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \ldots + (-1)^n \frac{x^{2n+1}}{(2n+1)!} + \ldots$$

Подобные ряды могут быть использованы для приближённого вычисления рассматриваемых величин и функций. Для этого достаточно взять сумму нескольких первых членов, при этом чем больше их взять, тем с большей точностью будет получено нужное значение. Для одних и тех же величин и функций имеются различные ряды, суммой которых они являются, например,

$$\ln \frac{3}{2} = \frac{1}{2} - \frac{1}{2} \cdot \frac{1}{2^2} + \frac{1}{3} \cdot \frac{1}{2^3} - \frac{1}{4} \cdot \frac{1}{2^4} + \ldots + (-1)^{n+1} \frac{1}{2^n n} + \ldots ,$$

$$\ln \frac{3}{2} = \frac{2}{5}\left(1 + \frac{1}{3} \cdot \frac{1}{5^2} + \frac{1}{5} \cdot \frac{1}{5^4} + \ldots + \frac{1}{2n+1} \cdot \frac{1}{5^{2n}} + \ldots\right).$$

При практических вычислениях в целях экономии числа операций (а следовательно, экономии времени и уменьшения накопления ошибок) целесообразно из имеющихся рядов выбрать ряд, который сходится «более быстро». Если даны два сходящихся ряда $\sum_{n=1}^{\infty} u_k$ и $\sum_{n=1}^{\infty} v_k$, и $r_n = u_{n+1} + u_{n+2} + \ldots$, $\rho_n = v_{n+1} + v_{n+2} + \ldots$ - их остатки, то 1-й ряд называется сходящимся быстрее 2-го ряда, если

$$\lim_{n \to \infty} \frac{r_n}{\rho_n} = 0 .$$

Например, ряд

$$1 + \frac{1}{2^3} + \frac{1}{3^3} + \ldots + \frac{1}{n^3} + \ldots$$

сходится быстрее ряда

$$1 + \frac{1}{2^2} + \frac{1}{3^2} + \ldots + \frac{1}{n^2} + \ldots .$$

Используются и другие понятия «более быстро» сходящихся рядов. Существуют различные методы улучшения сходимости рядов, то есть методы, позволяющие преобразовать данный ряд в «более быстро» сходящийся. Аналогично случаю рядов вводится понятие «более быстрой» сходимости и для несобственных интегралов, для которых также имеются способы улучшения их сходимость.

Большую роль понятие сходимость играет при решении всевозможных уравнений (алгебраических, дифференциальных, интегральных), в частности при нахождении их численных приближённых решений. Например, с помощью *последовательных приближений метода* можно получить последовательность функций, сходящихся к соответствующему решению данного обыкновенного дифференциального уравнения, и тем самым одновременно доказать существование при определённых условиях решения и дать метод, позволяющий вычислить это решение с нужной точностью. Как для обыкновенных дифференциальных уравнений, так и уравнений с частными производными существует хорошо разработанная теория различных сходящихся конечноразностных методов их численного решения (см. *Сеток метод*). Для практического нахождения приближённых решений уравнений широко используются ЭВМ.

Если изображать члены a_n последовательности $\{a_n\}$ на числовой прямой, то сходимость этой последовательности к a означает, что расстояние между точками a_n и a становится и остаётся сколь угодно малым с возрастанием n. В этой формулировке понятие сходимость обобщается на последовательности точек плоскости, пространства и более общих объектов, для которых может быть определено понятие расстояния, обладающее обычными свойствами расстояния между точками пространства (например, на последовательности векторов, матриц, функций, геометрических фигур и т. д., см. *Метрическое пространство*). Если последовательность $\{a_n\}$ сходится к a, то вне любой окрестности точки a лежит лишь конечное число членов последовательности. В этой формулировке понятие сходимость допускает обобщение на совокупности величин ещё более общей природы, в которых тем или иным образом введено понятие окрестности (см. *Топологическое пространство*).

В теории оптимизации используются различные виды сходимости последовательности функций $\{f_n(x)\}$ к функции $f(x)$ (на некотором множестве M). Если $\lim_{n\leftarrow\infty} f_n(x_0) = f(x_0)$ для каждой точки X_0 (из M), то говорят о сходимости в каждой точке [если это равенство не имеет места лишь для точек, образующих множество меры нуль (см. *Мера множества*), то говорят о сходимости почти всюду]. Несмотря на свою естественность, понятие сходимость в каждой точке обладает многими нежелательными особенностями [например, последовательность непрерывных функций может сходиться в каждой точке к разрывной функции; из сходимости функций $f_n(x)$ к $f(x)$ в каждой точке не следует, вообще говоря, сходимость интегралов от функций $f_n(x)$ к интегралу от $f(x)$ и т. д.]. В связи с этим было введено понятие равномерной сходимости, свободное от этих недостатков: последовательность $\{f_n(x)\}$ называется равномерно сходящейся к $f(x)$ на множестве M, если

$$\lim_{n\leftarrow\infty} \sup_{x\in M} \left| f_n(x) - f(x) \right| = 0$$

Этот вид сходимости соответствует определению расстояния между функциями $f(x)$ и $\varphi(x)$ по формуле

$$r(f,\varphi) = \sup_{x\in M} \left| f(x) - \varphi(x) \right| = 0$$

Д. Ф. Егоров доказал, что если последовательность измеримых функций сходится почти всюду на множестве M, то из M можно так удалить часть сколь угодно малой меры, чтобы на оставшейся части имела место равномерная сходимость.

В теории интегральных уравнений, ортогональных рядов и т. д. широко применяется понятие средней квадратической сходимости: последовательность $\{f_n(x)\}$ сходится на отрезке $[a, b]$ в среднем квадратическом к $f(x)$, если

$$\lim_{n\to\infty} \int_a^b \left[f(x) - f_n(x) \right]^2 dx = 0 .$$

Более общо, последовательность $\{f_n(x)\}$ сходится в среднем с показателем p к $f(x)$, если

$$\lim_{n\to\infty} \int_a^b |f(x) - f_n(x)|^p dx = 0 .$$

Эта сходимость соответствует заданию расстояния между функциями по формуле

$$\left[\int_a^b |f(x) - \varphi(x)|^p dx\right]^{1/p} .$$

Из равномерной сходимости на конечном отрезке вытекает сходимость в среднем с любым показателем p. Последовательность частичных сумм разложения функции $f(x)$ с интегрируемым квадратом по нормированной *ортогональной системе функций* может расходиться в каждой точке, но такая последовательность всегда сходится к $f(x)$ в среднем квадратическом. Рассматриваются также другие виды сходимости. Например, сходимость по мере: для любого $\varepsilon > 0$ мера множества тех точек, для которых $|f_n(x) - f(x)| < \varepsilon$, стремится к нулю с возрастанием n, слабая сходимость:

$$\lim_{n\to\infty} \int_a^b f_n(x)\varphi(x)dx = \int_a^b f(x)\varphi(x)dx$$

для любой функции $f(x)$ с интегрируемым квадратом (например, последовательность функций $\sin x$, $\sin 2x$,..., $\sin nx$,... слабо сходится к нулю на отрезке $[-p, p]$, так как для любой функции $f(x)$ с интегрируемым квадратом коэффициенты $b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} \varphi(x) \sin nx dx$ ряда Фурье стремятся к нулю).

Указанные выше и многие другие понятия сходимости последовательности функций систематически изучаются в функциональном анализе, где рассматриваются различные линейные пространства с заданной нормой (расстоянием до нуля) - так называемые банаховы пространства. В таких пространствах можно ввести понятия сходимость функционалов, операторов и т. д.,

определяя для них соответствующим образом норму. Наряду со сходимостью по норме (так называемой сильной сходимостью), в банаховых пространствах рассматривается слабая сходимость, определяемая условием $\lim_{n\to\infty} \varphi(f_n) = \varphi(f)$ для всех линейных функционалов; введённая выше слабая сходимость функций соответствует рассмотрению нормы $\left[\int_a^b |f(x)|dx\right]^{1/2}$. В математике рассматривается также сходимость по частично упорядоченным множествам (см. *Упорядоченные и частично упорядоченные множества*). В теории вероятностей для последовательности случайных величин употребляются понятия сходимость с вероятностью 1 и сходимость по вероятности.

Ещё математики древности (Евклид, Архимед) по существу употребляли бесконечные ряды для нахождения площадей и объёмов. Доказательством сходимости рядов им служили вполне строгие рассуждения по схеме *исчерпывания метода*. Термин «сходимость» в применении к рядам был введён в 1668 Дж. Грегори при исследовании некоторых способов вычисления площади круга и гиперболического сектора. Математики 17 в. обычно имели ясное представление о сходимостиь употребляемых ими рядов, хотя и не проводили строгих с современной точки зрения доказательств сходимости. В 18 в. широко распространилось употребление в анализе заведомо расходящихся рядов (в частности, их широко применял Л. Эйлер). Это, с одной стороны, привело впоследствии ко многим недоразумениям и ошибкам, устранённым лишь с развитием отчётливой теории сходимости, а с другой - предвосхитило современную теорию *суммирования* расходящихся рядов. Строгие методы исследования сходимости рядов были разработаны в 19 в. (О. Коши, Н. Абель, К. Вейерштрасс, Б. Больцано и др.). Понятие равномерной сходимости было введено Дж. Стоксом. Дальнейшие расширения понятия сходимость были связаны с развитием теории функций, функционального анализа и топологии.

15.2. Скорость сходимости

Скорость сходимости является одной из основных характеристикик методов оптимизации.

Пусть $\{x_n\}$ — последовательность приближений рассматриваемого алгоритма нахождения корня *x** некоторого уравнения, тогда:

Говорят, что метод обладает *линейной сходимостью*, если $\exists \alpha \in [0,\ 1]: \quad \exists N \in \mathbb{N},\ \forall n \geq N \quad ||x_n - x^*|| < \alpha ||x_{n-1} - x^*||$.

Говорят, что метод обладает *сходимостью степени* β, если $\exists \alpha \in [0,\ 1]: \quad \exists N \in \mathbb{N},\ \forall n \geq N \quad ||x_n - x^*|| < \alpha ||x_{n-1} - x^*||^{\beta}$.

Отметим, что обычно скорость сходимости методов не превышает квадратичной. В редких случаях метод может обладать кубической скоростью сходимости (метод Чебышева).

Пусть $\{x_n\}$— последовательность приближений рассматриваемого алгоритма нахождения корня *x** некоторого уравнения, тогда скорость сходимости β определяют из уравнения:

$$||x_i - x_n|| < \alpha ||x_{i-1} - x_n||^{\beta}$$

Для упрощения его переписывают в виде:

$$\log ||x_i - x_n|| < \log \alpha + \beta \log ||x_{i-1} - x_n||$$

Непосредственно скорость сходимости оценивают по тангенсу угла наклона логарифмического графика зависимости $||x_i - x_n||$ от $||x_{i-1} - x_n||$.

Рассмотрим использование методов сходимости на примерах методов безусловной оптимизации, а именно: градиентном и Ньютона.

Градиентный метод

1. Эвристические соображения. Проанализируем один из наиболее важных в идейном отношении метод безусловной оптимизации – градиентный. Это метод, редко применяемый на практике в «чистом виде», служат моделью для построения более реалистических алгоритмов. На примере данных методов будет подробно разобран вопрос о сходимости — будут даны различные доказательства сходимости, описана общая техника построения доказательств, обсуждены соотношения между теоретическими результатами о сходимости и практическим использованием методов.

Предположим, что в любой точке x можно вычислить градиент функции $\nabla f(x)$. В такой ситуации наиболее простым методом минимизации $f(x)$ является *градиентный,* в котором, начиная с некоторого начального приближения x^0, строится итерационная последовательность

$$x^{k+1} = x^k - \gamma_k \nabla f(x^k), \qquad (1)$$

где параметр $\gamma_k \geq 0$ задает длину шага. К методу (1) можно прийти из разных соображений.

Во-первых, при доказательстве необходимых условий экстремума можно использовать то обстоятельство, что если в точке x условие экстремума не выполняется ($\nabla(x) \neq 0$), то значение функции можно уменьшить, перейдя к точке $x - \tau \nabla(x)$ при достаточно малом $\tau > 0$. Итеративно применяя этот прием, приходим к методу (1).

Во-вторых, в точке x^k дифференцируемая функция $f(x)$ приближается линейной $f_k(x) = f(x^k) + (\nabla f(x^k), x - x^k)$ с точностью до членов порядка $o(x-x^k)$. Поэтому можно искать минимум аппроксимации $f_k(x)$ в окрестности x^k. Например, можно задаться некоторым ε_k и решить вспомогательную задачу

$$\min_{\|x - x^k\| \leq \varepsilon_k} f_k(x). \qquad (2)$$

Ее решение естественно принять за новое приближение x^{k+1}. Можно остаться в окрестности x^k и иначе, добавив к $f_k(x)$ «штраф» за отклонение от x^k. Например, можно решить вспомогательную задачу

$$\min [f_k(x) + \alpha_k \|x - x^k\|^2] \qquad (3)$$

и ее решение взять в качестве x^{k+1}. Читателю предоставляется убедиться в том, что решение задач (2), (3) задается формулой (1).

В-третьих, можно в точке x^k выбрать *направление* локального *наискорейшего спуска,* т. е. то направление y^k, $\|y^k\| = 1$, для которого достигается минимум $f'(x^k; y)$. Используя формулу

$$f'(x;\ y)=\varphi'(0)=(\ \nabla f(x), y\ )$$

для производной по направлению, получаем

$$y^k = \underset{\|y\|=1}{\operatorname{argmin}} (\nabla f(x^k),\ y) = -\nabla f(x^k)/\|\nabla f(x^k)\|. \qquad (4)$$

Таким образом, направление наискорейшего спуска противоположно направлению градиента,

Мы привели здесь столь подробно эти соображения, поскольку они же будут использоваться при построении методов оптимизации в более сложных ситуациях (например, при наличии ограничений). Однако в этих ситуациях они могут привести к различным методам.

2. Сходимость. Рассмотрим простейший вариант градиентного метода, в котором $\gamma_k \equiv \gamma$:

$$x^{k+1} = x^k - \gamma \nabla f(x^k). \qquad (5)$$

Нас будет интересовать поведение этого метода при различных предположениях относительно $f(x)$ и γ.

Теорема 1. *Пусть $f(x)$ дифференцируема на $\mathbf{R}^n$, градиент $f(x)$ удовлетворяет условию Липшица:*

$$\|\nabla f(x) - \nabla f(y)\| \leqslant L\|x - y\|, \qquad (6)$$

$f(x)$ ограничена снизу:

$$f(x) \geqslant f^* > -\infty \qquad (7)$$

и γ удовлетворяет условию

$$0 < \gamma < 2/L. \qquad (8)$$

Тогда в методе (5) *градиент стремится к* 0:

$$\lim_{k\to\infty} \nabla f(x^k) = 0,$$

а функция $f(x)$ монотонно убывает: $f(x^{k+1}) \leqslant f(x^k)$.

Доказательство. Подставим в формулу градиента функции

$$f(x+y) = f(x) + \int_0^1 (\nabla f(x+\tau y),\ y)\, d\tau =$$

$$= f(x) + (\nabla f(x),\ y) + \int_0^1 (\nabla f(y+\tau y) - \nabla f(x),\ y)\, d\tau.$$

$x = x^k$, $y = -\gamma \nabla f(x^k)$ и воспользуемся (6):

$$f(x^{k+1}) = f(x^k) - \gamma \|\nabla f(x^k)\|^2 - \gamma \int_0^1 (\nabla f(x^k - \tau\gamma\nabla f(x^k)) -$$

$$- \nabla f(x^k),\ \nabla f(x^k))\, d\tau \leqslant f(x^k) - \gamma \|\nabla f(x^k)\|^2 +$$

$$+ L\gamma^2 \|\nabla f(x^k)\|^2 \int_0^1 \tau\, d\tau = f(x^k) - \gamma\left(1 - \frac{1}{2} L\gamma\right) \|\nabla f(x^k)\|^2.$$

Суммируя неравенства

$$f(x^{k+1}) \leqslant f(x^k) - \alpha \|\nabla f(x^k)\|^2, \quad \alpha = \gamma(1 - L\gamma/2) \tag{9}$$

по *k* от 0 до *s*, получаем

$$f(x^{s+1}) \leqslant f(x^0) - \alpha \sum_{k=0}^{s} \|\nabla f(x^k)\|^2.$$

Поскольку $\alpha > 0$ в силу (8), то

$$\sum_{k=0}^{s} \|\nabla f(x^k)\|^2 \leqslant \alpha^{-1}(f(x^0) - f(x^{s+1})) \leqslant \alpha^{-1}(f(x^0) - f^*)$$

при всех *s*, т. е. $\sum_{k=0}^{\infty} \|\nabla f(x^k)\|^2 < \infty$. Отсюда $\|\nabla f(x^k)\| \to 0$.

Покажем, что все условия этой теоремы существенны. Нарушения условия (6) могут быть двух типов. Во-первых, функция *f(x)* может быть недостаточно гладкой в какой-либо точке. Пусть, например, $f(x) = \|x\|^{1+\alpha}$, $0 < \alpha < 1$. Эта функция дифференцируема, но ее градиент не удовлетворяет условию Липшица, так как $\|\nabla f(x) - \nabla f(0)\| / \|x - 0\| = (\alpha + 1)\|x\|^{\alpha-1} \to \infty$ при $||x|| \to 0$. В этом случае будет $\gamma\|\nabla f(x^k)\| \gg \|x^k - x^*\| = \|x^k\|$ при малых $||x^k||$, т. е. шаг в методе (5) получается большим и монотонность убывания *f(x)* нарушается. Во-вторых, (6) не выполняется для функций, растущих быстрее квадратичной. Пусть, например, $f(x) = \|x\|^{2+\alpha}$, $\alpha > 0$, тогда $\|\nabla f(x) - \nabla f(0)\| / \|x - 0\| = (2 + \alpha)\|x\|^{\alpha} \to \infty$ при $\|x\| \to \infty$. При этом для всякого $\gamma > 0$ можно указать такое x^0, что метод (5), примененный к функции $\|x\|^{2+\alpha}$, $\alpha > 0$, с начальным приближением x^0, расходится, поскольку будет $\|x^{k+1}\| > \|x^k\|$, $k = 0, 1, \ldots$

Если не выполнено условие (7), то функция *f(x)* не достигает минимума и градиент в методе (5) не обязан стремиться к 0 (например, если *f(x)* линейна: *f(x)* = (*c*, *x*), то $\|\nabla f(x)\| \equiv \|c\| > 0$).

Наконец, выбирать γ, нарушая условие (8), вообще говоря, также нельзя, что видно на примере функции $f(x) = Lx^2/2$, $x \in \mathbf{R}^1$.

Действительно, если $\gamma \geq 2/L$, то в методе (5) для этой функции будет

$$f(x^{k+1}) \geqslant f(x^k), \quad k = 0, 1, \ldots,$$

при любом x^0.

С другой стороны, при сделанных в теореме 1 предположениях нельзя доказать ничего большего, например, сходимость последовательности x^k. Примером может служить $f(x) = 1/(1 + \|x\|^2)$. Эта функция удовлетворяет условиям теоремы и при любом $x^0 \neq 0$ будет $\|x^k\| \to \infty$.

Если потребовать, чтобы множество $\{x: f(x) \leqslant f(x^0)\}$ было ограничено, то из x^k можно выбрать подпоследовательность, сходящуюся к некоторой стационарной точке x^*. Однако точка x^* не обязана быть точкой локального или глобального минимума. В частности, градиентный метод (5) (или даже (1) с произвольным выбором γ_k), начатый из некоторой стационарной точки x^0, останется в этой точке: $x^k = x^0$ для всех k. Иными словами, градиентный метод «застревает» в любой стационарной точке — точке максимума, минимума или седловой. Что же касается поиска глобального минимума, то градиентный метод «не отличает» точек локального минимума от глобального и никакой гарантии сходимости к глобальному минимуму он не дает.

Наконец, в условиях теоремы 1 скорость сходимости $\nabla f(x^k)$ к 0 может быть очень медленной. Например, для $f(x) = 1/x$ при $x \geq 1$ (вид $f(x)$ при $x < 1$ безразличен) метод (5) при $\gamma=1$, $x^0 = 1$ принимает вид $x^{k+1} = x^k + (x^k)^{-2}$, при этом можно показать, что

$$|f'(x^k)| = O(k^{-2/3}).$$

Рассмотрим поведение градиентного метода для более узкого класса функций — сильно выпуклых. Естественно, здесь удается доказать более сильные результаты, чем в теореме 1 - именно, сходимость итераций x^k к точке глобального минимума со скоростью геометрической прогрессии.

Нам понадобится несколько неравенств, относящихся к дифференцируемым, выпуклым и сильно выпуклым функциям.

Лемма 1. *Пусть f(x) дифференцируема, $\nabla f(x)$ удовлетворяет условию Липшица с константой Y и $f(x) \geq f^*$ для всех x.*
Тогда

$$\|\nabla f(x)\|^2 \leqslant 2L(f(x) - f^*). \qquad (10)$$

Доказательство. Сделаем из точки x шаг градиентного метода с $\gamma = 1/L$. Тогда (см. (9))

$$f^* \leqslant f(x - L^{-1}\nabla f(x)) \leqslant f(x) - (2L)^{-1}\|\nabla f(x)\|^2.$$

Лемма 2. *Пусть f(x) выпукла и дифференцируема, а* $\nabla f(x)$ *удовлетворяет условию Липшица с константой L. Тогда*

$$(\nabla f(x) - \nabla f(y),\ x - y) \geqslant L^{-1}\|\nabla f(x) - \nabla f(y)\|^2. \qquad (11)$$

Доказательство. Докажем (11) лишь для дважды дифференцируемых функций. Тогда

$$\nabla f(y) = \nabla f(x) + \int_0^1 \nabla^2 f(x + \tau(y - x))(y - x)\,d\tau = \nabla f(x) + A(y - x),$$

где матрица $A = \int_0^1 \nabla^2 f(x + \tau(y - x))\,d\tau$ симметрична и неотрицательно определена, т. е. $A \geq 0$. Кроме того, $\|A\| \leq L$, так как $\|\nabla^2 f(x)\| \leqslant L$, для всех x в силу условия Липшица на градиент. Поэтому

$$(\nabla f(x) - \nabla f(y),\ x - y) =$$

$$= (A(x - y),\ x - y) \geqslant \|A\|^{-1}\|A(x - y)\|^2 \geqslant L^{-1}\|\nabla f(x) - \nabla f(y)\|^2.$$

Лемма 3. *Пусть f(x)— дифференцируемая сильно выпуклая (с константой l) функция, x* — ее точка минимума (она существует). Тогда*

$$\|\nabla f(x)\|^2 \geqslant 2l(f(x) - f(x^*)).$$

Теорема 2. *Пусть f(x) дифференцируема на* $\mathbf{R}^n$, *ее градиент удовлетворяет условию Липшица с константой L и f(x) является сильно выпуклой функцией с константой l. Тогда при* $0 < \gamma < 2/L$ *метод* (5) *сходится к единственной точке глобального минимума x* со скоростью геометрической прогрессии:*

$$\|x^k - x^*\| \leqslant cq^k, \quad 0 \leqslant q < 1. \qquad (12)$$

Доказательство. Выполнены все условия теоремы 1, поэтому справедливо неравенство (9):

$$f(x^{k+1}) \leqslant f(x^k) - \gamma(1 - L\gamma/2)\|\nabla f(x^k)\|^2.$$

Используем лемму 3:

$$f(x^{k+1}) \leqslant f(x^k) - l\gamma(2 - L\gamma)(f(x^k) - f(x^*)).$$

Отсюда

$$f(x^{k+1}) - f(x^*) \leqslant (1 - l\gamma(2 - L\gamma))(f(x^k) - f(x^*)) = q_1(f(x^k) - f(x^*)),$$
$$f(x^k) - f(x^*) \leqslant q_1^k(f(x^0) - f(x^*)), \quad q_1 = 1 - 2l\gamma + Ll\gamma^2.$$

Поскольку $0 < \gamma < 2/L$, то $0 < q_1 < 1$, и следовательно, $f(x^k) \to f(x^*)$. Из неравенства

$$f(x) \geq f(x^*) + l\|x - x^*\|^2/2$$

следует

$$\|x^k - x^*\|^2 \leqslant (2/l)\, q_1^k (f(x^0) - f(x^*)).$$

Рассмотрим еще более узкий класс функций — сильно выпуклых дважды дифференцируемых.

Теорема 3. *Пусть f(x) дважды дифференцируема и*

$$lI \leqslant \nabla^2 f(x) \leqslant LI, \quad l > 0, \tag{13}$$

для всех x. Тогда при $0 < \gamma < 2/L$

$$\|x^k - x^*\| \leqslant \|x^0 - x^*\| q^k, \quad q = \max\{|1 - \gamma l|, |1 - \gamma L|\} < 1. \tag{14}$$

Величина q минимальна и равна

$$q^* = (L - l)/(L + l) \quad при \quad \gamma = \gamma^* = 2/(L + l). \tag{15}$$

Доказательство. По формуле

$$g(x + y) = g(x) + \int_0^1 g'(x + \tau y)\, y\, d\tau =$$

$$= g(x) + g'(x)\, y + \int_0^1 (g'(x + \tau y) - g'(x))\, y\, d\tau.$$

определяем

$$\nabla f(x^k) = \nabla f(x^*) + \int \nabla^2 f(x^* + \tau(x^k - x^*))(x^k - x^*)\, d\tau = A_k(x^k - x^*),$$

где в силу (13) $lI \leqslant A_k \leqslant LI$. Поэтому

$$\|x^{k+1} - x^*\| = \|x^k - x^* - \gamma \nabla f(x^k)\| =$$
$$= \|(I - \gamma A_k)(x^k - x^*)\| \leqslant \|I - \gamma A_k\| \|x^k - x^*\|.$$

Для всякой симметричной матрицы A имеем

$$\|I - A\| = \max\{|1 - \lambda_1|, |1 - \lambda_n|\},$$

где λ_1 и λ_n — наименьшее и наибольшее собственные значения A. Поэтому

$$\|x^{k+1} - x^*\| \leqslant q\|x^{k+1} - x^*\|, \; q = \max\{|1 - \gamma l|, |1 - \gamma L|\}.$$

Поскольку

$0 < \gamma < 2/L$, $0 < l \leqslant L$, то $|1 - \gamma l| < 1$, $|1 - \gamma L| < 1$, т. е. $q < 1$.

Минимизируя q по γ получаем (15).

Покажем, что оценка скорости сходимости, даваемая теоремой 3, точная, она достигается для любой квадратичной функцией. Пусть

$$f(x) = (Ax, x)/2 - (b, x), \quad A > 0, \quad 0 < l = \lambda_1 \leqslant \lambda_2 \ldots \leqslant \lambda_n = L,$$

где λ_i — соственные числа матрицы A. Возьмем произвольное $0 < \gamma < 2/L$. Предположим, что $|1 - \gamma l| \geqslant |1 - \gamma L|$. Выберем $x^0 = x^* + e^1$, где e^1 — собственный вектор. отвечающий λ_1, $\|e^1\| = 1$. Тогда $x^k - x^* = (I - \gamma A)^k (x^0 - x^*) =$

$$= (1 - \gamma\lambda_1)^k e^1, \quad \|x^k - x^*\| = |(1 - \gamma l)|^k = q^k \|x^0 - x^*\|.$$

Аналогичным образом, если $|1 - \gamma L| \geqslant |1 - \gamma l|$, то выберем $x^0 = x^* + e^n$, e^n — собственный вектор, отвечающий λ_n, $\|e^n\| = 1$, и получим так же

$$\|x^k - x^*\| = |(1 - \gamma L)|^k = q_k \|x^0 - x^*\|.$$

Таким образом, для всякого $0<\gamma\ 2/L$ найдется x^0 такое, что Оценку

$$\|x^n - x^*\| \leqslant (q^*)^k \|x^0 - x^*\|, \quad q^* = (L - l)/(L + l)$$

нельзя улучшить, даже если выбирать γ оптимальным образом для каждого x^0. Действительно, возьмем $x^0 = x^* + e^1 + e^n$ (обозначенияте же, что и выше). Тогда пои любом $0 < \gamma < 2/L$

$$x^k - x^* = (I - \gamma A)^k (x^0 - x^*) = (1 - \gamma l)^k e^1 + (1 - \gamma L)^k e^n,$$

$$\|x^k - x^*\| = [(1 - \gamma l)^{2k} + (1 - \gamma L)^{2k}]^{1/2} \|x^0 - x^*\| / \sqrt{2}.$$

Поэтому, если либо $|1 - \gamma l| > q^*$, либо $|1 - \gamma L| > q^*$, то $||x^k - x^*||$ убывает медленнее, чем $(q^*)^k$. Но $q = \max\{|1-\gamma l|, |1-\gamma L|\} \leq q^*$ лишь при $\gamma = \gamma^*$, при этом

$$|1 - \gamma^* l| = |1 - \gamma^* L| = q^* \text{ и } \|x^k - x^*\| = (q^*)^k \|x^0 - x^*\|.$$

Аналогичное расуждение справедливо для любой точки x^0 такой, что

$$(x^0 - x^*, e^1) \neq 0, \quad (x^0 - x^*, e^n) \neq 0.$$

Локальный аналог теоремы 3 справедлив и для невыпуклых функций.

Теорема 4. *Пусть x^* — невырожденная точка локального минимума $f(x)$. Тогда при $0 < \gamma < 2/\|\nabla^2 f(x^*)\|$ метод* (5) *локально сходится к x^* со скоростью геометрической прогрессии, т. е. для всякого $\delta > 0$ найдется $\varepsilon > 0$ такое, что при $\|x^0 - x^*\| \leqslant \varepsilon$ будет*

$$\|x^k - x^*\| \leqslant \|x^0 - x^*\|(q+\delta)^k,$$
$$q = \max\{|1-\gamma l|, |1-\gamma L|\} < 1, \quad 0 < lI \leqslant \nabla^2 f(x^*) \leqslant LI. \qquad (16)$$

Величина q минимальна и равна $q^* = (L-l)/(L+l)$ при $\gamma^* = 2/(L+l)$.

Метод Ньютона

1. Эвристические соображения. В градиентном методе основой является идея локальной линейной аппроксимации минимизируемой функции $f(x)$. Если же функция дважды дифференцируема, то естественно попытаться использовать ее квадратичную аппроксимацию в точке x^k, т. е. функцию

$$f_k(x) = f(x^k) + (\nabla f(x^k), x - x^k) + (\nabla^2 f(x^k)(x - x^k), x - x^k)/2. \qquad (17)$$

В градиентном методе следующее приближение x^{k+1} искалось из условия минимума линейной аппроксимации при дополнительных ограничениях на близость к x^k (так как линейная функция не достигает минимума на всем пространстве) — см. (2), (3) и (4). Для квадратичной аппроксимации можно попытаться не накладывать таких ограничений, так как при $\nabla^2 f(x^k) > 0$ функция $f_k(x)$ достигает безусловного минимума. Выберем точку минимума $f_k(x)$ в качестве нового приближения:

$$x^{k+1} = \operatorname*{argmin}_{x \in R^n} f_k(x).$$

Таким образом, мы получаем метод

$$x^{k+1} = x^k - [\nabla^2 f(x^k)]^{-1} \nabla f(x^k). \qquad (18)$$

К этому методу можно прийти и из несколько иных соображений. Точка минимума должна быть решением системы n уравнений с n переменными

$$\nabla f(x) = 0. \qquad (19)$$

Одним из основных методов решения таких систем является *метод Ньютона,* заключающийся в *линеаризации* уравнений в точке x^k и

решении линеаризованной системы (см. ниже п. 3). Эта линеаризованная система в данном случае имеет вид

$$\nabla f(x^k) + \nabla^2 f(x^k)(x - x^k) = 0 \tag{20}$$

и ее решение x^{k+1} дается формулой (18).

Сходимость.

Теорема 1. *Пусть f(x) дважды дифференцируема,* $\nabla^2 f(x)$ *удовлетворяет условию Липшица с константой L, f(x) сильно выпукла с константой l и начальное приближение удовлетворяет условию*

$$q = (Ll^{-2}/2)\|\nabla f(x^0)\| < 1. \tag{21}$$

Тогда метод (18) *сходится к точке глобального минимума x* с квадратичной скоростью:*

$$\|x^k - x^*\| \leqslant (2l/L)\, q^{2^k}. \tag{22}$$

Доказательство. Из условий Липшица на $\nabla^2 f(x^k)$ следует

$$\|\nabla f(x+y) - \nabla f(x) - \nabla^2 f(x)\, y\| \leqslant (L/2)\|y\|^2.$$

Возьмем здесь $x = x^k,\ y = -[\nabla^2 f(x^k)]^{-1}\nabla f(x^k)$, тогда $x+y = x^{k+1}$ и

$$\|\nabla f(x^{k+1})\| \leqslant (L/2)\left\|[\nabla^2 f(x^k)]^{-1}\nabla f(x^k)\right\|^2 \leqslant (L/2)\left\|[\nabla^2 f(x^k)]^{-1}\right\|^2 \|\nabla f(x^k)\|^2.$$

По поскольку $\nabla^2 f(x^k) \geqslant lI$ (условие сильной выпуклости), то

$$[\nabla^2 f(x^k)]^{-1} \leqslant l^{-1}I \quad \text{и} \quad \|[\nabla^2 f(x^k)]^{-1}\| \leqslant l^{-1},$$

т. е.

$\|\nabla f(x^{k+1})\| \leqslant (Ll^{-2}/2)\|\nabla f(x^k)\|^2$. Итерируя это неравенство, получаем

$$\|\nabla f(x^k)\| \leqslant \frac{2l^2}{L}\left(\frac{L}{2l^2}\|\nabla f(x^0)\|\right)^{2^k} = \frac{2l^2}{L}\, q^{2^k}.$$

Покажем, что все условия теоремы существенны, а усилить ее утверждение, вообще говоря, нельзя. Ясно, что существование второй производной требуется в самой формулировке метода, а условие сильной выпуклости гарантирует существование $[\nabla^2 f(x^k)]^{-1}$. Меньшие требования к гладкости (отказ от условия Липшица на $\nabla^2 f(x)$) могут привести к уменьшению скорости сходимости метода. Пусть, например, $f(x)=|x|^{5/2}$, $x \in R^1$. Тогда при $x > 0$ $f'(x) = (5/2)x^{3/2}$, $f''(x) = (15/4)x^{1/2}$ и $f''(x)$ не удовлетворяет условию Липшица. Метод принимает вид (при $x^0>0$)

$x^{k+1} = x^k - (4/15)(x^k)^{-1/2}\cdot(5/2)(x^k)^{3/2} = (1/3)x^k$, т. е. $x^k = (1/3)^k x^0$ и метод сходится к $x^* = 0$ со скоростью геометрической прогрессии (а не с квадратичной скоростью), Наконец, нельзя утверждать сходимость метода при любом начальном приближении (не

удовлетворяющем условию (21)). Пусть задача заключается в минимизации одномерной функции, производная которой изображена на рис. 1.

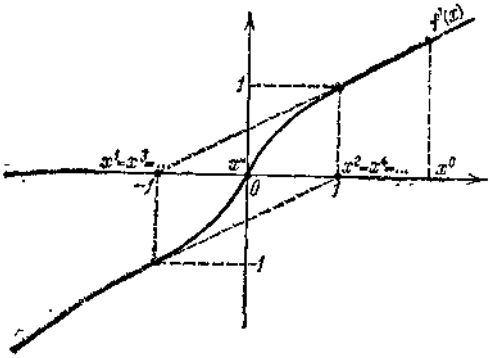


Рис. 1. Расходимость метода Ньютона.

Эта функция дважды дифференцируема, сильно выпукла (так как $f'(x) \geq 1/2 > 0$ для всех x), $f''(x)$ удовлетворяет условию Липшица, $x^* = 0$. Однако если начать итерационный процесс из любой точки x^0 с $|x^0|>1$, то метод не сходится: $|x^k| \equiv 1$ для всех $k \geq 1$.

Условия теоремы 1 можно несколько ослабить лишь в одном направлении — можно глобальные требования на $f(x)$ заменить на локальные.

Теорема 2. *Пусть $f(x)$ дважды дифференцируема в окрестности U точки невырожденного минимума x^*, и $\nabla^2 f(x)$ удовлетворяет условию Липшица на U. Тогда найдется $\varepsilon>0$ такое, что при $\|x^0 - x^*\| \leqslant \varepsilon$ метод* (18) *сходится к x^* с квадратичной скоростью.* А

Для квадратичной функции $f(x) = (Ax, x)/2 - (b, x)$ с $A > 0$ метод Ньютона сходится за 1 шаг, т. е. $x^1 = x^*$ при любом x^0. Это очевидно, так как аппроксимирующая функция $f_0(x)$ совпадает с $f(x)$. Чем ближе $f(x)$ к квадратичной, тем быстрее сходится метод Ньютона. Формально — чем меньше L, тем в соответствии с теоремой больше область сходимости, определяемая (21), и тем быстрее скорость сходимости, определяемая величиной q.

Метод Ньютона для уравнений. Метод Ньютона может применяться не только для задач минимизации, но и для решения произвольных нелинейных уравнений

$$g(x) = 0, \quad g: \mathbf{R}^n \to \mathbf{R}^n. \tag{23}$$

Он основан на той же идее линейной аппроксимации — на k-й итерации решается линеаризованное уравнение

$$g(x^k) + g'(x^k)(x - x^k) = 0,$$

откуда

$$x^{k+1} = x^k - g'(x^k)^{-1} g(x^k). \qquad (24)$$

Теорема 3. *Пусть уравнение* (23) *имеет решение* x^*, *функция* g: $\mathbf{R}^n \to \mathbf{R}^n$ *дифференцируема в окрестности* x^* *и* $g'(x)$ *удовлетворяет условию Липшица в этой окрестности. Пусть матрица* $g'(x^*)$ *невыождена. Тогда найдется* $\varepsilon > 0$ *такое, что при* $\|x^0 - x^*\| \leqslant \varepsilon$ *метод* (24) *сходится к* x^* *с квадратичной скоростью.*

Очевидно, что теорема 2 есть частный случай теоремы 3 при $g(x) = \nabla f(x)$; доказательство остается прежним.

Подчеркнем, что для сходимости (24) не нужно ни симметричности, ни положительной определенности $g'(x)$. В частности, метод Ньютона годится для отыскания стационарных точек функции $f(x)$, отличных от точек минимума.

15.3. Общие схемы исследования скорости сходимости

Результаты о сходимости и скорости сходимости алгоритмов минимизации в п. 15.2 были получены непосредственно, без привлечения каких-либо общих теорем. Такой подход был естествен, поскольку доказательства очень просты. Однако по мере усложнения задач и методов их обоснование становится более громоздким и трудоемким. Внимательный анализ применяемых доказательств показывает, что лежащие в их основе идеи просты и единообразны. Разумно выделить эти идеи «в явном виде», получить с их помощью ряд общих результатов о сходимости, а затем систематически использовать их при обосновании конкретных алгоритмов. Такого рода общие результаты и приводятся в данном разделе.

15.3.1. Первый метод Ляпунова

Идея этого подхода заключается в линеаризации итеративной процедуры, после чего вывод о сходимости удается сделать на основе анализа линеаризованного процесса. Предварительно приведем необходимые сведения из линейной алгебры.

1. Сведения из линейной алгебры. Пусть A — квадратная матрица $n \times n$. $\lambda_1, \ldots, \lambda_n$ — ее собственные значения. *Спектральным радиусом A* называется число

$$\rho(A) = \max_{1 \leqslant i \leqslant n} |\lambda_i|. \tag{1}$$

Другой важной характеристикой матрицы (не обязательно квадратной) является ее *норма*

$$\|A\| = \max_{\|x\|=1} \|Ax\|. \tag{2}$$

Используя тот факт, что у симметричной матрицы все собственные значения вещественны и существует полная ортогональная система собственных векторов, нетрудно доказать, что для симметричной матрицы $\rho(A) = \|A\|$. Для несимметричной матрицы $\rho(A) \leqslant \|A\|$..... и, вообще говоря, $\rho(A) \neq \|A\|$. Например, для матрицы $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ оба собственных значения равны 0, по-этому $\rho(A)=0$, однако $\|A\|=1$. Важная связь между $\|A\|$ и $\rho(A)$ устанавливается равенством

$$\rho(A) = \lim_{k \to \infty} \|A^k\|^{1/k}. \tag{3}$$

Из (3) вытекает следующий фундаментальный факт.

Лемма 1. *Чтобы*

$$\lim_{k \to \infty} A^k = 0,$$

необходимо и достаточно выполнение условия $\rho(A) < 1$, *при этом для всякого* $\varepsilon > 0$ *найдется* $c = c(\varepsilon)$ *такое, что* $\|A^k\| \leqslant c(\rho(A) + \varepsilon)^k$ *для всех натуральных k.*

Следствие. *Для того чтобы итерационная последовательность векторов* $x^{k+1} = Ax^k$ *сходилась к* 0 *при* $k \to \infty$ *при любом* x^0, *необходимо и достаточно выполнение условия* $\rho(A) < 1$.

Лемма 2. *Пусть* $\rho(A) < 1$. *Тогда матричное уравнение*

$$A^T U A = U - C \tag{4}$$

имеет решение U, которое симметрично, если матрица C симметрична, и $U \geq C$ *при* $C \geq 0$.

Доказательство. Поскольку $\|A^k\| \leqslant cq^k, \quad q < 1$ (лемма 1), то ряд

$$\sum_{k=0}^{\infty} (A^T)^k C A^k$$

сходится к некоторой матрице *U*.

Эта матрица симметрична при симметричной C, $U \geq 0$ при

$$C \geqslant 0, \quad A^T U A = \sum_{k=1}^{\infty} (A^T)^k C A^k = U - C, \quad U = C + A^T U A \geqslant C$$

при $C \geq 0$.

Назовем квадратную матрицу F с собственными значениями $\lambda_1, ..., \lambda_n$ *устойчивой (гурвицевой),* если

$$\operatorname{Re} \lambda_i < 0, \qquad i = 1, \ldots, n. \tag{5}$$

Лемма 3. *Для того чтобы*

$$\lim_{t \to \infty} e^{At} = 0,$$

необходимо и достаточно, чтобы A была устойчива. При этом для всякого $\varepsilon > 0$ *найдется* $c = c(\varepsilon)$ *такое, что* $\|e^{At}\| \leqslant c(\varepsilon) e^{(\gamma+\varepsilon)t}$ *для всех* $t \geq 0$,

$$\gamma = \max_i \operatorname{Re} \lambda_i.$$

Действительно, собственными значениями $B = e^A$ являются e^{λ_i}, поэтому

$$\rho(B) = \max_i e^{\operatorname{Re} \lambda_i} = e^{\gamma}.$$

Поскольку $e^{\gamma} < 1$ тогда и только тогда, когда $\gamma < 0$, то условие $\rho(B) < 1$ эквивалентно условию $\gamma < 0$. Теперь остается воспользоваться леммой 1.

