В. Н. САЛИЙ

МАТЕМАТИЧЕСКИЕ ОСНОВЫ ГУМАНИТАРНЫХ ЗНАНИЙ

Учебное пособие
для студентов гуманитарных направлений и специальностей
высших учебных заведений

Издательство Саратовского университета
2005

УДК [51-7: 009] (075.8)
ББК 22.1я73
С16

Салий В.Н.

С16 Математические основы гуманитарных знаний: Учеб. пос. для студентов гуманит. направлений и специальностей высш. учеб. заведений. – Саратов: Изд-во Сарат. ун-та, 2005. – 308 с.: ил.

ISBN 5-292-03355-3

Систематически и в доступной форме излагается теоретический материал, направленный на усвоение основных идей и важнейших понятий математики, а также полезные сведения прикладного характера. Текст сопровождается многочисленными диаграммами, таблицами и рисунками, иллюстрирующими содержание, интересными фактами из истории науки.

Данное учебное пособие признано победителем Всероссийского конкурса учебников нового поколения, проведенного Минобразования России в 1999 г.

Для студентов гуманитарных направлений и специальностей, может служить справочным пособием для научных работников и аспирантов, а также для всех интересующихся классической и современной математикой.

Рекомендуют к печати:

Кафедра теоретических основ
компьютерной безопасности и криптографии
Саратовского государственного университета
Профессор Московского государственного университета *Х. Д. Икрамов*

Издается при финансовой поддержке
гранта Саратовского государственного университета
«Лучшая книга СГУ – 2004» (номинация «Учебник или учебное пособие»)

Работа издана в авторской редакции

УДК [51-7: 009] (075.8)
ББК 22.1я73

ISBN 5-292-03355-3

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ

Цель предлагаемого пособия – дать студенту-гуманитарию общее представление об основных идеях математики, познакомить его с важнейшими ее понятиями, помочь ему получить первоначальные навыки применения математических методов.

Математическая идеология в главных своих чертах сформировалась при решении классических ныне проблем геометрии, алгебры и анализа. Фигура, число и функция – три символа, олицетворяющих математику, три краеугольных камня, на которые опирается интуиция ее творцов. Необходимый минимум сведений об этих базисных конструкциях содержат главы «Числа и уравнения», «Мир функций» и «Геометрические пространства». В каждой из них изложение ориентировано на несколько уровней восприятия, но так, чтобы, выбрав соответствующий своим возможностям «срез» через эти три главы, читатель мог получить достаточно цельное представление о сути вещей. Важно найти то расстояние, с которого детали уже видны, но еще не закрывают целого.

Глава «Что значит *доказать*?» посвящена основам математической логики, которая изучает правила доказательных рассуждений и общие свойства формальных теорий. Здесь математика тесно соприкасается с новейшей философией познания.

Математические понятия и методы встречаются во многих гуманитарных областях – либо непосредственно, либо в виде некоторых традиционных профессиональных приемов. Однако способ передачи этой методологии зачастую выражается в виде прямых рекомендаций, обосновываемых общепризнанной полезностью или ссылками на авторитеты. Преподавание курса «Математические основы гуманитарных знаний» должно показать, в чем суть применяемых «матметодов», и тем самым дать пользователю возможность проявлять творческую свободу и инициативу. В современных гуманитарных исследованиях наряду с неизбежной статистикой чаще всего используются алгебра множеств, комбинаторные и вероятностные расчеты, понятия из теории передачи сообщений, нечеткие конструкции. Глава «Математика неопределенного» трактует этот материал не только с идейных позиций, но и с прикладной точки зрения. Впрочем, при разнообразии специальностей, которым адресовано пособие, приходится ограничиваться лишь сведениями универсального характера.

В главе «Дискретные системы и их математическое описание» особого внимания заслуживают разделы, посвященные отношениям и графам. Эти объекты привлекают все большее внимание в самых различных гуманитарных науках. Понятие алгоритма является одним из центральных в дискретной математике и информатике. Раздел о формальных языках и грамматиках перебрасывает мост от естественных языков человеческого общения к искусственным языкам программирования, посредством которых мы вступаем в контакт с ЭВМ. В заключительной главе «Искусственный интеллект» рассказывается о современных возможностях и перспективах использования компьютерной техники в качестве усилителя человеческого интеллекта.

В небольшом списке литературы указаны лишь некоторые известные книги по математике, написанные для широкого круга читателей, и пособия, пользуясь которыми, можно углубить знания по темам, представляющим наибольший интерес.

Строя лекционный курс, преподаватель, конечно, примет во внимание специфику будущей профессиональной деятельности слушателей. В одних случаях упор нужно сделать на вопросы математического анализа, в других на первое место выдвигаются геометрические представления, в третьих – понятия алгебры и логики. Круг идей, связанных со статистикой, по-видимому, актуален для всех, а дискретная математика по сути своей является наиболее адекватным аппаратом для построения нестатистических моделей.

Обдумывая форму подачи материала, необходимо, с одной стороны, помнить об особенностях гуманитарного мышления, не терпящего окостеневших схем, а с другой – стремиться как можно более точно описать ключевые понятия. Разумеется, допустима известная свобода в интерпретации математических результатов и идей, но она не должна принимать характер произвольных толкований.

Что касается доказательств, то те из них, в которых цель достигается путем длинных технических выкладок, должны быть безусловно опущены, но там, где доминирующую роль играет логический элемент, стоит задержаться и обсудить применяемые схемы умозаключений.

Столь же осторожно следует подходить к упражнениям. Для студентов специальностей, в которых существенно используется тот или иной математический аппарат, полезно освоить его на примерах, соответствующих уровню приложений (построение графиков, техника дифференцирования, вычисление простейших интегралов, комбинаторика, диаграммы и т.п.). Вообще же, решением задач не следует злоупотреблять, оно способствует, главным образом, овладению техникой, но не идеями. Лучше обсудить на семинаре темы, не вошедшие в лекции, или провести коллоквиум по наиболее сложным вопросам. Тот, кто преподавал математику студентам гуманитарных направлений, согласится, что нет никакого противоречия между двумя фразами выдающегося русского философа Н.М.Бердяева «Я сносно знал теорию математики» и «Я никогда не мог решить ни одной математической задачи»...

Изучение математических курсов студентами гуманитарных специальностей имеет еще один важный аспект. Всякое знание в своем развитии рано или поздно достигает такого уровня абстракции, за которым дальнейший прогресс возможен лишь при соответствующей степени формализации. Современная математика не имеет готовых средств, пригодных для подобных потребностей гуманитарных наук. Математическое образование гуманитариев – необходимый этап в создании этих средств.

В работе над текстом автор опирался как на собственный опыт преподавания различных разделов математики студентам гуманитарных отделений, так и на опыт своих коллег, которым выражает искреннюю признательность за полезные обсуждения.

Бакалавр Самсон Карраско в известной беседе с Дон Кихотом справедливо заметил: «Кто отдает свое произведение в печать, тот подвергается величайшему риску, ибо совершенно невозможно сочинить такую книгу, которая удовлетворила бы всех». Автор с благодарностью примет отзывы о содержании предлагаемого пособия и об избранной в нем форме изложения.

ВВЕДЕНИЕ. **ГУМАНИТАРНАЯ ЦЕННОСТЬ МАТЕМАТИКИ**

Математика – это наука о количественных отношениях и пространственных формах действительного мира. Ее история уходит в глубины тысячелетий. Если первой абстракцией, созданной человеческим разумом, было Слово, то второй – несомненно Число. Как способ мышления, математика сформировала свои основные черты еще в античные времена: точность в определении понятий, логическая строгость доказательных рассуждений, однозначность в понимании смысла установленных фактов. Впоследствии эти черты стали восприниматься как обязательные свойства любой науки.

С другой стороны, в отвлеченных математических схемах оказались выразимыми бесчисленные ситуации и процессы, изучавшиеся в конкретных областях знания. Благодаря этому удавалось решить важные практические задачи в самых разных сферах человеческой деятельности. Математика стала не только идеалом, но и универсальным инструментом естественных и прикладных наук, наряду с языком приобрела роль могучей силы, способствующей развитию цивилизации.

«Чистая» математика, которая образует фундамент всех этих приложений, имеет много общего с гуманитарными науками и даже с искусством. Хотя ее конструкции и связаны с объектами реального мира, но связь эта весьма опосредована, и творческая мысль математика относительно свободна, как и творческая фантазия поэта, композитора или художника. Эстетические критерии всегда занимали важное место в оценке математических результатов. И не случайно среди выдающихся математиков было так много личностей типично гуманитарного склада, оставивших глубокий след в литературе, живописи, музыке, философии, психологии, педагогике, богословии, лингвистике, истории, юриспруденции, искусствоведении, политике.

В наши дни, более чем когда-либо, математические идеи и понятия, своеобразно преломляясь, становятся достоянием общей культуры, служат толчком к возникновению новых взглядов на те или иные явления, проникают в язык искусства и средств массовой информации, интеллектуализируют обыденный разум. Усиление роли математики в общественном сознании не только не противоречит идеям гуманитаризации науки, но, напротив, способствует их распространению на области, традиционно далекие от непосредственного интереса к человеку.

Краткую сводку, суммирующую высказанные в разное время соображения о месте достойнейшей из наук в системе человеческих ценностей, завершим прямой цитатой – из книги английского математика и философа Уайтхеда (1925 г.): «Чистая математика, в ее современном развитии, может претендовать на роль самого оригинального создания человеческого духа. Другим претендентом на это является музыка».

ГЛАВА I. **ЧИСЛА И УРАВНЕНИЯ**

§ 1. Натуральные числа

Число – важнейшее понятие математики. Потребовалось несколько тысячелетий, чтобы это понятие приобрело форму, которая в настоящий момент признается удовлетворительной подавляющим большинством математиков. Однако в соответствующих формулировках используется профессиональный язык столь высокого уровня, что попытка передать их точный смысл «простыми и понятными словами», по-видимому, безнадежна. Приходится довольствоваться лишь общими описаниями.

Простейший вид чисел – натуральные числа – исторически возник из потребностей счета: одна лодка, два человека, три дерева и т.д. Лишь на достаточно высоком интеллектуальном уровне было осознано, что у конкретных предметных групп «два камня», «две птицы» и «две руки» есть нечто общее: «два». Абстрактные, отвлеченные числа позволяли сравнивать количество предметов в разнородных совокупностях, что имело важное значение при обменных операциях типа «раковина за орех».

Развитие счета шло параллельно с изменением в психологическом восприятии понятия «много». Вначале было «один, два, много» или «один, два, три, много», но постепенно граница отодвигалась, формировался натуральный ряд чисел: 1, 2, 3, 4 и т.д. Естественный инструмент счета – пальцы на руках – установил первый предел: десять. Принцип группировки по десять позволял охватывать все большие количества объектов, объединяя их в новые единицы счета: десять десятков – сотня, десять сотен – тысяча, дальше десяти тысяч обыденный разум не заглядывал. Так сформировалась десятичная система счисления. Она позволяла с помощью небольшого количества слов называть все встречающиеся числа: например, триста шестьдесят пять – это три сотни и шесть десятков и пять единиц. Не у всех народов десяток стал основным числом счета: одни осознали в качестве первой границы пять (пальцы одной руки), другие – двадцать (все пальцы на руках и на ногах), в Вавилоне употреблялась система с загадочным основанием шестьдесят, в согласии с ней мы до сих пор делим окружность на триста шестьдесят градусов и измеряем время: в часе – шестьдесят минут, в минуте – шестьдесят секунд. Но в конце концов десятичный принцип стал общепризнанным.

С появлением письменности возникла проблема записи чисел. Древние греки и евреи применяли алфавитную систему нумерации: числа от единицы до девяти, а затем все десятки и сотни обозначались буквами в порядке алфавита, над которыми ставилась черта. Создатели славянского письма перенесли этот прием на новую почву: знаки кириллицы, соответствовавшие греческим буквам, получили те же числовые значения (но алфавитный порядок при этом нарушился), сверху ставилось титло. Таким образом, приходилось запоминать 27 (проверьте) числовых знаков – цифр.

В Западной Европе вплоть до XVIII века в официальных документах применялась римская буквенная нумерация. Она использовала всего семь цифр: I – 1, V – 5, X – 10, L – 50, C – 100, D – 500, M – 1000. Число также записывалось в виде последовательности цифр, но из эстетических соображений запрещалось четырехкратное повторение одной и той же цифры. Так что числа 4, 9, 40, 90, 400, 900 обозначались соответственно как IV, IX, XL, XC, CD, CM – меньшая по значению цифра оказывалась левее большей (но часовщики упорно писали на циферблатах IIII, чтобы не путать с шестеркой VI). Римские цифры используются до сих пор в обозначениях дат и в порядковых номерах. Примеры: 31.XII, Сонет CCLXIX, Petro primo Catharina secunda MDCCLXXXII, XIV год Республики, Leonardo Eulero academia petropolitana MDCCCXXXVII, Anno Domini MCMXXII, XIV съезд ВКП(б), папа Иоанн XXIII, аккорд VI ступени (какие ассоциации вызывают у вас эти записи? Как выглядит в римской нумерации текущий год?).

В процессе счета возникли и основные арифметические действия над числами – сложение и умножение, были осознаны основные законы, которым они подчиняются.

Для сложения имеют место

1) закон ассоциативности (или сочетательности):

$$(a+b)+c = a+(b+c);$$

2) закон коммутативности (или перестановочности):

$$a+b = b+a.$$

Эти законы выражают тот очевидный факт, что если мы имеем несколько групп предметов, то для вычисления общего количества этих предметов все равно, с какой группы начинать пересчет и как объединять те или иные из этих групп.

Для умножения справедливы

3) закон ассоциативности:

$$(ab)c = a(bc);$$

4) закон коммутативности:

$$ab = ba;$$

5) закон нейтральности числа 1:

$$a \cdot 1 = a.$$

Наконец, сложение и умножение связывает

6) закон дистрибутивности (или распределительности):

$$a(b + c) = ab + ac.$$

Эти знакомые нам с детства правила арифметики были сформулированы в явном виде лишь в первой половине XIX века.

До появления современной системы записи чисел выполнение арифметических операций было затруднено, так что, например, перемножить две достаточно большие величины было задачей, доступной лишь узкому кругу лиц (когда в 1000 году папа Сильвестр II был заподозрен в связях с дьяволом, одним из поводов к этому послужили его выдающиеся вычислительные способности). Открытие позиционной системы счисления освободило умственную энергию человека от этой утомительной работы, сведя дело к освоению нескольких шаблонных приемов (алгоритмов). Вы не забыли, как это выглядело в школе? Перемножьте «столбиком», например, 1234 на 567.

Основная идея позиционной системы состояла в том, чтобы одной и той же цифре можно было придавать разные значения в зависимости от места (позиции), которое она занимает в записи числа. Так, в обозначении 666 первая шестерка выражает количество сотен, вторая – десятков, третья – единиц. Но как быть в том случае, если в составе числа отсутствуют единицы какого-то разряда? Как отличить запись числа шестьсот шесть от записи числа шестьсот шестьдесят или числа шестьдесят шесть (везде две шестерки)? Чтобы обойти эту трудность, был изобретен символ 0, которым стали обозначать пропуск в каком-либо разряде. Этот технический знак стал впоследствии восприниматься как число, выражающее отсутствие предметов интересующего нас (или вообще всякого) вида. В таком понимании 0 вступает в арифметические действия с другими числами, подчиняясь правилам

$$a + 0 = 0 + a = a$$

(нейтральность относительно сложения),

$$a \cdot 0 = 0 \cdot a = 0.$$

Заметим, что ни мифический Пифагор (VI в. до н.э.), ни загадочный Эвклид (III в. до н.э.), ни Архимед (287–212 до н.э.) нуля не знали.

Позиционная десятичная система счисления возникла в Индии в начале нашей эры и в конце первого тысячелетия стала распространяться в арабских странах, к XII веку достигнув Европы. В русских текстах «арабские» цифры появляются начиная с XVI века. В «Арифметике» Леонтия Магницкого (1703 г.), по которой учился Ломоносов, все вычисления ведутся уже в новых обозначениях, но номера страниц, условий задач и т.п. указываются еще по буквенной системе.

В позиционной десятичной системе степени основания (т.е. числа десять) называются разрядами: $10^0 = 1$ (нулевой разряд – единицы), $10^1 = 10$ (первый – десятки), $10^2 = 100$ (второй – сотни), $10^3 = 1000$ (третий – тысячи), $10^4 = 10000$ (четвертый – десятки тысяч) и т.д. Всякое число записывается в виде последовательности цифр. Цифра (от 0 до 9), стоящая в этой записи на i-м справа месте ($i = 0, 1, 2, \dots$), показывает, сколько входит в состав числа единиц i-го разряда. Таким образом,

$$365 = 3 \cdot 10^2 + 6 \cdot 10^1 + 5 \cdot 10^0,$$

$$1001 = 1 \cdot 10^3 + 0 \cdot 10^2 + 0 \cdot 10^1 + 1 \cdot 10^0.$$

Другими словами, каждое число x представляется в виде разложения по степеням десятки:

$$x = a_0 \cdot 10^n + a_1 \cdot 10^{n-1} + \dots + a_{n-1} \cdot 10^1 + a_n \cdot 10^0 \qquad (1)$$

и коэффициенты этого разложения $a_0, a_1, \dots, a_{n-1}, a_n$ являются последовательными цифрами в записи числа.

Если бы наша система была не десятичной, а пятеричной, т.е. если бы основным числом счета была не десятка, а пятерка (пальцы одной руки), позиционный принцип позволил бы записать каждое число с помощью пяти цифр, разлагая его по степеням пятерки в соответствии с формулой (1). Принимая в качестве пятеричных цифр $0, 1, 2, 3, 4$, мы получили бы, например, для числа дней в году

$$365_{10} = 2 \cdot 5^3 + 4 \cdot 5^2 + 3 \cdot 5^1 + 0 \cdot 5^0 = 2430_5$$

(индекс указывает, в какой системе рассматривается число), а для «числа Шехерезады»

$$1001_{10} = 1 \cdot 5^4 + 3 \cdot 5^3 + 0 \cdot 5^2 + 0 \cdot 5^1 + 1 \cdot 5^0 = 13001_5.$$

С другой стороны,

$$1001_5 = 1 \cdot 5^3 + 0 \cdot 5^2 + 0 \cdot 5^1 + 1 \cdot 5^0 = 126_{10}.$$

Для обозначения чисел в двадцатеричной системе (основное число двадцать – все пальцы рук и ног) потребовалось бы двадцать цифр. В ней десятичные числа 365 и 1001 записывались бы короче: соответственно двумя и тремя знаками, ибо $365 = 18 \cdot 20^1 + 5 \cdot 20^0$ и $1001 = 2 \cdot 20^2 + 10 \cdot 20^1 + 1 \cdot 20^0$ (коэффициенты 18 и 10, конечно, должны быть заменены соответственно девятнадцатой и одиннадцатой двадцатеричными цифрами).

Еще пример:

$$1111_{20} = 1 \cdot 20^3 + 1 \cdot 20^2 + 1 \cdot 20^1 + 1 \cdot 20^0 = 8421_{10}.$$

Из систем с основанием, отличным от десяти, наиболее распространена (в практике, связанной с ЭВМ) двоичная система счисления. В ней всего две цифры – 0 и 1, и двоичная запись чисел получается по формуле (1) при замене основания 10 на основание 2. Например,

$$365_{10} = 1 \cdot 2^8 + 0 \cdot 2^7 + 1 \cdot 2^6 + 1 \cdot 2^5 + 0 \cdot 2^4 +$$

$$+1 \cdot 2^3 + 1 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0 = 101101101_2.$$

Для перехода от обычного представления чисел к их двоичной форме и наоборот нужно знать степени двойки: $2^0 = 1$, $2^1 = 2$, $2^2 = 4$, $2^3 = 8$, $2^4 = 16$, $2^5 = 32$, $2^6 = 64$, $2^7 = 128$, $2^8 = 256$, $2^9 = 512$, $2^{10} = 1024$ и т.д. В качестве упражнений найдите двоичную запись всех чисел от 0 до $16 (= 10000_2)$, числа 1001 и десятичную запись двоичного числа 1010011010. Числа в двоичном представлении чрезмерно длинны и удручающе однообразны.

Школьные алгоритмы сложения и умножения «столбиком» справедливы во всех позиционных системах счисления независимо от их основания. Доставьте себе удовольствие выполнить в двоичной системе умножение числа 11011 на число 101 (помните, что $1 + 1 = 10$).

Практическая деятельность человека и развитие его интеллекта раздвигали границы окружающего мира и вычленяли в нем все новые подробности. Вместе с этим возникали все бо́льшие числа: натуральный ряд то медленно, то сразу скачками удлинялся.

Алфавитные системы нумерации позволяли непринужденно обращаться с числами первой тысячи, а при помощи дополнительных знаков – в пределах десяти тысяч (это было последнее число, имевшее у греков свое имя – мириада). Классическая древность и не сталкивалась с необходимостью заглядывать дальше этой границы в каких-либо реальных или теоретических ситуациях. Неопределенные библейские выражения типа «тысячи тысяч», «тьма тем» (Дан. 7,10), «легион» (Лук. 8,30) или обыденное «как песчинок» отодвигали числовой горизонт в некоторую загадочную даль, невыразимую в конкретных количествах. В заметке «Псаммит» (т.е. исчисление песка) Архимед показал, как можно систематически строить и называть сколь угодно большие числа. В частности, размещая в маковом зерне 10 000 (мириада) песчинок, он находит, что во Вселенной (шар диаметром в мириаду диаметров Земли) поместилось бы (в наших обозначениях) не более чем 10^{63} песчинок. Любопытно, что современные подсчеты количества атомов в видимой Вселенной приводят к числу 10^{67} (всего в мириаду раз больше).

На Руси число 10 000 называлось «тьма» и тоже служило последним пределом естественного, т.е. соотносимого с какой-либо деятельностью счета. Впрочем, еще в XII веке новгородский дьякон Кирик в своем «Учении, им же ведати человеку числа всех лет» подсчитал, что в прошедших от сотворения мира 6644 годах содержится: месяцев – 79 728, недель – 346 673, дней – 2 426 721. В умозрительных построениях (так называемый «великий счет») под «тьмой» понимался нынешний миллион 10^6. Далее шли: легион 10^{12} (именно – «тьма тем»), леодр 10^{24}, ворон 10^{48} и, наконец, колода 10^{96} («более же сего не бывает»). Колода в полной записи выглядит как единица с девяносто шестью нулями. (Как назвать колоду в терминах воронов?)

Великий Архимед убедил, что он в состоянии указать некоторые числа, превосходящие число песчинок в объеме всей Вселенной. Но воображение его остановилось на жутком образе мира, утонувшего в пыли. Точно так же и безвестный служитель «цыфирной науки» ограничил полет своей терминологической фантазии колодой, устояв перед соблазном рассмотреть, скажем, «легион колод». Математики не хотели изобретать большие числа свыше количества их, необходимого для тех или иных конкретных нужд. Натуральный ряд мыслился лишь потенциально бесконечным, т.е. неограниченно продолжаемым, а не существующим актуально, в качестве завершенного объекта. Считалось, что мы создаем все новые натуральные числа, а не открываем их, как острова в безбрежном океане. На противоположной точке зрения стоял святой Августин (354–430), обличавший своих оппонентов в том, что они считают, «будто бесконечность превышает знание Господне». С конца прошлого века математики постепенно склонялись к признанию бесконечных множеств как актуально существующих – независимо от того, описан ли как-нибудь способ их образования. В современной математической практике эта точка зрения возобладала, но не все ее разделяют, и теоретические дискуссии об актуальной и потенциальной бесконечности продолжаются (а как вы воспринимаете натуральный ряд?).

Мостом между двумя пониманиями натурального ряда выступает

Аксиома индукции. Любое множество натуральных чисел A, обладающее следующими свойствами:

а) 1 принадлежит A,

б) если число n принадлежит A, то и следующее за ним число $n+1$ также принадлежит A, –

совпадает со всем натуральным рядом $\mathbb{N}$.

(Аксиома – это математическая истина, принимаемая без доказательства, в некотором смысле акт веры. Индукция – переход от частного к общему). Для обозначения множества всех натуральных чисел всюду в дальнейшем будем использовать введенный символ $\mathbb{N}$.

В 1955 году английский математик Скьюз показал, что существует натуральное число x, обладающее некоторым важным свойством (детали для нас несущественны), и что оно не превосходит величины $10^{10^{10^{964}}}$. Число $10^{10^{10^{964}}}$ в настоящий момент является наибольшим натуральным числом, использованным для какой-либо практической цели. Его полная запись представила бы собой единицу с количеством нулей, заполняющим многие тома. И архимедово число песчинок, и число атомов во Вселенной, и даже «великое славянское число» колода несопоставимы с этим монстром, обозначающим сегодняшнюю границу потенциально бесконечного натурального ряда.

Главная книга христианского мира Библия в полной мере отражает ту роль, которую играли разнообразные вычисления в жизни наших далеких предков. В пятой главе книги «Бытие» указываются потомки Адама от Сифа до Ноя. Стандартная конструкция этой главы имеет следующий вид:

«25 Мафусаил жил сто восемьдесят семь лет, и родил Ламеха.

26 По рождении Ламеха, Мафусаил жил семьсот восемьдесят два года, и родил сынов и дочерей.

27 Всех же дней Мафусаила было девятьсот шестьдесят девять лет; и он умер».

Согласимся, что арифметические примеры типа $187 + 782 = 969$ никак нельзя считать тривиальными для времени создания Библии. В главе 11 «Бытия» после рассказа о крушении Вавилонской башни приводится список потомков Сима, старшего сына Ноя, однако здесь общая схема дается в усеченном виде:

«24 Нахор жил двадцать девять лет, и родил Фарру.

25 По рождении Фарры, Нахор жил сто девятнадцать лет, и родил сынов и дочерей».

Финальный возраст Нахора не указывается, интересующемуся придется искать сумму $29 + 119$.

В книге «Числа» (название говорит само за себя) приводятся статистические сведения о числе всех сынов Израилевых, годных для войны, во всех коленах, о распределении воинов каждого колена по станам. Здесь уже приходится иметь дело с величинами вроде 46 500, 59 300, 64 400, а общее количество всех военнообязанных достигает 603 550 при первом обследовании (глава 1) и 601 730 при втором (глава 26). В главе 31 (стихи 26-47) рассматривается сложный пример деления военной добычи. Он не вполне завершен и мог бы послужить предметом интересных обсуждений.

Разнообразные подсчеты и измерения проводятся в книгах Иисуса Навина (глава 21), 1-й Паралипоменон (главы 12, 15), Ездры (главы 1, 2, 8), Неемии (глава 7), Иеремии (глава 52), Иезекииля (глава 40). Наибольшее конкретное число указывается во второй книге Царств (глава 24): 800 000. И, конечно, нельзя не упомянуть об Откровении святого Иоанна Богослова (Апокалипсис), где в заключительном стихе главы 13 указывается «число зверя»: 666 (или римскими цифрами: DCLXVI – шесть разных цифр в правильном порядке! Кроме того, 666 – это сумма первых 36 натуральных чисел). В различные исторические периоды пытливые умы, применяя реальные или изобретенные к случаю алфавитные нумерации, пытались разоблачить тех или иных деятелей путем «расшифровки» их имен и титулов так, чтобы получилось роковое число. Вот и Пьер Безухов («Война и мир», т. 3, часть 1, глава XIX), приписав числовые значения буквам французского алфавита, нашел, что L'empereur Napoléon дает 666 «и что поэтому Наполеон и есть тот зверь, о котором предсказано в Апокалипсисе». Несмотря на некоторые погрешности (пропуск буквы j в алфавите, арифметическая ошибка, исправляя которую, приходится писать Le empereur), апокалиптические вычисления Пьера, троекратно приводящие к числу 666, изумляют. Перечитайте это место у Толстого.

Наряду с задачами прикладного характера (статистические данные, распределение военных трофеев, подсчет числа недель, прошедших от сотворения мира, или песчинок в объеме Вселенной) древние математики рассматривали и проблемы совсем другого рода – относящиеся, так сказать, к «чистой» науке.

Много таких задач было связано с простыми числами – так называются натуральные числа, которые не имеют других делителей, кроме 1 и себя (2, 3, 5, 7, 11, 13, 17, 19, 23, 29 – первые десять простых). Каждое натуральное число представляется в виде произведения простых чисел, так что простые числа – это в некотором смысле атомы натурального ряда относительно умножения (по сложению атом один – единица). Эвклид доказал, что простых чисел бесконечно много (в его формулировке: «больше любого предложенного числа их», – он не признавал актуальной бесконечности).

Теорема. Множество простых чисел бесконечно.

ДОКАЗАТЕЛЬСТВО. Допустим, что простых чисел конечное число и p – наибольшее из них. Перемножим все простые числа: $2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot \dots \cdot p$ и обозначим это произведение через P. Тогда число $P + 1$ не делится ни на одно из указанных простых чисел, так как в остатке каждый раз получается 1. Значит, $P + 1$ либо само будет простым числом, бóльшим p, либо представляется в виде произведения каких-то простых чисел, не входящих в наш список. И то и другое невозможно в силу исходного предположения. Таким образом, из гипотезы о конечности числа простых чисел получилось противоречие. Значит, эта гипотеза ложна и истинным будет противоположное ей утверждение о бесконечности множества простых чисел. □

Здесь применен метод доказательства от противного – один из основных приемов установления истины в математике. Желая убедиться в справедливости некоторого утверждения П, мы берем противоположное ему не-П и логически строгими рассуждениями получаем из него некоторое абсурдное следствие (reductio ad absurdum – приведение к нелепости). Отсюда заключаем, что не-П ложно, а значит, истинно П – согласно принципу исключенного третьего: из двух противоположных высказываний верно либо одно, либо другое, третьего не дано, tertium non datur.

(Понятно ли вам доказательство знаменитой теоремы Эвклида, его логическая схема, смысл полученного в ходе рассуждений противоречия? Слегка перефразируя Литлвуда, заметим: «Опыт показывает, что некоторые непрофессионалы понимают доказательство Эвклида; с другой стороны, опыт показывает, что некоторые его не понимают,– это не должно их огорчать».)

В 1742 году петербургский академик Христиан Гольдбах в письме к Эйлеру, жившему тогда в Берлине, высказал гипотезу о том, что всякое нечетное число, начиная с 7, может быть представлено в виде суммы трех простых чисел. В течение почти двухсот лет эта проблема привлекала внимание выдающихся математиков, но не поддавалась их усилиям. Одним из крупнейших достижений теории чисел стала теорема, доказанная в 1937 году академиком И.М.Виноградовым: всякое достаточно большое нечетное число в самом деле представляется суммой трех простых чисел. Последующие исследования показали, что «достаточно большое» означает «больше, чем $C = 3^{3^{15}} = 3^{14348907}$». Так что для окончательного решения осталось проверить гипотезу для всех чисел, меньших C, но это уже не принципиально. (Иван Матвеевич Виноградов (1891–1983) получил свои замечательные теоретико-числовые результаты с помощью совершенно новых методов, которые оказались применимыми к весьма широкому кругу задач. С 1932 г. до своей кончины он возглавлял Математический институт Академии наук СССР, главный центр математических исследований страны).

К числу нерешенных до сих пор задач относится проблема «близнецов» – так называются простые числа, отличающиеся друг от друга на 2 (в первом десятке простых такими парами будут

3 и 5, 5 и 7, 11 и 13, 17 и 19). Неизвестно, оборвется ли когда-нибудь этот список или же он бесконечен, как и ряд простых чисел.

Во всех областях знания, в том числе и в математике, современные методы ни в какой мере не сравнимы с возможностями, которыми располагали ученые ранних эпох развития нашей цивилизации. Но вот есть же понятные всем задачи, которые были неприступными тогда и остаются непобежденными теперь, несмотря на систематические изощренные атаки. И, заметим, эти задачи касаются натуральных чисел – той области, с которой началась математика.

Часто спрашивают: а зачем решать такие задачи? Какие полезные следствия вызовет, скажем, установление истины в проблеме о количестве простых чисел-близнецов? Подобные сомнения возникают в каждой области творческой деятельности человека («Кому нужна т а к а я музыка, т а к а я живопись? Для кого пишутся т а к и е стихи?»). Оправдываясь, специалисты, работающие в теории чисел, обычно приводят слова Леонарда Эйлера (1707–1783): «Математика, вероятно, никогда не достигла бы такой великой степени совершенства, если бы древние не приложили столько усилий для изучения проблем, которыми сегодня многие пренебрегают из-за их мнимой бесплодности». Как часто новые методы, новая техника, новая форма, возникшие при решении, казалось бы, частных задач, приводили науку на новый, более высокий уровень развития! Точно так же обстоит дело и в искусстве (приведите примеры).

Здесь останавливается наше движение по натуральному ряду. Не слишком ли много внимания мы уделили начальным шагам в математику? Ответом на это мог бы послужить известный афоризм немецкого математика Леопольда Кронекера (1823–1891): «Бог создал натуральные числа, все остальное – дело рук человеческих».

§ 2. Кольцо целых чисел

С появлением позиционной системы счисления к натуральным числам присоединился ноль, который стал восприниматься как обозначение отсутствия какого-либо количества. Принцип расширения имеющегося числового множества многократно использовался в математической практике. С его помощью над фундаментом – натуральным рядом – постепенно выросло огромное здание, символизирующее все оттенки современного представления о числе.

Если ноль был введен для преодоления некоторых трудностей, связанных с созданием нового, универсального способа нумерации, то последующие обобщения понятия о числе возникли в основном на алгебраической почве. Можно сказать, что арифметика изучает свойства чисел и, в частности, производимые над ними операции, алгебра же изучает свойства операций, производимых над произвольными объектами и, в частности, над числами. Законы ассоциативности и коммутативности, которым подчиняются сложение и умножение чисел, в равной мере относятся и к арифметике, и к алгебре.

Операция сложения соответствует практическому действию «прибавить нечто к имеющемуся количеству», например, к трем апельсинам еще два. С ней связана обратная операция – вычитание, имеющая в основе не менее естественное действие – «отнять». Однако вычитание выполнимо не во всех случаях: если некто имеет три апельсина, то отнять у него пять апельсинов невозможно. Указанная неполноценность вычитания станет особенно очевидной, когда мы перейдем на язык уравнений. Если имеется a предметов, то сколько еще нужно прибавить к ним, чтобы получить b предметов? Другими словами,чему равен x, если $a + x = b$? Решение $x = b - a$ не имеет смысла в множестве натуральных чисел $\mathbb{N}$, если a больше b. Значит, нужно в указанном уравнении добавлять условие «при $a < b$». Чтобы избавиться от такого рода ограничений (в большом количестве они совершенно затемняют смысл производимых действий), были введены отрицательные числа -1, -2, -3 и т.д. Вместе с натуральными числами и нулем эти новые числа образуют множество $\mathbb{Z}$ всех целых чисел. В отличие от натурального ряда $\mathbb{N}$, в множестве $\mathbb{Z}$ уравнение $a + x = b$ всегда разрешимо: при $a < b$ решением будет натуральное число $b - a$, при $a = b$ получаем $x = 0$, при $a > b$ решение $x = b - a$ оказывается отрицательным целым числом, противоположным натуральному числу $a - b$.

Если 0 долгое время не признавался числом, то отрицательные числа при своем появлении встретили еще более решительное неприятие. Обыденный разум не допускал этих «ложных» чисел, по своему смыслу меньших нуля, т.е. «меньших, чем ничто». Полное признание наступило лишь в XVII веке, когда отрицательные числа были осознаны в практической деятельности как показатели убытка (в противоположность прибыли), движения вспять, т.е. в направлении, обратном заданному, уровней ниже фиксированного и т.п. Но в естественные языки явные названия отрицательных чисел («минус десять», «минус двадцать пять») вошли разве лишь в качестве температурных показаний.

Основным понятием алгебры является операция. Говорят, что на множестве объектов A задана операция, если каждой паре объектов (a, b) из A однозначно сопоставлен третий объект c, тоже принадлежащий множеству A.

Например, сложение и умножение являются операциями в множестве $\mathbb{N}$ всех натуральных чисел, а вычитание операцией в этом множестве не будет, так как не для каждой пары натуральных чисел (a, b) определена разность $a - b$. В множестве $\mathbb{Z}$ всех целых чисел вычитание становится полноправной операцией.

Конкретные операции обозначаются индивидуальными знаками: $+, \cdot, -$ и т.п. В общих рассуждениях результат операции, примененной к элементам (объектам) a, b, будем обозначать через $a \circ b$ (а читать, для краткости, будем «a умножить на b»).

Операция $\circ$ в множестве A называется ассоциативной, если $(x \circ y) \circ z = x \circ (y \circ z)$ для любых x, y, z из A. Умножение и сложение в $\mathbb{N}$ и в $\mathbb{Z}$ ассоциативны, операция вычитания в $\mathbb{Z}$ этими свойствами не обладает: $(3 - 2) - 1 = 0$, но $3 - (2 - 1) = 2$.

Операция $\circ$ в множестве A называется коммутативной, если $x \circ y = y \circ x$ для любых x, y из A. Это тождество выполняется для сложения и умножения в $\mathbb{N}$ и в $\mathbb{Z}$, но не выполняется для вычитания (например, $2 - 1 = 1$, но $1 - 2 = -1$).

Теперь введем одно из важнейших понятий современной математики.

Множество A называется группой, если в нем определена ассоциативная операция $\circ$ такая, что все уравнения вида $a \circ x = b$ и $y \circ a = b$ разрешимы в A, т.е. для любых a и b из A существуют в A элементы u и v, удовлетворяющие равенствам $a \circ u = b$ и $v \circ a = b$. Группа называется коммутативной, если операция $\circ$ коммутативна.

Поскольку сложение целых чисел ассоциативно и коммутативно (т.е. в $\mathbb{Z}$ выполняются тождества $(x + y) + z = x + (y + z)$ и $x + y = y + x$) и все уравнения вида $a + x = b$ разрешимы (решением будет $x = b - a$), то множество $\mathbb{Z}$ всех целых чисел образует относительно операции сложения коммутативную группу. Группа $(\mathbb{Z}, +)$ называется аддитивной группой целых чисел (от латинского additio – сложение).

Проверку того, что данное множество A с ассоциативной операцией $\circ$ является группой, можно существенно упростить. Сначала дадим необходимые для этого определения.

Элемент e множества A называется нейтральным относительно операции $\circ$, если $x \circ e = e \circ x = x$ для любого x из A. Например, число 0 будет нейтральным элементом относительно сложения в $\mathbb{Z}$, а число 1 – нейтральным элементом относительно умножения в $\mathbb{N}$ и в $\mathbb{Z}$ (мы уже выписывали соответствующие тождества).

Элемент $\bar{a}$ из A называется обратным для элемента a относительно операции $\circ$, если $a \circ \bar{a} = \bar{a} \circ a = e$. В $\mathbb{Z}$ последние равенства для сложения приобретают вид $a + \bar{a} = \bar{a} + a = 0$, и мы видим, что $\bar{a}$ – это не что иное, как противоположное для a число, т.е. $\bar{a} = -a$. Для умножения в $\mathbb{N}$ и в $\mathbb{Z}$ с обратными элементами дело обстоит неважно: если $a\bar{a} = \bar{a}a = 1$, то это может быть только когда $a = \bar{a} = 1$, т.е. только число 1 имеет в $\mathbb{N}$ обратный элемент, и им будет сама единица 1.

<u>Теорема 1</u>. Пусть на множестве A определена ассоциативная операция $\circ$, относительно которой в A существует нейтральный элемент e и каждый элемент a имеет обратный для него элемент $\bar{a}$. Тогда $(A, \circ)$ является группой.

ДОКАЗАТЕЛЬСТВО. Нужно установить, что все уравнения вида $a \circ x = b$ и $y \circ a = b$ разрешимы в A при любых a и b из A. Следующая цепочка равенств показывает, что $u = \bar{a} \circ b$ будет решением первого уравнения:

$$a \circ u = a \circ (\bar{a} \circ b) = (a \circ \bar{a}) \circ b = e \circ b = b$$

(во втором переходе мы воспользовались ассоциативностью операции $\circ$, в третьем – тем, что $\bar{a}$ является обратным для a элементом, в четвертом – нейтральностью элемента e). Аналогично доказывается, что $v = b \circ \bar{a}$ будет решением уравнения $y \circ a = b$ (проделайте соответствующие выкладки). Итак, в $(A, \circ)$ разрешимы все уравнения вида $a \circ x = b$ и $y \circ a = b$, и значит, $(A, \circ)$ – группа. $\square$

Можно показать, что справедлива и обратная теорема: если $(A, \circ)$ – группа, то в ней имеется нейтральный элемент и все элементы обладают обратными.

Итак, для того чтобы проверить, что множество A с ассоциативной бинарной операцией $\circ$ является группой, достаточно убедиться, что в A есть нейтральный относительно этой операции элемент e и что для каждого элемента a из A существует в A обратный для него элемент $\bar{a}$.

Важное значение имеет следующая

Теорема 2. В каждой группе $(A, \circ)$ любое из уравнений вида $a \circ x = b$ и $y \circ a = b$ имеет единственное решение.

ДОКАЗАТЕЛЬСТВО. В доказательстве теоремы 1 было установлено, что $u = \bar{a} \circ b$ является решением уравнения $a \circ x = b$. Пусть x_0 – произвольное решение этого уравнения, т.е. $a \circ x_0 = b$. Тогда

$$x_0 = e \circ x_0 = (\bar{a} \circ a) \circ x_0 = \bar{a} \circ (a \circ x_0) = \bar{a} \circ b,$$

т.е. $x_0 = \bar{a} \circ b$, других решений нет (объясните каждый из переходов в выписанной цепочке равенств). Аналогично доказывается. что $v = b \circ \bar{a}$ будет единственным решением уравнения $y \circ a = b$. $\Box$

Аддитивная группа целых чисел бесконечна в смысле количества составляющих ее элементов (целых чисел). Но существуют и совсем маленькие группы. Если взять всего одно число 0 и рассмотреть его вместе с операцией сложения (или умножения), то, как нетрудно заметить, получится группа. Такие группы – содержащие один элемент – называются единичными или тривиальными группами.

На множестве Z_2, состоящем из двух чисел 0 и 1, определим операцию $\oplus$, полагая

$$0 \oplus 0 = 1 \oplus 1 = 0, \;\; 0 \oplus 1 = 1 \oplus 0 = 1, \; -$$

так называемое «сложение по модулю 2». Несложная, но несколько утомительная проверка показывает, что эта операция ассоциативна: нужно убедиться в справедливости восьми равенств, получающихся из формулы $(x \oplus y) \oplus z = x \oplus (y \oplus z)$, если переменным x, y, z придавать значения 0 и 1. По своему определению операция $\oplus$ коммутативна. Нейтральным элементом является 0, противоположным для каждого элемента будет он сам. Таким образом, $(Z_2, \oplus)$ – коммутативная группа, содержащая всего два элемента.

Другой простой пример. Рассмотрим множество Z^*, состоящее из двух чисел 1 и -1 вместе с обычной операцией умножения: $1 \cdot 1 = (-1) \cdot (-1) = 1, \; 1 \cdot (-1) = (-1) \cdot 1 = -1$ (так называемое правило знаков, регулирующее умножение целых чисел. «Минус на минус дает плюс» и в наши дни представляется школьникам непостижимой тайной. Великий Эйлер безуспешно пытался дать какое-нибудь «естественное» толкование этому математическому соглашению).

Умножение чисел ассоциативно, коммутативно и имеет нейтральным элементом 1, так что в нашем случае остается решить вопрос с обратными элементами. Но, как видим, и здесь каждый элемент будет сам для себя обратным (слово «противоположный» употребляется, когда операция обозначается знаком + или похожим на него). Итак, мы имеем еще одну двухэлементную группу: $(Z^*, \cdot)$.

По образцу школьных таблиц сложения и умножения составим таблицы для операций в группах Z_2 и Z^*:

$\oplus$	0	1
0	0	1
1	1	0

$\cdot$	1	-1
1	1	-1
-1	-1	1

Неправда ли, они чем-то похожи? В самом деле, если во второй таблице заменить 1 на 0, а -1 на 1 и вместо точки, обозначающей операцию умножения, выставить знак $\oplus$, то получится в точности таблица, определяющая группу Z_2. Таким образом, группа $(Z^*, \cdot)$ – это не что иное, как группа $(Z_2, \oplus)$, но только с другими обозначениями элементов и другим знаком операции. В математике такие группы называются изоморфными. Точнее: две группы, по определению, изоморфны, если между их элементами можно установить попарное соответствие, переводящее операцию одной группы в операцию другой. Это соответствие называется изоморфизмом. В нашем примере изоморфизмом является указанное выше соответствие $0 \mapsto 1,\ 1 \mapsto -1$.

Имея этот образец, попробуйте установить, изоморфны ли группы $(Z_4, \oplus)$ и $(K_4, \cdot)$, представленные следующими таблицами операций:

$\oplus$	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

$\cdot$	e	a	b	c
e	e	a	b	c
a	a	e	c	b
b	b	c	e	a
c	c	b	a	e

Есть прямой путь «лобовой атаки»: придавать буквам e, a, b, c значения 0,1,2,3 в разных вариантах (их в данном случае 24) и каждый раз получающуюся таблицу, упорядочивая строчки, сравнивать с таблицей группы Z_4. Если на каком-то шаге получится совпадение – группы изоморфны, а если не получится никогда – не изоморфны. Но обычно пытаются сначала усмотреть какое-нибудь различие в устройстве таблиц. И в самом деле, в группе K_4

«квадраты» всех элементов равны: $ee = aa = bb = cc = e$, в группе же Z_4 это не так: $0 \oplus 0 = 2 \oplus 2 = 0$, но $1 \oplus 1 = 3 \oplus 3 = 2$. Значит, Z_2 и K_4 – неизоморфные группы.

Изоморфные группы в алгебре не различают, считая их просто разными реализациями одной и той же «абстрактной» группы (так же, как «два камня», «две птицы» и «две руки» – это разные конкретные представления абстрактного числа 2).

Идея изоморфизма, возникшая в математике в середине XIX века при изучении групп, со временем приобрела универсальный характер. Изоморфизм в широком понимании – это попарное соответствие между двумя системами, выражающее «одинаковость» их строения. Обнаружив, что две системы изоморфны, мы с удовлетворением говорим: «Оказывается, это одно и то же!» – ибо, зная структуру одной, можем перенести все связанные с этим достижения на другую. В химии под изоморфизмом понимается свойство веществ кристаллизоваться в одинаковых формах, в лингвистике этим термином обозначают параллелизм в организации звуковой и смысловой сторон языка, слово это все чаще используется в философских и искусствоведческих работах, входит в число общекультурных понятий.

Множество $\mathbb{Z}$ всех целых чисел является коммутативной группой по сложению. Но кроме этой операции, в $\mathbb{Z}$ имеется еще умножение – ассоциативная и коммутативная операция, дистрибутивная (распределительная) относительно сложения: $x(y + z) = xy + xz$. Система $(A, +, \cdot)$ такая, что $(A, +)$ – коммутативная группа, а операция $\cdot$ ассоциативна и дистрибутивна относительно операции $+$, называется кольцом. Следовательно, $(\mathbb{Z}, +, \cdot)$ – кольцо. Это кольцо целых чисел. Если в группе Z_2 наряду со сложением по модулю 2 рассмотреть еще обычное числовое умножение, то тоже получится кольцо, в нем всего два элемента: 0 и 1. Приведите доводы, убеждающие в том, что четные целые числа (0, 2, -2, 4, -4, 6, -6 и т.д.) образуют кольцо относительно обычных сложения и умножения. А как обстоит дело с нечетными числами?

Для колец, как и для групп, вводится понятие изоморфизма, и изоморфные кольца в соответствии с обычным алгебраическим принципом не различаются.

Пусть m – некоторое натуральное число, $m > 1$. На множестве Z_m, состоящем из чисел $0, 1, 2, \ldots, m - 1$, введем операцию «сложения по модулю m» (знак $\oplus$), понимая под $x \oplus y$ остаток от деления обычной суммы $x + y$ на m (сложение по модулю 2 – частный случай). Аналогично определяется и операция «умножения по модулю m»: $x \otimes y$ – это остаток от деления числа xy на m. Можно показать, что $(Z_m, \oplus, \otimes)$ является кольцом – кольцом остатков от деления на m. Например, в Z_3 (т.е. при $m = 3$) таблицы сложения и умножения имеют вид:

$\oplus$	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

$\otimes$	0	1	2
0	0	0	0
1	0	1	2
2	0	2	1

Удивительное на первый взгляд равенство $2 \oplus 2 = 1$ объясняется просто: $2 \oplus 2$, по определению, представляет собой остаток от деления на 3 числа $2 + 2 = 4$, а это и есть 1. Точно так же $2 \otimes 2 = 1$. Составьте таблицы умножения в кольцах Z_2 и Z_4 (таблицы сложения по модулям 2 и 4 у нас уже встречались), таблицы сложения и умножения в кольцах Z_5 и Z_6. Кольца вида Z_m – это арифметические миры, лежащие, в общем, вне нашего непосредственного опыта, связанного с кольцом $\mathbb{Z}$. Например, в Z_4 будет $2 \otimes 2 = 0$ – сомножители отличны от нуля, а произведение их – ноль. В $\mathbb{Z}$ это невозможно.

Если идея изоморфизма отражает тот всем известный по опыту факт, что одна и та же сущность может представать в разительно несхожих внешних формах, то понятие гомоморфизма происходит от известного приема изучать явления в главном, отвлекаясь от несущественных (или не интересующих в данном контексте) деталей.

Предположим, что некто, рассматривая кольцо целых чисел $\mathbb{Z}$, обращает внимание только на два свойства: четность и нечетность числа (или, скажем так, различает числа только по этим признакам, так что, например, 0, 2, -4 для него одно и то же: четное число, а 1 и 2 – разные объекты). При таком подходе сложение и умножение чисел предстают в следующем виде: четное+четное=нечетное+нечетное=четное, четное+нечетное=нечетное; четное на четное=четное на нечетное=четное, нечетное на нечетное=нечетное. Если для краткости вместо «четное» писать 0, а вместо «нечетное» 1, то бесконечные таблицы сложения и умножения целых чисел предстанут перед нашим наблюдателем в форме

+	0	1
0	0	1
1	1	0

$\cdot$	0	1
0	0	0
1	0	1

Но ведь это не что иное, как сложение и умножение по модулю 2, т.е. операции в кольце Z_2! Таким образом, отвлекаясь от всех индивидуальных свойств чисел, кроме их четности или нечетности, мы увидим бесконечное кольцо $\mathbb{Z}$ как двухэлементное кольцо Z_2. Этот огрубленный до предела образ, однако, полностью характеризует взаимодействие свойств четности и нечетности относительно сложения и умножения чисел.

В современной терминологии кольцо Z_2 является гомоморфным образом кольца $\mathbb{Z}$, а отождествление всех четных чисел с числом 0 и всех нечетных чисел с числом 1 есть гомоморфизм кольца $\mathbb{Z}$ на кольцо Z_2.

Вообще, любое из колец Z_m представляет собой гомоморфный образ кольца целых чисел $\mathbb{Z}$. Как бы вы описали гомоморфизм кольца $\mathbb{Z}$ на кольцо Z_3 (остатки от деления на 3)? Какие целые числа этот гомоморфизм отождествляет с числом 0, с числом 1, с числом 2?

Историки, описывающие жизнь средневекового города, юристы, составляющие уголовный кодекс, психологи, разрабатывающие теорию личности,– список можно продолжать без конца,– все они, по существу, ищут гомоморфные образы сложнейшей системы – человеческого общества, изучая его с тех или иных частных позиций.

Самой знаменитой задачей, относящейся к целым числам, является так называемая Великая теорема Ферма. Французский юрист Пьер Ферма́ (1601-1665) в часы, свободные от основной деятельности, занимался математикой. Читая латинский перевод древней «Арифметики» Диофанта, Ферма писал на полях книги свои комментарии. Особое его внимание привлекло то место, где обсуждается уравнение $x^2 + y^2 = z^2$. Целые числа x, y, z, удовлетворяющие ему, называются пифагоровыми тройками (например, $3^2 + 4^2 = 5^2, 5^2 + 12^2 = 13^2$), их бесконечно много. Ферма отметил, что, в противоположность этому, уравнение $x^n + y^n = z^n$, где показатель степени n больше 2, решений в ненулевых целых числах x, y, z не имеет: «Я нашел удивительное доказательство, но здесь его не уместить». Более трех столетий никто не мог ни доказать, ни опровергнуть сформулированное Ферма утверждение. Пытаясь решить эту проблему, выдающиеся умы создали новые разделы математики, открыли методы, с помощью которых нашли ответы на многие важные теоретические и прикладные вопросы, но Великая теорема Ферма по-прежнему оставалась загадкой. Она получила широчайшую известность, стала своего рода символом несовершенства человеческого разума. К началу 1990-х годов было, в частности, установлено, что уравнение $x^n + y^n = z^n$ не имеет целочисленных решений при всех показателях степени n в пределах $2 < n < 100000$. Неприступность задачи, казалось, должна была создать твердое убеждение, что дилетантам в этой области делать нечего. Тем не менее люди самых разных профессий упорно продолжали поиски доказательства или опровергающего примера. Среди заявлявших о своей победе «ферматистов» известны учителя, военнослужащие, музыканты, философы, священники,– впрочем, и сам Ферма, как уже отмечалось, размышлял на математические темы лишь в перерывах между судебными заседаниями.

В мае 1995 года английский математик Эндрю Джон Уайлс опубликовал доказательство теоремы Ферма, которое он получил, используя самые современные методы. Но его рассуждения доступны только очень узкому кругу специалистов, так что должно пройти какое-то время, прежде чем широкая общественность смирится с фактом падения неприступной интеллектуальной крепости.

Теперь рассмотрим тип уравнений, сыгравших основную роль в развитии алгебры и в классификации чисел. Уравнение вида

$$a_0 x^n + a_1 x^{n-1} + \cdots + a_{n-1} x + a_n = 0, \tag{1}$$

где $a_0, a_1, \ldots, a_{n-1}, a_n$ – целые числа, называется алгебраическим уравнением с целыми коэффициентами. Число n (старшая из степеней неизвестного x) называют степенью уравнения.

При $n = 1$ получаются линейные уравнения, их традиционная общая запись

$$ax + b = 0.$$

При $n = 2$ имеем квадратные уравнения:

$$ax^2 + bx + c = 0,$$

при $n = 3$ – кубические:

$$ax^3 + bx^2 + cx + d = 0.$$

В уравнении (1) те или иные коэффициенты (кроме старшего a_0) могут отсутствовать, т.е. равняться нулю. Число x_0, которое, будучи подставленным в (1) вместо x, превращает уравнение в тождество, называется его корнем. Проверьте, например, что корнем линейного уравнения $x + 3 = 0$ будет число $x_0 = -3$, у квадратного уравнения $x^2 - 3x + 2 = 0$ корнями будут $x_1 = 1$, $x_2 = 2$, кубическое уравнение $x^3 + x - 2 = 0$ одним из корней имеет $x_0 = 1$. Решить уравнение (1) значит найти все его корни.

Древние греки не знали буквенных обозначений для неконкретных чисел и поэтому не имели возможности записывать формулы. Они представляли себе числа в виде отрезков соответствующей длины и все рассуждения проводили в геометрических терминах. Первые шаги в области алгебры сделал Диофант (кроме двух книг, которые он написал, о нем ничего не известно; считается, что он жил в III веке н.э.), но по-настоящему ею занялись среднеазиатские математики, среди которых выделяются Мухаммед бен Муса аль-Хорезми (780–847) и Гиясаддин Абу-л-Фатх Омар ибн Ибрагим аль-Хайям (1048–1131). Аль-Хорезми (от его имени происходит термин «алгоритм») в книге «Алджебр» изложил теорию решения линейных и квадратных уравнений, а Омар Хайям одновременно с созданием поэтического шедевра «Рубаи» разработал способы нахождения корней некоторых кубических уравнений. С этого времени алгебра утвердилась как самостоятельная ветвь математики.

§ 3. Рациональные и иррациональные числа

Следующий шаг в обобщении понятия числа связан с операцией деления. В множестве натуральных чисел $\mathbb{N}$ она имеет ограниченный характер: если a и b натуральные числа, то не всегда найдется натуральное число x такое, чтобы $ax = b$ (приведите примеры). Другими словами, в том случае, когда b не делится нацело на a, уравнение $ax = b$ в $\mathbb{N}$ неразрешимо. Чтобы устранить это несовершенство, вводятся дроби, записываемые в виде отношений $\dfrac{m}{n}$, где m и n – натуральные числа. При этом число m называется числителем, а число n – знаменателем дроби $\dfrac{m}{n}$. Вспомним правила действий с дробями:

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd}, \qquad \frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd}, \qquad \frac{a}{b} : \frac{c}{d} = \frac{ad}{bc}. \tag{1}$$

Правило $\dfrac{ac}{bc} = \dfrac{a}{b}$ позволяет «сокращать» дробь на общий для числителя и знаменателя множитель. Например, $\frac{12}{20} = \frac{3 \cdot 4}{5 \cdot 4} = \frac{3}{5}$. Как видим, равные дроби могут очень различаться по внешнему виду (как вы убедитесь в том, что $\frac{333}{703}$ и $\frac{369}{779}$ равны?). Натуральное число n можно считать частным случаем дроби, отождествляя его с $\frac{n}{1}$.

Построенное расширение натурального ряда обозначим через $\mathbb{Q}^+$ – это положительные рациональные числа (от латинского ratio – отношение). Умножение в $\mathbb{Q}^+$ ассоциативно и коммутативно, уравнения вида $ax = b$, где a и b – любые числа из Q^+, разрешимы (решением будет $x = \frac{b}{a}$). Следовательно, $(\mathbb{Q}^+, \cdot)$ – коммутативная группа, это – мультипликативная группа положительных рациональных чисел (от латинского multiplicatio – умножение). Деление в ней осуществляется неограниченно.

Если ноль и отрицательные числа сначала появились как математические абстракции и лишь впоследствии нашли содержательное толкование, то дроби были известны с древнейших времен. Распределение некоторого общего достояния на индивидуальные доли было повседневной практикой (см., например, в библейской книге Чисел стихи 25-46 главы 31). Другим видом деятельности, приводившим к дробям, были измерения: если, например, стандарт длины не укладывался между двумя данными точками целое число раз, приходилось прибегать к более мелким его частям. Так, в библейской книге пророка Иезекииля (глава 40) мерная трость подразделяется на шесть локтей.

Присоединяя к положительным рациональным числам противоположные им величины и ноль, получаем все множество рациональных чисел $\mathbb{Q}$. Оно состоит, таким образом, из нуля, положительных и отрицательных целых чисел, положительных и отрицательных дробей. Сложение, умножение и деление в $\mathbb{Q}$ выполняются по формулам (1), вычитание осуществляется по правилу

$$\frac{a}{b} - \frac{c}{d} = \frac{ad - bc}{bd}.$$

(Проведите следующие действия с рациональными числами:

$$\frac{3}{4} + \frac{5}{6}, \qquad \frac{2}{7} - \frac{11}{8}, \qquad -\frac{9}{10} \cdot -\frac{1}{3}, \qquad -12 : \frac{1}{13},$$

придумайте другие примеры.)

Действия с дробями представляют собой вершину практической арифметики. Желая произвести должное впечатление на просвещенных экзаменаторов Митрофана, наставник его Цыфиркин скорбел: «С парнем третий год над ломаными бьемся, да что-то плохо клеится». Но учитывая, что у недоросля не получалось «один да один», ясно, что «ломаные», т.е. дроби, упомянуты тут всуе, для красного словца.

В множестве $\mathbb{Q}$ рациональных чисел все четыре арифметические операции выполняются беспрепятственно за одним досадным исключением: нельзя делить на ноль (один из доводов в пользу того, что 0 – ненастоящее число). Следовательно, в этом множестве разрешимы уравнения вида $a + x = b$ при любых a, b и уравнения вида $ax = b$ при всех $a \neq 0$ и при всех b. Таким образом, множество $\mathbb{Q}$ является кольцом, а его ненулевые элементы образуют коммутативную группу по умножению. Кольца, обладающие этим свойством, называются полями. Мы построили поле рациональных чисел $(\mathbb{Q}, +, \cdot)$. Оно расширяет кольцо целых чисел $(\mathbb{Z}, +, \cdot)$, позволяя неограниченно выполнять операцию деления (кроме деления на ноль).

Поле рациональных чисел бесконечно, но существуют и конечные поля. Таковым будет, например, любое кольцо Z_p остатков от деления натуральных чисел на простое число p. Имея перед собой таблицу умножения поля Z_5, легко решить в нем уравнения $2x = 1,\ 3x = 4,\ 4x = 5$. А вот в кольце Z_6 эти уравнения не имеют корней (сверьтесь с таблицами).

В поле рациональных чисел разрешимо любое линейное уравнение $ax + b = 0$,– решением будет $x = -\dfrac{b}{a}$.

Наглядное представление о рациональных числах дает числовая ось. На некоторой прямой линии выбирается точка 0 – начало отсчета, указывается единица масштаба. Направление, в котором откладывается от нуля единица, называется положительным направлением оси (вправо от нуля), а противоположное ему (влево от нуля) – отрицательным. Если дано положительное рациональное число $\dfrac{m}{n}$, то единица масштаба делится на n равных частей и вправо от нуля эта доля откладывается m раз. Полученная точка и есть изображение числа $\dfrac{m}{n}$. Если число отрицательное – откладывание производят влево от нуля. На рис. 1 представлен процесс построения числа $\dfrac{5}{4}$: единица масштаба разделена на четыре части и 5 долей отложены вправо от нуля. Чтобы изобразить число $-\dfrac{2}{3}$, разделим единицу на три части и отложим влево от нуля две таких доли.

После построения числового множества $\mathbb{Q}$ (поле рациональных чисел), в котором разрешимо любое линейное алгебраическое уравнение $ax + b = 0$, естественно перейти к исследованию квадратных уравнений $ax^2 + bx + c = 0$. В простейшем случае $x^2 - 1 = 0$, т.е. $x^2 = 1$, имеем два решения (корня): $x_1 = 1$, $x_2 = -1$. Однако уже следующий напрашивающийся шаг заводит нас в тупик.

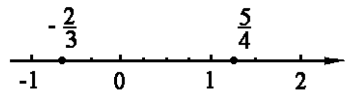


Рис. 1

Теорема 1. Уравнение $x^2 = 2$ не имеет решений в поле рациональных чисел.

ДОКАЗАТЕЛЬСТВО. Предположим противное: пусть рациональное число $\dfrac{a}{b}$ является решением уравнения $x^2 = 2$, т.е. $\left(\dfrac{a}{b}\right)^2 = 2$. Не нарушая общности, будем считать, что $\dfrac{a}{b}$ – несократимая дробь. (Слова «не нарушая общности» – одно из важнейших ритуальных выражений в математике. Весьма распространенной ошибкой в доказательствах является неосознанная подмена исследуемой ситуации некоторым ее частным случаем – нарушение общности. Словами «не нарушая общности, будем считать, что» доказывающий признает, что возможны и другие случаи, но они или очевидны или легко сводятся к данному: конечно, числитель и знаменатель в $\dfrac{a}{b}$ могут иметь общие множители, но, сократив на них дробь, мы получим тот случай, который и предлагается рассматривать.)

Так как $\left(\dfrac{a}{b}\right)^2 = \dfrac{a^2}{b^2} = 2$, то $a^2 = 2b^2$. Отсюда видно, что число a^2 – четное, а значит, и a – четное (почему?). Пусть $a = 2k$. Тогда $a^2 = (2k)^2 = 4k^2 = 2b^2$. Следовательно, $b^2 = 2k^2$, и, значит, b – четное число. Мы получили противоречие: с одной стороны, дробь $\dfrac{a}{b}$ несократима, а с другой, a и b – четные числа, т.е. имеют общий множитель 2. Это противоречие показывает, что исходное предположение неверно, и значит, уравнение $x^2 = 2$ не имеет рациональных решений. $\square$

В прежние времена доказательства заканчивались стандартным оборотом «что и требовалось доказать» – латинское «quod erat demonstrandum», сокращенно QED. Но структура доказательных рассуждений в современной математике бывает столь сложной, что заключительный этап в них может прямо не соотноситься с формулировкой теоремы, поэтому завершение доказательства стали отмечать знаком квадрата.

Факт неразрешимости в поле рациональных чисел $\mathbb{Q}$ уравнения $x^2 = 2$ (или в стандартной форме $x^2 - 2 = 0$) сам по себе не вызывает особых эмоций: уравнения типа $2+x = 1$ не имели решений в натуральных числах – и были введены отрицательные числа; неразрешимость в $\mathbb{N}$ уравнений типа $2x = 1$ привела к появлению дробей,– значит, придется расширять и поле $\mathbb{Q}$, придумав какие-то новые числа (название для них напрашивается само: иррациональные). Мы уже находимся в рамках некоторой философской системы – свобода в алгебраических действиях достигается ценой усложнения понятия числа, и очередной шаг вызывает разве лишь вопрос: до каких же пор это обобщение будет продолжаться? (Ответ: в этом направлении нам осталось сделать два шага).

Для древних греков, не знавших алгебры, факт неразрешимости в рациональных числах уравнения $x^2 = 2$ предстал в следующей поразившей их геометрической форме. Пусть имеется квадрат со стороной 1 и в нем проведена диагональ. Чему равна длина x этой диагонали? Так как две прилегающие стороны квадрата и соединяющая их концы диагональ образуют прямоугольный треугольник, а по теореме Пифагора сумма квадратов катетов равна квадрату гипотенузы, то $1^2 + 1^2 = x^2$, т.е. $x^2 = 2$, – длина диагонали и есть решение уравнения $x^2 = 2$. Если это решение выражается числом $\dfrac{m}{n}$, то мы разделим единицу длины (сторону квадрата) на n частей и, отложив такую долю m раз, получим на числовой оси точку, соответствующую длине диагонали квадрата. Но согласно доказанной теореме 1, такого решения уравнение $x^2 = 2$ не имеет. Следовательно, диагональ квадрата не может быть измерена ни в каких долях его стороны. Таким образом, теорема 1 приобретает следующий вид.

<u>Теорема 2</u>. Диагональ квадрата несоизмерима с его стороной.

Теоремы 1 и 2 при всем их внешнем несходстве представляют собой лишь разные интерпретации одного и того же математического факта: одна на языке алгебры, другая – в геометрических терминах. В математике это обычное явление. Иногда требуются

значительные усилия, чтобы во вновь доказанных утверждениях распознать истины, давно известные в других формулировках.

Обратимся теперь к числовой оси. Если взять на ней точки, соответствующие рациональным числам a и b (пусть $a < b$), то середина отрезка $[a, b]$ выражается числом $c = \dfrac{a+b}{2}$. Это тоже рациональное число, так что между любыми двумя рациональными числами лежит еще одно. Деля пополам отрезки $[a, c]$ и $[c, b]$, получим еще два рациональных числа между a и b и т.д. Поскольку этот процесс деления пополам (дихотомия) можно продолжать неограниченно, приходим к выводу, что между произвольными рациональными числами a и b находится бесконечно много других рациональных чисел. Представив себе все это, мы могли бы прийти к заключению, что рациональные числа заполняют сплошь всю числовую ось. Но нет – если от точки 0 отложить вправо диагональ единичного квадрата, то согласно теореме 2 другой конец диагонали не попадет ни в какую рациональную точку. Вот эти «дыры» на числовой прямой и были интерпретированы как иррациональные числа. (В словаре: «иррациональный – в философии: не постигаемый разумом, такой, который не может быть выражен в логических понятиях». Повседневный математический смысл проще: иррациональное число – это просто число, не являющееся рациональным).

Иррациональных чисел тоже бесконечно много: если α – иррациональное, а a – рациональное число, то сумма $\beta = \alpha + a$ и произведение $\gamma = \alpha a$ (при $a \neq 0$) тоже будут иррациональными (иначе, например, $\alpha = \beta - a = \dfrac{\gamma}{a}$ оказалось бы рациональным).

Первый конкретный пример иррационального числа – это длина диагонали единичного квадрата, т.е. положительный корень уравнения $x^2 = 2$, обозначим это число через $\sqrt{2}$. Рассуждения, проведенные в доказательстве теоремы 1, могут быть почти дословно повторены и для уравнений $x^2 = 3,\ x^2 = 5,\ x^3 = 2$ и т.д., что доказывает иррациональность чисел $\sqrt{3}$ (длина диагонали единичного куба,– проверьте!), $\sqrt{5},\ \sqrt[3]{2}$ и т.д.

После введения иррациональных чисел появилась возможность дать общую запись для решений любого квадратного уравнения $ax^2 + bx + c = 0$ с использованием радикала (знака извлечения корня), именно

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

Если величина $D = b^2 - 4ac$ (дискриминант) положительна, уравнение имеет два разных корня, рациональных или иррациональных, в зависимости от того, является ли дискриминант полным квадратом или нет; если $D = 0$, получается два совпадающих корня (т.е. один); если $D < 0$, – уравнение не имеет решений.

Золотым сечением называется деление отрезка длины 1 на две части, бо́льшая из которых x является средней пропорциональной величиной между всем отрезком и его меньшей частью $1 - x$, т.е. $\dfrac{1}{x} = \dfrac{x}{1-x}$. Принцип золотого сечения (название ввел Леонардо да Винчи в конце XV века) составлял, в частности, теоретическую основу архитектурных композиций классической древности и эпохи Возрождения (на практике, впрочем, редко использовавшуюся). Для нахождения x получаем квадратное уравнение $x^2 + x - 1 = 0$, откуда $x = \dfrac{1}{2}(\sqrt{5} - 1)$ (отрицательный корень отбрасываем). Для нас иррациональность этого числа очевидна (все-таки поясните), в XIII веке она была установлена весьма сложным путем, а сам факт произвел удручающее впечатление.

Пример квадратного уравнения возбудил надежду на то, что алгебраические уравнения и всех других, более высоких степеней тоже окажутся разрешимыми в радикалах, т.е. корни их можно будет выразить с помощью арифметических операций и извлечения корней. В середине XVI века итальянские математики Тарталья, Кардано и Феррари нашли подобные формулы для кубического уравнения и уравнения четвертой степени – происходило это в атмосфере ожесточенной полемики о приоритете, с публичными состязаниями в решении соответствующих задач, проклятиями и покаяниями. В последующие почти три столетия дальнейших существенных продвижений не было, и лишь в 1826 году норвежский математик Нильс Хенрик Абель (1802–1829) доказал, что для каждого натурального числа $n > 4$ существуют алгебраические уравнения степени n с целыми коэффициентами, неразрешимые в радикалах (простейшим примером для $n = 5$ является уравнение $x^5 - 4x - 2 = 0$). Окончательное решение проблемы, занимавшей умы лучших математиков, принадлежит французу Эваристу Галуа (1811–1832). Он ввел понятие группы и показал, что каждому алгебраическому уравнению соответствует некоторая группа, по свойствам которой и можно судить, разрешимо или нет данное уравнение в радикалах.

Абель и Галуа ушли из жизни совсем молодыми (первый скончался от туберкулеза, второй был убит на дуэли), их идеи не были должным образом восприняты современниками, но оказали впоследствии огромное влияние на развитие важнейших разделов математики. Имена этих выдающихся ученых носят многие математические объекты, например, коммутативные группы называются абелевыми, а конечные поля – полями Галуа.

§ 4. Поле действительных чисел. Комплексные числа

Рациональные и иррациональные числа вместе образуют множество действительных чисел $\mathbb{R}$. С рациональной его частью – полем $\mathbb{Q}$ – мы знакомы хорошо, но что представляют собой иррациональные числа? Некоторые из них можно записать с помощью радикалов: $\sqrt{2}$, $\sqrt[3]{5}$, $\sqrt{1+\sqrt[3]{2}+\sqrt[4]{3}}$; другие обозначаются специальными символами: π, e; третьи получаются как значения функций: lg 2, sin10°,– но все это ничего не говорит о величине этих чисел. Возьмем, например, число $\sqrt{2}$. Это иррациональное число, квадрат которого равен 2. Так как $1^2 = 1$, $2^2 = 4$, то $\sqrt{2}$ больше 1, но меньше 2, т.е. имеют место неравенства $1 < \sqrt{2} < 2$. Разделим отрезок $[1, 2]$ на 10 частей и посмотрим, в какой из них находится $\sqrt{2}$. Поскольку $1,4^2 = 1,96$, а $1,5^2 = 2,25$, заключаем, что $1,4 < \sqrt{2} < 1,5$. Теперь разобьем на 10 частей отрезок $[1,4; 1,5]$ и определим ту его часть, где расположена точка, соответствующая числу $\sqrt{2}$. По таблице квадратов (или с помощью калькулятора) находим: $1,41^2$=$1,988 < 2$, но $1,42^2 = 2,016 > 2$, так что $1,41 < \sqrt{2} < 1,42$. Продолжая это увлекательное занятие, получаем для числового значения символа $\sqrt{2}$ все более тесные границы: $1,414 < \sqrt{2} < 1,415$; $1,4142 < \sqrt{2} < 1,4143$; $1,41421 < \sqrt{2} < 1,41422$; $1,414213 < \sqrt{2} < 1,414214$ и т.д.

(В этом месте полезно напомнить правила округления десятичных дробей до определенного разряда. Округление состоит в отбрасывании, точнее в замене нулями, всех цифр, стоящих правее цифры данного разряда. Если при этом первой из отбрасываемых цифр является 0, 1, 2, 3 или 4, последняя остающаяся цифра не меняется; если же первой из отбрасываемых цифр окажется 5, 6, 7, 8 или 9, последняя остающаяся цифра увеличивается на 1, а если этой цифрой была 9, то увеличение на 1 произойдет в предшествующем ей разряде, а сама она заменится на 0.

Старинное «правило четной цифры»: когда отбрасывается только цифра 5, последняя остающаяся цифра не меняется, если она четная, и увеличивается на 1, если она нечетная. Пользуясь этими правилами, округлите до второго знака после запятой, т.е. до сотых долей числа $1,4142;\ 2,71828;\ 3,14159;\ 0,445;\ 0,435$. Аналогично округляются до десятков, сотен, тысяч и т.д. целые числа).

Рассматриваемый процесс поиска числового значения для $\sqrt{2}$ никогда не закончится, поскольку все получаемые в нем числа рациональны, а $\sqrt{2}$ – нет. На том этапе, где мы остановились, можно сказать, что имеет место приблизительное равенство $\sqrt{2} \approx 1,41421$, т.е. $\sqrt{2}$ вычислен с точностью до пяти знаков после запятой, или с точностью до стотысячных долей. Само же число $\sqrt{2}$ можно отождествить с бесконечной десятичной дробью, последовательные знаки которой отыскиваются в соответствии с описанной выше процедурой. Точно так же обстоит дело с каждым иррациональным числом: величина его считается известной, если указан способ вычисления соответствующей бесконечной десятичной дроби. Например, для знаменитого числа π, которое выражает отношение длины любой окружности к ее диаметру, Виета предложил удивительную формулу

$$\frac{\pi}{2} = \frac{1}{\sqrt{\frac{1}{2}} \cdot \sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2}}} \cdot \sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2}}}} \cdots},$$

где в знаменателе стоит бесконечное произведение (сможете вы написать следующий множитель?), а Лейбниц представил $\frac{\pi}{4}$ в виде бесконечной знакочередующейся суммы

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \frac{1}{11} + \ldots,$$

имеющей, впрочем, более эстетическую, чем вычислительную ценность. Беря все большее число множителей в первой формуле или число слагаемых во второй, можно находить все более точные значения для π.

Арифметические операции над иррациональными числами определяются с помощью одноименных действий над их приближенными значениями, выраженными в десятичных дробях: складывая все более точные значения иррациональных чисел α и β, мы будем получать все более точные значения их суммы $\alpha + \beta$ и т.п. Рациональные числа для единообразия тоже можно воспринимать как бесконечные десятичные дроби, например, $\dfrac{1}{2} = 0,500\ldots$ (бесконечное число нулей), $\dfrac{1}{3} = 0,333\ldots$ (бесконечное число троек), $\dfrac{8}{7} = 1,(142857)$ (бесконечная периодическая дробь). При таком подходе все действительные числа становятся равноправными. Относительно сложения и умножения множество действительных чисел $\mathbb{R}$ оказывается полем – это поле действительных чисел. В нем неограниченно осуществимы обе обратные арифметические операции – вычитание и деление (кроме, конечно, деления на ноль).

Геометрическим образом множества действительных чисел $\mathbb{R}$ служит числовая прямая – числовая ось, на которой мы изображали рациональные числа, пополненная точками, представляющими иррациональные числа. Каждой точке этой прямой в свою очередь соответствует некоторое число, рациональное или иррациональное. Таким образом, на числовой прямой нет «дыр», она непрерывна. Любая точка на ней определяется одной координатой – тем действительным числом, которое соответствует этой точке. В этом смысле числовая прямая $\mathbb{R}$ представляет одномерное пространство, оно называется одномерным арифметическим пространством. Первую исчерпывающую теорию действительных чисел построил в 1872 году немецкий алгебраист Рихард Юлиус Вильгельм Дедекинд (1831–1916). Другие подходы предложили позднее его соотечественники Кантор и Вейерштрасс. Таким образом, к концу XIX века завершилось формирование общего представления о числе как о мере вещей.

А как в $\mathbb{R}$ обстоит дело с алгебраическими уравнениями? Не всякое алгебраическое уравнение с целыми коэффициентами разрешимо в поле действительных чисел, например, $x^2 + 1 = 0$ неразрешимо: ясно, что нет такого рационального или иррационального числа, квадрат которого равнялся бы -1. Если же уравнение имеет действительный корень x_0, то он либо выражается в радикалах (и может оказаться рациональным числом, если все корни в его записи извлекаются), либо, не выражаясь в радикалах, может быть вычислен с какой угодно точностью одним из многочисленных методов приближенного решения уравнений.

Любое ли действительное число является корнем некоторого алгебраического уравнения с целыми коэффициентами? Для рациональных чисел это так: если $x_0 = \frac{a}{b}$, где a и b – целые числа, то x_0 будет корнем линейного уравнения $bx - a = 0$. Для иррациональных чисел, выразимых с помощью радикалов, ситуация аналогична, хотя сответствующее уравнение написать не всегда просто. Например, если $x_0 = \sqrt[3]{1 + \sqrt{2}}$, то (возведем в куб обе части) $x_0^3 = 1 + \sqrt{2}$, откуда $x_0^3 - 1 = \sqrt{2}$, а после возвышения в квадрат $(x_0^3 - 1)^2 = 2$, или (по формуле квадрата разности) ${x_0}^6 - 2{x_0}^3 + 1 = 2$, т.е. ${x_0}^6 - 2{x_0}^3 - 1 = 0$, и значит, x_0 является корнем уравнения $x^6 - 2x^3 - 1 = 0$.

Корни алгебраических уравнений с целыми коэффициентами называются алгебраическими числами, а числа, не являющиеся алгебраическими, называются трансцендентными. (В философском смысле «трансцендентный» – это недоступный познанию, находящийся за пределами опыта. В нашем случае – за пределами алгебраического опыта начальной математики). Вопрос о существовании трансцендентных чисел неожиданно оказался связанным с самой известной из математических проблем – задачей о квадратуре круга.

Задача эта заключалась в том, чтобы с помощью только циркуля и линейки построить квадрат, площадь которого равнялась бы площади данного круга. Квадратурой круга занимались во все времена и во всех математических цивилизациях, она стала символом непреодолимой трудности в любой сфере человеческой деятельности, вошла в качестве понятия в обыденный язык. (Как говорила императрица Екатерина II о своей тезке, президенте Российской Академии, «скорее найдут квадратуру круга, чем переспорят княгиню Дашкову».)

Не нарушая общности, будем считать, что данный круг имеет радиус r, равный 1. Площадь круга вычисляется по формуле $s = \pi r^2$, так что в нашем случае $s = \pi$, и значит, нужно построить квадрат, площадь которого равнялась бы π. Так как площадь квадрата равна квадрату его стороны, то сторона искомого квадрата должна иметь длину $\sqrt{\pi}$. Одна из простейших задач на построение позволяет по заданному отрезку длины a начертить с помощью циркуля и линейки отрезок длины $\sqrt{a}$. Следовательно, вопрос о квадратуре круга сводится в конце концов к тому, сможем ли мы в данном масштабе, пользуясь только циркулем и линейкой, изобразить отрезок длины π.

Но какое же числовое значение имеет π – константа, обозначающая отношение длины окружности любого круга к его диаметру? Древнейшее приближенное значение для π зафиксировано в Библии, во Второй книге Паралипоменон (глава 4, стих 2): «И сделал море литое,– от края его до края его десять локтей,– все круглое, вышиной в пять локтей; и снурок в тридцать локтей обнимал его кругом»,– речь идет о медном бассейне в сооружаемом Соломоном доме Господнем. Так как длина окружности бассейна составляла 30 локтей, а его диаметр 10 локтей, то получаем, что $\pi \approx 3$. Современное «бытовое» приближение $\pi \approx 3,14$ было получено Архимедом. Попытки найти точное значение для π (а значит, и разрешить проблему квадратуры круга) приводили к появлению все новых десятичных знаков этого числа, но конца им не было видно. В 1610 году голландец Лудольф ван Цейтен вычислил π с 32 знаками после запятой (π долгое время после этого называлось «лудольфовым числом». Обозначение π предложил для него Эйлер в 1737 году).

Квадратурой круга специально занимались Леонардо да Винчи и Альбрехт Дюрер, но они не вдавались в утомительные подсчеты, связанные с π. В начале XVII века, когда стало известно 127 десятичных знаков числа π, все ведущие математики уже были убеждены, что процесс уточнения числовой величины для π бесконечен, т.е. что π – иррациональное число. В 1766 году это доказал немецкий математик Иоганн Генрих Ламберт (1728–1777), после чего направление в квадратуре круга, связанное с поисками точного значения π, пресеклось. Но к тому времени появилась новая идея: анализируя геометрические образы, которые можно построить циркулем и линейкой, удалось установить, что если отрезок длины a получается таким путем, то число a обязательно будет корнем некоторого алгебраического уравнения, т.е. алгебраическим числом. В 1882 году Карл Луис Фердинанд фон Линдеман (1852–1939) показал, что число π таковым не является, т.е. что π – трансцендентное число. Тем самым задача о квадратуре круга была решена в отрицательном смысле: с помощью только циркуля и линейки нельзя построить квадрат, по площади равный данному кругу. (Все слышали о квадратуре круга, но даже среди математиков далеко не каждый назовет сейчас того, кто поставил последнюю точку в решении знаменитой задачи, – O quam cito transit gloria mundi, «О как скоро проходит земная слава».)

Какие же практические результаты имело решение задачи, на протяжении тысячелетий привлекавшей внимание самых выдающихся ученых? В непосредственном понимании этого вопроса – никаких. Но так же, как и Великая теорема Ферма, эта проблема внесла существенный вклад в формирование новых математических идей и методов.

Заметим, что достижение Линдемана не уменьшило количество любителей, продолжавших (и продолжающих!) сражаться с химерой, называемой «квадратурой круга»: сложнейшее математическое доказательство невозможности желанного построения им недоступно, а внешняя простота задачи вызывает обманчивую надежду найти к ней какой-то необычный, не замеченный профессионалами подход. (В одной из комедий Аристофана – IV век до н.э.! – герой восклицает:

Возьмем линейку, точно все разметим,
И вмиг из круга сделаем квадрат!)

До сих пор в печати время от времени появляются сочувственные публикации о том, что тот или иной энтузиаст справился, наконец, с квадратурой круга, но никак не может добиться признания из-за сопротивления официальной науки. Что тут сказать?..

Буквенные обозначения в алгебру ввел уже упоминавшийся Франсуа Виета (1540–1603). По профессии он был юристом и свои математические изыскания проводил в немногие свободные от королевской службы часы. Новый способ записи уравнений и формул позволил Виета увидеть скрытые до того соотношения между различными алгебраическими величинами. Все помнят из школьного курса математики следующую его теорему о корнях квадратного уравнения.

Теорема. Сумма корней квадратного уравнения $ax^2+bx+c = 0$ равна $-\frac{b}{a}$, а их произведение равно $\frac{c}{a}$.

ДОКАЗАТЕЛЬСТВО. Почленным делением преобразуем формулы для корней:

$$x_1 = \frac{-b+\sqrt{b^2-4ac}}{2a} = -\frac{b}{2a} + \frac{\sqrt{b^2-4ac}}{2a},$$

$$x_2 = -\frac{b}{2a} - \frac{\sqrt{b^2-4ac}}{2a}.$$

Складываем корни:

$$x_1 + x_2 = -\frac{b}{2a} + \frac{\sqrt{b^2-4ac}}{2a} - \frac{b}{2a} - \frac{\sqrt{b^2-4ac}}{2a} = -2\cdot\frac{b}{2a} = -\frac{b}{a},$$

т.е. $x_1 + x_2 = -\frac{b}{a}$.

Вычисляя произведение корней, воспользуемся известным тождеством для разности квадратов $(u-v)(u+v)=u^2-v^2$:

$$x_1x_2=\left(-\frac{b}{2a}+\frac{\sqrt{b^2-4ac}}{2a}\right)\left(-\frac{b}{2a}-\frac{\sqrt{b^2-4ac}}{2a}\right)=$$

$$=\left(-\frac{b}{2a}\right)^2-\left(\frac{\sqrt{b^2-4ac}}{2a}\right)^2=$$

$$=\frac{b^2}{4a^2}-\frac{b^2-4ac}{4a^2}=\frac{b^2}{4a^2}-\frac{b^2}{4a^2}+\frac{c}{a}=\frac{c}{a},$$

т.е. $x_1x_2=\dfrac{c}{a}$. $\square$

Доказательство теоремы проводится непосредственными вычислениями, которые шаг за шагом ведут к поставленной цели. Логически эта схема гораздо проще, чем доказательство от противного.

Выкладки, проделанные в доказательстве теоремы Виета, формально, без проникновения в суть производимых действий, проходят и в том случае, когда $b^2-4ac<0$, т.е. когда формулы для корней становятся бессмысленными. Заметим, что Виета не признавал отрицательных чисел и для него эти формулы еще чаще оказывались бессодержательными. Как все было бы хорошо и просто, если бы не было этих ограничений, если бы квадратное уравнение всегда имело два корня!

Подобная ситуация возникла и при решении кубических уравнений: в промежуточных формулах появлялись квадратные корни из отрицательных чисел, хотя в результате дальнейших формальных действий получались обычные величины.

И наконец, главное: в поле действительных чисел неразрешимо простейшее уравнение $x^2+1=0$, или $x^2=-1$. Что если применить и к данной ситуации уже использованный принцип расширения привычного числового мира? Введем новое число $i=\sqrt{-1}$. Тогда уравнение $x^2=-1$ будет иметь два корня: i и $-i$, ибо $i^2=(\sqrt{-1})^2=-1$, $(-i)^2=(-\sqrt{-1})^2=-1$. Но, более того, с появлением «мнимой единицы» (так было названо i) любое квадратное уравнение получает два корня! В самом деле, если $D = b^2 - 4ac < 0$, то $-D > 0$, и значит,

$$x_{1,2} = \frac{-b \pm \sqrt{D}}{2a} = \frac{-b \pm \sqrt{(-1)(-D)}}{2a} =$$

$$= \frac{-b \pm \sqrt{-1}\sqrt{-D}}{2a} = \frac{-b \pm i\sqrt{-D}}{2a}.$$

Например, для уравнения $x^2 - 2x + 2 = 0$ имеем $D = -4$, откуда $-D = 4$, и значит, $x_{1,2} = \dfrac{2 \pm i\sqrt{4}}{2} = \dfrac{2 \pm 2i}{2} = 1 \pm i$. Проверка: ${x_1}^2 - 2x_1 + 2{=}(1 + i)^2 - 2(1 + i) + 2{=}(1 + 2i + i^2) - (2 + 2i) + 2{=}1 + 2i + i^2 - 2 - 2i + 2{=}1 + i^2{=}1 + (-1){=}0$, т.е. $x_1 = 1 + i$ действительно является корнем. Аналогично для $x_2 = 1 - i$. (Вот оно, по словам Бальзака, «высокое безумство, которое пытается найти неизвестное из уравнения с мнимыми корнями».)

Конечно, соглашались математики, эти корни не настоящие, это лишь мнимые числа, но смотрите, как прекрасно все получается, если мы допустим их в качестве некоторых идеальных, воображаемых объектов. Или, в подлинных словах Лейбница: «Мнимые числа – это прекрасное и чудесное убежище божественного духа, сочетание бытия с небытием». (Готфрид Вильгельм Лейбниц (1646–1716) – немецкий математик и философ, один из основоположников математического анализа и математической логики, изобретатель счетной машины, ученый, внесший вклад также в историю, языкознание, биологию, педагогику, юриспруденцию и многие другие науки. «Я был, собственно говоря, самоучкой во всякой науке; как только я приобретал в ней первые понятия, я всегда искал нового, часто просто потому, что не успевал достаточно усвоить обыкновенное».)

Итак, очередной (и, как оказалось, завершающий) шаг в процессе обобщения понятия числа – в связи с решением алгебраических уравнений – был сделан. Появились комплексные числа (математики произносят только «компле́ксные» – с ударением на втором слоге – и даже добились соответствующей пометки в словарях). Символ вида $a + bi$, где a и b – действительные числа, а $i = \sqrt{-1}$ – мнимая единица, называется комплексным числом с действительной частью a и мнимой частью b. Сложение и умножение в множестве $\mathbb{C}$ комплексных чисел осуществляется по правилам

$$(a + bi) + (c + di) = (a + c) + (b + d)i$$

(т.е. отдельно складываются действительные и мнимые части),

$$(a + bi)(c + di) = (ac - bd) + (ad + bc)i$$

(раскрываем скобки, как при обычном алгебраическом умножении, но помним, что $ii = -1$).

Действительное число a можно трактовать как комплексное число $a + 0i$ (т.е. с отсутствующей мнимой частью), так что поле действительных чисел является подмножеством множества $\mathbb{C}$. При этом операции над действительными числами оказываются частным случаем операций, введенных в $\mathbb{C}$:

$$(a + 0i) + (b + 0i) = (a + b) + 0i,$$

$$(a + 0i)(b + 0i) = ab + 0i$$

(проделайте вычисления в деталях).

Комплексные числа вида $0 + bi = bi$ называются чисто мнимыми.

Нетрудно убедиться в том, что сложение и умножение, определенные в $\mathbb{C}$, ассоциативны и коммутативны и что умножение дистрибутивно относительно сложения.

Вычитание и деление в $\mathbb{C}$ осуществляются по формулам

$$(a + bi) - (c + di) = (a - c) + (b - d)i,$$

$$\frac{a + bi}{c + di} = \frac{ac + bd}{c^2 + d^2} + \frac{bc - ad}{c^2 + d^2}i.$$

Последняя формула получается применением замечательного тождества $x^2 + y^2 = (x + yi)(x - yi)$ – разложение суммы квадратов в произведение! Действительно,

$$\frac{a + bi}{c + di} = \frac{(a + bi)(c - di)}{(c + di)(c - di)} =$$

$$= \frac{(ac + bd) + (bc - ad)i}{c^2 + d^2} = \frac{ac + bd}{c^2 + d^2} + \frac{bc - ad}{c^2 + d^2}i.$$

Деление на ноль, т.е. на $0 + 0i$, и здесь запрещено.

Таким образом, комплексные числа образуют поле $\mathbb{C}$, включающее в себя поле действительных чисел $\mathbb{R}$.

Прежде чем подняться на вершину, окинем взглядом пройденный путь. Невозможность решить в натуральных числах линейные уравнения вида $a + x = b$ при $a > b$ привела к появлению отрицательных величин, которые вместе с натуральным рядом и нулем образовали кольцо целых чисел $\mathbb{Z}$. Однако в кольце $\mathbb{Z}$ все еще были неразрешимы линейные уравнения типа $ax = b$, где b не делилось нацело на a. Появились дроби, в результате присоединения которых к $\mathbb{Z}$ возникло поле рациональных чисел $\mathbb{Q}$. В нем уже любое линейное уравнение $ax + b = 0$ с целыми и даже рациональными коэффициентами a и b оказалось разрешимым. Открытие несоизмеримости диагонали квадрата с его стороной в алгебраических терминах означало, что в поле $\mathbb{Q}$ простейшее квадратное уравнение $x^2 = 2$ не имеет корней. Расширение поля рациональных чисел $\mathbb{Q}$ до поля действительных чисел $\mathbb{R}$ за счет иррациональных величин оказалось выходом из тупика, но лишь частичным: по-прежнему не имело решений уравнение $x^2 = -1$. Изобретение мнимой единицы i и комплексных чисел позволило построить новое числовое поле $\mathbb{C}$, в котором содержались все известные до того числа и в котором оказалось разрешимым любое квадратное уравнение.

Можно было ожидать, что при анализе уравнений более высоких степеней возникнет необходимость в дальнейшем обобщении понятия числа. Но в 1799 году Гаусс доказал так называемую Основную теорему алгебры.–

Теорема. Всякое алгебраическое уравнение степени n с действительными или комплексными коэффициентами

$$a_0x^n + a_1x^{n-1} + \cdots + a_{n-1}x + a_n = 0$$

разрешимо в поле комплексных чисел и имеет n корней (с учетом их кратностей).

Например, согласно этой теореме уравнение $x^4 + 4x^2 = 0$ имеет четыре корня. В самом деле, после разложения на множители $x^4 + 4x^2 = x^2(x^2 + 4) = x^2(x - 2i)(x + 2i)$ получаем $x_1 = 0,\ x_2 = 0,\ x_3 = 2i,\ x_4 = -2i$.

Карл Фридрих Гаусс (1777–1855) по праву считается одним из величайших математиков. Современник Гете, Бетховена и Гегеля, он получил выдающиеся результаты практически во всех областях теоретической и прикладной математики. Гаусс считал математику царицей всех наук, а теорию чисел – царицей математики. Среди всех своих замечательных открытий он особенно ценил полученное в 19-летнем возрасте решение древней геометрической задачи: построение правильного 17-угольника с помощью циркуля и линейки. Эта фигура, согласно завещанию Гаусса, была выгравирована на его могиле.

В XVIII веке мнимые числа (так их назвал Декарт) широко использовались в математической практике, но не имели никакого содержательного истолкования. Им не было места на числовой прямой, а выйти за пределы этого одномерного пространства не догадались ни Лейбниц, ни Бернулли, ни Даламбер, ни Эйлер (именно он обозначил мнимую единицу буквой i). Первым это сделал датский землемер Каспар Вессель в 1799 году. Он предложил отождествить комплексные числа с точками плоскости.

Возьмем на плоскости декартову прямоугольную систему координат, т.е. к горизонтальной оси (она называется осью абсцисс, или осью Ox) присоединим еще вертикальную ось (называемую осью ординат, или осью Oy) таким образом, чтобы начала отсчета на обеих осях совпадали, а положительное направление на вертикальной оси было бы указано вверх. Тогда каждой точке A плоскости можно соотнести пару действительных чисел (x, y), где x – проекция точки A на ось абсцисс Ox, а y – ее проекция на ось ординат Oy. Эти два числа называются координатами точки A. С другой стороны, каждой паре действительных чисел (x, y) соответствует однозначно определенная точка плоскости – именно та, координатами которой являются x и y. Совокупность всевозможных пар действительных чисел обозначается через $\mathbb{R}\times\mathbb{R}$ или $\mathbb{R}^2$ и называется числовой плоскостью, или двумерным арифметическим пространством.

Идея геометрического представления комплексных чисел заключается в том, что комплексному числу $a+bi$ сопоставляется точка плоскости с координатами (a, b). Таким образом, между точками числовой плоскости и комплексными числами устанавливается попарное соответствие (рис. 2). Так мнимые числа получили вполне осязаемую интерпретацию.

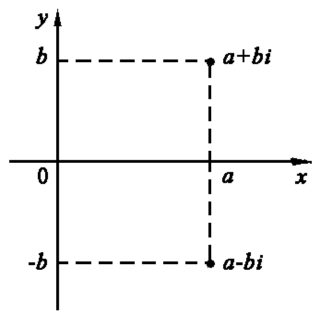


Рис. 2

Статья Весселя осталась незамеченной, и геометрическое представление комплексных чисел вошло в математику лишь после 1831 года, когда Гаусс переоткрыл его (и предложил отказаться от термина «мнимые числа»). С начала

XX века имя Весселя было воскрешено историками математики, но прижизненная известность его миновала,– далеко не единичный случай в науке.

Мнимые числа появились в то время, когда бурно развивавшаяся математика привлекла внимание многих выдающихся личностей, в общем, далеких от нее, и под влиянием их авторитета становилась предметом интереса в широких кругах общества. Не имеющее никакого реального смысла число $\sqrt{-1}$, ставшее вдруг необходимым элементом при решении важных прикладных задач,– в этом было что-то мистическое. Псевдонаучные рассуждения, связанные с мнимой единицей, были в ходу вплоть до конца XIX века, когда комплексные числа стали постепенно проникать в школьные программы и утратили ореол таинственности. Выражения «мнимая величина», «мнимая единица» перестали быть привнесенными в язык математическими оборотами и стали расхожими речевыми штампами.

§ 5. Сравнение бесконечностей. Кардинальные числа

При обмене, скажем, раковин на орехи (штука на штуку) участники сделки имеют две возможности для сравнительного сопоставления своих ценностей. Первый метод состоит в том, чтобы пересчитать сначала предметы одного рода, затем другого и далее иметь дело уже с отвлеченными числами: 5 против 7. Однако этот способ предполагает умение безошибочно считать до известного предела. Другой прием более примитивен, но зато и более надежен: выстроить в ряд раковины и напротив каждой класть по ореху. Процесс раскладки закончен, лишних раковин или орехов не оказалось – выставленные на обмен товары равноценны. Эта бесхитростная идея и была положена Кантором в основу важнейшего для теории множеств определения. Говорят, что два множества A и B равномощны, если составляющие их элементы могут быть приведены в попарное соответствие, т.е. если каждому элементу множества A будет сопоставлен однозначно определенный элемент из B и каждому элементу из B – однозначно определенный элемент из A (иначе говоря, если между элементами множеств A и B может быть установлено взаимно однозначное соответствие).

Множество называется конечным, если оно состоит из конечного числа элементов, и бесконечным в противном случае. Множество всех книг в библиотеке нашего университета конечно

– все его элементы (книги) перечислены в генеральном каталоге. Конечным будет и множество корней любого алгебраического уравнения: по Основной теореме алгебры их столько, какова степень уравнения. Числовые множества $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ бесконечны. К числу конечных множеств относят и пустое множество $\varnothing$, которое вовсе не имеет элементов.

Конечные множества удобно задавать перечислением: запись $A = \{a_1, a_2, \ldots, a_n\}$ означает, что множество A состоит из n элементов $a_1, a_2, \ldots, a_n$. В области конечных множеств понятие равномощности объясняет

<u>Теорема 1</u>. Конечные множества A и B равномощны тогда и только тогда, когда у них одинаковое число элементов.

ДОКАЗАТЕЛЬСТВО. Логическая связка «тогда и только тогда, когда» выражает равносильность двух соединяемых ею предложений, т.е. что из справедливости одного неизбежно вытекает истинность другого. Поэтому доказательство теорем с подобной формулировкой состоит из двух частей.

1) Покажем, что если конечные множества A и B равномощны, то у них одинаковое число элементов. Пусть между элементами множеств A и B имеется взаимно однозначное соответствие. Пересчитаем элементы множества A, вместе с каждым из них откладывая и соответствующий ему элемент множества B. Когда эта процедура завершится, будут одновременно исчерпаны оба множества. Значит, количество элементов в них одно и то же.

2) Покажем, что если в конечных множествах A и B одинаковое число элементов, то A и B равномощны. Пусть множества A и B имеют по n элементов каждое. Расположим элементы множества A в некотором порядке: $a_1, a_2, \ldots, a_n$. То же самое проделаем с элементами множества B: $b_1, b_2, \ldots, b_n$. Теперь каждому элементу множества A сопоставим элемент из множества B, имеющий такой же номер: первому – первый, второму – второй,. . . , последнему – последний. Это попарное соответствие и доказывает равномощность множеств A и B. $\square$

Теперь обратимся к бесконечным множествам. Перечислением элементов такое множество задать невозможно: процесс никогда не закончится. Поэтому бесконечное множество обычно описывают указанием свойства, характерного для всех его элементов, используя запись вида $A = \{x |\ \Pi(x)\}$, произносимую как «A есть множество всех элементов x, обладающих свойством П» или «A состоит из . . . ». Например, отрезок числовой прямой [0,1], образованный действительными числами, заключенными между

0 и 1, включая сами эти границы, можно записать в виде $[0, 1] = \{x|\ x$ – действительное число и $0 \leqslant x \leqslant 1\}$. Вместо слов «$x$ – действительное число», указывающих, из какого множества берутся элементы с обсуждаемым свойством, пишут кратко $x \in \mathbb{R}$. Это указание обычно делается сразу, и получается $[0, 1] = \{x \in \mathbb{R} | 0 \leqslant x \leqslant 1\}$. Еще пример: $\mathbb{N} = \{x \in \mathbb{Z}|\ x > 0\}$ (т.е. натуральные числа – это положительные целые числа). Разумеется, рассматриваемый способ задания множеств применим и к конечным множествам. Так, свойство $x^4 - 1 = 0$ определяет в зависимости от того, откуда берется число x, разные конечные множества: $A_1 = \{x \in \mathbb{N}|\ x^4 - 1 = 0\} = \{1\}$, т.е. A_1 состоит только из одного числа 1, $A_2 = \{x \in \mathbb{Z}|\ x^4 - 1 = 0\} = \{x \in \mathbb{R}|\ x^4 - 1 = 0\} = \{1, -1\}$, наконец, $A_3 = \{x \in \mathbb{C}|\ x^4 - 1 = 0\} = \{1, -1, i, -i\}$. Символ $\{\ldots|\ \ldots\}$ называется классификатором.

Иногда после указания нескольких элементов бесконечного множества объединяющее их свойство становится ясным, и тогда данное множество представляют более наглядно в виде последовательности:

$$
\begin{aligned}
&\mathbb{N} = \{1,\ 2,\ 3,\ 4,\ 5, \ldots\},\\
&A = \{2,\ 4,\ 6,\ 8,\ 10, \ldots\},\\
&B = \{1,\ 4,\ 9,\ 16,\ 25, \ldots\},\\
&P = \{2,\ 3,\ 5,\ 7,\ 11,\ 13,\ 17,\ 19, \ldots\},\\
&\mathbb{Z} = \{0,\ 1,\ -1,\ 2,\ -2,\ 3,\ -3, \ldots\},\\
&E = \{1,\ -\frac{1}{8},\ \frac{1}{27},\ -\frac{1}{64},\ \frac{1}{125},\ -\frac{1}{216}, \ldots\}, -
\end{aligned}
$$

распознав коллективизирующее свойство элементов множества, можно неограниченно продолжать их перечисление (продвиньтесь в каждом случае еще на три шага).

В 1683 году Галилей, записав последовательность натуральных чисел $\mathbb{N}$, а под нею последовательность их квадратов (в нашем примере это последовательность B), с недоумением заметил, что приходится признать: квадратов столько же, сколько всех натуральных чисел,– явное противоречие с принципом здравого смысла «целое больше своей части». Великий схоласт Фома Аквинский взывал из XIII столетия каноном Аристотеля: infinitum actu non datur – «актуальная бесконечность не дана», нет никаких бесконечных последовательностей и потому подобные нелепости не могут возникнуть. Но Галилей, так и не разрешив открытого им парадокса, заключил лишь, что в области бесконечного, по-видимому, имеют место какие-то иные, не свойственные конечным совокупностям отношения.

В нашей терминологии наблюдение Галилея означает, что множество натуральных чисел и множество их квадратов равномощны. Множества, равномощные натуральному ряду, называются счетными. Таким образом, счетное множество – это такое множество, все элементы которого можно расположить в виде бесконечной последовательности по типу натурального ряда (приведите пример еще какого-нибудь счетного множества, указав характеристическое свойство его элементов).

Выписанные выше бесконечные последовательности показывают, что счетными являются множество четных натуральных чисел (A), множество простых чисел (P), множество всех целых чисел $\mathbb{Z}$. Если множества A, B, P представляют собой части натурального ряда $\mathbb{N}$, то $\mathbb{Z}$, напротив, содержит в себе $\mathbb{N}$.

Согласно теореме 1, равномощность конечных множеств означает, что у них одинаковое число элементов. Поспешное перенесение такого понимания на случай бесконечных множеств привело бы к предложениям вроде «квадратов столько же, сколько всех натуральных чисел» (Галилей) или «всех целых чисел столько же, сколько натуральных», с которыми, конечно, трудно согласиться. Позднее будет предложена более тонкая формулировка. А пока – первая из серии знаменитых теорем, поразивших математиков конца XIX века.

<u>Теорема 2</u> (Кантор, 1873). Множество $\mathbb{Q}$ рациональных чисел счетно.

ДОКАЗАТЕЛЬСТВО. Рассмотрим сначала множество $\mathbb{Q}^+$ всех положительных рациональных чисел: $\mathbb{Q}^+ = \{x \in \mathbb{Q} | \; x > 0\}$. Каждое такое число можно представить в виде несократимой дроби $\dfrac{m}{n}$, где m и n – натуральные числа. Выпишем сначала все дроби с суммой числителя и знаменателя $m+n$, равной 2 (меньше не может быть), затем все, у которых $m+n=3$, далее такие, где $m+n=4$ и т.д. Получаем:

$$\frac{1}{1};\ \frac{1}{2},\ \frac{2}{1};\ \frac{1}{3},\ \frac{3}{1};\ \frac{1}{4},\ \frac{2}{3},\ \frac{3}{2},\ \frac{4}{1};\ \frac{1}{5},\ \frac{5}{1};\ \frac{1}{6},\ \frac{2}{5},\ \frac{3}{4},\ \frac{4}{3},\ \frac{5}{2},\ \frac{6}{1};\ \ldots$$

(продолжите список, перечислив далее несократимые дроби $\dfrac{m}{n}$ с $m+n=8$ и $m+n=9$). Понятно, что рано или поздно мы доберемся до любого положительного рационального числа.

Итак, множество $\mathbb{Q}^+$ представлено в виде бесконечной последовательности по типу натурального ряда, и значит, является счетным. Действуя теперь, как при перечислении целых чисел: $\mathbb{Z} = \{0, 1, -1, 2, -2, \dots\}$, перенумеруем все рациональные числа:

$$\mathbb{Q} = \{0,\ 1,\ -1,\ \frac{1}{2},\ -\frac{1}{2},\ 2,\ -2,\ \frac{1}{3},\ -\frac{1}{3},\ 3,\ -3,\ \dots\}.$$

Этим завершается доказательство. □

Представим себе числовую прямую. Как отмечалось в § 3, рациональные точки на ней расположены таким образом, что в любом сколь угодно малом отрезке их содержится бесконечно много. Натуральные же числа уходят вправо от нуля, отмечая концы монотонно откладываемой единицы масштаба. И вот, теорема 2 утверждает, что эти множества равномощны! Но оказывается, что и множество всех алгебраических чисел, т.е. множество чисел, являющихся корнями всевозможных алгебраических уравнений с целыми коэффициентами, тоже счетно. А ведь в него, кроме всех рациональных чисел, входят, в частности, все иррациональные числа, записываемые с помощью радикалов. Возникает естественный вопрос: «А может быть, все бесконечные множества счетны, т.е. равномощны натуральному ряду?» В этом предположении тоже есть своя логика, но его опровергает следующая

Теорема 3 (Кантор,1874). Множество $\mathbb{R}$ действительных чисел несчетно.

ДОКАЗАТЕЛЬСТВО. Проводимые далее рассуждения Кантор адресовал нематематикам, выступая в 1891 году на съезде естествоиспытателей.

Любое действительное число можно записать в виде бесконечной десятичной дроби $A, a_1a_2a_3\dots$, где A – это целое число, стоящее слева от запятой (вместе со знаком), а $a_1, a_2, a_3, \dots$ – цифры, выражающие количество соответственно десятых, сотых, тысячных и т.д. долей. Например, для числа $-10\pi = -31,4159\dots$ имеем $A = -31,\ a_1 = 4,\ a_2 = 1,\ a_3 = 5, \dots$.

Предположим, что множество $\mathbb{R}$ счетно. Тогда все действительные числа можно расположить в виде бесконечной последовательности:

$$\begin{gathered} A, a_1a_2a_3a_4\dots, \\ B, b_1b_2b_3b_4\dots, \\ C, c_1c_2c_3c_4\dots, \\ \dots \end{gathered}$$

Теперь определим число $0, x_1x_2x_3x_4\dots$ следующим образом. В качестве первой цифры после запятой, т.е. в качестве x_1, возьмем любую цифру, кроме цифры a_1, обозначающей десятые доли в первом по списку числе (по некоторым математическим причинам здесь и далее не будем использовать также цифры 0 и 9).

В качестве x_2 выберем любую цифру, кроме цифры b_2, выражающей сотые доли во втором числе. Цифра x_3 будет отлична от c_3, а в остальном произвольной (но не 0 и не 9) и т.д. Бесконечная десятичная дробь $0, x_1x_2x_3\ldots$ выражает некоторое действительное число, но его нет в нашем списке! В самом деле, оно отличается от первого числа первой цифрой после запятой, от второго – второй, от третьего – третьей и т.д. Таким образом, предположив, что все действительные числа могут быть расположены в виде бесконечной последовательности, мы тут же указали действительное число, отсутствующее в этой последовательности. Полученное противоречие показывает, что исходное предположение неверно и, следовательно, множество действительных чисел $\mathbb{R}$ несчетно. $\square$

Множество всех действительных чисел (и множество всех точек числовой прямой) называется континуумом (от лат. continuum – непрерывное, сплошное). В силу теоремы 3 континуум дает пример бесконечности нового типа, отличной от счетной бесконечности натурального ряда. Поскольку натуральный ряд весь размещается на числовой прямой, а действительные числа образуют несчетное множество, т.е. их нельзя рассредоточить по натуральному ряду (перенумеровать), можно сказать, что бесконечность континуума имеет более высокую ступень, чем счетная бесконечность.

Георг Кантор (1845–1918), который установил все изложенные в этом параграфе факты, родился в Петербурге. После окончания Берлинского университета с 1869 года преподавал в университете г. Галле. Его работы встретили неприятие со стороны большинства математиков того времени, оно приняло особенно резкие формы после того как Кантор в 1877 году доказал равномощность числовой прямой (одномерное пространство $\mathbb{R}^1$) и числовой плоскости (двумерное пространство $\mathbb{R}^2$), а вскоре и равномощность с ними трехмерного пространства $\mathbb{R}^3$ и любого $n-$мерного пространства $\mathbb{R}^n$. (Возникающую при этом растерянность лучше всего выразить словами самого Кантора: «Я это вижу, но не верю в это»). Теория бесконечных множеств получила всеобщее признание лишь в начале XX века, и теперь этот раздел математики образует фундамент большинства современных теорий.

Представление о равномощности множеств позволило по-новому взглянуть и на само понятие натурального числа. Термин «равномощные множества» имеет лингвистический эквивалент «множества, имеющие равную мощность». Однако что означает слово «мощность» в этом контексте? Классу всех множеств, равномощных некоторому данному множеству A, припишем символ, который будем называть мощностью или кардинальным числом множества A (и каждого множества, равномощного A). Согласно теореме 1, равномощность конечных множеств означает, что они имеют одинаковое число элементов. Следовательно, кардинальное число (мощность) конечного множества A можно отождествить с количеством элементов в A. Кардинальным числом пустого множества является 0, всем одноэлементным множествам соответствует кардинальное число 1, пара весел и чета белеющих на холме берез имеют общее кардинальное число 2 и т.д. Таким образом, натуральные числа – это не что иное, как кардинальные числа конечных множеств (если это определение покажется вам чересчур сложным, вспомните кронекеровское «Бог создал натуральные числа...» или, еще лучше, дефиницию Гегеля: «Число есть *чистая мысль* о самоотчуждении мысли»).

Все вместе натуральные числа составляют натуральный ряд $\mathbb{N}$. Множества, равномощные натуральному ряду, образуют класс счетных множеств. Этому классу приписывается кардинальное число $\aleph_0$ (древнееврейская буква алеф с индексом 0: «алеф-ноль»). Оно символизирует счетную бесконечность.

Свое кардинальное число имеет и континуум, т.е. множество всех действительных чисел $\mathbb{R}$. Мощность континуума обозначается через $\aleph$. Такая же мощность у двумерного континуума $\mathbb{R}^2$ (числовая плоскость) и трехмерного континуума $\mathbb{R}^3$ (арифметическая модель пространства, в котором мы живем). С другой стороны, континуальными множествами, т.е. имеющими мощность континуума $\aleph$, оказались множество всех иррациональных и множество всех трансцендентных чисел, а также любой отрезок числовой прямой (в распространенном нематематическом толковании: точек во всем пространстве $\mathbb{R}^3$ столько же, сколько их на единичном отрезке $[0, 1]$).

Понятно, что элементов в любом конечном множестве меньше, чем натуральных чисел, поэтому для каждого натурального числа n можно написать неравенство $n < \aleph_0$. Замечания, сделанные после доказательства теоремы 3, показывают,что $\aleph_0 < \aleph$. Есть ли какое-нибудь промежуточное кардинальное число между

$\aleph_0$ и $\aleph$, т.е. можно ли на числовой прямой найти такое бесконечное множество, которое не было бы равномощно ни натуральному ряду $\mathbb{N}$, ни всему множеству действительных чисел $\mathbb{R}$? Это знаменитая проблема континуума. В 1900 году, выступая на Международном конгрессе математиков в Париже, Гильберт сформулировал 23 проблемы, затрагивавшие основные направления математики, как они виделись на пороге XX века. Первой из них указана проблема Кантора о мощности континуума. Сам Кантор считал, что промежуточной мощности между мощностью натурального ряда и мощностью континуума нет, и упорно пытался доказать свою гипотезу (бесплодные умственные усилия, связанные с этим, и постоянные воинственные нападки математических противников во главе с Кронекером повергли в конце концов создателя теории множеств в глубокую душевную депрессию). В 1963 году американский математик Пол Джозеф Коэн установил, что при современном состоянии математики проблема континуума не разрешима: существующими средствами гипотезу Кантора нельзя ни доказать, ни опровергнуть. На Международном конгрессе математиков в Москве (1966 г.) Коэну была вручена золотая медаль с профилем Архимеда, присуждаемая раз в четыре года ученым не старше 40 лет за выдающиеся открытия в математике.

Мощность натурального ряда $\aleph_0$ и мощность континуума $\aleph$ были лишь первыми примерами бесконечных кардинальных чисел. Оказалось, что для любого заданного множества существует множество с большей мощностью: для каждой математической бесконечности можно указать бесконечность еще более высокой ступени. Наша интуиция в этой изощренной системе уверенно различает лишь ноль, натуральные числа по отдельности, дискретную бесконечность натурального ряда $\aleph_0$ и непрерывную бесконечность числовой прямой (и трехмерного пространства) $\aleph$, –

И аще кому треба
Счисляти что внутрь неба,
Довлеет числа сего
К вещам всем мира всего.

(В «Арифметике» Магницкого вирши эти относятся не к алефу, а к вполне заурядному числу 10^{24}, которое именуется там – «квадрилион». Но уже Ломоносову, в юности учившему их наизусть, квадрилиона было недостаточно:

Открылась бездна, звезд полна;
Звездам числа нет, бездне дна.)

ГЛАВА II. **МИР ФУНКЦИЙ**

§ 1. Что такое функция?

Слово «функция» в смысле «явление, зависящее от другого и изменяющееся по мере изменения этого другого явления» широко используется в современном общекультурном языке, вполне соответствуя математическому пониманию этого термина. Впервые его употребил Лейбниц в 1694 году, произведя от латинского глагола fungi – выполнять, выражать. Он назвал функциями отрезки линий, связанных с данной точкой на кривой (абсцисса, ордината, касательная и т.д.), меняющиеся при перемещении точки. В 1718 году Иоганн Бернулли завершает период первоначального осознания идеи следующим определением: «Функцией переменной величины называется количество, составленное каким угодно способом из этой переменной величины и постоянных» (т.е. функция, по существу, – это формула, выражающая одну переменную величину через другую). Следующий шаг сделал Эйлер (1755 г.): «Когда некоторые количества зависят от других таким образом, что при изменении последних и сами они подвергаются изменению, то первые называются функциями вторых» (эйлеровское толкование и вошло в общеязыковые словари).

Окончательное математическое уточнение связывают с именем немецкого математика Петера Густава Лежена Дирихле (1837 г.): «y называется функцией от x, если каждому значению x, взятому из некоторого интервала, соответствует определенное значение y». От современного самого общего определения функции по Дедекинду (1887 г.) это отличается лишь тем, что под x и y у Дирихле (так же как и у Бернулли и Эйлера) понимаются числа. Освободившись от этого ограничения, приходим к следующей формулировке одного из важнейших понятий математики.

Пусть X и Y – некоторые непустые множества произвольной природы. Говорят, что на множестве X задана функция f, принимающая значения в множестве Y (запись: $f : X \to Y$), если каждому элементу x из X сопоставлен однозначно определенный элемент $y = f(x)$ из Y. При этом x называется независимой переменной или аргументом, $f(x)$ – значением функции f в точке x, множество X – областью определения функции f, а множество $f(X)$, состоящее из всех y, которые соответствуют хотя бы одному x из X, называют областью значений функции f.

Синонимом слова «функция» является термин «отображение», так что запись $f : X \to Y$ можно прочитать также как «f является отображением множества X в множество Y» или «f отображает X в Y». Если $y = f(x)$, то y называют образом элемента x, а x – прообразом элемента y при отображении f. При отображении $f : X \to Y$ каждый элемент x из области определения X имеет единственный образ, но элемент y из области значений $f(X)$ может иметь много прообразов.

Если X и Y – числовые множества, т.е. состоят из чисел, то, как правило, говорят о функциях, а в случае нечисловых множеств X и Y – об отображениях, но, в общем, четкого разграничения в использовании этих слов нет, в разных областях математики сложились свои традиции.

Если $f : X \to X$, т.е. f отображает множество X в себя, то функцию f называют также преобразованием множества X.

Важнейшими преобразованиями плоскости, т.е. функциями вида $f : \mathbb{R}^2 \to \mathbb{R}^2$, являются вращения (повороты), параллельные переносы (сдвиги), осевые симметрии. Вращение с центром в точке O и углом вращения φ переводит точку плоскости M в точку M' такую, что $OM' = OM$ и $\angle MOM' = \varphi$ (угол отсчитывается против часовой стрелки). При параллельном переносе все точки сдвигаются в одном и том же направлении на одно и то же расстояние. При осевой симметрии каждая точка плоскости переходит в точку, симметричную ей относительно некоторой фиксированной прямой (оси).

Существуют различные способы задания функций. Числовые функции можно задавать с помощью формул, отправляясь от некоторых простейших функций, имеющих индивидуальные обозначения, и знаков арифметических и других операций. Например, $f(x) = x^2 + 3x + 2$, $g(x) = 2^x - 3\lg x$, $h(x) = \dfrac{\sin\ 3x}{1 + \cos\ x}$.

Некоторые функции задаются описанием, как, например, абсолютная величина числа:

$$|x| = \begin{cases} x, & \text{если} \quad x \geq 0; \\ -x, & \text{если} \quad x < 0. \end{cases}$$

При табличном способе задания функции для каждого из возможных значений аргумента x указывается соответствующее ему значение $f(x)$. Например, представлен список студентов группы с указанием дня рождения (число и месяц) каждого. Эта функция отображает указанное множество студентов в множество всех дат года.

Таблицу, задающую преобразование f конечного множества $X = \{x_1, x_2, \ldots, x_n\}$, обычно пишут в две горизонтальные строки:

$$f = \begin{pmatrix} x_1 & x_2 & \ldots & x_n \\ f(x_1) & f(x_2) & \ldots & f(x_n) \end{pmatrix}$$

и называют подстановкой на множестве X.

Перечислим в виде подстановок все преобразования двухэлементного множества $X = \{1,\ 2\}$. Их всего четыре:

$$f_1 = \begin{pmatrix} 1\ 2 \\ 1\ 2 \end{pmatrix}, \quad f_2 = \begin{pmatrix} 1\ 2 \\ 2\ 1 \end{pmatrix}, \quad f_3 = \begin{pmatrix} 1\ 2 \\ 1\ 1 \end{pmatrix}, \quad f_4 = \begin{pmatrix} 1\ 2 \\ 2\ 2 \end{pmatrix}.$$

Для наглядного представления числовых функций f: $\mathbb{R} \to \mathbb{R}$ используют графики, рисуемые в декартовой прямоугольной системе координат на плоскости. Рене Декарт (1596–1650), которому мы обязаны этим замечательным методом, учился в иезуитской коллегии, затем осваивал медицину и право, был офицером, много лет жил в Голландии, где и создал свои основные математические и философские сочинения. В конце 1649 г. по приглашению шведской королевы великий ученый прибыл в Стокгольм, чтобы организовать и возглавить академию наук, но вскоре простудился и умер от воспаления легких. Через 17 лет его останки были перевезены на родину, во Францию. Систему координат он увидел однажды утром в момент пробуждения.

Пусть $f : \mathbb{R} \to \mathbb{R}$ – некоторая числовая функция. Отложив по горизонтальной оси (оси абсцисс) Ox число x, а по вертикальной оси (оси ординат) Oy число $y = f(x)$, на пересечении перпендикуляров, восставленных в этих точках к соответствующим осям, получим точку плоскости с координатами $(x,\ f(x))$. Совокупность всех таких точек и образует график функции f.

Построим, например, график функции $f(x) = |x|$. По определению этой функции, если $x \geqslant 0$, то $f(x) = x$. Значит, откладывая вправо от нуля по оси абсцисс, скажем, число 2, на высоте $f(2) = 2$ над ним получим точку графика нашей функции $(2, 2)$. Аналогично на графике окажутся точки $(0, 0)$, $(3, 3)$, $(5, 5)$ и т.д. – все точки с совпадающими положительными абсциссой и ординатой. При $x < 0$ имеем $f(x) = -x$. Следовательно, откладывая влево от нуля число -2, на высоте $f(-2) = |-2| = 2$ над ним получим точку графика $(-2, 2)$. Аналогично на графике окажутся точки $(-1, 1)$, $(-3, 3)$, $(-5, 5)$ и т.д. – все точки с отрицательной абсциссой и противоположной ей положительной ординатой (рис. 3).

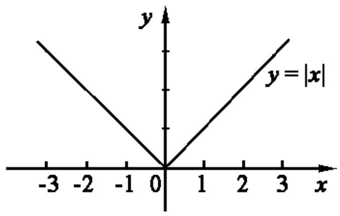


Рис. 3

В качестве упражнения нарисуйте график функции $f(x) = \text{sign } x$ («сигнум x», т.е. по латыни «знак x»), определяемой описанием

$$\text{sign } x = \begin{cases} 1, & \text{если} \quad x > 0; \\ 0, & \text{если} \quad x = 0; \\ -1, & \text{если} \quad x < 0. \end{cases}$$

А как выглядит график функции $f(x) = \text{sign}^2 x = (\text{sign } x)^2$?

Если спроектировать график функции на ось Ox, на ней выделится область определения функции, а проекция графика на ось Oy покажет область изменения. Например, область определения функции $f(x) = |x|$ – вся числовая ось $\mathbb{R}$, а область значений – неотрицательные действительные числа. Область определения функции $\text{sign } x$ – тоже $\mathbb{R}$, но область значений состоит всего из трех чисел: $\{-1,\ 0,\ 1\}$. А у функции $f(x) = \text{sign}^2 x$?

В школе графикам числовых функций уделяется достаточное внимание, в дальнейшем у нас они будут встречаться неоднократно. Чтобы показать, что в этом деле тоже не все просто, попытаемся представить себе график функции Дирихле

$$d(x) = \begin{cases} 1, & \text{если} \quad x \ - \ \text{рациональное число;} \\ 0, & \text{если} \quad x \ - \ \text{иррациональное число.} \end{cases}$$

Значения ее вычисляются легко: $d(1) = 1$, $d(-1,5) = 1$, $d(\sqrt{2}) = 0$, $d(\pi) = 0$ и т.д. Все точки графика расположены либо непосредственно на оси Ox, либо на прямой, параллельной этой оси и поднятой над ней на 1. «Визуально» это будут две параллельные прямые, но на самом деле на нижней прямой имеется счетное множество «дыр» (в рациональных точках), а на верхней число «дыр» континуально (в иррациональных точках). Если совместить эти «сита», получится полная числовая прямая.

Для отображений произвольного вида $f : X \to Y$ в том случае, если множества X и Y имеют небольшое число элементов, иногда рисуют так называемые графы, из точки x проводя стрелку в соответствующую ей точку $y = f(x)$. Например, на рис. 4 множество X объединяет девять выдающихся героев русской классической литературы, множество Y представляет собой список авторов соответствующих произведений, а отображение $f : X \to Y$ указывает создателя каждого из упомянутых персонажей.

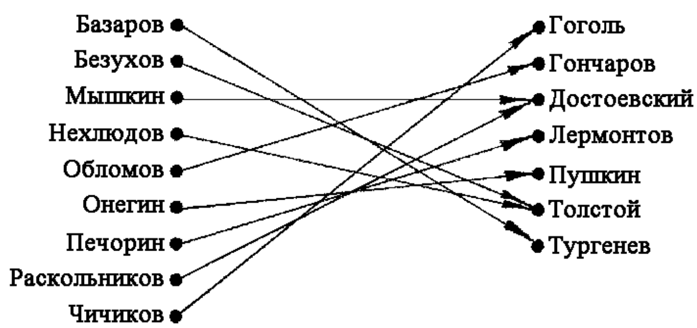


Рис. 4

Нарисуйте аналогичный граф авторства, имея следующий список знакомых вам с детства картин: X={«Бурлаки на Волге», «Владимирка», «Грачи прилетели», «Девятый вал», «Запорожцы пишут письмо турецкому султану», «Золотая осень», «Над вечным покоем», «Последний день. Помпеи», «Рожь», «Утро в сосновом лесу», «Явление Христа народу»} и алфавитный перечень создавших эти произведения художников: Y={Айвазовский, Брюллов, Иванов, Левитан, Репин, Саврасов, Шишкин}. Придайте первому списку такой порядок, чтобы стрелки на диаграмме не пересекались (это сделает ее более наглядной). Аналогичным образом упростите граф на рис. 4. Постройте подобный предыдущим пример для класических русских опер и их авторов.

Графы всех преобразований двухэлементного множества $X = \{1,\ 2\}$ показаны на рис. 5.

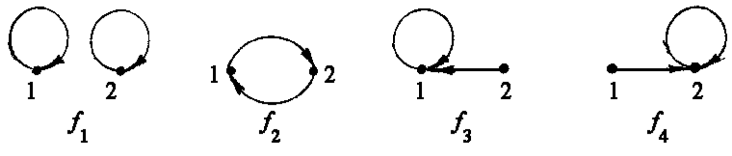


Рис. 5

Отображение $f : X \to Y$ называется сюръективным (или сюръекцией), если $f(X) = Y$, т.е. если область значений функции f совпадает со всем множеством Y: каждый элемент $y \in Y$ является образом некоторого элемента $x \in X$. Говорят еще, что f отображает множество X на множество Y. С давних времен известно сюръективное астрологическое отображение множества всех людей на множество зодиакальных созвездий (останется ли это отображение сюръективным, если ограничиться только множеством студентов вашей группы?). Важнейшей для человечества сюръекцией является его отображение на двухэлементное множество {м, ж} символов пола. Интересно, сколько начальных строф «Евгения Онегина» нужно прочесть, чтобы убедиться, что текст романа (в новой орфографии) отображается на весь русский алфавит?

Функция $y = \sin\ x$ отображает числовую ось $\mathbb{R}$ на отрезок $[-1,\ 1]$. Вообще, всякая функция $f : X \to Y$ является отображением ее области определения X на всю область значений $f(X)$.

Отображение $f : X \to Y$ называется инъективным (или инъекцией), если разным элементам из X оно ставит в соответствие разные элементы из Y, т.е. если при $x' \neq x''$ будет $f(x') \neq f(x'')$. Таким образом, каждый элемент y из области значений $f(X)$ имеет единственный прообраз в X. Инъекции называют также взаимно однозначными отображениями множества X в множество Y. Пересчитывая предметы из некоторого конечного множества (т.е. присваивая им номера), мы устанавливаем инъективное отображение этого множества в натуральный ряд $\mathbb{N}$. Продавая билеты, кассир осуществляет взаимно однозначное отображение множества пассажиров в множество мест в вагонах поезда (как болезненно мы относимся к нарушению этой инъективности – появлению «двойника»!). Если отображение $f : X \to Y$ одновременно сюръективно и инъективно, оно называется биективным (или биекцией). В этом случае говорят также, что f устанавливает взаимно однозначное соответствие между множествами X и Y. Вспомним, что множество X называется равномощным множеству Y, если между X и Y существует взаимно однозначное соответствие. В силу теоремы 1 из § I.5, если множества X и Y конечны и $f : X \to Y$ биекция, то X и Y имеют одинаковое число элементов. Биекциями являются изоморфизмы групп и колец. Вращения, параллельные переносы и осевые симметрии относятся к биективным преобразованиям плоскости.

Важным примером биективной функции является тождественное преобразование Δ произвольного множества X, т.е. отображение $\Delta : X \to X$, оставляющее неподвижной каждую точку: $\Delta(x) = x$.

На рис. 6 представлены пиктограммы, выражающие идею сюръективности (в каждую точку второго множества приходит хотя бы одна стрелка), инъективности (ни в какую точку не приходит больше одной стрелки), биективности (в каждую точку приходит точно одна стрелка) и отсутствие у отображения этих свойств.

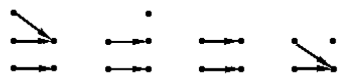


Рис. 6

Для преобразований конечных множеств понятия сюръективности, инъективности и биективности совпадают. Это вытекает из следующего предложения.

Теорема 1. Преобразование f конечного множества X сюръективно тогда и только тогда, когда оно инъективно.

ДОКАЗАТЕЛЬСТВО. 1) Пусть $X = \{x_1, x_2, \ldots, x_n\}$ и $f : X \to X$ – сюръекция. Запишем f в виде подстановки:

$$f = \begin{pmatrix} x_1 & x_2 & \ldots & x_n \\ f(x_1) & f(x_2) & \ldots & f(x_n) \end{pmatrix}.$$

Так как преобразование f сюръективно, во второй строчке стоят все элементы множества X. Их n штук, и они занимают n мест. Значит, никакой из них не повторяется. Это означает, что f инъективно.

2) Пусть $f : X \to X$ – инъекция. Тогда во второй строчке подстановки, соответствующей f, никакой элемент множества X не встречается более одного раза. Но поскольку в этой строчке заняты все n мест, значит, в ней перечислены в каком-то порядке все элементы множества X. Следовательно, f – сюръекция. $\square$

Таким образом, чтобы выяснить, является ли преобразование f конечного множества X биекцией, достаточно только проверить, будет ли f сюръективным, либо исследовать f только на инъективность. Выполнение любого из этих свойств автоматически означает и наличие другого и, значит, биективность преобразования f.

Для бесконечных множеств теорема 1 неверна. Пусть, например, функция $f : \mathbb{N} \to \mathbb{N}$ такова, что $f(x) = x + 1$ для любого натурального числа x. Очевидно, что она инъективна, но сюръективной не является: число 1 не входит в область ее значений. С другой стороны, функция $g : \mathbb{N} \to \mathbb{N}$, заданная описанием

$$g(x) = \begin{cases} 1, & \text{если} \quad x = 1; \\ x - 1, & \text{если} \quad x > 1, \end{cases}$$

сюръективна, но не инъективна: $g(1) = g(2)$.

Подстановки, задающие биективные преобразования конечного множества, называются его перестановками. Выпишите все перестановки трехэлементного множества $\{1,\ 2,\ 3\}$ и нарисуйте соответствующие им графы.

Пусть $f : X \to Y$ и $g : Y \to Z$ – произвольные функции (заметим, что у первой область значений содержится в множестве Y, а у второй Y является областью определения). Функция $h : X \to Z$, определенная для каждого $x \in X$ формулой $h(x) = g(f(x))$, называется суперпозицией (или композицией) функций f и g, или сложной функцией, составленной из f и g, символически: $h = f \circ g$. Таким образом, чтобы определить значение функции h в точке x, нужно сначала найти $f(x)$, а затем, применив к этому элементу из множества Y функцию g, вычислить $g(f(x))$, что и дает $h(x)$.

Например, суперпозицией числовых функций $y = f(x) = x+1$ и $z = g(y) = y^2$ будет $h(x) = g(f(x)) = g(x+1) = (x+1)^2$. Найдите суперпозицию функций $y = f(x) = 2x$ и $z = g(y) = \sin\ y$.

<u>Теорема 2</u>. Суперпозиция сюръективных (инъективных) функций является сюръективной (инъективной) функцией.

ДОКАЗАТЕЛЬСТВО. Пусть функции $f : X \to Y$, $g : Y \to Z$ сюръективны, $h = f \circ g$ – их суперпозиция, а z – произвольный элемент множества Z. Так как g сюръективна, то найдется $y \in Y$ такой, что $g(y) = z$. В свою очередь, для y, в силу сюръективности функции f, существует элемент $x \in X$, для которого $f(x) = y$.

Следовательно, $z = g(y) = g(f(x)) = h(x)$. Таким образом, для любого $z \in Z$ найдется $x \in X$ такой, что $h(x) = z$. Значит, h в самом деле отображает X на Z.

(Пусть теперь f и g инъективны и x', x'' – разные элементы множества X. Так как f инъективна, то $y' = f(x') \neq f(x'') = y''$. В свою очередь, поскольку g инъективна, $g(y') \neq g(y'')$, откуда $g(f(x')) \neq g(f(x''))$, т.е. $h(x') \neq h(x'')$. Следовательно, суперпозиция $h = f \circ g$ инъективна.) $\square$

Следствие. Если f и g – биекции, то их суперпозиция $h = f \circ g$ также является биекцией.

Для преобразований множеств термин «композиция» более употребителен. Композиция $f \circ g$ преобразований множества X – это последовательное выполнение сначала преобразования f, а затем преобразования g. Например, если f – поворот против часовой стрелки на 90° вокруг точки O (начало координат), а g – сдвиг на 1 в положительном направлении оси абсцисс Ox, то композиция $f \circ g$ переведет точку M с координатами $(1,\ 0)$ в точку M' с координатами $(1,\ 1)$, а композиция $g \circ f$ (сначала сдвиг, а потом вращение) передвинет точку M в точку M'' с координатами $(0,\ 2)$. Так что, как видим, композиция преобразований – некоммутативная операция: результат зависит от порядка компонент. Однако композиция преобразований обладает другим важнейшим свойством операций.

Теорема 3. Композиция преобразований ассоциативна, т.е. $(f \circ g) \circ h = f \circ (g \circ h)$ для любых преобразований $f,\ g,\ h$ множества X.

ДОКАЗАТЕЛЬСТВО. Пусть x – произвольный элемент множества X. Посмотрим, во что переводит его каждое из преобразований $(f \circ g) \circ h$ и $f \circ (g \circ h)$. Используя определение композиции преобразований (т.е. определение суперпозиции функций), получаем:

$$((f \circ g) \circ h)(x) = h((f \circ g)(x)) = h(g(f(x))),$$

$$(f \circ (g \circ h))(x) = (g \circ h)(f(x)) = h(g(f(x))).$$

Следовательно, на любой элемент x из X преобразования $(f \circ g) \circ h$ и $f \circ (g \circ h)$ действуют одинаково. Значит, они совпадают. $\square$

Всевозможные композиции вращений, параллельных переносов и осевых симметрий плоскости называются движениями (см. § III.2). Все движения являются биекциями (почему?).

Пусть $f : X \to Y$ – некоторая биективная функция. Обратной для нее функцией называется отображение $f^{-1} : Y \to X$, сопоставляющее каждому $y \in Y$ тот единственный $x \in X$, которому f соотносит y, т.е. если $y = f(x)$, то $x = f^{-1}(y)$. Когда вам выписывают студенческие билеты, осуществляется биективное отображение множества студентов первого курса на множество выделенных для них книжечек, а при вручении студенческих билетов используется обратное отображение: по готовой книжечке вызывается ее владелец. Очевидно, что $f \circ f^{-1} = \Delta_X$ (тождественная функция на множестве X), т.е. если для $x \in X$ сначала выполнить отображение f, а затем для $f(x)$ – обратное отображение f^{-1}, то мы вернемся в исходную точку x множества X. Аналогично $f^{-1} \circ f = \Delta_Y$ (тождественная функция на множестве Y).

Теорема 4. Совокупность $S(X)$ всех взаимно однозначных преобразований (биекций) произвольного множества X образует группу относительно композиции преобразований.

ДОКАЗАТЕЛЬСТВО. Согласно следствию из теоремы 2, совокупность $S(X)$ всех биекций множества X замкнута относительно композиции. По теореме 3 композиция преобразований ассоциативна. Тождественное преобразование Δ является биекцией и, кроме того, нейтральным элементом относительно композиции: если f – произвольное преобразование множества X, то $f \circ \Delta = \Delta \circ f = f$ (проделайте выкладки). Очевидно, что если f – биекция, то обратное преобразование f^{-1} – тоже биекция. Таким образом, для совокупности $S(X)$ всех биекций множества X и определенной на ней операции – композиции биекций – выполнены все условия теоремы 1 из § I.2. Следовательно, $(S(X), \circ)$ – группа. $\square$

Группа $S(X)$ называется симметрической группой на множестве X. Симметрическая группа на двухэлементном множестве $\{1,\ 2\}$ обозначается через S_2, она состоит из двух перестановок: $f_1=\begin{pmatrix}1\ 2\\1\ 2\end{pmatrix}$ и $f_2=\begin{pmatrix}1\ 2\\2\ 1\end{pmatrix}$. При этом $f_1 \circ f_1 = f_2 \circ f_2 = f_1$, $f_1 \circ f_2 = f_2 \circ f_1 = f_2$. Симметрическая группа S_3 на трехэлементном множестве $\{1,\ 2,\ 3\}$ содержит 6 элементов, это перестановки $\pi_1=\begin{pmatrix}1\ 2\ 3\\1\ 2\ 3\end{pmatrix}$, $\pi_2=\begin{pmatrix}1\ 2\ 3\\1\ 3\ 2\end{pmatrix}$, $\pi_3=\begin{pmatrix}1\ 2\ 3\\3\ 2\ 1\end{pmatrix}$, $\pi_4=\begin{pmatrix}1\ 2\ 3\\2\ 1\ 3\end{pmatrix}$, $\pi_5=\begin{pmatrix}1\ 2\ 3\\2\ 3\ 1\end{pmatrix}$, $\pi_6=\begin{pmatrix}1\ 2\ 3\\3\ 1\ 2\end{pmatrix}$. Симметрическая группа S_3 некоммутативна (т.е. не является коммутативной). В самом деле, так как $(\pi_2 \circ \pi_3)(1) = \pi_3(\pi_2(1)) = \pi_3(1) = 3$, $(\pi_2 \circ \pi_3)(2) = \pi_3(\pi_2(2)) = \pi_3(3) = 1$, $(\pi_2 \circ \pi_3)(3) = \pi_3(\pi_2(3)) = \pi_3(2) = 2$, то $\pi_2 \circ \pi_3=\begin{pmatrix}1\ 2\ 3\\3\ 1\ 2\end{pmatrix}=\pi_6$, но

$\pi_3 \circ \pi_2 = \pi_5$ (проверьте). Аддитивная группа кольца Z_6, которую образуют остатки от деления на 6 при сложении по модулю 6, имеет столько же элементов, что и симметрическая группа S_3, но в отличие от S_3 она абелева. Следовательно, группы $(S_3, \circ)$ и $(Z_6, +)$ не изоморфны.

В 1854 году английский математик Артур Кэли (1821–1895) доказал, что всякая группа изоморфна группе взаимно однозначных отображений некоторого множества на себя. В частности, элементы каждой конечной группы можно интерпретировать как перестановки на некотором множестве, а групповую операцию – как композицию перестановок.

Чтобы найти перестановку, обратную для данной перестановки f, нужно в f поменять местами первую и вторую строчки, а затем упорядочить элементы так, чтобы в первой строчке они стояли в стандартной последовательности. Например, в симметрической группе S_3

$$\pi_5{}^{-1} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 2 & 3 & 1 \\ 1 & 2 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix} = \pi_6.$$

Найдите обратные для остальных элементов группы S_3.

Составим теперь таблицу умножения симметрической группы $(S_3, \circ)$ (табл. 1).

Таблица 1

$\circ$	π_1	π_2	π_3	π_4	π_5	π_6
π_1	π_1	π_2	π_3	π_4	π_5	π_6
π_2	π_2	π_1	π_6	π_5	π_4	π_3
π_3	π_3	π_5	π_1	π_6	π_2	π_4
π_4	π_4	π_6	π_5	π_1	π_3	π_2
π_5	π_5	π_3	π_4	π_2	π_6	π_1
π_6	π_6	π_4	π_2	π_3	π_1	π_5

Таблица 2

$\circ$	Δ	s_1	s_2	s_3	t_1	t_2
Δ	Δ	s_1	s_2	s_3	t_1	t_2
s_1	s_1	Δ	t_2	t_1	s_3	s_2
s_2	s_2	t_1	Δ	t_2	s_1	s_3
s_3	s_3	t_2	t_1	Δ	s_2	s_1
t_1	t_1	s_2	s_3	s_1	t_2	Δ
t_2	t_2	s_3	s_1	s_2	Δ	t_1

Понятие группы появилось у нас в главе I в связи с вопросом о разрешимости числовых уравнений. Группы преобразований, т.е. функций, биективно отображающих произвольное множество на себя, нашли применение в связи с понятием симметрии плоских и пространственных фигур.

Пусть Φ – некоторая плоская фигура. Биективное преобразование плоскости $f : \mathbb{R} \to \mathbb{R}$ называется симметрией фигуры Φ, если оно переводит эту фигуру в себя: $f(\Phi) = \Phi$.

Теорема 5. Симметрии каждой плоской фигуры образуют группу.

ДОКАЗАТЕЛЬСТВО. Если f и g – симметрии плоской фигуры Φ, то $f(\Phi) = \Phi = g(\Phi)$, откуда $(f \circ g) = g(f(\Phi)) = g(\Phi) = \Phi$ и, значит, композиция $f \circ g$ также является симметрией фигуры Φ (здесь использовано еще следствие из теоремы 2, в силу которого $f \circ g$ будет биекцией). Очевидно, что тождественное преобразование Δ входит в число симметрий любой фигуры и что если f – симметрия фигуры Φ, то обратное преобразований f^{-1} также будет ее симметрией. Согласно теореме 1 из §I.2, совокупность всех симметрий фигуры Φ представляет собой группу относительно композиции преобразований. □

В практике под симметриями плоской фигуры обычно понимают ее симметрии, являющиеся движениями (жесткие симметрии), они также образуют группу. Рассмотрим одну из таких групп.

Пусть Φ – равносторонний треугольник с вершинами $1,\ 2,\ 3$ (немного непривычно, но удобно для дальнейшего) и центром O (рис. 7).

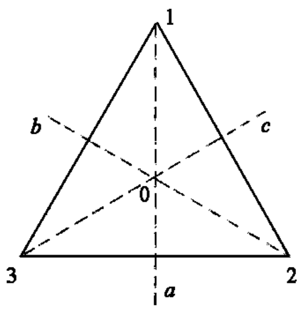


Рис. 7

Поворот t_1 плоскости вокруг точки O на 60° по часовой стрелке совместит треугольник с собой и, значит, будет симметрией этой фигуры. Ее симметриями являются также поворот t_2 с тем же центром на 120° по часовой стрелке и зеркальные отражения s_1, s_2, s_3 относительно осей Oa, Ob, Oc, проходящих через точку O и вершины треугольника. Присоединив к этому списку тождественную симметрию Δ, получим шестиэлементную группу симметрий правильного треугольника. Составим ее таблицу умножения. Что такое, например, $t_1 \circ s_2$? Сначала нужно повернуть треугольник на 60° по часовой стрелке, а затем произвести отражение относительно оси Ob (считаем, что оси симметрии неподвижны). Результирующее преобразование совпадет с симметрией s_3. Заметим, что t_1 и t_2 являются взаимно обратными элементами группы, поскольку $t_1 \circ t_2 = t_2 \circ t_1 = \Delta$, а каждое из отражений $s_1,\ s_2,\ s_3$ обратно самому себе: $s_1 \circ s_1 = s_2 \circ s_2 = s_3 \circ s_3 = \Delta$. Группа (жестких) симметрий правильного треугольника описывается табл. 2.

Внимательный читатель заметил, по-видимому, что каждая симметрия правильного треугольника связана с некоторой перестановкой на множестве его вершин. Например, вращению t_1 соответствует перестановка $\begin{pmatrix}1 & 2 & 3\\ 2 & 3 & 1\end{pmatrix}$=π$_5$, а отражению s_2 – перестановка $\begin{pmatrix}1 & 2 & 3\\ 3 & 2 & 1\end{pmatrix}$=π$_3$. С другой стороны, перестановка π$_4$=$\begin{pmatrix}1 & 2 & 3\\ 2 & 1 & 3\end{pmatrix}$ описывает отражение s_3 и т.д. Таким образом, между элементами группы симметрий правильного треугольника и перестановками трехэлементного множества устанавливается взаимно однозначное соответствие: $\Delta \leftrightarrow \pi_1$, $s_1 \leftrightarrow \pi_2$, $s_2 \leftrightarrow \pi_3$, $s_3 \leftrightarrow \pi_4$, $t_1 \leftrightarrow \pi_5$, $t_2 \leftrightarrow \pi_6$. Более того, произведя соответствующую замену букв в таблице умножения группы симметрий, мы получим таблицу умножения группы перестановок. Значит, эти группы изоморфны. Итак, группа симметрий правильного треугольника – это симметрическая группа S_3.

Изучая плоские орнаменты и кристаллические структуры, геолог академик Евграф Степанович Федоров (1853–1919) провел их классификацию по группам симметрий (так называемые кристаллографические группы). Он установил, что на плоскости существует 17 различных конфигураций, а в пространстве – 219. Результаты Е.С.Федорова явились замечательным примером нематематического приложения теории групп.

Группы с произвольными элементами (т.е. не являющимися числами или преобразованиями) и не обязательно конечные впервые исследовал в своей монографии «Абстрактная теория групп» (1916 г.) Отто Юльевич Шмидт (1891–1956), основатель московской алгебраической школы, прославившийся впоследствии своей многогранной деятельностью: математик, астроном, исследователь Арктики, главный редактор Большой Советской энциклопедии, создатель космогонической гипотезы о происхождении Солнечной системы. Именно с появлением книги О.Ю.Шмидта теория групп стала самостоятельной ветвью математики.

§ 2. Элементарные функции

Начиная с этого параграфа и до конца главы будем рассматривать только числовые функции, т.е. функции $f : X \to Y$, где X и Y – некоторые множества действительных чисел.

Пусть $f : X \to Y$ и $g : X \to Y$ – числовые функции с одинаковой областью определения X. Их суммой называется функция $f + g : X \to Y$ такая, что для любого $x \in X$

$$(f + g)(x) = f(x) + g(x),$$

т.е. значение функции-суммы в каждой точке равно сумме соответствующих значений функций-слагаемых.

Аналогично формулами

$$(f-g)(x)=f(x)-g(x),$$

$$(fg)(x)=f(x)g(x),$$

$$\left(\frac{f}{g}\right)(x)=\frac{f(x)}{g(x)}$$

определяются разность, произведение и частное двух функций (в последнем случае функция g должна быть такой, чтобы $g(x)\neq 0$ в любой точке $x\in X$). Тем самым для числовых функций введены арифметические операции сложения, вычитания, умножения и деления. В § 1 была определена также операция суперпозиции функций.

Приступая к изложению классических идей математического анализа, нужно сначала ответить на вопрос: анализом чего же занимается эта наука? В тех рамках, которыми ограничены наши возможности, математическому анализу подвергаются элементарные функции. Эти функции чаще всего встречаются в приложениях математики, с ними связаны наши интуитивные представления о процессах, происходящих в окружающем мире.

Прежде чем описать класс элементарных функций, т.е. определить, какие функции мы будем называть элементарными, рассмотрим некоторые важнейшие примеры. Для удобства будем обозначать область определения функции f через $D(f)$, а область ее значений символом $E(f)$.

I. Постоянные функции. Постоянными функциями называются функции вида $f(x)=c$, где c – некоторое фиксированное число. Область определения любой постоянной функции совпадает со всей числовой прямой: $D(f)=\mathbb{R}$, а область значений состоит из одной точки c, т.е. $E(f)=\{c\}$. График постоянной функции представляет собой прямую, параллельную оси абсцисс Ox и отстоящую от нее на c. На рис. 8 представлены графики постоянных функций $f(x)=1$ (сплошная линия) и $f(x)=-1$ (пунктир).

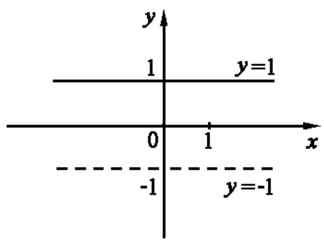


Рис. 8

II. Степенные функции. Степенными функциями называются функции вида $f(x) = x^{\alpha}$, где α – некоторое отличное от нуля фиксированное число. В общем виде область определения и область значений указать невозможно из-за большого разнообразия степенных функций. Рассмотрим некоторые конкретные примеры.

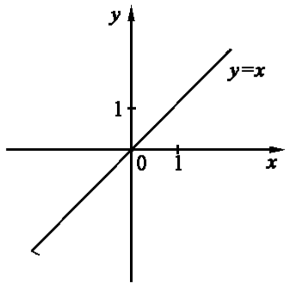


Рис. 9

1) $\alpha = 1$. В этом случае $f(x) = x$ (тождественная функция), $D(f) = \mathbb{R}$, $E(f) = \mathbb{R}$. График – прямая линия, биссектриса первого и третьего координатных углов (рис. 9).

2) $\alpha = 2$. В этом случае $f(x) = x^2$, $D(f) = \mathbb{R}$, $E(f) = \{x \in \mathbb{R} \mid x \geqslant 0\}$. График строится по точкам в соответствии с таблицей:

x	0	1	-1	2	-2	$\dots$
$y = x^2$	0	1	1	4	4	$\dots$

Во второй строчке встречаются повторы – функция $f(x) = x^2$ не является инъективной (взаимно однозначной). Кривая, изображенная на рис. 10а, называется (квадратичной) параболой. Примерно так же выглядит любая парабола четной степени – график функции $f(x) = x^n$ для четного $n > 0$. С увеличением n линия все теснее прижимается к оси Ox на интервале $(-1,\ 1)$ и все круче взмывает вверх при движении x вправо от точки 1 и влево от точки -1. Попробуйте представить себе параболу $y = x^{1000}$.

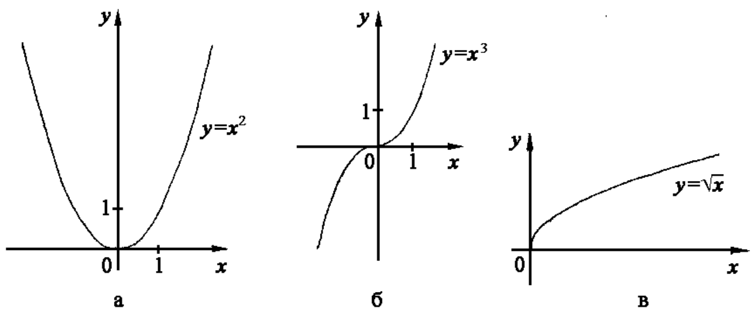


Рис. 10

3) $\alpha = 3$. В этом случае $f(x) = x^3$, $D(f) = \mathbb{R}$, $E(f) = \mathbb{R}$. График строится по точкам в соответствии с таблицей:

x	0	1	-1	2	-2	3	-3	...
$y = x^3$	0	1	-1	8	-8	27	-27	...

Разным значениям аргумента соответствуют разные значения функции: $f(x) = x^3$ осуществляет взаимно однозначное отображение числовой прямой на себя, т.е. является ее биективным преобразованием. Ее график – кривая, изображенная на рис. 10б,– называется кубической параболой. Примерно так же выглядит любая парабола нечетной степени – график функции $f(x) = x^n$ для нечетного $n > 1$. С увеличением n линия все теснее прилегает к оси Ox на интервале $(1,\ -1)$, все круче падает вниз на участке левее точки $x = -1$ и все круче взмывает вверх на участке правее точки $x = 1$.

4) $\alpha = \dfrac{1}{2}$. В этом случае $f(x) = x^{\frac{1}{2}} = \sqrt{x}$, $D(f) = \{\, x \in \mathbb{R} \mid x \geqslant 0 \,\}$ (в действительной области квадратный корень из отрицательного числа не существует), $E(f) = \{\, y \in \mathbb{R} \mid y \geq 0 \,\}$ (неправильно писать $\sqrt{4} = \pm 2$, при извлечении корня всегда получается неотрицательное число!). График (рис. 10в) строится по точкам в соответствии с таблицей:

x	0	1	4	9	16	...
$y = \sqrt{x}$	0	1	2	3	4	...

Это половина квадратичной параболы, расположенной вдоль оси Ox.

5) $\alpha = -1$. В этом случае $f(x) = x^{-1} = \frac{1}{x}$, $D(f) = \{\, x \in \mathbb{R} \mid x \neq 0 \,\}$ (нельзя делить на ноль), $E(f) = \{\, y \in \mathbb{R} \mid y \neq 0 \,\}$. График строится по точкам в соответствии с таблицей:

x	$\frac{1}{3}$	$-\frac{1}{3}$	$\frac{1}{2}$	$-\frac{1}{2}$	1	-1	2	-2	3	-3	...
$y = \frac{1}{x}$	3	-3	2	-2	1	-1	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{1}{3}$	$-\frac{1}{3}$	...

Полученная линия (рис. 11) называется гиперболой. При неограниченном удалении точки по оси Ox от начала координат соответствующая точка $f(x)$ на графике асимптотически (т.е. неограниченно) приближается к этой оси. Если же точка x движется в сторону начала координат, соответствующая точка на графике асимптотически приближается к оси Oy. Говорят, что прямые $y = 0$ и $x = 0$ (оси координат) являются асимптотами гиперболы $y = \dfrac{1}{x}$.

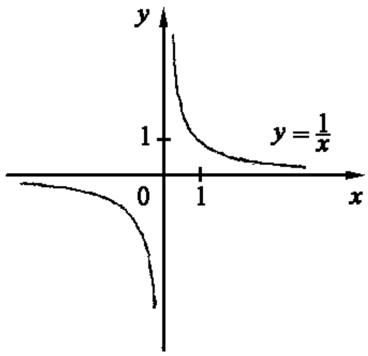


Рис. 11

Функция $f(x) = \dfrac{1}{x}$ биективна и совпадает со своей обратной функцией. Следовательно, $f \circ f = \Delta$ (тождественная функция), т.е. $(x^{-1})^{-1} = x$.

Нарисуйте по точкам графики степенных функций $f(x) = x^{\frac{1}{3}}$ (обратная для функции $y = x^3$), $f(x) = x^{-2}$.

Функция $f : X \to Y$ называется четной, если $f(-x) = f(x)$ для любого $x \in X$, т.е. если она принимает одинаковые значения в точках области определения, симметричных относительно начала координат O. График четной функции симметричен относительно оси ординат Oy. Таковы, например, все параболы четных степеней, функция $f(x) = x^{-2}$. Четными являются и все постоянные функции.

Функция $f : X \to Y$ называется нечетной, если $f(-x) = -f(x)$, т.е. если $f(x)$ изменяет знак с изменением знака аргумента. График нечетной функции симметричен относительно начала координат: если точка (a, b) лежит на графике, то ему принадлежит и точка $(-a, -b)$. Степенные функции $f(x) = x^n$ с нечетным целым показателем n нечетны.

Понятно, что когда мы говорим о четности или нечетности функции, ее область определения предполагается симметричной относительно нуля.

III. Показательные функции. Показательными функциями называются функции вида $f(x) = a^x$, где $a > 0$, $a \neq 1$. При любом основании a область определения $D(f) = \mathbb{R}$, область значений $E(f) = \mathbb{R}^+$ (положительные числа). Все показательные функции являются биекциями. Основные свойства показательной функции выражаются тождествами:

$$1)\quad a^{x+y} = a^x \cdot a^y,$$

$$2)\quad a^{-x} = \frac{1}{a^x},$$

$$3)\quad a^0 = 1.$$

Функция f называется возрастающей, если с увеличением аргумента x значение функции $f(x)$ не убывает, т.е. если при движении точки x вправо по оси абсцисс соответствующая точка $f(x)$ на графике не снижается. Возрастающими функциями будут степенные функции $f(x) = x^n$ при целых нечетных n, функция $f(x) = \sqrt{x}$. Возрастает и показательная функция $f(x) = a^x$ при $a > 1$. Постоянные функции тоже приходится признать возрастающими, как и функцию $\operatorname{sign} x$.

Функция f называется убывающей, если с увеличением аргумента x значение $f(x)$ функции не возрастает, т.е. если при движении точки x вправо по оси абсцисс соответствующая точка $f(x)$ на графике не поднимается. Убывающей функцией будет $f(x) = \dfrac{1}{x}$ (двигаясь по гиперболе вправо, точка все время идет вниз), а также любая показательная функция $f(x) = a^x$ при $a < 1$ (чем в большую степень возводится дробь, тем меньше результат). Всякая постоянная функция является убывающей. Функции могут на одних участках области определения возрастать, а на других убывать, как, например, $f(x) = x^2$.

Построим по точкам типичные графики показательных функций: $f(x) = 2^x$ и $f(x) = \left(\dfrac{1}{2}\right)^x$, пользуясь таблицами:

x	-2	-1	0	1	2	$\ldots$
2^x	$\frac{1}{4}$	$\frac{1}{2}$	1	2	4	$\ldots$

x	-2	-1	0	1	2	$\ldots$
$(\frac{1}{2})^x$	4	2	1	$\frac{1}{2}$	$\frac{1}{4}$	$\ldots$

График первой из этих функций изображен на рис. 12 сплошной линией, а второй – пунктирной.

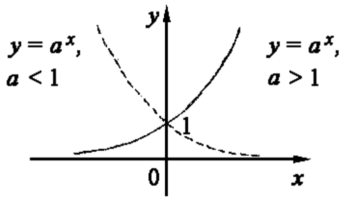


Рис. 12

Имея в виду Теорему 1, дадим точное математическое определение изоморфизма двух групп (на «наивном» уровне мы обсуждали его в § I.2). Изоморфизмом группы $(G, \circ)$ с операцией $\circ$ на группу $(H, *)$ с операцией $*$ называется взаимно однозначное отображение $f : G \to H$ множества G на множество H такое, что $f(x \circ y) = f(x) * f(y)$ для любых $x, y \in G$.

Теорема 1. Каждая показательная функция является изоморфизмом аддитивной группы $(\mathbb{R}, +)$ всех действительных чисел на мультипликативную группу $(\mathbb{R}^+, \cdot)$ положительных чисел.

ДОКАЗАТЕЛЬСТВО. Пусть $f(x) = a^x$ – произвольная показательная функция. Она взаимно однозначно отображает числовую прямую $\mathbb{R}$ на ее положительную часть $\mathbb{R}^+$. При этом

$$f(x + y) = a^{x+y} = a^x \cdot a^y = f(x)f(y).$$

Таким образом, если каждому действительному числу x сопоставить положительное число a^x и вместо знака сложения $+$ писать знак умножения $\cdot$, таблица сложения аддитивной группы $(\mathbb{R}, +)$ превратится в таблицу умножения мультипликативной группы $(\mathbb{R}^+, \cdot)$ (мы здесь, разумеется, условно говорим о таблицах: оба множества $\mathbb{R}$ и $\mathbb{R}^+$ континуальны). Это и означает, что рассматриваемые группы изоморфны. □

Итак, с алгебраической точки зрения две разительно несхожие структуры: аддитивная группа $(\mathbb{R}, +)$ и мультипликативная группа $(\mathbb{R}^+, \cdot)$,– могут рассматриваться как две реализации одной и той же абстрактной группы. Изоморфизм $f(x) = a^x$ «переводит» терминологию, связанную со сложением, в соответствующие понятия на языке умножения: $f(0) = a^0 = 1$ – образом нуля, нейтрального элемента по сложению, является 1 – нейтральный элемент относительно умножения; $f(-x) = a^{-x} = \dfrac{1}{a^x} = (a^x)^{-1} = [f(x)]^{-1}$ – образом противоположного элемента является обратный элемент и т.д. (Анри Пуанкаре: «Математика – это способ называть разные вещи одним именем»).

IV. Логарифмические функции. Логарифмическими называются функции, обратные для показательных. Функция $f(x) = \log_a x$, где $a > 0$, $a \neq 1$, каждому $x \in \mathbb{R}^+ = D(f)$ сопоставляет $y \in \mathbb{R} = E(f)$ такой, что $x = a^y$, т.е. логарифм числа x по основанию a – это показатель степени y, в которую нужно возвести a, чтобы получить x.

Основные свойства логарифмов выражаются тождествами

1) $\log_a(xy) = \log_a x + \log_a y$,

2) $\log_a \dfrac{1}{x} = -\log_a x$,

3) $\log_a 1 = 0$.

Алгебраические равенства $f \circ f^{-1} = \Delta$ и $f^{-1} \circ f = \Delta$, связывающие функцию с ее обратной, будучи применены к показательной и логарифмической функциям с одинаковым основанием, превращаются соответственно в тождества $\log_a a^x = x$ и $a^{\log_a x} = x$ («основное логарифмическое тождество»).

Логарифмическая функция $f(x) = \log_a x$ при $a > 1$ возрастает, а при $a < 1$ убывает.

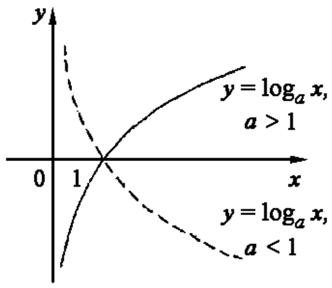


Рис. 13

Построим по точкам типичные графики логарифмических функций: $f(x) = \log_2 x$ и $f(x) = \log_{\frac{1}{2}} x$ (рис. 13), пользуясь таблицами:

x	$\frac{1}{4}$	$\frac{1}{2}$	1	2	4	...
$y = \log_2 x$	-2	-1	0	1	2	...

x	$\frac{1}{4}$	$\frac{1}{2}$	1	2	4	...
$y = \log_{\frac{1}{2}} x$	2	1	0	-1	-2	...

Специальные обозначения имеют логарифмы по основаниям 10 (десятичные логарифмы $\lg x$), 2 (двоичные логарифмы $\log x$) и $e \approx 2,71828$ (натуральные логарифмы $\ln x$). Десятичные логарифмы применяются в вычислительной практике, с двоичными логарифмами мы встретимся в теории информации (§ V.5), натуральные логарифмы играют основную роль в математическом анализе и его приложениях.

Логарифмы в начале XVIII века независимо друг от друга открыли шотландский лорд Непер и швейцарский часовщик Бюрги́. Первый стремился облегчить себе вычисления, связанные с астрологическими предсказаниями на основе Апокалипсиса, второму хотелось упростить расчеты, которые он проводил для Кеплера в астрономической обсерватории в Праге.

Логарифмы являются изоморфизмами мультипликативной группы $(\mathbb{R}^+, \cdot)$ положительных чисел на аддитивную группу $(\mathbb{R}, +)$ всех действительных чисел. Преобразуя действия, связанные с умножением чисел, в более простые для счета манипуляции с суммами и разностями, они сыграли революционную роль в естественнонаучных и технических приложениях математики. Как говорил Лаплас (1749–1827), «изобретение логарифмов, сокращая вычисления, словно удвоило жизнь астрономов».

Логарифмы сыграли неоценимую роль в создании современной двенадцатитоновой музыкальной шкалы. Настройка музыкальных инструментов по чистым интервалам (терция, квинта и т.п.) не давала возможности естественным образом переходить внутри произведения из одной тональности в другую. На рубеже XVII– XVIII веков немецкий теоретик музыки Андреас Веркмейстер предложил разделить октаву на 12 частей таким образом, чтобы разность двоичных логарифмов соседних частот была постоянной. Для рояля, настроенного в соответствии с этим принципом равномерной темперации, Иоганн Себастьян Бах (1685–1750) сочинил два тома прелюдий и фуг под названием «Хорошо темперированный клавир», где убедительно продемонстрировал преимущества новой системы, принятой после этого всеми композиторами.

V. Тригонометрические функции. Тригонометрическими функциями называются функции $\sin x,\ \cos x,\ \operatorname{tg} x,\ \operatorname{ctg} x$ (синус, косинус, тангенс и котангенс). Главной из них является $f(x) = \sin x$. Рассмотрим окружность с центром в начале координат и радиусом, равным единице (рис. 14).

Возьмем на окружности точку A и проведем в нее радиус OA. Тем самым определяется угол α между осью абсцисс Ox и отрезком OA. При этом величина угла считается положительной, если он отсчитывается от оси Ox против часовой стрелки, и отрицательной при движении точки A вдоль окружности по часовой стрелке. Синусом угла α называется ордината точки A, т.е. длина (со знаком) отрезка AB, а косинусом этого угла, по определению, считается абсцисса этой точки, т.е. длина (со знаком) отрезка OB (латинское sinus означает изгиб, извив). Величину угла α выражают не в градусах, а как длину дуги единичной окружности, на которую этот угол опирается. Длина всей окружности равна 2π, ей соответствует полный угол в 360°, половина длины окружности π символизирует развернутый угол в 180°, четверть длины окружности $\frac{\pi}{2}$ – это прямой угол 90°, восьмая часть $\frac{\pi}{4}$ выражает угол в 45° и т.д. Угол, соответствующий дуге длины 1, называется радианом, он равен примерно $57,3^\circ$. Переход от градусной меры углов к радианной сделал π главным числом тригонометрии, ее символом.

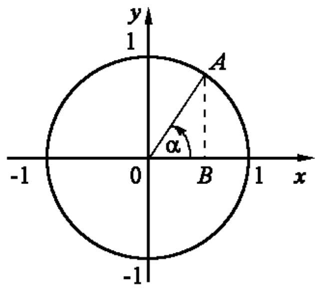


Рис. 14

Глядя на рис. 14 и представляя себе точку A в соответствующих положениях, без труда находим: $\sin 0 = 0$, $\sin\frac{\pi}{2} = 1$, $\sin\pi = 0$, $\sin\frac{3\pi}{2} = -1$, $\sin 2\pi = 0$. Если $\alpha = \frac{\pi}{4}(= 45^\circ)$, то $\triangle\, AOB$ – прямоугольный и равнобедренный (т.е. $AB = OB$) и, значит, по теореме Пифагора, $OA^2 = AB^2 + OB^2 = 2AB^2$, т.е. $1 = 2(\sin\frac{\pi}{4})^2$, откуда $\sin\frac{\pi}{4} = \sqrt{\frac{1}{2}} = \frac{\sqrt{2}}{2}$. При $\alpha = \frac{\pi}{6}(= 30^\circ)$ вспоминаем школьное заклинание: «катет, лежащий против угла в тридцать градусов, равен половине гипотенузы», что в нашем случае дает $AB = \frac{1}{2}OA$, т.е. $\sin\frac{\pi}{6} = \frac{1}{2}$. Завершив полный оборот, точка A возвращается в исходное положение, поэтому $\sin(\alpha + 2\pi) = \sin\alpha$, т.е. все значения функции $\sin x$ периодически повторяются, и период равен 2π.

Синус – нечетная функция: $\sin(-x) = -\sin x$. Построим по точкам график синуса (рис. 15), пользуясь таблицей:

x	0	$\frac{\pi}{6}$	$-\frac{\pi}{6}$	$\frac{\pi}{2}$	$-\frac{\pi}{2}$	$\frac{5\pi}{6}$	$-\frac{5\pi}{6}$	π	$-\pi$	…
$y = \sin x$	0	$\frac{1}{2}$	$-\frac{1}{2}$	1	-1	$\frac{1}{2}$	$-\frac{1}{2}$	0	0	…

Синусоида – одна из самых известных математических линий. Она соответствует нашему представлению о периодическом волнообразном движении. Первым нарисовал ее (в 1636 году) Роберваль, учитель Паскаля. Для функции $f(x) = \sin x$ имеем $D(f) = \mathbb{R}$, $E(f) = [-1, 1]$, (т.е. $-1 \leqslant \sin x \leqslant 1$ для любого $x \in \mathbb{R}$).

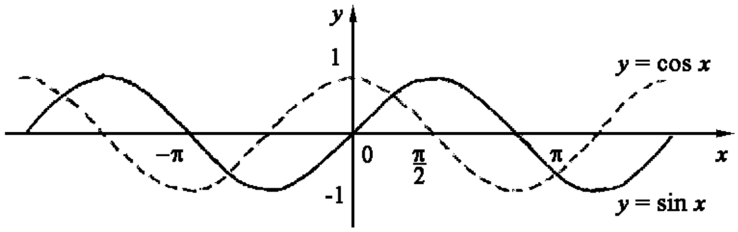


Рис. 15

Используя определение косинуса угла, составляем таблицу значений для функции $f(x) = \cos x$:

x	0	$\frac{\pi}{3}$	$-\frac{\pi}{3}$	$\frac{\pi}{2}$	$-\frac{\pi}{2}$	$\frac{2\pi}{3}$	$-\frac{2\pi}{3}$	π	$-\pi$	...
$y = \cos x$	1	$\frac{1}{2}$	$\frac{1}{2}$	0	0	$-\frac{1}{2}$	$-\frac{1}{2}$	-1	-1	...

Графиком косинуса является синусоида, сдвинутая влево на $\frac{\pi}{2}$ (т.е. $\cos x = \sin(x + \frac{\pi}{2})$), она показана на рис. 15 пунктиром. У $f(x) = \cos x$, как и у синуса, $D(f) = \mathbb{R}$, $E(f) = [-1, 1]$. Косинус – четная функция: $\cos(-x) = \cos x$.

Синус и косинус любого угла α связаны соотношением $\sin^2 x + \cos^2 x = 1$, которое называется основным тригонометрическим тождеством. Оно непосредственно усматривается из рис. 14: $\sin^2 \alpha + \cos^2 \alpha = AB^2 + OB^2 = OA^2 = 1$ (мы использовали теорему Пифагора и тот факт, что $OA = 1$).

Тригонометрические функции тангенс и котангенс вводятся формулами $\operatorname{tg} x = \frac{\sin x}{\cos x}$, $\operatorname{ctg} x = \frac{\cos x}{\sin x}$. Они играют вспомогательную, техническую роль, упрощая запись тригонометрических выражений. График тангенса показан на рис. 16. Он состоит из бесконечного множества одинаковых частей (ветвей), повторяющихся с периодом π. Пунктирные вертикальные линии являются асимптотами. Функция $y = \operatorname{tg} x$ нечетная, она возрастает на каждой части своей распадающейся области определения. Одна ветвь тангенса называется тангенсоидой.

Тригонометрией называется раздел математики, изучающий зависимость между сторонами и углами треугольника, а также соотношения между возникающими при этом функциями. Она появилась из чисто практических потребностей. Основоположником ее считается знаменитый астроном Гиппарх из Никеи (именно он во II веке до н.э. ввел географические координаты – широту и долготу). Существенно продвинувшись в трудах восточных энциклопедистов Мухаммеда ал-Бируни (Хорезм, 973–1048) и Насирэддина ат-Туси (Хорасан, 1201–1274), тригонометрия в XV веке перешла в руки европейских астрономов, решавших задачи, связанные с дальними плаваниями. Тригонометрические таблицы, появившиеся в то время, несомненно, вдохновляли навигаторов, пролагавших курс кораблей Колумба и Магеллана.