Лемма 4 (Ляпунов). *Пусть матрица A устойчива, а матрица C симметрична. Тогда уравнение*

$$AU + UA^T = -C \tag{6}$$

имеет решение, причем $U > 0$ ($U \geq 0$), *если* $C > 0$ ($C \geq 0$).

Доказательство. В соответствии с леммой 3 матрица $U = \int_0^{\infty} e^{At} C e^{A^T t} \, dt$ определена. Матрица $Z(t) = e^{At} C e^{A^T t}$ является решением дифференциального уравнения $\dot{Z}(t) = AZ + ZA^T$, $Z(0) = C$, т. е. $U = \int_0^{\infty} Z(t) \, dt$, поэтому

$$AU + UA^T = \int_0^{\infty} (AZ + ZA^T) \, dt = \int_0^{\infty} \dot{Z}(t) \, dt = -Z(0) = -C.$$

Отсюда

$$U = \int_0^{\infty} e^{At} C e^{A^T t} \, dt$$

является искомым решением, и из этой же формулы следует, что $U>0$ ($U\geq 0$), при $C>0$ ($C\geq 0$).

Связь между устойчивыми матрицами и матрицами с $\rho(A)<1$ устанавливается следующей леммой.

Лемма 5. *Пусть матрица A устойчива,* $B = I + \gamma A, 0 < \gamma < \min\left(-2\operatorname{Re}\lambda_i / |\lambda_i|^{-2}\right)$. *Тогда* $\rho(B)<1$.

Действительно, если λ_i — собственные значения A, μ_i — собственные значения B, то $\mu_i = 1 + \gamma\lambda_i$, $|\mu_i|^2 = (1 + \gamma \operatorname{Re}\lambda_i)^2 + \gamma^2(\operatorname{Im}\lambda_i)^2 = 1 + 2\gamma \operatorname{Re}\lambda_i + \gamma^2|\lambda_i|^2 < 1$, т. е. $\rho(B) < 1$.

2. Теоремы о линейной сходимости. Мы будем часто употреблять термин *линейная сходимость* как синоним сходимости со скоростью геометрической прогрессии. Аналогично *сверхлинейная сходимость* означает сходимость более быструю, чем определяемую любой геометрической прогрессией. Наконец, термин *квадратичная сходимость* используется для процессов, в которых справедлива оценка вида $u_{k+1} \leqslant cu_k^2$, где u_k — некоторая мера близости к решению на k-й итерации. Рассмотрим итерационный процесс вида

$$x^{k+1} = g(x^k), \tag{7}$$

где g — некоторое отображение из $\mathbf{R}^n$ в $\mathbf{R}^n$. Точку x^* будем называть *неподвижной точкой* (7), если $x^* = g(x^*)$. В этом случае при $x^k = x^*$ будет $x^s \equiv x^*$ для всех $s \geq k$.

Теорема 1. *Пусть* x^* *— неподвижная точка* (7), $g(x)$ *дифференцируема в* x^* *и спектральный радиус матрицы Якоби* $g'(x^*)$ *удовлетворяет условию* $\rho = \rho(g'(x^*)) < 1$. *Тогда процесс* (7) *локально линейно сходится к* x^*, *а именно, для всякого* $0<\varepsilon<1$ — ρ *найдутся* $\delta > 0$ *и* c *такие, что для всех* $k \geq 0$ *будет*

$$\|x^k - x^*\| \leqslant c(\rho + \varepsilon)^k \tag{8}$$

При $\|x^0 - x^*\| \leqslant \delta$.

Дадим краткую схему доказательства. Обозначим $A = g'(x^*)$, тогда в соответствии с определением производной $g(x) = g(x^*) + A(x - x^*) + o(x - x^*)$. Поэтому процесс (7) может быть записан в виде

$$z^{k+1} = Az^k + y^k, \quad z^k = x^k - x^*, \quad y^k = o(z^k).$$

Отсюда

$$z^{k+1} = A^{k+1}z^0 + \sum_{i=0}^{k} A^{k-i}y^i,$$

$$\|z^{k+1}\| \leqslant \|A^{k+1}\|\|z^0\| + \sum_{i=0}^{k} \|A^{k-i}\|\|y^i\|. \qquad (9)$$

Из леммы 1 $\|A^k\| \leqslant c(\varepsilon)(\rho+\varepsilon)^k$, подставляя эту оценку в (9) и используя тот факт, что $\|y^k\| = o(z^k)$, можно получить утверждение теоремы.

Теорема 1 гарантирует локальную сходимость метода (7). В некоторых случаях можно утверждать и глобальную сходимость. Один из таких случаев очевиден — это случай линейной функции *g(x)*. Приведем результат о глобальной сходимости и для нелинейных функций. При этом нам удобнее будет рассматривать итерационный процесс, заданный в виде

$$x^{k+1} = x^k - \gamma(Ax^k + \varphi(x^k)). \qquad (10)$$

Теорема 2. *Пусть матрица — A устойчива, а* $\varphi\colon \mathbf{R}^n \to \mathbf{R}^n$ *удовлетворяет условию*

$$\|\varphi(x)\| \leqslant L\|x\|.$$

Тогда, если

$$L < \frac{1}{2\|U\|}, \quad 0 < \gamma < \frac{\|U\|^{-1} - 2L}{(L+\|A\|)^2}, \qquad (11)$$

где U — решение матричного уравнения

$$UA + A^TU = I, \qquad (12)$$

то процесс (10) *сходится к* 0 *со скоростью геометрической прогрессии при любом* x^0*:*

$$\|x^k\|^2 \leqslant \|x^0\|^2\|U^{-1}\|\|U\|q^k,$$
$$q = 1 - (1/2)\gamma\|U\|^{-1} + \gamma L + (1/2)\gamma^2(\|A\| + L)^2. \qquad (13)$$

Для доказательства достаточно ввести

$$u_k = (Ux_k,\ x_k)$$

и получить соотношение $u_{k+1} \leqslant qu_k$.

Полученные выше результаты можно применить для исследования *уравнения в конечных разностях*

$$y_k = a_1y_{k-1} + a_2y_{k-2} + \ \dots\ + a_ny_{k-n} + \varphi(y_{k-1}, \dots, y_{k-n}), \quad (14)$$

где $y_i \in \mathbf{R}^1$. Для этого введем вектор

$$x^k = (y_{k-1}, \dots, y_{k-n}) \in \mathbf{R}^n, \quad x^{k+1} = (y_k,\ y_{k-1}, \dots, y_{k-n+1}) \in \mathbf{R}^n,$$

тогда

$$x^{k+1} = Ax^k + h(x^k),$$

где

$$A = \begin{pmatrix} a_1 & a_2 & \dots & a_n \\ 1 & 0 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & 1\ 0 \end{pmatrix}, \quad h(x) = \begin{pmatrix} \varphi(x) \\ 0 \\ \cdot \\ \cdot \\ 0 \end{pmatrix}. \tag{15}$$

Таким образом, итерационный процесс приведен к форме (7). Описанный прием типичен при исследовании многошаговых итеративных процессов, в которых каждое приближение зависит от нескольких предыдущих. Тогда увеличение размерности задачи позволяет свести ее к одношаговому процессу.

3. Теорема о сверхлинейной сходимости. В случае, когда $g'(x^*)=0$, из теоремы 1 следует, что метод (7) сходится быстрее любой геометрической прогрессии. Этот результат можно уточнить.

Теорема 3. *Пусть x^* — неподвижная точка* (7), *$g(x)$ дифференцируема в $S = \{x: \|x - x^*\| \leqslant \|x^0 - x^*\|\}$, $g'(x)$ удовлетворяет в S условию Липшица и $g'(x^*) = 0$. Тогда, если*

$$q = (L/2)\|x^0 - x^*\| < 1, \tag{16}$$

то

$$\|x^k - x^*\| \leqslant (2/L)\, q^{2^k}. \tag{17}$$

Доказательство. Очевидно, $x^0 \in S$. В силу формулы

$$\|g(x+y)-g(x)-g'(x)y\| \leq L\|y\|^2/2$$

запишем

$$\|x^1 - x^*\| = \|g(x^0) - g(x^*) - g'(x^*)(x^0 - x^*)\| \leqslant (L/2)\|x^0 - x^*\|^2 \leqslant q\,\|x^0 - x^*\|,$$

поэтому $x^1 \in S$. Аналогичным образом $x^k \in S$ для всех k. Поэтому мы имеем право пользоваться той же оценкой:

$$\|x^{k+1} - x^*\| = \|g(x^k) - g(x^*) - g'(x^*)(x^k - x^*)\| \leqslant (L/2)\|x^k - x^*\|^2.$$

Отсюда следует требуемый результат.

15.3.2. Второй метод Ляпунова

Этот метод является наиболее распространенным при обосновании сходимости итерационных процессов. Его идея заключается в том, что вводится некоторая скалярная неотрицательная функция $V(x)$ *(функция Ляпунова)* и рассматриваются ее значения на последовательных итерациях x^k. Если они монотонно убывают и ограничены снизу, то $V(x^k) - V(x^{k+1}) \to 0$. Отсюда при некоторых дополнительных предположениях следует сходимость метода.

Если посмотреть с этой точки зрения на приведенные выше результаты о сходимости, то окажется, что большинство из них получено именно по такой схеме. Так, при обосновании градиентного метода в качестве функции Ляпунова в теоремах 1, 2 выступала сама минимизируемая функция $f(x) — f^*$. И теоремах 3, 4 такую роль играло расстояние до точки минимума. При обосновании метода Ньютона (теорема 1) использовалось монотонное убывание нормы градиента (т. е. невязки в выполнении условия экстремума). Наконец, в теореме 2 п. 15.3.1 в доказательстве была построена специальная квадратичная функция Ляпунова. Эти же приемы выбора функции Ляпунова обычно применяются и для других, более сложных задач.

1. Леммы о числовых последовательностях. Для значений функции Ляпунова $u_k=V(x^k)$ на k-м шаге процесса обычно получается итерационное соотношение вида

$$u_{k+1} \leqslant \varphi_k(u_k). \tag{1}$$

Отсюда делается вывод, что $u_k \to 0$ и дается оценка скорости сходимости u_k. Поэтому важно исследовать поведение последовательностей вида (1) для нескольких «типовых» функций φ_k. С некоторыми простейшими соотношениями (1) мы уже сталкивались. Так, при доказательстве сходимости градиентного метода мы получали неравенство

$$u_{k+1} \leqslant q u_k, \quad 0 \leqslant q < 1 \tag{2}$$

(где $u_k = f(x^k) - f^*$, либо $u_k = \|x^k - x^*\|^2$, либо $u_k = \|\nabla f(x^k)\|$). Из (2) следует оценка $u_k \leqslant u_0 q^k$. При обосновании метода Ньютона было получено соотношение для

$$u_k = \|\nabla f(x^k)\|:$$

$$u_{k+1} \leqslant c u_k^2, \quad c > 0. \tag{3}$$

Отсюда $u_k \leqslant c^{-1}(cu_0)^{2^k}$ и, если $cu_0 < 1$, то $u_k \to 0$.

В других задачах, однако, соотношение (1) имеет более сложный вид, и его анализ не столь тривиален.

Начнем с линейных неравенств вида

$$u_{k+1} \leqslant q_k u_k + \alpha_k, \quad q_k \geqslant 0. \tag{4}$$

Отсюда

$$u_k \leqslant q_{k-1}q_{k-2}\dots q_0 u_0 + q_{k-1}\dots q_1\alpha_0 + \dots + q_{k-1}\alpha_{k-2} + \alpha_{k-1}. \tag{5}$$

Рассмотрим несколько частных случаев.

Лемма 1. *Пусть*

$$u_{k+1} \leqslant q u_k + \alpha, \quad 0 \leqslant q < 1, \quad \alpha > 0. \tag{6}$$

Тогда

$$u_k \leqslant \alpha/(1-q) + (u_0 - \alpha/(1-q))\, q^k. \tag{7}$$

Доказательство. Обозначая $v_k = u_k - \alpha/(1-q)$, из неравенства (6) получаем $v_{k+1} \leqslant v_k q$, что и дает (7).

Таким образом, u_k сходится в область $u \leqslant \alpha/(1-q)$ со скоростью геометрической прогрессии со знаменателем q.

Л е м м а 2. Пусть $u_k \geq 0$ *и*

$$u_{k+1} \leqslant (1+\alpha_k) u_k + \beta_k, \quad \alpha_k \geqslant 0, \quad \beta_k \geqslant 0,$$
$$\sum_{k=0}^{\infty} \alpha_k < \infty, \quad \sum_{k=0}^{\infty} \beta_k < \infty. \tag{8}$$

Тогда $u_k \to u \geqslant 0$.

Доказательство совпадает с приводимым ниже доказательством более общей леммы 9.

Лемма 3. *Пусть*

$$u_{k+1} \leqslant q_k u_k + \alpha_k, \quad 0 \leqslant q_k < 1, \quad \alpha_k \geqslant 0,$$
$$\sum_{k=0}^{\infty} (1-q_k) = \infty, \quad \alpha_k/(1-q_k) \to 0. \tag{9}$$

Тогда $\overline{\lim\limits_{k\to\infty}}\, u_k \leqslant 0$. *В частности, если* $u_k \geqslant 0$, *то* $u_k \to 0$.

Следствие. *Если в* (9)

$$q_k \equiv q < 1, \quad \alpha_k \to 0, \quad u_k \geqslant 0,$$

то $u_k \to 0$.

В условиях леммы 3 можно для ряда случаев оценить и скорость сходимости.

Лемма 4 (Чжун). *Пусть* $u_k \geqslant 0$ *и*

$$u_{k+1} \leqslant \left(1 - \frac{c}{k}\right) u_k + \frac{d}{k^{p+1}}, \quad d > 0, \quad p > 0, \quad c > 0. \tag{10}$$

Тогда

$$u_k \leqslant d\,(c-p)^{-1} k^{-p} + o\left(k^{-p}\right) \quad \textit{при} \quad c > p, \tag{11}$$
$$u_k = O\left(k^{-c} \ln k\right) \quad \textit{при} \quad p = c, \tag{12}$$
$$u_k = O\left(k^{-c}\right) \quad \textit{при} \quad p > c. \tag{13}$$

Доказательство. При любом соотношении c и p мы находимся в условиях применения леммы 3, так как $1 - q_k = c/k$, $\sum\limits_{k=0} (1-q_k) = \infty$, $\alpha_k (1-q_k)^{-1} = dc^{-1}k^{-p} \to 0$, поэтому $u_k \to 0$. Пусть $c > p$. Введем $v_k = k^p u_k - d\,(c-p)^{-1}$. Тогда

$$v_{k+1} = (k+1)^p u_{k+1} - \frac{d}{c-p} \leqslant k^p \left(1+\frac{1}{k}\right)^p \left(\left(1-\frac{c}{k}\right)u_k + \frac{d}{k^{p+1}}\right) -$$
$$- \frac{d}{c-p} = k^p u_k \left(1 - \frac{c-p}{k} + o\left(\frac{1}{k}\right)\right) +$$
$$+ \frac{d}{k}\left(1 + \frac{p}{k} + o\left(\frac{1}{k}\right)\right) - \frac{d}{c-p} =$$
$$= \left(v_k + \frac{d}{c-p}\right)\left(1 - \frac{c-p}{k} + o\left(\frac{1}{k}\right)\right) +$$
$$+ \frac{d}{k}\left(1 + \frac{p}{k} + o\left(\frac{1}{k}\right)\right) - \frac{d}{c-p} =$$
$$= v_k \left(1 - \frac{c-p}{k} + o\left(\frac{1}{k}\right)\right) + \frac{dp}{k^2} + o\left(\frac{1}{k^2}\right).$$

Применяя лемму 3, получаем $\overline{\lim\limits_{k\to\infty}}\, v_k \leqslant 0$, что и доказывает (11).

Пусть теперь $p \geq c$. Введем $v_k = u_k k^c$. Тогда

$$v_{k+1} = u_{k+1}(k+1)^c \leqslant$$
$$\leqslant \left[\left(1-\frac{c}{k}\right)u_k + \frac{d}{k^{p+1}}\right] \cdot k^c \left(1 + \frac{c}{k} + \frac{c^2}{2k^2} + o\left(\frac{1}{k^2}\right)\right) =$$
$$= \left(1 - \frac{c^2}{2k^2} + o\left(\frac{1}{k^2}\right)\right) v_k + \frac{d}{k^{p-c+1}}\left(1 + O\left(\frac{1}{k}\right)\right) \leqslant v_k + \frac{d'}{k^{p-c+1}}$$

для достаточно больших k. Суммируя по k, получаем, что v_k ограничено при $p > c$ (так как ряд $\sum\limits_{k=1}^{\infty} \frac{1}{k^{\alpha}}$ сходится при $\alpha > 1$)

и $v_k = O(\ln k)$ при $p=c$ (так как

$$\sum_{l=1}^{k} \frac{1}{l} = O(\ln k)\Big).$$

Это доказывает (12) и (13).

Лемма 5 (Чжун). *Пусть* $u_k \geqslant 0$ *и*

$$u_{k+1} \leqslant \left(1 - \frac{c}{k^s}\right)u_k + \frac{d}{k^t}, \quad 0 < s < 1, \quad s < t. \qquad (14)$$

Тогда

$$u_k \leqslant \frac{d}{c}\,\frac{1}{k^{t-s}} + o\left(\frac{1}{k^{t-s}}\right).$$

Перейдем к исследованию рекуррентных неравенств, задаваемых нелинейными соотношениями.

Лемма 6. *Пусть* $u_k > 0$ *и*

$$u_{k+1} \leqslant u_k - \alpha_k u_k^{1+p}, \quad \alpha_k \geqslant 0, \quad p > 0. \qquad (15)$$

Тогда

$$u_k \leqslant u_0 \left(1 + p u_0^p \sum_{i=0}^{k-1} \alpha_i\right)^{-1/p}. \qquad (16)$$

В частности, если $\alpha_k \equiv \alpha$, $p = 1$, *то*

$$u_k \leqslant u_0/(1 + \alpha k u_0). \qquad (17)$$

Докажем (16) лишь для случая $p \geq 1$. Разделим обе части (15) на $u_k^p u_{k+1}$:

$$u_k^{-p} \leqslant u_k^{1-p} u_{k+1}^{-1} - \alpha_k u_k u_{k+1}^{-1}.$$

Поскольку $u_{k+1} \leqslant u_k$, $u_k^{1-p} \leqslant u_{k+1}^{1-p}$ при $p>1$, то получаем $u_k^{-p} \leqslant u_{k+1}^{-p} - \alpha_k$. Суммируя неравенства, приходим к (16).

2. Леммы о случайных последовательностях. При исследовании итеративных методов, включающих элементы случайности (методы случайного поиска, задачи с помехами), обычно применяется та же техника, основанная на функциях Ляпунова. Однако здесь значения функции Ляпунова оказываются случайной величиной, поэтому нужно получить аналоги приведенных выше лемм для случайных последовательностей.

Напомним различные виды сходимости случайных величин. Пусть $v^1, \ldots, v^k, \ldots$ — последовательность n-мерных случайных векторов. Мы обычно не будем выписывать то вероятностное пространство $(\Omega, \mathfrak{F}, \mathsf{P})$, на котором заданы эти величины (т. е. не будем писать $v^1(\omega), \ldots, v^k(\omega), \ldots$, $\omega \in \Omega$, Ω — пространство элементарных событий, $\mathfrak{F}$—заданная на нем σ-алгебра измеримых множеств, P — вероятностная мера на $\mathfrak{F}$). Говорят, что последовательность v^k *сходится* к случайному вектору v:

а) *почти наверное* (с вероятностью 1), если

$$\mathsf{P}\left(\lim_{k\to\infty} v^k = v\right) = 1$$

(здесь и далее $\mathsf{P}(A)$ обозначает вероятность события A), при этом пишут $v^k \to v$ (п. н.).

б) *по вероятности,* если для каждого $\varepsilon > 0$ $\lim\limits_{k\to\infty} \mathsf{P}(\|v^k - v\| > \varepsilon) = 0$, что обозначается $v^k \xrightarrow{\mathsf{P}} v$.

в) *в среднем квадратичном,* если

$$\lim_{k\to\infty} \mathsf{M}\|v^k - v\|^2 = 0$$

(здесь и далее $\mathsf{M}\alpha$ обозначает математическое ожидание случайной величины α).

Основным инструментом при изучении сходимости случайных величин является теория полумартингалов. Последовательность скалярных случайных величин $v_0, \ldots, v_k \ldots$ называется *полумартингалом,* если $\mathbf{M}(v_{k+1} | v_1, \ldots, v_k) \leqslant v_k, \quad \mathbf{M}v_0 < \infty.$ Здесь $\mathbf{M}(v_{k+1} | v_0, \ldots, v_k)$ — условное математическое ожидание v_{k+1} при данных $v_0 \ldots v_k$. Часто в данном случае употребляют также термин *супермартингал,* для неравенства противоположного знака говорят о *субмартингале,* а для равенства — о *мартингале.* Полумартингал является обобщением на стохастический случай понятия монотонно убывающей последовательности. Ключевой результат о сходимости числовых последовательностей (ограниченная снизу монотонно убывающая последовательность имеет предел) для случайных величин приобретает следующий вид.

Лемма 7. *Пусть* $v_0 \ldots v_k \ldots$ *— полу мартингал, причем* $v_k \geq 0$ *для всех k. Тогда существует случайная величина* $v \geq 0$, $v_k \to v$ *(п. н.).*

Известное *неравенство Чебышева* (если $v \geqslant 0, \ \varepsilon > 0, \ \mathbf{M}v < \infty$, то $\mathbf{P}(v \geqslant \varepsilon) \leqslant \varepsilon^{-1}\mathbf{M}v$) для полумартингалов может быть усилено.

Лемма 8 (неравенство Колмогорова). *Пусть* $v_0 \ldots v_k \ldots$ *— полумартингал,* $v_k \geqslant 0, \ \varepsilon > 0$. *Тогда*

$$\mathbf{P}(v_k \geqslant \varepsilon \ \forall k) \leqslant \varepsilon^{-1}\mathbf{M}v_0. \text{ А} \tag{18}$$

Используя эти результаты, получим стохастические аналоги лемм 2 и 3.

Лемма 9 (Гладышев). *Пусть имеется последовательность случайных величин* $v_0, \ldots, v_k \geqslant 0, \ \mathbf{M}v_0 < \infty$ *и*

$$\mathbf{M}(v_{k+1} | v_0, \ldots, v_k) \leqslant (1 + \alpha_k) v_k + \beta_k,$$
$$\sum_{k=0}^{\infty} \alpha_k < \infty, \quad \sum_{k=0}^{\infty} \beta_k < \infty, \quad \alpha_k \geqslant 0, \quad \beta_k \geqslant 0. \tag{19}$$

Тогда $v_k \to v$ *(п. н),* *где* $v \geq 0$ *— некоторая случайная величина.*

Доказательство. Введем $u_k = \prod_{i=k}^{\infty} (1 + \alpha_i) v_k + \sum_{i=k}^{\infty} \beta_i \times$

$$\times \prod_{l=i+1}^{\infty} (1 + \alpha_l).$$

Тогда

$$u_k \geqslant 0, \ \mathbf{M}u_0 < \infty \left(\text{так как} \prod_{i=0}^{\infty} (1 + \alpha_l) < \infty, \ \sum_{i=0}^{\infty} \beta_i < \infty, \ \mathbf{M}v_0 < \infty \right).$$

При этом

$$\mathbf{M}(u_{k+1}|u_0, \ldots, u_k) =$$

$$= \prod_{i=k+1}^{\infty}(1+\alpha_i)\,\mathbf{M}(v_{k+1}|v_0, \ldots, v_k) + \sum_{i=k+1}^{\infty}\beta_i \prod_{j=i+1}^{\infty}(1+\alpha_j) \leqslant$$

$$\leqslant \prod_{i=k}^{\infty}(1+\alpha_i)\,v_k + \sum_{i=k}^{\infty}\beta_i \prod_{j=i+1}^{\infty}(1+\alpha_j) = u_k,$$

т. е. u_k — полумартингал, и по лемме 7 $u_k \to v$ (п. н.), $v \geqslant 0$.

Поэтому и $v_k = \left(u_k - \sum_{i=k}^{\infty}\beta_i\right) \Big/ \prod_{i=k}^{\infty}(1+\alpha_i) \to v$ (п. н.).

Лемма 10. *Пусть* $v_0, \ldots, v_k$ — *последовательность случайных величин,* $v_k \geqslant 0$, $\mathbf{M}v_0 < \infty$ *и*

$$\mathbf{M}(v_{k+1}|v_0, \ldots, v_k) \leqslant (1-\alpha_k)\,v_k + \beta_k, \tag{20}$$

$$0 \leqslant \alpha_k \leqslant 1, \quad \beta_k \geqslant 0, \quad \sum_{k=0}^{\infty}\alpha_k = \infty, \quad \sum_{k=0}^{\infty}\beta_k < \infty, \quad \frac{\beta_k}{\alpha_k} \to 0. \tag{21}$$

Тогда $v_k \to 0$ (*п. н*), $\mathbf{M}v_k \to 0$, *причем для всякого* $\varepsilon > 0$, $k > 0$

$$\mathbf{P}(v_j \leqslant \varepsilon \text{ для всех } j \geqslant k) \geqslant 1 - \varepsilon^{-1}\left(\mathbf{M}v_k + \sum_{i=k}^{\infty}\beta_i\right). \tag{22}$$

Доказательство. Беря безусловное математическое ожидание от обеих частей (20), получаем

$$\mathbf{M}v_{k+1} \leqslant (1-\alpha_k)\,\mathbf{M}v_k + \beta_k.$$

Отсюда по лемме 3 $\mathbf{M}v_k \to 0$. С другой стороны, $u_k =$

$= v_k + \sum_{i=k}^{\infty}\beta_i$ — полумартингал (ср. с доказательством леммы 9).

Используя леммы 8 и 9, получаем требуемый результат.

3. Основные теоремы. Рассматривается итеративный процесс вида

$$x^{k+1} = x^k - \gamma_k s^k, \tag{23}$$

где k — номер итерации, x^k, s^k — векторы в $\mathbf{R}^n$, $\gamma_k \geqslant 0$ — скалярный множитель, характеризующий длину шага. Мы объединим детерминированный и стохастический случаи — будет рассматриваться общая ситуация, когда x^k и s^k случайны, а детерминированный процесс включается в нее как частный случай. Основные предположения о процессе заключаются в следующем.

А. Процесс носит *марковский характер* — распределение s^k зависит только от x^k и k, s^k $s^k(x^k)$, величины s^k, s^{k-1}, ... взаимно независимы.

Б. Существует скалярная функция *(функция Ляпунова)*

$$V(x) \geqslant 0, \quad \inf_{x \in R^n} V(x) = 0, \quad V(x)$$

дифференцируема и $\nabla V(x)$ удовлетворяет условию Липшица с константой *L*.

В. Процесс (23) является *псевдоградиентным* по отношению к *V(x):*

$$(\nabla V(x^k), \ \mathbf{M}(s^k \mid x^k)) \geqslant 0, \tag{24}$$

т. е. — s^k в среднем является направлением убывания $V(x)$ в точке x^k.

Г. Выполняется следующее *условие роста* на s^k:

$$\mathbf{M}(\| s^k \|^2 \mid x^k) \leqslant \sigma^2 + \tau (\nabla V(x^k), \ \mathbf{M}(s^k \mid x^k)). \tag{25}$$

Величина σ^2 обычно характеризует уровень аддитивных помех. Случай $\sigma=0$ типичен для детерминированных задач.

Д. Начальное приближение удовлетворяет условию

$$\mathbf{M}V(x^0) < \infty. \tag{26}$$

Разумеется, это условие выполняется, если x^0 — детерминированный вектор.

Е. Длина шага такова, что

$$\gamma_k \geqslant 0, \quad \sum_{k=0}^{\infty} \gamma_k = \infty, \quad \overline{\lim_{k \to \infty}} \gamma_k < \frac{2}{L\tau}. \tag{27}$$

Приведем основные теоремы о сходимости. При условиях А—Е нельзя, вообще говоря, утверждать, что $V(x^k) \to 0$ для процесса (23) в каком-либо вероятностном смысле. Например, если $s^k \equiv 0$, то все условия выполняются, но $x^k \equiv x^0$. Однако некоторые утверждения о сходимости справедливы даже при этих минимальных предположениях.

Теорема 1. *Пусть выполнены условия* А — Е *и либо* $\sigma^2 = 0$, *либо* $\sum_{k=0}^{\infty} \gamma_k^2 < \infty$. *Тогда при любом* x^0 *в алгоритме* (23)

$$V(x^k) \to V \ (\text{п. н.}), \quad \overline{\lim_{k \to \infty}} (\nabla V(x^k), \ \mathbf{M}(s^k \mid x^k)) = 0 \ (\text{п. н.}). \tag{28}$$

Доказательство. Используя условие Б и формулу

$$\|g(x+y)-g(x)-g'(x)y\| \leq L\|y\|^2/2,$$

получаем

$$V(x^{k+1}) \leqslant V(x^k) - \gamma_k (\nabla V(x^k), \ s^k) + L\gamma_k^2 \| s^k \|^2/2.$$

Возьмем условное математическое ожидание обеих частей этого неравенства и применим условие Г:

$$\leqslant V(x^k) - \gamma_k (\nabla V(x^k),\ \mathbf{M}(s^k \mid x^k)) + L\gamma_k^2 \mathbf{M}(\|s^k\|^2 \mid x^k)/2 \leqslant$$

$$\leqslant V(x^k) - \gamma_k (1 - (1/2) L\tau\gamma_k)(\nabla V(x^k),\ \mathbf{M}(s^k \mid x^k)) + L\gamma_k^2\sigma^2/2. \quad (29)$$

В силу условий В и Г

$$\mathbf{M}(V(x^{k+1}) \mid x^k) \leqslant V(x^k) + L\gamma_k^2\sigma^2/2. \quad (30)$$

Применяя лемму 9, получаем, что $V(x^k) \to V$ (п. н.). Перейдем в (29) к безусловным математическим ожиданиям:

$$\mathbf{M}V(x^{k+1}) \leqslant \mathbf{M}V(x^k) - \gamma_k (1 - (1/2) L\tau\gamma_k) u_k + L\gamma_k^2\sigma^2/2,$$

$$u_k = \mathbf{M}(\nabla V(x^k),\ \mathbf{M}(s^k \mid x^k)).$$

Для достаточно больших k, в силу условия

$$E,\ 1 - (1/2) L\tau\gamma_k \geqslant \varepsilon > 0,$$

т. е.

$$\mathbf{M}V(x^{k+1}) \leqslant \mathbf{M}V(x^k) - \gamma_k \varepsilon u_k + L\gamma_k^2\sigma^2/2.$$

Поскольку $\mathbf{M}V(x^0) < \infty$ (условие Д) и $\sigma^2 \sum_{k=0}^{\infty} \gamma_k^2 < \infty$, то отсюда следует, что $\sum_{k=0}^{\infty} \gamma_k u_k < \infty$. Но так как

$$\sum_{k=0}^{\infty} \gamma_k = \infty,$$

то это означает, что

$$\varliminf_{k\to\infty} u_k = 0.$$

Из свойств сходимости в среднем следует, что если для случайных величин $z^k \geqslant 0$, $\mathbf{M}z^k \to 0$, то найдется подпоследовательность $z^{k_i} \to 0$ (п. н.). Поэтому

$$\varliminf_{k\to\infty} (\nabla V(x^k),\ \mathbf{M}(s^k \mid x^k)) = 0 \ \text{(п. н.)}.$$

Заменим условие В на условие В' *сильной псевдоградиентности:*

$$\text{В}'.\ (\nabla V(x^k),\ \mathbf{M}(s^k \mid x^k)) \geqslant lV(x^k),\quad l > 0.$$

Теорема 2. *Пусть выполнены условия* А— Е *и* В' *и либо* $\sigma^2 = 0$, *либо* $\sum_{k=0}^{\infty} \gamma_k^2 < \infty$. *Тогда при любом* x^0 *в алгоритме* (23)

$$V(x^k) \to 0 \ (п.\ н.),$$

$$\mathbf{P}(V(x^i) \leqslant \varepsilon \ \forall i \geqslant k) \geqslant 1 - \varepsilon^{-1}\left(\mathbf{M}V(x^k) + \tfrac{1}{2} L\sigma^2 \sum_{i=k}^{\infty} \gamma_i^2\right). \quad (31)$$

Доказательство. Из (29) и условия B' получаем

$$\mathbf{M}(V(x^{k+1}) \mid x^k) \leqslant (1 - l\gamma_k(1 - (1/2) L\tau\gamma_k)) V(x^k) + L\gamma_k^2\sigma^2/2. \quad (32)$$

Из леммы 10 и условия E следует требуемый результат.

Перейдем к условиям сходимости в среднем.

Теорема 3. *Пусть выполнены условия* A — E, B' *и либо* $\sigma^2 = 0$, *либо* $\gamma_k \to 0$. *Тогда в алгоритме* (23)

$$\mathbf{M}V(x^k) \to 0. \quad (33)$$

Доказательство. Беря безусловное математическое ожидание в (32), имеем

$$\mathbf{M}V(x^{k+1}) \leqslant (1 - l\gamma_k(1 - (1/2) L\tau\gamma_k)) \mathbf{M}V(x^k) + L\gamma_k^2\sigma^2/2. \quad (34)$$

Поскольку $1 - (1/2) L\tau\gamma_k \geqslant \varepsilon > 0$ для достаточно больших k, то

$$\mathbf{M}V(x^{k+1}) \leqslant (1 - l\varepsilon\gamma_k) \mathbf{M}V(x^k) + L\gamma_k^2\sigma^2/2.$$

По лемме 3 $\mathbf{M}V(x^k) \to 0$. A

Из неравенства (34) можно получать и другие результаты, в том числе оценки скорости сходимости. Приведем несколько примеров.

Теорема 4. *Пусть выполнены условия* A — E, B' *и* $\gamma_k \equiv \gamma$, $0 < \gamma < 2/(L\tau)$. *Тогда*

$$\mathbf{M}V(x^k) \leqslant \mathbf{M}V(x^0) q^k + \frac{L\gamma\sigma^2}{l(2 - L\tau\gamma)}(1 - q^k), \quad q = 1 - l\gamma(1 - (1/2) L\tau\gamma). \quad (35)$$

Этот результат следует из (34) и леммы 1.

Таким образом, если $\sigma^2 > 0$, то

$$\overline{\lim_{k\to\infty}} \mathbf{M}V(x^k) \leqslant L\gamma\sigma^2/[l(2 - L\tau\gamma)],$$

если же $\sigma^2 = 0$, то $\mathbf{M}V(x^k)$ стремится к 0 со скоростью геометрической прогрессии.

Теорема 5. *Пусть выполнены условия* A — E, B', $\sigma^2 > 0$ *и* $\gamma_k = \gamma/k$. *Тогда*

$$\mathbf{M}V(x^k) = \begin{cases} O(1/k) & при \quad l\gamma > 1, \\ O(1/k^{l\gamma}) & при \quad l\gamma < 1. \end{cases} \quad (36)$$

Этот результат легко можно получить из (34) и леммы 4.

4. Возможные модификации. Приведенные теоремы о сходимости не являются самыми общими и охватывающими все случаи. Они могут быть видоизменены в различных направлениях.

Во-первых, условия В, В' и Г могут быть обобщены следующим образом:

$$(\nabla V(x^k),\ \mathbf{M}(s^k \mid x^k)) \geqslant l_k V(x^k) - \beta_k, \tag{37}$$

$$\mathbf{M}(\|s^k\|^2 \mid x^k) \leqslant \sigma_k^2 + \tau_k (\nabla V(x^k),\ \mathbf{M}(s^k \mid x^k)) + \mu_k V(x^k). \tag{38}$$

При определенных условиях на l_k, β_k, σ_k, τ_k и μ_k можно с помощью лемм данного пункта доказать аналоги теорем 1—3.

Такого рода ситуации, когда выполняются условия (37) и (38), встретятся нам далее при изучении конечно-разностных вариантов градиентного метода, методов регуляризации и т. д.

Во-вторых, все приведенные до сих пор результаты носили глобальный характер — предполагалось, что условия на $V(x)$, $s^k(x)$ и т. д. выполняются для всех x, а начальное приближение x^0 могло быть любым. Однако нередко такого рода предположения выполняются лишь локально, в окрестности решения. Естественно, что при этом и утверждения о сходимости должны носить локальный характер. Примерами могут служить теоремы 4 и 1 о локальной сходимости градиентного метода и метода Ньютона. Наличие случайных помех вносит некоторые осложнения — возникает ненулевая вероятность выхода из области, в которой выполнены предположения. Поэтому локальные утверждения о сходимости могут выполняться лишь с некоторой вероятностью 1 — δ, δ>0. Приведем соответствующий аналог теоремы 2. Пусть

$$Q = \{x:\ V(x) \leqslant \varepsilon\},$$

где ε > 0 — некоторое число.

Теорема 6. *Пусть условия* А — Е, В' *выполнены для всех* $x,\ x^k \in Q$. *Тогда для метода* (23):

а) *если* x^0 *детерминировано,* $x^0 \in Q$, $\sigma^2 = 0$, s^k *детерминировано, то* $V(x^k) \to 0$;

б) *если то* $\sum_{k=0}^{\infty} \gamma_k^2 < \infty$,

$$\mathbf{P}(x^k \in Q\ \forall k) \geqslant 1 - \delta, \quad \mathbf{P}(V(x^k) \to 0) \geqslant 1 - \delta,$$

$$\delta = \varepsilon^{-1} \mathbf{M} V(x^0) + \frac{1}{2} L \sigma^2 \varepsilon^{-1} \sum_{k=0}^{\infty} \gamma_k^2. \tag{39}$$

Далее, можно рассматривать *непрерывные аналоги итеративных методов* — процессы, описываемые обыкновенными дифференциальными уравнениями

$$dx/dt = s(x,\ t), \quad x(0) = x^0. \tag{40}$$

Для них можно применить ту же технику, основанную на функции Ляпунова. При этом формулировки многих теорем о сходимости упрощаются и приобретают более наглядный смысл. Исторически метод функций Ляпунова и возник применительно к подобным задачам. Мы, однако, не будем приводить соответствующие результаты и рассматривать непрерывные методы. Дело в том, что развитие цифровой техники привело к тому, что теперь ЭВМ являются основными средствами решения вычислительных задач. Но при реализации процесса (40) на ЭВМ нужно переходить к его дискретной аппроксимации, т. е. вновь вернуться к итеративным методам. В то же время нужно иметь в виду, что переход к «предельной» форме дискретной траектории может быть целесообразен с методической точки зрения для упрощения формулировок и «угадывания» различных методов. Для обоснования сходимости подобный подход систематически используется в теории оптимизации.

Наконец, часто итерационный процесс рассматривается в форме

$$x^{k+1} = T(x^k), \quad T: \mathbf{R}^n \to \mathbf{R}^n, \tag{41}$$

а не в виде (23). Постулируется существование функции *V(x),* обладающей свойством

$$V(T(x)) < V(x), \quad x \neq T(x), \tag{42}$$

при этом ни дифференцируемости, ни гладкости *V(x)* и *T(x)* не требуется. Достаточно предположить, например, полунепрерывность снизу функции $\varphi(x)=V(T(x))$ и ограниченность множества $\{x: V(x) \leqslant V(x^0)\}$. При этих условиях удается доказать, что у последовательности (41) есть предельные точки, и каждая из них является неподвижной точкой *T(x).*

15.3. 3. Другие схемы

Не нужно думать, что первый и второй методы Ляпунова исчерпывают все многообразие схем исследования сходимости итерационных процедур. Иногда эти схемы основываются на несколько иных соображениях. Упомянем кратко некоторые Из них.

1. Принцип сжимающих отображений. Пусть $g: \mathbf{R}^n \to \mathbf{R}^n$ — некоторое отображение. Оно называется *сжимающим,* если

$$\|g(x) - g(y)\| \leqslant q\|x - y\|, \quad q < 1, \tag{1}$$

для всех $x, y \in \mathbf{R}^n$, т. е. если оно удовлетворяет условию Липшица с константой, меньшей 1. Рассмотрим итеративный процесс

$$x^{k+1} = g(x^k). \tag{2}$$

Теорема 1 (принцип сжимающих отображений). *Если*

g — сжимающее отображение, то оно имеет единственнцю неподвижную точку x, к которой сходится процесс* (2) *при любом* x^0 *со скоростью геометрической прогрессии*

$$\|x^k - x^*\| \leqslant q^k (1-q)^{-1} \|g(x^0) - x^0\|. \qquad (3)$$

Доказательство.

$$\|x^{k+1} - x^k\| = \|g(x^k) - g(x^{k-1})\| \leqslant q\|x^k - x^{k-1}\|,$$

$$\|x^{k+1} - x^k\| \leqslant q^k \|x^1 - x^0\|,$$

$$\|x^{k+s} - x^k\| \leqslant \sum_{i=k}^{k+s-1} \|x^{i+1} - x^i\| \leqslant$$

$$\leqslant (q^{k+s-1} + q^{k+s-2} + \ldots + q^k)\|x^1 - x^0\| \leqslant \frac{q^k}{1-q}\|x^1 - x^0\|. \qquad (4)$$

Следовательно, $\|x^{k+s} - x^k\| \to 0$ при $k \to \infty$ и любом s, т. е. x^k — последовательность Коши в $\mathbf{R}^n$. В силу полноты $\mathbf{R}^n$ x^k имеет предел x^*. Так как $g(x)$ непрерывна в силу (1), то из $x^k \to x^*$ следует $g(x^k) \to g(x^*)$, но $g(x^k) = x^{k+1} \to x^*$. Поэтому $x^* = g(x^*)$. Переходя в (4) к пределу при $s \to \infty$, получаем $\|x^* - x^k\| \leqslant (q^k/(1-q))\|x^1 - x^0\|$. Единственность неподвижной точки сразу следует из (1).

Принцип сжимающих отображением удобен тем, что он не только утверждает сходимость итеративного процесса, но и гарантирует существование неподвижной точки. Поэтому он традиционно применялся в математике для получения разнообразных теорем существования.

Принцип сжимающих отображений допускает различные обобщения и модификации. Однако, как показывают приводимые ниже упражнения 1—3, существенно расширить его нельзя.

Отметим еще, что попытка непосредственно применить принцип сжимающих отображений к задачам, рассмотренным в п. 15.3.1, не дает результатов. В самом деле, там было показано, что если спектральный радиус $\rho(A)$ матрицы A меньше 1, то итерации $x^{k+1} = Ax^k$ сходятся. Однако в этих условиях линейное отображение $g(x) = Ax$ не является, вообще говоря, сжимающим, так как не обязательно $\|A\| < 1$, см. п. 15.3.1.

2. Теорема о неявной функции. Удобным инструментом при исследовании итеративных методов, не разрешенных относительно x^{k+1}, является хорошо известная из анализа теорема о неявной функции. Пусть $F(x, y)$ — отображение из $\mathbf{R}^n \times \mathbf{R}^n$ в $\mathbf{R}^n$. Будем

обозначать $F'_x(x, y)$, $F'_y(x, y)$ производные F по соответствующим переменным.

Теорема 2 (о неявной функции). *Пусть* $F(x^*, y^*) = 0$, $F(x, y)$ *непрерывна по* $\{x, y\}$ *в окрестности* x^*, y^*, *дифференцируема по* x *в окрестности* x^*, y^*, $F'_x(x, y)$ *непрерывна в* x^*, y^* *и матрица* $F'_x(x^*, y^*)$ *невырождена. Тогда существует единственная непрерывная в окрестности* y^* *функция* $x = \varphi(y)$ *такая, что* $x^* = \varphi(y^*)$, $F(\varphi(y), y) = 0$. *Если, кроме того,* $F'_y(x^*, y^*)$ *существует, то* $\varphi(y)$ *дифференцируема в* y^* *и*

$$\varphi'(y^*) = -[F'_x(x^*, y^*)]^{-1} F'_y(x^*, y^*). \quad A \tag{5}$$

Иными словами, уравнение $F(x, y)=0$ может быть разрешено относительно x в окрестности y^*. Применим этот результат прежде всего для исследования существования и устойчивости решений уравнений.

Теорема 3. *Пусть уравнение* $g(x) = 0$, $g: \mathbb{R}^n \to \mathbb{R}^n$, *имеет решение* x^*, *причем* $g(x)$ *дифференцируема в окрестности* x^*, $g'(x)$ *непрерывна в* x^* *и матрица* $g'(x^*)$ *невырождена. Тогда уравнение*

$$g(x) = y \tag{6}$$

имеет решение $x(y)$ *при достаточно малых* y, *причем*

$$x(y) = x^* - g'(x^*)^{-1} y + o(y). \tag{7}$$

Приведенные результаты позволяют исследовать итерационные процессы, в которых новое приближение x^{k+1} задается неявным выражением, например оно является решением некоторой вспомогательной задачи безусловной минимизации. Именно так обстоит дело в методе регуляризации и многих методах решения задач с ограничениями (например, методах штрафных функций).

15.4. Роль теорем сходимости

Ниже излагается взгляд Б.Т. Поляка, известного специалиста в области теории оптимизации, на роль теорем сходимовти в задачах оптимизации

1. Две крайние точки зрения. Возьмем какую-нибудь книгу по методам оптимизации, написанную «математиком для математиков». Основную ее часть составляют теоремы о сходимости методов и их доказательства. Их формулировки максимально общи и абстрактны, используется аппарат современного функционального анализа. Критерии оценки результатов те же, что и в «чистой» математике — глубина, красота и простота утверждений и доказательств.

Комментарии и примеры почти отсутствуют; сравнительный анализ методов не производится; важность или эффективность методов не обсуждается; численных примеров нет. Читателю, который интересуется использованием методов, приходится самому догадываться о связи математических результатов с практикой вычислений, и зачастую такую связь установить не просто. При этом нередко (особенно в журнальной литературе) такому же формальному исследованию подвергаются методы малоинтересные, а иногда и заведомо неэффективные. Это дало повод для появления остроумной пародии на «наукообразные» работы по методам оптимизации, написанной Вульфом. Увы, эта пародия не исправила положения (более того, многие читатели восприняли статью всерьез, не поняв ее нарочитой нелепости).

Такая ситуация породила другой крайний взгляд, по существу отвергающий роль теории в разработке и изучении методов оптимизации. Его сторонники считают, что при создании метода достаточно эвристических соображений. Строгое доказательство сходимости излишне, так как условия теорем труднопроверяемы в конкретных задачах, сам факт сходимости мало что дает, а оценки скорости сходимости неточны и неэффективны. Кроме того, при реализации метода возникает масса обстоятельств, строгий учет которых невозможен (ошибки округления, приближенное решение различных вспомогательных задач и т. д.) и которые могут сильно повлиять на ход процесса. Поэтому единственным критерием оценки метода является практика его применения.

Не будем обсуждать эти точки зрения на абстрактном уровне, так как это потребовало бы решения общих проблем о предмете и стиле вычислительной математики. Вместо этого попытаемся на примере приведенных выше результатов о сходимости двух методов безусловной минимизации выяснить, в какой мере могут быть полезны теоремы о сходимости и почему они требуют к себе достаточно осторожного отношения.

2. Зачем нужны теоремы о сходимости? Ответ на этот «наивный» вопрос не так прост. Конечно, для математика, занимающегося теоретическим обоснованием методов, теоремы представляют самостоятельный интерес с точки зрения используемой в них техники, полноты исследования методов и т. д. Однако чем могут быть полезны такие теоремы человеку, собирающемуся решать практическую задачу?

Прежде всего, условия теорем выделяют класс задач, для которых можно рассчитывать на применимость метода. Эта информация нередко носит отрицательный характер — если условия теоремы не

выполняются, то метод может (но разумеется не обязан) оказаться неработоспособным. Так, наименее жесткие предположения, при которых можно обосновать сходимость градиентного метода в форме (5), заключаются в достаточной гладкости минимизируемой функции (теорема 1). При обсуждении теоремы мы видели, что нарушение этих предположений действительно может привести к расходимости процесса. Аналогичным образом более сильные условия гладкости функции для применимости метода Ньютона (теорема 1), как мы видели из примеров, также существенны. Удобно, когда подобные требования носят качественный характер (гладкость, выпуклость и т. п.) — это позволяет их проверять даже в сложных задачах. Важно также, чтобы требования в теоремах не были завышенными. Например, если судить по теореме 3, то для применимости градиентного метода нужно существование второй производной. Однако в действительности это требование излишне (см. теорему 1); оно нужно лишь для получения оценок скорости сходимости. Поэтому полезно иметь несколько теорем, в которых даются утверждения об одном методе при различных предположениях (таковы теоремы 1—4 для градиентного метода).

Теоремы о сходимости дают также важную информацию о качественном поведении метода: сходится ли он для любого начального приближения или только для достаточно хорошего, в каком смысле сходится (по функции, по аргументу или в пределе удовлетворяется условие экстремума и т. д.). Так, теорема 1 гарантирует применимость градиентного метода из любой начальной точки, в то же время утверждается лишь, что $\nabla f(x^k) \to 0$ (а сходимость по функции или аргументу может отсутствовать, что подтверждают рассмотренные там же примеры). В теореме 1 наоборот обосновывается сходимость метода Ньютона (по аргументу к глобальному минимуму) лишь для хорошего начального приближения, и как мы видели выше, это требование является существенным. Поэтому при практическом использовании метода Ньютона нужно иметь хорошее начальное приближение, в противном случае возможна расходимость метода.

Полезная информация нередко содержится и в самом доказательстве теорем о сходимости. Чаще всего они построены на той идее, что некоторая скалярная величина монотонно убывает в процессе итераций. В теоремах 1, 2 такой величиной является сама минимизируемая функция, в теоремах 3, 4 — расстояние до точки минимума, в теореме 1 — норма градиента. Часто эта величина доступна $(f(x), \|\nabla f(x)\|)$ и по ее фактическому поведению в процессе вычислений можно судить о сходимости или расходимости

метода — при нормальном течении процесса она должна убывать. Если же доказательство основано, например, на монотонном убывании $\|x^k - x^*\|$, то неразумно требовать монотонного убывания $f(x)$ на каждом шаге.

Особенно важные сведения о методе дает оценка скорости сходимости. Эти сведения могут быть как положительного, так и отрицательного характера. Например, оценка скорости сходимости метода Ньютона, содержащаяся в теореме 1, показывает, что метод сходится чрезвычайно быстро. Действительно, если начальное приближение достаточно близко к решению ($q< 1$), то в соответствии с (6) $\|x^k - x^*\| \leqslant 2q^{2^k}$ (так как $l \leq L$). Поэтому для q=0,5 будет $\|x^k - x^*\| \leqslant 2^{-2^k+1}$, так что $\|x^5 - x^*\| < 10^{-9}$, а для q = 0,1 имеем $\|x^k - x^*\| \leqslant 2 \cdot 10^{-2^k}$, так что $\|x^4 - x^*\| < 10^{-16}$. Иными словами, если метод Ньютона применим, то обычно требуется не более 4—5 итераций для получения решения с очень высокой точностью. С другой стороны, градиентный метод при оптимальном выборе γ в соответствии с терремой 3 сходится со скоростью геометрической прогрессии со знаменателем $q = (L - l)/(L + l)$, причем мы видели, что эта оценка является точной для случая квадратичной функции. Для больших чисел обусловленности $\mu = L/l$ знаменатель прогрессии $q \approx 1 - 2/\mu$ близок к 1. Как мы увидим в дальнейшем, нередко для самых простых задач среднеквадратического приближения полиномами величина μ достигает значений 10^8 и выше. Ясно, что при $\mu = 10^8$, нужно сделать порядка $2 \cdot 10^8$ итераций, чтобы уменьшить $\|x^0 - x^*\|$ в e раз. Иными словами, градиентный метод в такой ситуации неработоспособен. Этот отрицательный результат о поведении градиентного метода удается получить чисто теоретически, не прибегая ни к каким численным экспериментам. Применительно к другим задачам минимизации он дает основания для настороженного отношения к градиентному методу — вряд ли можно рассчитывать на этот метод как эффективное средство решения сложных задач.

Теоретическая оценка скорости сходимости показывает также, от каких факторов зависит поведение метода. Так, для градиентного метода «трудны» задачи плохо обусловленные, а выбор начального приближения не влияет на скорость сходимости, тогда как для метода Ньютона скорость определяется качеством начального приближения и близостью функции к квадратической, но не обусловленностью задачи. Для метода сопряженных градиентов, как мы увидим в дальнейшем, основную роль играет размерность задачи, тогда как в полученных выше оценках для методов градиентного и Ньютона размерность явно

не входит. На основе этих соображений можно делать ориентировочные выводы о целесообразности применения различных методов в той или иной конкретной ситуации.

Наконец, при достаточно полной информации о задаче можно с помощью результатов о скорости сходимости заранее выбрать (или оценить) требуемое число итераций для достижения необходимой точности. Так, если мы находимся в условиях теоремы 3 и известны оценки для l, L и $||x^0 - x^*||$, то можно указать число шагов k, гарантирующее точность $||x^k - x^*|| \le \varepsilon$ в градиентном методе с оптимальным $\gamma = 2/(L + l)$:

$$k \ge \ln \frac{\varepsilon}{\|x^0 - x^*\|} \Big/ \ln \frac{\mu - 1}{\mu + 1} \approx \frac{\mu}{2} \ln \frac{\|x^0 - x^*\|}{\varepsilon}, \quad \mu = \frac{L}{l}.$$

3. Необходима осторожность. Прислушаемся к другой точке зрения, критикующей теоретическое исследование методов как излишнюю, а иногда и вредную роскошь.

Сторонники этой точки зрения указывают, что сам факт сходимости метода ровно ничего не говорит об эффективности последнего. Безусловно, это так. Ошибочно считать, что данный метод можно применять на практике, если его сходимость доказана — ведь скорость сходимости может быть безнадежно медленной. Однако мы уже отмечали, что теоремы сходимости (даже не содержащие оценок скорости сходимости) дают важную информацию об области применимости метода, его качественном поведении и т. п. Разумеется, вся эта информация недостаточна для окончательных выводов о целесообразности и возможности применения метода для решения конкретной задачи.

Далее, результаты о сходимости часто критикуют за неконструктивный характер. Их предположения трудно проверить, входящие в них параметры неизвестны, оценки носят асимптотический характер и т. п. Такие обвинения во многом обоснованы. Нередко теоремы о сходимости чрезвычайно громоздки, и проверить их для какой-либо реальной задачи невозможно. Еще хуже, когда формулировки носят апостериорный характер («...пусть в процессе итераций выполняется такое-то соотношение...») Почему бы тогда просто не предположить, что $x^k \to x^*$? Однако не всегда ситуация столь мрачная. Как видно из приведенных теорем, предположения в них просты и носят общий характер (требуются гладкость, выпуклость, сильная выпуклость, невырожденность и тому подобные естественные и часто легко проверяемые условия). Константы L, l и q, входящие в формулировки теорем, обычно действительно неизвестны, поэтому конструктивный выбор γ в градиентном методе или явные оценки скорости сходимости невозможны. Однако существуют более сложные

способы выбора γ_k в градиентном методе, для которых приведенные теоремы служат основой. Что же касается скорости сходимости, то хотя ее количественная оценка не всегда доступна, ее качественный характер не вызывает сомнений. Наконец, оценки скорости сходимости совсем не обязательно носят асимптотический характер — так, в приведенных теоремах они верны для всех конечных *k.*

Еще один упрек теоремам о сходимости заключается в том, что они рассматривают идеализированную ситуацию, отвлекаясь от наличия помех, ошибок округления,, невозможности точного решения вспомогательных задач и т. п., а все эти факторы сильно влияют на поведение метода в реальных условиях. Действительно, в приведенных выше теоремах предполагалось, что градиент вычисляется точно, что обращение матрицы в методе Ньютона делается без погрешностей и т. д. Далее мы рассмотрим те же методы при наличии разного рода помех. Оказывается, их влияние заметно сказывается на эффективности методов. Поэтому оценки качества методов нужно делать с учетом более общих теорем о сходимости, рассчитанных на наличие помех.

Подводя итог, можно сказать, что теоретические исследования методов оптимизации могут дать много информации вычислителю-практику. Нужно лишь при этом проявлять разумную осторожность и здравый смысл.

4. О роли общих схем исследования сходимости. Общие теоремы типа приведенных в этом разделе берут на себя стандартную, рутинную часть доказательств сходимости и тем самым упрощают процесс обоснования алгоритмов. Однако не нужно преувеличивать их роль и считать, что они делают анализ сходимости элементарным. Во-первых, во многих случаях проверка их условий представляет самостоятельную нетривиальную проблему. Во-вторых, для простых задач непосредственное, «в лоб», доказательство ничуть не сложнее обращения к общим теоремам.

Таким образом, анализ сходимости остается творческим процессом, требующим исскуства и здравого смысла.

16. Устойчивость.

16.1. Устойчивость по Ляпунову

Понятие устойчивости как способности того или иного объекта, состояния или процесса сопротивляться не учитываемым заранее внешним воздействиям появилось еще в античной науке и сейчас

занимает одно из центральных мест в физике и технике. Существуют различные конкретные реализации этого общего понятия в зависимости от типа рассматриваемого объекта, характера внешних воздействий и г. д. Одна из таких реализаций появлялась у нас ранее. Сейчас мы рассмотрим понятие *устойчивости по Ляпунову,* одно из наиболее важных, введенное и систематически изученное А. М. Ляпуновым в 1892 г.

Пусть состояние некоторого объекта описывается конечным числом параметров, для определенности тремя параметрами, *x, y, z,* так что изменение этого объекта во времени задается тремя функциями $x = x(t), y = y(t), z = z(t)$ (*t*—время). Пусть закон этого изменения имеет вид системы дифференциальных уравнений

$$\left.\begin{aligned} \frac{dx}{dt} &= P(x, y, z), \\ \frac{dy}{dt} &= Q(x, y, z), \\ \frac{dz}{dt} &= R(x, y, z) \end{aligned}\right\} \qquad (1)$$

с заданными правыми частями, не содержащими явно независимой переменной *t.* Последнее условие означает, что дифференциальный закон развития процесса не меняется с течением времени.

Пусть состояние *равновесия* рассматриваемого объекта, когда он не меняется с течением времени, описывается постоянными значениями $x=x_0, y = y_0, z = z_0$; тогда эта система постоянных, рассматриваемых как функции времени, также должна удовлетворять системе (1). Из непосредственной подстановки в (1) следует, что для этого необходимо и достаточно, чтобы одновременно

$$P(x_0, y_0, z_0) = 0, \quad Q(x_0, y_0, z_0) = 0, \quad R(x_0, y_0, z_0) = 0. \qquad (2)$$

Пусть в некоторый момент t_0 объект под влиянием каких-то причин вышел из состояния равновесия, т. е. параметры *x, y, z* стали равными

$$x = x_0 + \Delta x_0, \quad y = y_0 + \Delta y_0, \quad z = z_0 + \Delta z_0.$$

Чтобы выяснить дальнейшее изменение рассматриваемого объекта, надо решить систему уравнений (1) при начальных условиях

$$x(t_0) = x_0 + \Delta x_0, \quad y(t_0) = y_0 + \Delta y_0, \quad z(t_0) = z_0 + \Delta z_0. \qquad (3)$$

Исследуемое состояние равновесия называется устойчивым по Ляпунову, если после бесконечно малого выхода из этого состояния объект продолжает оставаться в бесконечной близости от него на протяжении всего дальнейшего времени. Другими словами, при бесконечно малых $\Delta x_0, \Delta y_0, \Delta z_0$ для решения системы (1) при начальных условиях (3) разности

$$\Delta x = x(t) - x_0, \quad \Delta y = y(t) - y_0, \quad \Delta z = z(t) - z_0$$

должны быть бесконечно малыми на всем интервале времени $t_0 < t < \infty$.

На первый взгляд может показаться странным рассмотрение бесконечно малых отклонений параметров и бесконечно большого промежутка времени, так как на практике все эти величины конечны. Однако полезно вспомнить различие практической и математической бесконечностей. Практической бесконечно малой является просто малая в масштабах рассматриваемого процесса реальная величина, а практически бесконечным промежутком времени является время *переходного процесса,* т. е перехода от исследуемого состояния к состоянию иного типа (например, от одного состояния равновесия к другому или от состояния равновесия к разрушению объекта и т. п.). Таким образом, реально устойчивость по Ляпунову означает, что малый выход из состояния равновесия практически не нарушает этого состояния.

Для выяснения того, будет ли иметь место устойчивость, подставим в систему (1) $x = x_0 + \Delta x$, $y = y_0 + \Delta y$, $z = z_0 + \Delta z$, что даст

$$\left.\begin{aligned} \frac{d(\Delta x)}{dt} &= P(x_0 + \Delta x, \ y_0 + \Delta y, \ z_0 + \Delta z) = \\ &= (P'_x)_0 \Delta x + (P'_y)_0 \Delta y + (P'_z)_0 \Delta z + \ldots, \\ \frac{d(\Delta y)}{dt} &= (Q'_x)_0 \Delta x + (Q'_y)_0 \Delta y + (Q'_z)_0 \Delta z + \ldots, \\ \frac{d(\Delta z)}{dt} &= (R'_x)_0 \Delta x + (R'_y)_0 \Delta y + (R'_z)_0 \Delta z + \ldots, \end{aligned}\right\} \qquad (4)$$

где обозначено $(P'_x)_0 = P'_x(x_0, y_0, z_0)$ и т. п. Здесь при преобразовании правых частей мы воспользовались формулой Тейлора и формулами (2), многоточиями обозначены члены выше первого порядка малости.

Так как при выяснении устойчивости рассматриваются лишь малые Δx, Δy, Δz, то в правых частях системы (4) основную роль играют выписанные, линейные члены. Поэтому заменим систему (4) на *укороченную систему (систему первого приближения),* отбросив члены высшего порядка малости:

$$\left.\begin{aligned} \frac{d(\Delta x)}{dt} &= (P'_x)_0 \Delta x + (P'_y)_0 \Delta y + (P'_z)_0 \Delta z, \\ \frac{d(\Delta y)}{dt} &= (Q'_x)_0 \Delta x + (Q'_y)_0 \Delta y + (Q'_z)_0 \Delta z, \\ \frac{d(\Delta z)}{dt} &= (R'_x)_0 \Delta x + (R'_y)_0 \Delta y + (R'_z)_0 \Delta z \end{aligned}\right\} \qquad (5)$$

Система (5) — это линейная система с постоянными коэффициентами, которая решается известным методом. При этом решение системы (5)

получается как комбинация функций вида e^{pt}, где p удовлетворяет характеристическому уравнению

$$\begin{vmatrix} (P'_x)_0 - p & (P'_y)_0 & (P'_z)_0 \\ (Q'_x)_0 & (Q'_y)_0 - p & (Q'_z)_0 \\ (R'_x)_0 & (R'_y)_0 & (R'_z)_0 - p \end{vmatrix} = 0. \qquad (6)$$

При этом малым

$$\Delta x_0, \quad \Delta y_0, \quad \Delta z_0$$

отвечают малые значения произвольных постоянных C_1, C_2, C_3 и поэтому все дело в поведении функции e^{pt} при возрастании t. Так как при $p = r + is$ (случай $s = 0$ не исключен) будет

$$|e^{pt}| = e^{rt},$$

то

$$|e^{pt}|_{t \to \infty} \to 0 \text{ (при } r < 0), \quad |e^{pt}|_{t \to \infty} \to \infty \text{ (при } r > 0), \quad (7)$$

и мы приходим к следующим выводам. *Если все корни характеристического уравнения* (6) *имеют отрицательную вещественную часть* (в частности, они могут быть вещественными отрицательными), *то рассматриваемое состояние равновесия* x_0, y_0, z_0 *истойчиво по Ляпунову.* Кроме того, в этом случае при малых Δx_0, Δy_0, Δz_0, будет

$$x(t) \to x_0, \quad y(t) \to y_0, \quad z(t) \to z_0,$$

при $t \to \infty$; такая устойчивость называется *асимптотической, если же среди корней уравнения* (6) *имеется по крайней мере один с положительной вещественной частью, то рассматриваемое состояние равновесия неустойчиво по Ляпунову.*

Эти результаты мы вывели для системы (5), но согласно сказанному выше те же утверждения справедливы для полной системы (4). Отметим, что наличие кратных корней у уравнения (6) не нарушает наших утверждений, несмотря на то, что при этом в решении могут появиться степени t в качестве множителей, так как экспонента (7) при $r < 0$ стремится к нулю быстрее любой степени t.

Оба полученных вывода не охватывают случая, когда среди корней уравнения (6) нет корней с положительной вещественной частью, но имеется по крайней мере один с нулевой вещественной частью. Тогда в общем решении системы (5) появляются функции вида

$$e^{ist} = \cos st + i \sin st, \qquad |e^{ist}| = 1,$$

т. е. получается, будто бы рассматриваемый объект колеблется или остается неподвижным около состояния равновесия, не стремясь к нему. Но тогда из-за неограниченности времени начинают влиять отброшенные члены высшего порядка малости, которые могут

нарушить устойчивость. Итак, *в рассматриваемом особом случае по корням уравнения* (6) *нельзя заключить об устойчивости или неустойчивости состояния равновесия;* чтобы это сделать, надо привлечь какие-либо дополнительные соображения, например привлечь дальнейшие члены разложений (4).

Полученные результаты приобретают особенно простой вид для случая, когда изменение объекта описывается одной функцией *x(t),* удовлетворяющей дифференциальному уравнению

$$\frac{dx}{dt} = f(x). \qquad (8)$$

Мы получаем, что если $f(x_0) = 0,\ f'(x_0) < 0,$ то значению $x = x_0$ отвечает устойчивое состояние равновесия, а если $f(x_0) = 0,\ f'(x_0) > 0,$ то это состоя ние неустойчивое. (Получите этот вывод, исходя из расположения изоклин для уравнения (8) на плоскости *t, x.)*

16.2. Элементы теории устойчивости

Во многих задачах важно знать не *одно конкретное* решение задачи, отвечающей данным начальным условиям, а характер поведения решения при изменении начальных условий и при изменении аргумента. Этими вопросами занимается качественная теория дифференциальных уравнений, одним из основных разделов которой является *теория устойчивости решения,* или теория устойчивости движения.

Пусть некоторое явление описывается системой дифференциальных уравнений

$$\frac{dy_i}{dt} = f_i(t,\ y_1,\ \ldots,\ y_n) \qquad (i = 1,\ \ldots,\ n) \qquad (1)$$

с начальными условиями

$$y_i(t_0) = y_{i0} \qquad (i = 1,\ \ldots,\ n). \qquad (2)$$

Условия (2) обычно являются результатом измерения и, следовательно, получены с некоторой точностью.

Если сколь угодно малые изменения начальных данных способны сильно изменить решение, то решение системы (1), определяемое выбранными нами *не точными* начальными данными, не имеет никакого значения и даже приближенно не может описывать явление.

Поэтому важно знать условия, при которых малое изменение условий (2) влечет малое изменение решения системы (1).

Если t меняется в достаточно малом конечном промежутке $|t_0 - t| \leqslant T$, то ответ на этот вопрос можно получить на основании теоремы существования и единственности.

Теорема 1 (о непрерывной зависимости решения от начальных условий). *Если правая часть дифференциального уравнения*

$$\frac{dy}{dt} = f(t,\ y) \tag{3}$$

непрерывна и по переменному y имеет ограниченную частную производную ($|f'_y| \leqslant N$) *на прямоугольнике*

$$D = \{t_0 - a \leqslant t \leqslant t_0 + a,\ \ y_0 - b \leqslant y \leqslant y_0 + b\},$$

уравнения (3) $y(t) = y(t, t_0, y_0)$, *удовлетворяющее начальному условию* $y(t_0) = y_0$, *непрерывно зависит от начальных данных. Точнее, для всякого* $\varepsilon > 0$ *найдется такое число* $\delta > 0$, *что если* $|y_0 - \bar{y}_0| < \delta$, *то*

$$|y(t,\ t_0,\ y_0) - y(t,\ t_0,\ \bar{y}_0)| < \varepsilon$$

при

$$|t_0 - t| < T,\ T < T_0,\quad T_0 = \min\left\{a,\ \frac{1}{N},\ \frac{b}{M}\right\},$$
$$M = \max_{(t,\ y) \in D} |f(t,\ y)|.$$

Доказательство. При доказательстве теоремы существования мы получили, что

$$y(t) = y(t,\ t_0,\ y_0) = y_0 + \int_{t_0}^{t} f(t,\ y(t))\,dt,$$

$$\bar{y}(t) = y(t,\ t_0,\ \bar{y}_0) = \bar{y}_0 + \int_{t_0}^{t} f(t,\ \bar{y}(t))\,dt.$$

Отсюда, применяя теорему Лагранжа (пояснения ниже), получим

$$|y(t) - \bar{y}(t)| \leqslant |y_0 - \bar{y}_0| + \int_{t_0}^{t} |f(t,\ y(t)) - f(t,\ \bar{y}(t))|\,dt \leqslant$$
$$\leqslant |y_0 - \bar{y}_0| + N|t - t_0| \max_t |y(t) - \bar{y}(t)| \leqslant$$
$$\leqslant |y_0 - \bar{y}_0| + NT \max_t |y(t) - \bar{y}(t)|.$$

Так как $TN < 1$, то

$$(1 - TN)|y(t) - \bar{y}(t)| \leqslant |y_0 - \bar{y}_0|,$$

или

$$|y(t) - \bar{y}(t)| \leqslant |y_0 - \bar{y}_0|/(1 - TN).$$

Если теперь мы возьмем

$$|y_0-\bar{y}_0|<\delta, \quad \text{где} \quad \delta=\varepsilon(1-TN),$$

то

$$|y(t)-\bar{y}(t)|<\varepsilon.$$

Сделаем некоторые пояснения. Имеет место неравенство

$$|y(t,\ t_0,\ y_0)-y_0|\leqslant\left|\int_{t_0}^{t} f(t,\ y(t))\,dt\right|\leqslant|t-t_0|M\leqslant TM,$$

где $TM\leqslant T_0M<b$ или $b-TM=\delta_0>0$, показывающее, что интегральная кривая $y(t,\ t_0,\ y_0)$ для $|t—t_0|<T$ принадлежит к прямоугольнику, находящемуся строго внутри прямоугольника D. Отсюда видно, что кривая

$$y(t,\ t_0,\ \bar{y}_0),\ |t-t_0|<T,$$

тоже не выходит за пределы прямоугольника D, если только

$$|\bar{y}_0-y_0|<\delta_0.$$

Это показывает, что теорема Лагранжа была применена выше обоснованно, во всяком случае при $\delta < \delta_0$, — ведь функция $f(t, y)$ по условию имеет непрерывную производную

$$\frac{\partial f}{\partial y}$$

на D.

Подобная теорема верна и для системы (1).

При выполнении всех условий теоремы говорят, что *задача поставлена корректно.*

Мы изучали устойчивость решения на достаточно малом отрезке значений t.

Если же аргумент $t \in [t_0, \infty)$ может принимать какие угодно значения, то вопросом зависимости решения от начальных данных занимается теория устойчивости.

Определение. Пусть $\varphi(t)=\{\varphi_1(t),\ \ldots,\ \varphi_n(t)\}$ — решение системы (1). Решение $\varphi(t)$ системы (1) называется *устойчивым по Ляпунову*, если для всякого $\varepsilon > 0$ существует $\delta > 0$ такое, что для любого решения $y(t) =\{y_1(t),\ldots,y_n(t)\}$ той же системы, начальные значения которого удовлетворяют неравенствам

$$|y_i(t_0)-\varphi_i(t_0)|<\delta \quad (i=1,\ \ldots,\ n), \tag{4}$$

справедливы неравенства

$$|y_i(t)-\varphi_i(t)|<\varepsilon \quad (i=1,\ \ldots,\ n, \quad \forall t\in[t_0,\ \infty)). \tag{5}$$

Таким образом, решение $\varphi(t)$ устойчиво по Ляпунову, если близкие к нему по начальным условиям решения остаются близкими и для всех $t \geq t_0$ (рис. 1).

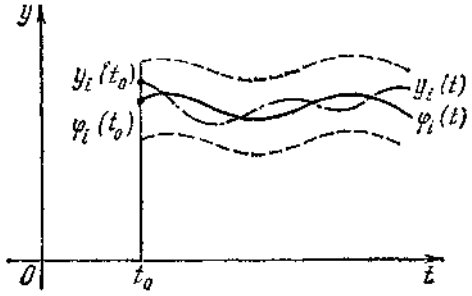


Рис. 1.

Если решение $\varphi(t)$ устойчиво по Ляпунову и, кроме того,

$$\lim_{t \to \infty} |y_i(t) - \varphi_i(t)| = 0 \qquad (i = 1, \ldots, n), \tag{6}$$

то оно называется *асимптотически устойчивым.*

Отметим, что из (6) не следует устойчивость по Ляпунову.

Пример 1. $\frac{dy}{dx} = -y, \; y(x_0) = y_0.$

Общее решение этого уравнения $y=C\exp(-x)$. Решение, удовлетворяющее начальному условию, имеет вид

$$y = y_0 \exp(x_0 - x).$$

Если теперь мы зададим другое начальное условие $\bar{y}(x_0) = \bar{y}_0$, то решение будет

$$\bar{y} = \bar{y}_0 \exp(x_0 - x).$$

Отсюда

$$|y - \bar{y}| = |y_0 - \bar{y}_0| \exp(x_0 - x) \leqslant |y_0 - \bar{y}_0|,$$

при $x \geq x_0$. Поэтому, если $|y_0 - \bar{y}_0| < \delta = \varepsilon$, то $|y - \bar{y}| \leqslant \varepsilon$, т. е. решение $y = y_0\exp(x_0 - x)$ устойчиво по Ляпунову при $x \geq x_0$. Это решение также и асимптотически устойчиво, так как

$$\lim_{x \to +\infty} |y - \bar{y}| = \lim_{x \to +\infty} |y_0 - \bar{y}_0| \exp(x_0 - x) = 0.$$

Пример 2. Для уравнения $y' = y$ аналогично можно показать, что

$$|y - \bar{y}| = |y_0 - \bar{y}_0| \exp(x - x_0)]$$

для $x \geq x_0$ при любом x_0.

Очевидно, каково бы ни было x_0 при $x \geq x_0$, решение y неустойчиво, так как сомножитель $\exp(x - x_0) \to +\infty$ при $x \to +\infty$.

Исследование на устойчивость по Ляпунову произвольного решения $\varphi(t) = \{\varphi_1(t), \ldots, \varphi_n(t)\}$ системы (1) можно свести к исследованию на устойчивость тривиального (тождественно равного нулю) решения некоторой другой системы. Для этого надо перейти к новым неизвестным функциям

$$x_i(t) = y_i(t) - \varphi_i(t) \qquad (i = 1, \ldots, n). \tag{7}$$

Отсюда

$$\frac{dy_i}{dt} = \frac{dx_i}{dt} + \frac{d\varphi_i}{dt}.$$

Поэтому система (1) перейдет в систему

$$\frac{dx_i}{dt} = f_i[t, x_1 + \varphi_1(t), \ldots, x_n + \varphi_n(t)] - f_i[t, \varphi_1(t), \ldots, \varphi_n(t)] \quad (i = 1, \ldots, n). \tag{8}$$

Система (8) имеет тривиальное решение

$$x_i(t) \equiv 0 \qquad (i = 1, \ldots, n). \tag{9}$$

Из сказанного следует теорема.

Теорема 2. *Решение* $\varphi(t) = \{\varphi_1(t), \ldots, \varphi_n(t)\}$ *системы* (1) *устойчиво по Ляпунову (асимптотически устойчиво) тогда и только тогда, когда устойчиво по Ляпунову (асимптотически устойчиво) тривиальное решение (точка покоя) системы* (8).

Это решение обладает тем свойством, что точка $(x_1(t), \ldots, x_n(t))$ в действительности не движется при изменении t, а стоит на месте. Само решение (9) и точка (0, .. , 0) в этом случае называется *положением равновесия системы* (1) или *точкой покоя.*

Условия устойчивости применительно к точке покоя $x_i = 0$ $(i = 1, \ldots, n)$ можно сформулировать так: точка покоя $x_i = 0$ $(i = 1, \ldots, n)$ системы (8) устойчива по Ляпунову, если $\forall \varepsilon > 0 \ \exists \delta(\varepsilon) > 0$ такое, что из неравенства

$$|x_i(t_0)| < \delta(\varepsilon) \qquad (i = 1, \ldots, n)$$

следует

$$|x_i(t)| < \varepsilon \qquad (i = 1, \ldots, n, \ \forall t \geq t_0),$$

т. е. траектория, начальная точка которой находится в некоторой δ-окрестности начала координат, при $t \geq t_0$ не выходит за пределы произвольной ε-окрестности начала координат. Здесь мы говорим об окрестностях прямоугольных, но можно перейти и к сферическим

окрестностям, что удобно особенно при векторной форме записи решения $\boldsymbol{x}(t) = \{x_1(t), \ldots, x_n(t)\}$:

$$\|\boldsymbol{x}(t_0)\| < \delta(\varepsilon) \Rightarrow \|\boldsymbol{x}(t)\| < \varepsilon, \quad \|\boldsymbol{x}\| = \sqrt{\sum_{i=1}^{n} |x_i|^2} \quad (t \geqslant t_0).$$

Замечание 1. Произвольное частное решение $y_0(t)$ линейной неоднородной системы дифференциальных уравнений

$$\frac{d\boldsymbol{y}}{dt} = A\boldsymbol{y} + \boldsymbol{f}(t) \tag{10}$$

устойчиво по Ляпунову (асимптотически устойчиво) тогда и только тогда, когда устойчива по Ляпунову (асимптотически устойчива) точка покоя соответствующей однородной системы (см. теорему 2)

$$\frac{d\boldsymbol{x}}{dt} = A\boldsymbol{x}. \tag{11}$$

В самом деле, система (10) является частным случаем системы (1), а система (11) есть частный случай системы (8). Здесь свободные члены исчезли, так как функция $f(t)$ зависит только от t и не зависит от искомых функций.

Теорема 3 (Ляпунова). *Пусть дана система*

$$\frac{dy_i}{dt} = f_i(t, y_1, \ldots, y_n) \quad (i = 1, \ldots, n), \tag{12}$$

имеющая тривиальное решение $y_i(t) \equiv 0 \ (i = 1, \ldots, n)$.

Пусть существует дифференцируемая функция $v(y_1,\ldots,y_n)$, *удовлетворяющая условиям:*

1) $v(y_1, \ldots, y_n) \geqslant 0$ *и* $v = 0$ *только при* $y_1 = \ldots = y_n = 0$, *т. е. функция* v *имеет строгий минимум в начале координат.*

2) *Полная производная функции* v *вдоль фазовой траектории (т. е. вдоль решения* $y_i(t)$ *системы* (1))

$$\frac{dv}{dt} = \sum_{i=1}^{n} \frac{\partial v}{\partial y_i} \frac{dy_i}{dt} = \sum_{i=1}^{n} \frac{\partial v}{\partial y_i} f_i(t, y_1, \ldots, y_n) \leqslant 0 \ \textit{при} \ t \geqslant t_0.$$

Тогда точка покоя $y_i \equiv 0 \ (i = 1, \ldots, n)$ *устойчива по Ляпунову.*

Если дополнительно потребовать, чтобы вне сколь угодно малой окрестности начала координат $(y_1^2 + \ldots + y_n^2) \geq \delta$

$$\frac{dv}{dt} \leqslant -\beta < 0 \quad (t \geqslant t_0),$$

где β *— постоянная величина, то точка покоя* $y_i(t) = 0 \ (i = 1, \ldots, n)$ *асимптотически устойчива.*

Функция v *называется функцией Ляпунова.*

Пример 3.

$$\frac{dy_1}{dt} = -y_1^5 - y_2,$$

$$\frac{dy_2}{dt} = y_1 - y_2^3.$$

Легко видеть, что точка покоя $y_1 = y_2 = 0$ является решением данной системы. Выясним, будет ли она устойчива.

Рассмотрим функцию $v(y_1, y_2) = y_1^2 + y_2^2$. Она удовлетворяет всем условиям теоремы:

1) $v(y_1, y_2) \geq 0$ и $v = 0$ только при $y_1 = y_2 = 0$.
2) Вдоль фазовой траектории

$$\frac{dv}{dt} = \frac{\partial v}{\partial y_1}\frac{dy_1}{dt} + \frac{\partial v}{\partial y_2}\frac{dy_2}{dt} = 2y_1(-y_1^5 - y_2) + 2y_2(y_1 - y_2^3) = -2(y_1^6 + y_2^4) \leqslant 0.$$

Кроме того, вне окрестности начала координат $(y_1^2 + y_2^2 \geqslant \delta > 0)$

$$\frac{dv}{dt} \leqslant -\beta < 0$$

(где β — минимум функции $2(y_1^6 + y_2^4)$ вне круга

$$y_1^2 + y_2^2 = \delta).$$

Значит, решение $y_1 = y_2 \equiv 0$ асимптотически устойчиво.

Замечание 2. Функцию Ляпунова рекомендуется искать в виде квадратичной формы от аргументов $y_1, \ldots, y_n$, т. е.

$$v = \sum_{i,\,j} a_{ij} y_i y_j.$$

Первое требование говорит о том, что v должна быть положительно определенной квадратичной формой. Каким образом подбирать коэффициенты a_{ij}, чтобы форма v была положительно определенной, указывается в теореме Сильвестра

$$\left(a_{11} > 0,\ \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0,\ \ldots,\ \begin{vmatrix} a_{11} & \ldots & a_{1n} \\ \ldots & \ldots & \ldots \\ a_{n1} & \ldots & a_{nn} \end{vmatrix} > 0\right).$$

16.3. Классификация точек покоя

Исследование на устойчивость системы двух линейных уравнений первого порядка с постоянными коэффициентами

$$\left.\begin{aligned} \frac{dx}{dt} &= a_{11}x + a_{12}y, \\ \frac{dy}{dt} &= a_{21}x + a_{22}y \end{aligned}\right\} \qquad (1)$$

можно провести на основании теоремы Ляпунова. Однако систему (1) можно исследовать на устойчивость и непосредственно, так как мы можем ее решить. Будем предполагать, что определитель

$$\Delta = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \neq 0. \qquad (2)$$

Легко видеть, что $y(t) = x(t) \equiv 0$ является решением системы (1) с нулевыми начальными условиями.

Для нахождения общего решения мы должны найти корни характеристического уравнения

$$\begin{vmatrix} a_{11}-\lambda & a_{12} \\ a_{21} & a_{22}-\lambda \end{vmatrix} = \lambda^2 - (a_{11}+a_{22})\lambda + \Delta = 0. \qquad (3)$$

Из условия $\Delta \neq 0$ следует, что $\lambda=0$ не является корнем характеристического уравнения (3).

1. Пусть корни характеристического уравнения λ_1 и λ_2 действительны и различны. Далее, пусть $\alpha = (\alpha_1, \alpha_2)$, $\beta = (\beta_1, \beta_2)$ — собственные векторы матрицы *A,* отвечающие корням λ_1 и λ_2 соответственно, т. е.

$$\left.\begin{aligned} (a_{11}-\lambda_1)\alpha_1 + a_{12}\alpha_2 &= 0, \\ a_{21}\alpha_1 + (a_{22}-\lambda_1)\alpha_2 &= 0, \end{aligned}\right\} \quad \left.\begin{aligned} (a_{11}-\lambda_2)\beta_1 + a_{12}\beta_2 &= 0, \\ a_{21}\beta_1 + (a_{22}-\lambda_2)\beta_2 &= 0. \end{aligned}\right\}$$

Тогда, как мы знаем, общее решение системы (1) имеет вид

$$\left.\begin{aligned} x &= C_1\alpha_1 e^{\lambda_1 t} + C_2\beta_1 e^{\lambda_2 t}, \\ y &= C_1\alpha_2 e^{\lambda_1 t} + C_2\beta_2 e^{\lambda_2 t}, \end{aligned}\right\} \qquad (5)$$

где C_1, C_2—произвольные постоянные.

Если $\lambda_1 < 0$, $\lambda_2 < 0$, то из (5) видно, что точка покоя $x=y\equiv 0$ асимптотически устойчива.

В самом деле, будем считать, например, что $t_0 = 0$, тогда решение (5), проходящее через точку (x_0, y_0) в момент времени t_0 определяется постоянными C_1 и C_2, которые вычисляются из уравнений

$$x_0 = C_1\alpha_1 + C_2\beta_1,$$
$$y_0 = C_1\alpha_2 + C_2\beta_2,$$

где $\alpha_1\beta_2 - \alpha_2\beta_1 \neq 0$. Но тогда

$$C_1 = Ax_0 + By_0, \quad C_2 = Dx_0 + Ey_0,$$

где *A, B, C, D, E* — некоторые константы. Следовательно,

$$|x(t)| \leqslant |Ax_0 + By_0||\alpha_1| + |Dx_0 + Ey_0||\beta_1|,$$
$$|y(t)| \leqslant |Ax_0 + By_0||\alpha_2| + |Dx_0 + Ey_0|\beta_2|,$$

потому что $|e^{\lambda_1 t}| \leqslant 1$, $|e^{\lambda_2 t}| \leqslant 1$ при $\lambda_1 < 0$, $\lambda_2 < 0$.

Отсюда следует, что для любого $\varepsilon > 0$ найдется $\delta > 0$ такое, что как только $|x_0|$, $|y_0| < \delta$, выполняются неравенства

$$|x(t)|, \quad |y(t)| < \varepsilon \quad (t > 0),$$

т. е. точка (0, 0) устойчива по Ляпунову. Кроме того, в силу того, что

$$\exp(\lambda_i t) \to 0 \quad (t \to +\infty),$$

из (5), очевидно, следует, что точка (0, 0) асимптотически устойчива.

Если мы исключим аргумент t из системы (5), то полученная функция $y=\varphi\delta(x)$ дает траекторию движения в системе xOy.