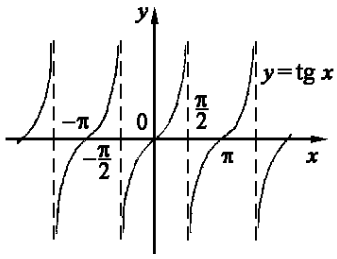


Рис. 16

VI. Обратные тригонометрические функции. Синус не является взаимно однозначной функцией и потому обратной функции иметь не может. Однако если ограничиться только отрезком $[-\frac{\pi}{2}, \frac{\pi}{2}]$ оси абсцисс, то на этой части своей области определения синус действует биективно и, следовательно, допускает обратную функцию, взаимно однозначно отображающую отрезок $[-1, 1]$ на отрезок $[-\frac{\pi}{2}, \frac{\pi}{2}]$. Эта функция называется арксинусом: $f(x) = \arcsin x$. Таким образом, арксинус числа $x \in [-1, 1]$ – это такой угол в пределах от $-\frac{\pi}{2}$ до $\frac{\pi}{2}$ (включая эти значения), синус которого равен x. Алгебраические равенства $f \circ f^{-1} = \Delta$ и $f^{-1} \circ f = \Delta$, связывающие функцию с ее обратной, будучи применены к синусу и арксинусу, превращаются соответственно в тождества $\arcsin(\sin x) = x$ и $\sin(\arcsin x) = x$. Аналогично определяются и другие обратные тригонометрические функции: арккосинус $\arccos x$, арктангенс $\mathrm{arctg}\, x$ и арккотангенс $\mathrm{arcctg}\, x$.

Обратные тригонометрические функции чаще всего возникают при решении тригонометрических уравнений и в интегральном исчислении. Постройте по точкам график арксинуса, это потребует известных усилий.

Рассмотренные функции типов I-VI называются основными элементарными функциями. Это постоянные, степенные, показательные, логарифмические, тригонометрические и обратные тригонометрические функции.

Теперь можно дать главное определение. Элементарными функциями называются функции, которые получаются из основных элементарных функций путем применения конечного числа арифметических операций и суперпозиций.

Укажем некоторые часто встречающиеся элементарные функции, не входящие в список основных.

Целой рациональной функцией называется функция

$$f(x) = a_0x^n + a_1x^{n-1} + \cdots + a_{n-1}x + a_n.$$

Она получается с помощью умножения и сложения из степенных и постоянных функций. При $n = 1$ (многочлен первой степени) получаем линейную функцию, ее записывают в виде $f(x) = ax + b$. График линейной функции – прямая линия. Она определяется любыми двумя своими точками. Например, прямая $y = 2x + 3$ проходит через точки $(0, 3)$ и $(-1, 1)$, так как координаты этих точек, будучи подставлены в уравнение прямой, превращают его в тождество. Постройте прямые $y = 2x$, $y = 2x - 3$, $y = -2x + 3$, найдя на каждой из них какие-нибудь две точки.

Многочлен степени $n = 2$ определяет квадратичную функцию $f(x) = ax^2+bx+c$ (как заметил американский писатель и поэт Эдгар По, изучавший высшую математику в военной академии,– написав квадратный трехчлен, хочется тут же приравнять его нулю). График квадратичной функции – парабола. Если $a > 0$, ветви параболы направлены вверх, при $a < 0$ – вниз. Вершина этой параболы имеет координаты $\left(-\frac{b}{2a}, c - \frac{b^2}{4a}\right)$. Например, если $y = -2x^2 + 4x + 1$, то $a = -2, b = 4, c = 1$ и, значит, вершина параболы лежит в точке $(1, 3)$ и ветви ее направлены вниз. При $x = 0$ и $x = 2$ получаем $y = 1$, т.е. кривая проходит через точки $(0, 1)$ и $(2, 1)$. Найдя на ней еще несколько точек, можно нарисовать приблизительный график. Изобразите параболу $y = x^2 - 2x - 3$. В 1638 году Галилей первым осознал, что траектория снаряда, выпущенного из орудия, представляет собой привершинную часть параболы. Известно, что воздушные воронки ураганов перемещаются по параболе, обращенной вершиной к западу.

Абсолютная величина числа $f(x) = |x|$ также является элементарной функцией, так как она представима в виде суперпозиции двух степенных функций: $|x| = \sqrt{x^2}$. (Известный математический софизм

$$1 = \sqrt{1^2} = \sqrt{(-1)^2} = -1$$

основан на ошибочном извлечении корня: на самом деле $\sqrt{(-1)^2} = | -1| = 1$). Основные свойства абсолютной величины:

1) $|x| \geqslant 0$,

2) $|x \pm y| \leqslant |x| + |y|$,

3) $|xy| = |x| \cdot |y|$,

4) $|\dfrac{x}{y}| = \dfrac{|x|}{|y|}$.

Дайте словесные формулировки этих правил.

Особо выделим следующий простой, но часто используемый факт.

Теорема 2. Неравенство $|x| \leqslant a$ имеет место тогда и только тогда, когда выполняются неравенства $-a \leqslant x \leqslant a$.

ДОКАЗАТЕЛЬСТВО. Абсолютная величина числа x выражает расстояние точки x на числовой оси от начала отсчета O. Неравенство $|x| \leqslant a$ означает, что точка x удалена от O не более чем на a. Значит, она не может лежать правее точки a (т.е. $x \leqslant a$) и левее точки $-a$ (т.е. $x \geqslant -a$). $\square$

График функции $f(x) = |x|$ приведен на рис. 3.

Элементарная функция $f(x) = A\sin(\omega x + \varphi)$ (постоянные обозначены в соответствии с традициями физики) получается как суперпозиция линейной функции и синуса, умноженная на постоянную. Эта синусоида описывает колебательный процесс с амплитудой A, частотой ω и фазой φ. Смысл этих констант следующий: амплитуда выражает наибольшее отклонение колеблющейся по оси Oy точки от положения равновесия $y = 0$, частота – количество колебаний за время $x = 2\pi$, а фаза – значение амплитуды в начальный момент времени: $y_0 = A\sin\varphi$. У простейшей синусоиды $y = \sin x$ имеем $A = 1,\ \omega = 1,\ \varphi = 0$. Две синусоиды, отличающиеся только фазой, получаются одна из другой сдвигом по оси времени Ox (см. рис. 15). Нарисуйте график функции $y = 2\sin(2x - \dfrac{\pi}{2})$.

Примеры элементарных функций можно неограниченно усложнять – вплоть до таких замысловатых конструкций, как

$$f(x) = \frac{2^{\sin 3x}\sqrt{\lg(1+x^2)}}{4\mathrm{tg}\,5x - \arcsin\dfrac{x}{2}}.$$

А можем ли мы привести пример какой-нибудь неэлементарной функции, т.е. указать функцию, которую невозможно получить из основных элементарных функций путем применения конечного числа арифметических операций и суперпозиций? В такой непосредственной постановке вопрос представляется чрезвычайно сложным: что значит «невозможно»? Однако уже в следующем параграфе будет указано некоторое простое свойство, которым обладает каждая элементарная функция. Следовательно, любая функция, не имеющая этого свойства, будет неэлементарной.

§ 3. Предел функции и непрерывность

Понятие предела функции является одним из центральных в математическом анализе. В своей современной форме оно было осознано к середине XIX века. Большая роль в этом принадлежит выдающемуся французскому математику Огюстену Луи Коши (1789–1857). В многочисленных (свыше 700) работах он получил фундаментальные результаты, относящиеся к самым различным областям теории функций и ее применений. Коши, наряду с Больцано и Вейерштрассом, ввел в математику тот дух логической строгости, который требовал отказа от интуитивных, наглядных представлений, господствовавших в формулировках и доказательствах. За полтора столетия, прошедших после открытий Декарта, Ньютона и Лейбница, было обнаружено столько разнообразных функций, что всякая языковая неточность грозила опасностью появления примеров, опровергающих очевидные на первый взгляд доводы. Возможности человеческого воображения оказались слишком ограниченными, чтобы вместить те фантастические ситуации, которые могут возникнуть в мире функций (попытайтесь представить себе, например, кривую, которая целиком заполняет квадрат). Поэтому, говорили поборники строгости, ссылки на интуицию и объяснения «на пальцах» допустимы лишь тогда, когда за ними стоят кем-то проверенные и где-то зафиксированные точные рассуждения. Конечно, те, кто работает в прикладных областях математики, действуют гораздо раскованнее, справедливо полагая, что химеры высшего анализа вряд ли встретятся им на пути. Как говорил академик А.Н.Крылов (кораблестроитель), утонченная строгость доказательств представляется иногда «торжеством науки над здравым смыслом». (Или, как говорил Козьма Прутков: «Многие вещи нам непонятны не потому, что наши понятия слабы; но потому, что сии вещи не входят в круг наших понятий»).

Теория пределов стала первым разделом анализа, вполне освобожденным от нечетких словесных построений.

Среди различных частей числовой прямой выделяются ее непрерывные, «сплошные» куски, которые называются интервалами. Если a, b – действительные числа и $a < b$, то с ними связаны следующие четыре интервала, определяемых при помощи неравенств:

$$[a,b] = \{\, x \in \mathbb{R} \mid a \leqslant x \leqslant b \,\} \text{ – замкнутый интервал,}$$

$$(a,b] = \{\, x \in \mathbb{R} \mid a < x \leqslant b \,\} \text{ – левый полуинтервал,}$$

$$[a,b) = \{\, x \in \mathbb{R} \mid a \leqslant x < b \,\} \text{ – правый полуинтервал,}$$

$$(a,b) = \{\, x \in \mathbb{R} \mid a < x < b \,\} \text{ – открытый интервал.}$$

Замкнутый интервал (или, в другой терминологии, отрезок) $[a, b]$ содержит все числа между a и b, включая эти концевые точки, полуинтервалы не имеют одного из концов, открытый интервал (или просто – интервал) открыт с обеих сторон. На рис. 17 условно изображены эти части числовой прямой.

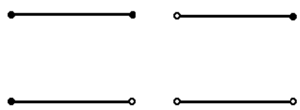


Рис. 17

Кроме указанных конечных интервалов, в записях используют и бесконечные интервалы $(-\infty, a)$, $(a, +\infty)$ и полуинтервалы $(-\infty, a]$, $[a, +\infty)$. Они соответствуют неравенствам $x < a$, $x > a$, $x \leqslant a$, $x \geqslant a$. Вся числовая прямая $\mathbb{R}$ иногда обозначается как интервал $(-\infty, +\infty)$. Символы $-\infty$ («минус бесконечность») и $+\infty$ («плюс бесконечность») не представляют собой какие-либо числа, а служат лишь для обозначения неограниченности интервала слева или справа.

Говорят, что функция $f(x)$ определена на некотором интервале, если она определена в каждой его точке, т.е. если все точки интервала содержатся в области определения функции $f(x)$.

Окрестностью точки a на числовой прямой называется любой открытый интервал, содержащий эту точку.

Пусть функция $f(x)$ определена в некоторой окрестности точки a за исключением, быть может, самой этой точки. Число b называется пределом функции $f(x)$ при x стремящемся к a, если для любого сколь угодно малого положительного числа ε существует такое положительное число δ, что при $x \neq a$ и $|x-a| < \delta$ будет выполняться неравенство $|f(x) - b| < \varepsilon$. В этом случае пишут $\lim\limits_{x\to a} f(x) = b$.

Символ $\lim$ – это начальные буквы латинского слова limes (предел, конечная цель), которое впервые в математическом смысле использовал Ньютон в 1687 году (наш «лимит» происходит от последующего французского limite).

Приведенное определение предела имеет следующую интуитивную основу: чем ближе точка x подходит к a, тем меньше соответствующее значение функции $f(x)$ отклоняется от b. Но как выразить эту физическую картину движения точки x в математических терминах числовой прямой, одномерного континуума $\mathbb{R}$ с застывшим временем? «Язык эпсилон-дельта» (от участвующих в определении предела греческих букв ε и δ) позволил Коши обойти эти неразрешимые трудности.

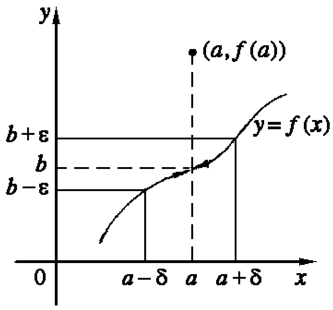


Рис. 18

Переписав неравенства $|x-a| < \delta$ и $|f(x) - b| < \varepsilon$ в виде $a - \delta < x < a + \delta$ и $b - \varepsilon < f(x) < b + \varepsilon$ (используется теорема 2 из § 2, проделайте выкладки), можно дать условную иллюстрацию к определению предела (рис. 18): $b = \lim\limits_{x\to a} f(x)$, если для любого $\varepsilon > 0$ можно указать число $\delta > 0$ такое, что на участке $(a - \delta, a + \delta)$ все точки графика функции $y = f(x)$, кроме, может быть, точки $(a, f(a))$, лежат между горизонтальными уровнями $b - \varepsilon$ и $b + \varepsilon$.

Глядя на графики функций, приведенные в §2, можно заключить, что, например, $\lim\limits_{x\to 0} x^2 = 0$, $\lim\limits_{x\to -2} x^2 = 4$, $\lim\limits_{x\to 9} \sqrt{x} = 3$, $\lim\limits_{x\to 3} \dfrac{1}{x} = \dfrac{1}{3}$, $\lim\limits_{x\to 0} a^x = 1$, $\lim\limits_{x\to 1} \log_a x =$, $\lim\limits_{x\to 0} \sin x = 0$, $\lim\limits_{x\to -\pi/2} \sin x = -1$, $\lim\limits_{x\to 0} \cos x = 1$.

Предел постоянной функции в любой точке a равен значению этой функции: $\lim\limits_{x\to a} c = c$. Функция $y = \dfrac{1}{x}$ не имеет предела в точке $a = 0$ – при приближении к нулю ветви ее графика расходятся в разные стороны. На рис. 19 изображены графики функций $y = \operatorname{sign} x$ и $y = \operatorname{sign}^2 x$ (стрелки показывают, что линия в соответствующем месте не доходит до оси ординат Oy). Первая из этих функций не имеет предела при $x \to 0$, а $\lim\limits_{x\to 0} \operatorname{sign}^2 x = 1$.

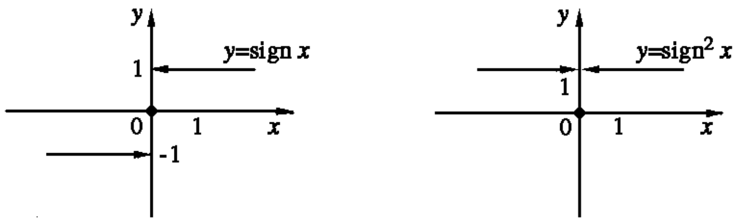


Рис. 19

Предельный переход согласован с арифметическими операциями над функциями в том смысле, что

1) $\lim\limits_{x\to a}[f(x) \pm g(x)] = \lim\limits_{x\to a} f(x) \pm \lim\limits_{x\to a} g(x)$

(предел суммы или разности функций равен соответственно сумме или разности их пределов),

2) $\lim\limits_{x\to a}[f(x)g(x)] = \lim\limits_{x\to a} f(x) \cdot \lim\limits_{x\to a} g(x)$

(предел произведения равен произведению пределов),

3) $\lim\limits_{x\to a} \dfrac{f(x)}{g(x)} = \dfrac{\lim\limits_{x\to a} f(x)}{\lim\limits_{x\to a} g(x)}$

(предел дроби равен пределу числителя, деленному на предел знаменателя. При этом должно быть $g(x) \neq 0$ и $\lim\limits_{x\to a} g(x) \neq 0$).

При предельном переходе знак неравенства между функциями не меняется:

4) если $f(x) \leqslant g(x)$, то $\lim\limits_{x\to a} f(x) \leqslant \lim\limits_{x\to a} g(x)$.

Используя эти свойства, находим пределы в более сложных случаях:

$$\lim_{x\to 0} \frac{2^x}{3\sin x - \cos x} = \frac{\lim\limits_{x\to 0} 2^x}{\lim\limits_{x\to 0} 3 \lim\limits_{x\to 0} \sin x - \lim\limits_{x\to 0} \cos x} = \frac{1}{3 \cdot 0 - 1} = -1$$

или

$$\lim_{x\to 1}\frac{x^2-\lg x}{\sqrt{x}}=\frac{\lim\limits_{x\to 1}x^2-\lim\limits_{x\to 1}\lg x}{\lim\limits_{x\to 1}\sqrt{x}}=\frac{1-0}{1}=1.$$

При этом мы, конечно, опираемся на свойства основных элементарных функций, которые можно извлечь из их графического представления. Найдите пределы $\lim\limits_{x\to -\frac{\pi}{2}}\dfrac{3\sin x}{1+\cos x}$, $\lim\limits_{x\to -2}\dfrac{x^3+1}{7x^2}$.

Если переменная x неограниченно возрастает (пишут: $x\to +\infty$) или неограниченно убывает ($x\to -\infty$), можно определить пределы $\lim\limits_{x\to +\infty}f(x)$ и $\lim\limits_{x\to -\infty}f(x)$. Например, $\lim\limits_{x\to +\infty}\dfrac{1}{x}=$ $=\lim\limits_{x\to -\infty}\dfrac{1}{x}=0$, $\lim\limits_{x\to -\infty}a^x=0$ при $a>1$. Факт неограниченного возрастания или убывания функции также выражают с использованием пределов: $\lim\limits_{x\to +\infty}x^3=+\infty$, $\lim\limits_{x\to -\infty}x^3=-\infty$, $\lim\limits_{x\to -\infty}a^x=+\infty$ при $a<1$, $\lim\limits_{x\to 0}\lg x=-\infty$, $\lim\limits_{x\to 0}\dfrac{1}{x^2}=+\infty$. Пример вычисления предела при x стремящемся к бесконечности:

$$\lim_{x\to +\infty}\frac{x+1}{x^2+1}=$$

(разделим числитель и знаменатель дроби на x^2 – дробь не изменится)

$$=\lim_{x\to +\infty}\frac{\frac{1}{x}+\frac{1}{x^2}}{1+\frac{1}{x^2}}=\frac{\lim\limits_{x\to +\infty}\frac{1}{x}+\lim\limits_{x\to +\infty}\frac{1}{x^2}}{\lim\limits_{x\to +\infty}1+\lim\limits_{x\to +\infty}\frac{1}{x^2}}=\frac{0+0}{1+0}=\frac{0}{1}=0.$$

Найдите пределы

$$\lim_{x\to +\infty}\frac{x^2+x+1}{2x^2-x-1},\ \lim_{x\to -\infty}\frac{x^3}{x^2+1}.$$

Функция $\alpha(x)$ называется бесконечно малой в точке $x=a$, если ее предел в этой точке существует и равен нулю: $\lim\limits_{x\to a}\alpha(x)=0$. Бесконечно малыми в точке $a=0$ являются функции $y=x$, $y=x^2$, $y=x^3$, в точке $a=1$ – все логарифмы. Синус будет бесконечно малым в точках 0, $\pm\pi$, $\pm 2\pi$ и т.д., а косинус – в точках $\pm\dfrac{\pi}{2}$, $\pm\dfrac{3\pi}{2}$, $\pm\dfrac{5\pi}{2}$ и т.д.

Бесконечно малые играли главенствующую роль при формировании математического анализа в XVII и XVIII веках (собственно, это и был анализ бесконечно малых). Их связь с пределами вскрывает следующее предложение.

Теорема 1. Число b тогда и только тогда является пределом функции $f(x)$ в точке a, когда разность $\alpha(x) = f(x) - b$ бесконечно мала в этой точке.

ДОКАЗАТЕЛЬСТВО. 1) Если $b = \lim\limits_{x\to a} f(x)$, то $\lim\limits_{x\to a} \alpha(x) = \lim\limits_{x\to a}[f(x) - b] = \lim\limits_{x\to a} f(x) - \lim\limits_{x\to a} b = b - b = 0$ (мы воспользовались формулой предела разности и тем, что предел постоянной равен ей самой). Итак, $\lim\limits_{x\to a} \alpha(x) = 0$, т.е. $\alpha(x)$ – бесконечно малая в точке a.

2) Пусть $\alpha(x) = f(x) - b$ – бесконечно малая в точке a, т.е. $\lim\limits_{x\to a} \alpha(x) = 0$. Тогда

$$0 = \lim_{x\to a} \alpha(x) = \lim_{x\to a}[f(x) - b] = \lim_{x\to a} f(x) - \lim_{x\to a} b = \lim_{x\to a} f(x) - b,$$

т.е. $\lim\limits_{x\to a} f(x) - b = 0$, откуда $b = \lim\limits_{x\to a} f(x)$. $\square$

Не имея точного определения предела и, следовательно, не осознавая до конца, что такое бесконечно малые, Лейбниц и Ньютон, а затем и их последователи смело оперировали с этими загадочными величинами, отличными от нуля, но в то же время сколь угодно близкими к нему. Именно неясности, связанные с бесконечно малыми, вызвали суровую критику математического анализа со стороны епископа Джорджа Беркли. В своем трактате «Аналист, или Рассуждение, обращенное к неверующему математику, где исследуется, более ли ясно воспринимаются и более ли очевидно выводятся предмет, принципы и умозаключения современного анализа, чем религиозные таинства и догматы веры» (1734) он на остроумных примерах продемонстрировал логические пробелы и противоречия в доказательствах ведущих математиков. Это вызвало ожесточенную полемику и сильно стимулировало деятельность по обоснованию анализа.

Вычисление пределов – непростая задача, оно требует известных навыков и изобретательности. Поэтому всякое усовершенствование, способствовавшее экономии мышления в этом деле, вызывало большой интерес. Следы соответствующих эмоциональных всплесков сохранились в названиях двух соотношений, упростивших решение многих задач, связанных с пределами.

Первый замечательный предел $\lim\limits_{x\to 0}\dfrac{\sin x}{x}=1$ имеет дело с функцией, не существующей в предельной точке: при $x = 0$ дробь $\dfrac{\sin x}{x}$ превращается в неопределенное выражение $\dfrac{0}{0}$. В раскрытии этой неопределенности и заключается смысл рассматриваемого предельного равенства. Первый замечательный предел «механизирует» процесс вычисления пределов, содержащих тригонометрические функции. Например,

$$\lim_{x\to 0}\frac{\operatorname{tg} x}{x}=\lim_{x\to 0}\frac{\left(\frac{\sin x}{\cos x}\right)}{x}=\lim_{x\to 0}\frac{\sin x}{x\cos x}=$$

$$=\lim_{x\to 0}\left(\frac{\sin x}{x}\cdot\frac{1}{\cos x}\right)=\lim_{x\to 0}\frac{\sin x}{x}\cdot\lim_{x\to 0}\frac{1}{\cos x}=1$$

(объясните действия, совершаемые при каждом переходе от одного выражения к другому).

Второй замечательный предел не только облегчает вычисления, но и вводит одну из важнейших математических констант – число e. Именно,

$$\lim_{x\to 0}(1+x)^{\frac{1}{x}}=e.$$

Функция $(1+x)^{\frac{1}{x}}$ не определена в точке 0, но предел при $x\to 0$ тем не менее существует и имеет приближенное значение 2,71828. Эйлер предложил обозначить его буквой e. Число e – трансцендентное, так же как и π. Показательная функция $y = e^x$ называется экспоненциальной функцией, или экспонентой (когда говорят, что нечто возрастает по экспоненте, имеют в виду круто идущую вверх при $x\to+\infty$ кривую $y = e^x$. Убывающая экспонента – это функция $y = e^{-x}$). Логарифмы по основанию e мы уже упоминали, само название «натуральный логарифм» показывает естественность использования при логарифмировании именно функции $\ln x$.

В 1748 году Эйлер нашел одну из самых знаменитых и красивых математических формул:

$$e^{i\pi}+1=0.$$

В ней участвуют пять главных чисел математики: единица 1, порождающая натуральный ряд; ноль 0, сделавший возможной позиционную запись чисел; мнимая единица i, олицетворяющая комплексные числа; основная тригонометрическая постоянная π и основание натуральных логарифмов e, а также два первоначальных знака – равенства и сложения. Даже не зная математического смысла формулы Эйлера, нельзя не восхититься ее эстетическим совершенством.

Пространство и время непрерывны, и большинство протекающих вокруг нас процессов носят тоже непрерывный характер, нарушение которого воспринимается как некоторое отклонение от нормы. Непрерывные функции играют основную роль в математике и ее приложениях.

Функция $f(x)$ называется непрерывной в точке a, если существует ее предел при стремлении x к a и этот предел равен $f(a)$ – значению функции в точке a, т.е. если $\lim\limits_{x\to a} f(x) = f(a)$.

Учитывая теорему 1, можно сказать, что функция $f(x)$ непрерывна в точке a, если разность $\alpha(x) = f(x) - f(a)$ является бесконечно малой в этой точке. Это вполне согласуется с интуитивным восприятием непрерывности процессов, зависящих от времени: малому изменению времени соответствует малое изменение в ходе самого процесса.

Локальное определение непрерывности – в одной точке – естественным образом расширяется на целые области числовой прямой: функция называется непрерывной на данном числовом множестве, если она непрерывна в каждой его точке. Графически свойство непрерывности функции на некотором интервале выражается тем, что график ее в пределах этого интервала можно начертить, не отрывая карандаш от бумаги.

Обращаясь к графикам основных элементарных функций (§ 2), можно заключить, что все они непрерывны в своей области определения. Сомнение могут вызвать разве лишь функции $f(x) = \dfrac{1}{x}$ и $f(x) = \operatorname{tg} x$: их графики состоят из отдельных ветвей, но здесь сама область определения не непрерывна, а представляет собой объединение отдельных интервалов, на каждом из которых рассматриваемая функция уже не имеет разрывов. На самом деле имеет место гораздо более общий факт, который выражает

Теорема 2. Всякая элементарная функция непрерывна в каждой точке ее области определения.$\square$

Таким образом, непрерывность – это свойство всех элементарных функций, необходимое условие элементарности функции. Следовательно, если в некоторой точке области определения функции $f(x)$ оно нарушается, то $f(x)$ – неэлементарная функция. Таковыми будут, например, функция $\operatorname{sign} x$ и ее квадрат $\operatorname{sign}^2 x$: первая вовсе не имеет предела в точке $a = 0$ (при подходе к этой точке справа значение функции все время равно 1, а при подходе слева получается -1), а вторая хотя и имеет здесь предел: $\lim_{x\to 0} \operatorname{sign}^2 x = 1$, но он не равен значению функции в нуле: $\operatorname{sign}^2 0 = 0$. Функция Дирихле представляет собой трудное испытание нашему воображению: ее график имеет разрыв в каждой точке числовой прямой! Используя определение элементарной функции, приходим, например, к выводу, что функция $f(x) = \operatorname{sign} x$ не может быть получена из основных элементарных функций применением конечного числа арифметических операций и суперпозиций. Однако стоит только убрать из ее графика точку $(0, 0)$, как картина резко меняется: полученная непрерывная функция

$$f^*(x) = \begin{cases} 1, & \text{если} \quad x > 0, \\ -1, & \text{если} \quad x < 0 \end{cases}$$

оказывается элементарной, именно $f^*(x) = \dfrac{|x|}{x}$ (выразите ее через основные элементарные функции). Заметим, что непрерывность сама по себе недостаточна для того, чтобы функция была элементарной: существует много непрерывных неэлементарных функций. Они изучаются в высших разделах анализа и в других математических дисциплинах.

Большой вклад в обоснование анализа внес чешский богослов Бернард Больцано (1781–1848). Отстраненный за свои неприятные властям проповеди от преподавания и публичных выступлений, он свыше двадцати лет прожил в деревне, где и создал свои основные произведения, большинство которых было опубликовано посмертно. В них Больцано доказывал бессмертие души (один из доводов – от противного: «Если не существует никакой другой жизни, то разрыв между добродетелью и счастьем на земле воспринимается как одно из сильных возражений против Божественной справедливости. Вера в бессмертие устраняет это возражение»), рассуждал о парадоксах актуальной бесконечности (предвосхищая идеи Кантора), построил теорию распространения волн. Следующая теорема Больцано находит бесчисленные приложения в математике.

Теорема 3. Если функция $f(x)$ непрерывна на отрезке $[a, b]$ и на концах его принимает значения $f(a)$ и $f(b)$, противоположные по знаку, то в интервале (a, b) существует точка ξ, в которой функция обращается в ноль: $f(\xi) = 0$. □

Сообщаемый в теореме факт интуитивно совершенно очевиден: чтобы перейти на плоскости из точки, лежащей по одну сторону от оси абсцисс Ox, в точку, лежащую по другую ее сторону, непрерывная кривая обязательно должна пересечь эту ось в некоторой точке (рис. 20). Именно вследствие этой очевидности никто из предшественников Больцано не только не доказывал, но даже специально не выделял предложение, ныне носящее его имя. На самом же деле при точном определении понятия непрерывности доказательство оказалось очень непростым, а использованные в нем идеи новыми и весьма плодотворными.

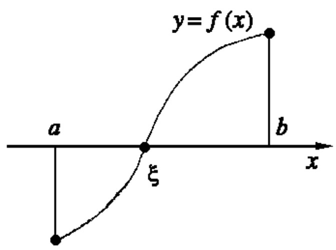


Рис. 20

В качестве примера применения теоремы Больцано можно указать следующее рассуждение: так как всякий многочлен $f(x)$ представляет собой непрерывную функцию, то из того факта, что в точках a и b он имеет значения противоположных знаков, следует наличие корня уравнения $f(x) = 0$ в интервале (a, b). Так, для $f(x) = x^5 - 4x - 2$ получаем $f(-1) = 1$, $f(0) = -2$ и, значит, уравнение $x^5 - 4x - 2 = 0$ (в §I.3 оно упоминалось как неразрешимое в радикалах) в интервале $(-1, 0)$ имеет корень. Используя теорему 3, покажите, что это уравнение имеет еще два действительных корня.

Занимаясь квадратурой круга, ректор Парижского университета Альберт Саксонский в 1353 году следующим образом доказывал существование квадрата, равного по площади данному кругу: «Пусть имеется квадрат, вписанный в круг, и пусть этот квадрат начнет непрерывно и равномерно увеличиваться, пока не станет больше круга. В какое-то время он обязательно окажется равным кругу, ибо проделанный переход от *меньше* к *больше* не может миновать *равно*». Видите ли вы в этих рассуждениях теорему Больцано?

Другое важнейшее свойство непрерывных функций устанавливает теорема Вейерштрасса.

Теорема 4. Если функция $f(x)$ непрерывна на отрезке $[a, b]$, то среди ее значений на этом отрезке есть наибольшее и есть наименьшее. $\square$

Очевидный интуитивный смысл этого утверждения заключается в том, что график функции, непрерывной в замкнутом интервале, имеет хотя бы одну наивысшую точку и хотя бы одну наинизшую. Кажущийся контрпример: функция $f(x) = \dfrac{1}{x}$ не имеет наибольшего значения на участке $(0, 1]$ и не имеет наименьшего значения на участке $[-1, 0)$, хотя она и непрерывна на обоих. Ответ: оба эти полуинтервала не замкнуты – первый слева, второй справа, что нарушает условия теоремы 4.

Выдающийся немецкий математик Карл Теодор Вильгельм Вейерштрасс (1815–1897) завершил логическое обоснование анализа, начатое Больцано и Коши, получил основополагающие результаты в теории функций, зависящих от комплексного аргумента (эта теория – один из самых мощных инструментов прикладной математики). Когда Софье Ковалевской (1850–1891) было отказано в посещении и Петербургского и Берлинского университетов (женщины не имели права на получение высшего образования), Вейерштрасс с 1870 по 1874 год занимался с ней индивидуально. «Занятия эти имели в высшей степени важное влияние на всю мою математическую карьеру... Все мои работы сделаны именно в духе вейерштрассовских идей»,– писала впоследствии благодарная ученица. Вейерштрассу принадлежат часто цитируемые слова: «В истинном математике всегда есть нечто от поэта». Мы не знаем, в какой мере они были вдохновлены образом Софьи Васильевны Ковалевской, но литературный талант ее общепризнан. Отвечая на вопросы читателей, как ей удается сочетать художественное творчество с профессиональными занятиями наукой, писательница – профессор Стокгольмского университета, приведя высказывание Вейерштрасса, пояснила: «Мне кажется, что поэт должен видеть то, чего не видят другие, видеть глубже других. И это же должен математик». Самым ярким литературным произведением Ковалевской является ее автобиографическая повесть «Воспоминания детства» (1889 г.), которую ставят в ряд с аналогичными сочинениями Аксакова и Толстого; самым значительным математическим результатом – открытие третьего (после случаев Эйлера и Лагранжа) случая разрешимости задачи о вращении твердого тела вокруг неподвижной точки (1888 г.). В XX веке появилось много женщин-математиков, достижения которых получили мировую известность, и все же среди этих звезд первое место по праву занимает Софья Ковалевская.

§ 4. Дифференциальное исчисление: идеология, техника, основные теоремы

С открытия дифференциального исчисления начинается совершенно новый период в истории математики. Из углубленной в себя, «чистой» науки, лишь в редких случаях снисходившей до нужд повседневного бытия, она, в течение одного лишь столетия, превратилась в мощный инструмент познания и преобразования мира. Исследователи и практики получили возможность изучать меняющиеся во времени процессы, записывать и решать с помощью дифференциальных уравнений задачи о всевозможных видах движения, будь то перемещение отдельной частицы, вращение волчка или полет «в пространство брошенных светил». Честь провозгласить идеи, медленно созревавшие в работах многих предшественников, выпала на долю Ньютона и Лейбница. В противоположность Лейбницу, блестящему и разностороннему дилетанту, человеку с гуманитарным складом ума, Исаак Ньютон (1643–1727) представлял собой тип кабинетного ученого, погруженного в науку и не проявлявшего никакого интереса к «роскоши человеческого общения». Он не заботился о том, кто и как будет читать его работы, насколько они будут понятны и в какой степени оценены. Обозначения и терминология Ньютона весьма затрудняли проникновение в смысл, в то время как простые и изящные формулы Лейбница сразу завоевали всеобщее признание.

Исходным понятием дифференциального исчисления является понятие производной. Пусть $f(x)$ – некоторая функция. Производной от нее называется такая функция $f'(x)$, значение которой в каждой точке a вычисляется по формуле

$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a}.$$

Если этот предел существует во всех точках некоторого интервала, функция $f(x)$ называется дифференцируемой на данном интервале, а процесс отыскания ее производной $f'(x)$ называется дифференцированием функции $f(x)$. Дифференцирование – это новая операция, в отличие от арифметических операций и суперпозиции она применяется к одной функции (как, например, извлечение квадратного корня – к одному числу). Совокупность понятий, методов и теорем, связанных с дифференцированием, и образует дифференциальное исчисление функций.

Рассмотрим пример непосредственного нахождения производной. Пусть $f(x) = x^2$. Тогда

$$f'(a) = \lim_{x\to a} \frac{x^2 - a^2}{x - a} = \lim_{x\to a} \frac{(x-a)(x+a)}{x-a} = \lim_{x\to a}(x+a) = 2a,$$

т.е. $f'(a) = 2a$. Так как точка a выбиралась произвольно, вместо нее можно написать x, и мы получаем $(x^2)' = 2x$.

Аналогичным образом найдите производные функций $f(x) = x^3$ и $f(x) = x^4$.

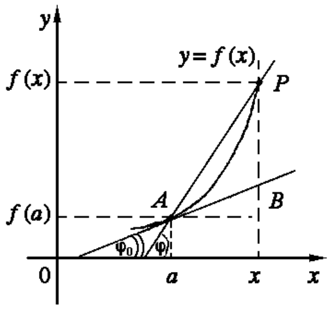


Рис. 21

Лейбниц пришел к понятию производной, решая задачу о проведении касательной к графику функции $y = f(x)$ в данной точке a (рис. 21). Чтобы построить касательную в точке A, имеющей координаты $(a, f(a))$, достаточно знать угол φ_0, который касательная образует с осью абсцисс Ox, а для этого, в свою очередь, достаточно найти, скажем, тангенс угла φ_0. Касательная в точке A является предельным положением секущей PA при стремлении точки $P(x, f(x))$ по кривой к точке A. В пределе угол φ, образуемый секущей с осью Ox, превращается в φ_0, так что

$$\operatorname{tg}\varphi_0 = \lim_{P\to A} \operatorname{tg}\varphi = \lim_{P\to A} \frac{PB}{AB} = \lim_{x\to a} \frac{f(x) - f(a)}{x - a} = f'(a).$$

Таким образом, производная $f'(x)$ равна тангенсу угла, который образует с осью абсцисс Ox касательная, проведенная к графику функции $f(x)$ в точке с абсциссой x. Это и есть геометрическое истолкование производной.

Для Ньютона производная означала скорость прямолинейного движения точки. Пусть функция $s = f(t)$ выражает зависимость пути, пройденного прямолинейно движущейся точкой, от времени t. Подсчитаем скорость движения в момент $t = a$. Разность $f(t) - f(a)$ – это путь, который точка прошла за время $t - a$. Значит, средняя скорость на этом участке была $\dfrac{f(t) - f(a)}{t - a}$. Мгновенная скорость – это предельное значение средней скорости при стягивании временно́го интервала в точку, т.е. при t, стремящемся к a:

$$v(a) = \lim_{t \to a} \frac{f(t) - f(a)}{t - a} = f'(a).$$

Таким образом, производная от пути по времени – это скорость движения. Таково механическое истолкование производной.

(При свободном падении точки под действием силы тяжести $s = \dfrac{gt^2}{2}$, где $g \approx 9,8$. Найдите скорость.)

Следующее предложение показывает, что все дифференцируемые функции непрерывны.

Теорема 1. Если функция $f(x)$ имеет производную в точке a, то $f(x)$ непрерывна в этой точке.

ДОКАЗАТЕЛЬСТВО. По определению производной, $f'(a) = \lim\limits_{x \to a} \dfrac{f(x) - f(a)}{x - a}$. По теореме 1 из §3, разность между функцией и ее пределом бесконечно мала в предельной точке, т.е. $\alpha(x) = \dfrac{f(x) - f(a)}{x - a} - f'(a)$ – бесконечно малая в точке A. Так как

$$f(x) - f(a) = [f'(a) + \alpha(x)](x - a),$$

то

$$\lim_{x \to a} [f(x) - f(a)] =$$

$$= \lim_{x \to a} \{[f'(x) + \alpha(x)](x - a)\} = \lim_{x \to a} [f'(a) + \alpha(x)] \lim_{x \to a} (x - a) =$$

$$= (\lim_{x \to a} f'(a) + \lim_{x \to a} \alpha(x)) \lim_{x \to a} (x - a) = (f'(a) + 0) \cdot 0 = 0,$$

т.е. $\lim\limits_{x \to a} [f(x) - f(a)] = 0$. Отсюда $\lim\limits_{x \to a} f(x) - \lim\limits_{x \to a} f(a) = 0$, и следовательно, $\lim\limits_{x \to a} f(x) = f(a)$, что и означает непрерывность функции $f(x)$ в точке a. □

Непрерывность функции является лишь необходимым условием существования производной, т.е. из непрерывности не обязательно следует дифференцируемость. Например, функция $f(x) = |x|$ непрерывна (см. рис. 3), но в точке $a = 0$ не имеет производной, так как в этой точке нельзя провести касательную к графику, а производная – это тангенс угла наклона касательной к оси Ox. Во всех остальных точках абсолютная величина дифференцируема:

$$|x|' = \begin{cases} \quad 1, \text{если } x > 0, \\ -1, \text{если } x < 0, \end{cases}$$

т.е. $|x|' = \dfrac{|x|}{x}$ (функция, получаемая из функции $\operatorname{sign} x$ удалением точки $(0, 0)$).

Одним из вызывающих изумление образов анализа является найденный Вейерштрассом пример функции, которая непрерывна, но ни в одной точке не имеет производной.

Вычисление производной непосредственно по ее определению в общем случае является чрезвычайно сложной задачей из-за отсутствия универсальных способов нахождения пределов. Однако уже к началу XVIII века была разработана техника дифференцирования, превратившая этот процесс из искусства в рутинную, механическую работу,– прекрасная иллюстрация к словам Альфреда Норса Уайтхеда (1861–1947): «Прогресс цивилизации выражается в увеличении числа важных действий, которые мы можем совершать, не думая о них».

Первые пять правил Исчисления касаются дифференцирования арифметических выражений.

<u>Теорема 2</u>. Если функции $f(x)$ и $g(x)$ дифференцируемы, то дифференцируемы также функции $cf(x)$, $f(x) + g(x)$, $f(x) - g(x)$, $f(x)g(x)$, $\dfrac{f(x)}{g(x)}$ (последняя при $g(x) \neq 0$), причем имеют место равенства:

1) $[cf(x)]' = cf'(x)$ (постоянная выносится за знак производной),
2) $[f(x) + g(x)]' = f'(x) + g'(x)$ (производная суммы),
3) $[f(x) - g(x)]' = f'(x) - g'(x)$ (производная разности),
4) $[f(x)g(x)]' = f'(x)g(x) + f(x)g'(x)$ (производная произведения),
5) $\left[\dfrac{f(x)}{g(x)}\right]' = \dfrac{f'(x)g(x) - f(x)g'(x)}{[g(x)]^2}$ (производная частного).

ДОКАЗАТЕЛЬСТВО. Ограничимся выводом первых двух правил.

1) Имеем:

$$[cf(x)]'_{x=a} = \lim_{x\to a} \frac{cf(x) - cf(a)}{x - a} =$$

$$= \lim_{x\to a} \left[c\frac{f(x) - f(a)}{x - a} \right] = \lim_{x\to a} c \lim_{x\to a} \frac{f(x) - f(a)}{x - a} = cf'(a).$$

В силу произвольности величины a можно заменить ее на привычный x, так что $[cf(x)]' = cf'(x)$.

2) «Производная суммы равна сумме производных». Доказательство ведется прямым вычислением с использованием определения производной в точке a и правила раскрытия предела суммы (§ 3):

$$[f(x) + g(x)]'_{x=a} = \lim_{x\to a} \frac{[f(x) + g(x)] - [f(a) + g(a)]}{x - a} =$$

$$= \lim_{x\to a} \frac{[f(x) - f(a)] + [g(x) - g(a)]}{x - a} =$$

$$= \lim_{x\to a} \left[\frac{f(x) - f(a)}{x - a} + \frac{g(x) - g(a)}{x - a} \right] =$$

$$= \lim_{x\to a} \frac{f(x) - f(a)}{x - a} + \lim_{x\to a} \frac{g(x) - g(a)}{x - a} = f'(a) + g'(a).$$

Заменяя a на x, получаем требуемое равенство.

Вполне аналогично доказывается правило 3 (проведите эти рассуждения). Более сложные выкладки приводят к цели в случаях 4 и 5. □

Правило дифференцирования сложной функции дает

Теорема 3. Если функции $f : X \to Y$ и $g : Y \to Z$ дифференцируемы, то дифференцируема и их суперпозиция $h = f \circ g : X \to Z$, причем $h'(x) = g'(y)f'(x)$. □

Поскольку любая элементарная функция получается из основных элементарных функций с помощью арифметических операций и суперпозиции, то, зная производные основных элементарных функций и правила дифференцирования, указанные в теоремах 2 и 3, мы сможем найти производную любой элементарной функции. Таблица производных от основных элементарных функций завершает, таким образом, решение задачи алгоритмизации процесса дифференцирования.

Теорема 4. Имеют место следующие равенства:

I. $c' = 0$ (производная постоянной).

II. $(x^\alpha)' = \alpha x^{\alpha-1}$ (производная степенной функции).

В частности, $x' = 1$ (производная аргумента).

III. $(a^x)' = a^x \ln a$ (производная показательной функции).

В частности, $(e^x)' = e^x$ (производная экспоненты).

IV. $(\log_a x)' = \dfrac{1}{x \ln a}$ (производная логарифма).

В частности, $(\ln x)' = \dfrac{1}{x}$ (производная натурального логарифма).

V. $(\sin x)' = \cos x$ (производная синуса).

VI. $(\cos x)' = -\sin x$ (производная косинуса).

VII. $(\arcsin x)' = \dfrac{1}{\sqrt{1-x^2}}$ (производная арксинуса).

ДОКАЗАТЕЛЬСТВО. Ограничимся только рассмотрением тривиального соотношения I и несколькими замечаниями.

Покажем, что производная постояной функции $f(x) = c$ всюду (т.е. в любой точке) равна нулю. В самом деле,

$$c' = f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a} = \lim_{x \to a} \frac{c - c}{x - a} = \lim_{x \to a} 0 = 0.$$

Производная синуса (формула V) вычисляется с помощью первого замечательного предела, при нахождении производной логарифма (формула IV) главную роль играет второй замечательный предел. Все остальные равенства следуют из этих двух формул. $\square$

Частным случаем формулы II является равенство $x' = 1$ (производная аргумента равна единице). Производная экспоненциальной функции e^x равна самой этой функции (III) – свойство, определяющее роль экспоненты в приложениях.

Покажем несколько примеров дифференцирования элементарных функций.

1. Производная многочлена $f(x) = a_0x^n + a_1x^{n-1} + \cdots + +a_{n-1}x + a_n$.

Используя правила дифференцирования 2 и 1 из теоремы 2, а также выражение для производной степенной функции (формула II в теореме 4), получаем:

$$f'(x) = na_0x^{n-1} + (n-1)a_1x^{n-2} + \cdots + a_{n-1}.$$

Применение к конкретным случаям: $(x^2 + x + 1)' = 2x + 1$, $(3x^4 - 2x^3 + x^2 - 4x + 5)' = 12x^3 - 6x^2 + 2x - 4$.

Продифференцируйте многочлены $2x^3 - 3x^2 + 5x - 7$, $x^5 - 4x - 1$. Найдите производные линейной функции $f(x) = ax + b$ и квадратичной функции $f(x) = ax^2 + bx + c$.

2. По правилу дифференцирования произведения,

$$(\sqrt{x}\, e^x)' = (\sqrt{x})'\, e^x + \sqrt{x}\, (e^x)' =$$

$$= \left(\frac{1}{2}x^{-\frac{1}{2}}\right) e^x + \sqrt{x}\, e^x = \frac{e^x}{2\sqrt{x}} + \sqrt{x}\, e^x.$$

(Напомним, что $\sqrt{x} = x^{\frac{1}{2}}$ и $x^{-\alpha} = \dfrac{1}{x^{\alpha}}$).

Найдите производные функций $x \ln x$, $e^x \sin x$.

3. Производная тангенса. Применяя правило дифференцирования дроби и формулы для производных синуса и косинуса, имеем:

$$(\operatorname{tg} x)' = \left(\frac{\sin x}{\cos x}\right)' = \frac{(\sin x)' \cos x - \sin x (\cos x)'}{\cos^2 x} =$$

$$= \frac{\cos x \cos x - \sin x(-\sin x)}{\cos^2 x} = \frac{\sin^2 x + \cos^2 x}{\cos^2 x} = \frac{1}{\cos^2 x},$$

т.е. $(\operatorname{tg} x)' = \dfrac{1}{\cos^2 x}$.

Найдите производную функции $f(x) = \operatorname{ctg} x$. Используя геометрическое истолкование производной, укажите различие в форме кубической параболы $y = x^3$ и тангенсоиды $y = \operatorname{tg} x$ вблизи начала координат.

4. Функция $\sqrt{2x+1}$ получена суперпозицией линейной функции $f(x) = 2x + 1$ и степенной функции $g(y) = \sqrt{y} = y^{\frac{1}{2}}$. Согласно теореме 3 и формуле II из теоремы 4,

$$(\sqrt{2x+1})' = (y^{\frac{1}{2}})'(2x+1)' = \frac{1}{2}y^{-\frac{1}{2}}\cdot 2 = 2\cdot\frac{1}{2}(2x+1)^{-\frac{1}{2}} = \frac{1}{\sqrt{2x+1}}.$$

Функция $\sin 2x$ представляет собой суперпозицию функций $f(x) = 2x$ и $g(y) = \sin y$. Применяем теорему 3:

$$(\sin 2x)' = (\sin y)'(2x)' = \cos y \cdot 2 = 2\cos 2x.$$

Функция $e^{\sin 2x}$ образована суперпозицией функций $f(x) = \sin 2x$ и $g(y) = e^y$. Следовательно,

$$(e^{\sin 2x})' = (e^y)'(\sin 2x)' = e^y \cdot 2\cos 2x = 2\cos 2x \cdot e^{\sin 2x}.$$

Продифференцируйте функции $(\sin x)^2$, $\ln \operatorname{tg} x$, $-\cos(2x + \frac{\pi}{6})$, $\arcsin 2x$.

Следующие три предложения относятся к числу основных результатов классического дифференциального исчисления. Теорема Лагранжа является эффективным средством исследования функций, используется в доказательстве многих теорем анализа.

Теорема 5. Если функция $f(x)$ дифференцируема на отрезке $[a, b]$, то внутри него существует точка ξ такая, что $f(x) - f(a) = f'(\xi)(b - a)$.

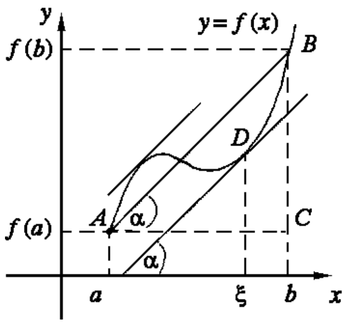


Рис. 22

ДОКАЗАТЕЛЬСТВО. Почти очевидны следующие геометрические соображения (точное доказательство требует, конечно, более убедительных мотивировок). Прямую, содержащую отрезок AB, соединяющий концы графика $y = f(x)$ в пределах отрезка $[a, b]$ (рис. 22), будем перемещать параллельно себе в направлении, где она будет пересекать график. В последний момент прямая эта займет положение касательной к графику в некоторой точке D. Пусть точка D имеет абсциссу ξ, а касательная образует с осью Ox угол α. Очевидно, что $\alpha = \angle BAC$. Значит, $\operatorname{tg}\angle BAC = \operatorname{tg}\alpha$.

Но $\operatorname{tg}\angle BAC = \frac{BC}{AC} = \dfrac{f(b)-f(a)}{b-a}$, а согласно геометрическому смыслу производной, $\operatorname{tg}\alpha = f'(\xi)$. Итак, $\dfrac{f(b)-f(a)}{b-a} = f'(\xi)$, откуда и следует требуемое равенство. □

Французский ученый Жозеф Луи Лагранж (1736–1813) добился выдающихся достижений во многих областях чистой и прикладной математики. В своем основном труде «Аналитическая механика» он свел все задачи о движении к некоторым общим уравнениям, что позволило подключить к решению проблем механики мощный аппарат математического анализа.

Большую роль в распространении идей анализа сыграл первый учебник по этому предмету, изданный в 1696 году маркизом де Лопиталем (1661–1704). В сущности, это была обработка лекций, которые читал ему (одному) Иоганн Бернулли. Написанная в легком стиле Лейбница, книга затрагивала все важнейшие вопросы исчисления бесконечно малых. Новым в ней был лишь весьма оригинальный прием нахождения предела дроби, числитель и знаменатель которой оба стремятся к нулю или к бесконечности. Этот эффектный метод «раскрытия неопределенностей вида $\dfrac{0}{0}$ или $\dfrac{\infty}{\infty}$» известен как правило Лопиталя. Впоследствии выяснилось, что и эту замечательную теорему знатный ученик получил за материальную компенсацию от Иоганна Бернулли,– но от переименования решили воздержаться.

<u>Теорема 6</u>. Пусть функции $f(x)$ и $g(x)$ дифференцируемы в некоторой окрестности точки a и $\lim\limits_{x\to a} f(x) = \lim\limits_{x\to a} g(x) = 0$ (или $\lim\limits_{x\to a} f(x) = \lim\limits_{x\to a} g(x) = \infty$). Тогда если существует предел отношения производных $\lim\limits_{x\to a} \dfrac{f'(x)}{g'(x)}$, то существует и предел отношения самих функций при $x \to a$, причем $\lim\limits_{x\to a} \dfrac{f(x)}{g(x)} = \lim\limits_{x\to a} \dfrac{f'(x)}{g'(x)}$. □

Производная $f'(x)$ функции $f(x)$ в свою очередь может иметь производную, которая называется второй производной функции $f(x)$ и обозначается $f''(x)$. Дифференцируя (если возможно) эту функцию, получают третью производную $f'''(x)$ функции $f(x)$ и т.д. Следующие две производные по традиции обозначаются римскими цифрами: $f^{IV}(x)$ и $f^{V}(x)$, а начиная с шестой – арабскими в скобках: $f^{(6)}(x)$, $f^{(7)}(x)$ и т.д., в общем виде пишут $f^{(n)}(x)$ – производная n-го порядка функции $f(x)$.

Найдем для примера производные высших порядков для функции $\sin x$. Согласно таблице производных (теорема 4, V), $(\sin x)' = \cos x$. Тогда $(\sin x)'' = (\cos x)' = -\sin x$, $(\sin x)''' = (-\sin x)' = -\cos x$, $(\sin x)^{IV}(x) = (-\cos x)' = \sin x$,– дальше все повторится, т.е. $(\sin x)^{(n+4)}(x) = (\sin x)^{(n)}$. Найдите соответствующие старшие производные для косинуса.

Если производная по Ньютону – это скорость движения, то вторая производная – скорость изменения скорости – есть не что иное, как ускорение.

В истории науки вообще и в истории математики в частности известно много случаев, когда первооткрыватели не вполне осознавали значение полученных ими результатов. Иоганн Бернулли недооценил возможности «правила Лопиталя» (и, как оказывается, впоследствии сильно сокрушался по поводу совершенной сделки), Ферма вообще не воспринимал всерьез свои математические работы, Лагранж не придал особого значения сделанным им первым шагам в теории групп. Точно так же и Брук Тейлор (1685–1731), секретарь Лондонского королевского общества (президентом был сэр Исаак Ньютон), опубликовав в 1715 году новый метод приближенного решения уравнений, лишь мимоходом упомянул о соотношении, известном ныне как формула Тейлора, которое уже вскоре по своей значимости было признано одной из главных теорем анализа.

Теорема 7. Если функция $f(x)$ имеет в некоторой окрестности точки a производные до порядка $n+1$ включительно, то имеет место следующее ее представление:

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \dots +$$

$$+\frac{f^{(n)}(a)}{n!}(x-a)^n + r_n(x),$$

где $r_n(x) \to 0$ при $x \to a$. $\Box$

Восклицательные знаки, стоящие в знаменателях дробей, использованы для краткой записи произведений: $2! = 1 \cdot 2$, $3! = 1 \cdot 2 \cdot 3, \dots,$ $n! = 1 \cdot 2 \cdot \dots \cdot n$; для единообразия (т.е. из эстетических соображений) используют и $1! = 1$. Подробнее о факториалах см. в §V.2.

Многочлен

$$T_n(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}}{n!}(x-a)^n$$

называется многочленом Тейлора степени n для функции $f(x)$, а $r_n(x)$ – остаточным членом.

В следующем параграфе мы увидим, как теорема Лагранжа, правило Лопиталя и формула Тейлора используются для исследования функций.

Первой публикацией по анализу была шестистраничная геометрическая статья Лейбница 1684 года. Вскоре к нему присоединились братья Якоб (1654–1705) и Иоганн (1667–1748) Бернулли, самые яркие представители швейцарской математической династии, давшей миру в трех поколениях восемь первоклассных ученых. Новая математическая наука, Анализ бесконечно малых, оказалась необычайно плодотворной и быстро распространялась по Европе. На фоне этих успехов и возник многолетний тягостный спор о приоритете между последователями Ньютона и Лейбница. Дело в том, что Ньютон пришел к основным идеям дифференциального исчисления еще в 1660-х годах (скрываясь в деревне от чумы) и, не публикуя своих результатов, вел о них переписку с Лейбницем. Это и дало повод к обвинению последнего в плагиате. В полемику были вовлечены не только математики, но и политики, возведшие чисто научные вопросы в ранг национального достоинства. Только в первой половине XIX века английские математики приняли, наконец, простые, не затемняющие сути дела обозначения и термины Лейбница. Историко-математические исследования нового времени не оставляют сомнений в том, что Ньютон и Лейбниц пришли к своему великому открытию независимо.

§ 5. Дифференциальное исчисление: примеры приложений

Все функции, встречающиеся в этом параграфе, считаются дифференцируемыми достаточное число раз, т.е. имеющими все производные до нужного порядка.

Одним из примеров использования производной является выяснение вопроса о характере изменения функции на заданном интервале. Напомним, что функция $f(x)$ называется возрастающей на некотором интервале, если для любых двух точек a, b из этого интервала при $a < b$ справедливо неравенство $f(a) \leqslant f(b)$. Если при $a < b$ выполняется неравенство $f(a) \geqslant f(b)$, то функция $f(x)$ называется убывающей на данном интервале.

Сначала установим критерий (т.е. необходимое и достаточное условие) постоянства дифференцируемой функции.

Теорема 1. Функция $f(x)$ постоянна на данном интервале тогда и только тогда, когда $f'(x) = 0$ во всех его точках.

ДОКАЗАТЕЛЬСТВО. 1) Если $f(x)$ постоянна, то $f'(x) = 0$ в любой точке.

2) Пусть $f'(x) = 0$ всюду на данном интервале и пусть a, b – произвольные точки этого интервала, $a < b$. Согласно теореме Лагранжа (теорема 5 из §4), между a и b лежит точка ξ такая, что $f(b) - f(a) = f'(\xi)(b - a)$. Из условия следует, что $f'(\xi) = 0$, откуда $f(b) - f(a) = 0$, т.е. $f(a) = f(b)$ для любых двух точек a, b – функция $f(x)$ постоянна. □

Критерием возрастания дифференцируемой функции является

Теорема 2. Функция $f(x)$ возрастает на данном интервале тогда и только тогда, когда $f'(x) \geqslant 0$ во всех его точках.

ДОКАЗАТЕЛЬСТВО. 1) Если $f(x)$ возрастает на данном интервале и a – произвольная его точка, то числа $x - a$ и $f(x) - f(a)$ имеют на этом интервале одинаковый знак. Следовательно, $\dfrac{f(x) - f(a)}{x - a} \geqslant 0$ и, по свойствам пределов, $\lim\limits_{x \to a} \dfrac{f(x) - f(a)}{x - a} \geqslant 0$, т.е. $f'(a) \geqslant 0$ в любой точке a интервала.

2) Пусть $f'(x) \geqslant 0$ всюду на некотором интервале и пусть a, b – произвольные точки этого интервала, $a < b$, т.е. $b - a > 0$. Согласно теореме Лагранжа, между точками a и b лежит точка ξ такая, что $f(b) - f(a) = f'(\xi)(b - a)$. Из условия следует, что $f'(\xi) \geqslant 0$, откуда $f(b) - f(a) \geqslant 0$, т.е. $f(a) \leqslant f(b)$ при $a < b$,– функция $f(x)$ возрастающая. □

Докажите следующий критерий убывания дифференцируемой функции на интервале.

Теорема 3. Функция $f(x)$ убывает на данном интервале тогда и только тогда, когда $f'(x) \leqslant 0$ во всех его точках. □

Применим доказанные условия возрастания и убывания дифференцируемой функции к конкретным случаям.

1. Степенные функции. Так как $(x^2)' = 2x$, то функция $y = x^2$ при $x > 0$ возрастает, а при $x < 0$ убывает (см. рис. 10а). Для кубической параболы $y = x^3$ имеем $(x^3)' = 3x^2 \geqslant 0$,– эта функция всюду возрастающая (см. рис. 10б). В случае гиперболы

$y = \dfrac{1}{x}$ получаем $\left(\dfrac{1}{x}\right)' = -\dfrac{1}{x^2} < 0$,– функция всюду убывает (кажущийся контрпример: $-2 < 2$ и $f(-2) = -\dfrac{1}{2} < \dfrac{1}{2} = f(2)$. Ответ: значения аргумента -2 и 2 лежат в разных частях распадающейся области определения. Функция $y = \dfrac{1}{x}$ убывает на каждом интервале своей области определения).

Применением теоремы 2 докажите, что $y = \sqrt{x}$ – возрастающая функция.

2. Показательные функции. Так как $(a^x)' = a^x \ln a$ и $a^x > 0$, то знак производной зависит от знака числа $\ln a$. Если $a > 1$, то $\ln a > 0$, если $a < 1$, то $\ln a < 0$. Следовательно, при $a > 1$ функция $y = a^x$ всюду возрастает, а при $a < 1$ – всюду убывает.

3. Логарифмические функции. Рассуждениями, аналогичными тем, которые были проведены для показательных функций, покажите, что функция $y = \log_a x$ при $a > 1$ возрастает, а при $a < 1$ – убывает.

4. Тригонометрические функции. Производной от синуса является косинус: $(\sin x)' = \cos x$. Поэтому синус возрастает во всех тех интервалах, где $\cos x > 0$ (т.е. где график косинуса лежит выше оси абсцисс Ox), и убывает там, где $\cos x < 0$ (т.е. где график косинуса расположен под осью абсцисс). Убедитесь в этом, сопоставляя графики синуса и косинуса; выпишите несколько интервалов возрастания и несколько интервалов убывания синуса. Проведите аналогичные рассуждения об интервалах возрастания и убывания косинуса. Что можно сказать о тангенсе и котангенсе?

5. Обратные тригонометрические функции. Так как $(\arcsin x)' = \dfrac{1}{\sqrt{1-x^2}} > 0$, то арксинус – возрастающая функция. Построим, наконец, ее график, используя характерные точки. Какой угол x из интервала $[-\frac{\pi}{2}, \frac{\pi}{2}]$ имеет синус, равный единице? Только один: $x = \dfrac{\pi}{2}$. Значит, $\arcsin 1 = \dfrac{\pi}{2}$. Ставя аналогичный вопрос о других известных значениях синуса, заполняем таблицу:

x	0	$\frac{1}{2}$	$-\frac{1}{2}$	$\frac{\sqrt{2}}{2}$	$-\frac{\sqrt{2}}{2}$	$\frac{\sqrt{3}}{2}$	$-\frac{\sqrt{3}}{2}$	1	-1
$\arcsin x$	0	$\frac{\pi}{6}$	$-\frac{\pi}{6}$	$\frac{\pi}{4}$	$-\frac{\pi}{4}$	$\frac{\pi}{3}$	$-\frac{\pi}{3}$	$\frac{\pi}{2}$	$-\frac{\pi}{2}$

Весь график арксинуса умещается в прямоугольнике, образованном вертикальными прямыми $x = -1,\ x = 1$ и горизонтальными прямыми $y = -\dfrac{\pi}{2},\ y = \dfrac{\pi}{2}$ (рис. 23).

Большое практическое значение имеет вопрос о нахождении экстремальных значений функции (лат. extremum – крайнее).

Говорят, что функция $f(x)$ имеет в точке a максимум, если в некоторой окрестности этой точки выполняется неравенство $f(x) \leqslant f(a)$, т. е. если значение функции в точке a является наибольшим из всех значений, принимаемых ею в этой окрестности. Функция $f(x)$, по определению, имеет в точке a минимум, если в некоторой окрестности этой точки справедливо неравенство $f(x) \geqslant f(a)$, т.е. если значение функции в точке a является наименьшим из всех значений, принимаемых ею в этой окрестности. Эти понятия носят локальный характер, т.е. относятся лишь к точке a и некоторой, быть может, очень малой ее окрестности. Неравенства, характеризующие максимум и минимум, вне указанных окрестностей могут не выполняться, т.е. максимум функции – это не обязательно ее наибольшее значение, а минимум – не обязательно наименьшее, причем некоторый минимум может оказаться больше какого-либо максимума (рис. 24).

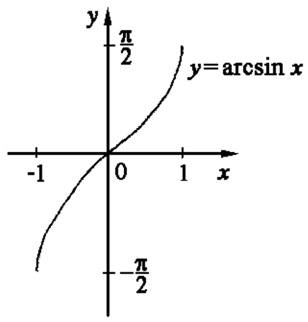


Рис. 23

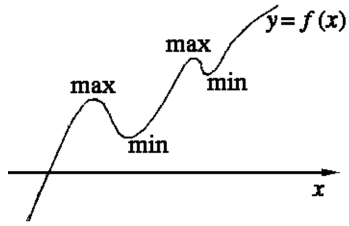


Рис. 24

Максимумы и минимумы функции называются ее экстремумами, а значения, принимаемые функцией в соответствующих точках,– ее экстремальными значениями.

В дифференциальном исчислении известны признаки, позволяющие выяснить вопрос о том, будет ли данная точка a точкой экстремума рассматриваемой функции $f(x)$. Следующее важное наблюдение показывает, в частности, что касательная к графику функции в точке экстремума параллельна оси абсцисс.

Теорема 4 (необходимое условие экстремума). Если функция $f(x)$ имеет в точке a экстремум, то ее производная в этой точке обращается в ноль: $f'(a) = 0$. $\square$

Обращение в ноль производной – это лишь необходимое условие экстремума, но оно не является достаточным. Например, у функции $f(x) = x^3$ в точке $a = 0$ производная $f'(x) = 3x^2$ обращается в ноль, но ни максимума, ни минимума кубическая парабола в нуле не имеет (см. рис. 10б).

Точки, в которых производная функции $f(x)$ обращается в ноль, называются критическими точками этой функции. В силу теоремы 4, все точки экстремума функции являются ее критическими точками.

Для многих (но не для всех) функций поведение в окрестности точки максимума соответствует следующему, кажущемуся интуитивно ясным представлению: слева от той точки a, где максимум достигается, функция $f(x)$ возрастает (график с увеличением x идет вверх), а справа от точки a – убывает (с ростом x график устремляется вниз). В силу теорем 2 и 3 получаем, что слева от точки максимума $f'(x) > 0$, а справа от нее $f'(x) < 0$, т. е., как говорят, при переходе через точку a производная $f'(x)$ меняет знак с плюса на минус. Таким образом, точка максимума отделяет интервал возрастания функции от интервала ее убывания.

И точно так же в окрестности точки минимума многие (но не все) функции ведут себя в соответствии с непосредственной интуицией: если функция $f(x)$ имеет в точке a минимум, то при подходе к этой точке слева $f(x)$ убывает и, значит, $f'(x) < 0$ по теореме 3, а при удалении от a вправо функция $f(x)$ возрастает и, следовательно, $f'(x) > 0$, – при переходе через точку a производная $f'(x)$ меняет знак с минуса на плюс, точка минимума отделяет интервал убывания функции от интервала ее возрастания.

(Вот пример «плохой» функции: $f(x) = 2 - x^2(2 + \sin\dfrac{1}{x})$ при $x \neq 0$ и $f(0) = 2$. В точке $a = 2$ она имеет максимум, но слева от этой точки $f(x)$ – из-за своих бесконечных колебаний – не является возрастающей функцией, а справа – не является убывающей.)

Сформулируем теперь правило отыскания экстремальных значений дифференцируемой функции $f(x)$ в некотором открытом интервале.

1. Найти производную $f'(x)$.

2. Найти все решения уравнения $f'(x) = 0$ в указанном интервале, т.е. найти все критические точки функции $f(x)$,

принадлежащие данному интервалу (будем считать, что их конечное число).

3. Для каждой критической точки a исследовать поведение производной в окрестности этой точки. Если при переходе через точку a производная $f'(x)$ меняет знак с + на $-$, то в точке a – максимум, если же при переходе через точку a знак производной меняется с $-$ на +, то в точке a – минимум. Если производная не меняет знак, в точке a экстремума нет.

4. В каждой точке экстремума a вычислить значение функции $f(a)$.

Для примера найдем все экстремальные значения функции $f(x) = x^5 - 5x + 1$. Дифференцируем: $f'(x) = 5x^4 - 5$. Приравниваем производную нулю: $5x^4 - 5 = 0$. Решая это уравнение: $x^4 - 1 = (x-1)(x+1)(x^2+1) = 0$, получаем два действительных корня: $x_1 = 1,\ x_2 = -1$, две критические точки функции $f(x)$. Так как $f'(0) = -5 < 0$, а $f'(2) = 75 > 0$, в точке $x_1 = 1$ имеем минимум. Поскольку $f'(-2) > 0$, а $f'(0) < 0$, в точке $x_2 = -1$ будет максимум. Минимальное значение функции достигается в точке $x_1 = 1$, оно равно $f(1) = -3$, максимальное значение функция принимает в точке $x_2 = -1$, и оно равно $f(-1) = 5$. Но, конечно, функция имеет и бóльшие и меньшие значения.

Если требуется найти наибольшее и наименьшее значение функции $f(x)$ на замкнутом интервале $[a, b]$ (они существуют по теореме Вейерштрасса), то к экстремальным значениям этой функции, найденным в открытом интервале (a, b), нужно добавить еще значения $f(a)$ и $f(b)$ на концах интервала и из всех полученных чисел выбрать наибольшее и наименьшее. Например, для функции $f(x) = x^5 - 5x + 1$ в замкнутом интервале $[-2;\ 1,5]$ имеем: $f(-2) = -21,\ f_{\max} = f(-1) = 5,\ f_{\min} = f(1) = -3,$ $f(1,5) \approx 1,1$ и, следовательно, наибольшим значением будет 5 (в точке максимума), а наименьшим -21 (на левом конце).

Найдите наибольшее и наименьшее значения 1) функции $f(x) = x^2 + x - 1$ на интервале $[-1,\ 1]$, 2) функции $f(x) = 2x^3 - 9x^2 + 12x - 4$ на интервале $[1,\ 3]$.

Точка экстремума квадратичной функции $f(x) = ax^2 + bx + c$ называется вершиной параболы, являющейся графиком этой функции. Найдите вершину этой параболы и сравните с приведенными в §2 данными.

В качестве приложения правила Лопиталя расмотрим первый замечательный предел. У дроби $\dfrac{\sin x}{x}$ при $x \to 0$ числитель и знаменатель оба стремятся к нулю, т.е имеем неопределенность вида $\dfrac{0}{0}$. По правилу Лопиталя предел такой дроби существует, если существует предел отношения производной числителя к производной знаменателя, и тогда эти пределы равны. В нашем случае

$$\lim_{x\to 0}\frac{\sin x}{x}=\lim_{x\to 0}\frac{(\sin x)'}{x'}=\lim_{x\to 0}\frac{\cos x}{1}=\lim_{x\to 0}\cos x=1.$$

Итак, $\lim\limits_{x\to 0}\dfrac{\sin x}{x}=1$.

Вычислите пределы $\lim\limits_{x\to 0}\dfrac{\mathrm{tg} x}{x}$, $\lim\limits_{x\to 1}\dfrac{1-\cos x}{1-x^2}$, $\lim\limits_{x\to 0}\dfrac{1-e^x}{\sin x}$.

Пусть теперь нужно найти предел $\lim\limits_{x\to\infty}\dfrac{x^2+x-1}{3x^2-2x+1}$. Числитель и знаменатель оба стремятся к бесконечности – имеем неопределенность вида $\dfrac{\infty}{\infty}$. Применяем правило Лопиталя:

$$\lim_{x\to\infty}\frac{x^2+x-1}{3x^2-2x+1}=\lim_{x\to\infty}\frac{(x^2+x-1)'}{(3x^2-2x+1)'}=\lim_{x\to\infty}\frac{2x+1}{6x-2}=$$

(полученное выражение снова является неопределенностью вида $\dfrac{\infty}{\infty}$, продолжим дифференцирование)

$$=\lim_{x\to\infty}\frac{(2x+1)'}{(6x-2)'}=\lim_{x\to\infty}\frac{2}{6}=\frac{1}{3}.$$

Итак, $\lim\limits_{x\to\infty}\dfrac{x^2+x-1}{3x^2-2x+1}=\dfrac{1}{3}$.

Анализируя этот пример, приходим к выводу, что при $x\to\infty$ предел отношения двух многочленов $\dfrac{P_m(x)}{Q_n(x)}$ степеней соответственно m и n будет равен нулю, если $m<n$ (дифференцирование числителя кончится раньше, в знаменателе еще останутся степени x), равен бесконечности, если $m>n$ (на каком-то этапе в знаменателе окажется постоянное число, а в числителе еще будет x, стремящийся к бесконечности), а при $m=n$ предел совпадет с отношением коэффициентов при старших степенях числителя и знаменателя. Используя это правило, раскройте следующие неопределенности вида $\dfrac{\infty}{\infty}$ при $x\to+\infty$:

$$\frac{x^3-2x+1}{5x^2-x+2},\quad \frac{-x^4+2x^3+x-1}{5x^4+1},\quad \frac{x^3+3x^2-2x+1}{x^5+2}.$$

Следующий пример показывает, что экспонента растет быстрее любой степенной функции. Найдем предел $\lim\limits_{x\to\infty}\dfrac{e^x}{x^n}$, где n – любое натуральное число. Правило Лопиталя придется применить многократно. Производные будем писать сразу, помня, что $(e^x)' = e^x$ и $(x^\alpha)' = \alpha x^{\alpha-1}$. Имеем:

$$\lim_{x\to\infty}\frac{e^x}{x^n} = \lim_{x\to\infty}\frac{e^x}{nx^{n-1}} = \lim_{x\to\infty}\frac{e^x}{n(n-1)x^{n-2}} = \cdots =$$

$$= \lim_{x\to\infty}\frac{e^x}{n(n-1)\cdot\dots\cdot 2\cdot x} = \lim_{x\to\infty}\frac{e^x}{1\cdot 2\cdot\dots\cdot n} = \infty,$$

так как в знаменателе последней дроби стоит постоянное число $n!$, а числитель по-прежнему стремится к бесконечности. Итак, $\lim\limits_{x\to\infty}\dfrac{e^x}{x^n} = \infty$ при любом n. В §2 мы пытались представить себе параболу $y = x^{1000}$. При $x > 1$ она почти вертикально уходит вверх. Начиная с некоторого x, график экспоненты будет лежать выше не только этой параболы, но и, скажем, выше параболы $y = x^{1000000}$. Трудно вообразить себе эти картины.

Выдающееся значение формулы Тейлора заключается в том, что с ее помощью функция может быть с какой угодно точностью приближена многочленом:

$$f(x) \approx f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n.$$

Следовательно, определение значений функции сводится к выполнению четырех арифметических операций: сложения, вычитания, умножения и деления,– единственных действий, которые вычисляющий человек может производить сам по себе, т.е. без вспомогательных математических средств. Природа не заложила в наш разум ни тригонометрических, ни логарифмических таблиц, но дала возможность создать их при необходимости, пользуясь, например, формулой Тейлора.

Возьмем функцию $f(x) = e^x$ и положим $a = 0$. Поскольку все производные экспоненты равны ей самой и в точке $a = 0$ превращаются в единицу, имеем:

$$e^x \approx 1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!}.$$

Беря n достаточно большим, будем находить значения экспоненты с какой угодно точностью. При $x = 1$ получаем замечательное представление числа e:

$$e \approx 1 + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!}.$$

Пусть $f(x) = \sin x$. В §4 мы нашли все производные синуса: $(\sin x)' = \cos x$, $(\sin x)'' = -\sin x$, $(\sin x)''' = -\cos x$, $(\sin x)^{IV} = \sin x$ и дальше все повторяется. Полагая в формуле Тейлора $a = 0$ и опуская равные нулю слагаемые, находим:

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \ldots$$

Если ограничиться только первыми двумя слагаемыми: $\sin x \approx x - \frac{x^3}{6}$, то на отрезке $[0, \frac{\pi}{4}]$ синус будет отличаться от указанного многочлена лишь на $\dfrac{1}{400}$. Покажите, что

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \ldots$$

Следующая важнейшая функция – натуральный логарифм. Чтобы сохранить стандартную точку $a = 0$, придется взять функцию $f(x) = \ln(1 + x)$, ибо $\ln x$ при $a = 0$ не существует. Оказывается,

$$\ln(x + 1) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \ldots$$

В частности, при $x = 1$ получаем открытую Лейбницем и поразившую его современников формулу:

$$\ln 2 = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \ldots$$

Геометрически формула Тейлора означает, что ход любой кривой $y = f(x)$ на некотором ее участке можно с какой угодно точностью изобразить с помощью наложения парабол различных порядков (например, синусоида получается вблизи нуля из прямой $y = x$ и парабол нечетных порядков).

Дифференциальным уравнением называется такое уравнение, в котором неизвестной является функция, и при этом в уравнение входит не только она сама, но и ее производные. Огромное число процессов, изучаемых в физике, астрономии, биологии, химии и других науках, записывается в виде дифференциальных уравнений. Приведем несколько простейших примеров.

1. Распад радиоактивного вещества. Известно, что скорость радиоактивного распада пропорциональна имеющемуся количеству вещества. Пусть $y(x)$ обозначает количество вещества в момент времени x. Скорость изменения функции $y(x)$ выражается ее производной $y'(x)$. В силу указанной пропорциональности имеем $y' = ky$, где k – некоторая постоянная. Очевидно, что $k < 0$, так как функция $y(x)$ – убывающая. Решением дифференциального уравнения $y' = kx$ является любая функция $y = ce^{kx}$, где c – произвольная постоянная. В самом деле, $y' = (ce^{kx})' = cke^{kx} = k(ce^{kx}) = ky$. Нам известно начальное количество вещества, т.е. значение функции y при $x = 0$. Пусть это будет y_0. Тогда из общего решения $y = ce^{kx}$ при $x = 0$ получаем $y_0 = ce^{k\cdot 0} = c$, т.е. $c = y_0$. Итак, количество вещества в произвольный момент времени x выражается формулой $y = y_0e^{kx}$. Для каждого радиоактивного вещества коэффициент k имеет свое конкретное значение, устанавливаемое экспериментально. Предположим, что мы имеем дело с радиоактивным изотопом углерода (он применяется для датировки археологических находок). Период полураспада этого вещества составляет 5600 лет (т.е. за такое время половина радиоактивного углерода превращается в обычный углерод). Таким образом, $\dfrac{y_0}{y_0e^{5600k}} = 2$, откуда $e^{-5600k} = 2$. Логарифмируя, получаем $-5600k = \ln 2 = 0,692$ и, значит, $k = -\dfrac{0,692}{5600} = -0,0001236$. Итак, если вначале было y_0 килограммов радиоактивного углерода, то через x лет останется $y = y_0e^{-0,0001236x}$ килограммов этого изотопа.

2. Рост капитала при непрерывном начислении процентов. Если на вложенный капитал начисляется k процентов и процесс непрерывен, то скорость возрастания капитала будет пропорциональна исходному вкладу с коэффициентом пропорциональности k, т.е. $y' = ky$. Мы приходим к тому же уравнению, которое описывает радиоактивный распад, но теперь $k > 0$,– капитал растет. Решением будет функция $y = y_0e^{kx}$, где y_0 – вложенный капитал. Например, найдем, за сколько лет t начальный капитал y_0 удвоится, если $k = 0,04$ (т.е. 4%). Имеем: $\dfrac{y_0e^{0,04t}}{y_0} = 2$, откуда $e^{0,04t} = 2$. Логарифмируя, получаем $0,04t = \ln 2$, т.е. $t = \dfrac{\ln 2}{0,04} = \dfrac{0,692}{0,04} = 17,3$ (года).

Примеры 1 и 2 являются прекрасной иллюстрацией того, что явления, по содержанию абсолютно не схожие, могут иметь одну и ту же математическую модель.

3. Гармонические колебания. Можно сказать, что важнейшие физические открытия были сделаны при изучении колебательных процессов разных типов. Пусть материальная точка с массой m движется вдоль вертикальной оси Oy под влиянием упругой силы пружины (такое движение называется гармоническим колебанием). Согласно второму закону Ньютона, $ma = F$, где a – ускорение, а F – действующая сила. Сила пружины пропорциональна смещению точки и направлена в противоположную сторону (закон Гука), т.е. $F = -ky$, где k – коэффициент, называемый жесткостью пружины. Ускорение, как мы в свое время отмечали,– это вторая производная от пути по времени. Следовательно, из упомянутого закона Ньютона получаем дифференциальное уравнение с неизвестной функцией $y = y(x)$ (смещение точки в зависимости от времени): $my'' = -ky$, или, полагая $\dfrac{k}{m} = \omega^2$ (из физических соображений), $y'' = -\omega^2 y$. Общим решением этого уравнения будет $y = A\sin(\omega x + \varphi)$ (проверьте!). Величина A называется амплитудой колебания, ω – его частотой, а φ – фазой (см. §2).

Кроме теории колебаний, наиболее важные технические приложения дифференциальные уравнения находят в теории автоматического управления, в частности, летательными и космическими аппаратами. Среди ученых, внесших основной вклад в решение этих проблем, одно из первых мест занимает академик Мстислав Всеволодович Келдыш (1911–1978), президент Академии наук СССР в 1961–1975 годы, ведущий теоретик космонавтики. Математические формулы и теоремы, полученные Келдышем, позволили также разработать методы борьбы с разрушительными вибрациями в различных частях самолета, создать скоростные речные суда на подводных крыльях.

Выдающуюся роль в математической теории оптимального управления сыграли работы академика Льва Семеновича Понтрягина (1908–1988). В 13-летнем возрасте в результате несчастного случая он потерял зрение и, после некоторых сомнений (как возможные варианты рассматривались музыка и история), выбрал путь математика. Всемирную известность принесли ученому результаты в теории дифференциальных уравнений, топологии, теории групп, во многих других разделах теоретической и прикладной математики. Большое значение имели его идеи в области реформирования школьного курса математики. Связь с практическими проблемами Л.С.Понтрягин считал совершенно необходимой для всякой математической дисциплины: «Приложения математики не только оправдывают ее существование, но и дают новые интересные задачи, которые невозможно получить из глубины разума».

§ 6. Интегральное исчисление

«Смысл там, где змеи интеграла»,– писал в одном из своих стихотворений Валерий Брюсов, интересовавшийся физическими и математическими теориями.

Понятие интеграла принадлежит к высшим достижениям человеческой мысли и, несмотря на известную сложность этой конструкции, в каком-то виде должно быть введено в общую культуру. В старых изданиях «Словаря русского языка» под интегралом понималась «величина, рассматриваемая как сумма своих бесконечно малых частей». При всей нечеткости этого выражения, главный смысл, как мы сейчас увидим, в нем сохранен.

Пусть $f(x)$ – некоторая положительная функция, непрерывная на отрезке $[a, b]$. Криволинейной трапецией, определенной этой функцией, называется фигура, ограниченная слева вертикальной прямой $x = a$, справа – вертикальной прямой $x = b$, снизу – осью абсцисс Ox, сверху – графиком функции $y = f(x)$.

Желая вычислить площадь этой фигуры, разобьем отрезок $[a, b]$ на n частей (не обязательно равных) точками $x_1,\ x_2, \ldots,\ x_{n-1}$, в каждой части произвольным образом выберем точку: $\xi_1,\ \xi_2, \ldots,$ ξ_n, – и составим сумму

$$S_n = f(\xi_1)(x_1 - a) + f(\xi_2)(x_2 - x_1) + \cdots + f(\xi_n)(b - x_{n-1}), \quad (1)$$

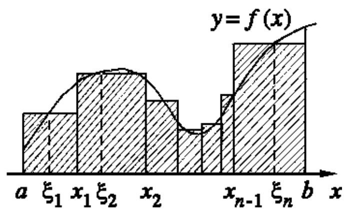


Рис. 25

называемую интегральной суммой для функции $f(x)$ на отрезке $[a, b]$, соответствующей данному разбиению этого отрезка и данному выбору точек $\xi_1, \xi_2, \ldots, \xi_n$. Геометрически S_n выражает площадь заштрихованной на рис. 25 ступенчатой фигуры, составленной из прямоугольников с основаниями $x_1 - a, x_2 - x_1, \ldots,$ $b - x_{n-1}$ и высотами соответственно $f(\xi_1), f(\xi_2), \ldots, f(\xi_n)$. Понятно, что чем мельче мы будем дробить отрезок $[a, b]$, тем ближе будут получающиеся интегральные суммы к площади S криволинейной трапеции. Обозначим через λ наибольшую из длин участков, на которые разбит отрезок $[a, b]$. Определенным интегралом от функции $f(x)$ на отрезке $[a, b]$ называется предел интегральных сумм при $\lambda \to 0$ (при этом, конечно, $n \to \infty$). Этот предел обозначается через $\int_a^b f(x)dx$.

Итак,

$$\int_a^b f(x)dx = \lim_{\lambda \to 0} S_n =$$

$$= \lim_{\lambda \to 0} [f(\xi_1)(x_1 - a) + f(\xi_2)(x_2 - x_1) + \dots + f(\xi_n)(b - x_{n-1})]. \quad (2)$$

По самому своему определению величина $\int_a^b f(x)dx$ совпадает с площадью криволинейной трапеции, образованной положительной функцией $f(x)$ на отрезке $[a, b]$, т.е. $S = \int_a^b f(x)dx$.

В записи $\int_a^b f(x)dx$ («интеграл от a до b от функции $f(x)$ по dx») выражение $f(x)dx$ называется подинтегральным выражением, $f(x)$ – подинтегральной функцией, числа a и b – пределами интегрирования: a – нижним, b – верхним. Сам знак интеграла – впервые его начертал в 1675 году Лейбниц – это стилизованная первая буква латинского слова Summa. Название «интеграл» (от латинского integer – целый) ввел Иоганн Бернулли в 1690 году.

В наших построениях предполагалось, что $f(x) > 0$ на интервале $[a, b]$. В общей конструкции определенного интеграла $\int_a^b f(x)dx$ это ограничение снимается: для произвольной непрерывной функции $f(x)$ строится интегральная сумма (1) и ищется предел (2), который может оказаться как положительным, так и отрицательным числом (или нулем). Это определение интеграла восходит к Коши.

Вычисление площадей, ограниченных кривыми, методом разбиения на (бесконечно) большое число (бесконечно) малых частей с последующим их суммированием производил еще Архимед, которого поэтому считают родоначальником интегрального исчисления. (В отличие от Пифагора и Эвклида, о которых не сохранилось никаких достоверных сведений, фигура Архимеда имеет более реальные очертания. Известно, что он жил в Сиракузах (Сицилия) и погиб от руки римского легионера в момент, когда что-то чертил на песке. Величайший математик древности, Архимед был также выдающимся механиком и изобретателем. Появившиеся в XVI веке латинские издания его трудов оказали большое влияние на формирование новой математики.)

Конструкция определенного интеграла возникает в очень многих физических и технических задачах, она оказалась тесно связанной с производной и стала основным инструментом при решении дифференциальных уравнений.

Непосредственный подсчет определенного интеграла как предела интегральных сумм оказывается чрезвычайно сложным делом даже для самых незатейливых функций. До великого открытия Ньютона и Лейбница (теорема 1) каждое достижение в этой области высоко ценилось и энергично обсуждалось. Лишь после установления связи между интегралом и производной дело было «поставлено на поток» и с новой вершины открылась безграничная перспектива приложений математики в самых разных областях человеческой деятельности.

Процесс отыскания интеграла от данной функции называется ее интегрированием, а совокупность понятий, методов и теорем, связанных с интегрированием, образует интегральное исчисление функций.