Материальная точка, находящаяся в начальный момент времени $t = t_0$ в δ-окрестности начала координат, при достаточно большом t переходит в точку, лежащую в ε-окрестности начала координат и при $t \to +\infty$ стремится к началу координат.

Такая точка покоя называется *устойчивым узлом* (рис. 1).

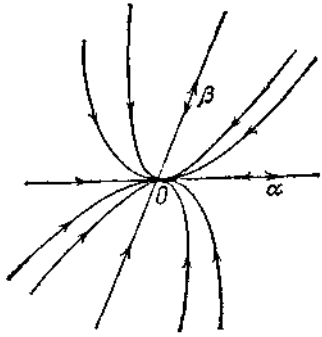


Рис. 1

На рис. 1 изображено расположение траекторий, соответствующих данному случаю. Стрелками мы указываем направление движения точки по траектории при $t \to +\infty$. Все траектории, кроме одной, в точке (0, 0) имеют общую касательную. Если считать, что $|\lambda_1| < |\lambda_2|$, то угловой коэффициент касательной равен α_2/α_1.

В самом деле, из (1) и (5) имеем ($C_1 \neq 0$)

$$\frac{dy}{dx}=\frac{a_{21}\left(C_1\alpha_1 e^{\lambda_1 t}+C_2\beta_1 e^{\lambda_2 t}\right)+a_{22}\left(C_1\alpha_2 e^{\lambda_1 t}+C_2\beta_2 e^{\lambda_2 t}\right)}{a_{11}\left(C_1\alpha_1 e^{\lambda_1 t}+C_2\beta_1 e^{\lambda_2 t}\right)+a_{12}\left(C_1\alpha_2 e^{\lambda_1 t}+C_2\beta_2 e^{\lambda_2 t}\right)}=$$

$$=\frac{a_{21}\left(C_1\alpha_1+C_2\beta_1 e^{(\lambda_2-\lambda_1)t}\right)+a_{22}\left(C_1\alpha_2+C_2\beta_2 e^{(\lambda_2-\lambda_1)t}\right)}{a_{11}\left(C_1\alpha_1+C_2\beta_1 e^{(\lambda_2-\lambda_1)t}\right)+a_{12}\left(C_1\alpha_2+C_2\beta_2 e^{(\lambda_2-\lambda_1)t}\right)}\longrightarrow$$

$$\xrightarrow[t\to+\infty]{}\frac{a_{21}C_1\alpha_1+a_{22}C_1\alpha_2}{a_{11}C_1\alpha_1+a_{12}C_1\alpha_2}=\frac{a_{21}\alpha_1+a_{22}\alpha_2}{a_{11}\alpha_1+a_{12}\alpha_2}=\frac{\lambda_1\alpha_2}{\lambda_1\alpha_1}=\frac{\alpha_2}{\alpha_1},$$

если $\alpha_1 \neq 0$, так как в силу (4)

$$a_{21}\alpha_1+a_{22}\alpha_2=\lambda_1\alpha_2,\quad a_{11}\alpha_1+a_{12}\alpha_2=\lambda_1\alpha_1.$$

Если же $\alpha_1 = 0$, то, рассуждая так же, получим

$$\frac{dx}{dy}\longrightarrow\frac{\alpha_1}{\alpha_2}=0\quad(\lambda_1\neq 0!).$$

Если $C_1 = 0$, то из (5) получаем одну траекторию

$$y=\frac{\beta_2}{\beta_1}x.$$

Касательная к этой траектории (прямой) имеет угловой коэффициент β_2/β_1. Таким образом, касательная к траекториям, у которых $C_1 \neq 0$, параллельна собственному вектору $\alpha=(\alpha_1,\ \alpha_2)$, отвечающему наименьшему по абсолютной величине соственному числу λ_1 (при $\alpha_1 = 0$ вектор направлен по оси *y*).

Кроме того, имеется одна траектория (при $C_1 = 0$), а именно, прямая

$$y=\frac{\beta_2}{\beta_1}x,$$

которая параллельна второму собственному вектору $\beta = (\beta_1, \beta_2)$, отвечающему большему но модулю собственному числу λ_2.

Если теперь $\lambda_1 > 0$ и $\lambda_2 > 0$, то из (5) видно, что точка покоя $x = y \equiv 0$ неустойчива, так как $\exp(\lambda_i t) \to +\infty$ при $t \to +\infty$. Такую точку покоя называют *неустойчивым узлом.* Данный случай получается из предыдущего заменой t на $(-t)$. Поэтому траектории будут иметь прежний вид, но движение точки по траектории будет происходить в противоположном направлении (рис. 2).

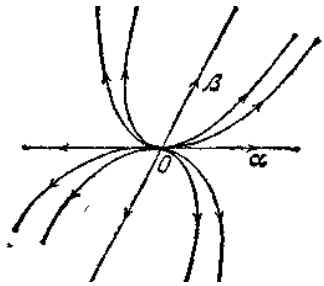


Рис. 2.

Наконец, если число $\lambda_1 < 0$, а $\lambda_2 > 0$ (или, наоборот, $\lambda_1 > 0$, $\lambda_2 < 0$), то точка покоя тоже неустойчива, так как $\exp(\lambda_2 t) \to +\infty$ при $t \to +\infty$. Точки, находящиеся в δ-окрестности начала координат, по траектории

$$x = C_2\beta_1 e^{\lambda_1 t}, \quad y = C_2\beta_2 e^{\lambda_1 t}$$

уходят в бесконечность. Отметим, что в данном случае имеется траектория, по которой движение точки происходит в направлении начала координат при $t \to +\infty$:

$$x = C_1\alpha_1 e^{\lambda_1 t}, \quad y = C_1\alpha_2 e^{\lambda_1 t}.$$

Это прямая $\alpha_1 y - \alpha_2 x = 0$. Точка покоя данного вида называется *седлом* (рис. 3).

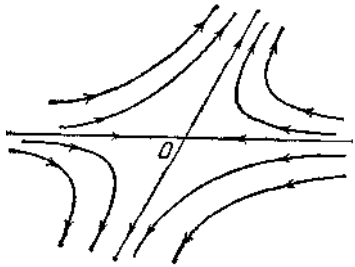


Рис. 3.

2. Пусть корни λ_1 и λ_2 комплексные (a_{kl} действительны!)

$$\lambda_1 = p + iq, \quad \lambda_2 = p - iq \qquad (q \neq 0).$$

Общее решение системы (1) можно записать в виде (5), где векторы α и $\beta = \bar{\alpha} = (\bar{\alpha}_1, \bar{\alpha}_2)$ будут уже с комплексными координатами.

Действительная и мнимая части этого решения также являются решением системы. Поэтому общее решение системы (1) можно записать в виде линейной комбинации этих решений:

$$\left.\begin{aligned} x &= e^{pt}(C_1 \cos qt + C_2 \sin qt), \\ y &= e^{pt}(a \cos qt + b \sin qt), \end{aligned}\right\} \qquad (6)$$

где C_1 и C_2 — произвольные постоянные и a, b — некоторые линейные комбинации этих постоянных

$$(a = kC_1 + lC_2, \quad b = mC_1 + nC_2).$$

Проиллюстрируем этот факт на конкретном примере.

Пример 1. Найти общее решение системы

$$\begin{cases} \dot{x} = x - y, \\ \dot{y} = 2x - y. \end{cases}$$

Характеристическое уравнение

$$\begin{vmatrix} 1-\lambda & -1 \\ 2 & -1-\lambda \end{vmatrix} = 0, \quad \text{или} \quad \lambda^2 + 1 = 0$$

имеет корни $\lambda_1 = i$, $\lambda_2 = -i$. Координаты векторов α и β находим из равенств

$$(1-i)\alpha_1 - \alpha_2 = 0, \quad (1+i)\beta_1 - \beta_2 = 0,$$

т. е. можно положить $\alpha_1 = 1$, $\alpha_2 = 1 - i$; $\beta_1 = 1$, $\beta_2 = 1 + i$. Тогда

$$x = \alpha_1 e^{it} = \cos t + i \sin t,$$
$$y = \alpha_2 e^{it} = (1-i)(\cos t + i \sin t) =$$
$$= (\cos t + \sin t) + i(\sin t - \cos t)$$

— решение системы.

Действительная и мнимая части этого решения тают являются решениями системы, причем линейно независимыми. Поэтому их линейная комбинация дает общее решение нашей системы:

$$x = C_1 \cos t + C_2 \sin t,$$
$$y = C_1(\cos t + \sin t) + C_2(\sin t - \cos t) =$$
$$= (C_1 - C_2)\cos t + (C_1 + C_2)\sin t.$$

Таким образом, в данном случае

$$a = C_1 - C_2, \quad b = C_1 + C_2.$$

При $p=0$ траектории (6) для различных C_1, C_2 в силу периодичности множителей в скобках являются замкнутыми кривыми — эллипсами в центром в точке (0, 0) (рис. 4).

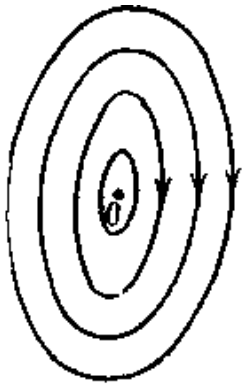


Рис. 4

Эта точка называется *центром.* При $p = 0$ нет асимптотической устойчивости — точка $(x(t), y(t))$ движется по одному из эллипсов указанного семейства, обходя его бесконечное число раз. Она, очевидно, не стремится ни к какому пределу при $t \to +\infty$. С другой стороны, при $p = 0$ точка покоя (0,0) является устойчивой по Ляпунову. Проверим это утверждение на разобранном выше примере 1. Найдем решение рассмотренной в примере 1 системы, проходящее в момент времени $t_0 = 0$ черезточку x_0, y_0. Очевидно,

$$x_0 = C_1, \quad y_0 = C_1 - C_2,$$

следовательно,

$$C_1 = x_0, \quad C_2 = x_0 - y_0$$

и решение имеет вид

$$x(t) = x_0 \cos t + (x_0 - y_0) \sin t,$$
$$y(t) = y_0 \cos t + (2x_0 - y_0) \sin t.$$

Имеем далее

$$|x(t)| \leqslant |x_0| + |x_0| + |y_0| = 2|x_0| + |y_0|,$$
$$|y(t)| \leqslant |y_0| + 2|x_0| + |y_0| = 2|x_0| + 2|y_0|.$$

Зададим $\varepsilon > 0$, и пусть $\delta = \varepsilon/4$. Тогда, очевидно, если $|x_0|, |y_0| < \delta$, то для всех t

$$|x(t)| \leqslant 2 \cdot \frac{\varepsilon}{4} + \frac{\varepsilon}{4} < \varepsilon,$$

$$|y(t)| \leqslant 2 \cdot \frac{\varepsilon}{4} + 2 \cdot \frac{\varepsilon}{4} = \varepsilon.$$

Мы доказали устойчивость по Ляпунову точки покоя. Из (6) видно, что при $p < 0$ точка (x, y) при $t \to +\infty$ стремятся к нулевой точке $x=0$, $y=0$, называемой *устойчивым фокусом.* Наличие множителя $\exp(pt) \to 0 \ (t \to +\infty)$ превращает замкнутые кривые в спирали, приближающиеся асимптотически при $t \to +\infty$ к началу координат (рис. 5).

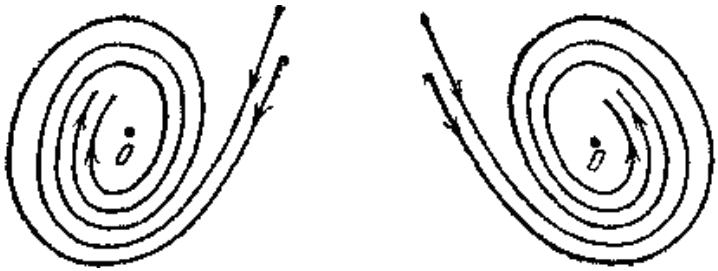


Рис. 5

Точки, находящиеся при $t = t_0$, в произвольной δ-окрестности начала координат, при достаточно большом t попадают в заданную ε-окрестность точки (0, 0).

Траектории, стремящиеся к фокусу, обладают тем свойством, что касательные к ним при $t \to +\infty$ не стремятся ни к какому пределу. Этим фокус отличается от узла.

В случае $p<0$ точка $x = y = 0$ асимптотически устойчива.

Если действительная часть p корней λ_1 и λ_2 положительна, то этот случай переходит в предыдущий при замене t на $(-t)$. Следовательно, траектории сохраняют форму как на рис. 5, только движение точки будет происходить в противоположном направлении. Так как $\exp(pt) \to +\infty$ при $t \to +\infty$, то точки, находящиеся в начальный момент времени в окрестности начала координат, затем уходят в бесконечность. Такая точка покоя носит название *неустойчивого фокуса* (рис. 6).

Рис. 6

3. Пусть корни λ_1, λ_2 равны между собой ($\lambda_1 = \lambda_2$, a_{kl} действительна!). Тогда они действительны и общее решение системы (1) имеет вид

$$\left.\begin{aligned} x &= (A + Bt)\, e^{\lambda_1 t}, \\ y &= (C + Dt) e^{\lambda_1 t}, \end{aligned}\right\} \qquad (7)$$

где A, B, C, D — константы, связанные между собой двумя линейными уравнениями, которые можно получить, если подставить функции x и y в систему и сократить на множитель $\exp(\lambda_1 t)$.

Если $\lambda_1<$и, то $\exp(\lambda_1 t) \to 0$, $t\exp(\lambda_1 t) \to 0$ при $t \to +\infty$, и следовательно, точка покоя $x = 0$, $y = 0$ асимптотически устойчива. Ее называют *устойчивым узлом* (как в п. 1).

Если же $\lambda_1 > 0$, то точка $x = 0$, $y=0$ неустойчива и называется она *неустойчивым узлом.*

Замечание 1. Если $\Delta=0$, то характеристическое уравнение (3) имеет корень $\lambda_1 = 0$ и $\lambda_2 = a_{11} + a_{22}$.

Пусть $\lambda_2 \neq 0$. Тогда общее решение системы (1) запишется так:

$$x = C_1\alpha_1 + C_2\beta_1 e^{\lambda_2 t},$$
$$y = C_1\alpha_2 + C_2\beta_2 e^{\lambda_2 t},$$

где C_1, C_2—произвольные постоянные и $a_{11}\alpha_1 + a_{12}\alpha_2 = 0$,
$-a_{22}\beta_1 + a_{12}\beta_2 = 0$.

Исключая параметр t, получаем семейство параллельных прямых

$$y - C_1\alpha_2 = \frac{\beta_2}{\beta_1}(x - C_1\alpha_1).$$

Если $\lambda_2<0$, то при $t\to+\infty$ на каждой траектории (на одном из параллельных лучей) точки приближаются к лежащей на этой траектории точке покоя $x=C_1\alpha_1$,

$y = C_1\alpha_2 \left(y = \frac{\alpha_2}{\alpha_1}x\right)$ (рис.7).

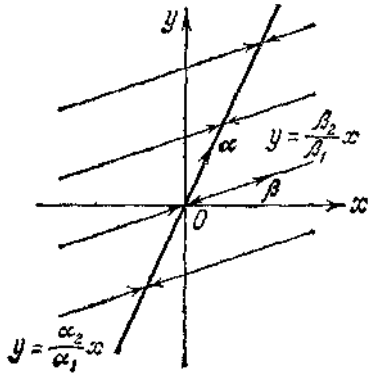


Рис.7

Точка (0, 0), так же как любая точка прямой

$$y = \frac{\alpha_2}{\alpha_1}x,$$

при $\lambda_2<0$ устойчива по Ляпунову, но не является асимптотически устойчивой.

Если $\lambda_2 > 0$, то точка покоя неустойчива.

Если же $\lambda_1 = \lambda_2 = 0$, то может быть два случая:

а) Общее решение системы (1) имеет вид $x = C_1$, $y = C_2$. В этом случае точка покоя $x=y=0$ устойчива по Ляпунову, но не является асимптотически устойчивой. Отметим, что данная ситуация имеет место, когда матрица A нулевая $a_{11} = a_{22} = a_{12} = a_{21} = 0$. В данном

случае все точки плоскости *(x, y)* являются точками покоя, устойчивыми по Ляпунову.

б) Общее решение системы (1) имеет вид

$$x = C_1 + C_2 t, \quad y = a + bt.$$

Точка покоя $x=y\equiv 0$ неустойчива.

В этом случае $a_{22} = -a_{11}$, $a_{12}a_{21} \leqslant 0$.

Пример 2. Выяснить характер точки покоя системы

$$\dot{x} = -x + ay,$$
$$\dot{y} = -2y.$$

Составим характеристическое уравнение

$$\begin{vmatrix} -1-\lambda & a \\ 0 & -2-\lambda \end{vmatrix} = 0.$$

Его корни $\lambda_1=-1$, $\lambda_2= -2$. Значит точка покоя $x=y\equiv 0$ является устойчивым узлом.

Пример 3. Какого типа точку покоя имеет система

$$\dot{x} = x - y,$$
$$\dot{y} = 2x + 3y.$$

Характеристическое уравнение

$$\begin{vmatrix} 1-\lambda & -1 \\ 2 & 3-\lambda \end{vmatrix} = 0$$

имеет комплексные корни $\lambda_1=2+i$, $\lambda_2=2-i$. Действительная часть этих сопряженных корней положительна, поэтому точка покоя $x=y\equiv 0$ является неустойчивый фокусом.

Замечание 2. Если матрица A симметрична, то, как мы знаем, характеристическое уравнение имеет только действительные корни. Кроме того, мы знаем, что

$$a_{11}a_{22} - a_{12}^2 = \lambda_1\lambda_2.$$

Так как симметричная матрица порождает квадратичную форму эллиптического, гиперболического или параболического типа, то мы систему дифференциальных уравнений (1) в этом случае будем называть

эллиптической, если $a_{11}a_{22} - a_{12}^2 = \lambda_1\lambda_2 > 0$,

гиперболической, если $a_{11}a_{22} - a_{12}^2 = \lambda_1\lambda_2 < 0$,

параболической, если $a_{11}a_{22} - a_{12}^2 = \lambda_1\lambda_2 = 0$.

Из изложенного выше ясно, что если система (1) эллиптическая, то точка покоя будет устойчивым узлом, если $\lambda_1 < 0$, $\lambda_2 < 0$. Это возможно, когда $a_{11} < 0$, $a_{22} < 0$,

Если же $\lambda_1 > 0$, $\lambda_2 > 0$, то точка покоя будет неустойчивым узлом $(a_{11} > 0, a_{22} > 0)$.

Отметим, что в эллиптическом случае числа a_{11} и a_{22} одного знака.

Если система (1) гиперболическая $(a_{11}a_{22}-a_{12}^2<0)$, то точка покоя всегда неустойчива (седло).

Если система (1) параболическая, то точка покоя устойчива, когда $\lambda_2=a_{11}+a_{22}\leqslant 0$, и неустойчива, когда

$$\lambda_2=a_{11}+a_{22}>0.$$

Пример 4. Выяснить характер точки покоя у систем:

а) $\dot{x}=-3x+2y$, $\dot{y}=2x-5y$; б) $\dot{x}=x+2y$, $\dot{y}=2x+3y$; в) $\dot{x}=x+\sqrt{3}y$, $\dot{y}=\sqrt{3}x+3y$.

Во всех примерах матрица A симметрична.

Система а) эллиптическая, потому что $a_{11}a_{22}-a_{12}^2=11>0$. Так как $a_{11}=-3<0$, $a_{22}=-5<0$, то точка покоя — устойчивый узел.

Система б) гиперболическая, так как

$$a_{11}a_{12}-a_{12}^2=-1<0.$$

Точка покоя — седло.

Система в) параболическая, потому что $a_{11}a_{22}-a_{22}^2=0$. Так как $a_{11}+a_{22}=4>0$, то точка покоя неустойчива.

Замечание 3. Можно доказать (так же как это сделано при n =2), что точка покоя заведомо устойчива по Ляпунову (асимптотически устойчива) для линейной однородной системы из n>1 уравнений о постоянными коэффициентами

$$\frac{dy}{dt}=Ay,$$

если все корни характеристического уравнения системы имеют отрицательные действительные части.

17.Теория разностных схем — понятия сходимости, аппроксимации и устойчивости

В данном разделе на примере явного метода Эйлера даются основные понятия теории разностных схем — понятия сходимости, аппроксимации и устойчивости. Доказывается теорема Лакса. Описываются основные приемы построения разностных схем.

17.1. Метод ломаных Эйлера

Напомним, что *метод ломаных Эйлера* – это метод нахождения аппроксимирующей интегральную кривую ломаной, который в образах парка со стрелками-указателями может быть представлен так (рис. 1). Из точки $(t_0,\ x_0)=(t_0,\ x^0)$ *расширенного фазового пространства* движемся τ "секунд", сообразуясь со стрелкой, помещенной в этой точке, и не обращая внимания на остальные стрелки. Придя (через время τ) в точку $(t_1,\ x_1)$, меняем направление, пользуясь указанием, задаваемым стрелкой в точке $(t_1,\ x_1)$; через τ секунд приходим в точку $(t_2,\ x_2)$, опять меняем направление и т. д. Полученная траектория и является *ломаной Эйлера*, аппроксимирующей решение задачи (E) – (C).

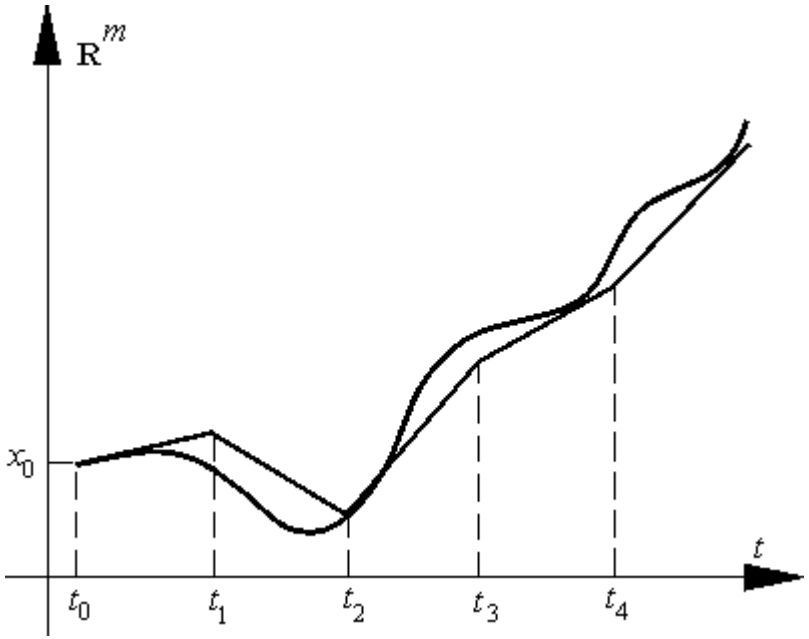


Рис. 1.

Легко видеть, что координаты i-й вершины ($i = 1,..., n$) ломаной Эйлера определяются формулами

$$t_i = t_0 + i\tau, \quad x_0 = x^0, \quad x_i = x_{i-1} + \tau f(t_{i-1}, x_{i-1}). \qquad (1)$$

Перепишем эти равенства в следующем виде:

$$\frac{x_i - x_{i-1}}{\tau} - f(t_{i-1}, x_{i-1}) = 0, \text{ если } i = 1,..., n,$$

$x_i = x^0$, если $i = 0$.

Операторная запись задачи (E) – (C).

Запишем метод ломаных Эйлера в общем виде, в котором могут быть записаны другие разностные методы и который позволяет укладывать исследование таких методов в общую схему. Для этого сначала запишем задачу (E) – (C) в операторной форме. Начальное условие (C) можно записать в виде

$$g(x) = 0,$$

где функция g имеет вид $g(x) = x(t_0) - x^0$. Тогда начальную задачу (E) – (C) можно записать в операторной форме

$$F(x) = 0, \qquad (O)$$

где $F(x) = (x'(\cdot) - f(\cdot, x), g(x))$. Не будем обсуждать подробно вопрос об областях определения и значений оператора F. В случае задачи (E) – (C) можно считать, например, что областью определения D(F) оператора F служит *пространство* $C^1([t_0, t_0 + T], \mathbf{R}^m)$ непрерывно дифференцируемых на отрезке $[t_0, t_0 + T]$ функций со значениями в $\mathbf{R}^m$, а областью значений R(F) — декартово произведение $C^1([t_0, t_0 + T], \mathbf{R}^m) \times \mathbf{R}^m$. Заметим, что решение φ задачи (E) – (C) есть прообраз нуля для оператора F: $\varphi = F^{-1}(0)$.

Операторная форма метода ломаных Эйлера.

Пусть x — *сеточная функция* из S_τ. Определим на S_τ оператор F_τ равенством

$$[F_\tau(x)]_i = x_i - x_{i-1} \tau - f(t_{i-1}, x_{i-1}), \text{ если } i = 1, ..., n,$$

$$x_i - x_0, \qquad \text{если } i = 0; \tag{2}$$

(Здесь $[F_\tau(x)]_i$ обозначает значение сеточной функции $[F_\tau(x)]$ в точке t_i.)

Задача о нахождении ломаной Эйлера, очевидно, эквивалентна задаче о нахождении решения φ_τ уравнения

$$F_\tau(x) = 0 \tag{S}$$

в пространстве S_τ сеточных функций в следующем смысле: значения $(\varphi_\tau)_i$ решения уравнения (S) в узлах сетки и только они являются ординатами вершин ломаной Эйлера в точках t_i (см. формулу (1)). Таким образом, вместо точного решения $\varphi = F^{-1}(0)$ мы ищем приближенное решение $\varphi_\tau = F_\tau^{-1}(0)$.

Разностные схемы.

Любое уравнение вида (S) для нахождения *сеточной функции x* будем называть *разностной схемой*, а его (уравнения) решение, которое мы всегда будем обозначать φ_τ, — *разностным* или *сеточным решением* (уравнения (S)). Оператор F_τ будем называть *разностным оператором*. Разумеется, в таком общем виде определение разностной схемы никак не связано с исходной задачей Коши (E) – (C) (уравнением (O)). В то же время, если в (S) оператор F_τ определен формулой (2), то разностная схема (S) имеет к задаче (E) – (C) самое непосредственное отношение (объяснить, что означают эти слова, и есть цель остальной части параграфа).

В последнем случае разностная схема (S) называется *явной (разностной) схемой Эйлера*. Явной она называется потому, что ее решение может быть выписано в явном виде с помощью рекуррентных соотношений:

$(\varphi_\tau)_0 = x^0$,

$(\varphi_\tau)_i = (\varphi_\tau)_{i-1} + \tau f[t_{i-1}, (\varphi_\tau)_{i-1}], \quad i = 1, \ldots, n.$

Единственное отличие метода ломаных Эйлера от явной схемы Эйлера заключается в том, что в первом ищется функция, заданная на

отрезке $[t_0,\ t_0 + T]$ (ломаная Эйлера), а во втором — на сетке G_τ (вершины этой ломаной).

Пример. Сходимость явной разностной схемы Эйлера.

Как правило разностное решение φ_τ аппроксимирует решение $\varphi(t) = x^0 e^{at}$ интересующей нас исходной (дифференциальной) задачи в следующем смысле: при всех $t \in [0, T]$

$$(\varphi_\tau)_i \to x^0 e^{at} \text{ при } \tau \to 0, \tag{3}$$

где $i = [t/\tau]$ (здесь $[t/\tau]$ — целая часть числа t/τ; ниже мы наряду с обозначением $[a]$ для целой части числа a используем обозначение $\{a\}$ для его дробной части: $a = [a] + \{a\}$). Действительно,

$$\lim_{\tau\to 0} (\varphi_\tau)_i = \lim_{\tau\to 0} x^0(1 + \tau a)^i = x^0 \cdot \lim_{\tau\to 0} (1 + \tau a)^{[t/\tau]} =$$

$$= x^0 \cdot \lim_{\tau\to 0} (1 + \tau a)^{t/\tau - \{t/\tau\}} =$$

$$= x^0 \cdot \lim_{\tau\to 0} (1 + \tau a)^{t/\tau} \cdot \lim_{\tau\to 0} (1 + \tau a)^{-\{t/\tau\}}.$$

Сомножитель $\lim_{\tau\to 0}(1 + \tau a)^{-\{t/\tau\}}$ равен единице, поскольку $\{t/\tau\} \in (0, 1)$. Второй сомножитель также вычисляется тривиально:

$$\lim_{\tau\to 0} (1 + \tau a)^{t/\tau} = \lim_{\tau\to 0} (1 + \tau a)^{[1/(\tau a)]\cdot at} = \left[\lim_{\tau\to 0} (1 + \tau a)^{1/\tau a}\right]^{at} = e^{at}$$

и соотношение (3) доказано.

Сходимость разностных схем.

Будем говорить, что разностная схема (S) *сходится* (к решению φ задачи (E) – (C)), если

$$\|\varphi_\tau - P_\tau\varphi\|_\tau \to 0 \text{ при } \tau \to 0. \qquad (4)$$

Сходимость разностной схемы означает, что при достаточно малом τ значения сеточного (приближенного) решения φ_τ и точного решения φ мало отличаются. Соотношение (4) на практике оказывается мало полезным, поскольку на основании его нельзя судить о том насколько малым мы должны выбрать шаг τ, чтобы в узлах сетки точное и приближенное решения отличались друг от друга не более, чем на ε (заранее заданную точность). Если удается доказать, что при достаточно малых $\tau > 0$

$$\|\varphi_\tau - P_\tau\varphi\|_\tau \le C\tau^k, \qquad (5)$$

где C — не зависящая от τ константа, то говорят, что схема (S) *сходится с порядком k* (или является *схемой k-го порядка* (*сходимости*)). Оценка (5), если в ней известна (для конкретной задачи (E) – (C)) константа C, позволяет по заранее выбранной точности ε *a priori* выбрать шаг так, чтобы приближенное решение аппроксимировало решение данной (дифференциальной) задачи Коши с точностью ε:

$$\|\varphi_\tau - P_\tau\varphi\|_\tau \le \varepsilon;$$

достаточно взять $\tau \le \sqrt[k]{\varepsilon/C}$.

Аппроксимация.

Явная схема Эйлера обладает двумя важными свойствами, из которых, как будет показано ниже, последует ее сходимость с первым порядком. Во-первых, при достаточно малых τ

$$\|F_\tau(P_\tau\varphi)\|_\tau \le C_a\tau, \qquad (6)$$

где C_a — константа, не зависящая от τ, а φ — как обычно, точное решение задачи (E) – (C). В этом случае говорят, что схема (S) имеет *первый порядок аппроксимации на решении*. (Если в правой части неравенства (6) стоит $C_a\tau^k$, то, соответственно, говорят о *схеме k-го порядка аппроксимации* (*на решении*).) Другими словами, неравенство (6) эквивалентно тому, что $\|F_\tau(P_\tau\varphi)\|_\tau = O(\tau)$ (в случае схемы k-го порядка аппроксимации — $O(\tau^k)$). Тот факт, что разностная схема обладает аппроксимацией на решении, означает, грубо говоря, что при подстановке точного решения дифференциальной задачи в разностную схему мы получаем невязку соответствующего порядка малости по τ. (Было бы идеально, если бы после такой подстановки мы получали в левой части нуль, однако в общем случае конструктивно такие схемы выписать нельзя.)

Часто вместо свойства аппроксимации на решении рассматривают формально более жесткое требование, которое называют *свойством аппроксимации* (в зарубежной литературе — *согласованностью*); именно, говорят, что схема (S) является *схемой k-го порядка аппроксимации на функции x*, если при достаточно малых τ

$$\|F_\tau(P_\tau x) - P_\tau F(x)\|_\tau \le C_a\tau^k.$$

Обычно требуется, чтобы схема обладала свойством аппроксимации на множестве функций из некоторого класса гладкости. Очевидно, *если решения дифференциального уравнения (E) m раз непрерывно дифференцируемы, а разностная схема обладает k-м порядком аппроксимации (согласованности) на m раз непрерывно дифференцируемых функциях, то она обладает k-м порядком аппроксимации на решении*.

Аппроксимация явной схемы Эйлера.

Покажем, что *если функция f в (E) непрерывно дифференцируема по t и x, то явная схема Эйлера имеет первый порядок аппроксимации на решении*. Действительно, пусть φ — решение задачи (E) – (C):

$$\varphi'(t) \equiv f[t, \varphi(t)], \quad t \in [t_0, t_0 + T].$$

Поскольку f дифференцируема по t и x, решение φ дважды непрерывно дифференцируемо (см. утверждение о гладкости решений). В частности, найдется M такое, что $\|\varphi''(t)\| \le M$ при всех $t \in [t_0, t_0 + T]$. Кроме того, в силу гладкости φ для любых $t = 1, ..., n$

$$\varphi(t_i) = \varphi(t_{i-1} + \tau) = \varphi(t_{i-1}) + \tau\varphi'(t_{i-1}) + \frac{\tau^2}{2}\varphi''(t_{i-1} + \Phi_i\tau),$$

где $\Phi \in (0, 1)$. Но тогда

$$[F_\tau(P_\tau\varphi)]_i = \frac{(P_\tau\varphi)_i - (P_\tau\varphi)_{i-1}}{\tau} - f[t_{i-1}, (P_\tau\varphi)_{i-1}] =$$

$$= \frac{\varphi(t_i) - \varphi(t_{i-1})}{\varphi} - f[t_{i-1}, \varphi(t_{i-1})] =$$

$$= \frac{\varphi(t_{i-1}) + \tau\varphi'(t_{i-1}) + \frac{\tau^2}{2}\varphi''(t_{i-1} + \Phi_i\tau) - \varphi(t_{i-1})}{\tau} -$$

$$- f[t_{i-1}, \varphi(t_{i-1})] = \varphi'(t_{i-1}) + \frac{\tau^2}{2}\varphi''(t_{i-1} + \Phi_i\tau) - f[t_{i-1}, \varphi(t_{i-1})] =$$

$$= \frac{\tau}{2}\varphi''(t_{i-1} + \Phi_i\tau).$$

(Если $i = 0$, то очевидно, $[F_\tau(P_\tau\varphi)]_i = 0$.) Из сказанного следует, что

$$\|F_\tau(P_\tau\varphi)\|_\tau = \max_{0\le i\le n} \|[F_\tau(P_\tau\varphi)]_i\| \le$$

$$\le \frac{\tau}{2} \max_{0\le i\le n} \|\varphi''(t_{i-1}+\Phi_i\tau)\| \le \tau \frac{M}{2} \stackrel{\text{def}}{=} C_a\tau.$$

Явная схема Эйлера обладает первым порядком аппроксимации (согласованности) на любой дважды непрерывно дифференцируемой функциию

Устойчивость.

Второе важное свойство, которым обладает явная схема Эйлера, называется *устойчивостью* и определяется так: если $z \in S_\tau$ и, кроме того, $\|z\|_\tau$ и τ достаточно малы, то уравнение

$$F_\tau(y) = z \tag{7}$$

однозначно разрешимо и существует такая не зависящая от τ и $\|z\|_\tau$ константа C_s, что

$$\|y - \varphi_\tau\|_\tau = \|F_\tau^{-1}(z) - \varphi_\tau\|_\tau \le C_s\|z\|_\tau. \tag{8}$$

Устойчивость разностной схемы означает, что малые возмущения z в начальных данных и правой части разностной схемы приводят к равномерно малому по τ изменению решения (напомним, что φ_τ — решение невозмущенной системы, а $F_\tau^{-1}(z)$ — возмущенной). Поскольку $\varphi_\tau = F_\tau^{-1}(z)$,неравенство (8), переписанное в виде $\|F_\tau^{-1}(z) - F_\tau^{-1}(0)\|_\tau \le C_s\|z\|_\tau$, означает, в частности, непрерывность обратного к разностному оператору оператора F_τ^{-1}в нуле.

Устойчивость — очень важное в приложениях свойство разностных схем. При практической реализации на ЭВМ разностных методов возникают, в частности, проблемы, связанные с невозможностью представления точных чисел в компьютере. В результате мы решаем не разностную схему (S), а несколько отличающееся от (S) уравнение. Все

такие возмущения в разностной схеме, грубо говоря, можно "перенести в правую часть" и, таким образом, считать, что в ЭВМ ищется решение не разностной схемы (S), но решение возмущенного уравнения (7). Свойство устойчивости разностной схемы гарантирует близость при достаточно малых τ между точным (теоретическим) решением φ_τ разностной схемы и его практической реализацией $F_\tau^{-1}(z)$(где z — суммарный вектор возмущений). Источником возмущений служит не только невозможность точного представления данных в ЭВМ, но и неточность определения физических параметров модели, погрешность измерений и т.п.

Пример. Устойчивость явной схемы Эйлера.

Докажем, что *явная схема Эйлера устойчива.*

Разрешимость уравнения (7) для любых τ и z очевидным образом следует из того, что явная схема Эйлера является явной: значения y_i решения $y = F_\tau^{-1}(z)$ этого уравнения определяются рекуррентными формулами

$$y_0 = x^0 + z_0,$$

$$y_i = y_{i-1} + \tau f(t_{i-1}, y_{i-1}) + \tau z_i, \quad i = 1, ..., n.$$

Обозначим $y - \varphi_\tau$ через ξ. Очевидно,

$$\xi_0 = z_0,$$

$$\xi_i = \xi_{i-1} + \tau f(t_{i-1}, y_{i-1}) - \tau f[t_{i-1},(\varphi_\tau)_{i-1}] + \tau z_i, \quad i = 1, ..., n.$$

Покажем теперь, что

$$\|\xi_i\| \le (1 + \tau L)^i \cdot \frac{L + 1}{L} \|z\|_\tau.$$

Для этого заметим сначала, что

$$\|\xi_i\| = \|\xi_{i-1} + \tau f(t_{i-1}, y_{i-1}) - \tau f[t_{i-1},(\varphi_\tau)_{i-1}] + \tau z_i\| \le$$

$$\le \|\xi_{i-1}\| + \tau\|f(t_{i-1}, y_{i-1}) - f[t_{i-1},(\varphi_\tau)_{i-1}]\| + \tau\|z_i\| \le$$

$$\le \|\xi_{i-1}\| + \tau L|y_{i-1} - (\varphi_\tau)_{i-1}\| + \tau\|z\|_\tau = (1 + \tau L)\|\xi_{i-1}\| + \tau\|z\|_\tau.$$

Проведя такие оценки i раз, получим

$$\|\xi_i\| \le (1 + \tau L)\|\xi_{i-1}\| + \tau\|z\|_\tau \le$$

$$\le (1 + \tau L)[(1 + \tau L)\|\xi_{i-2}\| + \tau\|z\|_\tau] + \tau\|z\|_\tau =$$

$$= (1 + \tau L)^2\|\xi_{i-2}\| + [(1 + \tau L) + 1]\tau\|z\|_\tau \le \ldots$$

$$\ldots \le (1 + \tau L)^i\|\xi_0\| + [(1 + \tau L)^{i-1} + \ldots + (1 + \tau L) + 1]\tau\|z\|_\tau =$$

$$= (1 + \tau L)^i\|\xi_0\| + \frac{(1 + \tau L)^i - 1}{(1 + \tau L) - 1}\tau\|z\|_\tau \le$$

$$\le (1 + \tau L)^i\|z\|_\tau + \frac{(1 + \tau L)^i - 1}{L}\|z\|_\tau =$$

$$= \frac{(1 + \tau L)^i L + (1 + \tau L)^i - 1}{L}\|z\|_\tau \le (1 + \tau L)^i \cdot \frac{L + 1}{L}\|z\|_\tau.$$

Воспользуемся теперь известным неравенством $(1 + \alpha)^{1/\alpha} < e$ (напомним также, что $\tau = T/n$):

$$(1 + \tau L)^i \le (1 + \tau L)^n = (1 + \tau L)^{[(TL)/(\tau L)]} \le [(1 + \tau L)^{[1/(\tau L)]}]^{TL} < e^{TL}.$$

Окончательно,

$$\|F_\tau^{-1} z - \varphi_\tau\|_\tau = \|y - \varphi_\tau\|_\tau = \|\xi\|_\tau = \max_{0 \le i \le n} \|\xi_i\| \le$$

$$\le \max_{0 \le i \le n} e^{TL} \cdot \frac{L+1}{L} \|z\|_\tau = e^{TL} \cdot \frac{L+1}{L} \|z\|_\tau \stackrel{\text{def}}{=} C_s \| z \|_\tau .$$

Итак, явная схема Эйлера устойчива.

Покажем теперь, что из аппроксимации и устойчивости следует сходимость разностной схемы.

Теорема Лакса.

Любая устойчивая разностная схема k-го порядка аппроксимации на решении является схемой k-го порядка сходимости.

Доказательство. Действительно, если разностная схема имеет *k*-й порядок аппроксимации на решении, то $\|F_\tau(P_\tau\varphi)\|_\tau \le C_a \tau^k$ и поэтому, в частности, при малых τ мала и $\|F_\tau(P_\tau\varphi)\|_\tau$. Следовательно, в силу устойчивости, $F_\tau^{-1}[F_\tau(P_\tau\varphi)]$ существует и $\|F_\tau^{-1}[F_\tau(P_\tau\varphi)] - \varphi_\tau\|_\tau \le C_s \|F_\tau(P_\tau\varphi)\|_\tau$. Но тогда, очевидно,

$$\|P_\tau\varphi - \varphi_\tau\|_\tau = \|F_\tau^{-1}[F_\tau(P_\tau\varphi)] - \varphi_\tau\|_\tau \le$$

$$\le C_s \| F_\tau(P_\tau\varphi)\|_\tau \le C_s C_a \tau \stackrel{\text{def}}{=} C\tau^k .$$

что и требовалось доказать. Эта теорема описывает наиболее распространенный способ доказательства сходимости разностных схем.

Комментарии некоторых методов построения разностных схем.

Явная схема Эйлера может быть построена, исходя из различных соображений. Каждый из описываемых ниже приемов порождает ряд

обобщений явной схемы Эйлера и может иллюстрировать основные методы построения разностных схем. В дальнейшем эти методы будут рассматриваться подробнее.

Попытаемся, отталкиваясь от уравнения (Е), заменить его приближенным в том или ином смысле уравнением так, чтобы в результате получилась разностная схема. Первая идея выглядит так. Заменим в уравнении

$$x'(t) = f[t, x(t)] \tag{9}$$

производную $x'(t)$ в точке t_{i-1} ее приближенным значением $[x(t_i) - x(t_{i-1}]/\tau$, а правую часть — ее значением в этой точке. В результате получим *приближенное* уравнение

$$\frac{x(t_i) - x(t_{i-1})}{\tau} \approx f[t_{i-1}, x(t_{i-1})]$$

для отыскания значений *точного* решения уравнения (Е) в точках сетки G_τ. Переход к сеточным функциям и замена приближенного равенства точным приводит к *точному* уравнению для *приближенных* значений решения, а именно, к явной схеме Эйлера. Использование других аппроксимаций производной в (9) (например, $x'(t_{i-1}) \approx [x(t_{i+1}) - x(t_{i-1})]/2\tau$, а также других аппроксимаций правой части (например, $f[t_i, x(t_i)]$) взамен $f[t_{i-1}, x(t_{i-1})]$ позволяет получать другие разностные схемы.