Рассмотрим определенный интеграл с закрепленным нижним пределом и меняющимся верхним. С изменением величины верхнего предела будет принимать разные значения и величина самого интеграла, т.е. он становится функцией своего верхнего предела: $\int_a^x f(x)dx = \Phi(x)$.

Заметим, что, например, $\Phi(a) = \int_a^a f(x)dx = 0$ (площадь над или под вырожденным в точку отрезком $[a, a]$), а $\Phi(b) = \int_a^b f(x)dx$.

Основная теорема анализа, установленная независимо Ньютоном и Лейбницем, гласит, что производная от интеграла, рассматриваемого как функция верхнего предела, существует и равна подинтегральной функции.

Теорема 1.

$$\Phi'(x) = \left(\int_a^x f(x)dx\right)' = f(x).$$

ДОКАЗАТЕЛЬСТВО. По определению производной,

$$\Phi'(x) = \lim_{\xi \to x} \frac{\Phi(x) - \Phi(\xi)}{x - \xi}.$$

Разность $\Phi(x) - \Phi(\xi)$ выражает площадь криволинейной трапеции, образованной функцией $f(x)$ на отрезке $[\xi, x]$. Так как

$f(x)$ – непрерывная функция, то, согласно теореме Вейерштрасса, на отрезке $[\xi, x]$ она имеет наименьшее значение m и наибольшее значение M. Упомянутая криволинейная трапеция содержит в себе прямоугольник с основанием $[\xi, x]$ и высотой m и, в свою очередь, содержится в прямоугольнике с тем же основанием и высотой M (рис. 26). Следовательно, между площадями этих трех фигур выполняются неравенства

$$m(x-\xi) \leqslant \Phi(x) - \Phi(\xi) \leqslant M(x-\xi),$$

откуда

$$m \leqslant \frac{\Phi(x) - \Phi(\xi)}{x-\xi} \leqslant M.$$

Устремляя ξ к x, видим, что в пределе $m = M = f(x)$, так что

$$\lim_{\xi \to x} \frac{\Phi(x) - \Phi(\xi)}{x-\xi} = f(x),$$

т.е. $\Phi'(x) = f(x)$. $\square$

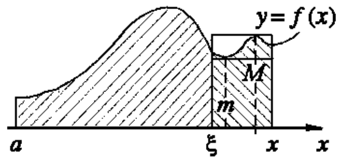


Рис. 26

Чтобы получить отсюда знаменитую формулу Ньютона–Лейбница, позволяющую эффективно вычислять определенные интегралы, потребуется ввести еще одно важное понятие.

Пусть $f(x)$ – непрерывная функция. Первообразной для нее называется такая функция $F(x)$, производная которой равна данной функции $f(x)$, т.е. $F'(x) = f(x)$. Имея перед собой таблицу производных (теорема 3§4), заключаем, например, что $F(x) = \sin x$ является первообразной для $f(x) = \cos x$, ибо $(\sin x)' = \cos x$. Первообразной для экспоненты $f(x) = e^x$ будет она сама, а для функции $f(x) = \frac{1}{x}$ в качестве первообразной можно предложить $F(x) = \ln x$.

Теорема 1 в новой терминологии утверждает, что интеграл с переменным верхним пределом является первообразной для подинтегральной функции. Таким образом, интегрирование оказывается действием, обратным дифференцированию: дифференцируя функцию $f(x)$, мы находим ее производную, а интегрируя – восстанавливаем одну из функций, для которых $f(x)$ является производной, т.е. одну из ее первообразных.

Производная данной функции находится однозначно, но первообразных у функции существует бесконечно много. В самом деле, если $F(x)$ – первообразная для функции $f(x)$, т.е. $F'(x) = f(x)$, то первообразной для $f(x)$ будет и любая функция вида $F(x) + C$, где C – произвольная постоянная, ибо $(F(x) + C)' = F'(x) + C' = F'(x) = f(x)$. Но оказывается, что этим приемом (прибавлением постоянной) и исчерпывается все многообразие первообразных для одной и той же функции.

Теорема 2. Две первообразные для одной и той же функции отличаются разве лишь на постоянную.

ДОКАЗАТЕЛЬСТВО. Пусть $F_1(x)$ и $F_2(x)$ – две первообразные для функции $f(x)$, т.е. $F_1'(x) = F_2'(x) = f(x)$. Тогда $[F_2(x) - F_1(x)]' = F_2'(x) - F_1'(x) = f(x) - f(x) = 0$, – функция $F_2(x) - F_1(x)$ имеет производную, всюду равную нулю. Согласно теореме 1§4, эта функция постоянна. Итак, $F_2(x) - F_1(x) = C$ и, значит, $F_2(x) = F_1(x) + C$. □

Следующая теорема и дает формулу Ньютона–Лейбница.

Теорема 3.

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

где $F(x)$ – произвольная первообразная функции $f(x)$.

(Таким образом, определенный интеграл от функции $f(x)$ на отрезке $[a, b]$ равен разности значений какой-либо первообразной для этой функции, соответствующих правому и левому концам отрезка).

ДОКАЗАТЕЛЬСТВО. Пусть $F(x)$ – какая-нибудь первообразная для функции $f(x)$. Согласно теореме 1, интеграл $\Phi(x) = \int_a^x f(x)dx$ тоже является первообразной для $f(x)$. По теореме 2 имеем: $\Phi(x) = F(x) + C$ для любого x. В частности, при $x = a$ получаем $\Phi(a) = F(a) + C$. Так как $\Phi(a) = 0$, то $C = -F(a)$. Итак, $\Phi(x) = F(x) - F(a)$ для любого x. Полагая $x = b$, находим: $\Phi(b) = F(b) - F(a)$. Но $\Phi(b) = \int_a^b f(x)dx$. Следовательно, $\int_a^b f(x)dx = F(b) - F(a)$, что и требовалось доказать. □

Для удобства разность $F(b) - F(a)$ обозначают через $F(x)|_a^b$, так что формула Ньютона–Лейбница приобретает вид

$$\int_a^b f(x)dx = F(x)|_a^b,$$

где $F(x)$ – какая-либо первообразная для $f(x)$.

Формула Ньютона–Лейбница сводит вычисление определенного интеграла к отысканию первообразной для подинтегральной функции.

Пример 1. Найдем площадь под параболой на отрезке $[0,1]$ (рис. 27). Искомая площадь s – это площадь криволинейной трапеции, левая граница которой вырождается в точку. По смыслу определенного интеграла $s = \int_0^1 x^2 dx$. Теперь нужно предложить какую-нибудь первообразную для функции $f(x) = x^2$. Это сделать нетрудно: $\left(\dfrac{x^3}{3}\right)' = \dfrac{1}{3} \cdot 3x^2 = x^2$. Следовательно,

$$s = \int_0^1 x^2 dx = \frac{x^3}{3}\Big|_0^1 = \frac{1}{3} - \frac{0}{3} = \frac{1}{3}.$$

Итак, $s = \dfrac{1}{3}$,– треть площади единичного квадрата. Этот результат получил Архимед непосредственным вычислением.

Пример 2. Найдем площадь под одной полуволной синусоиды $y = \sin x$ (рис. 28). Так как одной из первообразных для $f(x) = \sin x$ будет функция $F(x) = -\cos x$, то

$$s = \int_0^{\pi} \sin x dx = -\cos x|_0^{\pi} =$$

$$= -\cos \pi - (-\cos 0) = -(-1) - (-1) = 2.$$

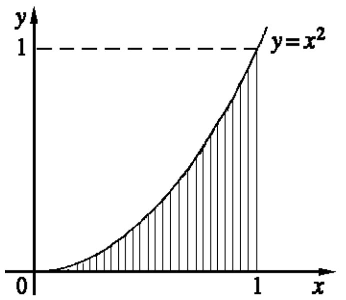


Рис. 27

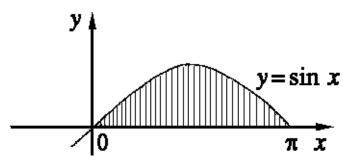


Рис. 28

Формула Ньютона–Лейбница сводит задачу вычисления определенного интеграла от функции $f(x)$ к нахождению первообразной для этой функции. Множество всех первообразных для данной функции $f(x)$ называют неопределенным интегралом от этой функции и обозначают через $\int f(x)dx$ («интеграл от $f(x)$ по dx»). Если $F(x)$ – конкретная первообразная для функции $f(x)$, то согласно теореме 2

$$\int f(x)dx = F(x) + C,$$

где C – произвольная постоянная. Процесс отыскания неопределенного интеграла (т.е. первообразной) также называется интегрированием. Следующая теорема содержит таблицу важнейших неопределенных интегралов.

Теорема 4. Имеют место следующие равенства:

$$
\begin{aligned}
\text{I.} &\quad \int x^{\alpha} dx = \frac{x^{\alpha+1}}{\alpha+1} + C \text{ при } \alpha \neq -1,\\
\text{II.} &\quad \int \frac{1}{x} dx = \ln x + C,\\
\text{III.} &\quad \int e^{x} dx = e^{x} + C,\\
\text{IV.} &\quad \int \sin x dx = -\cos x + C,\\
\text{V.} &\quad \int \cos x dx = \sin x + C,\\
\text{VI.} &\quad \int \frac{1}{\sqrt{1-x^2}} dx = \arcsin x + C.
\end{aligned}
$$

ДОКАЗАТЕЛЬСТВО. Проверка всех этих равенств осуществляется дифференцированием: производная правой части должна совпасть с подинтегральной функцией, стоящей в левой части. Например,

$$\left(\frac{x^{\alpha+1}}{\alpha+1} + C\right)' = \left(\frac{x^{\alpha+1}}{\alpha+1}\right)' + C' =$$

$$= \frac{1}{\alpha+1}(x^{\alpha+1})' = \frac{1}{\alpha+1}(\alpha+1)x^{(\alpha+1)-1} = x^{\alpha},$$

что доказывает формулу I. $\square$

Найдите площадь под гиперболой $y = \dfrac{1}{x}$ на отрезке $[1, e]$, площадь под кубической параболой $y = x^3$ на отрезке $[0, 1]$, площадь под кривой $y = \sqrt{x}$ на отрезке $[0, 1]$ (сравните с результатом примера 1).

В отличие от дифференцирования, представляющего из себя механическую процедуру, освоить интегрирование – не простое дело. Существует много приемов отыскания первообразных для функций тех или иных классов, но это уже чистая техника, и мы не будем касаться ее, оставаясь на «наивном» уровне угадывания первообразной с помощью теоремы 4.

Пусть, например, нужно найти $\displaystyle\int \sin 2x dx$. По теореме 4 первообразной для $\sin x$ является $-\cos x$. Если мы в нашем случае возьмем, руководствуясь этим образцом, функцию $-\cos 2x$, то получаем, что $(-\cos 2x)' = \sin 2x \cdot (2x)' = 2\sin 2x$, т.е. не совсем ту функцию, что стоит под интегралом. Но лишний множитель 2 можно убрать, добавляя $\dfrac{1}{2}$; и мы получаем: $\int \sin 2x dx = -\dfrac{1}{2}\cos 2x + C$.

Пользуясь аналогичными соображениями, найдите неопределенные интегралы

$$\int e^{-3x} dx,\ \int \frac{1}{x+1} dx,\ \int \cos 5x dx,\ \int \sqrt[3]{x+1} dx,\ \int \sqrt{2x+1} dx.$$

Следующий результат расширяет круг наших возможностей.

Теорема 5. Для неопределенного интеграла выполняются следующие свойства линейности:

1) $\int kf(x)dx = k\int f(x)dx$ (постоянная выносится за знак интеграла),

2) $\int [f(x) \pm g(x)]dx = \int f(x)dx \pm \int g(x)dx$ (интеграл от суммы или разности функций равен соответственно сумме или разности интегралов от этих функций).

ДОКАЗАТЕЛЬСТВО. Эти свойства выводятся из аналогичных свойств производных. В самом деле, если $F(x)$ – первообразная для функции $f(x)$, т.е. $F'(x) = f(x)$, то $kF(x)$ будет первообразной для $kf(x)$, ибо $[kF(x)]' = kF'(x) = kf(x)$, откуда следует соотношение 1). Если $F(x)$ и $G(x)$ – первообразные сооответственно для функций $f(x)$ и $g(x)$, то $[F(x) + G(x)]' = F'(x) + G'(x) = f(x) + g(x)$, т.е. $F(x) + G(x)$ будет первообразной для суммы $f(x) + g(x)$, откуда $\int [f(x) + g(x)]dx = F(x) + G(x) + C = \int f(x)dx + \int g(x)dx$. Аналогично для разности. □

Следующие простые задачи дают общее представление о процессе интегрирования. Найдите неопределенные интегралы

$$1)\ \int (x^2 - 3x + \frac{2}{x} - \frac{1}{\sqrt{x}} + 1)dx,$$

$$2)\ \int (2e^{-x} + 3\cos 2x - \sqrt[3]{x})dx,$$

3) $\int (\sqrt[3]{x^2} - \frac{2}{\sqrt{1-x^2}} + \sin 3x - \frac{1}{2x} + 5)dx$,

4) $\int \frac{x^2+x+1}{\sqrt[4]{x}} dx$,

5) $\int \frac{1+\cos^2 x}{\cos^2 x} dx$.

Вычислите площади: 1) под одной полуволной синусоиды $y = 3\sin 2x$; 2) между экспонентами $y = e^x$ и $y = e^{-x}$ на отрезке $[0, \ln 2]$; 3) между кривыми $y = x^2$ и $y = \sqrt{x}$ на отрезке $[0, 1]$; 4) между прямой $y = x$ и параболой $y = x^3$ на отрезке $[-1, 1]$. К каждому примеру сделайте схематический чертеж.

Интегрирование, как уже упоминалось, является основным приемом при решении дифференциальных уравнений. В примерах, рассмотренных в §5 (радиоактивный распад, рост капитала, гармонические колебания), решение уравнения просто предлагалось. Покажем, как оно находится в простейших ситуациях.

Пусть с башни высотой 490 метров упал камень. Через сколько секунд он достигнет земли (сопротивление воздуха не учитывается)? Пусть $y(x)$ обозначает высоту камня над поверхностью земли в момент времени x. Скорость падения камня выражается производной $y'(x)$, а ускорение – второй производной $y''(x)$. Поскольку на камень не действуют никакие силы, кроме силы притяжения земли, ускорение его падения будет постоянным и направленным вниз. Величина ускорения свободного падения известна, она равна $g = 9,8$ м/сек^2 (сколько раз пришлось ронять камни с башен, чтобы найти эту константу!). Таким образом, дифференциальное уравнение движения камня имеет вид $y'' = -g$. Отсюда и нужно найти функцию $y(x)$. Интегрируем: $y' = \int y'' dx = \int -g dx = -g \int dx = -gx + C_1$. Итак, $y' = -gx + C_1$, где C_1 - произвольная постоянная. Интегрируем еще раз: $y = \int y'(x)dx = \int (-gx + C_1)dx = -\frac{gx^2}{2} + C_1 x + C_2$. Таким образом, общим решением будет функция $y(x) = -\frac{gx^2}{2} + C_1 x + C_2$. Найдем конкретные значения постоянных C_1 и C_2 в наших обстоятельствах. В начальный момент времени $x = 0$ камень находился на высоте 490 м, т.е. $y(0) = 490$. Полагая в общем решении $x = 0$, получаем $C_2 = 490$. Так как камень падал сам по себе (а не был брошен чьей-то рукой), его начальная скорость $y'(0)$ была нулевой. Подставляя в выражение скорости $x = 0$, находим $C_1 = 0$. Окончательное уравнение движения камня выразится формулой $y(x) = -\dfrac{gx^2}{2} + 490$. В момент падения на землю $y(x) = 0$, откуда $\dfrac{gx^2}{2} = 490$ и, значит, $x^2 = 100$, т.е. $x = 10$,– камень упадет на землю через 10 секунд.

Все рассмотренные нами дифференциальные уравнения описывают одномерные процессы: радиоактивный распад можно вообразить себе как укорочение вертикально стоящего углеродного стержня, рост капитала – как увеличение столбика золотых монет, гармоническое колебание – это движение точки между ее крайними положениями по оси ординат, камень падает вертикально вниз. Что же изображают в таком случае экспоненты – убывающая в первой и возрастающая во второй задаче, синусоида в третьей и кусок перевернутой параболы в четвертой? Это так называемые мировые линии, т.е. линии, являющиеся траекториями движения соответствующих точек в двухмерном пространстве-времени (y – пространственная координата, x – временна́я). Точка на такой линии (мировая точка) описывает событие, состоящее в том, что в момент времени x материальная точка (конец стержня, верхушка столбика монет, камень) имеет пространственную координату y. Обитателю оси ординат Oy эти линии представляются такой же фантастической абстракцией, какой покажутся нам мировые линии в четырехмерном пространстве-времени теории относительности (см. §III.4).

Начав этот параграф восхвалительной строкой поэта и ученого Валерия Брюсова, закончим его скептическими словами ученого и поэта Огюстена Коши, внесшего величайший вклад в математический анализ и его приложения: «Существуют истины, отличные от истин алгебры, и реальности, не являющиеся чувственными объектами... Не надо воображать себе, что с помощью формул удастся успешно атаковать историю, а моральным одобрением подтверждать истинность теорем алгебры и интегрального исчисления».

ГЛАВА III. **ГЕОМЕТРИЧЕСКИЕ ПРОСТРАНСТВА**

§ 1. Аналитическая геометрия плоскости

Введением системы координат Декарт превратил плоскость в двухмерное арифметическое пространство $\mathbb{R}^2$: точки были поставлены во взаимно однозначное соответствие с парами действительных чисел – координатами. Вторая революционная идея Декарта состояла в следующем. Пусть имеется алгебраическое уравнение с двумя неизвестными $F(x, y) = 0$ (например, $2x - 3y + 1 = 0$, или $x^2 + y^2 - 1 = 0$, или $(x^2 + y^2)^2 - 2(x^2 - y^2) = 0$ и т.п.). Будем считать в нем x абсциссой некоторой точки, а y, который получается из уравнения при подстановке в него данного значения x,– ее ординатой. В результате множество точек плоскости, координаты которых превращают уравнение $F(x, y) = 0$ в тождество, образует некоторый геометрический образ, чаще всего линию (может оказаться, конечно, что уравнение совсем не имеет решений, как, например, $x^2 + y^2 + 1 = 0$, или что получится всего одна или несколько точек, как, например, для $x^2 + y^2 = 0$,– но это неинтересные, вырожденные случаи). С другой стороны, если линия задана каким-нибудь геометрическим описанием (т.е. как геометрическое место точек, удовлетворяющих определенным условиям), то, введя подходящую систему координат, можно попытаться написать уравнение этой линии, т.е. уравнение вида $F(x, y) = 0$, решениями которого были бы в точности координаты точек данной линии. Так возникла аналитическая геометрия – раздел математики, изучающий геометрические объекты средствами алгебры и математического анализа, примененными к уравнениям этих объектов, связывающим координаты образующих их точек.

Степенью алгебраического одночлена с двумя неизвестными $ax^m y^n$ называется сумма $m + n$ (здесь каждый из показателей

m, n – натуральное число или ноль), а степенью многочлена $F(x, y)$ – наибольшая из степеней составляющих его одночленов. Степень уравнения $F(x, y) = 0$ – это степень многочлена $F(x, y)$. Это же число называют порядком линии, задаваемой уравнением $F(x, y) = 0$.

Общий вид уравнения первой степени с двумя неизвестными $Ax+By+C = 0$, где $A,\ B,\ C$ – некоторые числа, причем A и B не могут одновременно равняться нулю. Общий вид уравнения второй степени с двумя неизвестными $Ax^2 + Bxy + Cy^2 + Dx + +Ey + F = 0$ – здесь $A,\ B$ и C не равны одновременно нулю.

К числу первоначальных, неопределяемых геометрических образов относится понятие прямой: все одинаково представляют себе, что это такое. Следующая теорема Декарта позволяет дать чисто алгебраическое толкование этому простейшему объекту.

<u>Теорема 1</u>. Прямые, и только они, являются линиями первого порядка. □

Таким образом, если у нас имеется уравнение первой степени с двумя переменными, то геометрическое место точек, координаты которых удовлетворяют этому уравнению, является прямой. И обратно, всякая прямая в произвольно выбранной декартовой системе координат может быть представлена уравнением первой степени.

Обычно уравнение прямой записывается в виде $y = kx + l$, где k – это тангенс угла φ, образуемого этой прямой с осью абсцисс Ox (число k называется угловым коэффициентом прямой), а l – отрезок, который прямая отсекает на оси ординат Oy (рис. 29). При этом угол φ изменяется от 0 до π; тангенс острого угла положителен, для тупого угла он имеет отрицательное значение. Число l также имеет знак: оно будет положительным, если точка пересечения прямой с осью ординат Oy лежит выше оси абсцисс Ox, и отрицательным – в противном случае. Чтобы построить прямую $y = kx + l$, нужно сначала провести прямую через точки $(0, 0)$ (начало координат) и $(1, k)$. Если $l = 0$, прямая построена, а если $l \neq 0$, первоначальную прямую нужно еще сдвинуть параллельно себе на l единиц вверх (при $l > 0$) или вниз (при $l < 0$).

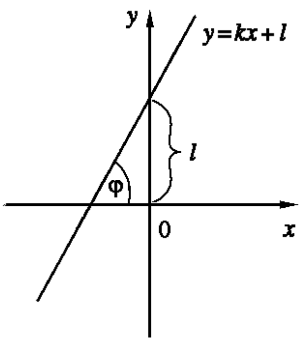


Рис. 29

Постройте прямые $y = 2x + 3,\ y = 3x - 1,\ y = -x + 1,$ $y = -2x - 3$.

Особый случай представляют не укладывающиеся в эту схему вертикальные прямые. Они задаются уравнениями вида $x = a$, где a – постоянное число. Такая прямая параллельна оси Oy и проходит на расстоянии $|a|$ от нее – справа при $a > 0$ и слева при $a < 0$. Прямая $x = 0$ совпадает с осью ординат.

Полное описание линий второго порядка также было получено Декартом. Ими, за исключением некоторых вырожденных случаев, являются так называемые конические сечения, которые мы сейчас определим.

Окружностью называется геометрическое место точек, равноудаленных от некоторой точки (центра). Расстояние от любой точки окружности до ее центра называется радиусом.

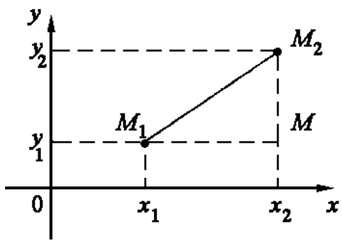


Рис. 30

Для вывода уравнения окружности воспользуемся формулой, выражающей расстояние между двумя произвольными точками плоскости: M_1 с координатами (x_1, y_1) и M_2 с координатами (x_2, y_2) (рис. 30).

По теореме Пифагора $M_1M^2 + M_2M^2 = M_1M_2^2$. Так как точка M имеет координаты (x_2, y_1), то $M_1M = x_2 - x_1,\ M_2M = y_2 - y_1$, откуда $M_1M_2 = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$,– расстояние между двумя точками плоскости равно квадратному корню из суммы квадратов разностей их одноименных координат.

Пусть данная окружность имеет центр в точке $M_0(x_0, y_0)$ и радиус r и пусть $M(x, y)$ – произвольная точка этой окружности. По определению, $MM_0 = r$, т.е. $\sqrt{(x - x_0)^2 + (y - y_0)^2} = r$, или $(x - x_0)^2 + (y - y_0)^2 = r^2$. Это и есть уравнение окружности с центром в точке (x_0, y_0) и радиусом r. Постройте окружности $x^2 + y^2 = 1,\ (x - 2)^2 + (y + 1)^2 = 4$. Напишите уравнение окружности с центром в точке $(-1, 2)$ и радиусом 3.

Эллипсом называется геометрическое место точек, сумма расстояний каждой из которых до двух данных точек F_1 и F_2, называемых фокусами, есть постоянная величина, равная $2a > F_1F_2$. Взяв нить длиной $2a$ и закрепив ее концы в точках F_1 и F_2, натягиваем ее острием карандаша и начинаем двигать карандаш по бумаге. Завершив полный оборот, получим чертеж эллипса (рис. 31).

Если выбрать систему координат так, чтобы ось Ox проходила через отрезок F_1F_2, а ось Oy – через его середину (центр эллипса), то уравнение эллипса можно записать в виде

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1,$$

где $2b$ – это расстояние между точками пересечения эллипса с осью Oy. Окружность можно считать специальным случаем эллипса – когда его фокусы совпадают.

Два других типа конических сечений – парабола и гипербола – уже встречались в §II.2, и теперь мы объясним общее название всех этих кривых. Пусть имеется прямой круговой конус, т.е. поверхность, образуемая при движении прямой линии l (образующая), проходящей через данную точку S (вершина) и пересекающей данную окружность C, в центр которой проектируется точка S (рис. 32).

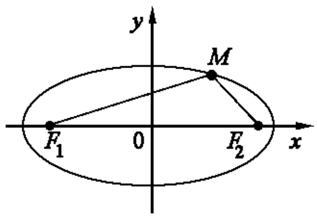


Рис. 31

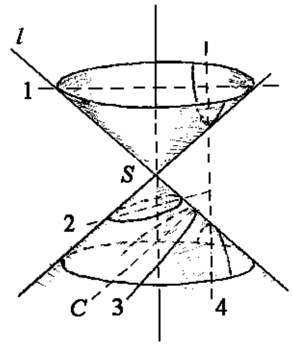


Рис. 32

Рассмотрим линии, которые получаются при пересечении конуса плоскостью. Если секущая плоскость перпендикулярна оси конуса и не проходит через его вершину, в сечении получается окружность (1). Когда плоскость не перпендикулярна оси конуса и не параллельна ни одной из его образующих, то сечением конуса будет эллипс (2). Плоскость, параллельная только одной из образующих, пересечет конус по параболе (3). Наконец, в том случае, когда секущая плоскость окажется параллельной двум образующим конуса, сечением оказывается гипербола (4).

Можно показать, что при надлежащем выборе системы координат уравнение любого из конических сечений может быть приведено к виду $\lambda x^2 - y^2 + 2px = 0$, где λ и p – некоторые постоянные. Если $p \neq 0$, то при $\lambda < 0$ получается эллипс, при $\lambda = 0$ – парабола и при $\lambda > 0$ – гипербола. Кроме эллипсов (с частным случаем окружностей), парабол и гипербол, в сечении конуса может получиться точка (когда секущая плоскость, например, перпендикулярна оси конуса и проходит через его вершину), пара пересекающихся прямых (укажите соответствующее сечение) и т.п., но эти случаи считаются вырожденными.

Конические сечения подробно изучались в древнегреческой математике. Систематический трактат на эту тему был написан в начале II века до н.э. Аполлонием. Ему принадлежат и названия: слово «эллипс» происходит от греческого «недостаток» (свойство, которое в вышеуказанном общем уравнении конических сечений выражается неравенством $\lambda < 0$), «парабола» – от «приравнивание» ($\lambda = 0$), «гипербола» – от «избыток» ($\lambda > 0$). Вплоть до конца XVI века конические сечения (кроме окружности, конечно) оставались лишь образами математической фантазии, не находившими никаких реальных приложений. С появлением системы Коперника, поставившего в центр мира не Землю, а Солнце, перемещения планет по небесной сфере получили более простое объяснение – их орбиты стали считать круговыми,– но все равно требовались неприятные поправки. В 1609 году Иоганн Кеплер (1571–1630), императорский математик в Праге, в своем труде «Новая астрономия» сформулировал, в частности, первый закон движения планет, сильно упростивший расчеты: их орбитами являются эллипсы, в одном из фокусов которых находится Солнце. (Кеплеру принадлежит следующее доказательство обитаемости Юпитера: «Какой смысл иметь четыре луны, вращающиеся вокруг планеты, если на ней нет никого, кто бы наблюдал их?») После открытия Ньютоном закона всемирного тяготения нашлось астрономическое толкование и для других конических сечений. Кометы, оказывается, могут двигаться по очень удлиненным эллипсам (например, комета Галлея, навещающая нас один раз в 75 лет), по параболам и по гиперболам. Астрономы получили для своих расчетов готовый математический аппарат, разработанный более чем за полторы тысячи лет до потребности в нем. Пример этот математики, работающие в чисто теоретических областях, часто приводят в качестве ответа на неквалифицированные обвинения в оторванности их исследований от практических нужд.

Классификацию кривых третьего порядка осуществил Ньютон. Конечно, она не имеет такого изящного вида, как декартово описание линий первого и второго порядков: слишком велико число возникающих возможностей. Еще более усложняется дело в случае алгебраических уравнений $F(x,y) = 0$ степеней выше третьей. Укажем лишь некоторые замечательные линии, сыгравшие заметную роль в истории математики и ее приложений. В их загадочных названиях и необычном виде есть нечто романтическое, соединяющее математическую эстетику с общим представлением о красоте.

1. Локон Аньези. Эта кривая третьего порядка задается уравнением $(x^2 + a^2)y - a^3 = 0$, или $y = \dfrac{a^3}{x^2 + a^2}$, где a – диаметр определяющей ее окружности (не будем вдаваться в детали) (рис. 33).

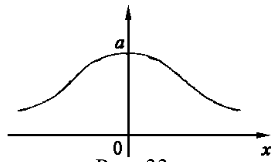


Рис. 33

Линия названа по имени Марии Гаэтаны Аньези (1718–1799), автора двухтомного руководства «Основания анализа» (Милан, 1748), которая изучала эту кривую, известную, впрочем, еще Ферма. Нарисуйте по точкам локоны Аньези, получающиеся при значениях параметра $a = 1$ и $a = 2$.

2. Строфоида (рис. 34) соответствует уравнению $x^2(x + a) + y^2(x - a) = 0$. Эту кривую открыл и исследовал итальянский физик и математик Эванжелиста Торичелли (1608–1647) в связи с изысканиями в оптике. До середины XIX века она называлась «крылом Торичелли» («строфоида» – от греческого «поворот»).

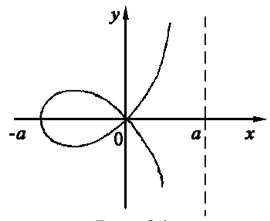


Рис. 34

3. Кардиоида описывается точкой, которая лежит на окружности, катящейся по неподвижной окружности с таким же радиусом (рис. 35). Это кривая четвертого порядка, ее уравнение $(x^2 + y^2 + 2ax)^2 = 4a^2(x^2 + y^2)$, где a – радиус упомянутых окружностей. Площадь, ограниченная кардиоидой, равна $6\pi a^2$, а длина ее $16a$. Название – от греческого «сердце».

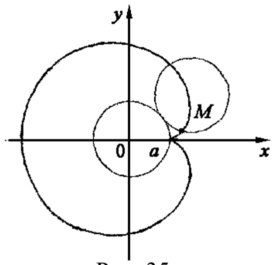


Рис. 35

4. Астроида – это траектория точки, лежащей на окружности, которая катится по внутренней части неподвижной окружности вчетверо большего радиуса (рис. 36). Астроида – кривая шестого порядка, ее уравнение имеет красивый вид в записи с дробными степенями: $x^{\frac{2}{3}} + y^{\frac{2}{3}} = a^{\frac{2}{3}}$, где a – радиус неподвижной окружности. Площадь, ограниченная астроидой, равна $\frac{3}{8}\pi a^2$, а ее длина $6a$. Название – от греческого «звезда».

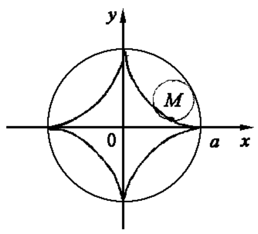


Рис. 36

5. Лемниската Бернулли представляет собой геометрическое место точек, произведение расстояний каждой из которых до двух данных точек F_1 и F_2, называемых фокусами, есть постоянная величина, равная a^2, где a – половина расстояния между фокусами (рис. 37). Уравнение лемнискаты $(x^2+y^2)^2 - 2a^2(x^2-y^2) = 0$ (определите его степень, т.е. порядок кривой) впервые получил Якоб Бернулли в 1694 году. Площадь, ограниченная каждой петлей, равна a^2. Лемниската используется в качестве переходной кривой на крутых поворотах железнодорожных линий. Название происходит от греческого слова «бант».

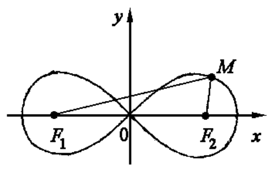


Рис. 37

Плоские кривые, которые могут быть заданы уравнением $F(x, y) = 0$, где левая часть представляет собой многочлен от двух переменных, называются алгебраическими, а в противном случае – трансцендентными. Приведем примеры известнейших трансцендентных кривых (самые известные – это графики основных элементарных функций, таких, как показательные, логарифмические или тригонометрические).

6. Спираль Архимеда – это траектория точки, равномерно движущейся по прямой, которая в свою очередь равномерно вращается вокруг одной из своих точек (полюс). Считая, что движущаяся точка в начальный момент находилась в полюсе и что вращение прямой происходит против часовой стрелки, получаем линию, изображенную на рис. 38 (здесь расстояние от полюса равно углу поворота, измеренному в радианах: $r = \varphi$).

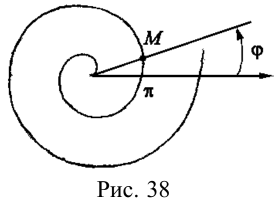


Рис. 38

Площадь первого витка этой спирали нашел Архимед – методом интегральных сумм, что и дает повод считать его провозвестником интегрального исчисления. Спираль Архимеда находит многочисленные приложения в теории механизмов.

7. Логарифмическая спираль $r = a^{\varphi}$, $a > 0$, $a \neq 1$, при $a > 1$ закручивается по ходу часовой стрелки, делая бесконечное число оборотов и асимптотически приближаясь к полюсу (рис. 39). Рисунок логарифмической спирали можно увидеть на некоторых раковинах и во многих технических устройствах. Эта кривая была одновременно открыта Декартом и Торичелли. Исследовавший ее свойства Якоб Бернулли был столь заворожен красотой и совершенством spira mirabilis («дивной спирали»), что завещал изобразить ее на своем надгробии (в соборе в Мюнстере, где похоронен великий математик, на памятной доске в самом деле высечена спираль, но, увы, не логарифмическая).

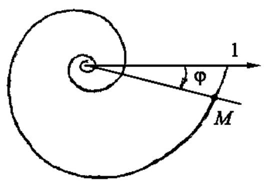


Рис. 39

8. Цепной линией называется кривая, форму которой принимает тяжелая нерастяжимая нить с закрепленными концами (рис. 40).

Уравнение цепной линии имеет вид $y = \frac{a}{2}(e^{\frac{x}{a}} + e^{-\frac{x}{a}})$, где a – высота наинизшей точки нити над осью Ox (она зависит от свойств материала и расстояния между точками крепления). В технике цепная линия используется при расчетах, связанных с устройством линий электропередач и т.п. Форма провисания нити была окончательно выяснена усилиями Лейбница, Христиана Гюйгенса (1629–1695) и Иоганна Бернулли (до того господствовало убеждение Галилея, что цепная линия – это обыкновенная парабола). Любопытно отметить, что в 1838 году на конкурсе студенческих научных работ по физико-математическому отделению философского факультета Московского университета золотую медаль за исследование о свойствах цепной линии получил А.В.Сухово-Кобылин, впоследствии известный драматург. При вращении цепной линии вокруг оси Ox получается поверхность, называемая катеноидом – о ней см. в §3.

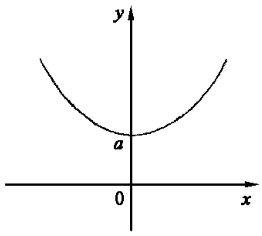


Рис. 40

9. Трактриса (линия влечения) – это след материальной точки, прикрепленной к нерастяжимой нити, другой конец которой движется по некоторой прямой (рис. 41). Уравнение трактрисы нашли Лейбниц и Гюйгенс:

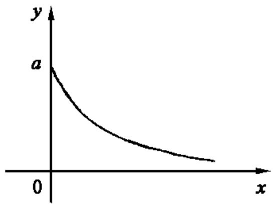


Рис. 41

$$x = a \ln \frac{a + \sqrt{a^2 - y^2}}{y} - \sqrt{a^2 - y^2},$$

где a – длина нити. Точка, движущаяся по линии влечения, асимптотически приближается к оси Ox. Площадь под трактрисой конечна и равна $\dfrac{\pi a^2}{4}$. При вращении трактрисы вокруг оси Ox получается поверхность, похожая на рупор,– знаменитая псевдосфера, о которой мы будем говорить в § 3.

10. Циклоидой называется кривая, которую описывает точка окружности, катящейся по неподвижной прямой (рис. 42).

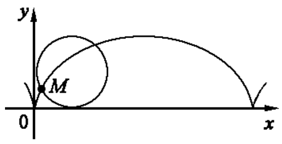


Рис. 42

Площадь, ограниченная одной аркой циклоиды, равна $3\pi a^2$, где a – радиус упомянутой окружности. Площадь эту вычислил Галилей, который дал кривой и ее название.

Циклоида является решением двух знаменитых задач, положивших начало новым разделам математики. Первая из этих задач состояла в следующем: для двух данных точек, не лежащих на одной вертикали и на одном уровне, найти такую соединяющую их кривую, чтобы материальная точка, двигаясь по ней под действием силы тяжести, за наименьшее возможное время скатилась бы из верхней точки в нижнюю. Искомая кривая была названа брахистохроной (линия кратчайшего времени). В 1696 году Иоганн Бернулли доказал, что брахистохроной является дуга (перевернутой) циклоиды. В другой задаче предлагалось найти в вертикальной плоскости кривую, двигаясь по которой под действием силы тяжести, материальная точка спускалась бы до заданного уровня за время, не зависящее от своего начального положения. Такая кривая была названа изохроной (линия постоянного времени). В 1673 году Гюйгенс установил, что изохрона – это циклоида (тоже, конечно, перевернутая), и создал на этой основе циклоидальный маятник, период колебания которого не зависит от амплитуды. (Перечитайте еще раз задачи о брахистохроне и изохроне и изобразите на чертеже описываемые в них ситуации.)

В наш краткий список не попали столь достойные объекты, как декартов лист, улитка Паскаля, трезубец Ньютона, излюбленные фантастами нового времени клотоида, конхоида и кохлеоида, розы с различным числом лепестков и многие другие линии, при исследовании которых на конкретном и наглядном материале оттачивались методы алгебры и анализа, решались задачи, возникавшие в различных сферах человеческой деятельности. И конечно, возникали новые математические идеи, новые математические миры.

§ 2. Геометрии и группы. Проективная геометрия

Тремя самыми известными теоремами античной математики являются доказанное Эвклидом предложение о бесконечности множества простых чисел (теорема из §I.1), утверждение о несоизмеримости диагонали квадрата с его стороной (теорема 2 §I.3) и теорема Пифагора. Первая из них относится к теории чисел, вторая – в равной мере к алгебре и анализу, третья по своим следствиям и приложениям представляет собой основополагающий результат геометрии.

Теорема 1. В прямоугольном треугольнике квадрат, построенный на гипотенузе, равновелик взятым вместе квадратам, построенным на катетах (коротко: квадрат гипотенузы равен сумме квадратов катетов).

ДОКАЗАТЕЛЬСТВО. Теорема Пифагора к настоящему времени имеет около 400 различных по своей идее доказательств. Мы обратимся к книге «Лилавати» (что значит «прекрасная») выдающегося индийского математика Бхаскары (1114–1185), где он в поэтической форме рассказывает своей дочери о способах решения тех или иных задач. По поводу теоремы Пифагора приводится только один чертеж (рис. 43, слева), под которым написано: «Смотри!». Выполняя это указание, мы видим квадрат, разрезанный на четыре равных прямоугольных треугольника и маленький квадрат. Большой квадрат построен на гипотенузе каждого из этих треугольников, сторона маленького квадрата равна разности между длинами большого и малого катетов треугольника (обозначим их соответственно через y и x). Из этих пяти частей квадрата сложим среднюю фигуру на рис. 43 и убедимся, что она может быть составлена также из двух квадратов: один со стороной x, а другой со стороной y (правая фигура). Таким образом, площади левой и правой фигур оказываются равными – теорема доказана. (Лев Толстой часто и с увлечением демонстрировал доказательство Бхаскары своим посетителям.) □

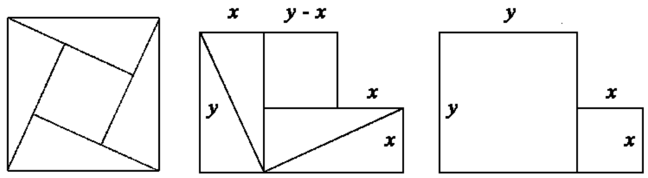


Рис. 43

Из теоремы Пифагора выводится формула для длины отрезка: если его концами являются точки $M_1(x_1, y_1)$ и $M_2(x_2, y_2)$, то $M_1M_2 = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$ (см. §1). Отрезок и окружность – основные геометрические образы эвклидовой геометрии плоскости. Из отрезков составляются многоугольники, окружность – неизбежный элемент в задачах на построение. Линейка и циркуль – инструментальное воплощение отрезка и окружности соответственно.

Решающим шагом в переходе от конкретных задач, связанных с данным участком земли (измерить площадь, разделить на некоторые части и т.д.), к соответствующим проблемам общего характера была идея о равенстве фигур: считать две фигуры равными, если они отличаются только своим расположением. Треугольник, сколоченный из трех реек, остается самим собой, куда бы мы его ни положили, и, значит, его разные отпечатки должны рассматриваться как одна и та же фигура. Дадим точную формулировку.

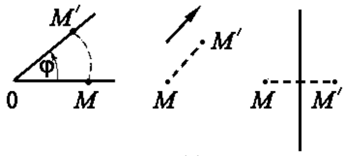


Рис. 44

Геометрическим преобразованием плоскости называется ее взаимно однозначное отображение на себя (т.е. всякая ее биекция). К числу геометрических преобразований относятся упомянутые в §II.1 вращения, параллельные переносы и осевые симметрии. Рис. 44 схематически изображает эти важнейшие геометрические преобразования (точка M переходит в точку M').

Геометрическое преобразование плоскости называется движением, если оно сохраняет расстояние между точками. Нетрудно заметить, что вращения, параллельные переносы и осевые симметрии являются движениями. Можно показать, что всякое движение плоскости представляет собой композицию вращений, параллельных переносов и осевых симметрий.

Фигура Φ' называется равной фигуре Φ, если существует движение f, которое переводит фигуру Φ в фигуру Φ', т.е. такое, что $f(\Phi) = \Phi'$. Это определение вполне соответствует первоначальной идее равенства фигур: две фигуры равны, если они отличаются только расположением на плоскости.

Теперь мы видим, что эвклидова геометрия плоскости (т.е. та геометрия, которую мы изучаем в школе) рассматривает свойства фигур, не меняющиеся при движениях.

Теорема 2. Совокупность всех движений плоскости образует группу.

ДОКАЗАТЕЛЬСТВО. Пусть f и g – движения, т.е. геометрические преобразования плоскости, сохраняющие расстояние между точками. Композиция $f \circ g$ – это геометрическое преобразование, состоящее в последовательном выполнении сначала преобразования f, а затем преобразования g. Ясно, что $f \circ g$ сохраняет расстояния. Тождественное преобразование Δ, переводящее каждую точку плоскости в себя, несомненно, сохраняет расстояния и, значит, является движением. Наконец, если f – движение, то обратное ему геометрическое преобразование f^{-1} возвращает точки, смещенные преобразованием f, в их исходное положение – понятно, что расстояния между точками f^{-1} не изменяет. Таким образом, операция композиции движений ассоциативна (как суперпозиция функций вообще), относительно нее имеется нейтральный элемент (тождественное движение Δ) и каждое движение имеет обратное для него – все групповые свойства выполнены. □

В 1870 году в Париже вышел обширный «Трактат о перестановках» Камилла Жордана (1838–1922), сделавший, наконец, идеи Галуа достоянием широких математических кругов. С этого времени понятие группы начинает играть важнейшую роль в анализе и геометрии, в механике и физике, преобразует алгебру. Уже через два года Феликс Клейн (1849–1925), вступая в должность профессора в Эрлангене, представил доклад, в котором раскрыл внутреннюю связь между различными геометриями и группами преобразований. Этот доклад, названный впоследствии «Эрлангенской программой», надолго определил направление развития геометрических теорий.

Основная идея Клейна заключалась в следующем. Эвклидова геометрия плоскости рассматривает свойства фигур, не изменяющиеся при движениях. Вместо движений возьмем совокупность геометрических преобразований какого-либо другого типа. «Равенство» фигур определим, как и в случае движений: фигура Φ' «равна» фигуре Φ, если существует преобразование f данного типа, переводящее Φ в Φ', т.е. такое, что $f(\Phi) = \Phi'$ (в конкретных геометриях вместо слова «равенство», как правило, употребляют какой-нибудь специфический термин. Даже в эвклидовой геометрии вместо «равные» о фигурах часто говорят «конгруэнтные»).

Понятие равенства фигур, как бы мы его ни определяли, должно обладать следующими естественными свойствами: 1) всякая фигура равна себе (рефлексивность); 2) если Φ' равна Φ, то и Φ равна Φ' (симметричность); 3) если Φ равна Φ', а Φ' равна Φ'', то Φ равна Φ'' (транзитивность равенства). Требование рефлексивности означает, что к числу преобразований данного типа должно относиться тождественное преобразование Δ, переводящее каждую фигуру в себя. Условие симметричности равенства реализуемо лишь тогда, когда для каждого преобразования f данного типа обратное ему преобразование f^{-1} принадлежит тому же типу. Наконец, для выполнения свойства транзитивности равенства необходимо, чтобы композиция преобразований рассматриваемого типа снова была преобразованием этого типа. Следовательно, определенное выше «равенство» фигур будет обладать естественными свойствами равенства лишь в том случае, когда геометрические преобразования данного типа образуют группу,– как это имеет место в случае движений.

Пусть теперь дана некоторая группа геометрических преобразований. Определив «равенство» фигур описанным выше образом, будем рассматривать те их свойства, которые одинаковы для «равных» фигур. Возникающая теория и называется геометрией данной группы преобразований. Например, эвклидова геометрия плоскости – это геометрия группы движений.

Другой пример геометрии дает группа подобий. Подобием называется геометрическое преобразование f, при котором для любых двух точек A, B и их образов $A' = f(A)$, $B' = f(B)$ длины соответствующих отрезков связаны соотношением $AB = k \cdot A'B'$, где число $k \neq 0$ называется коэффициентом подобия. Например, каждое движение плоскости является подобием (с $k = 1$). Нетрудно убедиться в том, что подобия образуют группу, т.е. что тождественное преобразование Δ является подобием, обратное для подобия f преобразование f^{-1} сам будет подобием (если f

имеет коэффициент k, то какой коэффициент имеет f^{-1}?), что композиция подобий f и g – подобие (установите связь между коэффициентами для f, g, и $f \circ g$). В соответствии с общим принципом назовем фигуру Φ' подобной фигуре Φ, если существует подобие f, переводящее Φ' в Φ, т.е. такое, что $f(\Phi) = \Phi'$. Свойства фигур, не меняющиеся при преобразованиях подобия, образуют геометрию группы подобий. Всякое подобие сохраняет величину углов и порядок точек на прямой; точки, не лежащие на одной прямой, переводит в точки, не лежащие на одной прямой; окружности преобразует в окружности, отрезки в отрезки (но с другим радиусом и другой длиной). Школьные теоремы о подобных треугольниках относятся к геометрии группы подобий.

(Что бы вы могли сказать о геометрии группы вращений? Определите «равенство» фигур, укажите примеры геометрических свойств, сохраняющихся и изменяющихся при вращениях.)

Гораздо больший интерес представляет проективная геометрия, которая изучает свойства фигур, не меняющиеся при преобразованиях проективной группы. Если эвклидова геометрия возникла из практики землемеров, то проективную геометрию породило воображение художников.

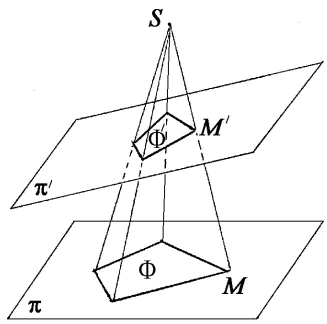


Рис. 45

Пусть в пространстве имеются две плоскости π и π', все равно, параллельные или пересекающиеся, и пусть точка S (центр) не принадлежит ни одной из них. Возьмем в плоскости π некоторую точку M и соединим ее с центром S. Прямая SM пересечет плоскость π' в точке M', которая называется проекцией точки M на плоскость π'. Сопоставляя каждой точке плоскости π ее проекцию на плоскость π', получаем отображение плоскости π на плоскость π', называемое перспективой с центром S. Предположим, что в плоскости π задана фигура Φ. Проектируя каждую ее точку на плоскость π', получим в этой плоскости фигуру Φ' – перспективную проекцию фигуры Φ, ее образ в данной перспективе (рис. 45). Если плоскости π и π' параллельны, то фигура Φ' будет подобна фигуре Φ. Когда художник изображает нечто, он тем самым проектирует оригинал на плоскость картины, глядя из центра перспективы.

Другой разновидностью перспективы является параллельная проекция – когда через все точки фигуры Φ проводятся параллельные между собой прямые, а затем берутся точки их пересечения с плоскостью π'. Параллельную проекцию можно считать частным случаем центральной, если центр перспективы поместить в бесконечно удаленную точку S_∞. Поэтому в дальнейшем, говоря о перспективах и проекциях, будем иметь в виду и эту возможность.

Учение о перспективе, возникшее в эпоху Возрождения, коренным образом реформировало технику живописи. Выдающуюся роль в этом процессе сыграли «Трактат о перспективе» Леонардо да Винчи (1452–1519) и «Руководство к измерению» Альбрехта Дюрера (1471–1528). Сочинение Дюрера, так же как и его «Четыре книги о пропорциях,» представляют собой типичные математические труды того времени.

Фигура Φ', перспективная проекция фигуры Φ на плоскость π', в свою очередь может быть спроектирована на плоскость π из какого-либо другого центра перспективы S', в результате чего на плоскости π появится фигура Φ'' (если S' совпадает с S, то, конечно, получится исходная фигура Φ). Фигуры Φ и Φ'' в этом случае называются перспективными. Таким образом, две фигуры плоскости перспективны, если они являются перспективными проекциями на эту плоскость одной и той же фигуры. Можно ли перспективные фигуры объявить «равными» и, основываясь на таком понимании равенства фигур, построить геометрию? Нет, потому что перспективность не транзитивна: если фигура Φ_1 перспективна фигуре Φ_2 (т.е. они получаются проекциями с центром S' на плоскость π некоторой фигуры Ψ', лежащей в какой-то плоскости π'), а фигура Φ_2 перспективна фигуре Φ_3 (т.е. они получаются проекциями с центром S'' на плоскость π некоторой фигуры Ψ'', лежащей в какой-то плоскости π''), то Φ_1 и Φ_3 могут оказаться не перспективными. Чтобы в рамках рассматриваемой схемы получить полноценное понятие «равенства» для фигур, придется ввести более сложное отношение между ними. Именно, фигуру Ψ назовем проективной фигуре Φ, лежащей в той же плоскости, если существует такая последовательность фигур $\Phi,\ \Phi_1,\ \Phi_2, \ldots,\ \Phi_n,\ \Psi$, начинающаяся с Φ и кончающаяся Ψ, в которой любые две соседние фигуры перспективны. Понятие проективности уже обладает всеми тремя основными свойствами обычного равенства (проверьте) и, следовательно, может быть положено в основу некоторой геометрии. Это и будет проективная геометрия плоскости.

Но плоскость, на которой реализуется проективная геометрия, необычна. Желание вместить параллельную проекцию в схему перспективы приводит к тому, что каждому семейству параллельных прямых придется сопоставить бесконечно удаленную точку плоскости, а все бесконечно удаленные точки считать лежащими на одной бесконечно удаленной прямой. Эвклидова плоскость, пополненная этими новыми, «идеальными» объектами, называется проективной плоскостью. В проективной плоскости бесконечно удаленные элементы равноправны со всеми остальными. В ней нет параллельных прямых, ибо любые две прямые пересекаются – в конечной или бесконечно удаленной точке.

Проективным преобразованием проективной плоскости называется такое ее геометрическое преобразование, при котором точки, лежащие на одной прямой, переходят в точки, также лежащие на одной прямой. Например, каждое подобие (а значит, и каждое движение) является проективным преобразованием. Проективные преобразования образуют группу – это проективная группа. Действуя по общей схеме эрлангенской программы, определим «равенство» фигуры Φ' фигуре Φ как равносильное равенству $f(\Phi) = \Phi'$ при некотором проективном преобразовании f. Оказывается, это условие выполняется в точности тогда, когда фигуры Φ и Φ' проективны в смысле учения о перспективе. Теперь, как говорится, « все линии сошлись», проективная геометрия определена: она изучает свойства фигур проективной плоскости, инвариантные (сохраняющиеся) при проективных преобразованиях.

Какими же особенностями обладает проективная геометрия по сравнению с эвклидовой? С помощью подходящего проективного преобразования любой четырехугольник может быть переведен в любой другой четырехугольник, т.е. все четырехугольники проективны (как в эвклидовой геометрии равны, например, все квадраты с данной длиной стороны). Если заданы два четырехугольника, то существует единственное проективное преобразование, переводящее один в другой.

Пусть в некоторой плоскости π выделена окружность и пусть центр перспективы S лежит на прямой, проходящей через центр окружности и перпендикулярной плоскости π. Проведя через S и каждую точку окружности прямую, получим прямой круговой конус. При пересечении этого конуса различными плоскостями возникают эллипсы, параболы и гиперболы в зависимости от расположения секущей плоскости относительно конуса (см. §1). Все эти кривые, как видим, будут перспективными проекциями исходной окружности. Отсюда следует, что все конические сечения проективны между собой, т.е. в проективной геометрии они воспринимаются как одна и та же кривая.

В 16-летнем возрасте Паскаль получил следующий важный результат проективной геометрии.

Теорема 3. Если шестиугольник вписан в коническое сечение, то точки пересечения его противоположных сторон лежат на одной прямой. □

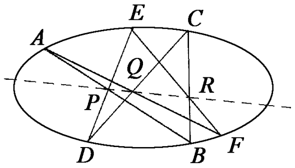


Рис. 46

Поскольку все конические сечения проективны, для иллюстрации выберем эллипс (рис. 46). Заметим, что шестиугольник может иметь самопересекающийся контур. Противоположными являются его первая и четвертая, вторая и пятая, третья и шестая стороны (номера – в порядке прохождения по контуру). Сделайте еще несколько рисунков на тему теоремы Паскаля.

Проективную геометрию как самостоятельную ветвь математики создал лейтенант инженерных войск наполеоновской армии Жан-Виктор Понселе (1788–1867), ставший впоследствии бригадным генералом и президентом Парижской академии наук. Свой «Трактат о проективных свойствах фигур», изданный в 1822 году, он написал в 1813 году, находясь в плену в Саратове. Конечно, у него были предшественники, среди которых должны быть упомянуты Паскаль, Ньютон и, в первую очередь, Жирар Дезарг (1593–1662), тоже французский военный инженер (в 1628 году он вместе с Д’Артаньяном и его приятелями-мушкетерами активно участвовал в осаде крепости Ларошель). Мы закончим рассказ о проективной геометрии формулировкой знаменитой теоремы Дезарга – она является одним из самых ярких результатов этой теории.

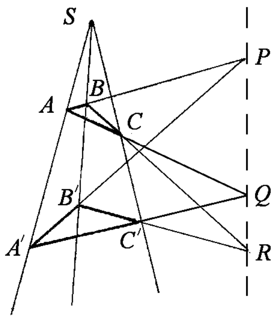


Рис. 47

Теорема 4. Пусть заданы два треугольника ABC и $A'B'C'$ такие, что прямые AA', BB', CC' пересекаются в точке S. Тогда пары соответствующих сторон AB и $A'B'$, AC и $A'C'$, BC и $B'C'$ пересекаются в трех точках P, Q, R, лежащих на одной прямой (рис. 47). □

Совокупность участвующих в теореме 10 точек и 10 прямых называется конфигурацией Дезарга. Это один из основных образов проективной геометрии. Придумайте несколько существенно различных конфигураций Дезарга (допуская и бесконечно удаленные элементы плоскости).

§ 3. Трехмерное эвклидово пространство. Векторы

Ни создатель аналитической геометрии Декарт, ни Ньютон, первым освоивший ее (он жаловался на трудности в понимании декартовой «Геометрии»), не выходили за пределы плоскости. Ввести третью координату они не догадались. Вообще, пространственные объекты – тела и поверхности – систематически не исследовались. Древние греки знали шар и цилиндр (шар, вписанный в цилиндр, видел Цицерон на могиле Архимеда); конус рассматривался лишь в связи с его сечениями плоскостью. Основное внимание было обращено на правильные многогранники – платоновы тела. Многогранник называется правильным, если все углы при его вершинах равны и все грани являются одинаковыми правильными (т.е. с равными сторонами) многоугольниками. Платоновых тел – пять: тетраэдр (4 вершины, 6 ребер, 4 грани), гексаэдр (куб: 8, 12, 6), октаэдр (6, 12, 8), додекаэдр (20, 30, 12) и икосаэдр (12, 30, 20) (рис. 48). Эвклид показал, что приведенный список не пополняем.

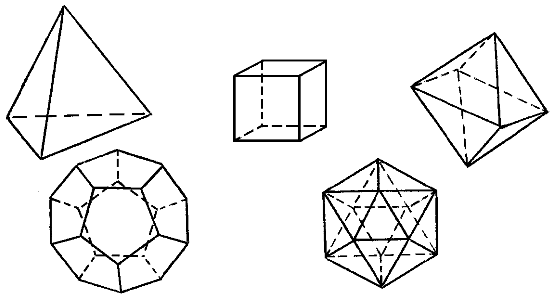


Рис. 48

Согласно Платону, атомы четырех стихий – огня, земли, воздуха и воды – имеют форму соответственно тетраэдра, куба, октаэдра и икосаэдра, а весь мир находится внутри додекаэдра. Несмотря на постоянный пристальный интерес к этим телам, считавшимся символом эстетического совершенства, до Эйлера никто не заметил, что между числом их вершин (В), ребер (Р) и граней (Г) существует удивительная закономерность: В-Р+Г=2, справедливая, впрочем, для гораздо более широкого класса объектов (см. §VI.2).

Если не считать отдельные попытки предшественников, систематически геометрию трехмерного пространства начал изучать Эйлер: ко второму тому «Введения в анализ» (1748) он присоединил обширное «Приложение о поверхностях». Леонард Эйлер (1707–1783), имя которого встречается практически во всех разделах математики, родился в Базеле (Швейцария). Его отец, пастор, был учеником Якоба Бернулли, а сам будущий великий математик занимался у Иоганна Бернулли. В 1726 году Эйлер прибыл в Петербург и работал в Академии наук до конца жизни (с 1741 по 1766 год жил в Берлине). В возрасте 60 лет полностью потерял зрение (после этого им было продиктовано свыше 400 статей и книг). Всего Эйлеру принадлежит более 850 научных трудов, в числе которых целый ряд больших монографий. Его «Opera omnia» – полное собрание сочинений, издаваемое в Швейцарии с 1909 года, рассчитано на 72 тома.

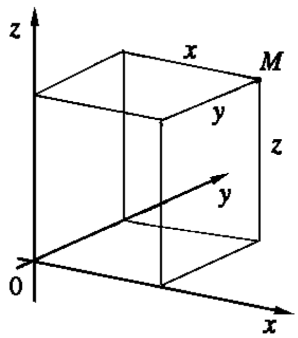


Рис. 49

Следуя Эйлеру, для задания точки M в пространстве выбирают три взаимно перпендикулярные оси Ox, Oy, Oz и в качестве координат этой точки принимают ее расстояния от плоскостей yz, xz и xy, взятые с соответствующими знаками. Эти координаты называются: x – абсциссой, y – ординатой и z – аппликатой точки $M(x,\ y,\ z)$ (рис. 49).

Чтобы найти расстояние между точками $M_1(x_1,\ y_1,\ z_1)$ и $M_2(x_2,\ y_2,\ z_2)$, заметим, что оно равно диагонали параллелепипеда со сторонами $x_2 - x_1$, $y_2 - y_1$, $z_2 - z_1$ (рис. 50).

Дважды применяя теорему Пифагора, получаем

$$M_1M^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2,$$

$$M_1M_2^2 = M_1M^2 + MM_2^2 = (x_2 - x_1)^2 + (y^2 - y_1)^2 + (z_2 - z_1)^2,$$

откуда $M_1M_2 = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$ – полная аналогия с формулой расстояния между точками на плоскости.

Обобщая идею Декарта, каждому уравнению $F(x,\ y,\ z) = 0$ с тремя неизвестными можно сопоставить поверхность – геометрическое место точек пространства, координаты которых $x,\ y,\ z$ удовлетворяют этому уравнению.

Поверхностями первого порядка, т.е. теми, которые определяются уравнением первой степени $Ax + By + Cz + D = 0$, оказываются плоскости. Важнейшими поверхностями второго порядка являются сферы, эллипсоиды, гиперболоиды, конусы и цилиндры.

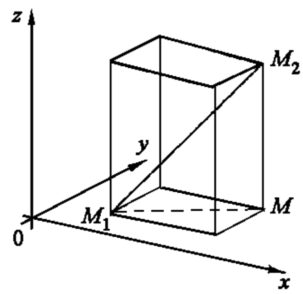


Рис. 50

Из формулы расстояния между двумя точками без труда получается уравнение сферы – геометрического места точек, равноудаленных от данной точки (центра). Если центр C имеет координаты $x_0,\ y_0,\ z_0$, а радиус сферы (т.е. расстояние ее точек от центра) равен r, то эта сфера описывается уравнением $(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 = r^2$. (Напишите уравнение сферы с центром в точке $(1, -1, -2)$ и радиусом 2. Найдите центр и радиус сферы $x^2 + y^2 + z^2 + 2x - 2y + 1 = 0$ – для этого уравнение нужно привести к стандартному для уравнения сферы виду.) Площадь сферы радиуса r равна $4\pi r^2$, а объем ограниченного ею шара составляет $\frac{4}{3}\pi r^3$.

Если вращать эллипс (см. рис. 31) вокруг оси Ox, его контур опишет поверхность – эллипсоид вращения. Увеличим (или уменьшим) аппликаты всех точек этого эллипса в одно и то же число раз, т.е. растянем (или сожмем) эллипсоид вращения вдоль оси Oz. Полученная поверхность, похожая на морской камешек, отшлифованный прибоем, называется трехосным эллипсоидом (рис. 51). Любое сечение эллипсоида плоскостью будет эллипсом.

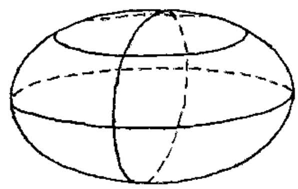


Рис. 51

Вращение гиперболы вокруг одной из ее двух осей симметрии с последующим растяжением или сжатием приводит к гиперболоидам: однополостному и двухполостному (рис. 52а,б). В сечениях гиперболоидов плоскостями получаются эллипсы и гиперболы. Аналогично парабола порождает эллиптический параболоид (рис. 53а). Форму параболоида вращения имеют зеркала в прожекторах и автомобильных фарах: параболическое зеркало отражает лучи источника света в направлении, параллельном его оси. Наоборот, свет далекой звезды, придя (почти) параллельными лучами, отражается от параболического зеркала телескопа-рефлектора и собирается в одной точке, которая и считается изображением светила. (Хотелось бы обсудить форму зеркал в гиперболоиде инженера Гарина, но чертежи и объяснения, представленные в романе А.Н.Толстого, не имеют реальной теоретической основы.)

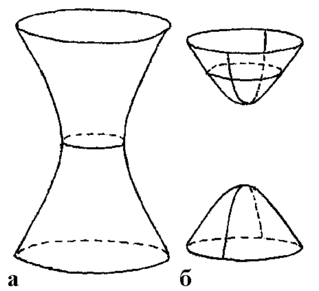


Рис. 52

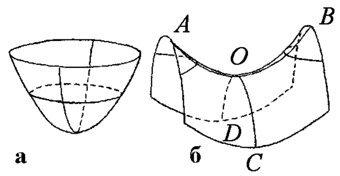


Рис. 53

Если в пространстве взять параболу AOB (см. рис. 53б) и в плоскости, перпендикулярной ее плоскости, другую параболу COD, то при скольжении вершины O по параболе AOB парабола COD, двигаясь параллельно себе, опишет седлообразную поверхность, называемую гиперболическим параболоидом. Сечения этой поверхности плоскостями, перпендикулярными оси неподвижной параболы AOB, дают два семейства гипербол (объясните, какие).

Прямой круговой конус изображен на рис. 32, прямой круговой цилиндр – это бесконечная в обе стороны круглая труба.

Кроме поверхностей первого и второго порядка, к числу самых известных образов трехмерного эвклидова пространства относятся тор, катеноид, псевдосфера, лист Мебиуса и бутылка Клейна.

Форму тора имеют бублик, спасательный круг и многие другие полезные предметы. Тор получается вращением окружности около прямой, лежащей в той же плоскости и не пересекающей окружность. Если через r обозначить радиус окружности, а через R расстояние от центра до оси вращения, то поверхность тора будет равна $4\pi^2 Rr$, а объем его как тела составит $2\pi^2 Rr^2$. Torus на латыни означает «выпуклость».

Катеноид образуется вращением цепной линии (см. рис. 40) вокруг оси Ox (лат. catena означает «цепь»). Форму катеноида принимает мыльная пленка между двумя раздвинутыми проволочными кругами (рис. 54).

Псевдосферу описывает вращающаяся вокруг оси Ox трактриса (рис. 55). Для существ, живущих на псевдосфере, естественной геометрией их мира является геометрия Лобачевского (а не Эвклида, как на обычной плоскости).

Лист Мебиуса весьма популярен в произведениях научной фантастики и любительских рассуждениях о непростоте окружающего мира. Если взять бумажную ленту и, перекрутив ее один раз (т.е. на 180°), склеить концы, то как раз и получится лист Мебиуса (рис. 56). Впервые это проделал в 1858 году немецкий математик Август Фердинанд Мебиус (1790–1868). (Одновременно и независимо ту же поверхность открыл его соотечественник Листинг, но терминологический памятник достался не ему.) Лист Мебиуса представляет собой одностороннюю поверхность: по ней, не переходя через край ленты, можно попасть из любой заданной точки в любую другую.

Бутылка Клейна (рис. 57) – тоже односторонняя поверхность, но без края. Несмотря на то что эта поверхность замкнутая, она не делит пространство на внутреннюю и внешнюю по отношению к ней части (в отличие, скажем, от сферы или эллипсоида). Если горлышко обычной бутылки сильно вытянуть, проткнуть им боковую стенку бутылки, затем ее дно и, выйдя наружу, «загладить» край горлышка с дном, то получится бутылка Клейна (воображение автора эрлангенской программы создало этот вызывающий изумление образ в 1874 году).

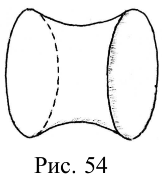


Рис. 54

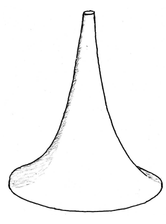


Рис. 55

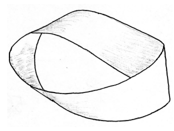


Рис. 56

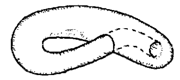


Рис. 57

Декартовы координаты явились могучим средством для исследования свойств линий и поверхностей, для изучения геометрии плоскости и пространства. Вместе с тем язык координат оказывается неудобным при формулировке, например, физических законов. Во-первых, в координатной форме они зачастую приобретают громоздкий вид, а во-вторых, формулы, выражающие данный закон, выглядят по-разному в зависимости от выбора осей координат. Когда линия задается как геометрическое место точек, обладающих некоторым определенным свойством, это определение не зависит от системы координат (мы вводим ее позже, чтобы изучать линию). Но если изъясняться только на координатном языке, то один и тот же геометрический образ может представиться в разных системах координат очень непохожими уравнениями. Например, уравнения $xy = 1$ и $x^2 - y^2 = 2$ задают одну и ту же гиперболу: в первом случае оси координат совпадают с ее асимптотами, а во втором – с ее осями симметрии.

В физике удобной формой выражения законов оказалась их векторная запись. Она компактна и инвариантна относительно координат в пространстве.

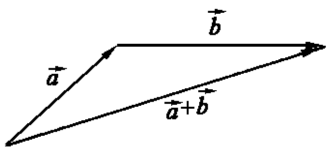


Рис. 58

Вектором называется направленный отрезок. Два вектора считаются равными, если они имеют одинаковую длину и одно и то же направление. Вектор $\vec{0}$, начало и конец которого совпадают, называют нуль-вектором; ему приписывается произвольное направление. Пусть $\vec{a}$ и $\vec{b}$ – векторы. Отложим вектор $\vec{b}$ от конца вектора $\vec{a}$. Вектор, начало которого совпадает с началом вектора $\vec{a}$, а конец – с концом вектора $\vec{b}$, по определению, является векторной суммой $\vec{a} + \vec{b}$ (рис. 58).

Вектор, равный по длине вектору $\vec{a}$, но противоположно направленный, называется противоположным для $\vec{a}$ вектором и обозначается $-\vec{a}$.

Теорема 1. Множество всех векторов пространства образует относительно операции сложения абелеву группу.

ДОКАЗАТЕЛЬСТВО. 1) Сложение векторов ассоциативно. В самом деле, пусть $\vec{a}$, $\vec{b}$, $\vec{c}$ – произвольные векторы. Отложим $\vec{b}$ от конца вектора $\vec{a}$, а $\vec{c}$ – от конца вектора $\vec{b}$. Вектор-сумма $(\vec{a}+\vec{b})+\vec{c}$ (сначала складываются $\vec{a}$ и $\vec{b}$, а затем к результату прибавляется $\vec{c}$) соединяет начало вектора $\vec{a}$ с концом вектора $\vec{c}$. Но точно так же выглядит и сумма $\vec{a} + (\vec{b} + \vec{c})$ (проделайте эти построения). Таким образом, $(\vec{a} + \vec{b}) + \vec{c} = \vec{a} + (\vec{b} + \vec{c})$.

2) Нуль-вектор $\vec{0}$ является нейтральным элементом. Действительно, чтобы построить сумму $\vec{a}+\vec{0}$, нужно отложить $\vec{0}$ от конца вектора $\vec{a}$. Очевидно, что при этом получится вектор $\vec{a}$, так что $\vec{a}+\vec{0}=\vec{a}$. Аналогично $\vec{0}+\vec{a}=\vec{a}$.

3) Для каждого вектора $\vec{a}$ существует вектор, дающий в сумме с $\vec{a}$ нуль-вектор. Этим свойством, легко догадаться, обладает противоположный для $\vec{a}$ вектор $-\vec{a}$, т.е. $\vec{a}+(-\vec{a})=\vec{0}=-\vec{a}+\vec{a}$.

Свойства 1)-3) означают, что векторы образуют группу по сложению.

4) Сложение векторов коммутативно. Для доказательства рассмотрим изображенный на рис. 59 параллелограмм. Его диагональ можно воспринимать как сумму $\vec{a}+\vec{b}$ или как сумму $\vec{b}+\vec{a}$ в зависимости от того, верхний или нижний треугольник мы рассматриваем, так что $\vec{a}+\vec{b}=$ $=\vec{b}+\vec{a}$.

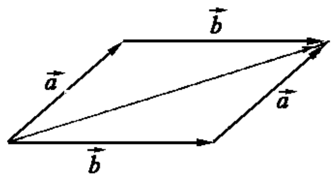


Рис. 59

Таким образом, аддитивная группа векторов пространства является коммутативной, или абелевой. □

Разностью $\vec{a}-\vec{b}$ векторов $\vec{a}$ и $\vec{b}$ называется сумма $\vec{a}+(-\vec{b})$ вектора $\vec{a}$ и вектора, противоположного вектору $\vec{b}$.

Длина вектора $\vec{a}$ обозначается символом $|\vec{a}|$.

Произведением вектора $\vec{a}$ на число λ называется вектор $\lambda\vec{a}$, длина которого равна $|\lambda||\vec{a}|$, а направление совпадает с направлением вектора $\vec{a}$ при $\lambda>0$ и противоположно ему, если $\lambda<0$. Например, вектор $3\vec{a}$ направлен как вектор $\vec{a}$ и имеет втрое большую длину; вектор $(-\frac{1}{2})\vec{a}$ имеет противоположное направление и вдвое короче вектора $\vec{a}$.

Второй закон Ньютона записывается в виде $\vec{F}=m\vec{a}$, где $\vec{F}$ - сила, действующая на тело, m – масса тела, $\vec{a}$ – его ускорение. Закон всемирного тяготения гласит, что каждая масса m_1 притягивается другой массой m_2 с силой

$$\vec{F}=\frac{Gm_1m_2}{|\vec{r}|^2}\vec{r}^0,$$

где $\vec{r}$ – вектор, идущий от m_1 к m_2, $\vec{r}^0$ – вектор единичной длины в направлении $\vec{r}$, G – гравитационная постоянная.

Пусть в пространстве задана некоторая декартова система координат. Через $\vec{i}$, $\vec{j}$, $\vec{k}$ обозначим векторы, имеющие длину 1 и направления, совпадающие с положительными направлениями осей сответственно Ox, Oy, Oz. Эти векторы называются базисными векторами данной системы координат. Координатами вектора $\vec{a}$ называются координаты его конца, если вектор $\vec{a}$ отложить от начала координат O (рис. 60).

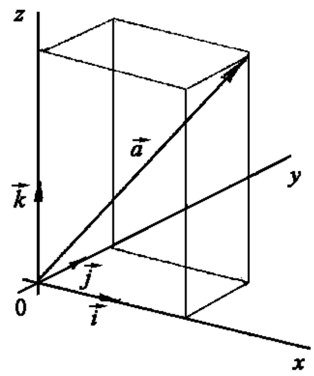


Рис. 60

Если a_x, a_y, a_z – координаты вектора $\vec{a}$, то получаем следующее его разложение по базисным векторам: $\vec{a} = a_x\vec{i}+a_y\vec{j}+a_z\vec{k}$. Для вектора $\lambda\vec{a}$ имеем $\lambda\vec{a} = \lambda(a_x\vec{i} + a_y\vec{j} + a_z\vec{k}) = (\lambda a_x)\vec{i} + (\lambda a_y)\vec{j} + (\lambda a_z)\vec{k}$. Отсюда следует, что координатами вектора $\lambda\vec{a}$ будут числа λa_x, λa_y, λa_z, т.е. при умножении вектора на число все его координаты умножаются на это число. Пусть $\vec{b} = b_x\vec{i}+b_y\vec{j}+b_z\vec{k}$. Тогда $\vec{a}+\vec{b} = (a_x\vec{i}+a_y\vec{j}+a_z\vec{k})+(b_x\vec{i}+b_y\vec{j}+b_z\vec{k}) = (a_x+b_x)\vec{i}+(a_y+b_y)\vec{j}+(a_z+b_z)\vec{k}$. Следовательно, при сложении векторов складываются их одноименные координаты.

Проделайте выкладки, показывающие, что при вычитании векторов происходит вычитание их одноименных координат. Найдите координаты векторов $3\vec{a}$, $-2\vec{b}$, $\vec{a}+\vec{b}$, $\vec{a}-\vec{b}$, $3\vec{a}-2\vec{b}$, зная, что $\vec{a}(1,\ -2,\ 3)$ и $\vec{b}(-3,\ 2,\ 1)$.

На плоскости каждый вектор будет иметь две координаты: коэффициенты его разложения по базисным векторам $\vec{i}$ и $\vec{j}$. Постройте на плоскости векторы $\vec{a} = -\vec{i} + 2\vec{j}$, $\vec{b} = 3\vec{i} + \vec{j}$, $\vec{c} = \vec{i} - \vec{j}$, откладывая их от начала координат.

Векторы, направления которых параллельны, называются коллинеарными. Такие векторы можно расположить на одной прямой.

Теорема 2. Два вектора коллинеарны тогда и только тогда, когда их одноименные координаты пропорциональны.

ДОКАЗАТЕЛЬСТВО. Пусть векторы $\vec{a}(a_x,\ a_y,\ a_z)$ и $\vec{b}(b_x,\ b_y,\ b_z)$ коллинеарны. Расположив их на одной прямой, можем записать, что $\vec{a} = \lambda\vec{b}$, где λ – подходящее число. Отсюда следует, что $a_x = \lambda b_x,\ a_y = \lambda b_y,\ a_z = \lambda b_z$, т.е. $\dfrac{a_x}{b_x} = \dfrac{a_y}{b_y} = \dfrac{a_z}{b_z} = \lambda$ – одноименные координаты векторов пропорциональны.

С другой стороны, если $\dfrac{a_x}{b_x} = \dfrac{a_y}{b_y} = \dfrac{a_z}{b_z} = \lambda$, то $a_x = \lambda b_x$, $a_y = \lambda b_y,\ a_z = \lambda b_z$, откуда $\vec{a} = \lambda\vec{b}$. Векторы $\vec{a}$ и $\vec{b}$ имеют либо одинаковое, либо противоположное направление. Значит, они коллинеарны. □

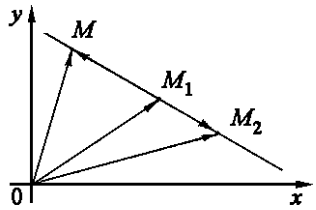


Рис. 61

Пользуясь теоремой 2, составим уравнение прямой, проходящей на плоскости xy через точки $M_1(x_1,\ y_1)$ и $M_2(x_2,\ y_2)$ (рис. 61).

Пусть $M(x,\ y)$ – произвольная точка прямой. Тогда векторы $\overline{M_1M}$ и $\overline{M_1M_2}$ коллинеарны. Так как $\overline{M_1M} = \overline{OM} - \overline{OM_1}$, а векторы $\overline{OM}$ и $\overline{OM_1}$ имеют координаты соответственно $(x,\ y)$ и $(x_1,\ y_1)$, то координатами вектора $\overline{M_1M}$ будут $(x - x_1,\ y - y_1)$. Аналогично $\overline{M_1M_2} = \overline{OM_2} - \overline{OM_1}$, и значит, $\overline{M_1M_2}$ имеет координаты $(x_2 - x_1,\ y_2 - y_1)$. В силу коллинеарности векторов $\overline{M_1M}$ и $\overline{M_1M_2}$ их одноименные координаты пропорциональны:

$$\frac{x - x_1}{x_2 - x_1} = \frac{y - y_1}{y_2 - y_1}.$$

Это и есть уравнение прямой, проходящей через точки $M_1(x_1,\ y_1)$ и $M_2(x_2,\ y_2)$. Например, если $M_1(-1,\ 2)$, $M_2(3,\ 1)$, то прямая, определяемая этими точками, имеет уравнение

$$\frac{x - (-1)}{3 - (-1)} = \frac{y - 2}{1 - 2},\ \text{или}\ \frac{x + 1}{4} = \frac{y - 2}{-1},\ \text{или}\ x + 4y - 7 = 0.$$

Так как отсюда $y = -\dfrac{1}{4}x + \dfrac{7}{4}$, то данная прямая имеет угловой коэффициент $k = -\dfrac{1}{4}$ и отсекает на оси Oy отрезок $\dfrac{7}{4}$. Напишите уравнение прямой, проходящей через точки $M_1(2,\ -3)$, $M_2(0,\ -1)$; через точки $M_1(2,\ 0)$, $M_2(-3,\ 1)$. Чему равны у этих прямых угловой коэффициент и отрезок, отсекаемый на оси ординат?

Важнейшей конструкцией векторного исчисления является скалярное произведение. Скалярным произведением ненулевых векторов $\vec{a}$ и $\vec{b}$ называется число, равное произведению длин этих векторов и косинуса угла φ между ними:

$$\vec{a}\vec{b} = |\vec{a}||\vec{b}|\cos\varphi.$$

(Скалярное произведение любого вектора и нуль-вектора полагают равным нулю.)

Например, работа A, совершаемая постоянной силой $\vec{F}$, приложенной к телу, на пути $\vec{s}$, вычисляется по формуле $A = \vec{F}\vec{s}$.

Если векторы $\vec{a}$ и $\vec{b}$ перпендикулярны, т.е. $\varphi = \dfrac{\pi}{2}$, то $\cos\varphi =$ $= 0$ и, значит, $\vec{a}\vec{b} = 0$. Обращение в ноль скалярного произведения – признак перпендикулярности векторов.

Если $\vec{a} = \vec{b}$, то $\varphi = 0$, откуда $\cos\varphi = 1$ и, значит, $\vec{a}^2 = |\vec{a}|^2$, т.е. скалярный квадрат вектора равен квадрату его длины. Отсюда получаем, что $|\vec{a}| = \sqrt{\vec{a}^2}$,– длина вектора равна квадратному корню из его скалярного квадрата. Формула

$$\cos\varphi = \frac{\vec{a}\vec{b}}{|\vec{a}||\vec{b}|}$$

выражает угол между векторами через их скалярное произведение и длины.

Очевидно, что скалярное произведение не зависит от порядка сомножителей: $\vec{a}\vec{b} = \vec{b}\vec{a}$. Можно показать, что оно обладает свойством дистрибутивности относительно сложения векторов, т.е. что $\vec{a}(\vec{b} + \vec{c}) = \vec{a}\vec{b} + \vec{a}\vec{c}$. Еще одно полезное свойство: $\lambda(\vec{a}\vec{b}) = (\lambda\vec{a})\vec{b} = \vec{a}(\lambda\vec{b})$ – числовой множитель можно отнести к любому из скалярно перемножаемых векторов. Следующая теорема дает выражение скалярного произведения в координатах.

<u>Теорема 3</u>. $\vec{a}\vec{b} = a_x b_x + a_y b_y + a_z b_z$
(т.е. скалярное произведение векторов равно сумме попарных произведений их одноименных координат).

ДОКАЗАТЕЛЬСТВО. Разложим векторы $\vec{a}$ и $\vec{b}$ по базисным векторам $\vec{i}$, $\vec{j}$, $\vec{k}$ пространственной системы координат: $\vec{a} = a_x\vec{i} + a_y\vec{j} + a_z\vec{k}$, $\vec{b} = b_x\vec{i} + b_y\vec{j} + b_z\vec{k}$. Так как векторы $\vec{i}$, $\vec{j}$, $\vec{k}$ попарно перпендикулярны, то $\vec{i}\vec{j} = \vec{i}\vec{k} = \vec{j}\vec{k} = 0$. Поскольку эти векторы имеют длину 1, их скалярные квадраты равны 1, т.е. $\vec{i}^2 = \vec{j}^2 = \vec{k}^2 = 1$. Используя эти замечания и свойства скалярного произведения, упомянутые перед формулировкой теоремы, имеем:

$$\vec{a}\vec{b} = (a_x\vec{i} + a_y\vec{j} + a_z\vec{k})(b_x\vec{i} + b_y\vec{j} + b_z\vec{k}) = a_xb_x + a_yb_y + a_zb_z,$$

что и требовалось. □

Умножая скалярно вектор $\vec{a} = a_x\vec{i} + a_y\vec{j} + a_z\vec{k}$ на базисный вектор $\vec{i}$, получаем: $\vec{a}\vec{i} = a_x$. Аналогично $a_y = \vec{a}\vec{j},\ a_z = \vec{a}\vec{k}$. Таким образом, координаты вектора – это его скалярные произведения на базисные векторы, так что $\vec{a} = (\vec{a}\vec{i})\vec{i} + (\vec{a}\vec{j})\vec{j} + (\vec{a}\vec{k})\vec{k}$. В этом представлении вектора $\vec{a}$ декартова система координат в ее стандартном виде не участвует: есть три взаимно перпендикулярных вектора единичной длины, все остальные векторы пространства выражаются через них.

В следующих упражнениях нужно применить свойства векторов.

1°. Докажите, что координаты фиксированного вектора равны разности одноименных координат его конца и начала (воспользуйтесь рис. 61).

2°. Будет ли прямоугольным треугольник, образованный точками $A(1,\ -1,\ 2)$, $B(0,\ 2,\ 3)$ и $C(5,\ 1,\ 0)$?

3°. Найдите угол между векторами $\vec{i} - \vec{j}$ и $\vec{k} - \vec{j}$.

4°. Докажите, что прямая, проходящая через точки $A(-1,\ 1)$ и $B(1,\ 4)$, перпендикулярна прямой, проходящей через точки $C(5,\ 0)$ и $D(2,\ 2)$.

Векторное исчисление предложил в 1844 году Герман Гюнтер Грассман (1809–1877), немецкий математик и филолог. Всю жизнь он работал учителем гимназии, преподавая математику, язык и Закон Божий. Изучив санскрит, перевел «Ригведу», наиболее древний и значительный памятник индийской литературы, собрание 1028 арийских гимнов (10 книг). Математические работы Грассмана не были поняты и лишь в конце 1860-х годов привлекли широкое внимание отчетливо высказанной в них идеей о многомерном пространстве.

Понятие вектора позволяет подойти к вопросу о размерности пространства с формальной точки зрения, не привлекая никаких физических или философских соображений. Всякий вектор $\vec{a}$, расположенный на прямой, может быть выражен через один единственный вектор $\vec{i}$ единичной длины: $\vec{a} = a_x\vec{i}$, где a_x – длина вектора $\vec{a}$, взятая с плюсом, если $\vec{a}$ одинаково направлен с $\vec{i}$, и с минусом, в случае противоположной направленности $\vec{a}$ и $\vec{i}$. Можно записать $a_x = \vec{a}\vec{i}$. На плоскости каждый вектор $\vec{a}$ представим в виде $\vec{a} = (\vec{a}\vec{i})\vec{i} + (\vec{a}\vec{j})\vec{j}$, где $\vec{i}$, $\vec{j}$ - два взаимно перпендикулярных единичных вектора. В пространстве $\vec{a} = (\vec{a}\vec{i})\vec{i} + (\vec{a}\vec{j})\vec{j} + (\vec{a}\vec{k})\vec{k}$,– любой вектор раскладывается по трем попарно перпендикулярным единичным векторам $\vec{i}$, $\vec{j}$, $\vec{k}$. Количество базисных векторов и есть размерность пространства. Прямая – одномерное пространство, плоскость – двумерное, собственно пространство – трехмерно. Присоединив формально к векторам $\vec{i}$, $\vec{j}$, $\vec{k}$ единичный вектор $\vec{l}$, перпендикулярный каждому из них, получим четырехмерное пространство, где каждый вектор $\vec{a}$ будет иметь представление $\vec{a} = (\vec{a}\vec{i})\vec{i} + (\vec{a}\vec{j})\vec{j} + (\vec{a}\vec{k})\vec{k} + (\vec{a}\vec{l})\vec{l}$. И так можно продолжать сколько угодно.