Вторая идея основывается на замене дифференциального уравнения (9) интегральным

$$x(t + \tau) = x(t) + \int_t^{t+\tau} f[s, x(s)]ds, \tag{10}$$

Если заменить в (10) t на t_{i-1}, а интеграл — приближенной квадратурной формулой (в данном случае прямоугольников), то мы получим приближенное уравнение

$$x(t_i) \approx x(t_{i-1}) + \tau f[t_{i-1}, x(t_{i-1})],$$

которое так же, как и выше приводит к явной схеме Эйлера. Если использовать другие квадратурные формулы (заменяя, например, интеграл на $\tau f[t_i, x(t_i)]$ или $\tau[f(t_{i-1}, x(t_{i-1})) + f(t_i, x(t_i))]/2$), то будут получаться другие разностные схемы.

Третья возможность построения разностных схем связана с разложением решения в ряд Тейлора:

$$x(t_i) = x(t_{i-1}) + \tau x'(t_{i-1}) + \frac{\tau^2}{2} x''(t_{i-1}) + \dots .$$

"Обрежем" этот ряд до второго члена и выразим производную $x'(t_{i-1})$ из (9). В результате получим все то же приближенное уравнение

$$x(t_i) \approx x(t_{i-1}) + \tau f[t_{i-1}, x(t_{i-1})],$$

приводящее к явной схеме Эйлера. Удлинение отрезка ряда и другие аппроксимации коэффициентов приводят к другим разностным схемам.

Наконец, четвертая возможность связана с поиском решения в виде полинома. Допустим, мы ищем решение в виде полинома первого порядка:

$$\psi(t) = x_{i-1} + a \cdot (t - t_{i-1})$$

с неизвестным коэффициентом a. Потребуем, чтобы этот полином точно удовлетворял уравнению (9) в некоторой точке $t_{i-1} + \alpha$:

$\psi'(t_{i-1} + \alpha) = a = f(t_{i-1} + \alpha, x_{i-1} + a\alpha)$.

Переходя к сеточным функциям, как и выше, получаем разностную схему:

$x_i = \psi(t_i) = \psi(t_{i-1} + \tau) = x_{i-1} + \tau f(t_{i-1} + \alpha, x_{i-1} + a\alpha)$.

При $\alpha = 0$ это явная схема Эйлера. Если выбирать отличные от нуля α, а также брать полиномы более высоких порядков, то получается класс различных разностных схем.

17.2. Методы Рунге — Кутты

В этом разделе описываются и исследуются явные методы Рунге — Кутты, а также кратко - неявные методы Рунге — Кутты.

Задача повышения порядка сходимости.

В приложениях особенно важна задача повышения порядка сходимости разностных схем. Разностные схемы более высокого порядка позволяют уменьшить шаг сетки. Например, если при заданной точности ε схема первого порядка требует шага $O(\varepsilon)$, то схема четвертого порядка — $O(\sqrt[4]{\varepsilon})$. Практика показывает, что разностные схемы высокого порядка оказываются особенно эффективными при проведении прецизионных (с большой требуемой точностью) расчетов. Более того, схемами низких порядков такие расчеты часто вообще провести невозможно даже при сколь угодно малом (допустимом в ЭВМ) шаге.

В соответствии с теоремой Лакса для повышения порядка сходимости, вообще говоря, нужно повышать порядок аппроксимации разностной схемы. В этом и последующих параграфах в случаях, когда это необходимо, мы считаем уравнение (E) скалярным, т. е. считаем, что $m = 1$.

Схема Хойна, или предиктор-корректор.

Попытаемся повысить порядок аппроксимации явной схемы Эйлера следующим образом. В исходном уравнении (E) заменим производную обычным конечно-разностным соотношением, а $f(t, x)$ постараемся аппроксимировать "с более высоким порядком". Например, возьмем вместо $f(t_{i-1}, x_{i-1})$ "среднее направление" между векторами поля направлений уравнения (E) в точках (t_{i-1}, x_{i-1}) (вектор f_1 на рис. 3) и (t_i, x^*_i), где (t_i, x^*_i)— следующая вершина ломаной Эйлера (вектор f_2). Интуитивно ясно, что получившееся направление (вектор f_3) ближе к "истинно нужному" направлению (вектор f_4).

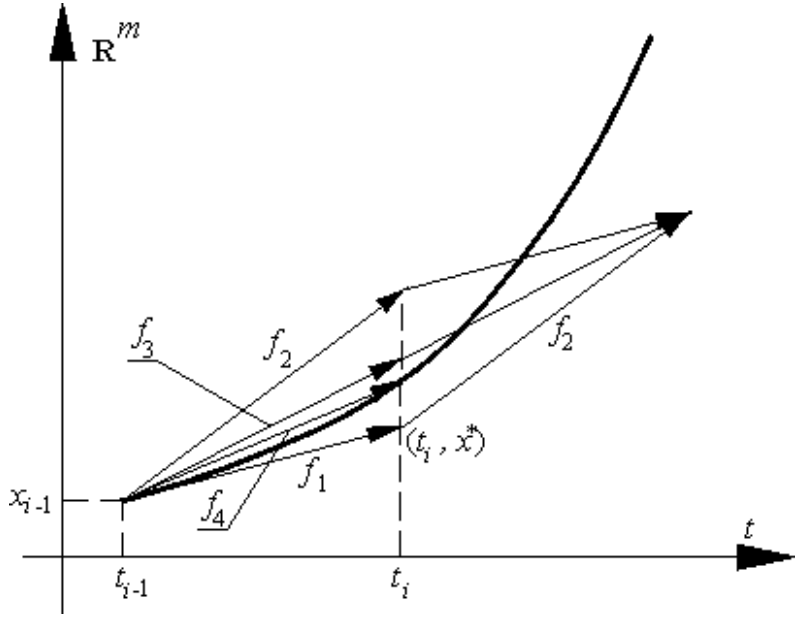


Рис. 3.

Аналитически эта идея реализуется в виде разностной схемы (S) с разностным оператором

$$[F_\tau(x)]_i = \{ x_i - x_{i-1} - \frac{\tau}{2} (f(t_{i-1}, x_{i-1}) + f[t_i, x_{i-1} + \tau f(t_{i-1}, x_{i-1})]$$

если $i = 1, ..., n$, $x_i - x_0$, если $i = 0$;

Эта схема (которую иногда называют разностной *схемой Хойна*) также является явной в том смысле, что ее решение вычисляется по рекуррентным формулам:

$(\varphi_\tau)_0 = x^0$,

$$(\varphi_\tau)_i = (\varphi_\tau)_{i-1} + \frac{\tau}{2}\left(f[t_{i-1}, (\varphi_\tau)_{i-1}] + f(t_i, (\varphi_\tau)_{i-1} + f[t_{i-1}, (\varphi_\tau)_{i-1}]) \right).$$

Разностную схему Хойна (вернее рекуррентный переход от x_{i-1} к x_i) часто записывают в виде двух полушагов:

$$x^*_i = x_{i-1} + \tau f(t_{i-1}, x_{i-1}), \qquad (1а)$$

$$x_i = x_{i-1} + \frac{\tau}{2}[f(t_{i-1}, x_{i-1}) + f(t_i, x^*_i)]. \qquad (1б)$$

Поэтому она обычно называется схемой *предиктор-корректор*: на первом полушаге (1а) приближенное решение "предсказывается" (от англ. *to predict* — предсказывать) с первым порядком точности, а на втором (1б) — "корректируется" (от *to correct* — исправлять, корректировать) с целью повышения точности.

Аппроксимация схемы предиктор-корректор.

Покажем, что *если f в (E) дважды непрерывно дифференцируема, то схема предиктор-корректор имеет второй порядок аппроксимации.*

Решение φ трижды непрерывно дифференцируемо (см. утверждение о гладкости решений) и поэтому

$$\varphi(t_i) = \varphi(t_{i-1} + \tau) = \varphi(t_{i-1}) + \tau\varphi'(t_{i-1}) + \frac{\tau^2}{2}\varphi''(t_{i-1}) + O(\tau^3).$$

Ниже мы обозначаем $\partial f(t, x)/\partial t$ через f_t, ($\partial f(t, x)/\partial x$ — через f_x, $f(t, x)$ — через f, а значения этих функций в точке $(t_i, \varphi(t_i))$ — через $f_{t,\,i}$, $f_{x,\,i}$ и f_i, соответственно. В силу уравнения (Е)

$$\varphi'(t_{i-1}) = f_{i-1},$$

$$\varphi''(t_{i-1}) = f_{t,\,i-1} + f_{x,\,i-1}\varphi'(t_{i-1}) = f_{t,\,i-1} + f_{x,\,i-1}f_{i-1}.$$

Далее, поскольку $f \in C^2$

$$f(t + \tau, x + h) = f + f_t\tau + f_x h + O(\tau^2) + O(\| h\|^2).$$

Поэтому

$$f(t_i, \varphi(t_{i-1}) + \tau f[t_{i-1}, \varphi(t_{i-1})]) =$$

$$= f_{i-1} + f_{t,\,i-1}\tau + f_{x,\,i-1}\tau f_{i-1} + O(\tau^2) + O(\| \tau f_{i-1}\|^2) =$$

$$= f_{i-1} + \tau\varphi''(t_{i-1}) + O(\tau^2) + O(\| \tau f_{i-1}\|^2).$$

Поэтому, фактически, правая часть последнего равенства равна $f_{i-1} + \tau\varphi''(t_{i-1}) + O(\tau^2)$.

Оценим теперь $\|[F_\tau(P_\tau\varphi)]_i\|$. При $i = 0$ очевидно $[F_\tau(P_\tau\varphi)]_i = 0$. Если же $i > 0$, то

$$[F_\tau(P_\tau\varphi)]_i = \frac{(P_\tau\varphi)_i - (P_\tau\varphi)_{i-1}}{\tau} - \frac{1}{2}\,(\,f[t_{i-1}, (P_\tau\varphi)_{i-1}] +$$

$$+ f(t_i, (P_\tau\varphi)_{i-1} + \tau f[t_{i-1}, \varphi(t_{i-1})]) =$$

$$= \frac{\varphi(t_{i-1}) + \tau\varphi'(t_{i-1}) + \frac{\tau^2}{2}\varphi''(t_{i-1}) + O(\tau^3) - \varphi(t_{i-1})}{\tau} -$$

$$- \frac{1}{2}[f_{i-1} + f_{i-1} + \tau\varphi''(t_{i-1}) + O(\tau^2)] = \varphi'(t_{i-1}) + \frac{\tau}{2}\varphi''(t_{i-1}) +$$

$$+ \frac{O(\tau^3)}{\tau} - \varphi'(t_{i-1}) - \frac{\tau}{2}\varphi''(t_{i-1}) + O(\tau^2) = O(\tau^2).$$

Таким образом, $\|[F_\tau(P_\tau\varphi)]_i\| \le C\tau^2$ при малых τ и при всех допустимых i.

Таким образом, схема предиктор-корректор является схемой второго порядка аппроксимации, и поэтому, если бы мы доказали ее устойчивость, то в силу теоремы Лакса он была бы схемой второго порядка точности.

Схемы Рунге — Кутты. Общие соображения.

Схема предиктор-корректор, так же, как и явная схема Эйлера, является представителем обширного семейства явных схем Рунге — Кутты. Их построение основано на следующей идее. Разложим φ в ряд Тейлора

$$\varphi(t_i) = \varphi(t_{i-1}) + \tau\varphi'(t_{i-1}) + \frac{\tau^2}{2}\varphi''(t_{i-1}) + ...,$$

оборвем ряд, взяв первые $k + 1$ его членов:

$$\varphi(t_i) \approx \varphi(t_{i-1}) + \tau\varphi'(t_{i-1}) + ... + \frac{\tau^k}{k!}\varphi^{(k)}(t_{i-1}), \qquad (2)$$

выразим все производные решения из уравнения (E) и подставим в правую часть (2):

$$\varphi(t_i) \approx \varphi(t_{i-1}) + \tau\delta(t_{i-1}, \varphi(t_{i-1}), \tau); \qquad (3)$$

здесь через δ обозначен результат такой подстановки. Например, для $k = 2$

$$\delta(t, x, \tau) = f + \frac{\tau}{2}(f_t + f_x f). \qquad (4)$$

Разложение (3) порождает очевидным образом класс разностных схем

$$x_i = x_{i-1} + \tau\delta(t_{i-1}, x_{i-1}, \tau). \qquad (5)$$

Этот класс обладает существенным недостатком — он требует вычисления высших производных функции f, что не всегда возможно или не всегда обходится дешево в вычислительном плане. Основное соображение, легшее в основу класса схем Рунге — Кутты, состоит в том, чтобы попытаться аппроксимировать функцию δ в (5) выражением, не содержащим производных функции f.

Схемы Рунге — Кутты. Пример.

Попытаемся аппроксимировать δ при $k = 2$ выражением вида

$$\Phi(t, x, \tau) = \alpha_1 f(t, x) + \alpha_2 f[t + \beta_2\tau, x + \gamma_{21}\tau f(t, x)],$$

(выбор обозначений для индексов у коэффициентов α, β, γ станет ясен чуть ниже), подбирая коэффициенты α_1, α_2, β_2, γ_{21} так, чтобы главные члены в разложениях δ и Φ по степеням τ были одинаковы. Имеем

$$\Phi(t, x, \tau) = (\alpha_1 + \alpha_2)f + \tau\alpha_2(\beta_2 f_t + \gamma_{21} f_x f) + O(\tau^2).$$

Сравнение с (4) приводит к системе уравнений на коэффициенты α_1, α_2, β_2, γ_{21}

$$\alpha_1 + \alpha_2 = 1, \quad \alpha_2\beta_2 = \frac{1}{2}, \quad \alpha_2\gamma_{21} = \frac{1}{2}.$$

Эта система имеет однопараметрическое семейство решений

$$\alpha_1 = 1 - \alpha, \quad \alpha_2 = \alpha, \quad \beta_2 = \gamma_{21} = \frac{1}{2\alpha},$$

Найденный набор решений порождает однопараметрическое семейство разностных схем

$$x_i = x_{i-1} + \tau \left((1 - \alpha) f(t_{i-1}, x_{i-1}) + \alpha f \left[t_{i-1} + \frac{\tau}{2\alpha}, x_{i-1} + \frac{\tau}{2\alpha} f(t_{i-1}, x_{i-1}) \right] \right). \quad (6)$$

Схема (6), естественно, дополняется начальным условием $x_0 = x^0$, которое мы в дальнейшем будем часто опускать в случаях, когда его наличие очевидно. При $\alpha = 1/2$ мы получаем рассмотренную выше схему предиктор-корректор. Схему (6) обычно записывают в виде

$$x_i = x_{i-1} + \tau\Phi(t_{i-1}, x_{i-1}, \tau), \quad (7)$$

где

$$\Phi(t, x, \tau) = \alpha_1 k_1 + \alpha_2 k_2 = \sum_{s=1}^{2} \alpha_s k_s, \qquad (8)$$

$$k_1 = f(t, x),\ k_2 = f[t + \beta_2\tau,\ x + \gamma_{21}\tau f(t, x)]. \qquad (9)$$

и называют *явной двухэтапной* (или *двухстадийной*) *схемой Рунге — Кутты* по числу слагаемых в представлении (8) для функции Ф (или, что то же, числу вычислений правой части уравнения).

Явные схемы Рунге — Кутты.

Общая *явная p-этапная схема Рунге — Кутты* по определению имеет вид

$$x_i = x_{i-1} + \tau\Phi(t_{i-1}, x_{i-1}, \tau),\quad x_0 = x^0, \qquad (10)$$

где

$$\Phi(t, x, \tau) = \sum_{s=1}^{p} \alpha_s k_s, \qquad (11)$$

$$k_1 = f(t, x),\quad k_s = f\left(t + \beta_s\tau,\ x + \tau \sum_{r=1}^{s-1} \gamma_{sr} k_r\right),\quad s = 2, ..., p. \qquad (12)$$

Коэффициенты α_s, β_s и γ_{sr} определяются (как и в предыдущем пункте) так, чтобы функция Ф наилучшим образом аппроксимировала функцию δ в (5). Подробнее эта процедура выглядит так. Вычисляются частные производные функции Ф порядков 0, ..., $p - 1$ по τ при $\tau = 0$ и приравниваются к производным точного решения. При этом для методов высокого порядка ($p \geq 3$) обычно предполагаются выполненными дополнительные условия вида

$$\beta_s = \sum_{r=1}^{s-1} \gamma_{sr}, \qquad (13)$$

которые сильно упрощают как решение, так и исследование системы уравнений на коэффициенты искомых схем.

Уравнения на коэффициенты явной трехэтапной схемы.

Для примера изложим план вывода уравнений на коэффициенты *явной трехэтапной схемы Рунге — Кутты*.

План вывода уравнений на коэффициенты *явной трехэтапной схемы Рунге — Кутты* при $k = 3$

$$\delta(t, x, \tau) = f + \frac{\tau}{2}(f_t + f_x f) + \frac{\tau^2}{6}(f_{tt} + 2f_{tx} f + f_{xx} f^2 + f_x f_t + f_x^2 f).$$

(здесь и ниже индексы у f обозначают соответствующие производные, например, $f_{tx} = \partial f^2(t, x)/\partial t \partial x$.

План вывода уравнений на коэффициенты *явной трехэтапной схемы Рунге — Кутты* при $p = 3$

$$\Phi(t, x, 0) = (\alpha_1 + \alpha_2 + \alpha_3)f,$$

$$\Phi'_\tau(t, x, 0) = (\alpha_2\beta_2 + \alpha_3\beta_3)f_t + (\alpha_2\gamma_{21} + \alpha_3\gamma_{31} + \alpha_3\beta_3\gamma_{32})f_x f,$$

$$\Phi''_{\tau\tau}(t, x, 0) = (\alpha_2\beta_2^2 + \alpha_3\beta_3^2)f_{tt} + 2(\alpha_2\beta_2\gamma_{21} + \alpha_3\beta_3\gamma_{31} + \alpha_3\beta_3\gamma_{32})f_{tx} f +$$

$$+ [\alpha_2\gamma_{21}^2 + \alpha_3(\gamma_{31} + \gamma_{32})^2]f_{xx} f_2 + 2\alpha_3\beta_2\gamma_{32} f_t f_x + 2\alpha_3\gamma_{21}\gamma_{32} f_t^2 f.$$

Приравнивая δ, δ'_τ,$\delta''_{\tau\tau}$ и Φ, Φ'_τ,$\Phi''_{\tau\tau}$, соответственно, в точке $(t, x, 0)$, а затем приравнивая коэффициенты при соотвествующих агрегатах из

производных функции f (например, при f_{tx} f) и учитывая соотношения (13), коэффициенты этой явной трехэтапной схемы Рунге — Кутты удовлетворяют системе уравнений

$$\begin{aligned}
&\alpha_1 + \alpha_2 + \alpha_3 = 1,\\
&\alpha_2\beta_2 + \alpha_3\beta_3 = \frac{1}{2},\\
&\alpha_2\beta_2^2 + \alpha_3\beta_3^2 = \frac{1}{6},\\
&\alpha_3\beta_2\gamma_{32} = \frac{1}{3},\\
&\gamma_{21} = \beta_2,\\
&\gamma_{31} + \gamma_{32} = \beta_3.
\end{aligned} \qquad (14)$$

Громоздкость вычислений сверхбыстро растет с ростом p. К настоящему времени придуманы изящные обозначения и приемы, позволяющие существенно упростить эти выкладки. В общем случае p-этапной схемы вопросы о построении и разрешимости системы уравнений на коэффициенты схемы и о нахождении ее решений весьма сложен.

Порядок аппроксимации явных схем Рунге — Кутты.

Система уравнений для определения коэффициентов схемы явной p-этапной схемы Рунге — Кутты в общем случае имеет семейство решений. Интересен вопрос о том, какая из соответствующих схем имеет наивысший порядок аппроксимации. (Здесь мы говорим о порядке аппроксимации, хотя имеем в виду, как это будет показано в следующем параграфе, порядок сходимости.) Полный ответ на этот вопрос не известен. Известно точное минимальное число этапов $p(k)$, необходимое для достижения порядка аппроксимации k явной схемы Рунге — Кутты для всех $k \le 7$:

$$p(k) = \begin{cases} k \text{ при } k = 1,2,3,4, \\ k+1 \text{ при } k = 5,6, \\ k+2 \text{ при } k = 7. \end{cases}$$

Максимальный достигнутый порядок аппроксимации для явных схем Рунге — Кутты равен 10. Такой порядок достигается на построенной в 1975 г. явной восемнадцатиэтапной схеме. Эта схема занесена в книгу рекордов Гиннесса. Позднее построена семнадцатиэтапная схема 10-го порядка.

Неявные методы Рунге — Кутты.

Прямым обобщением рассмотренных ниже методов являются так называемые *неявные p-этапные методы Рунге — Кутты.* Они определяются следующими формулами (ср. с (10) – (12))

$$x_i = x_{i-1} + \tau\Phi(t_{i-1}, x_{i-1}, \tau), \quad x_0 = x^0,$$

где

$$\Phi(t, x, \tau) = \sum_{s=1}^{p} \alpha_s k_s,$$

$$k_s = f\left(t + \beta_s\tau, x + \tau \sum_{r=1}^{p} \gamma_{sr} k_r \right), \quad s = 1, ..., p. \qquad (15)$$

Коэффициенты α_s, β_s и γ_{sr} находятся из тех же соображений, что и для явных методов. Простейшим представителем этого класса схем является *неявный метод Эйлера*:

$$x_i = x_{i-1} + \tau f(t_i, x_i).$$

В отличие от явных методов Рунге — Кутты формулы (15) не позволяют вычислять (m-мерные) векторы k_s "один за другим": они (формулы) представляют собой систему из p m-мерных уравнений для p m-мерных неизвестных k_1, ..., k_p, или, что то же, систему pm скалярных уравнений с pm неизвестными.

Эта система при достаточно малых τ всегда однозначно разрешима. Доказать это утверждение можно так. Положим $\mathbf{k} = (k_1, ..., k_p) \in \mathbf{R}^{mp}$ и $\mathbf{F}: \mathbf{R}^{mp} \to \mathbf{R}^{mp}$ ($= (\mathbf{R}^m)^p$) формулой

$$\mathbf{F}(\mathbf{k}) = (\mathbf{f}_1(\mathbf{k}), ..., \mathbf{f}_p(\mathbf{k})),$$

где

$$\mathbf{f}_s(\mathbf{k}) = \left(t + \beta_s \tau, x + \tau \sum_{r=1}^{p} \gamma_{sr} k_r \right), \quad s = 1, ..., p.$$

В этих обозначениях система (15) записывается в виде

$$\mathbf{k} = \mathbf{F}(\mathbf{k}). \tag{16}$$

Если теперь определить норму в $\mathbf{R}^{mp}$, например, равенством $\|\mathbf{k}\|_{mp} = \max_{1 \le s \le p} \|k_s\|$, то относительно этой нормы, как легко видеть, оператор $\mathbf{F}$ удовлетворяет условию Липшица с константой $\mathbf{l} = \tau L \cdot \max_{1 \le s \le p} \sum_{r=1}^{p} |\gamma_{sr}|$.

Поэтому при $\mathbf{l} < 1$ при достаточно малых τ, т. е. при таких τ оператор $\mathbf{F}$ является сжимающим. Но тогда в силу принципа сжимающих отображений (эквивалентное (15)) уравнение (16) имеет единственное решение, которое может быть приближенно вычислено методом простой итерации.

Необходимость решать на каждом шаге систему (15) резко увеличивает объем необходимой вычислительной работы.

Если в формулах (15) $\gamma_{sr} = 0$ при $s > r$ и хотя бы одно $\gamma_{ss} \neq 0$, то метод называется *полуявным p-этапным методом Рунге — Кутты*. В случае полуявных методов система уравнений (15) распадается на систему p (m-мерных) уравнений, которые можно решать поочередно: сначала уравнение

$$k_1 = f(t + \beta_1\tau, x + \tau\gamma_{11}k_1)$$

относительно k_1, затем уравнение

$$k_2 = f(t + \beta_2\tau, x + \tau\gamma_{21}k_1 + \tau\gamma_{22}k_2)$$

(с уже найденным k_1) относительно k_2, и т. д.

Возрастающий объем вычислительных затрат для неявных методов Рунге — Кутты частично компенсируется бóльшим в общем случае порядком сходимости. Например, доказано, что для любого $p \in \mathbf{N}_+$ существует неявный p-этапный метод Рунге — Кутты $2p$-го порядка сходимости. Кроме того, неявные методы по сравнению с явными обладают лучшими свойствами устойчивости (об этом мы будем говорить позже).

17.3. О сходимости явных методов

В этом разделе доказываются общие теоремы об условиях сходимости явных одношаговых методов. В качестве приложений рассматриваются явные методы Рунге — Кутты.

Мы часто будем использовать следующее вспомогательное утверждение.

Лемма. Пусть последовательность a_n неотрицательных чисел удовлетворяет условиям: $a_0 = 0$ и

$$a_i \leq (1 + \tau A)a_{i-1} + \tau B, \quad i = 1,2, \ldots \tag{1}$$

(A и B — неотрицательные константы). Тогда при $\tau i \le T$

$$a_i \le \frac{e^{AT} - 1}{A} B.$$

Доказательство. По индукции покажем, что при всех i

$$a_i \le \frac{(1 + \tau A)^i - 1}{A} B. \qquad (2)$$

При $i = 0$ неравенство (2) очевидно выполнено. Если же оно выполнено при некотором $i > 0$, то при том же i

$$a_{i+1} \le (1 + \tau A)a_i + \tau B \le (1 + \tau A) \frac{(1 + \tau A)^i - 1}{A} B + \tau B =$$

$$= \left(\frac{(1 + \tau A)^{i+1} - 1 - \tau A}{A} B + \tau \right) B = \frac{(1 + \tau A)^{i+1} - 1}{A} B$$

и (2) при всех i доказано.

Далее, поскольку $1 + \tau A \le 1 + \tau A + (\tau A)^2/ 2! + ... = e^{\tau A}$ и $\tau i \le T$,

$$\frac{(1 + \tau A)^i - 1}{A} B \le \frac{e^{i\tau A} - 1}{A} B \le \frac{e^{AT} - 1}{A} B,$$

что и требовалось.

Если $a > 0$, то из неравенства (1) вытекает неравенство

$$a_i \le e^{AT} a_0 + \frac{e^{AT} - 1}{A} B$$

при $i\tau \le T$.

В заключение отметим, что данная лемма является разностным аналогом классической леммы Гронуолла о дифференциальных неравенствах.

Явные одношаговые методы.

Явные методы Рунге — Кутты относятся к так называемым *явным одношаговым методам:* для того, чтобы вычислить значение сеточного решения в точке t_i необходимо знать его значение *только* в предшествующей точке t_{i-1}. Разностный оператор общего явного одношагового метода имеет вид

$$(F_\tau x)_i = \begin{cases} \dfrac{x_i - x_{i-1}}{\tau} - \Phi(t_{i-1}, x_{i-1}, \tau), & \text{если } i = 1, \ldots, n, \\ x^0, & \text{если } i = 0 \end{cases} \qquad (3)$$

(здесь Φ: $\mathbf{R}\times\mathbf{R}^m\times[0, +\infty) \to \mathbf{R}^m$). Разностную схему (S) метода мы, как обычно, будем записывать в виде рекуррентного соотношения

$$x_i = x_{i-1} + \tau\Phi(t_{i-1}, x_{i-1}, \tau), \qquad (4)$$

опуская начальное условие $x_0 = x^0$, которое мы считаем выполненным по умолчанию. Функция Φ часто называется *инкрементом*, или *приращением* метода.

Всегда будем предполагать, что *инкремент метода есть непрерывная по совокупности переменных и удовлетворяющая при*

достаточно малых τ *условию Липшица по второму аргументу с (универсальной) константой* L *функция.*

Теорема (необходимое условие сходимости явных одношаговых методов).

Если явный одношаговый метод сходится, то $\Phi(t, x, 0) \equiv f(t, x)$ *при всех* $(t, x) \in \mathbf{R}\times\mathbf{R}^m$.

Доказательство. Предположим противное: в некоторой точке $(t_0, x_0) \in \mathbf{R}\times\mathbf{R}^m$

$$\Phi(t_0, x_0, 0) \neq f(t_0, x_0).$$

В силу теоремы Коши — Пикара начальная задача

$$x' = \Phi(t, x, 0), \tag{5}$$

$$x(t_0) = x_0 \tag{6}$$

имеет единственное решение, скажем ψ (определенное на всей оси, поскольку Φ удовлетворяет условию Липшица по x). Через φ, как обычно, обозначается решение задачи (E) – (C).

Пусть теперь φ_τ — решение разностной схемы (S) с определенным формулой (3) разностным оператором, т. е. (см. (4))

$$(\varphi_\tau)_i = (\varphi_\tau)_{i-1} + \tau\Phi[t_{i-1}, (\varphi_\tau)_{i-1}, \tau]. \tag{7}$$

По условию теоремы

$$\|\varphi_\tau - P_\tau\varphi\|_\tau \to 0 \text{ при } \tau \to 0. \tag{8}$$

Если мы докажем, что

$$\|\varphi_\tau - P_\tau\psi\|_\tau \to 0 \text{ при } \tau \to 0. \qquad (9)$$

то тогда из (8) и (9) будет вытекать, что

$$\varphi = \psi. \qquad (10)$$

В самом деле, соотношения (8) и (9) означают, что функции φ и ψ равномерно аппроксимируются в узлах сетки одной и той же сеточной функцией φ_τ, а поскольку φ и ψ непрерывны на $[t_0, t_0 + T]$ и шаг сетки стремится к нулю, имеет место (10).

Но тогда, в частности, $\varphi(t_0) = \psi(t_0)$, а следовательно,

$$\varphi'(t_0) = f[t_0, \varphi(t_0)] = f(t_0, x_0) \neq \Phi(t_0, x_0, 0) = \Phi[t_0, \psi(t_0), 0] = \psi'(t_0),$$

что противоречит (10).

Осталось доказать соотношение (9). Положим

$$\rho_i = \frac{\psi(t_i) - \psi(t_{i-1})}{\tau},$$

или, что то же,

$$\psi(t_i) = \psi(t_{i-1}) + \tau\rho_i. \qquad (11)$$

По теореме о среднем (см. любой курс математического анализа)

$$\rho_i = \psi'(t_{i-1} + \theta_i\,\tau) = \Phi[t_{i-1} + \theta_i\tau, \psi(t_{i-1} + \theta_i\tau)], \qquad (12)$$

где $\theta_i \in (0, 1)$. Обозначим $(\varphi_\tau)_i - \psi(t_i)$ через ε_i. В этих обозначениях из (7) и (11) следует, что

$\varepsilon_i = \varepsilon_{i-1} + \tau(\Phi[t_{i-1}, (\varphi_\tau)_{i-1}, \tau] - \rho_i)$.

Добавляя и отнимая необходимые слагаемые и используя (12), получаем следующую оценку

$$\begin{aligned} \|\varepsilon_i\| \le \|\varepsilon_{i-1}\| &+ \tau\|\Phi[t_{i-1}, (\varphi_\tau)_{i-1}, \tau] - \Phi[t_{i-1}, \psi(t_{i-1}), \tau]\| + \\ &+ \tau\|\Phi[t_{i-1}, \psi(t_{i-1}), \tau] - \Phi[t_{i-1}, \psi(t_{i-1}), 0]\| + \\ &+ \tau\|\Phi[t_{i-1}, \psi(t_{i-1}), 0] - \Phi[t_{i-1} + \theta_i\tau, \psi(t_{i-1} + \theta_i\tau), 0]\|. \end{aligned} \qquad (13)$$

Обозначим сумму двух последних слагаемых в правой части (13) через $\tau[r(\tau)]_i$. Очевидно, $[r(\tau)]_i \to 0$ при $\tau \to 0$ равномерно по $i \in \{1, ..., n\}$, так как Φ и ψ непрерывны (и, следовательно, равномерно непрерывны на компактах).

Второе слагаемое в правой части оценивается с помощью условия Липшица величиной $\tau L\|(\varphi_\tau)_{i-1} - \psi(t_{i-1})\| = \tau L\|\varepsilon_{i-1}\|$. Тогда, продолжая (13), получаем

$\|\varepsilon_i\| \le (1 + \tau L)\|\varepsilon_{i-1}\| + \tau\|r(\tau)\|_\tau$.

По лемме $\|\varepsilon_i\| \le r(\tau)(e^{LT} - 1)/L$ и соотношение (9), а вместе с ним и теорема доказаны.

Теорема (достаточное условие сходимости явных одношаговых методов).

Условие $\Phi(t, x, 0) \equiv f(t, x)$ *при всех* $t \in \mathbf{R}\times\mathbf{R}^m$ *является и достаточным условием для сходимости метода* (4).

Д о к а з а т е л ь с т в о теоремы, по-существу, нами уже проведено: в силу теоремы Коши — Пикара решение φ задачи (E) – (C) и решение ψ задачи (5) – (6) единственны и поэтому $\varphi = \psi$, а следовательно, доказанное выше соотношение (9) равносильно нужному соотношению (8).

Сходимость схем Рунге — Кутты.

Поскольку для произвольной явной p-этапной схемы Рунге — Кутты

$$\Phi(t, x, 0) = \left(\sum_{s=1}^{p} \alpha_s\right) f(t, x),$$

в силу доказанных теорем для сходимости этих схем необходимо и достаточно, чтобы

$$\sum_{s=1}^{p} \alpha_s = 1.$$

Как уже говорилось выше, на практике гораздо более полезной оказывается информация о порядке сходимости схемы.

О порядке сходимости явных одношаговых схем.

Определим на $\mathbf{R}\times\mathbf{R}^m\times[0, +\infty)$ функцию ρ, положив для произвольных $t_0 \in \mathbf{R}$, $x_0 \in \mathbf{R}^m$ и $\tau \geq 0$

$$\rho(t_0, x_0, \tau) = \begin{cases} f(t_0, x_0), & \text{если } \tau = 0, \\ \dfrac{\varphi(t_0 + \tau) - \varphi(t_0)}{\tau}, & \text{если } \tau > 0, \end{cases}$$

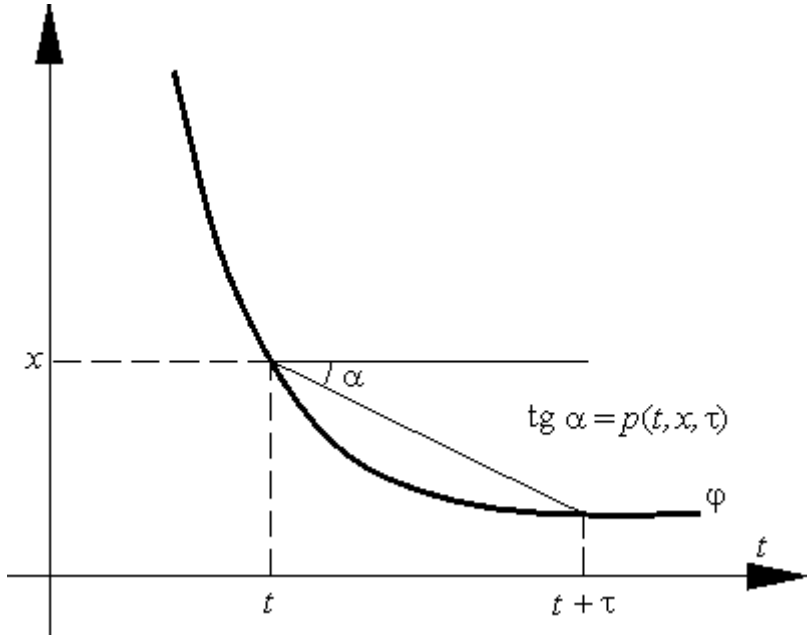


Рис. 4.

где φ — решение задачи (E) – (C) (рис. 4). Очевидно, при $\tau > 0$

$$\varphi(t_0 + \tau) = \varphi(t_0) + \tau\rho(t_0, x_0, \tau). \qquad (14)$$

Другими словами, явный одношаговый метод с инкрементом ρ является идеальным, т. е. точным на каждом шаге.

Теорема о порядке сходимости явных одношаговых схем. *Пусть для некоторых* $k > 0$ *и* $M \geq 0$

$$\|\Phi(t, x, \tau) - \rho(t, x, \tau)\| \leq M\tau^k \qquad (15)$$

при всех $(t, x) \in \mathbf{R}\times\mathbf{R}^m$ *и достаточно малых* τ. *Тогда явный одношаговый метод* (4) *сходится с k-м порядком, точнее, имеет место оценка*

$$\|\varphi_\tau - P_\tau\varphi\|_\tau \le \frac{e^{LT}-1}{L} M\tau^k. \qquad (16)$$

Суть этой теоремы другими словами можно выразить так: если инкремент метода (4) отличается от инкремента идеального метода на величину порядка $O(\tau^k)$, то метод сходится с k-м порядком.

Доказательство. Обозначим, погрешность $(\varphi_\tau)_i - \varphi(t_i)$ метода в точке t_i на решении φ через ε_i. Имеем (см. (4) и (14))

$$\|\varepsilon_i\| = \|\varepsilon_{i-1} + \tau(\Phi[t_{i-1},(\varphi_\tau)_{i-1},\tau] - \rho[t_{i-1},\varphi(t_{i-1}),\tau])\| \le$$

$$\le \|\varepsilon_{i-1}\| + \tau\|\Phi[t_{i-1},(\varphi_\tau)_{i-1},\tau] - \Phi[t_{i-1},\varphi(t_{i-1}),\tau]\| +$$

$$+\tau\|\Phi[t_{i-1},\varphi(t_{i-1}),\tau] - \rho[t_{i-1},\varphi(t_{i-1}),\tau]\| \le$$

$$\le \|\varepsilon_{i-1}\| + \tau L\|(\varphi_\tau)_{i-1} - \varphi(t_{i-1})\| + M\tau^{k+1} \le (1+\tau L)\|\varepsilon_{i-1}\| + \tau M\tau^{k+1}.$$

Применяя к последовательности $a_i = \|\varepsilon_i\|$ лемму , получаем неравенство

$$\|\varepsilon_i\| \le \frac{e^{LT}-1}{L} M\tau^k \text{ при } i\tau \le T,$$

которое, очевидно, эквивалентно нужному нам неравенству (16).

Пример. Сходимость метода предиктор-корректор. Теорема о сходимости позволяет устанавливать сходимость явных методов Рунге — Кутты и, более того, позволяет получать полезные на практике оценки близости между точным и сеточным решениями. Рассмотрим, например, метод предиктор-корректор. Для простоты будем считать уравнение (E) скалярным. Для того чтобы воспользоваться оценкой (16), нам нужны оценки константы Липшица L инкремента метода Ф по второму аргументу и константы M в неравенстве (15). Начнем с более простой оценки константы L. Для метода предиктор-корректор

$$\Phi(t, x, \tau) = \frac{1}{2}(f(t, x) + f[t + \tau, x + \tau f(t, x)]).$$

Тогда (напомним, что L — константа Липшица функции f по x)

$$\|\Phi(t, x, \tau) - \Phi(t, y, \tau)\| \le \frac{1}{2}\|f(t, x) - f(t, y)\| +$$

$$+ \frac{1}{2}\|f[t + \tau, x + \tau f(t, x)] - f[t + \tau, y + \tau f(t, y)]\| \le$$

$$\le \frac{L}{2}\|x - y\| + \frac{L}{2}\|x + \tau f(t, x) - y - \tau f(t, y)\| \le \frac{L}{2}\|x - y\| +$$

$$+ \frac{L}{2}(\|x - y\| + \tau\|f(t, x) - f(t, y)\|) \le L\left(1 + \frac{L}{2}\right)\|x - y\|.$$

Поэтому

$$\mathrm{L} \le L\left(1 + \frac{L}{2}\right).$$

Оценка константы M получается более громоздко. Пусть функция f дважды непрерывно дифференцируема и, кроме того, все частные производные от нулевого до второго порядков ограничены (скажем, константой M). Тогда, во-первых (напомним, что мы опускаем аргумент (t, x) у функции f и ее производных), раскладывая Φ в ряд Тейлора по степеням τ,

$$\Phi(t, x, \tau) = \Phi(t, x, 0) + \tau\Phi'_\tau(t, x, 0) + \frac{\tau^2}{2}\Phi''_{\tau\tau}(t, x, \theta\tau) =$$

$$= \frac{1}{2}f + \frac{1}{2}f + \frac{\tau}{2}f_t + \frac{\tau}{2}f_x f + \frac{1}{2}\cdot\frac{\tau}{2}[f_{tt}^{\theta} + 2f_{tx}^{\theta}f^{\theta} + f_{xx}^{\theta}(f^{\theta})^2] =$$

$$= f + \frac{\tau}{2}(f_t + f_x f) + \frac{\tau^2}{4}\mathbf{M}_1(\theta).$$

где верхний индекс θ означает, что аргумент, соответствующей функции равен $(t + \theta\tau,\ x + \theta\tau f(t, x))$ $(0 < \theta < 1)$, а обозначение $\mathbf{M}_1(\theta)$ очевидно. Во-вторых (здесь φ — решение уравнения (E), проходящее через точку (t, x), т. е. такое, что $\varphi(t) = x$),

$$\rho(t, x, \tau) = \frac{\varphi(t + \tau) - \varphi(t)}{\tau} =$$

$$= \frac{x + \tau\varphi'(t) + \frac{\tau^2}{2}\varphi''(t) + \frac{\tau^3}{6}\varphi'''(t + \xi\tau) - x}{\tau} =$$

$$= \varphi'(t) + \frac{\tau}{2}\varphi''(t) + \frac{\tau^2}{6}\varphi'''(t + \xi\tau) =$$

$$= f + \frac{\tau}{2}(f_t + f_x f) + \frac{\tau^2}{6}(f_{tt}^{\xi} + 2f_{tx}^{\xi}f^{\xi} + f_{xx}^{\xi}(f^{\xi})^2 + f_x^{\xi}f_t^{\xi} + (f_x^{\xi})^2 f^{\xi}) =$$

$$= f + \frac{\tau}{2}(f_t + f_x f) + \frac{\tau^2}{6}\mathbf{M}_2(\xi),$$

где $0 < \xi < 1$. Поэтому

$$\|\Phi(t, x, \tau) - \rho(t, x, \tau)\| \le \left(\left\| \frac{\mathbf{M}_1(\theta)}{4} \right\| + \left\| \frac{\mathbf{M}_2(\xi)}{6} \right\| \right) \tau^2.$$

Поскольку все частные производные функции f до второго порядка ограничены константой M,

$$\|\mathbf{M}_1(\theta)\| \le M + 2M^2 + M^3,$$

а

$$\|\mathbf{M}_2(\xi)\| \le M + 2M^2 + M^3 + M^2 + M^3 = M + 3M^2 + 2M^3.$$

Поэтому

$$\|\Phi(t, x, \tau) - \rho(t, x, \tau)\| \le \frac{5M + 12M^2 + 7M^3}{12} \tau^2,$$

и следовательно,

$$\mathrm{M} \le \frac{5M + 12M^2 + 7M^3}{12}.$$

Замечания.

а) Условие ограниченности производных функции f можно ослабить до требования их ограниченности только на некотором множестве, о котором *a priori* известно, что в нем лежит искомое решение. В частности, если это множество замкнуто и ограничено, то достаточно требовать непрерывности соответствующих производных.

б) Полученная оценка погрешности, поскольку она расчитана на широкий класс функций, разумеется, на конкретных задачах выполняется с запасом. Часто на конкретных задачах полученная фактическая погрешность в десятки раз ниже теоретической.

17.4. Анализ погрешностей

В этом разделе описываются, с одной стороны, более тонкие и, с другой — более практичные методы оценки погрешностей одношаговых методов. Кроме того, изучается возможность изменения длины шага в процессе вычислений.

Пример. Оценка (16), фигурирующая в теореме о порядке сходимости одношаговых методов, в ряде случаев оказывается весьма грубой. Рассмотрим, например явный метод Эйлера для уравнения

$$x' = \lambda x \tag{1}$$

с $\lambda = -10^3$. Упомянутая оценка содержит быстро растущий с ростом T множитель $\approx (e^{|\lambda|T} - 1)/|\lambda|$ (поскольку константа Липшица для правой части уравнения (1) равна, очевидно, $|\lambda|$ и, следовательно, $L = |\lambda|$). Поэтому для вычисления решения на больших промежутках требуется сильное уменьшение шага: $\tau = O(\varepsilon[M(e^{|\lambda|T} - 1)/|\lambda|]^{-1}) \approx O(\varepsilon e^{-\lambda T})$, где ε — требуемая точность. Например, при $\varepsilon \approx 1$ и $T \approx 1$ шаг должен быть порядка $e^{-1000} \approx e^{-400}$.

С другой стороны, в силу того, что уравнение (1) экспоненциально устойчиво, явный метод Эйлера обладает свойством уменьшения погрешности на каждом шаге. Для того чтобы понять это явление, рассмотрим простейшую ситуацию. Допустим мы ищем нулевое решение уравнения (1) и обозначим через ε_0 погрешность при выборе начального условия: $\varepsilon_0 = x_0 - 0 = x_0$. Проследим, как меняется эта погрешность в процессе счета (предполагая, для простоты, что мы ведем вычисления без ошибок округления). Имеем

$$\varepsilon_i = x_i - 0 = x_{i-1} + \lambda\tau x_{i-1} = (1 + \lambda\tau)x_{i-1} = (1 + \lambda\tau)(x_{i-1} - 0) = (1 + \lambda\tau)\varepsilon_{i-1}, \tag{2}$$

откуда получаем

$\varepsilon_i = (1 + \lambda\tau)^i \varepsilon_0$.

Так как $\lambda < 0$, при $0 < \tau \le -2/\lambda$ множитель $(1 + \lambda\tau)^i$ по модулю не превосходит единицы и, следовательно, погрешность не превосходит ε_0 независимо от длины T промежутка, на котором ведутся вычисления (более того, погрешность ε стремится к нулю при $i \to \infty$).

Суть этого явления, огрубляя ситуацию, можно описать так. При выводе оценки (2) мы считали, что погрешность с шага на шаг переносится, подчиняясь уравнению (1), а при выводе оценки (1.4.16) — подчиняясь уравнению

$\varepsilon' = L\varepsilon$,

где L — константа Липшица функции Ф (рис. 5).

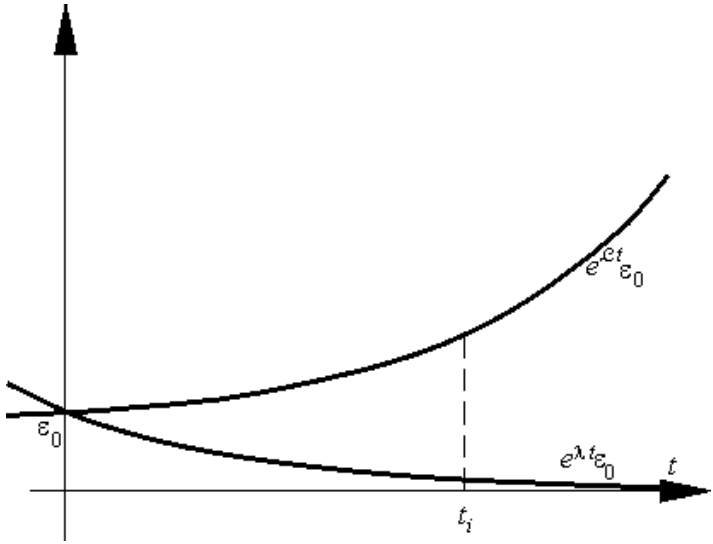


Рис. 5.

Таким образом, метод, использованный при доказательстве теоремы, не учитывает более тонких свойств уравнения (E), а именно, свойств устойчивости. Поэтому мы сейчас сначала оценим

погрешность на каждом шаге, а затем, учитывая свойства уравнения (E), оценим общую погрешность метода.

Локальная и глобальная погрешности.

Пусть $(t, x) \in \mathbf{R}\times\mathbf{R}^m$ — произвольная точка, φ — решение уравнения (E), проходящее через эту точку, а φ_τ — приближенное решение, начинающееся в точке (t, x), определяемое явным одношаговым методом

$$x_i = x_{i-1} + \tau\Phi(t_{i-1}, x_{i-1}, \tau). \qquad (3)$$

Локальной погрешностью метода (3) (в точке (t, x)) называется величина

$$\varepsilon(\tau) = \varepsilon(t, x, \tau) = \varphi(t + \tau) - (\varphi_\tau)_1$$

(рис. 6). Очевидно,

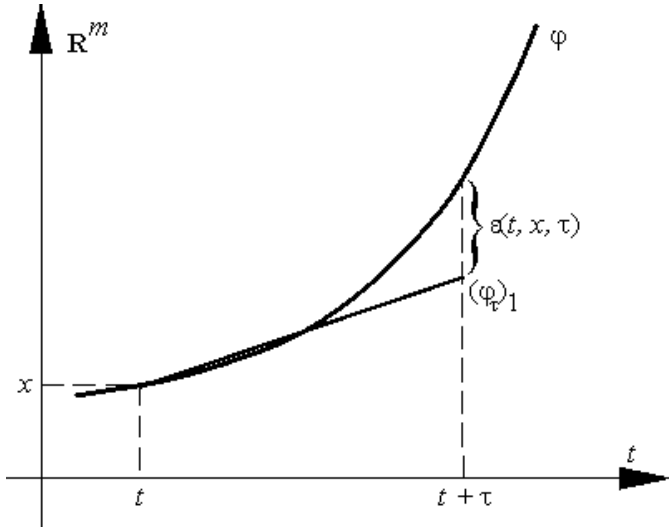


Рис. 6.

$\varepsilon(t, x, \tau) = x + \tau\rho(t, x, \tau) - x - \tau\Phi(t, x, \tau) = \tau[\rho(t, x, \tau) - \Phi(t, x, \tau)]$

(напомним, что $\rho(t, x, \tau) = [\varphi(t + \tau) - \varphi(t)]/\tau$, если $\tau > 0$). Условие (1.4.15) в теореме 1.4.6 в точности означает, что

$$\varepsilon(t, x, \tau) \le M\tau^{k+1}. \qquad (4)$$

Как следует из примера, локальная погрешность характеризует свойства аппроксимации метода.

Глобальной погрешностью метода (3) называют величину

$$E_n(\tau) = \varphi(t_n) - (\varphi_\tau)_n = \varphi(t_n) - \varphi_\tau(t_n).$$

Несколько огрубляя ситуацию, можно говорить, что в теореме утверждается, что

$$\|E_n(\tau)\| \le \frac{e^{LT} - 1)}{L} M\|\varepsilon(\tau)\|.$$

В следующем пункте мы выведем более тонкую оценку глобальной погрешности через локальные.

Перенос погрешностей.

Пусть φ — как обычно, решение задачи (E) – (C), а φ_τ — решение явной одношаговой схемы (3). Пусть, кроме того, для простоты, уравнение (E) скалярное. Обозначим через φ^i ($i = 1, ..., n$) решение

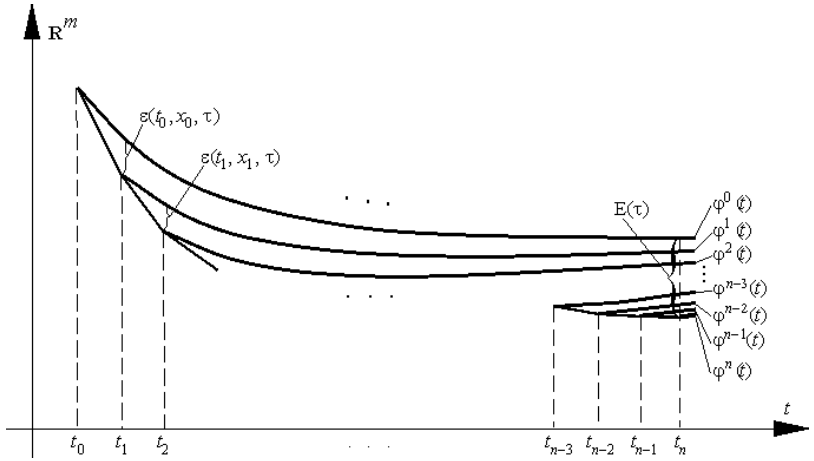


Рис. 7.

уравнения (E), удовлетворяющее начальному условию $x(t_i) = (\varphi_\tau)_i$ (рис. 7). На рисунке видно, что глобальная погрешность $E_n(\tau)$ получается в результате суммирования "перенесенных" вдоль решений уравнения (E) локальных погрешностей на каждом шаге:

$$E_n(\tau) = \varphi(t_n) - (\varphi_\tau)_n = \sum_{i=0}^{n} [\varphi^{i-1}(t_n) - \varphi^i(t_n)]. \qquad (5)$$

Оценим каждое слагаемое. Поскольку φ^i — решения уравнения (E),

$$(\varphi^{i-1})'(t) - (\varphi^i)'(t) \equiv f[t, \varphi^{i-1}(t)] - f[t, \varphi^i(t)]. \qquad (6)$$

Обозначим $\varphi^{i-1} - \varphi^i$ через ψ^i, а $(f[t, \varphi^{i-1}(t)] - f[t, \varphi^i(t)])/[\varphi^{i-1}(t) - \varphi^i(t)]$ через $a^i(t)$.

Отметим, что если $\varphi^i(t_i) - \varphi^{i-1}(t_i) \neq 0$, то $\varphi^i(t) - \varphi^{i-1}(t) \neq 0$ при всех t.

В этих обозначениях равенство (6) перепишется в виде

$(\psi^i)'(t) = a^i(t)\psi^i(t)$,

откуда

$$\psi^i(t_n) = \psi^i(t_i)\cdot\exp\left(\int_{t_i}^{t_n} a^i(s)\,ds\right),$$

или

$$\varphi^{i-1}(t_n) - \varphi^i(t_n) = \varepsilon(t_{i-1}, x_{i-1}, \tau)\cdot\exp\left(\int_{t_i}^{t_n} a^i(s)\,ds\right),$$

Поэтому, продолжая (5), получим

$$E_n(\tau) = \sum_{i=1}^{n} \varepsilon(t_{i-1}, x_{i-1}, \tau)\cdot\exp\left(\int_{t_i}^{t_n} a^i(s)\,ds\right).$$

Именно второй множитель в каждом слагаемом содержит информацию о свойствах дифференциального уравнения, позволяющую оценивать как переносятся погрешности вдоль его решений. В самом деле, по теореме Лагранжа

$a_i(s) = f_x[s, \xi^i(s)]$,

где $\xi^i(s) \in (\varphi^{i-1}(s), \varphi^i(s))$ и, таким образом,

$$E_n(\tau) = \sum_{i=1}^{n} \varepsilon(t_{i-1}, x_{i-1}, \tau)\cdot\exp\left(\int_{t_i}^{t_n} f_x[s, \xi^i(s)]ds\right). \qquad (7)$$

Например, если о функции f неизвестно ничего, кроме непрерывности и условия Липшица, то $|f_x(t, x)| \le L$ при всех (t, x). Поэтому, в случае выполнения оценки (4), оценка (7) может быть продолжена так:

$$\mathrm{E}_n(\tau) \le \sum_{i=1}^{n} \mathrm{M}\tau^{k+1} \cdot \exp\left(\int_{t_{i-1}}^{t_n} L \cdot ds\right).$$

Более тонкие свойства уравнения используются для получения оценки глобальной погрешности в следующей теореме.

Теорема об оценке глобальной погрешности для устойчивого уравнения.

Пусть правая часть (скалярного) уравнения (E) дифференцируема по x и для некоторого $\Lambda > 0$ *и всех* (t, x) *выполнено неравенство*

$$f_x(t, x) \le -\Lambda. \tag{8}$$

Пусть, кроме того, выполнена оценка (4) *для локальной погрешности. Тогда*

$$\mathrm{E}_n(\tau) \le \frac{e^{-\Lambda\tau}}{1 - e^{-\Lambda\tau}} \mathrm{M}\tau^{k+1}. \tag{9}$$

Доказательство. В силу (8)

$$\int_{t_i}^{t_n} f_x[s, \xi^i(s)]ds \le \int_{t_i}^{t_n} (-\Lambda)ds = -\Lambda\,(t_n - t_{i-1}) = -\Lambda\,(n - i + 1)\tau.$$

Поэтому, продолжая (7), с учетом (4), получаем

$$E_n(\tau) \le \sum_{i=1}^{n} M\tau^{k+1} e^{-\Lambda(n-i+1)\tau} = M\tau^{k+1} \sum_{i=1}^{n} e^{-\Lambda i\tau} <$$

$$< M\tau^{k+1} \sum_{i=1}^{\infty} (e^{-\Lambda\tau})^i = \frac{e^{-\Lambda\,\tau}}{1 - e^{-\Lambda\tau}} M\tau^{k+1},$$

что и требовалось.

Замечания.

а) Оценка (9) не ухудшается с ростом промежутка, на котором ищется решение.

б) (k+1)-й порядок малости по τ в правой части оценки (9) лишь кажущийся; на самом деле, поскольку $\lim_{\tau\to 0}[\tau/(1 - e^{-\Lambda\tau})] = -1/\Lambda$, правая часть этой оценки имеет порядок $O(\tau^k)$.

в) Условие (8) для скалярного уравнения в точности означает, что все решения уравнения (E) являются экспоненциально устойчивыми.

г) Аналогом условия (8) в многомерном случае ($m > 1$) может служить условие

$$\mathrm{Re}\,\lambda \le -\Lambda$$

при всех $\lambda \in \sigma[f_x(t, x)]$, где $\sigma(A)$ — спектр оператора A. Другими словами, все собственные значения оператора $f_x(t, x)$ при всех (t, x) лежат в полуплоскости $\{\lambda \in \mathbf{C}: \mathrm{Re}\,\lambda \le \Lambda\}$.

Главный член локальной погрешности.

Условие (4) означает (при достаточной гладкости входящих в наши рассмотрения функций), что

$$\frac{\partial^l \varepsilon(t, x, \tau)}{\partial \tau^l}\Big|_{\tau=0} = 0 \text{ при } 0 \le l \le k.$$

Поэтому, раскладывая ε по степеням τ по формуле Тейлора до (k+1)-го порядка в точке τ = 0, имеем

$$\varepsilon(t, x, \tau) = h(t, x)\tau^{k+1} + O(\tau^{k+2}). \qquad (10)$$

Величина $h(t, x)\tau^{k+1}$ называется *главным членом локальной погрешности*. Она может быть вычислена через решение и его производные, а следовательно, через производные правой части.

Можно показать, что для явного метода Эйлера

$$h(t, x) = \frac{1}{2}(f_t + f_x f). \qquad (11)$$

Задача контроля локальной погрешности.

При проведении вычислений весьма желательна информация о величине погрешности, вносимой на каждом шаге. С одной стороны, мы должны следить, чтобы погрешность не была слишком большой, чтобы в результате не получить решение с неудовлетворяющей нас точностью. С другой стороны, погрешность, следуя принципу "лучшее — враг хорошего", не должна быть слишком маленькой; иначе будет проведен большой объем излишней вычислительной работы.

Теорема 1.4.6, а для устойчивых уравнений — теорема 1.5.5, позволяет оценивать глобальную погрешность через локальную. Поэтому один из способов контроля точности вычислений заключается в контроле локальной погрешности. Если локальная погрешность излишне мала, шаг можно увеличить, если велика — уменьшить. Естественно считать, что за величину локальной погрешности отвечает ее главный член. Однако использовать для его вычисления формулы

типа формулы (11) крайне невыгодно, поскольку они требуют знания высших производных функции f (методы же Рунге — Кутты придумывались именно для того, чтобы избавиться от вычисления этих производных). Имеется ряд приемов, позволяющих обойтись без формул для вычисления главного члена локальной погрешности. Один такой прием мы описываем ниже.

Оценка локальной погрешности с помощью экстраполяции Ричардсона.

Нашей задачей является оценка выражения $h(t, x)\tau^{k+1}$ в произвольной точке (t, x). С помощью метода (3) сделаем сначала один шаг длины τ из точки (t, x), обозначив полученное значение решения через $(\varphi_\tau)_1$, а затем *из той же точки* (t, x) сделаем *два шага длины* $\tau/2$, обозначив результат через $(\varphi_{\tau/2})_2$. Обе полученные величины являются приближениями решения φ в точке $t + \tau$. Погрешность вычисления $(\varphi_\tau)_1$ есть локальная погрешность в точке (t, x) и, в соответствии с (10),

$$\varphi(t + \tau) = (\varphi_\tau)_1 + h(t, x)\tau^{k+1} + O(\tau^{k+2}).$$

Обозначим для краткости $h(t, x)$ через H:

$$\varphi(t + \tau) = (\varphi_\tau)_1 + H\tau^{k+1} + O(\tau^{k+2}). \qquad (12)$$

Погрешность же вычисления $(\varphi_{\tau/2})_2$ складывается (см. (7)) из "перенесенной" локальной погрешности $\varepsilon(t, x, \tau/2)$ и локальной погрешности метода в точке $(t + \tau/2, (\varphi_{\tau/2})_1)$:

$$\varphi(t + \tau) = (\varphi_{\tau/2})_2 + \mathrm{E}_2\left(\frac{\tau}{2}\right) = (\varphi_{\tau/2})_2 +$$

$$+ \varepsilon\left(t, x, \frac{\tau}{2}\right) \cdot \exp\left(\int_{t+\tau/2}^{t+\tau} f_x[s, \xi^1(s)]\, ds\right) + \varepsilon\left[t + \frac{\tau}{2}, (\varphi_{\tau/2})_1, \frac{\tau}{2}\right].$$

Заметим теперь, что в этой формуле (мы используем обозначение $f_x = f_x(t, x)$)

$$\varepsilon\left(t, x, \frac{\tau}{2}\right) = H\cdot\left(\frac{\tau}{2}\right)^{k+1} + O(\tau^{k+2}), \tag{13}$$

$$\exp\left(\int_{t+\tau/2}^{t+\tau} f_x[s, \xi^1(s)]\, ds\right) = \exp\left(\int_{t+\tau/2}^{t+\tau} [f_x + O(\tau)]\, ds\right) = \tag{14}$$

$$= \exp\left(\frac{\tau}{2} f_x + O(\tau^2)\right) = 1 + \frac{\tau}{2} f_x + O(\tau^2),$$

$$\varepsilon\left(t + \frac{\tau}{2}, (\varphi_{\tau/2})_1, \frac{\tau}{2}\right) = h\left(t + \frac{\tau}{2}, (\varphi_{\tau/2})_1\right)\left(\frac{\tau}{2}\right)^{k+1} + \tag{15}$$

$$+ O(\tau^{k+2}) = [H + O(\tau)]\left(\frac{\tau}{2}\right)^{k+1} + O(\tau^{k+2}).$$

Тогда, учитывая (13) – (15),

$$\varphi(t + \tau) = (\varphi_{\tau/2})_2 + E_2\left(\frac{\pi}{2}\right) = (\varphi_{\tau/2})_2 + [H\cdot\left(\frac{\pi}{2}\right)^{k+1} + O(\tau^{k+2})] \times$$

$$\times[+\frac{\pi}{2} f_x + O(\tau^2)] + [H + O(\tau)]\left(\frac{\pi}{2}\right)^{k+1} + O(\tau^{k+2}) =$$

$$= (\varphi_{\tau/2})_2 + 2H\left(\frac{\pi}{2}\right)^{k+1} + O(\tau^{k+2}). \tag{16}$$

Отбрасывая в (12) и (16) члены (k+2)-го порядка малости, можно, во-первых, "найти" значение H и, во-вторых, "найти" значение $\varphi(t + \tau)$. Кавычки здесь означают, что мы можем найти эти значения с точностью $O(\tau^{k+2})$. Для этого нужно решить систему

$$\varphi\sim = (\varphi_\tau)_1 + h\sim\tau^{k+1},$$

$$\varphi\sim = (\varphi_{\tau/2})_2 + h\sim\tau^{k+1}/2^k$$

относительно неизвестных $\varphi\sim$ и $h\sim$:

$$h\sim\tau^{k+1} = \frac{2^k}{2^k - 1}[(\varphi_{\tau/2})_2 - (\varphi_\tau)_1],$$

$$[(\varphi_{\tau/2})_2 - (\varphi_\tau)_1],$$

$$\varphi\sim = (\varphi_{\tau/2})_2 + \frac{(\varphi_{\tau/2})_2 - (\varphi_\tau)_1}{2^k - 1}.$$

Теперь величина $h\sim\tau^{k+1}$ может использоваться (с известной осторожностью) взамен главного члена погрешности $h(t, x)\tau^{k+1}$ и применяться для контроля погрешности на шаге, а величина $\varphi\sim$приближает значение $\varphi(t + \tau)$ с увеличенным на единицу порядком точности.

18. Постановка задачи оптимизации

Задачей оптимизации в теории оптимизации называется задача о нахождении экстремума (минимума или максимума) вещественной функции в некоторой области. Как правило, рассматриваются области, принадлежащие $\mathbb{R}^n$ и заданные набором равенств и неравенств.

Стандартная математическая задача оптимизации формулируется таким образом. Среди элементов χ, образующих множества *X*, найти такой элемент χ^*, который доставляет **минимальное значение** $f(\chi^*)$ **заданной функции** $f(\chi)$. Для того, чтобы корректно поставить задачу оптимизации необходимо задать:

1. *Допустимое множество* — множество $\mathbb{X} = \{\vec{x} |\ g_i(\vec{x}) \le 0,\ i = 1, \ldots, m\} \subset \mathbb{R}^n$;
2. *Целевую функцию* — отображение $f :\ \mathbb{X} \to \mathbb{R}$;
3. *Критерий поиска* (max или min).

Тогда решить задачу

$$f(x) \to \min_{\vec{x} \in \mathbb{X}}$$

означает одно из:

1. Показать, что $\mathbb{X} = \varnothing$.
2. Показать, что целевая функция $f(\vec{x})$ не ограничена снизу.
3. Найти $\vec{x}^* \in \mathbb{X} :\ f(\vec{x}^*) = \min_{\vec{x} \in \mathbb{X}} f(\vec{x})$.
4. Если $\nexists \vec{x}^*$, то найти $\inf_{\vec{x} \in \mathbb{X}} f(\vec{x})$.

Если минимизируемая функция не является выпуклой, то часто ограничиваются поиском локальных минимумов и максимумов: точек x_0 таких, что всюду в некоторой их окрестности $f(x) \ge f(x_0)$ для минимума и $f(x) \le f(x_0)$ для максимума.

Если допустимое множество $\mathbb{X} = \mathbb{R}^n$, то такая задача называется *задачей безусловной оптимизации*, в противном случае — *задачей условной оптимизации*.

Постановка задачи оптимизации. В процессе оптимизации ставится обычно задача определения наилучших, в некотором смысле, структуры или значения параметров объектов. Такая задача называется оптимизационной. Если оптимизация связана с расчетом оптимальных значений параметров при заданной структуре объекта, то она называется ***параметрической***. Задача выбора оптимальной структуры является ***структурной*** оптимизацией.

Постановка задачи параметрической оптимизации. Необходимые и достаточные условия экстремума

Говоря о задачах оптимизации выделяют несколько общих моментов

• Определяют некоторую «скалярную" (что важно для нас) меру *качества — целевую функцию* "Ф"

• Определяют набор *независимых* переменных и формулируются условия, которые характеризуют их приемлемые значения (размерность задачи и ее ограничения)

• Решение оптимизационной задачи - это приемлемый набор значений переменных, которому отвечает *оптимальное* значение целевой функции

Под *оптимальностью* (в нашем рассмотрении) обычно понимают *минимальность* целевой функции

Пусть $x \in M$ - элемент метрического пространства M и с помощью ограничений выделено множество $X \in M$.

Говорят, что целевая функция $\Phi(x)$ *имеет локальный минимум на элементе* $x^* \in loc \min\limits_{X} \Phi(x)$ *если существует некоторая конечная* ε*-окрестность точки* x^* *- шар* $K_\varepsilon(x^*)$, *такая, что*

$$\Phi(x^*) < \Phi(x), \ \forall x \in K_\varepsilon(x^*), \ 0 < \rho(x, x^*) < \varepsilon. \qquad (1)$$

В случае (1) говорят о *строгом минимуме* (в смысле неравенства), тогда как $\Phi(x^*) \le \Phi(x)$ при $\rho(x, x^*) < \varepsilon$ говорят о нестрогом минимуме.

У функции $\Phi(x)$ может быть несколько локальных минимумов - множесию $loc \min\limits_{X} \Phi(x)$. Если же в этом множестве существует точка $x^* \in loc \min\limits_{X} \Phi(x)$, в которой достигается *наименьшее* значение функции

$$\Phi(x^*) = \inf_{X} \Phi(x) \qquad (2)$$

то говорят о достижении в точке x^* *абсолютного* минимума.

Относительно целевой функции $\Phi(x)$ естественно тробовагь ее непрерывности, хотя и не всегда; а относительно множества X - *компактности* и *замкнутости* этого множества Напомним:

Множество X - компактно если из каждого его бесконечного и ограниченного подмножества можно выделить сходящуюся последовательность точек.

Множество X - замкнуто если предел любой сходящейся последовательности точек $\{x_n\}$ из X принадлежит X

В частности при $X = M$ само M должно быть *банаховым* пространством.

Задача (2) решается выбором наименьшего из соответствующих локальных минимумов.

Второе существенное ограничение — это рассмотрение задачи минимизации без ограничений, т. е. $X = M$.

1) $X = R^1$ задача минимизации функции одного переменного
2) $X = R^n$ — задача минимизации фуикциии n переменных,
3) X — гильбертово пространство и задача о минимизации функционала (скалярная целевая ф\икция)

С решением задачи (1) в предположении соответсвующей гладкости целевой функции $\Phi(x)$ связывают *необходимое* условие экстремума (Эйлера)

$$\left.\frac{\delta \Phi}{\delta x}\right|_{x^*} = 0. \qquad (3)$$

Для случая одного переменного это условие приводит к одному нелинейному уравнению

$$\Phi'(x) = 0.$$

В случае n мерной задачи мы получаем систему нелиненых уравнений

$$\frac{\partial \Phi}{\partial x_k}(x_1, ..., x_n) = 0.$$

В случае задачи минимизации функционала $\Phi(x)$ уравнение (3), как правило, дифференциальное или интегро-дифференциальное уравнение. Например, для функционала

$$\Phi(x) = \int_a^b F(t, x(t), \dot{x}(t) dt)$$

получаем

$$\delta\Phi = 0 \Leftrightarrow \begin{cases} \dfrac{d}{dt}\left(\dfrac{\partial F}{\partial x}\right) = 0 & \textit{уравнение Эйлера – Лагранжа} \\ + \textit{краевые условия} \end{cases}$$

Численное решение задачи (3) - самостоятельная проблема . Как правило здесь используются итерационные меюды, обладающие своими достоинствами и недостатками Нас же будут интересовть во второй книге базовой теории оптимизации методы безусловной минимизации (1), не связанные прямо с решением необходимого условия (3).

19. Классификация методов оптимизации

Методы оптимизации прежде всего делятся на ***параметрическую оптимизацию, структурную оптимизацию*** и ***структурно-параметрическую оптимизацию.***

Параметрическая оптимизация

Параметрическая оптимизация – это процесс определения параметров (номиналов) элементов математического объекта, при которых будут удовлетворены граничные условия. При параметрической оптимизации определяются именно оптимальные параметры элементов, так как структура должна быть задана. Оптимизация структуры производится в процессе структурной (морфологической) оптимизации, а при структурно-параметрической оптимизации определяются оптимальные и структура и параметры элементов, ее составляющих.

Если в результате параметрической оптимизации, оптимизируемый объект будет оптимальным (квазиоптимальным) по какому-либо критерию (критериям), то процесс оптимизации будет называться оптимальным (квазиоптимальным). Особый интерес представляет именно оптимальный (квазиоптимальный) параметрический процесс, и именно он используется в теории оптимизации для оптимизации объектов и систем.

Процессы параметрической оптимизация легко поддаются формализации, а, следовательно, и автоматизации.

Для автоматизации процесса параметрической оптимизации необходимы:

- математическая (компьютерная) модель;
- оптимизационный алгоритм;
- целевая функция, представляющая собой формализованное задание оптимизации параметра объекта.

При использовании целевой функции, оптимизируемый объект будет оптимальным (квазиоптимальным) по какому-либо критерию (критериям).

Исследованием и разработкой оптимизационных алгоритмов занимается базовая теория оптимизации. В рамках базовой теории оптимизации разрабатываются как алгоритмы глобального поиска, так и алгоритмы локального поиска. Алгоритмы глобального поиска позволяют найти самое наилучшее решение из возможных (глобальное оптимальное решение), в то время как алгоритмы локального поиска находят ближайший к начальной точке локальный экстремум. Обычно, на начальных этапах процесса параметрической оптимизации используются алгоритмы глобального поиска, а на завершающих – алгоритмы локальной оптимизации.

Структурная оптимизация

Структурная оптимизация – это процесс, в результате которого определяется оптимальная структура объекта таким образом, чтобы были удовлетворены условия задания на синтез объекта оптимизации. Если при этом оптимизируемый объект получается оптимальным (квазиоптимальным) по какому-либо критерию (критериям), то процесс оптимизации является оптимальным (квазиоптимальным).

Математические модели, применяемые при структурной оптимизации объектов, существенно отличаются от моделей, используемых при параметрической оптимизации. Так, если **при параметрической оптимизации структура объекта в процессе**

оптимизации остается постоянной, то в процессе структурной оптимизации изменяется его структура.

Моделями, которые удовлетворяют требованиям, предъявляемым к моделям для структурной оптимизации, являются универсальные модели.

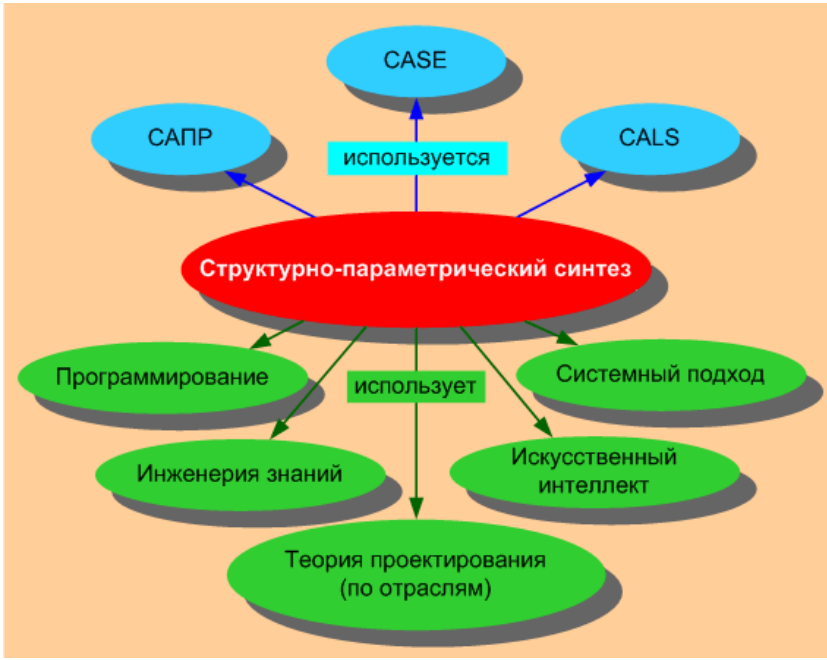
Модуль, осуществляющий структурную оптимизацию, является неотъемлемой частью систем оптимизации, точно также таким является модуль параметрической оптимизации. Так как все объекты и системы на определенном уровне рассмотрения имеют структуру, то любая задача оптимизации может быть сведена к задаче структурной оптимизации. Ввиду этого разработка общей теории структурной оптимизации, инвариантной к классу оптимизируемых объектов (технических, экономических, абстрактных), является особо актуальной.

- определение инвариантного ядра процесса структурной оптимизации, т.е. выявления того общего, что имеется при решении задач оптимизации объектов и систем любой природы;
- разработка удобных методов и средств представления теорий оптимизаций в предметных областях;
- дальнейшая разработка и совершенствование алгоритмов структурной (дискретной) оптимизации, а также методов многокритериальной оптимизации и теории принятия решений;
- разработка технологий программной реализации систем, поддерживающих процесс структурной оптимизации (модуль процесса структурной оптимизации).

Моделями процесса структурной оптимизации, отвечающими вышеперечисленным требованиям, являются **интегративные модели** и, в частности, четырехуровневые интегративные модели, на базе которых возможно построение распределенных (мультиагентных) систем автоматизации процесса структурной оптимизации.

Структурно-параметрическая оптимизация

Структурно-параметрическая оптимизация – это процесс, в результате которого определяется оптимальная структура объекта и находятся значения оптимальных параметров составляющих ее элементов, таким образом, чтобы были удовлетворены условия задания на синтез объекта оптимизации. Если при этом оптимизируемый объект получается оптимальным (квазиоптимальным) по какому-либо критерию (критериям), то процесс оптимизации является оптимальным (квазиоптимальным).



Структурно-параметрический синтез и другие дисциплины

Математические модели, применяемые при структурно-параметрической оптимизации объектов, существенно отличаются от моделей, используемых при параметрической и структурной оптимизациях. Так, если **при параметрической оптимизации структура объекта в процессе оптимизации остается постоянной,**

то в процессе структурно-параметрической оптимизации изменяются как параметры объекта, так и его структура. Сравнение характеристик моделей, применяемых при параметрической и структурно-параметрической оптимизации давно в таблице.

Параметрическая оптимизация	**Структурно-параметрическая оптимизация**
Структура модели фиксирована и не изменяется в процессе оптимизации	Структура модели заранее неизвестна и модель формируется автоматически
Поиск осуществляется в пространстве параметров, следовательно изменяются только параметры (номиналы элементов)	Поиск осуществляется в пространстве параметров и структур, следовательно изменяются как параметры элементов, так и структура
Размерность вектора параметров фиксирована	Размерность вектора параметров заранее неизвестна и может быть определена только после того как будет определена структура

Моделями, которые удовлетворяют требованиям, предъявляемым к моделям для структурно-параметрической оптимизации, являются универсальные модели.

Модуль, осуществляющий процесс структурно-параметрической оптимизации, является неотъемлемой частью систем оптимизации, точно также таким являются модули параметрической и структурной оптимизации.
Так как все объекты и системы на определенном уровне рассмотрения имеют структуру, а элементы, составляющие структуру, имеют параметры, то практически любая задача оптимизации может быть сведена к задаче структурно-параметрической оптимизации. Ввиду этого разработка общей теории структурно-параметрической оптимизации, инвариантной к классу оптимизируемых объектов (технических, экономических, абстрактных), является особо актуальной.

- определение инвариантного процесса ядра структурно-параметрической оптимизации, т.е. выявления того общего,

что имеется при решении задач оптимизации объектов и систем любой природы;

- разработка удобных методов и средств представления теорий оптимизаций в предметных областях;
- дальнейшая разработка и совершенствование алгоритмов структурной (дискретной), непрерывной (параметрической), а также дискретно-непрерывной (структурно-параметрической) оптимизации, а также методов многокритериальной оптимизации и теории принятия решений;
- разработка технологий программной реализации систем, поддерживающих процесс структурно-параметрический оптимизации (модуль структурно-параметрической оптимизации).

Моделями струкутрно-параметрической оптимизации, отвечающими вышеперечисленным требованиям, являются **интегративные модели** и, в частности, четырехуровневые интегративные модели, на базе которых возможно построение распределенных (мультиагентных) систем автоматизации процесса структурно-параметрической оптимизации.

Методы оптимизации классифицируют в соответствии с задачами оптимизации:

- Локальные методы: сходятся к какому-нибудь локальному экстремуму целевой функции. В случае унимодальной целевой функции, этот экстремум единственен, и будет глобальным максимумом/минимумом.
- Глобальные методы: имеют дело с многоэкстремальными целевыми функциями. При глобальном поиске основной задачей является выявление тенденций глобального поведения целевой функции.

Существующие методы поиска можно разбить на три группы:

1. детерминированные;
2. случайные (стохастические);
3. комбинированные.

По критерию размерности допустимого множества, методы оптимизации делят на методы *одномерной оптимизации* и методы *многомерной оптимизации*.

По виду целевой функции и допустимого множества, задачи оптимизации и методы их решения можно разделить на следующие классы:

- Задачи оптимизации, в которых целевая функция $f(\vec{x})$ и ограничения $g_i(\vec{x}),\ i = 1, \ldots, m$ являются линейными функциями, разрешаются так называемыми методами *линейного программирования*.
- В противном случае имеют дело с задачей *нелинейного программирования* и применяют соответствующие методы. В свою очередь из них выделяют две частные задачи:
  - если $f(\vec{x})$ и $g_i(\vec{x}),\ i = 1, \ldots, m$ — выпуклые функции, то такую задачу называют задачей *выпуклого программирования*;
  - если $\mathbb{X} \subset \mathbb{Z}$, то имеют дело с задачей *целочисленного (дискретного) программирования*.

По требованиям к гладкости и наличию у целевой функции частных производных, их также можно разделить на:

- методы нулевого порядка (прямые методы), требующие только вычислений целевой функции в точках приближений;
- методы первого порядка: требуют вычисления первых частных производных функции;
- методы второго порядка: требуют вычисления вторых частных производных, то есть гессиана целевой функции.

Помимо того, оптимизационные методы делятся на следующие группы:

- аналитические методы (например, метод множителей Лагранжа и условия Каруша-Куна-Таккера);
- численные методы;
- графические методы.

В зависимости от природы множества **X** задачи математического программирования классифицируются как:

- задачи дискретного программирования (или комбинаторной оптимизации) — если **X** конечно или счётно;
- задачи целочисленного программирования — если **X** является подмножеством множества целых чисел;
- задачей нелинейного программирования, если ограничения или целевая функция содержат нелинейные функции и **X** является подмножеством конечномерного векторного пространства.
- Если же все ограничения и целевая функция содержат лишь линейные функции, то это — задача линейного программирования.

Кроме того, разделами математического программирования являются параметрическое программирование, динамическое программирование и стохастическое программирование. Математическое программирование используется при решении оптимизационных задач исследования операций.

Способ нахождения экстремума полностью определяется классом задачи. Но перед тем, как получить математическую модель, нужно выполнить 4 этапа моделирования:

- Определение границ системы оптимизации
    - Отбрасываем те связи объекта оптимизации с внешним миром, которые не могут сильно повлиять на результат оптимизации, а, точнее, те, без которых решение упрощается
- Выбор управляемых переменных
    - «Замораживаем» значения некоторых переменных (неуправляемые переменные). Другие оставляем принимать любые значения из области допустимых решений (управляемые переменные)
- Определение ограничений на управляемые переменные
    - … (равенства и\или неравенства)
- Выбор числового критерия оптимизации
    - Создаём целевую функцию

На рис. 1 представлена схема классификации математических методов оптимизации.

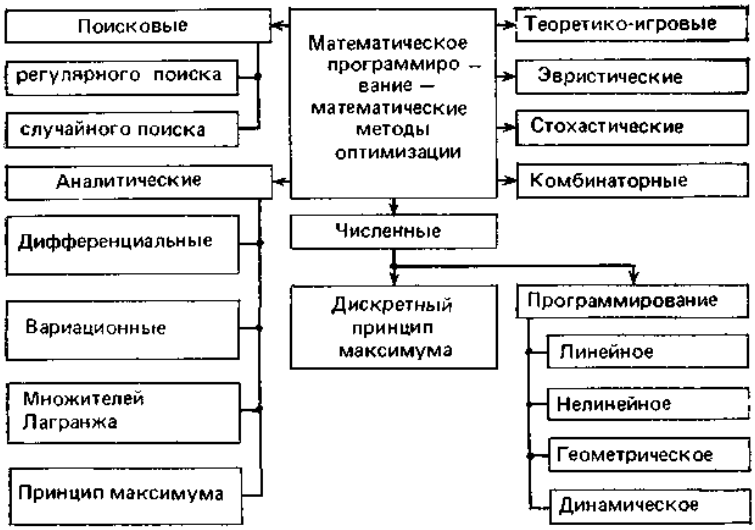


Рис.1. Классификация математических методов оптимизации

Поисковые методы применены для решения любых задач оптимизации как для унимодальных целевых функций, так и для функций многих переменных, однако весьма трудоемки. *Унимодальной* называется целевая функция, имеющая в заданном, интервале значений параметров одно экстремальное. Так, функция $f(x)$ на интервале $[0,a]$ будет унимодальной, если она строго возрастает (или убывает) при $x \le x_м$ и $x \ge x_м$, где $x_м$ — точка эстремума из интервала $[0,a]$, т. е. $0 \le x_м \le a$. На рис. 2.*а* показана одномерная целевая функция, имеющая два локальных максимума *A* и *B*. Очевидно, что всякая выпуклая функция на интервале изменения $[0,a]$ одновременно и унимодальна (см. рис. 2,*б*).

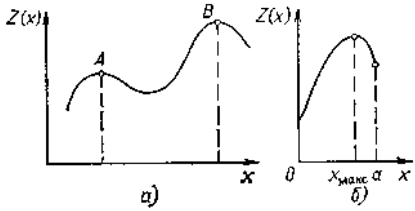


Рис. 2. Экстремумы функций

Однако обратное утверждение может быть и несправедливо. *Аналитические методы* предпочтительнее других, поскольку позволяют получить достаточно полное и общее представление об оптимизируемом объекте, наглядно установить влияние оптимизируемых параметров и ограничений на целевую функцию. Однако их применение во многих случаях ограничивается характером целевой функции и ограничений.