Формулы векторного исчисления, если не выражать их в координатах, не зависят от размерности пространства, которая, таким образом, приобретает второстепенное значение.

Другое дело, как идею о многомерном пространстве совместить с нашим воображением, ограниченным тремя измерениями физического пространства. Но это зависит от силы воображения. В послании святого апостола Павла к Ефесянам читаем: «Чтобы вы, укорененные и утвержденные в любви, могли постигнуть со всеми святыми, что́ широта и долгота, и глубина и высота» (глава 3, стих 18). Можно допустить, что здесь идет речь о четырехмерности пространства, дополненного атрибутом духовности. Впрочем, в некоторых рассуждениях средневековых схоластов куб как тело, изготовленное из некоторого материала, имел три измерения: длину, ширину и высоту, но с полым кубом связывалось еще одно пространственное свойство – глубина.

§ 4. Неэвклидовы геометрии и физическое пространство

Понятие n-мерного пространства было вполне освоено математиками во второй половине XIX века. Оно позволило по-новому взглянуть на традиционные геометрические конструкции, дало возможность облечь в ясную и простую форму алгебраические теоремы об уравнениях со многими неизвестными, открыло путь к трактовке различных совокупностей объектов как своеобразных пространств (появились функциональные пространства в анализе, фазовые пространства в физике и т.п.). Пространства различной (и даже бесконечной) размерности стали обыденными в научной практике. Вместе с тем в кругах, далеких от собственно математических проблем, идея о многомерном пространстве по-прежнему остается предметом пристального интереса и часто служит основой для толкования тех или иных непознанных явлений. Большую роль в этом сыграла появившаяся в теории относительности конструкция четырехмерного пространства-времени в качестве геометрической модели нашего мира. Человек с его «трехмерной» интуицией уподобился плоскому обитателю некоторой поверхности, не могущему представить себе третье измерение.

Образ разумного существа, обреченного жить в пространстве с той или иной «нечеловеческой» геометрией, одинаково популярен как в научной фантастике, так и в профессиональной литературе. Его привлекают для объяснения особенностей проективного пространства, для иллюстрации понятий внутренней геометрии различных поверхностей, при обсуждении парадоксов теории относительности и т.д. В романе английского педагога Эббота «Флатландия» (1884) описывается жизнь на плоскости, населенной многоугольниками, скользящими по ней, как тени. О приключениях плоских обитателей расширяющейся сферы рассказывается в книге «Сферландия» (1957), которую написал голландский ученый Бюргер. «Вот так же и мы», – должен подумать читатель, тщетно пытающийся вообразить четвертую ось координат.

Пусть в трехмерном эвклидовом пространстве задана некоторая поверхность. Изгибанием называется такая деформация этой поверхности, при которой не изменяются длины всех лежащих на ней линий. Лист бумаги, сворачиваемый в трубку, или полиэтиленовая пленка, колышущаяся на воде, дают наглядное представление об изгибании. Совокупность свойств, не меняющихся при изгибаниях поверхности, составляет ее внутреннюю геометрию. К ней, кроме уже упомянутых длин кривых, относятся углы между линиями, площади фигур, все, что может быть определено без выхода в объемлющее пространство. Например, планиметрия – это внутренняя геометрия плоскости. Важнейшую роль в планиметрии играют прямые и их конечные части – отрезки. Кратчайшим путем между двумя точками плоскости является отрезок проведенной через них прямой. На произвольной поверхности аналогом прямых являются геодезические линии. Так называются линии поверхности, дающие кратчайшее расстояние между любыми двумя ее точками. Геодезическая – это след материальной точки, движущейся по инерции на данной поверхности.

На сфере геодезическими линиями являются большие круги – окружности, полученные от пересечения сферы с плоскостями, проходящими через ее центр. Если на сфере даны две не диаметрально противоположные точки, то существует единственный большой круг, на котором они находятся. Расстояние между точками измеряется по меньшей из дуг этого большого круга. В сферической геометрии нет параллельных геодезических: любые два больших круга пересекаются – в двух точках. При этом получаются четыре сферических двуугольника – фигуры, не имеющие прямолинейных аналогов на плоскости.

Сферический треугольник образуется тремя точками сферы и тремя дугами больших кругов, меньшими полуокружности и имеющими концы в данных точках. Сумма углов сферического треугольника всегда больше $180°$, но не превосходит $540°$ (на плоскости сумма углов любого прямолинейного треугольника равна двум прямым углам, т.е. $180°$). Сферические треугольники считаются равными, если они могут быть совмещены после передвижения по сфере. Кроме обычных трех признаков равенства прямолинейных треугольников плоскости, в сферической геометрии есть еще один – по трем углам: два сферических треугольника равны, если равны их соответствующие углы. Это означает, что на сфере нет неравных подобных треугольников.

Очевидно, что если одна поверхность путем изгибания может быть наложена на другую, то эти поверхности имеют одинаковую внутреннюю геометрию. Например, кусок цилиндра можно разогнуть и уложить его на плоскости, но никакую часть сферы путем изгибания нельзя «распрямить». Это означает, что внутренние геометрии плоскости и цилиндра совпадают, но у плоскости и сферы они различны. Заметим, что когда мы говорим о внутренней геометрии поверхности, то предполагается, что исследуются геометрические свойства некоторой части этой поверхности, а не всей ее в целом. Так, на цилиндре имеются замкнутые геодезические – окружности, получающиеся при пересечении цилиндра плоскостями, перпендикулярными его оси. На плоскости же, где геодезическими линиями являются только прямые, двигаясь по геодезической в одном направлении, нельзя вернуться в исходную точку. Но в малом, как уже говорилось, внутренние геометрии плоскости и цилиндра не различимы.

Окружностью на поверхности называется геометрическое место точек этой поверхности, равноудаленных от некоторой ее точки C (расстояние измеряется по геодезическим). Пусть r – радиус такой окружности. Если бы дело происходило на плоскости, длина окружности равнялась бы $2\pi r$. Но в зависимости от рельефа поверхности вблизи данной точки, эта величина может оказаться больше или меньше числа $2\pi r$ (рис. 62): в случае выпуклости (C– вершина) меньше, при седлообразной форме (C на перевале) – больше. Степень отклонения истинной длины окружности на поверхности от длины окружности того же радиуса на плоскости выражается некоторым числом K, называемым полной кривизной поверхности в данной точке. Для плоскости $K = 0$ во всех ее точках.

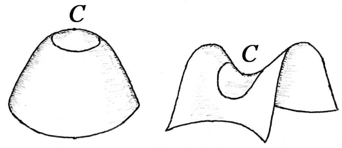


Рис. 62

Замечательным открытием Гаусса было то, что полная кривизна, определение которой существенно зависит от расположения поверхности в пространстве, на самом деле принадлежит внутренней геометрии поверхности, т.е. не меняется при изгибаниях. Theorema egregium («прекрасная теорема» – так назвал Гаусс доказанное им предложение) показывает, как можно вычислить полную кривизну с помощью измерений, производимых только на поверхности.

В общем случае полная кривизна поверхности меняется от точки к точке. Поверхность, во всех точках которой K принимает одно и то же значение, называется поверхностью постоянной кривизны. Плоскость и все развертывающиеся на нее поверхности (цилиндр, конус, колышущаяся на волнах полиэтиленовая пленка и т.п.) имеют постоянную нулевую кривизну: $K = 0$. Для сферы радиуса r во всех ее точках $K = \dfrac{1}{r^2}$ – полная кривизна здесь является постоянной и положительной величиной. Пример поверхности с постоянной отрицательной кривизной дает псевдосфера, для нее $K = -\dfrac{1}{a^2}$, где a – радиус «раструба». Так как полная кривизна не меняется при изгибаниях, то, например, кусок сферы невозможно наложить на сферу другого радиуса и никакую часть псевдосферы нельзя распрямить на плоскость. Разумные плоские обитатели сферы и цилиндра, произведя измерения в окрестности всего лишь одной точки своей поверхности, смогут установить, что живут в мирах, по-разному выглядящих в трехмерном пространстве, не представляя себе, впрочем, что это означает (уточните, какие это могли бы быть измерения).

Во все времена люди размышляли о том, как устроено в геометрическом смысле окружающее пространство и какое место занимает в нем Земля, местообитание человечества. Аристотель считал, что Земля шарообразна, а звезды и планеты расположены на хрустальных сферах, вращающихся вокруг нее как общего центра мироздания. Это представление по существу не менялось до появления в 1543 году системы Коперника, согласно которой центром вращения небесных сфер стало Солнце.

Радиус земной сферы равен примерно 6400 километров. Обыденная деятельность каждого человека ограничена столь малым участком этой сферы, что в пределах разумной точности измерений практическую геометрию земной поверхности можно считать геометрией плоскости. Именно поэтому в школах изучается планиметрия, а не сферическая геометрия.

В математике древнего мира все математические рассуждения облекались в геометрическую форму. Алгебраические соотношения воспринимались как имеющие смысл лишь постольку, поскольку они могли быть представлены в виде некоторых свойств геометрических объектов. Геометрия была единственной и всеобъемлющей наукой. Важнейшим среди творений математиков классической древности являются «Начала» Эвклида (300 г. до н.э.). В тринадцати книгах этого сочинения был подведен итог многовековых исследований, создана логическая система, в рамках которой математика развивалась последующие две тысячи лет. Изложение материала было основано на аксиоматическом методе: перечислялись основные объекты теории и объявлялись первоначальными, принимаемыми без доказательства истинами (аксиомами) некоторые свойства этих объектов. Все дальнейшие предложения (теоремы) получались путем строгих логических рассуждений – доказательств. Из пяти аксиом, взывающих к геометрической интуиции читателя (Эвклид называл эти положения постулатами), четыре абсолютно очевидны:

I. От всякой точки до всякой точки можно провести прямую.

II. Ограниченную прямую (т.е. отрезок) можно непрерывно продолжить по прямой.

III. Из всякого центра всяким раствором (циркуля) может быть описан круг.

IV. Все прямые углы равны между собой.

Далее идет знаменитый Пятый постулат. Две тысячи лет невероятного интеллектуального напряжения, связанного с попытками осознать его роль в геометрии, привели к открытию, изменившему взгляд человечества на физический мир. Итак,

V. Если прямая, падающая на две прямые, образует внутренние и по одну сторону углы, в сумме меньшие двух прямых углов, то, продолженные неограниченно, эти две прямые встретятся с той стороны, где углы меньше двух прямых (рис. 63).

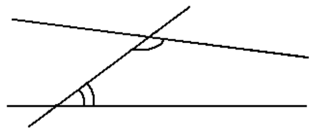


Рис. 63

Как видим, в отличие от предшествующих четырех действительно очевидных положений, Пятый постулат требует известных усилий для уяснения его смысла. И античные геометры, и математики последующих эпох не могли смириться с мыслью о том, чтобы столь сложное утверждение было признано первоначальной истиной. Начиная с Архимеда, на протяжении многих веков предпринимались попытки найти доказательство Пятого постулата, т.е. вывести его из других, более простых допущений. В V веке н.э. византийский математик Прокл установил, что Пятый постулат равносилен следующему предложению: «На плоскости через точку, не принадлежащую данной прямой, проходит единственная параллельная ей прямая». Именно эта формулировка впоследствии вошла в школьные учебники. (Напомним, что две прямые плоскости называются параллельными, если они не пересекаются). Хотя в приведенной форме Пятый постулат обладает достаточно высокой степенью очевидности, борьба за его удаление из числа аксиом продолжалась. Но все усилия оставались бесплодными: в каждом новом «доказательстве» очень скоро обнаруживалась ошибка. Чаще всего это была ссылка на какое-нибудь совсем уж «очевидное» свойство, которое при более внимательном анализе оказывалось равносильным злополучному постулату о параллельных. Например, его формами являются утверждение о том, что сумма углов каждого треугольника равна 180° или что существуют подобные треугольники, имеющие разную площадь.

Как и многие его предшественники, Николай Иванович Лобачевский (1792–1856) пытался доказать Пятый постулат методом от противного. Он предположил, что на плоскости через точку, не принадлежащую некоторой прямой, проходят по крайней мере две параллельные ей прямые. Отсюда нужно было вывести какоенибудь противоречие. Но противоречие не получалось. Длинной чередой появлялись удивительные теоремы, основанные на постулате о параллельных, противоположном эвклидову, и ни одна из них не опровергала другую. И тогда Лобачевский пришел к казавшемуся невероятным заключению: что, кроме эвклидовой, существует еще одна геометрия, которую он назвал «воображаемой». В этой геометрии сумма углов треугольника всегда меньше 180°; два треугольника равны, если равны их соответственные углы; через три точки, не лежащие на одной прямой, не всегда можно провести окружность; перпендикуляр, проведенный к одной стороне острого угла, может оказаться параллельным другой его стороне; линия, проходящая на постоянном расстоянии от данной прямой, не является прямой; если две прямые на плоскости не пересекаются, то они расходятся до бесконечности или в одну сторону, или в обе стороны.

Первое сообщение о новой геометрии Лобачевский сделал на заседании отделения физико-математических наук Казанского университета 23 (по новому стилю) февраля 1826 года. Эвклидова геометрия всегда воспринималась как учение о свойствах реального пространства. Так ее понимали и в физике Ньютона–Галилея. На что же в таком случае могла претендовать «воображаемая» геометрия? Развивая свою теорию, Лобачевский обнаружил, что в достаточно малых областях она почти не отличается от геометрии Эвклида. Он предположил, что физическое пространство в целом подчиняется законам неэвклидовой геометрии и что точные измерения в космических масштабах могут обнаружить это. На основе астрономических наблюдений им была вычислена сумма углов треугольника Земля–Солнце–Сириус. Она в самом деле оказалась меньше 180°, но, к сожалению, на величину, которую вполне можно было объяснить погрешностями приборов и расчетов.

По предложению Гаусса, Лобачевский в 1842 году был избран членом-корреспондентом Геттингенского королевского общества. Но ни это, ни личный авторитет (20 лет на посту ректора Казанского университета) не могли изменить резко отрицательного отношения научных кругов к главному достижению создателя неэвклидовой геометрии. Она по-прежнему воспринималась как плод болезненной фантазии («Что такое геометрия без аксиомы параллельных линий?» – недоумевал Н.Г.Чернышевский). Видимо, опасаясь подобной реакции, Гаусс, пришедший к тем же выводам, что и Лобачевский, не сообщал о них никому. Неэвклидову геометрию в 1832 году независимо открыл венгерский математик

Янош Бойяи (1802–1860), военный инженер («я сделал новый, другой мир»). Не получив поддержки от Гаусса, он приостановил дальнейшие публикации, а вскоре узнал о работах Лобачевского и, потрясенный утратой приоритета, отошел от научных занятий.

Последний свой труд («Пангеометрия», 1855) Лобачевский, потерявший зрение, диктовал. Ему так и не удалось доказать непротиворечивость построенной им геометрии, снять с нее затянувшийся статус «воображаемой». В 1868 году итальянский математик Бельтрами обнаружил, что внутренняя геометрия псевдосферы совпадает с внутренней геометрией плоскости Лобачевского, а в 1870 году Клейн установил, что если нет противоречий в геометрической системе Эвклида, то их нет и в системе Лобачевского. Так геометрия Лобачевского стала не только реальной, но и равноправной с геометрией Эвклида.

Но какая же из этих геометрий – Эвклида или Лобачевского – соответствует реальной структуре физического пространства? В малых частях мира, подвластных человеческому опыту, они фактически совпадают, и потому, в силу своей простоты (принцип экономии мышления), эвклидова геометрия более предпочтительна с практической точки зрения. Кроме того, она традиционно лежит в основе наших интуитивных представлений о пространственных формах мира, стала частью обыденного разума. Что касается глобальных проблем мироздания, то здесь дело оказалось значительно сложнее.

В 1854 году Георг Фридрих Бернгард Риман (1826–1866), вступая в должность доцента Геттингенского университета, прочел лекцию «О гипотезах, лежащих в основании геометрии». В общих чертах он представил геометрическую теорию, названную впоследствии римановой геометрией. В ней соединились идеи о многомерном пространстве, внутренней геометрии поверхности и неэвклидовой геометрии. Геометрии Эвклида и Лобачевского оказались частными случаями римановой геометрии. Присутствовавший на лекции Гаусс ничего не сказал, но ушел в глубокой задумчивости. Риману принадлежат выдающиеся достижения в теории чисел, во многих разделах анализа, в теоретической физике. Рано умерший от туберкулеза, он, как и Лобачевский, не дожил до того времени, когда новые геометрические идеи наконец были поняты и применены к описанию свойств физической Вселенной.

В 1898 году выдающийся французский математик, физик, астроном и философ Анри Пуанкаре (1854–1912), пытаясь объяснить опыты, показывавшие, что свет распространяется с одинаковой скоростью независимо от собственной скорости его источника, пришел к мысли о том, что не существует абсолютного времени, единого для всей Вселенной, что о событиях, разделенных большими растояниями, в общем случае нельзя сказать, какое из них произошло раньше другого. Пространство и время столь тесно связаны, что рассматривать их по отдельности невозможно. Математическое выражение формировавшейся физической теории (специальная теория относительности) было дано в терминах четырехмерного пространства-времени. Три пространственные координаты x, y, z и одна временна́я t фиксируют положение материальной точки. Точка, покоящаяся в пространстве, движется во времени по мировой линии, параллельной временно́й оси. Мировая линия останется прямой и при любом равномерном движении точки, но уже не будет параллельной оси времени. При неравномерном движении точки (даже если она перемещается в пространстве по прямой) мировая линия искривляется.

Эти математические построения приобрели твердую физическую основу после публикации в 1905 году статьи Альберта Эйнштейна, которую принято считать началом теории относительности. Геометрию четырехмерного пространства-времени детально разработал Герман Минковский (1864–1909), профессор Геттингенского университета. Она оказалась тесно связанной с геометрией трехмерного пространства Лобачевского.

Общая теория относительности, созданная Эйнштейном в 1916 году, трактовала тяготение не как силу, а как искривление пространства-времени вблизи материальных тел. Яблоко падает на Землю не потому, что она его притягивает: оно просто «скатывается» к ней в силу искривленности пространства, как скатывается маленький шарик к большому, лежащему на резиновой пленке и прогибающему ее. Математически все это означает, что пространство-время, геометрически неоднородное из-за распределенных в нем масс, представляет собой особый тип четырехмерного риманова пространства. Такова современная геометрическая модель мира, в котором мы живем.

ГЛАВА IV. **ЧТО ЗНАЧИТ «ДОКАЗАТЬ»?**

§ 1. Алгебра высказываний

Доказательство, в обычном понимании этого слова, представляет собой рассуждение, основанное на некоторых общепризнанных исходных положениях и призванное убедить кого-либо в справедливости доказываемого. Естественный язык и логика здравого смысла, с помощью которых строятся доказательства во всех сферах человеческих отношений, таят в себе, однако, известные опасности, могущие привести к парадоксальным ситуациям. Вот несколько примеров.

1°. Двигаясь по натуральному ряду, будем называть числа либо их стандартными десятичными именами, либо как-нибудь по-другому, лишь бы было ясно, о каком числе идет речь (например, «двадцать пять» и «пять в квадрате» – два названия числа 25). «Наименьшее натуральное число, которое нельзя назвать менее чем двумя русскими словами»,– очевидно, 21. Взятое в кавычки определение можно считать одним из названий числа 21. «Наименьшее натуральное число, которое нельзя назвать менее чем тремя русскими словами»,– одно из имен числа 121 и т.д. Где-то очень далеко появится в натуральном ряде «наименьшее натуральное число, которое нельзя назвать менее чем двенадцатью русскими словами». Обозначим это число через a. Одним из его названий будет последнее взятое в кавычки определение. Но в нем 11 слов, и значит, a названо меньше чем 12 словами. Возникающий логический тупик показывает, что в естественном языке существуют внешне безупречные, но внутренне противоречивые конструкции. (Рассмотренный парадокс открыл в 1905 году Жюль Ришар (1863–1956), доктор естественных наук и медицины, директор Океанографического музея в Монако.)

2°. (Вариация на тему архимедовского «Псаммита».) «Отдельно взятая песчинка не образует кучу песка. Если некоторое число песчинок не образует кучу песка, то и после добавления к ним еще одной песчинки куча песка не получится. Значит, никаким числом последовательно добавляемых песчинок нельзя образовать кучу песка». Правильные рассуждения приводят к выводу, не согласующемуся с реальностью и интуицией, из-за неопределенности понятия «куча песка».

3°. «Деревенский брадобрей бреет всех тех и только тех жителей своей деревни, которые не бреются сами. Бреет ли он сам себя?» Патовое положение, в которое мы попадаем при попытке ответить на этот вопрос, свидетельствует о несовершенстве логики здравого смысла.

Подобные парадоксы, конечно, не связаны с житейскими ситуациями, но в принципе могут возникнуть в тонких искусствоведческих или юридических доказательных рассуждениях, а тем более в таких неэкспериментальных областях, как «чистая» математика или богословие (известный вопрос схоластов: «Если Некто всемогущ, то может ли Он создать тяжесть, которую Сам не мог бы поднять?»). Желание оградить математические доказательства от искажений, вызванных особенностями естественного языка и интуитивной логики, привели к появлению формальной, или математической логики. Возникшая во второй половине XIX века, эта наука постепенно приобрела большое прикладное значение, далеко выходящее за рамки первоначальных сугубо математических целей. Ее идеи и методы проникли в лингвистику и психологию, в философию и информатику, она стала теоретической основой для создания компьютерной техники и искусственного интеллекта.

Начальным разделом математической логики является алгебра высказываний.

Высказыванием называется предложение, о котором можно судить, истинно оно или ложно. Например, «$\pi > 3$», «в слове ХЛЕБ четыре буквы», «функция $\operatorname{sign} x$ не является элементарной» – истинные высказывания, а суждения «орга́н – струнный инструмент», «в эвклидовой плоскости подобные треугольники равны», «функция $\sin x$ не является непрерывной» представляют собой ложные высказывания. Утверждение «в тексте повести А.С.Пушкина «Капитанская дочка» 29 343 слова» – тоже высказывание, хотя потребуются значительные усилия, чтобы убедиться, что оно истинно. Фразы типа «$x > 0$», «площадь этой фигуры равна 1» или «Василий Иванович поймал зайца» не будут высказываниями, поскольку истинность или ложность содержащихся в них фактов зависит от конкретного значения числа x, выбора упоминаемой фигуры, биографических сведений о ловце зайцев. Известный с древности парадокс заключается в том, что некто говорит: «Предложение, которое я сейчас произношу, – ложно» и нужно установить, истинным или ложным является этот тезис. Легко видеть, что, как и в деле брадобрея, при любом предположении получается противоречие. Выход один – признать, что произнесенные слова не составляют высказывания.

В каждом из следующих упражнений, пользуясь предложенными в них вариантами, составьте три истинных высказывания.

1°. На (плоскости, сфере, псевдосфере) сумма углов треугольника, образованного частями геодезических линий, (меньше, равна, больше) 180°.

2°. Оперу («Борис Годунов», «Евгений Онегин», «Иван Сусанин») сочинил композитор (М.И.Глинка, М.И.Мусоргский, П.И.Чайковский).

3°. Функция $(-x,\ \dfrac{2}{x},\ \dfrac{2}{x^2})$ является производной от функции $(2\ln x,\ -\dfrac{2}{x},\ -\dfrac{x^2}{2})$.

При доказательных рассуждениях из одних высказываний при помощи речевых конструкций образуются другие, более сложные. В математических текстах важнейшими связками являются союзы «и», «или», условный оборот «если – то», конструкция «тогда и только тогда, когда». В алгебре высказываний им соответствуют логические операции конъюнкция, дизъюнкция, импликация и эквивалентность, которые мы сейчас определим. Условимся обозначать высказывания малыми латинскими буквами p, q и т.д., может быть, с индексами. Эти буквы называются пропозициональными переменными, от латинского propositium – высказывание. Логическим значением высказывания p считается символ 1, если высказывание p истинно, и символ 0, если p ложно.

Конъюнкцией высказываний p, q называется высказывание «p и q», обозначаемое $p \wedge q$, логическое значение которого устанавливается по табл. 3.

Таблица 3

p	q	$p \wedge q$
0	0	0
0	1	0
1	0	0
1	1	1

Таким образом, конъюнкция истинна только в одном случае, когда оба составляющих ее высказывания истинны, и ложна во всех остальных случаях. Например, предложение «$2 \cdot 2 = 4$

и Луна – спутник Земли» является истинным высказыванием, а предложение «$2 \cdot 2 = 4$ и в слове ХЛЕБ три буквы» – ложным.

Дизъюнкцией высказываний p, q называется высказывание «p или q», обозначаемое $p \vee q$, логическое значение которого устанавливается по табл. 4.

Таблица 4

p	q	$p \vee q$
0	0	0
0	1	1
1	0	1
1	1	1

Как видим, дизъюнкция ложна только в одном случае, когда оба составляющих ее высказывания ложны, и истинна во всех остальных случаях. Например, предложение «$2 \cdot 2 = 4$ или в слове ХЛЕБ три буквы» является истинным высказыванием, а предложение «$2 \cdot 2 = 5$ или Солнце – спутник Земли» – ложным.

Заметим, что союз «или», определяющий дизъюнкцию, не имеет разделительного смысла («либо – либо»), дизъюнкция двух истинных высказываний истинна.

Импликацией высказываний p, q называется высказывание «если p, то q», обозначаемое $p \to q$, логическое значение которого устанавливается по табл. 5.

Таблица 5

p	q	$p \to q$
0	0	1
0	1	1
1	0	0
1	1	1

Высказывание p называется посылкой импликации, а высказывание q – следствием (или заключением). Импликация ложна только в одном случае, когда ее посылка истинна, а следствие ложно, и истинна во всех остальных случаях. Например, предложения «если $2 \cdot 2 = 5$, то Луна – спутник Земли» и «если $2 \cdot 2 = 5$, то Солнце спутник Земли» оба считаются истинными, а предложение «если $2 \cdot 2 = 4$, то в слове ХЛЕБ три буквы» – ложным.

Здесь, конечно, требуются какие-то объяснения. В естественном языке логические связки соединяют высказывания, принадлежащие одному смысловому контексту. Поговорка «В огороде бузина, а в Киеве – дядя» издевается именно над нарушениями этого принципа. В математической логике от него приходится отказаться: операции должны быть определены для любой пары высказываний. Получаемые с помощью связок предложения могут оказаться лишенными непосредственного смысла, так что об их истинности или ложности судить невозможно. Таблицы, входящие в определения логических операций, позволяют приписать этим предложениям то или иное логическое значение, после чего они становятся высказываниями. Понятно, что таблицы эти согласованы с обыденной логикой в тех случаях, когда сложные предложения получаются содержательными.

Импликацию $p \to q$ читают также как «из p следует q». При таком понимании формальное определение этой операции еще дальше отходит от интуитивного представления о причинно-следственных связях, но и с этим приходится смириться, чтобы получить свободу в применении логических операций во всех случаях. Вспомним, что с подобными целями были введены отрицательные и мнимые числа в алгебре, бесконечно малые величины в анализе, бесконечно удаленные точки и прямые в проективной геометрии. Важно было избавиться от поминутных сомнений в применимости тех или иных действий в конкретных ситуациях. Компьютер, носитель искусственного интеллекта, не может рассуждать о том, на самом ли деле из $2 \cdot 2 = 5$ следует, что Луна (или Солнце) является спутником Земли, но, обращаясь к табл. 5, без колебаний признает истинность соответствующих импликаций.

Эквивалентностью высказываний p, q называется высказывание «p тогда и только тогда, когда q», обозначаемое $p \leftrightarrow q$, логическое значение которого устанавливается по табл. 6.

Таблица 6

p	q	$p \leftrightarrow q$
0	0	1
0	1	0
1	0	0
1	1	1

Таким образом, эквивалентность истинна в том случае, когда логические значения составляющих ее высказываний одинаковы, и ложна, когда они различны. Например, предложение «$2 \cdot 2 = 5$ тогда и только тогда, когда в слове ХЛЕБ три буквы» признается истинным высказыванием, а предложение «$2 \cdot 2 = 5$ тогда и только тогда, когда Луна – спутник Земли» – ложным. Разумеется, в содержательных математических текстах эквивалентность связывает высказывания одного смыслового поля. Запись $p \leftrightarrow q$ произносят также «p эквивалентно q», «p равносильно q» и т.п.

Простейшим логическим действием является отрицание данного высказывания. Введем соответствующую операцию, применяемую в отличие от предыдущих операций, к одному высказыванию.

Отрицанием высказывания p называется высказывание «неверно, что p», обозначаемое $\neg p$, логическое значение которого устанавливается по табл. 7.

Таблица 7

p	$\neg p$
0	1
1	0

Таким образом, отрицание истинного высказывания ложно, а отрицание ложного высказывания истинно. Запись $\neg p$ читают также «не p». Найдите простые формулировки для предложений, являющихся отрицаниями высказываний «Ни в одном отделе этого магазина нет дешевых товаров», «В одной из групп нашего курса нет отличников».

Множество всех высказываний, рассматриваемое вместе с операциями конъюнкции, дизъюнкции, импликации, эквивалентности и отрицания, образует алгебру высказываний.

Формула алгебры высказываний состоит из пропозициональных переменных, знаков логических операций, соединяющих эти переменные, и скобок, которые указывают последовательность выполнения операций. Например, выражение $(p_1 \wedge p_2) \vee p_3$ является формулой алгебры логики, а $p_1 \wedge p_2 \vee p_3$ – не формула, поскольку неясно, в какой последовательности применены операции: $(p_1 \wedge p_2) \vee p_3$ или $p_1 \wedge (p_2 \vee p_3)$. В формулах типа $(\neg p) \rightarrow q$ скобки обычно опускают и пишут $\neg p \rightarrow q$, т.е. считается, что знак отрицания, стоящий перед буквой, относится только к ней.

Пользуясь таблицами истинности, входящими в определения логических операций, можно вычислить логическое значение любой формулы при тех или иных конкретных логических значениях входящих в нее пропозициональных переменных. Рассмотрим, например, формулу $(p \wedge \neg q) \vee \neg(q \rightarrow p)$, обозначим ее через F. При данных конкретных значениях переменных p, q для вычисления соответствующего значения формулы F нужно осуществить 5 действий: 1) $\neg q$, 2) $p \wedge \neg q$, 3) $q \rightarrow p$, 4) $\neg(q \rightarrow p)$, 5) $(p \wedge \neg q) \vee \neg(q \rightarrow p)$. Так как любое высказывание принимает два логических значения 0 и 1, то в нашем случае имеется четыре возможных варианта: $p = 0$, $q = 0$; $p = 0$, $q = 1$; $p = 1$, $q = 0$; $p = 1$, $q = 1$. В каждом из них нужно выполнить пять логических операций в указанном выше порядке. Эти действия сведены в четырех последовательно заполняемых строчках таблицы истинности для формулы F (табл. 8).

Таблица 8

p	q	$\neg q$	$p \wedge \neg q$	$q \rightarrow p$	$\neg(q \rightarrow p)$	$F : (p \wedge \neg q) \vee \neg(q \rightarrow p)$
0	0	1	0	1	0	0
0	1	0	0	0	1	1
1	0	1	1	1	0	1
1	1	0	0	1	0	0

Мы видим, что формула F истинна, когда логические значения высказываний p и q различны, и ложна, когда они совпадают.

Две формулы F_1 и F_2 называются равносильными, если при каждом наборе значений участвующих в них переменных логическое значение формулы F_1 равно соответствующему логическому значению формулы F_2. В этом случае пишут $F_1 \cong F_2$. Сопоставляя определение эквивалентности высказываний p и q с тем заключением, которое было сделано по поводу рассмотренной выше формулы F, приходим к выводу, что F равносильна отрицанию эквивалентности, т.е. что $(p \wedge \neg q) \vee \neg(q \to p) \cong \neg(p \leftrightarrow q)$.

Докажите следующие равносильности: $p \vee q \cong \neg(\neg p \wedge \neg q)$, $p \to q \cong \neg(p \wedge \neg q)$, $p \leftrightarrow q \cong \neg(p \wedge \neg q) \wedge \neg(\neg p \wedge q)$ (т.е. все логические операции могут быть выражены через конъюнкцию и отрицание); $p \wedge q \cong \neg(\neg p \vee \neg q)$, $p \to q \cong \neg p \vee q$, $p \leftrightarrow q \cong \neg(\neg(\neg p \vee q) \vee \neg(p \vee \neg q))$ (т.е. все логические операции могут быть выражены через дизъюнкцию и отрицание); $p \wedge q \cong$ $\cong \neg(p \to \neg q)$, $p \vee q \cong \neg p \to q$, $p \leftrightarrow q \cong \neg((p \to q) \to \neg(q \to p))$ (т.е. все логические операции могут быть выражены через импликацию и отрицание). Из этих и подобных упражнений выросла мощная ветвь математической логики – теория булевых функций (мы соприкоснемся с ней в §VI.3).

Формула алгебры высказываний называется тождественно истинной, или тавтологией, если она истинна при любых значениях входящих в нее переменных. Очевидными примерами тавтологий являются формулы $p \leftrightarrow p$ и $p \vee \neg p$ (поясните простыми рассуждениями).

Тавтология – это формула, истинная в силу своей логической структуры, независимо от конкретного смысла составляющих ее высказываний. Заметим, что в обыденном понимании термин «тавтология» означает повторение слов или выражений, имеющих один и тот же смысл, и чаще всего носит неодобрительный оттенок.

Основные схемы умозаключений, используемые в научных и житейских доказательствах, весьма немногочисленны и имеют универсальный характер, т.е. не зависят от конкретного содержания суждений, к которым они применяются. В математической логике эти схемы выражаются тавтологиями, названия которых соответствуют названиям законов и правил классической аристотелевой логики. Приведем важнейшие из них.

I. Закон тождества $p \to p$.
«Истина истинна сама по себе».

II. Закон отрицания противоречия $\neg(p \wedge \neg p)$.
Невозможно, чтобы одновременно были истинными данное

высказывание и ему противоположное. Формула вида $p \wedge \neg p$, ложная при всех значениях p, называется противоречием.

III. Закон исключенного третьего $p \vee \neg p$.

Если выдвинут некий тезис p, то истинным будет либо он сам, либо его отрицание,– третьего не дано (лат. tertium non datur).

IV. Закон двойного отрицания $\neg\neg p \leftrightarrow p$.

Дважды отрицая, возвращаемся к исходному.

V. Закон контрапозиции $(p \to q) \to (\neg q \to \neg p)$.

В импликации $p \to q$ следствие q называют также необходимым условием для посылки p, а p – достаточным условием для q. Рассматриваемый закон утверждает, что если q – необходимое условие для p, то при нарушении q не выполняется и p. Пример: если число n делится на 6 (высказывание p), то оно делится на 3 (высказывание q); следовательно, если число n не делится на 3 (т.е. $\neg q$), то оно не делится на 6 (т.е. $\neg p$).

VI. Закон транзитивности импликации (или правило цепного заключения) $((p_1 \to p_2) \wedge (p_2 \to p_3)) \to (p_1 \to p_3)$.

Сложные доказательства состоят из многих этапов, результатом каждого из которых является утверждение, вытекающее из некоторых предшествующих ему. Как соотнести финальный вывод доказательства с исходной посылкой, содержащей условия теоремы? Возможность переноса истины шаг за шагом по цепочке импликаций и обеспечивает закон транзитивности.

VII. Закон равносильности $((p \to q) \wedge (q \to p)) \leftrightarrow (p \leftrightarrow q)$.

На этом законе основан один из широко используемых приемов обоснования равносильности двух утверждений: нужно показать, что каждое из них является следствием другого. Примеры: доказательство теоремы 1 из §I.5 о критерии равномощности конечных множеств, теоремы 1 из §II.1, где устанавливается равносильность свойств инъективности и сюръективности для преобразований конечного множества. Укажите другие примеры доказательств, где используется закон равносильности.

VIII. Законы двойственности $\neg(p \wedge q) \leftrightarrow (\neg p \vee \neg q)$, $\neg(p \vee q) \leftrightarrow (\neg p \wedge \neg q)$.

Интуитивно ясен смысл этих законов: «Неверно, что одновременно выполняются условия p и q» означает, что «не выполняется хотя бы одно из этих условий»; «неверно, что выполняется хотя бы одно из условий p,q» означает, что «не выполняется ни p, ни q». Законы двойственности называют также законами Де Моргана.

IX. Правило приведения к противоречию $(p \to (q \wedge \neg q)) \to \to \neg p$.

Если из некоторого утверждения p может быть получено противоречие, то p – ложно. Это правило лежит в основе метода доказательства от противного (см. доказательство теоремы Эвклида в §I.1 и комментарий к нему, приведите другие примеры). Латинское название правила – reductio ad absurdum.

X. Правило заключения $(p \wedge (p \to q)) \to q$.

Если истинно некоторое положение p и из него правильными рассуждениями выводится q, то истинно и q. Это основное правило логики доказательств, оно обосновывает принцип получения новых истин. Латинское название правила – modus ponens.

Докажите, что формулы I-X в самом деле являются тавтологиями.

Основоположником логики как науки является древнегреческий философ Аристотель (384–322 до н.э.). Его учение безраздельно господствовало вплоть до середины XIX века, когда английские математики Джордж Буль (1815–1864) и Огастес Де Морган (1806–1871) опубликовали первые работы, в которых пытались очистить классическую аристотелеву логику от схоластических наслоений и придать ей черты математической теории. Свой вклад в этот процесс внес и профессор математики Оксфордского университета священник Чарлз Латуидж Доджсон (1832–1898), более известный под литературным псевдонимом Льюис Кэррол. В частности, именно он ввел таблицы истинности для логических формул. Его книга «Алиса в стране чудес» (1865) является безусловным лидером по количеству цитирований в математике. Логика парадоксов, виртуозно продемонстрированная в речах и действиях героев этой сказочной повести, оказалась весьма притягательной для профессионалов, пользующихся логикой, исключающей парадоксы.

§ 2. Формализованное исчисление высказываний

Логический парадокс, изложенный в §1 в форме брадобрея, был открыт в 1903 году и произвел сильнейшее впечатление на математиков, заставив их обратиться к основаниям своей науки. В сложных доказательствах интуиция играет лишь второстепенную роль, на первое место выдвигается логика, так что неблагополучие в ней самой подрывает доверие к полученным с ее помощью результатам. Автором знаменитого парадокса был Бертран Артур Уильям Рассел (1872–1970), английский математик и философ, лауреат Нобелевской премии по литературе. Он считал, что математические теории могут быть сведены к формальной логике и что, избавив логику от парадоксов, удастся поставить на прочный фундамент все здание математики. В 1910–1913 годах Рассел и его коллега по Кембриджскому университету Уайтхед в трехтомном труде «Основания математики» («Principia Mathematica») разработали логический аппарат, который, по замыслу авторов, должен был свести всякое доказательство к цепочке стандартно осуществляемых действий, подчиненных строгим правилам.

В качестве примера рассмотрим формализованное исчисление высказываний, представляющее в подобном виде алгебру высказываний.

Первым этапом построения формальной теории является создание ее внутреннего языка, на котором будут вестись записи полученных в ней результатов.

Алфавит формализованного исчисления высказываний состоит из трех групп символов: 1) счетного множества букв $p_1,\ p_2,\ \ldots,\ p_n,\ \ldots$ (пропозициональные символы); 2) двух логических знаков $\neg$ и $\to$; 3) двух скобок (и).

Из символов алфавита можно составлять цепочки конечной длины – слова: $p_1\ \to)$, $((p_1\ \to\ p_2)\ \to\ \neg p_3)$, $((\to\ \neg\neg$ и т.д. Не все они представляют интерес, так же как не все буквосочетания естественного языка имеют смысл. В формальной теории роль осмысленных слов играют формулы. В нашем случае они определяются следующим образом: 1) каждый отдельно взятый пропозициональный символ является формулой; 2) если F – формула, то и $\neg F$ считается формулой; 3) если F_1 и F_2 – формулы, то $(F_1\ \to\ F_2)$ – также формула; 4) других формул нет. Последний пункт означает, что формулы получаются только по рецептам 1)-3).

Согласно приведеному определению, слово $((p_1 \rightarrow p_2) \rightarrow \rightarrow \neg p_3)$ будет формулой, а слово $p_1 \rightarrow)$ – нет. Договоримся для простоты опускать в записи формул внешние скобки.

Следующим этапом является выделение некоторых формул, которые будут считаться исходными для построения теории. Эти формулы называются аксиомами. В формализованном исчислении высказываний аксиом бесконечно много, но по своему виду они разбиваются на три класса. Аксиомой объявляется всякая формула, имеющая один из следующих видов:

А 1. $F_1 \rightarrow (F_2 \rightarrow F_1)$,

А 2. $(F_1 \rightarrow (F_2 \rightarrow F_3)) \rightarrow ((F_1 \rightarrow F_2) \rightarrow (F_1 \rightarrow F_3))$,

А 3. $(\neg F_1 \rightarrow \neg F_2) \rightarrow (F_2 \rightarrow F_1)$,

где $F_1,\ F_2,\ F_3$ – произвольные формулы.

При естественном понимании теории в ней из одних положений логическим путем выводятся другие, это обеспечивает ее развитие. Смысл всякой теории состоит не в перечислении раз навсегда заданных истин, но в возможности получения все новых и новых результатов.

В формализованном исчислении высказываний имеется только одно правило вывода – modus ponens (кратко: MP). Оно гласит, что если имеются формулы F_1 и $F_1 \rightarrow F_2$, то из них выводится формула F_2. Символически это записывается в виде дроби:

$$\frac{F_1, F_1 \rightarrow F_2}{F_2}.$$

При этом говорят, что формула F_2 является непосредственным следствием формул F_1 и $F_1 \rightarrow F_2$.

Далее, формула F называется доказуемой, или теоремой (запись: $\vdash F$), если существует конечная последовательность формул $F_1,\ F_2,\ \ldots,\ F_n$ такая, что 1) F_n совпадает с F; 2) каждая из формул $F_1,\ F_2,\ \ldots,\ F_n$ либо является аксиомой, либо получается по правилу MP из каких-то двух формул, предшествующих ей в данной последовательности. Последовательность $F_1,\ F_2,\ \ldots,\ F_n$ называют доказательством теоремы F.

Множество всех теорем, которые могут быть доказаны в формализованном исчислении высказываний, и составляет теорию этого исчисления.

Приведем пример целой серии теорем с их единообразным доказательством.

Теорема 1. $\vdash F \to F$, какова бы ни была формула F.

ДОКАЗАТЕЛЬСТВО. Рассмотрим последовательность формул:

1) $F \to ((F \to F) \to F)$,

2) $(F \to ((F \to F) \to F)) \to ((F \to (F \to F)) \to (F \to F))$,

3) $(F \to (F \to F)) \to (F \to F)$,

4) $F \to (F \to F)$,

5) $F \to F$.

Первая формула последовательности представляет собой аксиому серии А 1, где вместо F_1 подставлена формула F, а вместо F_2 – формула $F \to F$.

Вторая формула – это аксиома серии А 2, где вместо F_1 и F_3 подставлена формула F, а вместо F_2 – снова $F \to F$.

Третья формула является непосредственным следствием первой и второй формул, т.е. получена из них по правилу вывода MP.

Четвертая формула – аксиома серии А 1, где вместо F_1 и F_2 подставлена формула F.

Последняя формула получена по правилу MP из четвертой и третьей. Она совпадает с доказываемой формулой.

Согласно определению доказательства в формализованном исчислении высказываний, предложенная последовательность формул является доказательством формулы $F \to F$, а сама она, следовательно, будет теоремой при любом конкретном выборе формулы F. $\square$

Подобным образом доказываются все теоремы формализованного исчисления высказываний.

Но какое отношение эта формальная теория имеет к алгебре высказываний, построенной на основе классической логики? Чтобы ответить на этот вопрос, придадим следующее истолкование символам и формулам формальной теории. Пропозициональные символы будут считаться обозначениями высказываний, знаки $\neg$ и $\to$ знаками логических операций отрицания и импликации. Вместо формулы $\neg(F_1 \to \neg F_2)$ будем писать сокращенно $F_1 \wedge F_2$, для формулы $\neg F_1 \to F_2$ введем запись $F_1 \vee F_2$, наконец, длинную формулу $\neg((F_1 \to F_2) \to \neg(F_2 \to F_1))$ будем заменять на $F_1 \leftrightarrow F_2$. Новые знаки $\wedge$, $\vee$ и $\leftrightarrow$ назовем соответственно конъюнкцией, дизъюнкцией и эквивалентностью. Вспомним, что в §1 соответствующие логические операции мы именно так выразили через отрицание и импликацию.

Таким образом, все формулы формализованного исчисления высказываний имеют смысл в алгебре высказываний и, наоборот, каждая формула алгебры высказываний после замены в ней операций конъюнкции, дизъюнкции и эквивалентности их выражениями через отрицание и импликацию приобретет равносильный вид, воспринимаемый как формула формализованного исчисления высказываний. Следующая теорема, которую в 1921 году доказал американский логик Эмиль Леон Пост (1897–1954), завершает процесс установления соответствия между содержательной алгеброй высказываний и построенной формальной теорией.

Теорема 2. Формула алгебры высказываний тогда и только тогда тождественно истинна (т.е. является тавтологией), когда соответствующая ей формула формализованного исчисления высказываний доказуема (т.е. является теоремой). □

Теорема Поста называется теоремой о полноте формализованного исчисления высказываний: в этой формальной системе понятия истинности и доказуемости совпадают, всякое истинное утверждение имеет доказательство.

Из теоремы Поста следует, что формализованное исчисление высказываний непротиворечиво, т.е. что в нем нельзя доказать некоторую формулу F и доказать ее отрицание $\neg F$. В самом деле, F и $\neg F$ не могут быть одновременно тавтологиями алгебры высказываний, а значит, не могут одновременно быть доказуемыми в формальной теории.

Формализованное исчисление высказываний обладает еще одним важным свойством – разрешимостью. Его объясняет

Теорема 3. Существует эффективная процедура, позволяющая разрешить вопрос о доказуемости каждой предъявленной формулы формализованного исчисления высказываний.

ДОКАЗАТЕЛЬСТВО. Пусть F – некоторая формула формализованного исчисления высказываний и нужно установить, доказуема ли она. Согласно теореме 2 все доказуемые формулы являются тавтологиями содержательной алгебры высказываний. Составим таблицу истинности для формулы F. Поскольку, как в каждой формуле, в F входит конечное число пропозициональных переменных, то и строк в этой таблице будет конечное число, так что за конечное число шагов мы получим ответ на вопрос о том, будет ли формула F тавтологией, а следовательно, и на вопрос о ее доказуемости в формализованном исчислении высказываний. □

Если бы каждую математическую теорию можно было формализовать наподобие алгебры высказываний, это дало бы математикам принципиальную возможность представить свои результаты в виде, допускающем почти механическую проверку, избавило бы творческий процесс от угрозы возникновения внутренне противоречивых и парадоксальных конструкций. Для этого сначала нужно было в полном объеме изучить тот логический аппарат, который используется в математике и простейшей частью которого является алгебра высказываний.

Одноместным предикатом на данном множестве A называется функция, определенная на этом множестве, значениями которой являются высказывания о соответствующих значениях аргумента. Например, предикатами на множестве $\mathbb{N}$ натуральных чисел будут «x – простое число», «$x > 10$», «x при делении на 5 дает остаток 1» и т.п. При подстановке вместо x конкретных натуральных чисел эти предложения превратятся в высказывания. Так, при $x = 13$ первый и второй предикаты станут истинными высказываниями, а третий – ложным. Одноместные предикаты соответствуют свойствам, которыми в принципе могут обладать элементы данного множества.

Различные отношения между элементами множества A описываются многоместными предикатами на этом множестве, они содержат в своей формулировке несколько переменных и превращаются в высказывания при конкретизации всех аргументов. Двухместными предикатами на множестве $\mathbb{N}$ будут «$x < y$», «x и y дают одинаковый остаток при делении на 5» и т.п. При $x = 3$, $y = 7$ первый предикат станет истинным высказыванием, а второй – ложным. Укажите какие-нибудь значения переменных, при которых оба эти предиката стали бы истинными высказываниями, ложными высказываниями, первый ложным, а второй истинным высказыванием.

Предикат называется тождественно истинным, если при любых значениях входящих в него переменных он превращается в истинное высказывание. Таковы, например, рассматриваемые на множестве $\mathbb{R}$ всех действительных чисел предикаты $\sin^2 x + \cos^2 x = 1$ и $|x + y| \leq |x| + |y|$. Предикат $x^2 + 1 \neq 0$ будет тождественно истинным на множестве $\mathbb{R}$, но не на множестве $\mathbb{C}$ всех комплексных чисел (почему?).

Высказывания тоже считаются предикатами – нульместными. Предикат от переменных $x_1,\ x_2,\ \ldots,\ x_n$ обозначается через $P(x_1, x_2, \ldots, x_n)$.

Из предикатов, соединяя их логическими связками, получают новые предикаты: $\neg P(x)$, $P_1(x,y) \wedge P_2(x,y,z)$, $P_1 \vee P_2(y)$, $P_1 \rightarrow P_2(x)$, $P_1(x,y) \leftrightarrow P_2$.

Принципиально новым способом образования одних предикатов из других является применение кванторов.

Квантором общности называется функция, сопоставляющая каждому одноместному предикату $P(x)$ высказывание $(\forall x)(P(x))$ (читается: «для любого x выполняется $P(x)$»), истинное, если $P(x)$ истинно для любого $x \in A$, и ложное в противном случае, и каждому n-местному ($n > 1$) предикату $P(x_1, x_2, \ldots, x_n)$ сопоставляющая $(n-1)$-местный предикат $(\forall x_1)(P(x_1, x_2, \ldots, x_n))$, истинный при данных значениях $x_2, \ldots, x_n$, если $P(x_1, x_2, \ldots, x_n)$ истинно для любого $x_1 \in A$, и ложный в противном случае.

Квантором существования называется функция, сопоставляющая каждому одноместному предикату $P(x)$ высказывание $(\exists x)(P(x))$ (читается: «существует x такой, что выполняется $P(x)$»), истинное, если $P(x)$ истинно для некоторого $x \in A$, и ложное в противном случае, и каждому n-местному ($n > 1$) предикату $P(x_1, x_2, \ldots, x_n)$ сопоставляющая $(n-1)$-местный предикат $(\exists x_1)(P(x_1, x_2, \ldots, x_n))$, истинный при данных значениях $x_2, \ldots, x_n$, если $P(x_1, x_2, \ldots, x_n)$ истинно для некоторого $x_1 \in A$, и ложный в противном случае.

На множестве $\mathbb{Z}$ рассмотрим предикат $x > y$. Какими будут логические значения высказываний $(\forall x)(\exists y)(x > y)$ и $(\exists y)(\forall x)(x > y)$? Истинным или ложным будет высказывание $(\exists x)(x^2 + x - 1 = 0)$ на множестве $\mathbb{Z}$ целых чисел, на множестве $\mathbb{R}$ действительных чисел?

Множество всех предикатов, рассматриваемое вместе с логическими операциями и кванторами, образует алгебру предикатов.

Из предикатных символов (с переменными), логических связок, кванторов и скобок разумным образом составляются формулы алгебры предикатов. Если участвующие в формуле символы предикатов заменить конкретными предикатами на некотором множестве A, она превратится в предикат на этом множестве.

Формула алгебры предикатов называется тавтологией, если при всех конкретизациях предикатных символов, участвующих в ней, она превращается в тождественно истинный предикат.

В отличие от алгебры высказываний проверка тавтологичности формул алгебры предикатов в общем случае не сводится к механическим процедурам типа составления таблиц истинности. Приведем примеры тавтологий алгебры предикатов.

Так как любая тавтология алгебры высказываний является также и тавтологией алгебры предикатов, то справедлива

Теорема 4. Если формула F является тавтологией алгебры высказываний, то при замене в ней пропозициональных переменных произвольными предикатными символами получается тавтология алгебры предикатов. $\square$

Теорема 5. Следующие формулы («законы отрицания») являются тавтологиями алгебры предикатов:

$$\neg(\forall x)(P(x)) \leftrightarrow (\exists x)(\neg P(x)),$$

$$\neg(\exists x)(P(x)) \leftrightarrow (\forall x)(\neg P(x)).$$

ДОКАЗАТЕЛЬСТВО. Проведем доказательство для первой формулы. Пусть предикатный символ $P(x)$ интерпретирован как конкретный предикат на некотором множестве A. Он выражает некоторое свойство, имеющее смысл для элементов множества A. Высказывание $\neg(\forall x)(P(x))$ означает, что не все элементы множества A обладают этим свойством, а высказывание $(\exists x)(\neg P(x))$ означает, что по крайней мере один элемент множества A не обладает указанным свойством. Очевидно, что эти высказывания либо оба истинны, либо оба ложны, ибо по смыслу они означают одно и то же. Значит, эквивалентность $\neg(\forall x)(P(x)) \leftrightarrow (\exists x)(\neg P(x))$ истинна в рассматриваемой произвольной интерпретации. $\square$

Применяя теорему 5, докажите, что формулы $(\exists x)(P(x)) \leftrightarrow \leftrightarrow \neg(\forall x)(\neg P(x))$ и $(\forall x)(P(x)) \leftrightarrow \neg(\exists x)(\neg P(x))$ являются тавтологиями (они выражают кванторы друг через друга).

Проведите рассуждения, показывающие, что формула

$$(\exists y)(\forall x)(P(x, y)) \to (\forall x)(\exists y)(P(x, y))$$

является тавтологией алгебры предикатов, и найдите конкретный пример двухместного предиката на каком-либо числовом множестве, опровергающий обратную импликацию.

Алгебра предикатов является логическим базисом, на котором строятся все классические разделы математики.

Построение формализованного исчисления предикатов (ФИП), конечно, представляло более сложную задачу, чем формализация алгебры высказываний. Аналогично тому, как это делалось в формализованном исчислении высказываний, определяется язык формализованного исчисления предикатов, фиксируются аксиомы и правила вывода, вводится понятие доказуемости формулы. Доказуемые формулы называются теоремами, они образуют теорию ФИП. Теорему о полноте для ФИП в 1930 году доказал австрийский логик и философ Курт Гёдель (1906–1978).

Теорема 6. Формула алгебры предикатов тогда и только тогда тождественно истинна (т.е. является тавтологией), когда соответствующая ей формула формализованного исчисления предикатов доказуема (т.е. является теоремой). □

Таким образом, и в формализованном исчислении предикатов понятия истинности и доказуемости совпадают, всякое истинное утверждение имеет доказательство.

Отсюда вытекает непротиворечивость формализованного исчисления предикатов: в нем нельзя доказать некоторую формулу F и доказать ее отрицание $\neg F$.

Однако в отличие от формализованного исчисления высказываний свойство разрешимости в ФИП не имеет места. В 1936 году американский логик Алонсо Чёрч (1903–1995) получил следующий отрицательный результат.

Теорема 7. Не существует эффективной процедуры, позволяющей разрешить вопрос о доказуемости каждой предъявленной формулы формализованного исчисления предикатов. □

Следовательно, для выяснения вопроса об истинности утверждения, записанного в виде формулы алгебры предикатов, в общем случае нужно либо строить его доказательство, либо искать опровергающий пример (в алгебре высказываний достаточно составить для соответствующей формулы таблицу истинности). В математической практике обычно начинают с попыток найти такой контрпример. Сколько искусно построенных и, по видимости, убедительных доказательств превращалось в никому не нужную словесную шелуху перед лицом простого и всем понятного контрпримера! Никто из математиков не застрахован от ошибок, и долгое время после опубликования интересного и сложно доказанного результата его автор угнетается мыслью о возможной катастрофе. С предельной откровенностью об этой стороне жизни творческих математиков говорит «Краткое жизнеописание Л.С.Понтрягина, составленное им самим»: «Различного рода страхи, связанные с профессиональной работой, всегда преследовали и продолжают преследовать меня теперь. Каждое новое начинание вызывает тревогу, так как неясно, справлюсь ли я с ним. Незаконченная научная работа вызывает страх, что я вообще не сумею ее закончить и несколько лет тяжелого труда пропадут даром. Законченная научная работа вызывает страх тем, что в ней может обнаружиться ошибка. Все эти страхи перед возможной неудачей составляют тяжелую эмоциональную сторону профессиональной работы». Заметим, что это сказано одним из самых выдающихся математиков.

§ 3. Аксиоматический метод

С созданием формализованного исчисления предикатов математика получила в свое распоряжение логику, свободную от парадоксов и противоречий, свойственных интуитивной логике здравого смысла. Но чтобы эффективно применить этот аппарат в некоторой конкретной области, нужно, очевидно, чтобы и те математические конструкции, с которыми придется работать, в свою очередь были надежными и не таили в себе потенциальных угроз, могущих проявиться на том или ином этапе развития теории. Средством избежать подобных неприятных ситуаций представлялся аксиоматический метод построения теорий, согласно которому выделяются некоторые исходные положения, называемые аксиомами, а все остальные предложения теории получаются из них как логические следствия.

Аксиоматический метод впервые в полном объеме был продемонстрирован Эвклидом в 13 книгах его трактата «Начала», где вся геометрия выводилась из пяти аксиом и пяти постулатов. О постулатах, носивших геометрический характер, мы говорили в §III.4, аксиомы, по-видимому, представлялись Эвклиду общенаучными положениями: «I. Равные одному и тому же равны между собой; II. И если к равным прибавляются равные, то и целые будут равны; III. И если от равных отнимаются равные, то остатки будут равны; IV. И совмещающиеся друг с другом равны между собой; V. И целое больше части». (Противоречие с последней аксиомой привело в замешательство Галилея, обнаружившего, что квадратов натуральных чисел столько же, сколько всех натуральных чисел,– см. §I.5.)

Первое предложение первой книги «Начал» описывает построение равностороннего треугольника с данной стороной, последнее предложение последней книги завершает теорию правильных многогранников: «Вот я утверждаю, что кроме упомянутых пяти тел нельзя построить другого тела, заключенного между равносторонними равноугольными равными друг другу многоугольниками». Самоотверженный читатель, пробившийся сквозь все тринадцать книг, должен был испытывать глубокое эстетическое удовлетворение от такого финала: он постиг гармонию, с которой устроен мир.

Теорема о бесконечности множества простых чисел – это предложение 20 в девятой книге, теорема Пифагора и обратная ей («если в треугольнике квадрат на одной стороне равен вместе взятым квадратам на остальных сторонах, то заключенный между остальными сторонами треугольника угол есть прямой») имеют номера 47 и 48 в первой книге и завершают ее. Они же завершали и математическое образование бакалавров средневековых университетов. Pons asinorum – «мост ослов» – так называли теорему Пифагора школяры времен «Декамерона», для которых она была камнем преткновения. Сейчас смысл названия не вполне ясен, но, согласимся, есть в нем что-то привлекательное.

Совершенство «Начал» на два тысячелетия затормозило развитие геометрии: до Декарта и Ньютона геометры были лишь комментаторами Эвклида (подобно тому, как современные им логики вели бесконечные схоластические дискуссии на темы Аристотеля). До недавнего времени во всех школах мира геометрию изучали по учебникам, восходящим к труду Эвклида. «Это удивительнейшее произведение мысли дало человеческому разуму ту уверенность в себе, которая была необходима для его последующей деятельности. Тот не рожден для теоретических исследований, кто в молодости не восхищался этим творением»,– писал о «Началах» создатель теории относительности Альберт Эйнштейн.

По образцу великого сочинения Эвклида были написаны «Математические начала натуральной философии» («Philosophiae Naturalis Principia Mathematica», 1687) Ньютона, где предлагалось аксиоматическое построение механики, и уже упоминавшийся трехтомник Рассела и Уайтхеда «Principia Mathematica» (тоже «Начала»), в котором аксиоматизировались логика и теория множеств. В 1939 году в Париже вышла первая книга гигантского трактата «Элементы математики» (или «Начала математики» – и так можно перевести французское название «Éléments de Mathématique»), автором которого обозначен Николай Бурбаки. В нескольких десятках томов, появившихся с тех пор, на строгой и единой аксиоматической основе излагаются результаты, полученные в различных областях математики. Эта выдающаяся справочная библиотека энциклопедического характера стала неотъемлемой частью современной математической культуры. Способ изложения и общая идеология Бурбаки оказали существенное влияние на развитие математики второй половины XX века, стимулировали перестройку ее школьного преподавания. Понятно, что одному человеку осуществление подобного проекта было бы не под силу. «Николай Бурбаки» – это коллективный псевдоним, за которым скрывается группа ведущих французских математиков, меняющаяся по своему составу (например, в связи с обязательным выходом из нее участников, достигших 50 лет). Если в 13 книгах «Начал» Эвклида подводился итог предыдущего развития всей математики, то «Начала» Бурбаки, несмотря на свою грандиозность, конечно, не могут претендовать на соответствующую роль в современной науке, – как многократно повторял Козьма Прутков, другой коллективный псевдоним: «Никто не обнимет необъятного!» Открытие неэвклидовой геометрии Н.И.Лобачевским и Бойяи поставило не возникавшую до того проблему непротиворечивости математических теорий. Эвклидова геометрия на протяжении тысячелетий воспринималась как учение о пространственных свойствах физического мира. «Воображаемая» геометрия поначалу представлялась игрой чистого разума, но после обнаружения конкретных пространств, для которых она оказалась внутренней геометрией (например, псевдосфера), и в ходе переосмысления геометрической структуры реального пространства она тоже стала претендовать на роль его математической модели. Таким образом, аксиоматическая теория «Начал» утратила свою извечную «земную» основу, ту естественную модель, которая гарантировала экспериментальную проверку всем логически выведенным утверждениям. Чем же теперь можно было мотивировать непротиворечивость эвклидовой геометрии? В более широком смысле: как вообще можно доказать непротиворечивость той или иной аксиоматической теории, обеспечить уверенность в том, что, логически обосновав некоторое предложение p, мы не найдем доказательства и противоположного ему утверждения $\neg p$? В противоречивой теории все доказуемо, и потому она бессмысленна.

Предположим, что нам удастся построить модель интересующей нас теории T в рамках некоторой другой теории T'. Это означает, что все исходные понятия и связи теории T истолковываются как некоторые понятия и связи теории T', после чего теоремы теории T «переводятся» на язык теории T'. Тогда если в T будет получено противоречие, то ему будет соответствовать и противоречие в T'. Значит, если теория T' непротиворечива, то непротиворечивой будет интерпретированная в ней теория T. Например, внутренняя геометрия плоскости Лобачевского может быть истолкована как внутренняя геометрия псевдосферы – поверхности в эвклидовом пространстве. Если эвклидова геометрия непротиворечива, то непротиворечива и внутренняя геометрия псевдосферы, и, следовательно, непротиворечива внутренняя геометрия плоскости Лобачевского. Таким способом и была доказана относительная непротиворечивость геометрии Лобачевского: она непротиворечива, если непротиворечива геометрия Эвклида. Но этот последний факт в свою очередь требует доказательства – мы не можем ссылаться на эвклидовость физического пространства.

В конце XIX века была построена модель эвклидовой геометрии внутри арифметики. Арифметические модели вскоре после этого были предложены и для других разделов математики. По существу, к вопросу о непротиворечивости арифметики свелась проблема непротиворечивости всей классической математики. Следовательно, непротиворечивость арифметики должна была быть доказана без моделирования в других системах.

В знаменитом докладе Гильберта на II Международном математическом конгрессе (Париж, 1900) проблема непротиворечивости арифметики стоит второй после канторовской проблемы континуума. (Давид Гильберт (1862–1943), профессор Геттингенского университета, оказал большое влияние на развитие математики. Ему принадлежат важные открытия в алгебре и анализе, полный список аксиом эвклидовой геометрии, исследования в области математической физики и оснований математики. Он был, по-видимому, последним из великих математиков, которые видели все величественное здание своей науки в целом.)

Система аксиом для арифметики была предложена в 1891 году итальянским математиком Джузеппе Пеано (1858–1932). В ней множество $\mathbb{N}$ всех натуральных чисел с присоединенным нулем 0 рассматривается вместе с тремя операциями: сложением, умножением и «следованием». Эта последняя операция означает переход от натурального числа n к непосредственно следующему за ним натуральному числу n' (например, $0' = 1$, $1' = 2$, $99' = 100$ и т.д.). В качестве аксиом принимаются следующие предложения:

I. $(\forall x)(\forall y)(x' = y' \to x = y)$,

II. $(\forall x)(\neg(x' = 0))$,

III. $(\forall x)(x + 0 = x)$,

IV. $(\forall x)(\forall y)(x + y' = (x + y)')$,

V. $(\forall x)(x \cdot 0 = 0)$,

VI. $(\forall x)(\forall y)(x \cdot y' = x \cdot y + x)$,

VII. $(P(0) \wedge (\forall x)(P(x) \to P(x'))) \to (\forall y)(P(y))$,
каково бы ни было свойство P.

Аксиомы I-VI достаточно просты и интуитивно понятны, если под x' подразумевать $x + 1$. Аксиома VII представляет собой на самом деле бесконечный список утверждений – для каждого конкретного свойства (одноместного предиката) P отдельно. Запишем ее словами: «если свойством P обладает число 0 и для любого числа x из того, что x обладает свойством P, следует, что и число x' обладает этим свойством, то всякое число обладает свойством P». Аксиома VII называется принципом полной индукции.

Аксиомы Пеано в сочетании с формализованным исчислением предикатов образуют формальную арифметику. Все ее предложения записываются в виде формул, т.е. «слов», составленных по специальным правилам из символов переменных, знаков арифметических и логических операций, знака равенства и скобок. Всякое доказательство представляет собой последовательность формул, образованную по правилам формальной логики и завершающуюся формулой, выражающей доказываемое предложение (как в формализованном исчислении высказываний). Если арифметика противоречива, то в ней можно доказать некоторое утверждение p вместе с противоположным ему утверждением $\neg p$, т.е. в качестве теоремы получить противоречие $p \wedge \neg p$, из которого следует все что угодно, например, $1 = 0$. Таким образом, чтобы доказать непротиворечивость арифметики, достаточно убедиться в том, что в ней невозможно вывести равенство $1 = 0$. (Можно представить себе, какой казуистикой выглядят все эти рассуждения с позиций обыденного разума.)

Программа обоснования математики, предложенная Гильбертом, состояла в том, чтобы все разделы этой науки формализовать по образцу арифметики. Это не отрицает творческой свободы исследователя в создании новых математических объектов и не отрицает роли интуиции в доказательствах. Но каждый раз, завершив свой вдохновенный труд, математик в идеале должен был перевести содержательные рассуждения в бессловесные последовательности формул или, по крайней мере, убедить себя и других, что это можно сделать.

Перспектива преобразования математики по рецептам гильбертовского формализма вызвала большой энтузиазм у одних ученых и столь же сильное неприятие со стороны других. Всем хотелось получить некоторый универсальный метод проверки правильности доказательств, но пугала угроза превращения математики в бессодержательную игру с цепочками символов.

Доказательство полноты, а следовательно, и непротиворечивости формализованного исчисления предикатов, полученное Гёделем, было весомым вкладом в осуществление программы Гильберта. Но уже через год этим планам был нанесен роковой удар – была опубликована знаменитая теорема Гёделя о неполноте. Она разбивается на следующие два утверждения.

Теорема 1. Если непротиворечивая формальная система S содержит в качестве своей части арифметику, то в этой системе существуют предложения F такие, что ни само F, ни его отрицание $\neg F$ не являются доказуемыми в S. $\square$

Поскольку одно из предложений F, $\neg F$ является истинным, теорема 1 означает, что если, согласно идее Гильберта, и удастся превратить математику в формальную систему, то в ней найдутся истинные утверждения, которые невозможно будет доказать. Как тут не вспомнить вдохновенные слова Гильберта из его доклада: «Убеждение в разрешимости каждой математической проблемы является для нас большим подспорьем в работе; мы слышим внутри себя постоянный призыв: вот проблема, ищи решение. Ты можешь найти его с помощью чистого мышления; ибо в математике не существует Ignorabimus!» (Последнее латинское слово означает «не будем знать».)

Прежде чем сформулировать вторую часть теоремы Гёделя, заметим, что ему удалось представить в виде арифметической формулы утверждение о непротиворечивости упоминаемой в теореме 1 системы S.

Теорема 2. Если непротиворечивая формальная система S содержит в качестве своей части арифметику, то предложение F о том, что S непротиворечива, не доказуемо в S. $\square$

Таким образом, если, согласно идее Гильберта, и удастся превратить математику в формальную систему, то средствами этой системы невозможно будет установить, является ли она непротиворечивой. Одновременно как частный случай получается, что непротиворечивость арифметики нельзя доказать непосредственно в ней самой.

Результаты Гёделя рассматриваются как высшее достижение логики (впрочем, по свидетельству Бурбаки, на конференции в Принстоне один известный ученый в присутствии Гёделя сказал, что «в логике не было сделано ничего нового со времен Аристотеля»).

Как и неэвклидова геометрия Лобачевского, теорема Гёделя о неполноте стала предметом многочисленных нематематических суждений. Оба эти открытия оказали огромное влияние на философию естествознания. С Лобачевским музыковеды сравнивали композитора Арнольда Шёнберга (1874–1951), в противовес традиционной тональной системе музыки предложившего новый метод композиции – додекафонию. С Гёделем литературными критиками сопоставляется писатель Франц Кафка (1883–1924), загадочная жизнь героев которого ставит под сомнение естественную логику.

Теорема Гёделя о неполноте подорвала всеобщую веру в аксиоматический метод как универсальный способ построения математических теорий. Но было бы нелепо отказываться от столь мощного средства только потому, что оно не идеально. Математики по-прежнему работают с аксиоматическими теориями, и деятельность Бурбаки – лучшее подтверждение жизненности идей Эвклида.

Аксиоматический метод неоднократно пытались применить и в гуманитарных науках. Философ Спиноза (1632–1677) таким образом строил свою «Этику», филолог А.Ф.Лосев (1893–1988) предложил своеобразную систему аксиом для знаковой теории языка. Очевидным примером учения, основанного на аксиомах, является догматическое богословие. О трудностях, возникающих при аксиоматизации, хорошо сказано у А.Ф.Лосева: «Аксиоматика дается математикам легче, чем гуманитариям, потому, что они оперируют понятиями, раз навсегда обладающими определенным и бесспорным содержанием. Таблица умножения, пользуемся ли мы ею в простейших расчетах или космических полетах, все равно остается одной и той же, единообразной и не допускающей никаких сомнений или споров. Но возьмем любое понятие исторических дисциплин – и мы встретимся здесь с множеством разных подходов, разных оттенков мысли, разных субъективных тенденций, вообще с такими трудностями, разрешить которые весьма нелегко, а тем не менее без их разрешения невозможна никакая историография как точная дисциплина. Поэтому при современном состоянии гуманитарных наук, а в том числе и языкознания, приходится совершенно отказываться от аксиоматики этих дисциплин в точном смысле слова... Однако совершенно необходимо предпринимать посильные шаги в этом направлении».

ГЛАВА V. **МАТЕМАТИКА НЕОПРЕДЕЛЕННОГО**

§ 1. Алгебра множеств

Понятие множества лежит в основе современной математики. Оно считается интуитивно ясным и как таковое принимается без определения (мы свободно пользовались им в предыдущих главах). Всякая попытка дать точное математическое описание привела бы к появлению объясняющих терминов типа «совокупность», «собрание» и т.п., которые в свою очередь пришлось бы признать общепонятными и потому неопределяемыми.

Множества в общем случае обозначаются прописными латинскими буквами: A, B, C и т.д. Множества состоят из элементов. Тот факт, что объект a является элементом множества A, записывают как $a \in A$ («a – элемент множества A», «a принадлежит множеству A» и т.п.). Символ принадлежности $\in$ – это стилизованная греческая буква ε, первая в слове $\varepsilon\lambda\varepsilon\mu\varepsilon\nu\tau o\nu$. Если a не принадлежит множеству A, пишут $a \notin A$. Пустое множество – так называется множество, не содержащее ни одного элемента, – обозначается символом $\emptyset$.

Два множества A и B считаются равными, если у них одни и те же элементы. Таким образом, $A = B$ тогда и только тогда, когда истинно высказывание

$$(\forall x)(x \in A \leftrightarrow x \in B).$$

Говорят, что множество A является подмножеством множества B или что A включается в B, если каждый элемент множества A содержится в B. Символически это записывается в виде $A \subseteq B$. Таким образом, $A \subseteq B$ тогда и только тогда, когда истинно высказывание

$$(\forall x)(x \in A \rightarrow x \in B).$$

Подмножествами произвольного множества A являются, например, $\emptyset$ и A. Подмножества, отличные от A, называются собственными подмножествами множества A. Совокупность всех подмножеств множества A обозначается через $P(A)$. Например, если $A = \{1,\ 2,\ 3\}$, то в состав $P(A)$ в качестве элементов входят подмножества $\emptyset$, $\{1\}$, $\{2\}$, $\{3\}$, $\{1,\ 2\}$, $\{1,\ 3\}$, $\{2,\ 3\}$, A (понятно, что $\{1\}$ обозначает подмножество множества A, состоящее только из элемента 1).

Для наглядного изображения множеств используют так называемые диаграммы Венна (Джон Венн (1834–1923), профессор математики и философии в Кембридже). Пусть Ω – фиксированное непустое множество («универсум»). На рис. 64 представлены диаграммы Венна для Ω, его пустого подмножества $\emptyset$ и произвольного собственного непустого подмножества A. Здесь прямоугольник символизирует универсум Ω, а под ним указывается то подмножество, которое выделено в Ω штриховкой.

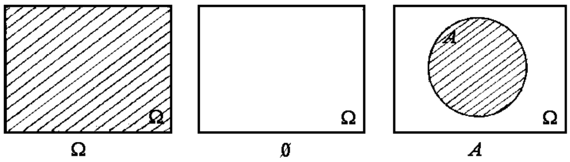


Рис. 64

Определим теперь некоторые действия (операции) над множествами. Пусть A и B – подмножества множества Ω, т.е. $A, B \subseteq \Omega$ или $A, B \in P(A)$. Пересечением подмножеств A и B называется подмножество, обозначаемое $A \cap B$, которое состоит из элементов множества Ω, одновременно входящих в A и в B. Диаграмма для пересечения представлена на рис. 65.

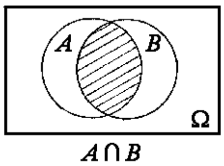


Рис. 65

Используя символ классификатора и логическую операцию конъюнкции, пересечение можно представить в следующей форме:

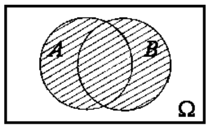
$$A \cap B = \{\, x \in \Omega \mid x \in A \wedge x \in B \,\}.$$

Если $A \cap B = \emptyset$, то говорят, что подмножества A и B не пересекаются. В этом случае круги, изображающие A и B на диаграмме, не имеют общих точек.

Объединением подмножеств A и B называется подмножество, обозначаемое $A \cup B$, которое состоит из элементов множества Ω, входящих хотя бы в одно из подмножеств A или B. Диаграмма для объединения показана на рис. 66. Используя символ классификатора и логическую операцию дизъюнкции, объединение можно представить в следующей форме:

$$A \cup B = \{\, x \in \Omega \mid x \in A \vee x \in B \,\}.$$

Если $A \subseteq \Omega$, то дополнением подмножества A называется подмножество, обозначаемое $\bar{A}$, которое состоит из элементов множества Ω, не входящих в A. Диаграмма для дополнения изображена на рис. 67.

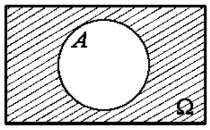


$A \cup B$

Рис. 66

Используя символ классификатора и логическую операцию отрицания, дополнение можно представить в следующей форме:

$$\bar{A} = \{\, x \in \Omega \mid \neg(x \in A) \,\}.$$



$\bar{A}$

Рис. 67

Заметим, что вместо $\neg(x \in A)$ обычно пишут $x \notin A$.

Как видим, соотношения между множествами и операции над ними оказываются тесно связанными с логическими операциями.

Исследование свойств множеств и определенных для них операций пересечения, объединения и дополнения составляет предмет алгебры множеств. Ее создатель – Джордж Буль (в соответствии с традицией отметим, что он был самоучкой). В книге «Исследование законов мышления» (1854) он разработал основы символической логики и изучал связанные с ней системы, называемые теперь булевыми алгебрами. С развитием электронной вычислительной техники и соответствующих разделов математики идеи Буля приобрели новый смысл, и теперь он считается одним из классиков науки.

Важнейшим примером булевой алгебры является совокупность всех подмножеств некоторого непустого множества, рассматриваемая вместе с указанными выше тремя операциями над подмножествами. Рассмотрим некоторые важнейшие тождества алгебры множеств.

Законы коммутативности:

$$A \cap B = B \cap A, \quad A \cup B = B \cup A.$$

Это аналоги арифметических правил: «от перемены мест слагаемых (сомножителей) сумма (произведение) не меняется».

Законы ассоциативности:

$$A \cap (B \cap C) = (A \cap B) \cap C, \quad A \cup (B \cup C) = (A \cup B) \cup C).$$

Свойством ассоциативности (сочетательности) обладают сложение и умножение чисел, ассоциативна операция в любой группе. В данном случае мы получаем, например, возможность писать без скобок, обозначающих порядок действий: $A \cap B \cap C$, понимая, что речь идет об общей части трех множеств A, B, C.

Большую роль в задачах типа «упростить выражение» играют *законы поглощения*

$$A \cap (A \cup B) = A, \quad A \cup (A \cap B) = A.$$

Например, выражение $A \cap (A \cup (A \cap A))$ согласно первому закону поглощения (если вместо B подставить $A \cap A$) равно A, и оно же, если в нем, в силу второго закона поглощения, заменить $A \cup (A \cap A)$ на A, дает $A \cap A$. Таким образом, мы получаем новое тождество $A \cap A = A$. Аналогично $A \cup A = A$ (проведите соответствующие рассуждения). Это так называемые *законы идемпотентности*.

Не вызывает сомнений истинность правил действий с пустым множеством $\emptyset$ и универсумом Ω, именно

$$A \cap \emptyset = \emptyset, \quad A \cup \emptyset = A, \quad A \cap \Omega = A, \quad A \cup \Omega = \Omega$$

для любого $A \subseteq \Omega$.

Законы дистрибутивности:

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C), \quad A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$$

Из арифметики известно, что умножение чисел дистрибутивно (распределительно) относительно сложения: $a(b + c) = ab + ac$, но не наоборот. В алгебре множеств операции пересечения и объединения в этом смысле равноправны.

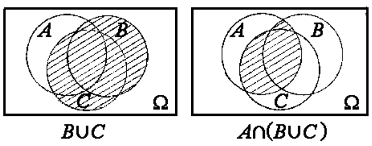
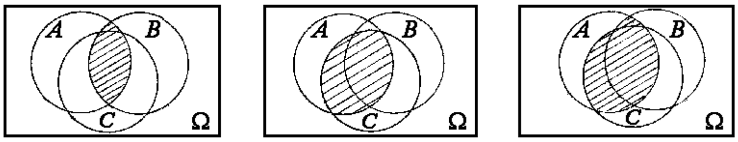


Рис. 68

Доказательства тождеств в алгебре множеств можно проводить методом диаграмм. Последовательно (в порядке производимых операций) строятся диаграммы Венна для левой и правой частей доказываемого тождества. Если финальные диаграммы совпали – тождество справедливо. При этом на каждой диаграмме должны быть представлены все участвующие в тождестве множества. Установим, например, истинность первого закона дистрибутивности $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$. Две диаграммы для левой части изображены на рис. 68. Для правой части придется строить три диаграммы (рис. 69).



$A \cap B$ $\qquad$ $A \cap C$ $\qquad$ $(A \cap B) \cup (A \cap C)$

Рис. 69

Как видим, финальные диаграммы совпали – тождество доказано.

Глядя на диаграмму для дополнения (см. рис. 67), непосредственно убеждаемся в истинности тождеств $A \cap \bar{A} = \emptyset$, $A \cup \bar{A} = \Omega$, а также $\bar{\bar{A}} = A$ (*закон двойного дополнения*).

Законы Де Моргана:

$$\overline{A \cap B} = \bar{A} \cup \bar{B}, \quad \overline{A \cup B} = \bar{A} \cap \bar{B}$$

(на языке элементов: если элемент не входит в общую часть A и B, то он не входит в A или не входит в B; если элемент не входит хотя бы в одно из множеств A, B, то он не входит ни в A, ни в B). Де Морган был первым президентом Лондонского математического общества.

Рассмотренные тождества являются основными законами алгебры множеств. В соответствующей записи они выполняются в любой булевой алгебре. С ними мы еще раз встретимся в §VI.3, где будем знакомиться с двоичной булевой алгеброй, лежащей в основе компьютерной логики.

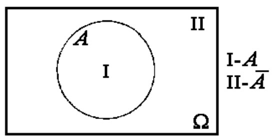
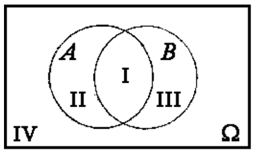


Рис. 70

Теоретико-множественные операции применяются при описании различных классификаций. Пусть для элементов множества Ω имеет смысл обсуждать некоторое свойство (признак) A. По отношению к этому свойству все элементы множества Ω разбиваются на два класса: один образуется элементами, обладающими свойством A (обозначим это подмножество тоже буквой A), а другой объединяет элементы, не имеющие данного свойства (очевидно, это будет дополнение $\bar{A}$). Таким образом, классификация элементов множества Ω по одному свойству A имеет вид, показанный на рис. 70. При этом $\mathrm{I} \cap \mathrm{II} = \emptyset$, $\quad \mathrm{I} \cup \mathrm{II} = \Omega$.

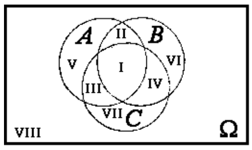
Если мы обсуждаем для элементов множества Ω два независимых свойства A и B, то по отношению к ним Ω разобьется на четыре класса: I – элементы, обладающие обоими свойствами A и B; II – элементы, имеющие только свойство A; III – элементы, имеющие только свойство B; IV – элементы, не обладающие ни одним из указанных свойств. Так что классификацию элементов множества Ω по двум свойствам A и B в самом общем случае иллюстрирует рис. 71.



I - $A \cap B$
II - $A \cap \bar{B}$
III - $\bar{A} \cap B$
IV - $\bar{A} \cap \bar{B}$

Рис. 71

При наличии трех независимых свойств A, B, C множество Ω аналогичным образом подразделяется на 8 классов в соответствии со схемой рис.72.

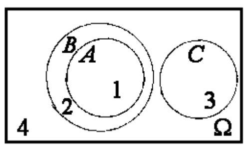


I - $A \cap B \cap C$
II - $A \cap B \cap \bar{C}$
III - $A \cap \bar{B} \cap C$
IV - $\bar{A} \cap B \cap C$
V - $A \cap \bar{B} \cap \bar{C}$
VI - $\bar{A} \cap B \cap \bar{C}$
VII - $\bar{A} \cap \bar{B} \cap C$
VIII - $\bar{A} \cap \bar{B} \cap \bar{C}$

Рис. 72

Если вообразить себе подмножество A окрашенным в красный цвет (т.е. если свойство A означает «быть красным»), B – в желтый, а C – в синий, то классу I будет соответствовать коричневый цвет, II – оранжевый, III – фиолетовый, IV – зеленый, V – красный, VI – желтый, VII – синий, VIII – белый.

При различных зависимостях между свойствами (это выражается специфическим расположением представляющих их областей на диаграмме Венна) соответствующие классификации могут быть неполными в том смысле, что некоторые классы окажутся пустыми. Например, если наличие свойства A обязательно влечет за собой и свойство B, а свойство C не совместимо с B, то классификационная схема в этом случае имеет вид рис. 73.



1 - A (II)
2 - $\bar{A} \cap B$ (VI)
3 - C (VII)
4 - $\bar{B} \cap \bar{C}$ (VIII)

Рис. 73

В скобках указаны соответствующие классы полной классификации по трем признакам, причем ее области I, III, IV, V оказались пустыми. Понятно, что класс 1 имеет оранжевый цвет, 2 – желтый, 3 – синий, 4 – белый.

Рассмотрите другие неполные классификации по трем признакам и представьте себе их цветные диаграммы. С помощью диаграммы Венна изобразите условное взаиморасположение следующих множеств чисел: натуральные, целые, рациональные, иррациональные, алгебраические, трансцендентные, положительные, отрицательные, действительные, чисто мнимые, комплексные. На сколько классов разобьет множество всех комплексных чисел их классификация по указанным 11 свойствам? В каждом классе выберите по одному представителю. Обратите внимание на число 0.

§ 2. Комбинаторика

Часто говорят, что если гуманитариям и нужна какая-нибудь математика, то это непременно должна быть «математика без чисел». Конечно, разумное начало в подобных суждениях есть, но тем не менее всем нам приходится считать и не всегда при этом удается обойтись простой арифметикой. Рассмотрим несколько формул, полезных для вычисления простейших вероятностей, при обработке результатов наблюдений, да и в обычном житейском общении с числами.

Раздел математики, занимающийся подсчетом количества элементов в конечных множествах, называется комбинаторикой.

Через $n(A)$ обозначим количество элементов в конечном множестве A. Например, если $A = \{$ север, юг, восток, запад $\}$, то $n(A) = 4$.

Пусть A и B – конечные множества. Подсчитаем $n(A \cup B)$ – количество элементов в объединении $A \cup B$. В это число входят все элементы из A (их $n(A)$ штук) и все элементы из B (их $n(B)$). Но при этом элементы, находящиеся в пересечении $A \cap B$, учитываются дважды: в составе A и в составе B. Значит,

$$n(A \cup B) = n(A) + n(B) - n(A \cap B). \tag{1}$$

Если множества A и B не пересекаются, т.е. $A \cap B = \emptyset$, то $n(A \cap B) = n(\emptyset) = 0$ и, значит, $n(A \cup B) = n(A) + n(B)$. Выпишем этот важный частный случай формулы (1) отдельно:

$$A \cap B = \emptyset \to n(A \cup B) = n(A) + n(B). \tag{2}$$

Правило (2) применяется для подсчета числа элементов при классификациях.