Численные методы являются наиболее универсальными, хорошо ориентированы на использование ЭВМ, весьма продуктивны при оптимизации многошаговых процессов. Особенно хорошо развиты методы оптимизации линейных целевых функций при линейных ограничениях.

Комбинаторные методы получили бурное развитие начиная с 60-х годов прошлого столетия в связи с решением задачи о коммивояжере. Эти методы хорошо ориентированы на использование ЭВМ.

Эвристические методы направлены на решение задач, недостаточно четко формализованных, используются также в случаях, когда применение других методов ограничивается возможностями вычислительной техники.

Теоретико-игровые методы используются в конфликтных ситуациях, т. е. для решения задач с неполной информацией, когда исход процесса зависит от действия двух или более сторон, преследующих различные цели. При этом результат действия одной из сторон зависит от образа действия других.

Стохастические методы позволяют решать условно экстремальные задачи при вероятностной информации о параметрах исследуемого процесса. Как отмечалось выше, исходная информация для процесса оптимизации часто бывает недостаточно достоверна и изменяется во времени. При этом отдельные или все оптимизируемые параметры, критерии качества и ограничения могут оказаться неопределенными и случайными.

Охарактеризуем более подробно математические методы оптимизации, которые приведенны на рис.1.

19.1. Аналитические методы оптимизации

Отыскание экстремальных значений целевой функции *методами дифференциального исчисления* возможно только в том случае, если, целевая функция дифференцируема и при этом, что особенно существенно, отсутствуют ограничения.

Если целевая функция Z представляет собой функцию многих переменных $x_1, x_2, ..., x_n$, то необходимо решить систему

$$\partial Z(x_1, x_2, \ldots, x_n)/\partial x_i = 0 \quad (i = 1, \ldots, n) \tag{1}$$

Однако это трудоемко и не обеспечивает гарантированного решения. Этим способом нельзя найти максимум, если он лежит не внутри, а на границе исследуемой области. При большом числе переменных — параметров оптимизации, задача становится практически неразрешимой. Необходимые условия применения методов дифференциального исчисления к задачам оптимизации сводятся к следующему: исследуемый процесс должен быть одпошаговым; целевая функция должна быть дифференцируемой по всем переменным, хотя бы двукратно; ограничения на решение отсутствуют.

Методы вариационного исчисления являются обобщением методов дифференциального исчисления на случай бесконечного числа переменных. Они позволяют найти экстремальное значение ***функционалов, т. е. функции, аргумент которой также функция***. Если в выражении для целевой функции $Z(x_1, x_2, ..., x_n)$ число переменных становится бесконечно большим, то это выражение можно записать как $Z[x(t)]$, где t — непрерывная переменная. В этом случае функция $x(t)$ рассматривается в качестве бесконечномерного аналога переменных $x_1, x_2, ..., x_n$.

Задание функционала $K[x(t)]$ равносильно заданию закона, по которому каждой функции $x(t)$ из некоторого класса ставится в соответствие определенное число

$$K = \int_a^b Z\,[x\,(t)]\,dt. \tag{2}$$

В функционале значение интеграла, т. е. действительное число, ставится в соответствие каждой интегрируемой функции из данного класса функций.

Простейшая задача вариационного исчисления, называемая также первой или фундаментальной задачей, состоит в нахождении экстремума функционала вида

$$K = \int_a^b F(x, \dot{x}, t)\,dt. \tag{3}$$

Для того чтобы функция $x(t)$ обращала функционал (3) в максимум или минимум, необходимо, чтобы она удовлетворяла уравнению Эйлера

$$\frac{\partial F}{\partial x} - \frac{d}{dt}\left(\frac{\partial F}{\partial \dot{x}}\right) = 0. \qquad (4)$$

Однако даже в этом простейшем случае уравнение (4) решается не всегда.

Для решения вариационных задач используются прямые и непрямые методы. Сущность последних состоит в сведении вариационной задачи к исследованию дифференциального уравнения или системы уравнений. Прямые методы заключаются в построении минимизирующей последовательности функций (кривых) $x_1, x_2, \ldots, x_n$, таких, что $\lim K(x_n) = K_э$, где $K_э$ — экстремум K; кроме того, необходимы доказательства, что у этой последовательности существует предельная кривая $x_{пред}$ и предельный переход $K[x_{пред}] = \lim\limits_{n\to\infty} K(x_n)$.

Необходимые условия применения методов вариационного исчисления к задачам оптимизации: ***наличие аналитического выражения для целевой функции; непрерывность и дифференцируемость этой функции; отсутствие ограничений.***

Метод множителей Лагранжа применим при наличии функциональных ограничений вида

$$f_i = f_i(x_1, x_2, \ldots, x_n) = 0. \qquad (5)$$

Для решения задачи составляют функцию Лагранжа

$$F = Z(\bar{x}) + \sum_{i=1}^{m} \lambda_i f_i(x), \qquad (6)$$

где λ_i — неопределенные множители Лагранжа; $\bar{x}$ — вектор с компонентами $x_1,\ldots..x_n$.

Оптимальное решение находят из системы $n+m$ уравнений. Первые m уравнений — это ограничения (5), остальные n уравнений получают приравниванием нулю частных производных

$$\partial F/\partial x_i = 0 \quad (i=1, \ldots, n). \qquad (7)$$

В результате решения системы (5), (7) вычисляют n значений $x_1, \ldots, x_n$ и m значений $\lambda_1, \ldots, \lambda_m$.

Пример. В состав аппаратуры управления (автопилота) тяжелого транспортного самолета входят: гиростабилизированная платформа (ГСП) и бортовая ЭВМ. Энергетические параметры ГСП и ЭВМ соответственно обозначим через x_1 и x_2. О данной аппаратуре известно, что ее масса в функции энергетических параметров выражается соотношением $G=x_1+2x_2$ и что критерий, характеризующий точность работы автопилота, выражается в виде $E(x_1, x_2) = 64/(x_1, x_2)$.

Необходимо найти оптимальные параметры аппаратуры управления, т. е. оптимальные значения x_1 и x_2, минимизирующие погрешности (максимизирующие точность, т. е. E), при условии, что предельно допустимая масса аппаратуры управления не превышает 16 усл. ед., т. е. $G_{макс}$=16:

$$x_1+2x_2=16. \qquad (8)$$

Решение. Составим целевую функцию

$$Z=E(x_1,\ x_2)+\lambda(G-G_{макс})=64/(x_1x_2)+\lambda(x_1+2x_2-16).$$

Продифференцируем Z по x_1 и x_2 и приравняем результаты нулю, тогда

$$-64/(x_1^2x_2)+\lambda=0;\ \ -64/(x_1x_2^2)+2\lambda=0. \qquad (9)$$

Решая совместно уравнения (8) и (9), находим искомые значения x_1=8 и x_2 = 4.

При этом $E_{макс}=64\cdot2/(4\cdot8)=4$.

Принцип максимума Понтрягина (основные результаты получены . в 1956—1961 гг. *Л. С. Понтрягиным, Б. Г. Болтянским, Р. В. Гамрекелидзе и Е. .Ф. Мищенко)* представляет собой обобщение методов вариационного исчисления и позволяет решать задачи при наличии ограничений в виде неравенств.

Необходимо найти функции $\mathbf{U}(t)\in\mathbf{U}(t_{доп})$ так, чтобы функционал принял максимальное значение:

$$Z=\int_{t_0}^{t_к} f_0[\mathbf{x}(t),\ \mathbf{U}(t)]\,dt. \qquad (10)$$

Здесь $\mathbf{U}(t)$ — вектор-функция, переводящая фазовую точку из начального положения t_0 в конечное $t_к$, $\mathbf{U}(t_{доп})$ — область допустимых значений вектор-функции; знак $\in$ включение; $\mathbf{x}(t)$ — вектор состояния (положения) системы, характеризуемый фазовыми координатами.

Связи между фазовыми координатами и управлениями описываются системой дифференциальных уравнений

$$\dot{x}_i=f_i(x,\ u)\ \ (i=1,\ \ldots,\ n). \qquad (11)$$

Согласно принципу максимума необходимым условием оптимальности $U(t)$ и траектории $x(t)$, т. е. обеспечения минимума функционала, будет существование такой ненулевой вектор-функции $\Psi(t)$, при которой для любого t, принадлежащего отрезку t_0, $t_к$, функция H достигает в фиксированной точке $U=\mathbf{U}(t)$ максимума.

Введем дополнительную переменную — текущее значение функционала

$$x_0 = \int_{t_0}^{t} f_0 [\mathbf{x}(t), \mathbf{U}(t)]\, dt. \tag{12}$$

Беря от обеих частей (12) производную и дополняя уравнением (11), запишем полную систему уравнений задачи оптимизации функции

$$\dot{x}_i = f_i[\mathbf{x}(t), \mathbf{U}(t)] \quad (i=0, 1, \ldots, n). \tag{13}$$

Для оптимизации функционала (10) составим вспомогательную функцию типа функции Гамильтона

$$H = \sum_{i=0}^{n} \Psi_i f_i. \tag{14}$$

Здесь вспомогательная вектор-функция $\Psi(\Psi_0, \ldots, \Psi_n)$ определяется следующим образом:

$$d\Psi_i/dt = -\partial H/\partial x_i, \quad (i=0, 1, \ldots, n).$$

Если удастся проинтегрировать систему, то решение получим в аналитической форме.

Рассмотрим геометрическую интерпретацию принципа максимума. Предположим, что необходимо перевести изображающую точку из начального положения $x_{нач}$ в конечное $x_{кон}$ (рис. 3) за минимальное время.

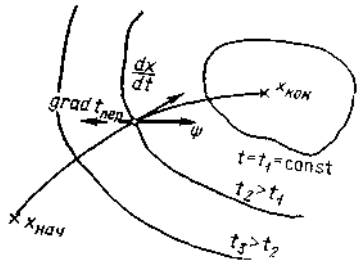


Рис. 3

Геометрическая интерпретация принципа максимума Понтрягина

Каждой точке фазового пространства, окружающего $x_{кон}$, соответствуют определенная оптимальная траектория и отвечающее ей минимальное время перехода в эту точку. Вокруг $x_{кон}$ построим поверхности, являющиеся геометрическим местом точек с одинаковым минимальным временем перехода в эту точку, т. е. *изоповерхности* или *изохроны*. Оптимальная по быстродействию траектория из $x_{нач}$ в $x_{кон}$ должна быть предельно близка нормалям к изохронам настолько, насколько это позволяют ограничения, налагаемые на координаты. Дей-

ствительно, всякое движение вдоль изохрон увеличивает время процесса без уменьшения отрезка времени, оставшегося до момента достижения $x_{кон}$. Математически условие оптимальности траектории означает, что на протяжении всей траектории скалярное произведение вектора скорости *dx/dt* на вектор, обратный градиенту времени перехода в конечную точку, должно быть максимально Обозначим это произведение через *H,* что соответствует (14), а вектор, обратный градиенту времени перехода, — через Ψ, т. е. $\Psi = -\mathrm{grad}\,(t_{пер})$. Тогда это условие можно записать так:

$$H = \Psi f = \sum_{i=1}^{n} \Psi_i f_i = \max_{u \in U},$$

где Ψ_i и f_i — соответственно координаты векторов $\Psi(\Psi_1, \Psi_2, \ldots, \Psi_n)$ и $f(\dot{x}_1, \dot{x}_2, \ldots, \dot{x}_n)$.

Таким образом, условием оптимальности является максимум проекции вектора скорости *dx/dt* на направление Ψ, что и составляет существо принципа максимума.

Принцип максимума дает лишь необходимое условие оптимальности, помогающее сузить траекторий, среди которых следует искать оптимальные.

19.2. Численные методы оптимизации

Численными методами математического программирования называют методы приближенного или точного решения задач нахождения экстремума целевой функции, основанные на построении конечной последовательности действий над конечным множеством чисел. Численные методы представляют собой последовательность однотипных шагов, или итераций.

Напомним, что ***итерацией называют регулярно повторяющийся в ходе реализации алгоритма состав процедур***. В основе итерации лежат рекуррентные соотношения, определяющие новые значения каждой переменной через прежние значения ее и других переменных одной или нескольких предшествующих итераций. Итеративное представление алгоритма особенно эффективно при реализации его на ЭВМ, поскольку реализация различных итераций осуществляется одной и той же частью программы, что уменьшает затраты на программирование и сокращает объем необходимой памяти ЭВМ.

Численные методы подразделяются на конечные и бесконечные (итеративные). Конечные методы позволяют получать решение за конечное, обычно заранее неизвестное число шагов. Таким, например,

является симплексный метод линейного программирования. В итеративных методах строится последовательность более точных приближений к решению. Процесс решения прекращается на том или ином числе итераций, когда достигается заданная точность решения.

Большинство задач оптимизации относится к нелинейным. Решение нелинейных задач — сложная вычислительная проблема, поэтому практически для решения используют приближенные методы, сущность которых состоит в том, что исходная постановка задачи сводится к одной линейной задаче или их совокупности. Основу многих процедур решения составляет линейное программирование.

Линейное программирование представляет собой численный метод решения задач нахождения экстремума линейной функции многих переменных при наличии линейных ограничений, т. е. линейных равенств или неравенств, связывающих эти переменные. Первые исследования по линейному программированию были проведены в 30-е годы прошлого века в Ленинградском университете *Л. В. Канторовичем.* Термин «линейное программирование» появился в 1951 г. в работах американских ученых *Дж. Б. Данцига и Т. Купманса.* Математически задача линейного программирования ставится следующим образом: необходимо найти неотрицательные значения переменных $x_1, x_2, ..., x_n$, минимизирующие целевую функцию

$$Z = c_1x_1 + c_2x_2 + \ldots + c_nx_n = \sum_{i=1}^{n} c_ix_i \qquad (1)$$

при наличии ограничений

$$\left.\begin{array}{l} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1; \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m. \end{array}\right\} \qquad (2)$$

Целевую функцию называют также *линейной формой,* а ограничения, представляющие систему линейно независимых уравнений, — *системой ограничений.*

На переменные x_i наложено ограничение неотрицательности, т. е. $x_i \geq 0$, однако и таких решений имеется бесчисленное множество. Задача линейного программирования состоит в выборе из этого множества одного решения, которое обращает в минимум линейную форму.

В ряде случаев все или некоторые ограничения задаются в виде неравенств

$$a_{j1}x_1 + \ldots + a_{jn}x_n \leqslant b_j, \quad \text{или} \quad a_{j1}x_1 + \ldots + a_{jn}x_n \geqslant b_j, \qquad (3)$$

которые превращаются в равенства введением дополнительной переменной $x_{n+1} \geq 0$.

При решении задач линейного программирования часто пользуются геометрическими интерпретациями.

Неравенства (3) определяют в пространстве $x_1, x_2, ..., x_n$ выпуклую область — выпуклый многогранник или многоугольник. Рассмотрим для простоты случай двух переменных: x_1 и x_2, для которых справедлива система ограничений

$$\left.\begin{aligned} a_{11}x_1 + a_{12}x_2 + b_1 \leqslant 0; \\ a_{21}x_1 + a_{22}x_2 + b_2 \leqslant 0; \\ \cdots\cdots\cdots\cdots \\ a_{m1}x_1 + a_{m2} + b_m \leqslant 0. \end{aligned}\right\} \qquad (4)$$

Каждое из этих соотношений определяет область, лежащую по одну сторону от прямой $a_{i1}x_1 + a_{i2}x_2 + b_i = 0$.

На рис. 4,*а* заштрихована выпуклая область значений x_1 и x_2.

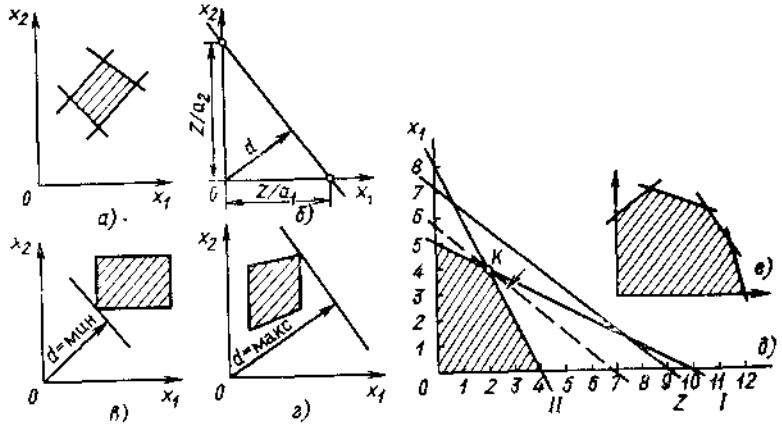


Рис. 4. Геометрическая интерпретация линейного программирования

Линейная форма в случае двух переменных запишется в виде

$$Z = a_1x_1 + a_2x_2. \qquad (5)$$

Это уравнение прямой в плоскости x_1, x_2, пересекающее оси x_1 и x_2 соответственно в точках Z/a_1 и Z/a_2 (рис. 4,*б*).

Величины a_1 и a_2 определяют угол наклона прямой α. Действительно,

$$\cos\alpha = a_1/\sqrt{a_1^2 + a_2^2}. \qquad (6)$$

Расстояние от начала координат до прямой

$$d = Z/\sqrt{a_1^2 + a_2^2}. \quad (7)$$

Геометрически задача линейного программирования интерпретируется следующим образом. Если требуется найти такие x_1 и x_2, которые придали бы линейной форме минимальное значение, то геометрически это означает, что необходимо провести прямую Z (5), проходящую хотя бы через одну точку области и имеющую минимальное расстояние d от начала координат (рис. 4,*в*). В случае нахождения максимума целевой функции это расстояние должно быть максимальным (рис. 4,*г*).

Для решения многомерных задач линейного программирования обычно используется *симплексный метод* (симплекс-метод), представляющий собой специальный способ оптимального последовательного (направленного) перебора, называемый также методом последовательного улучшения плана, так как решение задачи осуществляется итерациями, при этом на каждой последующей итерации получают план лучше полученного на предыдущей итерации. В геометрической интерпретации симплекс-метод состоит в переходе от одной вершины области допустимых значений к другой, соседней, в которой значение целевой функции лучше, чем в исходной точке. Движение происходит по периметру контура двумерной области, а для случая двух переменных — по ребрам многомерного многогранника.

Пример. Завод выпускает два вида узлов Y_1 и Y_2 системы управления, используя для этой цели два вида технологических линеек $ТЛ_1$ и $ТЛ_2$. На производство одного Y_1 на $ТЛ_1$ затрачивается 2 ч, а на $ТЛ_2$ — 1 ч; на изготовление одного Y_2 затрачивается соответственно 1 и 2 ч. Завод может использовать $ТЛ_1$ в течение 10, а $ТЛ_2$ в течение 8 ч. Прибыль от реализации одного Y_1 составляет 5, а от реализации одного Y_2 — 4 руб. Определить количество x_1 узлов Y_2 и количество x_2 узлов Y_2, которое необходимо выпустить заводу с тем, чтобы:

1) был полностью использован весь фонд времени двух технологических линеек;

2) завод получил максимальную прибыль.

Решение. Целевую функцию запишем в виде

$$Z = 5x_1 + 4x_2.$$

Ограничения имеют вид

$$2x_1 + x_2 = 10; \quad x_1 + 2x_2 = 8 \quad (8)$$

Для наглядности решим задачу геометрически. Отложим на осях координат x_1 и x_2 количество узлов (рис. 4.*д*). Здесь прямая *I* представляет производственные возможности $ТЛ_1$, а прямая *II*—$ТЛ_2$. Заштрихованная область, ограниченная прямыми *I, I I* и осями коорди-

нат, дает представление о множестве допустимых планов, т. е. множестве значений x_1 и x_2, удовлетворяющих ограничениям (8).

Целевая функция — прямая Z — передвигается параллельно самой себе по стрелке. Первая точка заштрихованной области, которой она коснется при таком перемещении, — точка *K;* она и будет решением задачи.

Таким образом, оптимальное решение всегда находится в точке пересечения граней многогранника ограничивающих условий типа уравнений (8). Если число ограничений уравнений типа (8) больше двух, то это приведет к увеличению числа граней многоугольника (рис. 4,*е*). Для случая трех видов продукции (трехмерное пространство) геометрическая интерпретация задачи затруднена и вообще невозможна для *n*-мерного пространства *(n*>3*).*

Можно показать аналитически, что экстремум в задачах линейного программирования — единственный, т. е. локальный экстремум одновременно является и глобальным, и достигается на границе области допустимых значений, как правило, в вершине.

Обычно процедура отыскания экстремума с помощью симплекс-метода оформляется в виде специальной таблицы, содержащей коэффициенты при переменных системы линейных уравнений. Это позволяет избежать громоздких преобразований системы из одной формы в другую. Тогда переход от одной системы уравнений к другой сводится к пересчету коэффициентов в таблицах, что осуществляется по формальным правилам, хорошо приспособленным для решения на ЭВМ. Данный метод нахождения оптимального решения получил название *табличного.*

Следует иметь в виду, что большинство задач оптимизации относится к нелинейным.

Однако решение нелинейных задач представляет собой сложную вычислительную проблему. Поэтому для решения используются приближенные методы. Сущность этих методов состоит в том, что исходная постановка задачи сводится к одной линейной задаче или их совокупности. Линейное программирование в этом смысле является основой для решения многих оптимизационных задач.

Многие процессы оптимизации являются многоэтапными (многошаговыми). Весьма эффективным методом оптимизации сложных многошаговых процессов является *динамическое программирование (планирование).* Этот метод был предложен и развит американским математиком *Р. Беллманом* в 60-х годах прошлого столетия. В основу метода положен интуитивно очевидный принцип, названный *принципом оптимальности,* который можно сформулировать следующим образом: ***оптимальное поведение в данный***

момент времени определяется только состоянием объекта (системы) в этот момент времени и конечным желательным состоянием и не зависит от поведения в прошлом.

При использовании этого метода можно заменить исходную сложную задачу многошаговой оптимизации последовательным решением некоторого количества существенно более простых одношаговых задач оптимизации. Основной метод динамического программирования — *метод рекуррентных соотношений,* базирующийся на принципе оптимальности.

Задача динамического программирования обычно формулируется следующим образом: из множества допустимых решений $U_{доп}$ необходимо найти такое U, которое переводит объект (систему) из начального состояния $x_0 \in x_{доп}$ в конечное $x_k \in x_{k\ доп}$ так, чтобы целевая функция (критерий качества) принимала максимальное значение

$$Z = \max_{U} [Z(U)]. \qquad (9)$$

При этом система должна находиться в допустимой области состояний $x_{доп}$. Для фазового пространства эта задача может быть сформулирована так: найти оптимальное решение U, под действием которого точка фазового пространства x переместится из начальной области в конечную, не выходя из допустимой области $x_{доп}$, так, чтобы при этом критерий K обратился в максимум (рис. 5).

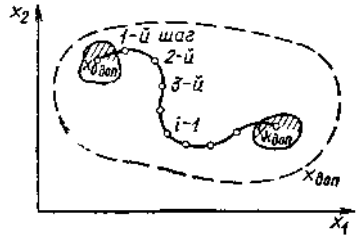


Рис. 5
Геометрическая интерпретация динамического программирования

Оптимизация решения многошагового процесса, таким образом, сводится к нахождению такой последовательности решений U_0, U_1, ... U_{n-1}, которая обеспечивает достижение максимального значения целевой функции или, что то же, минимального значения затрат (потерь, штрафов), т. е.

$$\sum_{i=0}^{n-1} Q(x_i, U_i).$$

По определению

$$f_n(x_0) = \min_{U_0} \min_{U_1} \dots \min_{U_{n-1}} [Q(x_0, U_0) + \dots + Q(x_{n-1}, U_{n-1})].$$

Первое слагаемое $Q(x_0, U_0)$ зависит только от решения U_0 на первом шаге, остальные — как от U_0, так от решений на других шагах, т. е.

$$f_n(x_0) = \min_{U_0} \left\{ Q(x_0, U_0) + \min_{U_1} \dots \min_{U_{n-1}} [Q(x_1, U_1) + \dots \right.$$
$$\left. \dots + Q(x_{n-1}, U_{n-1})] \right\}. \quad (10)$$

Обозначим

$$f_{n-1}(x_1) = \min_{U_1} \dots \min_{U_{n-1}} [Q(x_1, U_1) + \dots + Q(x_{n-1}, U_{n-1})], \quad (11)$$

тогда

$$f_n(x_0) = \min_{U_0} [Q(x_0, U_0) + f_{n-1}(x_1)]. \quad (12)$$

В общем случае для *(n*—1)-шагового процесса, начинающегося с состояния x_i, получим уравнение *Беллмана*

$$f_{n-l}(x_l) = \min_{U_l} [Q(x_l, U_l) + f_{n-(l+1)}(x_{l+1})], \quad (13)$$

представляющее собой рекуррентное соотношение, позволяющее последовательно определять оптимальное решение на каждом шаге оптимизационного процесса.

Характерной особенностью метода динамического программирования является совмещение простоты решения задач оптимизации на отдельном шаге с учетом самых отдаленных последствий этого шага. Действительно, выбор решения на каждом шаге производится не только исходя из минимизации потерь (или максимизации выигрыша) на данном шаге, т. е. минимизации величины $Q(x_i, U_i)$, но и из минимизации суммарных потерь $Q(x_l, U_l) + f_{n-(l+1)}(x_{l+1})$ на всех последующих шагах.

В уравнении Беллмана *n*—1 означает число шагов до конца процесса. Введем новые обозначения:

$$n - l = K; \quad x_l = x_{n-k} = x; \quad U_l = U_{n-k} = U.$$

Здесь *x и U* означают состояния объекта и управление за *K* шагов до конца процесса.

С учетом новых обозначений уравнение Беллмана представим в виде

$$f_k(x) = \min_{U} [Q(x, U) + f_{k-1}(x')]; \quad (14)$$

здесь x' означает то состояние, к которому переходит объект из состояния *x* при применении управления *U*.

Для расчетов на ЭВМ последнее соотношение удобнее записать в виде

$$F_k(x,\ U) = Q(x,\ U) + f_{k-1}(x'); \qquad (15)$$

$$f_k(x) = \min_U F_k(x,\ U). \qquad (16)$$

Значения $Q(x, U)$ вычисляют заранее и представляют в виде таблицы, которая хранится в памяти ЭВМ.

Метод динамического программирования применим не для всех задач, а лишь для задач с определенной структурой зависимости оптимизируемых параметров на различных шагах процесса и для целевых функций специального вида — *аддитивных функционалов от траектории.*

19.3. Поисковые методы оптимизации

Поисковые методы оптимизации являются численными методами нахождения экстремума функций одной или многих переменных; позволяют решать задачи нелинейного программирования; служат для отыскания экстремума произвольной функции Z, относительно которой отсутствует полная информация, например вид функции Z неизвестен и имеется возможность измерять или вычислять значения функции в отдельных точках.

Под задачей поиска будем понимать отыскание экстремума неизвестной функции $Z(x_1,\ x_2,\ \ldots,\ x_n)$ или значений переменных $x_{10}, x_{20}, \ldots, x_{n0}$, соответствующих оптимальному значению функции

$$Z = Z(x_{10},\ x_{20},\ \ldots,\ x_{n0}).$$

При этом аналитическая зависимость между функцией и аргументами либо неизвестна, либо сложна.

Градиентными методами оптимизации называют методы, в которых направление движения в точке экстремума функции F определяется с точностью до знака направлением градиента этой функции. При использовании этих методов строится минимизирующая последовательность $x_1, x_2, \ldots, x_m \ldots$ по формуле

$$x_{m+1} = x_m - a\ \mathrm{grad}\ (x_m), \qquad (1)$$

где grad $(x) \in N$ представляет собой градиент функции $Z(x)$ в точке x. Задачи нелинейного программирования, решаемые поисковыми методами, в зависимости от числа аргументов делят на одномерные и многомерные. К одномерным методам относятся такие изящные и эффективные методы, как дихотомии, «золотого сечения», Фибоначчи; к многомерным методам — градиентные, позволяющие создавать

процедуры поиска, универсальные для одномерных и многомерных функций.

Задача оптимизации градиентным методом может быть сформулирована следующим образом: ***пусть оптимизируемая система характеризуется некоторой зависимостью между выходной величиной и входными переменными***

$$Z = Z(x_1,\ x_2,\ \ldots,\ x_n),$$

при этом вид функции Z может быть заранее неизвестен. Требуется найти совокупность переменных x_i, для которой величина функции Z минимальная, а допустимые значения переменных x_i удовлетворяют системе ограничений-неравенств

$$L_i(x_1,\ x_2,\ \ldots,\ x_n) \leqslant 0.$$

Градиентные методы являются большей частью итеративными, т. е. строится последовательность приближений x_0, x_i ..., сходящаяся к точке минимального значения функции *Z*. Этим методом вычисляют частные производные целевой функции по всем переменным, направление градиентов и шаги по этому направлению.

Метод наискорейшего подъема (спуска) является развитием градиентного метода и отличается тем, что градиент вычисляется не в каждой точке. Шаги в направлении градиента продолжаются до тех пор, пока функция увеличивается (уменьшается), когда же это увеличение (уменьшение) прекращается, то вычисляют градиент в соответствующей точке, и процедура повторяется.

Метод сканирования, или слепого поиска, заключается в последовательном переборе всех возможных значений аргументов и фиксирования наибольшего значения функции. Метод позволяет отыскивать глобальный экстремум независимо от вида функции, однако требует больших вычислительных затрат.

Метод покоординатного подъема (спуска), или *метод Гаусса — Зайделя,* состоит в последовательной, поочередной оптимизации функции по каждой из переменных. Характеризуется простыми алгоритмами.

Следует отметить, что практический интерес к поисковым методам появился в связи с развитием вычислительной техники.

19.4. Оптимизация в конфликтных ситуациях

При решении задач оптимизации возникают ситуации, характеризуемые противоположностью интересов двух (или более) сторон, и тогда результат действия одной из сторон зависит от образа действия других. Такие ситуации называются *конфликтными* и

изучаются *теорией игр,* позволяющей формализовать и анализировать количественно конфликтные ситуации и дающей рекомендации о наилучшем поведении объекта в таких ситуациях.

Остановимся на некоторых определениях.

***Игра* — это упрощенная формализованная модель реальной конфликтной ситуации.**

Формализация конфликтных ситуаций заключается в том, что действия сторон подчинены определенным правилам, которые называют правилами игры.

***Правила игры* предопределяют возможные варианты действия, или стратегии, сторон.**

Рассмотрим *парные игры,* т. е. игры двух сторон (игроков) *А* и *В.* Результат игры, т. е. выигрыш или проигрыш, характеризуется числом (ценой игры). *Джон фон Нейман* исследовал игры с «нулевой суммой», когда выигрыш стороны *А* есть проигрыш стороны *В,* т. е. результат от реализации принятого решения распределяется между сторонами. Или, точнее, алгебраическая сумма выигрышей сторон равна нулю. Поскольку в игре с нулевой суммой интересы сторон прямо противоположны, достаточно рассматривать выигрыш одной стороны. Согласно *основной теореме теории игр* игрок *В* может выиграть в среднем сумму, равную *К* за одну игру, а игрок *А* может ему помешать выиграть большую сумму. Утверждается также, что для игрока *В* существует оптимальная стратегия, обеспечивающая выигрыш этой суммы, а при применении оптимальной стратегии игроком *А* он может проиграть не более чем *К.* Это так называемый *принцип (теорема) минимакса,* который запишем в виде

$$\max_a [\min_b V(a, b)] \leqslant \min_b [\max_a V(a, b)], \qquad (1)$$

где *a* и *b* — соответственно характеристики действия игроков *А* и *В; V*(*a, b)* — функция потерь, или платежная функция.

В случае если имеет место равенство

$$\max_a \left[\min_b V(a, b)\right] = \min_b \left[\max_a V(a, b)\right],$$

то соответствующее значениефункции *V*(*a, b)* называется *седловой точкой игры,* которая является точкой пересечения оптимальных стратегий игроков *А* и *В.*

В этой точке минимум максимума потерь одного игрока совпадает с максимумом минимума потерь другого. На рис. 6 представлена поверхность, имеющая седловидную форму, где точка *К* — седловая точка игры.

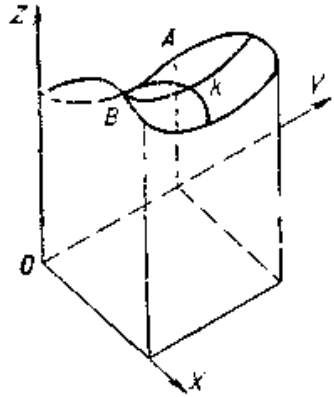


Рис.6. Геометрическая интерпретация теоремы минимакса.

Например, при проектировании системы управления летательного аппарата игроком A можно считать конструктора. Его цель заключается том, чтобы путем выбора соответствующего алгоритма управления получить наибольший эффект, например, минимизировать средний квадрат ошибки управления. Природа — игрок B в наименее благоприятном случае имеет прямо противоположную цель — максимизировать средний квадрат ошибки управления, для чего располагает выбором характеристик входных воздействий. В нашем примере стратегия игрока A —выбор оптимального алгоритма управления, а стратегия игрока B — выбор реализации входного сигнала. Ограничимся рассмотрением конечной игры, т. е. такой игры, в которой игроки A и B располагают только конечным числом стратегий (в отличие от бесконечных игр). Игрок A располагает стратегиями $A_1, A_2, ..., A_n$. Игрок B располагает стратегиями $B_1, B_2.....B_m$. Это так называемая игра $m \times n$.

Выигрыш игрока A при стратегиях A_i и B_j обозначим через a_{ij}.

В общем случае выигрыш является случайной величиной, т. е. a_{ij} обозначает средний выигрыш. Значения a_{ij} образуют *матрицу,* иначе называемую *матрицей игры* или *эффективности.* Назовем меру выигрыша a_{ij} показателем эффективности варианта стратегии A_i в условиях B_j. Матрица эффективности в этом случае аналогична табл. 1, в которой вместо варианта решения записывают вариант стратегии и заменяют W_{ij} на a_{ij}. Решить игру $m \times n$, значит найти для каждого игрока такую стратегию, чтобы его средний выигрыш за большое число игр был наибольшим.

Таблица 1

	1	2	...	i
1	W_{11}	W_{12}	...	W_{1i}
2			...	
. . .	. . .	. . .		. . .
j	W_{j1}	W_{j}	...	W_{ji}

Теория игр рекомендует каждому игроку выбирать такую стратегию, при которой получается максимально возможный выигрыш при наименее благоприятном действии противника. Такую стратегию называют *стратегией минимакса.* Оптимальная стратегия игрока F определяется из матрицы эффективности путем отыскания такого ее элемента, который удовлетворяет условию

$$a_{\text{опт } A} = \max_i \min_j a_{ij} \tag{2}$$

иными словами, такой элемент выбран как максимальный по строкам i из минимальных в каждой строке по столбцам j. Оптимальная стратегия игрока B определяется по элементу

$$a_{\text{опт } B} = \min_j \max_i a_{ij}, \tag{3}$$

иными словами, по такому элементу матрицы, который является минимальным по столбцам j из максимальных по строкам i каждого столбца.

Пример. Предприятие F приобретает гироскопические приборы для систем управления, потребность в которых зависит от спроса B. Потребность в гироскопических приборах при пониженном спросе может составить 100, при нормальном спросе — 150 и при повышенном спросе — 200 шт. Цена гироскопических приборов при пониженном спросе составляет 1000, при нормальном спросе — 1500 и при повышенном спросе — 2000 руб.

Какую стратегию следует выбрать предприятию F, а именно: закупить 100, 150 или 200 гироскопических приборов, чтобы получить наибольшую прибыль?

Решение. Составим матрицу затрат (табл. 2) и вычислим девять возможных сочетаний стратегий игроков.

Таблица 2

Запас гироскопических приборов A	Элементы a при разном спросе B		
	пониженный	нормальный	повышенный
100	$a_{11} = -100$	$a_{12} = -175$	$a_{13} = -300$
150	$a_{21} = -150$	$a_{22} = -150$	$a_{23} = -250$
200	$a_{31} = -200$	$a_{32} = -200$	$a_{33} = -200$

Элемент a_{ij} соответствует пониженному спросу и минимальному запасу 100·1000= —100 000 руб. Знак минус означает затраты. Таким образом, a_{11}= —100 тыс. руб.

Элемент a_{12} соответствует приобретению 100 шт. гироскопических приборов по 1000 руб. и еще 50 шт. по 1500 руб., т. е. a_{12}=100·100+50·1500= —175 тыс. руб.

Элемент a_{13}= 100·1000+100-2000= —300 тыс. руб.

Элемент $a_{21} = a_{22}$= 150·1000= —150 тыс. руб.

Элемент a_{23}= 150·1000+50·2000= —250 тыс. руб.

Элемент $a_{31} = a_{32} = a_{33}$=200·1000= —200 тыс. руб.

Из таблицы следует, что минимумы строк составляют —300, —250 и —200 тыс. руб., а максимумы столбцов —100, —150 и —200 тыс. руб. и максимальный минимум строк (по столбцам) совпадает с минимальным максимумом столбцов по строкам. Оба они равны a_{33}. Таким образом, оптимальным решением является приобретение 200 гироскопических приборов по цене 1000 руб. за штуку. В нашем примере

$$a_{\text{опт}A} = a_{\text{опт}B} = a_{33} = a_{AB},$$

что свидетельствует о наличии в матрице седловой точки, являющейся одновременно максимином для игрока A и минимаксом для игрока B. Подобное совпадение не всегда имеет место, и матрица в общем случае может не содержать седловой точки a_{AB}.

19.5. Комбинаторные методы оптимизации

Комбинаторика как ветвь математики возникла в XVII в., хотя первые упоминания о вопросах, близких к комбинаторным, встречаются в китайских рукописях, относящихся к XII—XIII вв. до н. э. Развитие комбинаторных методов связано с именами *Паскаля,*

Ферма, Бернулли, Лейбница и Эйлера. Комбинаторные методы являются основными при решении ряда важных задач оптимизации. Так, решения задач целочисленного программирования основаны на той или иной идее направленного перебора вариантов, в результате которого сокращается число допустимых решений, отыскивается оптимальное решение, когда исключаются подмножества вариантов, не содержащих оптимум. Основное содержание комбинаторных методов составляет совокупность способов решения, объединенных общим термином *метод ветвей и границ.*

Общая идея метода достаточно проста. Множество допустимых планов разбивается на подмножество. В свою очередь каждое подмножество снова разбивается на подмножества до тех пор, пока каждое подмножество не обратится в точку многомерного пространства. В силу конечности наборов значений переменных дерево подмножества (схема ветвлений) конечно. Построение схемы ветвления есть не что иное, как формирование процедуры перебора. Перебор может осуществляться различными способами. Возможность оценки образуемых подмножеств по наибольшему (наименьшему) значению позволяет сократить перебор, поскольку одно из подмножеств при выполнении определенных соотношений исключается и в дальнейшем не анализируется. Таким образом, специфика комбинаторных методов состоит в применении двух видов операций: *выборки,* состоящей в отборе подмножеств, и *упорядочения.*

Для этого используются методы: непосредственного подсчета числа выборок, производящих функций, логические, экстремальные, геометрические и др.

Методы непосредственного подсчета числа выборки составляют содержание элементарной комбинаторики. Этими методами находят числа r выборок, получаемых из n элементов соответствующего множества.

Метод производящих функций сформировался в работах *Эйлера и Лапласа.* Этот метод позволяет оперировать не с отдельными комбинаторными объектами, а с их классами, что дает определенные практические преимущества.

Логические методы состоят в попеременном отбрасывании и возвращении подмножеств, обладающих определенными свойствами или весами. Основную формулу, выражающую сущность метода, можно записать в виде

$$n(\bar{P}_1, \bar{P}_2, \ldots, \bar{P}_N) = n - \sum_{i=1}^{N} n(P_i) +$$

$$+ \sum_{i,j} n(P_i, P_j) + \ldots + (-1)^N n(P_1, P_2, \ldots, P_N).$$

Здесь n — множество элементов; N — множество их свойств $P_1, P_2, \ldots, P_N$, которыми каждый элемент n множества обладает в некоторой комбинации, P_k — отсутствие свойства P_k.

Экстремальные методы включают методы локальной оптимизации, случайного поиска, ветвей и границ.

Геометрические методы основаны на геометрической интерпретации комбинаторных ситуаиий с помощью множества точек, отрезков и др. Простейшими геометрическими комбинаторными системами являются конечные плоскости, т. е. системы инцидентности двух конечных множеств (точек и линий), подчиненных системе аксиом проективной геометрии.

Следует иметь в виду, что термин *комбинаторные методы* не является вполне точным, так как комбинаторные методы применяются в основном для решения задач математического программирования, которые не попадают в сферу приложения классических методов. Такими задачами являются многие нелинейные, в частности многоэкстремальные, задачи. К комбинаторным методам тесно примыкают эвристические, использующие интуитивные соображения и основывающиеся на индивидуальных особенностях конкретной задачи.

Остановимся подробней на *методе ветвей и границ*. Этот метод относится к группе комбинаторных методов дискретного программирования и впервые был предложен в 1961 г. в работе *Лэнга и Дойга* для решения задач целочисленного программирования. Второе рождение метод получил в 1963 г в связи с решением задачи о коммивояжере.

Основная идея метода ветвей и границ состоит в разбиении всего множества допустимых решений задачи на некоторые подмножества, внутри которых осуществляется упорядоченный просмотр решений с целью оптимального выбора Для всех решений, входящих в выделенные подмножества, вычисляется нижняя граница G_H минимального значения целевой функции. Как только G_H становится больше Z, для наилучшего из ранее известных решений подмножество решений, соответствующих этой границе, исключается из исходной области решений. Это обеспечивает сокращение перебора. Процесс поиска оптимума сопровождается разбиением поля решений и вычислением G_H до тех пор, пока не будут исключены все решения,

кроме оптимального. Для описания процесса поиска оптимального решения строится дерево решений (рис. 7,*а*).

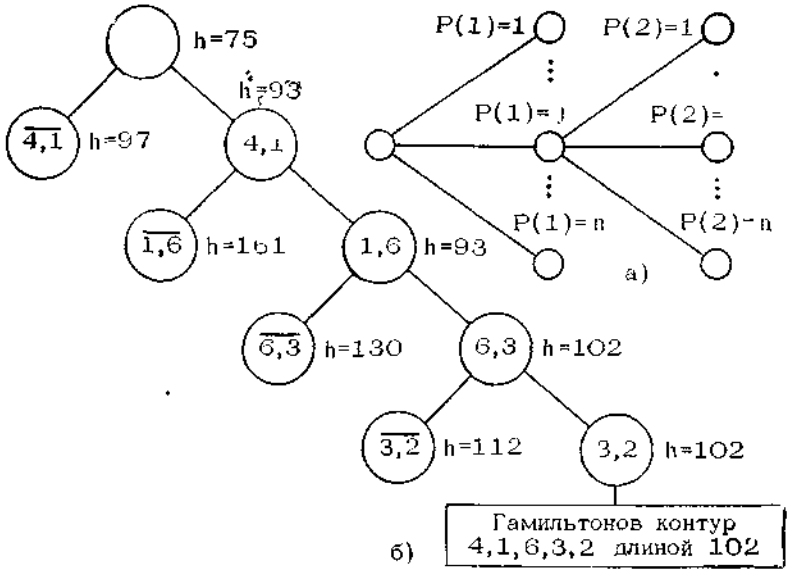


Рис. 7. Дерево решений

Идея алгоритма состоит в том, что вначале оценивают длину маршрута снизу для множества всех гамильтоновых контуров (маршрутов), т. е. контуров (путей), проходящих через все вершины графа. После этого множество всех гамильтоновых контуров разбивают на два подмножества: первое состоит из гамильтоновых контуров, включающих некоторую дугу (*i, j*). Обозначим ее через $[i, j]$. Второе состоит из контуров, не включающих эту дугу. Обозначим ее через $\overline{[i, j]}$. Для каждого из подмножеств определяют нижнюю границу оценок — НГО. Процесс разбиения подмножеств сопровождается построением некоторого дерева (рис. 7,*б*). Как следует из рисунка, ветви *h*, для которых НГО окажется больше длины полученного маршрута, исключают из рассмотрения.