ПРИМЕР. 50 студентов первого курса в течение семестра выполнили три контрольные работы. По первой получили зачет 34 человека, по второй – 31 и по третьей – 30. При этом первую и вторую работы успешно выполнили 24 студента, первую и третью – 21, вторую и третью – 17, все три – 14 студентов. Сколько студентов не получили зачета ни по одной работе? Сколько человек справились только с одной работой? Точно с двумя работами?

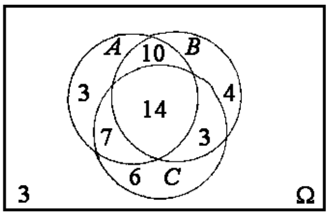


Рис. 74

Прежде всего проведем классификацию студентов первого курса по трем признакам: A – успех при выполнении первой работы, B – второй, C – третьей (см. рис. 72). Согласно условию, класс I содержит 14 элементов. Далее, $n(A \cap B) = 24$. Но $A \cap B = \mathrm{I} \cup \mathrm{II}$, причем $\mathrm{I} \cap \mathrm{II} = \emptyset$. Следовательно, по формуле (2), $24 = n(\mathrm{I}) + n(\mathrm{II})$, откуда $n(\mathrm{II}) = 10$. Аналогично, $21 = n(A \cap C) = n(\mathrm{I}) + n(\mathrm{III})$, откуда $n(\mathrm{III}) = 7$, и $17 = n(B \cap C) = n(\mathrm{I}) + n(\mathrm{IV})$ и, значит, $n(\mathrm{IV}) = 3$. В множестве A 34 элемента, причем $34 = n(\mathrm{I} \cup \mathrm{II} \cup \mathrm{III} \cup \mathrm{V}) = n(\mathrm{I}) + n(\mathrm{II}) + n(\mathrm{III}) + n(\mathrm{V})$, откуда $n(\mathrm{V}) = 3$. Аналогично, $n(\mathrm{VI}) = 4$ и $n(\mathrm{VII}) = 6$. Так как всего в множестве Ω студентов первого курса 50 элементов, а в класах I-VII содержится 14+10+7+3+3+4+6=47, то $n(\mathrm{VIII}) = 3$, и мы получаем сводку о количественном составе классов, в виде диаграммы на рис. 74.

Теперь можно ответить на вопросы: 3 студента не получили зачета ни по одной из работ (класс VIII), только с одной работой справились 3+4+6=13 студентов (классы V, VI, VII), точно две работы зачтены 10+7+3=20 студентам (классы II, III, IV).

Через $A \times B$ обозначим множество всевозможных пар вида (a, b), где $a \in A$, $b \in B$. Если $A = \{a_1, a_2, \ldots, a_m\}$, $B = \{b_1, b_2, \ldots, b_n\}$, то все элементы множества $A \times B$ («каждый из A с каждым из B») можно перечислить в следующей таблице:

$$\begin{matrix} (a_1, b_1) & (a_1, b_2) & \ldots & (a_1, b_n) \\ (a_2, b_1) & (a_2, b_2) & \ldots & (a_2, b_n) \\ \ldots & \ldots & \ldots & \ldots \\ (a_m, b_1) & (a_m, b_2) & \ldots & (a_m, b_n) \end{matrix}$$

В этой таблице m строк (по числу элементов в A) и n столбцов (по числу элементов в B), так что $A \times B$ имеет mn элементов. Итак,

$$n(A \times B) = n(A) \cdot n(B). \tag{3}$$

Следующие три задачи приводят к важнейшим формулам комбинаторики (их решение мы получим, руководствуясь обычным здравым смыслом).

Задача 1. Сколькими способами можно разместить m предметов в n ящиках по одному?

РЕШЕНИЕ. Обозначим искомое число через A_n^m («число размещений m (предметов) в n (ящиках)»). По смыслу задачи $m \leqslant n$. Для примера рассмотрим случай, когда $m = 2$, $n = 3$, т.е. найдем A_3^2. Имеем два предмета и три ящика. Для размещения первого предмета есть три возможности. С каждой из них связаны две возможности для размещения второго предмета. Так что, согласно формуле (3), $A_3^2 = 3 \cdot 2 = 6$ (рис. 75).

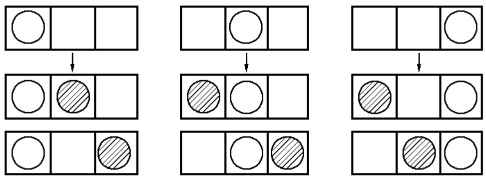


Рис. 75

В общем случае действуем аналогично. Размещаем предметы по очереди. Для первого предмета можно выбрать любой из n ящиков: $A_n^1 = n$. Тогда для второго предмета остается $n - 1$ пустых ящиков, и значит, возможностей разместить два предмета будет $A_n^2 = n(n-1)$. С каждой из этих возможностей связано $n-2$ способа размещения третьего предмета. По формуле (3) имеем: $A_n^3 = A_n^2 \cdot (n-2) = n(n-1)(n-2)$. Продолжая эти рассуждения, получаем:

$$
\begin{aligned}
&A_n^1 = n,\\
&A_n^2 = n(n-1),\\
&A_n^3 = n(n-1)(n-2),\\
&\dots\\
&A_n^m = n(n-1)(n-2)\dots(n-(m-1)).
\end{aligned}
$$

Итак,

$$A_n^m = n(n-1)(n-2)\dots(n-(m-1)) \tag{4}$$

(всего m множителей – их количество указывает верхний индекс, первым является n – нижний индекс, каждый следующий множитель на единицу меньше предыдущего).

Например, $A_5^3 = 5 \cdot 4 \cdot 3 = 60$, $A_{10}^4 = 10 \cdot 9 \cdot 8 \cdot 7 = 5040$.

Задача 2. Сколькими способами можно m объектов расположить в ряд?

РЕШЕНИЕ. Искомое число обозначим через P_m («число перестановок m объектов»). Если в задаче 1 положить $n = m$, т.е. число ящиков сделать равным числу предметов, то размещения будут отличаться друг от друга только порядком расположения предметов и фактически превращаются в перестановки. Так что $P_m = A_m^m = m(m-1)\dots(m-(m-1)) = m(m-1)\cdot\dots\cdot 2\cdot 1 =$ $= 1\cdot 2\cdot\dots\cdot m$.

Итак,

$$P_m = 1\cdot 2\cdot\dots\cdot m. \tag{5}$$

Произведение первых m натуральных чисел $1\cdot 2\cdot\dots\cdot m$ обозначается специальным символом $m!$ («m факториал»). Вот первые шесть факториалов: 1!=1, 2!=1·2=2, 3!=1·2·3=6, 4!=1·2·3·4=24, 5!=1 · 2 · 3 · 4 · 5=120, 6!=1 · 2 · 3 · 4 · 5 · 6=720.

Как видим, факториал – очень быстро растущая функция. Заметим, что $m! = (m-1)!\cdot m$, и продолжим список факториалов: 7!=6!·7=720·7=5040, 8!=7!·8=5040·8=40 320, 9!=40320·9=362 880, 10!=3 628 800.

Задача 3. Сколькими способами можно выбрать m предметов из имеющихся n?

РЕШЕНИЕ. Обозначим искомое число через C_n^m («число сочетаний из n (предметов) по m», «цэ из n по m»). По смыслу задачи $m \leqslant n$.

Чтобы найти число C_n^m, предложим новый способ подсчета величины A_n^m, найденной в задаче 1. Именно, из имеющихся ящиков выберем произвольно m штук (по числу предметов). Это можно сделать C_n^m способами. С каждым таким выбором m ящиков связано P_m способов размещения в них m предметов (это будут их всевозможные перестановки). Понятно, что, действуя таким образом, мы получим все размещения наших m предметов в n ящиках. Согласно принципу (3), $A_n^m = C_n^m \cdot P_m$, откуда $C_n^m = \dfrac{A_n^m}{P_m}$.

Итак,

$$C_n^m = \frac{n(n-1)\dots(n-(m-1))}{1\cdot 2\cdot\dots\cdot m} \tag{6}$$

(синтаксически: в числителе стоит m множителей, начиная с n, в порядке убывания; в знаменателе тоже m множителей, но начиная с 1 и в порядке возрастания).

Например,

$$C_{10}^{3} = \frac{10 \cdot 9 \cdot 8}{1 \cdot 2 \cdot 3} = 120, \qquad C_{6}^{2} = \frac{6 \cdot 5}{1 \cdot 2} = 15.$$

Отметим полезную *формулу приведения*:

$$C_n^m = C_n^{n-m} \tag{7}$$

(когда мы n предметов разделим на две группы: m штук в одной и $n - m$ штук в другой, то – какую группу считать выбранной? Другими словами, каждому конкретному выбору m предметов соответствует выбор (оставшихся) $n - m$ предметов).

Формула (7) используется в ситуациях, когда верхний индекс в C_n^m (т.е. число выбранных предметов) больше половины нижнего (т.е. общего числа предметов). Например,

$$C_{10}^{7} = C_{10}^{10-7} = C_{10}^{3} = 120,$$

$$C_{100}^{98} = C_{100}^{100-98} = C_{100}^{2} = \frac{100 \cdot 99}{1 \cdot 2} = 4950$$

(без формулы (7) пришлось бы писать 98 множителей в числителе и столько же в знаменателе,– чтоб потом почти все зачеркнуть при сокращении).

Запомним: когда мы выбираем m предметов из n имеющихся, они представляются камешками на ладони – нет среди них ни первого, ни последнего; при размещениях же приходится учитывать еще и взаимное расположение предметов – они как бы выстраиваются в ряд.

В следующих упражнениях нужно применить подходящие из формул (1)-(7).

1. Первокурсники писали две контрольные работы. Зачет по первой получили 43 человека, по второй – 46, по обеим работам – 40 человек. Сколько студентов справилось хотя бы с одной работой? (Формула (1); 49).

2. Сколькими способами могут сесть за круглый стол 3 женщины и 3 мужчины так, чтобы женщины и мужчины чередовались? (Формулы (3) и (5); 36).

3. Некто забыл последние пять цифр шестизначного телефонного номера и только помнил, что все они разные. Действуя методом случайного набора, скольких абонентов он побеспокоит в наихудшем случае? (Формула (4); 30 240).

4. Группа А может выставить для участия в соревнованиях 5 человек, группа Б – 7 и группа В – троих. Сколькими способами можно составить команду курса, беря из каждой группы 3 человека? (Формулы (3), (6) и (7); 350).

5. «Двести фунтов – не бог весть какая сумма; голова у него раскалывалась, и цифры гудели в ней, как перезвон колоколов: «200, 002, 020»; его раздражало, что он не может подобрать четвертой комбинации – 002, 200, 020...» (Грэм Грин. Суть дела). Как бы вы действовали в этой задаче на месте полицейского чиновника Сотби?

Одним из приемов, использовавшихся средневековыми учеными для публикации результатов, в которых не было полной уверенности, служили анаграммы: буквы, составлявшие сообщение об открытии, выстраивали в произвольном порядке и получившуюся абракадабру рассылали коллегам, чтобы потом можно было отстаивать приоритет. Галилей, принявший кольца Сатурна за два его спутника, осторожно сообщал:

$$Smaismrmmielmepoetaleumibuvnenugttaviras$$

(«Altissimum planetam tergeminum observavi» – высочайшую планету тройною наблюдал» плюс две лишние буквы: e, u). Заинтригованный Кеплер, отбросив две буквы (e, v), упорядочил остальные следующим образом:

$$Salve,\ umbistineum\ geminatum\ Martia\ proles$$

(«Привет вам, близнецы, Марса порождение») и объявил, что Галилей обнаружил, наконец, предсказанные им, Кеплером, два спутника Марса.

Умы студентов-богословов многих поколений, начиная с XV века, занимала анаграмма знаменитого вопроса, который Пилат задал Христу: Quid est veritas? («Что есть истина?»). Христос не ответил, ибо ответ якобы уже содержался в формулировке вопроса – если переставить буквы: Vir, qui ad est, est («Это муж, который перед (тобой)»).

Взглянув еще раз на табличку факториалов, легко представить себе, каких усилий (или озарений) требовали поиски и расшифровки подобных перестановок-анаграмм.

§ 3. Вероятность

Теория вероятностей как раздел математики возникла в XVII веке первоначально из потребностей азартных игр. Уловить закономерности, управляющие случаем, первым попытался Блез Паскаль (1623–1662). С двадцати пяти лет он жил в монастыре, занимаясь философией, теологией и математикой. Некий кавалер де Мере обратился к нему с вопросами, которые возникли в связи с игрой в кости. В завязавшейся по этому поводу переписке Паскаля и Ферма и были получены первые научные результаты о вероятностях (1654 г.). В последующие столетия теория вероятностей и связанная с ней математическая статистика своими достижениями внесли существенный вклад в формирование общественного взгляда на математику как необходимый инструмент в человеческой деятельности. Однако, несмотря на это, математический смысл основных понятий, связанных с вероятностью, оставался не вполне ясным вплоть до 1933 года, когда А.Н.Колмогоров предложил их аксиоматическое описание. (Андрей Николаевич Колмогоров (1903–1987), профессор Московского унмверситета, академик. Получил основополагающие результаты в области математического анализа, теории вероятностей, логики, механики, теории информации. Автор многочисленных работ по применению математических методов в военном деле, биологии, лингвистике. Инициатор реформирования школьного преподавания математики. «В основе большинства математических открытий лежит какая-либо простая идея: наглядное геометрическое построение, новое элементарное неравенство и т.п. Нужно только применить надлежащим образом эту простую идею к решению задачи, которая с первого взгляда кажется недоступной», – эти слова выдающегося ученого находят яркое подтверждение и в его собственном творчестве.)

Пусть проводится некоторый эксперимент, например, бросается монета или игральная кость. С ним связаны те или иные события. В случае с монетой это выпадение герба, выпадение цифры (научные синонимы «орла» и «решки»), при упражнении с игральной костью – большее разнообразие: выпадение единицы, выпадение двойки,..., шестерки, появление четного числа очков, числа очков $\geqslant 3$ и т.д.

Событие, которое всегда происходит при проведении данного эксперимента, называется достоверным и обозначается символом Ω. Например, при бросании монеты достоверным будет событие «выпадение герба или цифры», а проводя очередной эксперимент с игральной костью, мы неизбежно столкнемся с событием «число выпавших очков равно одному из чисел 1, 2, 3, 4, 5, 6».

Событие, которое никогда не наступает в данном эксперименте, называется невозможным (или пустым) и обозначается символом $\emptyset$. Бросая монету, невозможно встретить событие «герб и цифра одновременно»; событие «нечетное число, большее пяти» никогда не появится в эксперименте с кубиком.

События, отличные от достоверного и невозможного, называются случайными событиями. «Выпадение герба» или «выпадение четного числа очков» – случайные события в опыте соответственно с монетой и с игральной костью. В общих рассуждениях события обозначаются прописными латинскими буквами A, B, C и т.д. События, которые в данном эксперименте не могут появиться одновременно, называются несовместными. «Герб» и «цифра» (монета), «чет» и «нечет» (кубик) – примеры несовместных событий.

В аксиоматической теории вероятностей события считают подмножествами некоторого фиксированного множества Ω. Оно само отождествляется с достоверным событием, а пустое подмножество $\emptyset$ – с невозможным событием (отсюда и введенные выше обозначения). Пересечение $A \cap B$ подмножеств A и B интерпретируется как одновременное наступление соответствующих событий, а их объединение $A \cup B$ – как появление хотя бы одного из событий A, B. Дополнение $\bar{A}$ подмножества A, представляющего одноименное событие, символизирует противоположное для A событие: оно состоит в непоявлении A. Таким образом, события и действия с ними можно иллюстрировать диаграммами Венна (см. §1).

А что же такое вероятность? Каждому событию A приписывается число $p(A)$ – вероятность события A. При этом должны выполняться следующие требования (аксиомы вероятности):

I. $p(A) \geqslant 0$ (вероятность любого события неотрицательна);

II. $p(\Omega) = 1$ (вероятность достоверного события равна 1);

III. $A \cap B = \emptyset \to p(A \cup B) = p(A) + p(B)$ (если события A и B несовместны, то вероятность появления хотя бы одного из них равна сумме их вероятностей).

Вся теория вероятностей выводится из этих трех простых условий. Приведенная математическая схема явилась итогом трехсотлетнего научного поиска закономерностей, связанных с понятием случайности: от простейших азартных игр (орлянка и кости) до сложнейших процессов, описывающих физический мир и развитие общества.

Посмотрим, как подсчитываются вероятности событий в некоторых конкретных ситуациях.

Под исходом эксперимента понимается такое связанное с ним событие, которое не является следствием появления каких-либо других событий. Например, выпадение двойки – исход бросания игральной кости, но выпадение четного числа очков – не исход: это событие наступает, скажем, при появлении двойки (или четверки, или шестерки), т.е. оказывается следствием других событий.

Эксперимент называется классическим, если он имеет конечное число исходов и все исходы равновозможны. Такими экспериментами будут, очевидно, бросание монеты и бросание игральной кости. Пусть событие A связано с классическим экспериментом. Говорят, что некоторый исход эксперимента благоприятствует событию A, если A наступает каждый раз при появлении этого исхода.

Вероятностью события A в классическом эксперименте называется число

$$p(A) = \frac{n(A)}{N}, \tag{1}$$

где N – общее число исходов эксперимента, $n(A)$ – число исходов, благоприятствующих событию A.

Рассмотрим несколько примеров вычисления вероятностей в классических экспериментах.

1°. При бросании монеты $N = 2$ (Г – герб, Ц – цифра), $n(\text{Г}) = 1$, так что $p(\text{Г}) = \dfrac{n(\text{Г})}{N} = \dfrac{1}{2}$. Аналогично $p(\text{Ц}) = \dfrac{1}{2}$.

2°. Для игральной кости $N = 6$ (выпадение 1, 2, 3, 4, 5, 6 очков). Если A – «появление числа очков, кратного 3», то $n(A) = 2$ (тройка или шестерка) и, значит, $p(A) = \dfrac{n(A)}{N} = \dfrac{2}{6} = \dfrac{1}{3}$.

3°. В урне 10 шаров: 4 белых и 6 черных. Наугад извлекается один шар. Какова вероятность того, что этот шар будет белым (событие Б)? Здесь некоторую трудность представляет определение числа N. Правильное значение $N = 10$ (исход – это извлечение каждого отдельного шара. Для нас все белые шары одинаковы, но каждый из них сам для себя – яркая индивидуальность). Так как, очевидно, $n(\text{Б}) = 4$, то $p(\text{Б}) = \dfrac{n(\text{Б})}{N} = \dfrac{4}{10} = 0,4$.

4°. Бросаются две игральные кости. Что вероятнее: получить сумму 7 (событие A) или сумму 8 (событие B)? Для каждой игральной кости имеем шесть исходов. По принципу «каждый с каждым» (формула (3) из §2) для двух игральных костей $N = 6 \cdot 6 = 36$. При этом сумма 7 получается шестью способами (первое число – результат бросания для одной игральной кости, второе – для другой): 1+6, 2+5, 3+4, 4+3, 5+2, 6+1, так что $n(A) = 6$. Для суммы 8 имеем пять благоприятствующих исходов: 2+6, 3+5, 4+4, 5+3, 6+2. Следовательно, $n(B) = 5$. Итак, $p(A) = \dfrac{n(A)}{N} = \dfrac{6}{36}$, $p(B) = \dfrac{n(B)}{N} = \dfrac{5}{36}$, т.е. $p(A) > p(B)$,– сумма 7 вероятнее, чем сумма 8. (Как мы понимаем вероятность $\dfrac{5}{36}$? Можно считать, что в среднем при 36 бросаниях двух игральных костей 5 раз будет появляться сумма 8. Но это, конечно, не означает, что так оно и будет в каждых конкретных 36 бросаниях!)

5°. Какова вероятность угадать все пять номеров в лотерее спортлото «5 из 36» (событие A)? Мы зачеркиваем 5 произвольных чисел из 36. Чему равно N, т.е. сколькими способами можно выбрать 5 чисел из 36? По формуле (6) из §2 это будет $C_{36}^5 = \dfrac{36 \cdot 35 \cdot 34 \cdot 33 \cdot 32}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5} = 376992$. Очевидно, что $n(A) = 1$ (именно та пятерка чисел, которая появится в тираже). Значит,

$$p(A) = \frac{n(A)}{N} = \frac{1}{376992} \approx 0,000003.$$

Рассмотрим некоторые свойства классических вероятностей.

Теорема 1 (Базисные свойства). Вероятности, подсчитываемые по методике (1), удовлетворяют требованиям I-III математической теории.

ДОКАЗАТЕЛЬСТВО. I. Действительно, так как $N > 0$ (эксперимент не может иметь отрицательное или равное нулю число исходов) и $n(A) \geqslant 0$ (может случиться, что событию A не благоприятствует ни один исход: когда $A = \emptyset$), то $p(A) = \dfrac{n(A)}{N} \geqslant 0$.

II. Поскольку $n(\Omega) = N$ (каждый исход благоприятствует достоверному событию), то $p(A) = \dfrac{n(\Omega)}{N} = \dfrac{N}{N} = 1$.

III. Если события A и B несовместны, т.е. $A \cap B = \emptyset$, то согласно формуле (2) из §2, $n(A \cup B) = n(A) + n(B)$, откуда

$$p(A \cup B) = \frac{n(A \cup B)}{N} = \frac{n(A) + n(B)}{N} =$$
$$= \frac{n(A)}{N} + \frac{n(B)}{N} = p(A) + p(B).$$

Теорема доказана. □

Отметим еще два полезных свойства.

Теорема 2 (Вероятность противоположного события).

$$p(\bar{A}) = 1 - p(A).$$

ДОКАЗАТЕЛЬСТВО. Очевидно, что $n(\bar{A}) = N - n(A)$ (событию $\bar{A}$ благоприятствуют в точности те исходы, которые не благоприятствуют A). Тогда

$$p(\bar{A}) = \frac{n(\bar{A})}{N} = \frac{N - n(A)}{N} = 1 - \frac{n(A)}{N} = 1 - p(A),$$

что и требовалось. □

(Приведите несколько соображений в пользу того, что вероятность невозможного события равна нулю).

Теорема 3 (Границы для вероятности).

$$0 \leqslant p(A) \leqslant 1.$$

ДОКАЗАТЕЛЬСТВО. Нижняя граница для вероятности (т.е. неравенство $p(A) \geqslant 0$) установлена в теореме 1. Для доказательства второго неравенства заметим, что $n(A) \leqslant N$, откуда делением обеих частей на положительное число N получаем: $\frac{n(A)}{N} \leqslant 1$, т.е. $p(A) \leqslant 1$. □

Теорема 3 позволяет выражать вероятности в процентах – умножением их на 100. Например, можно сказать, что вероятность выпадения герба при бросании монеты равна 50%, а вероятность угадать все 5 номеров в лотерее спортлото «5 из 36» составляет примерно 0,0003% (три десятитысячных процента).

Не всякий эксперимент является классическим. Пусть, например, на плоскость случайным образом бросается точка. Известно, что она оказалась в области Ω (рис. 76). Какова вероятность того, что эта точка попадет в подобласть A (событие A)? Исходами эксперимента являются попадания в конкретные точки области Ω. Эти исходы равновозможны (мы бросаем точку наугад), но в классическую схему эксперимент не укладывается: у него бесконечно много исходов (столько, сколько точек в Ω).

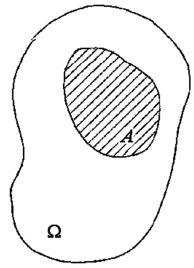


Рис. 76

В данной и похожих ситуациях предлагается следующий «геометрический принцип» подсчета вероятностей:

$$p(A) = \frac{S(A)}{S(\Omega)}, \tag{2}$$

где $S(\Omega)$ – площадь области Ω, а $S(A)$ – площадь подобласти A.

Пусть, для примера, внутри круга Ω радиуса $R = 10$ лежит круг радиуса $r = 5$. Какова вероятность того, что точка, случайным образом брошенная в круг Ω, окажется в круге A?

По схеме (2),

$$p(A) = \frac{S(A)}{S(\Omega)} = \frac{\pi r^2}{\pi R^2} = \frac{r^2}{R^2} = \frac{5^2}{10^2} = \frac{25}{100} = 0,25,$$

т.е. вероятность попадания в меньший круг равна 25%.

Назовем событие A невероятным, если $p(A) = 0$. Пусть в некотором квадрате Ω выделена точка A. Какова вероятность того, что при случайном бросании точки в квадрат Ω мы попадем именно в точку A? Так как площадь области, состоящей из одной точки, равна нулю, то, согласно геометрическому принципу, $p(A) = 0$, т.е. попадание в каждую конкретную точку квадрата – невероятное событие. Но это событие не является невозможным: при любом бросании мы ведь попадем в какую-то точку! Аналогично событие, имеющее вероятность 1 (=100%), не обязательно достоверно, т.е. не обязательно наступает при каждом проведении эксперимента (приведите пример).

Геометрический принцип подсчета вероятностей можно распространить на линейные области (например, Ω – отрезок и A – его часть) и на пространственные области (скажем, Ω – куб, а A – расположенный внутри него шар). В первом случае вероятность вычисляется как отношение длин, а во втором – как отношение объемов.

Нетрудно проверить, что все свойства, установленные нами для классического случая (теоремы 1-3), выполняются и для геометрических вероятностей.

Классическая и геометрическая схемы не охватывают всего многообразия вероятностных задач. Пусть, например, бросается игральная кость со смещенным центром тяжести. Исходов в этом эксперименте 6 – конечное число, но они не являются равновозможными: если дробинка «встроена» под грань с шестеркой, то шестерка будет выпадать гораздо реже, чем противоположная ей единица. Как теперь подсчитать вероятность выпадения шестерки? На помощь приходит так называемая статистическая схема.

Эксперимент называется статистическим, если его можно проводить неограниченное число раз и если результат каждого его проведения не зависит от предыдущих результатов (и, значит, не влияет на последующие). Бросание «правильной» и «выбитой» монеты, «честной» и «нечестной» игральной кости, выбор наугад слова из текста романа «Евгений Онегин» – все это статистические эксперименты.

Пусть статистический эксперимент проведен N раз и появление интересующего нас события A было зарегистрировано при этом $n(A)$ раз. Число $p^* = \dfrac{n(A)}{N}$ называется частотой события A в данной серии проведений эксперимента. Например, если игральная кость брошена 1000 раз и при этом шестерка появилась 123 раза, то частота появлений шестерки была $p^* = 0,123$. Это не означает, конечно, что при следующих 1000 бросаниях мы опять получим шестерку 123 раза,– частота может меняться от серии к серии.

Одной из важнейших теорем о вероятностях считается следующий *Закон Больших Чисел*, доказанный Якобом Бернулли.

<u>Теорема 4</u>. Пусть в статистическом эксперименте событие A наступает с вероятностью p. Тогда частота p^* события A, вычисленная в достаточно длинной серии проведений эксперимента, почти наверное (т.е. с вероятностью сколь угодно близкой к единице) отклонится от p меньше чем на данное число, как бы мало оно ни было. $\Box$

Таким образом, если эксперимент проведен много раз, то в качестве заменителя неизвестной вероятности события с большой уверенностью можно взять его частоту. Закон больших чисел дает теоретическое обоснование широко употребляемому экспериментаторами статистическому способу подсчета вероятностей.

Заметим: теорема Бернулли не дает оснований для вывода о том, что чем больше число испытаний, тем меньше частота отличается от вероятности. В принципе, бросив правильную монету 1000 раз, мы можем получить, скажем, 900 гербов. Однако вероятность такого происшествия, согласно Закону больших чисел, чрезвычайно мала, оно может быть отнесено к категории чудес (как и главные выигрыши в лотереях).

Закон больших чисел, открытый Якобом Бернулли еще в 80-х годах XVII века, был опубликован в 1713 году, уже после смерти ученого. В 1913 году в России отмечалось 200-летие этой знаменитой теоремы. Заканчивая свою речь на торжественном заседании Императорской Академии Наук, академик А.А.Марков сказал: «Со времени смерти Бернулли прошло более 200 лет; однако он живет и будет жить в своей теореме».

Вероятности, вычисляемые по статистической схеме, удовлетворяют аксиомам I-III общей математической теории. Это не удивительно: аксиомы эти как раз и выражают то общее, что присуще всем вероятностям, по каким бы методикам они ни вычислялись.

В заключение предложим три несложные задачи на подсчет классических вероятностей.

1. Из ваших имени и фамилии наугад выбраны три буквы. Какова вероятность того, что все они будут согласными? (Ответ: если вы Иван Иванов, то $\frac{1}{12}$; если вы Петр Петров, то $\frac{7}{24}$).

2. Карточки с буквами Т,Р,О,В,А случайным образом раскладываются в ряд. С какой вероятностью при этом получится осмысленное слово? (Ответ: $\frac{1}{20}$).

3. Какова вероятность при бросании двух игральных костей хотя бы на одной получить не меньше двух очков? А если игральных костей три? Какова вероятность того, что при бросании десяти монет хотя бы на одной будет отмечен герб? (Ответ: $\frac{35}{36}$, $\frac{215}{216}$, $\frac{1023}{1024}$).

После решения третьей задачи полезно поразмышлять над следующей цитатой, показывающей, как причудливо может использовать сухие математические схемы раскованное воображение:

«Еще существует где-то в архивах Французской Академии знаменитый закон вероятий, выработанный известными математиками путем алгебраического процесса... Он гласит так: «Если два лица дают свое показание о факте и, таким образом, каждый передает ему $\frac{5}{6}$ достоверности, этот факт будет иметь тогда $\frac{35}{36}$ достоверности, т.е. его вероятие будет относиться к невероятию в пропорции 35:1. Если три согласных показания соединены вместе, вероятие дает $\frac{215}{216}$. Показание десяти лиц, каждое равняющееся $\frac{1}{2}$ вероятия, даст $\frac{1023}{1024}$ и т.д... Можно удовлетвориться подобной достоверностью, не заботясь о большей» (Елена Блаватская. Тайная доктрина. Перевод Елены Рерих).

Насколько объективно оцениваем мы вероятности тех или иных событий, свидетелями которых являемся? Многочисленные опыты на этот счет показали, что даже в самых простых случаях имеется значительное расхождение между истинными вероятностями и их субъективным восприятием. Например, все мы постоянно сталкиваемся с печатными текстами и, конечно, замечаем, что одни буквы встречаются в них чаще, а другие реже. О каждой из 32 букв русского алфавита попробуйте ответить на вопрос: сколько раз данная буква попадется в случайно выбранном отрывке, состоящем из 1000 букв (в случае равных вероятностей – 31 раз). При этом всякий раз руководствуйтесь интуицией, связанной с каждой буквой, и не заботьтесь о точном распределении числа 1000 между всеми буквами алфавита. Более простой эксперимент: выпишите в порядке убывания частот три, по вашему мнению, самых «частых» буквы и в порядке возрастания частот – три самых «редких». Сравните свои результаты с данными табл. 14 из § 5.

§ 4. Статистика

Статистические данные представляли интерес во все времена. В библейской Книге Чисел рассказывается, как Моисей в Синайской пустыне исчислил «все общество сынов Израилевых по родам их, по семействам их, по числу имен, всех мужеского пола поголовно от двадцати лет и выше, всех годных для войны». Точно указывается распределение всех 603 550 зарегистрированных лиц по двенадцати коленам, кроме левитского, которое в свою очередь подверглось отдельному подробному обследованию.

В новое время количественные демографические наблюдения регулярно проводились с середины XVII века. В замечательной работе Дж. Арбатнета 1712 года было, например, сообщено, что ежегодно на протяжении многих десятилетий в Лондоне рождалось больше мальчиков, чем девочек,– закономерность, являвшаяся, по мнению автора, несомненным доводом в пользу Божественного Провидения, учитывая объективно преобладающие опасности в жизни мужчин.

В наши дни, по существу, никакая практическая деятельность общественного масштаба не обходится без статистики.

Статистика изучает массовые явления, т.е. свойства (признаки), присущие объектам более или менее обширных совокупностей. Методы сбора статистических данных, их систематизация, обработка и содержательное истолкование – вот основные задачи статистики. Каждая экспериментальная область имеет свою традиционную статистику, поэтому здесь мы остановимся лишь на общих понятиях, которые используются в статистиках гуманитарных наук.

Пусть изучается количественный (т.е. выражаемый числом) или качественный признак ξ, присущий некоторым объектам. Это может быть, например, рост или коэффициент интеллектуальности человека, цвет или форма предмета и т.п. Совокупность всех носителей данного признака называется генеральной совокупностью. Как правило, генеральная совокупность в целом недоступна наблюдению, и экспериментатору приходится иметь дело с отдельными ее частями, они называются выборками. Число N объектов, попавших в выборку, называется ее мощностью (или объемом). Чтобы иметь представление об общем состоянии данного признака ξ, мы должны сформировать такую выборку, которая правильно отражала бы все пропорции генеральной совокупности относительно ξ. Такие выборки называют представительными (или репрезентативными). Но как узнать, является ли данная выборка представительной, если мы только еще приступаем к изучению свойства ξ? Принцип, основанный на Законе больших чисел, утверждает, что выборка будет представительной, если все объекты генеральной совокупности имеют равные шансы попасть в эту выборку.

Из многочисленных способов случайного отбора отметим следующие.

1. Простой случайный отбор. Каждый объект генеральной совокупности получает индивидуальный числовой код, и вызов объектов для обследования осуществляется по таблице случайных чисел. Например, каждому ученику данной средней школы может быть приписано пятизначное число, первые две цифры которого обозначают параллель (от 01 до 11), третья – номер класса в параллели (пусть их будет меньше десяти), четвертая и пятая – номер ученика в алфавитном списке класса.

2. Механический отбор. Для формирования выборки мощности N генеральная совокупность механически делится на N частей, в каждой из которых выбирается один объект. Например, с конвейера за смену сходит 5000 стандартных изделий, проверяется каждое сотое ($N = 50$).

3. Серийный отбор. Генеральная совокупность естественным образом разбивается на части (серии), каждая из которых достаточно адекватно отражает общее состояние признака. Случайным образом выбирается какое-то число серий, и они подвергаются сплошному обследованию. Пример – выборочная перепись населения.

4. Типический отбор. Состояние признака может сильно колебаться в разных частях генеральной совокупности по отношению друг к другу. В этом случае генеральную совокупность делят

на некоторые типические части, выборки составляются отдельно из объектов каждой части. Например, проверяя уровень грамотности учащихся школ города, естественно составлять выборки из школьников одной параллели.

Основные понятия, возникающие при обработке наблюдений, рассмотрим на конкретном примере, содержание которого хотя и относится к лингвистике, но в достаточной степени общезначимо.

Пусть исследуется количественный признак (случайная величина) ξ – буквенная длина слова, извлеченного наугад из текста романа А.С.Пушкина «Евгений Онегин» (в современной орфографии). Генеральная совокупность – множество всех слов, имеющихся в тексте. В силу специфической организации поэтического текста романа, можно считать, что сериями, на которые естественно разбивается генеральная совокупность, являются строфы (знаменитая онегинская строфа, состоящая из 14 строк). Закодируем каждую строфу трехзначным числом: номер главы + номер строфы в ней (арабскими цифрами), сложим в урну свернутые листки с кодами и затем наугад вынем один из них – 701:

Гонимы вешними лучами,
С окрестных гор уже снега
Сбежали мутными ручьями
На потопленные луга.
Улыбкой ясною природа
Сквозь сон встречает утро года;
Синея блещут небеса.
Еще прозрачные, леса
Как будто пухом зеленеют.
Пчела за данью полевой
Летит из кельи восковой.
Долины сохнут и пестреют;
Стада шумят, и соловей
Уж пел в безмолвии ночей.

1. *Мощность выборки*. Наша строфа содержит 53 слова. Для удобства расчетов отбросим последние три слова («в безмолвии ночей») и зафиксируем $N = 50$.

2. *Последовательная регистрация значений признака*. Подсчитывая количество букв в каждом слове, получим (табл. 9):

Таблица 9

6,	7,	6,	1,	9,	3,	3,	5,	7,	7,
7,	2,	11,	4,	7,	5,	7,	6,	3,	9,
4,	4,	5,	6,	6,	3,	10,	4,	3,	5,
5,	8,	5,	2,	5,	7,	5,	2,	5,	8,
6,	6,	1,	8,	5,	5,	1,	7,	2,	3.

3. *Значения признака* $x_1, x_2, ..., x_n$ *в данной выборке* записываются обычно в порядке их возрастания. В нашем примере ряд значений имеет вид: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 , 11.

4. *Вариационный размах выборки* – это число $v = x_{\max} - x_{\min}$. У нас $v = 11 - 1 = 10$.

5. *Подсчет частостей.* Частостью значения x_i называется число m_i, равное количеству объектов выборки, значение признака у которых равно x_i. В нашем случае, например, $m_1 = 3$ (три слова длины 1). Подсчет частостей производится последовательным заполнением второго столбца следующей табл. 10 (поочередно вычеркивая числа в табл. 9, ставим палочки в соответствующих строчках):

Таблица 10

x_i												m_i
1	|	|	|									3
2	|	|	|	|								4
3	|	|	|	|	|	|						6
4	|	|	|	|								4
5	|	|	|	|	|	|	|	|	|	|	|	11
6	|	|	|	|	|	|	|					7
7	|	|	|	|	|	|	|	|				8
8	|	|	|									3
9	|	|										2
10	|											1
11	|											1

Подсчитывая в каждой строке количество палочек, заполняем столбец частостей m_i. Сумма всех частостей должна равняться мощности выборки: $m_1 + m_2 + \cdots + m_n = N$.

6. *Вариационный ряд* фиксирует состояние признака в данной выборке. Это двухстрочечная таблица, в первой строчке которой перечисляются значения признака в данной выборке (пункт 3), а во второй – соответствующие им частости. Случайная величина ξ – длина слова в романе «Евгений Онегин» – в нашей выборке предстает в следующем эмпирическом виде:

$$X = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\ 3 & 4 & 6 & 4 & 11 & 7 & 8 & 3 & 2 & 1 & 1 \end{pmatrix}.$$

7. *Частоты*. Частотой значения x_i называется число $p_i = \dfrac{m_i}{N}$. Это доля объектов, имеющих значение признака x_i, среди всех объектов выборки. Частоты называют также эмпирическими вероятностями. В рассматриваемом примере имеем: $p_1 = \dfrac{m_1}{N} = \dfrac{3}{50} = 0,06$; $p_2 = \dfrac{m_2}{N} = \dfrac{4}{50} = 0,08$; $\ldots$; $p_{11} = \dfrac{m_{11}}{N} = \dfrac{1}{50} = 0,02$.

Сумма всех частот равна единице: $p_1 + p_2 + \cdots + p_n = 1$.

8. *Ранжирование*. Рангом значения x_i называется число r_i – место, которое занимает x_i среди других значений, если их расположить в порядке убывания частостей (и частот). Если частости двух или нескольких значений совпадают, то либо всем им присваивается одинаковый ранг, либо ранги индивидуализируются в сответствии с порядком этих значений в вариационном ряде. Мы примем вторую точку зрения.

9. *Накопленные частости*. Накопленной частостью значения x_i называется число $s_i = m_1 + m_2 + \cdots + m_i$, т.е. сумма частостей всех значений, предшествующих x_i, плюс частость самого x_i.

10. *Итоговая таблица* для эмпирической случайной величины X, описывающей состояние признака ξ в данной выборке, в рассматриваемом примере представлена табл. 11.

Таблица 11

x_i	m_i	p_i	r_i	s_i
1	3	0,06	7	3
2	4	0,08	5	7
3	6	0,12	4	13
4	4	0,08	6	17
5	11	0,22	1	28
6	7	0,14	3	35
7	8	0,16	2	43
8	3	0,06	8	46
9	2	0,04	9	48
10	1	0,02	10	49
11	1	0,02	11	50

В целях наглядности при статистической обработке наблюдений используются графические образы: полигон, кумулята и круговая диаграмма.

11. *Полигон*. На плоскости выбирается прямоугольная система координат. По горизонтальной оси x_i откладываются значения признака, по вертикальной оси m_i – их частости. Масштабы на осях, как правило, разные. Начало отсчета (ноль) на горизонтальной оси выбирается так, чтобы все x_i (даже если среди них есть отрицательные числа) лежали правее вертикальной оси; начало отсчета на вертикальной оси совпадает с точкой пересечения осей. В этой системе координат строим точки $(x_1, m_1), (x_2, m_2), \ldots, (x_n, m_n)$. Последовательно соединяя их

прямолинейными отрезками, получаем ломаную, которая и называется полигоном (рис. 77).

12. *Кумулята.* В системе координат (x_i, s_i) последовательно соединяются отрезками точки $(x_1, s_1), (x_2, s_2), \ldots, (x_n, s_n)$, где s_i – накопленная частость значения x_i. Полученная ломаная называется кумулятой (рис. 78).

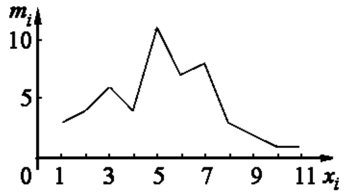


Рис. 77

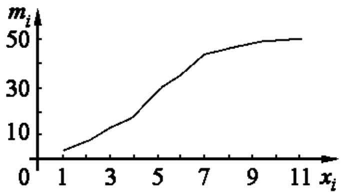


Рис. 78

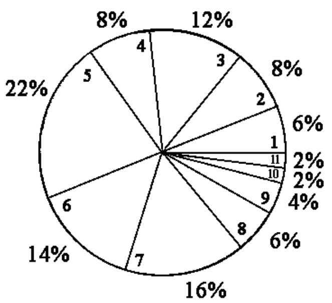


Рис. 79

13. *Круговая диаграмма.* Круг разбивается на n секторов (по числу различных значений признака в данной выборке), причем площадь сектора, соответствующего значению x_i, равна $p_i \cdot S$, где S – площадь круга. Частоты p_i обычно выражают на круговой диаграмме в процентах. Для рассматриваемого примера см. рис. 79.

Следующие три числовых показателя характеризуют типичные объекты выборки.

14. *Среднее значение.* Средним значением эмпирической случайной величины

$$X = \begin{pmatrix} x_1 & x_2 & \ldots & x_n \\ m_1 & m_2 & \ldots & m_n \end{pmatrix}, \quad m_1 + m_2 + \cdots + m_n = N \quad (1)$$

называется число

$$\bar{X} = \frac{x_1 m_1 + x_2 m_2 + \cdots + x_n m_n}{N}. \quad (2)$$

Какова средняя длина наугад выбранного из текста «Евгения Онегина» слова? Имеем:

$$\bar{X} = \frac{1{\cdot}3 + 2{\cdot}4 + 3{\cdot}6 + 4{\cdot}4 + 5{\cdot}11 + 6{\cdot}7 + 7{\cdot}8 + 8{\cdot}3 + 9{\cdot}2 + 10{\cdot}1 + 11{\cdot}1}{50} = \frac{261}{50},$$

т.е. $\bar{X} = 5,22$.

Среднее значение, конечно, не обязано быть одним из значений случайной величины: среднестатистический человек – это математическая абстракция. Так что в «Евгении Онегине» типичными по длине (ближайшая к среднему значению величина) будут пятибуквенные слова.

15. *Медиана*. Медианой случайной величины X называется число meX, относительно которого все объекты выборки разбиваются на две равные группы: в одну входят объекты, у которых значение признака не превышает meX, а в другую – те, у которых значение признака не меньше этого числа. Если все объекты выборки выстроить в ряд по возрастанию значений признака, то носителем медианы будет объект, стоящий в центре этой шеренги: медиана равна отмеченному у него значению. Медиана ищется по столбцу накопленных частостей s_i. Просматривая этот столбец, отыскиваем ту его строчку, где накопленная частость впервые достигнет или превзойдет половину мощности выборки $\frac{N}{2}$. Соответствующее значение x_i и есть медиана. В нашем примере $\frac{N}{2} = 25$, в столбце s_i последовательно стоят: 3, 7, 13, 17,– числа меньшие $\frac{N}{2}$, а затем 28 – величина, превосходящая $\frac{N}{2}$. Значит, $meX = 5$.

Иногда используют так называемую точную медиану, которая определяется равенством

$$MeX = x_i + \frac{\frac{N}{2} - s_{i-1}}{m_i} \cdot (x_{i+1} - x_i), \qquad (3)$$

где $x_i = meX$; s_{i-1} – накопленная частость предшествующего значения x_{i-1}; m_i – частость значения x_i; разность $x_{i+1} - x_i$ равна расстоянию между x_i и следующим значением признака x_{i+1} (напомним, что значения расположены по возрастанию). В рассматриваемом примере

$$MeX = 5 + \frac{25 - 17}{11} \cdot (6 - 5) \approx 5 + 0,73 = 5,73.$$

Так что типичным представителем выборки по методике медианы естественно считать слова длины 6.

16. *Мода*. Мода – это значение случайной величины, которое имеет наибольшую частость (т.е. мода – это то, что чаще всего встречается). В нашем случае $moX = 5$. Иногда используют так

называемую точную моду, которая определяется равенством

$$MoX = x_i + \frac{m_i - m_{i-1}}{(m_i - m_{i-1}) + (m_i - m_{i+1})} \cdot (x_{i+1} - x_i), \qquad (4)$$

где $x_i = moX$, а $m_{i-1},\ m_i,\ m_{i+1}$ – это частости соответственно значений $x_{i-1},\ x_i,\ x_{i+1}$. Для длины слова в «Евгении Онегине» будет

$$MoX = 5 + \frac{11-4}{(11-4)+(11-7)} \cdot (6-5) \approx 5 + 0,64 = 5,64.$$

Так что и по показателю моды типичная длина слова в романе равна 6.

Исследуемый признак может иметь одну моду (и тогда он называется унимодальным) или несколько мод (такие признаки называют полимодальными).

17.*Квантили* – это уровневые характеристики выборки. Пусть k – натуральное число и $1 \leqslant j \leqslant k-1$. Под j-й квантилью понимается число q_k^j, относительно которого все объекты выборки разбиваются на две группы: в одну входят $\dfrac{j}{k}N$ объектов – те, у которых значение признака не превосходит число q_k^j, а в другую – остальные $\dfrac{k-j}{k}N$ объектов, у которых значение признака не меньше чем q_k^j. Например, q_2^1 – это медиана.

Наиболее употребительны случаи $k = 4, 10, 100$. При $k = 4$ квантили называются квартилями (их три: q_4^1 – первая, q_4^2 – вторая, это медиана, q_4^3 – третья). При $k = 10$ получаем децили (от первой q_{10}^1 до девятой q_{10}^9; пятая q_{10}^5 – медиана). Если $k = 100$, квантили называют процентилями, их 99. Что такое, например, третья квартиль? Согласно общему определению, это такое число q_4^3, что три четверти объектов выборки имеют значение признака не больше q_4^3, а у четверти объектов значение признака превосходит этот уровень. Разумеется, $q_4^3 = q_{100}^{75}$, так что q_4^3 задает семидесятипятипроцентный уровень признака.

Квантили находят по столбцу накопленных частостей s_i. Просматривая этот столбец, фиксируют ту строку, где накопленная частость впервые достигнет или превзойдет контрольное число $\dfrac{j}{k}N$. Значение признака, стоящее в этой строке, и есть квантиль q_k^j. Найдем в примере с «Евгением Онегиным» квартили q_4^1 и q_4^3. Для этого отметим контрольные числа $\frac{1}{4}N = \frac{1}{4}50 = 12,5$ и $\frac{3}{4}N = \frac{3}{4}50 = 37,5$. Первое из них впервые «перешагиваем» в третьей строке, второе – в седьмой. Значит, $q_4^1 = 3$, $q_4^3 = 7$ (по случайности значения совпадают с номерами строк). Межквартильный размах выборки (есть и такое понятие) равен $q_4^3 - q_4^1 = = 7 - 3 = 4$.

Иногда используют так называемые точные квантили, применяя формулу

$$Q_k^j = x_i + \frac{\frac{j}{k}N - s_{i-1}}{m_i} \cdot (x_{i+1} - x_i), \tag{5}$$

где $x_i = q_k^j$, s_{i-1} – накопленная частость значения x_{i-1}, а m_i – частость значения x_i. В нашем случае

$$Q_4^1 = 3 + \frac{12,5 - 7}{6} \cdot (4 - 3) \approx 3 + 0,92 = 3,92$$

и

$$Q_4^3 = 7 + \frac{37,5 - 35}{8} \cdot (8 - 7) \approx 7 + 0,31 = 7,31.$$

Учитывая целозначность случайной величины ξ – длины слова в «Евгении Онегине», можно сказать, что примерно четверть всех слов в романе состоит меньше чем из четырех букв, а три четверти слов имеют в своем составе не более семи букв.

18. *Дисперсия* – мера рассеяния значений случайной величины относительно ее среднего значения. Дисперсия случайной величины

$$X = \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ m_1 & m_2 & \dots & m_n \end{pmatrix}, \qquad m_1 + m_2 + \dots + m_n = N$$

определяется формулой

$$DX = \frac{(x_1 - \bar{X})^2 \cdot m_1 + (x_2 - \bar{X})^2 \cdot m_2 + \ldots + (x_n - \bar{X})^2 \cdot m_n}{N}. \tag{6}$$

По своему определению дисперсия всегда неотрицательна, $DX \geqslant 0$.

Несложными преобразованиями дисперсию можно привести к виду

$$DX = \frac{x_1^2 m_1 + x_2^2 m_2 + \cdots + x_n^2 m_n}{N} - \bar{X}^2 \tag{7}$$

или, подставляя значение $\bar{X}$, представить в форме

$$DX = \frac{x_1^2 m_1 + x_2^2 m_2 + \cdots + x_n^2 m_n}{N} - \left(\frac{x_1 m_1 + x_2 m_2 + \cdots + x_n m_n}{N}\right)^2. \tag{8}$$

Последние две формулы часто используются при вычислениях.

В упражнениях с романом «Евгений Онегин» среднее значение $\bar{X}$ уже было найдено, так что дисперсию считаем по формуле (7):

$$DX = \frac{1^2 \cdot 3 + 2^2 \cdot 4 + 3^2 \cdot 6 + \cdots + 9^2 \cdot 2 + 10^2 \cdot 1 + 11^2 \cdot 1}{50} - (5,22)^2 =$$

$$= \frac{3+16+54+64+265+252+392+192+162+100+121}{50} - (5,22)^2 =$$

$$= 32,42 - 27,25 = 5,17.$$

Итак, $DX = 5,17$.

Таковы основные статистические показатели, применяемые при обработке экспериментальных данных или наблюдений, для признака, выраженного в данной выборке случайной величиной, представленной вариационным рядом (1).

Может случиться так, что при измерении интерес представляет не точное значение признака у объекта, а некоторый интервал, куда попадает это значение. К числу таких интервальных признаков могут быть отнесены, например, рост человека, коэффициент интеллектуальности, возраст и т.п. Так, скажем, случайная величина ξ – «рост студента первого курса (в см.)» – в выборке мощности $N = 100$ может быть представлена в виде табл. 12.

Таблица 12

№	x_i	m_i	p_i	r_i	s_i
1	155-160	1	0,01	8	1
2	160-165	8	0,08	5	9
3	165-170	17	0,17	3	26
4	170-175	29	0,29	1	55
5	175-180	25	0,25	2	80
6	180-185	11	0,11	4	91
7	185-190	7	0,07	6	98
8	190-195	2	0,02	7	100

При обработке наблюдений, связанных с интервальными случайными величинами, возникают некоторые особенности, модифицирующие схему, изложенную выше для дискретных признаков.

В качестве начала отсчета интервалов x_0 обычно выбирают ближайшее слева к x_{min} (наименьшее значение признака в выборке) «удобное» число. Например, если минимальный рост, отмеченный среди измеряемых, равен 158, то естественно положить $x_0 = 155$.

Длина интервалов h обычно выбирается из тех соображений, чтобы никакой интервал не содержал более 30% объектов выборки и чтобы число интервалов не превосходило 15. Одной из формальных рекомендаций является

$$h = \frac{v}{1 + \log_2 N} = \frac{x_{max} - x_{min}}{1 + \log_2 N}.$$

Если, например, самый высокий студент в данной выборке имеет рост 194, то $v = x_{max} - x_{min} = 194 - 158 = 36$ и

$$h = \frac{36}{1 + \log_2 100} = \frac{36}{1 + 6,67} = \frac{36}{7,67} \approx 4,7.$$

Естественно положить $h = 5$.

Считается, что интервалы имеют вид $[x_i, x_{i+1})$, т.е. что у них нет правых концов. Студент с ростом 175 см попадает в интервал 175-180. Последний по счету интервал содержит значение x_{max}.

Под частостью интервала понимается количество объектов выборки, значение признака у которых находится в данном интервале.

Для наглядного представления интервальных признаков используются столбиковые диаграммы, называемые гистограммами. Гистограмма строится в уже описанной системе координат (x_i, m_i): над каждым интервалом $[x_i, x_{i+1})$ рисуется прямоугольник высотой m_i. Если последовательно соединить отрезками середины верхних оснований прямоугольников гистограммы, получится полигон. Гистограмма и полигон для роста студентов изображены на рис. 80.

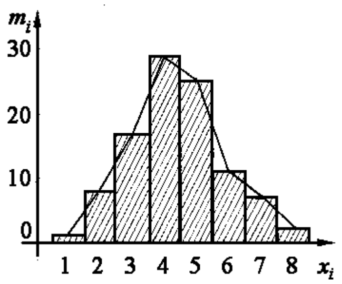


Рис. 80

Заметим, что при подсчете частостей палочки в соответствующей таблице типа табл. 10 выстраиваются в гистограмму, повернутую на $90°$ по часовой стрелке.

Среднее значение интервальной случайной величины вычисляется по формуле

$$\bar{X} = \frac{x_1^* m_1 + x_2^* m_2 + \cdots + x_n^* m_n}{N},$$

где x_i^* –середины интервалов (т.е. X заменяется дискретной случайной величиной, изображаемой полигоном). Существуют различные приемы, облегчающие «ручной» подсчет величины $\bar{X}$, но, пользуясь калькулятором, мы без труда справимся и с прямыми выкладками для среднего роста студентов:

$$\bar{X} = \frac{157,5 \cdot 1 + 162,5 \cdot 8 + \cdots + 187,5 \cdot 7 + 192,5 \cdot 2}{100} = 174,5.$$

По табл. 12, просматривая столбец накопленных частостей интервалов, находим медианный интервал: 170-175 – здесь накопленная частость впервые достигает уровень $\dfrac{N}{2} = 50$. Далее применяется формула, по существу совпадающая с формулой точной медианы (3):

$$MeX = x_i + \frac{\frac{N}{2} - s_{i-1}}{m_i} \cdot h,$$

где x_i – начало медианного интервала, m_i – его частость, s_{i-1} – накопленная частость предмедианного интервала, h – длина интервалов. В рассматриваемом примере

$$MeX = 170 + \frac{50 - 26}{29} \cdot 5 \approx 170 + 4,1 = 174,1.$$

Модальный интервал распознается как интервал с максимальной частостью, и мода ищется в нем по формуле точной моды (4):

$$MoX = x_i + \frac{m_i - m_{i-1}}{(m_i - m_{i-1}) + (m_i - m_{i+1})} \cdot h,$$

где x_i – начало модального интервала, m_{i-1}, m_i, m_{i-1} – частости соответственно предмодального, модального и следующего за модальным интервалов, h – длина интервалов.

При обследовании роста студентов получаем

$$MoX = 170 + \frac{29 - 17}{(29 - 17) + (29 - 25)} \cdot 5 = 170 + 3,75 \approx 173,8.$$

Как и в случае дискретных случайных величин, интервальный признак может оказаться унимодальным или полимодальным.

Для нахождения квантили Q_k^j сначала по столбцу накопленных частостей s_i отыскивают Q_k^j-интервал как интервал, где s_i впервые достигнет или превзойдет число $\dfrac{j}{k}N$. Затем используется формула

$$Q_k^j = x_i + \frac{\frac{j}{k}N - s_{i-1}}{m_i} \cdot h,$$

где x_i – начало Q_k^j-интервала, m_i – его частость, s_{i-1} – накопленная частость предыдущего интервала, h – длина интервалов.

Найдем первую и третью квартили для роста студентов. Здесь $\dfrac{1}{4}N = \dfrac{1}{4} \cdot 100 = 25$, $\dfrac{3}{4}N = \dfrac{3}{4} \cdot 100 = 75$. Интервалом первой квартили будет $[165, 170)$, третьей – $[175, 180)$. Следовательно,

$$Q_4^1 = 165 + \frac{25 - 9}{17} \cdot 5 \approx 165 + 4,7 = 169,7$$

и

$$Q_4^3 = 175 + \frac{75 - 55}{25} \cdot 5 = 175 + 4 = 179.$$

У четверти студентов рост не превышает 170 см, семидесятипятипроцентным уровнем является 179 см. Межквартильный размах составляет всего 9 см.

Для вычисления дисперсии DX воспользуемся непосредственно определяющей ее формулой (6), где под значениями интервальной случайной величины, как и в случае среднего значения, понимаются середины интервалов. Заготовим разности $x_i - \bar{X}$: $157,5 - 174,5 = -17$ и далее -12, -7, -2, 3, 8, 13, 18. Следовательно,

$$DX = \frac{(-17)^2 \cdot 1 + (-12)^2 \cdot 8 + \cdots + 13^2 \cdot 7 + 18^2 \cdot 2}{100} = 51,5.$$

Для статистического исследования качественных признаков арсенал средств существенно сужается. После составления итоговой таблицы со столбцами x_i, m_i, p_i, r_i (значения, частости, частоты и ранги) рисуют гистограмму, основаниями прямоугольников которой являются условные интервалы, соответствующие значениям признака, и строят круговую диаграмму.

Пусть, например, при анкетировании 1000 жителей города на вопрос «Каково в общем и целом Ваше мнение о передачах местного радио?» респонденту предлагалось выбрать один из пяти вариантов ответа: 1 – очень доволен, 2 – доволен, 3 – не очень доволен, 4 – совсем не доволен, 5 – никакого мнения. Результат опроса представлен табл. 13.

Таблица 13

№	x_i	m_i	p_i (%)	r_i
1	очень доволен	32	3,2	4
2	доволен	376	37,6	2
3	не очень доволен	480	48,0	1
4	совсем не доволен	91	9,1	3
5	никакого мнения	21	2,1	5

На рис. 81 показаны гистограмма и круговая диаграмма, характеризующие данный качественный признак.

Еще раз заметим, что традиции конкретных областей знания могут отличаться формой представления статистических объектов, их именами и обозначениями (то, что у нас называется «частостью», иногда именуют «частотой», при этом вместо слова «частота» используя термин «относительная частота», и т.п.). Но основные приемы обработки наблюдений во всех статистиках совпадают.

Конечно, в разделе «Статистика» читатель мог ожидать хотя бы упоминания о таких ее важных для приложений областях, как

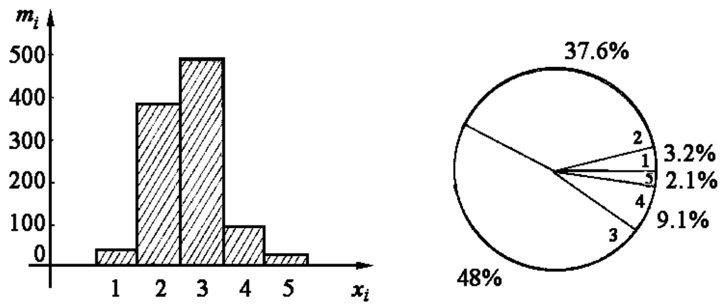


Рис. 81

теория корреляции, факторный анализ, проверка статистических гипотез (вот, они упомянуты), но это совсем другой уровень, и в непрерывном изложении до него не добраться. Кроме того, полезно помнить слова известного польского математика Штейнгауза, разрабатывавшего математические методы не только для статистики, но и для биологии, медицины, правоведения: «Математический аппарат, служащий для развития какой-либо отрасли знания, должен быть простейшим». Усложнения, связанные с технологией, идейно обогащенные специалисты найдут сами: они уже знают, где искать. (Кроме своих научных трудов и прикладных исследований, Гуго Дионисий Штейнгауз (1887–1972) энергично занимался популяризацией математики. Всеобщее восхищение вызывает его «Математический калейдоскоп» – книжка с картинками, в 125 коротких заметках которой математика предстает в удивительно поэтическом виде.)

§ 5. Теория передачи сообщений (теория информации)

В 1948 году, занимаясь конкретными задачами увеличения пропускной способности технических каналов связи и их помехоустойчивости, американский инженер Клод Элвуд Шеннон (1916–2001), работавший в лаборатории телефонной компании «Белл» (с 1958 профессор Массачусетского технологического института), предложил использовать для этого вероятностно-статистический подход, основанный на понятии энтропии и введенной им количественной мере информации. Возникшая из этих идей математическая теория передачи сообщений (теория информации) создала логическую базу для решения множества теоретических и технических проблем связи. Вместе с тем она обладала такой степенью общности, которая позволила интерпретировать ее результаты в самых различных областях знания. В основе применения идей и методов теории информации в гуманитарных науках лежит представление об аналогии процесса переработки информации человеком с процессами передачи информации по техническим линиям связи.

Под линией связи понимается система устройств для передачи сообщения (информации). Основными объектами линии связи являются: источник сообщения, кодер (кодирующее устройство), передатчик, канал связи, приемник, декодер, получатель сообщения. В реальных линиях связи некоторые из перечисленных устройств могут совмещаться. Общий вид условной линии связи представлен на рис. 82.

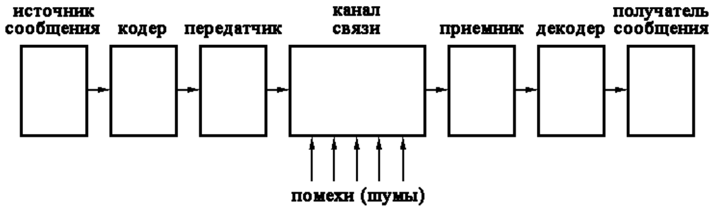


Рис. 82

Схема передачи по линии связи выглядит следующим образом. Источник сообщения выдает текст в некотором естественном для него алфавите. Этот текст кодируется – переводится кодером в последовательность технических сигналов. Сигналы подаются передатчиком на вход канала связи, транспортируются по нему, и на выходе они – или их в той или иной мере искаженные под действием помех образы – регистрируются приемником. Затем полученный текст переводится декодером на язык, доступный адресату. Классический пример линии связи – телеграф.

Широкие возможности в истолковании приведенной абстрактной схемы способствовали проникновению идей теории передачи сообщений в самые различные области знания.

Приведем математическое описание объектов линии связи и относящихся к ним важнейших понятий.

Поскольку мы собираемся передавать по линии связи некоторую информацию, выясним сначала, что понимается под этим словом.

1. *Энтропия* – величина, характеризующая степень неопределенности в предсказании исхода эксперимента. Если опыт имеет m исходов, вероятности которых суть $p_1, p_2, \dots, p_m$, то под энтропией этого опыта (обозначим его через ξ) понимается число $H(\xi)$, определяемое формулой Шеннона

$$H(\xi) = -p_1 \log p_1 - p_2 \log p_2 - \dots - p_m \log p_m \qquad (1)$$

(логарифм берется по основанию 2, но оно обычно не указывается – в теории информации, как правило, используются только двоичные логарифмы).

Для опыта с m равновероятными исходами $p_1 = p_2 = \dots = = p_m = \dfrac{1}{m}$ и, значит, $H(\xi) = \log m$.

Единицей для измерения энтропии является бит. Если эксперимент ξ имеет два равновероятных исхода (например, бросание монеты), то $H(\xi) = \log 2 = 1$. Это и есть бит. Термин образован стяжением английских слов «двоичная единица»: bit=bi(nary digi)t. Бит – это неопределенность, испытываемая нами при необходимости предсказать результат опыта, имеющего два равновероятных исхода.

Энтропия опыта η, состоящего в бросании игральной кости, тоже вычисляется по формуле для равновероятных исходов: $H(\eta) = \log 6 = 2,58$ бита (таблица двоичных логарифмов приводится в некоторых книгах по теории информации). Примером применения общей формулы Шеннона (1) может служить подсчет энтропии эксперимента ζ, состоящего в извлечении наугад одного шара из урны, где находятся 2 красных, 2 желтых и 4 синих шара. Фиксируя результат проведения эксперимента цветом извлеченного шара, получаем три исхода – «красный», «желтый», «синий» – с вероятностями соответственно $\frac{1}{4}, \frac{1}{4}, \frac{1}{2}$. Следовательно,

$$H(\zeta) = -\frac{1}{4}\log\frac{1}{4} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{2}\log\frac{1}{2} = \frac{1}{4}\cdot 2 + \frac{1}{4}\cdot 2 + \frac{1}{2}\cdot 1 = 1,5.$$

Как видим, из трех рассмотренных экспериментов наибольшее беспокойство вызывает необходимость угадать результат бросания игральной кости.

Энтропия, определенная формулой Шеннона (1), всегда неотрицательна. Действительно, вероятности $p_1, p_1, \dots, p_m$ – это правильные дроби, так что $\log_2 p_i < 0$, а $-\log_2 p_i > 0$ для всех $i = 1, 2, \dots, m$. Отсюда $H(\xi) \geqslant 0$. Если эксперимент имеет всего один исход, то этот исход появляется при любом проведении эксперимента и, значит, его вероятность $p = 1$. Отсюда

$H(\xi) = \log 1 = 0$,– и в самом деле, мы не испытываем никакой неопределенности в предсказании исхода.

Пусть эксперименты ξ и η имеют каждый m исходов, но все исходы эксперимента η равновероятны, а у ξ это не так. Понятно, что предсказывать исход опыта η труднее (при проведении ξ естественно ставить на исход с максимальной вероятностью). Теоретические расчеты показывают, что и математически $H(\xi) \leqslant H(\eta)$. Так как $H(\eta) = \log m$, то получаем неравенство

$$H(\xi) \leqslant \log m, \tag{2}$$

где m – количество исходов эксперимента ξ. Например, у любого эксперимента ξ с 32 исходами $H(\xi) \leqslant \log 32 = 5$ битов.

2. *Количество информации*, получаемой в результате проведения эксперимента ξ, обозначается символом $I(\xi)$. Оно численно равно энтропии $H(\xi)$: проведя опыт, мы получаем о его результате ровно столько информации, какова была неопределенность в предсказании этого результата. Единицей количества информации тоже является бит. Бит – это количество информации, которое мы получаем, когда становится известным результат опыта с двумя равновероятными исходами.

ПРИМЕР. Некто задумал целое число между 1 и 32. Каково минимальное количество вопросов с ответами «да» и «нет», гарантирующее отгадывание?

Энтропия этого опыта (отгадать одно из 32 чисел) равна, очевидно, $\log 32 = 5$ битам. Всякий ответ, устанавливающий альтернативный признак неизвестного числа, дает не более одного бита информации. Таким образом, меньше чем пятью вопросами рассматриваемого типа неопределенность в 5 битов не исчерпать. С другой стороны, если вопросы таковы, что каждый из них предполагает равное количество «да»-чисел и «нет»-чисел, то первый шаг оставит для таинственного x 16 возможностей, следующий – 8 и т.д. После четвертого такого вопроса останется только два подозрительных числа и, значит, пятое испытание наверняка фиксирует x. Практическое правило для чисел: дихотомия (деление пополам) по признаку «больше–меньше». – Это число больше 16? – Нет. – Оно больше 8? – Да. – Больше 12 (середина между 8 и 16)? – Да. – Больше 14? – Нет. – Больше 13? – Нет.– Значит, $x = 13$.

Учитывая, что количество атомов в известной нам части Вселенной оценивается примерно числом 2^{270}, получаем, что двухсот семидесяти дихотомических вопросов достаточно для локализации любого атома во Вселенной. Вот что такое информация в 270 битов.

3. *Источник сообщения* – это система Ξ, состояния которой меняются с течением времени случайным образом. Каждому физическому состоянию источника сообщения можно взаимно однозначно сопоставить некоторый символ и говорить не о переходе рассматриваемой системы в данное состояние, а о появлении соответствующего символа. Эти символы образуют алфавит источника сообщения, и всякая цепь последовательных переходов системы Ξ из одного состояния в другое (в дискретные моменты времени) предстает в виде текста в указанном алфавите. Этот текст и есть то, что называется сообщением.

Например, если источником сообщения является бросаемая над плоскостью монета, а его состоянием в данный момент наблюдения считается выпавшая сторона (герб или цифра), то алфавит состоит из символов Г и Ц, а сообщением может оказаться, скажем, такой текст: ГГЦГЦЦЦГЦЦГЦГЦГ (пятнадцать последовательных наблюдений состояния системы).

Важнейшей величиной, связываемой с источником сообщения Ξ, является его энтропия $H(\Xi)$, характеризующая неопределенность в предсказании следующего символа в прерванном сообщении. Если источник сообщения Ξ генерирует случайный текст в русском алфавите (32 буквы), то в качестве следующей может появиться с равной вероятностью любая буква, и значит, $H(\Xi) = \log 32 = 5$ битов. Для латинского алфавита получается в этом случае $H(\Xi) = \log 26 = 4,76$ бита.

Величину $H(\Xi)$ можно истолковывать также как среднее количество информации, содержащееся в одной букве текста. Разумеется, в структурированных (не хаотичных) сообщениях энтропия меньше (мы ссылаемся на формулу (2)), но, следовательно, меньше и количество информации, приходящееся в среднем на один алфавитный символ, использованный в тексте. В осмысленных сообщениях на естественных языках эта величина составляет примерно 1,5 бита. Так что строка «Долины сохнут и пестреют» содержит $1,5 \cdot 24 = 36$ битов информации (пробел тоже отнесем к буквам).

Для нахождения энтропии источника сообщения нужно знать, в частности, вероятности встречаемости в его текстах отдельных букв алфавита. В табл. 14 представлены частоты появления букв в русских текстах. Рекордсменом является О: в отрывках длиной 1000 букв О встречается в среднем 105 раз, а вот твердый знак Ъ (лидер в старой орфографии), скорее всего, не попадется. Как показывают эксперименты, вероятности встречаемости наиболее «частых» букв в субъективной оценке сильно занижаются, а у самых «редких» букв – сильно завышаются. Практически никому не удается правильно угадать первые три и последние три по рангу буквы.

Таблица 14

А	0,0808		И	0,0700		Р	0,0584		Ш	0,0050
Б	0,0170		Й	0,0158		С	0,0466		Щ	0,0035
В	0,0569		К	0,0264		Т	0,0625		Ъ	0,0003
Г	0,0111		Л	0,0284		У	0,0202		Ы	0,0179
Д	0,0388		М	0,0337		Ф	0,0017		Ь	0,0168
Е	0,0836		Н	0,0723		Х	0,0132		Э	0,0017
Ж	0,0096		О	0,1047		Ц	0,0029		Ю	0,0063
З	0,0174		П	0,0371		Ч	0,0180		Я	0,0249

В исследовании творческих процессов предполагают, что вероятность перехода источника сообщения в очередное состояние зависит только от того, какими были r предыдущих его состояний. Такой источник сообщения называется системой марковского типа, а число r – его эргодичностью. (Андрей Андреевич Марков (1850–1922), академик, профессор Петербургского университета. Основные работы относятся к теории чисел, теории вероятностей и математическому анализу. Первым применил статистический метод для анализа художественного произведения – романа А.С.Пушкина «Евгений Онегин».)

Представить себе сообщения марковского типа самого общего вида можно на примере известной игры в составление фраз на заданную тему, когда очередной игрок дописывает слово, подходящее по смыслу, видя только r предыдущих слов. При $r = 0$ (т.е. когда закрыты все ранее написанные слова) обычно получается полная бессмыслица. Но уже при $r = 1$ (марковская цепь) могут возникнуть забавные предложения, как, например, в эксперименте Шеннона:

The head and in frontal attack on an english writer that the character of this point is therefore another method for the letters that the time who ever told the problem for an unexpected

(«Голова и в фронтальной атаке на английского писателя что характер этой точки – следовательно другой метод для букв что время кто когда-либо сообщил эту проблему для непредвиденного». См. конец §VII.2.)

4. *Избыточность* – это величина, характеризующая информативность источника сообщения:

$$R_\Xi = 1 - \frac{H(\Xi)}{H_{\max}}, \tag{3}$$

где H_{max} – энтропия источника сообщения, алфавит которого имеет то же число символов, что и алфавит Ξ, но все они в сообщениях равновероятны. Избыточность показывает долю «лишних» букв в текстах. Она колеблется в пределах от 0 (когда $H(\Xi) = H_{\max}$ – следующим символом в прерванном сообщении с равной вероятностью может быть любая буква) до 1 (когда $H(\Xi) = 0$ – все время передается один символ) и часто выражается в процентах. Снижение избыточности делает передачу сообщений более экономичной (при составлении телеграмм мы опускаем предлоги), но зато менее надежной: в сообщении источника с нулевой избыточностью пропущенный символ не восстановим.

Поскольку во всех реальных каналах связи имеются помехи, разумная избыточность в сообщениях необходима. В естественных языках избыточность превышает 70%. Попробуйте восстановить сообщение, в котором пропущено около 50% букв – все согласные: *еее ау оиоий уае оя оуо*.

А вот как выглядит текст, когда в нем вычеркнуты все гласные, – тоже примерно 50% букв: *блт прс днк в тмн мр глбм*.

Если в последовательности радиосигналов, дошедшей до нас из некоторой точки космического пространства, имеется избыточность, не объясняемая физическими причинами, можно подозревать, что мы имеем дело с некоторым осмысленным сообщением.

5. *Кодирование* – это запись сообщения в виде последовательности различимых сигналов для передачи по каналу связи. Источник сообщения, как правило, представляет текст в алфавите, не удобном для передачи. Поэтому символы алфавита или целые их группы (слова) заменяются символами или группами символов (словами) нового алфавита. Правило, по которому осуществляется эта замена, называется кодом. Всем известен телеграфный код «азбука Морзе», сопоставляющий буквам алфавита наборы коротких («точки») или длинных («тире») звуковых сигналов.

Другими словами, код – это словарь, позволяющий переводить исходное сообщение в текст на кодовом языке.

Закодированное сообщение передается по каналу связи. В целях экономии сообщение кодируют так, чтобы передача занимала меньше времени. Например, часто встречающимся элементам сообщения должны сответствовать по возможности более короткие слова в кодовом алфавите (самая «частая» английская буква «e» в коде Морзе передается одним коротким сигналом – «точкой»). Этот принцип лежит в основе достаточно эффективного приема кодирования – кода Шеннона. Пусть $x_1, x_2, \ldots, x_m$ – кодируемые части («кванты») сообщения, расположенные в порядке убывания вероятностей их встречаемости. Кодовый алфавит будет состоять из двух букв: 0 и 1. Разобьем наш список на две части так, чтобы суммы вероятностей в них были по возможности равны. Всем словам, стоящим в «верхней» половине, присваивается символ 1, а всем попавшим в «нижнюю» половину списка – символ 0. Аналогичную процедуру проводим с каждой из полученных двух групп и т.д. Этот процесс прекращается, когда после очередного разбиения каждая группа будет содержать лишь один из квантов $x_1, x_2, \ldots, x_m$. После этого выписываются символы кодового алфавита, стоящие в строке, соответствующей каждому кванту сообщения x_i,– это и будет кодовое слово для x_i.

ПРИМЕР. Методом Шеннона закодируем сообщение *дадымдумдамдомумодад*.

В сообщении используется шесть символов: *а, д, м, о, у, ы*. Общая длина текста – 20 единиц, *а* встречается 3 раза, значит, вероятность появления этого элемента в тексте равна 0,15. Аналогично для *д, м, о, у, ы* получаем соответственно 0,35; 0,25; 0,10; 0,10; 0,05. Составляем таблицу кодирования (табл. 15).

Таблица 15

Кванты	Вероятности	Разбиение			Коды		
д	0,35	1	1		1	1	
м	0,25	1	0		1	0	
а	0,15	0	1	1	0	1	1
о	0,10	0	1	0	0	1	0
у	0,10	0	0	1	0	0	1
ы	0,05	0	0	0	0	0	0

Пользуясь полученным кодом, легко перевести наш текст на «двоичный» язык:

11011111000101100110110111011010100011001011011111.

Код Шеннона, конечно, допускает однозначное декодирование (этим свойством, очевидно, должен обладать любой код, применяемый для передачи сообщений). Например, последовательность 100001101010011 может представлять в нашем случае лишь сообщение *мыдома*.

При прохождении текста по каналу связи шумы могут исказить сообщение. Существуют коды, которые позволяют при декодировании установить наличие ошибки и даже исправить ее. Простейший пример для исправления одиночной ошибки – код с троекратным повторением. Двоичный текст сообщения передается три раза подряд, и при дешифровке каждый символ восстанавливается «по большинству» – из трех его образов на приемнике. Но, конечно, имеются и более экономичные коды с подобным свойством.

6. *Помехи (шумы)* – это общее обозначение факторов, искажающих сообщение при передаче по каналу связи. В технических линиях связи они, как правило, возникают вследствие физических причин. Если под каналом связи понимать, например, центральную нервную систему человека, работающего у компьютера, то шумами можно считать все, что усложняет процесс восприятия: плохое освещение, посторонние звуки, усталость, неправильное понимание инструкции и т.п.

К шумам можно отнести и многочисленные погрешности, возникающие при неоднократном переписывании древних рукописей. Именно они создали большинство «темных» мест в текстах, дошедших до нас из глубины веков. Искажения при современном компьютерном наборе хотя и не разрушают смысловой ткани, все же сильно влияют на эмоциональное восприятие полиграфической продукции (чего стоят, например, строчки в книге, после переноса начинающиеся с мягкого или твердого знака!).

7. *Поток информации* – скорость выдачи информации устройством линии связи, работающим во временно́м режиме.

Пусть алфавит, в котором записано сообщение, содержит m единиц, вероятности встречаемости которых $p_1, p_2, \ldots, p_m$, а время, необходимое для передачи этих частей текста, соответственно равняется $t_1, t_2, \ldots, t_m$. Тогда средняя длительность t передачи сообщения вычисляется по формуле $t = t_1p_1 + t_2p_2 + \cdots + t_mp_m$.

Количественным выражением потока информации является величина

$$C = \frac{H(\Xi)}{t}, \tag{4}$$

т.е. C – это количество информации, пропускаемой данным устройством линии связи за одну секунду. В часто употребляемых словах «все увеличивающийся поток информации» сетуют не на количественный рост информации, а на скорость, с которой она поступает.

ПРИМЕР. В эксперименте по вычислению скорости выдачи информации читался вслух связный (т.е. понятный читающему) текст с максимально возможной быстротой. Энтропия текста была 1,5 бита. Средняя длина отрывка, прочитываемого за одну секунду, составляла 20 букв. Поток информации $C = 1,5 \cdot 20 = 30$ бит/с. Когда испытуемый читал тот же, но зашифрованный текст (казавшийся ему бессмысленным набором букв), в среднем за секунду произносилось 6 букв, но поскольку энтропия такого текста равна $\log 32 = 5$ битов, то скорость передачи информации остается примерно такой же.

8. *Пропускная способность* – максимальная возможная скорость выдачи информации устройством линии связи, работающим во временно́м режиме. Таким образом,

$$C_0 = maxC = max\frac{H(\Xi)}{t}.$$

Из этого определения видно, что увеличение пропускной способности может быть достигнуто либо увеличением энтропии сообщения, т.е. количества информации в его элементарных частях, либо уменьшением среднего времени передачи этих элементарных частей.

Если число элементарных единиц, при помощи которых записывается сообщение, фиксировано и равно m, то его энтропия не может, как мы знаем, превышать величину $\log m$. С другой стороны, физические возможности данного устройства линии связи ограничивают количество сигналов, пропускаемых им за одну секунду, некоторой величиной W. Значит,

$$C_0 = W \log m. \tag{5}$$

Пропускная способность всей линии связи зависит от величин пропускных способностей ее отдельных устройств и совпадает с наименьшей из них.

В качестве примера приведем расчет пропускной способности органа зрения – глаза. Исходя из того, что 1) зрительный нерв имеет 830 000 волокон, 2) глаз воспринимает 55 мельканий в секунду, 3) каждое волокно за одно мелькание пропускает один бит информации («свет–тень») и 4) все волокна действуют независимо,– получаем пропускную способность органа зрения как канала связи равной $C_0 = 830000 \cdot 55 \cdot 1 = 45000000$ бит/с.

Понятно, что это фантастическое число никак не связано с сознательным восприятием информации зрительной системой в целом. Пропускная способность зрительного анализатора («глаз плюс мозг») была найдена экспериментально следующим образом.

Испытуемому предъявлялись быстро сменяющиеся телевизионные изображения знакомых предметов (или их силуэтов). Расчет велся по формуле (5), представленной в виде

$$C_0 = \frac{n}{T} \log N,$$

где N – количество возможно предъявляемых предметов, T – время предъявления, n – количество правильно опознанных за это время предметов. Полученные по этой и другим методикам величины пропускной способности зрительной системы колеблются в пределах от 20 до 90 бит/с.

Введение количественной меры информации было большим научным достижением. Суть в том, что результаты одного и того же опыта могут вызывать совершенно разные субъективные оценки, а еще сложнее обстоит дело, когда приходится сопоставлять далекие по своему содержанию эксперименты.

Вместе с тем количество информации никак не связано с ее содержательной ценностью: исход праздного бросания монеты и последствие применения лекарства, в половине случаев приносящего больному роковое ухудшение, дают одинаковое количество информации – один бит.

Кроме того, шенноновская мера информации далеко не универсальна: круг ее применений – массовые эксперименты, явления с большим разнообразием. Поэтому словесные рассуждения о количестве информации должны быть достаточно осторожными в тех областях, которые не допускают пока точного формального описания.

§ 6. Нечеткие множества

Математические методы, успешно применяемые в исследовании сложных систем, допускающих точное описание, зачастую оказываются мало пригодными в гуманитарных науках и искусстве. Однозначность в толковании используемых понятий и логическая строгость рассуждений, принятые в математике, находятся в противоречии с естественными языками, не говоря уже о языке литературы, или музыки, или живописи. Вместе с тем системный подход, например, к изучению процессов творчества требует привлечения некоторых общих конструкций, которые не могут возникнуть в конкретных областях, отличающихся большим своеобразием. Для создания абстракций нового типа, которые вместили бы в себя объекты и отношения, не поддающиеся количественным измерениям в обычном понимании, необходима разработка соответствующего математического языка, не скованного постулатами точности в определениях и однозначности в выводах.

Естественно, что первым подверглось пересмотру основное в математике коллективизирующее понятие – «множество». В классическом понимании множества мы о каждом объекте должны точно знать, принадлежит он или не принадлежит данному множеству. Выражения типа «держу пари десять против одного, что вы не найдете Иванова в этом списке» или «на 80 процентов уверен, что там его нет» в теории множеств не обсуждаются.

В 1965 году американский математик Лотфи Заде предложил концепцию «нечеткого множества», признанную на сегодняшний день наиболее отвечающей указанной выше задаче абстрактного изучения сложных гуманитарных систем.

Пусть Ω – некоторое фиксированное непустое множество (универсум). Нечеткое подмножество A в нем выделяется функцией принадлежности $\mu_A : \Omega \to [0,1]$, заданной на Ω и принимающей значения в числовом отрезке $[0,1]$. Таким образом, каждому элементу a универсума Ω сопоставляется число $\mu_A(a)$, которое истолковывается как степень принадлежности этого элемента нечеткому подмножеству A.

ПРИМЕР. Контрольный тест, выполненный студентами одной из групп первого курса за неделю до экзамена, состоял из 5 вопросов. Свою уверенность в том, что данный студент сдаст экзамен, преподаватель выражал числом $0,2{\cdot}n$, где n – количество правильных ответов студента в указанном тесте. Считая студентов, сдавших экзамен, успевающими по предмету, преподаватель своим прогнозом выделяет в группе нечеткое подмножество A потенциально успевающих студентов. Студент Иванов, не решивший ни одной задачи, в это подмножество не входит; студент Петров с двумя правильными ответами может быть отнесен к числу потенциально успевающих с уверенностью 0,4; гораздо выше степень принадлежности нечеткому подмножеству A у студента Сидорова: 0,8; наконец, блестяще выполнившая тест студентка Яковлева признается потенциально успевающей с абсолютным показателем 1. Общая ситуация представлена табл. 16.

Таблица 16

1	1	0,8	0,8	0,6
1	1	0,8	0,6	0,6
0,8	0,8	0,8	0,6	0,4
0,8	0,6	0,6	0,4	0,2
0,6	0,6	0,4	0,2	0

В каждом из квадратиков, соответствующих 25 студентам группы, проставлено число, характеризующее степень их принадлежности нечеткому подмножеству A потенциально успевающих студентов.

Поскольку выделение всякого подмножества в универсуме Ω связано с рассмотрением некоторого свойства для элементов Ω, то можно сказать, что нечетким подмножествам соответствуют нечеткие свойства. В нашем примере нечеткая классификация студентов осуществляется по нечеткому свойству «быть потенциально успевающим студентом».

Чем дальше от точных наук находится область знания, тем менее четкими являются ее классификационные принципы. Термины естественного языка «молодой», «высокий», «богатый», «несколько», «недавно» и т.п. имеют весьма нечеткие границы. Возьмем, например, выражение «несколько предметов». Ясно, что одна книга – это еще не несколько, а двадцать – уже не несколько книг. С какой уверенностью вы отнесете к понятию «несколько» число 2? А 7? Решая этот вопрос, мы тем самым оцениваем степень $\mu(n)$ принадлежности натурального числа n к нечеткому подмножеству, ассоциированному с числительным «несколько» (в Словаре С.И.Ожегова: «некоторое, небольшое количество»). Можно предложить такой вид функции μ:

$$\mu = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 0 & 0,2 & 0,5 & 0,8 & 1 & 1 & 0,8 & 0,4 & 0,1 & 0 \end{pmatrix}$$

и $\mu(n) = 0$ при $n > 10$. Но возможны и другие варианты.

Пусть A и B – нечеткие подмножества универсума Ω, определяемые функциями принадлежности соответственно μ_A и μ_B. Пересечением нечетких подмножеств A и B называется нечеткое подмножество, обозначаемое $A \wedge B$, функция принадлежности которого следующим образом выражается через μ_A и μ_B:

$$\mu_{A\wedge B}(x) = \min\left(\mu_A(x), \mu_B(x)\right)$$

для любого $x \in \Omega$.

Например, если $\Omega = \{1,\ 2,\ 3,\ 4,\ 5\}$,

$$\mu_A = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 0,2 & 0 & 0,8 & 1 & 0,3 \end{pmatrix},\ \mu_B = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 0,1 & 0,3 & 0,9 & 0,6 & 0,6 \end{pmatrix},$$

то

$$\mu_{A\wedge B} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 0,1 & 0 & 0,8 & 0,6 & 0,3 \end{pmatrix}.$$

Объединением нечетких подмножеств A и B называется нечеткое подмножество $A \vee B$ такое, что

$$\mu_{A \vee B}(x) = \max(\mu_A(x), \mu_B(x))$$

для любого $x \in \Omega$.

В нашем примере

$$\mu_{A \vee B} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 0,2 & 0,3 & 0,9 & 1 & 0,6 \end{pmatrix}.$$

Дополнение нечеткого подмножества A обозначается через $\tilde{A}$ и задается функцией принадлежности

$$\mu_{\tilde{A}}(x) = 1 - \mu_A(x)$$

для любого $x \in \Omega$.

Например, для рассматриваемого A будет

$$\mu_{\tilde{A}} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 0,8 & 1 & 0,2 & 0 & 0,7 \end{pmatrix}.$$

Таблица 17

0	0	0,2	0,2	0,4
0	0	0,2	0,4	0,4
0,2	0,2	0,2	0,4	0,6
0,2	0,4	0,4	0,6	0,8
0,4	0,4	0,6	0,8	1

Нечеткое понятие «потенциально неуспевающий студент» (именно оно представляет интерес для деканата) по материалам контрольного теста может быть представлено табл. 17, получающейся из табл. 16.

Как видим, только студент Иванов без сомнений признается потенциально неуспевающим.

Обычные подмножества универсума Ω можно воспринимать как его нечеткие подмножества, функция принадлежности которых принимает только два значения: 1 (да) и 0 (нет). Тогда операции пересечения, объединения и дополнения нечетких подмножеств, примененные к обычным подмножествам в указанной интерпретации, превращаются по существу в одноименные операции над подмножествами, введенные в §1.

Почти все тождества алгебры множеств имеют место и для нечетких подмножеств (с соответствующей заменой операций). К сожалению, не сохраняются важнейшие свойства дополнения: на примере нечетких подмножеств, рассмотренных при определении операций, убеждаемся, что в общем случае $A \wedge \tilde{A} \neq \emptyset$ и $A \vee \tilde{A} \neq \Omega$ (здесь $\emptyset$ истолковывается как нечеткое подмножество с нулевой функцией принадлежности, а Ω – как нечеткое подмножество, у которого функция принадлежности во всех точках равна 1). Следовательно, алгебра нечетких подмножеств не является булевой алгеброй. Но, конечно, $\tilde{\tilde{A}} = A$ и, как ни удивительно, остаются справедливыми и законы Де Моргана: $(A \wedge B)^{\sim} = \tilde{A} \vee \tilde{B}$ и $(A \vee B)^{\sim} = \tilde{A} \wedge \tilde{B}$ (попробуйте придать им содержательное толкование).

В реальных ситуациях довольно часто на основе нечетких представлений приходится делать вполне однозначные заключения. Вряд ли деканат, озабоченный накануне сессии прогнозом успеваемости студентов, удовлетворится информацией типа табл. 17. Преподавателю, составлявшему сводку, придется каждого студента четко отнести либо к потенциально успевающим, либо к потенциально неуспевающим. Он может сделать это чисто механически при помощи следующего процесса детерминизации нечетких подмножеств. Пусть A – нечеткое подмножество с функцией принадлежности μ_A и пусть ε – положительное число, не превосходящее 1 (т.е. $0 < \varepsilon \leqslant 1$). Под ε-детерминизатором нечеткого подмножества A понимается нечеткое подмножество $A(\varepsilon)$ такое, что

$$\mu_{A(\varepsilon)}(x) = \begin{cases} 1, \text{если } \mu_A(x) \geqslant \varepsilon; \\ 0, \text{если } \mu_A(x) < \varepsilon \end{cases}$$

для любого $x \in \Omega$.

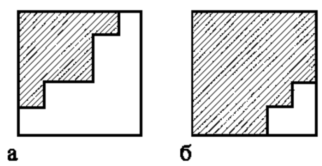


Рис. 83

Таким образом, $A(\varepsilon)$ – это обычное (не нечеткое) подмножество универсума Ω, состоящее в точности из тех элементов $x \in \Omega$, которые могут быть отнесены к A с уверенностью не меньшей, чем ε. Если, например, положить $\varepsilon = 0,7$, то подмножество потенциально успевающих студентов будет представлено на рис. 83 диаграммой Венна а), а при $\varepsilon = 0,4$ – диаграммой Венна б). Вторая точка зрения намного оптимистичней.

Число ε называется уровнем детерминизации. В нейтральных ситуациях обычно полагают $\varepsilon = 0,5$. Например, при таком уровне детерминизатором нечеткого понятия «несколько» (см. выше) будет четкое «от трех до семи включительно».

На основе понятия нечеткого подмножества за последние 30 лет развита обширная «нечеткая математика», включающая в себя сответствующую логику, теорию отношений, алгоритмов, принятия решений и т.д. Она применяется в качестве языка описаний во многих гуманитарных областях.

ГЛАВА VI. **ДИСКРЕТНЫЕ СИСТЕМЫ И ИХ МАТЕМАТИЧЕСКОЕ ОПИСАНИЕ**

§ 1. Отношения

В отличие от классической математики в современных исследованиях большое место занимают системы дискретного характера, т.е. такие, при описании которых не используется идея непрерывности. Чаще всего это системы с конечным числом элементов.

Язык теории отношений является одним из самых разработанных способов описания дискретных систем. Само понятие отношения между двумя объектами носит столь универсальный характер, что невозможно представить себе область человеческой деятельности, где оно не проявлялось бы в виде тех или иных конкретных связей.

Пусть A и B – некоторые непустые множества. Их декартовым произведением называется множество $A \times B$ (математики читают: «A крест B»), состоящее из всевозможных упорядоченных пар вида (a, b), где на первом месте стоит элемент множества A, а на втором – элемент из множества B. Декартово произведение $A \times A$ множества A на себя называют декартовым квадратом множества A. Например, числовая плоскость $\mathbb{R}^2$ – это декартов квадрат числовой прямой: $\mathbb{R}^2$=$\mathbb{R}\times\mathbb{R}$. Если множества A и B конечны и имеют соответственно m и n элементов (т.е. $|A| = m$, $|B|$=n), то, как было установлено в §V.2, в декартовом произведении $A\times B$ содержится mn элементов, т.е. $|A \times B| = mn = |A| \cdot |B|$.

ПРИМЕР. Составим декартово произведение множеств $A = \{a, b, c\}$ и $B = \{1, 2\}$. Чтобы перечислить все пары, образующие множество $A\times B$, выпишем сначала те, у которых на первом месте стоит элемент a, затем пары с первым элементом b и, наконец, пары, начинающиеся с c. Пары с одинаковым первым элементом располагаем в порядке следования их вторых элементов в записи множества B. Итак, $A \times B$=$\{(a, 1), (a, 2), (b, 1), (b, 2), (c, 1), (c, 2)\}$.

Отношением между элементами множеств A и B называется всякое подмножество декартова произведения $A \times B$. Другими словами, отношение – это некоторый набор упорядоченных пар. Отношения будем обозначать малыми греческими буквами: ρ, σ, τ и т.п.

Среди всех отношений между элементами множеств A и B выделяются два крайних случая: пустое отношение $\emptyset$, которое не содержит ни одной пары, и универсальное отношение, в состав которого входят все пары вида (a, b), где $a \in A$, $b \in B$. Очевидно, что универсальное отношение совпадает с декартовым произведением $A \times B$. Для любого отношения ρ между элементами множеств A и B имеют место включения

$$\emptyset \subseteq \rho \subseteq A \times B.$$

Если пара (a, b) входит в состав отношения ρ, т.е. если $(a, b) \in \rho$, то говорят, что элемент a находится в отношении ρ с элементом b.

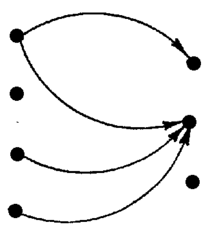


Рис. 84

Отношения удобно представлять с помощью графов. Для этого элементы множеств A и B изображают точками и из точки a в точку b проводят ориентированную линию (дугу) в том и только том случае, когда $(a, b) \in \rho$. Вспомним, что графы уже появлялись у нас в §II.1 в связи с понятием функции. Рассматривая приведенные там примеры функций авторства для литературных персонажей и картин, приходим к выводу, что функции эти, как и функции вообще, можно считать частным случаем отношений: когда каждый элемент множества A находится в данном отношении точно с одним элементом множества B. Общую идею отношения представляет пиктограмма рис. 84. Из нее видно, что элемент может находиться в данном отношении с одним или несколькими элементами, а может и не быть в данном отношении ни с одним элементом. С другой стороны, с каждым данным элементом в рассматриваемом отношении могут состоять один или несколько элементов, но может и вообще не оказаться таких элементов (укажите на пиктограмме все эти ситуации).

Подмножества декартова квадрата $A \times A$ называются отношениями на множестве A. Выделим основные свойства, которые обычно рассматриваются в практике при исследовании отношений на тех или иных конкретных множествах.

Пусть A – некоторое непустое множество и ρ – отношение на нем.

1) Отношение ρ называется рефлексивным, если каждый элемент множества A состоит в этом отношении сам с собой, т.е. если $(\forall x \in A)((x, x) \in \rho)$.

Например, какие бы свойства ни обсуждались для элементов множества A, отношение «иметь одинаковые свойства» рефлексивно, ибо каждый элемент находится в этом отношении прежде всего сам с собой. (Из математических афоризмов Штейнгауза: «Пан А. избегает учебников алгебры с тех пор, как в одном из них он обнаружил аксиому: A=A».)

Граф рефлексивного отношения имеет в каждой вершине петлю, т.е. дугу с совпадающими началом и концом.

2) Отношение ρ называется антирефлексивным, если никакой элемент множества A не состоит в этом отношении сам с собой, т.е. если $(\forall x \in A)((x, x) \notin \rho)$.

Например, «быть потомком» или «иметь больший доход» – антирефлексивные отношения, ибо никакой субъект не является своим потомком и не имеет большего дохода, чем тот, который он имеет. Всякое отношение, устанавливающее различие между элементами множества по каким бы то ни было признакам, будет антирефлексивным.

Граф антирефлексивного отношения не имеет петель.

Заметим, что свойство антирефлексивности несовместимо с рефлексивностью, но не является противоположным для нее свойством, ибо (вспомним исчисление предикатов) высказывание $\neg(\forall x \in A)((x, x) \in \rho)$, отрицающее рефлексивность, равносильно высказыванию $(\exists x \in A)((x, x) \notin \rho)$, что, конечно, не совпадает с антирефлексивностью (средневековые логики оживленно обсуждали эти проблемы).

3) Отношение ρ называется симметричным, если из того, что элемент x находится в отношении ρ с элементом y, следует, что и y состоит в этом отношении с x, т.е. если $(\forall x, y \in A)((x, y) \in \in \rho \to (y, x) \in \rho)$.

Житейские отношения «быть знакомым» (т.е. «x знаком с y») или «состоять в родстве» симметричны, но этим свойством, увы, не обладает одно из важнейших человеческих отношений «любить» (мировая литература дает бесчисленные трагические примеры несимметричности этого отношения).

В графе симметричного отношения каждая дуга имеет встречную дугу, т.е. дугу между теми же вершинами, но в противоположном направлении (петля считается встречной дугой сама для себя).

4) Отношение ρ называется антисимметричным, если ни для каких двух различных элементов $x, y \in A$ не могут одновременно выполняться связи $(x, y) \in \rho$ и $(y, x) \in \rho$, т.е. если $(\forall x, y \in A)((x, y) \in \rho \wedge (y, x) \in \rho \rightarrow x = y)$.

Например, не существует двух человек, которые были бы старше друг друга по возрасту, так что «быть старше» – антисимметричное отношение, так же как и отношение $\leqslant$ на числовой прямой $\mathbb{R}$ (но первое отношение, кроме того, еще и антирефлексивно, а второе, напротив, рефлексивно).

В графе антисимметричного отношения нет встречных дуг, кроме, может быть, петель.

5) Отношение ρ называется транзитивным, если из того, что в отношении ρ находятся x с y и одновременно y с z, следует, что x состоит в данном отношении с z, т.е. если $(\forall x, y, z \in A)((x, y) \in \in \rho \wedge (y, z) \in \rho \rightarrow (x, z) \in \rho)$.

Транзитивным будет, например, отношение подчиненности в служебных иерархиях. Свойство транзитивности тех или иных отношений часто провозглашалось в качестве общественных или моральных принципов («вассал моего вассала – мой вассал» или «друг моего друга – мой друг»). В каких из приведенных выше примеров отношений имеет место транзитивность?

На графе отношения свойство транзитивности распознается следующим образом: если из вершины x в вершину z есть двузвенный (т.е. составленный из двух дуг) путь, то должен быть и однозвенный (т.е. состоящий из одной дуги) путь из x в z.

Комбинируя в разных вариантах перечисленные свойства, получают важнейшие типы отношений на множествах.

Пусть A – непустое множество. Отношение τ на нем называется отношением толерантности или, для краткости, толерантностью, если оно рефлексивно и симметрично. Стандартный пример толерантности получается следующим образом. Допустим, что для элементов множества A рассматривается некоторый набор свойств (признаков). Говорят, что элемент a похож на элемент b, если a имеет с b хотя бы одно общее свойство из указанного списка свойств. Так определенное отношение сходства, очевидно, является толерантностью.

Пара (A, τ), где A – непустое множество, а τ – толерантность на нем, называется пространством толерантности.

Студенческую группу A можно превратить в пространство толерантности многими способами. Для студентов $a, b \in A$ положим $(a, b) \in \tau_1$ тогда и только тогда, когда a состоит в приятельских отношениях с b. С некоторой натяжкой считая, что каждый сам себе приятель, получаем пространство толерантности (A, τ_1). Другой способ введения структуры толерантности на A: считать $(a, b) \in \tau_2$ тогда и только тогда, когда в прошедшую сессию студенты a и b получили одинаковую оценку хотя бы по одному предмету. Еще одна толерантность: $(a, b) \in \tau_3$ означает, что a имеет ту же группу крови, что и b.

Пусть (A, τ) –пространство толерантности. Под классом толерантности τ, или τ-классом, понимается подмножество $X \subseteq A$ такое, что 1) любые два элемента a, b из X толерантны, т.е. $(a, b) \in \tau$; 2) всякий элемент из A, не входящий в X, нетолерантен хотя бы с одним элементом из X. Второе условие означает, что X нельзя расширить, сохраняя свойство всеобщей толерантности между элементами.

Граф отношения толерантности рисуют, не изображая петли (они имеются во всех вершинах) и заменяя каждую пару встречных дуг одной неориентированной линией (ребром). Классам толерантности соответствуют так называемые полные графы, в которых каждые две вершины соединены ребром. На рис. 85 показаны полные графы, имеющие 2, 3, 4, 5 и 6 вершин.

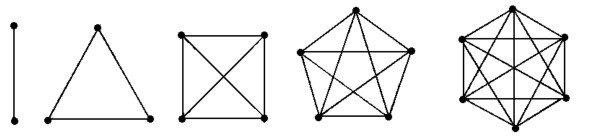


Рис. 85

Покрытием множества A называется такой набор его подмножеств, объединение которого дает все множество A. Подмножества, образующие покрытие, называют его блоками. Классы любой толерантности τ образуют покрытие множества A. Это следует из того, что каждый элемент множества A содержится по крайней мере в одном τ-классе.

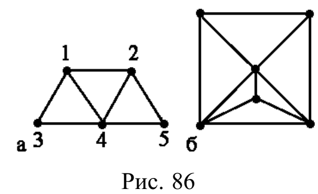


Рис. 86

На рис. 86а покрытие множества $A = \{1, 2, 3, 4, 5\}$ образуют классы толерантности $A_1 = \{1, 2, 4\}$, $A_2 = \{2, 4, 5\}$, $A_3 = \{1, 3, 4\}$. При этом мы замечаем, что класс A_1 является как бы «лишним»: входящие в него элементы уже распределены по другим классам. Исключая A_1, получим так называемое минимальное покрытие множества A – классами толерантности A_2 и A_3. Минимальные покрытия замечательны тем, что в каждом блоке такого покрытия найдется элемент, не входящий ни в какой другой блок (в A_2 это будут 2 и 5, а в A_3 – элементы, обозначенные символами 1 и 3). Рис. 86б показывает, что толерантность может давать минимальные покрытия с разным числом блоков (в данном случае с двумя или тремя блоками – укажите эти покрытия).

К числу важнейших математических конструкций принадлежит понятие эквивалентности. Отношение ε на непустом множестве A называется отношением эквивалентности, или, для краткости, эквивалентностью, если оно рефлексивно, симметрично и транзитивно. Другими словами, эквивалентности – это транзитивные толерантности. Из рассмотренных примеров толерантностей в студенческой группе A транзитивным будет только отношение «иметь ту же группу крови». Приятель вашего приятеля не обязательно окажется приятелем и для вас; нетрудно придумать пример, показывающий нетранзитивность отношения «получить одинаковую оценку хотя бы по одному предмету», если предметов не меньше двух.

Примеры эквивалентностей можно указывать без конца в любой области, везде, где есть какая-либо классификация объектов. Распределение людей по полу, возрасту, национальности, гражданству, знаку зодиака, вероисповеданию, группе крови – за всем этим стоит следующая математическая схема: имеется непустое множество A, и для его элементов рассматривается некоторый признак, принимающий определенные значения; элемент a считается находящимся в отношении ε с элементом b, если значение признака у a – такое же, как и у b. Нетрудно понять, что так определенное отношение ε между элементами множества A рефлексивно, симметрично и транзитивно, т.е. является отношением эквивалентности.

Важнейшая геометрическая эквивалентность – равенство плоских фигур: если зафиксирована некоторая группа преобразований плоскости, то фигура Φ считается равной фигуре Ψ, если существует преобразование f данной группы, переводящее Φ в Ψ, т.е. такое, что $f(\Phi)=\Psi$. Понятие равномощности множеств также обладает всеми свойствами эквивалентности. Пример из логики – равносильность формул. (В каждом из этих случаев проверьте выполнимость условий рефлексивности, симметричности и транзитивности.)

Пространством эквивалентности называется пара (A, ε), где A – непустое множество, а ε – отношение эквивалентности на нем.

Так как эквивалентность ε представляет собой толерантность (с дополнительным свойством транзитивности), то ε-классы образуют покрытие множества A. Но в отличие от общего случая здесь имеется одна важная особенность. Покрытие множества A называется разбиением этого множества, если блоки покрытия попарно не пересекаются, т.е. не имеют общих элементов.

Теорема. Пусть ε – отношение эквивалентности на множестве A. Тогда ε-классы образуют разбиение множества A. С другой стороны, если задано некоторое разбиение множества A, то отношение «находиться в одном блоке» является эквивалентностью на множестве A.

ДОКАЗАТЕЛЬСТВО. 1) Пусть ε – эквивалентность на множестве A. Предположим, что элемент c входит одновременно в ε-класс A_1 и в ε-класс A_2. Возьмем произвольные элементы $a \in A_1$ и $b \in A_2$. Так как a и c находятся в одном ε-классе A_1, то $(a, c) \in \varepsilon$. Так как c и b находятся в одном ε-классе A_2, то $(c, b) \in \varepsilon$. В силу транзитивности, из $(a, c) \in \varepsilon$ и $(c, b) \in \varepsilon$ следует, что $(a, b) \in \varepsilon$, т.е. что a и b находятся в одном ε-классе. Поскольку a и b были выбраны произвольно в A_1 и A_2, получаем, что A_1 и A_2 состоят из одних и тех же элементов, т.е. что A_1=A_2. Таким образом, разные ε-классы не могут иметь общих элементов. Тем самым доказано, что покрытие множества A, образованное ε-классами, является разбиением.

2) Пусть имеется некоторое разбиение множества A и $(a, b) \in \varepsilon$ означает, что a находится в том же блоке разбиения, что и b. Рефлексивность и симметричность этого отношения очевидны. Пусть теперь $(a, b) \in \varepsilon$ и $(b, c) \in \varepsilon$. Это означает, что a и b находятся в одном блоке, скажем, в A_1, и b и c находятся в одном блоке, например, в A_2. Как видим, блоки A_1 и A_2 имеют общий элемент b, а так как разные блоки разбиения не пересекаются, то A_1=A_2. Отсюда следует, что a и c находятся в одном блоке, т.е. что $(a, c) \in \varepsilon$. Доказана и транзитивность отношения ε, которое, таким образом, будет эквивалентностью. □

Рассматривая классы эквивалентности, мы отвлекаемся от индивидуальных особенностей объектов, выделяя лишь то общее, что объединяет их в один класс. Тем самым достигается новый, более высокий уровень абстракции. Так, за понятием «число 2» стоит необозримый класс равномощных множеств, сходных лишь тем, что в каждом из них два элемента. Абстрактное понятие «единичный квадрат» реализуется в каждом из квадратов со стороной 1, в каком бы положении на плоскости он ни находился. Биологические термины «мужчина» и «женщина»; демографические статусы «гражданин России», «москвич», «русский», «служащий»; астрологические символы «телец», «лев», «скорпион» и т.д. – это не что иное, как обозначения классов соответствующих эквивалентностей.

К эквивалентностям приводят и классификации по отношению к нескольким признакам, когда в один класс объединяются объекты, обладающие одинаковым набором признаков. Так, если для элементов множества Ω имеет смысл обсуждать три признака A, B, C, то соответствующая эквивалентность разобьет Ω на 8 классов – подобные классификации обсуждались в §V.I.

Третий тип отношений, широко распространенных в практике, – это порядки. Отношение ω на множестве A называется отношением порядка, или, для краткости, порядком, если оно рефлексивно, антисимметрично и транзитивно. Стандартный пример – отношение $\leqslant$ («не больше») на числовой прямой. Действительно, для любых чисел $x, y, z \in \mathbb{R}$ имеем: 1) $x \leqslant x$ (рефлексивность: каждое число не больше себя), 2) $x \leqslant y \ \wedge \ y \leqslant x \to x = y$ (антисимметричность), 3) $x \leqslant y \ \wedge \ y \leqslant z \to x \leqslant z$ (транзитивность).

Пара (A, ω), где A – непустое множество, а ω – порядок на нем, называется упорядоченным множеством. Числовые упорядоченные множества $(\mathbb{N}, \leqslant)$, $(\mathbb{Z}, \leqslant)$ и $(\mathbb{R}, \leqslant)$ всем знакомы со школьных лет.

Антирефлексивная часть отношения порядка называется строгим порядком. Для чисел строгий порядок обозначается символом $<$ («меньше»).

Отношение порядка дает математическое описание идеи сравнения объектов по тем или иным критериям.

Пусть, например, дано некоторое множество Ω и пусть для его подмножеств X, Y запись $X \subseteq Y$ (читается «X включается в Y» или «X содержится в Y») означает, что каждый элемент из X входит в Y, т.е. подмножества сравниваются по составу элементов. Для отношения включения $\subseteq$ без труда проверяются свойства порядка: 1) $X \subseteq X$, 2) $X \subseteq Y \wedge Y \subseteq X \to X = Y$, 3) $X \subseteq Y \wedge Y \subseteq Z \to X \subseteq Z$.

Естественный порядок $\leqslant$ на числовых множествах обладает свойством линейности: для любых чисел x, y обязательно выполняется одно из неравенств $x \leqslant y$ или $y \leqslant x$. Отношение включения $\subseteq$ не является линейным порядком: из двух произвольно выбранных подмножеств X, Y не обязательно одно содержится в другом. (Будет ли линейным порядком отношение служебной подчиненности в учреждении?)

Для наглядного представления конечных упорядоченных множеств (A, ω) используют так называемые диаграммы, которые рисуются следующим образом. Точки, изображающие элементы множества A, располагаются так, что чем больше элемент, тем выше находится соответствующая точка. При этом точка a соединяется прямолинейным отрезком с другой точкой b, если в (A, ω) элемент b непосредственно следует за элементом a в том смысле, что $(a,b) \in \omega$ и не существует отличного от a и b элемента x такого, что $(a,x) \in \omega$ и $(x,b) \in \omega$ (т.е. если между a и b нельзя «вставить» никакой промежуточный по порядку ω элемент). На рис. 87 показаны диаграммы упорядоченных множеств $(N_6, \leqslant)$, $(N_6, |)$ и $(P(\Omega), \subseteq)$, где $N_6 = \{1,2,3,4,5,6\}$ (первые шесть натуральных чисел); $x|y$ означает, что число x является делителем числа y; $\Omega = \{1,2,3\}$, $P(\Omega)$ – множество всех подмножеств множества Ω.

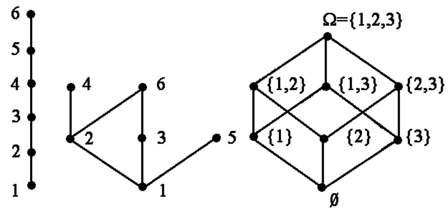


Рис. 87

При построении диаграмм существует произвол в расположении точек на плоскости, их стараются разместить так, чтобы рисунок был понятным и выглядел, по возможности, привлекательным (изобразите диаграммы рис. 87 другими способами).

Если $(a,b) \in \omega$, то на диаграмме это распознается тем, что от вершины a к вершине b имеется восходящая ломаная.

Диаграмма отношения «быть потомком» на множестве лиц, ведущих свое происхождение от какого-либо человека, называется родословным деревом. Отрезки родословного дерева соединяют детей с их родителями.

На рис. 88 изображено родословное дерево математической династии Бернулли (два Николая не были профессиональными

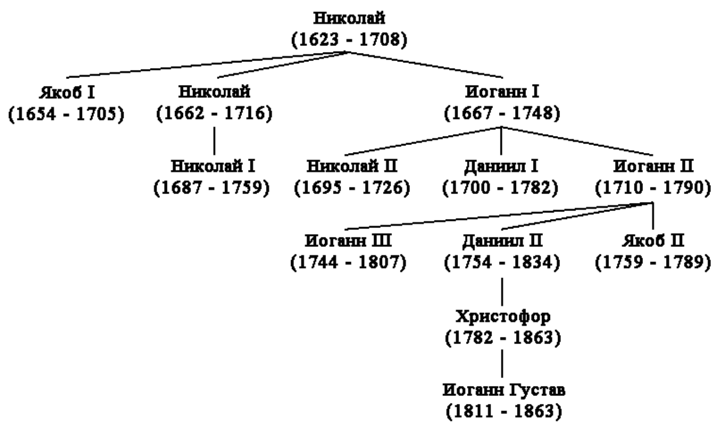


Рис. 88

математиками, и потому традиция не включает их в нумерацию одинаковых имен). Из 11 представителей этого клана Якоб I, Иоганн I и Даниил I принадлежат к числу выдающихся ученых, оказавших огромное влияние на развитие математики и физики. Почетный член Петербургской академии наук Даниил I Бернулли имел много громких титулов, но писал только один: «сын Иоганна Бернулли».

§ 2. Графы

Графы уже встречались у нас как вспомогательное средство для задания функций и отношений. Однако теория графов в настоящее время является вполне самостоятельной ветвью математики, одним из мостов между ее теоретическими и прикладными разделами.

Под графом понимается пара $G = (V, \alpha)$, где V – конечное непустое множество, а α – отношение на нем. Элементы множества V называются вершинами графа G, а пары, входящие в состав α,– дугами. При этом, если (u, v) – дуга, то говорят, что вершина u является ее началом, а вершина v – концом. Дуги, у которых начало и конец совпадают, называются петлями.

Графы с небольшим числом вершин удобно представлять в виде рисунка: вершины изображаются точками (или кружками, квадратиками и т.п.), а дуги – стрелками, соединяющими – от начальной к конечной – соответствующие вершины.

Графы используют в качестве моделей систем в самых разных областях человеческой деятельности. Вершины соответствуют элементам системы, а дуги – связям между элементами. Рассмотрим, например, систему, состоящую из пяти элементов, функциональная схема которой приведена на рис. 89.

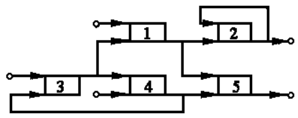


Рис. 89

Эта схема интерпретируется каждым из двух графов, изображенных на рис. 90. Вершины сохраняют названия (или номера) соответствующих элементов, входные и выходные каналы отсутствуют, информацию о связи между элементами несут дуги.

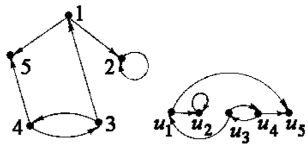


Рис. 90

Желание (или необходимость) по-разному изображать один и тот же граф, а также присваивать новые имена его вершинам заставляет искать принцип идентификации графов более широкий, чем тождественное совпадение. Так возникает понятие изоморфизма графов. Графы $G_1 = (V_1, \alpha_1)$ и $G_2 = (V_2, \alpha_2)$ называются изоморфными (в записи: $G_1 \cong G_2$), если между их вершинами (т.е. между множествами V_1 и V_2) можно установить взаимно однозначное соответствие $\varphi : V_1 \to V_2$ таким образом, чтобы из вершины u в вершину v в графе G_1 вела дуга тогда и только тогда, когда в графе G_2 дуга соединяет вершину $\varphi(u)$ с вершиной $\varphi(v)$. Биекция φ называется изоморфизмом графа G_1 на граф G_2. Можно сказать, что два графа изоморфны в том и только в том случае, когда они допускают одно и то же немое (т.е. без обозначения вершин) изображение. (Укажите изоморфизм одного из графов, изображенных на рис. 90, на другой.)

Если отношение α симметрично, то каждая дуга (u, v) графа $G = (V, \alpha)$ имеет встречную дугу (v, u). По соглашению, в этом случае, изображая граф, вершины u и v соединяют одной неориентированной линией, которая называется ребром. Графы с симметричным и антирефлексивным отношением α называются неориентированными графами. Обычно слово «граф» относят именно к неориентированнным графам, применяя к ориентированным графам термин «орграф». Но из контекста, как правило, ясно, о каких именно графах идет речь.

Введение нового объекта исследования вызывает интерес теми задачами, которые при этом возникают. Простая конструкция: некоторое множество точек плоскости, выборочно соединенных линиями,– оказалась весьма содержательной как в теоретическом, так и в прикладном аспекте.

Первой теоремой о графах было решение Эйлером задачи о кенигсбергских мостах (1736 г.). На реке Прегель в Кенигсберге имелось семь мостов, соединявших берега с островами. Вопрос, волновавший просвещенных жителей столицы Восточной Пруссии, состоял в том, можно ли пройти последовательно через все мосты, побывав на каждом из них только один раз. Удастся ли сделать это так, чтобы в конце маршрута оказаться в его исходной точке?

Изобразив острова и берега точками, а мосты между ними линиями, Эйлер получил неориентированный граф и общую формулировку задачи: как узнать, можно ли пройти по всем ребрам графа в точности по одному разу? Такой путь в графе теперь называют эйлеровым, а граф, допускающий эйлеров путь с возвратом в исходную вершину,– эйлеровым графом. Заметим, что, двигаясь по эйлерову пути, мы имеем право посещать одну и ту же вершину несколько раз: важно, чтобы не повторялись ребра.

Эйлер нашел простой и эффективный критерий существования требуемых путей в графе. Чтобы сформулировать соответствующий результат, введем еще несколько понятий. Степенью вершины в неориентированном графе называется количество ребер, которым она принадлежит (иначе говоря, количество вершин, с которыми она соединена). Вершину называют четной или нечетной в зависимости от того, четным или нечетным числом будет ее степень. Под связным графом понимается такой неориентированный граф, в котором можно пройти по ребрам из любой данной вершины в любую другую.

Теорема 1. В связном графе G тогда и только тогда существует эйлеров путь, соединяющий две данные вершины, когда эти вершины являются единственными нечетными вершинами графа G. Связный граф будет эйлеровым тогда и только тогда, когда все его вершины – четные. □

В кенигсбергском графе, объяснил Эйлер, четыре нечетные вершины – желанная прогулка по мостам неосуществима.

Плоская линия называется уникурсальной, если ее можно нарисовать «одним росчерком пера», т.е. не отрывая карандаш от бумаги, и притом не проходя никакого участка более одного раза (точки самопересечения кривой допускаются). Эйлеров путь в графе образует уникурсальную линию, проходящую по всем его ребрам. Какие из графов, изображенных на рис. 91, являются эйлеровыми, в каких существует эйлеров путь между двумя разными вершинами? Нарисуйте соответствующие уникурсальные линии (известная салонная игра XIX века).

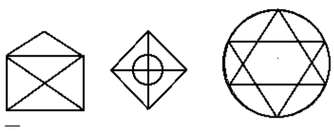


Рис. 91

Эйлеровы пути играют важную роль при конструировании сложных технических устройств и систем наземных коммуникаций.

Совсем по-другому сложилась судьба другой классической задачи теории графов. Выдающийся ирландский математик и механик Уильям Роуан Гамильтон (1805–1865), один из создателей векторного исчисления, в 1859 году предложил игру: пройти по ребрам додекаэдра так, чтобы посетить каждую вершину точно один раз и вернуться в исходную (все ребра проходить не обязательно). В связи с этим возник вопрос о существовании в графах подобных путей (они называются гамильтоновыми циклами). Гамильтонов граф – это граф, в котором имеется гамильтонов цикл.

Всякий полный граф с числом вершин не меньше трех является гамильтоновым: если взять правильный n-угольник и провести все его диагонали, получится полный n-вершинный граф; гамильтоновым циклом в нем будет, например, исходный многоугольник.

С поиском гамильтонова цикла в графе связана так называемая задача о коммивояжере: приказчик, путешествующий для распространения изделий фирмы (принимая заказы на основании образцов), должен найти кратчайший путь, чтобы объехать некоторое количество городов и вернуться в исходный пункт.

Несмотря на большое прикладное значение и связанные с этим значительные усилия специалистов, проблема существования гамильтонова цикла в графе до сих пор не получила эффективного решения (прямой путь – перебирать все возможные варианты – в графах с большим числом вершин и ребер практически неосуществим).

Изображение графа называется плоским, если в нем ребра не имеют общих точек, кроме, может быть, вершин. Если граф допускает плоское изображение, его называют планарным. Рис. 92 дает примеры неплоского и плоского изображения двух планарных графов. В примере б) проставьте номера вершин, чтобы убедиться в изоморфности графов.

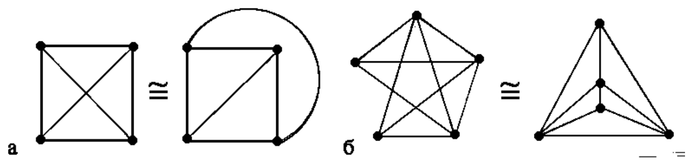


Рис. 92

Электрическую схему можно нанести печатным способом на одностороннюю плату тогда и только тогда, когда граф, представляющий схему, является планарным. Таким образом, прежде чем приступать к изготовлению печатной схемы, нужно найти плоское изображение соответствующего графа. Но всегда ли оно существует? Многим памятны бесплодные детские усилия провести непересекающиеся тропинки от каждого из трех сказочных домиков к каждому из трех волшебных колодцев, т.е. найти плоское изображение графа, обозначаемого $K_{3,3}$ (рис. 93а). Столь же безуспешными окажутся ваши попытки нарисовать без самопересечений полный пятивершинный граф K_5 (рис. 93б).

Графы $K_{3,3}$ и K_5 оказались ключевыми конфигурациями для решения проблемы планарности. Назовем графом типа $K_{3,3}$ граф, полученный из $K_{3,3}$ добавлением произвольного числа вершин на его ребрах вне точек их пересечения. Аналогично определяется граф типа K_5. Критерий планарности графа, полученный независимо Л.С.Понтрягиным и польским математиком Казимежем Куратовским (1896–1980), утверждает, что граф тогда и только тогда будет планарным, когда из него нельзя удалить какие-нибудь вершины и ребра так, чтобы остался граф типа $K_{3,3}$ или граф типа K_5. Отсюда следует, в частности, что все графы типов $K_{3,3}$ и K_5, а значит, и сами эти исходные графы не являются планарными.

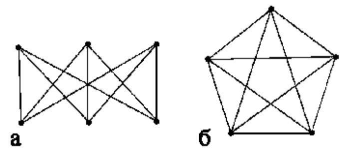


Рис. 93

Широко применяются в технических, естественнонаучных и гуманитарных приложениях планарные графы, называемые деревьями. Дерево – связный граф, не имеющий в своем составе циклов (т.е. графов, которые можно нарисовать в виде правильного многоугольника). На рис. 94 показаны все существенно различные (т.е. неизоморфные) деревья с числом вершин 1, 2, 3, 4 и 5.

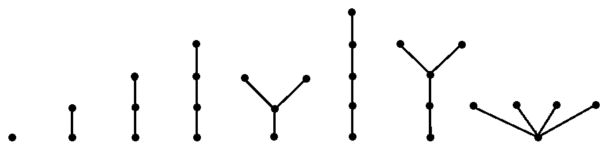


Рис. 94

Нарисуйте все (их 6) существенно различные деревья с 6 вершинами. Родословное дерево потомков некоторого человека как граф не обязательно будет деревом, но если ограничиться нисходящими только одного пола, циклы не появятся (поясните).

Гранью в плоском изображении планарного графа называется область, ограниченная какими-либо его ребрами. К числу граней относится и бесконечная часть плоскости, охватывающая граф. На рис. 92а плоский граф имеет 4 грани, а на рис. 92б число граней равно 6. У любого дерева одна грань.

Теперь мы имеем все необходимое для формулировки и доказательства знаменитой формулы Эйлера.

Теорема 2. В плоском изображении связного планарного графа число вершин В, число ребер Р и число граней Г связаны соотношением $\text{В} - \text{Р} + \Gamma = 2$.

ДОКАЗАТЕЛЬСТВО. Здесь у нас появляется новый тип математического доказательства – методом сведения к некоторому простейшему частному случаю.

1) Пусть рассматриваемый граф – дерево. Так как в этом случае $\Gamma = 1$, а $\text{В} = \text{Р} + 1$ (вершин на одну больше, чем ребер), то $\text{В} - \text{Р} + \Gamma = (\text{Р} + 1) - \text{Р} + 1 = 2$. Таким образом, формула Эйлера верна в простейшем частном случае деревьев.

2) Пусть G – произвольный связный планарный граф, предъявленный в плоском изображении, и пусть он имеет В вершин, Р ребер и Г граней. Если G – дерево, то формула Эйлера для него верна, как показано в 1). Предположим, что G не дерево. Тогда в нем обязательно есть хотя бы один цикл. Удалим из графа G какое-нибудь ребро, входящее в один из циклов (концевые вершины ребра остаются!). В полученном графе G_1 число вершин В_1 то же, что и в G, т.е. $\text{В}_1 = \text{В}$; число ребер Р_1 на 1 меньше (за счет удаленного), т.е. $\text{Р}_1 = \text{Р} - 1$, число граней Γ_1 на 1 меньше (при удалении ребра сольются разделявшиеся им грани), т.е. $\Gamma_1 = \Gamma - 1$. Таким образом, $\text{В}_1 - \text{Р}_1 + \Gamma_1 = \text{В} - (\text{Р} - 1) + (\Gamma - 1) = \text{В} - \text{Р} + \Gamma$. Граф G_1 связный (удаление ребра, входящего в цикл, не нарушает связности). Если он не дерево, найдем в нем цикл и удалим из этого цикла одно ребро. В полученном графе G_2 будет $\text{В}_2 = \text{В}_1$, $\text{Р}_2 = \text{Р}_1 - 1$, $\Gamma_2 = \Gamma_1 - 1$, откуда $\text{В}_2 - \text{Р}_2 + \Gamma_2 = \text{В}_1 - \text{Р}_1 + \Gamma_1$. Продолжим этот процесс до тех пор, пока на k-м шаге не получится граф G_k, не имеющий циклов и, значит, являющийся деревом. В силу 1), $\text{В}_k - \text{Р}_k + \Gamma_k = 2$. Но тогда

$$\text{В} - \text{Р} + \Gamma = \text{В}_1 - \text{Р}_1 + \Gamma_1 = \text{В}_2 - \text{Р}_2 + \Gamma_2 = \cdots = \text{В}_k - \text{Р}_k + \Gamma_k = 2,$$

т.е. $\text{В} - \text{Р} + \Gamma = 2$,– формула Эйлера выполняется для исходного графа G. $\square$

Проверьте формулу Эйлера для плоских изображений всех планарных графов, которые встречаются в этом параграфе. Частным случаем доказанной формулы являются соотношения между числом вершин, ребер и граней правильных многогранников (см. §III.2; с них Эйлер и начал). Если каркас правильного многогранника представить себе сделанным из гибкой и растяжимой проволоки, то непрерывной деформацией этот пространственный граф можно уложить на плоскость. При этом укладка будет происходить внутри одной соответствующим образом растянутой грани, которую в подсчетах заменит бесконечная внешняя грань. На рис. 95 мы видим плоские укладки каркасов тетраэдра, куба и додекаэдра (проверяя для этих графов формулу Эйлера, заодно найдите в них гамильтоновы циклы).

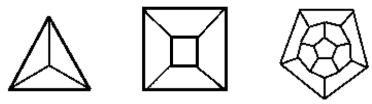


Рис. 95

Графы в прикладных исследованиях впервые появились в 1847 году, когда немецкий физик и механик Густав Роберт Кирхгоф (1824–1887) стал изучать свойства электрических цепей, представляя эти цепи в виде графов. В 1857 году известный нам алгебраист Кэли применил деревья к задаче классификации изомеров предельных углеводородов, занимавшей в то время химиков. В конце 1930-х годов Курт Левин (1890–1947) и другие американские психологи разработали систему представления межличностных отношений в виде графов (в духе §1).

В экономических и управленческих приложениях чаще всего встречаются так называемые сети. Сеть – это ориентированный граф, каждой дуге которого приписано некоторое положительное число. Это число называют длиной, или весом, или стоимостью дуги в зависимости от ситуации. Путем в сети называется последовательность дуг, в которой конец каждой дуги (кроме последней) совпадает с началом следующей. Под длиной пути понимается сумма длин составляющих его дуг. Одна из стандартных задач для сетей – отыскание кратчайших путей из какой-либо фиксированной вершины до других выделенных (или всех) вершин (см. §5). Сетевые модели для сложных производственных проектов начали разрабатываться в начале 1950-х годов и в настоящее время представляют собой мощный инструмент оптимального планирования и оперативного управления.

Математические методы исследования экономических объектов и процессов составляют самостоятельную дисциплину – математическую экономику. В ней широко применяются результаты как классических разделов математики (анализ, алгебра, геометрия), так и сравнительно новых ее ветвей (теория вероятностей, математическая статистика и др.). За основополагающие достижения в этой области академик Леонид Витальевич Канторович (1912–1986) был удостоен Нобелевской премии по экономике (1975 г.).

В исторических и филологических работах графы используют с целью выяснения картины взаимодействия тех или иных событий, для установления индивидуальных авторских особенностей в текстах, при объяснении структурных свойств произведений.

Точки и соединяющие их линии – эта простая и предельно общая идея, обозначенная словом «граф», на протяжении двух с половиной столетий давала яркие примеры конкретных воплощений как в области чистого интеллекта, так и в сфере материальной деятельности человека, и вдохновляющие возможности ее все еще кажутся неисчерпаемыми.

§ 3. Двоичная булева алгебра

Как уже упоминалось в §V.1, булевы алгебры появились во второй половине XIX века в работах логиков. Главным объектом изучения в то время были алгебры множеств, т.е. совокупности подмножеств некоторых универсумов, рассматриваемые вместе с операциями объединения, пересечения и дополнения.

Простейшим примером булевой алгебры, которая не связана непосредственно с множествами, является так называемая двоичная булева алгебра. Она тоже возникла из чисто логических потребностей, но в середине XX века стала одним из главных инструментов для описания разнообразных управляющих систем и основой компьютерной логики.

Двоичная булева алгебра – это двухэлементное множество $B = \{0, 1\}$, в котором имеются три операции: логическое сложение $x + y$, логическое умножение $x \cdot y$ и дополнение x', определяемые следующими таблицами:

$+$	0	1
0	0	1
1	1	1

$\cdot$	0	1
0	0	0
1	0	1

x	x'
0	1
1	0

Как видим, логическая сумма $x + y$ равна 1, когда равно 1 хотя бы одно из слагаемых; логическое произведение xy равно 1 только тогда, когда оба сомножителя равны 1.

Операции двоичной булевой алгебры тесно связаны с логическими операциями над высказываниями: если через $\lambda(p)$ обозначить логическое значение высказывания p, то легко усматривается, что $\lambda(p \vee q) = \lambda(p) + \lambda(q)$, $\lambda(p \wedge q) = \lambda(p)\lambda(q)$, $\lambda(\neg p) = [\lambda(p)]'$ («логическое значение дизъюнкции двух высказываний равно сумме их логических значений» и т.д.).

Для того чтобы утверждать, что система $(B, +, \cdot, ')$ является булевой алгеброй, нужно убедиться в истинности для нее всех тождеств, аналогичных тождествам алгебры множеств.

В двоичной булевой алгебре выкладки осуществляются обращением к таблицам операций. Например, проделывая в порядке, указанном скобками, сначала сложение, затем умножение, дополнение и снова умножение, получаем

$$[(1+0)\cdot 1]'\cdot 0 = [1\cdot 1]'\cdot 0 = 1'\cdot 0 = 0\cdot 0 = 0.$$

Проверка истинности некоторого тождества в B сводится к вычислению левой и правой его части при всевозможных конкретных значениях переменных (0 или 1) и сравнению полученных результатов. Удобнее всего это делать путем составления таблиц, аналогичных таблицам истинности для логических формул (см. §IV.1). Проверим, например, справедливость закона дистрибутивности $x(y+z)=xy+xz$,– в алгебре множеств он выглядел как $A\cap(B\cup C)=(A\cap B)\cup(A\cap C)$. Составим вычислительную таблицу:

x	y	z	$y+z$	$x(y+z)$	xy	xz	$xy+xz$
0	0	0	0	0	0	0	0
0	0	1	1	0	0	0	0
0	1	0	1	0	0	0	0
0	1	1	1	0	0	0	0
1	0	0	0	0	0	0	0
1	0	1	1	1	0	1	1
1	1	0	1	1	1	0	1
1	1	1	1	1	1	1	1

В первых трех столбцах перечислены все восемь вариантов возможных значений для x, y, z. Далее идут результаты выполняемых действий в соответствии с их порядком, указанным расстановкой скобок. Пятый столбец содержит значения левой части тождества, а восьмой – правой при тех же значениях переменных. Эти столбцы совпадают, значит, тождество доказано: логическое умножение распределительно относительно логического сложения.

Покажем еще выполнимость закона Де Моргана $(xy)'=x'+y'$:

x	y	xy	$(xy)'$	x'	y'	$x'+y'$
0	0	0	1	1	1	1
0	1	0	1	1	0	1
1	0	0	1	0	1	1
1	1	1	0	0	0	0

Таким же образом устанавливается истинность в B аналогов и других тождеств алгебры множеств.

Для каждого тождества алгебры множеств выпишите соответствующее тождество двоичной булевой алгебры. Для этого знаки теоретико-множественных операций $\cup, \cap$ и $^{-}$нужно заменить на символы операций в B – соответственно $+$, $\cdot$ и $'$, а вместо $\emptyset$ и Ω писать соответственно 0 и 1.

В конце 1930-х годов двоичная булева алгебра была применена к задачам анализа и синтеза релейно-контактных схем. Этот важный в истории технического обеспечения информатики шаг сделали Виктор Иванович Шестаков и, независимо, Клод Шеннон (создатель теории информации – см. §V.5).

Релейно-контактная схема (для краткости РКС) – это электрическая цепь, состоящая из проводников, переключателей (реле) и двухпозиционных контактов. Каждое реле (будем обозначать их буквами x, y, z и т.п.) может управлять несколькими контактами. Одни из них при срабатывании реле замыкаются и пропускают ток, а при отключении реле размыкаются и ток не пропускают. Такие контакты помечаются на схеме той же буквой, что и управляющее ими реле. Другие контакты ведут себя противоположным образом: при срабатывании соответствующего реле размыкаются, а при его отключении – замыкаются. Они называются инверторами и обозначаются символом своего реле, но со штрихом.

Пусть запись $x = 1$ будет означать, что реле x сработало (и, следовательно, все связанные с ним контакты x пропускают ток), а запись $x = 0$ – что реле x отключено (и, значит, все управляемые им контакты x разомкнуты и ток не пропускают).

С каждой РКС связана ее функция проводимости F. Она зависит от входящих в схему реле и принимает значение 1, когда схема проводит ток, и значение 0 в противном случае. Найдем функции проводимости простейших РКС.

1) Параллельное соединение контактов x и y (рис. 96а) пропускает ток, когда хотя бы один из этих контактов замкнут, т.е. функция проводимости $F(x, y)$ принимает значение 1, если

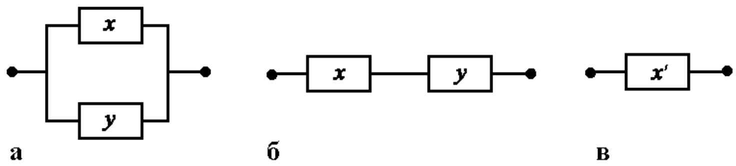


Рис. 96

хотя бы одна из переменных x, y имеет значение 1. Сравнивая эту ситуацию с таблицей логического сложения, получаем, что $F(x, y) = x + y$. Таким образом, логическое сложение реализуется как параллельное соединение двух контактов, управляемых каждый своим реле.

2) Последовательное соединение контактов x и y (рис. 96б) пропускает ток только когда оба контакта замкнуты. Следовательно, для этой РКС функция проводимости может быть представлена в виде $F(x, y) = xy$ – логическое умножение реализуется последовательным соединением двух независимых (т.е. управляемых разными реле) контактов.

3) Инвертор (рис. 96в) пропускает ток при отключенном реле и размыкается при включении реле. Ясно, что $F(x) = x'$,– операция дополнения в двоичной булевой алгебре моделируется схемой инвертора.

Каждая РКС может быть построена с помощью параллельного и последовательного соединения контактов и инверторов, управляемых реле. Следовательно, функция проводимости каждой РКС допускает запись в виде формулы двоичной булевой алгебры. Найдите, например, функцию проводимости для схемы, изображенной на рис. 97.

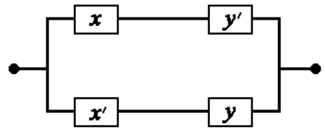


Рис. 97

На простом примере покажем, как методы двоичной булевой алгебры применяются в теории релейно-контактных схем.

Предположим, что нужно построить схему с тремя реле, пропускающую ток только в тех случаях, когда первое реле отключено, а из двух других хотя бы одно сработало. Обозначим реле буквами x, y, z и запишем условия работы схемы в терминах ее функции проводимости $F(x, y, z)$. Получаем: $F(0, 0, 1)$=$F(0, 1, 0)$=$F(0, 1, 1)$=1 (во всех остальных случаях схема ток не проводит). Условие $F(0, 0, 1)$=1 означает, что схема пропускает ток, когда реле x и y отключены, а z сработало. Но именно так действует последовательное соединение инверторов x', y' и контакта z. Так что требование $F(0, 0, 1)$=1 равносильно выполнению в B условия $x'y'z = 1$. Аналогично равенства $F(0, 1, 0)$=1 и $F(0, 1, 1)$=1 могут быть заменены соответственно на $x'yz'$=1 и $x'yz$=1. Таким образом, $F(x, y, z) = 1$ (т.е. схема пропускает ток) тогда и только тогда, когда в B выполняется хотя бы одно из равенств $x'y'z = 1$, или $x'yz' = 1$, или $x'yz = 1$, т.е. когда $x'y'z + x'yz' + x'yz = 1$. Отсюда получаем, что $F(x, y, z) =$

$= x'y'z + x'yz' + x'yz$. Схема с такой функцией проводимости реализует заданные условия работы, она изображена на рис. 98а.

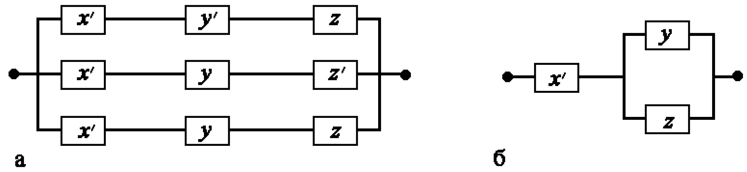


Рис. 98

Первый этап закончен: мы построили схему с нужными условиями работы. Второй этап состоит в поисках возможностей упростить найденную схему, сохраняя ее функциональные свойства. Для этого, используя тождества двоичной булевой алгебры, производят преобразования функции проводимости по типу школьного задания «упростить выражение». В нашем случае получаем:

$$F(x, y, z) = x'y'z + x'yz' + x'yz =$$

(используя тождество $x+x = x$, добавим еще одно слагаемое $x'yz$, после чего, применяя законы коммутативности, ассоциативности и дистрибутивности, проводим обычные алгебраические выкладки)

$$= x'y'z + x'yz' + x'yz + x'yz = (x'y'z + x'yz) + (x'yz' + x'yz) =$$
$$= x'z(y' + y) + x'y(z' + z) =$$

(далее ссылаемся на тождества $x + x' = 1$ и $x \cdot 1 = 1$)

$$= x'z \cdot 1 + x'y \cdot 1 = x'z + x'y = x'(y + z),$$

т.е. $F(x, y, z) = x'(y + z)$.

Эта функция проводимости реализуется РКС, представленной на рис. 98б. Новая схема равносильна по условиям функционирования первоначальной, но имеет всего три контакта. Она и дает окончательное решение задачи (проверьте).

Конечно, в рассмотренном примере тот же результат можно было получить и непосредственными рассуждениями, но в схемах, где задействованы сотни переключателей и тысячи контактов, простыми соображениями не обойтись. Приняв это к сведению, предложите релейно-контактную схему, которая позволила бы зажигать и гасить электрический фонарь над входом в двухквартирный коттедж из каждой квартиры своим, независимо действующим переключателем.

§ 4. Автоматы

С помощью графов обычно описывают взаимодействие частей сложной системы, ее внутреннее устройство, отвлекаясь от связей с внешней средой. Однако в вопросах, относящихся к восприятию, переработке и передаче информации, главный интерес представляет как раз реакция системы на поступающие извне сигналы. В этом смысле наиболее адекватной моделью дискретной системы является автомат. Это понятие оформилось в 50-е годы XX века, с одной стороны, под влиянием работ в области формальной теории нейронных сетей (конечная цель – выяснить законы функционирования человеческого мозга), а с другой – в связи с появлением электронных вычислительных машин, когда возникла необходимость разработки общих математических концепций, идеологии этого нового направления научно-технического прогресса.

Конструкторы вычислительного устройства работают со схемами, основываясь на теории графов и двоичной булевой алгебре, но те, кто будет использовать его, в первую очередь интересуются иным: что мы будем иметь в ответ на введенные в машину данные? Для пользователя компьютер представляется загадочным «черным ящиком» (художественное название системы с неизвестной внутренней структурой), с которым можно вести обмен информацией, подавая на его вход одни сведения и снимая на выходе другие. Именно такое восприятие отражается в следующем определении.

Автоматом называется пятерка $\mathcal{A} = (S, X, Y, \delta, \lambda)$, где S, X, Y – конечные непустые множества, $\delta : S \times X \to S$ – отображение декартова произведения $S \times X$ в множество S, а $\lambda : S \times X \to Y$ – отображение того же декартова произведения в множество Y. При этом элементы множества S называются состояниями автомата, элементы множеств X и Y соответственно входными и выходными сигналами, δ – функцией переходов, а λ – функцией выходов. Считается, что автомат работает в дискретно измеряемом времени по следующему правилу: если s – состояние автомата $\mathcal{A}$ в данный момент и на его вход подан сигнал x, то в следующий момент автомат $\mathcal{A}$ перейдет в состояние $t = \delta(s, x)$ и выдаст при этом на выходе сигнал $y = \lambda(s, x)$. Таким образом, очередное состояние автомата зависит от того, каким было его предыдущее состояние и какой сигнал поступил на вход. От этого же зависит и то, какой сигнал появится на выходе, т.е. какова будет реакция автомата на входное воздействие. Например, вы нажимаете клавишу пишущей машинки (многие еще помнят это устройство), каретка сдвигается на одну позицию влево (механизм перешел в новое состояние), а на бумаге появляется, скажем, буква А (сигнал на выходе). Другая реализация той же идеи: в прорезь автомата (вот откуда абстрактный термин), продающего напитки, опускается жетон, внутри устройства что-то щелкает и жужжит – совершается замысловатый переход в состояние готовности принять очередной заказ, на выходе – наполненный жидкостью стакан. Ну и, конечно, персональный компьютер: перед вами клавиатура и манипулятор «мышь», с помощью которых в машину вводится информация, базовый блок, вмещающий сложнейшую электронную схему искусственного мозга, и два варианта выходного устройства – монитор (дисплей) с экраном и принтер.

Автомат $\mathcal{A}$ обычно задают таблицей, состоящей из двух подтаблиц – переходов и выходов. Строки таблицы помечаются элементами множества состояний S, а столбцы – символами входных сигналов. На пересечении строки, помеченной состоянием s, и столбца, соответствующего входному сигналу x, в первой подтаблице проставляется состояние $\delta(s, x)$, а во второй – выходной сигнал $\lambda(s, x)$.

Пусть $S = \{s_1, s_2, s_3, s_4\}$, $X = \{x_1, x_2\}$, $Y = \{0, 1\}$ и функции δ и λ заданы табл. 18.

Из этой таблицы видно, например, что, находясь в состоянии s_2, под действием сигнала x_1 автомат перейдет в состояние s_3 и выдаст сигнал 1.

Удобной формой представления автоматов с небольшим числом состояний и входных сигналов являются так называемые автоматные диаграммы. Это ориентированные графы, дуги которых помечены символами входных и выходных сигналов. Вершины диаграммы соответствуют состояниям автомата, из вершины s в вершину t проводится дуга с меткой $x|y$, если входной сигнал x переводит состояние s в состояние t и при этом на выходе появляется сигнал y. Рис. 99 изображает диаграмму автомата, заданного табл. 18.

Таблица 18

	x_1	x_2	x_1	x_2
s_1	s_2	s_1	0	1
s_2	s_3	s_2	1	0
s_3	s_3	s_2	1	0
s_4	s_4	s_3	0	1

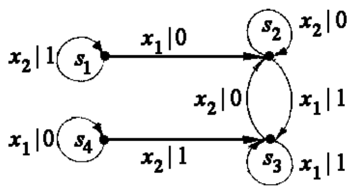


Рис. 99

Заметим, что в отличие от графов, описывающих функциональные схемы систем, автоматная диаграмма реально работающего устройства, в общем, не связана с его технической конструкцией, а лишь символизирует наше представление о том, как оно осуществляет переработку информации, преобразуя входные сигналы в сигналы на выходе. Пусть, например, торговый автомат рассчитан на продажу трех порций пива (в математических мирах допускаются сильные отклонения от нормы). Мы можем трактовать его как абстрактный автомат, имеющий четыре состояния: s_0, s_1, s_2, s_3, где s_i означает «содержать i порций пива» (возможные значения для i: $0, 1, 2, 3$), один входной сигнал x (опускание жетона в прорезь) и два выходных сигнала: 1 – выдача порции пива, 0 – появление световой надписи «пива нет». На рис. 100 представлена диаграмма этого автомата, которую мы и вручим конструктору для воплощения в металле.

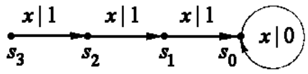


Рис. 100

Работая с компьютером, вы, конечно, не будете после каждого введенного в него сигнала смотреть, что получается на выходе: многие промежуточные сообщения, мелькающие на экране монитора, отражают изменения в состоянии автомата и не содержат интересующей нас финальной информации, к тому же не все этапы работы машины фиксируются в виде реакций на выходе, доступных для наблюдения. Рассмотрим пример кажущегося несоответствия между тем, что делается на входе и на выходе автомата. Пусть в ответ на введенную в него двоичную (т.е. состоящую из символов 0 и 1) последовательность автомат должен сообщить, имеется ли в ней четное или нечетное число единиц. Такую реакцию может осуществить устройство с двумя состояниями s_0 и s_1, тремя входными сигналами $0, 1$ и e, тремя выходными сигналами e, *ч* и *н*, работа которого описывается диаграммой рис. 101.

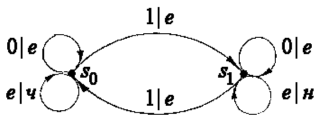


Рис. 101

Символ e является сигналом об окончании ввода последовательности, составленной из 0 и 1; это «пустой» символ – его появление на выходе наблюдателем не фиксируется. В начальный момент автомат находится в состоянии s_0. Пусть на входной ленте появилась последовательность $011100101e$. Считывая ее, автомат пройдет цепочку состояний $s_0s_1s_0s_1s_1s_1s_0s_0s_1$ и напечатает на выходной ленте последовательность *eeeeeeeeeн*. Первые 9 букв невидимы, так что в качестве ответа на введенную информацию мы прочтем сообщение, содержащее только один символ $н$, обозначающий нечетность количества единиц во входном слове. (Измените конструкцию автомата так, чтобы на выходе полностью повторялась входная последовательность вместе с заключением о четности или нечетности количества нулей в ней.)

Автомат по продаже пива (см. рис. 100) «помнит», сколько порций напитка содержится в нем в данный момент, автомат по проверке четности (см. рис. 101) запоминает, четное или нечетное число единиц прошло через его входной канал. Информация, хранимая в автомате, зашифрована в его состояниях. Чем разнообразнее информация, тем больше состояний должен иметь автомат, обрабатывающий ее. Поскольку число состояний всегда предполагается конечным, то не для всякой задачи распознавания может быть построен решающий ее автомат. Например, невозможно придумать автомат, который для любой предъявленной ему двоичной последовательности вида 1^m0^n (сначала m единиц, а затем n нулей) отвечал бы на вопрос, будет ли в ней единиц столько же, сколько нулей. В самом деле, прежде чем приступить к сравнению этих количеств, наше устройство должно зафиксировать, сколько единиц предшествует первому нулю. Так как число m в принципе может быть сколь угодно большим, конечным числом состояний здесь не обойтись. Другое дело, если возможное количество единиц заранее ограничить сверху («не больше чем...»),– предложите конструкцию автомата, который решал бы обсуждаемую задачу для последовательностей указанного вида, содержащих не более 3 единиц (постарайтесь минимизировать число состояний).

Используя математические понятия, введенные в этом и двух предыдущих параграфах, разработку схемы электронного преобразователя информации условно можно разбить на следующие этапы: 1) построение абстрактного автомата, реализующего все вход–выходные реакции, предусмотренные для проектируемого устройства; 2) создание графовой модели функциональных связей между реальными физическими элементами будущей системы; 3) подбор с помощью двоичной булевой алгебры самих этих элементов, осуществляющих простейшие логические шаги в процессе переработки и передачи информации.

Практическое воплощение многих идей и конструкций прикладной математики было бы невозможно без вычислительных автоматов.

Первый механизм – для сложения натуральных чисел – сконструировал в 1641 году Паскаль, желавший облегчить вычислительные занятия отца, ведавшего сбором налогов в Руане.

Несмотря на сложность обращения с машиной и неторопливость вращения ее шестеренок, преимущество перед имевшимися в то время счетными приспособлениями было несомненным. Поскольку операции с многозначными числами во времена Паскаля все еще относились к высшим сферам интеллектуальной деятельности, и сам юный изобретатель, и те, кто видел в работе его детище, не сомневались в том, что сделан важный шаг в механизации умственного труда.

Прошло всего три десятилетия, и в 1673 году Лейбниц продемонстрировал свою счетную машину, производившую уже все четыре арифметических действия. Уровень тогдашней техники не позволял воплотить идею в виде массового продукта: изготовление каждого экземпляра требовало величайшего искусства механиков. Лишь к концу XIX века дело было поставлено на поток, и арифмометры петербургского инженера Вильгольдта Теофиловича Однера (1845–1905) распространились по всему миру.

Новый принцип арифметической машины предложил в 1878 году академик Петербургской академии наук Пафнутий Львович Чебышёв (1821–1894). Необычайно многосторонней была деятельность выдающегося ученого-математика и механика. Теория чисел, математический анализ, теория вероятностей, математическая теория механизмов – во всех этих научных направлениях П.Л.Чебышёв получил фундаментальные результаты. В своем творчестве он постоянно стремился соединить теорию с реальной практикой: «В старину задавали математические задачи боги... Далее наступил второй период, когда задачи задавали полубоги: Ньютон, Эйлер, Лагранж. Теперь третий период, когда задачи задает практика...» И все же истинным прообразом современных компьютеров послужил не арифмометр, каким бы могущественным вычислителем он ни представал в последних по времени моделях. Эта роль выпала на долю «аналитической машины», проект которой в первой половине XIX века разработал профессор математики Кембриджского университета Чарлз Бэббидж (1792–1871). По его замыслу, механизм должен был состоять из трех частей, которые в современной терминологии можно было бы назвать «памятью», «арифметическим устройством» и «устройством управления». Ввод информации в машину предполагалось осуществлять с помощью перфокарт – картонных карточек с пробитыми в них отверстиями. На них же записывалась и «программа» – последовательность арифметических операций, которые нужно было проделать для достижения конечного результата. Получив задание, вычислительная машина должна была далее действовать без вмешательства человека. Идея Бэббиджа приобрела широкую известность в гуманитарных кругах, но несмотря на многие восторженные отзывы (например, Эдгара По), не заинтересовала ни его коллег, ни техников. Она была реализована лишь в 1944 году (т.е. через сто лет), когда профессор Гарвардского университета Говард Хетауэй Эйкен (1900–1973), используя релейно-контактные схемы, построил первую автоматическую вычислительную машину «Марк-I» (в модели 1950 года «Марк-III» уже появились электронные схемы).

В роли главного идеолога современной вычислительной техники выступил выдающийся математик XX века американец венгерского происхождения Джон фон Нейман (1903–1957). Его ставшие классическими результаты охватывают столь далекие по своему содержанию области, как теория множеств, функциональный анализ, теория групп, квантовая механика, математическая экономика, теория автоматов. Активный участник американской атомной программы, он в 1945 году вошел также в состав Бюро по проектированию счетной техники и вскоре распространил доклад, в котором изложил свое понимание основных принципов устройства и функционирования компьютеров. Эти принципы, с теми или иными естественными видоизменениями, продолжают оставаться «путеводной звездой» для конструкторов вычислительных машин.

В 1948 году американский математик Норберт Винер (1894–1964), воскресив восходящий к Платону (427–347 до н.э.) термин «кибернетика», предложил назвать этим словом формировавшуюся в то время новую науку об управлении сложными системами – как в технических, так и в биологических и социальных сферах. Идеи кибернетики и возникшей одновременно с ней теории информации (см. §V.5) получили вскоре широкое распространение, выйдя далеко за пределы их первоначально намечавшейся математической проблематики. Весьма свободное толкование понятий и результатов этих наук зачастую приводило к парадоксальным и поверхностно-сенсационным выводам в гуманитарных областях, традиционно далеких от формальной математики. Слова «кибернетика» и «информация» стали приобретать неясное содержание, а связанные с ними математические предложения оказались как бы одной из струй в широком потоке разнообразных суждений, далеко не все из которых носили научный характер. Поиски нового названия: «теоретическая кибернетика», «математическая кибернетика», «системотехника», «системный анализ» – отражают стремление математиков обособить свои исследования сложных систем и процессов управления ими в отдельное русло. В той части, где в качестве сложной системы выступает ЭВМ, установилось в конце концов английское (скорее, американское) computer science и его французский эквивалент informatique – информатика. Заметим, что в отличие от кибернетики шенноновская теория информации (математическая теория передачи сообщений), имевшая более конкретные исходные понятия и задачи, успешно пережила аналогичные попытки придать ей некоторое универсальное, наднаучное толкование.

Одним из основоположников математической кибернетики (а значит, и информатики) в СССР был академик Виктор Михайлович Глушков (1923–1982). Известный алгебраист, успешно работавший в теории групп, в конце 1950-х годов он решительно изменил направление своего творчества, обратившись к теории автоматов и, в частности, к проектированию электронных вычислительных машин. С 1961 года В.М.Глушков возглавлял Институт кибернетики в Киеве, ставший одним из ведущих научных центров в области вычислительной техники.

Достижения в промышленной электронике (изобретение транзисторов – 1948 год, интегральных схем – 1959 год, наконец микропроцессоров – 1970 год) позволили уменьшить физические размеры ЭВМ, первоначально занимавших огромные площади, до малогабаритных устройств настольного типа. В 1981 году появились первые персональные компьютеры.

Благодаря успехам в деле программного обеспечения ЭВМ непрограммирующие пользователи (так именуются лица, которым при работе с компьютером не нужно составлять собственные программы) получили непосредственный доступ к новым информационным технологиям, что существенно сблизило естественнонаучное и гуманитарное знание.

ЭВМ перестали быть только вычислителями в собственном понимании этого слова и проникли буквально во все сферы человеческой деятельности. Этому победному шествию способствовало и увеличение быстродействия процессоров, очень условно измеряемого количеством элементарных операций, выполняемых за одну секунду. В новейших компьютерах преодолен фантастический порог – миллиард операций в секунду. Как говорил с гордостью Бэббидж, «за одну минуту моя машина сможет осуществлять, например, 60 актов сложения».

§ 5. Алгоритмы и машина Тьюринга

Одним из основных понятий дискретной математики (т.е. математики, изучающей дискретные системы) и информатики является понятие алгоритма. Широкая компьютеризация многих видов человеческой деятельности способствовала его проникновению в естественные и гуманитарные науки, в повседневный язык средств массовой информации. В самом общем истолковании алгоритм – это точное общепонятное описание действий, последовательное выполнение которых ведет к достижению цели. («Так в чем же он, по-вашему,– алгоритм успеха?»– спрашивает тележурналист у директора процветающего предприятия. «Ну, во-первых»,– начинает директор...)

Математик, в основном согласившись с приведенным выше толкованием, добавит, что речь идет не о способе решения какой-то одной конкретной задачи, а о едином методе, позволяющем решить любую из бесконечной серии однотипных задач. Инструкции по сборке игрушечных моделей или предметов мебели; церковные, дипломатические и военные церемониалы; рецепты приготовления необычных кулинарных изделий; нотные тексты музыкальных произведений – все эти и подобные им предписания подразумевают многократное их использование в соответствующих типовых ситуациях.

Подкрепив этими примерами интуитивное понимание термина «алгоритм», можно было бы ожидать его точной формулировки. Но ее не будет. Алгоритм – это первоначальное понятие, оно не выражается через какие-либо более простые идеи и в этом смысле стоит в одном ряду с такими ключевыми словами математики, как множество, точка, прямая, плоскость, событие, истина, ложь, информация и т.д.

Школьные алгоритмы сложения, вычитания, умножения чисел «столбиком» и деления «углом» были выдающимися достижениями позиционной системы счисления, превратившими искусство счета в формальное выполнение некоторых простых операций, не требовавших никакой изобретательности. Поскольку десятичная система пришла в Европу через переводы сочинений восточных математиков и, в первую очередь, аль-Хорезми, от его имени и произвели слово «алгоритм», которым называли учебники арифметики, излагавшие новый подход к обращению с числами. Современное понимание термина сложилось в первой трети XX века.

Каким же требованиям должна удовлетворять система предписаний для решения некоторого класса задач, чтобы ее можно было признать алгоритмом? Обычно в качестве характерных признаков алгоритма принимаются следующие пять.

1. *Массовость*. Исходные данные для применения алгоритма представляют собой некоторую конечную систему величин, выбираемых из потенциально бесконечного множества.

2. *Дискретность*. Алгоритм состоит из конечного числа предписаний (шагов), выполняемых в дискретном времени.

3. *Детерминированность*. Описание действий, предусмотренных на каждом шаге алгоритма, не должно допускать разночтений.

4. *Элементарность*. Действия, предписываемые алгоритмом, должны быть простыми, требующими от исполнителя чисто механической работы.

5. *Результативность*. Алгоритм должен останавливать работу по истечении некоторого конечного времени (зависящего от исходных данных), указывая при этом, что считается результатом.

Рассмотрим эти признаки для конкретного алгоритма построения кратчайших путей в сетях специального вида (см. §2).

Предположим, что некоторый процесс в своем развитии проходит несколько стадий. Из каждого возможного на данной стадии состояния допускается переход в несколько состояний следующей стадии, причем каждому из этих переходов соответствует некоторое характеризующее его число (время, стоимость затрат и т.п.). Требуется найти оптимальное по данному показателю развитие процесса от начального состояния до конечной стадии. Покажем на примере один из алгоритмов поиска кратчайших путей из начальной вершины в конечную в сетях, моделирующих многостадийные процессы. Пусть в сети, изображенной на рис. 102, начальная вершина помечена символом s, конечная – f, верхний индекс в названии промежуточных вершин обозначает номер стадии процесса, а нижний – порядковый номер состояния на этой стадии. На дугах выставлены числа, обозначающие их длины.

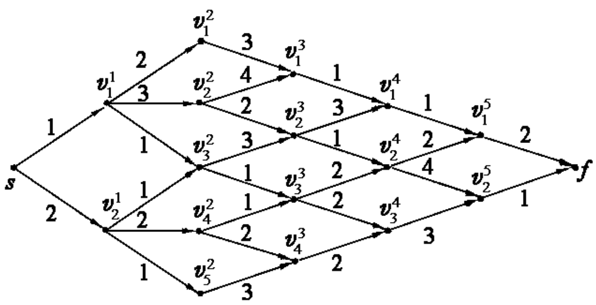


Рис. 102

АЛГОРИТМ

Вершины просматриваются последовательно от стадии к стадии в порядке возрастания номеров.

Шаг 1. Присвоить начальной вершине метку $\lambda(s) = 0$.

Шаг 2. Для каждой дуги (u, v), ведущей в очередную в порядке просмотра вершину v, найти число $\lambda(u) + w(u, v)$, где $\lambda(u)$ – метка начала дуги, а $w(u, v)$ – длина дуги.

Шаг 3. Из полученных чисел выбрать наименьшее и сделать его меткой $\lambda(v)$ вершины v.

Шаг 4. Стереть все дуги (u, v), для которых $\lambda(u) + w(u, v) > \lambda(v)$.

Шаг 5. Если $v = f$, т.е. v – конечная вершина, работа алгоритма завершена: кратчайшие пути из начальной вершины s в конечную вершину f прослеживаются по оставшимся дугам и каждый из них имеет длину, равную $\lambda(f)$ – метке вершины f. В противном случае перейти в 2.

Действуя согласно предписаниям алгоритма, последовательно находим $\lambda(s) = 0$ (шаг 1), а затем $\lambda(v_1^1) = 1$, $\lambda(v_2^1) = 2$, $\lambda(v_1^2) = 3$, $\lambda(v_2^2) = 4$ – во всех случаях после выполнения шага 2 выбирать, как это предусмотрено в шаге 3, не приходится и стирание дуг (шаг 4) не происходит, так как в каждую из указанных вершин входит всего одна дуга. Дойдя до вершины v_3^2, мы увидим, что в нее ведут две дуги: из вершины v_1^1 и из вершины v_2^1 – и обе имеют длину 1. Следовательно, для вершины v_3^2 получаем две суммы: $\lambda(v_1^1) + w(v_1^1, v_3^2)$=1 + 1=2, $\lambda(v_2^1) + w(v_2^1, v_3^2) = 2 + 1 = 3$. Первая меньше, значит, полагаем $\lambda(v_3^2) = 2$ (шаг 3) и (шаг 4) стираем дугу (v_2^1, v_3^2). Если бы v_3^2 была конечной вершиной, мы бы остановились и увидели, что кратчайший путь из s в v_3^2 идет по дугам (s, v_1^1) и (v_1^1, v_3^2), он имеет длину 2. Но $v_3^2 \neq f$, и мы снова переходим к выполнению шага 2. Продолжите действия, описанные в алгоритме, до конца его работы и сравните полученный результат со схемой кратчайших путей в данной сети на рис. 103.

Рассмотренный алгоритм применим к любой сети указанного типа, т.е. к бесконечному множеству однотипных задач; другими словами, он решает массовую проблему. Дискретность алгоритма проявляется в том, что он состоит из пяти предписаний (шагов), разбивающих процесс решения задачи на некоторое число последовательно осуществляемых этапов. Действия, которые нужно совершить на каждом этапе, вполне детерминированы (т.е. истолковываются однозначно) и элементарны (найти сумму

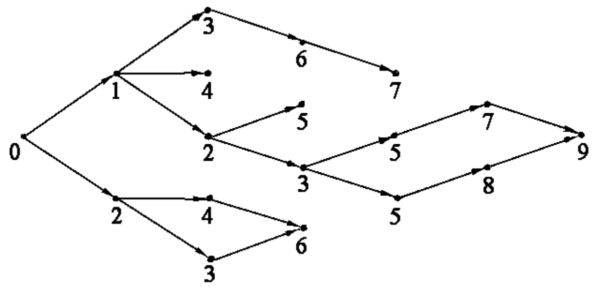


Рис. 103

двух чисел, или выбрать из нескольких чисел наименьшее, или стереть дугу, или проверить, не является ли очередная вершина конечной). После окончания работы, т.е. при выполнении условия $v = f$, из текста алгоритма уясняем результат наших действий: как выглядят всевозможные кратчайшие пути из начальной вершины в конечную и чему равна их длина.

Для наглядности алгоритмы изображают в виде ориентированных графов, которые называются блок-схемами. Вершины блок-схемы соответствуют шагам алгоритма, а ее дуги указывают порядок переходов от шага к шагу в процессе работы. Каждый шаг алгоритма может быть отнесен к одному из двух типов: выполнение определенных действий или проверка некоторых условий. Шаги первого типа изображают на блок-схеме в виде прямоугольников, шаги второго типа – в виде ромбов. Еще нужно предусмотреть этап предъявления исходных данных и завершающий этап. Блок-схема алгоритма поиска кратчайших путей в сети, соответствующей многостадийному процессу, изображена на рис. 104.

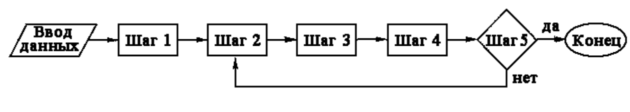


Рис. 104

Один из подходов к уточнению математического понятия алгоритма состоит в том, чтобы считать алгоритмами такие вычислительные предписания, которые могла бы выполнить некоторая подходящим образом устроенная «машина». Эту идею высказал в 1936 году английский математик и инженер Алан Матисен Тьюринг (1912–1954) и одновременно с ним – уже упоминавшийся нами американский логик Пост. Если машина Бэббиджа предвосхищала конструктивные принципы ЭВМ, то умозрительные машины Тьюринга и Поста стали теоретической основой программного обеспечения вычислительной техники.

Машина Тьюринга – это автомат, подсоединенный своими входом и выходом к бесконечной ленте и снабженный механизмом, меняющим положение ленты. Лента разбита на ячейки, в каждую из которых может быть записан один символ из некоторого конечного алфавита. В начальный момент автомат принимает на вход символ из стартовой ячейки, переходит в новое состояние, заносит в эту ячейку появившийся на его выходе символ, после чего сдвигает ленту на одну ячейку влево или вправо или оставляет ее на месте. В следующем такте все повторяется: считывается содержимое поданной к входу ячейки, меняется внутреннее состояние автомата, выданный им сигнал фиксируется в обозреваемой ячейке, срабатывает лентопротяжный механизм.

В качестве математического объекта машина Тьюринга определяется как пятерка $\mathcal{M} = (S, X, \delta, \lambda, \zeta)$, где S, X – конечные непустые множества (соответственно множество внутренних состояний и алфавит, одновременно входной и выходной), $\delta : S \times X \to S$ – отображение декартова произведения $S \times X$ в множество S (функция переходов), $\lambda : S \times X \to X$ – отображение того же декартова произведения в множество X (функция выходов), $\zeta : S \times X \to \{\text{л}, \text{п}, \text{останов}\}$ – третье отображение множества $S \times X$, на этот раз в трехэлементное множество $\{\text{л}, \text{п}, \text{останов}\}$ (функция управления механизмом ленты). Таким образом, четверка $\mathcal{A} = (S, X, \delta, \lambda)$ представляет собой автомат с совпадающими входным и выходным алфавитами.

Работу машины Тьюринга в дискретном времени можно описать последовательностью пятерок (s, x, t, y, c), где s – внутреннее состояние машины в данный момент, x – считываемый ею символ на ленте, $t = \delta(s, x)$ – состояние, в котором окажется машина, $y = \lambda(s, x)$ – выходной символ, который она запишет в ячейку ленты вместо x, а $c = \zeta(s, x)$ – команда для перемещения ленты. После сдвига ленты наступает следующий рабочий такт.

В §4 была предложена конструкция автомата для проверки четности числа единиц в двоичной входной последовательности. Построим машину Тьюринга, решающую эту задачу. Она будет иметь два внутренних состояния s_0, s_1 и пять алфавитных символов 0, 1, e, *ч*, *н*. В начальный момент машина находится в состоянии s_0 и обозревает ячейку, содержащую первый символ последовательности. Если это 0, состояние машины не изменяется, содержимое ячейки тоже, лента сдвигается влево, машина обозревает ячейку со следующим символом последовательности... Символ e обозначает пустые ячейки ленты. После окончания работы в пустой ячейке, примыкающей к последней из ячеек, содержащих символы исследуемой двоичной последовательности, будет записано *ч*, если количество единиц четное, или *н*, если оно нечетно. Искомая машина Тьюринга работает по программе представленной в табл. 19.

Запишите на ленте последовательность 001101 и проследите по тактам работу машины. Что будет записано в непустых ячейках ленты после остановки (технический термин – «останов»)?

Таблица 19

s_0	0	s_0	0	*л*
s_0	1	s_1	1	*л*
s_0	*e*	s_0	*ч*	останов
s_1	0	s_1	0	*л*
s_1	1	s_0	1	*л*
s_1	*e*	s_1	*н*	останов

Внимательный анализ показывает, что построенная машина Тьюринга, в сущности, функционирует так же, как упомянутый автомат по проверке четности из §4 (вернее, как его модификация, которая дублирует на выходе входную двоичную последовательность, завершая ее сообщением о четности или нечетности числа единиц). Это лишь одно из проявлений следующего общего факта.

Теорема. Для всякого автомата $\mathcal{A}$ существует машина Тьюринга, перерабатывающая входные последовательности автомата $\mathcal{A}$ в его выходные последовательности в точности так, как это делает $\mathcal{A}$.

ДОКАЗАТЕЛЬСТВО. Пусть $\mathcal{A} = (\mathcal{S}, \mathcal{X}, \mathcal{Y}, \delta, \lambda)$ – произвольный автомат. Присоединим его вход и выход к тьюринговой ленте, механизм которой может либо сдвигать ее на одну ячейку влево, либо останавливать. Алфавит полученной машины $\mathcal{M}$ будет состоять из входных и выходных сигналов автомата $\mathcal{A}$ и специального указателя пустой ячейки, например, e. Встретив в просматриваемой ячейке символ $x \in X$ (т.е. входной сигнал автомата $\mathcal{A}$), машина $\mathcal{M}$, находящаяся в некотором состоянии $s \in S$, перейдет в новое состояние $\delta(s, x)$, заменит в указанной ячейке x на $y = \lambda(s, x) \in Y$ и сдвинет ленту на один шаг влево. Чтобы предусмотреть все возможности, предпишем еще машине $\mathcal{M}$ делать останов при появлении в обозреваемой ячейке любого символа, не принадлежащего к числу входных сигналов автомата $\mathcal{A}$, не меняя при этом содержимое ячейки. Построенная машина Тьюринга $\mathcal{M}$ полностью имитирует поведение автомата $\mathcal{A}$ в смысле переработки входной информации. В самом деле, предположим, что автомат $\mathcal{A}$, находясь в начальном состоянии s_0 и получая на вход последовательность $x_1x_2\dots x_n$, печатает на выходе последовательность $y_1y_2\dots y_n$. Запишем $x_1x_2\dots x_n$ в n последовательных ячейках на ленте машины Тьюринга $\mathcal{M}$, остальные ячейки предполагая пустыми. Приведем машину $\mathcal{M}$ в состояние s_0 и подадим к ее входу ячейку, содержащую символ x_1. Начав работу, машина просмотрит n ячеек, запишет в них последовательно символы $y_1, y_2, \dots, y_n$, наткнется на пустую ячейку (символ e) и остановится. На ленте мы увидим (в окружении пустых ячеек) требуемую последовательность $y_1y_2\dots y_n$. $\square$

Доказанная теорема означает, что если некоторую задачу может решить автомат, то ее может решить и подходящая машина Тьюринга. Но, может быть, верно и обратное? Следующий пример показывает, что это не так: существуют задачи, решаемые машинами Тьюринга, но недоступные для автоматов. Как отмечалось в §4, невозможно построить автомат, который сравнивал бы число единиц и нулей в двоичных последовательностях вида 1^m0^n. В силу конечности числа состояний, автомат не может запомнить, сколько единиц прошло до первого нуля, если число единиц в рассматриваемой последовательности не ограничено заранее каким-нибудь конкретным числом. Но лента машины Тьюринга бесконечна, и подобных проблем с запоминанием здесь не возникает. Приведем программу (табл. 20), решающую обсуждаемую задачу на машине с 6 состояниями $s_0, s_1, s_2, s_3, s_4, s_5$ и алфавитом, состоящим из 5 символов $0, 1, e$ (пустой символ), $\neq$ (нулей и единиц разное количество),$=$ (нулей и единиц поровну). Начальное состояние – s_0, первая обозреваемая ячейка содержит начальную единицу последовательности.