Основное достоинство метода состоит в указании способа вычисления НГО и указании дуги (*i, j*), включение или невключение которой в маршрут разбивает множество гамильтоновых путей на подмножества. Для определения НГО можно использовать задачу назначения, однако это связано с громоздкими вычислениями. Другой

способ использует матрицу расстояний, он прост в вычислениях, однако дает менее точную оиенку.

Осуществим с этой целью так называемое приведение матрицы, для чего воспользуемся теоремой, которая формулируется следующим образом: *если из какой-нибудь строки или столбца матрицы вычесть произвольное положительное число, то решение задачи о коммивояжере с этой неизменной матрицей расстояний совпадает с прежним решением, а длина маршрута изменится на это же самое число.*

Для приведения в каждой строке (или столбце) матрицы находят минимальный элемент и вычитают его из всех элементов этой строки (столбца). Полученную матрицу называют приведенной по строкам (столбцам). Матрица содержит по крайней мере один нуль в каждой строке. Длина оптимального маршрута в задаче с неприведенной матрицей

$$L=L_n+h, \qquad (1)$$

где L_n — длина оптимального маршрута в задаче с приведенной матрицей ($L_n \geq 0$, поскольку в приведенной матрице все элементы неотрицательны); h — константа приведения.

Здесь уместно сформулировать правило, являющееся целью наших рассуждений: сумма констант приведения может служить НГО длины гамильтонова маршрута.

Пример. Задана матрица расстояний I (табл. 3).

Таблица 3

I

	1	2	3	4	5	6	h_i ↓
1	∞	8	6	2	4	1	1
2	8	∞	3	2	6	5	2
3	6	3	∞	1	3	4	1
4	2	2	1	∞	5	6	1
5	4	6	3	5	∞	2	2
6	1	5	4	6	2	∞	1

II

	1	2	3	4	5	6
1	∞	7	5	1	3	0
2	6	∞	1	0	4	3
3	5	2	∞	0	2	3
4	1	1	0	∞	4	5
5	2	4	1	3	∞	0
6	0	4	3	5	1	∞
h_j →	1				1	

III

	1	2	3	4	5	6
1	∞	6	5	1	2	0
2	6	∞	1	0	3	3
3	5	1	∞	0	1	3
4	1	0	0	∞	3	5
5	2	3	1	3	∞	0
6	0	3	3	5	0	∞

Произведем приведение матрицы по строкам и столбцам с целью получения НГО гамильтонова маршрута.

Решение. Справа от матрицы I проставлены столбцом константы приведения по строкам. Результаты приведения по столбцам даны в матрице II. Снизу этой матрицы против соответствующих столбцов проставлены константы приведения по столбцам. Результаты приведения по столбцам даются в матрице III. Таким образом, матрица III приведена как по строкам, так и по столбцам.

Сумма констант приведения составляет

$$\Sigma h = \underbrace{(1+2+1+1+2+1)}_{\text{по строкам}} + \underbrace{(1+1)}_{\text{по столбцам}} = 10.$$

Эта сумма не что иное, как нижняя граница оценки гамильтонова маршрута.

Известные алгоритмы, реализующие метод ветвей и границ, осуществляют последовательный анализ ветвей дерева решений с отсечением ветвей, содержащих заведомо неоптимальные решения.

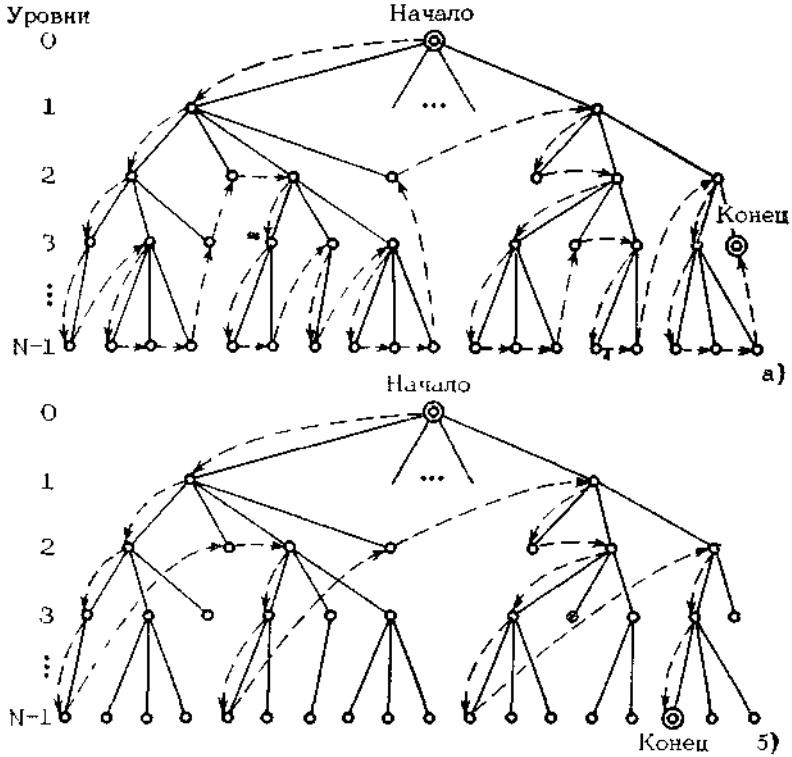


Рис. 8. Геометрическая интерпретация принципа обязательного просмотра всех ветвей дерева решений

При ограниченном времени решения задачи, имеющей большую размерность, остается неисследованным большинство ветвей, расположенных справа (рис. 8,*а*), т. е. нахождение экстремума не гарантируется.

Остановимся на процедуре анализа ветвей дерева решений, в основу которой положен *принцип обязательного просмотра всех ветвей*

дерева. Очевидно, что при ограниченном времени решения полный анализ каждой ветви дерева решений невозможен и будет ограниченным. Введем понятие оператора глубины анализа ветвей решений

$$U_{\Gamma} = \overline{0, N-2}. \qquad (2)$$

Оператор определяет глубину полного просмотра ветвей дерева. На рис. 8,*б* показано прохождение по дереву частичных решений при U_{Γ}=2. Из рисунка следует, что при заданном значении U_{Γ} осуществляется полный поиск в узлах дерева нулевого, первого и второго уровней и поиск только в узлах одной цепочки ветвей дерева, корнями которых являются узлы второго уровня дерева решений. При U_{Γ} =0 анализируется только N узлов дерева решений, и алгоритм вырождается в известный метод последовательной оптимизации, позволяющей получить локальный экстремум.

При U_{Γ} = N—*2* объем анализа узлов дерева решений становится эквивалентным объему анализа в известных алгоритмах, построенных на методе ветвей и границ. Таким образом, изменяя значения оператора глубины анализа ветвей дерева решений, можно изменять объем просмотра от N до $1{,}26^N$ узлов.

Ниже описывается алгоритм, реализующий изложенный метод, на примере решения задачи разрезания графа на подграфы с ограничением числа вершин в них.

Пример. Задан граф $C(V, T)$, где V—множество вершин; T — множество ребер, имеющих веса S. Необходимо найти такое разрезание этого графа $B(C_i)$ на подграфы C_i, чтобы вес S_{ij} разрезаемого графа был минимальным.

Решение. Пусть

$$\forall [C_i,\ C_j \in B(C_i)] \left[\sum_{ij} S_{ij} \to \min,\ i \neq j,\ \sum_{i} V_i \leqslant M \right]. \qquad (3)$$

Алгоритм сводится к последовательному выполнению трех операторов: ветвления, исключения и установления оптимального решения.

Ветвление процесса решения задачи состоит в построении узлов *дерева наборов ребер* (ДНР). Отобразим множество ребер $T=\{t_n\}$ в множество целых неотрицательных чисел $S=\{s_n\}$, которые определяют вес ребер исходного графа. Упорядочим множество T по невозрастанию весов ребер. В результате получим последовательность $<t_m>^0$, которой поставим в соответствие последовательность индексов:

$$I = <i_n> = <1, 2, \ldots, N>. \qquad (4)$$

Очевидно, что любому подмножеству $<t_m>$ из $P^x = n \times m$ индексов

соответствует набор ребер $<t_m>$. Назовем последовательность $<i_m>$ индексным выражением набора ребер $<t_m>$, а множество возможных наборов, определяющее число узлов ДНР, обозначим через *H*. Корень ДНР представляет собой набор ребер, построенный на первых P^x номерах ребер упорядоченной последовательности $<t_m>^0$. Индексное выражение этого набора имеет вид $i_{P^x} = <1, 2, \ldots, P^x>$.

Правило построения остальных *H*—1 узлов ДНР состоит в следующем. Поскольку узел ДНР может иметь два соседних узла — справа и снизу, то, для того чтобы образовать соседний справа узел, необходимо в текущем индексном выражении ДНР увеличить на единицу последний индекс. Соседний справа узел не может быть образован, если $i_j = N$. Для образования соседнего снизу узла необходимо найти в текущем индексном выражении ДНР, начиная со старших индексов, пару индексов, для которой справедливо отношение $i_j - i_{j-1} = 2$, и увеличить меньший индекс на единицу. Если в текущем узле не найдется пары соседних индексов, удовлетворяющих данному отношению, то соседний снизу узел не может быть образован.

Остановимся на свойствах ДНР, вытекающих из правил его построения.

Свойство 1. Значение показателя эффективности набора, соответствующего узлу, достижимому из текущего вправо вниз, не меньше значения эффективности набора, соответствующего данному узлу.

Свойство 2. Для любого узла ДНР все наборы, соответствующие узлам, из которых достижим данный, имеют в индексном выражении общую часть.

Свойство 3. ДНР содержит все наборы индексов и никакой набор в нем не встречается дважды. Отметим, что ДНР не хранится в памяти ЭВМ, а строится последовательно оператором ветвления по рассматриваемым ниже правилам. Назовем процедуру последовательного построения ДНР *естественным ветвлением.*

Оператор исключения *U* состоит из четырех операторов: $<U_1, U_2, U_3, U_4>$, определенная последовательность выполнения которых обеспечивает решение следующих задач: исключение на каждом шаге построения индексного набора ребер графа *C(V, T),* нарушающих ограничение по числу вершин в подграфах; исключение из рассмотрения на оптимальность узлов ДНР, соответствующие наборы ребер которых заведомо неоптимальны; коррекция процесса естественного ветвления; обеспечение заданной глубины анализа ветвей ДНР.

На первом шаге оптимизации Ψ=l, перед включением очередного i_j-го ребра в набор оператор U_1, производит проверку условия допу-

стимости ребра t_j исходя из ограничения на число вершин во всех образованных подграфах. Если ограничение не выполняется, то это означает, что данное ребро в набор не входит.

Согласно свойству 2 то же условие не будет выполняться и для множеств наборов, соответствующие узлы ДНР которых достижимы из рассматриваемого естественным ветвлением вправо. В этом случае необходим переход на нижнюю ветвь ДНР, который обеспечивает оператор U_2. Если условие допустимости выполняется, то ребро t_j включается в набор, построение которого рассматривается, образуя новое звено индексного выражения. После этого анализируется окончание набора ребер.

Если $i_j = N$, то набор $<t_m>^1$ является первым опорным решением разрезания и соответствует решению задачи методом последовательной оптимизации. На втором шаге оператор U_3 осуществляет определение опорного решения второго шага оптимизации $\Psi = 2$ с помощью исключения узлов ДНР, значение показателей эффективности которых заведомо хуже решения $<t_m>^1$. В индексном выражении, соответствующем узлу ДНР, в котором получено опорное решение, начиная со старших разрядов ищется пара соседних индексов, для которых выполняется условие

$$\Delta i = i_j - i_{j-1} > 1.$$

Если $\Delta i > 2$, то меньший из сравниваемых индексов увеличивается, на единицу. Если $\Delta i = 2$, то в следующей паре соседних индексов, для которых разность больше единицы, меньший из сравниваемых индексов увеличивается на единицу. Значение измененного индекса запоминается и считается границей набора F_1. Все индексное выражение, находящееся слева от границы, входит в новый набор, соответствующий исходному узлу второго шага оптимизации. Этот набор дополняется индексами до индекса $i_j = N$ по правилам построения дерева наборов, приведенным выше.

Процедура второго шага оптимизации аналогична обычно используемому алгоритму, реализующему метод ветвей и границ. В описываемый алгоритм введен оператор глубины анализа ветвей ДНР — $U_г$, который является оператором U_4, т. е.

$$U_г = U_4 = \overline{0,\ N-2}. \qquad (5)$$

Ниже дается описание работы алгоритма для $U_г = U_4 = 1$. В индексном выражении опорного решения начиная от границы F_Ψ в направлении младших индексов ищется первый индекс, для которого выполняется условие $\Delta i = i_j - i_{j-1} \geqslant 2$. Меньший из сравниваемых индексов увеличивается на единицу. Это значение индекса считается новым значе-

нием границы $F_{\Psi+1}$ и заносится в память ЭВМ на место старого значения F_{Ψ}.

Все индексы, находящиеся слева от измененного, соответствуют исходному звену набора, входящему в опорную вершину (Ψ+1)-го шага оптимизации. Включение очередных ребер, находящихся в множестве индексов 1 за границей $F_{\Psi+1}$, осуществляется в соответствии с правилами построения индексного набора первого шага оптимизации. При значениях оператора U_4>1 вводятся понятия новых уровней локальных границ набора — F_{Ψ}^{φ}, которые обеспечивают работу алгоритма при анализе узлов ДНР для индексов набора за границей верхнего уровня $F_{\Psi}^{\varphi-1}$.

Оператор установления оптимального решения включается в процесс оптимизации после получения очередного опорного решения. Основная характеристика этого оператора — верхняя граница

$$\widehat{V}(\Psi) = \min\{V_0, V_1, \ldots V_{\Psi}\}, \; V_0 = \infty. \qquad (6)$$

Очевидно, что использование этой характеристики позволяет на Ψ-м шаге процесса оптимизации, конечность которого следует из конечности ДНР, определить оптимальное решение. Как показали исследования, глубина анализа U_4 = 1 обычно достаточна для решения задачи разрезания сложной электрической сети с числом разъемов более 100.

19.6. Эвристическое программирование

Эвристическое программирование объединяет методы оптимизации, основанные на применении правил, приемов, упрощений, обобщающих опыт человека. Эвристическое программирование использует в основном аналогии и неполные индукции, базируется на изучении процессов решения задач оптимизации человеком и представляет собой попытку промоделировать на ЭВМ процесс индуктивного умозаключения. Под индуктивным умозаключением понимается методика поиска решения, когда на основе некоторой гипотезы делаются выводы или решается задача и при этом используется *метод проб и ошибок.* Эвристическое программирование нацелено на решение особенно сложных задач, которые трудно или невозможно решить другими способами.

Эвристическое решение задачи начинается с выбора гипотез, которые необходимо проверить В качестве критерия пригодиости гипотез используется имеющаяся информация, при этом делается попытка

отделить возможные решения от невозможных. Такие критерии принято называть эвристиками задачи.

Эвристики — это своего рода эмпирические алгоритмы, получаемые на основе опыта решения вполне определенных задач, поэтому их нельзя применять при решении более широкого класса задач, так как небольшие изменения в условиях задач могут привести к тому, что эвристики, найденные на основе ограниченного эмпирического опыта, окажутся неэффективными. Эвристики подразделяют на синтаксические и семантические.

Обычно человек решает задачу не целиком, а разбивает ее на подзадачи, т. е. осуществляет декомпозицию задачи. Естественно, что совокупность подзадач эквивалентна исходной задаче. Такое разбиение общей задачи на подзадачи получило название *метода регрессии* и не всегда может привести к решению. Это составляет одно из свойств эвристических методов — нет гарантии, что решение будет получено и цель достигнута. Эвристическое программирование не гарантирует нахождения оптимального решения, а только определяет квазиоптимальное решение, что в большинстве случаев достаточно для ряда задач. Укажем на некоторые характерные приемы эвристического программирования: сужение области исследования. Специалист, опираясь на опыт и интуицию, может ввести дополнительные ограничения на область допустимых решений, чем облегчит отыскание оптимальной с точки зрения специалистов, что позволяет ограничиться проверкой критерия качества не во всей области решения, а только в окрестности самого решения.

Эвристическое программирование представляет собой метод *реализации эвристических алгоритмов,* которые основываются на сокращении перебора вариантов путем введения определенных эвристически найденных процедур. Пусть при решении задачи необходимо проанализировать *дерево решений,* его вершинам соответствуют рассматриваемые позиции (ситуации), между которыми устанавливается иерархия. Для исследования любой позиции a_i дерева решений достаточно оценить все непосредственно подчиненные ей позиции, т. е. те из них, в которые можно прийти из этой позиции за один ход.

Для организации иерархического перебора можно составить программу решения, однако ее реализация связана с большими вычислительными затратами. Целесообразней изыскать метод, обеспечивающий отсечение заведомо бесперспективных ветвей. Такой метод известен под названием *метода граней и оценок.* Оценка исходной позиции a_0 определяется как максимум оценок позиций a_1, a_2, ..., a_n, непосредственно ей подчиненных. Метод отсечения

особенно эффективен, если анализирует лучшие варианты. Быстрые способы определения оценки рассматриваемой позиции (варианта, ситуации) целесообразны, хотя дают приближенный и даже не всегда правильный результат. Для сокращения перебора запоминают ранее рассмотренные подобные позиции, которые могут встретиться в других ситуациях.

Однако сокращение перебора с использованием только этих общих методов недостаточно для удовлетворительного решения задач эвристического программирования, и возникает необходимость разработки методов, специфических для данного класса задач или данной конкретной задачи. Для упрощения разработки таких методов создают семантические модели ситуаций (позиций). Семантические модели включают в себя фиксированный круг понятий, а также средства его расширения. Для автоматизации построения таких понятий применяют методы *теории распознавания образов,* являющейся частью общей проблемы искусственного интеллекта.

Эвристическое программирование весьма целесообразно для решения задач большой размерности, даже в том случае, когда имеются «строгие» методы решения. К задачам такого класса следует отнести и задачи проектирования систем управления. Здесь возможно большое число вариантов, но опытный проектировщик сразу отбрасывает бесперспективные. При проектировании широко и порой недостаточно осознанно используются разного рода эвристические оценки, правила и алгоритмы. Достаточно указать на то, что опытный конструктор дает окончательную оценку конструкции исходя из критерия смотрится — не смотрится.

19.7. Стохастическое программирование

Во многих случаях исходная информация для решения задач оптимизации недостаточно определена, например уровень вибрационных нагрузок или аэродинамические характеристики для систем управления летательными аппаратами, уровень фоновых засветок в системах астроориентации и астронавигации, помехи на входе управляющих ЭВМ в автоматизированных системах управления и др. Таким образом, при решения задач оптимизации характеристики, параметры, критерии качества и ограничения могут оказаться неопределенными и случайными. Как следствие возникает проблема поиска оптимальных решений в условиях неопределенности.

Оптимальное решение в условиях неопределенности является не безусловно самым лучшим, а лучшим лишь в некоторых условиях, на-

пример при многочисленных повторениях процесса или большой его длительности. Поэтому целесообразно попытаться избавиться от неопределенности, максимально использовать априорную и текущую информацию, сократить дисперсии распределений или сузить множества возможных значений неопределенных величин и т. д.

Стохастическое программирование дает методы решения условных экстремальных задач при неполной информации о параметрах исследуемого процесса. Обычно при решении подобных задач, когда целевая функция и (или) ограничения являются случайными величинами, рассматривают математические ожидания этих величин. ***Иными словами, сводят стохастическую задачу к детерминированной.*** Однако такой подход не является строгим, так как при усреднении параметров может быть нарушено соответствие математической модели изучаемому процессу. Другой подход состоит в вычислении математического ожидания критерия с использованием метода статистических испытаний, но это связано с серьезными трудностями.

Рассмотрим еще один подход. Пусть задан функционал $F(x, B, U, T)$, характеризующий эффективность управления системой за время T. Здесь x, B, U — функции времени. Функционирование системы описывается в каждый момент времени t ($0 \le t \le T$) наблюдаемой вектор-функцией состояния $x(t)$ и ненаблюдаемой вектор-функцией условий $B(t)$. Вектор функционалов управления $U=U[t, x(t)]$ наблюдаем до момента t реализации вектор-функции состояния $x(\tau)$, где $0 \le \tau \le t$. Векторы-функции $x(t)$ и $B(t)$ связаны определенными закономерностями между собой и с выбранным управлением $U[t, x(t)]$.

Управление стремятся выбрать так, чтобы оптимизировать функцию, представляющую собой математическое ожидание показателя эффективности управления, т. е.

$$J(U, x) = M_{x,B} F(x, B, U, T), \qquad (1)$$

где $M_{x,B}$ — математическое ожидание показателя эффективности управления по процессам x и B.

В такой общей постановке задачи построить эффективные алгоритмы ее решения практически невозможно, поэтому задачу решают для частных случаев, например $B(t)$ представляет собой случайный процесс, заданный своими статистическими характеристиками, $x(t)$—детерминированный процесс, протекающий независимо от $B(t)$, либо условия $B(t)$ не изменяются и заданы априорным распределением $P(B)$ или плотностью распределения $f(B)$ и стохастически связаны с $x(t)$. Следует, однако, иметь в виду, что указанные выше статистические характеристики не всегда известны.

19.8. Методы формализации качественных характеристик

При решении задач оптимизации возникают задачи оптимизации качественных характеристик. Для решения подобных задач известными количественными методами необходимо качественным характеристикам придать некоторые количественные оценки, т. е. решить *задачу оценивания*. Смысл последней состоит в сопоставлении рассматриваемой системе (альтернативе, критерию) вектора пространства E_m. Пространство E_m назовем m-мерной шкалой (при m=1 — просто шкалой), операцию сопоставления системе вектора — *оцениванием,* нахождение этого вектора — *задачей оценивания.* Простейшей задачей оценивания является задача измерения при m=1, когда оценивание сводится к сравнению с эталоном. Сложнее, если эталон отсутствует.

Определения множества допустимых оценок (МДО) и наиболее точной оценки являются характерными этапами процесса оценивания. Если на первом этапе определяется подмножество множества $E = \bigcup_{m=1}^{\infty} E_m$, в котором ищется оценка системы, то на втором, этапе выбирается оценка, наиболее точно выражающая свойства оцениваемой системы. Общая схема экспертизы изображена на рис. 9.

Рис. 9. Построение множества допустимых оценок

Важной разновидностью определения МДО является *задача ранжирования,* состоящая в упорядочении объектов, образующих систему, по убыванию или возрастанию значения некоторого признака, количественно неизмеримого. Ранг x_i указывает то место, которое занимает *i*-й объект среди других *n*—1 объектов, ранжированных в соответствии с признаком *x*. Построение МДО для экспертов существенно зависит от *формы опроса эксперта.* Здесь можно указать на *опрос типа интервью,* представляющий собой беседу исследователя с экспертом, в ходе которой исследователь ставит эксперту вопросы в соответствии с заранее разработанной программой (сценарием). Большую роль играет взаимопонимание между исследователем и экспертом. Другая форма опроса состоит в *анкетировании.* Вопросы в анкете должны быть сформулированы так, чтобы исключить их неоднозначное восприятие. Сложные вопросы разбиваются на простые. Качественным вопросам дается предпочтение перед количественными.

При взаимодействии экспертов возможны варианты полной изоляции экспертов, регламентированного обмена информацией между ними и свободного обмена информацией. Для объективизации оценок большое значение имеет организация *обратной связи в экспертизе.* Так, для *метода Дельфи,* разработанного в корпорации РЭНД *Хелмером и Далки* в 1963 г., обратная связь организуется путем анонимного ознакомления экспертов с мнениями, высказанными их коллегами на предыдущих турах опроса.

Важное место занимает обработка результатов экспертных оценок. Если рассматривать результаты оценок каждого из экспертов как реализации некоторой случайной величины, то к ним можно применять методы математической статистики. Степенью согласованности мнений экспертов служит дисперсия

$$\sigma^2 = \sum_i^N (a - a_i)^2 \alpha_i \,/\, \sum_i^N \alpha_i, \qquad (1)$$

где a — результирующая оценка; a_i — оценка *i*-го эксперта; α_i — веса экспертов.

Степень согласованности мнений экспертов для случая строгого ранжирования, т. е. отсутствия равных рангов в ранжировке каждого эксперта, можно определить при помощи коэффициента конкордации

$$\Phi = 12 \sum_{i=1}^{n} [r_i - N(n+1)/2]^2 / [N^2 n(n^2-1)], \qquad (2)$$

где *n, N* — число объектов и экспертов соответственно. Если Ф =0, то это означает, что связь между ранжировками экспертов отсутствует. Если Ф=1, то все эксперты одинаково ранжируют объекты по данному признаку.

Метод ПАТТЕРН, или метод прогнозного графа, заключается в построении на основе экспертных оценок дерева решений как модели сложной сети взаимосвязей. Существенно, что при этом сложная задача разбивается на относительно простые подзадачи, каждая из которых подвергается обработке на ЭВМ.

В последние десятилетия, главным образом в работах *Л. Заде* и его школы, создан математический аппарат, получивший название *теории расплывчатых (размытых, нечетких) множеств*. Центральным понятием этой теории является понятие расплывчатого множества, отличающегося от обычного «нерасплывчатого» (жесткого, четкого, неразмытого) множества тем, что, если произвольный элемент *x* рассматриваемой предметной области *X* может либо принадлежать данному «жесткому» множеству *M,* т. е. имеет место $x \in M$, либо не принадлежать ему $(x \notin M)$, расплывчатое множество $M_р$ допускает принадлежность элементов множеству различной степени, оцениваемой на бесконечной шкале действительных чисел от 0 до 1 (0 означает полную непринадлежность, 1 — полную принадлежность). Промежуточные оценки записываются в виде μx, где μ — *оператор расплывания,* значения которого находятся в области [0, 1]. Таким образом, оценки переходят из области *точных значений* в область *размытых значений.* Применительно к размытости значений можно сформулировать следующий подход: эксперт указывает, какие значения параметров заведомо неприемлемы и какие значения оцениваются по высшему баллу. Далее полагают, что нулевое и единичное значения соединяются в диапазоне (x_0, x_1) плавной кривой. Так, например, в диапазоне $(0, x)$ можно использовать функцию

$$y = 2\left[1 - 1 \Big/ \left(1 + \exp \frac{x - x_0}{x}\right)\right]. \qquad (3)$$

Тогда оценка эксперта типа «очень» будет определяться через y^2, а «гораздо хуже» — через $y^{1/2}$ и т. п. Существенно здесь то, что принадлежность параметра к классам «хорошо» и «плохо» определяется экспертом и четкой границы между плохим и хорошим он установить не может. Метод *Л. Заде* частично позволяет

использовать нечеткие определения. По мысли автора метода, теория расплывчатых множеств при анализе и проектировании *гуманистических систем,* т. е. систем, в которых существенная роль принадлежит суждениям и решениям человека, может оказаться (и оказалась) гораздо эффективнее, чем классическая математика с ее идеей математической непрерывности или даже конечная математика, непосредственно переводимая на цифровой язык ЭВМ. Нечеткие множества особенно полезны при количественном анализе особо сложных гуманистических систем. Этому служит и *принцип несовместимости Л. Заде,* согласно которому высокая точность невозможна для систем большой сложности.

Список обозначений

$\mathbf{R}^n$ — n-мерное вещественное евклидово пространство.

$\{x_1, ..., x_n\}$ — компоненты вектора $x \in \mathbf{R}^n$.

$\| \bullet \|$ — норма в $\mathbf{R}^n$: $\|x\|^2 = x^2_1 + ... + x^2_n$.

$(\bullet, \bullet)$ — скалярное произведение в $\mathbf{R}^n$: $(x, y) = x_1y_1 + ... + x_ny_n$

I — единичная матрица

A^{T} — матрица, транспонированная к A.

A^+ — псевдообратная матрица к A.

$A \geq B$ — матрицы A и B симметричны и $A - B$ неотрицательно определена

$A > B$ — матрицы A и B симметричны и $A—B$ положительно определена.

$\|A\|$ — норма матрицы A: $\|A\| = \max\limits_{\|x\|=1} \| Ax \|$.

$\rho(A)$ — спектральный радиус матрицы A .

$x \geq y$ — все компоненты вектора $x \in \mathbf{R}^n$ не меньше соответствующих компонент вектора $y \in \mathbf{R}^n$: $x_i \geq y_i$, $i = 1, ..., n$.

$\mathbf{R}^n_+$ — неотрицательный ортант в $\mathbf{R}^n$: $\mathbf{R}^n_+ = \{ x \in \mathbf{R}^n : x \geq 0 \}$.

x_+ — положительная часть вектора $x \in \mathbf{R}^n$: $(x_+)_i = \max \{0, x_i\}$, $i = 1, ... n$.

$x^* = \arg\min\limits_{x \in Q} f(x)$ —любая точка глобального минимума $f(x)$ на Q: $x \in Q$, $f(x^*) = \min\limits_{x \in Q} f(x)$.

$X^* = Arg \min\limits_{x \in Q} f(x)$ — множество точек глобального минимума $f(x)$ на Q: $X^* = \{x^* = \arg\min\limits_{x \in Q} f(x)\}$.

$\nabla f(x)$, $f'(x)$ — градиент скалярной функции $f(x)$.

$\nabla g(x)$, $g'(x)$ — производная векторной функции $g(x)$, матрица Якоби.

$\nabla^2 f(x)$, $f''(x)$—матрица вторых производных, гессиан.

$L'_x(x, y)$, $L''_{xx}(x, y)$ — градиент и матрица вторых производных $L(x, y)$ по переменной x.

$df(x)$ — субградиент выпуклой функции.

$\partial_\varepsilon f(x)$ — ε-субградиенг выпуклой функции.

$f'(x; y)$ — производная функции $f(x)$ в точке x по направлению y.

$D(f)$ — область определения функции $f(x)$.

Conv Q — выпуклая оболочка множества Q.

Q — внутренность множества Q.

$\varnothing$ — пустое множество.

$P_Q(x)$ — проекция точки x на множество Q .

$\rho(x, Q)$ — расстояние от точки x до множества Q: $\rho(x,Q)=\inf\limits_{y\in Q} \| x-y \|$

$o(h(x))$ — если $g: \mathbf{R}^n\to\mathbf{R}^m$, $h: \mathbf{R}^n\to\mathbf{R}^s$ и $\|g(x)\|/\|h(x)\| \to 0$ при $\|x\|\to 0$.

$O(h(x^*))$ — если $g: \mathbf{R}^n\to\mathbf{R}^m$, $h: \mathbf{R}^n\to\mathbf{R}^S$ и найдутся $\varepsilon > 0$, α такие, что $\|g(x)\| \le \alpha \| h(x)\|$ при $\|(x)\| \le \varepsilon$, то $g(x) = O(h(x))$.

$o(u_k)$ — если последовательности $u_k \in \mathbf{R}^n$, $v_k \in \mathbf{R}^m$, $k = 1, 2, \ldots$, таковы, что $\| v_k \|/\| u_k \| \to 0$ при $k\to\infty$, то $v_k = o(u_k)$.

$O(u_k)$ — если для последовательностей $u_k \in \mathbf{R}^n$, $v_k \in \mathbf{R}^m$, $k = 1, 2, \ldots$, найдутся $\alpha > 0$, k_0 такие, что $\| v_k \| \le \alpha \| u_k \|$ при $k \ge k_0$, то $v_k = O(u_k)$.

$\mathsf{M}\xi$ — математическое ожидание случайной величины ξ

$\mathsf{M}(\xi | x)$ — условное математическое ожидание случайной величины ξ, зависящей от x, при фиксированном значении x.

$\forall$ — квантор общности: $\forall x \in Q$ — «для всех $x \in Q$».

Литература

1.Основная

1. Аоки М. Введение в методы оптимизации. — М.: Паука, 1977.
2. Бахвалов Н. С. Численные методы.—М.: Наука, 1973.
3. В а й н б е р г М. М. Вариационный метод и метод монотонных операторов в теории нелинейных уравнений —М.: Наука, 1972.
4. Габасов Р., Кириллова Ф. М. Методы оптимизации. — Минск: Б ГУ, 1975.
5. Демьянов В. Ф., Рубинов А М. Приближенные методы решения экстремальных задач —Л.: ЛГУ, 1968.
6. Зангвилл У. Нелинейное программирование. Единый подход --М.: Сов. Радио, 1973.
7. Зойтендейк Г. Методы возможных направлений —М.: ИЛ, 1963.
8. Карманов В. Г. Математическое программирование. — М.: Наука, 1975.
9. Кононюк А.Ю. Вища математика. К.1. — К.: КМТ, 2009.
10.Кононюк А.Ю. Вища математика. К.2. — К.: КМТ, 2009.
11.Кононюк А.Е. Дискретная математика. К.1, ч.1 — К.: Освіта України, 2010.
12.Кононюк А.Е. Дискретная математика. К.1, ч.2 — К.: Освіта України, 2010.
13.Кононюк А.Е. Дискретная математика. К.2, ч.1 — К.: Освіта України, 2011.
14.Кононюк А.Е. Дискретная математика. К.2, ч.2 — К.: Освіта України, 2011.
15.Кононюк А.Е. Дискретная математика. К.2, ч.3 — К.: Освіта України, 2011.
16.Кононюк А.Е. Дискретная математика. К.3, ч.1 — К.: Освіта України, 2011.
17.Кононюк А.Е. Дискретная математика. К.3, ч.2 — К.: Освіта України, 2011.
18. М о и с е е в Н. Н., И в а н и л о в Ю. П., Столярова Е. М. Методы оптимизации.—М.: Наука, 1978.
19. Ортега Дж., Рейнболдт В. Итерационные методы решения нелинейных систем уравнений со многими неизвестными — М.: Мир, 1975.
20. Полак Э. Численные методы оптимизации. Единый подход — М: Мир, 1974.
21. Поляк Б. Т. Введение в оптимизацию. —М.: Наука, 1983.

22. Пшеничный Б. Н., Данилин Ю. М. Численные методы в экстремальных задачах.— М.: Наука, 1975.
23. Растригин Л. А. Системы экстремального управления —М.: Наука, 1974.
24. С е а Ж. Оптимизация. Теория и алгоритмы.— М.: Мир, 1973
25. Уайлд Д. Дж. Методы поиска оптимума.—М.: Науки 1967.
26. Федоренко Р. П. Приближенное решение задач оптимального управления. — М.: Наука, 1978.
27. Фиакко А., Мак-Кормик Дж. Нелинейное программирование: методы последовательной безусловной минимизации.—М: Мир 1972.
28. Хи м м е л ь б л а у Д. Прикладное нелинейное программирование М.: Мир, 1975.
29. Численные методы условной оптимизации /Под ред. Ф. Гилла, У Мюр рея. — М.: Мир, 1977.
30. Э р р о у К. Дж., ГурвицЛ.УдзаваХ. Исследования по лпигП ному и нелинейному программированию. — М.: ИЛ, 1962.

2. Дополнительная

1. *Абакаров А.Ш., Сушков Ю.А.* Статистическое исследование одного алгоритма глобальной оптимизации. — Труды ФОРА, 2004.
2. *Акулич И.Л.* Математическое программирование в примерах и задачах: Учеб. пособие для студентов эконом. пец. вузов. — М.: Высшая школа, 1986.
3. *Гилл Ф., Мюррей У., Райт М.* Практическая оптимизация. Пер. с англ. — М.: Мир, 1985.
4. *Жиглявский А.А., Жилинкас А.Г.* Методы поиска глобального экстремума. — М.: Наука, Физматлит, 1991.
5. *Карманов В.Г.* **Математическое программирование** = Математическое программирование. — Изд-во физ.-мат. литературы, 2004.
6. *Корн Г., Корн Т.* Справочник по математике для научных работников и инженеров. — М.: Наука, 1970. — С. 575-576.
7. *Коршунов Ю.М., Коршунов Ю.М.* Математические основы кибернетики. — М.: Энергоатомиздат, 1972.
8. *Максимов Ю.А.,Филлиповская Е.А.* Алгоритмы решения задач нелинейного программирования. — М.: МИФИ, 1982.
9. *Максимов Ю.А.* Алгоритмы линейного и дискретного программирования. — М.: МИФИ, 1980.
10. *Огирко И. В.* Расчет и оптимизация термоупругого состояния тел с учетом геометрической и физической нелинейности :

Автореф. дис. на соиск. учен. степ. д-ра физ.-мат. наук : (01.02.04) / Казан. гос. ун-т им.— Казань, 1989.
11. *Плотников А.Д.* Математическое программирование = экспресс-курс. — 2006. — С. 171. — ISBN 985-475-186-4
12. *Растригин Л.А.* Статистические методы поиска. — М.: 1968.
13. *Хемди А. Таха* Введение в исследование операций = Operations Research: An Introduction. — 8 изд.. — М.: «Вильямс», 2007. — С. 912. — ISBN 0-13-032374-8
14. Никайдо Х. Выпуклые структуры и математическая экономика. — М.: Мир, 1972
15. Кини Р. Л., Райфа Х. Принятие решений при многих критериях: предпочтения и замещения.- М.: Радио и связь, 1981
16. Соболь И. М., Статников Р. Б. Выбор оптимальных параметров в задачах со многими критериями. — М.: Наука, 1981
17. Подиновский В. В., Ногин В. Д. Парето-оптимальные решения многокритериальных задач. — М.: Наука, 1982
18. Морозов В. В., Сухарев А. Г., Федоров В. В. Исследование операций в задачах и упражнениях. — М.: Высшая школа, 1986
19. Юдин Д. Б. Вычислительные методы теории принятия решений. — М.: Наука, 1989
20. Емеличев В. А., Мельников О. И., Сарванов В. И., Тышкевич Р. И. Лекции по теории графов. — М.: Наука, 1990
21. Штойер Р. Многокритериальная оптимизация. — М.: Радио и связь, 1992
22. Батищев Д. И., Коган Д. И. Вычислительная сложность экстремальных задач переборного типа. — Изд. ННГУ, Н. Новгород, 1994
23. Коротченко А. Г., Тихонов В. А. Методические указания (сборник задач) по курсу «Модели и методы принятия решений» — Изд. ННГУ, Н. Новгород, 2000
24. Коротченко А. Г., Бобков А. Н. Принципы оптимальности в задачах принятия решений (методическая разработка) — Изд. ННГУ, Н. Новгород, 2002
25. Батищев Д. И. Задачи и методы векторной оптимизации. — Изд. ГГУ, Горький, 1979
26. Розен В. В. Цель- оптимальность- решение: Математические модели принятия оптимальных решений. — М.: Радио и связь, 1982
27. Батищев Д. И. Методы оптимального проектирования. — М.: Радио и связь, 1984

28. Г. М. Уланов и др. Методы разработки интегрированных АСУ промышленными предприятиями. М.: Энергоатомиздат – 1983.
29. А. М. Анохин, В. А. Глотов, В.В. Павельев, А.М. Черкашин. Методы определения коэффициентов важности критериев "Автоматика и телемеханика", №8, 1997, с3-35.
30. Таха, Хэмди А. Введение в исследование операций – М.:Мир,2001, с354-370.
31. Р. Штойер. Многокритериальная оптимизация: теория, вычисления, приложения. М.:Наука, 1982, с14-29, 146-258.
32. Многокритериальная оптимизация. Математические аспекты. М.:Наука, 1989, с116-123.
33. В.В. Подиновский, В.Д. Ногин. Парето-оптимальные решения многокритериальных задач. М.: Наука, 1982, с9-64.
34. В. В. Хоменюк. Элементы теории многокритериальной оптимизации. М.: Наука, 1983, с8-25.
35. Д.И.Батищев, С.А.Исаев, Е.К.Ремер. Эволюционно-генетический подход к решению задач невыпуклой оптимизации. /Межвузовский сборник научных трудов «Оптимизация и моделирование в автоматизированных системах», Воронеж, ВГТУ, 1998г, стр.20-28.
36. Д.И.Батищев, С.А.Исаев. Оптимизация многоэкстремальных функций с помощью генетических алгоритмов. /Межвузовский сборник научных трудов «Высокие технологии в технике, медицине и образовании», Воронеж, ВГТУ, 1997г, стр.4-17.
37. С.А.Исаев. Популярно о генетических алгоритмах. Интернет-ресурс http://bspu.ab.ru./Docs/~saisa/ga/ga-pop.html.
38. С.А.Исаев. Обоснованно о генетических алгоритмах. Интернет-ресурс http://bspu.ab.ru/Docs/~saisa/ga/text/part1.html.
39. С.А.Исаев. Решение многокритериальных задач. Интернет-ресурс http://bspu.ab.ru/Docs/~saisa/ga/idea1.html.
40. Раздел «Математика\Optimization Toolbox». Интернет-ресурс http://www.matlab.ru/optimiz/index.asp.
41. Система СИМОП для автоматизации выбора рациональных решений в комплексах САПР и АСНИ. Интернет-ресурс. http://www.software.unn.ac.ru/mo_evm/research/symop.html
42. Интегрированный пакет многокритериальной оптимизации «МАЛТИ». Интернет-ресурс http://ksu.kst.kz/emf/kafkiber.htm
43. Комплексный инженерный анализ - прочность, динамика, акустика. Интернет-ресурс http://osp.admin.tomsk.ru/ap/1998/02/31.htm
44. Программы семейства COSMOS – универсальный инструмент конечно-элементного анализа.
Интернет-ресурс http://cad.com.ru/7/Info/cosmos_3.html

Научно-практическое издание

Кононюк Анатолий Ефимович

Базовая теория оптимизации

Книга 1
Начала теории оптимизации

Авторская редакция

Подписано в печать 21.02.2011 г.
Формат 60x84/16.
Усл. печ. л. 16,5. Тираж 300 экз.

Издатель и изготовитель:
Издательство «Освита Украины»
04214, г. Киев, ул. Героев Днепра, 63, к. 40
Свидетельство о внесении в Государственный реестр
издателей ДК №1957 от 23.04.2009 г.
Тел./факс (044) 411-4397; 237-5992
E-mail: osvita2005@ukr.net, www.rambook.ru