Таблица 20

s_0	1	s_1	e	*л*
s_1	1	s_1	1	*л*
s_1	0	s_2	0	*л*
s_1	e	s_1	$\neq$	останов
s_2	0	s_2	0	*л*
s_2	e	s_3	e	*п*
s_3	0	s_4	e	*п*
s_4	0	s_4	0	*п*
s_4	e	s_4	$=$	останов
s_4	1	s_5	1	*п*
s_5	1	s_5	1	*п*
s_5	e	s_0	e	*л*

Проследите по тактам работу машины для последовательностей 1110, 1100, 1000.

Напишите для подходящей машины Тьюринга программу, после выполнения которой в предъявленной двоичной последовательности «стирались» бы первый и последний символы. Постройте также автомат, решающий эту задачу.

Математическая практика показала, что для всех известных вычислительных (в широком смысле) алгоритмов удается построить реализующие их машины Тьюринга. Уверенность в том, что это наблюдение отражает объективную закономерность, формулируется в виде так называемого тезиса Тьюринга: «Всякий алгоритм может быть реализован подходящей машиной Тьюринга». Доказать это утверждение невозможно из-за отсутствия точного математического определения понятия «алгоритм». Принимая тезис Тьюринга как аксиому, мы, по существу, утверждаем: алгоритмы – это такие системы предписаний, которые могут быть осуществлены машинами Тьюринга.

Как уже говорилось, алгоритмы создаются для решения массовых проблем, т.е. бесконечных классов однотипных задач. Но всегда ли искомый алгоритм существует? С давних времен в математике были известны массовые проблемы, которые никак не удавалось решить с помощью единого метода. Собственно из рассмотрения подобных проблем и возникли понятие алгоритма и связанные с ним теории. Массовая проблема называется алгоритмически неразрешимой, если не существует приводящего к ее решению алгоритма. В 1936 году Чёрч установил алгоритмическую неразрешимость проблемы распознавания тождественной истинности формул исчисления предикатов (см.§IV.2). Большой энтузиазм вызвали в 1950-х годах результаты академика Петра Сергеевича Новикова (1901–1975), доказавшего алгоритмическую неразрешимость некоторых проблем теории групп. В частности, было показано, что не существует алгоритма, который для любых двух предъявленных групп распознавал бы, изоморфны они или нет.

Одним из самых известных достижений в этой области явилось решение 10-й проблемы Гильберта (о проблемах Гильберта см. §I.5). В ней предлагалось указать общий метод, который бы позволил в конечное число шагов узнать для данного уравнения с несколькими неизвестными и целыми коэффициентами, имеет ли оно решение в целых числах. Решением уравнений в целых числах математики занимались с древнейших времен. Достижения античных ученых в этом направлении были подытожены Диофантом, по имени которого эти уравнения и называются. К числу диофантовых уравнений относятся, в частности, уравнения вида $x^n + y^n = z^n$, $n > 1$, с которыми связана Великая теорема Ферма (см. §I.2). Сам Гильберт был уверен, что искомый метод существует, и до конца 1930-х годов продолжались безуспешные попытки найти его. Однако с появлением теории алгоритмов и понятия алгоритмической неразрешимости возникли сомнения в самом существовании искомого алгоритма. Исследования группы американских математиков, предпринятые в 1950-60-х годах, значительно усилили эти подозрения. Завершающий результат в истории 10-й проблемы Гильберта получил в 1970 году ленинградский математик Юрий Владимирович Матиясевич: алгоритм для распознавания наличия решений у произвольного диофантова уравнения не существует.

§ 6. Формальные языки и грамматики

Машина Тьюринга, так же как и автомат, перерабатывает знаковые последовательности, поступившие на вход, в знаковые последовательности, которые появляются на выходе. Принимая тезис о том, что каждый алгоритм может быть реализован подходящей машиной Тьюринга, мы тем самым предполагаем, что любой алгоритмический процесс можно представить в виде элементарных шагов, записываемых строчками символов. Считая символы эти буквами некоторого алфавита, а строчки – словами, в итоге получаем, что алгоритм, по существу, является набором правил, по которым в некотором искусственном языке из одних слов получаются другие. Изучение искусственных языков имеет важное значение, поскольку оно охватывает, в частности, и языки программирования, на которых пишутся программы, реализующие алгоритмы на ЭВМ.

В 1842 году графиня Ада Аугуста Лавлейс (1815–1852), дочь Байрона и ученица Де Моргана, в статье, посвященной машине Бэббиджа, писала: «Аналитическая машина не претендует на то, чтобы создавать что-то действительно новое. Машина сможет выполнить все то, что мы сумеем ей предписать. Она может следовать только программе». Принято считать, что это было первым замечанием, относящимся к программированию.

Электронный мозг компьютера состоит из огромного числа элементов, которые могут находиться только в двух устойчивых состояниях, определяемых тем, достигает или нет электрический сигнал некоторого порогового значения. Состояние элемента, соответствующее достаточному уровню тока, обозначается символом 1, а то, в котором сигнал не воспринимается, символом 0. Этими сугубо техническими причинами объясняется та роль, которую играют в информатике двоичная система счисления и двоичная булева алгебра. Входной алфавит ЭВМ содержит всего два символа – 0 и 1, и вся информация, поступающая в машину, должна быть представлена в виде двоичных последовательностей, числовых или логических. Легко вообразить себе, в какой колоссальный труд (и с каким количеством возможных ошибок) вылилась бы попытка выразить списками двоичных слов то, что мы хотим предписать машине. К счастью, это сложнейшее препятствие на пути общения человека с компьютером удалось преодолеть с помощью компромисса: человек пишет свои указания (команды) на некотором искусственном, но вполне доступном ему языке, а машина по специальной программе, приспособленной к этому языку, переводит команды в двоичные последовательности, понятные ей. Искусственные языки, на которых записываются команды, управляющие работой ЭВМ, называются языками программирования, а программы, осуществляющие перевод на машинный язык,– трансляторами.

В настоящее время известно несколько тысяч языков программирования (среди них есть и АДА – универсальный язык высокого уровня, появившийся в США в 1979 году и названный в честь леди Лавлейс). Одни из этих языков удобны для проведения сложных технических расчетов, другие предназначены для работы с нечисловыми данными, третьи ориентированы на решение каких-либо специфических классов задач.

Вопросами создания и использования искусственных языков для ЭВМ, а также разработкой методов и средств составления, проверки и улучшения компьютерных программ занимается теоретическое программирование – ветвь информатики, тесно связанная с теорией алгоритмов и теорией формальных языков и грамматик. Среди многих ученых, внесших решающий вклад в теорию и практику программирования, особое место занимают Эдсгер Дейкстра (Голландия) и Дональд Эрвин Кнут (США). Признанным лидером отечественного программирования был академик Андрей Петрович Ершов (1931–1988), работавший в Новосибирском научном центре. (Ему, выдающемуся специалисту в той области, где главная цель – расчленить человеческую мысль на последовательность компьютерных команд, принадлежат следующие строки:

На пригорке церковь без креста.
И, тревогой старою объятый,
Я молюсь, чтоб эта красота
Навсегда осталась неразъятой.)

Пусть задано некоторое конечное непустое множество X, которое будем называть алфавитом, а его элементы – буквами. Конечная последовательность $x_1x_2\ldots x_n$, составленная из букв алфавита X, называется словом. Длина слова – это количество букв в нем. Например, если $X = \{0,1\}$, то слово 01001110 имеет длину 8. В число слов включают и пустое слово ε – не содержащее ни одной буквы, его длина равна нулю. Совокупность всех слов в алфавите X обозначим через X^*. Множество X^* является счетным, его элементы, т.е. слова, можно расположить в виде бесконечной последовательности в порядке возрастания длины слов. В частности, для двоичного алфавита получаем

$$X^* = \{\varepsilon, 0, 1, 00, 01, 10, 11, 000, 001, \dots\}$$

(напишите следующие 7 слов этого бесконечного списка).

Конкатенацией (от латинского catena – цепь) слов $u = x_1x_2 \dots x_m$ и $v = y_1y_2 \dots y_n$ называется слово $w = uv = x_1x_2 \dots x_my_1y_2 \dots y_n$ длины $m + n$, в котором первые m букв образуют слово u, а следующие n букв – слово v. Говорят, что в слове $w = uv$ слово u является префиксом, а слово v – суффиксом. Например, в двоичном слове 1001 префиксами будут 5 слов: $\varepsilon, 1, 10, 100, 1001$. Выпишите все суффиксы слова 1001. Если $w = utv$, слово t называют подсловом слова w, причем префикс u и суффикс v могут быть пустыми словами. Укажите все двухбуквенные подслова слова 1001.

Формальным языком (или просто – языком) над алфавитом X называется всякое подмножество L множества X^*. Другими словами, язык – это произвольный набор слов в данном алфавите.

Как видим, привычные, казалось бы, понятия «слово», «суффикс», «язык» получают в математической лингвистике (а именно она занимается искусственными языками) весьма расширительное толкование. В его основе – отказ от осмысленности в ее интуитивном понимании. Должны быть заключены какие-то дополнительные соглашения, которые позволили бы классифицировать слова на имеющие смысл и не имеющие смысла, не вникая в их значение. С подобными соглашениями мы уже встречались. В алгебре имеющие смысл цепочки символов, составленные из букв, обозначающих числа, знаков операций и скобок, распознаются исключительно по их внешнему виду. Всем ясно, что набор символов $)a + b]^2$ бессмыслен, а «слово» $(a + b)^2$ представляет собой содержательное выражение. Точно так же обстоит дело и в математической логике (§IV.1): слова, называемые формулами, выделяются здесь среди прочих знаковых последовательностей тоже своей структурой. Можно сказать, что в обеих рассматриваемых ситуациях признаком осмысленности слова является его «грамматическая» правильность.

Под грамматикой естественного языка понимается свод правил, регулирующих образование и изменение слов, соединение их в предложения. Грамматики математической лингвистики решают примерно те же задачи, но на более абстрактном материале искусственных, формальных языков. Формальные грамматики делятся на два класса: распознающие и порождающие. Распознающая грамматика дает возможность для любого предъявленного набора алфавитных символов установить, является ли он словом данного языка. Порождающая грамматика позволяет построить любое слово данного языка, и все построенные ею слова входят в этот язык. Основной интерес для программирования представляют порождающие грамматики. Их мы и будем рассматривать, называя просто грамматиками.

Под грамматикой понимается четверка $\Gamma = (N, T, P, S)$, где N – конечное множество, элементы которого называются вспомогательными символами; T – конечное множество основных символов, причем $N \cap T = \emptyset$ (нет общих элементов); S – выделенный вспомогательный символ, называемый начальным; P –конечное множество выражений вида $\alpha \to \beta$ (правила), где α и β – некоторые слова над объединенным алфавитом $N \cup T$. Вспомогательные символы будем обозначать большими латинскими буквами, основные символы – малыми латинскими, слова над алфавитом $N \cup T$ – малыми греческими буквами.

Говорят, что слово δ непосредственно выводимо в грамматике Γ из слова γ, если 1) в Γ есть правило $\alpha \to \beta$; 2) α является подсловом в γ; 3) δ получается из γ заменой подслова α на слово β. Последовательность слов $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ называется выводом слова ε_n из слова ε_1, если каждое входящее в нее слово, начиная со второго, непосредственно выводимо из предыдущего.

Язык $L(\Gamma)$, порождаемый грамматикой Γ,– это множество слов в основном алфавите T, которые можно вывести из начального символа S.

Рассмотрим примеры порождающих грамматик и соответствующих языков.

1°. Пусть грамматика Γ_1 такова, что ее вспомогательный алфавит N_1 состоит только из начального символа S, т.е. $N_1 = \{S\}$; в основной алфавит T_1 входит только один символ 1, т.е. $T_1 = \{1\}$; список правил P_1 представлен двумя выражениями: $S \to 1$ и $S \to 1S$. Язык $L(\Gamma_1)$ состоит из слов 1, 11, 111, ... В краткой записи, $L(\Gamma_1) = \{1^n | n \in \mathbb{N}\}$. Как, например, выводится из начального символа S слово $1^3 = 111$? Дважды применяя второе правило, имеем: $S, 1S, 11S$. Заменяя согласно первому правилу в последнем слове подслово S на слово 1, получаем 111.

Если последовательность, состоящую из n единиц, считать записью натурального числа n, как это принято для машин Тьюринга, то можно сказать, что грамматика Γ_1 порождает натуральный ряд чисел. Изменив второе правило на $S \to 11S$, получим грамматику Γ_2, порождающую нечетные числа.

(Пусть нужно сложить два натуральных числа m и n. Запишем на ленте машины Тьюринга m единиц, затем знак «+», еще n единиц и знак «=». После исполнения программы (табл. 21) на ленте появится сумма $m + n$.)

Таблица 21

s_0	1	s_0	1	л
s_0	+	s_1	1	л
s_1	1	s_1	1	л
s_1	=	s_2	е	п
s_2	1	s_2	е	останов

2°. Грамматика Γ_3=(N_3, T_3, P_3, S), где $N_3 = \{S\}$, T_3=$\{0, 1\}$, P_3=$\{S \to 1S, S \to S0, S \to 10\}$ порождает язык $L(\Gamma_3)$= =$\{ 1^m0^n | m, n \in \mathbb{N}\}$. Напишите, например, вывод слова 11100. Как можно изменить правила грамматики Γ_3, чтобы в порождаемом новой грамматикой языке допускались и слова, состоящие только из единиц (только из нулей)?

3°. Грамматика Γ_4 с вспомогательным алфавитом $\{S, A\}$, основным алфавитом $\{0, 1\}$ и правилами $S \to 1S$, $S \to S0$, $S \to A$, $1A0 \to 10$, так же как и Γ_4, порождает двоичные слова вида 1^m0^n, т.е. $L(\Gamma_4)$=$L(\Gamma_3)$ (напишите вывод в Γ_4 слова 11100). Если две грамматики порождают один и тот же язык, они называются эквивалентными.

Грамматики классифицируются по виду их правил. Если все правила грамматики Γ имеют вид $A \to \alpha$, где A – вспомогательный символ, а α – слово в алфавите $N \cup T$, она называется контекстно свободной, а в противном случае – контекстно зависимой. В рассмотренных примерах грамматики Γ_1, Γ_2, Γ_3 являются контекстно свободными, а Γ_4 – контекстно зависимой. В правилах грамматик $\Gamma_1 - \Gamma_3$ символ вспомогательного алфавита S заменяется некоторым словом независимо от своего окружения (контекста), в грамматике Γ_4 правило $1A0 \to 10$ разрешает заменять вспомогательный символ A (пустым словом) только когда слева от A стоит символ 1, а справа 0. Контекстно свободная грамматика Γ называется регулярной, если все ее правила имеют вид $A \to aB$, где $A, B \in N$, $a \in T$. Такова, например, грамматика Γ_1.

Язык L называется контекстно свободным, если существует контекстно свободная грамматика Γ, порождающая его: $L = L(\Gamma)$. Если язык может быть порожден регулярной грамматикой, он называется регулярным. Например, язык $\{ 1^m0^n | m, n \in \mathbb{N}\}$ является контекстно свободным, так как существует контекстно свободная грамматика Γ_3, порождающая его (он порождается и контекстно зависимой грамматикой Γ_4). Будет ли этот язык регулярным? Оказывается, будет: его порождает также и регулярная грамматика Γ_5, у которой $N_5 = \{S, A\}$, $T_5 = \{0, 1\}$, $P_5 = \{S \to 1S,\ S \to 1A,\ A \to 0A,\ A \to 0\}$. Как вы получите в ней слово 11100?

Теперь у нас снова появятся автоматы. Однако, в отличие от предыдущих рассмотрений (§4), мы не будем интересоваться реакцией на выходе, а сосредоточимся на том, что происходит под внешним воздействием в «памяти» автомата, т.е. в множестве его внутренних состояний. Пусть дан автомат без выхода $\mathcal{A}=(S, X, \delta)$. Если он находится в состоянии $s \in S$ и на вход подан сигнал $x \in X$, то в следующий тактовый момент автомат окажется в состоянии $t = \delta(s, x)$. Функция переходов δ описывает реакцию автомата на одиночный входной символ. В реальных ситуациях, когда на вход поступает слово $p = x_1 x_2 \ldots x_n \in X^*$, важно знать то финальное состояние, в котором окажется автомат, «прочитав» эту последовательность. Если начальным состоянием было s_0, то получится цепочка состояний: $s_1 = \delta(s_0, x_1)$, $s_2 = \delta(s_1, x_2)$, ... , $s_n = \delta(s_{n-1}, x_n)$. Последнее состояние s_n и будем считать реакцией автомата на входное слово p, записывая $s_n = \delta(s_0, p)$. Теперь функция переходов определена для всех входных последовательностей, а не только для однобуквенных. Для пустого слова ε положим $\delta(s, \varepsilon) = s$ (т.е. состояние не меняется).

Пусть L – некоторый язык, алфавитом которого является множество X входных символов автомата $\mathcal{A} = (S, X, \delta)$, т.е. $L \subseteq X^*$. Говорят, что язык L представим в автомате $\mathcal{A}$ начальным состоянием s_0 и финальным подмножеством $F \subseteq S$, если этот язык состоит из всевозможных слов $p \in X^*$, переводящих s_0 в одно из состояний, входящих в состав F, т.е. если $L = \{\, p \in X^* | \, \delta(s_0, p) \in F\}$.

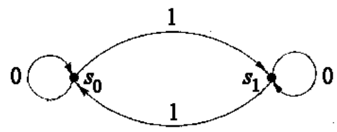


Рис. 105

Пусть, например, $\mathcal{A} = (S, X, \delta)$ – автомат без выхода, где $S = \{s_0, s_1\}$, $X = \{0, 1\}$, а функция переходов представлена графом на рис. 105. Язык L_1, состоящий из всех слов, содержащих нечетное число единиц, представи́м в автомате $\mathcal{A}$ начальным состоянием s_0 и финальным подмножеством $F_1 = \{s_1\}$. Язык L_2, состоящий из всех входных слов, в каждом из которых четное число единиц, представи́м в $\mathcal{A}$ начальным состоянием s_0 и финальным подмножеством $F_1 = \{s_0\}$. Иначе говоря, подмножество F_1 «запоминает», что количество прошедших через вход автомата единиц было нечетным, а подмножество F_2 является индикатором четности.

Если в некотором автомате $\mathcal{A}=(S, X, \delta)$ выделено состояние $s_0 \in S$ и подмножество $F \subseteq S$, то слова языка $L \subseteq X^*$, представимого в $\mathcal{A}$ начальным состоянием s_0 и финальным подмножеством F, можно выписывать как последовательности меток на дугах диаграммы вдоль путей, ведущих из s_0 в различные состояния из F. Например, какой язык представляют в автомате, изображенном на рис. 106, начальное состояние s_0 и финальное подмножество $F = \{s_2\}$? Попасть из s_0 в s_2 можно лишь путями, состоящими из дуги s_0s_1, какого-то числа петель в вершине s_1, дуги s_1s_2 и какого-то числа петель в вершине s_2. Выписывая метки дуг вдоль подобного пути, будем получать двоичные слова вида 1^m0^n, где $m, n \in \mathbb{N}$. Искомым языком будет знакомый нам язык $\{\,1^m0^n |\, m, n \in \mathbb{N}\}$. А какой язык представляют в рассматриваемом автомате s_0 и $F = \{s_1, s_3\}$?

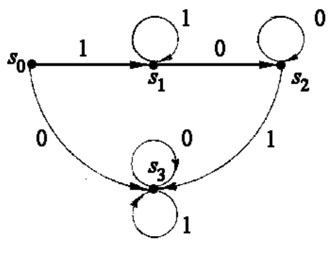


Рис. 106

Одним из центральных результатов теории формальных языков является следующая теорема, которую доказал американский логик Стивен Коул Клини (1909–1994).

Теорема 1. Язык L тогда и только тогда представим в некотором автомате $\mathcal{A}$ подходящим выбором начального состояния s_0 и финального подмножества F, когда этот язык регулярен. □

Другими словами, регулярные языки – это в точности языки, представимые в автоматах.

Существуют ли нерегулярные языки, т.е. языки, не порождаемые регулярными грамматиками? Теорема Клини позволяет дать утвердительный ответ на этот вопрос.

Следствие. Язык $L = \{\,1^n0^n |\, n \in \mathbb{N}\}$ не является регулярным.

ДОКАЗАТЕЛЬСТВО. Предположим, что L – регулярный язык. Тогда, по теореме Клини, он представим в некотором автомате $\mathcal{A} = (S, X, \delta)$ начальным состоянием s_0 и финальным подмножеством $F \subseteq S$. На вход автомата $\mathcal{A}$, находящегося в состоянии s_0, будем подавать символ 1. Так как множество S конечно, состояния $s_1 = \delta(s_0, 1)$, $s_2 = \delta(s_1, 1)$=$\delta(s_0, 1^2)$, $s_3 = \delta(s_2, 1) = \delta(s_0, 1^3)$, ... не могут быть все разными. Наступит момент, когда очередное состояние этой последовательности совпадет с одним из предыдущих: $s_n = s_m$ при $m < n$. После этого пойдет циклический повтор. Слово 1^m0^m принадлежит языку L и, значит, $s = \delta(s_0, 1^m0^m) \in F$. Как нетрудно заметить, $s = \delta(s_m, 0^m)$. С другой стороны, $\delta(s_0, 1^n0^m)$= $\delta(s_n, 0^m)$=$\delta(s_m, 0^m)$=$s \in F$, т.е. слово 1^n0^m переводит начальное состояние s_0 в финальное состояние s. Это невозможно, так как слово 1^n0^m не принадлежит языку L. Полученное противоречие показывает, что исходное предположение ложно, и следовательно, L – нерегулярный язык. $\square$

В §4 приводились соображения, показывавшие, что нельзя построить автомат, распознающий, содержится ли в двоичной последовательности вида 1^m0^n одинаковое число нулей и единиц. В новой терминологии это и означает, что язык $L = \{\,1^n0^n|\,n \in \mathbb{N}\}$ нерегулярен, – факт, строго доказанный выше. В §5 была предложена машина Тьюринга, распознающая среди слов вида 1^m0^n те, у которых число нулей и единиц совпадает. Можно доказать более общий факт, то, что в машинах Тьюринга представимы все вообще контекстно свободные языки.

Примером контекстно зависимого языка служит язык в алфавите $\{a, b, c\}$, состоящий из всевозможных слов вида $a^nb^nc^n$, где $n \in \mathbb{N}$. Впрочем, контекстно зависимые языки имеют больше теоретическое, чем практическое значение. Это может показаться удивительным, ибо, на первый взгляд, грамматики естественных языков, как правило, учитывают контекст (например, винительный падеж дополнения в фразах типа «я вижу стол» заменяется родительным падежом в том случае, когда перед глаголом появляется отрицание: «я не вижу стола»). Однако, как выяснилось, ценою увеличения числа правил контекстно зависимые части грамматик в естественных языках могут быть заменены эквивалентными им контекстно свободными грамматиками.

В терминах контекстно свободных грамматик выражается основная часть синтаксических правил в языках программирования, где грамматическая правильность является единственным критерием осмысленности. Компьютер останется равнодушным к драматическим событиям, описанным в знаменитой фразе академика-языковеда Л.В.Щербы, но признает ее безусловно понятной, т.е. соответствующей всем нормам грамматики русского языка:

Глокая куздра штеко будланула бокра и курдячит бокренка.

ГЛАВА VII. **ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ**

§ 1. Математическое моделирование и вычислительные эксперименты

Один из эффективных методов получения новых знаний – математическое моделирование изучаемых процессов и явлений. На определенном этапе исследования сложной системы возникает необходимость выделить в множестве полученных конкретных фактов то существенное, что, собственно, и определяет внутреннюю логику развития этой системы, ее основные структурные взаимосвязи. Создается некоторый идеализированный, абстрактный образ, как правило, допускающий описание в математических терминах. Так возникшая математическая модель позволяет, с одной стороны, отвлечься от несущественных (или кажущихся на данном этапе несущественными) явлений, сосредоточиться на сути дела, а с другой стороны, дает возможность подключить к решению представляющих интерес задач мощный аппарат, разработанный в соответствующих областях математики.

Результаты, полученные формальными методами при исследовании модели, истолковываются на языке исходной системы, сопоставляются с наблюдениями над ней. Некоторые из них оказываются действительно новыми фактами, другие же вступают в противоречие с реальностью и приводят к корректировке модели. Следующая, уточненная модель подвергается дальнейшему изучению, обнаруженные ее свойства снова интерпретируются как возможные свойства системы – процесс продолжается до построения такой модели, которая на данном этапе будет удовлетворительно объяснять поведение системы.

Следует учитывать, что никакая математическая модель не может вполне адекватно описывать все особенности моделируемой сложной системы, она отражает лишь определенный уровень в познании этих особенностей. Кроме того, некоторые свойства модели могут оказаться чисто логическими следствиями ее математической структуры и не иметь истолкования в понятиях исследуемой системы. Попытки придать этим свойствам некоторый содержательный смысл приводят к ложному представлению о сути изучаемых явлений – многочисленные примеры такого рода возникали при использовании методов кибернетики и теории информации, особенно в гуманитарных областях, что вызывало естественный протест специалистов. Здесь уместно вспомнить предостережение Гёте: «Математики – это некоторый род французов: если говоришь им что-нибудь, они переводят это на свой язык, и тогда это становится тотчас же чем-то совсем другим». И все же польза от применения математического моделирования неизмеримо превышает его возможные издержки.

Как уже отмечалось (§III.4), большой резонанс вызвали в свое время математические открытия, связанные с поисками геометрической модели окружающего нас мира (неэвклидова геометрия Лобачевского, многомерные пространства, внутренняя геометрия, риманова геометрия и т.д.). Но если эти обсуждения носили в значительной мере общефилософский характер, то «открытие планеты на кончике пера» буквально поразило современников. В первой половине XIX века астрономам стало ясно, что ньютоновская модель Солнечной системы не может объяснить наблюдаемое перемещение Урана – самой далекой в то время планеты. Англичанин Джон Коуч Адамс (1819–1892) и независимо от него француз Жан Жозеф Леверье (1811–1877) пришли к заключению о том, что неправильности в движении Урана, по-видимому, вызваны воздействием на него какой-то неизвестной планеты. Проведя соответствующие вычисления, они смогли установить ее массу и орбиту. В Гринвичской обсерватории к сообщению Адамса (21 октября 1845 г.) отнеслись хладнокровно и не спешили провести поисковые наблюдения, письмо Леверье получили в Берлинской обсерватории 23 сентября 1846 г. и в тот же вечер, направив телескоп в расчетную точку, увидели вблизи нее тусклую звездочку, оказавшуюся Восьмой планетой (Леверье уже имел для нее имя – Нептун). Ни одно научное достижение предшествующих времен не вызывало такой сенсации и не приносило ученым такой славы, которая выпала на долю Леверье. Если о неэвклидовой геометрии говорили, не вполне понимая суть дела («Лобачевский доказал, что параллельные пересекаются в бесконечности»), а идея многомерности пространства отдавала чем-то мистическим, то в случае с Нептуном триумф математики был совершенно очевиден: решая какие-то уравнения, «вычислили» неведомую планету.

Через сто лет после публикации работ Адамса и Леверье выдающийся английский математик Джон Иденсор Литлвуд (1885–1977) показал, что на самом деле задача была не такой уж сложной – царивший вокруг нее многолетний энтузиазм не вполне соответствовал научному уровню открытия. Но эти рассуждения доступны только специалистам и ни в коей мере не умаляют исторической значимости события. (Полистайте «Математическую смесь» Литлвуда. Она в чем-то сродни «Математическому калейдоскопу» Штейнгауза, но адресована более искушенным читателям. Есть в вашей профессиональной области такая книга?)

Исследования математических моделей атомного ядра привели к открытию позитрона (частица с массой электрона, но положительно заряженная) и некоторых других элементарных частиц. Важную роль сыграли при этом методы теории групп.

Первым результатом по математическому моделированию в биологии была модель системы кровообращения, предложенная Эйлером. В настоящее время с использованием новейших методов математики и информатики построены, например, модели нейронных сетей, сетчатки глаза, процесса синтеза белка в клетке и др.

Целая наука – математическая экономика – сформировалась на основе метода математического моделирования экономических процессов. В задачах выявления эффективных способов производства, прогноза развития экономики, о рациональном поведении потребителей, об экономическом равновесии не только использовались уже готовые математические методы, но и возникали ситуации, для изучения которых требовалось создание новых математических дисциплин. Так сформировались линейное программирование, теория игр, теория полезности и т.д.

Теория информации явилась математической моделью процесса передачи сообщений в самых разнообразных линиях связи. Это позволило с новой и зачастую совершенно неожиданной точки зрения интерпретировать многие явления в таких гуманитарных науках, как психология, искусствоведение, лингвистика.

«Тридцать лет непрерывных военных действий доказали теорему о том, что математика является прочнейшим фундаментом подлинной военной науки»,– писал Георг Вега (1756–1802), словенский математик и артиллерист, составитель многократно переиздававшихся логарифмических и тригонометрических таблиц. Еще более определенно высказался академик М.В.Остроградский: «Военное дело само является не чем иным, как одним из важнейших приложений математической науки». (Михаил Васильевич Остроградский (1801–1862), выдающийся русский математик и механик. Внес большой вклад в интегральное исчисление и теорию дифференциальных уравнений, создал отечественную школу прикладной механики. Решил ряд проблем, связанных с артиллерийской стрельбой. Резко отрицательно относился к Лобачевскому и его «воображаемой» геометрии.) Если бы авторы приведенных афоризмов заглянули в исследовательские лаборатории и вычислительные центры военных ведомств нашего времени, они были бы глубоко удовлетворены той ролью, которую играет здесь «королева наук». На математических моделях не только проверяются те или иные гипотезы, связанные с потенциальным противником, но и проводятся полномасштабные «войны», по результатам которых корректируется оборонная или инициативная стратегия.

Примеры можно продолжать неограниченно – они охватывают все сферы научной и практической деятельности человека. Но несмотря на различие в содержании и используемых теоретических средствах, есть одно общее, что объединяет современные математические модели сложных систем: их изучение невозможно без мощной компьютерной поддержки. Реальные эксперименты в экономике, военном деле, метеорологии, физике плазмы и многих других областях чрезвычайно затруднены. Они либо непомерно дороги, либо грозят возможными неприятными последствиями, либо практически не осуществимы. Эти препятствия удается преодолеть с помощью метода машинного, или вычислительного, эксперимента, одного из самых значительных достижений практической информатики.

После выбора математической модели составляются компьютерные программы, с помощью которых машина просчитывает различные варианты развития системы в зависимости от диктуемых экспериментатором условий. Результаты вычислений тщательно анализируются и сопоставляются друг с другом и с уже известными фактами, что приводит к уточнению модели и к новому уровню понимания свойств моделируемой системы. Таким образом, машинный эксперимент стал необходимой частью математического моделирования на современном этапе. Полученные с его помощью данные, как правило, не могут быть добыты никакими другими методами. Человеческий интеллект получил возможность переложить бремя недоступных по объему и сложности вычислений на плечи электронного коллеги и сосредоточиться на более тонких задачах. (Как тут не вспомнить пифагорейца Архита (428–365 до н.э.), который, отвлекшись от занятий в теории чисел, астрономии и музыке, «в соответствии с математическими правилами вырезал из дерева голубя, подвесил его и надул своим дыханием так, что он полетел», – непревзойденное достижение докомпьютерной прикладной математики.)

Подавляющее большинство вычислительных экспериментов остается вне сферы общественного внимания, хотя зачастую они затрагивают интересы многих миллионов людей. На общедоступном языке трудно объяснить и саму постановку задачи, и полученные результаты. Поэтому популяризаторы информатики стараются подробно рассказать о тех немногих реальных достижениях, которые получены с помощью компьютерного моделирования в гуманитарных областях. Отметим наиболее известные нетривиальные примеры.

1. Дж. Хокинс. Разгадка тайны Стоунхенджа (1964 г.).

Среди мрачных болот Солсберийской равнины в юго-западной Англии высятся громадные камни, расставленные по нескольким окружностям в период 1900 – 1600 годов до н.э. Назначение этого величественного сооружения на протяжении столетий оставалось загадкой, породившей множество легенд. В начале XX века появилась гипотеза о том, что Стоунхендж (так называется этот памятник позднего каменного века) как-то связан с астрономическими явлениями. Это предположение основывалось на том, что в ключевые для определения сезонов дни восход и заход Солнца происходит точно над вершинами или в арках главных объектов. Астроном Смитсоновской обсерватории Джералд Хокинс, проведя большое число измерений, построил математическую модель Стоунхенджа как системы, предназначенной для слежения за основными небесными светилами. Компьютерный анализ модели показал, что каменная астрономическая обсерватория «позволяет с удивительной точностью вести календарный счет, отмечать начала времен года и предсказывать наступление солнечных и лунных затмений».

2. А.Н.Колмогоров. Языковая модель стихотворного размера (1968 г.).

Статистический метод анализа поэтических текстов возник в XIX веке. Одним из главных вопросов было выявление ритмических предпочтений в произведениях поэтов разных эпох. В 1917 году известный филолог и литературовед Б.В.Томашевский, анализируя стих «Евгения Онегина», вычислял для каждой конкретной ритмической фигуры 4-стопного ямба, которым написан роман, теоретическую частоту ее встречаемости и сопоставлял полученное число с действительной частотой в тексте. В случае малого отличия этих величин предлагалось считать данную ритмическую форму присущей не автору, а языку в целом; внимание же нужно обращать на те случаи, где имеется существенное расхождение частот, ибо именно здесь проявляется индивидуальный вкус поэта. В начале 1960-х годов, заинтересовавшись математическими вопросами стиховедения, А.Н.Колмогоров с сотрудниками возродил и усовершенствовал этот забытый метод, выразив его суть в точных терминах так называемой языковой модели стихотворного размера, органично вошедшей в современную теорию стиха.

3. Г.Хьетсо и С.Гил. Кто написал «Тихий Дон»? (1984 г.).

Попытки найти «истинного автора» романа Михаила Шолохова «Тихий Дон» предпринимались неоднократно (этому способствовало загадочное исчезновение рукописей первого и второго томов романа). Очередное, наиболее солидное сочинение на эту тему появилось в Париже в 1974 году без указания имени автора, но с предисловием известного писателя А.И.Солженицына, полностью разделявшего вывод о том, что Шолохов якобы воспользовался рукописью умершего в 1920 году казацкого писателя Федора Крюкова (других серьезных претендентов нет). Норвежские слависты Гейр Хьетсо и Стейнар Гил вместе с коллегами из Швеции в 1975–1984 годах провели компьютерный эксперимент по сравнению разработанных ими статистических моделей стиля Шолохова, стиля Крюкова и стиля автора «Тихого Дона». В результате они пришли к однозначному выводу о том, «что Крюков совершенно отличен от Шолохова по своему творчеству и что Шолохов пишет поразительно похоже на автора «Тихого Дона». (Заметим, что в конце 1990-х годов утраченные авторские рукописи были найдены, но ни этот факт, ни – тем более – компьютерные доказательства не умерили энтузиазма антишолоховских поисковиков.)

Успехи, достигнутые в вычислительных экспериментах с математическими моделями, с полным основанием можно воспринимать как наиболее значительные завоевания искусственного интеллекта. Ведь речь здесь идет не просто об экономии мышления, а о расчетах, в принципе не осуществимых «невооруженным» разумом. Новые знания, полученные на основе этих расчетов,– плод совместных усилий естественного и искусственного интеллекта.

И все же роль машины в исследовании математических моделей сравнима с ролью гигантского экскаватора при производстве земляных работ: она лишь механически выполняет действия, предписанные ей программами. Все алгоритмы, лежащие в основе этих программ, созданы человеком, и все содержательные выводы, базирующиеся на результатах вычислений, будет делать человек.

Так что слово «интеллект» в этой ситуации можно ассоциировать с компьютером несколько условно: сама машина никаких суждений о смысле расчетов не имеет (во всяком случае, не высказывает). Совсем иначе обстоит дело в экспертных системах – другой сфере приложений искусственного интеллекта.

Экспертная система – это пакет прикладных программ, позволяющих имитировать действия эксперта, когда он принимает решения в своей специализированной области знаний. В основе экспертной системы лежит база знаний – представленная в приемлемой для машинного использования форме информация, которой владеют специалисты. Она включает в себя понятия данной предметной области, связанные с ними факты и зависимости. Кроме исходных знаний, машине должны быть сообщены те логические приемы, с помощью которых эксперты на основе имеющихся данных приходят к тем или иным выводам. Типовые схемы профессиональных умозаключений обычно представляются импликациями вида $p_1 \wedge p_2 \wedge \cdots \wedge p_n \to q$ («если одновременно имеют место факты $p_1, p_2, \ldots, p_n$, то можно допустить результат q»). Экспертная система должна обеспечивать также диалог пользователя с компьютером, содержащим базу знаний и связанную с ней структуру принятия решений.

Рассмотрим, например, в общих чертах экспертную систему предварительной медицинской диагностики, используемую в Японии. База знаний здесь содержит список примерно 300 заболеваний, классифицированных по 11 видам (сердечно-сосудистые, дыхательных органов, органов пищеварения и т.д.), и около 300 различных симптомов. При этом указаны возможные взаимосвязи между болезнями и симптомами. Естественно, что они не выражаются категорическими утверждениями, а лишь отражают уровень уверенности медицинских экспертов – прекрасный пример использования математической концепции нечеткости (см. §V.6). В базе знаний записано: простуда и кашель взаимосвязаны (степень уверенности 0,9); если есть кашель, то допустима простуда (0,8); если есть простуда, то возможен кашель (0,6).

Механизм выводов представлен нечеткими импликациями вида «если есть симптом p, то наличие болезни q оценивается как μ», «если признать наличие заболевания p, то проявление симптома q оценивается как μ». При этом допустимыми значениями для μ принимаются: «очень возможное», «довольно возможное», «возможное», «сомнительное», «довольно сомнительное», «очень сомнительное», «неясное». Формальный математический аппарат – нечеткое исчисление высказываний.

В диалоге компьютер исполняет роль врача, а пользователь – роль больного. Врачу сообщаются жалобы: кашель, головная боль, температура и т.д. (допускаются градации нечеткости «немного», «очень» и т.п.). Финал – предварительный диагноз – с указанной выше степенью уверенности μ.

Согласимся,что действия компьютера в этой ситуации выглядят гораздо «интеллектуальнее», чем в процессе решения им сложнейшей системы дифференциальных уравнений. И результат подан в более эффектной форме: непосредственно сформулировано новое знание, а не просто указано число или функция, которые должен интерпретировать человек. Но в отличие от машинных экспериментов экспертные системы пока решают задачи не столь высокого уровня.

Способ представления знаний и механизм принятия решений в конкретных экспертных системах реализуются с помощью инструментальных средств, приспособленных к проблемам данной предметной области. Существуют разнообразные языки представления знаний: логические, основанные на исчислении высказываний и исчислении предикатов; реляционные, где главную роль играют отношения, заданные графами; продукционные, в которых информация о процедурах принятия решения задается в виде описания возможных условий и указания действий, которые следует осуществить в этих условиях. Инженеры по представлению знаний (так называются программисты, конструирующие экспертные системы) стремятся максимально приблизить возникающие формальные понятия и методы к привычным для потенциального пользователя образам и способам рассуждения. Взаимодействие человека с машиной будет наиболее продуктивным в том случае, когда он сможет обмениваться с ней информацией на своем стандартном профессиональном языке.

Главное назначение экспертной системы и других компьютерных систем, базирующихся на знаниях,– быть консультантом, оказывать квалифицированному специалисту помощь в принятии оптимального решения в условиях многовариантности или неопределенности. Она может выполнять также и функции обучающей системы. Пользователь-ученик повысит свои профессиональные навыки, проводя с помощью компьютера те или иные умозрительные эксперименты, приобретая способность приходить к правильным выводам в незнакомых ему до того обстоятельствах. В свою очередь, экспертная система не является окончательным продуктом, заложенные в нее знания соответствуют лишь некоторому уровню изучения предметной области и заведомо ограничены. Выступая в роли учителя, опытный эксперт поможет системе расширить ее кругозор путем совместного анализа новых для нее ситуаций.

Один из способов представления знаний и содействия в выработке решений демонстрирует программа Excel фирмы Microsoft, широко применяемая в экономической и управленческой деятельности (непременный – наряду с текстовым редактором Word – объект практических занятий по информатике). Excel может оказаться полезной всем, кому приходится иметь дело с табличной формой записи данных и связанными с этим вычислениями.Она охватывает также и область деловой графики, позволяя выводить на экран и на печать различные виды диаграмм, наглядно представляющих полученные результаты.

Табличные процессоры, такие, как Excel, не только обрабатывают введенную в виде таблиц информацию, но и дают возможность проводить вычислительные эксперименты с построенной на основе этой информации моделью соответствующей предметной ситуации. Представим себе, что знания о некоторой прикладной области зафиксированы в виде многих таблиц с десятками параметров и что результирующие величины рассчитываются по каким-нибудь очень сложным формулам. Как, управляя изменением параметров, приблизить интересующие нас показатели к оптимальным значениям? Что будет, если, например, такой-то параметр будет неуклонно расти, а некоторый другой колебаться в заданных пределах? Программа с готовностью будет просчитывать все предлагаемые ей варианты «А что если...», предоставляя пользователю выбрать на основе полученных данных эффективную стратегию в применении доступных знаний о моделируемой области. (В одном из очерков Глеба Успенского среди картин уездной жизни 1880-х годов есть и такая: «Живет отставной военный человек, изобретающий статистическую машину. Можно будет взять на ней, как на пианино, аккорд известных цифр по известным правилам, и вместо звука получатся выводы, также цифровые».)

§ 2. Распознавание образов

Далеко не все знания могут быть выражены с помощью уравнений, числовых таблиц или точных логических конструкций. Обширные области человеческого интеллекта, не допускающие прямой формализации, все еще остаются вне сферы машинного разума с его огромными потенциальными возможностями. Может быть поэтому специалистов по искусственному интеллекту интересует не столько расширение круга успешно решаемых машиной задач, сколько ее возможности в имитации творческого процесса создания и переработки информации человеческим мозгом.

«Может ли машина мыслить?» – эту проблему обсуждал Тьюринг в своей статье 1950 года, имея перед собой первые образцы ЭВМ. Чтобы избежать трудностей, связанных с толкованием понятия «мыслить», он предложил сформулировать вопрос в терминах придуманной им «игры в имитацию», где испытатель общается по телеграфу с двумя адресатами, один из которых – машина. Не вдаваясь в подробности, можно сказать, что, по Тьюрингу, машину допустимо признать мыслящей, если она сумеет достаточно долго поддерживать у экспериментатора впечатление, будто он имеет дело с человеком.

Следующий восхитительный диалог взят не из современной пьесы абсурда и не из знаменитой «Алисы» Кэррола, он приведен в статье Тьюринга. Испытатель С хочет понять, с кем он имеет дело.

С: Напишите, пожалуйста, сонет на тему о мосте через реку Форт.

А: Увольте меня от этого. Мне никогда не приходилось писать стихи.

С: Прибавьте 34 957 к 70 764.

А (молчит примерно 30 секунд, затем дает ответ): 105 621.

С: Вы играете в шахматы?

А: Да.

С: У меня только король на е8 и других фигур нет. У Вас только король на е6 и ладья на h1. Как Вы сыграете?

А (после 15 секунд молчания): Лh8. Мат.

(Что сказали бы вы, будучи на месте С, о своем собеседнике А,– не человек ли он в самом деле? Или это хитрости компьютера?)

Собственные прогнозы Тьюринг выразил в следующей весьма осторожной форме: «Я уверен, что лет через пятьдесят станет возможным программировать работу машин с емкостью памяти 10^9 так, чтобы они могли играть в имитацию настолько успешно, что шансы среднего человека установить присутствие машины через пять минут после того, как он начнет задавать вопросы, не поднимались бы выше 70%».

Трудно сказать, реализуемо ли сейчас предсказание Тьюринга, но к ответу на вопрос «Может ли машина мыслить?» за прошедшие полвека мы не приблизились ни на шаг. И это несмотря на то что во многих областях современный компьютер, носитель искусственного интеллекта, добивается конечных результатов, намного превышающих возможности «среднего человека». Почему ребенок, с трудом осваивающий таблицу умножения, правила передвижения шахматных фигур, названия других городов, – разумное существо, а суперкомпьютер, вычисляющий траектории космических аппаратов, выигравший официальный матч у чемпиона мира, способный выдать любую справку об административно-территориальном делении страны,– разумным не считается? Если интеллектом назвать способность распознавать, обучаться, понимать и знать, т.е. умственную способность, то ученик первого класса обычной школы уже обладает интеллектом и дальше будет лишь развивать его. Возможно, пока он не знает, что картинку «Золотая осень» (в его букваре) нарисовал художник Левитан (см. §II.1). Знает ли об этом компьютер, который в ответ на запрос мгновенно извлечет фамилию из заложенной в него базы данных (алфавитный список создателей 100 шедевров отечественной литературы и искусства, годы жизни автора, область его творчества, название произведения и год окончания работы над ним:

Левитан 1861 1900 живопись «Золотая осень» 1895)?

«Да,– скажет критик искусственного интеллекта,– но точно в таком же смысле, в каком это знает справочник, откуда введены в базу данных соответствующие сведения». Не совсем так. Компьютер может сообщить, например, что свою знаменитую картину художник написал в возрасте 34 лет. В исходных данных факт этот явным образом не отмечался, он представляет собой хотя и без труда получаемое, но формально новое знание.

Если попросить малолетнего школьника найти среднюю длину слова в первой строфе седьмой главы романа А.С.Пушкина «Евгений Онегин», он вряд ли поймет, что от него требуется. Компьтер же после запуска соответствующей программы и ввода данных сразу даст ответ: 5,22 (см. §V.4). Понимает ли он содержание поставленной перед ним задачи ? «Точно в таком же смысле, в каком настенные часы понимают, что наступил полдень и нужно пробить 12 раз». Не будем возражать?

Кто знает, насколько продвинется наш маленький герой в обучении шахматной игре, но успехи машинных программ в этой области поразительны. Чтобы представить себе хотя бы на самом примитивном уровне, как может обучаться компьютер, рассмотрим следующую игру $\mathbb{G}$. Игрок A пишет на бумаге 0 или 1, игрок B добавляет к написанному символу 0 или 1. Далее аналогичное действие снова осуществляет игрок A, партия заканчивается ходом игрока B, второй раз выбирающего 0 или 1. Судья C, рассматривая полученную четверку из нулей и единиц как двоичную запись некоторого числа, объявляет победителем B, если это число делится на 3, и объявляет победителем A в противном случае. Игрок может отказаться от своего хода, и тогда ему засчитывается поражение. Например, в партии, где последовательность ходов выглядит как 1001, C назовет победителем B, ибо 1001 – это двоичное представление числа 9, в партии же 0101 выигрывает A, поскольку 0101 является двоичной записью числа 5. Партия 101 засчитывается в пользу A, так как B отказался от своего второго хода. Первокласснику B правило присуждения выигрыша представляется в высшей степени загадочным, и, целиком доверяясь судье C, он надеется приобрести необходимые навыки в многократных сражениях со своим партнером A.

Если бы на месте B был компьютер, обучающее правило имело бы очень простую, почти тривиальную форму: «Ход, приведший в данной позиции к проигрышу, больше в этой позиции не применяй». Преимущество машины перед ребенком в том, что она без труда запоминает все игровые ситуации. Изобразим возможные варианты течения игры $\mathbb{G}$ в виде дерева (рис. 107). Его вершины соответствуют конкретным позициям, а на ребрах указываются ходы, сделанные игроками. Полная партия в игре $\mathbb{G}$ состоит из пяти стадий: нулевая содержит начальную позицию (четыре пустые клетки), четвертая – позицию после второго хода игрока B. Последовательность ходов, приведшая к возникнове-

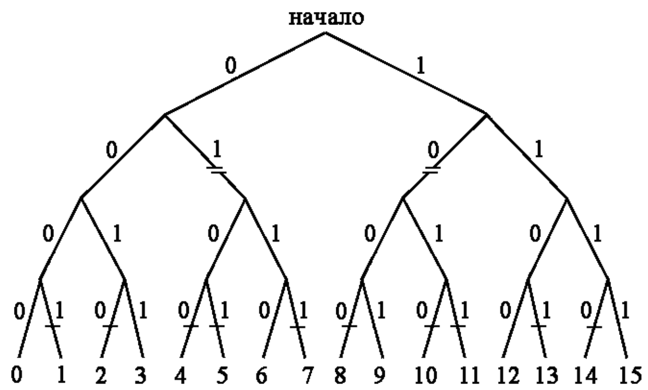


Рис. 107

нию конкретной позиции, получается, если выписать метки ребер вдоль пути, ведущего от начальной позиции к данной. Финальные положения обозначены соответствующими числами в десятичной записи. Например, число 10 соответствует партии 1010.

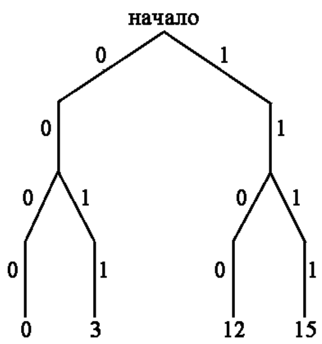


Рис. 108

В партии 0001 компьютер проиграет и, в соответствии с заложенным в него правилом, в позиции 000 больше никогда не будет делать ход 1. Перечеркнем на рис. 107 соответствующее ребро и так же поступим со всеми другими ребрами из третьей стадии в четвертую, ведущими к проигранным позициям. Таким образом, в ситуации 110 машина всегда будет делать ход 0 и выигрывать, поскольку 1100 – это 12. Однако после ходов 010 оба варианта оказываются отсеченными, компьютер не может сделать ход и, следовательно, сдает партию. Этот горький опыт заставит его, согласно правилу обучения, отказаться от хода 1 после начального 0, на рис. 107 принятое решение показано двойным перечеркиванием соответствующего ребра. Аналогичная ситуация приводит к запрету хода 0 в ответ на начальный выпад противника 1. После длительной и разнообразной практики обучение завершится, и модифицированное дерево игры $\mathbb{G}$ примет вид, показанный на рис. 108. Компьютер непобедим, игра теряет смысл. Заметим, что в исходном дереве игрок B имел среди 16 возможных финальных позиций лишь 6 выигранных и, следовательно, вероятность его победы при случайном выборе ходов противниками составляет всего $\frac{6}{16}=\frac{3}{8}=0,375$, т.е. 37,5%.

В аналогичном дереве, которое мы захотели бы составить для шахмат, из начальной вершины выходило бы 20 ребер, соответствующих 20 возможностям для первого хода белых, и из каждой полученной вершины первой стадии пришлось бы провести 20 ребер по числу вариантов первого хода черных. Таким образом, уже после первого хода партнеров в принципе может возникнуть 400 различных позиций. Трудно вообразить себе дальнейшее развитие этого ветвящегося процесса. Искусство шахматиста состоит, в частности, в том, чтобы отсекать огромное число заведомо неприемлемых путей, оставляя для дальнейшего рассмотрения лишь некоторое обозримое число вариантов. Выдающееся достижение шахматных программистов и заключается в создании формальных критериев, пользуясь которыми, компьютер отбраковывает варианты развития партии практически на уровне профессиональной интуиции мастера.

Анализируя стратегию, выработанную компьютером в игре $\mathbb{G}$ (см. рис. 108), мы замечаем, что он просто-напросто повторяет ходы противника. А что если игроки будут делать не по два, а по пять или по двадцать ходов? Приведет ли в этих случаях обнаруженный прием к победе? Оказывается, да! В самом деле, двоичное представление числа – это его разложение на сумму степеней двойки (см. §I.1). Но сумма двух соседних степеней двойки всегда делится на три: $2^{k+1} + 2^k = 2^k(2+1) = 3 \cdot 2^k$. Так что если в двоичной записи натурального числа n единицы и нули распределены в соответствии со стратегией компьютера, то n обязательно будет делиться на 3. Например, $1100111100 = 1\cdot2^9+1\cdot2^8+0\cdot2^7+0\cdot2^6+1\cdot2^5+1\cdot2^4+1\cdot2^3+1\cdot2^2+0\cdot2^1+0\cdot2^0 = (2^9+2^8)+(2^5+2^4)+(2^3+2^2) = 3\cdot2^8+3\cdot2^4+3\cdot2^2 = 3\cdot(2^8+2^4+2^2) = 3\cdot276$. (Апологет искусственного интеллекта скажет: «Таким образом, метод, открытый машиной, получил теоретическое подтверждение». Критик же заметит: «Идея повторения ходов – по-видимому, первое, что пришло бы в голову первокласснику B, напуганному правилом подведения итогов в игре $\mathbb{G}$». Весьма возможно. Но и компьютер стоит похвалить.)

Так или иначе, но такие основные признаки человеческого интеллекта, как «знать», «понимать» и «обучаться», допускают своеобразную интерпретацию в терминах возможностей искусственного интеллекта. «Знать» – значит хранить в памяти известные и выявлять новые факты, «понимать» – правильно реагировать на команды, «обучаться» – обнаруживать на основе конкретных примеров некоторые общие правила. Гораздо хуже обстоит дело с распознаванием.

В основе человеческого интеллекта лежит способность распознавать в хаосе окружающей информации те ее комплексы (образы), которые имеют или могут иметь отношение к решаемой в данный момент задаче, в чем бы она ни состояла. Процесс обучения, по существу, направлен на создание эталонных образов, по которым можно было бы классифицировать сложные и зачастую нечеткие образы материальной и духовной реальности. Поскольку подавляющая часть информации поступает в мозг человека через органы зрения и слуха, ключевой проблемой в области искусственного интеллекта является создание компьютерных программ для распознавания зрительных образов и естественной речи.

Раздел информатики, называемый «распознаванием образов», имеет целью разработку методов классификации и идентификации объектов. При этом считается, что каждый объект допускает описание при помощи некоторого конкретного набора свойств.

Задача классификации иллюстрируется следующим примером. Первокласснику предлагается распределить 12 пластмассовых шаров на группы так, чтобы в одной группе оказались одинаковые, по его мнению, шары. Первое, что отмечает испытуемый,– это то, что шары различаются размером, – в самом деле, одни величиной с теннисный мяч, другие – как мячик для пинг-понга. Второй бросающийся в глаза признак – цвет: одни шары красные, другие – синие. Так что шары делятся на четыре группы: большие красные, большие синие, маленькие красные и маленькие синие. При этом в каждой группе оказалось три шара. Раскладывая шары на четыре кучки, ребенок замечает, что внутри некоторых шаров заключен какой-то небольшой предмет (он тарахтит, если шар потрясти), а другие шары – пустые. Это позволяет ему уточнить классификацию. Окончательно он распределит 12 шаров на 8 классов по трем признакам: величина (Большой–Малый), цвет (Красный–Синий), содержимое (Есть–Нет). Если класс обозначить трехмерным вектором, компонентами которого являются значения этих признаков для данного класса, то состав классов оказался таким: (Б, К, Е) – 2, (Б, К, Н) – 1, (Б, С, Е) – 2, (Б, С, Н) – 1, (М, К, Е) – 2, (М, К, Н) – 1, (М, С, Е) – 2, (М, С, Н) – 1.

Решая предложенную задачу, мальчик сам обнаружил признаки, по которым проводил распознавание сходства и различия предъявленных ему объектов. Утомление или отсутствие соответствующего опыта не позволили ему в этом процессе самообучения выделить еще одно свойство исследуемых им образов: находящийся внутри шара предмет является или круглой дробинкой, или плоским диском, что существенно влияет на характер звука при сотрясении шара. С учетом этого дополнительного классификационного признака 12 шаров на самом деле образуют эталонный ряд, т.е. набор образов, обеспечивающий идентификацию по данным свойствам любого объекта из тех, для которых эти свойства имеют смысл.

Задача идентификации, т.е. распознавания предъявленного объекта как идентичного, тождественного с одним из эталонов, для мальчика с шарами выглядит так: ему вручается маленький красный шар с дробинкой внутри и предлагается найти такой же среди лежащих на столе 12 шаров. Довольно быстро исключив из рассмотрения все большие шары, все маленькие синие и пустой красный шарик, распознаватель остановится в нерешительности: какому из двух маленьких красных тарахтящих шаров отдать предпочтение. Он будет трясти поочередно врученный ему шар и эти два эталонных, пока не уловит сходство и различие в характере производимых звуков. Цель обучения достигнута: юный исследователь приобрел некоторое новое знание.

Переходя к реальным ситуациям, мы видим, что после нескольких лет обучения на примерах квалификация школьника в распознавании зрительных образов достигнет такого уровня, что он, например, без труда будет узнавать буквы в их самых разнообразных рукописных начертаниях (впрочем, жалуясь иногда на неразборчивость почерка). Распознаванием устной речи на своем родном языке он овладел гораздо раньше, а теперь может вести незатейливый диалог и что-то писать по-английски.

Электронному ученику до этого еще очень далеко. И дело здесь не в трудностях программирования, а в чем-то гораздо более существенном: человеческий интеллект является продуктом сложной системы, о которой мы имеем, по-видимому, лишь самые первоначальные наивные представления.

Итак, искусственный интеллект не будет признан интеллектом, пока его не научат распознавать зрительные образы и общаться с человеком на естественном языке. («А как же тогда воспринимать достижения Елены Келлер, Ольги Скороходовой, доктора наук Александра Суворова и других замечательных личностей, с раннего возраста лишенных и зрения и слуха?» – На это можно ответить так: Было бы неразумно, пытаясь моделировать человеческий интеллект, начинать с его предельных возможностей.)

И Тьюринг, и Шеннон, и фон Нейман, стоявшие у истоков информатики, считали, что информация должна поступать на вход перерабатывающего ее устройства в виде последовательности символов, или сигналов. По этому принципу «работают» машины Тьюринга, линии связи и автоматы. Он же лежит и в основе функционирования ЭВМ. Однако человеческий глаз не укладывается в рамки этой схемы: распознавание обозреваемого объекта осуществляется путем параллельного действия огромного числа независимых элементов сетчатки и нервных волокон, показания которых синтезируются мозгом в единый образ, сопоставляемый затем с хранящимися в памяти эталонами. Попытки создать нечто подобное из набора компьютерных процессоров к сколько-нибудь значительным результатам пока не привели.

Другое дело – распознавание речи. Произносимые человеком слова и фразы являются именно последовательностями звуковых сигналов и могут анализироваться компьютером в порядке поступления в его слуховой блок. Существующие системы письменной фиксации речи надежно работают под диктовку размеренным голосом, с четкой дикцией и выраженными паузами между словами. Говорить в таком режиме – достаточно утомительное дело, проще обратиться к клавиатуре. Но уже сейчас вполне реально создание компьютера, который в ответ на устные вопросы по темам, отраженным в его программном обеспечении, будет бесстрастно сообщать нужные сведения (с синтезатором речи, в общем, всё в порядке). Можно считать, что сделаны первые шаги в реализации мечты о беседах с всезнающим компьютером на нормальном человеческом языке. Осталось только разобраться в деталях того, что же он собой представляет – Язык.

К числу задач, связанных с распознаванием образов, можно отнести и проблему машинного перевода с одного естественного языка на другой. Наряду с шахматами, эта тематика привлекла внимание конструкторов искусственного интеллекта практически со времен появления ЭВМ. Понятна и практическая значимость компьютерного переводчика. Публичная демонстрация первой программы автоматического перевода русских фраз на английский язык состоялась в 1954 году (фирма IBM). С тех пор много раз чередовались периоды энтузиазма и разочарования, но дело не стояло на месте. Была создана теория формальных языков и грамматик (см. §VI.6), математические методы органично вошли в лингвистические исследования, постоянно совершенствовалось искусство программирования языковых процессов. Каковы же предварительные итоги? Полноценный перевод с одного языка на другой для искусственного интеллекта по-прежнему является далекой мечтой, трудности воплощения которой в настоящий момент осознаются как непреодолимые. Но в роли практического помощника машина выступает достаточно успешно. Вы не можете доверить ей перевод на английский язык своего делового письма, но, введя в нее текст полученного из-за рубежа сообщения, без труда разберетесь в его содержании, с улыбкой отмечая забавные сбои в русских фразах электронного толмача. Современные высококачественные программы-переводчики поддерживаются словарями, включающими около 50 000 слов.

Вот как переводит с английского языка программа, ориентированная на тексты из области информатики. Подлинник (книга по прикладной алгебре):

«One meets automata in various forms such as computers, money changing devices, telephone switch boards etc. All of the above have one aspect in common, namely a «box» which can assume various «states». These states can be transformed into other states by outside influence (called «inputs»), for instance by electrical or mechanical impulses. Often the automaton «reacts» and produces «outputs» like results of computations or change».

Компьютерный перевод (мгновенно после стартового сигнала):

«Этот встречает автоматы в различных формах типа компьютеров, деньги заменяющие устройства, телефонные платы переключателя и т.д. Все выше имеют один аспект в общем, а именно, «блок» который может принимать различные «состояния». Эти состояния могут преобразовываться в другие состояния снаружи влияния (называемый «вводами»), например электрическими или механическими импульсами. Часто автомат «реагирует» и производит «выводы» подобно результатам вычислений или изменений».

Используя материал §VI.4, отредактируйте эту заготовку-подстрочник. С такой же уверенностью был переведен и бессмысленный текст из эксперимента Шеннона (см. §V.5). Особенно трогательно выглядит выставленное компьютером тире.

§ 3. Компьютер и жизнь

Развитие компьютерной техники не только способствует автоматизации умственного труда, но и повышает качество жизни человека. Ушли в прошлое взволнованные дискусии о возможностях искусственного интеллекта и таящихся в нем опасностях. Незатейливые стихи и мелодии, сочиненные машинами под руководством и при участии программистов, остались символами романтической мечты найти в роботах младших братьев по разуму. Практическая информатика сняла ореол загадочности с достижений, завоеванных при помощи ЭВМ, ввела компьютеры в повседневную действительность как массовый продукт, предназначенный для удовлетворения информационных потребностей человека. И вот уже в популярной газете появляется вполне прозаическая рубрика «Компьютер и жизнь», содержащая полезные советы по приобретению и эксплуатации электронных устройств с программным управлением, сферой применения которых является офисная деятельность, телекоммуникация, деловая информация, домашняя экономика, перевод текстов, изготовление и копирование печатной продукции, организация досуга и т.д.

Приспосабливаясь к массовому потребителю, конструкторы стремятся, с одной стороны, уменьшить размеры персональных компьютеров, а с другой – увеличить их функциональные возможности. Последние модели блокнотного типа (ноутбуки) используют процессор Пентиум, снабжены цветным экраном, дисководом для чтения компакт-дисков, встроенными колонками. Так же как и настольные ЭВМ, ноутбук позволяет использовать современные средства связи, такие, как электронная почта, телефон, факс, модем. Карманные компьютеры в сочетании с сотовым телефоном также обеспечивают все указанные виды коммуникации, но имеют более скромное аппаратное и программное обеспечение, впрочем, достаточное для ведения личных записей, финансовых расчетов и получения разнообразных справочных сведений общего характера.

Примерно треть из продаваемых на отечественном рынке компьютеров приобретается для личного пользования.В настоящее время имеются программы, позволяющие применить электронную технику в таких естественных направлениях, как организация домашнего бюджета, создание разветвленной картотеки и архива, общее образование. Однако программы эти пока еще мало востребованы, и основную часть своего активного времени домашние компьютеры проводят в играх и развлечениях. Человеку, сидящему у экрана, предоставляется широчайшее меню удовольствий: от простейших тестов на скорость реакции до возможности ощутить себя составным элементом некой иной реальности. Дополнительные устройства позволят вам погрузиться в музыкальные миры, закодированные на компакт-дисках.

Выдающимся достижением современных информационных технологий явилось создание глобальной компьютерной сети Интернет, охватывающей все страны и континенты. Через нее владелец ЭВМ может вступить в контакт с любым из сотен миллионов зарегистрированных пользователей, получает доступ к электронным версиям самых разнообразных печатных изданий, от развлекательных газет до солидных научных журналов. По Интернет можно передавать текст, изображение и звук – все, что допускает цифровое кодирование. Диалог в режиме реального времени создает предпосылки для плодотворного сотрудничества ученых, деятелей культуры и искусства, политиков, предпринимателей, общественных и религиозных организаций, многочисленных групп по интересам и т.д. Понятное беспокойство вызывает бесконтрольность циркулирующей по Интернет информации, и официальные органы во многих странах уже ставят вопрос о необходимости какого-то ее ограничения. Но пока обсуждается эта юридически трудно разрешимая задача, миллионы людей ежедневно выходят в информационный космос со своими проблемами и откровениями. Может быть, и вы захотите оформить личную www-страницу, в которой раскажете о себе и интересующих вас вопросах. Она будет записана в память специальной мощной ЭВМ и окажется доступной всем клиентам Интернет – ждите откликов (World Wide Web – «всемирная паутина»).

Полезность компьютера в земных делах, в том числе и в отвлечении от земных дел, несомненна. С развитием производства программной продукции зависимость человека от услуг искусственного интеллекта будет неуклонно возрастать. Электронные банки данных уже сейчас сосредоточивают в себе такой объем жизненно важной информации, хранение и обработка которого в условиях «бумажного века» представляли бы почти неразрешимую проблему. Как искусный лоцман, проведет вас подходящая справочная система, например, через рифы налогового законодательства. Не без робости вы обращаетесь к компьютеру: «Каков порядок взимания НДС при производстве лекарственных средств из давальческого сырья?» – и немедленно получаете на экране ответ: «Согласно Закону Российской Федерации «О налоге на добавленную стоимость» лекарственные средства (в том числе лекарства-субстанции) налогом на добавленную стоимость не облагаются. Согласно пп. «у» п. 12 Инструкции Государственной налоговой службы Российской Федерации от 11.10.95 № 39 «О порядке исчисления и уплаты налога на добавленную стоимость» от этого налога освобождаются также услуги по реализации лекарственных средств. Оплата услуг по переработке давальческого сырья при производстве лекарственных средств должна производиться с учетом налога на добавленную стоимость. Основание: Письмо Минфина от 05.05.96 № 04-03-07». Принтер бесстрастно скопирует эту справку на выбранный вами твердый носитель (бумага, картон, пластмассовая пленка и т.п.).

Информационные технологии стали необходимой составной частью в общем развитии нашей цивилизации.

Как и в практической жизни, в вопросах творчества проблема конкуренции машины и человека ныне всерьез не обсуждается. И здесь речь идет, главным образом, о том, чтобы эффективно использовать ЭВМ для преодоления неизбежных трудностей, связанных с технической стороной воплощения художественной идеи. Полученные результаты трудно переоценить. Чего стоит одно то, что композитор с помощью компьютера может услышать свое произведение в реальном звучании – на любом инструменте и в любом оркестровом составе. Производство видеофильмов уже немыслимо без участия специализированных суперкомпьютеров. Конструктор на разных этапах проектирования объекта получает возможность увидеть его трехмерное изображение в целом или в каких-то частях. Системы компьютерной графики стали необходимым инструментом в работе художников-полиграфистов.

Можно ли сегодня говорить о собственно машинном искусстве? Это зависит от критериев, с которыми мы подходим к оценке произведений. Многое из того, что создается сейчас представителями музыкального, живописного и литературного авангарда, вполне может быть предъявлено как продукция электронных устройств и в этом качестве не вызовет никакого интереса. Во все времена внимание привлекали не сами по себе странные наборы звуков, унылые черные квадраты на белом холсте или стихи вроде

Дыр бул щыл
Убещур
Скум! –

а те идеи, которые они символизировали (додекафония, супрематизм, заумь). У компьютера нет собственных идей, и все его музыкальные, живописные и литературные поделки, выполненные по программам, имитирующим оригинальное творчество, будут восприниматься именно как достижения в области конструирования таких программ. Другое дело, когда за этим стоит нечто такое, что само по себе имеет объективную ценность. Начиная с 1984 года огромный успех сопутствовал художественной выставке «Границы хаоса», представлявшей около сотни картин, синтезированных компьютерами в университете г. Бремена (Германия) при изучении математиками и физиками так называемых фракталов. Математическую идею фрактальности высказал в 1975 году американский математик Бенуа Мандельброт, изучавший геометрию природных объектов неопределенной, хаотической, формы, подобных себе в том смысле, что каждая малая часть такого объекта как бы воспроизводит структуру целого: линия морского берега, дерево, облако, языки огня и т.п. Бременские ученые исследовали фрактальные геометрические фигуры, образуемые на плоскости точками, возникающими в процессе построения некоторых последовательностей комплексных чисел. Цветные образы на экране монитора выглядели так необычно, что решено было показать их широкой общественности. Ганс-Отто Пайтген и Петер Рихтер пишут в своей книге «Красота фракталов (образы комплексных динамических систем)»: «Мы думали, что достаточно будет эстетической привлекательности самих картин. Но как наивны мы были и как недооценивали нашу публику! То, что казалось нам просто забавой в контексте нашей научной работы, вдруг стало темой серьезных дискуссий. Зрители требовали объяснить этот контекст и желали знать, какова его важность. Мы вдруг почувствовали себя обязанными прояснить содержание, символизируемое нашими картинами». Упомянутая книга и дает доступное для широкого круга читателей толкование помещенных в ней репродукций картин и рисунков, объясняя, что возникли они не «просто так», а как побочный продукт при решении конкретных математических задач. Это придает красоте фракталов, построенных компьютером, дополнительное значение, не включающееся в эстетику «чистой», неодухотворенной формы.

Попытки вовлечь ЭВМ в высшие сферы интеллектуальной деятельности предпринимались уже с конца 1950-х годов. Казалось, что в такой формализованной области мышления, как математика, это может произойти наиболее успешно. Начало, в самом деле, было многообещающим: машина IBM-704 по программе, составленной американским логиком Хао Ваном, за несколько минут доказала свыше 350 теорем из «Principia Mathematica» Рассела и Уайтхеда (см. §§IV.2-3). Но в этом трактате, где рассматриваются лишь аксиоматическая логика и теория множеств, доказательные рассуждения уже заменены цепочками формул, механически получаемых по специальным правилам. Все дальнейшие попытки расширить сферу автоматического доказательства теорем на другие математические дисциплины заметных успехов пока не принесли.

В теоретической математике, как и в других творческих областях, компьютер оказался полезным, а иногда и незаменимым помощником в выполнении технически сложных, но вполне программируемых рутинных работ. В связи с этим нельзя не упомянуть о решении проблемы четырех красок. Первоначальная история этой задачи связана с именами таких ученых, как Мёбиус, Де Морган, Гамильтон и Кэли. Они пытались выяснить, каково наименьшее число красок, которыми можно закрасить любую мыслимую географическую карту так, чтобы каждые две страны, имеющие общий участок границы, были окрашены в разные цвета. После выступления Кэли в Лондонском географическом обществе (1878 г.) задача эта приобрела широкую известность, ею занялись многие профессионалы и любители. Легко придумать карту, которую нельзя раскрасить менее чем четырьмя красками (приведите пример). С другой стороны, в 1890 году английский математик Перси Джон Хивуд доказал, что для раскрашивания любой карты достаточно пяти красок. Гипотеза о том, что на самом деле можно ограничиться четырьмя красками, на протяжении ста лет оставалась одной из самых волнующих проблем математики. Простота самого вопроса и бессилие выдающихся умов в поисках ответа на него сделали проблему четырех красок одним из символов общей культуры (другие примеры: квадратура круга, вечный двигатель, теорема Ферма, машина времени). В 1976 году американские математики Кеннет Аппель и Вольфганг Хакен объявили, что с помощью ЭВМ им удалось решить знаменитую задачу. Сообщение это, вызвавшее большой энтузиазм в средствах массовой информации (его усиливала загадочная роль электронного соавтора), математиками было встречено весьма сдержанно. Идея доказательства состояла в том, чтобы свести вопрос о возможности требуемой раскраски любой карты к анализу некоторого, пусть большого, но конечного набора карт. Сведение к конечному случаю и осуществили два математика, а затем три компьютера IBM-360 путем перебора огромного числа вариантов за 1200 часов работы сумели найти раскраску в четыре цвета во всех предусмотренных теорией случаях. Полное доказательство было опубликовано только в 1989 году. Оно занимало 744 страницы и большинством специалистов было оценено как недоступное для проверки, главным образом, в части, выполненной машинами. Но в общественном сознании уже закрепился выписанный на почтовых штемпелях и пластмассовых значках тезис: «Четырех красок достаточно». К настоящему времени найдены в общем обозримые доказательства теоремы о четырех красках, но обойтись без помощи компьютера пока не удается. Математическое сообщество не считает такие доказательства полноценными и воспринимает их как существенные продвижения на пути к тексту, который можно будет построчно прочесть с карандашом в руке.

Вычислительные способности компьютеров (т.е. вычислителей) не идут ни в какое сравнение с возможностями человека, и результаты, полученные в этой области, вызывают изумление. Ну, например, в начале 1996 года доцент и аспирант Токийского университета после 113 часов работы на суперЭВМ нашли 6 442 450 000 (шесть миллиардов четыреста сорок два миллиона четыреста пятьдесят тысяч) знаков после запятой в записи числа π (в начале прошлого века гимназисты заучивали двустишие

Кто и шутя и скоро пожелаетъ
«Пи» узнать число – ужъ знаетъ,

дававшее значение $\pi \approx 3,1415926536$). Еще в середине

1960-х годов был известен миллион цифр числа e и миллион цифр числа $\sqrt{2}$. Возьмите школьную тетрадь по арифметике и подсчитайте, сколько таких тетрадей в 12 листов потребуется доценту и аспиранту, чтобы, помещая по одной цифре в клетке, переписать из памяти компьютера полученное ими приближенное значение для π. Сколько времени займет эта процедура, если, чередуясь, они будут без перерыва писать по одной цифре в секунду? (Для разминки полезно решить задачу известного педагога конца XIX века С.А.Рачинского: «Мальчик прочел по 4 раза Верую (149 слов), Отче наш (52), Достойно (30), Богородице (23) и Пресвятая Троице (21). Говорит он по 100 слов в минуту. Сколько минут он читал?» Художник Н.П.Богданов-Бельский, ученик Рачинского, на картине «Устный счет» изобразил своего учителя в окружении детей, пытающихся вычислить «в уме» написанное на доске выражение

$$\frac{10^2 + 11^2 + 12^2 + 13^2 + 14^2}{365}.$$

Юные счетчики должны были сильно удивиться, осознав в процессе работы, что $10^2 + 11^2 + 12^2 = 13^2 + 14^2 = 365$ – прикосновение к математической эстетике.)

Смысл рекордов, связанных с бесконечным уточнением величины числа π, и подобных им достижений можно с натяжкой объяснить стремлением совершенствовать методы и технические средства вычислений. Возможно также, что это когда-нибудь пригодится для каких-то пока неизвестных целей (история математики знает множество таких примеров). Но ценность результатов, полученных с помощью компьютеров в теории чисел, уже сейчас не подвергается никаким сомнениям.

Далее мы будем опираться на материал из §I.1. Знаменитая теорема Эвклида утверждает, что простых чисел бесконечно много (кстати, это одна из немногих «настоящих» теорем, для которых получено машинное доказательство). Разбросаны они по натуральному ряду бессистемно, и эффективных способов их обнаружения в далеких его областях нет. Поэтому открытие каждого нового простого числа вызывает живой интерес. В 1883 году сельский священник с Урала Иван Михайлович Первушин (1827–1900) представил в Петербургскую Академию наук заметку «Число $2^{61} - 1 = 2\ 305\ 843\ 009\ 213\ 693\ 951$ есть простое число», а к 1897 году он завершил сорокалетний труд составления сплошной таблицы простых чисел от 1 до 10 000 000 (она занимает 750 листов большого формата, исписанных мелким почерком). Составлением таблиц простых чисел занимались феноменальные вычислители XIX века Захарий Дазе в Германии и Якуб Филип

Кулик в Чехии. Но, как и в случае с Первушиным, их титанический труд не мог быть опубликован из-за очень большого объема рукописей и в связи с невозможностью проверки уникальных расчетов. Начиная с 1951 года к поискам больших простых чисел привлекли ЭВМ. В таблице рекордов вместе с именами математиков стали указывать и компьютеры, на которых производились вычисления. Например: $180 \cdot (2^{127} - 1)^2 + 1$ (простое число), 79 (количество цифр в нем), 1951 (год открытия), Миллер+Уиллер+EDSAC 1 (авторы); или $2^{2281} - 1$, 687, 1952, Лемер+Робинсон+SWAC; или $2^{19\ 937} - 1$, 6002, 1971, Таккерман + IBM 360. В 1983 году с использованием многопроцессорного компьютера Крей-1 Дэвид Словински установил, что простым является число $2^{86\ 243} - 1$. В его десятичной записи 25 962 цифры. Машина работала 1 час 3 минуты и 22 секунды. Что считать и как считать, конечно, определяет человек, но без электронного ассистента он был бы беспомощен перед такими огромными величинами. (Кроме увлекательных расчетов, связанных с простыми числами, сверхскоростной гигант конструкции Сеймура Крея (1926–1996) занимался моделированием ядерных войн, дешифровкой секретных кодов, глобальными метеорологическими прогнозами и другими задачами, непосильными для обычных ЭВМ.) В 1996 году в сети Интернет был объявлен проект поиска больших простых чисел вида $2^n - 1$, объединивший вскоре усилия 75 000 добровольцев разных возрастов и профессий из многих стран и 250 000 компьютеров. Седьмым в серии рекордных достижений коллектива стало число, простоту которого 15 мая 2004 года установил Джош Файндли (США) после 14 дней непрерывной работы его электронного соавтора 2.4 GHz Pentium 4 Windows XP PC, – самым большим из известных к тому времени простых чисел стало $M_{41} = 2^{24036583} - 1$ с количеством цифр 7 235 733 (премия в $ 100 000 ждет того, кто первым предъявит простое число с 10 000 000 цифр).

В связи с компьютеризацией жизненно важных областей человеческой деятельности весьма актуальной стала проблема защиты информации. Широко известны случаи злонамеренного проникновения в секретные банковские и даже оборонные сети. Одна из древнейших наук – криптография («тайнопись») – получила новые стимулы для разработки удобных в применении и максимально сложных для разгадки шифров. В конце 1970-х годов группой американских математиков была предложена принципиально новая идея шифрования. Не вдаваясь в детали, скажем, что в каждом конкретном шифре такого рода основную роль играет некоторое число N, разлагающееся в произведение двух простых чисел: $N = pq$. Описав свой метод, авторы его предложили желающим прочесть зашифрованное ими сообщение, пообещав за это премию (впрочем, весьма скромную – всего 100 долларов). Для успешного дешифрования достаточно было найти простые делители p и q объявленного числа N, содержавшего 129 цифр. Разлагать на множители такие большие числа никто тогда не умел, а непосредственный подбор при реализации на компьютере с процессором Пентиум занял бы более 20 000 лет непрерывной работы. Надежность нового способа шифрования (шифр RSA) привлекла к нему внимание структур, ответственных за сохранение в тайне важнейшей информации финансового и делового характера. Понятно, какой эффект произвело на них появившееся в апреле 1994 года сообщение о том, что шифр RSA, основанный на упомянутом выше 129-значном числе N (его назвали RSA-129), разгадан. За восемь месяцев до того американский программист Арьен Ленстра обратился через Интернет к своим коллегам, призывая их принять участие в разложении на множители числа RSA-129 с использованием последних достижений теории чисел. Откликнулось более 600 человек из 26 стран, которые получили индивидуальные задания. Работая на 1600 компьютерах (от миниатюрных моделей до 20-тонных мастодонтов) в свободное от их основной нагрузки время, волонтеры помогли заполнить нулями и единицами рабочую таблицу, содержащую около трехсот миллиардов клеток. Заключительный этап вычислений осуществил суперкомпьютер MasPar. Число RSA-129 оказалось произведением простых чисел p и q, одно из которых содержало 64, а другое 65 цифр. Стодолларовой фразой оказалась следующая: «Магические слова – чувствительная скопа» (скопа – это хищная птица, питающаяся рыбой). Могущественным организациям и ведомствам, применяющим шифр RSA, пришлось подбирать ключевые числа N с еще большим количеством цифр. Сплошной перебор возможных делителей 200-значного числа N займет у самого быстродействующего персонального компьютера около миллиарда лет, а метод, позволивший справиться с RSA-129, в этом случае современной вычислительной технике пока не по силам.

Дело Чувствительной Скопы (The Case of the Squeamish Ossifrage) трактуется как одно из самых ярких достижений новых информационных технологий. Удивительно, что в нем оказались равноправными партнерами детище конца XX века практическая информатика и всегда воспринимавшаяся как антипод прикладной математики древняя теория чисел. Завершая, мы возвращаемся к началу.

МЫСЛИ О МАТЕМАТИКЕ

Ни одно человеческое исследование не может назваться истинной наукой, если оно не прошло через математическое доказательство.

Леонардо да Винчи (1452–1519)

Математические науки, достигшие высокой степени совершенства, во многом должны служить образцом состояния, к которому надлежит стремиться и остальным наукам.

Н.Г.Чернышевский (1828–1889)

Математика – наука великая, замечательнейший продукт одной из благороднейших способностей человеческого разума.

Д.И.Писарев (1840–1868)

Все правильно полученные математические построения и открытия общеобязательны для всех. Борьба с ними, которая не раз повторялась в истории человеческой мысли, есть борьба с ветряными мальницами. Философская (или религиозная) мысль, если она сталкивается с математическими достижениями, должна была в конце концов их признать и из них делать следствия, а не изменять по своим построениям.

В.И.Вернадский (1863–1945)

Математика – это величественное здание, созданное воображением человека для постижения Вселенной.

Ле Корбюзье (1887–1965)

Не дерзаю чем-либо умалить прекраснейшую науку высокой математики, за которой признаю первенство в человеческом знании, но полагаю, что она должна применяться в своем месте...

М.В.Ломоносов (1711–1765)

Утверждая, что геометрический метод применим не ко всему, ошибаются,– но правы, когда говорят, что его не следует применять ко всему. Всякий предмет должен быть трактуем посвоему. Геометрический метод слишком сух, чтобы применять его к обучению манерам, и наш язык слишком несовершенен, чтобы целиком ему следовать... Но если приходится порою отказываться от применения его, то все же следует помнить о нем; это своего рода компас для ума и узда для воображения.

Д.Дидро (1713–1784)

Я уважаю математику как самую возвышенную, полезную науку, когда ее применяют там, где она уместна, но не могу одобрить, чтобы ею злоупотребляли, применяя ее к вещам, которые совсем не входят в ее область и которые превращают благородную науку в бессмыслицу.

И.-В.Гете (1749–1832)

Математика не может дать более точных доказательств, чем те, что дает художнику его творческое чутье.

Э.По (1809–1849)

В оный день, когда над миром новым
Бог склонял лицо свое, тогда
Солнце останавливали словом,
Словом разрушали города...
А для низкой жизни были числа,
Как домашний, подъяремный скот,
Потому что все оттенки смысла
Умное число передает.

Н.С. Гумилев (1886–1921)

СПИСОК ЛИТЕРАТУРЫ

Адамар Ж. Исследование психологии процесса изобретения в области математики.– М.: Сов. радио, 1970.

Березина Л.Ю. Графы и их применение.– М.: Просвещение, 1979.

Биркгофф Г. Математика и психология.– М.: Сов. радио, 1977.

Боголюбов А.Н. Математики и механики: Биографический справочник.– Киев: Наукова думка, 1983.

Варга Б., Димень Ю., Лопариц Э. Язык, музыка, математика.– М.: Мир, 1981.

Виленкин Н.Я. Рассказы о множествах.– М.: Наука, 1969.

Виленкин Н.Я. Комбинаторика.– М.: Наука, 1969.

Волошинов А.В. Математика и искусство.– М.: Просвещение, 2002.

Заде Л. Понятие лингвистической переменной и его применение к принятию приближенных решений.– М.: Мир, 1976.

Зенкевич И.Г. Не интегралом единым.– Тула: Приок. кн. изд-во, 1971.

Игошин В.И. Логика с элементами математической логики (лекции для студентов гуманитарных специальностей). – Саратов: Научная книга, 2004.

Кемени Дж., Снелл Дж., Томпсон Дж. Введение в конечную математику.– М.: Мир, 1965.

Клайн М. Математика: утрата определенности.– М.: Мир, 1984.

Клайн М. Математика: поиск истины.– М.: Мир, 1988.

Колмогоров А.Н. Математика в ее историческом развитии.– М.: Наука, 1991.

Курант Р., Роббинс Г. Что такое математика?– М.: Просвещение, 1967.

Левитин К. Геометрическая рапсодия.– М.: Знание, 1984.

Литлвуд Дж. Математическая смесь. – М.: Наука, 1973.

Мичи Д., Джонстон Р. Компьютер – творец.– М.: Мир, 1987.

О квадратуре круга.– М.;Л.: Гос. технико-теорет. изд-во, 1934.

Пойя Д. Математическое открытие.– М.: Наука, 1970.

Пойя Д. Математика и Правдоподобные рассуждения.– М.: Наука, 1975.

Пуанкаре А. О науке.–М.: Наука, 1983.

Реньи А. Трилогия о математике. – М.: Мир, 1980.

Салий В.Н. Введение в вероятность и теория информации для психологов. – Саратов: Изд-во Сарат. ун-та, 1977.

Стройк Д.Я. Краткий очерк истории математики.– М.: Наука, 1990.
Стюарт Я. Концепции современной математики.– Минск: Вышэйшая школа, 1980.
Тьюринг А. Может ли машина мыслить?– М.: Физматгиз, 1960.
Уайтхед А.Н. Введение в математику.– Пг., 1916.
Фоменко А.Т. Методы статистического анализа нарративных текстов и приложения к хронологии.– М.: Изд-во Моск. ун-та, 1990.
Шрейдер Ю.А. Равенство, сходство, порядок.– М.: Наука, 1971.
Штейнгауз Г. Математический калейдоскоп.– М.;Л.: Гос. технико-теорет. изд-во, 1949.

Учебное издание

Салий Вячеслав Николаевич

Математические основы гуманитарных знаний

Учебное пособие
для студентов гуманитарных направлений и специальностей
высших учебных заведений

Технический редактор Л. В. Агальцова
Корректор А. И. Яровинская
Оригинал-макет подготовлен И. А. Пономаревой

Подписано в печать 14.11.2005. Формат $60 \times 84\ 1/16$.
Бумага офсетная. Гарнитура Таймс.
Печать офсетная. Усл. печ. л. 17,90 (19,25). Уч.-изд. л. 18,2.
Тираж 200. Заказ 212.

Издательство Саратовского университета.
410012, Саратов, Астраханская, 83.
Типография Издательства Саратовского университета.
410012, Саратов, Астраханская, 83.